Tomasz Pirowski*, Jadwiga Pomietłowska*

Distribution of Krakow's Population by Dasymetric Modeling Method Using Urban Atlas and Publicly Available Statistical Data**

Summary: This publication attempts to adapt the dasymetric method of mapping statistical areas to study population density and carry out the mapping of population distribution of the city of Krakow. Population studies of cities are an important issue, both from a spatial and urban planning point of view and in the preparation of market or crisis studies. Existing maps of the population of Krakow come down to simple cartograms referenced in administrative divisions. Based on the approaches given in literary sources (and modifying them appropriately with regard to the nature and detail of the data already in place), three population distribution maps of Krakow were developed using the binary method and – in two variable options – surface-weight aggregation.

Data on land cover and use from the 2009 Urban Atlas was used as a limiting variable, available under the GMES Urban Audit program. *The statistical data on population that required conversion from administrative units into built-up areas came from 2009 data developed for the 141 urban units of Krakow.*

Arbitrarily adopted numbers of people assigned to particular types of land cover (the first two implemented approaches) generalize the population distribution of the city and/or assume the possibility of a direct transfer of experience from other areas to the analyzed area. Therefore, in the third solution, an optimization of the surface-weight aggregation method was made, obtaining new characteristic proportions for Krakow in the population density for inhabited categories of land cover. When analyzing cartograms and histograms of errors noted within the urban units, the validity of this approach was confirmed. The map generated provides a better reflection of the actual spatial diversity of the population of Krakow than the previously used simple cartograms (or the dasymetric modeling, which did not take into account the specifics of the mapped area).

Keywords: dasymetric map, census data, population density, Urban Atlas

---

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

1.   Introduction

In the era of globalization and the information society, distribution and population density maps are important sources of information. In the basic descriptive dimension, they provide a showcase for population growth, migration rate, and population density. More importantly, they allow for the dynamic use of this information in design studies, spatial analyses, crisis scenarios, predictable phenomena modeling and its associated risks, and even in public policy [9, 11]. In classic cartography, population mapping was usually carried out using the isoline and cartogramic methods, mainly through dot maps or cartograms. With the help of geographic information systems (GIS), options for the processing of capabilities and presentation of spatial data have increased over the past several years. Using access to digital data (including remote sensing data), computer cartography nowadays provides a better representation of population distribution by using the dasymetric method (which was used sporadically in the past). The main difference in this approach is the departure from artificially imposed territorial units such as municipalities or districts when referencing areas directly related to the distribution of the researched phenomenon (in the case of this study, for population in built-up zones or areas). This kind of solution can be implemented at different levels of detail depending on the spatial and statistical data used. As a result, a dasymetric cartogram can be generated that expresses the population density on a map with the number of people present in certain areal units (smaller than cartogram units) that are geometrically closely related to the issue in concern.

In GIS systems with raster (cell) data representation, such areas can serve as single cells in extreme cases. Although their geometrical form is artificial, the relatively high spatial resolution provides a correct representation (better than in classic cartograms) of the range and spatial distribution of a population. This form of data is of particular relevance to modeling and performing an advanced spatial analysis. Depending on the accuracy of the input data, the cell value is the estimated number of people in the cell area or average value for a larger area containing this cell.

The dasymetric method is based on the proposal to separate data between surface units based on additional spatial information [17]. Bielecka [2] defines it as the interpolation of data in reference to a surface area. Eicher and Brewer [7] understand it more broadly as a method used in most forms of surface interpolation, particularly in the smart interpolation method. Langford et al. [10] describes it as a process of transferring data from source units to target units, spatially overlapping each other. In classic cartography, this is considered as one of the mapping methods that allows the presentation of a statistical surface using surface signs [12]. The increase in popularity of the dasymetric method is reflected in the form of implementation of selected algorithms in commonly used GIS class programs [13].

Dasymetric modeling algorithms vary, especially in two key stages of map development: showing limiting variables and variable links as well as in the adoption of redistribution methodology for the statistical data of counted input into newly designated units. A common problem in population mapping (reported by all of the referenced authors using this method) is the correct indication of significant additional data that closely affects population distribution. This entails the need to establish the physiographic conditions for settlement. In the past, classical information of this type was provided by topographic maps. Today, a more-important role is being played by remote sensing sources such as aerial and satellite imaging as well as the products of their processing created during the interpretation or classification process – land use and land cover (LULC) maps [1, 14]. The possibility of a dynamic estimation of the number of people based on the data gathered from mobile device users is also tested [5].

National experiments are based on solutions that differ from the methodical side, primarily by defining the potential places of the presence of a population. In the case of works by Bielecka [2] and Bielecka et al. [4], the basic material was the CORINE Land Cover 2000 product (CLC 2000). In a work by Pirowski and Drzewiecki [11], orthophoto aerial photographs at a scale of 1:26000 were used. The common denominator of these projects is the methodological approach, which allows for the use of open, free remote sensing data as well as the products derived from their processing to detect areas of development. The authors performed their own interpretation of the imagery or used the available LULC maps. They differ in the level of detail, both in spatial terms and the segmentation of the types of built-up areas. Both studies were different, but both based on free statistical data on the population number published on a regular basis.

In the present studies, the primary goal was to use the LULC data from the Urban Atlas project (UA) as a source of data with greater detail than in the CLC product as well as statistical data from urban units characterized by smaller census units than those published by the CSO. Thanks to this set of data, it is possible to prepare a map of the population of urban-industrial areas. To develop a methodology, a complementary experience reported in the two above-mentioned national experiments was used. The research problems faced included choosing a specific area (like a city) and choosing the appropriate option of the dasymetric method. Additional difficulties were linked to the choice of the number and types of LULC categories, including assigning them with appropriate weights associated with the population density on the given type of land.

2. Statistical and Spatial Data

Research experiments were carried out within the city of Krakow. Since 1991, the city has been administratively divided into eighteen districts. Smaller urban

units were also introduced (141 in total) taking into account, among others, the old cadastral divisions belonging to the parish (the division into settlements or historical units derived from cadastral municipalities) that, in essence, meant the division of the city for tax purposes.

Demographic data for individual urban units published on the StatKrak site maintained by the Department of City Strategy and Development, the Municipality of Krakow, and the Department of Analysis and Urban Development Forecasting were used for modeling [15]. This web-service was created as a permanent supportive instrument for city management. Sources of data include Statistical Office publications, the REGON (National Official Register of National Economy Entities) database maintained by the Statistical Office in Krakow, CSO's Regional Data Bank, CSO's Local Data Bank, Reports on the State of the City, and the population census database maintained by the Office of the City of Krakow. The data collected in the service is updated annually and is more accurate than that which is provided by the CSO. They take into account the people actually living in the city; i.e., registered for permanent residence or staying temporarily for a period of more than two months [15].

Data on settlements and their borders came from the Urban Atlas, which is part of the Urban Audit initiative. The program provides LULC maps for the larger urban areas of Europe (whose number of residents exceeds 100,000). The pilot program took place in 1999; four years later, a full-scale program was launched, covering 15 European Union countries. Subsequent editions took place in 2006/2007 and 2009, where 321 cities from European Union countries along with a total of 36 cities from Norway, Turkey, and Switzerland took part. Currently, GIS data with maps and metadata for 282 European cities (including 27 in Poland) has been made available on the GMES Urban Atlas program website. In total, 34 LULC classes were distinguished in the Urban Atlas program, and the nomenclature was taken from the CORINE and MOLAND programs [3, 6, 16].

3. Population Mapping Methodology

The article presents the results of two varieties of dasymetric modeling and population mapping: the binary method, and (in two options) the surface-weight aggregation method.

The binary method belongs to the simplest variety of dasymetric methods. The variable limiting population distribution is adequate area coverage – administrative unit areas are divided into two classes: inhabited and uninhabited (assigning a population number to the former). The advantages of this method are the volume of data it retains in the administrative units and the reduction of subjectivity resulting from the relatively simple choice of uninhabited areas from the list of categories of

LULC. For example, in the Corine Land Cover 2000 project, Bielecka and Ciołkosz [3] distinguished 31 classes of land cover, 18 of which were uninhabited areas.

In the surface-weight aggregation method, one limiting variable is the data on land cover, due to which the area being worked on has to be divided into a selected number of land cover types. The second limiting variable is the a priori adopted percentage of population inhabiting the mentioned types of land cover. For example, such an approach was used by the authors using CLC data [4, 7, 8]. When developing the population map of Warsaw and its surroundings, Bielecka [2] divided the area into four types of land cover. These were high-density built-up, low-density built-up, sparsely populated, and uninhabited areas. On the other hand, Pirowski and Drzewiecki [11] introduced the location of points of residential space as limiting variables, dividing them into four categories: single-family homes, townhouses, low-rise apartment buildings, and high-rise buildings. This method also allows for the introduction of variable connections, thanks to which it is possible to apply the segmentation of the study area and, thus, take into account its local characteristics. Bielecka [2] proposed dividing the city into eight types of municipalities, and Pirowski and Drzewiecki [11] divided the city into districts (seeing elements of the historical conditions of its development in the administrative division).

4. Mapping Krakow's Population

None of methodological approaches presented in the previously reported studies was appropriate to the development of a detailed dasymetric map for the city. Because of this, the workflow described in the following paragraph was applied.

The first step was the selection of limiting variables. In the binary method, inhabited areas were included in the following classes given in the Urban Atlas: high-density built-up, low-density built-up (four source data categories merged together), single buildings, commercial-services, and industrial areas. Out of 19 LULC classes distinguished in the UA for the city of Krakow, as many as 12 areas are unpopulated. These include railroads and related land, rapid transit routes and related areas, other roads and areas related to them, the airport, open pit exploitation and landfills, construction sites, sites without specific use, urban green areas, sport and leisure areas, agricultural land, and areas of semi-natural vegetation, forests, and waters. The surface of each unit was divided into inhabited and uninhabited areas, and the population density was then calculated (independently within each of them) by dividing the statistical data on the population and surface area of the built-up areas. Such an approach works well in units with a single type of building; when buildings vary significantly, it results in a large generalization of information.

By analyzing the created map in detail, it was noted that the commercial and industrial areas (classified in the first approach as 'inhabited') distort the assessment of population distribution (Fig. 1a). These are mainly large compact areas usually

occupying more space than others built-up categories. As they average the density value in the unit in which they appear, they give a misleading impression of large high-density residential complexes (although they may contain a single household at most, which was confirmed by analyzing the terrain based on orthophotomaps). Taking this into account, the calculations were repeated after excluding these sites from the inhabited area classes (Fig. 1b).
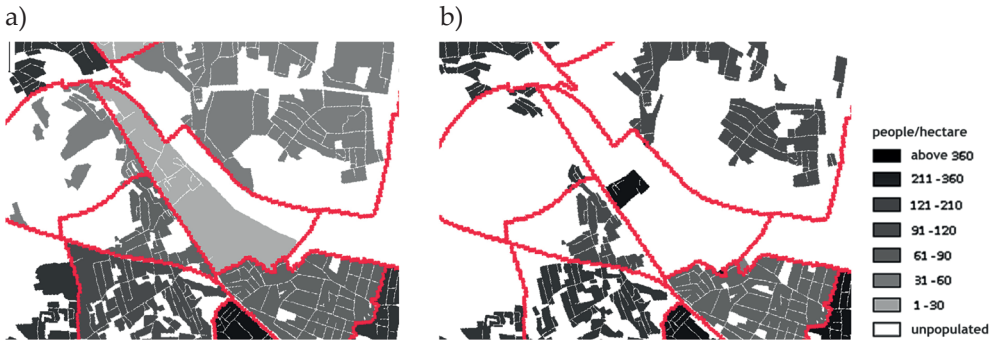
a)   b)

Fig. 1. Binary method: a) including commercial and industrial areas as occupied;
b) excluding them (red – urban area boundaries). Płaszów, Krakow Area

In only one case, mapping the population living there was impossible using the typical industrial urban unit (9 people). In other cases, this LULC class was accompanied by other classes of built-up areas, and no shifting of residents previously assigned to them occurred. This kind of map provides a better representation of population distribution and, thus, was subjected to further analysis.

According to the study [4], data on land cover regarded as limiting variables were aggregated into four types in the surface-weight aggregation method. Categories that arose through a combination of classes stated in the Urban Atlas are as follows: high-density built-up areas (UA: high-density built-up), low-density built-up areas (UA: low-density built-up, highly compacted, medium compacted, low compacted, very low compacted), sparsely populated areas (UA: single objects built), and uninhabited land (UA: the remaining 13 categories). Combining four categories of scattered built-up areas into one was done due to the lack of detailed information on population density (which would enable differentiating these categories correctly). In spite of the data aggregation, information from the UA (1:25,000) retained an advantage over the CLC map (1:1,000,000) due to its higher spatial detail. The division into four categories of land was treated as the first limiting variable. An arbitrarily defined percentage of population inhabiting three of these types of LULC in each unit was assumed as a second variable. Since demographic and urban planning studies for the city of Krakow and its parts do not exist (which would give

the percentage of the population living in the different types of LULC), it was assumed [2] that 70% of people live in high-density built-up, 28% in scattered built-up, and 2% in sparsely populated areas.

In the first stage, the proportions of population density among the three categories of terrain indicated above were calculated using information on the total number of city inhabitants and total areas of particular types of LULC. In the second stage, the appropriate values of population density were allocated within each of the urban areas while maintaining relative variation between the different built-up areas defined above and maintaining population volume in each unit. A map developed in this way was compared with a binary map (Fig. 2).
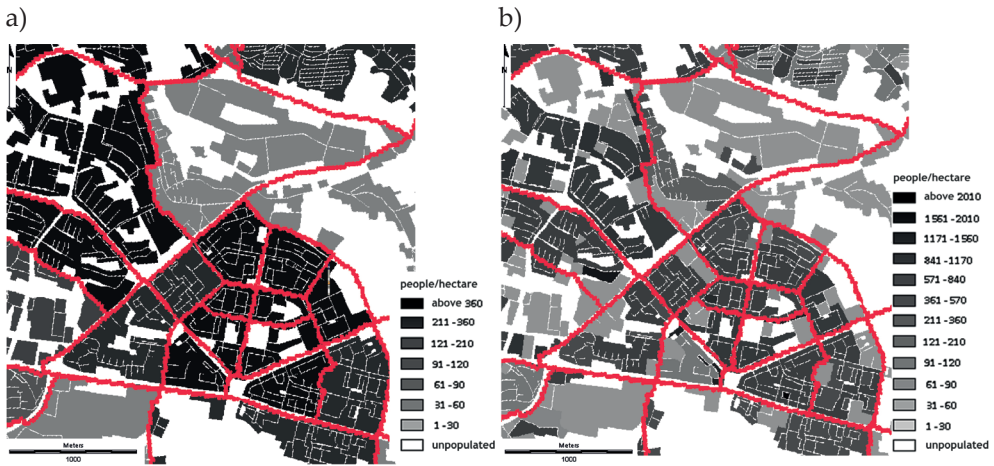

Fig. 2. Comparison method:
a) binary; b) surface-weight aggregation. Center of Nowa Huta, Krakow

An important advantage of this solution is the differentiation of population density within the statistical units in the presence of different types of built-up areas. Such situations can easily be seen in Figure 2b by areas occupied by detached houses (large areas with lower density) and geometrical areas covered with more-densely-built apartment buildings. The highest values are recorded in small, tight areas occupied by high-rise apartment buildings (skyscrapers). The main disadvantage of the above-described model is the rigid proportion of the population imposed depending on the LULC categories. The adopted values used in Warsaw and the Mazowieckie voivodeship [2, 4] may not be reflected in other areas of the country (particularly in Krakow). Therefore, an attempt was made in the third experiment to statistically estimate the weights (ratio of density) of the different types of land cover. For this purpose, the data on the density of population in each of the four separate categories was established in the first stage based on the aggregated number of Krakow inhabitants and the proportion assumed as before (70-28-2).
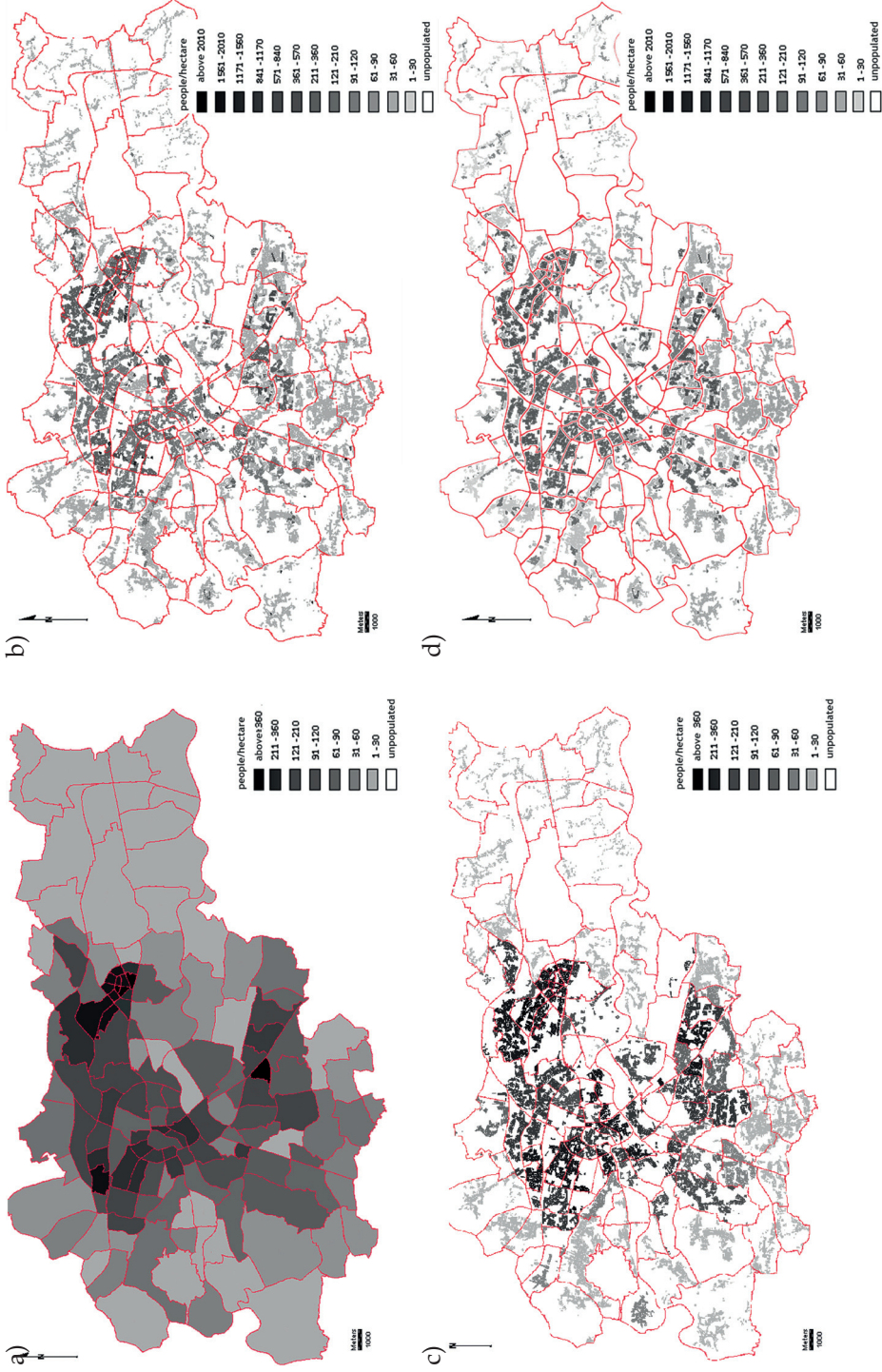
Fig. 3. Mapping population of city of Krakow by the following methods:
a) cartogram; b) binary; c) surface-weight aggregation; d) surface-weight aggregation with optimization

With this approach, the data retention volume only occurs for all of Krakow in general, whereas the differences between the actual value (from the statistical data) and the value calculated as a result of the dasymetric modeling has been noted for each individual urban unit. In the next step, an iterative approach enabled us to search for the optimal ratio (the number of residents allocated to each of the four categories) by minimizing the occurring differences noted for the 141 urban units using the Least Squares Method. This corresponds to the optimizing solutions used in works [4, 8]. In the latter case, a voivodeship (province) would constitute the equivalent of a city, while a municipality would constitute an urban unit.

As a result, a new percentage of proportions for Krakow was obtained that varied significantly from the initial 70-28-2 figure (correspondingly amounting to 92.9-6.9-0.2). In the next step, the experience reported in [11] was used. After stabilizing the parameters by using an iterative process, the numerical results obtained for each unit in relation to the statistical data were compared. From the ratio of the two values, a factor correcting for local differences was obtained, and the iteratively calculated values of population density were modified by this value. Thanks to this effort, data consistency at the level of individual urban areas was achieved while still retaining the proportions calculated for different types of built-up areas. The final score for all of the tested methods and the classical cartogram are illustrated in Figure 3.

5.　Discussion of Results

To confirm the legitimacy of this iterative method as a way of obtaining the correct proportion of population density for different LULC types, a simulation of error distribution for the first two methods was made (i.e., binary and surface-weight aggregation without optimization). This was achieved using the same assumption as used in the third method; i.e., the data on the number of inhabitants for the whole city was aggregated while the surface of the built-up area (limiting variables) was calculated globally for the area of analysis. Next, the average population density values falling in the different options for Krakow were calculated. In the next step, the number of inhabitants in each approach noted within the individual urban units was calculated, and those values were referenced with the actual ones derived from statistical data. The distribution of errors in the form of cartograms and histograms for all three options is shown in Figure 4.

When analyzing the cartograms, it is easy to notice that the third approach is characterized by the least number of mistakes, and their distribution is normal. This observation is confirmed by the global values of absolute percentage errors for the individual urban units: 191.7% (binary method), 119.8% (surface-weight aggregate), and 67.2% (with optimization).
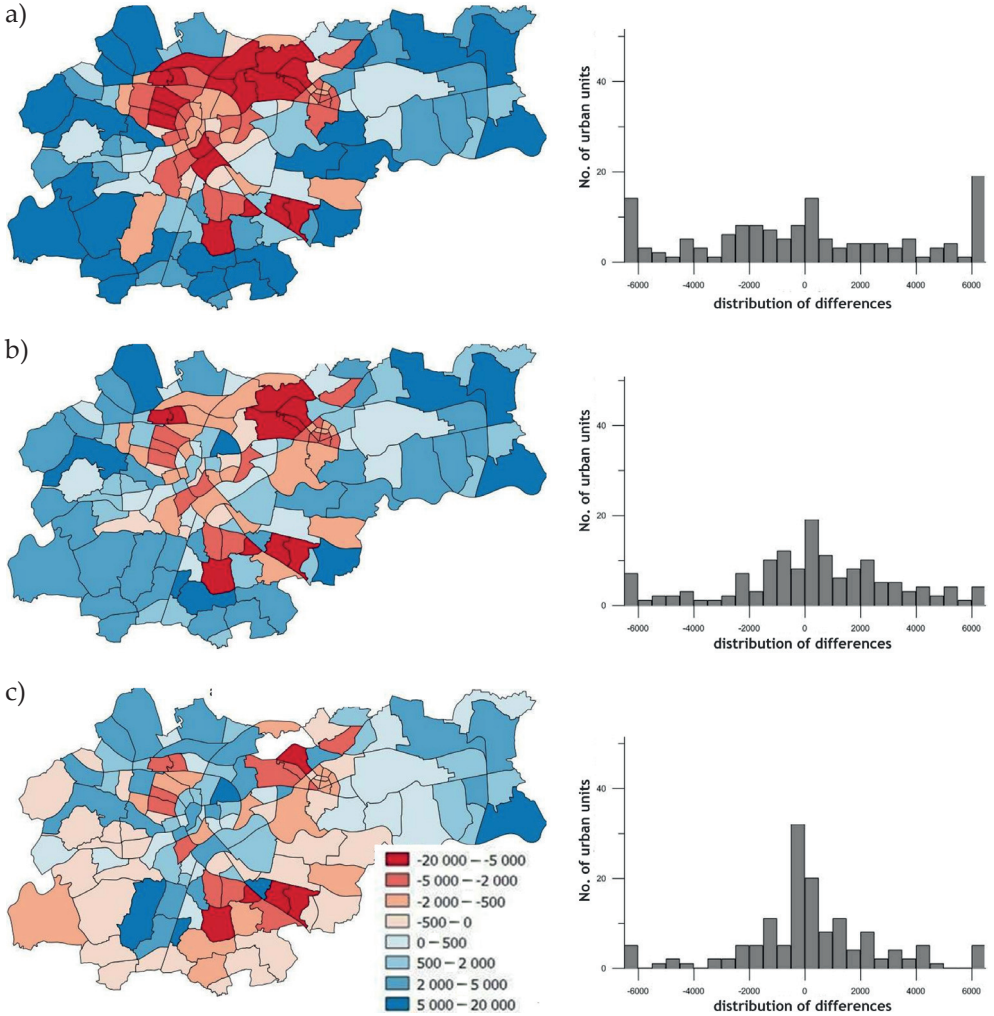
a)

b)

c)

Fig. 4. Cartograms and distribution of errors for urban units for the following methods:
a) binary; b) surface-weight aggregation; c) with optimization

6.   Conclusions

The use of the data from the Urban Atlas project together with the detailed statistical data facilitated the creation of a detailed map of the current population distribution for the city of Krakow. The data used and the manner of its processing allow for the relatively easy and systematic updating of information. This kind of product constitutes a valuable layer of information in the GIS database, particularly important in the case of spatial analyses, decision support, simulations, and conflict and crisis resolutions in the city [11].

From the tested methods, the surface-weight aggregation with the optimization method should be regarded as the best. Striving to achieve even better results would require the introduction of additional variable connections that allow for the segmentation of urban units (within such, designating each to their individual group in order to appropriately weigh the individual types of built-up areas). Such a segmentation, however, is difficult because it requires capturing the differentiation in the distribution of the population that can be associated with a variety of conditions of the terrain (historical, sociological, and physiographical). These tests carried out under domestic conditions have produced limited results [2, 11]. Another way to improve the results is the differentiation of population density for all available classes associated with built-up areas that are mapped in the Urban Atlas project.

However, to be able to carry out such detailed experiments, we should exclude the so-called equifinality phenomenon, where the global result (map) seems to be correct, but detailed analyses indicate that it is, in fact, not. This danger occurs when conducting an optimization of the surface-weight method of aggregation and increases in the case of the aforementioned possibility of introducing more classes of built-up areas or spatial segmentation of the conducted modeling. Given these objections, final verification of the results should rely on detailed reference data describing the population distribution within the urban units. Therefore, under continuation of research in this area, there are plans to obtain demographic data for other types of spatial units and to develop a detailed population map using interpretive methods with support from BDOT. Such reference material will allow estimations of the actual precision of the products derived based on power mean data such as Urban Atlas or CORINE Land Cover and should facilitate the proper execution of the segmentation and optimization processes. It may be interesting to compare the possibility of developing population maps through multispectral image classification using various classifiers, including decision-making trees, neural networks, and object-oriented classification.

References

[1] Azar D., Engstrom R., Graesser J., Comenetz J.: *Generation of fine-scale population layers using multi-resolution satellite imagery and geospatial data*. Remote Sensing of Environment, vol. 130, 2013, pp. 219–232, DOI: 10.1016/j.rse.2012.11.022.

[2] Bielecka E.: *A dasymetric population density map of Poland*. ICC2005 – International Cartographic Conference, Corunna, 2005, [on-line:] http://www.cartesia.org/geodoc/icc2005/pdf/oral/TEMA5/Session%209/ELZBIETA%20BIELECKA.pdf [access: 28.09.2017].

[3] Bielecka E., Ciołkosz A.: *Pokrycie terenu w Polsce. Bazy danych CORINE Land Cover*. Biblioteka Monitoringu Środowiska, Państwowa Inspekcja Ochrony Środowiska Warszawa 2005.

[4] Bielecka E., Kuczyk A., Witkowska E.: *Modelowanie powierzchni statystycznej przedstawiającej gęstość zaludnienia w Polsce przy pomocy metody dozymetrycznej*. Roczniki Geomatyki, t. 3, z. 2, 2005, pp. 9–16.

[5] Deville P., Linard C., Martin S., Gilbert M., Stevens F.R., Gaughan A.E., Blondel V.D.. Tatem A.J.: *Dynamic population mapping using mobile phone data*. Proceedings of the National Academy of Sciences of the United States of America, vol. 111, no. 45, 2014, pp. 15888–15893.

[6] Drzewiecki W.: *Monitoring zmian pokrycia i użytkowania terenu na podstawie wieloczasowych obrazów teledetekcyjnych*. Roczniki Geomatyki, t. 6, z. 3, 2008, pp. 131–142.

[7] Eicher C.L., Brewer C.A.: *Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation*. Cartography and Geographic Information Science, vol. 28, no. 2, 2001, pp. 125–138.

[8] Gallego F.J., Peedell S: *Using CORINE Land Cover to map population density.* [in:] *Towards agri-environmental indicators: Integrating Statistical and Administrative Data with Land Cover Information*, EEA Topic Report, 6, European Environment Agency, Copenhagen 2001, pp. 94–105.

[9] Gregory I.N.: *The accuracy of areal interpolation techniques: standardizing 19th and 20th century census data to allow long-term comparisons*. Computers, Environment and Urban Systems, vol. 26, 2002, pp. 292–314.

[10] Langford M., Unwin D.: *Generating and mapping population density surfaces within a geographical information system*. The Cartographic Journal, vol. 31, issue 1, 1994, pp. 21–26.

[11] Pirowski T., Drzewiecki W.: *Mapa gęstości zaludnienia Krakowa, propozycja metodyki opracowania oraz przykładowe zastosowania*. Roczniki Geomatyki, t. 10, z. 3, 2012, pp. 95–106.

[12] Robinson A.H., Sale R.D., Morrison J.L. Muehrcke P.C.: *Podstawy kartografii*. PWN, Warszawa 1998.

[13] Sleeter R., Gould M.L.: *Geographic information system software to remodel population data using dasymetric mapping methods*. U.S. Geological Survey Techniques and Methods, 11-C2, U.S. Geological Survey, 2007.

[14] Ural S., Hussain E., Shan J.: *Building population mapping with aerial imagery and GIS data*. International Journal of Applied Earth Observation and Geoinformation, vol. 13, issue 6, 2011, pp. 841–852.

[15] Web sites StatKrak, Wydział Strategii i Rozwoju Miasta Urząd Miasta Krakowa, Referat Analiz i Prognoz Rozwoju Miasta, www.msip2.um.krakow.pl/statkrak/ [access: 28.10.2016].

[16] Web sites Copernicus – Land Monitoring Service, http://land.copernicus.eu /local/urban-atlas [access: 30.10.2016].

[17] Wright J.K.: *A Method of Mapping Densities of population with Cape Code as an Example*. Geographic Review, vol. 26, no. 1, 1936, pp. 103–110.

Modelowanie rozmieszczenia ludności Krakowa metodą dazymetryczną z wykorzystaniem Urban Atlas i ogólnodostępnych danych statystycznych

Streszczenie: Tematem publikacji jest zaadaptowanie dazymetrycznej metody mapowania powierzchni statystycznej do badania gęstości zaludnienia oraz skartowania rozmieszczenia ludności Krakowa. Opracowania ludnościowe dla miast są ważne w aspekcie planowania przestrzennego i urbanistycznego oraz sporządzania analiz rynkowych bądź kryzysowych. Dotychczasowe mapy ludności Krakowa sprowadzają się do kartogramów prostych odniesionych do jednostek podziału administracyjnego. Bazując na podejściach podanych w źródłach literaturowych oraz odpowiednio je modyfikując z uwzględnieniem charakteru i szczegółowości posiadanych danych, opracowano trzy mapy rozmieszczenia ludności Krakowa: metodą binarną oraz, w dwóch wariantach, metodę powierzchniowo-wagowej agregacji.

Jako zmienne ograniczające wykorzystano dane o pokryciu i użytkowaniu terenu Urban Atlas z 2009 roku, dostępne w ramach programu GMES Urban Audit. Dane statystyczne o ludności poddane przeliczeniu pochodziły z 2009 roku i były opracowane dla 141 jednostek urbanistycznych Krakowa.

Przyjęte arbitralnie wartości dotyczące liczby ludności, przypisane poszczególnym typom pokrycia terenu (dwa pierwsze realizowane podejścia), upraszczają przeliczenie ludności miasta, gdyż zakładają możliwość skutecznego przeniesienia wprost doświadczeń z innych obszarów testowych. W rozwiązaniu trzecim zoptymalizowano więc metodę powierzchniowo-wagowej agregacji, uzyskując nowe proporcje w zagęszczeniu ludności dla wydzielonych charakterystycznych kategorii pokrycia terenu na obszarze Krakowa. Analizując dla poszczególnych trzech testowanych rozwiązań kartogramy błędów oraz ich rozkład na histogramach, potwierdzono zasadność takiego podejścia. Uzyskana mapa stanowi lepsze odzwierciedlenie rzeczywistego zróżnicowania przestrzennego ludności Krakowa niż stosowane dotychczas kartogramy proste czy modelowanie dazymetryczne nieuwzględniające specyfiki kartowanego obszaru.

Słowa kluczowe: mapa dazymetryczna, spis ludności, gęstość zaludnienia, Urban Atlas