A DASYMETRIC POPULATION DENSITY MAP OF POLAND

Elzbieta Bielecka
Institute of Geodesy and Cartography
ul. Modzelewskiego 27, 02-679 Warsaw, Poland
tel. (48 22) 32 91 984 e-mail: elzbieta.bielecka@igik.edu.pl

Abstract

This paper presents research results on the dasymetric mapping method, the methodology adopted and a *Dasymetric Population Density Map of the Northeastern Part of Poland*. After studying certain published dasymetric mapping methods, a decision was made to start from the binary dasymetric procedure as described by Langdorf and Unwin (1994) in order to eliminate unpopulated regions. The CORINE Land Cover (CLC) database was utilized as ancillary data because the land cover pattern best reflects population density and provides useful georeferenced information for disaggregating census data. The first issued addressed was "how is population distributed among CLC classes in Poland?" Special attention was made to estimate the coefficients for the population ratio in land cover classes. Several different approaches were tested and eventually modified areal weighting was adopted. The dasymetric mapping method is based on the assumption that the ratio between the population density of two land cover categories is the same for any given commune.

INTRODUCTION

Population mapping, in general, has two purposes: firstly, to cartographically portray the extent and density of population across an area of interest, and secondly, to derive a quantitative estimation of population density for use in subsequent spatial analytical modeling tasks. A cartographic portrayal of population traditionally has the form of a choropleth map. This kind of mapping is very simple, but despite its simplicity, choropleths maps have limited utility for detailed spatial analysis of population data, especially where human populations are concentrated in relatively small numbers of villages, towns and cities. Langford and Unwin (1994) noticed that three significant problems associated with choropleth maps must be addressed. First, the enumeration areas are arbitrary determined for data collection convenience, subsequently creating large polygons for regions where population density is low and small polygons for regions where there is high population density. Second, the enumeration areas act as low-pass filters, generalising the highs and low values, and removing the spatial specialization of the mapped phenomena. As boundaries are arbitrary, they will not convey any significance to the data, therefore, attention is paid towards the size of the polygons rather than on the data conveyed. Lastly, enumeration areas do not differentiate physiographic entities such as lakes, rivers, forests, beaches, rocks and agricultural areas. Consequently, the portrayal is incorrect as these areas are populated.

Openshaw (1983) described these limitations as the modifiable areal unit problem (MAPU), defined as a situation in which modifying the boundaries and scale of data aggregation significantly affects the result of spatial data analysis. Choropleth maps of population by administrative units give the impression that population is distributed homogeneously throughout each areal unit, even when portions of the region are, in actuality, uninhabited. Openshaw (1983) states that it is often unclear whether the results of statistical data analysis indicate some reality about the individuals living in that region or are rather strictly a function of the particular areal unit used in the analysis.

One way to avoid this limitation is to transform the administrative units into smaller and more relevant map units through the process known as dasymetric mapping. The dasymetric technique maps a quantitative variable according to boundaries derived from the character of the data distribution. It is a form of areal interpolation that uses ancillary data to transform population data from one set of spatial units to another.

HISTORICAL OUTLINE

The method was firstly utilized and termed *dasymetric* by the Russian cartographer Tian-Shansky, who developed the multi-sheet population density map of European Russia, scale 1:420 000, published in the 1920s (Preobrazenski 1954). As populated territory, Tian-Shansky mapped areas within the equidistant of one verst (1067 m) from built-up terrains. The first cartographer who popularised dasymetric mapping was Wright (1936). He set forth a new method of presenting population density based upon the division of a given administrative unit into smaller areas complying with

different types of geographical environments. Wright derived population density values subjectively, which is considered as the main disadvantage of the dasymetric method.

Although the technique was first described in the 1930s, dasymetric mapping lacks a standardized methodology. Recent advancements in geographic information systems, as well as increased availability of digital datasets, have revitalized the concept of dasymetric mapping. One difficulty has been finding ancillary data that corresponds well with population distribution. In the most often quoted publication concerning the dasymetric method (Wright 1936), topographic maps were used as ancillary data. In cartographic literature also, a few other sources are mentioned: satellite images and land use maps (Langford and Unwin 1994; Eicher and Brewer 2001; Mennis 2002, 2003). These sources allow an analysis of the suitability of the natural environment for settlement.

Night-time city lights imagery, for example, has been shown to demonstrate a reasonable correlation with population (Sutton 2003). To achieve high spatial resolution as well as more reliable characterization of ancillary data, many researchers turn to land-cover maps, such as those produced from satellite imagery. Working with various land cover classes the key question becomes how to distribute the population among those classes. The basic technique is known as binary classification, wherein all classes are designated as either inhabitable or non- inhabitable, and the population is distributed by areal weighting into the inhabitable areas of each enumeration district. This simple method can improve accuracy of traditional choropleth maps.

Langford and colleagues (1991, 1994) described a dasymetric mapping procedure using land cover data derived from Landsat Thematic Mapper (TM) multispectral imagery to build a series of predictive models regressing population density on land use. These models were then used to redistribute U.K. census-ward population data for Leicestershire to a 1 km resolution raster surface. The models do not preserve the *pycnophylactic* property defined by Tobler (1979) as when the summation of population data to the original set of areal units is preserved in the transformation to a new set of areal units. They underestimate population for suburban and rural areas and overestimate it for urban areas.

Eicher and Brewer (2001) present three methods (binary, three-class and limiting variable) for dasymetric mapping of population density. They also evaluated map accuracy using both statistical analyses and visual presentation of errors. A repeated-measure analysis of variance showed that the traditional limiting variable method had significantly lower error than the other four methods. Furthermore, polygon methods had lower errors than grid-based counterparts, although the difference was not statistically significant.

Mennis (2003) proposes computing density values for each land cover classes based upon the density of all census block groups that lie entirely within that class. One difficulty that this method faces is that, it is difficult to harmonize spatial resolution of these two sources of data: census data and land cover classification. While spatial resolution of the remotely sensed land cover data has improved it is difficult to find even one block group that lies entirely within one land cover type.

A modified version of areal weighting interpolation was set forth by Gallego and Peedel (2001) to map population density in Europe. Gallego and Peedel disaggregated population data with the help of the CORINE Land Cover database assuming that the ratio between the population densities of two land cover classes is the same for any commune in the European Community in the 1990s (15 countries). They obtained rather robust results and concluded that results can be improved if Europe is divided into regional grouping communes with similar characteristics.

Although dasymetric mapping has existed for well over a century, the method for producing these maps has not been thoroughly examined. Subjectivity, diversity of criteria for selecting territorial units as well as the fact that the cartographer is forced to estimate density for these units, are often considered as the drawbacks of this method. Conversely, dasymetric mapping enables straying from the administrative boundaries, which do not usually reflect the patterns of population density of the given area.

This paper describes the results of GIS application that uses ArcGIS to map population density more precisely in the Podlasie region of Poland. This research presents a dasymetric mapping technique based on the areal weighting and regression model. This approach takes into account land cover information and the typology of administrative units in order to provide a more realistic view of population distribution in the Podlasie region.

DASYMETRIC MAPPING METHODS AND RESULTS

The implemented dasymetric mapping method utilizes population data (sources zones) as input and land cover data (ancillary zones) and redistributes the population data to a set of target zones formed from the intersection of the source

and the ancillary zones. Two methods were tested: the so-called binary method and a method based on the aeral weighting and regression model.

Study area

At present about 61.9% of Poland's population lives in towns, with 38.1% residing in rural areas. The average population density in urban areas amounts to 1107 inhabitants per km^2 in contrast to 50 people per km^2 recorded in rural areas. The settlement network, consisting of 864 towns and more than 40 000 rural settlements units, varies geographically. One of the regions where populations are distributed unevenly is the northeastern part of Poland with a significant acreage of lakes, forests and protected areas limiting inhabited areas. For these reasons the Podlasie Voivodship was chosen as the study area. It is located in the northeast part of Poland (Fig.1) and has an area constituting 6.4 % of the total area of the country (20 180 km^2), which makes it the sixth largest voivodship in Poland.
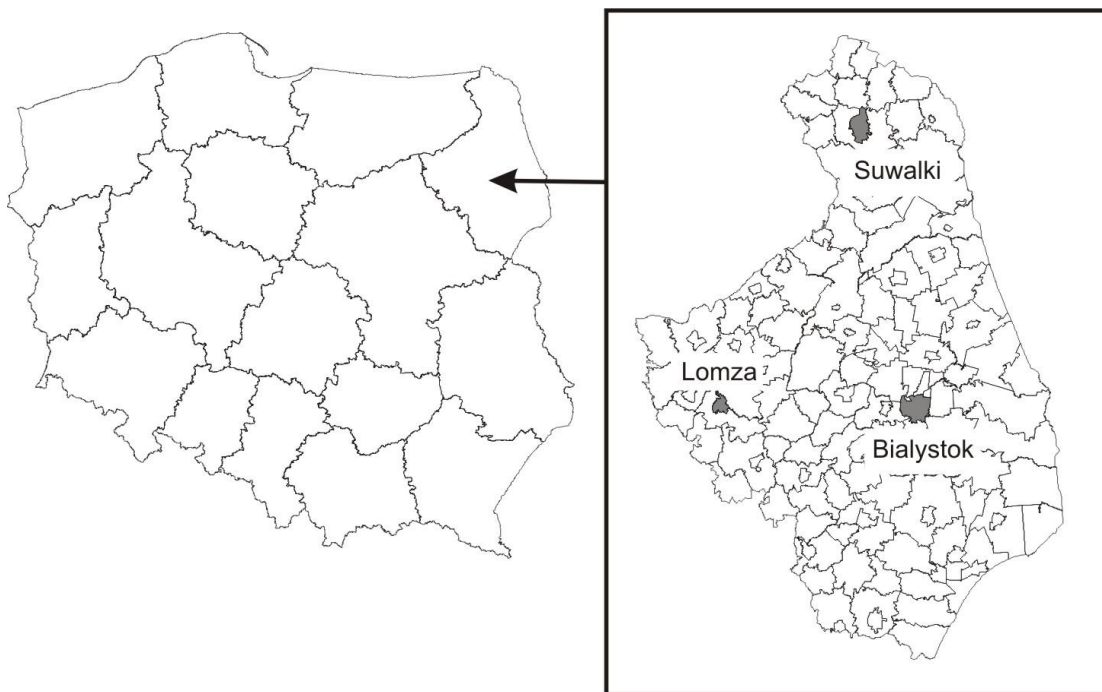

Fig. 1. Location of the study area.

The Podlaskie voivodship is divided into 17 counties (poviats), and 118 communes (13 urban, 23 urban-rural and 82 rural). The largest town in the region is Bialystok with 300 thousand inhabitants. The other major towns are: Lomza, around 66 thousand inhabitants, and Suwalki, around 70 thousand inhabitants. Podlaskie is inhabited by 1 224 000 people, of which 57.7% live in urban areas. Rural settlement is dispersed and fragmented. Over 120 thousand farms are registered, the majority of which are small farms of 1-5 ha and medium-sized farms of 5 – 10 ha. Arable land constitutes around 60 % of the total area of the region – most of which is plough land (around 40 %), forests occupy over 30%, water and wetlands – nearly 3%.

Around 30% of the area of the voivodship is under protection where settlement is limited or even prohibited. There are four national parks, three landscape parks, 88 nature reserves and 15 protected landscape areas. The voivodship constitutes a part of the ecologically clean area known as the 'the Green Lungs' of Poland.

Data

Population data was obtained from the 2002 Poland Census. The mapping unit for population density was communes, the lowest level of administrative division in Poland.

In the Podlaskie voivodship, area communes and population density vary significantly from urban to rural communes (fig. 2). Urban communes are relatively small and of homogeneous population density, while rural communes are typically much larger, and have a much more heterogeneous population distribution. Dasymetric mapping may, therefore, be used to generate a surface model that provides a more accurate representation of population in this region.
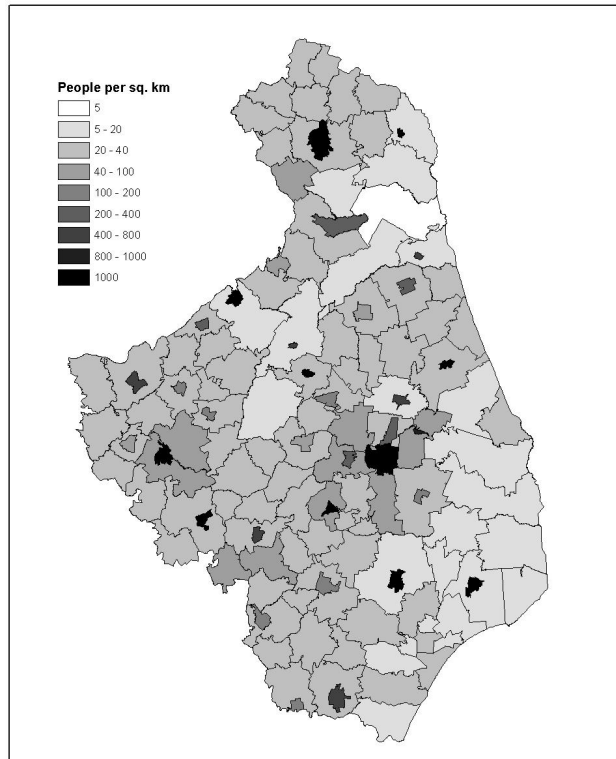
Fig. 2. Choropleth population density map (2002 Census). Podlaskie voivodship.

The ancillary land cover data used was acquired during the CORINE Land Cover project in the 1990s and its updating in 2000 (Perdigao and Annoni 1997). CORINE Land Cover (CLC) is a geographic land cover/land use database encompassing most of the countries of Europe.

Land cover data, stored in the CLC database, was acquired based on the visual interpretation of satellite images (LANDSAT TM and MSS, SPOT). Ancillary data (aerial photographs, topographic or vegetation maps, statistics, local knowledge) was used to refine the interpretation and the assignment of the territory into the categories of the CORINE Land Cover nomenclature. CLC describes land cover according to a nomenclature of 44 classes organised hierarchically in three levels. The first level (5 classes) corresponds to the main categories of the land cover/land use (artificial areas, agricultural land, forests and semi-natural areas, wetlands, water surfaces). The second level (15 classes) covers physical and physiognomic entities at a higher level of detail (urban zones, forests, lakes, etc), and the third level is composed of 44 classes (31 in Poland). Figure 3 shows the CORINE Land Cover classes in the Podlaskie voivodship. The smallest areas mapped correspond to 25 hectares. Linear features less than 100 m in width are not considered. The scale of the output product was fixed at 1:100 000. Thus, the location precision of the CLC database is 100 m.

The CORINE Land Cover database was originally made available in a vector format. This categorical polygon dataset exhaustively partitions the territory of Poland into 31 land cover categories. The CLC database cannot indicate population density directly, although it is possible to estimate the positive relationship, which occurs between population density and the degree of urban development.

Binary method

The binary dasymetric procedure, described by Langford and Unwin (1994), as well as Eicher and Brewer (2001), was applied. Areas inhabited by people were identified by selecting sixteen land cover classes out of 31 appearing in Poland. These included, among others: water courses and water bodies, wetland and peat bogs, beaches, bare rocks, semi-natural vegetation, road and rail networks, ports, airports, green urban and sport areas, industrial and open mining areas. Population density in the binary dasymetric method is calculated by dividing the number of inhabitants in the commune by the area of the inhabited part. The results are shown below in figure 4.
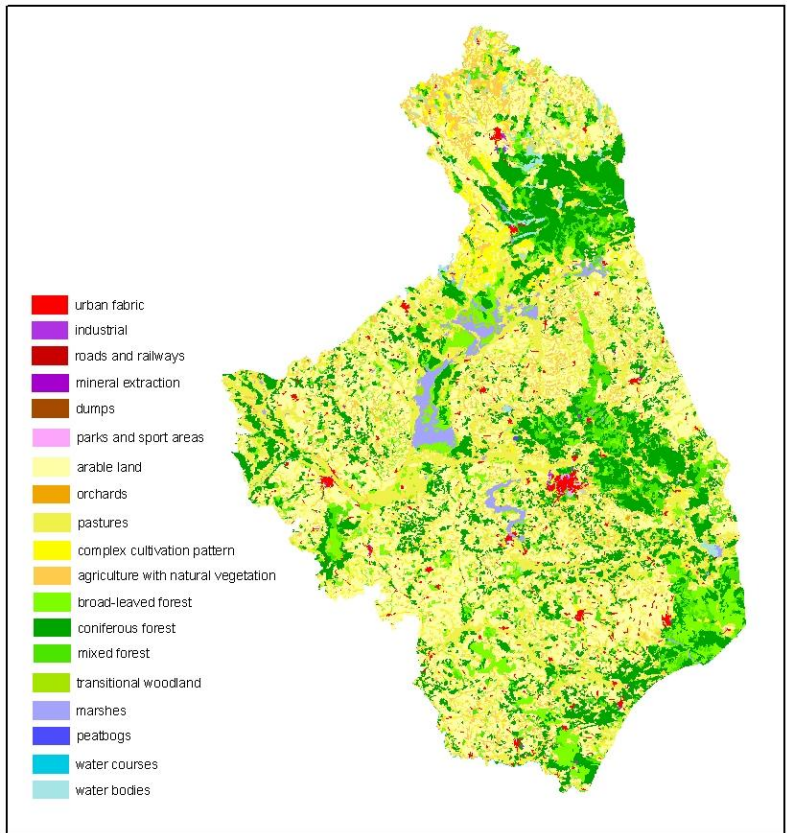
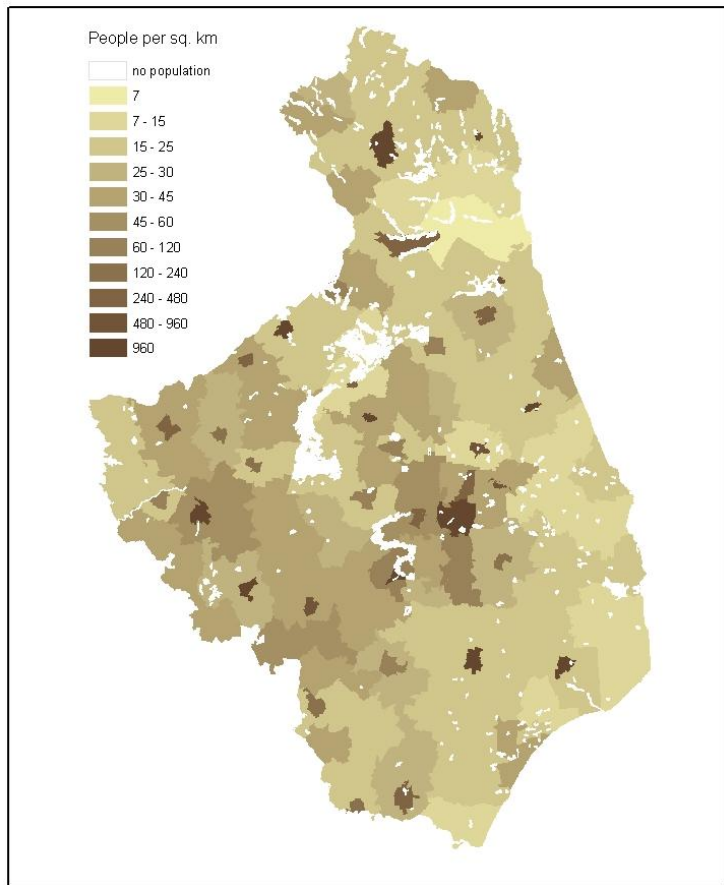Fig. 3. CORINE Land Cover. Podlaskie Voivodship.

Fig. 4. Dasymetric population density map. Binary Dasymetric Method.

The primary advantage of the binary method, emphasized by all authors, is its simplicity. It is only necessary to reclassify land cover data into two classes: inhabitable and uninhabitable. The disadvantage of this method, however, is the subjectivity of reclassification. The decision as to which areas are inhabitable and which are uninhabitable depends on land cover information (the smallest mapping unit and number of classes) and on the information relating to the mapped area.

Modifying Areal Weighting Method

Under this approach, it was assumed, after Gallego and Peedell (2001), that the ratio between the population density of two land cover categories is the same for any given commune. Land cover categories were obtained after grouping CORINE Land Cover classes into six categories, listed below in Table 1. This reclassification was made to decrease the number of land cover classes by grouping together those having similar population density.

Population data was redistributed based on the combination of areal weighting and assumed relative densities of land cover categories. Hence the densities were computed according to the following equation:

$$D_c = U_c \frac{P}{\sum_c A_c U_c}$$

where: D_c : density of population for the land cover category in the voivodship,
   U_c : weighting coefficient (diversifying population according to land cover category in the voivodship),
   P : population in the voivodship,
   A_c : area of the land cover category in the voivodship.
The disaggregation of population data was carried out with an initial set of coefficients computed for Europe by Gallego and Peedell (2001) and is listed in Table 1as follows.

Table 1. Grouped CORINE Land Cover classes and initial coefficients.

Land cover category	Number of CORINE Land Cover classes	Initial coefficient U_c
1 – urban areas	2	176
2 – industrial and commercial units	1	10
3 – agriculture areas	4	3
4 - complex cultivation pattern (scattered settlement)	1	6
5 – forests and semi-natural vegetation	5	0.5
6 – uninhabited areas	18	0

Although this method meets the requirements for preserving the *pycnophylactic* property in the voivodship (Tobler 1979), discrepancies between attributed population and statistical data at lower level (commune) were very large and unacceptable.

Analysing the structure of land cover and population density in all communes in the region it was obvious that stratification of communes is necessary. A simple regression model was used to explore the relationship between land cover classes and population in each commune in the voivodship. The results of the communes fall into six regions: (1) cities, (2) a urban section in a rural-urban commune, (3) a rural section in a rural-urban commune with the urban land cover category, (4) a rural section in a rural-urban commune without the urban land cover category, (4) rural communes with the urban land cover category, and (6) rural communes without the urban land cover category.

For each region a set of coefficients weighting population to the land cover category was computed in an iterative way until the difference indicator δ become stable. The difference indicator δ_r for the region r was computed as the sum of the absolute values of the differences between the population attributed to each commune P_m^0 and the known population P_m (statistical data):

$$\delta_r = \sum_{m \in r} \left| P_m^o - P_m \right|$$

In order to adjust the correlation weighting coefficients, within a region, a correlation ρ_{cr} between the ratio of the population attributed to each commune P_m^0 and the known population P_m and the ratio between the area of land cover category A_{cm} in the region and the total area A_m of the region was computed.

$$\rho_{cr} = corr\left(\frac{P_m^o}{P_m}, \frac{A_{cm}}{A_m}\right)$$

This correlation was thereafter used to compute the new value of the U_c according to equation:

$$U_{cr}^{'} = U_c\left(1 - \frac{\rho_{cr} \times \delta_r}{2 \times P_r}\right)$$

Thirty iterations were necessary to obtain satisfying results. The final coefficients are listed below in Table 2.

Table 2. Weighting coefficients U_c after thirty iterations

	Regions	Urban	Industrial and commercial	Agriculture	Complex cultivation	Forest
1	Cities	18.3668	2.7116	0.5486	5.5919	0.0548
2	Urban section in a rural-urban commune	17.7929	2.1829	0.1035	12.9384	0.3427
3	Rural section in a rural-urban commune with urban land cover category	7.2842	0.6166	0.6931	4.3405	0.2946
5	Rural communes with urban land cover category	8.0743		0.7163	2.2202	0.2084
4	Rural section in a rural-urban commune without urban land cover category			0.8735	2.5411	0.0832
6	Rural communes without urban land cover category			0.8737	2.5411	0.0892

Figure 5 clearly shows that the weighting coefficients U_c for urban and complex cultivation land cover categories fluctuate significantly for urban to rural communes.
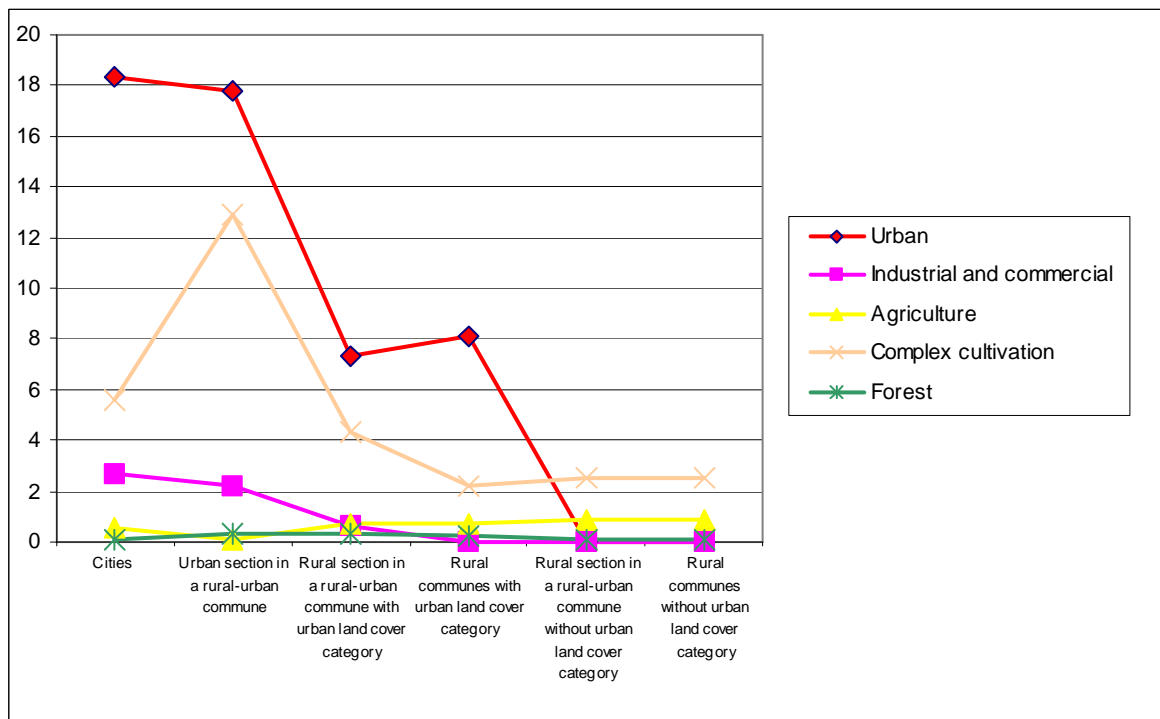

Fig.5. Weighting coefficients U_c for land cover categories in each region.

The results of disaggregation of the population data in the Podlaskie voivodship with these sets of coefficients are illustrated in figure 6. To better demonstrate the population distribution in the region, part of the study area has been enlarged.

Quality assessment of dasymetric methods

A variety of the methods for assessing the quality of results of GIS analyses can be used, which include, among others: root mean squared (RMS) error, coefficient of variations, mean percent error and relative error. Relative errors were computed in this analysis for evaluating the modified areal weighting method. These errors were then attributed to the communes and are shown in figure 7.

The results of a comparison between attributed population and population data known from statistical measurements indicate that the population value attributed to most of the communes is approximately in agreement with the statistical data. The coefficients seem to be too high for urban communes and too low for rural areas.
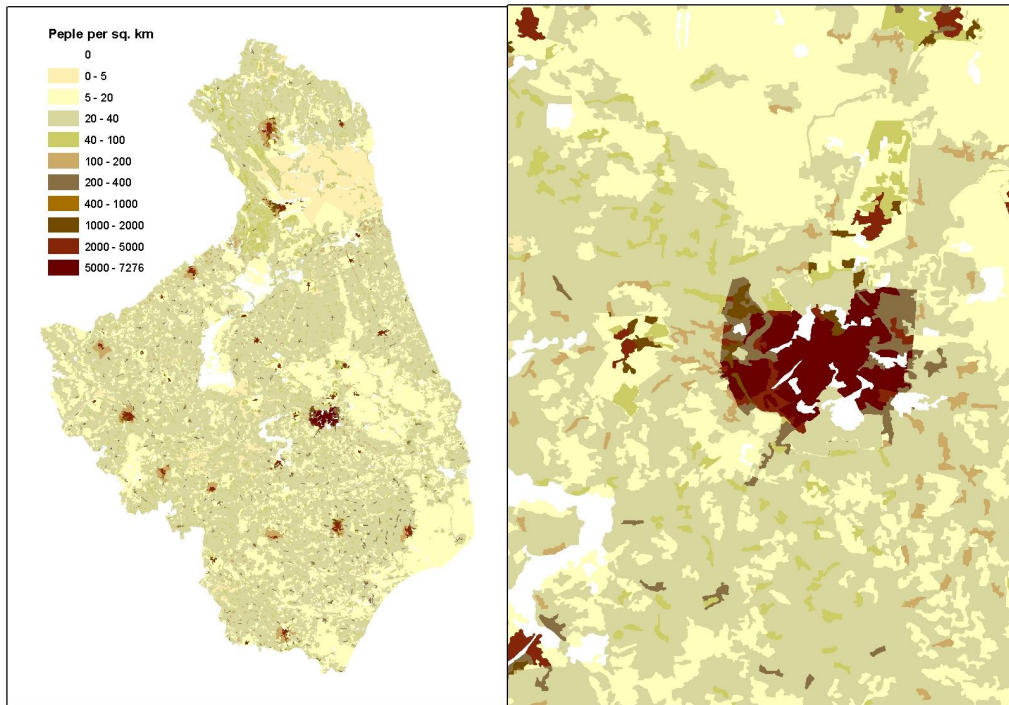

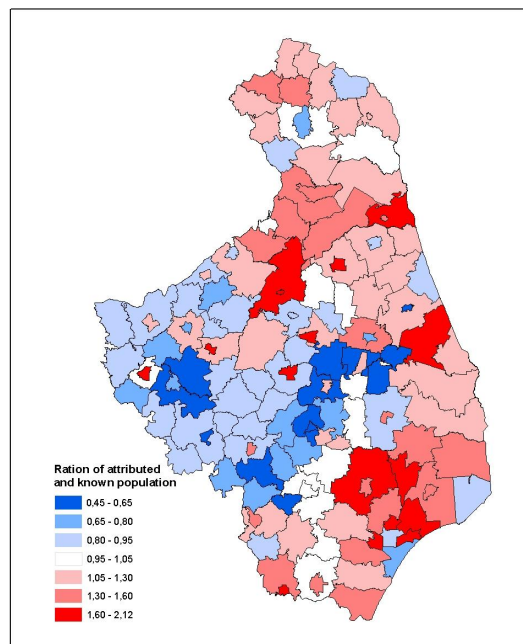Fig.6. Dasymetric population density map. Modifying Areal Weighting Method.

Fig. 7. Relative errors of population in communes.

CONCLUSION

This paper demonstrates the use of satellite derived ancillary land cover data to map population densities using dasymetric mapping. The two dasymetric methods presented revealed the inter-regional variation in population density more realistically, in particular, among urban and rural areas.

The binary method, the simplest, is easy to implement in GIS and gives a better view of population distribution over a given area than conventional choropleth maps. The only drawback is the delimitation of uninhabited areas.

The modifying areal weighting method presents a new way of calculating weighting coefficients. This is based on the method proposed by Gallego and Peedell (2001), but is adapted to Polish conditions by grouping land cover classes, stratifying communes and computing coefficients.

Results of this research will be currently extended to all voivodships in Poland to present a more realistic population distribution over the whole country.

The dasymetric population map was visualized according to choropleth map rules in such a way that it focuses on the relationship between settlement and the natural environment. The information generated from a dasymetric population density map could provide useful assistance to district administrations, especially those responsible for regional or city development and land management.

REFERENCES

Bielecka E. and A. Ciolkosz. 2004. Land cover structure in Poland and its changes in the last decade of the 20th century, *Annals of Geomatics* 2 (1): 81-88. Warsaw.

Gallego J., S. Peedell. 2001. Using CORINE Land cover to map population density. *Towards agri-environmental indicators.* EEA Topic report 6/2001:94-105.

Mennis, J. 2003. Generating surface models of population using dasymetric mapping. *The Professional Geographer* 55(1):31-42.

Mennis, J. 2002. Using geographic information systems to create and analyse statistical surface of population and risk for environmental justice analysis. *Social Science Quarterly*, 83(1).

Eicher, C. L., and C. A. Brewer. 2001. Dasymetric mapping and areal interpolation: implementation and evaluation. *Cartography and Geographic Information Science* 28 (2): 125-38.

Langford M. and D. J. Unwin. 1994. Generating and mapping population density surface within a geographical information system. *The Cartographic Journal* 31:21-26.

Openshaw S. 1983. The modifiable areal unit problem. *Concepts and Techniques in Modern Geography* vol. 38. Norwich: Geobooks.

Perdigão V. A. Annoni. 1997. *Technical and Methodological Guide for Updating CORINE Land Cover Data Base.* JRC-EC.

Preobrazenski, A. J. 1954. Dorewolucjonnyje i sovietskije karty razmieszczenija nasielenia. *Woprosy Geografii. Kartografia* 34:134-149, Moscow.

Robinson A., R. Sale, J. Morrison.1988. Elements of Cartography. *PWN,* Warsaw 1988 (Polish edition)

Sutton P. 2003. Estimation of human population parameters using night-time satellite imagery. *Remotely Sensed Cities*, London: Taylor and Francis, 301-334.

Tobler W. 1979. Smooth pycnophylactic interpolation for geographic regions. *Journal of the American Statistical Association* 74: 519-530.

Wright J. K. 1936. A method of mapping densities of population with Cape Cod as an example. *Geographical Review* 26: 103-110.

BIOGRAPHY
Elzbieta BIELECKA

Education

Institution:	Faculty of Geography and Regional Study University of Warsaw
Date	11/1990
Diploma obtained:	Doctor of Natural Sciences

Institution:	Department of Geodesy and Cartography Warsaw Technical University
Date	09/1973 to 03/1978
Diploma obtained:	M. Sc. Eng. in Geodesy and Photogrammetry

Present position: Head of GIS Department at the Institute of Geodesy and Cartography

Professional experience

Date:	January 1994- present
Location	Warsaw, Poland
Company	Institute of Geodesy and Cartography
Position	Head of GIS Department from May 1997
Description	Participation in the research works on the application of using IT with special attention to LIS/GIS and remote sensing Specialist of digital mapping and data transferring Consulting experience related to LIS/GIS systems GIS application for environment, agriculture, statistics
Date:	June 1992 - December 1993
Location	Warsaw, Poland
Company	Neokart GIS
Position	Remote Sensing and GIS specialist
Description	Consulting experience related to spatial information systems and computer-aid cartography Executing of training course in the theory and practice of image processing
Date:	June 1978 - September 1992
Location	Warsaw, Poland
Company	Faculty of Geography, University of Warsaw
Position	Lecturer
Description	Teaching of remote sensing and geodesy Participation in the research works on the application of using remote sensing data in thematic mapping Studies of mountain ecosystems Consulting and training in using remote sensing data in the environmental study

Membership of professional bodies

Polish Geographic Society, Remote Sensing and Geographical Environment Club

AGILE Association of Geographic Information Laboratories for Europe

Polish Standardisation Technical Committee Geographical Information

Polish Committee of CODATA

Polish National ICA Committee

Polish Association for Spatial Information