https://doi.org/10.7494/geom.2024.18.4.97

Sylwia Borkowska[1], Elżbieta Bielecka[2], Krzysztof Pokonieczny[3]

Weight Impact on Comparative Evaluation of Topographic Data

Abstract: The paper addresses the problem of weighting in an analysis that supports the selection of a categorical data set according to user needs. Using the Relative Change (RC) of the Compound Correspondence Index (CCI), it is shown that weights have a significant impact on user choice – reaching extreme values in both urbanized and forested areas. Decreasing the weights from 0.25 to 0.17 in forested and built-up areas resulted in the maximum variations that were seen in the hot spot maps, with cold areas generally corresponding to built-up regions and hot areas to forested areas. The analysis covers seven counties that are located in different regions of Poland: Pomerania, Podlasie, Mazovia, Greater Poland and the Beskidy Mountains.

Keywords: quantitative data, OSM, sensitivity analysis, MCDA, TOPSIS

Received: May 6, 2024; accepted: July 22, 2024

[1] Military University of Technology, Faculty of Civil Engineering and Geodesy, Institute of Geospatial Engineering and Geodesy, email: sylwia.borkowska@wat.edu.pl (corresponding author), https://orcid.org/0000-0003-3183-1512

[2] Military University of Technology, Faculty of Civil Engineering and Geodesy, Institute of Geospatial Engineering and Geodesy, email: elzbieta.bielecka@wat.edu.pl, https://orcid.org/0000-0003-3255-1264

[3] Military University of Technology, Faculty of Civil Engineering and Geodesy, Institute of Geospatial Engineering and Geodesy, email: krzysztof.pokonieczny@wat.edu.pl, https://orcid.org/0000-0001-9114-5317

1. Introduction

The comparative analysis of spatial quantitative data is often used to select data sets that are suitable for a user's purpose. This generally uses multi-criteria evaluation [1] based on available MCDA (multi-criteria decision analysis) applications. As a decision-support tool, the main objective of MCDA is to assist decision-makers by providing decision options according to accepted criteria. As noticed by [2], however, the criteria should be rational, transparent and non-overlapping. Despite their high diversity, multi-criteria decision applications share some characteristics: (1) a determinate number of comparable alternatives; (2) many criteria against which the alternatives are compared; (3) measurable values that define the quality of the alternative with respect to each of its criteria; and (4) weights for each of the criteria that determine the importance of each of them. Researchers [e.g., 3, 4] have claimed that weights and the choice of how to measure the distances between given criteria are, in general, fundamental and predominantly influence the results.

Many criteria-weighting rules have been presented in the MCDA literature [5, 6]. Their variety leads to the following question: how does the choice of weights affect the final ranking of decision alternatives? Hence, this study aims to analyze the weight impact in a fit-for-purpose assessment of topographical data. It uses the TOPSIS (technique for order of preference by similarity to ideal solution) methodology as well as the Comparative Compound Index (CCI) that was previously introduced in [1]. The CCI was calculated separately for each county in our study; hence, it was demarcated as local. The presented research used and summarized the results of the suitability analysis of the topographical data that was published in [1, 7, 8]. Therefore, the sensitivity analysis of the TOPSIS ordering was carried out on the same seven counties and two topographical data sets; namely, official data that is maintained by the Head Office of Geodesy and Cartography – Database of Topographic Objects (BDOT10k) as well as volunteer data – OpenStreetMap (OSM). This work is part of the discourse on the importance of attribute weights in final TOPSIS ratings. The study confirms the significant influence of the adopted weights on the usability evaluation of the data and the final decision that is made. The novelty of the research lies in the complex universal methodological approach that allows for an evaluation of categorical data; i.e., qualitative data grouped into categories [1] rather than measured data that refers to a form of information that is stored and identified by names or labels (e.g., forest, river, lake, city) according to user-defined criteria. To the best of our knowledge, this research concerns a problem that has not yet been addressed by researchers regarding changes in final TOPSIS rankings as related to changes in attribute weights at the pixel level as well as the relationship between changes in TOPSIS rankings and land use.

The paper is structured as follows: Section 2 describes selected publications on MCDA sensitivity analysis, focusing on the use of TOPSIS and weighting methods.

Section 3 describes the materials and methods that were used, Section 4 presents the results of the sensitivity analysis, and Section 5 is a scientific discussion of the obtained results. Finally, Section 6 concludes the paper.

2. Literature Review

TOPSIS is one of the most popular multi-criteria decision techniques [7, 9–11]. Based on a thorough literature review of TOPSIS applications, Behzadian et al. [9] found that the TOPSIS model had been used mainly in technical and socio-economic research but still needed a broader focus on environmental issues. A similar opinion was shared by Zyoud and Fuchs-Hanusch [10], who found that TOPSIS was mostly used in supply chain management and sustainability research, while analytic hierarchy process (AHP) was predominant in risk modeling and analysis in Geographic Information Systems [10]. The traditional TOPSIS model suffers from correlations between criteria [11] because it uses Euclidean distance, which does not take correlation into account; therefore, its results are affected by overlapping information. To overcome this, the correlation of the attributes should be checked a priori [12]. Furthermore, Li et al. [4] observed that TOPSIS studies generally assumed that parameter weights were invariant and mostly subjectively determined by experts. Yet, only a few studies have included TOPSIS sensitivity analyses based on weight changes [e.g., 4, 13–18], although the results of previous analyses are difficult to generalize today. Criteria weights have various interpretations and implications that are misunderstood and neglected – not only by decision makers, but also by academics. Kobryń and Prystrom [17] found that rating alternatives in TOPSIS strongly depended on the nature of the accepted criteria and the version of TOPSIS (classical, interval, or fuzzy). Choo et al. [15] identified several plausible interpretations of criteria weights and their appropriate roles in decision models, such as scale validity, commensurability, criteria importance, and rank consistency. They also insisted on defining the concept of criteria importance, noting that the “proper interpretation and application of criteria weights would improve the quality of results obtained by using the variety of MCDM models” [15]. Based on investigations of some MCDA applications and available weighting methods on the objectivity of the resulting rankings, Bączkiewicz et al. [16] observed that (1) a proper method for the problem to be adequately solved was essential, (2) a comparative analysis of the results was strongly recommended, and (3) a selection of criteria weights that reflected the preferences of the decision-maker were essential parts of MCDM.

Więckowski and Zwiech [19] used TOPSIS and entropy for selecting energy-efficient materials. The results of analyzing the correlations between weighting and MCDA methods came to the conclusion that, although there were similarities between the rankings, they were not so significant that the weighting methods could be applied equally without changes in the final rankings. Chen et al. [20] used sensitivity analysis to examine the dependence of model results on input parameters, identify criteria that are particularly vulnerable to weight changes, and show the impact of changing the criteria weights on the model results in the spatial dimension as well as their relative impact on the final evaluation results. The study was carried out using the example of assessing the suitability of irrigated farmland in Australia. The authors of [20] altered and examined the original weights for the five different criteria over a range of 40 simulations using a method of deviating the weights from a base range, defined as a limited set of discrete percentage changes (±20%) in which the weight of each criterion was varied by 1%. A similar study of parameter-sensitivity analysis for determining the variability in the results caused by different input weights for four criteria (climate, soil, slope and erosion) was conducted for land suitability for sorghum cultivation in the Republic of Yemen by [21]. Sixteen weighting schemes were constructed and related to the layers of the criteria map. The results showed that slope and soil were highly sensitive elements in the suitability classification, while climate and erosion were moderately sensitive. Liern and Pérez-Gladish [22] proposed a new TOPSIS approach in which the weights were not determined a priori in an exact way. Weights were considered to be decision variables in a set of optimization problems whose goal was to maximize the relative closeness of each alternative to the ideal solution. The result was a new index of relative proximity that was a function that depended on the values of the weights. The method [22] can be useful in such decision-making situations where it is difficult to determine precise subjective weights.

Undoubtedly, the data and parameter weight burdened the final results of the analysis; hence, wide-ranging and thoughtful TOPSIS sensitivity is still challenging. Our research contributes to the relatively recent discussion of the influence of initial parameters on the results of multi-criteria and multi-attribute analyses (exemplified by TOPSIS).

3. Material and Methods

The study focuses on a sensitivity analysis of TOPSIS weighting in ranking the local Compound Correspondence Index (CCI) that was previously described in detail in [1] expressed as the fitness-for-purpose of six types of topographical objects that are stored in OpenStreetMap (OSM) and the National Database of Topographic Objects (BDOT10k); namely, buildings, forests, water bodies, roads, railroads and rivers. The rationale behind the choice of topographical objects is their clear and unambiguous definition in both databases and their importance in analyses of sustainable development and crisis management. They are also consistent with the authors' previous research on assessing the usability of topographical data. The research was carried out via the following three steps: (1) ranking the local CCI using the TOPSIS method and equal weighting; (2) comparing the CCI ranking results of expert subjective and equal weighting by the Relative Change (RC_{CCI}); and (3) spatial and statistical analyses of RC_{CCI}. The priority of this study is to answer the following research questions:

1. Does the combination of weights that are used in the local CCI calculation affect the final hexagonal pixel ranking? If so, by how much?
2. Do changes in the local CCI values that are expressed as relative change RC_{CCI} cluster spatially?
3. Are high and low RC_{CCI} values related to land cover types?

3.1. Study Area

Studies were conducted in seven Polish counties; these were characterized in Borkowska et al. [1] and are shown in Figure 1. Słupski County, the largest of the counties (2,300 km²), is situated in the northern part of Poland (along the Baltic Sea coastline), and Sokólski County (along the Polish-Belarusian border). In the central part of Poland (and belonging to the Warsaw agglomeration) are located Otwocki and Piaseczno (Piaseczyński) Counties (each with an area of more than 600 km²). Ostrowski and Międzyrzecki Counties (with areas of more than 1,100 km² each) are located in the western part of Poland. With an area that is comparable to each of the previous two, Sanocki County is situated in southern Poland (near the border with Slovakia).

Fig. 1. Study area: locations of analyzed counties

Source: [1]

The geographical and geopolitical locations of the counties, their sizes, different use structures and levels of urbanization determined their representativeness in the conducted analyses.

3.2. CCI Rating by TOPSIS

The TOPSIS technique aims at gaining an order preference that is similar to an ideal solution; i.e., a hypothetical solution with maximum benefits and minimum costs of the criteria that are used (attributes or alternatives). The best alternative is that which is nearest to the positive ideal solution and furthest from the negative one [11]. The similarity (or difference) is described by the Euclidean or Mahalanobis geometric distance; in this study the Euclidean distance was applied. The ideal solution and the negative one are examined based on the maximum (or minimum) values of the distances. As mentioned by [18], the TOPSIS method allows for trade-offs between criteria, as it allows a poor performance on one criterion to be ignored in favor of a good performance on another. When choosing the best alternative, the TOPSIS technique is comprised of the following main steps – normalizing the decision matrix, calculating the weighted normalized matrix, calculating the ideal positive and negative solutions, calculating the separation measure and calculating the relative closeness and alternative rankings. The steps that are mentioned above are followed by establishing non-overlapping criteria and their weights [11, 18].

The following subsection provides a comparison of CCI ratings. The CCI values are based on criteria such as differences in the areas that are covered by buildings, forests, and water bodies as well as the lengths of roads, railroads and rivers that are assigned to a 1 km^2 hexagonal grid [1, 8]. The CCI synthetic indicator, which describes the differences between the two studied topographical data sets (OSM and BDOT10k), was developed by using the classical TOPSIS method in two approaches. The first assumes varied weights, while the second assumes equal weights of the studied topographical objects. For each 1 km^2 hexagonal grid, the differences for the lengths or areas of the OSM topographical objects that were studied against the BDOT10k objects were calculated in ArcGIS Pro environment according to Equation (1):

$$x_i = \left| \mathrm{BT}_i - \mathrm{OSM}_i \right| \tag{1}$$

where:

x_i, $i = 6$ – difference value of topographical objects (buildings, forests, water bodies [area] or roads, railroads, rivers [length]) that is assigned to the hexagonal grid attribute,
BT_i – BDOT10k object,
OSM_i – OSM object.

The adopted weighting assumes that all of the criteria are equally important; hence, each criterion takes on a weight amount of 0.167 (the second combination in Equation (2)). In order to express the relative percentage change, the Relative Change (RC) between the local CCI is subsequently determined with two variants of weights according to Equation (2).

$$\mathrm{RC}_{\mathrm{CCI}} = \frac{\mathrm{CCI}_{\mathrm{W2}} - \mathrm{CCI}_{\mathrm{W1}}}{\mathrm{CCI}_{\mathrm{W2}}} \cdot 100\% \tag{2}$$

where:

$\mathrm{RC}_{\mathrm{CCI}}$ – RC of CCI values,
$\mathrm{CCI}_{\mathrm{W1}}$ – value of CCI using various weights (first combination),
$\mathrm{CCI}_{\mathrm{W2}}$ – value of CCI using equal weights (second combination).

The first combination ($\mathrm{CCI}_{\mathrm{W1}}$) utilizes object weighting due to the objects' recognizability in the satellite and aerial images from which they were obtained; i.e., aerial ortoimages (10 m pixel) and SPOT 5 ortoimages in the EU border zone. Thus, buildings and forests were each given a weight of 0.25, paved roads and railroads – 0.15, water bodies and streams – 0.10. These weighting rules are also used in accessibility analyses and are extremely important in emergency management [23]. In the second combination ($\mathrm{CCI}_{\mathrm{W2}}$), the weights were equal and amount to 0.167.

In order to compare the RC values that were obtained for the studied counties, four class divisions were defined; these were created with ranges of values that represented the proportions of the standard deviation. The negative RC values were analyzed in two classes; for these, each range was defined according to the interval of half of the standard deviation (0.5σ) that was calculated as the average value for the analyzed counties. The positive RC values were also divided into two classes according to the value of one standard deviation (σ) as the interval of the ranges.

3.3. Hot Spot and Statistical Analysis

Hot spot analysis was used to indicate the spatial relationships and identify the spatial clustering of the RC values. The resulting values showed where objects with high or low values were spatially clustered [24]. A hot spot can be described as an area with a higher concentration of events as compared to an expected number after considering the random distribution of events. A feature with a high value is interesting but may not be a statistically significant hot spot. For an object to be a statistically significant active point, the object will have a high value and be surrounded by other objects with high values as well. The local sum of an object and its neighbors is compared proportionally with the sum of all of the objects. When the local sum is different from the expected local sum and when the difference is too large to be due to random chance, a statistically significant "*z*" result is obtained [25] according to Equations (3)–(5):

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2}{n-1}}} \tag{3}$$

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{4}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X}\right)^2} \tag{5}$$

where:

x_j – RC value of CCI feature,
$w_{i,j}$ – spatial weight between CCI features i and j,
n – total number of features.

The Getis-Ord Gi* statistic provides a z-score, p-value and confidence interval with an interpretation according to Table 1.

Table 1. Hot spot analysis parameter interpretation

Statistics	Description	Implication
$z > 0$ and p-value is small	high-high spatial cluster (the larger the z-score, the greater the clustering degree)	$CCI_{W1} < CCI_{W2}$ $RC_{CCI} > 0$
z is closer to 0	no obvious spatial clustering	–
$z < 0$ and p-value is small	low-low spatial cluster (the smaller the z-score, the greater the clustering degree)	$CCI_{W1} > CCI_{W2}$ $RC_{CCI} < 0$

Source: own elaboration based on [25]

A statistical analysis that was based on descriptive statistics and Pearson correlations was used to provide an overall overview of county-level results.

4. Results

4.1. CCI with Equal Weights Overview

Otwocki and Piaseczno Counties were characterized by the highest inter quantile range (IQR) values as well as the highest standard deviations this indicated the high dispersion of their local CCI values (Table 2). The standard deviation took values that were lower than the mean of the CCI values, which indicated that the CCI values were more concentrated; i.e., the consistency of the data was relatively high in these instances. Otwocki and Piaseczno Counties were characterized by standard deviations of 0.065 and 0.062, respectively; these were nearly double the lowest value that was recorded in Sokólski County (0.036).

Table 2. Descriptive CCI statistics for TOPSIS analysis with equal weights

Statistics	Mean	Median	Minimum	Maximum	Q1	Q3	σ	IQR
Międzyrzecki	0.0426	0.0328	0.0000	0.4607	0.0152	0.0553	0.0425	0.0401
Ostrowski	0.0565	0.0470	0.0000	0.5170	0.0277	0.0720	0.0455	0.0442
Otwocki	0.0989	0.0887	0.0000	0.4172	0.0555	0.1268	0.0619	0.0713
Piaseczno	0.0864	0.0756	0.0000	0.4954	0.0425	0.1153	0.0645	0.0728
Sanocki	0.0714	0.0615	0.0000	0.5102	0.0347	0.0946	0.0572	0.0600
Sokólski	0.0355	0.0281	0.0000	0.5855	0.0140	0.0479	0.0356	0.0339
Słupski	0.0389	0.0314	0.0000	0.4572	0.0157	0.0482	0.0382	0.0325

The descriptive statistics of the CCI with the different weights are presented below in Table 3 for comparison purposes (a detailed analysis is described in [1]).

Table 3. Descriptive CCI statistics for TOPSIS analysis with various weights

Statistics	Mean	Median	Minimum	Maximum	Q1	Q3	σ	IQR
Międzyrzecki	0.0353	0.0271	0.0000	0.5899	0.0129	0.0427	0.0016	0.0405
Ostrowski	0.0427	0.0314	0.0000	0.5765	0.0181	0.0538	0.0017	0.0406
Otwocki	0.0989	0.0832	0.0000	0.5069	0.0542	0.1246	0.0046	0.0680
Piaseczno	0.0915	0.0754	0.0000	0.4828	0.0420	0.1196	0.0052	0.0723
Sanocki	0.0678	0.0533	0.0000	0.4995	0.0315	0.0790	0.0039	0.0627
Słupski	0.0414	0.0313	0.0000	0.4764	0.0165	0.0507	0.0018	0.0426
Sokólski	0.0390	0.0304	0.0000	0.5329	0.0159	0.0499	0.0015	0.0386

Source: own elaboration based on [1]

According to the five data-compliance ranges that were defined by Borkowska et al. [1], the percentages of the local CCI classes that were calculated with equal weights are presented in Table 4. The significant predominance of areas with low and very low differentiations between BDOT10k and OSM (the first and second classes of the CC compliance) could be observed in almost all of the analyzed counties – from 81.7% in Słupski County to 76.6% in Otwocki County. The exception was Piaseczno County; such areas accounted for slightly more than half of the county's size (55%). In this county, the highest diversity (defined as a semi-compliance [the third class]) could be noted, with a value of 24.7%; the noncompliance (the fourth and the fifth classes) amounted to as high as 20.2%. The other counties in the semi- and non-compliance classes ranged from approximately 13% to 15% (semi-compliance) and 6% to 8% (noncompliance).

Table 4. County area percentages in CCI_L classes for equal weights

Class	Description	County area percentage [%]						
		Międzyrzecki	Ostrowski	Otwocki	Piaseczno	Sanocki	Słupski	Sokólski
1	maximum compliance	33.2	32.1	33.5	24.0	32.4	31.5	32.3
2	moderate compliance	46.9	47.5	43.1	31.0	46.4	50.2	48.8
3	semi-compliance	14.0	12.9	15.1	24.7	15.2	13.1	14.0
4	moderate noncompliance	5.9	5.2	6.7	16.4	6.0	2.3	4.9
5	maximum noncompliance	–	2.4	1.6	3.8	–	2.9	–

Table 5 below shows the percentages of the local CCI classes calculated with various weights (an analysis is widely described in [1]).

Table 5. County area percentages in CCI_L classes for various weights

Class	Description	County area percentage [%]						
		Międzyrzecki	Ostrowski	Otwocki	Piaseczno	Sanocki	Słupski	Sokólski
1	maximum compliance	29.1	33.2	34.1	35.5	30.6	30.6	32.1
2	moderate compliance	56.2	46.5	43.5	42.3	52.2	52.9	49.6
3	semi-compliance	9.6	13.3	13.1	13.1	10.6	10.3	12.7
4	moderate noncompliance	2.5	4.6	9.3	5.6	3.1	3.1	2.7
5	maximum noncompliance	2.6	2.5	–	3.4	3.5	3.2	2.9

Source: own elaboration based on [1]

4.2. Relative Changes of CCI

The results of the RC between the local CCIs for the variant of differentiated and equal weights were quite diverse (Tables 6, 7, Fig. 2). A similar range of the minimum (from −34.4% to −30.5%) and maximum (from 64.6% to 74.6%) RC values could be observed in Piaseczno, Sokólski, Sanocki and Otwocki Counties. However, the high values that were obtained in Ostrowski and Międzyrzecki Counties (where the minimum values of the percentage changes in the local CCIs were −12.7% and −23.1%, respectively, and the maximum values were 120.6% and 98.2%, respectively) significantly exceeding the obtained maximum results for the other counties. Ostrowski County also had the highest median (44.0%) and variance (1,429.1%) among the studied counties. The standard deviation values in the analyzed counties ranged from 25.6% (Sokólski County) to 29.6% (Międzyrzecki County), with relatively low means (from 5.5% to 1.1%); these indicated greater variability. Ostrowski County achieved the highest σ value (37.8%) with a mean value of 45.4%.

Table 6. Descriptive statistics of RC of CCI

Statistics	Piaseczno	Sokólski	Sanocki	Słupski	Ostrowski	Otwocki	Międzyrzecki
Mean	1.1	−5.5	13.1	0.1	45.4	5.8	25.8
Median	−1.4	−8.1	15.3	0.1	44.0	6.3	29.6
Minimum	−34.5	−34.2	−30.5	−34.4	−12.7	−31.9	−23.1
Maximum	65.0	64.6	74.6	64.7	120.6	73.5	98.2
Q1	−25.1	−29.5	−8.2	−23.5	15.1	−17.7	6.6
Q3	16.9	9.1	28.9	9.8	72.7	17.7	32.1
Variance (σ^2)	761.7	657.8	691.8	712.4	1429.1	717.1	877.0
Std. dev. (σ)	27.6	25.6	26.3	26.7	37.8	26.8	29.6

Table 7. Percentages of county areas for RC of CCI ranges

Class	Range of RC	Percentage of the county's area [%]						
		Piaseczno	Sokólski	Sanocki	Słupski	Ostrowski	Otwocki	Międzyrzecki
1	min ≤ −15%	36.2	44.1	19.3	34.3	–	27.6	11.8
2	−15% < RC ≤ 0	14.9	14.3	11.1	15.4	16.3	16.1	10.3
3	0 < RC ≤ 30%	32.2	31.0	45.5	36.5	17.4	39.5	29.7
4	30% < RC ≤ max	16.7	10.6	24.1	13.7	66.4	16.8	48.2

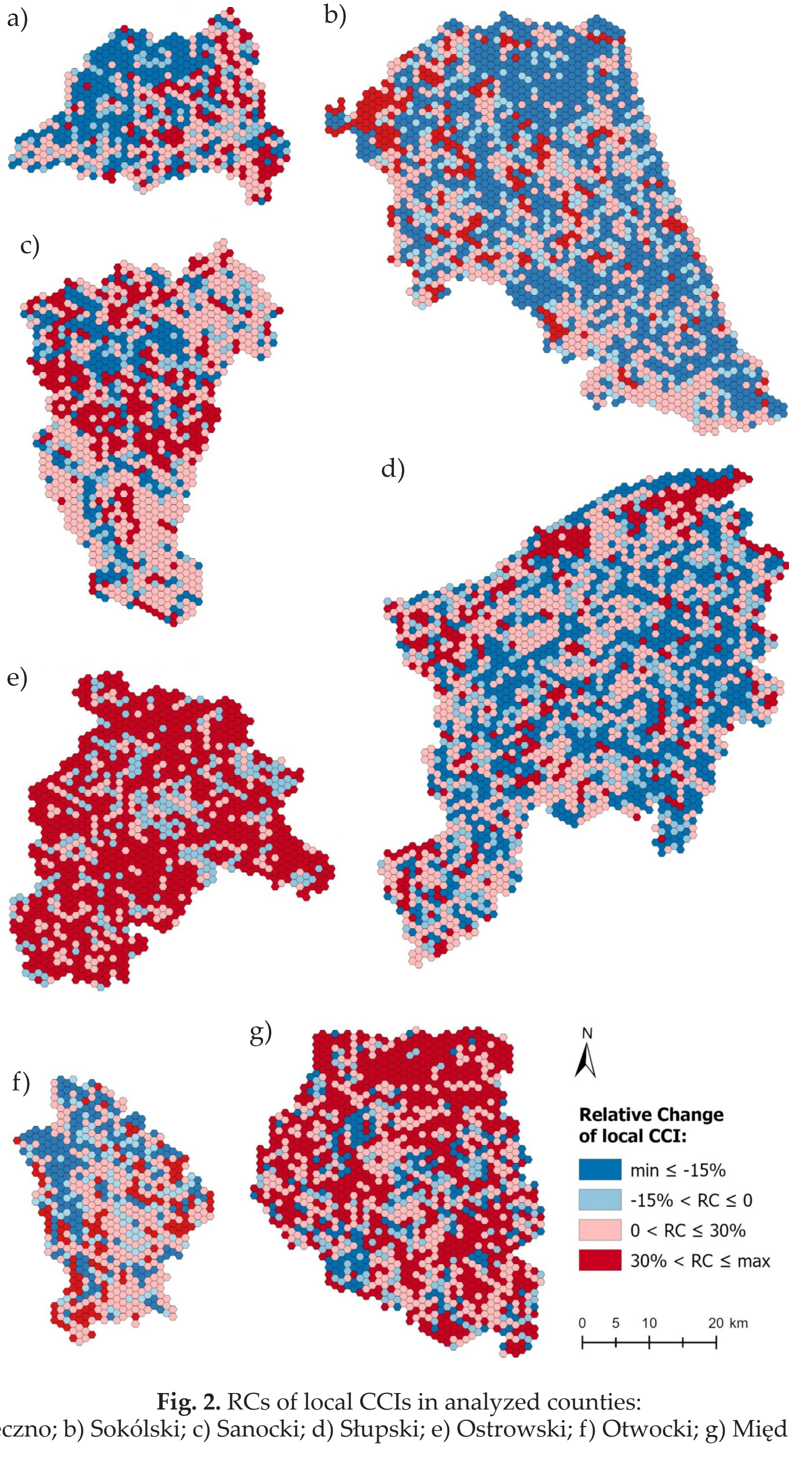

Fig. 2. RCs of local CCIs in analyzed counties:
a) Piaseczno; b) Sokólski; c) Sanocki; d) Słupski; e) Ostrowski; f) Otwocki; g) Międzyrzecki

Piaseczno, Sokólski, Słupski, and Otwocki Counties (Table 7) achieved similar proportions of negative RC values (accounting for about half of each county), with a clear predominance of values of up to −15% RC (the maximum being 44.1% of the area of Sokólski County). Values from −15% to 0% RC for these districts represented from 14.3% of the area in Sokólski County to 16.1% of the area of Otwocki County. In Ostrowski, Międzyrzecki and Sanocki Counties, negative RC values account for 16.3%, 22.1%, and 30.4%, respectively. Positive values of up to 30% of RC predominated in Sanocki (45.5%) and Otwocki (39.5%) Counties. However, the shares of Piaseczno, Sokólski, and Międzyrzecki Counties were similar, amounting to about one-third of the analyzed set. The largest shares of a county's area (within a range of more than 30% RC) were represented by Ostrowski (66.4%) and Międzyrzecki (48.2%) Counties, and the smallest shares were those of Słupski (13.7%) and Sokólski (10.6%) Counties.

The land use of the analyzed areas was dominated by agricultural land (53% on average) and forests (36% on average). The relative sizes of the built-up areas varied from less than 2% in Międzyrzecki County to 13.3% in Piaseczno County (Fig. 3).

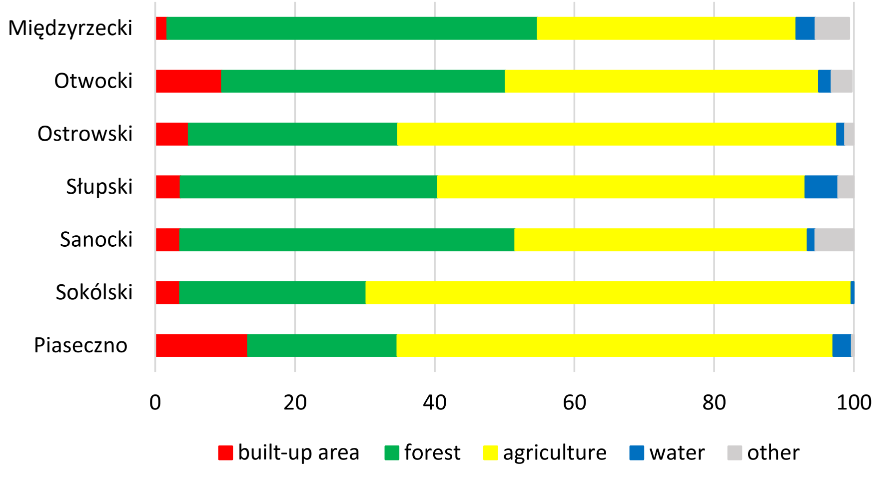

Fig. 3. Land use in study area

Source: own elaboration based on official cadastral statistics from 2021

Pearson linear correlations (r) provide insight into the associations of land use and RC. At a significance level of $p < 0.0500$, the Pearson correlation varies depending on the range of the RC levels. A moderate negative correlation (−0.52) can be observed between a forest and an RC level that is less than 0, while a strong positive correlation (0.76) can be observed between a forest and an RC level that is greater than 30%. A strong negative correlation (−0.82) was recorded between agriculture and an RC level that was greater 30%. Built-up areas are moderately negatively (−0.61) correlated with an RC range that is between 0% and 15%.

4.3. Variants of Local CCI Weights – Hot Spot Analysis

A Getis-Ord Gi* analysis identified the statistically significant hot and cold spots that are shown in Figure 4.

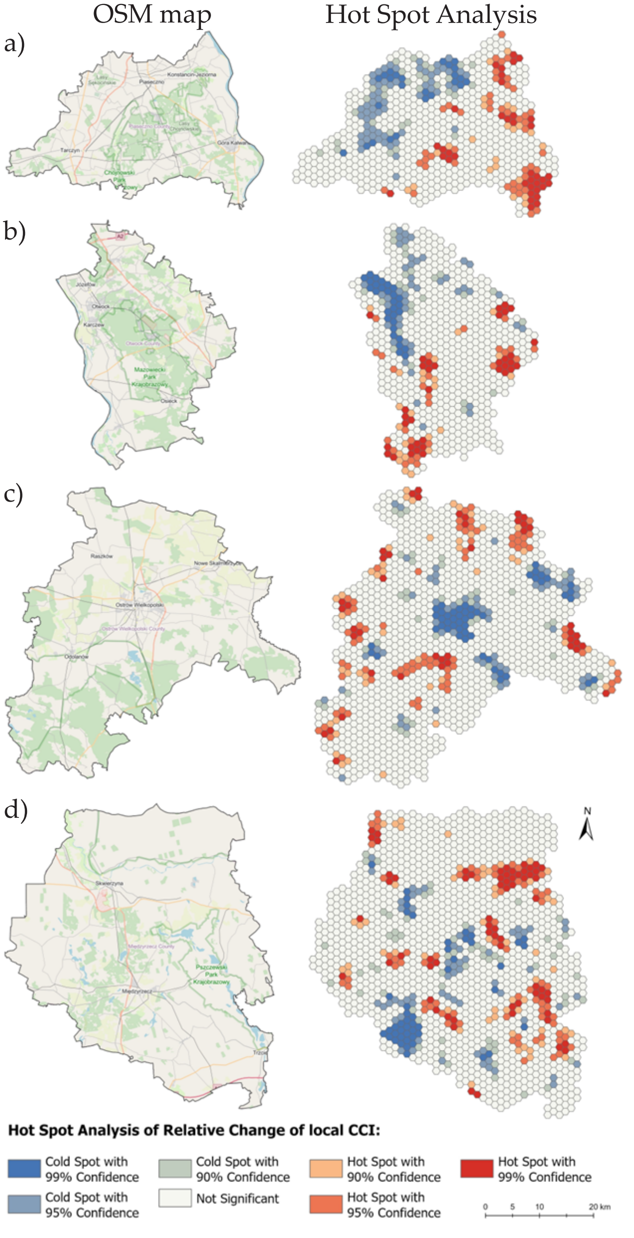

Fig. 4. Hot spot analysis of RC of local CCIs in analyzed counties:
a) Piaseczno; b) Otwocki; c) Ostrowski; d) Międzyrzecki (compared to OSM map)

Based on the hot spot maps (Fig. 2), a visual analysis of the relationship between the landscape and the hot and cold clusters was performed. In Piaseczno and Otwocki Counties, those areas that were identified as hot spots were clustered mainly in open areas (meadows, farmlands), in the vicinity of water bodies (ponds in Żabieniec) and (less frequently) in areas of dispersed settlements (mainly rural areas) and forested areas (Chojnowskie Forests, Masovian Landscape Park). However, those areas that were identified as cold spots occurred in urban areas (the cities of Piaseczno [along with its neighboring towns south to the city of Tarczyn] and Otwock [with its neighbors Józefów and Karczew]) as well as along major transportation lines (Krakowska Avenue). In Ostrowski and Międzyrzecki Counties, the hot spots were similarly concentrated in open areas (the northeastern part of the county), large water bodies, and forests (Pszczewski Landscape Park, Barycz Valley Landscape Park). Cold spots also occurred in areas of compact development – the cities of Skierzyna, Międzyrzecz and Ostrów Wielkopolski, the towns of Odolanów and Nowe Skalmierzyce as well as in the forests in the northern part of Międzyrzecki County (Nietoperek Nature Reserve).

5. Discussion

Fitness for purpose is a principle that is widely accepted among analysts as the correct approach for obtaining a quality data set [26, 27]. However, only a few analysts or end users of data can accurately determine what data quality is required for a specific task. When selecting a particular spatial data set, the user should be very attentive, as it is impossible to evaluate all of the strengths and weaknesses of available data. One aspect that is difficult to assess is the up-to-dateness, which is given for an entire data set, while its parts could be characterized by a different topicality [28]. Topographical data are updated periodically according to the rules in force (which vary from country to country). In Poland, this used to be a ten-year period [29, 30]; however, it was recently changed to a two-year period. The OSM data is updated by users (mappers), so the data up-to-dateness depends on their activities. The BDOT10k data that was used in this research was from March 2020, while it was not possible to determine the year of the OSM data update for the analyzed areas. A literature research [31, 32] showed that the most up-to-date data was on roads and buildings.

Nevertheless, an important aspect that significantly influences the TOPSIS ranking results is the selection of topographical objects and their prioritization, which is usually associated to the overarching objective; i.e., answering the question about the purpose of an analysis. In the present study, it was assumed that this objective was related to crisis management (i.e., floods, fires, terrorist attacks) for which the identification of populated areas, access routes and hazard areas is important. The second application area was sustainable development according to

Agenda 2030. The greatest weights were assigned to buildings and forests, whose importance in both applications was indisputable [1]. The research did not consider object attributes due to the relatively small number of objects that were described with attributes in OSM [1, 31].

It is worth noting that a CCI can be calculated for any map unit (MU), whether natural (e.g., catchments, ecotopes), administrative, or geometric. The universality of a CCI also lies in the facts that any objects can be included in analyses and mapping units can be ranked by considering other (non-topographical) categorical data (even MUs within one data set).

In the presented research, two variants of the weights of the CCI, a comparative measure of OSM, and the BDOT10k quantitative data were analyzed. In the first approach (in accordance with [1]), differentiated weights were adopted, which corresponded to the relevance of the objects under study that were adopted by the authors; i.e., buildings, forests (with the greatest value of the weights), communication networks (a moderate value of the weights), and watercourses/water bodies (the lowest value of the weights). In contrast, all of the analyzed objects were considered to be equally important in the presented variant, and their weights were assumed to be equal. The CCI values with different and equal weights differed, as was previously mentioned in [18, 19]. The local CCI values showed clustering in all of the analyzed counties. According to the adopted gradual scale of compliance, the CCI in two combinations of weights occupied similar shares of the area of each county. The greatest differences in the occupied areas could be seen in the case of Piaseczno County – the sizes of the maximum and moderate compliance areas decreased by a 22.8% share of the county's area after equalizing the CCI weights (amounting to 55%). However, the area that was occupied by semi-compliance doubled to a 25% share of Piaseczno County's area. Similarly, the share of the areas that were assessed as being of moderate and maximum noncompliance increased from 9% to 20.2% of the share of the county's area after equalizing the weights. This allowed us to conclude that, as in the previous studies, those areas with high degrees of urbanization showed the greatest variability between the BDOT10k data and the OSM data.

At the pixel level, the Pearson correlation analysis did not show a significant relationship between the land cover type and the CCI for the equal weights (CCI_{W2}); this was similar to the CCI that was analyzed in the previous article (CCI_{W1}) [1]. Also, no significant statistical relationship was shown in the Relative Changes between the CCI weights that were used (significance level $p < 0.0500$). For this reason, a hot spot analysis was performed in order to identify clusters of spatial phenomena. The hot spot detection evolved from studying the distribution of the points or the spatial distribution of the points in space in order to comprehend the spatial patterns [33]. A visual dependency analysis revealed observations that, in the counties that were studied, the clusters that were defined as hot spots and cold spots included similar land cover types, thus allowing them to be characterized in terms of settlement type and land use.

6. Conclusions

The Relative Changes of the CCI showed the effect of the weights on the obtained results. The negative RC values revealed the predominance of the variant of the weighting with different weights (CCI_{W1}). According to the results, the highest share of negative RC values was shown in Sokólski (58.4%) and Piaseczno (51.1%) Counties. However, positive RC values proved the prevalence of the equal-weight variant (CCI_{W2}); this was especially evident in Ostrowski (83.7%), Międzyrzecki (77.9%) and Sanocki (69.9%) Counties. The county with the most equal shares of the different weighting variants was Slupski (49.8% and 50.2%, respectively). The equal weights in the TOPSIS method influenced the number and, thus, the area, while both of the topographical data sets (BDOT10k and OSM) had the highest and moderate compliances. These differences varied from county to county, taking 4.1% (Międzyrzecki) and a minimum of 0.2% (Sokólski) in the maximum-compliance and from 9.3 and 0.8% for the same counties in the moderate-compliance CCI classes. For 56% of the total area, the change in the weights altered the ranking by half a standard deviation. Relatively large changes of more than 2.5 standard deviations could be observed in 4% of the analyzed area. The demonstrated analyses prove that the studied data sets of OSM and BDOT10k were quite sensitive to the adopted weighting combinations.

A hot spot analysis of the CCI's Relative Changes indicated spatial relationships between the studied data sets despite the absence of a statistically significant Pearson correlation. Those areas that were identified as hot spots were mainly clustered in forests, open areas, cultivated areas, neighborhoods of water bodies, and (less frequently) areas with low building density. However, those areas that were identified as cold spots were found in the areas of urban-rural development and along major transportation lines.

Funding

The paper preparation was founded be the statutory research at the Military University of Technology, Faculty of Civil Engineering and Geodesy, Institute of Geospatial Engineering and Geodesy, Grant Number USG 531–4000–22–705.

CRediT Author Contribution

S. B.: conceptualization, methodology, formal analysis, data curation, writing – original draft preparation, writing – review and editing.

E. B.: conceptualization and result discussion, writing – review and editing.

K. P.: conceptualization, validation, writing – review and editing, supervision, funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Public Data:
OpenStreetMap data: https://download.geofabrik.de.
BDOT10k data: https://mapy.geoportal.gov.pl.

Use of Generative AI and AI-assisted Technologies

No generative AI or AI-assisted technologies were employed in the preparation of this manuscript.

References

[1] Borkowska S., Bielecka E., Pokonieczny K.: *Comparison of land cover categorical data stored in OSM and authoritative topographic data.* Applied Sciences, vol. 13(13), 2023, 7525. https://doi.org/10.3390/app13137525.

[2] Aksoy E., San B.T.: *Geographical information systems (GIS) and multi-criteria decision analysis (MCDA) integration for sustainable landfill site selection considering dynamic data source*. Bulletin of Engineering Geology and the Environment, vol. 78(4), 2019, pp. 779–791. https://doi.org/10.1007/s10064-017-1135-z.

[3] Anthoff D., Tol R.S.J.: *On international equity weights and national decision making on climate change*. Journal of Environmental Economics and Management, vol. 60(1), 2010, pp. 14–20. https://doi.org/10.1016/j.jeem.2010.04.002.

[4] Li P., Qian H., Wu J., Chen J.: *Sensitivity analysis of TOPSIS method in water quality assessment. Sensitivity to the parameter weights*. Environmental Monitoring and Assessment, vol. 185(3), 2013, pp. 2453–2461. https://doi.org/10.1007/s10661-012-2723-9.

[5] Odu G.O.: *Weighting methods for multi-criteria decision making technique.* Journal of Applied Sciences and Environmental Management, vol. 23(8), 2019, pp. 1449–1457. https://www.ajol.info/index.php/jasem.

[6] Roszkowska E.: *Rank ordering criteria weighting methods – a comparative overview*. Optimum. Studia Ekonomiczne, vol. 5(65), 2013, pp. 14–33. https://doi.org/10.15290/ose.2013.05.65.02.

[7] Borkowska S., Bielecka E., Pokonieczny K.: *OpenStreetMap – building data completeness visualization in terms of 'Fitness for purpose'.* Advances in Geodesy and Geoinformation, vol. 72(1), 2023, pp. 2–20. https://doi.org/10.24425/agg.2022.141922.

[8] Borkowska S., Pokonieczny K.: *Analysis of OpenStreetMap data quality for selected counties in Poland in terms of sustainable development*. Sustainability, vol. 14(7), 2022, 3728. https://doi.org/10.3390/su14073728.

[9] Behzadian M., Khanmohammadi Otaghsara S., Yazdani M., Ignatius J.: *A state-of the-art survey of TOPSIS applications*. Expert Systems with Applications, vol. 39(17), 2012, pp. 13051–13069. https://doi.org/10.1016/j.eswa.2012.05.056.

[10] Zyoud S.H., Fuchs-Hanusch D.: *A bibliometric-based survey on AHP and TOPSIS techniques*. Expert Systems with Applications, vol. 78, 2017, pp. 158–181. https://doi.org/10.1016/j.eswa.2017.02.016.

[11] Çelikbilek Y., Tüysüz F.: *An in-depth review of theory of the TOPSIS method: An experimental analysis.* Journal of Management Analytics, vol. 7(2), 2020, pp. 281–300. https://doi.org/10.1080/23270012.2020.1748528.

[12] Wang Z.X., Wang Y.Y.: *Evaluation of the provincial competitiveness of the Chinese high-tech industry using an improved TOPSIS method.* Expert Systems with Applications, vol. 41(6), 2014, pp. 2824–2831. https://doi.org/10.1016/j.eswa.2013.10.015.

[13] Kusumadewi S., Hartati S.: *Sensitivity analysis of multi-attribute decision making methods in Clinical Group Decision Support System.* [in:] *2007 International Conference on Intelligent and Advanced Systems: Kuala Lumpur, Malysia: 25–28 November 2007*, IEEE, pp. 301–304. https://doi.org/10.1109/ICIAS.2007.4658395.

[14] Dalalah D., Hayajneh M., Batieha F.: *A fuzzy multi-criteria decision making model for supplier selection.* Expert Systems with Applications. 2011, vol. 38(7), pp. 8384–8391. https://doi.org/10.1016/j.eswa.2011.01.031.

[15] Choo E.U., Schoner B., Wedley W.C.: *Interpretation of criteria weights in multicriteria decision making*. Computers & Industrial Engineering, vol. 37(3), 1999, pp. 527–541. https://doi.org/10.1016/S0360-8352(00)00019-X.

[16] Bączkiewicz A., Wątróbski J., Kizielewicz B., Sałabun W.: *Towards objectification of multi-criteria assessments: A comparative study on MCDA methods.* [in:] Ganzha M., Maciaszek L., Paprzycki M., Ślęzak D. (eds.), *Proceedings of the 16th Conference on Computer Science and Intelligence Systems: September 2–5, 2021*, Annals of Computer and Information Systems, vol. 25, IEEE, pp. 417–425. https://doi.org/10.15439/2021F61.

[17] Kobryń A., Prystrom J.: *A data pre-processing model for the TOPSIS method*. Folia Oeconomica Stetinensia, vol. 16(2), 2016, pp. 219–235. https://doi.org/10.1515/foli-2016-0036.

[18] Pavić Z., Novoselac V.: *Notes on TOPSIS Method*. International Journal of Engineering Research and General Science, vol. 1(2), 2013, pp. 5–12.

[19] Więckowski J., Zwiech P.: *Can weighting methods provide similar results in MCDA problems? Selection of energetic materials study case.* Procedia Computer Science, vol. 192, 2021, pp. 4592–4601. https://doi.org/10.1016/j.procs.2021.09.237.

[20] Chen Y., Yu J., Khan S.: *Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation*. Environmental Modelling & Software, vol. 25, 2010, pp. 1582–1591. https://doi.org/10.1016/j.envsoft.2010.06.001.

[21] Al-Mashreki M. H., Akhir J.B.M., Abd Rahim S., Tukimat L., Haider A.R.: *GIS-based sensitivity analysis of multi-criteria weights for land suitability evaluation of sorghum crop in the Ibb Governorate, Republic of Yemen*. Journal of Basic and Applied Scientific Research, vol. 1(9), 2011, pp. 1102–1111.

[22] Liern V., Pérez-Gladish B.: *Multiple criteria ranking method based on functional proximity index: Un-weighted TOPSIS*. Annals of Operations Research, vol. 311, 2022, pp. 1099–1121. https://doi.org/10.1007/s10479-020-03718-1.

[23] Dawid W., Pokonieczny K., Wyszyński M.: *The methodology of determining optimum access routes to remote areas for the purposes of crisis management*. International Journal of Digital Earth, vol. 15(1), 2022, pp. 1905–1928. https://doi.org/10.1080/17538947.2022.2134936

[24] Getis A., Ord K.: *The analysis of spatial association by use of distance statistics*. Geographical Analysis, vol. 24, 1992, pp. 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

[25] Ord K., Getis A.: *Local spatial autocorrelation statistics: distributional issues and an application*. Geographical Analysis, vol. 27(4), 2010, pp. 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x.

[26] Jimenez J.: *Fitness for purpose in relation to specification limits*. Accreditation and Quality Assurance, vol. 17(1), 2012, pp. 27–34. https://doi.org/10.1007/s00769-011-0825-7.

[27] Sheng J., Wilson J.P., Chen N., Devinny J.S., Sayre J.M.: *Evaluating the Quality of the National Hydrography Dataset for Watershed Assessments in Metropolitan Regions*. GIScience & Remote Sensing, vol. 44(3), 2007, pp. 283–304. https://doi.org/10.2747/1548-1603.44.3.283.

[28] Bielecka E., Jenerowicz A.: *Intellectual structure of CORINE land cover research applications in web of science: A Europe-wide review*. Remote Sensing, vol. 11(17), 2019, 2017. https://doi.org/10.3390/rs11172017.

[29] Bielecka E.: *Geographical data sets fitness of use evaluation*. Geodetski Vestnik, vol. 59(2), 2016, pp. 335–348. https://doi.org/10.15292/geodetski-vestnik.2015.02.335-348.

[30] Bac-Bronowicz J., Dygaszewicz J., Grzempowski P., Nowak R.: *Bazy danych referencyjnych jako źródła zasilania i aktualizacji warstw dotyczących budynków w Wielorozdzielczej Topograficznej Bazie Danych*. Roczniki Geomatyki, t. 8, z. 5(41), 2010, pp. 7–22.

[31] Biljecki F., Chow Y. S., Lee K.: *Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes.* Building and Environment, vol. 237, 2023, 110295. https://doi.org/10.1016/j.buildenv.2023.110295.

[32] Marczak S.: *Ocena zaangażowania społeczeństwa w tworzenie danych przestrzennych w Polsce na przykładzie projektu OpenStreetMap*. Roczniki Geomatyki, t. 13, z. 3(69), 2015, pp. 239–253.

[33] Chakravorty S.: *Identifying crime clusters: The spatial principles*. Middle States Geographer, vol. 28, 1995, pp. 53–58.