TELEDETEKCJA ŚRODOWISKA
dawniej „Fotointerpretacja w geografii"
Tom 66 (2026) s. 1–20
ISSN 0071-8076
POLSKIE TOWARZYSTWO GEOGRAFICZNE
ODDZIAŁ TELEDETEKCJI
http://ptgeo.org.pl/teledetekcja/

An attempt to landscape segmentation based on remote sensing LULC classification: a case study of Khmelnytskyi Oblast, Ukraine

Próba segmentacji krajobrazu na podstawie danych teledetekcyjnych i klasyfikacji LULC: studium przypadku obwodu chmielnickiego, Ukraina

Krzysztof BĘDKOWSKI[1], Uliana KOMAROVA[2]
[1] University of Lodz, Faculty of Geographical Sciences, Institute of Urban Geography, Tourism Studies and Geoinformation, Kopcińskiego 31 St., 90-142 Lodz; krzysztof.bedkowski@geo.uni.lodz.pl
[2] University of Lodz, Admission Centre, Uniwersytecka 3 St., 90-137 Lodz; uliana.komarova@uni.lodz.pl

Abstract: Satellite imagery provides an effective tool for large-scale analysis and supports physico-geographical regionalization. In this study, we investigated the possibility of delineating the landscape of the Khmelnytskyi Oblast (Ukraine) into spatial units based on the results of unsupervised classification of Landsat 8 imagery. Five land cover classes were identified: bare soil, forests, shrublands, agricultural crops, and water bodies. The study area was subsequently divided into 6 × 6 km grid cells, and the proportion of each land cover class within each cell was calculated. In addition, three landscape metrics were determined: IJI, DIVISION, and SHDI. In total, five variables describing land cover proportions and three additional variables describing the spatial structure of their distribution within each grid cell were obtained. The analyses were conducted in five variants using five, eight, and two variables. To reduce interdependencies between variables and to obtain an alternative representation of the data, principal component analysis (PCA) was applied, transforming the variables into a set of uncorrelated components. The transformed variables were then used to perform spatial segmentation using a selected segmentation algorithm. The approximate number of 17 segments was not treated as an optimization target, but rather as a reference level enabling comparison between the independently obtained segmentation results and the physical-geographical regionalization of Khmelnytskyi Oblast proposed by Herenchuk (1980). Parameter selection aimed to achieve a comparable level of spatial detail rather than reproduce the reference division. The obtained spatial divisions showed similarity to the reference regionalization. The results were influenced both by the segmentation algorithm parameters and by the number of variables used in the analysis. Full agreement between our segmentation and the reference regionalization was not possible because the reference regionalization also considered physical-geographical features not reflected in the structure and texture of the satellite imagery. We demonstrated that segmentation in the proposed approach, especially when supplemented with additional data, can be used for the delineation of spatial landscape units.

Streszczenie: Obrazy satelitarne stanowią skuteczne narzędzie do analiz wielkoobszarowych oraz wspierają regionalizację fizycznogeograficzną. W tym studium badaliśmy możliwość wykonania podziału krajobrazu Obwodu Chmielnickiego (Ukraina) na jednostki przestrzenne na podstawie wyników nienadzorowanej klasyfikacji obrazów Landsat 8. Zidentyfikowaliśmy pięć klas pokrycia terenu: gleby bez pokrywy roślinnej, lasy, zakrzaczenia, uprawy rolne oraz wody. Następnie analizowany obszar podzielono na kwadraty o wielkości 6 × 6 km i określono udział każdej z klas pokrycia terenu. Ponadto wyznaczyliśmy trzy metryki krajobrazowe: IJI, DIVISION oraz SHDI. Łącznie uzyskaliśmy pięć zmiennych opisujących udziały form pokrycia terenu oraz trzy dodatkowe zmienne opisujące strukturę przestrzenną rozmieszczenia tych klas w każdym z kwadratów. Analizy prowadziliśmy w pięciu wariantach, z wykorzystaniem pięciu, ośmiu oraz dwóch zmiennych. W celu zmniejszenia współzależności między zmiennymi oraz uzyskania alternatywnej reprezentacji danych zastosowano analizę głównych składowych (PCA), przekształcając zmienne do postaci nieskorelowanych komponentów. Przekształcone zmienne wykorzystaliśmy do wykonania podziału

przestrzeni za pomocą wybranego algorytmu segmentacji. Przybliżona liczba 17 segmentów nie była celem optymalizacji, lecz punktem odniesienia umożliwiającym porównanie wyników niezależnie uzyskanej segmentacji z regionalizacją fizyczno-geograficzną obwodu chmielnickiego według Herenchuka (1980). Dobór parametrów miał na celu osiągnięcie porównywalnego poziomu szczegółowości przestrzennej, a nie odtworzenie podziału referencyjnego. Uzyskaliśmy podziały przestrzeni wykazujące podobieństwo do regionalizacji referencyjnej. Wpływ na uzyskiwane wyniki miały zarówno parametry algorytmu segmentacji, jak i liczba zastosowanych zmiennych. Uzyskanie pełnej zgodności między naszą segmentacją a regionalizacją referencyjną nie było możliwe, ponieważ regionalizacja ta uwzględniała również cechy fizycznogeograficzne niewidoczne w strukturze i teksturze obrazu satelitarnego. Wykazaliśmy, że segmentacja w zaproponowanym przez nas ujęciu, a także wzbogacona o inne dane, może być wykorzystywana do wyodrębniania przestrzennych jednostek krajobrazowych.

Keywords: remote sensing, landscape metrics, principal components analysis, image segmentation

Słowa kluczowe: teledetekcja, metryki krajobrazowe, analiza głównych składowych, segmentacja obrazu

Introduction

Owing to the wide availability and high analytical potential, satellite imagery has become a fundamental source of information in environmental research and regionalization studies. Continuous improvements in spatial, temporal, radiometric, and spectral resolution enable analyses across various fields, including geology, oceanography, forestry, and agriculture. Remote sensing enables the analysis of environmental processes from a different perspective compared to traditional methods, without direct contact with the study area, thereby facilitating the interpretation of landscapes as integrated systems of natural components. (Haglauer, 1996; Olędzki, 2007). However, it should be noted that not all landscape elements can be fully identified or reliably interpreted based solely on satellite data.

Satellite imagery can also be used as a data source for physico-geographical regionalization. It may serve as the primary source of information for conducting analyses, but can also be integrated with other datasets (Komarova, Będkowski, 2025; Nita et al., 2016; Olędzki, 2007).

In physico-geographical regionalization, the concept of a patch is used to represent the spatial structure of the landscape. A patch represents an area of the landscape with relatively homogeneous, according to the adopted criteria, features of land cover and spatial structure (Komarova, Będkowski, 2025). The arrangement of patches forms a mosaic that reflects the spatial organization of the landscape.

Landscape metrics are quantitative indices used to describe the spatial structure, composition, and configuration of landscape elements, as well as the relationships between them (McGarigal and Marks, 1995; Turner et al., 2001). Landscape metrics are calculated based on a thematic map representing the landscape as a mosaic of patches classified into different categories.

The literature provides numerous examples of the application of landscape metrics in environmental studies, including analyses of land use change, habitat fragmentation, and landscape heterogeneity.

Landscape indices are widely used to assess spatiotemporal changes based on multi-temporal satellite data, such as analyzing the fragmentation of urban green spaces resulting from rapid urbanization processes (Zabihi et al., 2025) or evaluating changes in landscape patterns to support sustainable regional planning (Dadashpoor et al., 2019). Landscape metrics are also widely applied across various environmental fields, including studies on soils (Wei et al., 2025; Wang et al., 2024; da Silva et al., 2015), deforestation (Frohn and Hao, 2006), and hydrology (Jeitany et al., 2024).

At the same time, many studies focus on the properties and limitations of the metrics themselves, examining their sensitivity to scale (Zhang and Li, 2013), the size of the analytical unit (Saura and Martínez-Millán, 2001), thematic resolution, and inter-metric correlations (Szabó et al., 2012).

Image data analysis often employs advanced segmentation algorithms that analyze the spatial variability of spectral responses recorded in individual spectral channels (Lasalle et al. 2015). We reasoned that the same algorithms could be applied to spatial segmentation based on data of a slightly different nature. In this study, we investigated the feasibility of dividing the landscape of Khmelnytskyi Oblast (Ukraine) into spatial units based on unsupervised classification of Landsat 8 imagery. We used variables describing the share of various land cover forms and landscape metrics describing their spatial structure. We used these data to segment the landscape using the segmentation algorithm implemented in TerrSet® software (Clark 2025). The research flow diagram is presented in Figure 1.

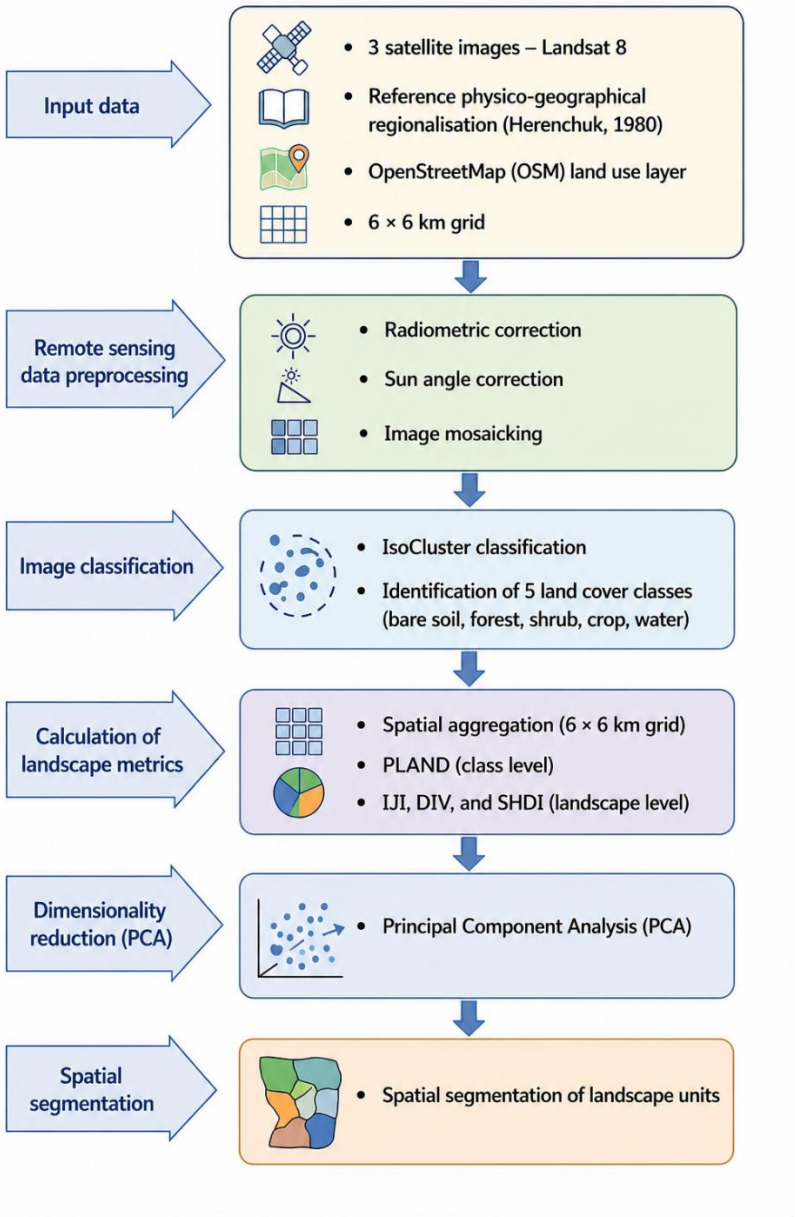


Fig. 1. Research workflow diagram. Source: own elaboration

Materials and Methods

The Area of interest

The study area concerns the Khmelnytskyi Oblast, located in western Ukraine within the southwestern part of the East European Plain. The region covers approximately 20,629 km² and is situated between 48°26′–50°35′ N and 26°08′–27°54′ E.

Geologically, it lies within the transition zone between the western margin of the Ukrainian Shield and the Volyn–Podolian Plate. The terrain is predominantly lowland with an average elevation of about 275 m a.s.l., although local variations in relief are evident. The highest elevations occur in the central part of the region (Verkhobuzka Upland), reaching up to ~380 m, while northern areas (Shepetivka Plain) are lower and characterized by sandy and loamy soils. Toward the south, the landscape becomes more dissected, particularly within the Dniester Foreland, where fluvial processes have created a system of deeply incised valleys. A distinctive geomorphological feature is the Tovtry Ridge, composed of reef-origin formations rising above the surrounding terrain.

In the Khmelnytskyi Oblast, landscape types exhibit a clear north–south gradient, with the delineation and number of zones varying depending on the methodological approach adopted by different authors. In this study, the division proposed by Herenchuk (1980) is adopted (Fig. 7). A detailed description of this classification and the short characteristics of individual landscape types are provided in Komarova and Będkowski (2025).

Satellite images and their processing

Three Landsat 8 satellite scenes were used in this study, two of which were recorded on 2019-08-21 and one on 2019-08-28, scenes 183025, 183026 and 184025, respectively. Their analysis was based on colour composites known as CIR (Color Infrared), where the near-infrared band (Band 5) is displayed in red, the red band (Band 4) in green, and the green band (Band 3) in blue. This type of composite is particularly useful when vegetation constitutes the main focus of the analysis. In the literature dedicated to the theoretical aspects of physical-geographical regionalization, vegetation is often considered an integrative indicator reflecting the combined influence of various environmental factors shaping the landscape (Richling, Solon 2011; Kondracki, 1976). Therefore, CIR composites were employed in the subsequent stages of the analysis.

Prior to the main analysis, radiometric correction was applied (Jakomulska and Sobczak, 2001; Osińska-Skotak, 2007). First, the dimensionless DN values were converted to spectral radiance, according to formula (1):

$$L_{sat} = c_0 + c_1 DN \tag{1}$$

The values of the unique calibration constants c_0 and c_1 (offset and gain) were read in the metadata files of the respective satellite scenes (Landsat 2024).

Subsequently, corrections accounting for the Sun's position at the time of image acquisition were performed. Given the considerable north–south extent of the study area (over 200 km), this step was deemed essential before proceeding with further analysis (Komarova and Będkowski, 2025). The calculation of the corrected radiance value was made according to formula (2):

$$L_{sun} = L_{sat} * sinSunElevation$$

The SunElevation value was taken from the metadata of each satellite scene.

Since the study area is covered by three satellite scenes, so it was necessary to create a mosaic of images before proceeding with subsequent analyses.

The next step involved the classification of the resulting mosaic. All attempts to perform supervised classification based on training samples produced unsatisfactory results, as land cover classes from different parts of the study area (southern, central, and northern) were frequently confused and aggregated into a single category. Owing to the considerable spatial extent of the study area (over 200 km in the north–south direction), its heterogeneity, and the lack of suitable reference data for supervised approaches, an unsupervised IsoCluster classification (ArcMap Desktop, 2024) was applied.

During the parameterization of the unsupervised classification, different values of the number of clusters (k) were tested. For $k < 10$, excessive spectral fragmentation of land cover classes was observed, with pixels representing the same land cover type being assigned to multiple clusters. This effect was particularly pronounced in the southern and northern parts of the study area. In contrast, for $k \geq 20$, the resulting clusters were overly fragmented (over-segmentation), which hindered their interpretation and further analysis at the scale of the entire study area. Therefore, an intermediate value of $k = 16$ was selected as a compromise between spectral separability and the interpretability of the classification results. Subsequently, based on the analysis of cluster spectral characteristics and their comparison with the source data, a reclassification step was performed, in which the clusters were merged into five final land cover classes. To identify the existing land use classes, OpenStreetMap data - specifically the Landuse layer - were analyzed. However, due to its incomplete spatial coverage (approximately 45% of the Khmelnytskyi Oblast), the dataset was treated as auxiliary information. The layer contained features assigned to 20 land use categories, which were subsequently reclassified in terms of land cover, corresponding to the objective of the image classification. Based on this analysis, five land cover classes were distinguished: (1) bare soil, (2) forest, (3) shrubs, (4) crop, and (5) water bodies (Fig. 2). Within built-up areas, pixels representing urban features exhibited a heterogeneous, mosaic-like pattern, which prevented their unambiguous classification as a distinct land use class. Despite numerous attempts to delineate built-up areas as an independent class using supervised classification, the results remained unsatisfactory. Consequently, a separate built-up class was not included in the final classification scheme.

Due to the large extent of the study area and the incomplete spatial coverage of the OSM land use layer, a dedicated sampling approach was developed to ensure that classification accuracy was assessed using validation points distributed across different parts of the study area.

Random validation points were generated within the area of interest and spatially filtered using OSM data, retaining only points intersecting reference polygons. Raster-based land cover classes were assigned to these points using the *Extract Values to Points* tool, while reference classes were attributed via a *spatial join (point-in-polygon)* in ArcMap. The resulting paired dataset was used to construct the confusion matrix (Table 1) and assess classification accuracy.

Table 1. Error matrix of the classified land cover map.

	Bare soil	Forests	Shrubs	Crop	Water	Σ classification
Bare soil	**29**	0	3	12	0	44
Forests	0	**27**	2	7	1	37
Shrubs	2	3	**22**	4	1	32
Crop	10	6	7	**42**	0	65
Water	0	2	0	0	**12**	14
OSM (ref)	41	38	34	65	14	**192**

The accuracy assessment of the unsupervised classification indicated an overall accuracy of 68.75%. The relatively low performance was mainly associated with confusion between bare soil and crop classes. In the classification process, these categories were separated based on spectral pixel values, whereas the reference dataset (OSM Landuse layer) included only the crop class

Additional misclassifications were observed between forest, shrubland, and crop classes, reflecting spectral similarities among these categories.

As a substantial number of dark pixels representing water were initially misclassified as forest, all water bodies were manually verified and corrected in ArcGIS Pro following the classification procedure and prior to the accuracy assessment. Consequently, the water class achieved the highest accuracy as a result of this post-classification correction.

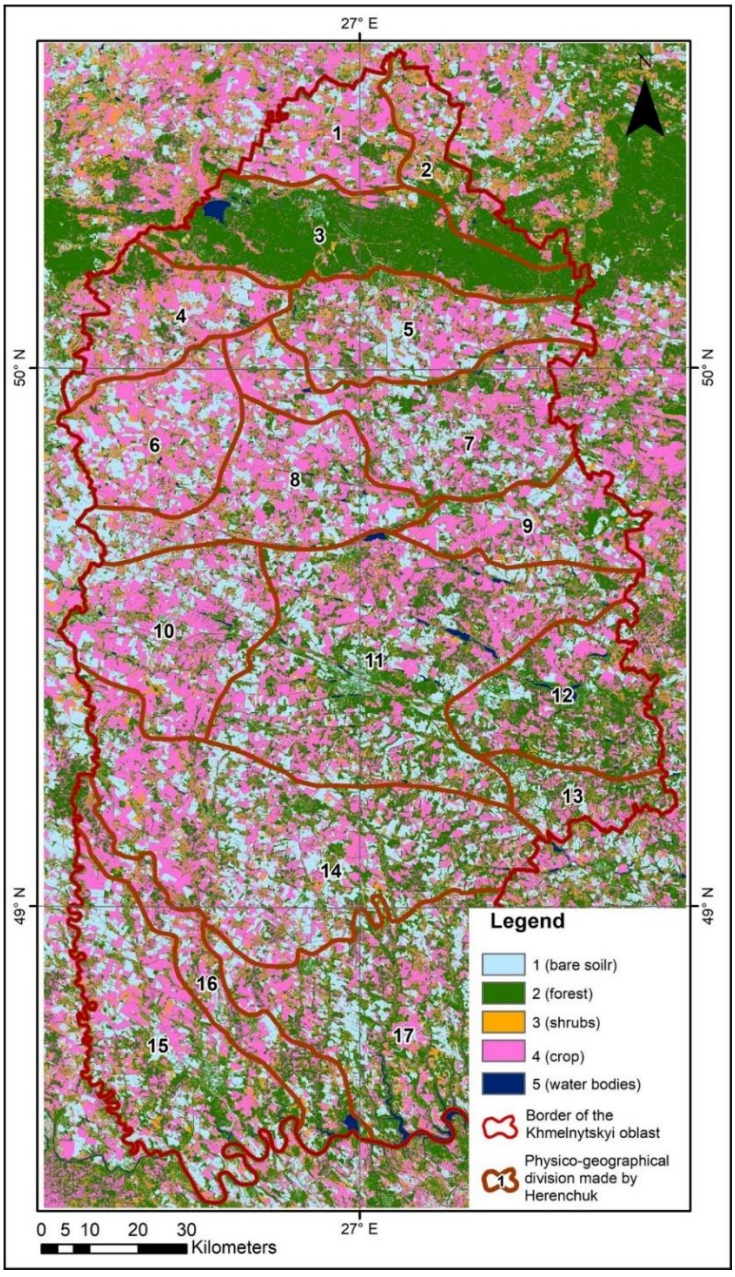


Fig. 2. The result of unsupervised classification of images of Khmelnytskyi oblast. Source: (Komarova, Będkowski 2025)

Landscape characteristics

Landscape metrics were calculated based on patches representing individual land cover classes using the FRAGSTATS software (https://fragstats.org/index.php).

For the purpose of the analysis, the study area was divided into 920 squares of 6 × 6 km. The choice of spatial scale is inherently a matter of debate, as there is no single, universally optimal size for this type of analysis. When selecting the grid size, we sought to avoid units that were too small, in which the landscape within a given cell would often lack sufficient heterogeneity, and at the same time, units that were too large, which would reduce the level of detail and obscure local variability. The chosen size therefore represents a compromise that allows for a significant level of landscape diversity while maintaining adequate resolution for analysis. Landscape analyses of the study area using different grid sizes can also be found in Komarova and Będkowski (2025).

The selected grid size corresponds to the 30 m spatial resolution of Landsat imagery, ensuring that pixels can be fully allocated to individual grid cells without being intersected by grid boundaries. For each square, the proportion of five land cover classes was calculated (PLAND metric calculated on the class level): *Bare soil*, *Forest*, *Crop*, *Shrub*, and *Water*. These indicators sum to unity.

Additionally, three further metrics— Interspersion and Juxtaposition Index (*IJI)*, Landscape Division Index *(DIVISION)*, and Shannon's Diversity Index (*SHDI)*—were introduced and calculated on the landscape level, with the aim of describing the spatial structure of the landscape within each square. The Interspersion and Juxtaposition Index, also reflecting the degree of aggregation, is based on the adjacency relationships between patches of different classes. The DIVISION is a measure of aggregation and describes the probability of an event in which any two base cells (pixels of a raster image) are not within the same patch (McGarigal and Marks, 1995). The Shannon's Diversity Index is one of the indicators that determines landscape diversity. It takes into account two factors: the number of land cover classes in a given landscape and their mutual proportions of landscape structure and spatial patterns (Komarova and Będkowski, 2025).

The metrics selected for this study describe both the composition and configuration of the landscape, taking into account its diversity and degree of fragmentation. Mean values of the variables and basic measures of their variability are presented in Table 2.

Table 2. Characteristics of variables describing the landscape in 920, 6 × 6 km squares

Variable	Mean	Stand. Dev.	Min.	Max.
Bare soil	21.79	12.37	0.05	70.38
Forest	35.44	18.38	3.82	97.45
Shrub	9.55	4.06	0.41	30.20
Crop	32.08	14.77	0.69	74.81
Water	1.14	3.10	0	29.91
IJI	71.04	6.92	31.67	96.76
DIVISION	0.83	0.17	0.05	0.98
SHDI	1.20	0.21	0.14	1.53

To achieve the research objectives, it is important that the proposed variables describing the landscape have a high explanatory value (Parysek, Ratajczak 2002), i.e., that each of them brings new information relevant to the identification of landscape units. To assess the variables from this perspective, a matrix of correlation coefficients between the original variables was calculated (Table 3). It is easy to notice that there is high collinearity between some variables, e.g., *SHDI* and *DIVISION*, and *Forest* and *DIVISION*, which indicates that they contribute similar information to the model of observed relationships.

Table 3. Correlation matrix of variables describing the landscape in 920, 6 × 6 km squares

Variable	Bare soil	Forest	Shrub	Crop	Water	IJI	DIVISION	SHDI
Bare soil	1							
Forest	**-0.5379**	1						
Shrub	-0.0042	-0.3284	1					
Crop	-0.1481	-0.7127	0.1605	1				
Water	-0.0908	0.0458	-0.1115	-0.1599	1			
IJI	0.4656	-0.5507	0.2534	0.1936	0.1521	1		
DIVISION	0.5000	**-0.7252**	0.4376	**0.3499**	0.0628	0.5711	1	
SHDI	0.4593	-0.6588	**0.4787**	0.2569	**0.2211**	**0.6156**	**0.9341**	1

Principal Components Analysis (PCA) was used to transform the original variables into an alternative set of

uncorrelated variables (Ratajczak 2003, Gramacki J, Gramacki A 2008). In this study, PCA was applied primarily as an exploratory tool and as a method of data representation aimed at reducing interdependencies among variables, rather than as a strict dimensionality reduction technique. The PCA analysis allowed for the determination of a set of new variables, C1, C2, …C8, which are not correlated with each other (Tables 4 and 5).

For landscape segmentation a selection of five sets of variables was performed: (A) variables 1–5 (Bare soil, Forest, Shrub, Crop, Water); (B) variables 1–8 (Bare soil, Forest, Shrub, Crop, Water, IJI, DIVISION, SHDI); (C) PCA components 1–5 derived from variables 1–5; (D) PCA components 1–8 derived from variables 1–8; and (E) PCA component C1 derived from variables 1–5 combined with PCA component C1 derived from variables 6–8. The number of new variables, so-called components, is equal to the number of source variables. The new variables are ordered in descending order, which means that they explain successively less and less of the total variability of the variables.

It should be emphasized that the variables representing land cover class proportions have a compositional nature, being constrained by a constant-sum condition. As a consequence, they are inherently interdependent, and part of the observed collinearity may result from this structural property rather than from substantive relationships between variables. Therefore, the results of correlation analysis and PCA should be interpreted with this limitation in mind. In this study, these methods were applied primarily in an exploratory and descriptive context, aimed at identifying dominant patterns of variability rather than performing strict statistical inference.

Table 4. Eigenvalues of the correlation matrix arranged in descending order (5 variables)

	C1	C2	C3	C4	C5
% Variance	39.57	23.35	19.66	17.42	0
Eigenvalue	1.98	1.17	0.98	0.87	0

Table 5. Eigenvalues of the correlation matrix arranged in descending order (8 variables)

	C1	C2	C3	C4	C5	C6	C7	C8
% Variance	47.48	16.92	13.35	11.68	6.04	3.96	0.57	0
Eigenvalue	3.80	1.35	1.07	0.93	0.48	0.32	0.05	0

The distributions (clouds) of points representing 920 squares in the space reduced to the first two components determined by PCA are presented in Fig. 3 and Fig. 4. It can be seen that the clouds are clearly stretched along the axis of the first component *C1*, which describes 39.57% of the total variability, while component *C2* describes 23.35% of the variability (Fig. 3, variant for five variables) and 47.48% and 16.92%, respectively (Fig. 4, variant for eight variables). Taking into account eight variables in the calculations led, compared to the five-variable variant, to a clear strengthening of the first component, while the second component was weakened in favor of the subsequent components (see Tables 4 and 5). From both figures, it can be concluded that the squares are similar in terms of the variables describing them, as well as whether they form clusters, which may indicate the existence of larger areas distinguished by similar landscape properties. Such clusters are clearly visible in Fig. 4. However, it should be remembered that Figs. 3 and 4 show the similarity of squares in terms of the values of the variables describing them and are not related to their spatial location in the Khmelnytskyi oblast.

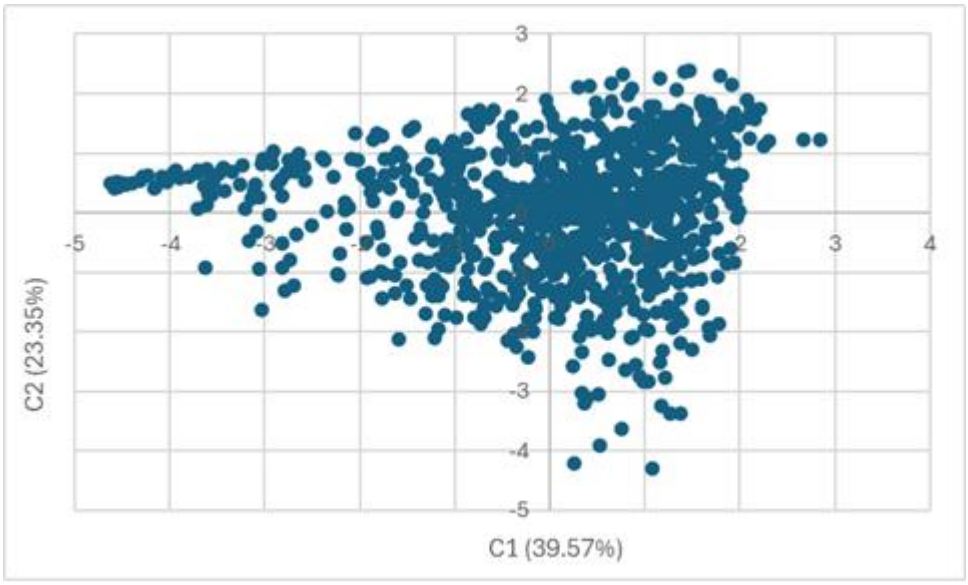


Fig. 3. Squares (920) in the coordinate system of the first two principal components C1 and C2 (in the 5-variable variant)

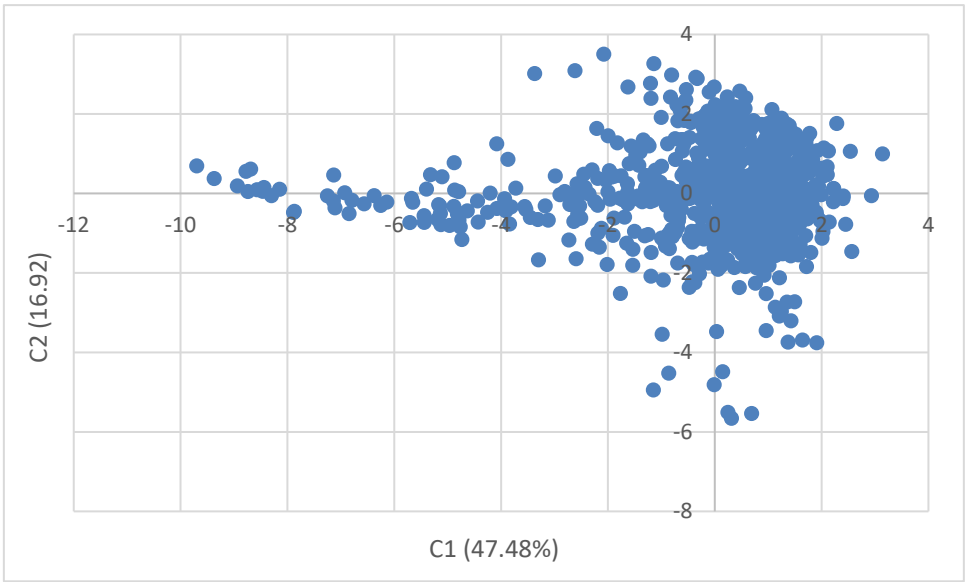


Fig. 4. Squares (920) in the coordinate system of the first two principal components *C1* and *C2* (in the 8-variable variant)

To assess the extent to which the original variables influenced the components, so-called Loadings matrices were used (Tables 6 and 7). The Loadings matrices (first two columns) provide the source of information for the graphs depicting the relationships between the original variables and the first two components (Figs. 5 and 6). Each column of Tables 5 and 6 displays the coefficients with the highest absolute value, indicating which of the original variables are most strongly associated with the new variables, i.e., the components. In other words, which features of the squares most influenced the individual components.

Table 6. Matrix of correlation coefficients, so-called Loadings, of original variables and principal components (in the 5-variable variant)

	C1	C2	C3	C4	C5
Bare soil	0.4366	**-0.8928**	-0.1091	0.0214	0.0019
Forest	**-0.9613**	0.1604	-0.2165	0.0580	0.0001
Shrub	0.4745	0.2765	-0.1653	**0.8192**	**-0.0366**
Crop	**0.7548**	0.5002	0.2127	-0.3672	0.0102
Water	-0.2629	-0.1337	**0.9229**	0.2475	-0.0090

Table 7. Matrix of correlation coefficients, so-called Loadings, of original variables and principal components (in the 8-variable variant)

	C1	C2	C3	C4	C5	C6	C7	C8
Bare soil	0.5788	-0.4494	**-0.6334**	-0.0402	-0.1094	-0.2192	0.0046	0
Forest	**-0.8850**	-0.2622	0.1159	0.2689	0.0544	**0.2435**	-0.0043	0
Shrub	0.4936	0.2554	0.3016	**0.7296**	0.0805	**-0.2473**	-0.0133	0
Crop	0.4695	**0.7545**	0.1553	-0.4310	0.0175	-0.0099	-0.0100	0
Water	0.0525	-0.5825	**0.7070**	-0.3359	-0.0748	-0.1976	-0.0233	0
IJI	0.7377	-0.2643	-0.0077	-0.0889	**0.6064**	0.1011	-0.0118	0
DIVISION	**0.9223**	-0.0614	0.0475	0.0837	-0.2375	**0.2432**	-0.1443	0
SHDI	**0.9050**	-0.1879	0.1902	0.1210	-0.1777	0.1995	**0.1530**	0

In principle, each component can be assigned one of the original variables (or, exceptionally, more, such as *Forest* and *Crop* for *C1*, Table 6, or *Forest*, *DIVISION*, and *SHDI* for *C1*, Table 7). The components located in the rightmost parts of the tables explain only the residual values of the overall variability and show practically no correlations with the original features. The strong correlations demonstrated between *C1* and more than one of the original variables indicate that they may have a similar impact on this component. This is particularly true for variables from the Landscape metrics group, i.e., *IJI*, *DIVISION*, and *SHDI*, which occupy a similar position in Fig. 6.

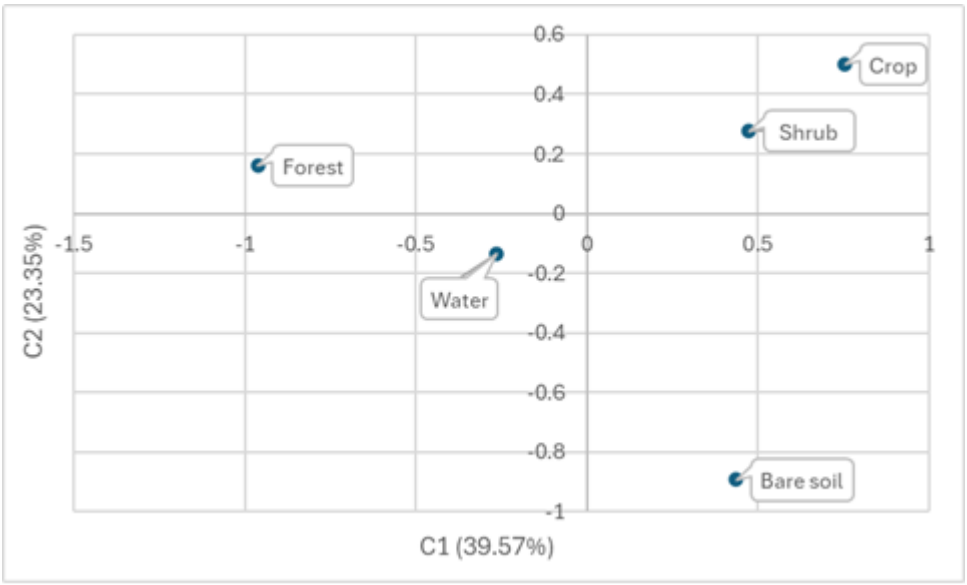


Fig. 5. Loadings describing the relationships between the original variables and the principal components C1 and C2 (five-variable variant)

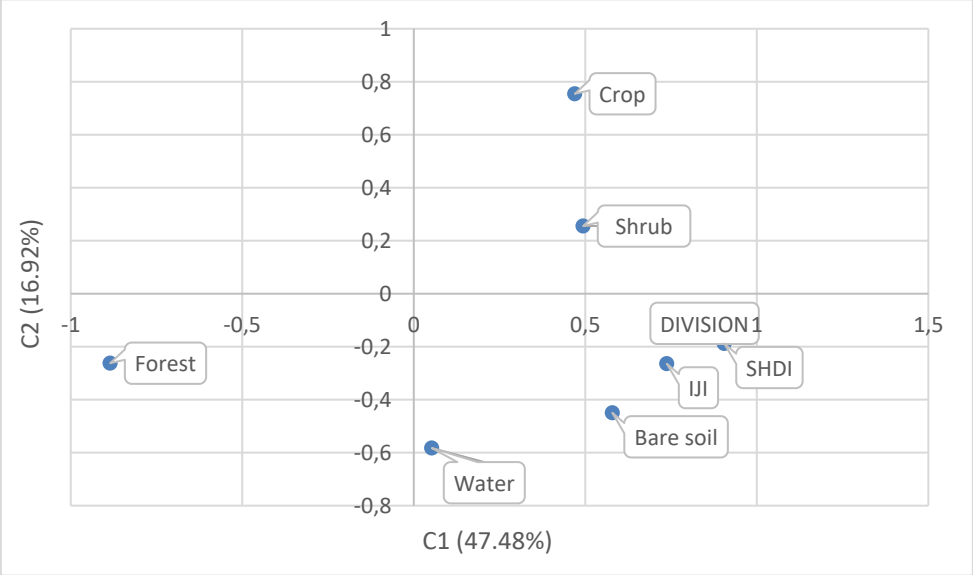


Fig. 6. Loadings describing the relationships between the original variables and the principal components C1 and C2 (eight-variable variant)

The comparison of both figures shows that the factors that most influence the first component, i.e., *C1*, are the variables *Forest* and *Crop* (five-variable variant, Fig. 5), and in the eight-variable variant (Fig. 6), these factors are *Forest* and three Landscape metrics. *Crop* and *Shrub*, as well as *Water* and *Bare soil*, are most closely related to the second components, i.e., *C2*. It should also be noted that the close location of the points on the graphs, resulting from the Loadings values, indicates a similar importance of the variables they represent. Furthermore, the location on opposite sides of the C2 axis indicates a negative correlation between them (an increase in the value of one variable is associated with a decrease in the other).

Given that the analyzed variables describing the proportions of individual land cover classes have the characteristics of compositional data, i.e., they are constrained by a constant-sum condition and are mutually dependent, an additional analysis was conducted to reduce the dimensionality of the data. For this purpose, the first principal component (PC1) was derived for five variables representing land cover class proportions, and separately the first principal component was derived for three landscape metrics (IJI, DIVISION, SHDI).

The resulting components were intended to synthesize the information contained in the original variables and to reduce the issue of interdependence arising from their structure. Subsequently, the relationship between the obtained components was examined, revealing a moderate correlation (r = 0.66), which suggests that they describe partially different but not entirely independent aspects of landscape structure. Consequently, test segmentation variants based on these components were performed (Table 8: Variant E); however, they did not show significant differences compared to the results obtained using the full set of variables.

Landscape segmentation

Landscape segmentation, or the division of the Khmenytsky Oblast into areas with similar variable properties describing 6 × 6 km squares, was performed using an algorithm implemented in TerrSet® software (Clark 2025). The segmentation algorithm takes into account the values of the variables describing the squares, as well as the relative positions of the squares. To be included in the same segment, squares must be adjacent to each other and meet similarity criteria based on variance image analysis, which is an image of the spatial differentiation of variables (TerrSet® Help System). The workflow of the segmentation algorithm is as follows:

- Computing an image representing the variance distribution (variance image)
- Delimiting segments using the watershed algorithm
- Merging adjacent segments if they meet specified criteria.

A significant feature of the algorithm is that it allows for the introduction of weights describing the importance of individual variables (*C1*, *C2*, …*Cn*, the sum of weights = 1). Furthermore, weights indicating the importance of the *mean* and *variance* values (the sum of weights = 1) can be selected, as well as the *Similarity* parameter value. Its value significantly determines the number of segments distinguished. It is recommended to experiment with a starting value of *Similarity* = 10, and then, after evaluating the results, adjust this parameter accordingly. The algorithm makes a decision for each square by analyzing values in a square window, sliding column by column and row by row within the analyzed area.

It should be noted that the application of PCA in this study does not aim at strict dimensionality reduction, as all resulting principal components were retained for further analysis. Instead, the transformation was used to

reduce interdependencies among the original variables and to obtain an alternative representation of the data that may better reflect underlying landscape structure in the context of segmentation. This approach was also treated as a test of whether PCA-based components lead to substantially different segmentation outcomes compared to the use of the full set of original variables. In the calculations, the same weight values were assumed for each variable (0.2 and 0.125, respectively, in the calculation variants based on 5 or 8 components). The same weights were assumed for the *mean* and *variance* values, determined in the segmentation procedure (0.5).

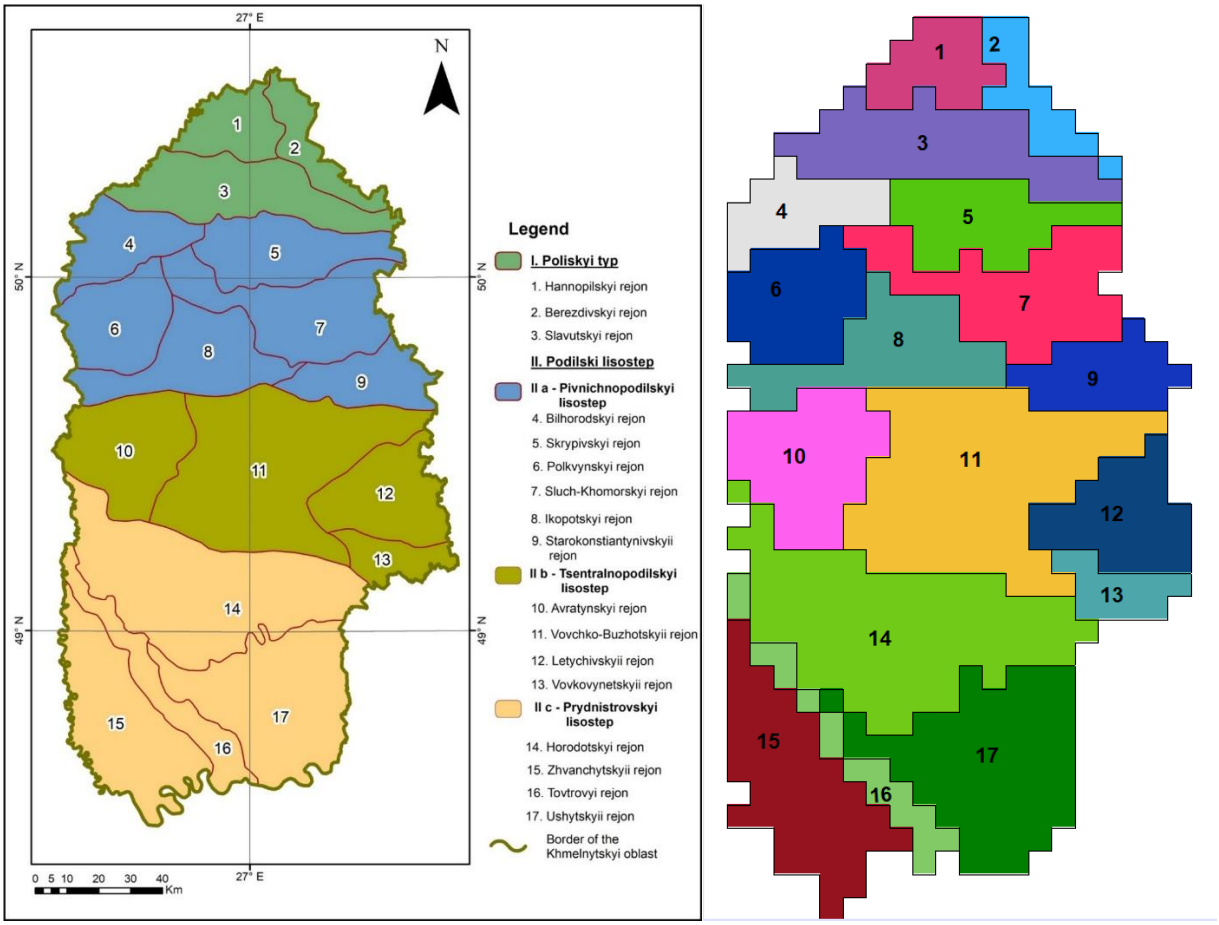
In order to ensure comparability of the obtained segmentation results with an established regional framework, a reference level of spatial aggregation was considered during the parameter selection process. The moving window in which the analyses were performed had a size of 3 × 3 pixels. Experiments were carried out using several values of the *Similarity* parameter in order to explore the sensitivity of the segmentation results to this setting. The approximate number of 17 segments was not treated as an optimization target, but rather as a reference level enabling comparison with the physico-geographical regionalization of Khmelnytskyi Oblast proposed by Herenchuk (1980), Fig 7a. In this sense, the parameter tuning was used to achieve a comparable level of spatial granularity, rather than to reproduce the reference division. This approach allowed for a controlled comparison between the independently derived segmentation and the existing regionalization scheme, while acknowledging that different segmentation levels may lead to different degrees of structural correspondence. Depending on the variant of variable selection for the segmentation algorithm and the chosen *Similarity* parameter, different numbers of segments were generated. The exact numbers are presented in Table 8.

Tab. 8. Number of segments generated in the analyzed variable selection variants

Similarity tolerance	Number of segments in the individual variants				
	A	B	C	D	E
10	41	48	56	46	47
20	28	33	55	46	43
25	20	19			
26	18	18			
30	10	14	48	43	29
40			32	31	19
42					18
45				19	13
50			20	13	
53			18		
60			11		

Variants of variable selection for the segmentation algorithm:

A. Variables 1-5: Bare soil, Forest, Shrub, Crop, Water
B. Variables 1-8: Bare soil, Forest, Shrub, Crop, Water, IJI, DIVISION, SHDI
C. PCA Components 1-5, derived from variables 1-5
D. PCA Components 1-8, derived from variables 1-8
E. PCA Component C1 (derived from variables 1-5) and PCA Component C1 (derived from variables 6-8)

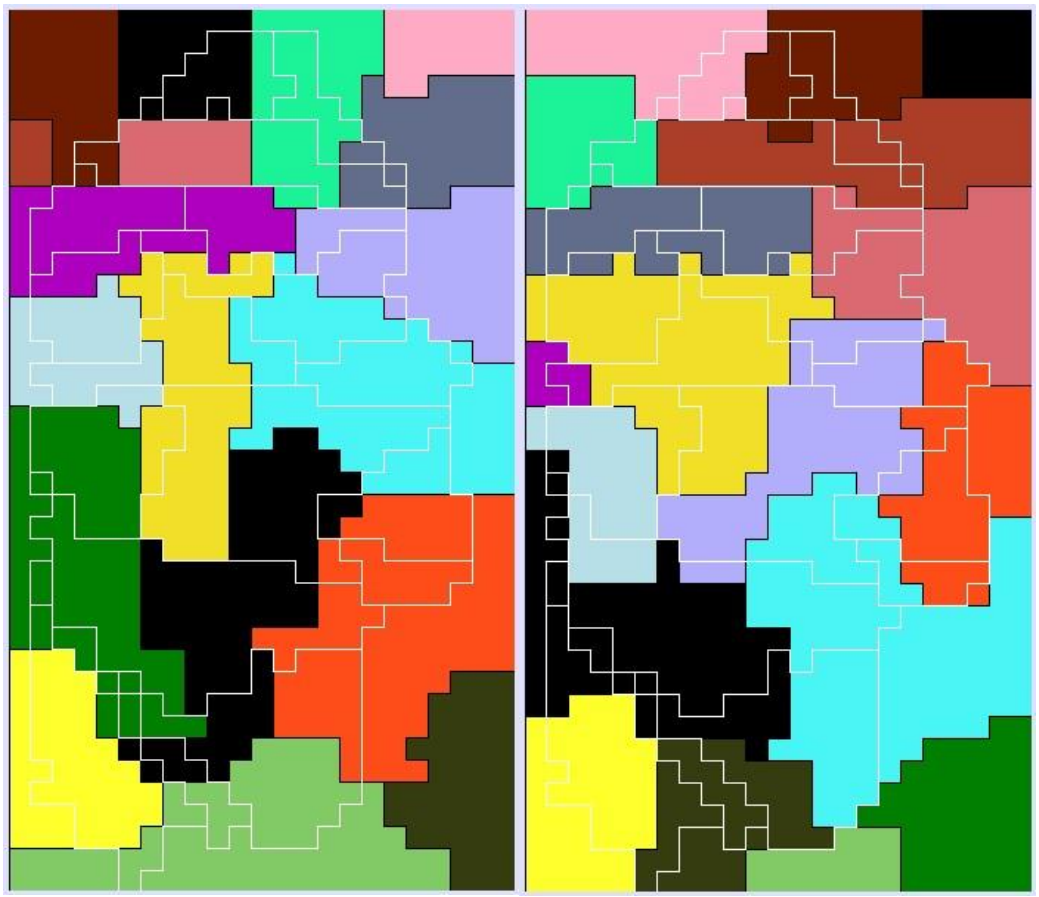


a. b.

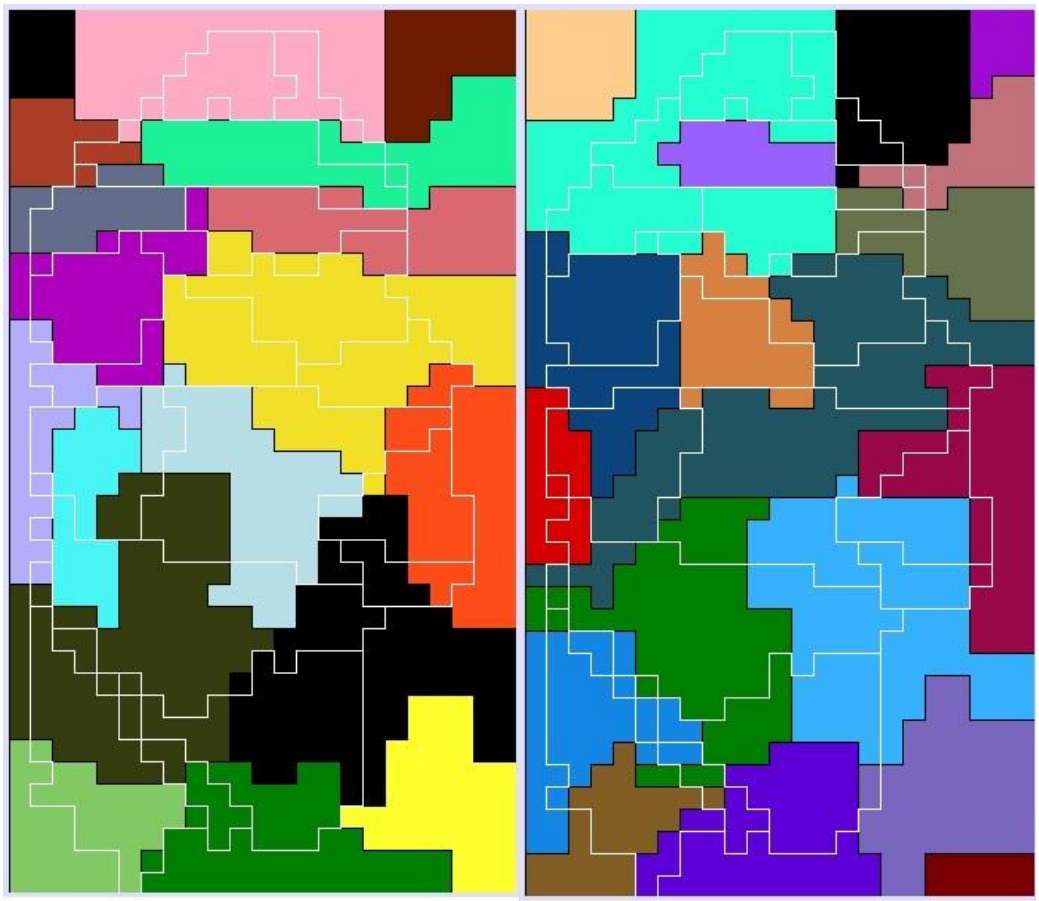
Fig. 7. a) Physico-geographical regionalization proposed by Herenchuk (1980). Source: (Komarova, Będkowski 2025) b) Raster version of the physico-geographical regionalization proposed by Herenchuk (1980), generalized to 6 × 6 km grid cells. Source: own elaboration.

The segmentation results for all variants mentioned above are presented in Fig. 8A–E. The presentation of results is limited to those variants of the *Similarity* parameter value that resulted in divisions of the Khmelnytsky oblast into the number of segments similar to that proposed by Herenchuk (1980).

A. **B.**



C. D.



E.

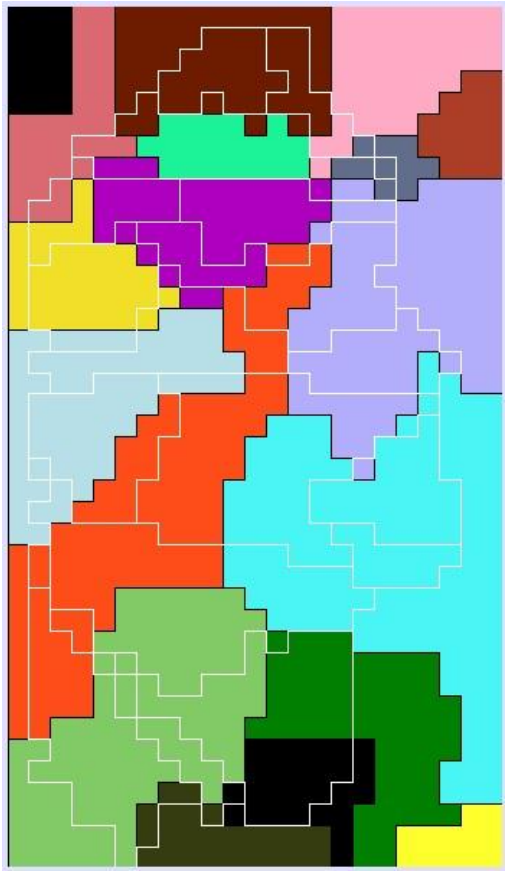


Fig. 8. Khmelnytskyi oblast (23 × 40 squares, i.e. 138 × 240 km) segmented by the applied algorithm based on: A.) Variables 1-5; B.) Variables 1-8, C.) PCA Components 1-5, derived from variables 1-5; D) PCA Components 1 -8, derived from variables 1-8; E) PCA Component C1 (derived from variables 1-5) and PCA Component C1 (derived from variables 6-8). White polygons show the extents of the regions delimited by Herenchuk (1980), generalized to 6 × 6 km grid cells. Source: own elaboration.

To further analyze the results obtained for the different segmentation variants, the tables (9-13) below present how the pixels of each Herenchuk (1980) region are distributed across the resulting segments (i.e., the distribution of counts within columns), as well as which Herenchuk regions are represented within each segment (i.e., the distribution of counts across rows). In addition, the Intersection over Union (IoU) index was calculated as a quantitative measure of agreement (Tables 9a-13a), integrating both perspectives by accounting for pixel shares from the standpoint of both the Herenchuk regions and the derived segments.

Tab 9. Distribution of pixels from Herenchuk regions across the generated segments (Variant A)

Herenchuk regions / Resulting segments	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	∑ pixels in the segment
1	12	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	5	11	11	0	2	0	0	0	0	0	0	0	0	0	0	0	0	29
6	0	3	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
7	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18
8	0	0	0	18	14	8	3	0	0	0	0	0	0	0	0	0	0	43
9	0	0	0	0	1	4	7	15	0	6	26	0	0	0	0	0	0	59
10	0	0	0	0	8	0	11	0	0	0	0	0	0	0	0	0	0	19
11	0	0	0	0	0	15	0	9	0	4	0	0	0	0	0	0	0	28
12	0	0	0	0	1	0	17	7	20	0	21	8	0	0	0	0	0	74
13	0	0	0	0	0	0	0	0	0	0	3	15	11	11	0	0	17	57
14	0	0	0	0	0	0	0	0	0	0	26	3	0	36	2	4	12	83
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	32
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	5	22	34
18	0	0	0	0	0	0	0	0	0	24	0	0	0	29	4	8	5	70
∑ pixels in Herenchuk regions	17	14	39	19	26	27	38	31	20	34	76	26	11	76	45	17	56	

Tab. 9a. Intersection over Union (IoU) index calculated for segmentation (Variant A)

Herenchuk regions / Resulting segments	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	**0,6**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	**0,1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	**0,1**	**0,3**	**0,2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	**0,2**	**0,1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	**0,5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	**0,4**	**0,3**	**0,1**	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	**0,1**	**0,2**	0	**0,1**	**0,2**	0	0	0	0	0	0
10	0	0	0	0	**0,2**	0	**0,2**	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	**0,4**	0	**0,2**	0	**0,1**	0	0	0	0	0	0	0
12	0	0	0	0	0	0	**0,2**	**0,1**	**0,3**	0	**0,2**	**0,1**	0	0	0	0	**0**
13	0	0	0	0	0	0	0	0	0	0	0	**0,2**	**0,2**	**0,1**	0	0	**0,2**
14	0	0	0	0	0	0	0	0	0	0	**0,2**	0	0	**0,3**	0	0	**0,1**
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,7**	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,1**	**0,1**	**0,3**
18	0	0	0	0	0	0	0	0	0	**0,3**	0	0	0	**0,2**	0	**0,1**	0

Tab 10. Distribution of pixels from Herenchuk regions across the generated segments (Variant B)

Herenchuk regions / Resulting segments	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	∑ pixels in the segment
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
2	**9**	0	**2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**11**
3	**8**	**8**	**2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**18**
4	0	**6**	**28**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**34**
5	0	0	**6**	**2**	0	0	0	0	0	0	0	0	0	0	0	0	0	**8**
6	0	0	0	**17**	**16**	**5**	**5**	0	0	0	0	0	0	0	0	0	0	**43**
7	0	0	**1**	0	**9**	0	**15**	0	0	0	0	0	0	0	0	0	0	**25**
8	0	0	0	0	0	0	0	**3**	0	0	0	0	0	0	0	0	0	**3**
9	0	0	0	0	**1**	**22**	**10**	**26**	0	**10**	**17**	0	0	0	0	0	0	**86**
10	0	0	0	0	0	0	**8**	**2**	**13**	0	**39**	**3**	0	**3**	0	**0**	0	**68**
11	0	0	0	0	0	0	0	0	0	**23**	0	0	0	**10**	0	0	0	**33**
12	0	0	0	0	0	0	0	0	0	0	**16**	**5**	**5**	**21**	0	0	**28**	**75**
13	0	0	0	0	0	0	0	0	0	**7**	0	**4**	**18**	**6**	0	0	0	**35**
14	0	0	0	0	0	0	0	0	0	1	0	0	0	**42**	**8**	**8**	**14**	**73**
15	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	**10**	**9**	**14**	**33**
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**27**	0	0	**27**
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
∑ pixels in Herenchuk regions	**17**	**14**	**39**	**19**	**26**	**27**	**38**	**31**	**20**	**34**	**76**	**26**	**11**	**76**	**45**	**17**	**56**	

Tab. 10a. Intersection over Union (IoU) index calculated for segmentation (Variant B)

Herenchuk regions / Resulting segments	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	**0,5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	**0,3**	**0,3**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	**0,1**	**0,6**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	**0,1**	**0,1**	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	**0,4**	**0,3**	**0,1**	**0,1**	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	**0,2**	0	**0,3**	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	**0,1**	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	**0,2**	**0,1**	**0,3**	0	**0,1**	**0,1**	0	0	0	0	0	0
10	0	0	0	0	0	0	**0,1**	0	**0,2**	0	**0,4**	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	**0,5**	0	0	0	**0,1**	0	0	0
12	0	0	0	0	0	0	0	0	0	0	**0,1**	**0,1**	**0,1**	**0,2**	0	0	**0,4**
13	0	0	0	0	0	0	0	0	**0,1**	0	0	**0,4**	**0,2**	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,5**	**0,1**	**0,1**	**0,1**
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,2**	**0,3**	**0,2**
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,6**	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tab 11. Distribution of pixels from Herenchuk regions across the generated segments (Variant C)

Herenchuk regions / Resulting segments	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	∑ pixels in the segment
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	17	9	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4
5	0	5	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	33
6	0	0	3	16	0	0	0	0	0	0	0	0	0	0	0	0	0	19
7	0	0	2	0	18	0	5	0	0	0	0	0	0	0	0	0	0	25
8	0	0	0	2	2	26	5	5	0	0	0	0	0	0	0	0	0	40
9	0	0	0	0	6	0	28	20	16	0	16	1	0	0	0	0	0	87
10	0	0	0	0	0	1	0	5	0	7	0	0	0	2	0	1	0	16
11	0	0	0	0	0	0	0	1	0	6	38	2	0	9	0	0	0	56
12	0	0	0	0	0	0	0	0	0	15	0	0	0	12	0	0	0	27
13	0	0	0	0	0	0	0	0	4	0	5	16	4	0	0	0	0	29
14	0	0	0	0	0	0	0	0	0	0	3	7	7	13	0	0	31	61
15	0	0	0	0	0	0	0	0	0	6	14	0	0	40	22	10	9	101
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	21
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6	16	24
∑ pixels in Herenchuk regions	17	14	39	19	26	27	38	31	20	34	76	26	11	76	45	17	56	

Tab. 11a. Intersection over Union (IoU) index calculated for segmentation (Variant C)

Herenchuk regions / Resulting segments	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0,6	0,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0,1	0,6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0,1	0,7	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0,5	0	0,1	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0,6	0,1	0,1	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0,1	0	0,3	0,2	0,2	0	0,1	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0,1	0	0,2	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0,1	0,4	0	0	0,1	0	0	0
12	0	0	0	0	0	0	0	0	0	0,3	0	0	0	0,1	0	0	0
13	0	0	0	0	0	0	0	0	0,1	0	0,1	0,4	0,1	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0,1	0,1	0,1	0	0	0,4
15	0	0	0	0	0	0	0	0	0	0	0,1	0	0	0,3	0,2	0,1	0,1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,5	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,2	0,3

Tab 12. Distribution of pixels from Herenchuk regions across the generated segments (Variant D)

Herenchuk regions / Resulting segments	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	∑ pixels in the segment
1	0	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19
5	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
6	17	7	13	17	18	1	3	0	0	0	0	0	0	0	0	0	0	76
7	0	0	3	0	5	0	4	0	0	0	0	0	0	0	0	0	0	12
8	0	0	0	2	0	26	0	10	0	12	0	0	0	0	0	0	0	50
9	0	0	0	0	1	0	9	21	0	0	3	0	0	0	0	0	0	34
10	0	0	0	0	0	0	0	0	0	6	0	0	0	5	0	0	0	11
11	0	0	0	0	2	0	22	0	16	16	38	0	0	5	0	1	0	100
12	0	0	0	0	0	0	0	0	4	0	8	11	1	0	0	0	0	24
13	0	0	0	0	0	0	0	0	0	0	0	0	0	2	20	8	3	33
14	0	0	0	0	0	0	0	0	0	0	16	15	10	19	0	0	16	76
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	1	0	17
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	4	22
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	11	0	0	45	3	3	13	75
∑ pixels in Herenchuk regions	17	14	39	19	26	27	38	31	20	34	76	26	11	76	45	17	56	

Tab. 12a. Intersection over Union (IoU) index calculated for segmentation (Variant D)

Resulting segments \ Herenchuk regions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	**0,4**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	**0,5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	**0,1**	**0,1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	**0,2**	**0,1**	**0,1**	**0,2**	**0,2**	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	**0,1**	0	**0,2**	0	**0,1**	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	**0,5**	0	**0,1**	0	**0,2**	0	0	0	0	0	0	0
9	0	0	0	0	0	0	**0,1**	**0,5**	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	**0,2**	0	0	0	**0,1**	0	**0**	**0**
11	0	0	0	0	0	0	**0,2**	0	**0,2**	**0,1**	**0,3**	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	**0,1**	0	**0,1**	**0,3**	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,3**	**0,2**	0
14	0	0	0	0	0	0	0	0	0	0	**0,1**	**0,2**	**0,1**	**0,1**	0	0	**0,1**
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,3**	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,1**	**0,1**	**0,3**
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	**0,1**	0	0	**0,4**	0	0	**0,1**

Tab 13. Distribution of pixels from Herenchuk regions across the generated segments (Variant E)

Resulting segments \ Herenchuk regions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	∑ pixels in the segment
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
2	0	**3**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**
3	**17**	**9**	**6**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**32**
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
5	0	0	**20**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**20**
6	0	**2**	**3**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**5**
7	0	0	**3**	**4**	0	0	0	0	0	0	0	0	0	0	0	0	0	**7**
8	0	0	**4**	**9**	**20**	**2**	**10**	**3**	0	0	0	0	0	0	0	0	0	**48**
9	0	0	0	**6**	0	**21**	0	**2**	0	0	0	0	0	0	0	0	0	**29**
10	0	0	2	0	**4**	0	**19**	0	**18**	0	**12**	**1**	0	0	0	0	0	**56**
11	0	0	0	0	0	**4**	0	**19**	0	**22**	**4**	0	0	**4**	0	0	0	**53**
12	0	0	0	0	0	0	0	0	**2**	0	**30**	**25**	**11**	**19**	0	0	0	**87**
13	0	0	0	0	**2**	0	**9**	**7**	0	**12**	**30**	0	0	**25**	**14**	**5**	0	**104**
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	**17**	**18**
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0	**5**
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
17	0	0	0	0	0	0	0	0	0	0	0	0	0	**27**	**27**	**10**	**19**	**83**
18	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	**20**	**21**
∑ pixels in Herenchuk regions	**17**	**14**	**39**	**19**	**26**	**27**	**38**	**31**	**20**	**34**	**76**	**26**	**11**	**76**	**45**	**17**	**56**	

Tab 13a. Intersection over Union (IoU) index calculated for segmentation (Variant E)

Resulting segments \ Herenchuk regions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	**0,2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	**0,5**	**0,2**	**0,1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	**0,5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	**0,1**	**0,1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	**0,1**	**0,2**	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	**0,2**	**0,4**	**0**	**0,1**	0	0	0	0	0	0	0	0	0	0
9	0	0	0	**0,1**	0	**0,6**	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	**0,1**	0	**0,3**	0	**0,3**	0	**0,1**	0	0	0	0	0	0
11	0	0	0	0	0	**0,1**	0	**0,3**	0	**0,3**	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	**0,2**	**0,3**	**0,1**	**0,1**	0	0	0
13	0	0	0	0	0	0	**0,1**	**0,1**	0	**0,1**	**0,2**	0	0	**0,2**	**0,1**	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,3**
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,1**	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,2**	**0,3**	**0,1**	**0,2**
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0,4**

Results and Conclusions

The segmentation attempts based on the analyses described above did not lead to fully consistent results. All segmentation variants proved to be highly sensitive to the input parameter settings. However, it is possible to observe characteristic similarities both between the physico-geographical division proposed by Herenchuk (1980) and the resulting segmentations, as well as among the individual segmentation variants themselves. The similarity between the physico-geographical division proposed by Herenchuk and the resulting segmentations is particularly evident in the northern part of the Khmelnytskyi Oblast, which is most likely associated with the presence and dominance of forest land cover. Across all segmentation variants, it can be observed that region no. 3, representing this continuous forest belt, is the only one to achieve IoU index values in the range of 0.5–0.6 (Tables 9a–13a). This indicates that a substantial portion of pixels from the original region no. 3 were assigned to the newly created segment, where they still constitute a majority compared to pixels originating from other landscape types (Tables 9-13). In other areas, especially further south, the delineation of spatial units was based less on the proportion of land cover types and more on variables describing their spatial configuration and heterogeneity. Notably, some reference units distinguished by Herenchuk (1980), such as the Tovtry region (No. 16), were not identified in the present analysis. In his description of individual regions, Herenchuk referred to their physiognomic structure, which is difficult to identify using the analyzed satellite imagery. This limitation may be attributed to anthropogenic transformations that have altered the landscape and increased its fragmentation, potentially leading to a loss of the original distinctiveness of individual regions. This may explain why the results of the conducted analyses do not fully correspond to all landscape units distinguished by Herenchuk.

With regard to the similarity of results among the individual segmentation variants, it can be stated that in all divisions at least seven of Herenchuk's regions can be identified, the majority of whose pixels were assigned to a single new segment (Tables 9–13).

The segmentation algorithm used is very sensitive to the control parameter settings, particularly the *Similarity* parameter. This can be observed when analyzing the resulting segmentations: although each of them includes Herenchuk's regions for which the majority of pixels were assigned to new segments, this does not necessarily mean that pixels from these regions constitute the majority within those new segments. In other words, a clear one-to-one correspondence between an original region and a newly created segment occurs only rarely. Moreover, the dominant classes also vary across the analyzed divisions. For example, in variant B these are regions 1, 3, 10, 14, and 15 (Table 10a), whereas in variant C they are regions 1, 3, 4, 5, 6, and 15 (Table 11a).

At the same time, there are no clearly established criteria for determining the optimal value of the *Similarity* parameter. The algorithm's creators (see TerrSet® Help System) recommend starting with a *Similarity* value of 10 and then modifying it until satisfactory results are achieved. It should be acknowledged that the choice of producing approximately 17 segments was guided by the intention to enable comparison with the reference regionalization of Herenchuk (1980), rather than being treated as a strictly optimized or independent segmentation target. While this introduces a certain degree of comparability bias, the aim of the study was not to replicate the reference division, but to assess the degree of structural correspondence between an independently derived segmentation and an established regionalization scheme. An alternative approach could involve generating a more detailed segmentation and subsequently aggregating it to a level comparable with the reference system, which may provide a more flexible framework for future comparative analyses.

In addition to the sensitivity related to segmentation parameters, the influence of the number and type of input variables on the resulting divisions was also examined. Including eight components in the calculations, i.e., those created with Landscape metrics in mind, led to a division similar to that generated based on five components. Differences between these variants occur throughout the study area, but mainly concern boundary squares, i.e., those located on the edges of the identified spatial units. This may indicate the existence of distinct spatial features that allow for the identification of given fragments as truly distinct from their surroundings. Another explanation is the strong correlation between the original values of the three landscape metrics and the variables describing the share of the Forest and Crop fractions in the landscape. It is also clearly visible that these variables play a significant role in the creation of the first component (C1).

The study did not test the influence of weights determining the importance of individual components. Equal weights were also assumed for the mean and variance, two important criteria for combining segments into larger spatial units. This is an interesting area for further research, especially considering new landscape metrics.

The regionalization adopted by Herenchuk (1980), as our reference, is not necessarily the best one. Errors, simplifications, and the influence of generalization are possible. There are also border squares (Komarova, Będkowski 2025), whose affiliation with neighboring regions may change depending on the importance assigned to individual features. Furthermore, the division of Khmelnytskyi Oblast into regions with rectangular boundaries, which approximate Herenchuk's (1980) analogous reference division, may be a source of doubt regarding the subsequent classification of grid cells. Most importantly, however, the Herenchuk regionalization also relies on many other physical features of the analyzed area, some of which do not affect the structure and texture of satellite images. Therefore, they are, in a sense, invisible to the procedure we propose. Therefore, it can be proposed that

the segmentation also take into account other physical and geographical features of the analyzed area, information about which may come from other sources.

The presented methodology can be applied to analyses of smaller areas, where the nature of the input data may differ—for instance, remote sensing data may be acquired from lower altitudes, providing higher spatial resolution. Furthermore, supplementing remote sensing data with field data can improve classification accuracy, which in turn may positively affect the overall results. More broadly, the proposed approach can serve as an additional component in regionalization studies, as it enables a unique assessment of the composition and configuration of the study area, which directly influences the resulting division. However, it should be emphasized that the outcomes of such analyses also depend on a number of methodological assumptions, the most important being the selection of variables, as well as, for example, the size of the basic spatial units (grid cells) used to divide the study area.

References

Clark 2025: Clark University. Center for Geospatial Analytics. https://www.clarku.edu/centers/geospatial-analytics/terrset/. (access: 17.12.2025)

Dadashpoor H., Azizi P., Moghadasi M. Land use change, urbanization, and change in landscape pattern in a metropolitan area. Science of The Total Environment 2019, 655, 707-719, https://doi.org/10.1016/j.scitotenv.2018.11.267

Fragstats. A Spatial Pattern Analysis Program for Categorical Maps. https://fragstats.org/index.php (access 04.03.2026).

ArcMap. Executing the Iso Cluster Unsupervised Classification tool. https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/image-classification/executing-the-iso-cluster-unsupervised-classification-tool.htm (access 03.02.2026).

Frohn R.C., Hao Y. Landscape metric performance in analyzing two decades of deforestation in the Amazon Basin of Rondonia, Brazil, Remote Sensing of Environment, 2006, 100, 237-251, https://doi.org/10.1016/j.rse.2005.10.026

Gramacki J., Gramacki A. Statystyczne odkrywanie zależności w danych. Przegląd Telekomunikacyjny 2008, LXXXI, 6, 711-713.

Haglauer D. Zdjęcia lotnicze jako podstawa regionalizacji geograficznej. Fotointerpretacja w Geografii 1996, 3, 45-53.

Herenchuk K.I. (ed.). Pryroda Khmelnytskoi oblasti. Vischa shkola 1980, Lvov.

Jakomulska A., Sobczak M. Korekcja radiometryczna obrazów satelitarnych – metodyka i przykłady. Teledetekcja Środowiska 2001, 32,152-171.

el Jeitany J., Nussbaum M., Pacetti T., Schröder B., Caporali E. Landscape metrics as predictors of water-related ecosystem services: Insights from hydrological modeling and data-based approaches applied on the Arno River Basin, Italy, Science of The Total Environment 2024, 954, 176567, https://doi.org/10.1016/j.scitotenv.2024.176567

Komarova, U., Będkowski, K. Assessment of the internal consistency of physico-geographical units using landscape metrics and statistical methods: case study of the Khmelnytskyi Oblast, Ukraine. Acta Geographica Lodziensia 2025, 118, 67-89, https://doi.org/10.26485/AGL/2025/118/5

Kondracki J. Podstawy regionalizacji fizycznogeograficznej. Warszawa: PWN, 1976.

Lassalle P., Inglada J., Michel J., Grizonnet M., Malik J. A Scalable Tile-Based Framework for Region-Merging Segmentation. IEEE Transactions on Geoscience and Remote Sensing 2015, 53 (10), 1-13, https://doi.org/10.1109/TGRS.2015.2422848

McGarigal, K., Marks, B. J. FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure. General Technical Report PNW-GTR-351, U.S. Forest Service, 1995

Nita Jerzy, Myga-Piątek Urszula, Pukowiec-Kudra Katarzyna; Propozycja mikroregionalizacji województwa śląskiego – weryfikacja metody na wybranych mezoregionach; Prace Komisji Krajobrazu Kulturowego 2016, 31.

Olędzki J. R. Regiony geograficzne Polski. Teledetekcja Środowiska 2007, 38, 1-337.

Osińska-Skotak K. Znaczenie korekcji radiometrycznej w procesie przetwarzania zdjęć satelitarnych. Archiwum Fotogrametrii, Kartografii i Teledetekcji 2007, 17B, 577-590.

Parysek J., Ratajczak W. Analiza składowych głównych, jej korzyści i ograniczenia z punktu widzenia badań geograficznych. In: H. Rogacki (ed.) Możliwości i ograniczenia zastosowań metod badawczych w geografii społeczno-ekonomicznej i gospodarce przestrzennej (61-73). Poznań: Bogucki Wyd. Naukowe, Poznań, 2002.

Ratajczak W. Analiza regresji a składowe główne. In: H. Rogacki (ed.) Problemy interpretacji wyników metod badawczych stosowanych w geografii społeczno-ekonomicznej i gospodarce przestrzennej (61-70). Poznań: Bogucki Wyd. Naukowe, 2003.

Richling A., Solon J. Ekologia krajobrazu, Warszawa: PWN, 2011.

Saura S., Martínez-Millán J. Sensitivity of Landscape Pattern Metrics to Map Spatial Extent. Photogrammetric Engineering and Remote Sensing 2001, 67 (9), 1027-1036.

da Silva A. M., Huang C.H., Francesconi W., Saintil T., Villegas J. Using landscape metrics to analyze micro-scale soil erosion processes, Ecological Indicators 2015, 56, 184-193, https://doi.org/10.1016/j.ecolind.2015.04.004

Szabó S., Csorba P., Szilassi P. Tools for landscape ecological planning—scale, and aggregation sensitivity of the contagion type landscape metric indices. Carpathian Journal of Earth Environmental Sciences 2012, 7, 127–136

Turner, M. G., Gardner, R. H., O'Neill, R. V. Landscape Ecology in Theory and Practice: Pattern and Process. Springer, 2001.

Wang J., Chen Y., Wu Z., Wei Y., Zhang Z., Wang X., Huang J., Shi Z. On the effectiveness of multi-scale landscape metrics in soil organic carbon mapping, Geoderma 2024, 449, 117026, https://doi.org/10.1016/j.geoderma.2024.117026

Wei C., Chen K., Yang Z., Yao S., He C., Zhang L. Performance of multiscale landscape metrics as indicators of soil C, N, P, and physicochemical properties of NFV in ML reaches of Yellow River Basin, China, Ecological Indicators 2025, 173, 113391, https://doi.org/10.1016/j.ecolind.2025.113391

Zabihi M., Mostafazadeh R., Gonabadi I.S. Analyzing the spatial patterns and changes in urban green spaces of an under rapid urbanization area through landscape metrics. Advances in Space Research 2025, 76 (5), 2779-2794, https://doi.org/10.1016/j.asr.2025.06.063.

Zhang N., Li H. Sensitivity and effectiveness of landscape metric scalograms in determining the characteristic scale of a hierarchically structured landscape. Landscape Ecology 2013, 28, 343–363, https://doi.org/10.1007/s10980-012-9837-x

Krzysztof Będkowski
Uliana Komarova
Teledetekcja Środowiska
Polskie Towarzystwo Geograficzne
Oddział Teledetekcji i Geoinformatyki – Warszawa 2026