Tomasz Pirowski[1]
ORCID: 0000-0003-0095-0316

Małgorzata Timek[2]
ORCID: 0000-0002-6682-8228

ANALYSIS OF THE USEFULNESS LAND USE AND LAND COVER MAPS FOR ESTIMATING THE POPULATION OF URBAN AREAS – VALORISATION OF MULTI-VARIANT POPULATION MAPS BASED ON THE GUS KILOMETRE NETWORK[1]

[1,2] AGH University of Science and Technology,
Faculty of Mining Surveying and Environmental Engineering

Keywords: Census data of the Central Statistical Office, dasymetric modeling, error rate correlation, LULC

Abstract

The paper is a continuation and summary of a series of publications related to the dasymetric estimation of the distribution of the population of Krakow. The conversion of the population from the original census units is based on the development data from three sources, the Corine Land Cover project (CLC), the Urban Atlas project (UA) and the object classification (OBIA) of the RapidEye data. The experiment was conducted using archival statistical data from 2009 from 141 urban units (u.u.) of the city.

In the first two parts of the cycle (Pirowski and Timek, 2018; Pirowski et al., 2018) population conversion was presented on the basis of CLC, UA and OBIA maps, obtaining a total of 12 maps of Krakow's population. The obtained error distributions were presented and the calculated weights of population density for each category of residential buildings were discussed. In the third part of the cycle (Pirowski and Berka, 2019) the results were analyzed in detail by reference to the reference, high-resolution population map of the Bronowice district (north-western part of the city).

In this publication, ending the cycle, population maps were verified on the basis of a kilometre grid of the Central Statistical Office (GUS), which is an aggregation of data from the National Census of Population and Housing 2011, made available by the Office in 2017. The results of high-resolution verification carried out in the Bronowice district were compared with the data of the CSO (GUS). In the GUS grid the best results were obtained for surface and weight UA methods (RMSE 908–917 people; MAPE 42-46%). The estimation of population distribution using OBIA data (RMSE 1115–2073 people; MAPE 121–184%) was found to be incorrect. After the correction of OBIA by UA data, a significant improvement in the results for surface-weighted methods was obtained (RMSE 930–1067 people; MAPE 53–68%), however, the error rate was still higher than for UA itself, which eliminates the OBIA method from practical applications in this area.

A correlation was found between the RMSE and MAPE errors recorded in UC at the stage of weight selection and the RMSE and MAPE errors recorded in the GUS grid, respectively $R2(RMSE)=91\%$, $R2(MAPE)=65\%$. Therefore, the correlation detected indicates that the low errors obtained at the selection stage translate into reliable population estimates. The proposed weighting methodology limits the subjectivity of the method, based on the minimisation of RMSE and MAPE in the original census units. The disadvantage of the method is that it is necessary to define the boundary conditions for the selection of weights, in case of obtaining unreal weights and the possibility of occurrence of equifinality phenomenon, difficult to detect in the absence of additional reference data.

ANALIZA PRZYDATNOŚCI MAP POKRYCIA I UŻYTKOWANIA TERENU DO SZACOWANIA LICZBY LUDNOŚCI OBSZARÓW MIEJSKICH – WALORYZACJA WIELOWARIANTOWYCH MAP LUDNOŚCI W OPARCIU O SIATKĘ KILOMETROWĄ GUS

Słowa kluczowe: dane spisowe GUS, modelowanie dazymetryczne, mapy pokrycia i użytkowania terenu, korelacja poziomu błędów

Abstrakt

Artykuł jest kontynuacją i podsumowaniem cyklu publikacji związanych z dazymetrycznym szacowaniem rozmieszczenia ludności Krakowa. Przeliczanie ludności z pierwotnych jednostek spisowych oparto na danych o zabudowie z trzech źródeł, z projektu Corine Land Cover (CLC), z projektu Urban Atlas (UA) oraz z klasyfikacji obiektowej (OBIA) danych RapidEye. Eksperyment przeprowadzono wykorzystując archiwalne dane statystyczne z roku 2009 ze 141 jednostek urbanistycznych (j.u.) miasta.

W pierwszych dwóch częściach cyklu (Pirowski i Timek, 2018; Pirowski i in., 2018) zaprezentowano przeliczanie populacji na bazie map CLC, UA oraz OBIA, łącznie uzyskując 12 map zaludnienia Krakowa. Przedstawiono uzyskane rozkłady błędów, poddano dyskusji obliczone wagi zagęszczenia ludności dla każdej kategorii zabudowy mieszkalnej. W trzeciej części cyklu (Pirowski i Berka, 2019) opracowane wyniki poddane zostały szczegółowej analizie poprzez odniesienie się do referencyjnej, wysokorozdzielczej mapy zaludnienia dzielnicy Bronowice (północno-zachodni obszar miasta).

W niniejszej publikacji, kończącej cykl, zweryfikowano mapy zaludnienia w oparciu o siatkę kilometrową GUS, będącą agregacją danych Narodowego Spisu Powszechnego Ludności i Mieszkań z 2011, udostępnioną przez Urząd w 2017 roku. Porównano wyniki weryfikacji wysokorozdzielczej prowadzonej na dzielnicy Bronowice z weryfikacją na danych GUS. W siatce GUS uzyskano najlepsze wyniki dla metod powierzchniowo-wagowych UA (RMSE 908–917 osób; MAPE 42–46%). Za błędne uznano szacowanie rozmieszczenia ludności przy użyciu danych OBIA (RMSE 1115–2073 os.; MAPE 121–184%). Po korekcie OBIA poprzez dane UA uzyskano znaczącą poprawę wyników dla metod powierzchniowo-wagowych (RMSE 930–1067 osób; MAPE 53–68%), jednak poziom błędów był nadal wyższy niż dla samej UA, co eliminuje metodę OBIA z zastosowań praktycznych w tym obszarze.

Stwierdzono zależność pomiędzy notowanymi błędami RMSE i MAPE w j.u. na etapie doboru wag a notowanymi błędami RMSE i MAPE w siatce GUS, odpowiednio $R^2_{(RMSE)} = 91\%$, $R^2_{(MAPE)} = 65\%$. Zatem wykryta korelacja wskazuje, że niskie błędy uzyskane na etapie doboru wag przekładają się na wiarygodne szacowanie liczby ludności. Proponowana metodyka doboru wag ogranicza subiektywizm metody, opierając się na minimalizacji RMSE i MAPE w pierwotnych jednostkach spisowych. Wadą metody jest konieczność definiowania warunków brzegowych doboru wag, w przypadku uzyskiwania nierzeczywistych wag oraz możliwość wystąpienia ekwifinalności, trudnej do wykrycia przy braku dodatkowych danych referencyjnych.

INTRODUCTION

Population maps are an important source of data for spatial analyses carried out in geographical information systems (GIS). They are included in indicator maps, in which the way data are presented determines the methodology of product development. Until the 1990s, presentation in the form of isolines, dot maps and simple cartograms prevailed (Robinson et al., 1988). With the development of data processing, mainly due to the popularization of GIS systems, progress in cartography and availability of digital spatial data, more and more dasymetric methods were used. They consist in the adoption of other reference fields – in cartograms these are imposed areas, usually administrative, in a dasymetric approach – these are areas related to the spatial distribution of the phenomenon (Goleń and Ostrowski, 1994). As a result, the representation of the spatial distribution of the depicted phenomenon with the use of cheap dasymetric methods is more reliable than in the case of a cartogram (Longley et al., 2006). In the case of population maps in a simple cartogram, the reference units are census units, in dasymetric methods – the range of development.

The first, basic problem is to find a correct way to convert the population from census units to units related to buildings (so-called limiting variables), including the diversity of buildings (so-called variable relationships) resulting from their nature (density, height, location, function, etc.). A more extensive overview of the

conversion methods used can be found, for example, in works by Eicher and Brewer (2001), Wu et al. (2005), Wang and Wu (2010), Całka et al. (2016).

The second problem is the correct determination of limiting variables and variable relationships. The implementation at the national level of detailed databases on buildings and monitoring of the coverage of overland areas (Corine Land Cover, Urban Atlas) were quickly reflected in population mapping (Gallego and Peedell, 2001; Bielecka, 2005; Kunze and Hecht, 2015; Cockx and Canters 2015; Batista and Poelman, 2016; Całka et al., 2016; Pirowski and Pomietłowska, 2017; Pirowski and Bartos, 2018). These databases are in many cases insufficient – they do not cover a given area, they have a limited time range and too low spatial accuracy. Therefore, experiments are still being undertaken to correlate the distribution of population with other data. This is most often achieved through remote registration – aerial photographs, high-resolution satellite data, laser scanning are used (e.g. works by Chen, K., 2002; Liu, 2004; Azar et al., 2010; Ural et al., 2011; Pirowski and Drzewiecki, 2012; Upegui and Viel, 2012). An example of analysis in areas devoid of systemic support by land cover databases, i.e. requiring the use of satellite data, are publications Nagle and Rose (2017), Wei and Blaschke (2018), Weber et al. (2018). Therefore, understanding the limitations of CLC and UA databases, the publication cycle also includes object-oriented mapping of the coverage. Confirmation of the suitability of an automatic cardholder for population conversion would enable the use – under national conditions – of vivid archival data for long-term analyses of population distribution.

The third problem is the reliable verification of the converted population maps. Only in a few publications has a detailed valorisation of methods or estimation of the correctness of the obtained detailed distribution been carried out. An example are works by Liu (2005), Zandberg and Ignizio (2010), Suddenly et al. (2014), Rose and Suddenly (2017), Weber et al. (2018).

The research objectives of this publication, which concludes the research cycle, are:

– valorisation of twelve developed variants of methods (Pirowski and Timek, 2018; Pirowski et al., 2018) on the basis of a set of independent statistical data; these data were published by the CSO (GUS) at the beginning of 2017 for the whole of Poland in the form of a kilometre grid; these data were not used to calculate the population based on CLC/UA/OBIA; they allow for an objective assessment and the answer which of the generated population maps is the best, i.e. after re-aggregation, this time to the „meshes” of the GUS grid, the lowest deviations are recorded – after re-aggregation, this time to the „meshes” of the GUS grid – in the smallest deviations from the GUS grid;

– due to the possibility of the occurrence of the equifinality phenomenon – reference to the analysis made on the basis of a high-resolution photo-interpretation map of the population of the Bronowice district; detailed methodology and results were published in the previous article of the cycle (Pirowski and Berka, 2019);

– development of a ranking of 12 analysed variants of population maps;

– providing advantages and limitations of the proposed methodology for determining weights in the dasymetric method, in the form of a set of recommendations.

The double verification of results carried out in different resolutions and the related discussion is a novelty against the background of published methodological experiences and attempts to provide usefulness of dasymetric population calculation.

METHODOLOGY OF POPULATION MAPPING

Population maps, which have been valorised, were prepared on the basis of three sources of data on the development of Corine Land Cover (CLC), Urban Atlas (UA) and object classification (OBIA) of the Rapid-Eye image. Population conversions were based on the dasymetric method, treating building zones as limiting variables and division into building types as connecting variables. The general scheme of action is presented in Fig. 1.

The stage of data preparation was related to data acquisition (downloading from dedicated CLC/UA services) or their processing (RapidEye data classification), then their spatial calibration. The next stages included separating the buildings, obtaining information

Fig. 1. Main steps in the development of dasymetric maps
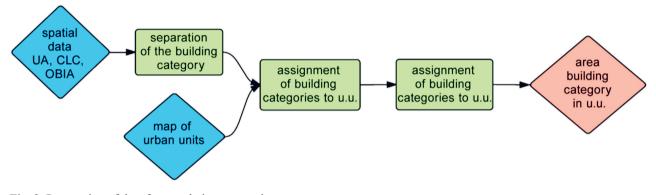**Rys. 1.** Główne etapy opracowania map dazymetrycznych

Fig. 2. Preparation of data for population conversion
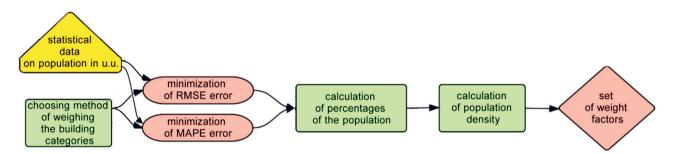**Rys. 2.** Przygotowanie danych do przeliczania ludności

Fig. 3. Determination of weighting factors for surface/weight methods
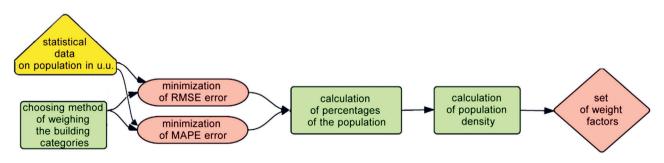**Rys. 3.** Wyznaczanie współczynników wagowych dla metod powierzchniowo-wagowych

Fig. 4. Development of a dasymetric map
Rys. 4. Opracowanie mapy dazymetrycznej

on their segmentation, surface areas in individual urban units of the city (u.u.) (Fig. 2)

The population density weights, adopted for each land cover category, were calculated iteratively, minimizing deviations in individual caves. (Fig. 3). Details of the methodology were included in the previous publications of the cycle.

The last stage is to use the calculated weight coefficients and maps of the development range to determine the population maps. This process still needs to be corrected in order to remove the observed deviations in individual u.u. (Fig. 4).

A total of 12 variants of maps of Krakow's population have been developed, i.e: 3 maps from CLC, 3 maps from UA, 3 maps from OBIA and 3 maps from OBIA+UA (in each subgroup variants: binary, surface-weight variants with RMSE minimization and MAPE minimization).

PREPARATION OF VERIFICATION DATA FROM THE GUS GRID

Verification of dasymetric population maps for the whole city was based on the kilometre grid of the Central Statistical Office (GUS) made available in 2017 (Fig. 5). The information contained therein is a result of aggregation of data from the National Census of Population and Housing 2011 (https://geo.stat.gov.pl). Among a series of information on population, only the total number of inhabitants was used, without any division into sex or age groups.

Since there is no information about the method of registration, it is difficult to say how big the difference between data in urban units and data in a kilometre grid can be. Analytically, this can only be done with a rough approximation, as part of the grid meshes lies on the administrative border of Krakow and the surrounding

Fig. 5. Population in individual meshes of the CSO grid on the example of the centre of Krakow
Rys. 5. Liczba ludności w poszczególnych oczkach siatki GUS na przykładzie centrum Krakowa
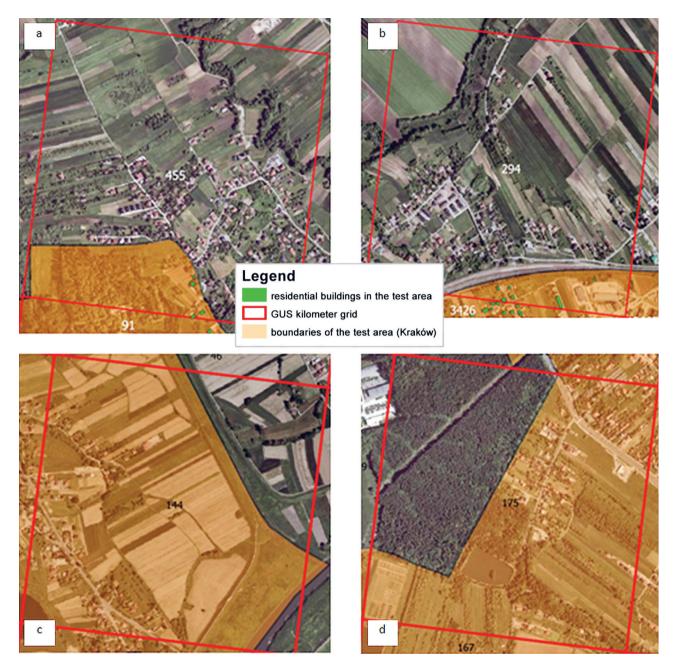
Fig. 6. Development of a GUS network for the needs of analysis. Example of border mesh meshes: a, b) eliminated; c, d) not eliminated

Rys. 6. Opracowanie siatki GUS na potrzeby analizy. Przykładowe graniczne oczka siatki: a, b) wyeliminowane; c, d) pozostawione

communes. Therefore, for statistical verification of methods, only those meshes of the GUS grid were used which were within or crossed the administrative boundaries of Krakow, not including the population of the neighbouring towns and cities (Fig. 6). Ultimately, there were 300 meshes left, of which 292 km² within Krakow,

i.e. 97.3% of the city's area. For this defined area, the total population amounted to 756503 people, which is a value similar to the population from statistical data in urban planning units amounting to 744254 people (the difference in both sets of statistical data amounted to 2.1%).

VERIFICATION OF DASYMETRIC MAPS BASED ON GUS GRID

Raster population maps, obtained dasymetrically from CLC / UA / OBIA / OBIA / OBIA+UA, were again aggregated in new spatial units, i.e. in the kilometre grid of the Central Statistical Office (GUS) and referred to statistical data published by the Office. The distribution of errors, obtained for particular variants, is presented in Fig. 7. As a reference point, the results obtained on the basis of a simple cartogram, based only on population data from urban units, without the use of limiting variables (Fig. 7a), are additionally illustrated.
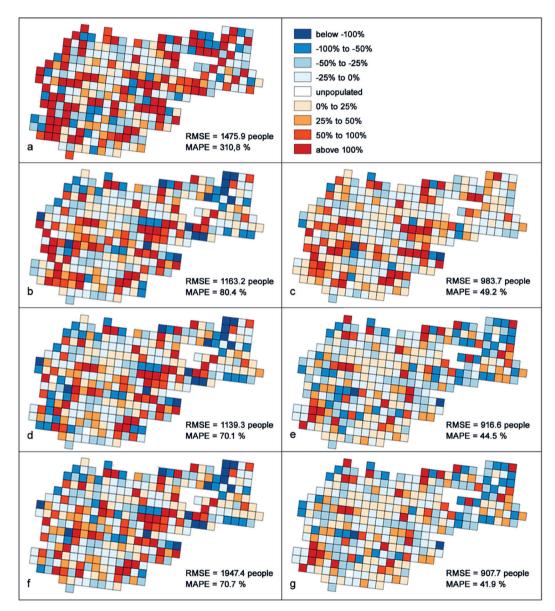

Fig. 7. Percentage errors in the kilometer grid of GUS for maps of populations based on: a) simple cartogram; b) CLC (binary method); c) UA (binary method); d) CLC (RMSE minimization in u.u.); e) UA (RMSE minimization in u.u.); e) UA (RMSE minimization in u.u.); f) CLC (min. MAPE in u.u.); g) UA (min. MAPE in u.u.)

Rys. 7. Błędy procentowe w siatce kilometrowej GUS dla map populacji 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 (min. MAPE w j.u.); g) UA (min. MAPE w j.u.)
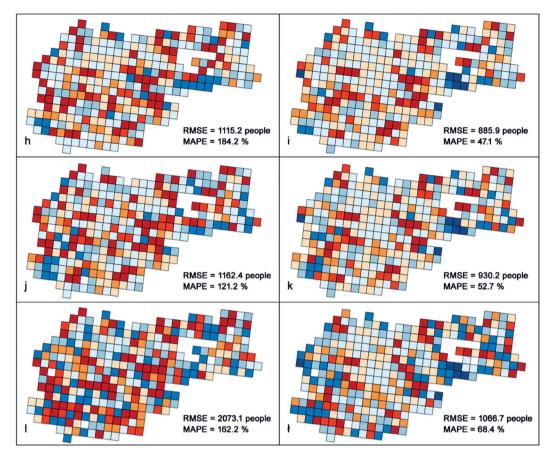
Fig. 7. h) OBIA (binary method); i) OBIA+UA (binary method); j) OBIA (min. RMSE in u.u.); k) OBIA+UA (min. RMSE in u.u.); l) OBIA (min. MAPE in u.u.); ł) OBIA+UA (min. MAPE in u.u.)
Rys. 7. h) OBIA(metoda binarna); i) OBIA+UA (metoda binarna); j) OBIA (min. RMSE w j.u.); k) OBIA+UA (min. RMSE w j.u.); l) OBIA (min. MAPE w j.u.); ł) OBIA+UA (min. MAPE w j.u.).

ANALYSIS OF RESULTS

For population maps based on CLC, despite a large generalization of development zones, each of the tested variants of population conversion significantly reduced the average percentage errors recorded in the GUS grid, from about 310% (simple cartogram, Fig. 7a) to 70–80% (Fig. 7b, 7d, 7f). The result of the area-weight method based on MAPE minimisation (Fig. 7f) gives false weight coefficients (85% of the population of Krakow is assigned to the Old Town, which is only 3 km² in size), which results in a very high population density for compact buildings. This error is manifested by a high RMSE error, higher than for a simple cartogram. For all population maps based on UA, lower errors were obtained by about 30% than for CLC. In UA variants (Fig. 7c, 7e, 7g) there was a compliance between MAPE and RMSE parameters: the worst result was obtained for the binary method (Fig. 7c), the best for the surface-weight MAPE method (Fig. 7g, RMSE=908 people, RMSE=42%).

For both CLC and UA, the greatest effect of improvement in population redistribution was observed through the use of limiting variables (Fig. 7b, 7c; MAPE decreased from 310% to about 80% for CLC, to 50% for UA). The introduction of variable relationships (Fig. 7d, 7e, 7f, 7g) improves the results by a further 7–10%. Analyzing the distribution of deviations in the GUS grid, it is easy to notice that for specific areas of the city („mesh" grid), larger deviations are recorded, regardless of the method of population calculation. The location of these meshes depends on the LULC data used. It should be concluded that the results are determined primarily by the distribution of limiting variables

(presence or absence of building zones), and the division into types of buildings (and the weight of population density assigned to them) is secondary. Places with higher errors are likely to have incorrectly delimited or outdated built-up areas.

For maps based solely on OBIA (Fig. 7h, 7j, 7l), percentage errors in the GUS grid are about 2–3 times higher than in relation to the previously discussed CLC and UA. The location of places with greater deviations is variable depending on the variant. The RMSE parameter shows the lowest errors in the binary method (Fig. 6h). Surface-weight method with RMSE minimization (Fig. 7j) gives questionable results – the improvement of MAPE parameter is accompanied by the increase of RMSE value. The last variant with MAPE minimization (Fig. 7l) gives very high errors and different spatial distribution of „meshes" with increased deviations. Much better results are obtained with complementary use of OBIA and UA (Fig. 7i, 7k, 7ł). Elimination of industrial areas by UA causes a large drop in errors to lower than for CLC, but higher than for UA variants. Building segmentation performed by OBIA is not correct from the population calculation point – therefore, the lowest errors are obtained for the binary method (Fig. 7i).

VALORISATION OF METHODS

In total, 12 variants of maps of Krakow's population were developed in the conducted research. They were subjected to detailed analyses based on two independent sets of reference data. The basic parameters obtained in the process of generating population maps and the valorisation of the variants are presented in Table 1.

The conversion of population was a step-by-step process: in the first step, maps were generated taking into account limiting variables and changing connections, using only the total population of the city. As a result, for each u.u., values of population were obtained that were different from those resulting from statistical data. By appropriately selecting the proportions of „weights" for particular types of buildings, the surface-weighting methods minimized the square or percentage deviations

Tab. 1. Errors for 12 population map variants obtained at different stages of their generation
Tab. 1. Błędy dla 12 wariantów map populacji uzyskane na różnych etapach ich generowania

LULC	variant of population mapping	RMSE [people]	MAPE [%]	MMAPE [%]	MMAPE [%]	RMSE [people]	MAPE [%]
		before Toble's cond.		subject to Tobler's condition (er. in u.u.=0)			
		(errors in u.u.)		50m	100m	errors in the CSO grid	
cartogram	simple	8684.57	4005.96	31.64	26.55	1475.88	310.77
CLC	binary	6002.75	447.12	29.54	24.35	1163.20	80.43
	surf.-weight RMSE	5689.75	339.77	29.54	24.35	1139.34	70.13
	surf.-weight MAPE	18627.23	167.78	29.54	24.35	1947.37	70.67
UA	binary	6206.62	186.60	26.42	22.15	983.67	49.25
	surf.-weight. RMSE	**3442.29**	67.80	24.80	19.52	916.61	44.52
	surf.-weight MAPE	3447.46	**66.21**	**24.59**	**19.30**	**907.75**	**41.88**
OBIA	binary	7222.35	6611.68	28.18	22.79	1115.18	184.20
	surf.-weight RMSE	5957.68	2427.42	30.76	25.61	1162.39	121.22
	surf.-weight MAPE	7987.04	2262.45	42.19	39.33	2073.15	162.24
OBIA + UA	binary	4540.26	108.09	26.44	21.58	885.89	47.08
	surf.-weight RMSE	4269.61	85.46	27.02	22.16	930.24	52.71
	surf.-weight MAPE	4468.83	74.80	28.97	24.30	1066.72	68.38

recorded in the u.u. (Pirowski and Timek, 2018; Pirowski et al., 2018). The measure of this stage are the values of RMSE and MAPE achieved (Table 1, „RMSE and MAPE before Tobler's condition").

In the second step, in each of the urban units an appropriate coefficient was applied to remove the deviations. In this way, the final values of population density in a given variant were calculated for a given u.u. This formed the basis for the final population maps, which were double-checked on independent data. The analysis with a detailed reference map resulted in a percentage of population error at a given resolution, from 5m to 500m. Table 1 presents the results for two key resolutions, i.e. 50m and 100m („MMAPE 50m and MMAPE 100m, subject to Tobler's condition in u.u.") (Pirowski and Berka, 2019). The second control was aggregative: population maps were converted to a kilometre grid and compared with data published in this resolution by the Central Statistical Office („RMSE and MAPE, errors in the CSO grid, subject to Tobler's condition in u.u.").

The binary method of map conversion is simple and subjective. The reliability of the obtained population map is unknown, depending on the correct definition and spatial range of built-up areas (inhabited areas) and internal diversification of building density, which is not taken into account in the binary method. The surface-weight method attempts to solve this last problem by selecting the density coefficients („weights") for

particular types of buildings. However, this process is burdened with subjectivism. The attempt made in this publication cycle concerned the answer to the question: can the quality of obtained final maps of the population be predicted by applying the rules of statistical, i.e. objective, search for these proportions by minimizing deviations in the original census units? Will the recorded errors be lower than for binary methods when using surface-weighted methods, and will this effect not be achieved through the use of equifinality phenomenon (low deviations for aggregated data for larger spatial units, with an incorrect distribution of the phenomenon within these units)? To what extent does the accuracy of determination of development zones (i.e. reliability and timeliness of the source of data on land cover) influence the quality of population conversion? Thanks to the preparation and proper use of two independent reference maps, the answers to the above questions were obtained.

Reference to the reference map of Bronowice (Pirowski and Berka, 2019), characterized by high spatial resolution of population mapping, made it possible to detect the equifinality phenomenon for the OBIA(MAPE) variant. For this population map an incorrect selection of weights was obtained for particular categories of buildings, difficult to catch through the analysis of MAPE and RMSE errors in the u.u. and GUS grid (calculated errors are relatively high, but
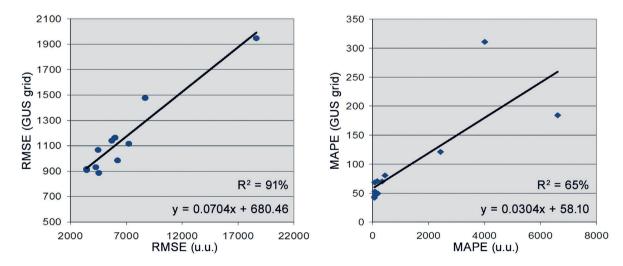

Fig. 8. Correlation of error rate for particular variants, calculated between errors quoted in u.u. and errors quoted in the GUS grid, for minimizing deviations: a) RMSE; b) MAPE
Rys. 8. Korelacja poziomu błędów dla poszczególnych wariantów, obliczona pomiędzy błędami notowanymi w j.u. a błędami notowanymi w siatce GUS, dla minimalizacji odchyłek: a) RMSE; b) MAPE
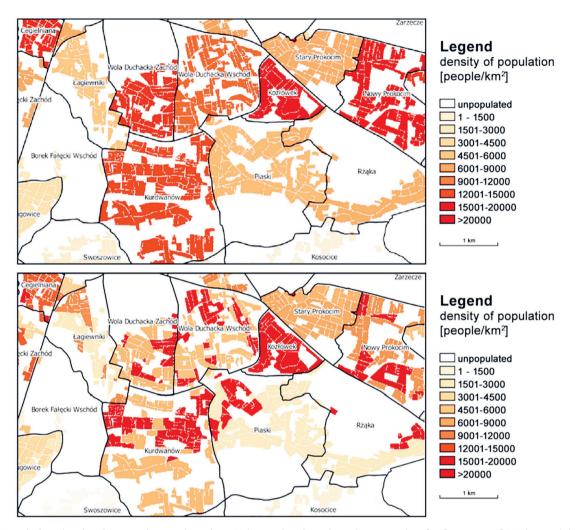
Fig. 9. Population density dasymetric map based on Urban Atlas, based on the example of a fragment of Krakow: a) binary method, b) surface-weighted aggregation method with minimization MAPE

Rys. 9. Mapa dazymetryczna gęstości zaludnienia oparta o Urban Atlas, na przykładzie fragmentu Krakowa: a) metoda binarna, b) powierzchniowo-wagowa metoda agregacji z minimalizacją MAPE

do not differ from some other variants). A very high MMAPE parameter (Table 1) indicates incorrect selection of scales.

For the remaining 11 variants and the simple cartogram, the RMSE and MAPE parameters were compared with the GUS grid (Fig. 8). There is a relation between low errors at the stage of scales selection and product quality, understood as the recorded level of errors in the GUS grid. Greater dependence was detected for RMSE (R^2=91%) than MAPE (R^2=65%).

Among the tested variants, the best results were obtained for surface-weight methods based on UA data (bold values in Table 1). He is responsible for the good

result: (1) adequate level of detail; (2) correct segmentation of buildings; (3) use of additional databases and/or photo-interpretations to determine the functions of buildings. The results, obtained on the basis of UA, are illustrated in Figure 9.

Coverage data, based on automatic classification, do not give good results. The following problems were identified: (1) No possibility of correct elimination of areas of industrial and commercial development unnecessary in the process of calculating the city's night population; (2) doubtful segmentation of development. These conclusions are confirmed by a significant improvement in the binary method due to the modifica-

tion of development areas by the UA and worse results for the OBIA+UA surface-weight variants. To sum up, in the OBIA method the mapping of the development range is carried out correctly, but it is not sufficient to calculate the population.

The proposed methodology of population conversion to new spatial units is mainly determined by limiting variables, which is manifested by large drops in errors already recorded for binary methods. The introduced segmentation of buildings is of secondary importance – carried out correctly, as in the data from the UA, gives a further and stable improvement in the estimation of the population distribution, regardless of the optimisation variant applied.

SUMMARY AND CONCLUSIONS

1. The proposed methodology of optimization of weights of surface-weight methods can be applied on condition that a map is available with building categories that have a real connection with the number of inhabitants. Urban Atlas data meet this condition, which is confirmed by the results presented in the works of Pirowski and Timek (2018), Pirowski and Pomietłowska (2017). CLC data are available for use, but population conversion must be based on a binary method or the method of determining weights must be arbitrary, which is consistent with the conclusions of Bielecka et al. (2005). OBIA data do not allow for correct population calculation (Pirowski et al., 2018). The limited number of surveys does not allow to formulate what boundary conditions should be met by LULC maps in order to be considered reliable for use.

2. In case of lack of reference data it is recommended to use RMSE as a measure of weight calculation. This parameter is more related to the expected quality of the final map. Relatively high RMSE values also indicate incorrect body segmentation and/or equifinality phenomenon. If the RMSE parameter for the surface-weight method is increased in relation to the binary method, it is recommended to use the latter.

3. Determining weights based on statistical measurements may cause the equifinality phenomenon and/or incorrect selection of weights. In the absence of reference data, a large discrepancy between the results of MAPE and RMSE optimisation and the

control of the weights established allows the partial exclusion of these problems. In some cases, it is necessary to provide boundary conditions for certain types of buildings, because automatically calculated weights, in an iterative process, are unrealistic, e.g. negative. This is a problem in case of lack of additional data. The problems observed are not always „sharp" in nature, hence the difficulty in distinguishing a good result from an erroneous one. The small number of studies conducted does not allow to formulate unambiguous solutions.

To summarise: the improvement in the estimation of population distribution, obtained by the surface-weight method, is subjective. The minimisation of this disadvantage can be achieved through expert experience, by supporting the conversion process with additional data, e.g. allowing to estimate the population density per given type of land cover. The second solution, applied here, is to use statistical measures to calculate weights. The validity of this approach has been sought by using extensive reference data. At this stage of the research it was confirmed that having different sets of data on land cover, the best solution can be selected and weights can be automatically selected, based only on RMSE errors (secondly MAPE errors) recorded in the original census units. However, the calculation of weights must be controlled by the operator, as a fully automatic approach, without boundary conditions and without analysing the quality of the building data, can lead to incorrect selection of weights and, in extreme cases, equifinality phenomenon.

The application of the presented method is possible without detailed reference data on population. Then it is possible to estimate the population on the basis of archival data on land cover (e.g. aerial or satellite images), where a limited amount of statistical data is available. The main problem then becomes the development of a sufficiently reliable map of limiting variables and variable connections, i.e. a properly categorised map of zones of building type. The conducted analyses indicate that the methods of automatic image classification fail. The best data from the Urban Atlas shown in the analysis confirm that this process should be supported by interpretation and additional data, especially when it comes to determining the function of buildings (residential/non-residential). Using the UA data as a benchmark, it is planned to improve the result: (1) using a larger num-

ber of building classes present in the project, (2) differentiating spatial densities of buildings within particular classes by analyzing impermeable zones and vegetation indices, (3) introducing geographically weighted regression into the calculation of weights.

LITERATURA

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

Batista F., Poelman S.H. (2016). *Mapping population density in Functional Urban Areas, A method to downscale population statistics to Urban Atlas polygons*, JRC Technical Raport

Bielecka E. (2005). *A Dasymetric population density map of Poland*. Materiały XXII Międzynarodowego Kongresu Kartograficznego, 11–16 lipca La Coruna, Hiszpania

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

Chen, K. (2002). *An approach to linking remotely sensed data and areal census data*. International Journal of Remote Sensing, 23(1), pp. 37–48

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

Kunze C., Hecht R. (2015). *Semantic enrichment of building data with volunteered geographic information to improve mappings of dwelling units and population*, Computers Environment and Urban Systems, vol. 53, DOI: 10.1016/ j.compenvurbsys.2015.04.002

Liu, X. (2004). *Dasymetric mapping with image texture*. ASPRS annual conference proceedings, Denver, Colorado, May

Liu X. (2005). *Uncertainty assessment in dasymmetric mapping*. ICC2005: International Cartographic Conference, La Coruña, España. 9–16 de Julio

Longley P.A., Goodchild M.F., Maguire D.J., Rhind D.W. (2006). *GIS. Teoria i praktyka*. Wydawnictwo Naukowe PWN

Nagle N., N., Buttenfield B.P., Leyk S., Speilman S. (2014). *Dasymetric Modeling and Uncertainty*, Ann Assoc Am Geogr. Jan 1; 104(1): 80–95. doi: 10.1080/00045608.2013.843439

Pirowski T., Bartos K. (2018). *Detailed mapping of the distribution of a city population based on information from the national database on buildings*. Geodetski Vestnik. ISSN 0351-0271. – vol. 62, no. 3, DOI: https://doi.org/10.15292/geodetski-vestnik.2018.03.458-471

Pirowski T., Berka K. (2019). *Analysis of land use and land cover maps suitability for estimating population density of urban areas – exclusion of the equifinality phenomenon on population maps developed from CLC, UA and OBIA data*; in print

Pirowski T., Drzewiecki W. (2012). *Mapa gęstości zaludnienia Krakowa, propozycja metodyki opracowania oraz przykładowe zastosowania*. Roczniki Geomatyki, Tom 10, zeszyt 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

Pirowski T., Timek M. (2018). *Analysis of land use and land cover maps suitability for modeling population density of urban areas – redistribution to new spatial unit based on CLC and UA databases*, Geoinformatica Polonica, ISSN 1642-2511. – vol. 17, s. 53–64, DOI 10.4467/21995923GP.18.004.9162

Pirowski T., Wietrzykowska K., Timek M. (2018). *Analysis of land use and land cover maps suitability for modeling population density of urban areas – redistribution to new spatial unit based on the object classification of RapidEye data*, Geoinformatica Polonica, ISSN 1642–2511. – vol. 17, s. 65–75, DOI 10.4467/21995923GP.18.004.9163

Robinson A.H., Sale R.D., Morrison J.L. Muehrcke P.C. (1998). *Podstawy kartografii*, PWN, Warszawa

Rose A.N., Nagle N.N. (2017). *Validation of spatiodemographic estimates produced through data fusion of small area census records and household microdata*. Computers, Environment and Urban Systems, Vol. 63, s. 38–49

Wang L., Wu C. (2010). *Population estimation using remote sensing and GIS technologies*. International Journal of Remote Sensing. Volume 31 – Issue 21. https://doi.org/10.1080/01431161.2010.496809

Weber E.M., Seaman V.Y., Stewart R.N., Bird T.J., Tatem A.J., McKee J.J., Bhaduri B.L., Moehl J.J., Reith A.E. (2018). *Census-independent population mapping in northern Nigeria*. Remote Sensing of Environment, Volume 204, January, pp. 786–798

Wei C., Blaschke T. (2018). *Pixel-Wise vs. Object-Based Impervious Surface Analysis from Remote Sensing: Correlations with Land Surface Temperature and Population Density*. Urban Sci. 2, 2; doi:10.3390/urbansci2010002

Wu, S.-S., Qiu, X. and Wang, L. (2005). *Population Estimation Methods in GIS and Remote Sensing: a review*. GIScience and Remote sensing, 42 (1):58–74; doi: 10.2747/1548-1603.42.1.80

Zandbergen, P., Ignizio, D. (2010). *Comparison of dasymetric mapping techniques for small-area population estimates*. Cartography and Geographic Information Science, 37 (3):199–214. DOI:10.1559/152304010792194985