Teka Kom. Ochr. Kszt. Środ. Przyr. – OL PAN, 2013, 10, 5–13

REGIONALIZATION OF CATCHMENTS WITH USE HIERARCHICAL CLUSTER ANALYSIS METHODS

Agnieszka Cupak

Department of Sanitary Engineering and Water Management, Agricultural University in Kraków
Mickiewicza Av. 24-28, 30-059 Kraków, a.cupak@ur.krakow.pl

Summary. In the paper, the attempt of hydrological homogenous region regards specific discharge q_{95} and chosen catchment's parameters was made. The analysis was made with use chosen methods of cluster analysis. Research was conducted for 34 catchments located in Upper Vistula river basin. On the basis of conducted research it was stated, that the best result was got for Ward's method, for which an average silhouette width equal 0.63. The lowest ASW was got for single and complete linkage methods. In case of pair-group method using the centroid average and weighted pair-group method using the centroid average should not be used for purpose of hydrological homogenous region determination.

Key words: low flow, cluster analysis, catchment, homogenous region

INTRODUCTION

Low flows, its regime and influence on biological function and stability of water ecosystem is a very important issue in hydrology [Števková *et al*. 2012]. The information about low flow, with a given warranty is essential in water resources management among other things to assess the energy production in water power plant, for water intake, for systems of fish breeding, for agricultural irrigation, for biological flow and to calculate dilution of contaminants off into the river after sewage treatment plant. According to the Water Frame Directive, for purpose of characterization an ecological condition of river the low flow characteristics are needed to estimate [Laaha and Blöschl 2006, Mamun *et al*. 2010]. Most of methods of low flow estimation require suitable observational data, which only can be got from controlled catchments, while uncontrolled catchments present a separate problem. Recently, more often to evaluation of low flow characteristics in uncontrolled catchments the regional regression methods are used [Laaha and Blöschl 2006, Mamun *et al*. 2010]. The basis of regionalization is a group of hydrological factors, which the best describe a specific hydrological information. Generation of hydrological homogeneous regions, can be based on many procedures, including regression tree, seasonal index or cluster analysis [Lin and Wang 2006].

The aim of the research was to evaluate the possibility of use statistical method for data agglomeration – cluster analysis for determination of hydrological homogeneous regions for the sake of the low flow and chosen catchments' physiographic and meteorological characteristics.

MATERIALS AND METHODS

The initial material for analysis was daily flows for 34 catchments located in Upper Vistula river basin and chosen physiographical and meteorological characteristics (Tab. 1). It was assumed, as a criterion of catchments' selection, that for analysis will be taken only these one, for which daily streamflows are available with a minimum record length of 10 years. Catchments chosen for analysis are diverse in respect of area, which approximates from 66.3 km^2 to 2093 km^2, mean slope is in range from 0.002 for Łęg river in Kępie Zaleszańskie gauging station to 0.091 for Biała river in Bielsko station. For interpretation of hydrological occurrences also medium elevation is useful, which is in range from 836 m o.s.l. (Dunajec river, cross section: Nowy Targ – Kowaniec) to 191 m o.s.l. (Breń river, cross section: Wampierzów). Also, as regards the precipitation, it is a diverse area. The average annual precipitation amounted more than 1000 mm for Carpathian inflows of Vistula river, and in case of other catchments about 600–800 mm. In analysis 11 physiographic and meteorological characteristics of catchments were used, such as: drainage area A (km^2), main channel length L (km), a mean slope ψ and medium elevation H_{me} (m o.s.l.), cross section elevation P.z. (m o.s.l.), soils (describes by Bołdakov coefficient – n), average annual precipitation P (mm), average annual temperature of the air T (°C), area percentage of evergreens (%). Daily streamflows and average annual precipitation were taken from Hydrological Yearbook for Vistula river basin, physiographical characteristics, average annual temperature of the air and land use were established on the basis of Hydrological Atlas of Poland [1987] and from Chełmicki [1991].

Low flows were quantified by the $Q_{95\%}$, i.e. the discharge that is exceeded on 95% of all days of the measurement period. This low flow characteristic is widely used in Europe and was chosen because of its relevance for multiple choices of water management, among other things in case of projection of water supply systems. Then, $Q_{95\%}$ was subsequently standardized by the catchment area and resulting specific low flow discharges q_{95} (dm^3·s^{-1}·km^{-2}).

Table 1. Information about cross sections

Catchment's code	Cross section	River	Main channel length, km	Catchment's area, km^2	q_{95} dm^3·s^{-1}· km^{-2}
C1	Skoczów	Wisła	71.1	297	3.03
C2	Nowy Targ – Kowaniec	Dunajec	75.4	681	5.33
C3	Stróża	Raba	51.4	644	3.57
C4	Zesławice	Dłubnia	42.4	264	2.65
C5	Dwikozy	Opatówka	47.1	256	1.32
C6	Koszyce	Biała Tarnowska	95.2	957	1.89
C7	Biskupice	Szreniawa	71	682	3.00
C8	Rudze	Wieprzówka	27.1	154	1.00
C9	Wampierzów	Breń	44.2	661	2.45
C10	Koprzywnica	Koprzywianka	54.7	499	1.50
C11	Kępie Zaleszańskie	Łęg	71.6	822	1.21
C12	Harasiuki	Tanew	95.2	2034	2.80
C13	Bielsko	Biała	8.8	7.03	5.54
C14	Międzyrzecze	Pszczynka	34.8	285	4.21
C15	Osielec	Skawa	27.6	244	3.48
C16	Skawica	Skawica	17.6	139	7.98
C17	Nowy Sącz	Kamienica Nawojowska	26	238	3.15
C18	Jakubkowice	Łososinka	33.2	343	2.51
C19	Mniszek	Biała Nida	34.8	439	2.61
C20	Morawica	Czarna Nida	32.4	755	1.91
C21	Wilkowa	Wschodnia	40.6	650	1.08
C22	Topoliny	Ropa	40.6	970	2.16
C23	Jasło	Jasiołka	56.2	513	1.89
C24	Terka	Solinka	25.4	310	5.3
C25	Szczawne	Osława	25.2	505	1.78
C26	Brzuska	Stupnica	16.8	173	3.54
C27	Gorliczyna	Mleczka	29	529	1.19
C28	Nowy Sącz	Łubinka	11.2	66.3	2.1
C29	Sarzyna	Trzebośnica	24.8	249	2.69
C30	Raków	Czarna	17	221	2.52
C31	Soła	Cięcina	28.4	413	4.1
C32	Grabinianka	Grabiny	24.8	180	2.44
C33	Wielopolka	Brzeźnica	24.6	484	2.11
C34	Wisłoka	Żółków	35.2	581	2.25

For purpose of homogenous area determination, following algorithms of hierarchical cluster analysis were used: complete and single linkage, pair group average, weighted pair group average, pair-group method using the centroid average, weighted pair-group method using the centroid average and Ward's method. For determination number of clusters the method proposed by Hellwig [1968] was use. For critical value estimation, the minimum value in each line of distance matrix has to be found. Then, for these variables an average $\bar{x}$ and standard deviation σ are calculated [Królczyk and Tukiendorf 2005]. The critical value is calculated with use following equation [Pluta 1977]:

$$W_k = \overline{x} + 2\sigma$$

The global goodness of clustering was checked on the basis of an average silhouette width (ASW) [Rao and Srinivas 2006]:

$$\mathrm{ASW} = \frac{(\sum_i sw_i)}{n}$$

$$sw = \frac{b_i - a_i)}{\max\{a_i, b_i\}}$$

where: sw – a silhouette width, and is in range: $-1 \leq sw \leq 1$

a_i – the average distance from point i to all other points in i's cluster,

b_i – the minimum average distance from point i to all points in another cluster.

ASW is in range: $0 < \mathrm{ASW} < 1$, and when the larger ASW the better split.

RESULTS AND DISCUSSION

On Figures 1 to 3, results of catchments' grouping as regards of low flow q_{95} formation and chosen physiographic and climatic parameters of catchments are shown. The critical value ($W_k = 3.6$), evaluated on the basis of the matrix of tacsonomic distances, serves to define a number of clusters, and catchments belong to each cluster, for applied grouping methods.

Chains created from objects characterize agglomeration with use single linkage method (Fig. 1a). For evaluated linkage distance ($W_k = 3.6$) four clusters were separated, from which three assemble by one catchment. Separation three clusters with one object is resulted, that these catchments have parameters, which differ from others. In case of first cluster (Dunajec river, section: Nowy Targ – Kowaniec) the biggest medium elevation (Hme = 836 m o.s.l.) and cross section elevation – 574.3 m o.s.l., from all objects were decided. In case of second cluster: Tanew river; station: Harasiuki, the biggest drainage area (A = 2034 km^2) decided, and in case of third cluster (Biała river; station Bielsko) the highest mean slope ($\psi = 0.091$). Fourth cluster includes others catchments, regardless of their physiographic and climatic characteristics. A large differential of chosen parameters in analyzed catchments causes, that single linkage method should not be use for a proper appointment of hydrological homogenous regions. Above statement is confirmed by evaluated for this method average silhouette width ASW, which amounted 0.27. It means that clusters made with use single linkage method probably has an accidental character.

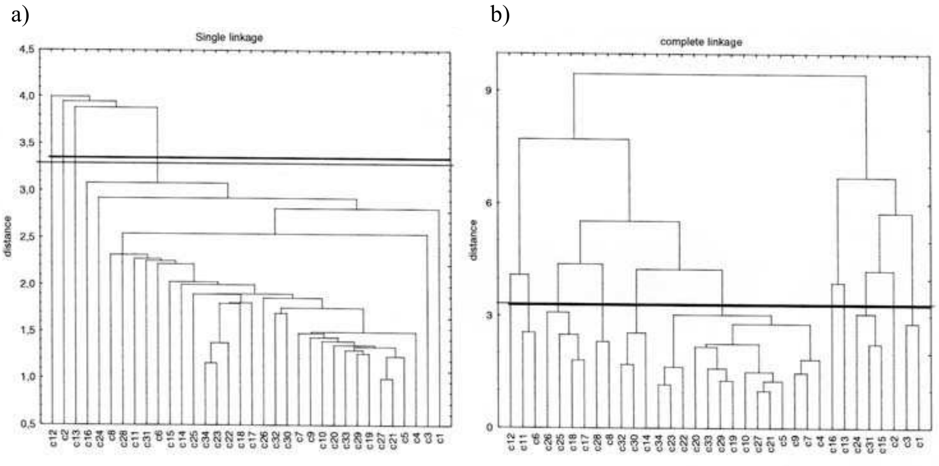

Fig. 1. Agglomeration with use single linkage method (a) and complete linkage method (b)

In a different manner, the agglomerative process in case of complete linkage method is forming (Fig. 1b). For evaluated binder distance, eleven clusters were separated, from which three assemble by one catchment, two with tree catchments, one with four and the last cluster with fourteen catchments. In complete linkage method, similarly like in single linkage method, in case of clusters with only one catchment, it was the same catchments, characterizes catchments with the biggest medium elevation, cross section elevation, drainage area and the highest mean slope from all objects. Fourth cluster contains Skawica river in Skawica gauging section, which is characterized by the biggest value of q_{95} discharge (7.98 $m^3 \cdot s^{-1}$). Next cluster includes rivers: Raba and Wisła, which are characterized by nearing value of: q_{95} discharge (about 3 $m^3 \cdot s^{-1}$), medium elevation (about 570 m o.s.l.), mean slope ($\psi = 0.03$) and precipitation above 950 mm. The sixth cluster contains rivers: Biała Tarnowska and Łęg, which have similar drainage area (above 800 km^2). Rivers Wieprzówka and Łubinka made another cluster, for the sake of similar gauging elevation (an average 250 m o.s.l). In the eight cluster, there are rivers, like: Skawa – station Osielec, Solina – station Terka and Soła – station Cięcina with similar cross section elevation – about 400 m o.s.l. In another cluster the catchments with similar medium elevation (an average 250 m o.s.l.) and land cover by evergreen (about 50%) were included. The tenth group includes rivers with similar as regards soils (the Bołdakow coefficient were on an average level 0.38), precipitation (an average 867 mm) and mean slope ($\psi = 0,026$). The last, eleventh cluster includes remaining fourteen catchments. Average silhouette width for complete linkage method equaled 0.35 what shows weakness of created clusters. The value of the coefficient for each cluster (in case of those where there more than 2 catchments) were varying from 0.15 for the eleventh cluster to 0.45 for the clusters five and seven.

In pair group average method for evaluated linkage distance ($W_k = 3.6$) seven clusters were separated (Fig. 2a). In this method, like in single and complete linkage method, in case of clusters with only one catchment, it was the same catchments. The fifth cluster includes river Skawica in section Skawica and Solinka in section Terka, which are characterized by similar value of specific low flow discharges q_{95} above 5 $m^3 \cdot s^{-1}$, the high value of average annual precipitation, about 1000 mm, similar medium elevation (above 740 m o.s.l.), cross section elevation above 400 m o.s.l. and percentage area covers by evergreen (< 5%). Another cluster includes catchments characterize by specific low flow discharges q_{95}, which range from 2 $m^3 \cdot s^-$ to 4 $m^3 \cdot s^{-1}$, medium elevation varying from 390 tp 720 m o.s.l., mean slope ($0.021 < \psi < 0.055$), cross section elevation to 400 m o.s.l. and an average annual precipitation to 1000 mm. The last, seventh cluster includes remaining rivers characterize by the lowest specific low flow discharges q_{95} below 4 $m^3 \cdot s^{-1}$, medium elevation (Hme < 460 m o.s.l.), mean slope ($\psi < 0.04$), cross section elevation (p.z. < 280 m o.s.l.). Average silhouette width ASW for pair group average method equaled 0,4 what suggests, that created clusters could have an accidental character.

In weighted pair group average method (Fig. 2b) for evaluated linkage distance eight clusters were separated. In this method, like previous methods, in case of clusters with only one catchment, it was the same catchments characterize by depart from other catchments parameters.

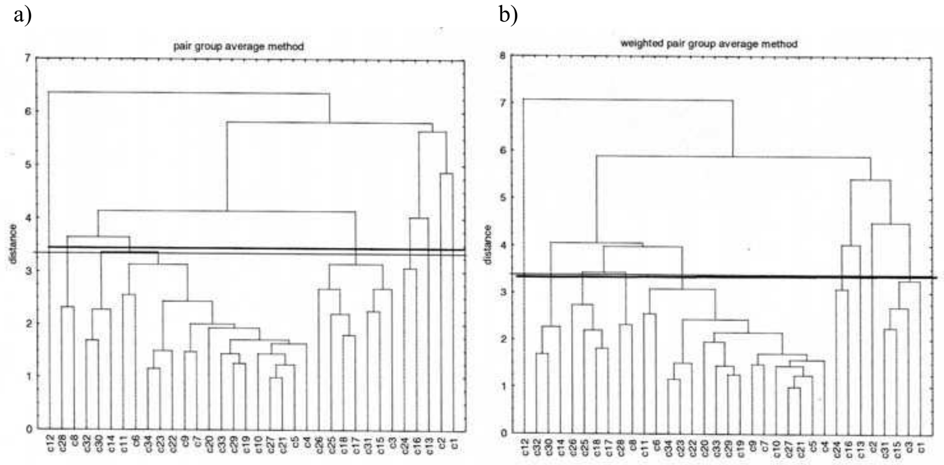

Fig. 2. Agglomeration with use pair group average method (a) and weighted pair group average method (b)

The fourth cluster includes Skawica river in Skawica section and Solina in Terka secion, characterize by similar value of specific low flow discharges $q_{95,}$ the high value of average annual precipitation, medium elevation, cross section elevation and percentage area covers by evergreen. Another cluster includes catchments with similar value of specific low flow discharges q_{95} varying from 3 $m^3 \cdot s^-$ to $4m^3 \cdot s^{-1}$, medium elevation range from 530 to 720 m o.s.l., mean slope ($0.03 < \psi < 0.055$), cross section elevation to 400 m o.s.l. and an average annual precipitation to 1000 mm. The next cluster assembles six catchments similar as regards of the main channel length amount to 34 km and mean slope (about 0.03). The last, eighth cluster includes remaining catchments characterize by the lowest value of: q_{95} (below 3 $m^3 \cdot s^{-1}$), medium elevation (Hme < 250 m o.s.l.), mean slope ($\psi < 0.04$), and cross section elevation (p.z. < 260 m o.s.l.). Average silhouette width ASW for weighted pair group average method equaled 0.4 what suggests, that created clusters could have an accidental character. The value of the coefficient for each cluster were varying from 0.2 for the second cluster to 0.6 for the clusters five, which includes Skawica and Solinka rivers.

In case of Ward's method, similarly like in complete linkage method, for evaluated linkage distance ($W_k = 3.6$) eleven clusters were separated (Fig. 3). In Ward's method, like in case of previous cluster methods, the cluster with only one catchment were characterized catchments with the biggest medium elevation, the highest cross section elevation, the biggest drainage area and the highest mean slope. The fourth cluster includes Skawica and Solinka rivers characterize by similar value if specific discharge equaled above 5 $m^3 \cdot s^{-1}$, high value of annual precipitation, about 1000 mm, similar medium elevation (Hme above 740 m o.s.l.), cross section elevation above 400 m o.s.l. and percentage area covers by evergreen (< 5%). Another cluster includes Biała Tarnowska and Łęg rivers, which are characterized by similar drainage area (above 800 km^2). The next cluster was made by catchments with similar medium elevation (an average for this group 250 m o.s.l.) and percentage area covers by evergreen (about 50%). The eighth cluster was created by catchments characterize by an average soil pervious (Bołdakov coefficient equal about 0.55), an average annual precipitation about 800 mm, mean slope (in range 0.011–0.018). Also the cross section elevation for catchments in this cluster was almost identical and equal about 220 m o.s.l. Another cluster was created by rivers: Wisła, Raba, Skawa and Soła.

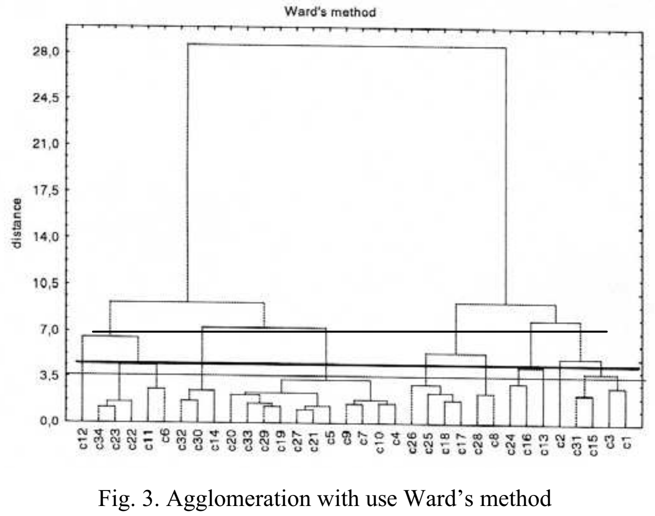

Fig. 3. Agglomeration with use Ward's method

The average medium elevation in this group was 601 m o.s.l., but the mean slope 0.039. The average annual precipitation equal 1034 mm, an average specific low flow discharges q_{95} equal 3.5 $m^3 \cdot s^{-1}$. Catchments in this cluster are also similar regards to percentage area covers by evergreen (about 50%).

The tenth cluster includes rivers: Kamienica Nawojowska, Łososinka, Osława and Zagórz. The Bołdakov coefficient for this cluster was varying from 0.4 to 0.44. Catchments in this group are similar as regards of main channel length (rivers in this cluster have the shortest length from all analyzed catchments, which equal 33 km), mean slope is 0.027. The last, eleventh cluster is characterized by rivers with the lowest value of following parameters: specific discharge is below 3 $m^3 \cdot s^{-1}$, medium elevation < 325 m o.s.l., mean slope ($\psi < 0.015$), cross section elevation below 260 m o.s.l. and annual precipitation equal to 700 mm. Average silhouette width for Ward's method equal 0.63 what means that cluster were created correct and was the highest from all analyzed methods for cluster analysis.

The next two methods of clustering: pair-group method using the centroid average and weighted pair-group method using the centroid average the agglomeration course determinationaly differ from remaining methods. In case of those two methods it was impossible to separate clear clusters of objects, because in both methods dendrites are like „chain" shape.

CONCLUSIONS

On the basis of conducted research the following statements were formulated:

1. The best results of grouping was got with use Ward's method, for which an average silhouette width equal 0,63 and was the highest from all analyzed cluster methods. The coefficient shows, that that cluster were created correct.

2. The lowest average silhouette width ASW (0.3) was got for single and complete linkage methods.

3. The pair-group method using the centroid average and weighted pair-group method using the centroid average should not be used for purpose of hydrological homogenous region determination. In case of these methods the agglomeration course determinationaly differ from remaining methods, what causes problems with clear separation of clusters.

REFERENCES

Hydrological Atlas of Poland. 1987. J. Stachý (ed.), vol. 1 (in Polish). Wyd. Geologiczne, Warszawa.

Chełmicki W., 1991. Location, division and characteristics of basin, in: Upper Vistula river basin I. Dynowska, M. Maciejewski (eds) (in Polish). PWN, Kraków.

Hellwig Z., 1968. Use of tacsonomical method to typological division of Poland for the sake of their development and resources and structure of qualified stuff (in Polish). Przegl. Statys. 4.

Kaya E., Demirel M.C., 2007. A Comparison of Low-Flow Clustering Methods: Streamflow Grouping. J. Eng. App. Sci. 2(3), 524–530.

Królczyk J., Tukiendorf M., 2005. Assesment of the relations between the run of mixing of granular, multi-component and their fractions using the cluster analysis (in Polish), Acta Sci. Pol., Technica Agraria 4(2), 21–30.

Laaha G., Blöschl G., 2006, A comparison of flow regionalization methods – catchment grouping. J. Hydrol. 323, 193–214.

Lin G.W., Wang C.M., 2006. Performing cluster analysis and discrimination analysis of hydrological factors in one step. Adv. Water Resour. 29, 1573–1585.

Mamun A.A., Hashim A., Daoud J.I., 2010. Regionalization of low flow frequency curves for Peninsular Malaysia. J. Hydrol. 381, 174–180.

Pluta W., 1977. Multidimensional comparative analysis in economic research (in Polish). Uniwersytet Śląski, 469 pp.

Rao A.R., Srinivas V.V. 2006. Regionalization of watersheds by hybrid-cluster analysis. J. Hydrol. 318, 37–56.

Števková A., Sabo M., Kohnová S., 2012. Pooling of low flow regimes using cluster and principal component analysis. Slovak J. Civil Eng., 20, 2, 19–27.

REGIONALIZACJA ZLEWNI Z WYKORZYSTANIEM HIERARCHICZNYCH METOD ANALIZY SKUPIEŃ

Streszczenie. W pracy podjęto próbę wyznaczenia obszarów hydrologicznie homogenicznych ze względu na odpływ jednostkowy q_{95} a wybranymi parametrami zlewni. Analizę wykonano z zastosowaniem wybranych metod analizy skupień. Badania przeprowadzono dla 34 zlewni zlokalizowanych na obszarze dorzecza górnej Wisły. Na podstawie przeprowadzonej analizy stwierdzono, ze najlepsze rezultaty grupowania zlewni podobnych ze względu na informację hydrologiczną uzyskana za pomocą metody Warda, w przypadku której średni współczynnik sylwetki ASW wyniósł 0,63. Najniższy średni współczynnik sylwetki ASW uzyskano w metodzie pełnego i pojedynczego wiązania. W przypadku metody środków ciężkości oraz ważonych środków ciężkości, ze względu na trudności z jednoznacznym wyznaczeniem grup obiektów, nie zaleca się ich stosowania do wyznaczenia obszarów jednorodnych.

Słowa kluczowe: przepływ niski, analiza skupień, zlewnia, obszar homogeniczny