VERSITA

Cent. Eur. J. Chem. • 7(1) • 2009 • 20-30
DOI: 10.2478/s11532-008-0082-8

Central European Journal of **Chemistry**

Application of linear discriminant analysis to the study of dew chemistry on the basis of samples collected in Poland (2004-2005)

Research Article

Kamila Klimaszewska[1]*, Costel Sârbu[2], Żaneta Polkowska[1], Dariusz Lech[3]
Piotr Pasławski[3], Stanisław Małek[4], Jacek Namieśnik[1]

[1] Department of Analytical Chemistry, Chemical Faculty, Gdansk University of Technology, 80-952 Gdańsk, Poland

[2] Faculty of Chemistry and Chemical Engineering Babeş-Bolyai University, 400028 Cluj-Napoca, Romania

[3] Central Chemical Laboratory, Polish Geological Institute, 00-975 Warsaw, Poland

[4] Forest Faculty, Agricultural University of Cracow, 31-425 Krakow, Poland

Received 23 July 2008; Accepted 23 September 2008

Abstract: This paper presents the results of determination of selected characteristics (anions, cations, formaldehyde, hydrogen peroxide, phenols, TC, TIC, TOC and metals) in dew samples collected in six different sites in Poland. The influence of local parameters (*e.g.* wind speed, humidity) was investigated. Discriminant analysis was applied to the study of several dew samples collected from different sampling sites covering six agglomerations in Poland. Discriminant function analysis was used not only for classifying samples into different groups with a better than chance accuracy, but also for detecting the most important variables that discern between the groups of samples considered. In this way it was possible to identify which ions or other physicochemical features are responsible for the similarities or differences observed between different groups of dew samples. A good agreement with their origin and location was observed. It is interesting to note that the classification of all samples was dominated by pH, wind direction, pressure and temperature with a significant contribution of Na^+ and Cl^- ions.

Keywords: *Dew Analysis of dew • Chemical composition • Discriminant analysis*

1. Introduction

Dust, gases, particles and aerosols are removed from the atmosphere both by dry and wet deposition (rain, snow, drizzle snow pellets, hail, cloud, fog, dew, rime, hoarfrost). The wet deposition process is a major way for pollutants to be transfered from the atmosphere to the biosphere; it plays an important role in controlling the pollutants concentration [1].

There are some problems connected with the chemistry of wet deposition. The chosen technique of collecting such specific environmental samples has a strong influence on the whole analytical procedure. Analytical errors made at this stage are difficult to evaluate, and minimization their influence on the final result is very hard. The choice of a proper sampler mostly depends on the type of atmospheric precipitation. Such problems as: specific composition and heterogenous character of the samples, small volumes of the collected samples,

* E-mail: chemanal@pg.gola.pl

low concentrations of the analytes determined in the samples, interactions between pollutants present in the samples are also important in the studies of atmospheric deposition processes.

Recently, an increased interest is being observed in the chemistry of dew - one of form of wet deposition.

Dew is a local phenomenon, significantly influenced by microclimatic ambiance, land profile, and favourable meteorological conditions. Studies revealed that concentrations of the chemical species in the dew samples were much higher than in the rain samples collected in the same areas [2-6]. The dew settles can be good indicators of the level of atmospheric pollution in a geographical region of interest, because of the types and quantities of the chemicals and materials transported by them and the range of their interactions.

This study presents the results of determination of selected characteristics (anions, cations, formaldehyde, hydrogen peroxide, phenols, TC, TIC, TOC and metals) in dew samples collected in six different agglomerations in Poland. The influence of local parameters (*e.g.*, wind speed, humidity) has been investigated. Discriminant function analysis was used not only for classifying samples into different groups with a better than chance accuracy, but also for detecting the most important variables that differ for the groups of samples considered. In this way it was possible to identify which ions or other physicochemical features are responsible for the similarities or differences observed between different groups of dew samples in a good agreement with their origin and location.

2. Experimental Procedures

2.1 Sampling

Samples of dew were collected using dedicated samplers of known geometry exposed to the atmosphere. Design of the sampler was based on one described by Muselli *et al.* [6]. The collecting surface of this sampler (2000 mm by 2000 mm) was made of rigid polyethylene mounted on a wooden frame. The sampler was ascended at a 30-degree angle to promote the flow of condensation droplets to a collector (groove) and subsequently to a collection vessel (50 cm^3 flask). Droplets remaining on the collection surface were transferred to the collector using a polyethylene scraper. The dew samples were collected early in the morning. Before the expected appearance of dew, the collecting surface was flushed with deionized water and subsequently dried. Atmospheric deposition samples were collected during or immediately after a precipitation event. Dew collection took place only on rainless nights to eliminate any influence of rain droplets on collected samples. They were stored at low temperature (4-7ºC) without chemical preservatives, since the analysis was performed either directly on-site, or immediately after the samples were delivered to the laboratory. Due to the high contents of solid particles (sand, dust, etc.) in the samples, inclusion of a filtration stage was necessary in some analyses (0.45 mm, Millex®-HV).

Dew samples were collected in Gdynia, Gdańsk (August 2004 - October 2004), Mława, Sopot and Krakow (August 2005 - October 2005). Table 1 summarizes the general characteristics of the sampling sites.

2.2 Laboratory analysis work

Once collected, samples were analyzed immediately for pH, volume and conductivity. Selected anions and cations were determined by using ion suppressed chromatography (Dionex Corporation, USA) and quantified against synthetic rain standards (RAIN-97, CRM 409). Formaldehyde was determined spectrophotometrically (Merck, Germany) based on the reaction with chromotropic acid. In a solution acidified with sulphuric acid, formaldehyde reacted with chromotropic acid to form a violet dye that was measured [7]. Hydrogen peroxide was measured using a photometric technique based on the reaction with titanic acid ester. In a solution acidified with sulphuric acid, hydrogen peroxide and compounds containing hydrogen peroxide reacted with titanic acid ester to form yellow peroxotitanic acids, the concentration of which was determined photometrically. In buffered solution, in the presence of an oxidizing agent phenol and its ortho- and meta-substituted compounds reacted with 4-aminoantipyrine and formed a red compound that is determined photometrically. Total carbon was determined in the samples by converting all carbon compounds into CO_2 (mineralization) and determining its amount coulometrically. Total organic carbon was determined as the difference between the total carbon and the total inorganic carbon. Total carbon was determined by using CM 5300 Furnace Apparatus Version 1.0 (UIC INC.COULOMETRICS) and a coulometric detector (CM 5014 CO_2 Coulometer). Total inorganic carbon was determined by using CM 5130 Acidification Module (UIC INC.COULOMETRICS) and the coulometric detector. Metals were determined by using mass spectrometry with the ICP ionisation (ICP-MS, Elan DRC, PerkinElmer). Determination of metals was performed in the Central Chemical Laboratory of Polish Geological Institute in Warsaw. The analytical techniques usedin this study are summarized in Table 2.

Table 1. Description of the sampling sites

Sampler No.	Sampling site	Short description of the site	Short description of the area
1	Gdynia N = 54°29', E = 18°32'	100 m from main street, high traffic intensity, areas of small houses	Tricity area is a large municipality in the eastern part of the Baltic coast in Poland. The population of the region is about 500,000 in the three cities of Gdańsk, Gdynia and Sopot. Major point sources of pollution in the region include harbours, shipyards, oil refinery, power plants and phosphate fertilizer plant. Traffic and combustion of coal and oil in small residential furnaces are the main diffuse sources of air pollution in the region.
2	Gdynia N = 54°30', E = 18°32'	100 m from main street, high traffic intensity, areas of woodland	
5	Sopot N = 54°27', E = 18°34'	500 m from Baltic Gulf, the allotment situated near apartement complexes	
3	Gdańsk N = 54°20', E = 18°36'	200 m from Gdańsk ring road, a place isolated by the acoustic baffle	
4	Mława N = 53°15', E = 20°20'	the allotment situated near apartement complexes, 5 km from the city centre	The Mława is the Green Lungs of Poland. The population of the city is about 23,000 people. Major point source of pollution in the region is the E7 international road. The specific character of the climate in this city is connected both with marine and continental influences.
6	Kraków N = 50°03', E = 19°57'	the allotment situated near apartement complexes	Sampling point is located in Zielonki village - 14,5 km north from the Kraków centre (in straight line). This village is composed of one family houses situated close to each other. Potential sources of pollution are house heating systems (on wood, coal and gas), small traffic, and a big Cracow town agglomeration which lays about 2 km from the sampling point.

Table 2. The characteristics of the analytical techniques used in the study

Analyte	Technique	Analytical parameters	Limit of detection	Precision [% RSD]
Anions	IC	AS9-HC column (2 × 250 mm), AutoSuppression Recycle Mode ASRC®-ULTRA (2 mm), conductivity detection, eluent 9.0 mM Na_2CO_3, flow rate 0.25 mL min^{-1}	Br^-, F^-, Cl^-, NO_3^-, SO_4^{2-} = 0.01 mg dm^{-3} NO^{2-} = 0.05 mg dm^{-3} PO_4^{3-} = 0.04 mg dm^{-3}	1
Cations		CS12A column (2 × 250 mm), AutoSuppression Recycle Mode CSRS ®-ULTRA (2 mm), conductivity detection, eluent 20 mM Methanesulfonic Acid, flow rate 0.25 mL min^{-1}	0.01 mg dm^{-3}	
TC	Coulometry	High - temperature mineralization (950°C), carrier gas- O_2, flow rate 100 mL min^{-1}	0.1 mg C dm^{-3}	2
TIC		temperature 100°C, carrier gas - air, flow rate 100 mL min^{-1}		
H_2O_2	Spectrophotometry	Absorbance measured at 405 nm	0.10 mg dm^{-3}	5
Phenols		Absorbance measured at 495 nm	0.001 mg dm^{-3}	
HCHO		Absorbance measured at 585 nm	0.005 mg dm^{-3}	5
Metals	ICP-MS	Sample introduction system: sample uptake rate-1 $mLmin^{-1}$, nebulizer-Cross flow, spray chamber- Scott double-pass Plasma: nebulizer gas flow-Ar, 0,98 L min^{-1}, plasma gas flow-Ar, 15 L min^{-1}, RF power-1300 W Measure: scan mode-peak hopping, dwell time-100 ms, number of repetitions-10 sweeps / 3 replicates, total measuring time~100 s	Be, U = 0.004 μg dm^{-3}; Cs = 0.009 μg dm^{-3}; Sb, Pb = 0.01 μg dm^{-3}; Co, Se, Tl = 0.04 μg dm^{-3}; Li, Rb, Mo, Cd, Bi = 0.05 μg dm^{-3}; As, Ag = 0.09 μg dm^{-3}; V, Cr, Mn, Cu, Sn = 0.1 μg dm^{-3}; Ni, Ba = 0.2 μg dm^{-3}; B, Zn, Sr = 0.5 μg dm^{-3}; Fe = 1 μg dm^{-3}; Al = 2 μg dm^{-3}; Mg = 5 μg dm^{-3}; Ca = 10 μg dm^{-3}	2

2.3 Discriminant function analysis

Discriminant function analysis or discriminant analysis (DA) is based on the extraction of the linear discriminant functions of the independent variables by means of the qualitative dependent variables and several quantitative independent variables [8-11]. DA can be formulated as follows:

let $\mathbf{X} = \{x_1, ..., x_n\} \subset \mathbf{R}^p$ be a finite set of characteristic vectors, where n is the number of the samples (measurements) and p is the number of the original variables (predictors);

let $\mathbf{x}_i^j = [x_i^1, x_i^2, ..., x_i^p]^T$ and k be a nominal characteristic (grouping variable) each of which characterizes one of the k partition composing the partition substructure of the data set.

The partition of **X** into k groups is computationally very similar to analysis of variance (ANOVA/MANOVA), sharing many of the same assumptions and tests; the most important variables are selected, and variables contributing only marginally to the differentiation of groups will be removed. In a similar way as with principal component analysis [12-14], first the total variance/covariance matrix is calculated according to the following expression:

$$\mathbf{V} = {}^T\mathbf{XDX} \tag{1}$$

where **X** is the centered data matrix, $^T\mathbf{X}$ is the transpose matrix, **D** is the diagonal matrix (in most cases is the unity matrix). Considering a new characteristic defined as $\mathbf{c} = Xu$, one can calculate its variance by applying the relation (2).

$$||c||^2 = {}^T\mathbf{cDc} = {}^T\mathbf{u}^T\mathbf{XDXu} = {}^T\mathbf{uVu} \tag{2}$$

The total variance **V** can be divided into two components: the between-group variance **B** and within-group variance **W**, namely

$$\mathbf{V} = \mathbf{B} + \mathbf{W}, \tag{3}$$

and, as a consequence, the variance of the characteristic **c** becomes

$$||c||^2 = {}^T\mathbf{uVu} = {}^T\mathbf{uBu} + {}^T\mathbf{uWu} \tag{4}$$

In this case, it is very easy to observe that eq. (4) can be rewritten in the following form

$$\frac{{}^T\mathbf{uBu}}{{}^T\mathbf{uVu}} + \frac{{}^T\mathbf{uWu}}{{}^T\mathbf{uVu}} = 1 \tag{5}$$

and because any term from the left side is positive, equivalent results will be obtained indifferent of the maximum/minimum condition. However, in practice the first ratio in eq. (5) is maximized

$$\lambda = \frac{{}^T\mathbf{uBu}}{{}^T\mathbf{uVu}} \quad (0 \leq \lambda < 1) \tag{6}$$

and finally a similar matrix equation to that obtained in the case principal component analysis results in

$$\mathbf{V}^{-1}\mathbf{Bu} = \lambda\mathbf{u}, \tag{7}$$

where **λ** and **u** represent the eigenvalues (characteristic roots) and eigenvectors of the matrix $\mathbf{V}^{-1}\mathbf{B}$. The vector $\mathbf{u}^1$, the first discriminant factor, corresponds to the highest value of **λ**; the higher this value the higher will be the discriminant power of this factor. After obtaining the first discriminant characteristic $\mathbf{c}_1 = \mathbf{Xu}^1$, in a similar way can be obtained the discriminant characteristic $\mathbf{c}_2 = \mathbf{Xu}^2$, uncorrelated with the first one and so on. It appears clearly that eigenvectors corresponding to the matrix $\mathbf{V}^{-1}\mathbf{B}$ namely $\mathbf{u}^1$, $\mathbf{u}^2$, .., $\mathbf{u}^{k-1}$, ranked in decreasing order of the positive values $\lambda_1, \ldots, \lambda_2, \ldots, \lambda_{k-1}$, are successive solutions of the above matrix equation. If the vector of the discriminant function is $\mathbf{u} = (\mathbf{u}^1, \ldots, \mathbf{u}^2, \ldots, \mathbf{u}^p)$, then the projection of sample i on this axis represents the distance to the origin:

$$c_i = x_i^1u^1 + x_i^2u^2 + \ldots + x_i^p u^p. \tag{8}$$

The vector **u** is called the discriminant factor and the vector **c** represents the discriminant scores. The linear function described by eq. (8) is called discrimination function. Finally, we have to emphasize that even if the power of discrimination does not depend on standardization of data, generally standardized data are used.

The quality of discrimination and the selection of the most discriminant independent variables can be evaluated by applying different criteria. A F test (Wilks' lambda) is used to check if the discriminant model as a whole is significant; the larger the lambda, the more likely it is. In the same order λ^* statistic defined by the eq. (8) can be used.

$$\lambda^* = \frac{{}^T\mathbf{uWu}}{{}^T\mathbf{uVu}} (0 \leq \lambda < 1) \tag{9}$$

The smaller the value of λ^*, the more the model is discriminating.

Concerning the contribution of the independent variables to the discrimination of groups, this can be appreciated either by the assay of the classes homogeneity using statistic F like in the case of ANOVA/MANOVA method, or by using Wilks' lambda (λ) for each variable. Wilks' lambda is the standard statistic used to express the significance of the overall discriminatory power of the variables in the model. The value 1.0 indicates no discriminatory power, whereas 0 indicates a perfect one. The partial Wilks' lambda (λ^*) describes the unique contribution of each variable to the discriminatory power of the model. The closer the partial lambda is to 0, the better the discriminatory force of the variable is. In addition, the tolerance value gives information about redundancy of the respective variable in the model, and is computed as 1 minus R-squares of the respective variable, with all other variables included in the model. In other words, it is the proportion of the variance contributed by respective variable. If variable is completely redundant, the squared tolerance value approaches zero.

This kind of information can be obtained from value of the discriminant coefficients associated to the descriptive variables x_i, and also from the correlation coefficients between each variable x_i and the vector score. The higher the discriminant coefficient (absolute value) and the closer the correlation coefficient is to one, the more the variable importance for the samples separation in defined groups. Also, the standardized discriminant coefficients, like beta weights in regression, are used to asses the relative classifying importance of the independent variables.

3. Results and discussion

At the beginning it should be mentioned that the results obtained by classical PCA and the cluster analysis were unsatisfactory concerning, for example, the classification of samples by comparing with their origin and location of sampling sites.

The forward stepwise discriminant function analysis [15-17] was used to select the quantitative variables that enhance discrimination of the groups established by the dependent variable defining location sampling sites. The first computed data set included all 160 samples and 30 characteristics: F^-, Cl^-, Br^-, SO_4^{2-}, NO_3^-, PO_4^{3-}, NO_2^-, HCO_3^-, $HCOO^-$, Na^+, K^+, Ca^{2+}, Mg^{2+}, NH_4^+, pH, conductivity, phenols, HCHO, H_2O_2, TC, TIC, TOC, ambient temperature I, ambient temperature II, humidity, pressure, wind direction, wind speed, rainfall and volume. Table 3 shows the basic statistical parameters and illustrates a large variation in data. After application of the forward stepwise DA to the matrix data (160 × 30) the variables presented in Table 4 were retained in the model. Clearly the greatest observable contribution comes from pH (λ^* = 0.553; F = 21.36), followed by wind direction (λ^* = 0.652; F = 14.11), pressure (λ^* = 0.689; F = 11.89), Cl^- (λ^* = 0.752; F = 8.72), Na^+ (λ^* = 0.756; F = 8.48), and ambient temperature II (λ^* = 0.750; F = 8.77).

Interestingly, the contributions of some characteristics are quite similar: Mg^{2+}, SO_4^{2-}, TOC, and wind speed, for instance. The similar behavior of the variables indicates a high correlation and suggests the same origin/source. The smallest contributor is F^- (λ^* = 0.961; F = 1.06), followed by TIC, humidity, NH_4^+, HCO_3^-, and ambient temperature I. The values of tolerance (R^2) and 1 minus R-square (Table 4) represent the correlation of the given variable with all other variables included in the model. One can observe that the most redundant variables appear to be ambient temperature I (R^2 = 0.898) and wind direction (R^2 = 0.851), while the most informative variables seem to be Na^+ (R^2 = 0.003) and Mg^{2+} (R^2 = 0.005). The eigenvalues and the corresponding standardized canonical discriminant function coefficients are also shown in Table 4. The first function generated a relatively high eigenvalue of 6.767. The eigenvalue drops to 2.128 for the second axis, and further to 0.975 for the third axis. The variance calculated for the first two axes is 82.19%. The total explained variance expressed along the first four axes is 96.95%. The highest standardized discriminant coefficients correspond to Na^+ (6.399), pH (-3.041), Cl^- (-2.992), and SO_4^{2-} (2.090) in root 1; TC (-2.244), TOC (2.205), pH (-1.475), and Ca^{2+} (1.029) in root 2; and Na^+ (6.399) and Mg^{2+} (-4.696) in root 3; Na^+ and Mg^{2+} have also the most contribution along the root 4 and root 5. The two-dimensional scatter plot using the discriminant scores of the samples along root 1, root 2, root 3 and root 4, as can be seen in Fig. 1, indicates a satisfactory separation of samples according to their origin. The group of samples from Gdansk, for example, showed about 97% of well classified samples. Only one sample was erroneously included in the group from Mława. The poorest classification was obtained for samples from Gdynia.

A subsequent analysis was carried out using the groups of samples from Gdynia 1, Gdynia 2 and Sopot as the qualitative (dependent) variables and all characteristics already mentioned above as the quantitative (independent) variables. By applying a forward stepwise algorithm to the matrix (83 × 30), the variables presented in Table 5 were retained in the model. In this case the greatest contributor is pressure (λ^* = 0.542; F = 27.49), followed by Na^+ (λ^* = 0.789; F = 9.14) and wind speed (λ^* = 0.835; F = 6.39). Again, we observe very similar contributions from several characteristics: Mg^{2+} and Cl^-, and also wind direction and rainfall. The smallest contributors were humidity (λ^* = 0.978; F = 0.73), ambient temperature II (λ^* = 0.973; F = 0.91), and

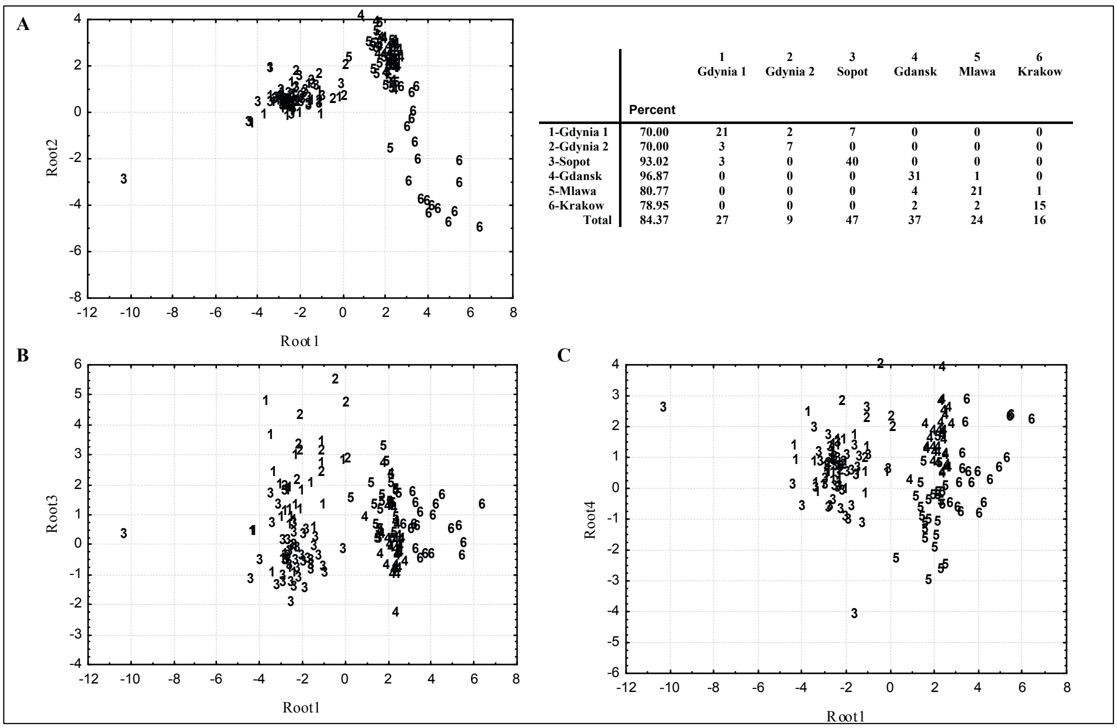
NO_2^- (λ^* = 0.968; F = 1.07). From Table 5 it is also evident that the most redundant variable is humidity (R^2 = 0.665) and the most informative variable is Na^+ (R^2 = 0.003). The results concerning the discriminant functions and the canonical roots are depicted in Table 5. The eigenvalue of the first axis is 1.609 and explains more than 83% of the variation in sample distribution along this axis. The eigenvalue drops to 0.322 for the second axis, and the variation explained is less than 17%.

Table 3. Basic statistical parameters of all characteristics (160 samples)

Variable	Mean	Median	Min.	Max.	Range	Std.Dev.	Skewness	Kurtosis
F^-	25.76	0.28	0.03	104.00	103.97	44.33	1.16	-0.65
$HCOO^-$	95.09	101.00	0.04	103.00	102.96	23.65	-3.69	11.83
Cl^-	21.99	5.80	0.07	103.00	102.93	34.79	1.77	1.38
NO_2^-	52.67	101.00	0.08	104.00	103.92	50.25	-0.05	-2.02
NO_3^-	23.68	6.24	0.41	103.00	102.59	36.09	1.64	0.89
PO_4^{3-}	44.82	3.67	0.14	103.00	102.86	49.41	0.27	-1.94
SO_4^{2-}	25.85	11.37	0.11	103.00	102.89	34.29	1.57	0.87
Br^-	97.36	101.00	0.11	102.00	101.89	19.21	-4.91	22.42
HCO_3^-	75.69	101.00	3.81	155.85	152.04	37.89	-0.43	-1.19
Na^+	31.21	6.06	0.31	102.00	101.69	42.62	1.02	-0.99
NH_4^+	70.76	101.00	0.53	103.00	102.47	46.07	-0.83	-1.34
K^+	30.05	4.36	0.37	102.00	101.63	43.03	1.06	-0.86
Mg^{2+}	28.38	2.63	0.28	102.00	101.72	43.94	1.08	-0.85
Ca^{2+}	37.95	17.08	1.71	102.00	100.29	39.69	0.85	-1.06
pH	20.12	6.64	4.90	102.00	97.10	33.28	2.05	2.23
HCHO	23.73	0.19	0.05	103.00	102.95	42.89	1.29	-0.35
Conductivity	157.27	112.65	5.84	881.00	875.16	124.93	2.58	9.83
TIC	30.49	14.67	0.75	102.00	101.25	36.49	1.36	0.07
TC	55.04	45.86	4.16	102.00	97.84	31.51	0.33	-1.24
TOC	45.31	34.47	0.21	102.00	101.79	34.02	0.73	-0.94
Phenols	36.03	0.42	0.01	103.00	102.99	48.81	0.63	-1.62
H_2O_2	50.88	8.70	0.20	104.00	103.80	50.07	0.02	-2.02
Ambient temperature I	12.61	12.25	3.00	20.00	98.00	7.62	9.89	115.20
Ambient temperature II	11.58	12.00	2.00	20.00	18.00	3.03	-0.26	0.30
Humidity	84.60	85.00	50.00	98.00	48.00	8.88	-0.76	0.61
Pressure	1015.7	1017.0	988.00	1032.00	44.00	9.21	-0.59	0.33
Wind direction	10.78	5.00	1.00	101.00	100.00	23.43	3.62	11.32
Wind speed	12.87	12.00	0.95	36.00	35.05	6.77	0.66	0.56
Rainfall	8.43	0.00	0.00	102.00	102.00	27.65	3.09	7.66
Volume	72.69	74.75	3.00	308.00	305.00	47.64	0.95	2.70

Table 4. Variables in the model (all samples) and the standardized canonical discriminant function coefficients

Variable	Wilks' λ	Partial λ*	F-remove	p-level	Tolerance	1-Tolerance	Root 1	Root 2	Root 3	Root 4	Root 5
Wind direction	0.015	0.652	14.11	0.0000	0.851	0.149	0.149	-0.753	-0.025	-0.100	-0.004
Wind speed	0.011	0.858	4.39	0.0009	0.585	0.415	-0.320	0.137	-0.465	-0.242	0.223
Ambient temperature I	0.010	0.956	1.20	0.3132	0.898	0.100	0.068	0.012	0.031	-0.301	0.189
Ambient temperature II	0.013	0.750	8.78	0.0000	0.728	0.272	-0.048	-0.012	-0.550	0.698	-0.126
Pressure	0.014	0.689	11.89	0.0000	0.469	0.533	-0.018	0.112	-1.111	-0.095	0.417
Humidity	0.010	0.939	1.71	0.1363	0.692	0.308	-0.125	0.116	-0.169	0.357	0.092
Rainfall	0.011	0.908	2.68	0.0241	0.676	0.324	-0.049	-0.020	-0.151	-0.072	0.698
Volume of sample	0.010	0.923	2.18	0.0596	0.546	0.454	0.068	-0.318	-0.209	0.272	-0.262
Conductivity	0.011	0.916	2.41	0.0394	0.416	0.584	-0.015	-0.508	-0.050	-0.247	0.058
pH	0.017	0.553	21.36	0.0000	0.044	0.956	-3.041	-1.475	-0.122	-0.169	1.559
F^-	0.010	0.961	1.06	0.3834	0.374	0.625	-0.103	-0.027	-0.099	-0.022	-0.597
Cl^-	0.013	0.752	8.72	0.0000	0.027	0.973	-2.992	-0.265	-0.598	-1.360	-1.307
SO_4^{2-}	0.011	0.847	4.76	0.0004	0.032	0.967	2.090	-0.210	-0.186	1.343	0.822
$HCOO^-$	0.011	0.917	2.37	0.0426	0.747	0.253	0.190	0.134	-0.351	-0.049	-0.138
HCO_3^-	0.010	0.930	1.98	0.0860	0.279	0.721	0.209	0.525	0.008	0.072	0.299
Ca^{2+}	0.011	0.906	2.74	0.0217	0.019	0.980	-1.920	1.029	-0.045	-0.554	-1.761
Mg^{2+}	0.011	0.846	4.81	0.0004	0.005	0.995	-0.644	0.343	-4.696	-6.676	-3.382
Na^+	0.013	0.757	8.48	0.0000	0.003	0.997	6.399	0.739	5.256	7.041	4.387
NH_4^+	0.010	0.939	1.71	0.1358	0.536	0.464	-0.182	-0.200	-0.075	-0.158	-0.426
TIC	0.010	0.959	1.13	0.3466	0.152	0.848	-0.237	-0.562	0.014	-0.119	0.065
TOC	0.011	0.861	4.27	0.0012	0.034	0.966	0.115	2.204	-0.021	1.269	0.812
TC	0.011	0.879	3.63	0.0041	0.031	0.969	0.106	-2.244	0.309	-0.713	-0.881
H_2O_2	0.011	0.901	2.91	0.0159	0.566	0.434	-0.198	0.067	0.408	0.021	-0.473
						Eigenvalue	6.767	2.128	0.975	0.622	0.330
						Cumulative Proportion	0.625	0.822	0.912	0.969	1.000



	Percent	1 Gdynia 1	2 Gdynia 2	3 Sopot	4 Gdansk	5 Mlawa	6 Krakow
1-Gdynia 1	70.00	21	2	7	0	0	0
2-Gdynia 2	70.00	3	7	0	0	0	0
3-Sopot	93.02	3	0	40	0	0	0
4-Gdansk	96.87	0	0	0	31	1	0
5-Mlawa	80.77	0	0	0	4	21	1
6-Krakow	78.95	0	0	0	2	2	15
Total	84.37	27	9	47	37	24	16

Figure 1. Scatterplot of canonical scores on the plan described by: A - root 1 and root 2; B - root 1 and root 3; C - root 1 and root 4 (all samples)

Table 5. Variables in the model (Gdynia 1, Gdynia 2 and Sopot) and the standardized canonical discriminant function coefficients

Variable	Wilks' λ	Partial λ*	F-remove	p-level	Tolerance	1-Tolerance	Root 1	Root 2
Wind direction	0.322	0.901	3.56	0.0340	0.597	0.403	-0.504	-0.191
Wind speed	0.347	0.836	6.39	0.0029	0.458	0.542	0.743	0.272
Ambient temperature II	0.298	0.973	0.91	0.4100	0.639	0.361	0.214	-0.241
Pressure	0.535	0.542	27.49	0.0000	0.359	0.641	1.407	0.470
Humidity	0.296	0.978	0.73	0.4873	0.665	0.335	0.155	0.273
Rainfall	0.325	0.892	3.92	0.0247	0.626	0.373	0.312	0.677
Volume of sample	0.305	0.951	1.67	0.1967	0.431	0.569	0.426	0.066
pH	0.302	0.959	1.37	0.2600	0.039	0.960	1.228	0.623
Cl^-	0.330	0.879	4.48	0.0149	0.016	0.983	2.823	-3.133
SO_4^{2-}	0.309	0.934	2.18	0.1206	0.019	0.981	-1.939	2.032
NO_2^-	0.299	0.968	1.07	0.3475	0.405	0.595	0.273	-0.368
Ca^{2+}	0.318	0.912	3.15	0.0494	0.013	0.987	2.874	-2.525
Mg^{2+}	0.334	0.868	4.95	0.0099	0.0045	0.995	6.492	-3.549
Na^+	0.371	0.780	9.14	0.0003	0.0026	0.997	-10.693	6.672
TOC	0.316	0.917	2.94	0.0598	0.292	0.707	-0.657	0.267
HCHO	0.320	0.907	3.34	0.0416	0.343	0.656	-0.472	-0.741
						Eigenvalue	1.609	0.322
						Cumulative Proportion	0.833	1.000

The highest discriminant coefficients correspond to Na^+ and Mg^{2+} in both roots (-10.693 and 6.492 in root 1 and 6.672 and -3.549 in root 2, correspondingly). The two-dimensional scatter plot using the discriminant scores of the samples along root 1 and root 2 presented in Fig. 2 indicates a highly satisfactory separation of samples with regard to their origin. It is interesting to mention that the separation of these samples is very similar to the separation obtained by computing all samples.

The third computed data set included samples collected in Gdansk, Mława and Krakow, and many more characteristics. The following metals were added to the characteristics mentioned above: Li, Be, B, Al, Ca, V, Cr, Fe, Mn, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo, Ag, Cd, Sn, Sb, Cs, Ba, Tl, Pb, Bi, U. Consequently, a 77 × 58 matrix was obtained. The variables retained in the model using the stepwise DA are presented in Table 6.

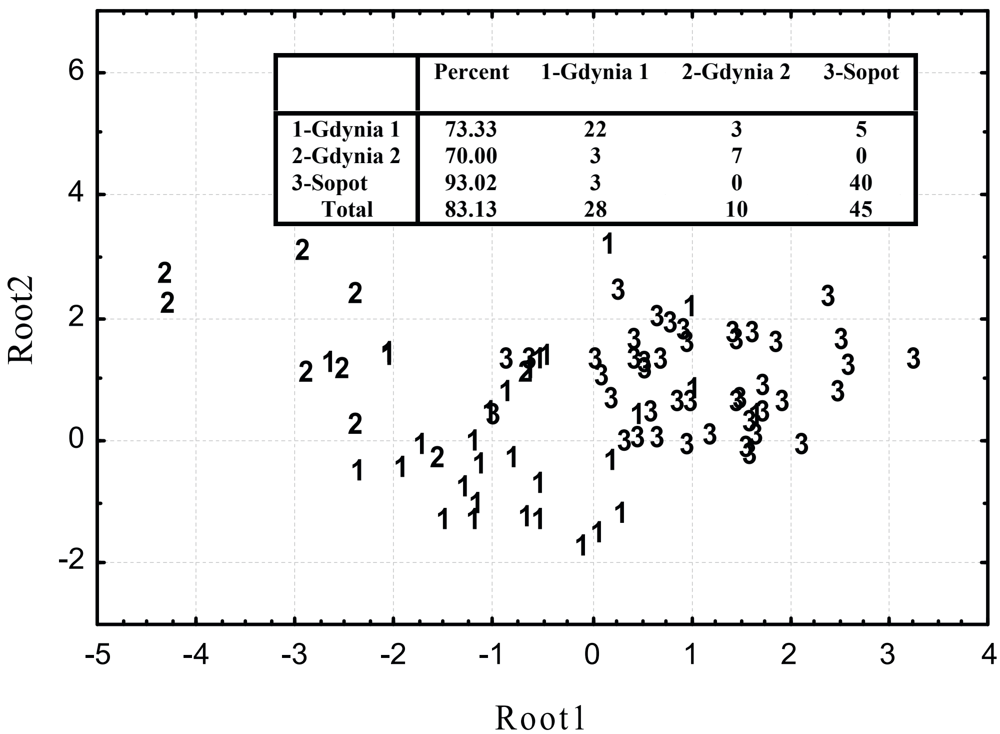


Figure 2. Scatterplot of canonical scores on the plan described by root 1 and root 2 (Gdynia 1, Gdynia 2 and Sopot)

The results obtained in this case indicate difference of the contribution to discrimination of different samples. The greatest contribution is from ambient temperature I (λ^* = 0.396; F = 39.62) and ambient temperature II (λ^* = 0.417; F = 36.39), followed by wind direction (λ^* = 0.841; F = 4.90), Ag (λ^* = 0.842; F = 4.89), humidity (λ^* = 0.847; F = 4.69), B (λ^* = 0.848; F = 4.64), and Bi (λ^* = 0.861; F = 4.18). Again, the contributions of some metals are similar (U, Be, Se, Sn, Ba and Al). The smallest contribution was obtained for Tl (λ^* = 0.974; F = 0.70). The eigenvalue of the first axis in this case is 6.815 and explains more than 57% of the variation in sample distribution along this axis.

Table 6. Variables in the model (Gdansk, Mława, Krakow) and the standardized canonical discriminant function coefficients

Variable	Wilks' λ	Partial λ^*	F-remove	p-level	Tolerance	1-Tolerance	Root 1	Root 2
Wind speed	0.023	0.941	1.62	0.2068	0.540	0.459	0.352	-0.025
Wind direction	0.025	0.841	4.90	0.0111	0.479	0.520	-0.597	-0.156
Ambient temperature I	0.054	0.396	39.62	0.0000	0.242	0.758	-0.751	1.738
Ambient temperature II	0.051	0.417	36.39	0.0000	0.194	0.806	0.589	-1.622
Humidity	0.025	0.847	4.69	0.0134	0.596	0.404	0.516	-0.170
Rainfall	0.024	0.897	3.22	0.0479	0.393	0.607	0.005	-0.580
Volume of sample	0.024	0.871	3.83	0.0279	0.449	0.550	-0.505	-0.276
pH	0.024	0.892	3.14	0.0517	0.542	0.458	-0.460	0.130
F^-	0.023	0.903	2.79	0.0705	0.609	0.391	-0.132	-0.416
Br^-	0.022	0.949	1.38	0.2590	0.647	0.353	-0.123	-0.279
NO_3^-	0.023	0.926	2.06	0.1376	0.681	0.319	0.048	-0.356
Tl	0.022	0.974	0.70	0.5029	0.669	0.331	0.208	-0.041
Ag	0.025	0.842	4.89	0.0113	0.727	0.273	0.499	-0.019
U	0.022	0.946	1.48	0.2372	0.501	0.499	0.223	0.277
Bi	0.025	0.861	4.18	0.0207	0.731	0.269	0.406	0.234
Be	0.022	0.952	1.29	0.2820	0.523	0.477	0.128	0.303
Fe	0.023	0.904	2.74	0.0735	0.471	0.529	-0.415	0.252
Se	0.023	0.941	1.64	0.2035	0.495	0.505	0.142	0.351
Sn	0.023	0.946	1.64	0.2038	0.553	0.447	-0.328	-0.126
B	0.025	0.848	4.64	0.0139	0.148	0.851	-1.081	0.053
Ba	0.023	0.933	1.88	0.1631	0.198	0.801	0.619	0.082
Al	0.022	0.952	1.30	0.2815	0.276	0.724	0.365	-0.260
H_2O_2	0.023	0.906	2.68	0.0776	0.557	0.443	0.417	0.139
						Eigenvalue	6.815	5.004
						Cumulative Proportion	0.577	1.000

The eigenvalue for the second axis is 5.004 and represents less than 43% (Table 5) of sample distribution variation. The highest discriminant coefficients are ambient temperature II (-0.751), Ba (0.619), wind direction (-0.597), ambient temperature I (0.589), humidity (0.516), volume (-0.505), Ag (0.499), pH (-0.460) in root 1, ambient temperature II (1.738), ambient temperature I (-1.622), rainfall (-0.580), F^- (-0.416), NO_3^- (-0.356), Se (0.351), and Br^- (-0.279) in root 2. A graphic representation of the two discriminant functions is shown in Fig. 3. The samples belonging to the group from Gdansk were well separated from samples originating in Mława and Krakow (100.00%). The graph clearly illustrates that some samples from Krakow and Mława appear as outliers. All the samples from Gdansk (100.00%) were well classified. Only one sample from Mława and two from Krakow seem to be erroneously classified.

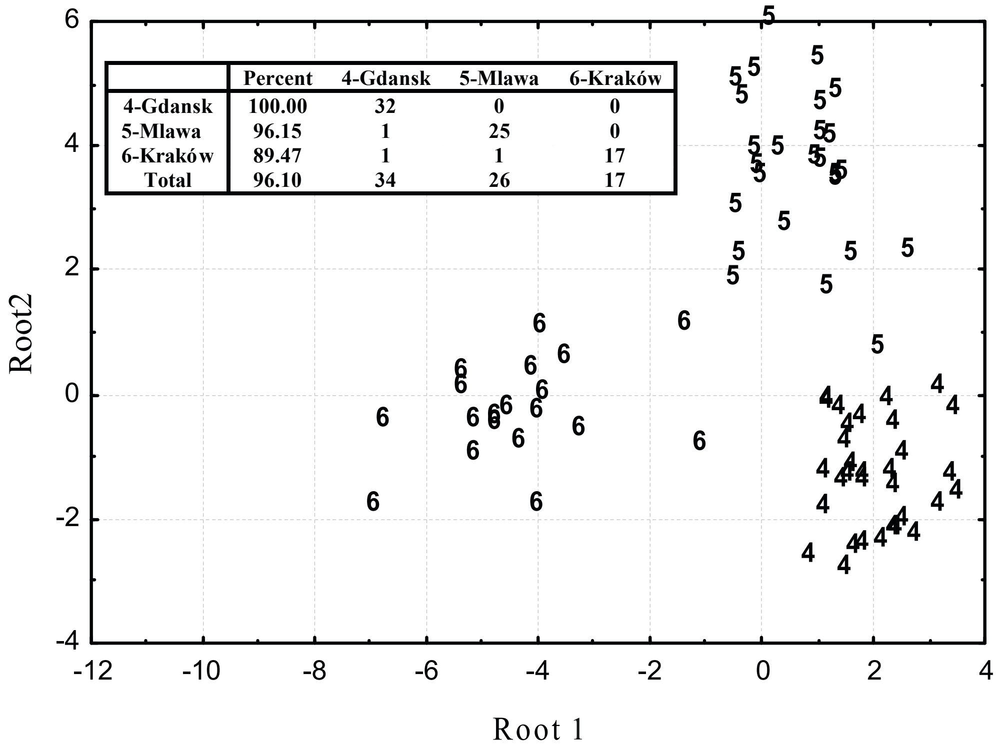


Figure 3. Scatterplot of canonical scores on the plan described by root 1

4. Conclusions

Dew samples collected in six different sites in Poland from August 2004 to October 2004 in Gdynia, Gdańsk and from August 2005 to October 2005 in Mława, Sopot and Krakow were analysed. The best classification results showed samples collected in Mława, Sopot and Krakow. The meteorological parameters (temperature, wind direction, and humidity) were found to be the most discriminant variables. The concentrations of Ag, B, and Bi also appeared to have significant effects on the differentiation of dew samples. It is interesting to note that the classification of all samples was dominated by pH, wind direction, pressure, and temperature, while Na^+ and Cl^- were significant contributors. Pressure, wind speed, and the concentrations of Na^+, Cl^-, and Mg^{2+} contributed highly to the separation of samples from Gdynia and Gdansk (location on the coast).

References

[1] S.P. Singh, P. Khare, K. Maharaj Kumari, S.S. Srivastava, Atmos. Res. 80, 239 (2006)

[2] P. Khare, S.P. Singh, G.S. Satsangi, K. Maharaj Kumari, A. Kumar, S.S. Srivastava, J. Atmos. Chem. 37, 231 (2000)

[3] M. Takeuchi, H. Okochi, M. Igawa, Dew water chemistry and its dominants factors. Acid Rain 2000, 6th International conference on acidic deposition, 10–16 December (Tsukuba, Japan, 2000) 65

[4] A. Jiries, Atmos. Res. 57, 261 (2001)

[5] M.A. Rubio, E. Lissi, G. Villena, Atmos. Environ. 36, 293 (2002)

[6] M. Muselli, D. Beysens, J. Marcillat, I. Milimouk, T. Nilsson, A. Louche, Atmos. Res. 64, 297 (2002)

[7] L.M. Cardenas, D.J. Brassington, B.J. Allan, H. Coe, B. Alicke, U. Platt, K.M. Wilson, J.M.C. Plane, S.A. Penkett, J. Atmos. Chem. 37, 53 (2000)

[8] B.F.J. Manly, Multivariate Statistical Methods (Chapman and Hall, London, 1986)

[9] D.L. Massart, E.B.G.M. Vandeginst, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics: a Textbook (Elsevier, Amsterdam, 1980)

[10] R.G. Brereton, Chemometrics: Applications of Mathematics and Statistics to the Laboratory (Ellis Horwood, Chichester, 1990)
[11] J. Einax, H. Zwanziger, S. GeiSS, Chemometrics in Environmental Analysis (John Wiley & Sons Ltd, Chichester, 1997)
[12] Thomas P.E. Auf der Heyde, J. Chem. Ed. 67, 461 (1990)
[13] S. Wold, K. Esbensen, P.Q. Geladi, Chem. Intell. Lab. Syst. 15, 37 (1987)
[14] C. Sârbu, H. Pop, Talanta 65, 1215 (2005)
[15] G. Scarponi, I. Moret, G. Capodaglio, P. Cescon, J. Agrlc. Food Chem. 30, 1135 (1982)
[16] J.F. Diáz-Flores, F. Diáz-Flores Estévez, C. Hernández Calzadilla, E.M. Rodríguez Rodríguez, C. Diáz Romero, L. Serra-Majem, Eur. J. Clin. Nutr. 58, 449 (2004)
[17] S. Mikkonen, K.E.J. Lehtinen, A. Hamed, J. Joutsensaari, M.C. Facchini, A. Laaksonen, Atmos. Chem. Phys. Discuss. 6, 8485 (2006)