Special Issue

Journal of CHEMOMETRICS

Received: 8 June 2007, Revised: 26 February 2008, Accepted: 18 April 2008, Published online in Wiley InterScience: 11 June 2008

(www.interscience.wiley.com) DOI: 10.1002/cem.1156

Multivariate Modeling and Exploration of Environmental *n*-Way Data From Bulk Precipitation Quality Control

Aleksander Astel[a]* and Stanisław Małek[b]

This paper describes the results of study on modeling and exploration of a three-way environmental data set acquired from monitoring bulk precipitation chemistry collected in the Dupniański Stream Catchment (Silesian Beskid, Southern Poland) using Tucker3 modeling and self-organizing map approach. It appears that the constructed Tucker3 model is appropriate for the studied data and explains more than 91% of the data variance. The Tucker3 model allows to distinguish 'heavy metals-dust particles' and 'anthropogenic' factors responsible for chemical profiles of bulk precipitation and to identify seasonal variations in bulk precipitation quality. A self-organizing map approach confirms Tucker3 modeling results and additionally allows to identify a strong, temporary impact of remote pollution sources located in the vicinity of Polish-Czech Republic border and indicates a cyclical impact of remote pollution sources located in highly industrialized Katowice and Bełchatów regions. Copyright © 2008 John Wiley & Sons, Ltd.

Keywords: inorganic analytes; Tucker3 modeling; self-organizing map; environmetrics; Silesian Beskid

1. INTRODUCTION

This study aimed at exploration of a three-way monitoring data set from the Dupniański Stream Catchment (Silesian Beskid, Southern Poland) making use of the chemometric *n*-way Tucker3 method and a self-organizing map approach. The main objective was to reveal the relationship between chemical composition of bulk precipitation and seasonality (winter and growing period) or specific meteorological conditions and events which facilitate defining the source of air pollution.

Southern Poland, especially the Silesian Beskid is heavily polluted by various sources [1–3] mainly from the Ostrava-Karvina region (Czech Republic) as well as the Rybnik and Katowice regions (Polish territory) [4–7]. Multi-year bulk precipitation is widely used in the assessment of air pollution quality and quantity (dry and wet deposition). A fairly good assessment is often accomplished despite a great number of variables involved: data from various periods for many chemical elements affected by various factors (wind speed and direction, temperature, distance from main sources – natural or/and artificial) [8]. In environmental studies, chemometrics can offer a variety of techniques that can be successfully applied to data exploration and modeling. This is due to the fact that environmental data sets are multidimensional and complex. Mining of relevant information hidden in the data is required in order to provide an overview of the processes in the system under study [9].

An assessment of the bulk precipitation or throughfall chemistry at sites of interest requires a consistent and constant or at least periodical (short time intervals) monitoring of carefully selected parameters. As a result, data with multiway structure are obtained. As an example, a data set obtained for bulk precipitation quality control described by several variables measured in monthly intervals over a few years, can be given. Such data can be presented as a three-way array arranged as *variables* × *months* × *years* and thus explored using approaches dealing with the three-way data structure, that is one of the most popular in chemometrics—Tucker3 [10–18]. It should be emphasized at this point that the most typical environmental data sets have the following form of a three-way matrix: *sampling points* × *variables* × *time*, but examples in which two separate modes of the data array are formed by the time dimension are also widely known [19,20]. The use of multiway models provides a better insight into the data structure, reduces the noise, and shows which of the original variables are correlated and which of them are most significant for a certain environmental problem description [12]. Unfortunately, in some cases multiway models provide only general information and application of other chemometric techniques becomes necessary [20]. A self-organizing map (SOM) approach offers specific 'resolving power' and allows simultaneous observation of both intercorrelations between variables and temporal changes in variability of chemical profiles (i.e. bulk precipitation) of a sample. Moreover, an SOM-based exploration allows detecting natural clusters of monitoring locations with a similar type of chemical profile of a

* *Correspondence to: Environmental Chemistry Research Unit, Biology and Environmental Protection Institute, Pomeranian Academy, 22a Arciszewskiego Str., 76-200 Słupsk, Poland.*
E-mail: astel@pap.edu.pl

a *A. Astel*
Environmental Chemistry Research Unit, Biology and Environmental Protection Institute, Pomeranian Academy, 22a Arciszewskiego Str., 76-200 Słupsk, Poland

b *S. Małek*
Forest Faculty, Department of Forest Ecology, Agricultural University of Cracow, Al. 29 Listopada 46, 31-425 Kraków, Poland

sample and identifying important discriminant variables responsible for the clustering and, thus, identifying possible sources of pollution or modeling the contribution of the identified sources to the formation of total concentration of the monitored chemical tracers. The possibility of identification and investigation of various items originating from long distance sources (aerosols, acid rain precursors, etc.) is still widely investigated [21]. The objective of the present study was to examine complex processes taking place in the local area of interest.

2. EXPERIMENTAL

2.1. Sampling sites and sampling procedure

The Dupniański Stream Catchment with an area of 1.68 km^2 is located in Southern Poland in the Silesian Beskid Mts. (49°34′N, 18°50′E) not far from the main industrial centers (Figure 1). The catchment is covered with Norway spruce (*Picea abies* Karst) stands of different ages growing on Dystric Cambisols developed on Istebna sandstone.

The studies were conducted in 1999–2003 following the methods described in the ICP-Forest Manual [22] and by Małek [23]. A bulk precipitation (BP) sampler was installed in the middle of the catchment at an elevation of 725 m above sea level within 500 m from the throughfall sampling point. During the growing season, that is from 1 May to 30 October of the same year, BP samples directly reaching the catchment were collected in special polyethylene collectors (five units with a 15-cm inlet diameter each) installed 0.5 m above ground in an open area, and connected to a polyethylene tube with an outlet joining a container and a measuring device (collector measuring the water amount each 6 min by a sound detector) installed in a bunker. In winter, that is from 1 November of the previous year to 30 April of the following year, six collectors (polyethylene, chemically neutral snow bags with a 15-cm inlet diameter each) were installed at 1.3 m above ground in the open area due to possible heavy snowfall up to 1 m in this region. Snow coming from individual snow bags was melted and combined. Comparable sampling conditions were ensured by using identical material (polyethylene) for collectors, the same inlet diameter, and by mixing water/snow collected in individual units for the two different collector systems to obtain the monthly average sample.

Characterization of meteorological conditions was accomplished on the basis of two sets of source data. One of them, related to temperature and annual rainfall, were delivered from the Carpathian Regional Gene Bank called 'Wyrchczadeczka' in the Wisła Forest District, located 5 km to the west from the Dupniański Stream Catchment [24], while those related to wind direction were obtained from the Institute of Meteorology and Water Management [25]. The annual rainfall (mm) determined in the consecutive years 1999, 2000, 2001, 2002 and 2003 was equal to 1376, 1641, 1557, 1196 and 1506, respectively. A 5-year average was equal to 1455 mm. The mean rainfall of vegetation seasons during the investigated time period was 828 mm (range: 673 mm in 2003 and 973 mm in 2001). The mean precipitation (snowfall or rainfall) of winter seasons was equal to 578 mm (range: 443 mm in 2002 and 839 mm in 2000). The annual

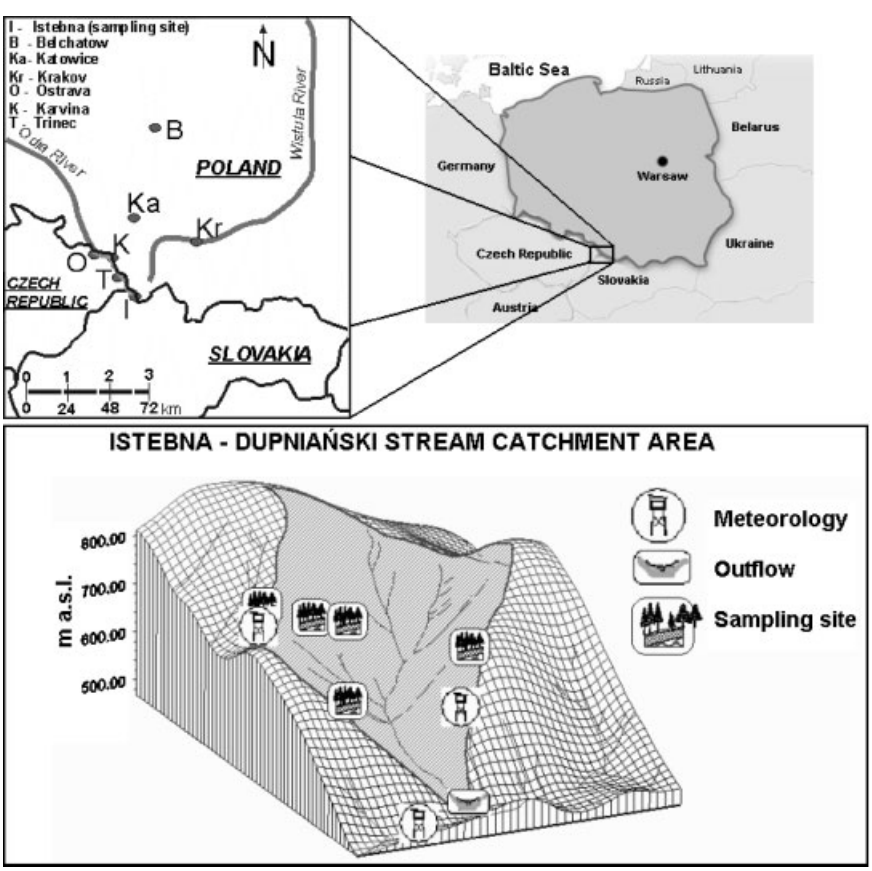

Figure 1. 3D visualization of the Dupniański Stream Catchment located in Southern Poland (Silesian Beskid Mts.).

temperature (°C) determined in the consecutive years 1999, 2000, 2001, 2002 and 2003 was equal to 7.1, 7.8, 6.5, 7.1 and 7.0, respectively. A 5-year average was equal to 7.1°C. The mean temperature of vegetation seasons during the investigated time period was 12.6°C (range: 12.4°C in 2001–2003 and 13.2°C in 2000). The mean temperature of winter seasons was equal to −0.7°C (range: −1.8°C in 2001 and 0.2°C in 2000).

2.2. Analytical procedures

Water samples were analyzed using a Dionex-320 ion chromatograph (Dionex Corp., Sunnyvale, CA, USA) to determine the concentrations of: Cl^-, NO_3^-, SO_4^{2-}, F^-, NH_4^+, Na^+, K^+, Ca^{2+}, Mg^{2+}, Fe^{2+}, Mn^{2+} and Zn^{2+}. The volume of bulk precipitation (V) was also determined. The procedure was validated using a certified reference material (CRM): a low-pH acid rain sample from southern Ontario (Canada), RAIN.97 - No 409. When the analyte was not detected in a sample, the value of one-third of the limit of detection was used in the data set due to chemometric requirements [26]. The limit of detection, expressed as concentration or mass, is derived from the smallest quantity that can be detected with a reasonable certainty for a given procedure [27]. The Tucker3 algorithm used for the calculations cannot handle missing values and, therefore, imputations for the non-detected values had to be used. Otherwise, either the samples or variables with non-detected values would have to be rejected and some valuable information would have been lost. The percentage of imputed values was for Zn^{2+} (8.3%), Mn^{2+} (21.7%), Fe^{2+} (31.7%) and F^- (61.7%). In spite of a high fraction of non-detected values of Fe^{2+} (small value of the mean and the median), it was included in data analysis while F^- was excluded.

2.3. Theory

2.3.1. Tucker3 model

The so-called Tucker3 model is one of the most basic multi-way models used in chemometrics. The model is defined by the decomposition of a three-way Table **X** into a three-way *core matrix* **Z** and three two-way loading matrices **A**, **B**, **C** (one for each mode):

$$x_{ijk} = \sum_{f}^{r}\sum_{g}^{s}\sum_{h}^{t} a_{if} b_{jg} c_{kh} z_{fgh} + e_{ijk},$$

where e_{ijk} represents the residual error term.

Tucker3 model algorithms were widely described before and the reader is referred to the referenced literature [10–18,28–32].

2.3.2. Self-organizing map (SOM) algorithm

The Self-Organizing Map (SOM) algorithm has been proposed by Kohonen [30] and is a neural-network model that implements a characteristic nonlinear projection from the high-dimensional space of sensory or other objects onto a low-dimensional array of neurons [34]. The term 'self-organizing' refers to the ability to learn and organize information without being given the associated-dependent output values for the input pattern [35]. An SOM shares with the conventional ordination methods the basic idea of displaying a high-dimensional objects manifold onto a much lower dimensional network in an orderly fashion (usually a two-dimensional space). An SOM consists of neurons organized on a regular low-dimensional grid. The number of neurons may vary from a few dozen up to several thousand. The neurons are connected to adjacent neurons by a neighborhood relation, which dictates the topology, or structure of the Kohonen map and, thus, similar objects (in our case sampling points), should be mapped close together on the grid. A training algorithm constructs the nodes in an SOM in order to represent the whole data set and their weights are optimized at each iteration step. In each step, one object from the input data set is chosen randomly and the distance between it and all the weight vectors of the SOM are calculated using some distance measure. Thus, the optimal topology is expected. In our study, the non-hierarchical K-means clustering algorithm was applied. Different values of k (predefined number of clusters) were tried and the sum of squares for each run was calculated. Finally, the best classification with the lowest Davies–Bouldin index was chosen (it is a function of the ratio of the sum of within-cluster scatter and between-cluster separation) [36].

The network organizes itself by adjusting the synaptic weights as the objects are presented to it; hence, the discovery of a new pattern is possible at any instant. Moreover, a SOM is noise tolerant; this property is highly desirable when site-measured data are used. Interesting SOM applications have been reported mainly in three fields: exploratory data analysis or data mining, the identification and monitoring of complex process states, and pattern exploration [37]. For the SOM algorithm, there are no precise rules for the choice of various parameters [37]. In this work, the Kohonen map was chosen as a rectangular grid with the number of nodes (n) determined from the following formula: $n = 5 \times \sqrt{\text{number of samples}}$ [38]. Furthermore, a hexagonal lattice was preferred because it does not favor horizontal or vertical direction [37]. Subsequently, for monthly averages the dimensionality of Kohonen's map was determined as 5×8 ($n = 5 \times \sqrt{60} = 38.7$; 60 is derived from 5 years multiplied by 12 months per year). Basically, the two largest eigenvalues of the training data were calculated and the ratio between side lengths of the map grid was set to the ratio between the two maximum eigenvalues. The actual side lengths were then set so that their product was close to the determined number of map units as stated above. Similar conditions were applied in the study presented previously by Park *et al.* [39].

All calculations in this study were performed using Statistica 6.0 and Matlab 6.5 software running on Windows 2000/XP platform. To perform SOM-based exploration, a free Teuvo Kohonen toolbox (SOM Toolbox 2.0) was applied [40], while in case of Tucker3, a free Rasmus Bro and Claus Anderson toolbox (The N-way toolbox for Matlab ver. 2.11) was used [41,42].

2.4. Data arrangement

The Tucker3 method can be successfully used for three-dimensional arrays [10–12]. The analyzed data were arranged in a three-way array of dimensionality 13 (variables) × 12 (months) × 5 (years). The initial data set contains mean values of the aforementioned variables measured in triplicate and presented as monthly averages measured for bulk precipitation. Each dimension was coded as *variables* (V) × *months* (M) × *years* (Y). Chemical variables, months and years constitute the modes of the array. The arrangement is presented schematically in Figure 2.

In environmental data the measured parameters are usually in different units, which is also the case here. Therefore, the data were first preprocessed by the use of scaling to unit standard

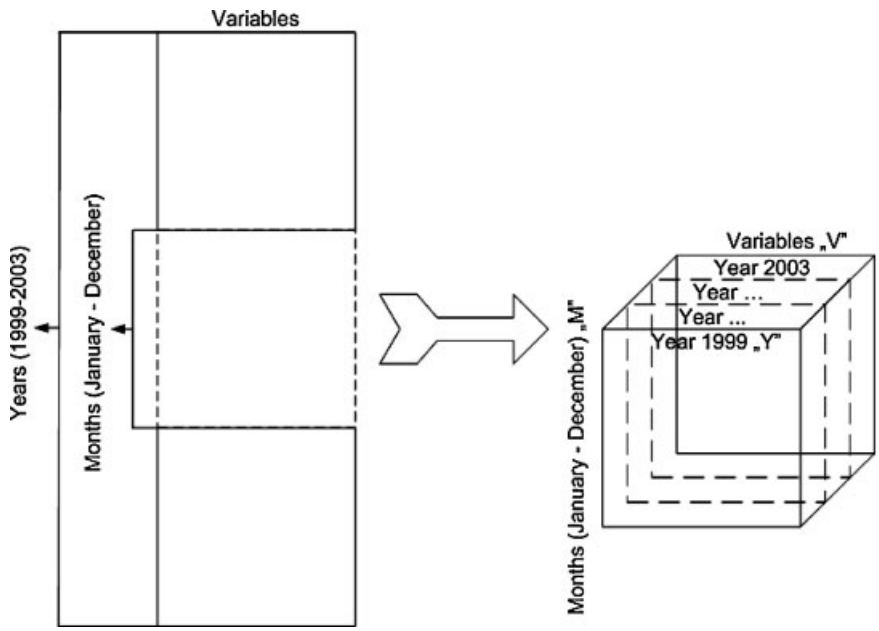

Figure 2. A graphical representation of the three-way data array.

deviation within the mode containing parameters [43]. The preprocessed three-way data were input to the Tucker3 method. To select a model with an optimal complexity, the variance of each combination of model complexities ($V \times M \times Y$) starting from the model with one factor in each mode, [111], to model with complexity [555] was evaluated (Figure 3). In general, the optimal model is the one with a small number of factors in each of the modes, but explaining a large part of the data variance. In practice, a trade-off between both requirements is needed. In Figure 3, the *x*-axis presents the models sorted according to $V \times M \times Y$, starting from those with the smallest *V*, followed by the models with the smallest *M* and *Y*. The *y*-axis gives the percentage of the explained data variance. It has to be pointed out that a line connecting models of certain complexities was drawn for illustrative purpose only and does not convey any mathematical meaning.

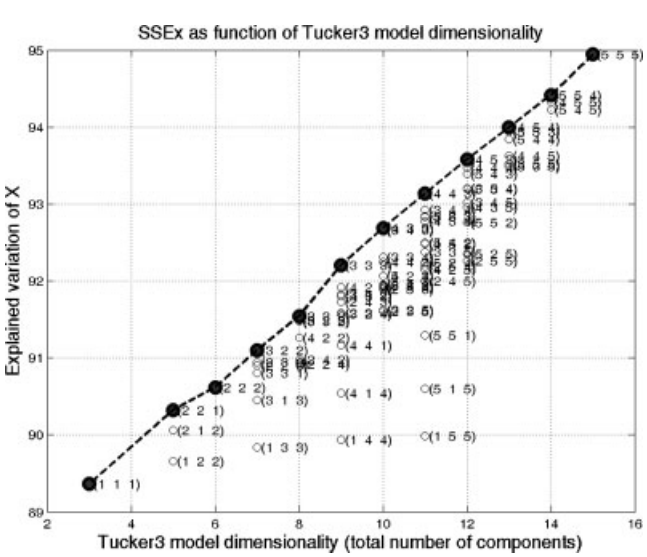

Figure 3. SSEx as function of Tucker3 model dimensionality.

3. RESULTS AND DISCUSSION

As shown in Figure 3, for bulk precipitation the percent of data variance explained increases almost linearly from 89.3% for the model with the lowest complexity [111] up to 94.9% for the model with the highest complexity [555]. Because no clear cut or bend is seen on the curve in Figure 3, several models ([1xx], [2xx], [3xx], [4xx] and [5xx]) were taken into account when selecting the final one. Models marked as [1xx] have a relatively low percentage of explained data variance, which does not exceed 90%. In contrast, models of the highest complexities marked as [4xx] and [5xx] explain more than 92%, but their interpretation becomes significantly more difficult. Based on a close inspection of Figure 3, and having in mind that the models with the highest complexity do not considerably change the explained variance, it can be argued that the appropriate dimensionality of the Tucker3 model can be described by the following rules: $1 < V < 4$; $1 < M < 3$ and likewise $Y < 3$. Using these rules, several models (e.g. [221], [222], [322], [233] and others) were preliminarily interpreted and finally the decomposition model with complexity [322], which explains 91.1% of the total data variance, was chosen to be interpreted in more detail. At the same time, a set of possible Tucker3 models was validated according to the procedure implemented in N-way Matlab toolbox. The cross-validation was performed in such a way that part of the data were set to missing, the models were fitted to the remaining data, and the residuals between fitted and true left-out elements were calculated. As presented in Figure 4, the validation procedure confirmed a high agreement between the fitted and cross-validated model with complexity [322]. To clarify all terms of variance used in the paper (data variance, core variance) it should be mentioned that data variance combines all the values in a data set to produce a measure of spread. A percentage of explained data variance indicates a part of total information related to a particular research object carried out in a spread data. In contrast, core variability is limited only to all elements of the core array and summation of the variance explained by all of the elements in the

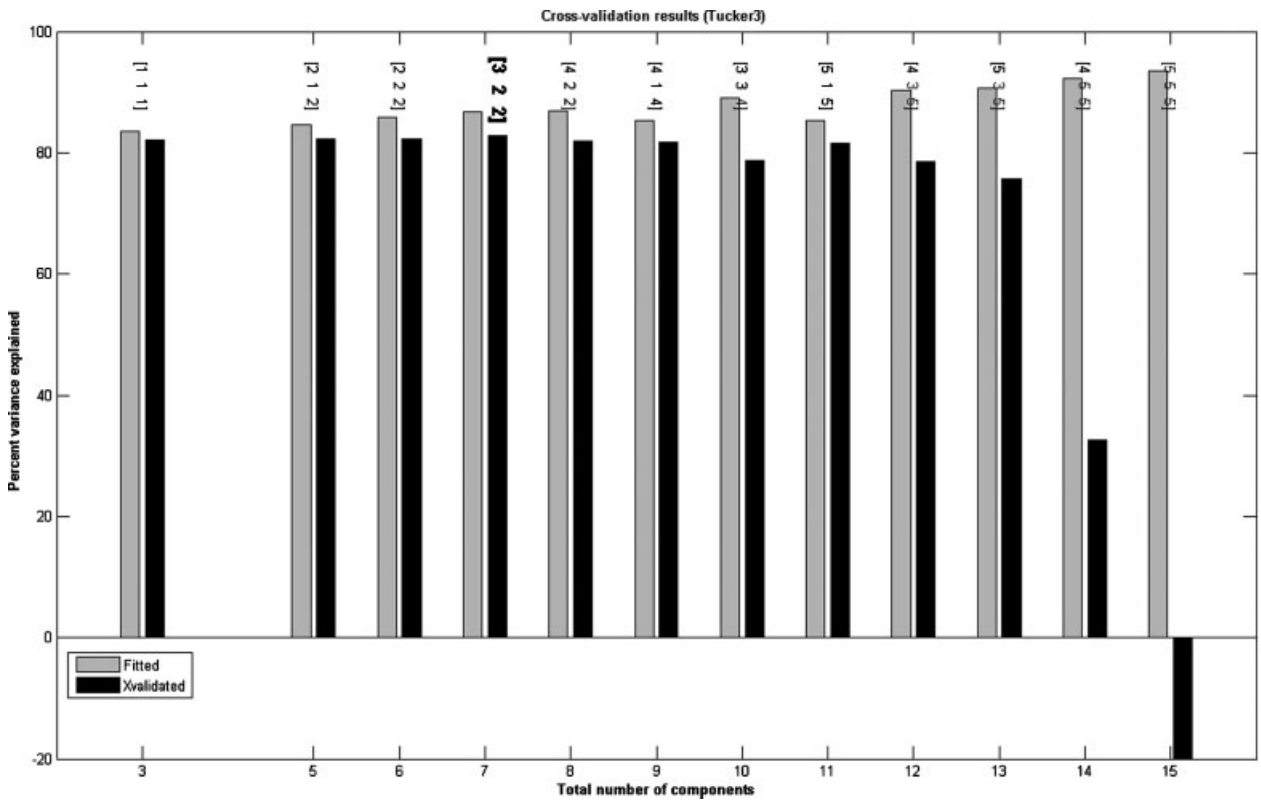

Figure 4. Cross-validation results of the best Tucker3 models arranged with an ascending number of total number of components included in it.

core array should give 100%. For an ordinary orthogonally constrained Tucker3 model applied in this study, the core variance is distributed among a few elements of the core array. Moreover, the rotation to optimal diagonality with conceivable important off diagonal elements of the core was not considered. Only in case of cubic cores (e.g. $3 \times 3 \times 3$), a diagonal optimization is efficient, because otherwise superdiagonality is not that meaningful. The distribution of core variance among few elements of the core array implies an easier interpretation of the obtained mode, due to consideration of a small number of interactions between modes. The three most important elements of the core array are listed in Table I in a descending order of the explained core variance. They are: [111], [221] and [222] accounting for 99.2% of the total core variance. Thus, for the interpretation concerning bulk precipitation, three loading vectors (factors) in the first mode, *V*, describing the variables, two loading vectors in the second mode, *M*, describing months and two loading vectors in the third mode, *Y*, describing years are needed.

Table I. Three most important core elements of a three-way Tucker3 model with complexity [322]

Core element	Explained core variance (%)	Core element value
[111]	98.05	−70.95
[221]	0.78	6.34
[222]	0.41	4.59

A description of loading plots will be given with the interpretation of the three most important elements of the core array. The results interpretation algorithm is proposed by Henrion [10] and will be demonstrated based on an example concerning the studied data. The loading plots for different modes are presented in Figure 5.

The first prominent core element [111] explains 98.05% of the core variance and reflects the interaction between the first factors in each of the modes (*V*1, *M*1, *Y*1). *V*1 is mainly connected with the volume of bulk precipitation with a strong positive loading and Zn^{2+}, Fe^{2+} and Mn^{2+} with negative loadings. The variables along *M*1 are ranked according to their positive values (Figure 4). The highest values are observed for May, June and July (growing season). Along *Y*1 all loadings have negative values. Taking into account the sign of bulk precipitation volume along *V*1(+), *M*1(+) and *Y*1(−) and the sign of the core element [111], which is '−', it can be deduced that the sign of the product is '+' and thus the volume of bulk precipitation increases during the growing period of the year. Taking into account the sign of Zn^{2+}, Fe^{2+} and Mn^{2+} along *V*1(−), *M*1(+) and *Y*1(−) and the sign of the core element [111], which is '−', it can be deduced that the sign of their product is '−'. Thus, it can be pointed out that generally the concentrations of Zn^{2+}, Fe^{2+} and Mn^{2+} in the dust particles rich in heavy metals are lower during the winter season than during the growing season of the same year.

The second important core element [221] explains 0.78% of the core variance. It reflects now the second factor in the first and

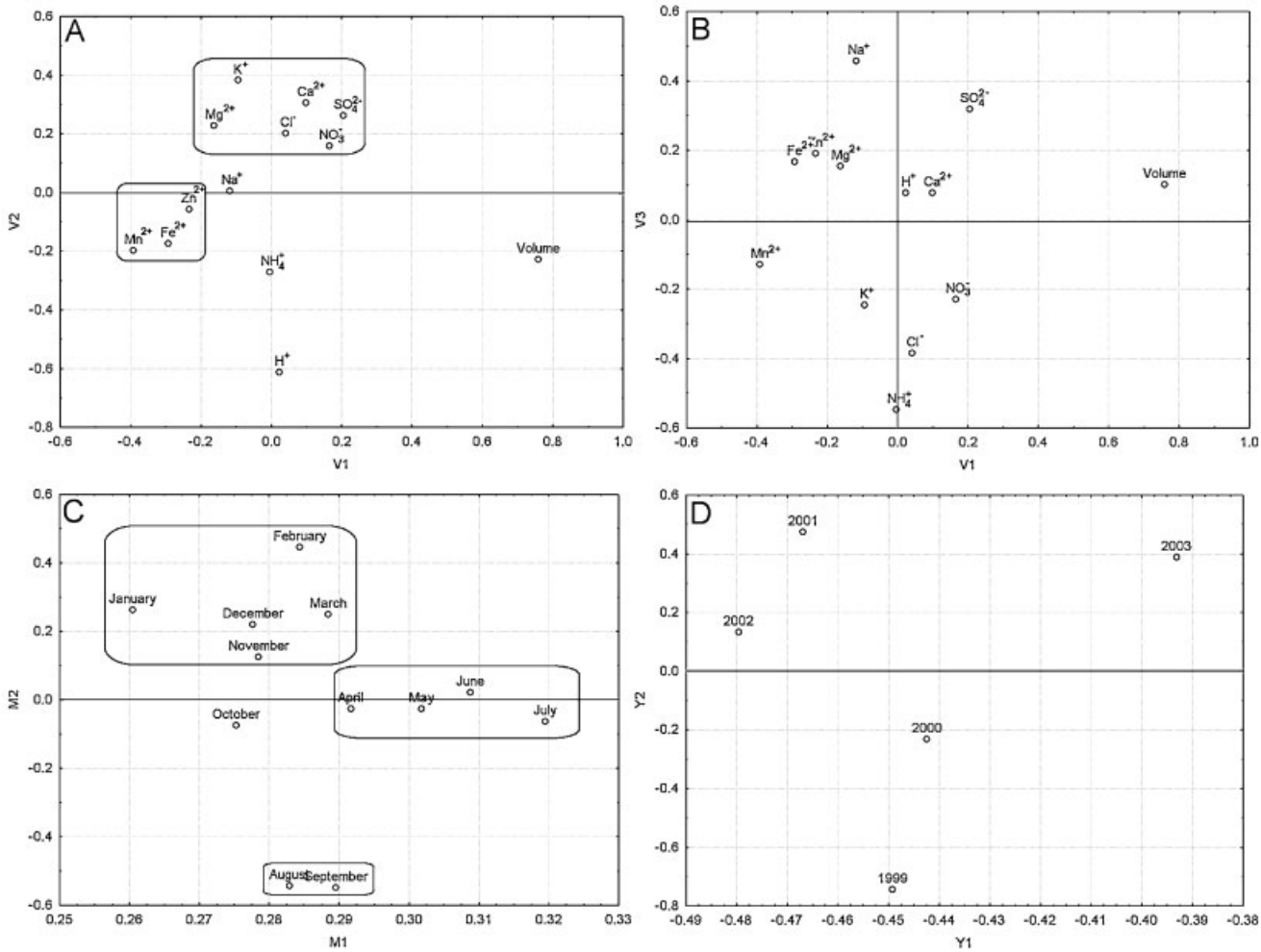

Figure 5. Loading plots of the three-way Tucker3 model of complexity [322]: (A) projection of chemical parameters on the plane defined by *V*1 and *V*2; (B) projection of chemical parameters on the plane defined by *V*1 and *V*3; (C) projection of months on the plane defined by *M*1 and *M*2; (D) projection of years on the plane defined by *Y*1 and *Y*2.

second modes as well as the first factor in the third mode (*V*2, *M*2, *Y*1). Factor *V*2 reflects the co-emission of SO_4^{2-} and NO_3^- precursors (SO_2, NO_x) originating from the combustion of fossil fuels and biomass and from automobile exhausts. Moreover, positive loading values for K^+, Mg^{2+} and Ca^{2+} prove that both SO_4^{2-} and NO_3^- from atmospheric sulfur dioxide and nitrogen oxides emission combine with K^+, Ca^{2+} and Mg^{2+} ions, and in the form of K_2SO_4, $MgSO_4$ and $CaSO_4$ are deposited on the ground together with precipitation. A similar explanation was presented previously by Plaisance *et al*. [44]. Factor *M*2 segregates months into two clearly distinguished groups. One is related to the winter season and combines January, November, December, March and February, while the second combines the two months from the late part of the growing season: August and September. Along *Y*1, all loadings have negative values. An explanation of the meaning of core element [221] requires two calculations. The first one takes into account the sign of Mg^{2+}, K^+, Ca^{2+}, Cl^-, NO_3^- and SO_4^{2-} along *V*2(+), *M*2(+) and *Y*1(−) and the sign of the core element [221], which is '+', resulting in the '−' sign of their product. It can thus be pointed out that, in general, the concentration of salts formed in the reaction between sulfur dioxide and nitrogen oxides with alkali cations decreases during the winter periods of the year, while increases substantially during August and September. The second observation is based on calculation of the final sign and can be obtained as a sum of the following factor signs: *V*2(+), *M*2(−), *Y*1(−) and the core element sign, which is '+'.

The third important core element [222] explains 0.41% of the core variance. It reflects the second factors of each mode (*V*2, *M*2, *Y*2). This core element explains fine annual variations in the concentration of anthropogenic pollutants. Taking into account the sign of Mg^{2+}, K^+, Ca^{2+}, Cl^-, NO_3^- and SO_4^{2-} along *V*2(+), *M*2(+) and *Y*1(+), and the sign of the core element [222], which is '+', it results in the '+' sign of their product. It means that, in general, for all the investigated variables higher concentrations were observed during the winter months over the period between 2001 and 2003, while the lowest concentrations were found in 1999.

To verify the preliminary results obtained by Tucker3 modeling, a self-organizing map based (SOM) exploration approach was applied. For monthly averages, the dimensionality of a unified distance matrix (U-matrix), which visualizes the distances

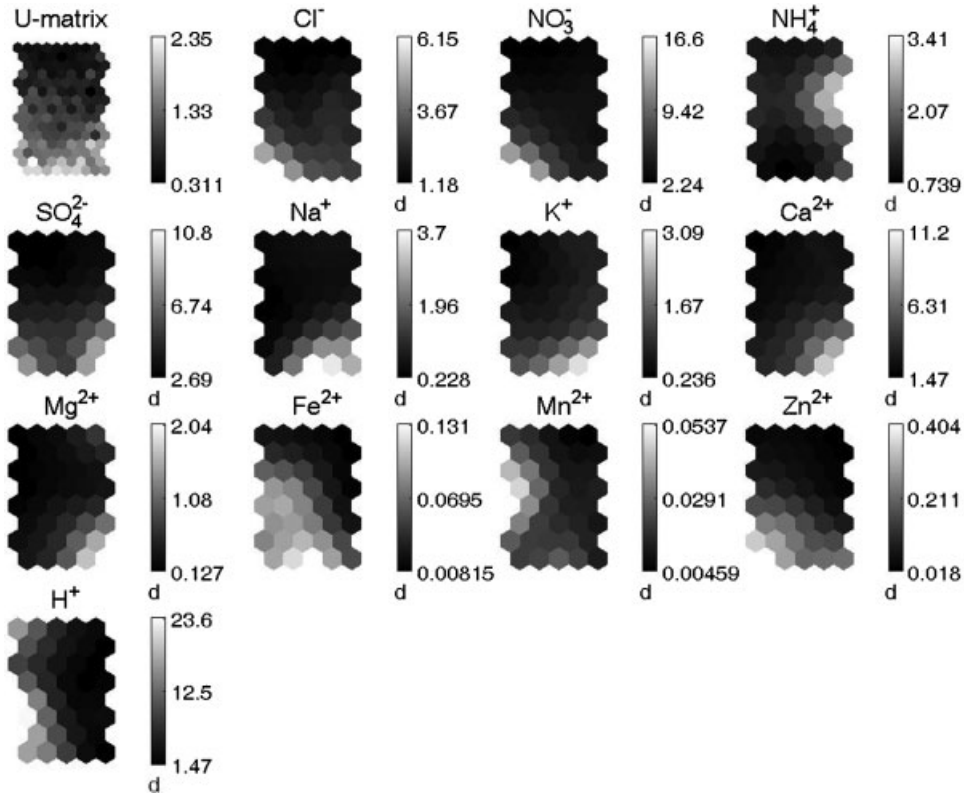

Figure 6. Visualization of the similarities in the unified distance matrix and variability of individual parameters in the space of objects.

between neighboring map units and helps to see the cluster structure on the map, was 9×15. The dimensionality of a Kohonen map presenting the variability of individual inorganic analytes in the space of objects was 5×8. Large values of the U-matrix indicate a cluster border while uniform areas of a small value indicate the clusters themselves [45]. Both the U-matrix and Kohonen's map presenting the variability of individual inorganic analytes in the space of objects are shown in Figure 6.

A visual assessment of similarities in gray-scale patterns for chemical variables allows identification of the correlated variables as follows ($p = 0.95$, $R_{crit} = 0.25$): $Cl^-NO_3^-$ (0.80), Cl^-Zn^{2+} (0.66), Na^+–K^+ (0.54), Mg^{2+}–SO_4^{2-} (0.57), Ca^{2+}–SO_4^{2-} (0.38), $NO_3^-SO_4^{2-}$ (0.41), $NO_3^-Zn^{2+}$ (0.65), $SO_4^{2-}Zn^{2+}$ (0.47), Na^+–K^+ (0.54), Na^+–Ca^{2+} + (0.40), Na^+–Mg^{2+} (0.45), K^+–Ca^{2+} (0.59), K^+–Mg^{2+} (0.43), Fe^{2+}–Mn^{2+} (0.42) and Fe^{2+}–Zn^{2+} (0.44). The obtained correlation coefficients are in good agreement with the results presented as the explanation of meaning of [221] core element in a Tucker3 approach. After that, the objects were clustered based on the K-means mode offered by a SOM. Different values of k (predefined number of clusters) were tried and the sum of squares for each run was calculated. Finally, the best classification with the lowest Davies–Bouldin index value (presented in Figure 7) was chosen. It can be seen that a four clusters configuration has the lowest index. As mentioned above for monthly averages, the dimensionality of the Kohonen's map was 5×8, and it is clear that more than one case from the initial data set ($n = 60$) was related to a particular hexagon. Cases included into each hexagon were grouped in agreement with the cluster borders. Setting up the initial mean values of chemical parameters along with the classification results allows obtaining a relatively reasonable interpretation of clustering pattern (Figure 8) associated with a temporal variation in the chemical profile of bulk precipitation in the Dupniański Stream Catchment. Clusters I–IV include different numbers out of a total of 60 cases as follows: I-3, II-5, III-15, IV-37. Cluster I (CI) groups bulk precipitation samples collected in September 2000, October 2000 and July 1999. Compared to other groups, these samples are characterized by the highest mean concentration of Zn^{2+} (0.4 mg dm^{-3}) and NO_3^- (11.1 mg dm^{-3}). The concentration

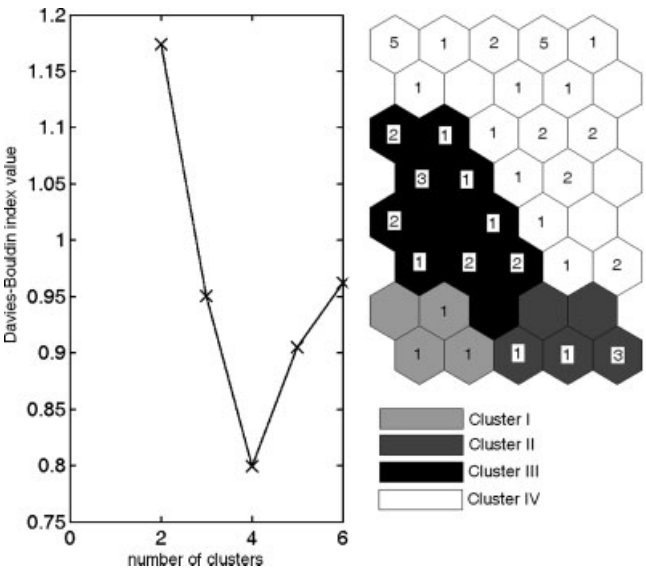

Figure 7. SOM exploration of chemical variables and clustering pattern according to the Davies–Bouldin index minimum value.

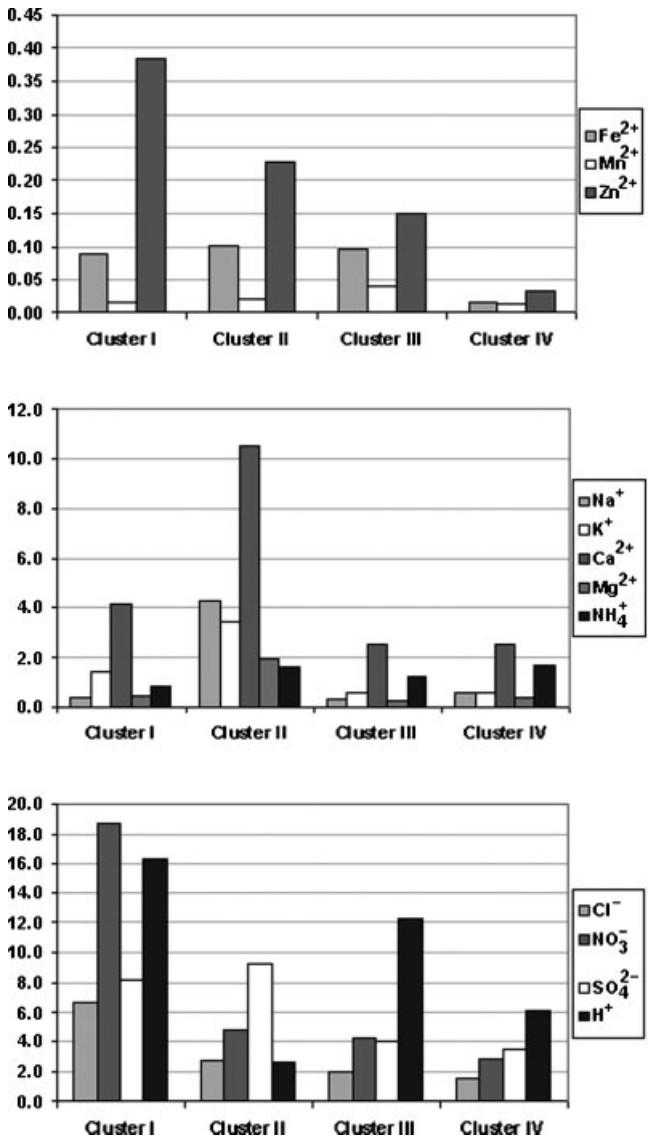

Figure 8. Mean values of chemical parameters Fe^{2+}, Mn^{2+}, Zn^{2+}, Na^{+}, K^{+}, Ca^{2+}, Mg^{2+}, NH_4^{+}, Cl^{-}, NO_3^{-}, SO_4^{2-} (mg dm^{-3}) and H^{+} (μg dm^{-3}) determined in bulk precipitation samples collected at Dupniański Stream Catchment divided into four clusters obtained by SOM clustering.

of SO_4^{2-} in CI (8.1 mg dm^{-3}) is similar to that in CII (9.2 mg dm^{-3}). In CI, two successive samples were grouped and this can indicate a strong, temporary impact of a pollution source located in the vicinity of the Polish-Czech Republic border. Moreover, an assessment of air mass movement trajectories at this time indicates that the majority of winds have blown from the south and southwest direction [24,25]. These two premises imply a possibility of pollution caused by the emission from the Ostrava-Karvina industrial region located 40 km and Trinec located 15 km from the Dupniański Stream Catchment. Cluster II (CII) groups bulk precipitation samples collected mainly in the late growing period (August 1999, 2000, 2002 and September 1999). Compared to other groups, these samples are characterized by the high concentration of Zn^{2+} (0.2 mg dm^{-3}) and SO_4^{2-} (9.2 mg dm^{-3}) and the highest concentration of Na^{+} (4.3 mg dm^{-3}) and K^{+} (3.5 mg dm^{-3}). Objects grouped in CII indicate a possibility of a cyclic impact of remote pollution sources of highly industrialized Katowice and Bełchatów regions, located 70 and 200 km, respectively, in the north direction and also highly urbanized Cracow metropolitan area, which is located 80 km north-east from the Dupniański Stream Catchment. Such a possibility seems to be reasonable because of prevailing air mass movement trajectories from the north and north-east [24,25]. Twenty-five out of 37 samples grouped in CIV were collected in the winter season, characterized by higher concentrations only of Cl^{-} (2.3 mg · dm^{-3}), NO_3^{-} (3.1 mg · dm^{-3}), NH_4^{+} (2.2 mg · dm^{-3}), K^{+} (0.9 mg · dm^{-3}) and Ca^{2+} (3.3 mg · dm^{-3}), during which the westerly air mass movements prevailed [24,25]. During the growing season (11 out of 15 samples), north and northwesterly winds prevailed [24,25], resulting in higher concentrations of heavy metals: Fe^{2+} (0.2 mg dm^{-3}), Mn^{2+} (0.13 mg dm^{-3}) and Zn^{2+} (0.038 mg dm^{-3}) as well as SO_4^{2-} (3.6 mg dm^{-3}), H^{+} (6.4 μg dm^{-3}), Na^{+} (0.5 mg dm^{-3}) and Mg^{2+} (0.4 mg dm^{-3}).

4. CONCLUSIONS

Tucker3 modeling allows identification of general factors responsible for the variability of bulk precipitation quality in a multidimensional space.

Preliminary results obtained by a Tucker3 model were subsequently verified by a self-organizing map approach. Moreover, a SOM proves its 'resolving power', allowing identification of both a cyclic impact of remote sources of pollutants and temporarily 'hot-spot' events.

Chemical composition of bulk precipitation depends on specific meteorological conditions, which allows defining major directions of air-mass movement and thus identifying sources of air pollution. Specific meteorological conditions are in a good agreement with the existence of maximum values of pollutants determined in samples of bulk precipitation.

Samples of bulk precipitation collected in September 2000, October 2000 and July 1999, characterized by the highest concentration of Zn^{2+} and NO_3^{-}, indicate a strong, temporary impact of a pollution source located in the vicinity of the Polish-Czech Republic border (pollution from Ostrava-Karvina and Trinec industrial regions).

Compared to other clusters obtained by an SOM, the bulk precipitation samples collected during the late growing period (August 1999, 2000, 2002 and September 1999) indicate an impact of highly industrialized Katowice and Belchatów regions as well as highly urbanized Cracow metropolitan area region because of high concentrations of Zn^{2+}, SO_4^{2-}, Na^{+} and K^{+}.

The concentration of elements in bulk precipitation depends on the movements of the air masses, with the prevailing westerly direction in winter, which brings mainly: Cl^{-}, NO_3^{-}, NH_4^{+}, K^{+} and Ca^{2+}, and north and northwesterly direction during the growing season, which delivers mainly heavy metals (Fe, Mn and Zn), SO_4^{2-}, H^{+}, Na^{+} and Mg^{2+}.

Acknowledgements

This research was supported financially within the framework of the project: 'Optimization of chemometric techniques of exploration and modelling results originating from environmental constituents pollution monitoring' (1439/T02/2007/32).

The authors would like to thank Prof. Rasmus Bro (University of Copenhagen, Faculty of Life Sciences) and Prof. Andrzej Przyjazny (Science & Mathematics Department, Kettering University, USA) for their invaluable suggestions and comments.

REFERENCES

1. Bytnerowicz A, Godzik S, Poth M, Anderson I, Szdzuj J, Tobias C, Macko S, Kubiesa P, Staszewski T, Fenn M. Chemical composition of air, soil and vegetation in forests of the Silesian Beskid Mountains, Poland. *Water Air Soil Pollut.* 1999; **116**: 141–150.
2. Bytnerowicz A, Godzik B, Frączek W, Grodzińska K, Krywult M, Bada O, Barančok P, Blum O, Čzerny M, Godzik S, Maňkovská B, Manning W, Moravčik P, Musselman R, Oszlányi J, Postelnicu D, Szdżuj J, Varšavova M, Zota M. Ozone, sulphur dioxide and nitrogen dioxide air pollution in forests of the Carpathian Mountains. In Effects of Air Pollution on Forest Health and Biodiversity in Forests of the Carpathian Mountains, Szaro RC, *et al*. (eds). IOS Press: 2002; 138–160.
3. Maňkovská B, Čzerny M, Moravčik P, Godzik B, Grodzińska K, Badea O, Barančok P, Oszlányi J, Varšavova M, Fleischer P, Blum O, Parpan V, Bytnerowicz A, Szaro R. Chemical and morphological changes in Carpathian Mountains trees caused by air pollution. In Effects of Air Pollution on Forest Health and Biodiversity in Forests of the Carpathian Mountains, Szaro RC, *et al*. (eds). IOS Press: 2002; 173–184.
4. Kulhavý J, Drápelová I, Remeš M, Lesná J. Depozični toky látek, stav půdy a minerální výživy ve smrkové monokulturře v horské oblasti (Deposition flow, soil condition and mineral nutrition in spruce monoculture of mountain region). In Trvale udržitelné hospodařeni v lesích a v krajiné od koncepce k realizacji (Sustainable forest and landscape management from concept to reality), Kuhlavý J, Skoupý A, Kantor P, Simon J (eds). eborník významných výsledků institucionálního výzkumu LDF MOLU v Brně řešeného v letech 1999–2004. Brno: 2005; 215–223.
5. Florek M, Maňkovská B, Oszlányi J, Frontasyeva MV, Ermakova E, Pavlov SS. The Slovak heavy metals survey by means the bryophyte technique. *Ecologia (Bratislava)* 2007; **26**(1): 99–114.
6. Małek S, Martinson L, Sverdrup H. Modeling future soil chemistry at a highly polluted forest site at Istebna in Southern Poland using the "SAFE" model. *Environ. Pollut.* 2005; **3**(137): 568–573.
7. Małek S, Astel A. The effect of stand age on throughfall chemistry in spruce stands in the Potok Dupniański Catchment in the Silesian Beskid Mountains, Southern Poland. *Sci. World J.* **7**(S1): 181–191.
8. Brimblecombe P, Hara H, Houle D, Novak M. Acid rain – Deposition to recovery. *Water Air Soil Pollut. Focus* 2007; **7**: 1–3.
9. Stanimirova I, Simeonov V. Modeling of environmental four-way data from air quality control. *Chemom. Intell. Lab. Syst.* 2005; **77**: 115–121.
10. Henrion R. N-way principal component analysis, Theory, algorithms and applications. *Chemom. Intell. Lab. Syst.* 1994; **25**: 1–23.
11. Geladi P. Analysis of multi-way (multi-mode) data. *Chemom. Intell. Lab. Syst.* 1989; **7**: 11–30.
12. Smilde A, Bro R, Geladi P. Multi-way analysis. Applications in the chemical sciences. John Wiley & Sons, Ltd.: Chichester, UK, 2004.
13. Bro R. Review on multiway analysis in chemistry – 2000–2005. *Crit. Rev. Anal. Chem.* 2006; **36**(3–4): 279–293.
14. Morais H, Ramos C, Forgács E, Jakab A, Cserháti T, Oliviera J, Illés T, Illés Z. Comparison of Principal Analysis and the Tucker3 Model. *QSAR Combin. Sci.* **2003**(22): (4): 449–455.
15. Giordani P, Henk A, Kiers L. Three-way component analysis of interval-valued data. *J. Chemometr.* 2005; **18**(5): 253–264.
16. Møller SF, Frese J, Bro R. Robust methods for multivariate data analysis. *J. Chemometr.* 2005; **19**(10): 549–563.
17. Henrion R. On global, local and stationary solutions in three-way data analysis. *J. Chemometr.* 2000; **14**(3): 397–413.
18. Jos MF, Berge T. Simplicity and typical rank of three-way arrays, with applications to Tucker3 analysis with simple cores. *J. Chemometr.* 2004; **18**(1): 17–21.
19. Geladi P, Xie Y-L, Polissar A, Hopke P. Regression on parameters from three-way decomposition. *J. Chemometr.* 1998; **12**: 337–354.
20. Singh KP, Malik A, Singh VK, Basant N, Sinha S. Multi-way modelling of hydro-chemical data of an alluvial river system – a case study. *Anal. Chim. Acta* 2006; **571**: 248–259.
21. Häsänen E, Lipponen M, Minkkinen P, Kattainen R, Markkanen K, Brujkhanov P. Elemental concentration of aerosol samples from the Baltic Sea area. *Chemosphere* 1990; **21**(3): 339–347.
22. ICP-Forest Manual. Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests, 4th edn. UN-ECE, Fed. Res. Centre for Forestry and Forest Products (BFH): Hamburg, Germany, 1998.
23. Małek S. The effect of the age of spruce stands on the balance of elements in the Potok Dupniański catchment. *Dendrobiology* 2004; **51**: 61–66.
24. Feliksik E, Durło G. Climatological characterisation of the area of the Carpathian Regional Gene Bank in the Wisła Forest District. *Dendrobiology* 2004; **51**: 47–55.
25. Institute of Meteorology and Water Management (IMGW – Poland) http://www.imgw.pl [15 April 2007].
26. Astel A, Mazerski J, Polkowska Ż, Namieśnik J. Application of PCA and time series analysis in studies of precipitation in Tricity (Poland). *Adv. Environ. Res.* 2004; **8**: 337–349.
27. IUPAC Compendium of Chemical Terminology, 2nd (1997) www.iupac.org/goldbook/L03540.pdf [24 April 2007].
28. Andersson CA, Bro R. Improving the speed of multi-way algorithms: Part I. Tucker3. *Chemom. Intell. Lab. Syst.* 1998; **42**: 93–103.
29. Paatero P, Andersson CA. Further improvements of the speed of the Tucker3 three-way algorithm. *Chemom. Intell. Lab. Syst.* 1999; **47**: 17–20.
30. Vandeginste BGM, Massart LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J. Handbook of Chemometrics and Qualimetrics: Part B, vol. 20B. Elsevier: Netherlands, 1998; 154–156.
31. An interactive introduction to the Tucker3 model in chemometrics http://www.models.kvl.dk/courses/tucker/ttext.htm [5 May 2007].
32. Introduction to Chemometrics and Statistics http://dissertations.ub.rug.nl/FILES/faculties/science/1998/j.nilsson/c2.pdf [5 May 2007].
33. Kohonen T. Self-organizing formation of topologically correct feature maps. *Biol. Cybern.* 1982; **43**: 59–69.
34. Kohonen T, Oja E, Simula O, Visa A, Kangas J. Engineering applications of the self-organizing map. *Proc. IEEE* 1999; **84**(10): 1358–1384.
35. Mukherjee A. Self-organizing neural network for identification of natural modes. *J. Comput. Civil. Eng.* 1997; **11**(1): 74–77.
36. Davies DL, Bouldin DW. A cluster separation measure. *Proc. IEEE Trans. on Pattern Recog. Mach. Intell.* 1979; **1**(2): 224–227.
37. Kohonen T. Self-organizing maps, 3rd edn. Springer: Berlin, 2001.
38. Vesanto J. Neural network tool data mining: SOM Toolbox. *Proceedings of Symposium on Tool Environments and Development Methods for Intelligent Systems (TOOL-MET2000)*, Oulu, Finland, 2000; 184–196.
39. Park YS, Tison J, Lek S, Giraudel JL, Coste J, Delmas F. Application of a self-organizing map to select representative species in multivariate analysis: A case study determining diatom distribution pattern across France. *Ecol. Inform.* 2006; **1**: 247–257.
40. Vesanto J, Himberg J, Alhoniemi E, Parhankagas J. SOM Toolbox for Matlab 5, Report A57, 2000; http://www.cis.hut.fi/projects/somtoolbox/ [27 April 2007].
41. Bro R, Andersson CA, N-way Toolbox for MATLAB. Version 2.10; http://www.models.kvl.dk/source/nwaytoolbox/download.asp [15 April 2007].
42. Andersson CA, Bro R. The N-way Toolbox for MATLAB. *Chemom. Intell. Lab. Syst.* 2000; **52**: 1–4.
43. Bro R, Smilde A. Centering and, scaling in component analysis. *J. Chemometr.* 2003; **17**: 16–33.
44. Plaisance H, Galloo JC, Guillermo R. Source identification and variation in the chemical composition of precipitation at two rural sites in France. *Sci. Total Environ.* 1997; **206**: 79–93.
45. Ultsch A, Siemon HP. Kohonen's self organizing feature maps for exploratory data analysis. In Proceedings of International Neural Network Conference (INNC'90). Kluwer: Dordrecht, Netherlands, 1990; 305–308.

