

Katarzyna PIETRUCHA-URBANIK¹
Anna NOGAJ²
Justyna STECKO³

ASSESSMENT OF WATER SUPPLY INFRASTRUCTURE STATE IN RURAL AREAS IN POLAND USING MULTIVARIATE ANALYSIS

In the study the condition of the water supply infrastructure in rural areas in Poland was assessed. Research period covered the years 2003 and 2014. Through this period, the water supply infrastructure state was considerably improved. The improvement was mainly due to Polish accession to the European Union, therefore, the proper standards of water supply should be ensured, in order to adapt them to of the EU rights. The European community has supported Poland through all kinds of subsidies thanks to which water supply was expanded and a procedure of improvement was implemented. For performing research the multivariate comparative analysis – the agglomeration method and the k-means method - were used. The analysis covered the following parameters characterizing the development of the water supply infrastructure in rural areas: changes in length of water supply network, number of water supply connections, number of water treatment plants. In the last decade, the percentage of the population using water supply network increased by 13%, with biggest increase in Masovian (up to 32%) and in Lesser Poland (up to 26%) and the lowest was observed in the Opole province (up to 6%). In contrast, the costs associated with the development of water supply infrastructure in rural areas increased by 33%. The largest costs investment were incurred in the Greater Poland Region (380%) and in the Silesian province (260%). Increasing the development level of water supply infrastructure and consequently, the sewage infrastructure development will influence the tourism and investment attractiveness of rural areas.

Keywords: water supply infrastructure, rural areas, multivariate analysis.

1. INTRODUCTION

The state of water supply infrastructure indicates the standard of living of people in the given area. Its development ensures an increase in the standard of living, as well as improving the environment. The water infrastructure state determines the form and the

¹Katarzyna Pietrucha-Urbanik, PhD, Eng., Politechnika Rzeszowska, Wydział Budownictwa, Inżynierii Środowiska i Architektury, al. Powstańców Warszawy 6, 35-959 Rzeszów, autor korespondencyjny: e-mail: kpiet@prz.edu.pl

²Anna Nogaj, MSc, Eng., Politechnika Rzeszowska, Wydział Budownictwa, Inżynierii Środowiska i Architektury, al. Powstańców Warszawy 6, 35-959 Rzeszów

³Justyna Stecko, PhD, Politechnika Rzeszowska, Wydział Zarządzania, al. Powstańców Warszawy 10, 35-959 Rzeszów

level of all kinds of economic activities, also decides about increasing or decreasing the area attractiveness, both in terms of investment and tourism.

Rural localities in Poland are significantly differentiated, next to the poorer and less developed areas there are also the richer ones which grow quite quickly. The state aims to reduce these differences as much as possible. It is not easy because it requires high financial expenditures. After 2004, that is after the accession of Poland to the European Union, Poland began to receive different types of subsidies to support the development of different activities, among others, rural development^{4,5,6}. Also analysis proposed in this work can be helpful in decision making about possible renewal of water network, also in terms of costs incurred for failure removal^{7,8,9}.

For this reason, the assessment of the water supply infrastructure state in rural areas in particular Polish provinces, was presented. The scope of the work covers the analysis of changes in the water supply network length, costs incurred for the construction or extension of the water supply network, water treatment plants, changes in the number of water supply connections, in rural areas in Poland in the years 2003 and 2014.

As to assess the water supply infrastructure state the multivariate analysis was used which allows to describe complex phenomena in an intelligible manner. Description of the multivariate analysis methods was presented in many publications¹⁰.

Multivariate analysis was performed through the cluster analysis method and the k-means method. Calculations were performed using Microsoft Excel and Statistica 12.5 GB. Data for the analysis were taken from the statistical yearbooks of CSO.

2. DESCRIPTION OF THE RESEARCH METHOD AND SELECTION OF DIAGNOSTIC CHARACTERISTICS

The cluster analysis involves data segregation so that the variables from the same group were the most related among themselves, wherein their relation with data from other groups was minimum.

Grouping of these variables is carried out in order to reduce the large number of data to the basic groups, which in later analysis are treated as variables. Then a classification of

⁴Gorczyca M., Regionalne dysparytety rozwoju infrastruktury wodociągowej (Regional disparities of water supply infrastructure development). „Gaz, Woda i Technika Sanitarna”, 2011, No 5, pp. 166-170;

⁵Kwietniewski M., Rak J., Niezawodność infrastruktury wodociągowej i kanalizacyjnej w Polsce (Reliability of water and sewerage infrastructure in Poland). PAN, Warszawa 2010;

⁶Mays W.L., Water Distribution Systems. Handbook. McGraw-Hill. New York 1999;

⁷Kleiner Y., Adams B., Rogers J., Water distribution network renewal planning, Journal of Computing in Civil Engineering, 2001, No 1, pp. 15-26;

⁸Pietrucha-Urbanik K., Studziński A., Wybrane aspekty kosztów oraz awaryjności przewodów przykładowego systemu wodociągowego (Selected aspects of costs and failure of pipes in exemplary water supply system), Rocznik Ochrona Środowiska, 2016, No 2, pp. 623-634;

⁹Studziński A., Pietrucha-Urbanik K., Preventive maintenance and reliability of water supply system elements. „Czasopismo Inżynierii Łądowej, Środowiska i Architektury - Journal of Civil Engineering, Environment and Architecture, JCEEA”, 2015, No 3/I, pp. 429-436. DOI: 10.7862/rb.2015.126;

¹⁰Marek T., Analiza skupień w badaniach empirycznych (Cluster analysis in empirical research). PWN, Warszawa 1989;

typical objects, the comparison of the events or objects with many features and the distinction of the main features of the examined objects or events, are made¹¹.

The final step is to link all the variables into one group and the result of applying the agglomeration method is a hierarchical tree.

The selection of appropriate method for measuring the distance between variables, which determines the differentiation, that is the degree of similarity or dissimilarity of the elements, is the basis for the division in the cluster analysis^{12,13}.

In the study the Euclidean distance was used. In this way the Euclidean distance matrix was obtained which is described by the numerical values of the distance, the value closer to zero the similarity of the examined characteristics is higher and if the numerical value is higher the similarity is lower¹⁴.

Before measuring the distance between variables the data standardization was made which allows to their unification due to the fact that they are expressed in different units, allowing for the elimination of differences in scale¹⁵.

In the case of the non-hierarchical method of k-means agglomeration the division of objects is obtained, where no cluster is a part of another, with the previously assumed number of clusters¹⁶.

As a result of grouping, the clusters in which objects are maximally similar to each other are obtained, while objects in different clusters differ from each other significantly¹⁷.

The analysis covered the parameters characterizing the development of water supply infrastructure:

- the number of water supply connections per km of water pipe, no/km, (v_1),
- the length of the water supply network per one rural locality, km/locality, (v_2),
- investment for the water network construction per one rural locality, thousand PLN/locality, (v_3),
- the number of water supply connections, no (v_4),
- the number of water supply connections per one rural locality, no/locality, (v_5),
- the number of water treatment plants, no (v_6),
- percent of the population using water supply system, %, (v_7).

¹¹Jajuga K., *Statystyczna analiza wielowymiarowa (Statistical multivariate analysis)*. PWN, Warszawa 1993;

¹² Migut G., *Zastosowanie technik analizy skupień i drzew decyzyjnych do segmentacji rynku (Application of cluster analysis and decision trees for of for market segmentation)*, StaSoft Polska 2009;

¹³ Ward J.H., Hierarchical grouping to optimize an objective function, „*Journal of the American Statistical Association*”, 1963, No 3, pp. 236-244;

¹⁴Jajuga K., *Statystyczna analiza wielowymiarowa (Statistical multivariate analysis)*. PWN, Warszawa 1993;

¹⁵Pluta W., *Wielowymiarowa analiza porównawcza w badaniach ekonomicznych (Multidimensional comparative analysis in economic research)*. PWN, Warszawa 1977;

¹⁶Jajuga K., Sokołowski A., Bock H.H., *Classification, clustering and data analysis: recent advances and applications*. Springer, Berlin 2002;

¹⁷Everitt B.S., Landau S., Leese M., Stahl D., *Cluster Analysis*, John Wiley & Sons, Ltd, Chichester 2011;

Data for the analysis were obtained from the Central Statistical Office. Variables were selected in such a way that the increase of their values indicated the change in the complex data. All analyses were conducted in the programs Statistica 12.5 PL and Microsoft Excel.

3. RESULTS OF MULTIVARIATE ANALYSIS OF VARIABLES CHARACTERIZING THE WATER SUPPLY INFRASTRUCTURE

Descriptive statistics characterizing the water supply infrastructure and so their arithmetic means, standard deviations and variation coefficients are presented in Table 1.

Table 1. Selected descriptive statistics of variables describing the state of water supply infrastructure.

Variable symbol	Average		Standard deviation		Variation coefficient	
	2003	2014	2003	2014	2003	2014
v ₁	14,5	15,2	3,8	4,0	0,259	0,27
v ₂	3,9	4,9	2,3	2,8	0,584	0,57
v ₃	13,2	18,1	9,2	11,0	0,698	0,61
v ₄	164720,6	218502,8	76727,3	110190,7	0,466	0,50
v ₅	3,5	21,8	5,6	16,0	1,599	0,73
v ₆	455,9	436,4	310,4	265,8	0,681	0,61
v ₇	73,8	86,2	9,6	8,3	0,129	0,10

In 2003, three clusters were identified, with the critical value of 0,85, according to¹⁸. The first cluster consists of the Kuyavian-Pomeranian Voivodeship. Most indicators remain at the average level, except for the investment for the water network construction per one rural locality (v₃). This value is the highest in the country (41.83 thousand of PLN/locality).

The second cluster consists of the Lower Silesia Voivodeship. The values of the indicators generally remain at above average level. It is worth to notice that the number of water supply connections per one rural locality, 19 no/locality, is the highest in Poland.

The third cluster is composed of the Voivodeships: Opole, Greater Poland, Lodz, Silesia, WarmianMasurian, West Pomeranian, Pomeranian, Podlaskie, Lubusz, Świętokrzyskie, Subcarpathian, Lesser Poland, Mazovian and Lublin. The mean values for each indicator remained mostly at below average level. Above average ranked the following indicators: the number of water supply connections per kilometre of water pipe, the number of water supply connections, as well as the number of water treatment plants. The average values of these indicators are, respectively: 14,9 no/km, 167211,3 no, 2,57 no/locality.

As a result of the analysis for 2014 three clusters were distinguished, based on the determined critical value of 0,86. The composition of the individual clusters differed in comparison to 2003. In the first cluster was found the Masovian Voivodeship. Only the indicators: the number of connections per kilometre of water pipe and the percentage of

¹⁸Hellwig Z., Zarys ekonometrii (Outline of econometrics). PWE, Warszawa 1973;

the population using the water, had the values below the average calculated for the whole country, respectively, 13 no/locality and 69,7%. In case of other indicators, their values exceeded the average.

The second cluster included Lesser Poland and Subcarpathian. The values of three indicators, the length of the water network per one rural locality, a number of water treatment plants and the percentage of the population using the water supply system, were lower than the average calculated for the whole country, the average values of those indicators were, respectively, 4,71 km/locality, 206,5 no and 61,3%. The rest of the indicators reached above average values. It is worth to mention that the value of the indicator defining the investment for the construction of water falling per one rural locality in the Subcarpathian was the highest in Poland and amounted 42,85 thousand PLN/locality.

The third cluster was formed from the following provinces: Podlaskie, Lublin, Lodz, Opole, West Pomeranian, Pomeranian, Lubusz, Warmian-Masurian, Kuyavian Pomeranian, Greater Poland, Silesia, Świętokrzyskie and Lower Silesia. The values of only two indicators, the number of water treatment plants and the percentage of the population using water supply system were higher than the average value calculated for the whole country, respectively, 455 no/locality and 80,38%. The values of other indicators were lower than the average.

Similar results were obtained using the k-means method, Polish provinces again were divided into three clusters in 2014, however, in 2003 different composition of the three clusters was obtained, as shown in Table 2.

Provinces from the first cluster were characterized in 2003 by the highest average values of the indicators v_1 , v_2 , v_3 and v_5 . The number of water supply connections per kilometre of water pipeline is 18 no/km, the length of water supply network per one rural locality (6,15 km/locality). The average values of v_6 and v_7 (the number of water treatment plants - 179 and the percentage of the population using the tap water - 67,68%) were lower than the values for the other clusters. The state of water supply system in this group of provinces is better than in other regions of the country, however, still many rural residents do not use tap water.

Provinces from the second cluster are characterized by the lowest average values of coefficients describing the number of water supply connections per kilometre of water pipeline (13 no/km), the length of water supply network per one rural locality (2,41 km/locality), number of water supply connections (87285). After analysing the values of these indicators it was found that they are lower than their average calculated for the whole country. However, it should be noted that the average value of the indicator of population percentage using tap water (78,82%) was the highest for this cluster. It can be stated, that the second cluster is characterized by worse level of water supply infrastructure, in spite of it more rural residents use tap water.

Table 2. Average values of the coefficients calculated for different clusters.

Cluster	Variables						
	v ₁	v ₂	v ₃	v ₄	v ₅	v ₆	v ₇
2003							
1 st cluster: Lower Silesian, Lesser Poland, Subcarpathian, Silesian, Świętokrzyskie	0,707	0,742	0,287	0,473	0,333	0,106	0,412
2 nd cluster: Kuyavian-Pomeranian, Lubusz, Opole, Podlaskie, pomorskie, Warmian Masovian, West Pomeranian	0,323	0,201	0,231	0,151	0,008	0,446	0,721
3 rd cluster: Lublin, Lodz, Masovian, Greater Poland	0,401	0,644	0,087	0,765	0,125	0,629	0,556
2014							
1 st cluster: Lesser Poland, Sycarpathian	0,777	0,555	0,884	0,445	0,500	0,121	0,290
2 nd cluster: podlaskie, Lublin, łódzkie, Opole, West Pomeranian, pomorskie, Lubusz, Warmian Masovian, KuyavianPomeranian, Greater Poland, Silesian, Świętokrzyskie, Lower Silesian	0,299	0,219	0,150	0,164	0,103	0,464	0,778
3 rd cluster: Masovian	0,435	0,848	0,271	0,552	0,676	0,530	0,789

Provinces from the third cluster have the highest average values of coefficients v₄ and v₆ (respectively: 260122 and 696) and the lowest investment costs for the construction of water supply network (for one rural locality about 8,63 thousand PLN). It can be concluded that the third cluster consists of the provinces in which water supply infrastructure is in good condition and more people than in the first cluster use tap water, however, the authorities of these rural areas spend much less money on construction and modernization of water supply system in comparison with the rest of the country.

Changes in the values of mentioned variables are shown in Table 3. Negative numbers represent a decrease of the value, while the positive the growth.

Table 3. Changes in the value of variables describing the state of water supply infrastructure for rural areas of particular provinces in 2003 and 2014.

Provinces	Variables						
	v ₁	v ₂	v ₃	v ₄	v ₅	v ₆	v ₇
Lower Silesian	0,65	1,91	-1,08	48579	35	9	11,8
KuyavianPomeranian	0,87	0,88	-33,16	45820	11	2	13,2
Lublin	-0,18	0,84	-0,69	61410	11	-43	13,9
Lubusz	0,17	0,87	-2,81	22665	10	16	14,0
Lodz	2,24	0,87	1,29	70400	20	-46	13,7
Lesser Poland	0,62	0,86	18,25	76706	19	8	13,9
Masovian	0,58	3,06	9,97	167284	44	76	20,2
Opole	-0,68	0,39	13,28	14703	5	-5	6,0
Subcarpathian	0,65	0,79	24,98	34640	22	27	9,3
Podlaskie	-0,55	0,31	1,67	22558	3	11	8,6
Pomeranian	0,56	0,52	4,88	46143	7	-26	11,2
Silesian	2,27	0,99	20,10	52419	35	47	10,2
Swietokrzyskie	0,66	1,98	-3,40	41889	35	14	16,5
Warmian Masovian	-0,58	1,54	2,39	34177	9	-111	15,1
Greater Poland	1,51	1,06	17,31	90595	20	-96	9,7
West Pomeranian	1,20	0,60	5,51	30527	7	-194	9,5

The detailed analysis showed:

- in the rural areas in Masovian in the years 2003-2014 the most water supply connections were made (167 284) and the average increase in the length of the water supply network was the highest (915,08 km·year⁻¹) in relation to rural areas in other provinces. It should be mentioned that in 2014 this province incurred the biggest costs associated with the development of water supply infrastructure in rural areas (194 436 thousand PLN),
- the smallest average increase of the water supply network length in the years 2003 and 2014 was observed in the Opole province (80,72 km·year⁻¹),
- the indicator of equipping in water supply system increased on average by 18 no/locality,
- the development of water supply infrastructure is possible thanks to funds, the degree of development of rural areas depends on their height,
- the average percentage of the population using water supply network increased by 13% over the analysed years (the highest value of this indicator was recorded in 2014, in the Opole province, it was 94,9%),
- the length of the water supply network increased by 50334,6 km,
- the number of water supply connections increased by 774148,
- due to extension of the network supplying water and sewage disposal, there has been a reduction in the difference between the percentage of people using tap water and sewerage by 7,8%.

The determined values lead to the conclusion that there were big changes associated with the improvement of water supply infrastructure.

4. CONCLUSIONS

More and more of the rural population uses the water supply and sewage, which will undoubtedly affect the improvement of living standards. The improvement of water supply infrastructure also increases the attractiveness of rural areas, in terms of both tourism and investment. The greater flow of tourists allows the rural population developing farm tourism. The investors will have a huge impact on the development of rural areas by building plants where the rural population could find work.

The significant improvement of the infrastructure supplying water was observed, which was connected with large financial outlays intended for the construction and modernization of the water supply infrastructure in rural areas in Poland.

The multivariate comparative analysis with the use of k-means method and cluster analysis conducted for the indicators characterizing the water supply infrastructure in rural areas in 2014 allowed for the distinction of the following clusters:

- 1st cluster, consists of Lesser Poland and Subcarpathian, the individual indicators are at the average level, with the exception of the indicator related to capital expenditures on the construction and modernization of water supply network and also the number of water supply connections per km of water pipeline..
- 2nd cluster, consists of provinces: Podlaskie, Lublin, Lodz, Opolskie, West Pomeranian, Pomeranian, Lubusz, Warmian-Masurian, Kuyavian Pomeranian, Greater Poland, Silesian, Świętokrzyskie and Lower Silesian. The analysis of each indicator is not satisfactory in comparison with other clusters, whose averages reach the lowest values among the other clusters.
- 3rd cluster, consists of Masovian, which is the most economically developed, resulting in the highest indicators of the degree of equipping in the water supply network, compared to other provinces. The analysis showed that all the considered parameters in the third cluster had been improved over the analysed years.

In both analyses used to describe the water supply infrastructure state, by the agglomeration method and k-means method, the similar results of grouping were obtained.

REFERENCES

- [1] Everitt B.S., Landau S., Leese M., Stahl D., *Cluster Analysis*, John Wiley & Sons, Ltd, Chichester 2011;
- [2] Górczyca M., *Regionalne dysparytety rozwoju infrastruktury wodociągowej (Regional disparities of water supply infrastructure development)*. „Gaz, Woda i Technika Sanitarna”, 2011, No 5, pp. 166-170;
- [3] Hellwig Z., *Zarys ekonometrii (Outline of econometrics)*. PWE, Warszawa 1973;
- [4] Jajuga K., Sokołowski A., Bock H.H., *Classification, clustering and data analysis: recent advances and applications*. Springer, Berlin 2002;

- [5] Jajuga K., *Statystyczna analiza wielowymiarowa (Statistical multivariate analysis)*. PWN, Warszawa 1993;
- [6] Kleiner Y., Adams B., Rogers J., *Water distribution network renewal planning*, Journal of Computing in Civil Engineering, 2001, No 1, pp. 15-26;
- [7] Kwietniewski M., Rak J., *Niezawodność infrastruktury wodociągowej i kanalizacyjnej w Polsce (Reliability of water and sewerage infrastructure in Poland)*. PAN, Warszawa 2010;
- [8] Marek T., *Analiza skupień w badaniach empirycznych (Cluster analysis in empirical research)*. PWN, Warszawa 1989;
- [9] Mays W.L., *Water Distribution Systems. Handbook*. McGraw-Hill. New York 1999;
- [10] Migut G., *Zastosowanie technik analizy skupień i drzew decyzyjnych do segmentacji rynku (Application of cluster analysis and decision trees for of for market segmentation)*, StaSoftPolska 2009;
- [11] Pietrucha-Urbanik K., Studziński A., *Wybrane aspekty kosztów oraz awaryjności przewodów przykładowego systemu wodociągowego (Selected aspects of costs and failure of pipes in exemplary water supply system)*, Rocznik Ochrona Środowiska, 2016, No 2, pp. 623-634;
- [12] Pluta W., *Wielowymiarowa analiza porównawcza w badaniach ekonomicznych (Multidimensional comparative analysis in economic research)*. PWN, Warszawa 1977;
- [13] Studziński A., Pietrucha-Urbanik K., *Preventive maintenance and reliability of water supply system elements*. „Czasopismo Inżynierii Lądowej, Środowiska i Architektury - Journal of Civil Engineering, Environment And Architecture, JCEEA”, 2015, No 3/I, pp. 429-436. DOI: 10.7862/rb.2015.126;
- [14] Ward J.H., *Hierarchical grouping to optimize an objective function*, „Journal of the American Statistical Association”, 1963, No 3, pp. 236-244.

OCENA STANU INFRASTRUKTURY ZAOPATRUJĄCEJ W WODĘ NA OBSZARACH WIEJSKICH W POLSCE METODĄ WIELOWYMIAROWEJ ANALIZY

W pracy przeprowadzono ocenę stanu infrastruktury zaopatrującej w wodę i odprowadzającą ścieki na obszarach wiejskich w Polsce. Badanym okresem były lata 2003 oraz 2014. Na przestrzeni tych lat stan infrastruktury wodociągowej uległ znacznemu polepszeniu. Poprawa nastąpiła głównie dzięki wstąpieniu Polski do Unii Europejskiej, zatem musiano zadbać o podwyższenie poziomu świadczonych usług wodociągowych, tak aby dopasować je do prawa unijnego. Wspólnota europejska wspierała Polskę wszelkiego rodzaju dotacjami, dzięki którym rozbudowywano, a także poddawano modernizacji infrastrukturę wodociągową oraz kanalizacyjną. Do przeprowadzenia analizy użyto wielowymiarowej analizy porównawczej-metod aglomeracyjnej oraz metody k-średnich. Analizą objęto następujące parametry charakteryzujące rozwój infrastruktury zaopatrującej w wodę, do których należą: długość sieci wodociągowej, liczby przyłączy wodociągowych, liczby powstałych stacji uzdatniania wody, stopnia zwodociągowania obszarów wiejskich. W ostatnim dziesięcioleciu procent ludności korzystającej z sieci wodociągowej zwiększył się o 13%, w tym największy wzrost w województwach mazowieckim (wzrost o 32%) oraz

małopolskim (wzrost o 26%), a najmniejszy zaobserwowano w województwie opolskim (wzrost o 6%). Natomiast koszty związane z rozbudową infrastruktury wodociągowej na obszarach wiejskich wzrosły o 33%. Największe nakłady inwestycyjne zostały poniesione w województwie wielkopolskim (wzrost o 380%) oraz w województwie śląskim (wzrost o 260%). Podniesienie poziomu rozwoju infrastruktury wodociągowej oraz w konsekwencji kanalizacyjnej będzie wpływało na turystyczną oraz inwestycyjną atrakcyjność obszarów wiejskich. Również rozwój gospodarstw agroturystycznych jest uwarunkowany stanem infrastruktury wodociągowej, która bezpośrednio wpływa na powstawanie nowych miejsc pracy na terenach wiejskich.

Słowa kluczowe: długość sieci wodociągowej, obszary wiejskie, wielowymiarowa analiza porównawcza.

DOI:10.7862/rz.2016.hss.29

Przesłano do redakcji: styczeń 2016

Przyjęto do druku: lipiec 2016