

Raster images generalization in the context of research on the structure of landscape and geodiversity

Tomasz Bartuś

AGH University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection, Department of the General Geology and Geotourism; al. Mickiewicza 30, 30-059 Krakow, Poland; e-mail: bartus@agh.edu.pl

© 2014 Authors. This is an open access publication, which can be used, distributed and reproduced in any medium according to the Creative Commons CC-BY 4.0 License requiring that the original work has been properly cited.

Received: 22 October 2014; accepted: 20 January 2015

Abstract: Generalization is one of the most important stages of work on cartographic data. It has a particular importance in the study of landscape structure, especially geodiversity. In raster images, it is based on modifying the structure of the image while maintaining its general characteristics. In ArcGIS software, the most important tools for generalization of raster images include: Boundary Clean and Majority Filter. Fragstat software was used for the analysis of structural modifications of the output images and assessment of the effects of generalization. Depending on the options used, both tools (Boundary Clean and Majority Filter) cause different types of modifications in rasters. Elimination of the so-called noise using one of the variants of Majority Filter is the most suitable if we wish to introduce only subtle modifications to the final image. If, however, we expect a greater level of interference in the structure of the source images, using Boundary Clean becomes necessary.

Keywords: generalization, raster, landscape, geodiversity, Boundary Clean, Majority Filter

INTRODUCTION

The end of the twentieth century led to the rapid development of various information technologies, including the creation of mapping documents and the implementation of spatial analysis using Geographic Information Systems (GIS). The dissemination of such research would not be possible without enormous resources of high quality reference data (digital elevation models, topographical maps, geological maps, sociological maps, soil maps, aerial and satellite images and others).

Among the many stages of work involving editing and spatial analysis of cartographic materials, generalization is worth noting (McMaster 1992, Weibel & Jones 1998, Chrobak 2005). This term refers to a set of operations aimed at optimizing the processed content, depending on its purpose. Generalization is a complex and multi-threaded issue

(Beat & Weibel 1999). Regardless of which data model one adopts (vector or raster); both models are used in the process of creating a composition of maps, as well as in various analyses and modelling. In the map edition section, generalization should be understood as a process of simplifying reference data, whose quality (resolution) is higher than necessary for the input materials in a particular scale (Daley et al. 1997). Generalization can be found at all stages of work with digital maps, from data collection through proper analysis to final processing and generation of results (McMaster 1992). The typical tasks performed during work using vector data include: selection of classes, reclassification, resymbolisation, simplification of lines and boundaries of objects, replacement of point objects and small areas (such as buildings) with larger area objects (e.g. built-up areas), omitting some information for greater map clarity, and

more (McMaster 1992, Harrie & Sarjakoski 2002). In terms of raster graphics, the purpose of generalization is usually filtering artefacts and highlighting general phenomena. Raster images, which are the results of GIS analyses, modelling, or processing of remote sensing data, are often characterized by the presence of various bugs and micro-structures. These are individual pixels or their small groups, surrounded by cells with different values of the analysed attributes. The processing of continuous spatial data (e.g. DEM) and the use of primary and secondary attributes calculated on that basis (e.g. topography) often results in the presence of groups of pixels of two categories in the characteristic chequerboard pattern. Such situations obscure the image of the analysed phenomena and are treated as the so-called noise (McMaster 1992, Beat & Weibel 1999). In this case, generalization aims at reviewing and deliberately discarding unwanted information (Veregin & McMaster 1997). This aim can be achieved by increasing the resolution of the raster (increasing basic cell size), or by using a variety of filter tools that will remove unnecessary parts of information from the images in a controlled manner.

Raster analysis, used e.g. in landscape ecology, is the study of spatial structures of variations of the observed separations. The most important factors that describe these structures include: the number, size and perimeter of units, their shapes as well as the relationships between the units, e.g.: adjacencies, diversity, uniformity, aggregation, isolation and fragmentation (McGarigal et al. 1995, 2012). The structure can be described at the level of individual units (patch level), their classes (types) meaning groups of pixels with the same values as the analysed attribute, or the entire landscape (study area level). Changes that are made using the processing tools may, to a greater or lesser extent, blur the characteristics of the original structures and thus hinder or even prevent correct interpretation of the results (Veregin & McMaster 1997).

In modern GIS systems, there are numerous tools used to generalize the geometry of lines and areas (vector model) as well as patch geometry (raster model).

The aim of the analysis was determining the impact of selected generalization tools on the structure of raster images with particular emphasis on the diversity of the observed classes.

DATA

All studies used a single set of data, hereinafter referred to as raw or primary data. The raster map of topoclimatic variability of the Ojców National Park (southern Poland) and its immediate buffer zone belongs to these data (Fig. 1). The map was generated using the J. Paszyński method (1980) with modifications by Kicińska et al. (2001), Kot (2005, 2006) and some solutions by the author (Bartuś 2014). The map highlights 23 classes (topoclimate types). The resolution of rasters has been determined at 10 m in accordance with Hengl's (2006) methodology.

METHODS

There are many image processing methods applicable to removing artifacts from raster images. Most often they use all kinds of point transformations (image elements are modified irrespective of the state of the neighboring elements), the context transformations (convolution, logical, low-pass or high-pass, laplacian and median filters), spectral transformations (using Fourier transformation) or morphological transformations (image elements are modified depending on the logical conditions) (Tadeusiewicz & Korochocka 1997). The effective noise filtering methods use a variety of interpolation methods in the analysis of continuous variables. These methods use, for example, inverse distance weighting method, geostatistical methods (in particular factorial kriging) and others.

The preparation of the raw image of the topoclimatic variation map in the ONP region, as well as all generalization operations were performed in the ArcGIS environment version 10.1. A raster image of the topoclimatic map was processed using two basic tools available in the Spatial Analyst module: Boundary Clean and Majority Filter.

Boundary Clean (hereinafter: BC), is a tool that smooths the jagged boundaries by eliminating small-surfaces that is less significant groups of pixels of the same class (patches), therefore simplifying the structure of raster images. The filter uses the expansion method, followed by shrinking. Performing operations determine the variables responsible for the expansion priority of patches. By default, it assumes the value "No sort" (hereinafter: NS).

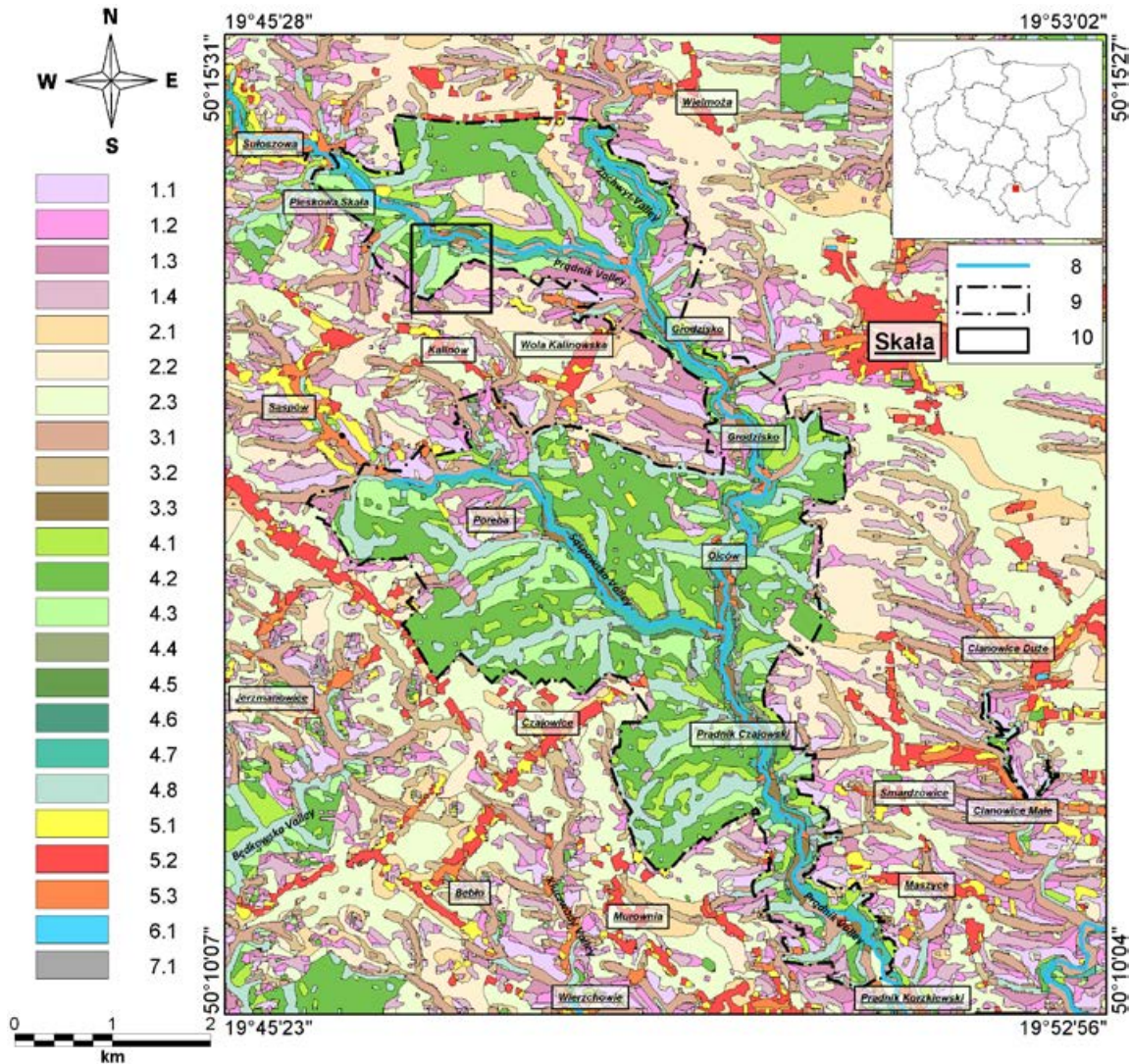


Fig. 1. Topoclimatic raster map of the Ojców National Park (ONP); topoclimates of unwooded-convex areas: 1.1 – with slopes 5–30° and S, SW or SE exposure; 1.2 – with slopes 5–30° and W or E exposure and with slopes 0–5° and N or S exposure; 1.3 – with slopes 5–30° and N, NW or NE exposure; 1.4 – areas with highly diverse relief with slopes 0–5° and various exposures: NW, NE, SW, SE, W or E; Topoclimates of unwooded-flat areas (except for flat sections of valleys): 2.1 – areas used in agriculture, characterized by a low relative level of total soil porosity; 2.2 – areas used in agriculture, without dense vegetation and with an average level of relative total soil porosity; 2.3 – areas used in agriculture characterized by a high relative level of total soil porosity and areas with dense vegetation; Topoclimates of unwooded-concave areas: 3.1 – wide and humid valley bottoms; 3.2 – areas located in upper parts of wide valley bottoms (lower slopes); 3.3 – narrow steep bottoms of valleys and gorges; Topoclimates of forest areas: 4.1 – with slopes 5–30° and S, SW or SE exposure; 4.2 – with slopes 5–30° and W or E exposure and flat areas with slopes 0–5°; 4.3 – with slopes 5–30° and N, NW or NE exposure; 4.4 – with slopes 30–90° and S, SW or SE exposure; 4.5 – with slopes 30–90° and N, NW or NE exposure; 4.6 – with slopes 30–90° and W or E exposure; 4.7 – humid valley bottoms; 4.8 – upper parts of wide valley bottoms (lower slopes); Topoclimates of anthropogenically transformed areas: 5.1 – in built-up and industrial areas located on convex morphology; 5.2 – built-up and industrial areas in plains; 5.3 – built-up and industrial areas located in all types of morphological depressions (valleys, closed depressions and others); Topoclimates of reservoirs: 6.1 – areas located in various water bodies and their immediate vicinity; Topoclimates of rock outcrops: 7.1 – areas observed in rock outcrops and denudation monadnocks; 8 – surface streams; 9 – the ONP area; 10 – fragment analysed in Figure 7

In such case, the sorting order of patches is only determined by the value of the attribute describing the pixels. Patches with higher attribute values have a higher priority to expand than patches with lower values. The process is completely independent

of patches size. In the case of the “Descend” option (hereinafter: *Desc*), groups of pixels will be sorted by size, in descending order. All patches with larger surface areas will have a higher priority to expand than patches with smaller surface areas.

Using the “Ascend” option (hereinafter: *Asc*) will result in sorting pixel groups by size in ascending order. All patches with smaller surface areas will have a higher priority to expand than ones with larger surface areas (ESRI 2012).

The BC tool procedure can be performed once (*One Way*) or twice (*Two Way*). Launching the appropriate mode is determined by selecting the option “*Expansion and shrinking twice*”. In “*Two Way*” mode (default option), the expansion and contraction of the patches is carried out depending on the particular type of sorting, and then an additional expansion and contraction is performed with reversed priority. In “*One Way*” mode, the expansion and contraction of patches is performed once (ESRI 2012).

Majority Filter (hereinafter MF) is a tool used to modify raster images by eliminating selected pixels and replacing their values with attributes present in the background. The operation of the filter is determined by the results of observations of the values of cells found around the so-called central cell and their continuity. MF has two criteria for its substitution mechanism. Firstly, the number of cells adjacent to the central pixel having a certain value must be sufficient for the mechanism to work. Depending on the replacement threshold value, the number of neighboring cells allowing the value of the central cell to be changed includes the majority (“*Majority*” selected), or at least half (“*Half*” selected). Selecting the majority options means that three of the four neighboring cells or five of the eight cells must have the same value. In the second case (when set to “*Half*”), two out of four cells or four out of eight cells are necessary. The second criterion conditioning replacements can be reduced to stating that the filter should aim at simplifying the structure of the map and minimizing the number of patches with a small number of pixels. Failure to meet any of the criteria will result in the lack of substitution of the central pixel (ESRI 2012).

The tool enables two ways to perform filtration. Procedure selection is carried out by defining the number of neighboring pixels involved in the analysis (*Number of neighbors to use*). The variable can have two values: “*Four*” (“4”) or “*Eight*” (“8”). If set to “4”, this means that the calculations will only involve pixels adjacent to the edges of the central cell, and the values of pixels located in the

corners of the neighborhood will not be taken into account. Using the “8” variable will apply the analysis to all eight pixels surrounding the central cell (ESRI 2012). The tool can be used repeatedly. After a few more iterations, the raster image reaches a stable form, which needs no further change.

LANDSCAPE METRICS

The configuration possibilities of generalization tools allow for the generation of ten processed images (six using the BC tool and four using the MF tool). The obtained output images were analysed in detail using landscape metrics. The study used Fragstats software version 4.1 (McGarigal et al. 1995, 2012). Moore’s principle of neighborhood was applied using eight pixels surrounding the central cell while defining the patches boundaries.

The Fragstats program, which was created and is being developed at the University of Oregon (McGarigal et al. 1995, 2012), is currently the most involved tool for quantitative analysis of the structure of raster spatial data (Raines 2002, Zwierzchowska et al. 2010). This application allows for the use of several landscape metrics (indicators) (McGarigal et al. 1995, 2012, Kot & Leśniak 2006, Urbański 2012). In its current version, the program divides available metrics into six groups: area and edge metrics, shape metrics, core area metrics, contrast metrics, aggregation metrics and diversity metrics. Depending on their nature, the available metrics may refer to the level of the element (patch), the class, or the entire landscape.

Percentage of Landscape (PLAND) is a relative area metric which quantifies the proportional abundance of each patch type in the landscape (eq. 1) (McGarigal et al. 1995, 2012, Urbański 2012):

$$PLAND_i = \sum_{j=1}^n a_{ij} \cdot \frac{10^2}{A} [\%] \quad (1)$$

where:

- i – class,
- j – item patch number,
- n – number of patches each class (type),
- a_{ij} – area of patch ij [m²],
- A – total landscape area [m²].

Largest Patch Index (LPI) is a simple measure of dominance. Calculated at the landscape level, it quantifies the percentage of total landscape area comprised by the largest patch (eq. 2) (McGarigal et al. 1995, 2012):

$$\text{LPI} = \frac{\max(a_{ij})}{A} \cdot 100 [\%] \quad (2)$$

Symbols as in eq. (1).

Total Edge (TE) is calculated at the class and total landscape level. At the landscape level, it is an absolute measure of total edge length of a whole patches (McGarigal et al. 1995, 2012, Urbański 2012).

Edge Density (ED) is calculated at the class and total landscape level. At the landscape level, it equals the sum of the lengths of all edge segments in the landscape divided by the total landscape area and converted to hectares (eq. 3) (McGarigal et al. 1995, 2012, Urbański 2012):

$$\text{ED} = \frac{E}{A} \cdot 10\,000 \text{ [m/ha]} \quad (3)$$

where E – total length of edge in landscape [m]. Other symbols as in eq. (1).

Shape Index (SHAPE) is a measure of the shape complexity. The metric is calculated at the patches level as patch perimeter divided by the square root of patch area, adjusted by a constant to adjust for a square standard (eq. 4). The values of the metric can be changed in the range from 1 to infinity. The metric can reach minimum value when the patch is square. Measure values can increase without limit as patch shape becomes more irregular. The advantage of the parameter is the resistance to the size of the patch (McGarigal et al. 1995, 2012). At the class and total landscape level, the simple descriptive statistics of the patches level parameter are analyzed:

$$\text{SHAPE} = \frac{0.25 \cdot p_{ij}}{\sqrt{a_{ij}}} [-] \quad (4)$$

where p_{ij} – perimeter of patch ij [m]. Other symbols as in eq. (1).

Fractal Dimension Index (FRAC) is a metric that reflects the complexity of the patches shape. The parameter is calculated at the patches level

(eq. 5). At the class and total landscape level, the analysis is based on simple descriptive statistics. The values of the metric can be changed in the range from 1 to 2. The value parameter reaches minimum for shapes with very simple perimeters. The gradual increase to the value of 2 indicates the increasing complexity of shapes. FRAC reflects shape complexity across a range of spatial scales (patch sizes) (Milne 1988, McGarigal et al. 1995, 2012, Urbański 2012):

$$\text{FRAC} = \frac{2 \cdot \ln(0.25 \cdot p_{ij})}{\ln a_{ij}} [-] \quad (5)$$

Symbols as in eq. (1) and (4).

Patch Richness (PR) is probably the simplest landscape composition metric. It is calculated at the total landscape level as the number of patch types (classes) (McGarigal et al. 1995, 2012).

Shannon's Diversity Index (SHDI) is a parameter which measures the degree of spatial class's diversity (Shannon & Weaver 1949). It is calculated at the total landscape level (eq. 6). The metric values can be changed in the range from 0 to $\ln m_{\max}$, where: m_{\max} means a maximum number of patches types. SHDI value is 0, when the landscape contains only one patch (no diversity). Metric values increase, with the increasing degree of uniform coverage area by different classes and with the increase of the number of different patch types. Shannon's Diversity Index is somewhat more sensitive to rare patch types and small patches than Simpson's diversity index (SIDI) (McGarigal et al. 1995, 2012, Kot & Leśniak 2006, Urbański 2012):

$$\text{SHDI} = - \sum_{i=1}^m (P_i \cdot \ln P_i) [-] \quad (6)$$

where:

P_i – proportion of the landscape occupied by patch type (probability of a particular patch in the landscape),

m – number of patch classes present in the landscape.

Simpson's Diversity Index (SIDI) is a parameter, which is calculated at the total landscape level. It describes the probability of an event in which two randomly chosen cells belong to different

patch types (eq. 7). The metric values can be changed in the range from 0 to 1. The metric can reach the minimum value, when the landscape contains only one patch (no diversity). Index values increases with increasing the number of different patch types and with increasing degree of uniform coverage area by different classes. Simpson's index is less sensitive to the presence of rare types of patches (Simpson 1949, McGarigal et al. 1995, 2012, Urbański 2012):

$$SIDI = 1 - \sum_{i=1}^m P_i^2 \quad [-] \quad (7)$$

Symbols as in eq. (6).

Shannon's Evenness Index (SHEI) is a measure of the proportional distribution of patches of different classes in the analyzed area (eq. 8). The metric is calculated at the total landscape level. The metric values can be changed from 0 to 1. The metric can reach the minimum value when the landscape contains only one patch (no diversity) or when the distribution of area among the different patch types becomes increasingly uneven (i.e., dominated by 1 type). The index can reach the maximum value when distribution of area among patch types is perfectly even. The difference between the maximum SHEI value equal to 1 and the actual indicator level is called dominance (McGarigal et al. 1995, 2012, Urbański 2012):

$$SHEI = \frac{-\sum_{i=1}^m (P_i \cdot \ln P_i)}{\ln m} \quad [-] \quad (8)$$

Symbols as in eq. (6).

Number of Patches (NP) is a simple measure, which is calculated at the class and the total landscape level. It equals to the number of patches in the whole analyzed area at the total landscape level. The metric often has limited interpretive value by itself because it conveys no information about area, distribution or density of patches (McGarigal et al. 1995, 2012, Urbański 2012).

Percentage of Like Adjacencies (PLADJ) can be used at the class and total landscape level. At the total landscape level, it equals to the sum of the number of like adjacencies for each patch type divided by the total number of cell adjacencies in the landscape (eq. 9). The index value is expressed in percentage and changes in the range

of 0 to 100%. The metric can reach the minimum value when the patch types are maximally disaggregated and there are no like adjacencies. Index value increases with the degree of aggregation of cells the same classes. The parameter can reach the maximum value when all patch types are maximally aggregated (McGarigal et al. 1995, 2012, Urbański 2012):

$$PLADJ = \left(\frac{\sum_{i=1}^m g_{ii}}{\sum_{i=1}^m \sum_{k=1}^m g_{ik}} \right) \cdot 100 \quad [\%] \quad (9)$$

where:

g_{ii} – number of like adjacencies between cells of patch type (class) i based on the double-count method,

g_{ik} – number of adjacencies between cells of patch types (classes) i and k based on the double-count method.

Aggregation Index (AI) can be calculated at the class and the total landscape level. In the second case, the metric is computed as an area-weighted average class aggregation index, where each class is weighted by its proportional area in the landscape (eq. 10). The index value is expressed in percentage and changes in the range of 0 to 100%. The metric can reach the minimum value when the patch types are maximally disaggregated. The AI value can be increases as the landscape is increasingly aggregated and equals 100 when the landscape consists of a single patch (McGarigal et al. 1995, 2012, Urbański 2012):

$$AI = \left[\sum_{i=1}^m \left(\frac{g_{ii}}{\max \rightarrow g_{ii}} \right) \cdot P_i \right] \cdot 100 \quad [\%] \quad (10)$$

where $\max \rightarrow g_{ii}$ – maximum number of like adjacencies between cells of patch type (class) i based on the single-count method. Other symbols as in eq. (6) and (9).

Landscape Division Index (DIVISION) is used at the class and the total landscape level. At the total landscape level, it describes the probability that two randomly chosen cells in the landscape are not situated in the same patch (eq. 11). The value of the metric can be changed in the range from 0 to 1. The metric can reach the minimum value when the whole analyzed area consists of a single patch. Parameter achieves its maximum value when the landscape is maximally

subdivided. That is, when every cell is a separate patch (McGarigal et al. 1995, 2012, Jaeger 2000, Urbański 2012):

$$\text{DIVISION} = \left[1 - \sum_{i=1}^m \sum_{j=1}^n \left(\frac{a_{ij}}{A} \right)^2 \right] [-] \quad (11)$$

Symbols as in eq. (1).

Interspersion and Juxtaposition Index (IJI) is calculated at the class and the total landscape level. It is a measure of the spatial patches variability structure. The metric calculates the observed interspersion over the maximum possible interspersion for the given number of patch types (eq. 12). The index value is expressed in percentage and changes in the range of 0 to 100%. The metric can reach the minimum value when the distribution of adjacencies among unique patch types becomes increasingly uneven. The parameter approaches value 100 when all patch types are equally adjacent to all other patch types and the length of the borders between all pairs of patches types are the same (McGarigal et al. 1995, 2012, Kot & Leśniak 2006):

$$\text{IJI} = \frac{-\sum_{i=1}^m \sum_{k=i+1}^m \left[\left(\frac{e_{ik}}{E} \right) \cdot \ln \left(\frac{e_{ik}}{E} \right) \right]}{\ln(0.5[m(m-1)])} \cdot 100 [\%] \quad (12)$$

where e_{ik} – total length of edge in landscape between patch types (classes) i and k [m]. Other symbols as in eq. (3) and (6).

Euclidean Nearest-Neighbor Distance (ENN) is landscape parameter which is calculated at the patch level. It is a measure of the patches isolation. ENN equals the minimal distance to the nearest neighboring patch of the same type, based on shortest edge-to-edge distance. In the studies based on the total landscape level, the simple descriptive statistics such as measures of an average and a variability are analyzed (McGarigal et al. 1995, 2012, Urbański 2012).

RESULTS

The raw, unprocessed image of the topoclimatic map of the ONP region (Fig. 1) included 23 different topoclimatic separations, with a total of 13,789 patches in the analyzed area of approximately 90 km² (Tab. 1). Three groups were distinguished in the total area occupied by patches of the same type, the first characterized by a large spread (5.0–25%), second by a medium spread (2.5–5.0%) and third with spread 0–2.5%, rare and covering small areas. The first group includes areas with topoclimate types: 1.4 (1.4.1 & 1.4.2), 2.2, 2.3, 3.1, 4.2 (4.2.1 & 4.2.2) (Paszyński 1980) and 4.8 (Kot 2005, 2006) (Tab. 1), the second: 1.1, 1.2 (1.2.1 & 1.2.2), 1.3, 2.1, 3.2, 4.1, 4.3, and 5.2. The third group includes areas with topoclimate types: 3.3, 4.4, 5.1, 5.3, and 6.1 (Paszyński 1980), 4.5, 4.6, and 4.7 (Kicińska et al. 2001) and 7.1 (Bartuś 2014).

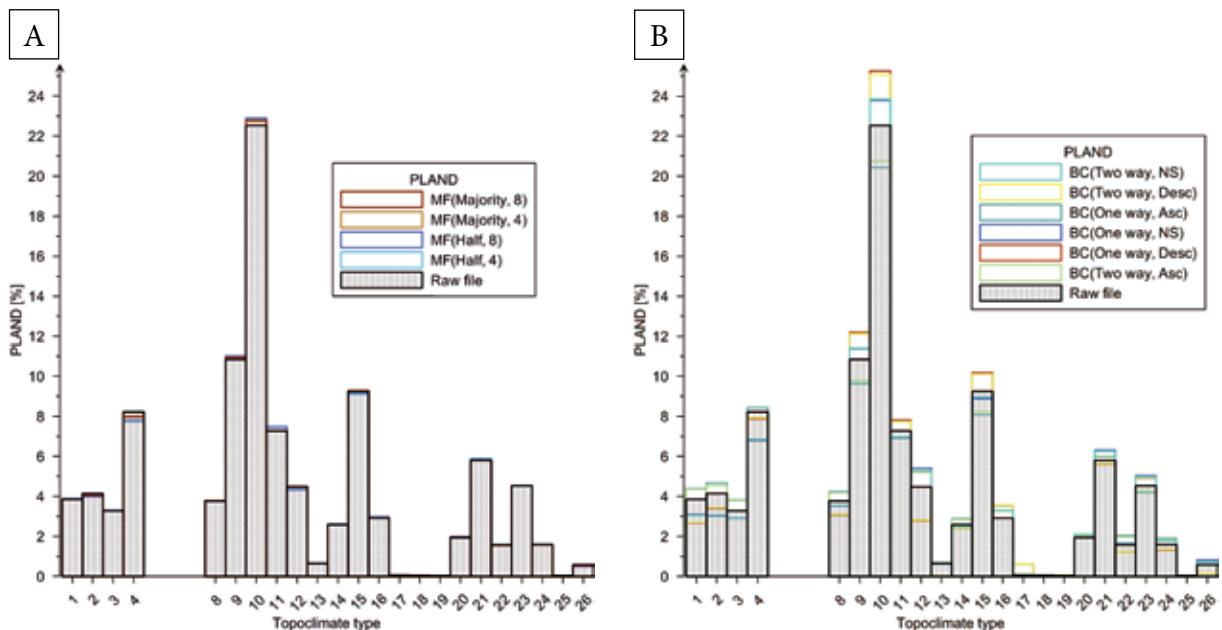


Fig. 2. Areas of certain type of topoclimate in the ONP region, calculated on the basis of: Raw Data (Raw file), Boundary Clean tool (A) and Majority Filter tool (B)

Table 1

The percentage of the study area (PLAND) occupied by various patches types in the raw image and in the images processed using the Boundary Clean and Majority Filter tools

Topoclimate type	ID	PLAND [%]										
		Raw file	Boundary Clean						Majority Filter			
			One Way			Two Way			Half		Majority	
			Asc	Desc	NS	Asc	Desc	NS	4	8	4	8
1.1	1	3.85	4.39	2.66	3.09	4.37	2.68	3.05	3.89	3.88	3.86	3.86
1.2.1 & 1.2.2	2	4.14	4.67	3.4	3.03	4.54	3.44	3.07	3.99	3.98	4.05	4.05
1.3	3	3.28	3.83	2.91	2.9	3.8	2.92	2.92	3.30	3.31	3.28	3.28
1.4.1 & 1.4.2	4	8.22	8.43	7.87	6.8	8.24	7.92	6.85	7.79	7.76	8.03	7.97
2.1	8	3.77	4.24	3.06	3.53	4.19	3.1	3.55	3.76	3.77	3.79	3.79
2.2	9	10.84	9.63	12.21	11.38	9.77	12.13	11.4	10.96	11.00	10.92	10.95
2.3	10	22.54	20.43	25.25	23.79	20.75	25.08	23.84	22.82	22.90	22.73	22.78
3.1	11	7.27	6.91	7.81	6.93	6.95	7.74	6.96	7.48	7.47	7.27	7.26
3.2	12	4.47	5.24	2.78	5.41	5.2	2.82	5.29	4.33	4.34	4.48	4.5
3.3	13	0.64	0.69	0.03	0.63	0.68	0.03	0.64	0.65	0.65	0.64	0.64
4.1	14	2.58	2.89	2.4	2.46	2.85	2.42	2.48	2.62	2.60	2.58	2.58
4.2.1 & 4.2.2	15	9.25	8.07	10.18	8.88	8.23	10.12	8.96	9.10	9.14	9.29	9.28
4.3	16	2.92	3.3	3.54	3.28	3.26	3.56	3.31	2.94	2.97	2.91	2.92
4.5	17	0.06	0.09	0.61	0.06	0.09	0.62	0.06	0.07	0.06	0.06	0.06
4.6	18	0.05	0.07	0.01	0.05	0.07	0.01	0.05	0.05	0.05	0.05	0.05
4.7	19	0.04	0.06	0.03	0.04	0.06	0.03	0.04	0.04	0.04	0.04	0.04
4.4	20	1.94	2.09	1.94	2.06	2.11	1.97	2.07	1.98	1.99	1.95	1.96
4.8	21	5.8	5.99	5.61	6.31	5.96	5.64	6.26	5.89	5.85	5.79	5.8
5.1	22	1.58	2.04	1.22	1.66	1.98	1.24	1.65	1.58	1.56	1.55	1.55
5.2	23	4.53	4.2	4.92	5.04	4.26	4.95	4.99	4.54	4.53	4.54	4.53
5.3	24	1.59	1.91	1.33	1.84	1.87	1.36	1.8	1.61	1.59	1.58	1.58
6.1	25	0.04	0.05	0.04	0.05	0.05	0.04	0.05	0.04	0.04	0.04	0.04
7.1	26	0.58	0.81	0.18	0.82	0.71	0.18	0.73	0.57	0.51	0.54	0.53

In the table: Topoclimate type – topoclimate type number in Paszyński (1980) classification with Kicińska et al. (2001), Kot (2005, 2006) and Bartuś (in preparation) modifications; Raw file – raw raster image; Boundary Clean (BC), Majority Filter (MF) – raster ArcGis generalization tools; One way – BC tool procedure which performs expansion and shrinking once; Two way – BC tool procedure which performs expansion and shrinking once, then performs an additional shrinking and expansion with the priority reversed; Asc – sorts patches in ascending order by size; Desc – sorts zones in descending order by size; NS – does no sorting by size; Half – MF tool option where half of the cells must have the same value and be contiguous; Majority – MF tool option where a majority of cells must have the same value and be contiguous; 4 – The kernel of the filter will be the four direct (orthogonal) neighbors to the present cell; 8 – The kernel of the filter will be the eight nearest neighbors (a 3-by-3 window) to the present cell.

An analysis of the results of generalization made using the BC tool (Tab. 1, Fig. 2), showed a strong correlation of the observed effects of

processing with the method of sorting patches. As expected, Asc sorting caused the patches classes with larger total surface areas to lose some of their

space in favor of classes with lower total surface areas. *Desc* sorting resulted in the opposite effect. Classes of patches with larger total surface areas were expanded at the expense of the classes with smaller total surface areas.

Image analysis of the results of generalization that was carried out using a procedure independent of the size of planes (*NS*) showed a growth in the areas of classes with a higher value of the analysed attribute ($ID > 8$) at the expense of the classes with lower attribute values (Tab. 1).

For both available procedures (*One Way* and *Two Way*), generalization carried out with the same priorities (*Asc* or *Desc*), gave similar results. For both sets of *Asc* procedures (*One Way* and *Two Way*) and *Desc* procedures (*One Way* and *Two Way*), the total surface areas of certain types of topoclimates are almost identical (Fig. 2A). A similar effect can be observed for generalization procedures dependent on the value of the attribute describing the patches (*ID*).

Even a superficial analysis of the effects of generalizations conducted using the MC tool demonstrates the different (compared to BC) nature of the modifications. In terms of total changes in the areas, the processed images are slightly different from the raw image (Fig. 2B). The pairs of images obtained by modes: *Majority* (calculated for 4 and 8 adjacent cells) and *Half* (calculated for 4 and 8 adjacent cells), one can observe a similarity in the area occupied by each class.

Analyses of the area of the largest patches index (LPI), absolute (TE) and the relative (ED) length of the edge of the patches indicators, carried out at landscape level, and are among the most intuitive measures of the intensity of generalization processes. A graph of the TE index variation (Fig. 3) indicates that using the BC tool reduces the length of the plane edge (original level 2902.23 km) by 23–30%, and in the case of the MF tool – by 13–19%. The highest level of reduction was observed while using the BC tool with the options: *Two Way* & *Desc*. The average edge density in this case decreased from about 324.5 m/ha to 227.3 m/ha.

An interesting insight is provided by observing the variability of the LPI index. As seen in Figure 3, the size of the largest patch is the same, typically slightly higher than in raw data, with

a value of 4.65–5.10% of the total study area. The exception is the low value of the index resulting from generalization performed using the BC tool (*One Way, Asc*). The sharp reduction in the surface area of the largest patch is forced by the increasing sorting priority and development of patches with smaller area at the expense of those with larger area. The lack of a similar phenomenon in the case of BC generalization (*Two Way, Asc*) was caused by additional sorting with reversed priority, which takes place in the second step of the procedure.

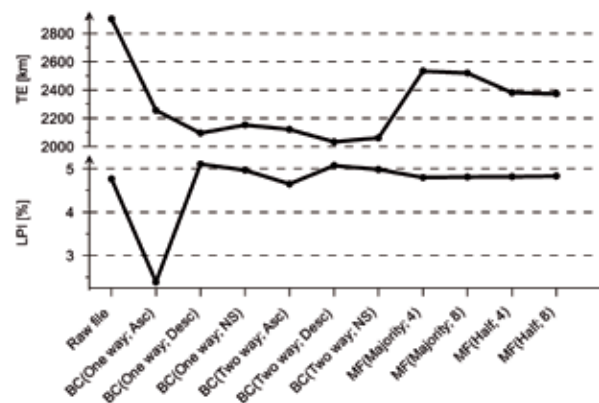


Fig. 3. Variability of the TE and LPI metrics

The measures of shape (SHAPE FRAC) are usually calculated at the levels of patches or classes. At the landscape level, the analysis includes simple statistical parameters. The analysis used arithmetic means and standard deviations of both measures of shapes analysed. In a similar way, the values of indicators illustrate the differences in the shape of patches (Fig. 4). All generalized images are characterized by slight, relative simplifications of patch shapes compared to the raw image. Detailed analysis helped to reveal the differences in the effects of generalization. The highest level of simplification in the shapes of separations is obtained using the BC tool combined with sorting by size in descending order (*Desc*) (Fig. 4). For both analysed variants (single and double processing) a similar level of generalization of the output images was achieved. Using the MF tool (*Half, 8*) gave only slightly worse results. The weakest effect simplifying the shapes of patches can be observed for tools: MF (*Majority, 8*) and BC (*One Way, Asc*).

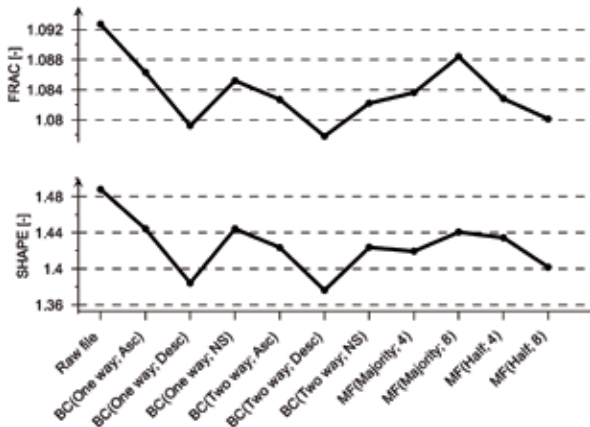


Fig. 4. Variability of the SHAPE and FRAC metrics

Landscaping measures of diversity belong to the key indicators used in the ecological analyses of bio- and geodiversity. Generalization of images, which are then subjected to such research, is a potential source of errors and may cause interpretation problems. The performed analysis showed that the use of the BC tool causes changes in the structure of the image, which may affect diversity assessments (Fig. 5). The results show that the use of the Asc parameter causes an increase in the relative class variation by about 3%, and that using the Desc parameter results in a diversity decline of about 4%. The probability of two randomly chosen cells in the image belonging to two planes of different classes when using the Asc parameter is 0.91, and when using Desc is 0.89. Using the NS option and generalization using the MF tool, did not cause significant changes in diversity. High relative uniformity values of the index indicate a very proportional distribution of the classes of the analysed characteristics.

The most intuitive parameter illustrating the possibilities of generalization of each of the analysed tools is the total number of patches (NP). The original image had a total of 13,789 units. Modification of the spatial structure of images carried out with the use of the BC tool resulted in a decrease to a minimum level of 8,162 planes (*Two Way, NS*) and a maximum of 9,392 planes (*One Way, Asc*) (Fig. 6). This corresponded to a total reduction of the number of planes of around 32–41%. An analysis of NP variability observed for the MF tool showed a much lower relative level of reduction. The greatest simplification was observed in the image obtained by MF (*Half, 4*). The total number of planes in this case was 10,084 units. MF reduced the total number of units by about 10–27% (Fig. 6).

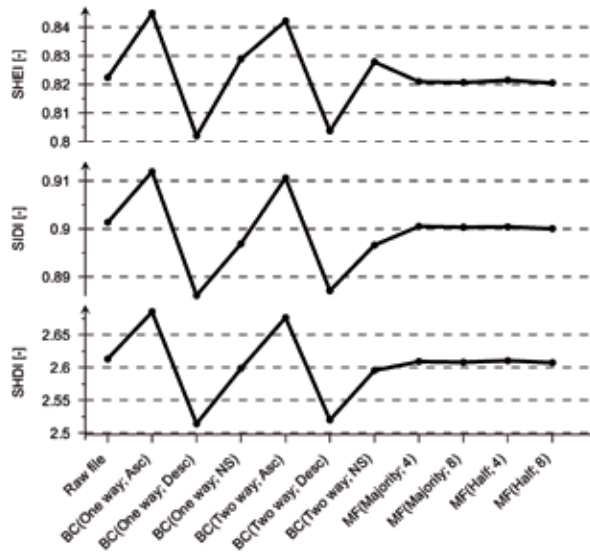


Fig. 5. Variability of the SHDI, SIDI and SHEI metrics

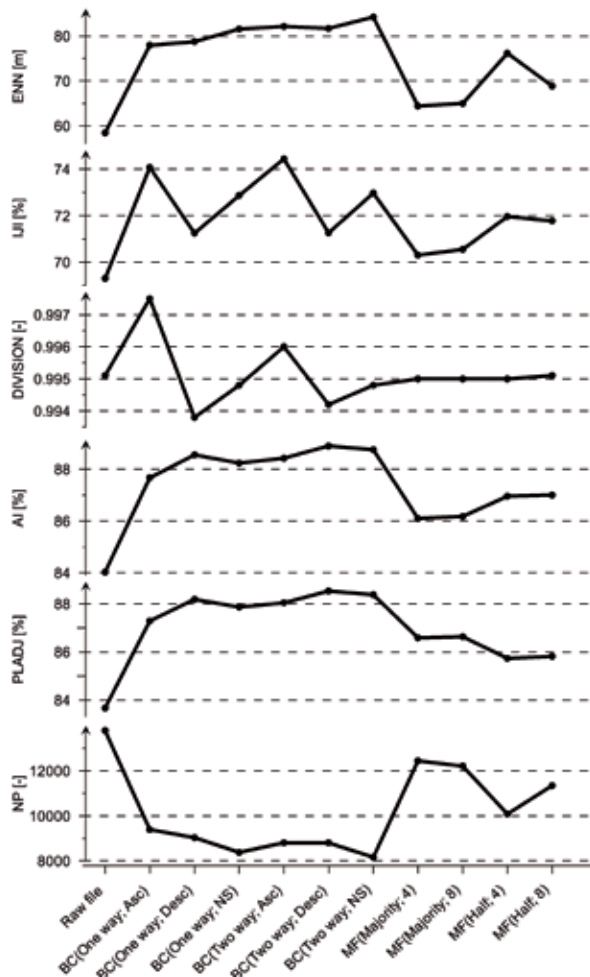


Fig. 6. Variability of the NP, PLADJ, AI, DIVISION, IJI and ENE metrics

The indicators that best reflect generalization performance are also the group of parameters

that use information about the cells adjacent to each other. Pixels may be adjacent to cells of the same class or other classes. Generalization should always cause an increase in the number of adjacent cells of the same type. PLADJ and AI are the parameters that illustrate this dependence most explicitly (Fig. 6). For all analysed variants of calculations, significant increases in homogeneous neighborhoods were observed. Two levels of generalization were also clearly visible. The raw image was characterized by cells neighboring other cells of the same class at approximately 84%. All variants of calculations using the BC tool caused an increase in the value of this parameter to about 88%, and in the case of MF tool, to about 86%.

The DIVISION fragmentation ratio analysis (Fig. 6), carried out for all output images analysed, indicated a very high probability of two randomly selected cells not belonging to the same patch. A slight increase in the probability observed in images generated using the BC tool with the *Asc* option, may be associated with a higher level of observed diversity compared to other options (Fig. 5).

IJI ratio analysis (Fig. 6) revealed that for all tested variants of generalization, there was a change in the ratios of the plane boundaries. In the raw image, the value of the IJI index, which measures the degree of sustainability of the common length of boundaries between the classes, was 69.3%. Using the BC tool resulted in obtaining a relatively higher uniformity in the length of boundaries than using the MF tool. Changing the tool options resulted in further diversification of the results. The highest degree of boundary sustainability of characterized images obtained using the tools were: BC with the *Asc* option (about 74%) and BC with the *NS* option (about 73%). The analyses performed for the MF tool showed a correlation between the observed IJI values and the assumed number of neighboring cells, allowing for the change of the value of the central cell. Selecting the Majority option resulted in a lower degree of homogeneity in boundary length of separations (about 70%) than selecting the *Half* option (about 72%).

The final analysed structural indicator was the distance to the nearest element of the class (ENN) (Fig. 6). The interpretation used arithmetic average values of distances calculated for each class of elements. Studies have shown that in all variants analysed, there was a significant increase in the isolation of elements. In the original image, the

average distance between the boundaries of the elements of one class was about 58 m. Regardless of the options used, generalization made with BC resulted in an increase in the distance to about 80 m, which corresponds to a growth of over 30%. The highest distance value between patches of the same class (about 84 m) was observed in an image created by the tools *Two Way*, and *NS*. Using MF provided a significantly lower level of element isolation growth. Using the MF resulted in an increase in the distance to approximately 65–75 m. Among the different computational variants of the MF tool, the highest growth was characterized by an image generated using the options *Half* and 4.

SUMMARY

Boundary Clean (BC) is a powerful tool that provides a high degree of reduction of the number of items and the total length of the patch boundaries (Fig. 7, Tab. 2). The most characteristic structural features of the images processed with the help of this tool include: a relatively large increase in the isolation of elements and a clear increase in cell aggregation under the same class. The lack of a clear impact of the tool on changing the size of the largest element of the image is fairly typical. The exception is the *One Way, Asc* procedure, which results in a rapid decrease in its size. The increased focus of pixels of the same type results in a change of shapes of patches that are more regular. The highest level of change in the shape of these elements is consistently observed in images generated with the *Desc* priority. Structural changes of images are also reflected in the level of landscape fragmentation. Research has shown that choosing ASC sorting caused an increase in the level of patch fragmentation, while *Desc* sorting caused the opposite effect and resulted in the greater focus. The described changes result also in increased level of sustainability of common boundary lengths between elements of successive classes. The highest level of homogenization was observed in images created with the *Asc* priority. Interesting results were observed by analyzing the level of landscape diversity. It turns out that this may be influenced by the type of sorting method used. The analysis showed that choosing *Asc* sorting caused a significant increase in the level of diversity, while *Desc* sorting resulted in decreased diversity.

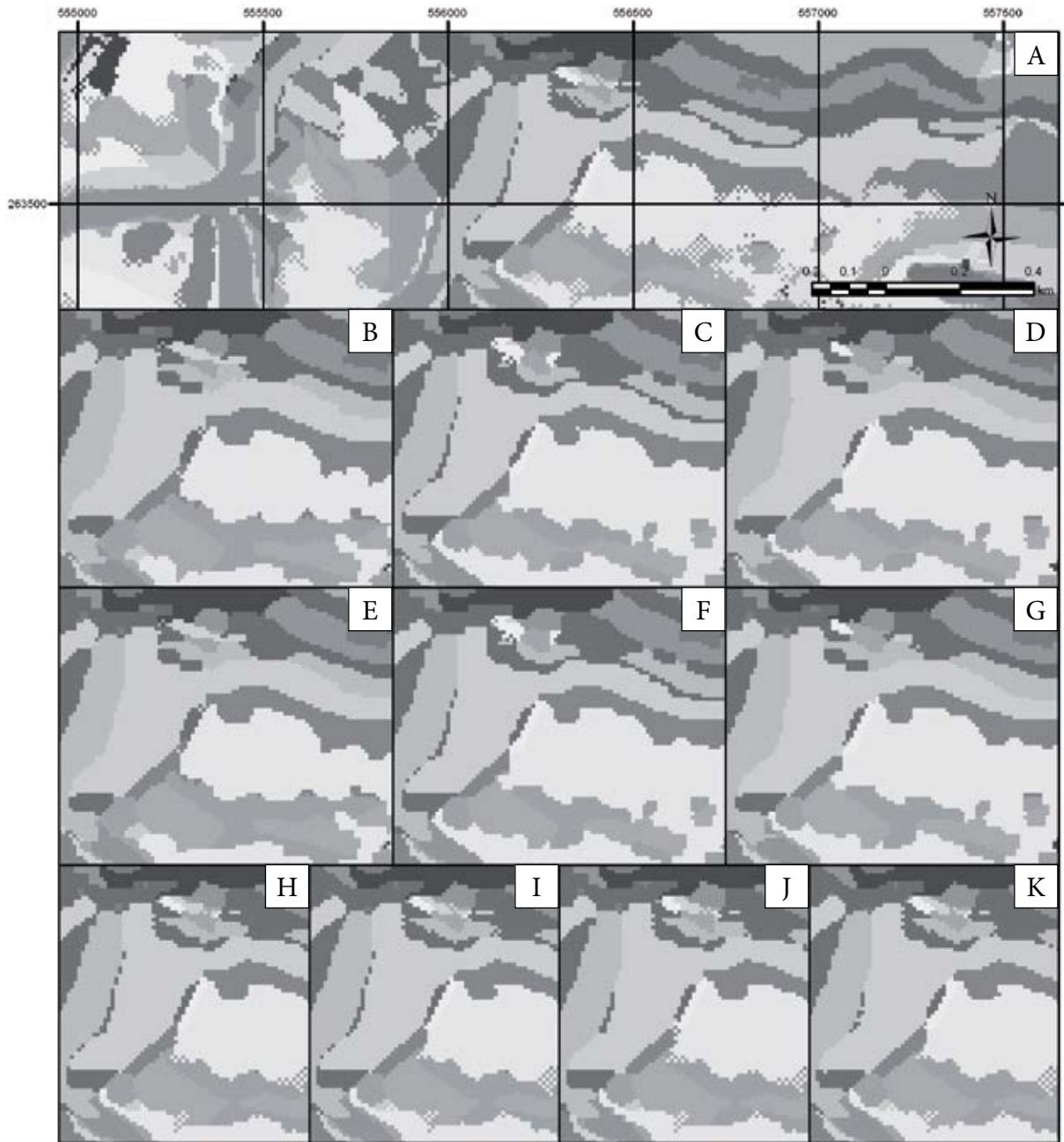


Fig. 7. Comparison of fragment of the raw raster image (A) and effects of generalization obtained using the Boundary Clean; B) One Way, Asc; C) One Way, Desc; D) One Way, NS; E) Two Way, Asc; F) Two Way, Desc; G) Two Way, NS and Majority Filter; H) Majority, 4; I) Majority, 8; J) Half, 4; K) Half, 8

Using the Majority Filter (MF) tool causes changes in the output images, which are less pronounced in the values of the analysed landscape indicators (Fig. 7, Tab. 2). The property that can be seen is a clear consequence of the results obtained, which to a small extent depends on the adopted analysis variant. Even though the tool reduces the number of patches and their boundary lengths, as well as causing a slight

simplification of their shape, it practically does not change the size of the largest element or the degree of landscape fragmentation. Its diversity also remains unchanged. However, one can observe a slight increase in the level of pixel aggregation, a slight increase in the distance between the elements of the same type, and a slight increase in boundary sustainability between successive image classes.

Table 2
Relative landscape metrics variability for each variant analysis

Generalisation type			Landscape metrics variability relative to the raw image								
			NP	TE	LPI	FRAC/SHAPE	SHDI/SIDI	ENN	PLADJ/AI	DIVISION	IJI
Boundary Clean	One Way	Asc	↓	↓	↓	↓	↑	↑	↑	↑	↑
		Desc	↓	↓	–	↓	↓	↑	↑	↓	↑
		NS	↓	↓	–	↓	↓	↑	↑	–	↑
	Two Way	Asc	↓	↓	–	↓	↑	↑	↑	↑	↑
		Desc	↓	↓	–	↓	↓	↑	↑	↓	↑
		NS	↓	↓	–	↓	↓	↑	↑	–	↑
Majority Filter	Majority	4	↓	↓	–	↓	–	↑	↑	–	↑
		8	↓	↓	–	↓	–	↑	↑	–	↑
	Half	4	↓	↓	–	↓	–	↑	↑	–	↑
		8	↓	↓	–	↓	–	↑	↑	–	↑

The arrows indicates the relative (high/medium/low) level of increase/decrease values of the landscape metrics.

CONCLUSIONS

Generalization is one of the most important stages of work with cartographic data. Every modern GIS system has a variety of tools for generalization procedures of raster and vector images. In the case of ArcGIS software developed by ESRI, which is a world leader in the processing of spatial data, the most important tools for generalization of raster images include: Boundary Clean and Majority Filter. All changes resulting from their operations are irreversible and cause the loss of the original information. The operation of generalization tools is reflected in changes in the raster structure. Generalization tools should therefore be used in a conscious way, taking into account the consequences stemming from their use.

Fragstat software was used analyzing the structural modifications of the output images and assessing the effects of generalization. A variety of landscape metrics allowed a very efficient and in-depth analysis of the results.

Generalization of raster images, understood in a conventional way, involves modifying the structure of the image, while maintaining its general characteristics. In practice, depending on the quality of the source image, it requires the use of more or less complex processing procedures. These procedures most often involve simplifying patch boundaries, removing non-essential “islands” and some narrow but elongated elements.

If we only wish to introduce subtle modifications in the output image, involving removal of individual strange pixels or their groups, i.e. so-called noise, then using one of the variants of the MF tool (e.g. *Half*, 4) is the most suitable. The image after modifications made using this tool was slightly different from the raw image. The reduction of the total length of all edges was approximately 13–19% and the reduction in the number of patches was about 10% (up to 27%). If, however, we expect a greater level of interference in the structure of the source image, e.g. simplification of a group of patches that form a checkerboard pattern, using the BC tool becomes necessary.

BC interferes with the structure of images to a much greater extent than MF. Analyses show that, depending on the options used, it allows the reduction of the total length of all edges by 23–30% and a reduction of the number of patches by 32–41%. While using BC, the key is to choose the type of sorting (*Asc* or *Desc*). Choosing the right option will respectively result in an increase or decrease in the importance (size) of patches with small surface areas and a decrease or increase in the importance of patches with relatively large sizes. The total size of space added or removed depends on the size of the patches. Patches that are frequent in a given area, or occupy more space, gain or lose more surface area than rare classes with low total surface areas. The type of sorting selected is important for assessing

diversity. Using *Asc* sorting consistently resulted in a 3% increase in diversity indicators, while *Desc* sorting caused their 4% decrease. The results should be taken into account in future assessments of bio- and geodiversity.

Studies have shown little effect from a BC tool procedure being performed once or twice. The results obtained for both variants were not significantly different from each other. An exception is the clear effect of BC generalization (*One Way, Asc*) on the size of the patches.

In some situations, using the *NS* tool option in the BC tool may be more beneficial than a generalization that uses sorting, whose criteria are based on the size of the separations. For example, it may be necessary to use a more subtle type of sorting, one that can be declared by the user using the ID value.

Execution of this work was possible due to the financial support provided by the Polish Scientific Research Committee, as part of the statutory works No 11.11.140.173.

REFERENCES

- Bartuś T., 2014. Topoklimaty Ojcowskiego Parku Narodowego. [in:] Kalinowska-Szymczak A. (ed.), *Kalejdoskop GIS, tom 3*, Esri Polska sp. z o.o., 102–103.
- Beat P. & Weibel R., 1999. Using Vector and Raster-Based Techniques in Categorical Map Generalization. *Third ICA Workshop on Progress in Automated Map Generalization, Ottawa, 12–14 August 1999*, 1–14.
- Chrobak T., 2005. Komputerowa redakcja kartograficzna wspomagana automatyczną generalizacją. *Geoinformatica Polonica*, 7, 7–14.
- Daley N., Goodenough D.G., Bhogal A.S., Bradley Q., Grant J. & Zin Y., 1997. Comparing Raster and Object Generalization, *Proc. IGARSS'97, Singapore*, 677–679.
- ESRI, 2012. ArcGIS Desktop: Release 10.1. Redlands, CA: Environmental Systems Research Institute.
- Harrie L. & Sarjakoski T., 2002. Simultaneous Graphic Generalization of Vector Data Sets. *GeoInformatica*, 6, 3, 233–261.
- Hengl T., 2006. Finding the right pixel size. *Computers and Geosciences*, 32, 1283–1298.
- Jaeger J.A.G., 2000. Landscape division, splitting index, and effective mesh size: new measures of landscape fragmentation. *Landscape Ecology*, 15, 115–130.
- Kicińska B., Olszewski K. & Żmudzka E., 2001. Uwagi o wykorzystaniu klasyfikacji J. Paszyńskiego do kartowania topoklimatycznego (z doświadczeń Zakładu Klimatologii Uniwersytetu Warszawskiego). [in:] Kuchcik M. (ed.), *Współczesne badania topoklimatyczne. Dokumentacja Geograficzna*, 23, IG i PZ PAN, Warszawa, 143–151.
- Kot R., 2005. Ocena georóżnorodności klimatu na przykładzie fordonskiego odcinka doliny dolnej Wisły i jej otoczenia w skali 1:25.000. [in:] Szponar A. & Horskak-Schwarz S. (eds), *Struktura przestrzenno-funkcyjna krajobrazu. Problemy ekologii krajobrazu*, XVII, 257–271.
- Kot R., 2006. *Georóżnorodność – problem jej oceny i zastosowania w ochronie i kształtowaniu środowiska na przykładzie fordonskiego odcinka doliny dolnej Wisły i jej otoczenia*. Towarzystwo Naukowe w Toruniu, Uniwersytet Mikołaja Kopernika, Toruń.
- Kot R. & Leśniak K., 2006. Ocena georóżnorodności za pomocą miar krajobrazowych – podstawowe trudności metodyczne. *Przegląd Geograficzny*, 78, 1, 25–45.
- McGarigal K., Cushman S.A. & Ene E., 2012. *FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps*. Computer software program produced by the authors at the University of Massachusetts, Amherst, [on-line:] <http://www.umass.edu/landeco/research/fragstats/fragstats.html> (access: 10.06.2013).
- McGarigal K. & Marks B.J., 1995. *FRAGSTATS: spatial pattern analysis program for quantifying landscape structure*. USDA Forest Service, Technical Reports, PNW-GTR-351, Portland.
- McMaster R., 1992. *Generalization in Digital Cartography*. Association of American Geographers, Washington
- Milne B.T., 1988. Measuring the fractal geometry of landscapes. *Applied Mathematics and Computation*, 27, 67–79.
- Paszyński J., 1980. Metody sporządzania map topoklimatycznych. [in:] *Metody opracowań topoklimatycznych. Dokumentacja Geograficzna*, 3, 13–28.
- Raines G.L., 2002. Description and comparison of geological maps with FRAGSTATS – a spatial statistics program. *Computers and Geosciences*, 28, 169–177.
- Shannon C.E. & Weaver W., 1949. *The mathematical theory of communication*. Univ. Illinois Press, Urbana.
- Simpson E.H., 1949. Measurement of diversity. *Nature*, 163, 688–688.
- Tadeusiewicz R., Korohoda P., 1997. Algorytmy i metody komputerowej analizy i przetwarzania obrazów. *Materiały do Szkoły Letniej, Tempus S_JEP-07181-94*, Kraków.
- Urbański J., 2012. *GIS w badaniach przyrodniczych*. Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.
- Veregin H. & McMaster R., 1997. Data Quality Implications of Raster Generalization. *ACSM ASPRS Annual Convention and Exposition*, 5, 267–276.
- Weibel R. & Jones Ch.B., 1998. Computational Perspectives on Map Generalization. *GeoInformatica*, 2, 4, 307–314.
- Zwierzchowska I., Stępniewska M. & Łowicki D., 2010. Możliwość wykorzystania programu Fragstats w badaniach środowiska przyrodniczego. *Przegląd Geograficzny*, 82, 1, 85–102.