Journal of Ecological Engineering
Volume 18, Issue 2, March 2017, pages 92–96
DOI: 10.12911/22998993/68141
Research Article

PRINCIPAL COMPONENT ANALYSIS AND CLUSTER ANALYSIS IN MULTIVARIATE ASSESSMENT OF WATER QUALITY

Jolanta Jankowska[1], Elżbieta Radzka[1], Katarzyna Rymuza[1]

[1] Department of Life Sciences, University of Natural Sciences and Humanities in Siedlce, Prusa 14, 08-110 Siedlce, Poland, e-mail: elzbieta.radzka@uph.edu.pl

Received: 2016.10.25
Accepted: 2016.12.28
Published: 2017.03.01

ABSTRACT
This paper deals with the use of multivariate methods in drinking water analysis. During a five-year project, from 2008 to 2012, selected chemical parameters in 11 water supply networks of the Siedlce County were studied. Throughout that period drinking water was of satisfactory quality, with only iron and manganese ions exceeding the limits (21 times and 12 times, respectively). In accordance with the results of cluster analysis, all water networks were put into three groups of different water quality. A high concentration of chlorides, sulphates, and manganese and a low concentration of copper and sodium was found in the water of Group 1 supply networks. The water in Group 2 had a high concentration of copper and sodium, and a low concentration of iron and sulphates. The water from Group 3 had a low concentration of chlorides and manganese, but a high concentration of fluorides. Using principal component analysis and cluster analysis, multivariate correlation between the studied parameters was determined, helping to put water supply networks into groups according to similar water quality.

Keywords: principal component analysis, water quality, cluster analysis

INTRODUCTION

Indispensable for life, water is a primary natural and renewable resource, undergoing a constant process in the hydrological cycle. The renewal of water resources can be slowed down or hindered due to anthropogenic factors [Chełmicki 2002]. In result, less water is available both for extraction and use. Poland has small resources of water, and, because of that, its rational use in accordance with legal regulations as well as sustainable and balanced development is crucial [Kundzewicz 2000]. Water supplied in adequate quantity and quality is essential for all life forms [Brzozowska and Gawęcki 2008]. That is why, water distribution companies are obliged to monitor chemical composition and properties of water for human consumption. Whether it comes from surface water or ground water, it should contain minerals, and its chemical, physical, and microbiological composition should meet Polish standards, which is why quality of drinking water is often tested [Blicharska et al. 2010, Kot 2001, Kot et al. 2000, Kregiel et al. 2011, Macioszczyk 2011, Soylak et al. 2002, Wichrowska et al. 2001]. In Poland Regulation of the Minister of Health of 29 March 2007 sets quality standards for water intended for human consumption. This regulation complies with the Drinking Water Directive of the European Union Council (Council Directive 98/83/EC of 3 November 1998) on quality of water intended for human consumption), laying down limits for hazardous substances and setting standards for colour, taste, odour, bacteria count, TOC, (Total Organic Carbon), and turbidity [Granops and Kaleta 2002, Gromiec 2004]. Evaluation of the quality of water for human consumption requires determination of physical, chemical and biological parameters, with all of them usually measured separately [Radzka et al. 2013]. However, multivariate methods allow determination of all those parameters together, and this way it is possible to establish multiple correlations between all parameter, and to put them into groups of similar

features [Boyacioglu 2006, Boyacioglu 2008, Rymuza and Radzka 2013].

The aim of this paper is to assess chemical properties of drinking water in the Siedlce County, using multivariate analysis.

MATERIALS AND METHODS

The District Sanitary Epidemiological Station in Siedlce provided data on drinking water quality in the Siedlce County. Altogether, between 2008 and 2012, 11 supply networks were studied to determine chemical properties of water. Water samples were taken four times a year. The following chemical compounds and elements were tested: nitrates (III), nitrates (V), chlorides, fluorides, sulphates, copper, sodium, iron, and manganese.

The presence of nitrates (III) and nitrates (V) in water is a consequence of using mineral fertilisers in agriculture, discharging sewage to watercourses, and natural mineralisation of organic substances (nitrification). Chlorides present in water can be of natural origin or they have been introduced together with contaminants. In drinking water they most often come from natural sources, industrial and municipal sewage, but also from rainwater.

The concentration of nitrates (V), chlorides, fluorides, sulphates, copper, and sodium are measured in mg/l, while iron and manganese in µg/l. The studies were done in accordance with the above mentioned *Regulation* of the *Minister of Health* of *29 March 2007* on quality of water intended for human consumption.

The results were statistically processed in three stages. In the first stage, means but also minimum and maximum values of chemical parameters were calculated. Additionally, coefficient of variance and the number of failures to meet standards were measured. Then standardization of variables was done because different units of measurements had been used. Principal component analysis was applied to determine maximum variation in the dataset. Using the Kaiser Criterion, the number of principal components to be analysed was chosen, all with the eigenvalue higher than 1 [Czernyszewicz 2008, Stanisz 2007]. During the last, third stage, data clustering was used to put water supply systems into groups according to the quality of drinking water. Taking into account the Euclidian distance, the grouping was done using the Ward method. The Caliński Harabsz index was used to determine the cut level of the dendrogram [Walesiak and Dudek 2006]. To check whether the grouping done with the clustering method was correct, another grouping was done with the k-means clustering. The groups were the same for each method.

RESULTS AND DISCUSSION

Tap water should meet quality standards. Using multivariate methods it is possible to estimate many parameters of water quality together. It is possible to check multivariate correlations between parameters, and to group items of similar parameters. In all the water supply systems, water comes from groundwater, the main source of providing it. To assess water, minimum, maximum, and median values of those parameters are used. In Table 1 median, minimum, and maximum values chemical parameters of drinking water in the Siedlce Commune are presented. The highest value of coefficient of variation was for iron (V = 280.08%) and manganese (V = 90.75%). The lowest values were for copper (V = 0.07%) and fluorides (V = 0.10%). Maximum values of some parameters indicate that they exceeded the limits. Iron excess was found in 21 water supply systems, while manganese in 11 (Table 1). Con-

Table 1. Mean, minimum, maximum, and permissible values of some chemicals in the water supply networks

Chemical parameter	$\bar{x}$	Coefficient of variation (V) [%]	Min.	Max.	Permissible standard	Number of failures to meet standards
Nitrates V	4.47	4.41	1.80	98.7	50	-
Chlorides	10.04	9.38	5.00	93.49	250	-
Fluorides	0.17	0.10	0.04	58.72	1.5	-
Sulphates	19.88	18.04	2.50	90.75	250	-
Cooper	0.39	0.07	0.00	18.46	2.0	-
Sodium	34.17	13.43	0.04	39.31	200	-
Iron	210.43	280.08	7.40	133.10	200	21
Manganese	50.40	90.75	0.02	180.06	50	12

centration of iron in water ranged from 7.40 to 133.10 units, and manganese from 0.02 to 180.06 units. Analyses of principal components showed that the first three of them, with eigenvalues higher than 1, explained over 85% part of multivariate variability of accessions, that is multivariate variability of concentration of chemical elements in analysed water supply systems (Table 2).

With the first principal component, concentration of chlorides (r = -0.973), sulphates (r = -0.913), and manganese (r = -0.896) was negatively correlated in the strongest way. Those parameters have the highest effect on multivariate variability of water quality. Those mutual correlations mean that in those networks where water contained a high amount of chlorides and sulphates, it also contained a relatively high amount of nitrates V, iron, and manganese (Table 3). Iron and manganese give water a yellow hue, and they both stimulate development of bacteria causing musty and earthy taste and odour [14]. The second principal component explaining over 18% of multivariate variability of chemical element concentration was negatively correlated with the concentration of copper (-0.841) and sodium (-0.786). In the supply networks where water contained more copper, it also contained more sodium. The third principal component only explained 14.54% of multivariate variability and was positively correlated with the concentration of fluorides (0.800) and iron (0.645).

Table 2. Eigenvalues, variance percentage, and cumulated variance percentage of components obtained

Principal components	Eigenvalues	Explained part of multivariate variability of accessions [%]	Cumulative part of multivariate variability [%]
PC 1	4.19	52.40	52.4
PC 2	1.46	18.27	70.7
PC 3	1.16	14.54	85.2
PC 4	0.68	8.50	93.7
PC 5	0.36	4.47	98.2
PC 6	0.10	1.30	99.5
PC 7	0.03	0.43	99.9
PC 8	0.01	0.09	100.0

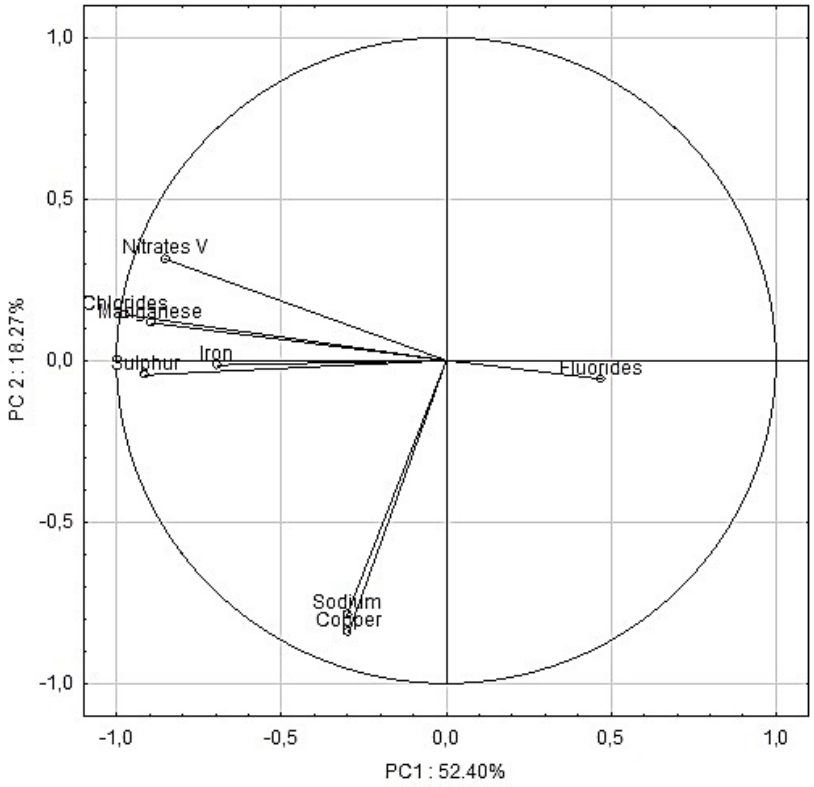

Figure 1. Distribution of water quality parameters in the first and second principal component space

Table 3. Factor loads depicting impact of water quality parameters on three main components

Parameter	PC 1	PC 2	PC 3
Nitrates V	-0.853	0.314	-0.181
Chlorides	-0.973	0.141	0.024
Fluorides	0.471	-0.057	0.800
Sulphur	-0.913	-0.045	0.147
Copper	-0.296	-0.841	0.063
Sodium	-0.301	-0.786	-0.219
Iron	-0.694	-0.014	0.645
Manganese	-0.896	0.116	-0.030

Table 4. Mean values of chemical parameters content in designated groups of water supply

Chemical parameters	Group 1	Group 2	Group 3
Nitrates V	2.6	4.7	10.8
Chlorides	5.7	9.9	43.0
Fluorides	0.2	0.2	0.1
Sulfur	8.8	22.5	59.0
Cooper	0.4	0.4	0.4
Sodium	40.0	32.0	40.0
Iron	77.2	252.9	502.8
Manganese	26.7	34.1	252.0

Cluster analysis resulted in putting water supply networks into three groups with different water quality. Group 1 constituted networks in the following places: Czerniejew, Kotuń, Radomyśl, Siedlce, with Group II networks in: Dąbrowa, Korczew, Domanice, Jesionka, Paprotnia, Krynica. The water from all the latter networks was of similar quality but different from Group 1 and 3. The network supplying water in Seroczyn constituted Group 3 (Fig. 2).

The water from Group 1 had the lowest content of fluorides and copper. Group 2, with six networks, supplied water with a high concentration of iron, while the concentration of fluorides and copper was similar to that of Group 1. Group 3, with one network in Seroczyn, provided water with a high concentration of sulphates, manganese, and iron, much higher than the average concentration from all networks in the Siedlce County.

CONCLUSIONS

1. Between 2008 and 2012 water in the Siedlce County was of quite good quality. However, iron content and manganese content exceeded the limits (21 times and 12 times, respectively).
2. The networks supplying water were combined into three groups according to water quality. The water in Group 1 had a high content of iron, sodium, and manganese. Group 2 networks supplied water of low content of fluorides and copper. The highest content of iron, manganese and sulphates had water form Group 3.
3. The techniques used to classify water were principal component analysis and cluster analysis. They made it possible to examine correlations among the variables and allowed grouping networks according to drinking water quality.

REFERENCES

1. Blicharska E., Komsta R., Kocjan R., Gumieniczek A., Wiśniewska A. 2010. Chemometric processing of iron chromatograms application to comparative analysis of Polish bottled mineral and spring waters. Pol. J. Environ. Stud., 19(5), 1071–1075.
2. Boyacioglu H. 2006. Surface water quality assessment using factor analysis. Water SA, 32(3), 389–393.
3. Boyacioglu H. 2008. Water pollution sources assesment by multivariate statistical methods in the Tahtali Basin. Turkey, Evironmental Geology, 54(2), 275–282.
4. Brzozowska A., Gawęcki J. 2008. Woda w żywieniu i jej źródła. Wyd. U.P. Poznań.
5. Chełmicki W. 2002. Woda, zasoby, degradacja, ochrona. PWN. Warszawa.
6. Czernyszewicz E. 2008. Zastosowanie analizy składowych głównych do opisu konsumenckiej struktury jakości jabłek. Żywność. Nauka. Technologia Jakość. 2(57), 119–127.
7. Granops M., Kaleta J. 2002. Woda – uzdatnianie i

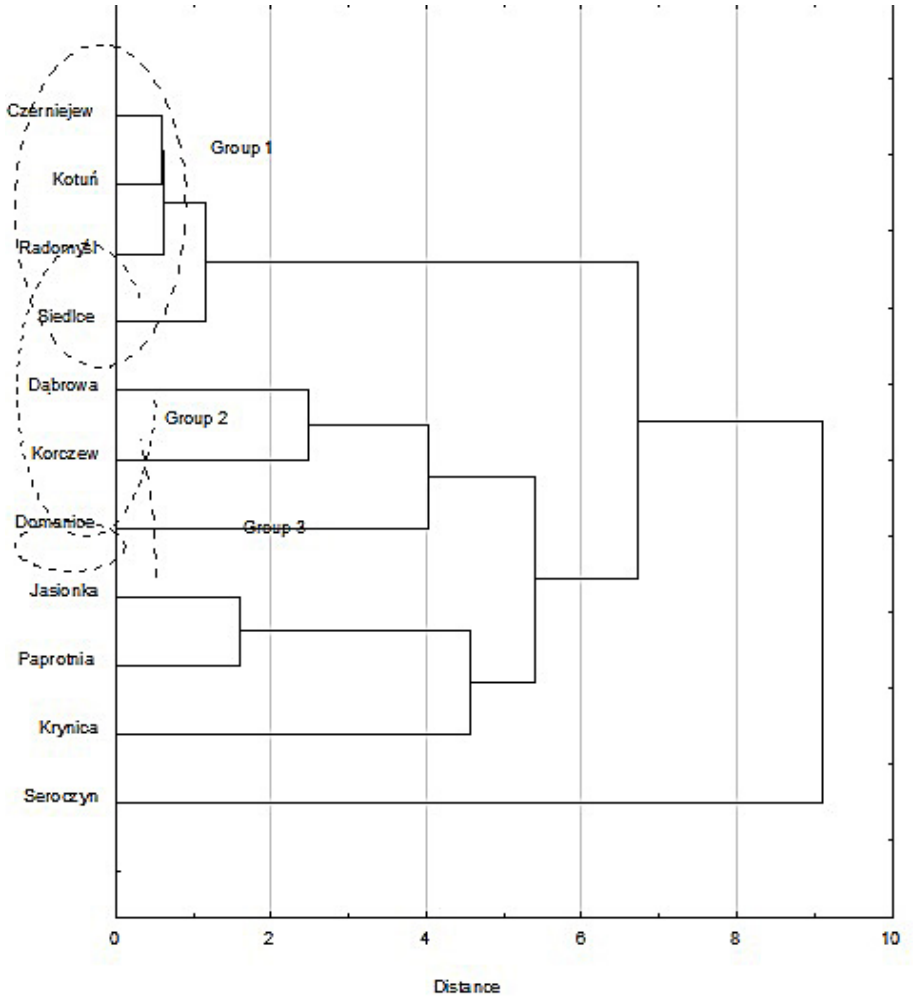

Figure 2. Dendrogram representing groups of water supply systems determined by using cluster analysis

odnowa. Wyd. SGGW, Warszawa.

8. Gromiec M. 2004. Postanowienia traktatu akcesyjnego i zobowiązania wynikające z przystąpienia Polski do UE w zakresie jakości zasobów. Gospodarka Wodna, 4, 129–132.
9. Kot A. 2001. Determination of content of zinc and cooper in table and therapeutic mineral water. Przegl. Lek., 58(7), 14–17.
10. Kot B., Baranowski R., Rybak A. 2000. Analysis of mine waters using X-ray fluoroscence spectrometry. Pol. J. Environ. Stud., 9(5), 429–431.
11. Kregiel D., Rygała A., Libudzisz Z. 2011. Bakterie z rodzaju Asaia – nowe zanieczyszczenie smakowych wód mineralnych. Żywność. Nauka. Technologia. Jakość, 2(75), 5–16.
12. Kundzewicz Z. 2000. Zasoby wodne dla trwałego rozwoju. PWN, Warszawa.
13. Macioszczyk A. (Ed.) 2011. Podstawy hydrogeologii stosowanej. Wyd. Nauk. PWN, Warszawa.
14. Michalik A. 2008. The use of chemical and cluster analysis for studying spring water quality in Świętokrzyski National Park. Pol. J. Environ. Stud., 17(3), 357–362.
15. Radzka E., Jankowska J., Rak J. 2013. Ocena jakości wody pitnej w powiecie mińskim. Gospodarka Wodna, 7, 284–288.
16. Rymuza K., Radzka E. 2013. Zastosowanie analiz wielowymiarowych do oceny jakości wody pitnej. Żywność. Nauka. Technologia. Jakość, 6(91), 165–174.
17. Soylak M., Armagan Aydin F., Saracoglu S., Elci L., Dogan M. 2002. Chemical analysis of drinking water samples from Yozgat, Turkey. Pol. J. Environ. Stud., 11(2), 151–156.
18. Stanisz A. 2007. Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. T. 3 Analizy wielowymiarowe. StatSoft. Kraków.
19. Walesiak M., Dudek A. 2006. Symulacyjna optymalizacja wyboru procedury klasyfikacyjnej dla danego typu danych – charakterystyka problemu. Zesz. Nauk. Uniwersytetu Szczecińskiego. 450, 634–646.
20. Wichrowska B., Kozłowski J., Jankowska D. 2001. Ocena ryzyka zdrowotnego w świetle przepisów Unii Europejskiej dotyczącej jakości wody do picia. Ochrona Środowiska, 83(4), 19–22.