FRIEND- Flow Regimes from International Experimental and Network Data (Proceedings of Braunschweig Conference, Octotxr 1993). IAHS Publ. no. 221. 1994

Identification of homogeneous hydrological regional types of basins

Ľ. SOLIN

Institute of Geography, Slovak Academy of Science, Stefdnikova 49, 814 73 Bratislava, Slovakia

Š. **POLÁČIK**

Institute of Archaeology, Slowik Academy of Science, Samova 5, 949 21 Nitra, Slovakia

Abstract By means of methods of numerical classification, homogeneous physical regional types of basins are delimited. On the basis of evaluation of differences in the average annual total runoff existing among them and similarity within them, homogeneous, hydrological regional types of basins are identified.

INTRODUCTION

Research into spatial variability of hydrological response of basins is oriented, above all, to the delimitation of homogeneous hydrological regions (hydrological regionalization) or homogeneous hydrological regional types of basins (hydrological typification). Based on general definition of region and regional types in regional taxonomy (Fischer, 1987), a homogeneous hydrological region is defined as an open system consisting of the neighboring spatial units or basins that show a high degree of similarity from the point of view of hydrological response. Definition of homogeneous hydrological regional type differs only in one thing - the basins belonging to the homogeneous regional types are not the neighboring ones in the geographical space.

If the assumption that basins with similar hydrological response have similar physical properties is valid, homogeneous hydrological types or regions can be delimited not only on the basis of the hydrological response data (Mosley, 1981; Gottschalk, 1985) but also on the basis of data on physical properties of the basins (Acreman & Sinclair, 1986; Wiltshire, 1986). Such an approach creates a necessary prerequisite for the causal explanation of spatial variability of the hydrological response and makes it possible to solve the problem of extrapolating hydrological data.

The use of physical regional types of basins as a source for the identification of homogeneous hydrological regional types of basins confront us with four problems: (a) creating a network of mutually comparable basins; (b) selecting the hydrological response and physical properties of basins; (c) selecting a method for the delimitation of homogeneous physical regional types of basins; and (d) identifying homogeneous hydrological regional types of basins on the basis of the evaluation of the differences in the selected hydrological response existing among, and similarity, within physical regional types of basins.

The identification of homogeneous hydrological regional types of basin in Slovakia on the basis of the above mentioned steps, is the aim of the present contribution.

DELIMITATION OF HOMOGENEOUS PHYSICAL REGIONAL TYPES OF BASINS

The process of delimiting homogeneous physical regional types is done by numerical classification. According to Fischer (1987) it contains the following basic steps: (a) determining the basic spatial units; (b) selecting the properties characterizing spatial units; (c) constructing a matrix of empirical data; (d) transforming the data; (e) selecting measure of similarity and choosing the clustering method; and (f) analyzing the resulting taxonomic structure.

DELIMITATION OF THE SPATIAL UNITS AND CHOICE OF VARIABLES

The basic spatial units are considered to be the basins. The population contains about 1300 basins and it was arranged in such way as to fulfill the following conditions: (a) the area of basins was situated in only one macro form of relief (mountain, basin, hilly land and lowland); and (b) its size did not exceed 300 km^2. The aim of these conditions is to prevent occurrence of a population of basins that might contain either a basin with an extension of about 10 000 km^2 with several macro forms of relief and lot of physical types of natural landscape or a basin with an area about 60 km^2 in one macro form of the relief and with a few physical types of landscape. These basins are not mutually comparable because they represent two different spatial hierarchical levels. 139 basins with hydrological observation were selected from the population. Each basin is characterized by ten variables that are easily detectable from topographical and thematic maps and are supposed to influence the value of selected hydrological response. The variables are as follows: (a) average total precipitation from 1971-1980; (b) permeability of soil- substrate complex; (c) basin area; (d) forest cover; (e) mean slope of basin; (f) mean altitude of basin; (g) maximum altitude of basin; (h) relative altitude of the basin; (j) shape of basin; (k) drainage network density. On the basis of these data a 139 X 1,0 data matrix was created.

DATA TRANSFORMATION

A basic condition of the use of numerical classification procedures based on the principle of Euclidean distance is the orthogonality of the vector columns of the data matrix. This condition can be fulfilled either by principal components analysis or by factor analysis. The method of principal components assumes that the total variance of original variables is wholly explained by principal components themselves. It does not take into account the fact that part of the total variance can be caused by deviations in the measurement of the values of the original variables or by other incidental effects. Factor analysis takes this into consideration. Therefore identifying the links of variables to the extracted common factors provides more information. Clear differentiation of their loading values facilitates better interpretation of factor score values and the resulting classification scheme. For these reasons, transformation of original variables was done by factor analysis. The simple structure of factor matrix was reached by its rotation using the varimax method. The loadings of two extracted factors (F) are shown in Table 1.

Table 1 Factor matrix.

Variable	*F1l*	*F2l*	h^2
Annual precipitation	0.650	-0.06	0.427
Permeability of soil-substrate complex	0.141	-0.48	0.255
Basin area	0.077	0.740	0.554
Forest cover	-0.01	-0.28	0.074
Mean slope of basin	0.474	-0.69	0.704
Mean altitude of basin	0.901	-0.26	0.879
Maximum altitude of basin	0.966	-0.15	0.955
Relative altitude of basin	0.926	0.063	0.862
Shape of basin	-0.04	-0.39	0.156
Drainage network density	0.088	0.043	0.010

The first common factor shows a linkage to the total precipitation, maximum, mean and relative altitude of the basin. It reproduces almost 56% of the common variance. Objectively, it is possible to interpret it as a factor of vertical zonality. From the point of view of interpretation of the second common factor, values of factor loading of the mean slope and area of the basin as well as the permeability of soil-substrate complex and shape of the basin, are significant. The second common factor elucidates 27% of common variance of the original variables. Objectively, it can be interpreted as a factor of azonality.

CLASSIFICATION

The basis of the classification procedure is the expression of the measure of similarity among individual basins. There are several ways to express it. The simplest way is

Euclidean distance expression:

$$D(x_i, x_j) = \sqrt{\sum_{k=1}^{p} (x_{i,k} - x_{j,k})^2} \qquad (1)$$

where:

x_{ik} = value of *k* variable in basin i;

and x_{jk} - value of *k* variable in basin j.

The variables within the Euclidean distance formula are factor score values of extracted factors for each basin. From the existing classification procedures the centroid method is used. It was also applied for instance in Berry's work (1961). This method is based on the distance of means between the clusters. Two clusters are united into one cluster with the shortest distance of the mean. Analysis of the aggregation process on the basis of loosing information showed that the optimum point of termination of the clustering process would be the one set of 139 basins divided into 11 groups representing 11 physical regional types. Original delimitation of the basins into created physical regional types was reclassified by discriminant analysis. However, only one basin fell into each of groups 10 and 11, and such small numbers can bias the remaining analysis. Therefore the basin of group 10 was added to group 4 and that of group 11 to group 9.

The number of basins and arithmetical means of the factor score (FS) of the first and second common factor of identified physical regional types of basins are cited in Table 2. These values show the existence of a certain hierarchical structure. The nine factor score values of the first common factor create four groups in which the factor score values of the second common factor are in descending order. A more concrete picture can be obtained by objective interpretations of the factor score of common factors. For the factor score values of the first common factor the maximum altitude of the basin is decisive and for the second factor the mean slope of the basin is decisive. Substitution of the factor score by arithmetical means of these variables (Table 2) makes possible the following interpretations of the quoted hierarchical structure: in the first group are included physical regional types of basins with the value of arithmetical means of the maximum altitude 821, 826 and 822 m a.m.s.l. and with the mean slope of the basin of 3, 5 and 8% respectively. In the second, third and fourth groups quoted variables have the following values: 1197, 1264 and 1371 m a.m.s.l., 5, 8 and 15%; 1634, 1746 m a.m.s.l. and 5, 9%; 2278 m a.m.s.l. and 14% respectively.

IDENTIFICATION OF HOMOGENEOUS HYDROLOGICAL REGIONAL TYPES OF BASINS

The hydrological response of each physical regional type of basin is expressed as an average annual total runoff. Its regional value was determined from hydrological observation in the period 1971-1980 on the basis of selected groups of basins. If physical regional types are to be homogeneous hydrological regional types as well, then differences in the hydrological response among them indicated by the sample of basins

Table 2 Basic characteristics of delimited physical regional types of basins.

Physical regional type	Number of basins	FS1	FS2	Maximum altitude (m)	Mean slope (%)
1	10	-1.06	1.97	821	2.88
2	18	-1.06	0.66	826	4.68
3	15	-0.99	-0.32	822	8.05
4	13	-0.35	0.97	1197	4.78'
5	34	-0.11	-0.20	1264	7.56
6	14	0.30	-1.22	1371	14.53
7	5	0.61	1.08	1634	5.18
8	10	0.86	-0.30	1746	8.58
9	20	1.82	-0.89	2278	14.07

have to be adequately contrasted. A suitable method forjudging the significance of these differences is analysis of variance. The total variance of average annual total runoff is expressed as follows:

$$\sum_{j=1}^{k}\sum_{i=1}^{nj}(y_{ij}-\bar{y})^2 = \sum_{j=1}^{k} n_j(\bar{y}_j-\bar{y})^2+\sum_{j=1}^{k}\sum_{i=1}^{nj}(y_{ij}-\bar{y}_j)^2 \qquad (2)$$

where:

y_{ij} = value of average annual total runoff of ***i*** basin in ***j*** physical regional types;

$\bar{y}$ = total arithmetical mean of average annual total runoff; and

$\bar{y}_j$ = arithmetical mean of average annual total runoff of ***j*** physical regional type.

The expression on the left side of the equation is the total variance; the first and second expressions on the right side are the variance between and within the physical regional types of basins.

The null hypothesis $\boldsymbol{H_0}$: $\bar{y}_1 - \bar{y} = \bar{y}_2 - \bar{y} = \dots \bar{y}_9 - \bar{y} = 0$ is tested on the basis of the F-ratio. This is the initial step in analyzing the differences of hydrological response among the delimited physical regional types. It only provides information that the differences are significant without specifying which of the differences between physical regional types are sources of this significance. Are all or only some of them significant? The solution lies in mutual couple comparison. In this case the null hypothesis $\boldsymbol{H_0}$: $\bar{y}_1 - \bar{y}_2 = \bar{y}_1 - \bar{y}_3 = \bar{y}_8 - \bar{y}_9 = 0$ is tested. This hypothesis is rejected in the selected level of significance $\alpha_{0,05}$ if zero is not a part of the 95% confidential interval for the difference between the arithmetical means of the two analyzed physical regional types of basins. For the calculation of confidential interval a method of least significant differences was used. Only in those cases in which the zero hypothesis is rejected will physical regional types be homogeneous hydrological regional types of basins as well.

RESULT OF EVALUATION

The F-test confirmed the significance of the differences of the average annual total runoff between the physical regional types of basins. Testing couple differences in the arithmetic means of the average annual total runoff between physical regional types showed, though, that not all differences are statistically significant. Physical regional types of basins are clustered, from the viewpoint of significance of these differences, into four homogenous hydrological regional types, but they are not disjunctive (Table 3). Fulfilling the condition of disjunction requires the classification of the physical regional types of basins situated at the intersection of two hydrological regional types into one of them. Classification was done on the basis of a simple rule, the essence of which will be explained with an example. Physical regional types 3 and 4 are the intersection of the regional hydrological types 1 and 2. Physical regional type 3 was adjunct to the hydro- logical regional type 1 because the difference between the arithmetical means of the average annual total runoff of the physical regional types 3 and 2 is smaller than the difference between 3 and 5. Similarly because of a smaller difference in the values of the average annual total runoff between the arithmetical mean of the physical regional types 4 and 5 than between 4 and 2, the physical regional type 4 was adjunct to the hydrological type 2. Thus the physical regional types of basins 8 and 7 were also classified to the hydrological regional types. After this reclassification of the questionable

Table 3 Non disjunct homogeneous hydrological regional types of basin.

Physical regional type	Number of basins	Total runoff (mm)	Hydrological regional types
1	10	291	1
2	18	324	1
3	15	386	12
4	13	393	1 2
5	34	462	2
6	10	531	23
7	14	616	3
8	5	688	34
9	20	872	4

physical regional types the identified hydrological regional types fulfill the condition of disjunction. Physical regional types of basins 1, 2 and 3, are classified as hydrological regional type I, physical regional types 4, 5 and 8 as hydrological regional type II, physical regional types 6 and 7 as hydrological regional type III and physical regional type 9 as hydrological regional type IV (Table 4).

The significance of the differences of the arithmetical means of the average annual total runoff among identified hydrological regional types of basins was again judged by analysis of variance. By the rejection of each tested null hypotheses at the significance level of$_{003}$ in both cases statistical significance of the originated differences was confirmed.

Assertions based on the tests of null hypothesis can be taken as unbiased only if the selected sets of basins fulfill the condition of being independent random selections. This was verified by the analysis of residuals. If the chosen sets fulfill the quoted condition the residual have normal distribution $N\{0, 6^2)$. The degree of agreement of residual distribution with normal distribution was verified by the x^2 test. The result of the test confirmed a satisfactory agreement of residuals with the normal distribution.

Table 4 Disjunct homogeneous hydrological regional types after reclassification by discriminant analysis.

(a) Original			(b) After reclassification by discriminant analysis		
Type	Number of basins	Total runoff (mm)	Type	Number of basins	TotaI runoff (mm)
1	43	338	1	47	267
2	57	458	2	55	481
3	19	634	3	29	742
4	20	872	4	8	1203

EVALUATION OF SIMILARITY OF HYDROLOGICAL RESPONSE WITHIN HOMOGENEOUS HYDRO- LOGICAL REGIONAL TYPES

Classification of the basins into four hydrological regional types explained 47% of the total variance of the values of the average annual total runoff. The remaining part of variance is caused by the properties of landscape that were not taken into consideration and also partially by deviations in the flow measurement. However, the effect of unconsidered elements and influences can cause the hydrological response value of some basins in hydrological regional type to be too different from its regional arithmetical mean and to be closer to the mean of another hydrological regional type. Justification of the similarity of basins from the point of hydrological response values within identified hydrological regional types was examined by the method of discriminant analysis. A discriminant score on the basis of a discriminant function, whose parameters were determined by the Fisher method (Johnson & Wichern, 1982) was calculated for each basin. The discriminant score of hydrological regional types is expressed as a mean value. The arithmetical means between the mean values of discriminant score of hydro- logical regional types create classification intervals into which individual basins were ordered on the basis of their discriminant score values. Correct ordering of each basin into hydrological regional types was reached after several iterative cycles (Table 4). Reclassification of the basins by discriminant analysis resulted in a reduced variability of the average annual total runoff values in identified hydrological regional types, but on the other hand they became more heterogeneous from the point of view of the physical regional types. According to the map presentation of the homogeneous hydro- logical regional types of basins, it is possible to say that the effect of the first common factor that is interpreted as a factor of zonality is slightly modified by the geographical location of the basin and a karst environment.

REFERENCES

Acreman, M. C. & Sinclair, C. D. (1986) Classification of drainage basins according lo their physical characteristics, an application for flood frequency analysis in Scotland. *J. Hydrol.* 84(3), 365-380.

Berry, B. J. L. (1961) A method for deriving multi-factor uniform regions. *Przeglad Gtograficzny* 33(2), 263-282.

Fischer, M. M. (1987) Some fundamental problems in homogeneous and functional regional taxonomy. *Bremer Beiirage zur Geographic und Rampianung* 11, 267-282.

Gottschalk, L. (1985) Hydrological regionalization of Sweden. *Hydrol. Set. J.* 30(1), 65-83.

Johnson, R. A. & Wichern, D. W. (1982) *Applied Multivariate Statistical Analysis.* Prentice-Hall, New Jersey.

Mosley. M. P. (1981) Delimitation of New Zealand hydrologic regions. *J. Hydrol.* 49(1), 173-192

Wiltshire, S. E. (1986) Identification of homogeneous regions for flood frequency analysis . *J. Hydrol.* 84(3), 387-402.