Bulletin of the *Transilvania* University of Braşov • Vol. 3 (52) - 2010
Series II: Forestry • Wood Industry • Agricultural Food Engineering

THE INFLUENCE OF THE SIZE OF WATERSHEDS ON THE HIERARCHICAL CLASSIFICATION OF MORPHOMETRIC AND MORPHOHYDROLOGICAL PARAMETERS

I.C. PETRIȚAN[1] I. CLINCIU[1]

Abstract: *The principal component analysis (PCA) allowed us to highlight the primordial part played by morphometric and morphohydrological parameters, which are highly dependant on the size of watersheds, in their hierachical classification according to the weight they contribute to the total variance. We proved as well the existence of a strong correlation between the 14-dimensional space and the two (three) dimensional one as a result of operating with this multivariate analysis method.*

Key words: *principal component analysis, morphometric and morphohydrological parameters, small watersheds.*

1. Introduction

The large number of variables involved in the study and prediction of hydrological measures makes it difficult to construct a hierarchical classification of the importance of these measures with respect to explaining the pursued causal connections, because: (1) obtaining the information for a large number of variables requires great efforts and costs; (2) the use of a large number of variables leads to the diminishing of their significance; (3) the existence probability of intercorrelated variables is high; and (4) when the initial variables are strongly intercorrelated, it is difficult to determine a structural dependence that should highlight the contribution of each variable to forming the variability of the entire space. Given that it is precisely the number of variables that defines the dimension of the analysis space, these are the reasons why we intend to simplify this space, without any important information loss.

Thus this can be achieved by redefining the variables, more precisely by defining other variables lower in number than the original ones, but generating themselves a new reduce ed space, equivalent to the first one in terms of the preserved amount of information. Within the multidimensional analysis, the variables redefining the analysis space are known as “principal components”, while the method for reducing the dimension of the original space borrowed the name of “principal

[1] Dept. of Forest Management and Terrestrial Measurement, *Transilvania* University of Braşov.

component analysis" (PCA) [5], [6]. The main inconvenient of this method is that the new variables introduced in the analysis are most of the times difficult to interpret. They are expressed as linear combinations of the original variables and are characterized by a maximum variability. The first principal component is a normalized linear combination of maximum variance; the second component is a linear combination uncorrelated with the first one, with as high variance as possible, but lower than that of the first one. Thus, by means of the principal component analysis (PCA), the preservation of the information contained in the original causal space can be ensured to a large extent.

The PCA operates by transforming a set of correlated variables into a new set of variables, this time, uncorrelated; in other words, the method attempts to synthesize the initial variables into less latent (non-manifest) variables, called principal components. These can extract the common variability of the variables: the first component extracts the maximum of the variability; the second component extracts the remaining variance and so forth. The variability extracted by each component (the variance of the component) is called an eigenvalue [4], [7]. Applying PCA in order to obtain uncorrelated components can be achieved by means of a technique borrowed from linear algebra, technique which implies generating a rotation matrix by which covariance is reduced to zero. The eigenvalues of the new covariance matrix correspond to the variances of the rotated variables (now called principal components). Correspondingly, these eigenvalues (variances) can be used to determine the weight of a certain principal component to the total variance (numerically expressed by the sum of all eigenvalues). The above-mentioned weight can be interpreted as a measure of the importance of that component, at least in terms of how much it accounts for out of the total information [5].

The aim of the research we undertook was to perfect the application of the methodology regarding the statistical study of morphometric and morphohydrological parameters of small, predominantly forested watersheds, by applying the principal component analysis (PCA) to the stratified data at the level of the hydrographic orders as well to the data unstratified by orders, corresponding to the entire Bârsa Groşetului sub-basin, in order to classify the studied parameters according to the weight they contribute to the total variance.

2. Material and Research Method

In order to achieve the aforementioned objective, we had at our disposal the morphometric and hydrological data regarding the Bârsa Superioară basin, located upstream the Zărneşti town - Braşov County, basin with a surface of 18,780 ha and which was decomposed during previous research [1] into four characteristic sub-basins: Valea Prăpastiei, Bârsa Groşetului, Bârsa lui Bucur and Bârsa Fierului.

In the present study, we dealt only with the data for Bârsa Groşetului sub-basin (the largest one) with a surface of 6,070 ha, where the 304 component watersheds were delineated (in the Strahler system) on 1 : 25,000 scale site plans, with the equidistance between the contour lines of ΔH = 50 m. Given that the hydrographic order (in the Strahler system) could represent a criterion for the stratification of watersheds within the framework of morphometric and hydrological studies performed at the level of statistical populations [1], [2], [3], following the operation of stratification by orders, we identified 234 first order watersheds; 55 second order watersheds; 12 third order watersheds; 2 fourth order watersheds and a fifth order watershed (the entire sub-basin).

For all the 301 component watersheds belonging to the first three orders, the values included in the present paper for research purposes were determined [1] as follows:

- surface (F, in ha): through planimetric operations;
- perimeter (P, in m): on the site plan, using the distance meter or the curvimeter, two or more times;
- length (L_b, in m): analitically, by assimilating the watershed to a rectangle, with both the surface and perimeter equal to those of the watershed;
- Gravelius coefficient (G_r, adimensional): as the ratio of the perimeter of the studied watershed to the perimeter of a hypothetical circular watershed of the same surface;
- average altitude (H_m, in m): as the arithmetic mean of the heights of the strips delimited by two successive contour lines, weighted by their surface;
- average slope (I_b, in %): as the mean of the slopes of the strips delimited by two successive contour lines, weighted by their surface;
- slope average length (B_v, in m): by multiplying by 5.5 the ratio of the surface to length of the hydrographic network;
- hydrographic network length (L_r, in m): by measurement on the site plan;
- hydrographic network density (D_r, in $m \cdot ha^{-1}$): as the ratio of the hydrographic network total length to the surface;
- main bed length (L_a, in m): by measurement on the site plan;
- main bed slope (I_a, in %): by relating the difference in elevation between its two extremities (spring and confluence) to the horizontal length of the main bed;
- concentration time (T_c, in min): by summing up the slope run-off (average) time (estimated by taking into consideration the average slope and the slope average length) and the main bed run-off time (estimated by taking into consideration the average slope of the main bed and its length);
- flood maximum liquid discharge ($Q_{e.1\%}$ in $m^3 \cdot s^{-1}$): by applying the rational formula methodology, for ensuring the 1%, with the torrential rain parameters as approached by Maria Platagea (corresponding to the M4 pluvial mountain area);
- specific maximum liquid discharge ($q_{e.1\%}$ in $m^3 \cdot ha^{-1} \cdot s^{-1}$): by relating the flood maximum liquid discharge to the surface.

We mention that the last three parameters (hereinafter called morphohydrological parameters) were established on the basis of a methodology for small watersheds, mainly covered with forests and meadows, approved by the National Institute of Hydrology and Water Management. The estimates approached in the present study regarding these parameters refer to the simplifying hypothesis of a “standard morphometric” torrential watershed [1], such a watershed being identical with the studied real watershed from the point of view of the morphological configuration (of the morphometry), but from which it differs in terms of the complete absence of soil and vegetal covering (precipitation loss through storage and evapotranspiration is neglected and the lithological substrate is considered as impermeable). Within this schematized approach, we can admit the value $c = 1.0$ for the run-off coefficient.

All the 301 watersheds delineated within the researched sub-basin form a population of watersheds distributed by size orders. Using the values of the researched parameters, we further on proceeded to classify the parameters according to the weight each one contributes to the total variance. This was carried out by means of the principal component analysis, which operates with eigenvectors and eigenvalues. All these statistical analyses were performed by means of the Statistics 7.1 software package.

3. Results and Discussion

3.1. First Order Watersheds

In Table 1 the eigenvalues as well as the percentage of the total variance for each component are shown. One can observe that the first two components explain over 70% of the total variance of the initial data and, thus, they are considered as principal components, while the following 11 components account together for a weight lower than 30%.

Table 1

Eigenvalues, explained variance and cumulated explained variance for the principal component analysis in the case of the first order watersheds (*the first 13 axes*)

Component	Eigen-value	% of the variance	% of the cumulated variance
1	7.671	54.792	54.792
2	2.297	16.407	71.199
3	1.506	10.761	81.959
4	0.837	5.976	87.936
5	0.739	5.281	93.217
6	0.460	3.287	96.504
7	0.257	1.833	98.337
8	0.114	0.817	99.155
9	0.056	0.401	99.556
10	0.035	0.253	99.810
11	0.017	0.121	99.932
12	0.008	0.061	99.992
13	0.001	0.008	100.000

In Table 2 the first two eigenvectors are presented. As they are related to the unit of length, they can be used as coordinates in a bidimensional graph, such as the one represented in Figure 1a. The perimeter, the surface and the maximum liquid discharge had the highest values on the first axis (in absolute value), while the watershed slope and Gravelius coefficient present minimum values. The concentration time as well as the length of the network, watershed and slopes also registers high values on the first axis. The Gravelius coefficient and the hydrographic network density have the highest values on the secondary axis (Axis 2), while the slope length and the concentration time register the lowest values (Table 2). There are only two parameters with positive values registered on both axes (the network density and the specific maximum liquid discharge), whereas the registered value for some other two (the slope average length and concentration time) is negative.

Table 2

Results of the principal component analysis in the case of the first order watersheds (*eigenvectors*)

Variable	PCA Axes	
	1	2
F	-0.954712	0.068702
P_b	-0.968740	0.207059
L_b	-0.884299	0.365822
H_m	-0.365822	-0.326077
L_a	-0.815068	0.433889
L_r	-0.815068	0.433889
D_r	0.665215	0.605131
B_v	-0.737894	-0.466649
G_r	-0.211245	0.673439
I_b	-0.173937	0.423237
I_a	0.429475	0.007670
T_c	-0.884734	-0.403750
$Q_{e.1\%}$	-0.957905	0.109753
$q_{e.1\%}$	0.784868	0.490910

3.2. Second Order Warersheds

As in the previously presented case, over 70% of the total variance is taken over by the first two components and for this reason they are considered the principal components of the system.

From Table 3, it is clear that, although the surface, perimeter and maximum liquid discharge preserve very high values on the first axis (like in the case of the first order), they are nevertheless overtaken by the concentration time which records the highest

Table 3

Results of the principal component analysis in the case of the second order watersheds (eigenvectors)

Variable	PCA Axes	
	1	**2**
F	-0.955338	0.008881
P_b	-0.980168	0.108794
L_b	-0.877314	0.200909
H_m	-0.267691	-0.681632
L_a	-0.809574	0.198672
L_r	-0.844025	0.093973
D_r	0.782377	0.243733
B_v	-0.668840	-0.065624
G_r	-0.441843	0.636847
I_b	-0.102967	0.600014
I_a	0.593350	0.263742
T_c	-0.974629	-0.116613
$Q_{e.1\%}$	-0.942978	0.014860
$q_{e.1\%}$	0.896397	0.224813

value. The average altitude and the watershed slope present minimum values on the first axis. On the secondary axis, the average altitude registers the highest value in absolute expression, while the following parameters: surface, maximum liquid discharge and slope average length have the minimum values. There are three parameters with only positive values (network density, bed slope and specific maximum liquid discharge) and other three with only negative values (average altitude, slope average length and concentration time).

3.3. Third Order Watersheds

Unlike the populations of the first two orders, where the principal components explained together around 70% of the total variance, in the case of the third order population, the first two components (principal components) account for more than 80% of the total variance.

The perimeter, watershed length, surface and concentration time hold, in this order, the maximum values on the first axis, while the watershed slope registers the minimum value. On the second axis, the watershed slope and Gravelius coefficient (absolute value) present the highest values, the hydrographic network length and the perimeter register lower values and the main bed length the minimum one (Table 4). In Figure 1c, the positions of all the studied parameters are represented, relating to the two new variables obtained by applying the principal component analysis. Considering the way the points are distributed in the four quadrants, we can state that half of the total number of the parameters comprise values with different signs on the two axes, while the negative values recorded on both axes are predominant in the case of the other half (4 cases out of 7).

Table 4

Results of the principal component analysis in the case of the third order watersheds (eigenvectors)

Variable	PCA Axes	
	1	**2**
F	-0.955495	0.226230
P_b	-0.989957	0.077853
L_b	-0.969882	-0.141474
H_m	-0.661398	-0.141145
L_a	-0.862772	0.005576
L_r	-0.955856	0.050134
D_r	0.664844	-0.393205
B_v	-0.623331	0.557707
G_r	-0.477348	-0.732130
I_b	0.014149	0.779328
I_a	0.793916	0.469831
T_c	-0.931472	-0.258145
$Q_{e.1\%}$	-0.869190	0.468323
$q_{e.1\%}$	0.966207	0.163056

3.4. Watersheds Unstratified By Orders

If for the first two orders the two axes explain over 70% of the total variance and more than 80% in the case of the third order, the explained percentage is slightly lower than 70% (Table 5) within the framework of

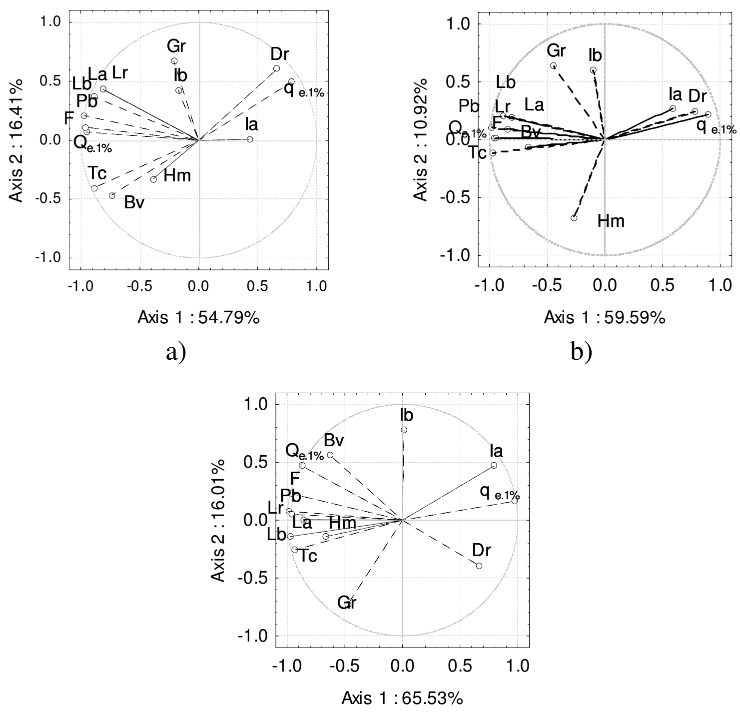

Fig. 1. *Principal component analysis for first* (a)*, second* (b) *and third* (c) *order watersheds*

applying the principal component analysis at the level of the entire sub-basin (without the stratification of watersheds by orders).

Before going on to the examination of the eigenvectors, resulted in the present case (Table 6), we revert to the similar data, previously established and analysed (Tables 2, 3 and 4). By comparing them, we can state that the first axis can be considered as the axis of the size of watersheds; indeed, from all the three tables, one may note that the surface and the perimeter of the watersheds, which are the most significant parameters in terms of expressing the size of watersheds, present values of the eigenvectors that are very close to one (between 0.95-0.98).

This can be interpreted as a strong correlation between the size of watersheds, on the one hand, and the first analysis axis, on the other hand. Implicitly, all the other parameters, which are more or less directly influenced by the watershed size, hold high values of the eigenvectors for the first axis.

Given the above-mentioned remarks, the data included in Table 6 also point out a very interesting fact: namely that, although all the parameters which are directly

Table 5

Eigenvalues, explained variance and cumulated explained variance for the principal component analysis in the case of the entire "Bârsa Groşetului" sub-basin

Component	Eigen-value	% of the variance	% of the cumulated variance
1	7.558	53.991	53.991
2	2.144	15.317	69.308
3…	1.389	9.928	79.236

Table 6

Results of the principal component analysis in the case of the entire "Bârsa Groşetului" sub-basin (eigenvectors)

Variable	PCA Axes	
	1	2
F	-0.3254	0.1456
P_b	-0.3443	0.0598
L_b	-0.3265	0.0742
H_m	-0.0824	-0.3646
L_a	-0.3186	0.1696
L_r	-0.3126	0.2287
D_r	0.1645	0.5317
B_v	-0.1412	-0.5416
G_r	-0.0092	0.1402
I_b	-0.0378	0.0508
I_a	0.2223	-0.0443
T_c	-0.3282	-0.1185
$Q_{e.1\%}$	-0.3277	0.1291
$q_{e.1\%}$	0.2826	0.3066

correlated with the surface present high values of the eigenvectors specific for the first axis, these values are nevertheless much more reduced than those obtained through the stratification of watersheds by hydrographic orders.

For instance, in the case of the surface, perimeter and liquid discharges, the eigenvectors are reduced from 0.95-0.98 to 0.32-0.35, which implies a reduction of more than 60%. We interpret this result as a clear new proof of the necessity to proceed to the stratification of watersheds by hydrographic orders when we consider the populations for the statistical study of morphometric and morphohydrological parameters.

Another aspect that deserves being discussed is the coefficient of determination corresponding to the regression between the distances ordered in bidimensional space (the first two axes in the principal component analysis which explain most of the total variance) and the distances in the original 14 - dimensional space. The measure of the distance for both types of spaces is the Euclidian distance.

Through the analysis of the obtained data regarding this aspect (Table 7), one may remark: (*i*) the existence of a strong correlation between the distances ordered in two (three) dimensional space and those in the original 14 - dimensional space, expressed by values of the coefficient of determination higher than 0.80; (*ii*) the higher the order, the more emphasized this dependance, on the first and third axes; and (*iii*) as for the effect of the stratification by hydrographic orders on the coefficient of determination, we cannot draw a firm conclusion as there is a high degree of closeness between the R^2 values corresponding to each order separately and that obtained for the total (unstratified) population.

Table 7

The coefficient of determination (R^2) corresponding to the regression between the distances in 2(3) dimensional space and those in 14-dimensional space

	Axis 1	Axis 2	Axis 3
First order (n = 13)	0.87	0.84	0.82
Second order (n = 14)	0.85	0.80	0.83
Third order (n = 11)	0.99	1.00	1.00
All orders (n = 14)	0.92	0.85	0.82

4. Conclusions

In order to classify the studied parameters regarding the explanation of the causal connections, we have resorted to "the principal component analysis". By replacing the initial variables (in this case, 14) with only two latent, non-manifest variables (principal components), this multidimensional analysis method can successively extract the variability which is common to all the parameters, indicating the eigenvalues and the percentage of the total variance for each component.

1. In the case of the first and second orders, the variance explained by the first two components is around 70% and it exceeds 80% in the case of the third order. In exchange, the weight of the first two

components in the gear system is below 70% for the data regarding the entire studied sub-basin (Bârsa Groşetului) and unstratified according to the hydrographic order criterion.

2. On the first axis of the analysis system, the surface and perimeter hold the highest values (of the eigenvectors) for all the three orders. Other two parameters, the maximum liquid discharge and the concentration time, present high values for the first two orders. In exchange, the bed slope, in the case of the third order, and the watershed average slope for the second order reach the minimum values.

3. On the secondary axis of the analysis system, the Gravelius coefficient has the maximum values for the first and third order watersheds, while for the second order watersheds this position is held by the average altitude. If for the first order, the concentration time and slope average length register reduced values, in the case of the second order, the minimum values belong to the surface, maximum liquid discharge and slope average length. For the third order, we find the minimum value in the case of the bed length.

4. On the basis of the results provided by the principal component analysis, we can state that the axis of the size of watersheds can be considered the first axis, as the most significant parameters of its expression (the surface and perimeter of watersheds) comprise values of the eigenvectors which are close to one (0.95, respectively 0.98). Implicitly, all the other parameters, which are more or less influenced by the size of watersheds, hold high values of the eigenvectors for the first axis.

5. The reduction of the eigenvalues specific for the first axis, including the parameters directly related to the size of the watersheds, by more than half, when the analysis is performed at the level of orders, represents one more argument in favour of the stratification of watersheds according to the hydrographic order when the operator determines the populations for the statistical study.

6. Finally, as the expression of the distances ordered in two (three) dimensional space by means of the ones in the original 14-dimensional space is characterized by extremely high values (over 0.80) of the coefficient of determination, the statements included in this paper are validated, namely that the reduction of the 14 initial variables to only two principal components did not result in a significant loss of information.

References

1. Clinciu, I.: *Contribuţii la studiul morfometriei şi hidrologiei bazinului hidrografic torenţial Bârsa Superioară* (*Contributions to the Study of the Morphometry and Hydrology of the Bârsa Superioară Torrential Basin).* In: Ph.D. Thesis. University of Braşov, 1983.
2. Clinciu, I.: *Corectarea torenţilor* (*Torrent Correction*). *Transilvania* University of Braşov, 2001.
3. Clinciu, I., Lazăr, N.: *Bazele amenajării torenţilor* (*Bases of Torrent Fitting*). Braşov. Lux Libris Publishing House, 1999.
4. Jolliffe, I.T.: *Principal Component Analysis*. 2nd Edition, Series in Statistics, Springer Verlag New York, Inc., 2002.
5. Lay, D.: *Linear Algebra and its Applications*. New York. Addison-Wesley, 2000.
6. Sachs, L.: *Angewandte Statistik: Anwendung statistischer Methoden - Zehnte überarbeitete und aktualisierte Auflage* (*Applied Statistics: Application of Statistical Methods - Tenth Revised and Updated Edition*). Springer-Verlag Berlin Heidelberg, 2002.
7. StatSoft, Inc.: *Electronic Statistics Textbook.* Tusla, OK, StatSoft. http://www.statsoft.com/textbook/stathome.html, 2007.