Bulletin of Geography. Socio-economic Series, No. 54 (2021): 113–122
http://doi.org/10.2478/bog-2021-0037

ISSN 1732–4254 quarterly

BULLETIN OF GEOGRAPHY. SOCIO–ECONOMIC SERIES

journal homepages:
https://content.sciendo.com/view/journals/bog/bog-overview.xml
http://apcz.umk.pl/czasopisma/index.php/BGSS/index

The potential of data exploration methods in identifying the relationship between short-period (daily) water consumption and meteorological factors

Adam Piasecki[1, CDMR], Agnieszka Pilarska[2, CFMR], Radosław Golba[3, DFM]

[1,2,3]*Nicolaus Copernicus University in Toruń*, Faculty of Earth Sciences and Spatial Management, Toruń, Poland, [1]e-mail: piasecki@umk.pl *(corresponding author)*, ORCID ID https://orcid.org/0000-0002-8787-7520; [2] ORCID ID https://orcid.org/0000-0003-0725-2756; [3] ORCID ID https://orcid.org/0000-0002-9854-4000

How to cite:
Piasecki, A, Pilarska, A., & Golba R. (2021). The potential of data exploration methods in identifying the relationship between short-period (daily) water consumption and meteorological factors. *Bulletin of Geography. Socio-economic Series*, 54(54): 113-122. DOI: http://doi.org/10.2478/bog-2021-0037

Abstract. The purpose of the work was to identify the hidden relationship between water consumption and meteorological factors, using principal component analysis. In addition, clusters of similar days were identified based on relationships identified by k-means. The study was based on data from the city of Toruń (Poland). The analysis was based on daily data from 2014–2017 divided into three groups. Group I included data from the entire period, Group II- from warm half-years (April–September), and Group III-from cold half-years (January–March and October–December). For Groups I and II the extent of water consumption was explained by two principal components. PC1 includes variables that increase water consumption, and PC2 includes variables that lessen water demand. In Group III, water consumption was not linked to any component.
The *k*-means method was used to identify clusters of similar days. In terms of PC1, the most numerous days were Saturdays, and in terms of PC2 Sundays and holidays. It was determined that further research aimed at explaining the specificity of water consumption on particular days of the week is appropriate.

Article details:
Received: 27 July 2021
Revised: 29 September 2021
Accepted: 15 November 2021

Key words:
PCA,
k-means method,
water consumption,
meteorological factors,
Toruń,
Poland

Contents:

1. Introduction

Water has always been a fundamental determinant of life and socio-economic development. Ensuring the availability of the right amount and quality of water is currently a major problem for many countries and regions around the world. In its World Water Development Report (2018), the United Nations predicts that international conflicts over access to water resources will increase in the coming years. Poland is one of those countries that has relatively few water resources (Piasecki & Marszelewski, 2014). Calculated per capita, they amount to an average of 1,580 m^3, while this volume decreases to around 1,000 m^3 in years with low total precipitation. As a result, Poland's water resources are almost one third the European average (4,560 $m^3 \cdot M^{-1}$) and about 4.5 times smaller than the global average (7,300 $m^3 \cdot M^{-1}$) (Grabowska, 2010).

At the beginning of the 1990s, a period of political and economic transformation began in Poland. The introduction of the free market economy had a number of consequences, including a significant decrease in water consumption. In the following years, with the financial support of the European Community (later from the European Union), large investments in water and wastewater management began with pre-accession funding. These investments grew after 2004 with Poland's accession to the EU and the inflow of additional resources as part of structural funds (Piasecki & Marszelewski 2016). In the years 1995–2017, a total of over 148,400 km of water supply network (16.7% in cities) was laid, and 123,200 km of sewerage network (29.7% in cities). As with many countries, Poland considers water supply infrastructure as a critical infrastructure (Gorączko & Pasela, 2015). Ensuring it's proper and trouble-free operation is a major priority for local authorities. It is therefore important to determine the factors that influence the demand for mains water. The purpose of the work is to identify the hidden relationship between water consumption and meteorological factors, using the principal component analysis. Another incidental goal, however, is to idientify clusters of days that are similar based on identified relationships, using the *k*-means method. The study was based on data from the city of Toruń (Poland).

Principal Component Analysis (PCA) and modifications thereof are widely used in various scientific fields, not only socio-economic, but also environmental (Di Salvo et al., 2015; Săndică et al., 2018; He et al., 2018; Fernández et al., 2018), agriculture and biological sciences (Hilman et al., 2007; Dong et al., 2015; Otitoju and Enete, 2016; Moraetis et al., 2016), and medicine and health (Berglund et al., 2017; Khadra et al., 2017; Hellwege et al., 2017; Zhang et al., 2018). The PCA method is also used in works on water consumption (Nosvelli and Musolesi, 2009; Zhou et al., 2009; Wentz et al., 2014; Wright et al., 2018). It is common to use the principal component method alongside cluster analysis, including in geology and geochemistry (Gazley et al., 2015; Jiang et al., 2015; Iwamori et al., 2017; Lindsey et al., 2018), and in research related to water quality (Oketola et al., 2013; Zou et al., 2015; Jankowska et al., 2017; Marín Celestino et al., 2018), and economics and tourism (Czillingová et al., 2012; Xhafaj & Nurja, 2015; Argüelles et al., 2014; Vajčnerová et al., 2016).

2. Methods

The work uses daily water consumption data (in 2014–2017) in Toruń, as provided by the company Toruńskie Wodociągi Sp. z o.o. The study also included meteorological data (from 2014–2017) from the Toruń-Wrzosy station belonging to the Institute of Meteorology and Water Management - National Research Institute (IMWM NRI) (Fig.1). The following set of meteorological parameters (daily values) was used: average temperature, minimum temperature, maximum temperature, precipitation, humidity and insolation. These were also used to calculate evaporation using the Penman–Monteith method (Allen, 2000).

Meteorological measurements and observations are carried out at a first-order synoptic station in Toruń, which belongs to the network of meteorological stations of the National Hydrological and Meteorological Service. The station carries out 24-hour meteorological measurements and observations with the use of standard and automatic equipment, as well as complementary visual measurements. SYNOP reports are generated every

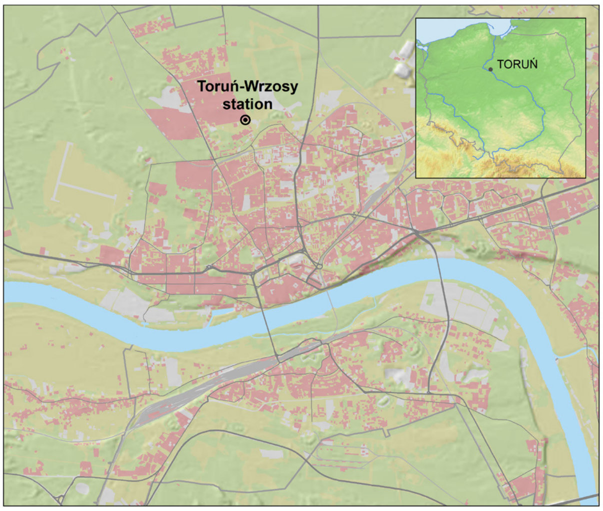

Fig. 1. Localization of Toruń - Wrzosy station.
Source: Own elaboration based on data from Head Office of Geodesy and Cartography (HOGC, 2020: BDOT10k, PRG, NMT, shaded model)

hour on the hour. Observations last for 24 hours every day, throughout the year.

Currently, measurements are performed with the use of the following equipment:

- the current air temperature (T) as well as daily maxims and minims of air temperature (Tmax, Tmin), are collected by means of automatic measuring equipment (MAWS). Sensors collecting data for the above-mentioned parameters are placed in a meteorological cage, 2 metres above the ground covered with low grass;
- relative air humidity (U) is determined by automatic measuring equipment - MAWS;
- measurements of wind direction and speed are performed by Vaisala sensors in the MAWS station, and the backup devices include AVIOMET and YOUNG (05103 and W863) anemometers;
- precipitation height (Rd) is measured 4 times daily, at: 00, 06, 12 and 18 UTC. The amount of precipitation is measured with a Hellmann rain gauge, with a reception area of 200 cm^2;
- sunshine duration (S) is measured with the use of electronic sensors connected to MAWS automatic stations and expressed in hours.

In the paper the principal component analysis (PCA) was used, which compared to canonical variates analysis (CVA), maximizes the total variance of a sample (in which there is no known group structure), by developing linear combinations (Thalib et al., 1999). The canonical variates analysis (CVA) is described as a technique for multivariate data which have known (a priori) group structure (Krzanowski, 1992). It should also be emphasized that in the case of descriptive sensory studies, CVA and PCA maps are very similar (Peltier et al., 2015).

The main task of the principal component method is to reduce dimensionality (Jolliffe, 2002). The method also allows patterns to be detected and verified in relationships between variables (Stanisz, 2007). These PCA properties constitute the methodological basis for this article. Many adaptations of PCA are currently being developed, including functional principal component analysis, simplified principal components, robust principal component analysis and symbolic data principal component analysis (Jolliffe & Cadima, 2016). These adaptations of PCA usually improve its effectiveness in solving a specific research problem. At the same time, the modifications often reduce the method's universality.

Bearing in mind that the purpose of the work is to identify a hidden relationship between water consumption and meteorological factors, a correlation matrix was used as the starting point of the analysis as part of a standard PCA procedure. Bartlett's sphericity test was carried out and the Kaiser–Mayer–Olkin coefficient (KMO) was determined to verify the validity of using PCA for the set of variables in question. According to the literature, KMO values below 0.500 are unacceptable (Hutcheson and Sofroniou, 1999). The analysed variable sets broken down for the whole year, cold half-year and warm half-year, exceeded 0.500. Furthermore, Bartlett's sphericity test showed significant correlations between variables. The use of principal component analysis is thus justified in this article.

The principal components were identified using the eigenvalue criterion – the so-called the Kaiser (1960) criterion – eigenvalues should be greater than or equal to 1. After the principal components were isolated, they were profiled. Based on the correlation values obtained between the identified components and the analysed variables, the hidden relationship between water consumption and meteorological factors was defined. The PCA analysis was performed in the PS IMAGO 5 based on the IBM SPSS Statistics analytical engine.

The *k*-means method is a non-hierarchical grouping methods, characterised by the fact that isolated clusters differ from one other, but objects in a given cluster are maximally similar (Stanisz, 2007). In the *k*-means method, the grouping procedure is simple and the number of groups is fixed (Balicki, 2009). Among the variants of the *k*-medium method, Hartigan's (1975) variant is worth mentioning, as is the centre of gravity method and a variant that takes into account the problem of inter- and intra-group variance (Panek & Zwierzchowski, 2013).

The assumed incidental goal was achieved using the *k*-means method. Appropriate calculations were made in the PS IMAGO 5 based on the IBM SPSS Statistics analytical engine. In this environment, the *k*-means algorithm uses simple Euclidean distance (IBM Knowledge Center 2018). Analysis was based on the iteration and classification method (not using moving averages) for the three data groups. Group I included data from the entire period, while Group II contained data from warm half-years (April–September), and Group III was of data from cold half-years (January–March and October–December). In the elaboration of graphical content, the ArcGIS 10.x, Inkscape 0.x and GIMP 2.x were used.

3. Results and discussion

The results in Table 1 indicate that the eigenvalue criterion (above 1) was met by just two components. This applies to all the analysed groups, for which two components explained approx. 69–79% of variance.

Based on the results in Table 2, individual variables were assigned to the two principal components. The first principal component (PC1) for Groups I and II was constructed from five variables: water consumption, maximum temperature, average temperature, insolation and evaporation. For Group I, an additional variable was used – minimum temperature. PC1 can be described as "conditions raising water demand". The second principal component (PC2) were constructed from the variables of precipitation and humidity. PC2 can be defined as "conditions lessening water demand". In Group III, water consumption was not related to any component.

These distinguished principal components accurately reflect the impact of meteorological conditions on water consumption. The results largely confirm those obtained using other research tools (Piasecki et al., 2018; Piasecki & Górski, 2018).

Table 1. Eigenvalues and percentage of variance accounted for analysed groups

Principal component (PC)	Group I			Group II			Group III		
	Eigenvalue	Explained variance [%]	Explained cumulative variance [%]	Eigenvalue	Explained variance [%]	Explained cumulative variance [%]	Eigenvalue	Explained variance [%]	Explained cumulative variance [%]
1	4.926	61.57	61.57	4.293	53.66	53.66	3.338	41.73	41.73
2	1.414	17.68	79.25	1.921	24.02	77.68	2.172	27.15	68.88
3	0.755	9.44	88.69	0.835	10.44	88.12	0.994	12.43	81.30
4	0.569	7.11	95.80	0.448	5.60	93.71	0.852	10.65	91.96
5	0.215	2.68	98.48	0.293	3.67	97.38	0.434	5.43	97.38
6	0.064	0.80	99.28	0.122	1.53	98.91	0.136	1.70	99.08
7	0.051	0.64	99.92	0.070	0.88	99.78	0.064	0.80	99.88
8	0.007	0.08	100	0.017	0.22	100.00	0.01	0.12	100.00

Source: own elaboration.

The relationships identified between water consumption and meteorological factors are confirmed by studies on other cities in Poland and worldwide. These studies pointed to air temperature as one factor increasing water consumption (Hotloś, 2013; Haque et al., 2015a; Haque et al., 2015b; de Souza et al., 2015; Xiao-jun et al., 2015; Xenochristou et al., 2018). Research in Germany has shown that lower water demand in the summer months is often associated with a higher average number of days with rainfall of >1 mm (Schleich & Hillenbrand, 2009). A similar relationship has been found in Poland for the city of Wrocław (Hotloś, 2013). Interesting results on the relationship between water consumption and precipitation have also been obtained for the Greater Melbourne region of Australia. Researchers have determined that daily water consumption decreases as precipitation rises, but that this relationship does not always pertain. They determined that the threshold value at which precipitation causes no further decrease in daily water consumption is 4.08 mm (Sarker et al., 2013). The same authors also indicated an analogous air temperature threshold of 15.53°C. Below this value, air temperature was not observed to affect water consumption.

One important issue is that the correlation between water consumption and principal component PC1 was highest for Group II (r=0.739). The lower correlation for Group I (r=0.620) was mainly due to data from the cold half-year (Group III). This confirms the lack of relationship between water consumption and the identified principal components for Group III. It can therefore be concluded that the data from the cold half-year distorts the relationship between water consumption and meteorological variables in the analysis of the year as a whole. Similar studies carried out for other cities in Poland and around the world have confirmed the insignificant impact of meteorological conditions on water consumption in the cold half-year (Akuoko-Asibey et al., 1993; Kolendo, 2016). This is very noteworthy, and should always be mentioned in works modelling or forecasting water consumption.

The high values of the components PC1 and PC2 are associated with different days. For PC1 in Groups I and II there were two Saturdays (4/07/2015, 8/08/2015), and for PC2 it was a Wednesday (13/07/2016) and a Thursday (14/07/2016).

Similar days in terms of the principal component PC1, were grouped in cluster 1 for Group I, and in

Table 2. Correlation between the two distinguished components and the analysed variables

Variable	Group I		Group II		Group III	
	PC1	PC2	PC1	PC2	PC1	PC2
Water consumption	0.62	-0.354	0.739	-0.294	-0.077	-0.142
Minimum temperature	0.794	0.499	0.585	0.741	0.754	0.586
Maximum temperature	0.945	0.241	0.889	0.372	0.952	0.175
Insolation	0.767	-0.425	0.695	-0.538	0.398	-0.754
Humidity	-0.806	0.397	-0.702	0.571	-0.48	0.783
Precipitation	0.01	0.729	-0.093	0.534	0.076	0.486
Evaporation	0.966	-0.069	0.944	-0.141	0.815	-0.433
Average temperature	0.921	0.327	0.859	0.477	0.894	0.416

Source: own elaboration.

Table 3. Cluster centres

PC	Group I			Group II			Group III		
Cluster	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
PC 1	1.202	-0.785	0.092	-0.380	-0.809	1.032	-1.034	0.183	1.191
PC 2	-0.490	-0.357	1.201	0.936	-0.894	-0.268	-0.392	0.692	-1.108
Number of observations	414	677	370	269	204	259	228	358	143

Source: own elaboration.

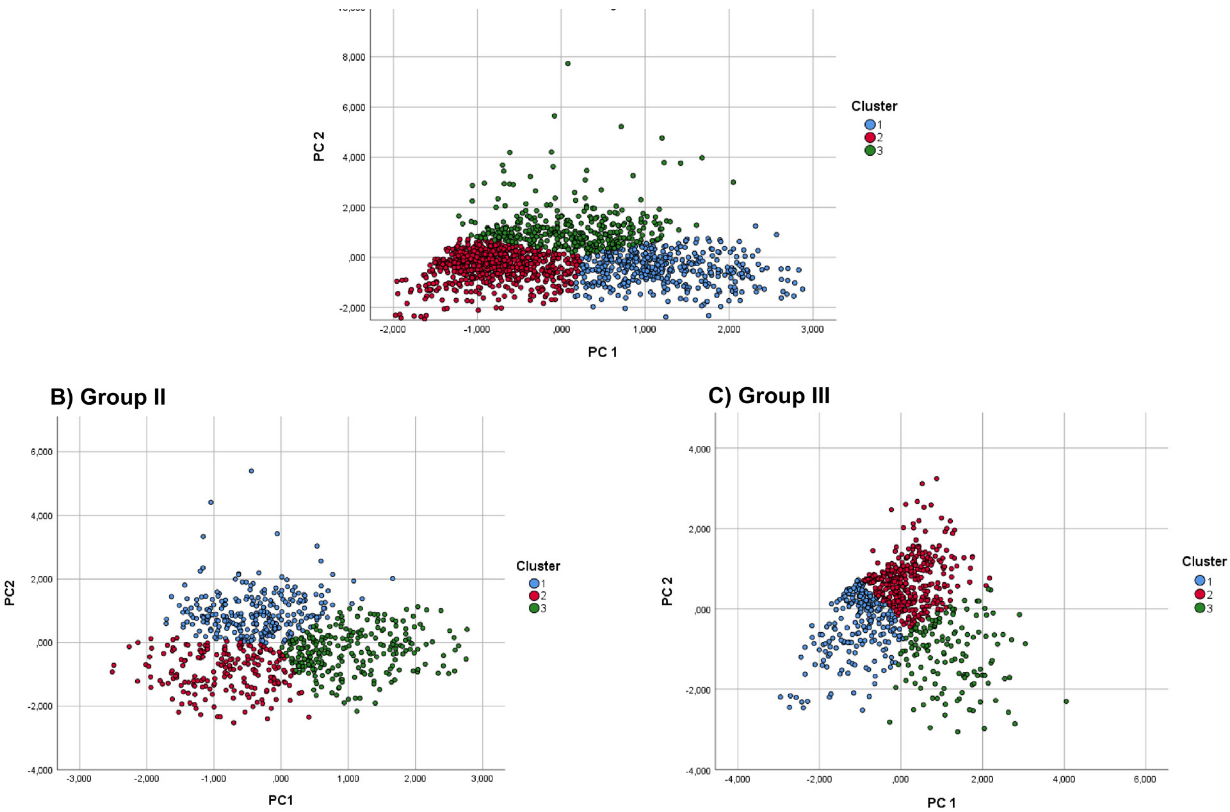

Fig. 2. Distribution of clusters in PC1 and PC2 component spaces
Source: Own elaboration.

cluster 3 for Group II (Table 3, Fig. 2). In clusters both 1 and 3, the most numerously occurring similar days were Saturdays (Group I – 67 days; Group II – 40 days). The least numerous similar days were Wednesdays for Group I (54 days), and Thursdays for Group II (34 days).

Principal component PC2 is represented in cluster 3 for Group I, and in cluster 1 for Group II. In these clusters, the most frequent similar days were Sundays and holidays (Group I – 72; group II – 52 days). The least numerous similar days were Saturdays (Group I – 45 days) and Thursdays (Group II – 29 days). In the cold half-year (Group III), water consumption was not positively correlated with any component.

The analysis showed that the largest number of similar days in terms of PC1 and PC2 occurs in the warm half-year (Group II). Non-work days were the most common similar days. The reasons for this can be found in the fact that, on such days, the city's residents primarily account for water consumption. In favourable weather, many residents devote time on such days to watering their lawns or allotments.

4. Summary

Using principal components identified the hidden relationships between water consumption and meteorological factors. The amount of water consumption was found to be explained by two principal components. The principal component PC1 is defined by variables that equate with increased water demand. The meteorological variables forming PC1 include air temperature (average, maximum and minimum), insolation and evaporation. In the case of principal component PC2, its component variables (humidity and precipitation) are associated with conditions that lessen water demand.

The *k*-means method was used to identify clusters of similar days. Of those, Saturdays were the most numerous for PC1, and Sundays and holidays for PC2. The least numerous similar days were Wednesdays and Thursdays for PC1, and Thursdays and Saturdays for PC2. In light of this, further research to explain the specific characteristics of water consumption on these days is appropriate.

The results obtained in this work confirm the utility of the PCA method in studying the internal structure of data and hidden relationships between variables. It should be emphasised that the results obtained using PCA and k-means can also be used as the basis for further, more advanced analyses. This applies in particular to modelling and forecasting water consumption using the independent variables determined by these methods.

References

Akuoko-Asibey, A., Nkemdirim, L.C., & Draper, D.L. (1993). The impacts of climatic variables on seasonal water consumption in Calgary, Alberta. *Canadian Water Resources Journal*, 18(2): 107-116. DOI: https://doi.org/10.4296/cwrj1802107

Allen, R.G. (2000). Using the FAO-56 dual crop coefficient method over an irrigated region as part of an evapotranspiration intercomparison study. *Journal of Hydrology*, 229(1-2): 27-41. DOI: https://doi.org/10.1016/S0022-1694(99)00194-8

Argüelles, M., Benavides, C., & Fernández, I. (2014). A new approach to the identification of regional clusters: hierarchical clustering on principal components. *Applied Economics*, 46(21): 2511-2519. DOI: https://doi.org/10.1080/00036846.2014.904491

Balicki, A. (2009). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne* (Statistical multivariate analysis and its socio-economic applications – in Polish). Gdańsk: Wydawnictwo Uniwerytetu Gdańskiego.

Berglund, A.E., Welsh, E.A., & Eschrich, S.A. (2017). Characteristics and validation techniques for PCA-based gene-expression signatures. *International Journal of Genomics*, 2017: 2354564. DOI: https://doi.org/10.1155/2017/2354564

Czillingová, J., Petruška, I., & Tkáč, M. (2012). Financial and economic analysis of steel industry by multivariate analysis. *Ekonomický časopis* (*Journal of Economics*), 60(4): 388-405. Available at: https://www.sav.sk/journals/uploads/0622125104%2012%20Tkac-RS.pdf (3.12.2021).

de Souza, A., Aristone, F., Sabbah, I., da Silva Santos, D.A., de Souza Lima, A.P., & Lima, G. (2015). Climatic Variations and Consumption of Urban Water. *Atmospheric and Climate Sciences*, 5(3): 292-301. DOI: http://dx.doi.org/10.4236/acs.2015.53022

Di Salvo, F., Ruggieri, M., & Plaia, A. (2015). Functional principal component analysis for multivariate multidimensional environmental data. *Environmental and Ecological Statistics*, 22(4): 739-757. DOI: https://doi.org/10.1007/s10651-015-0317-8

Dong, F., Mitchell, P.D., & Colquhoun, J. (2015). Measuring farm sustainability using data envelope analysis with principal components: The case of Wisconsin cranberry. *Journal of Environmental Management*, 147: 175-183. DOI: https://doi.org/10.1016/j.jenvman.2014.08.025

Fernández, S., Cotos-Yáñez, T., Roca-Pardiñas, J., & Ordóñez, C. (2018). Geographically weighted principal components analysis to assess diffuse pollution sources of soil heavy metal: application to rough mountain areas in Northwest Spain. *Geoderma,* 311: 120-129. DOI: https://doi.org/10.1016/j.geoderma.2016.10.012

Gazley, M.F., Collins, K.S., Roberston, J., Hines, B.R., Fisher, L.A., & McFarlane, A. (2015). Application of principal component analysis and cluster analysis to mineral exploration and mine geology. *AusIMM New Zealand Branch Annual Conference*, 131-139.

Gorączko, M., & Pasela, R. (2015). Causes and effects of water consumption drop by the population of cities in Poland - selected aspects. *Bulletin of Geography. Socio-economic Series*, 27: 67-79. DOI: https://doi.org/10.1515/bog-2015-0005

Grabowska, M. (2010). Wodne bariery rozwoju gospodarczego Polski (Water-related barriers in the economic development of Poland). *Socio-Economic Problems and the State* 1(3): 55-61. Available at: http://elartu.tntu.edu.ua/bitstream/123456789/648/1/SEPS_2010_v1_No3_M_Grabowska-Water-related_barriers_in_the_economic_development_of_Poland__55.pdf (3.12.2021).

Haque, M.M., Egodawatta, P., Rahman, A., & Goonetilleke, A. (2015a). Assessing the significance of climate and community factors on urban water demand. *International Journal of Sustainable Built Environment,* 4(2): 222-230. DOI: https://doi.org/10.1016/j.ijsbe.2015.11.001

Haque, M., Rahman, A., Goonetilleke, A., Hagare, D., & Kibria, G. (2015b). Impact of climate change on urban water demand in future decades: An Australian case study. In: Daniels, J.A. (ed.) *Advances in environmental research.* United States of America: Nova Science Publishers, 57-70. Available at: https://eprints.qut.edu.au/84018/21/Impact%2Bof%2Bclimate%2Bchange%2Bon%2Burban%2Bwater%2Bdemand%2Bin%2Bfuture%2Bdecades_An%2BAustralian%2Bcase%2Bstudy.pdf (3.12.2021).

Hartigan, J.A. (1975). *Clustering Algorithms*. New York: John Wiley & Sons. Inc.

He, Y., Pang, Y., Zhang, Q., Jiao, Z., & Chen, Q. (2018). Comprehensive evaluation of regional clean energy development levels based on principal component analysis and rough set theory. *Renewable Energy,* 122: 643-653. DOI: https://doi.org/10.1016/j.renene.2018.02.028

Hellwege, J.N., Jeff, J.M., Wise, L.A., Gallagher, C.S., Wellons, M., Hartmann, K.E., Jones, S.F., Torstenson, E.S., Dickinson, S., Ruiz-Narváez, E.A., Rohland, N., Allen, A., Reich, D., Tandon, A., Pasaniuc, B., Mancuso, N., Im, H.K., Hinds, D.A., Palmer, J.R., Rosenberg, L., Denny, J.C., Roden, D.M., Stewart, E.A., Morton, C.C., Kenny, E.E., Edwards, T.L., & Velez Edwards, D.R. (2017). A multi-stage genome-wide association study of uterine fibroids in African Americans. *Human Genetics,* 136(10): 1363-1373. DOI: https://doi.org/10.1007/s00439-017-1836-1

Hilman, Y., Rahim, A.A., Musa, M.H., & Hashim, A. (2007). Principal component analysis of factors determining phosphate rock dissolution on acid soils. *Indonesian Journal of Agricultural Science,* 8(1): 10-16. DOI: http://dx.doi.org/10.21082/ijas.v8n1.2007.p10-16

HOGC. 2020. Head Office of Geodesy and Cartography. BDOT10k - Baza Danych Obiektów Topograficznych (Database of Topographic Objects), NMT – Numeryczny Model Terenu (Digital Terrain Model) and shaded model from: https://mapy.geoportal.gov.pl/imap/Imgp_2.html (November 2020).

PRG – Państwowy Rejestr Granic i Powierzchni Jednostek Podziałów Terytorialnych Kraju (State Register of Borders and Area of Territorial Units of the Country) from: http://www.gugik.gov.pl/pzgik/inne-dane-udostepniane-bezplatnie (November 2020).

Hotloś, H. (2013). Analiza wpływu czynników meteorologicznych na zmienność poboru wody w miejskim systemie wodociągowym (Analysis of influence of meteorological factors on water demand variations in municipal water supply system– in Polish). *Ochrona Środowiska*, 35(2): 57-62. Available at: http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-07b84ec1-a4d4-4920-abdc-1f3250257942 (3.12.2021).

Hutcheson, G.D., & Sofroniou, N. (1999). The multivariate social scientist: Introductory statistics using generalized linear models. London-Thousand Oaks-New Dehli: SAGE Publications Ltd.

IBM Knowledge Center. 2018. Available at: https://www.ibm.com/support/knowledgecenter/pl/SSLVMB_sub/statistics_mainhelp_ddita/spss/base/idh_quic.html (18.11.2018)

Iwamori, H., Yoshida, K., Nakamura, H., Kuwatani, T., Hamada, M., Haraguchi, S., & Ueki, K. (2017). Classification of geochemical data based on multivariate statistical analyses: Complementary roles of cluster, principal component, and independent component analyses. *Geochemistry, Geophysics, Geosystems,* 18(3): 994-1012. DOI: https://doi.org/10.1002/2016GC006663

Jankowska, J., Radzka, E., & Rymuza, K. (2017). Principal Component Analysis and Cluster Analysis In Multivariate Assessment of Water Quality. *Journal of Ecological Engineering,* 18(2): 92-96. DOI: https://doi.org/10.12911/22998993/68141

Jiang, Y., Guo, H., Jia, Y., Cao, Y., & Hu, C. (2015). Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia. *Geochemistry,* 75(2): 197-205. DOI: https://doi.org/10.1016/j.chemer.2014.12.002

Jolliffe, I. (2002). *Principal Component Analysis. 2nd Edition.* New York: Springer.

Jolliffe, I.T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 374 (2065): 20150202. DOI: https://doi.org/10.1098/rsta.2015.0202

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement,* 20(1): 141-151. DOI: https://doi.org/10.1177/001316446002000116

Khadra, C., Le May, S., Ballard, A., Théroux, J., Charette, S., Villeneuve, E., Parent, S., Tsimicalis, A., &

MacLaren, Chorney J. (2017). Validation of the scale on Satisfaction of Adolescents with Postoperative pain management - idiopathic Scoliosis (SAP-S). *Journal of Pain Research,*10: 137-143. DOI: https://doi.org/10.2147/JPR.S124365

Kolendo, M. (2016). Czynniki ekonomiczno-środowiskowe w modelowaniu miesięcznego zapotrzebowania na wodę na przykładzie Białegostoku (Economic and environmental aspects in modeling monthly water demand: the case study of Bialystok – in Polish). *Ekonomia i Środowisko*, 2(57): 162-174. Available at: http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-d75cfd59-7213-4d11-9f1e-1c436997f941 (3.12.2021).

Krzanowski, W.J. (1992). Ranking principal components to reflect group structure. *Journal of Chemometrics,* 6(2): 97-102. DOI: https://doi.org/10.1002/cem.1180060207

Lindsey, C.R., Neupane, G., Spycher, N., Fairley, J.P., Dobson, P., Wood, T., McLing, T., & Conrad, M. (2018). Cluster analysis as a tool for evaluating the exploration potential of Known Geothermal Resource Areas. *Geothermics,* 72: 358-370. DOI: https://doi.org/10.1016/j.geothermics.2017.12.009

Marín Celestino, A.E., Martínez Cruz, D.A., Otazo Sánchez, E.M,. Gavi Reyes, F., & Vásquez Soto, D. (2018). Groundwater Quality Assessment: An Improved Approach to K-Means Clustering, Principal Component Analysis and Spatial Analysis: A Case Study. *Water,* 10(4): 437. DOI: https://doi.org/10.3390/w10040437

Moraetis, D., Lydakis-Simantiris, N., Pentari, D., Manoutsoglou, E., Apostolaki, C., & Perdikatsis, V. (2016). Chemical and physical characteristics in uncultivated soils with different lithology in semiarid Mediterranean clima. *Applied and Environmental Soil Science*, 2016: 3590548. DOI: https://doi.org/10.1155/2016/3590548

Nosvelli, M., & Musolesi, A. (2009). Water consumption and long-run socio-economic development: An intervention and a principal component analysis for the city of Milan. *Environmental Modeling & Assessment,* 14(3): 303-314. DOI: https://doi.org/10.1007/s10666-007-9127-1

Oketola, A.A., Adekolurejo, S.M., & Osibanjo, O. (2013). Water quality assessment of River Ogun using multivariate statistical techniques. *Journal of Environmental Protection,* 4(5): 466-479. DOI: http://dx.doi.org/10.4236/jep.2013.45055

Otitoju, M.A., & Enete, A.A. (2016). Climate change adaptation: Uncovering constraints to the use of adaptation strategies among food crop farmers in South-west, Nigeria using principal component analysis (PCA). *Cogent Food & Agriculture,* 2(1): 1178692. DOI: https://doi.org/10.1080/23311932.2016.1178692

Panek, T., & Zwierzchowski, J.K. (2013). *Statystyczne metody wielowymiarowej analizy porównawczej: teoria i zastosowania* (Statistical methods of multivariate comparative analysis: theory and applications – in Polish). Warszawa: Oficyna Wydawnicza, Szkoła Główna Handlowa.

Lewandowska, A., & Piasecki A. (2019). Selected aspects of water and sewage management in Poland in the context of sustainable urban development. *Bulletin of Geography. Socio-economic Series*, 45(45): 149-157. DOI: https://doi.org/10.2478/bog-2019-0030

Piasecki, A., & Górski, Ł. (2018). Analysis of water consumption in 2014-2017 in Toruń. *Infrastructure and Ecology of Rural Areas,* IV/1/2018: 973–984. DOI: https://doi.org/10.14597/INFRAECO.2018.4.1.067

Piasecki, A., Jurasz, J., & Kaźmierczak, B. (2018). Forecasting Daily Water Consumption: a Case Study in Torun, Poland. *Periodica Polytechnica Civil Engineering,* 62(3): 818-824. DOI: https://doi.org/10.3311/PPci.11930

Piasecki, A., & Marszelewski, W. (2016). Development of Water and Sewage Infrastructure in Poland During the Cooperation with the EU. *Wasserwirtschaft,* 106(4): 34-38.

Peltier, C., Visalli, M., & Schlich, P. (2015). Comparison of canonical variate analysis and principal component analysis on 422 descriptive sensory studies. *Food Quality and Preference,* 40 Part B: 326-333. DOI: https://doi.org/10.1016/j.foodqual.2014.05.005

Săndică, A.M., Dudian, M., & Ştefănescu, A. (2018). Air Pollution and Human Development in Europe: A New Index Using Principal Component Analysis. *Sustainability,* 10(2): 312. DOI: https://doi.org/10.3390/su10020312

Sarker, R.C., Gato-Trinidad, S., & Imteaz, M. (2013). Temperature and rainfall thresholds corresponding to water consumption in Greater Melbourne, Australia. *20th International Congress on Modelling and Simulation (MODSIM2013),* 2576-2582. Available at: https://researchbank.swinburne.edu.au/file/4a3e1173-0254-4ae9-b086-656a525311bd/1/PDF%20%28Published%20version%29.pdf (3.12.2021)

Schleich, J., & Hillenbrand, T. (2009). Determinants of residential water demand in Germany. *Ecological Economics,* 68(6): 1756-1769. DOI: https://doi.org/10.1016/j.ecolecon.2008.11.012

Stanisz, A.(2007). *Przystępny kurs statystyki: z zastosowaniem STATISTICA PL na przykładach z medycyny. Analizy wielowymiarowe* (An affordable statistics course: with the use of STATISTICA PL on examples from medicine. Multivariate analyzes – in Polish). Kraków: StatSoft Polska.

Thalib, L., Kitching, R.L., & Bhatti, M.I. (1999). Principal component analysis for grouped data—a case study. *Environmetrics: The official journal of the International Environmetrics Society,* 10(5): 565-574. DOI: https://doi.org/10.1002/(SICI)1099-095X(199909/10)10:5%3C565::AID-ENV360%3E3.0.CO;2-R

Vajčnerová, I., Šácha, J., Ryglová, K., & Žiaran, P. (2016). Using the cluster analysis and the principal component analysis in evaluating the quality of a destination. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis,* 64(2): 677-682. DOI: http://dx.doi.org/10.11118/actaun201664020677

Wentz, E.A., Wills, A.J., Kim, W.K., Myint, S.W., Gober, P., & Balling Jr, R.C. (2014). Factors influencing water consumption in multifamily housing in Tempe, Arizona. *The Professional Geographer,* 66(3): 501-510. DOI: https://doi.org/10.1080/00330124.2013.805627

Wright, C.J., Sargeant, J.M., Edge, V.L., Ford, J.D., Farahbakhsh, K., Shiwak, I., Flowers, C., Gordon, A.C., RICG., IHACC Research Team., & Harper, S.L. (2018). How are perceptions associated with water consumption in Canadian Inuit? A cross-sectional survey in Rigolet, Labrador. *Science of the Total Environment,* 618: 369-378. DOI: https://doi.org/10.1016/j.scitotenv.2017.10.255

WWAP (United Nations World Water Assessment Programme)/UN-Water. 2018. The United Nations World Water Development Report 2018: Nature-Based Solutions for Water. Paris, UNESCO.

Xenochristou, M., Blokker, M., Vertommen, I., Urbanus, J.F.X., & Kapelan, Z. (2018). Investigating the Influence of Weather on Water Consumption: A Dutch Case Study. *WDSA / CCWI Joint Conference 2018,* 1. Available at: https://ojs.library.queensu.ca/index.php/wdsa-ccw/article/view/12048 (3.12.2021)

Xhafaj, E., & Nurja, I. (2015). The principal components analysis and cluster analysis as tools for the estimation of poverty, an Albanian Case Study. *International Journal of Science and Research,* 4(1): 1240-1243. Available at: https://www.ijsr.net/archive/v4i1/SUB15305.pdf (3.12.2021)

Xiao-jun, W., Jian-yun, Z., Shamsuddin, S., Rui-min, H., Xing-hui, X., & Xin-li, M. (2015). Potential impact of climate change on future water demand in Yulin city, Northwest China. *Mitigation and Adaptation Strategies for Global Change,* 20(1): 1-19. DOI: https://doi.org/10.1007/s11027-013-9476-9

Zhang, Y., Li, X., Mao, L., Zhang, M., Li, K., Zheng, Y., Cui, W., Yin, H., He, Y., & Jing, M. (2018). Factors affecting medication adherence in community-managed patients with hypertension based on the principal component analysis: evidence from Xinjiang, China. *Patient Preference and Adherence,* 12: 803-812. DOI: https://dx.doi.org/10.2147%2FPPA.S158662

Zhou, D., Zhang, R., Liu, L., Gao, L., & Cai, S. (2009). A study on water resources consumption by principal component analysis in Qingtongxia irrigation areas of Yinchuan Plain, China. *Journal of Food, Agriculture & Environment,* 7(3-4): 734-738.

Zou, H., Zou, Z., & Wang, X. (2015). An enhanced K-means algorithm for water quality analysis of The Haihe River in China. *International Journal of Environmental Research and Public Health,* 12(11): 14400-14413. DOI: https://doi.org/10.3390/ijerph121114400