



Tomasz Pirowski, Małgorzata Timek

ANALYSIS OF LAND USE AND LAND COVER MAPS SUITABILITY FOR MODELING POPULATION DENSITY OF URBAN AREAS – REDISTRIBUTION TO NEW SPATIAL UNITS BASED ON CLC AND UA DATABASES*

AGH University of Science and Technology,
Faculty of Mining Surveying and Environmental Engineering

Keywords: population data, dasymetric modeling, LULC, Urban Atlas, Corine Land Cover

Abstract

The series of articles contains a comparison of the possibilities of using data from three sources for mapping people, with different spatial, thematic and time accuracy. These are data from Corine Land Cover (CLC) and Urban Atlas (UA) projects and the result of object classification (OBIA) of RapidEye data. The information on the existence of building zone included on the land use and land cover maps (LULC) constituted a limiting variable in the dasymetric method of population mapping. Categories related to building types allowed for the introduction of variable relationships, diversifying population density. These treatments enabled multi-variant development of maps of spatial population occurrence at a higher level than the original census units.

The experiment was carried out in the area of Krakow. Statistical data from 141 urban units (u.u.) of the city were used. Generation of population maps was carried out in several variants. Divisions of buildings were made depending on its characteristics and functions. The results of population conversion were analyzed on Central Statistical Office (hereafter referred as CSO, in Polish: GUS) data in a kilometer grid and on a specially prepared map of the population including a part of Krakow. The applied double verification allowed to rank the obtained population maps and provide border spatial accuracy of their cellular representation.

The first part of the cycle presents the state of knowledge about population mapping and population conversion using the dasymetric method. The area of research is described. Spatial and statistical data used in the research were characterized. Works related to population conversion based on CLC and UA were presented. Six maps of the population distribution of Krakow were obtained. A multi-variant process of recalculating and setting weights for various types of buildings is described by providing for urban units the values of RMSE and MAPE. Population using the surface-weight method based on UA data was considered the best (MAPE 66%, RMSE 3442 people/u.u.). On CLC data, these errors were: MAPE 168%, RMSE 5690 people/u.u.

In the subsequent parts of the cycle, the population conversion will be presented using object-oriented classification. The methodology for the verification of results will be described based on a photointerpretation map of the population and the GUS perimeter grid. A discussion will be conducted related to the use of RMSE and MAPE measures. The ranking of methods and recommendations improving the results of population redistribution based on CLC, UA and OBIA will be given.

* This article has been prepared within the scope of the AGH UST statutory research no. 11.11.150.949

ANALIZA PRZYDATNOŚCI MAP POKRYCIA I UŻYTKOWANIA TERENU DO MODELOWANIA GĘSTOŚCI ZALUDNIENIA OBSZARÓW MIEJSKICH – PRZELICZANIE MAP DO NOWYCH JEDNOSTEK PRZESTRZENNYCH, OPARTYCH O BAZY DANYCH CLC I UA

Słowa kluczowe: dane demograficzne, modelowanie dazymetryczne, Urban Atlas, Corine Land Cover

Streszczenie

Cykl artykułów zawiera porównanie możliwości wykorzystania do kartowania ludności danych z trzech źródeł, o różnej dokładności przestrzennej, tematycznej i czasowej: dane z projektów Corine Land Cover (CLC) i Urban Atlas (UA) oraz wynik klasyfikacji obiektowej (OBIA) danych RapidEye. Zawarta na mapach pokrycia i użytkowania terenu informacja o występowaniu zabudowy stanowiła zmienną ograniczającą w dazymetrycznej metodzie kartowania ludności. Kategorie związane z typami zabudowy pozwoliły na wprowadzenie zmiennych powiązań, różnicujących zagęszczenie ludności. Te zabiegi umożliwiły wielowariantowe opracowanie map przestrzennego występowania ludności na poziomie wyższym niż pierwotne jednostki spisowe.

Eksperyment przeprowadzono na obszarze Krakowa, wykorzystując dane statystyczne ze 141 jednostek urbanistycznych (j.u.) miasta. Generowanie map ludności przeprowadzono w kilku wariantach, dokonując podziałów zabudowy w zależności od jej charakterystyki i funkcji. Wyniki przeliczania ludności na nowe jednostki przestrzenne odniesiono na etapie weryfikacji do danych o ludności podanych przez GUS w siatce kilometrowej oraz do specjalnie przygotowanej przez autorów szczegółowej mapy ludności obejmującej fragment Krakowa. Zastosowana podwójna weryfikacja pozwoliła na uszeregowanie według jakości uzyskanych map populacji oraz podanie granicznych dokładności przestrzennych ich komórkowej reprezentacji.

W pierwszej części cyklu zaprezentowano zarys stanu wiedzy o kartowaniu ludności i zasadach przeliczania populacji metodą dazymetryczną. Opisano obszar badań, scharakteryzowano wykorzystane w badaniach dane przestrzenne i statystyczne. Przedstawiono prace związane z przeliczeniem populacji w oparciu o CLC i UA, uzyskując łącznie 6 map rozkładu ludności Krakowa. Wielowariantowy proces przeliczania i ustalania poprawnych wag dla różnych typów zabudowy scharakteryzowano poprzez podanie dla jednostek urbanistycznych, sprzed realizacji warunku Toblera, wartości średniego błędu kwadratowego (RMSE) oraz średniego absolutnego błędu procentowego (MAPE). W oparciu o te parametry kartowanie ludności metodą powierzchniowo-wagową, bazującą na danych UA, uznano za najlepszą (MAPE 66%, RMSE 3442os./j.u.), podczas gdy na danych CLC błędy te wyniosły: MAPE 168%, RMSE 5690 os./j.u.

W kolejnych częściach cyklu przedstawione zostanie przeliczanie populacji z zastosowaniem klasyfikacji obiektowej. Opisana zostanie metodyka weryfikacji wyników w oparciu o fotointerpretacyjną mapę ludności oraz siatkę kilometrową GUS. Przeprowadzona będzie dyskusja nad zasadnością stosowania miar optymalizacyjnych RMSE i MAPE. Podany zostanie ranking metod oraz rekomendacje poprawiające wyniki redystrybucji ludności w oparciu o CLC, UA i OBIA.

INTRODUCTION

The integration of many scientific disciplines has now also encompassed the field of knowledge, which is demography. It has been related to broadly understood geodesy and cartography through the use of tools implemented in Geographic Information Systems (GIS). Statistical offices dealing with the collection and making available of statistical information, demographic, social and economic information, are increasingly taking care of their spatial visualization. The most popular methods of cartographic presentation are based on the cartogram and on various types of cartodiagrams. The population per area unit is presented in relation to the top-down borders, most often according to the administrative division. This generates a falsified result of population

distribution. It does not take into account uninhabited areas, it assigns averaged values to diversified areas of population density. The dasymetric diagram method reduces these errors, taking into account the spatial variability of the phenomenon. The key in the method is to correctly define the areas of occurrence and concentration of buildings as the border and intensity of the population. Such information can be obtained based on maps of coverage and land use. They are currently one of the most important information about the state of the environment. As a result, it is possible to obtain a spatial distribution of population density at a higher level than the original census units.

Today's need to visualize the distribution of population results from descriptive causes such as showing population growth, migration, and population density.

However, first and foremost, the need to conduct spatial analyzes, modeling phenomena and risks associated with it, and supporting spatial decision-making by administration at various levels (Gregory, 2002). Information on the distribution of population is important in spatial planning, in crisis analyzes related to the security of citizens, in the assessment of the quality of life of residents. It also allows you to gain a competitive advantage by trade and service companies that use geo-marketing, data passporting, and logistics in their operations. Accurate data on the distribution of population also require constant updating and hence partial automation. Therefore, remote sensing data and GIS tools play an increasingly important role in acquiring information on building areas. The series of articles contains a comparison of the possibilities of using data from this type of three sources for mapping people with various spatial, thematic and time accuracy, and various methods of their elaboration. These are data on Land Use and Land Cover (LULC) – from the Corine Land Cover project (CLC), Urban Atlas project (UA) and from the object classification (OBIA) of RapiEye data.

The dasymetric method is a variant of the cartogram in which the system of basic units depends on the variability of the intensity of the phenomenon being studied. Their selection allows a more reliable presentation of the presented data. Due to the time-consuming nature of the study, its development falls into the late twentieth century. It is connected with increased computing possibilities and greater availability of GIS techniques. According to several sources (Goleń and Ostrowski, 1994; Maantay et al., 2007), the first recognized creator of the dasymetric map was the American geographer John K. Wright. In 1936, during the development of a standard North American population density map based on the state breakdown, he noticed in his product the masking of relevant data. Population density categories largely depended on wastelands and wetlands, neglecting huge areas that were heavily urbanized and at the same time gathering large population centers. To improve the credibility of his map, Wright omitted completely uninhabited regions. He divided the studied areas depending on the land use and type of building. In this way, he distinguished several classes of land cover type and assigned them the appropriate value of population density. However, until the mid-nineties of the twentieth century, the most popular

and the most commonly used methods of presenting the population remained proper cartograms. Only the computerization of cartography, the development of GIS and the increase in the availability of digital cartographic data in the recent years have popularized the dasymetric method.

In 1998 Robinson et al. (1998) divided the variables in the method into limiting (defining the absolute limits of the value of the presented phenomenon) and the variables of connections (affecting its intensity). An example of determination of limiting variables (LULC maps, impervious surfaces) based on satellite images are the works of Eicher and Brewer (2001), Gallego and Peedell (2001), Harley (2002), Mennis (2003), Bielecka et al. (2005), Wu and Murray (2005), Azara et al. (2010), Pirowski and Pomietłowska (2017). Significant trends related to population mapping run in three directions: spatial accuracy (Tapp, 2010; Pirowski and Drzewiecki, 2012; Sridharan and Qiu, 2013; Bakillah, 2014; Cockx and Canters 2015; Całka et al., 2016), time accuracy related to from the dynamics of population distribution changes (Sleeter and Wood, 2006; Horanont and Shibasaki, 2010; Smith et al., 2015; Lin and Cromley, 2015), process automation (Mennis and Hultgren, 2006; Sleeter and Gould, 2007). In Poland, works related to dasymetric population mapping concern Warsaw (Bielecka et al., 2005), the Opatów district (Całka et al., 2016), Kraków (Pirowski and Drzewiecki, 2012, Pirowski and Pomietłowska, 2017).

METHODOLOGY

The adopted method is based on the assumption that the amount of the night population is closely related to the existence of mixing (binary method) or the occurrence of residential development divided into several categories related to concentration and type of building dominance (surface-weight aggregation method). The idea of these methods, in reference to the simple cartogram, is illustrated in Fig. 1.

LULC (CLC, UA) used in the studies define limiting variables (occurrence of buildings) and variable relations (types of buildings). In the simplest binary method (Fig. 1e), only the presence or not of residential development is taken into consideration when redistributing the population. Subjectively divisions of buildings and related diversification in the attribution of population

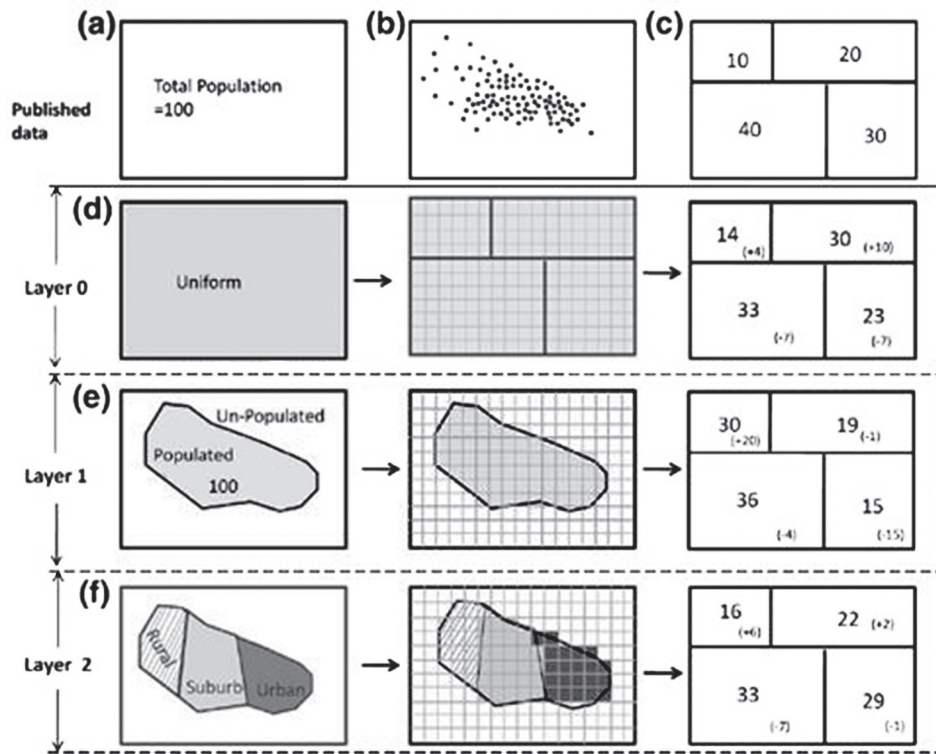


Fig. 1. Scheme of the dasymetric method; for Su et al. (2010), modified. Demographic data: a) total population in the area; b) dotted map of the population; c) population in the region divided into smaller statistical units. Spatial data – idea: d) simple cartogram; e) binary method; f) surface-weight aggregation method

Rys. 1. Schemat metody dazymetrycznej; za: Su i in. (2010), zmodyfikowany. Dane demograficzne: a) całkowita liczba ludności w rejonie; b) mapa kropkowa ludności; c) ludność w rejonie podzielona na mniejsze jednostki statystyczne. Dane przestrzenne – idea: d) kartogramu prostego; e) metody binarnej; f) powierzchniowo-wagowej metody agregacji

density are avoided. The disadvantage is the lack of the possibility of mapping local zones with an increased population of residents, associated with the concentration of close, high, multi-dwelling buildings.

The basic problem in the dasymetric surface-weight aggregation method (Fig. 1f) is to determine the correct proportions (weights) of population density per area unit, attributed to specific building categories. Optimization methods were used to determine them, which are based on the minimization of mean square error (RMSE) (1) and – in the second variant – mean absolute error percentage (MAPE) (2). These errors are counted as differences between the volume of population in u.u. from statistical data, and calculated using the dasymetric method (3). The possibilities of calculating these differences occur at the stage of implementing the Tobler condition (Tobler, 1979) for the entire city (the global value of the population is taken into ac-

count). In the next stage, after determining the correct proportions of building density, the errors are corrected. In every u.u. the number of inhabitants with statistical data is consistent. Details of this methodology are described in Pirowski and Pomietłowska (2017) and Pirowski and Bartos (2018).

$$\sum_{i=1}^{141} \delta(ju_i)^2 \rightarrow \min \quad (1)$$

$$\sum_{i=1}^{141} \left| \frac{\delta(ju_i)}{Pop(ju_i)} \right| \rightarrow \min \quad (2)$$

$$\delta(ju_i) + Pop_e(ju_i) = Pop(ju_i) \quad (3)$$

where:

$Pop(ju_i)$ – the population in the i -th urban unit;

$Pop_e(ju_i)$ – estimated total population in the i -th urban unit.

Research topics (part 1 of the publication cycle) concerned the answers to questions on how to correctly convert people into new spatial units. Can selected classes from CLC, UA reliably represent constraint variables and / or variable relationships? How to properly calculate the differentiation in population density for particular building classes for surface-weight methods? Is it confirmed that the surface-weight aggregation method can get lower errors than the binary method?

RESEARCH AREA, POPULATION DATA

The tested area is the city of Krakow within the administrative boundaries. The total area of the city is 326.8 km², 758334 inhabitants (state in 2012). The area is characterized by a large variety, population density

is variable. The city center is dominated by compact and commercial buildings. Due to the socio-economic changes, the city's strict center has practically already had service functions – tenement houses constitute premises for office and service activities, and the number of inhabitants is decreasing. With the distance from the center, block and dispersed buildings are becoming more and more popular (Kucharczyk et al., 2012). On the outskirts of the city there are scattered and loose buildings. In some areas of the city concentration of a particular type of building takes place, for example eastern regions are the dominance of industrial buildings.

Basic information about the city's population is given by the website of Cracow City Hall supervised by the City Development Department (Web sites StatKрак, <http://msip2.um.krakow.pl>). The data collected there

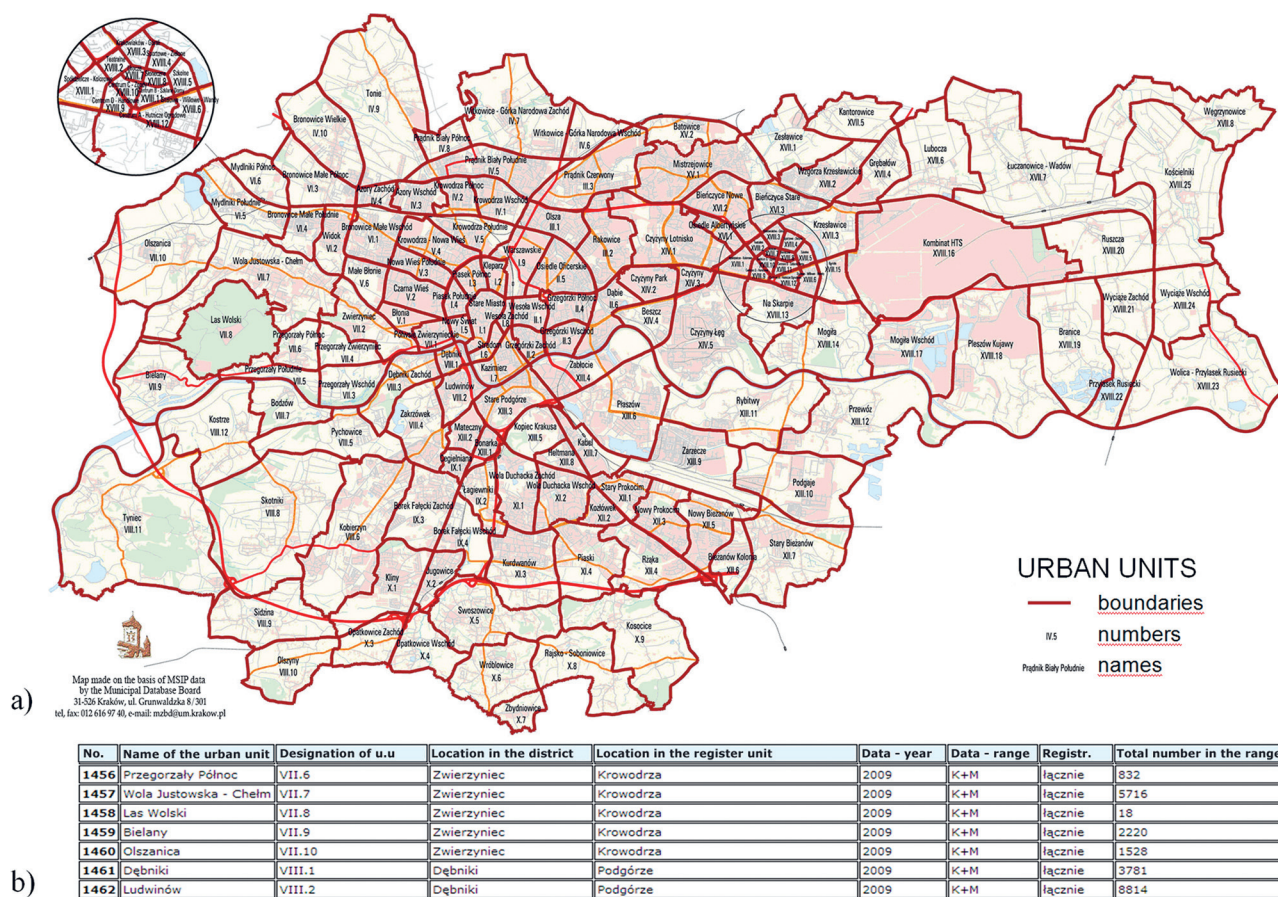


Fig. 2. Data used on Krakow's population: a) division into districts and urban units; b) a fragment of the demographic database published on the StatKрак portal (for www.msip2.um.krakow.pl)

Rys. 2. Wykorzystane dane o ludności Krakowa: a) podział na dzielnice i jednostki urbanistyczne; b) fragment bazy danych demograficznych opublikowanej na portalu StatKрак (za www.msip2.um.krakow.pl)

comes from the Local Data Bank of GUS, the Regional Data Bank of GUS and databases of census records. Among the available demographic data, the aggregate number of residents registered for permanent and temporary stay from 2009, in 141 urban units of Kraków (Fig. 2) was used for counting the population.

SPATIAL DATA ABOUT BUILDING ZONES

CORINE Land Cover 2012 (CLC2012) is a project run by the European Environment Agency as part of Copernicus GIO Land Monitoring 2011–2013. This is a continuation of projects from 1990 (CLC1990), 2000 (CLC2000) and 2006 (CLC 2006). The project is carried out at six-year intervals throughout almost all of Europe. In Poland, the institution responsible for the development of the CLC2012 database, verification and improvement of high-resolution HRL layers is the Institute of Geodesy and Cartography (clc.gios.gov.pl, www.igik.edu.pl). The main objectives of the project are a list of changes in land cover, creating a spatial database with information on land cover. The spatial database was created on the basis of satellite images RapidEye and IRS-P6 from 2011 and 2012. These data are available in raster and vector format on the Copernicus website <http://land.copernicus.eu>. The third, the most detailed classification level, has 44 classes. However, only 2 building classes were separated. The accuracy of the data is estimated at approximately 100m, and the smallest objects reach the area of 25 ha (Hościło, 2014; www.igik.edu.pl).

Urban Atlas (UA) is implemented by the Copernicus Program (formerly GMES – Global Monitoring for Environment and Security). The European Municipal Atlas provides high resolution land cover maps for 305 large European cities with a population of over 100,000. The website <http://www.eea.europa.eu> provides downloadable vector data for the land cover, a map of the area and a report with metadata. Development of the Urban Atlas database consists in the automatic classification of high resolution satellite images (SPOT 2.5m, ALOS 2.5m, RapidEye 5m), which allows the separation of main coverage classes. Then a visual interpretation is made using the available additional data. The smallest area of the mapped element is 0.25 ha, and the assumed 5 m accuracy (Urban Atlas, 2011, Borkowski et al., 2015). There are 20 coverage classes, of which 6 are for urban development and one for industrial and service buildings. There are 5 urban development class-

es and industrial buildings in Krakow. City buildings are divided according to the percentage value of the share of the impermeable area of a given element (Mapping Guide, 2011).

POPULATION REDISTRIBUTION BASED ON LULC MAPS: CLC AND UA

Based on the analysis of available materials concerning land development in Krakow and research works aimed at dasymetric mapping of the city (Kruszyńska and Wendel, 2001; Pirowski and Drzewiecki, 2012), it was decided to distinguish three categories from available LULC maps: compact, loose and dispersed. In the CLC project, two categories of buildings were distinguished: urban compact and urban loose. The analysis of additional data (orthophotomaps, BDOT) allowed to conclude that the two separate classes do not properly reflect the city's architecture. As a compact building, the Old Town and neighboring units were defined. The loose urban construction is the remaining areas where dense buildings exist. Areas of scattered development, and first of all single-family houses, have not been included as buildings. Therefore, it was decided to include scattered areas referred to as "complex cultivation patterns" as scattered development. Finally, 3 categories of buildings were identified (Figure 3a, Table 1), which allowed to make population maps in three variants: binary (by unification of building zones), surface-weight aggregation method based on RMSE minimization in u.u. and surface-weight aggregation method based on MAPE minimization in u.u.

Urban Atlas allows you to map 6 building types – five types of residential and one industrial – service. Using the Pirowski and Pomietłowska (2017) experiments, 5 residential housing classes were reduced to three (Figure 3b, Table 1). Based on such a map, two methods presented by the above-mentioned authors (binary and surface-weight aggregation method based on RMSE minimization) were reproduced, and a new, third variant based on MAPE minimization was developed.

In each case of surface-weight aggregation methods, both for CLC and UA, negative values for scattered development were obtained in the process of minimizing RMSE / MAPE errors. Therefore, it was necessary to apply boundary conditions and apply iterative calculations. Using the results of population mapping based on BDOT reported by Pirowski and Bartos (2017), the

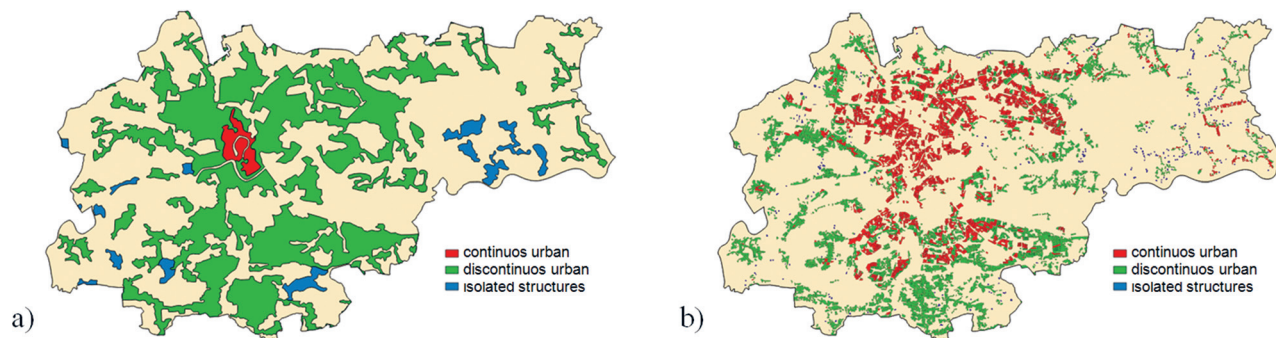


Fig. 3. Construction categories adopted in Krakow based on a) CLC; b) UA

Rys. 3. Przyjęte na obszarze Krakowa kategorie zabudowy na podstawie a) CLC; b) UA

Table 1. Source classes CORINE Land Cover / Urban Atlas and target division of buildings into 3 categories

Tabela 1. Źródłowe klasy CORINE Land Cover / Urban Atlas i docelowy podział zabudowy na 3 kategorie

LULC	Code	Classes in projects	Adopted categories for population conversion
CLC	111	Continuos urban fabric	I – continuos urban
	112	Discontinuos urban fabric	II – discontinuos urban
	242	Complex cultivation patterns	III – isolated structures
UA	11100	Continuous Urban Fabric (impervious surf.> 80%)	I – continuos urban
	11210	Discontinuous (i.s. 50% – 80%)	II – discontinuos urban
	11220	Dense (i.s. 30% – 50%)	
	11230	Urban Fabric (i.s. 10% – 30%)	
	11300	Isolated structures	III – isolated structures

Table 2. Calculated values of population density, in [people/ha] for binary and surface-weight aggregation variants

Tabela 2. Obliczone wartości zagęszczenie ludności, w [os./ha] dla wariantów binarnych oraz powierzchniowo-wagowych

LULC	Adopted categories for population redistribution	binary method	min. RMSE	min. MAPE	binary method	min. RMSE	min. MAPE
		before Tobler's condition			when the condition Tobler u.u.		
CLC	I – continuos urban	62,0	154,8	2017,6	0,25–281,6	61,6–443,0	78,5–5151,1
	II – discontinuos urban		64,0	9,7		0,29–281,6	0,28–281,6
	III – isolated structures		1,8	1,8		0,01–5,1	0,05–11,4
UA	I – continuos urban	107,7	257,2	251,7	10,3–491,4	19,3–1493,5	19,3–1390,9
	II – discontinuos urban		11,6	15,2		0,87–67,1	1,16–83,9
	III – isolated structures		11,4	11,4		0,86–25,6	0,87–24,3

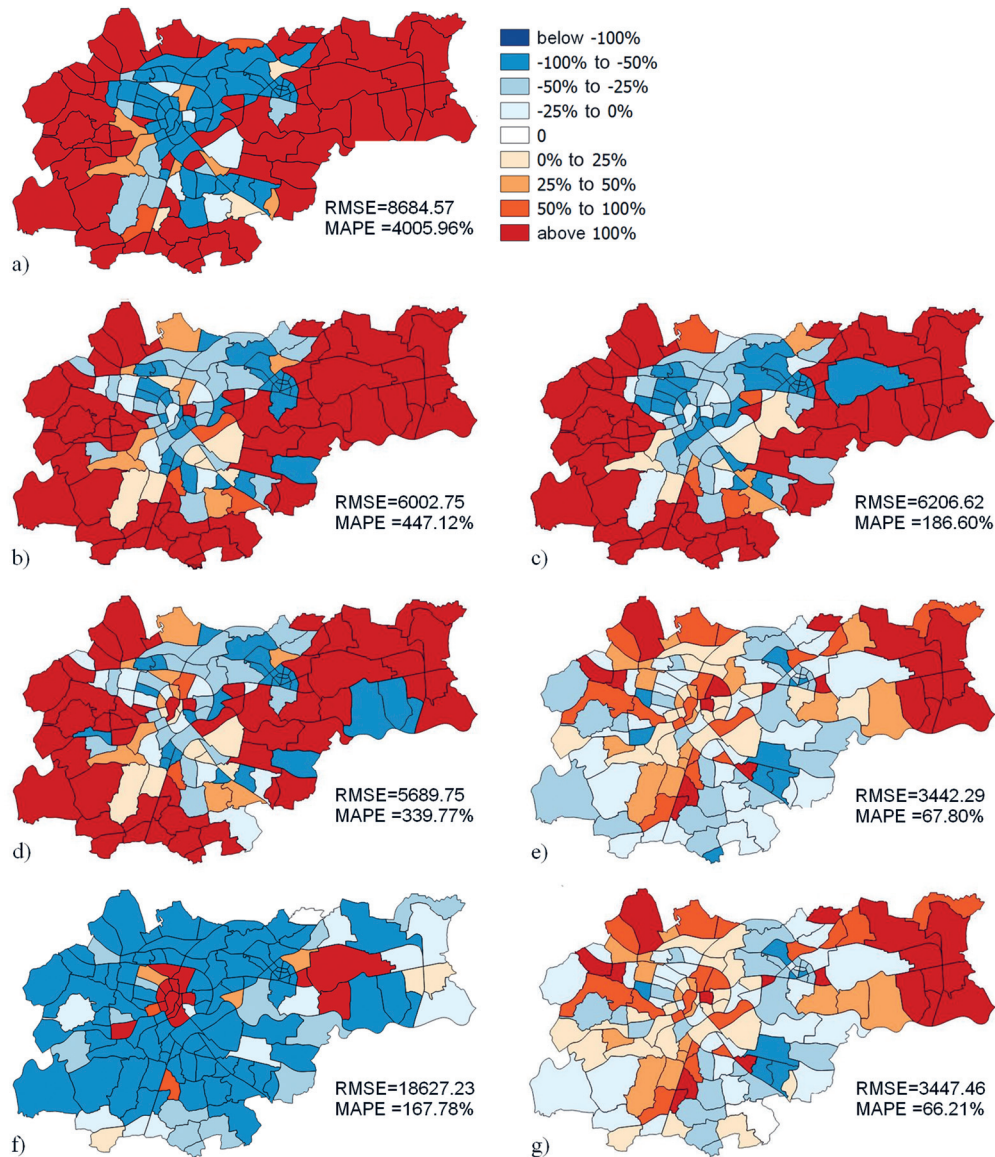


Fig. 4. Percentage errors in urban units for population maps implementing the Tobler condition for the whole city, based on: a) simple cartogram; b) CLC (binary method); c) UA (binary method); d) CLC (minimization of RMSE in u.u.); e) UA (minimization of RMSE in u.u.); f) CLC (minimization of MAPE in u.u.); g) UA (minimization of MAPE in u.u.). RMSE expressed in people/u.u.

Rys. 4. Błędy procentowe w jednostkach urbanistycznych dla map populacji realizujących warunek Toblera dla całego miasta, opartych o: a) kartogram prosty; b) CLC (metoda binarna); c) UA (metoda binarna); d) CLC (minimalizacja RMSE w j.u.); e) UA (minimalizacja RMSE w j.u.); f) CLC (minimalizacja MAPE w j.u.); g) UA (minimalizacja MAPE w j.u.). RMSE wyrażony w osobach/j.u.

value of 0.2% of Krakow's population was assumed as the minimum value assigned for scattered development. The calculated weights for each category of land cover, for individual surface-weight variants, are summarized in Table 2:

The adoption of such compaction values for particular categories enables the first stage of generation of population maps. They meet Tobler's condition, ie the compliance of the population's volume in census units, only globally, for the entire population of the city.

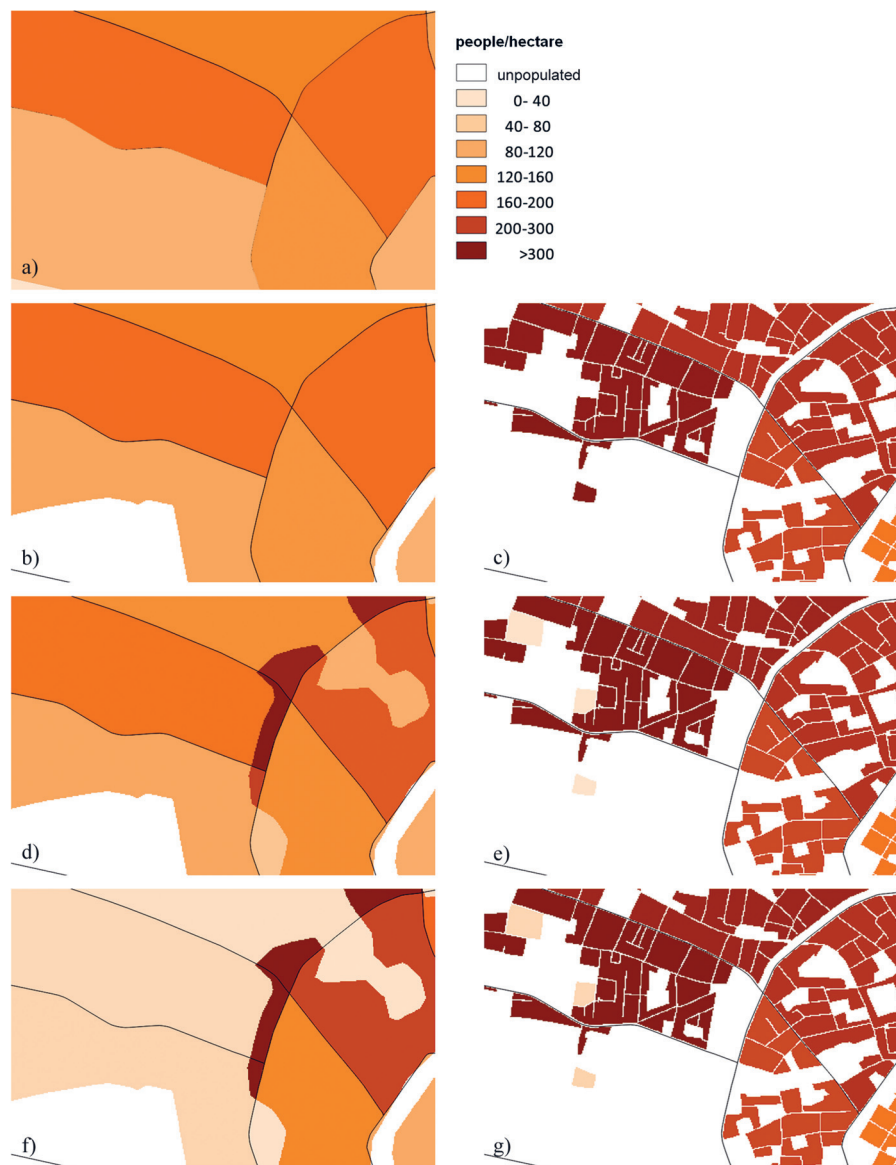


Fig. 5. Dasymeric maps of Krakow – a fragment of the center: a) simple cartogram; b) CLC (binary method); c) UA (binary method); d) CLC (minimization of RMSE in u.u.); e) UA (minimization of RMSE in u.u.); f) CLC (minimization of MAPE in u.u.); g) UA (minimization of MAPE in u.u.).

Rys. 5. Dazymericzne mapy Krakowa – fragment centrum: a) kartogram prosty; b) CLC (metoda binarna); c) UA (metoda binarna); d) CLC (minimalizacja RMSE w j.u.); e) UA (minimalizacja RMSE w j.u.); f) CLC (minimalizacja MAPE w j.u.); g) UA (minimalizacja MAPE w j.u.).

The distribution of errors and average errors for u.u. obtained at this stage, is shown in Figure 4.

The last stage is the implementation of Tobler's condition for each of the u.u. through the application of appropriate coefficients correcting discrepancies between the estimated distribution and statistical data. This treatment maintains the proportion of compac-

tion („weights”) for individual construction categories, where their absolute values are variable for individual census units.

In each of the variants, the population density values calculated using the dasymeric method have been converted into a raster model with a resolution of 5m (each cell is assigned the appropriate number

of population per such area, expressing this in the unit [people/ha]). Figure 5 shows the results for the center of Krakow.

DISCUSSION

The comparative analysis of the various variants is based on the assessment of the correctness of the limiting variables used (correctness of the distribution of housing development defined by the CLC / UA selected from the CLC / UA projects) and on the assessment of the variable relationships (adopted „weights” population density for individual sub-category of residential buildings). Without having additional verification data at this stage, the comparison of options is based on four interrelated information: (1) obtained RMSE / MAPE errors and (2) their distribution in u.u. (on the stage of estimating weights density of buildings, before the correction of the population volume), (3) the stability and realness of the obtained absolute values of population density, when the condition Tobler u.u. and (4) visual analysis of the obtained population maps:

1. Studying Figure 4, it is easy to see that reliance on the same value of population density for the entire city, without information on the extent of development, results in a large underestimation of the population in urban units in the city center, and vice versa – its large overestimation on the outskirts (Fig 4a). Each of the applied dazymetric variants (Fig. 4b–g) reduces errors, which indirectly is the first (intuitively easy to predict without detailed analyzes) argument for the validity of using limiting variables.
2. Decrease of RMSE and MAPE and distribution of errors in u.u. it is different for individual variants, however, there are some tendencies: in binary methods (Fig. 4b, 4c), underestimation in the center and overestimation is still – as in case of 4a, while the number of units with relatively low errors increases. These are the areas between the strict center and the areas lying at the city’s administrative boundaries.
3. UA-based surface-weight aggregation variants reduce global errors clearly, which also manifests in visual distribution (Fig. 4e, 4g). Amount of u.u. with relatively small percentage errors it dominates. The phenomenon of overestimation of the population in u.u. on the outskirts of the city, and undervaluation in the center.
4. For surface-weight maps based on CLC, the improvement, very small, occurs only for the variant based on the determination of scales with minimization of RMSE (Figure 4d). For the CLC method with MAPE minimization, the distribution of errors in u.u. deviates from all other variants (there is a reversal of the spatial location of areas with overestimated and underestimated population), and the density parameters for particular types of buildings are unrealistic (Table 2).
5. Comparing the results with CLC and UA it is clearly visible that in each of the subgroups of methods (binary, surface-weight aggregation with minimization of RMSE, surface-weight aggregation with minimization of MAPE), definitely better results are noted for the UA. This is confirmed by the synthetic parameters, where for the CLC the range of RMSE 5690 people/u.u.-18627 people/u.u is noted; MAPE 168%–447%, and for UA RMSE: 3442 people/u.u. – 6206 people/u.u, MAPE: 66% –186%. While for the binary method the RMSE error is comparable for CLC and UA, the percentage error for Urban Atlas is three times smaller. For surface-weight aggregation methods, the difference is even greater: two to five times lower RMSE and MAPE errors in favor of UA.
6. Optimization of weights by minimizing RMSE and MAPE errors is stable and similar for UA-based maps (Figures 4e and 4g, Fig. 5e, 5g) while the two developed CLC-based maps are completely different despite meeting the minimization conditions (Fig. 4d and 4f, Fig. 5d, 5f). The observed variability is confirmed by the analysis of population density parameters summarized in Table 2, where, depending on the CLC variant, the assigned weight for compact building changes up to 12 times, and the loose construction changes 6 times.
7. When analyzing population maps, it is clearly visible that the extent of CLC development, compared to UA, is very much generalized. It is inconsistent with the network of streets (including the boundaries leading to them) and with the actual spatial distribution of the urban develop-

ment intensity. Conversion of the population to such defined variables of connections, and even only to limiting variables, causes local jumps of population density, not reflected in reality. This is easily seen in the examples of population maps, in the middle of Fig. 5d and 5f.

CONCLUSIONS

The use of limiting variables (urban zones) from the Urban Atlas project provides better mapping of the city's population than the Corine Land Cover project. For a simple binary method, the generation of data on the distribution of building zones in CLC is acceptable for population conversion – the level of RMSE error is kept at a similar level to the AU, although percentage errors are clearly higher. However, the proposed weight optimization approach for surface-weight aggregation methods has proved effective only for Urban Atlas data, where the results based on minimization of RMSE and MAPE are similar, the calculated weights for different types of building are real. For CLC data the results were very divergent, and the variant based on percentage errors gave incorrect weights, despite the good minimization of the MAPE error. The reasons should be attributed to the division of buildings into categories that are inadequate for converting population, and even to determine its extent (Figures 5d, 5f). In the CLC project only a very small area, about 3km², located in the historical center of the city of Krakow, is described as compact. Meanwhile, in this region there is mainly old buildings – tenements, church areas, offices, characterized by a relatively small population density. Other areas of the city, densely populated housing estates, do not fall into this category and are mapped identically, as rural areas, located within the administrative boundaries of the city.

To sum up: to the proposed method of weight optimization, it is necessary to have development categories having a real connection with the number of people living in it. Failure to meet this condition may result in the so-called equifinality phenomenon. Global statistical results seem correct, while this is achieved by an unrealistic set of partial parameters. In the presented research, this situation occurred for the CLC method, with minimizing the MAPE error. The condition of linking the building category with the population seems to meet the AU data, which is confirmed by the research conducted here and presented in the publication of Pirowski

and Pomietlowski (2017). CLC data is possible to use, but it must be based on a binary method or the method of determining the weights must be arbitrary, which was effectively carried out, for example, in the works of Bielecka and others (2005) or Gallego and Peedell (2001). Confirmation of these applications requires in-depth studies of population distribution and its verification, using a set of control, independent statistical data. Such goals will be implemented in the subsequent parts of the publication cycle.

REFERENCES

- Azar, D., Graesser, J., Engstrom, R., Comenetz, J., Leddy JR, R.M., Schechtman, N.G. and Andrews, T. (2010). *Spatial refinement of census population distribution using remotely sensed estimates of impervious surfaces in Haiti*. International Journal of Remote Sensing, 31 (21): 5635–5655, doi:10.1080/01431161.2010.496799
- Bakillah, M., Liang, S., Mobasheri, A., Arsanjani J.J. and Zipf A. (2014). *Fine-resolution population mapping using OpenStreetMap points-of-interest*, International Journal of Geographical Information Science, vol. 28(9):1940–1963, doi: 10.1080/13658816.2014.909045
- Bielecka E., Kuczyk A., Witkowska E. (2005). *Modelowanie powierzchni statystycznej przedstawiającej gęstość zaludnienia w Polsce przy pomocy metody dozymetrycznej*. Polskie Towarzystwo Informatyki Przestrzennej, Roczniki Geomatyki, T. III, Z. 2, 9–16
- Borkowski A., Głowienka E., Hermanowska B., Kwiatkowska-Malina J., Kwolek M., Michałowska K., Mikrut S., Pękala A., Pirowski T., Zabrzaska-Gąsiorek B., (red. Głowienka E.), 2015. *GIS i teledetekcja w monitoringu środowiska*. Rzeszów, WSIE. <http://wsie.edu.pl/wp-content/uploads/2014/06/GIS-srodek.pdf>, ISBN: 978-83-60507-27-8
- Całka B., Bielecka E., Zdunkiewicz K. (2016). *Redistribution population data across a regular spatial grid according to buildings characteristics*, Geodesy and Cartography; Vol. 65, no. 2, pp. 149–162
- Cockx K., Canters F. (2015). *Incorporating spatial non-stationarity to improve dasymetric mapping of population*. Applied Geography, Vol. 63, s. 220–230
- Eicher, C.L., Brewer C.A. (2001). *Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation*. Cartography and Geographic Information Science, Vol. 28, No. 2, April, pp. 125–138(14)
- Gallego F.J., Peedell S. (2001). *Using CORINE Land Cover to map population density*. Towards agri-environmental indicators. EEA Topic report 6/2001, 94–105
- Goleń i Ostrowski (1994). *Metoda dazymetryczna – rys historyczny*, Polski Przegląd Kartograficzny 26(1): 3–16
- Gregory, I.N. (2002). *The accuracy of areal interpretation techniques: standardizing 19th and 20th century census data to allow long-term comparisons*. Computer, Environment and Urban Systems 26: s. 293–314.

- Harvey J. (2002). *Estimating Census District Populations from Satellite Imagery: Some Approaches and Limitation*, International Journal of Remote Sensing 23(10):2071–2095, May
- Horanont, T. and Shibasaki, R. (2010). *Estimate ambient population density: discovering the current flow of the city*. Available on line: https://www.academia.edu/2004297/estimate_ambient_population_density_discovering_the_current_flow_of_the_city
- Hościło A., Tomaszewska M. (2014). *CORINE Land Cover 2012–4th CLC inventory completed in Poland*. Geoinformation Issues, Vol. 6, No. 1 (6), s. 49–58
- Kruszyńska A., Wendel I. (2001). *Dzielnice Krakowa*, Urząd Miasta Krakowa, Kancelaria Rady Miasta i Dzielnic Krakowa, Kraków, s. 42
- Kucharczyk M. (2012). *Analiza cech podmiejskiej zabudowy mieszkaniowej na przykładzie inwestycji deweloperskich zrealizowanych w okolicy Krakowa*. Politechnika Krakowska, Czasopismo Techniczne. Architektura, z. 29
- Lin J., Cromley R.G. (2015). *Evaluating geo-located Twitter data as a control layer for areal interpolation of population*. Applied Geography, Vol. 58, s. 41–47
- Maantay J.A., Maroko A.R., Herrmann Ch. (2007). *Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymetric System (CEDS)*. Cartography and Geographic Information Science
- Mapping Guide for European Urban Atlas*, 2011; <http://www.eea.europa.eu/data-and-maps/data/urban-atlas#tab-methodology>
- Mennis J. (2003). *Generating Surface Models of Population Using Dasymetric Mapping*. The Professional Geographer, Vol. 55, No. 1
- Mennis, J., Hultgren, T. (2006). *Intelligent dasymetric mapping and its application to area interpolation*. Cartography and Geographic Information Science, 33(3): 179–194
- Pirowski T., Drzewiecki W. (2012). *Mapa gęstości zaludnienia Krakowa, propozycja metodyki opracowania oraz przykładowe zastosowania*. Roczniki Geomatyki, t. 10, z. 3
- Pirowski T., Pomietłowska J. (2017). *Distribution of Krakow's Population by Dasymetric Modeling Method Using Urban Atlas and Publicly Available Statistical Data*, Geomatics and Environmental Engineering, vol. 11/4, 83–95
- Robinson A.H., Sale R.D., Morrison J.L. Muehrcke P.C. (1998). *Podstawy Kartografii*, PWN, Warszawa
- Sleeter, R., Wood N. (2006). *Estimating daytime and nighttime population density for coastal communities in Oregon: Urban and Regional Information Systems Association. Annual Conference, Proceedings, Vancouver, BC, September 26–29*
- Smith, A., Newing, A., Quinn, N., Martin, D., Cockings, S. and Neal, J. (2015). *Assessing the Impact of Seasonal Population Fluctuation on Regional Flood Risk Management*. ISPRS Int. J. Geo-Inf. 2015, 4:1118–1141, DOI:10.3390/ijgi4031118
- Sridharan, H. and Qiu, F. (2013). *A Spatially Disaggregated Areal Interpolation Model Using Light Detection and Ranging-Derived Building Volumes*. Geographical Analysis, 45:238–258
- Tapp A.F. (2010). *Areal Interpolation and Dasymetric Mapping Methods Using Local Ancillary Data Sources*. Cartography and Geographic Information Science, Vol. 37, No. 3
- Tobler R.T. (1979). *Smooth pycnophylactic interpolation for geographic regions*. Journal of The American Statistical Association 74, 519–530
- Urban Atlas – delivery of land use/cover maps of major European Urban Agglomerations – Final Report (V 2.0)*, 2011; http://ec.europa.eu/regional_policy/sources/tender/pdf/2012066/urban_atlas_final_report_112011.pdf
- Web sites *StatKraK* prowadzony przez Wydział Strategii i Rozwoju Miasta Urząd Miasta Krakowa, *Referat Analiz i Prognoz Rozwoju Miasta*; www.msip2.um.krakow.pl/statkrak/, dostęp: 22.11.2016
- <http://www.igik.edu.pl/pl/corine-projekt> (dostęp: 26.05.2017)
- Wu, C. and Murray, A.T. (2005). *A cokriging method for estimating population density in urban areas*. Computers, Environment and Urban Systems, 29 (5), 558–579, doi:10.1016/j.compenvurbsys.2005.01.006