



UDK 528.914

SELECTED ASPECTS OF SETTLEMENT GENERALIZATION IN THE GENERAL GEOGRAPHIC DATABASE IN POLAND

Izabela Karsznia

*Department of Cartography, Faculty of Geography and Regional Studies, University of Warsaw,
Krakowskie Przedmieście St. 30, 00-927 Warsaw, Poland
E-mail: i.karsznia@uw.edu.pl*

Submitted 07 May 2013; accepted 18 September 2013

Abstract. The aim of the study was to investigate and highlight the specifics of settlements generalization in the General Geographic Database in Poland. As all the settlements are presented by signatures and selected, bigger ones by outlines – two important aspects were taken into consideration. First of all an attempt of formulating a formal criteria of settlements selection presented by signatures was conducted. The main challenge in this aspect was to achieve a high degree of formalization and automation in the generalization process with inclusion of the specifics of generalized features. In order to fulfil those conditions a concept of regional differentiation of parameters in the settlements selection process was proposed. Additionally the possibilities of contextual settlements' outlines aggregation and simplification with the application of mathematical morphology operations were also investigated. By applying functions of modified erosion and dilation in the settlements' outlines aggregation as well as using a simplification algorithm it was possible to achieve results which are cartographically correct in the terms of shape generalization.

Keywords: spatial data generalization, automation of the generalization process, General Geographic Database, settlement selection process, context settlements' outlines aggregation, mathematical morphology operations.

Reference to this paper should be made as follows: Karsznia, I. 2013. Selected aspects of settlement generalization in the General Geographic Database in Poland, *Geodesy and Cartography* 39(3): 129–137.

Introduction

Along with the dynamic development of geographic information systems (GIS) and furtherly underlined by Poland's joint of the European INSPIRE initiative, the demands on GIS data infrastructures have significantly increased. Current challenges of creating spatial data infrastructures include automatic updating of database content, the possibility of generating data at different levels of detail, as well as data access through web services (Burghardt *et al.* 2005; Neun *et al.* 2009). Consequently, generalization processes have been widely discussed and researched as one of the main elements of map production and visualization derived from databases at multiple scales. According to Ware and Jones (2005) a map generalization should be balanced between satisfying the purpose of a map and maintaining the readability with respect to the target scale and level of detail.

So far, research on map generalization has almost exclusively focused on maps and spatial data at large scales (i.e. 1:10 000 to 1:50 000). This is mainly due to the fact that basic spatial databases with a nationwide coverage have normally been collected at the scales of 1:10 000, 1:25 000, or 1:50 000 and hence the need for their automated generalization has appeared most urgently. At the same time, neither coherent standards nor homogeneous principles of small scale map generalization have been elaborated. Hence, in a current production of small-scale topographic maps, cartographic generalization was largely carried out manually in interactive systems and thus relied on the fundamental decisions taken by a cartographer. The result of the process depended thus largely on the – not always consistent – decisions of its author. Consequently, small scale maps at the same scale may differ considerably

between different authors (Stoter *et al.* 2009). Thus, after a wave of research on large-scale map generalization, the automation of generalization procedures at medium and small scales is necessary, highly innovative, and challenging.

The aim of the study was to investigate and highlight the specifics of settlements generalization in the Polish General Geographic Database. As all the settlements are presented by signatures and selected, bigger ones by outlines – two important aspects were taken into consideration. First of all, an attempt of formulating of formal criteria of settlements selection presented by signatures was conducted. As a second step, some notions in the field of mathematical morphology, in particular erosion and dilation in the application to settlements' outlines generalization were discussed. Owing to their properties, these operations can be used as an excellent tool in cartographic generalisation, particularly in the process of aggregation and simplification of the settlements' outlines at medium and small scales. The paper tackles some methodological problems associated with the generalisation of settlements at general maps and offers some examples on how they could effectively be solved.

The scope of the presented research covers the attempts of the digital landscape model of the settlements generalization within two selected Polish provinces that is Lower Silesia and Lodz Province. The General Geographic Database is one of the main elements of the Polish Spatial Data Infrastructure. It has been elaborated at the basic level of detail of 1:250 000. The target level of detail of the generalization processes presented in the article is 1:500 000. The choice of research areas is not accidental. In both regions the settlement and the road network have been developing under the influence of different historical and geographical conditions. The Lodz Province is a typical region for central Poland with the road network much less developed here than in the south-west of the country, as it is in Lower Silesia. Similarly, the settlement pattern in both regions varies greatly. In the Lodz Province the settlement network is more fragmented and diverse, with a fairly well-developed urban centres, while weakly urbanized, even backward peripheries. However, in Lower Silesia predominant type of settlement is much concentrated, it consists of a network of equally spaced medium-sized and large villages, and nodal system of roads, mostly feudal origin.

1. Present state of the research

In the process of settlements generalization at medium and small scales we can point at four stages:

- selection of categories of objects (object classes) presented on a map and their classification,
- selection of objects within particular categories,
- change in a cartographic method of representation by means of replacing an outline of area feature with a signature (Ratajski 1967),
- aggregation and simplification of settlement's outlines.

One of the key stages in the process of settlement generalization is the selection of settlements, which will be presented at a smaller scale (Topfer, Pilewizer 1966; Kadmon 1972; Langran, Poiker 1986; Richardson, Muller 1991). Intuitively, we assume that the larger settlements (measured primarily with population values) should take precedence in terms of their appearance on the map in relation to the smaller ones. However, it is not fully clear that if we choose five settlements of the whole set, this would be the biggest ones. A large city located in close proximity to other large one can be omitted in the target scale. At the same time, a small village located in isolation from other cities may be saved, according to its "relative importance" (van Kreveld *et al.* 1997; Li 2007). Thus, in order to achieve logical and contextual selection of settlements at geographical scales it is crucial that geographical knowledge has to be taken into account to decide which objects are important enough to keep them on the target map as well as cartographic knowledge has to be considered, as it will allow to make a decision related to the number of presented objects. In this light, the goal in the selection process is extracting the most characteristic elements of map content while maintaining the relationship between geographical objects (Flewelling, Egenhofer 1993). One of the key issues is therefore an indication of the quantitative criteria of generalization.

Several algorithms have been proposed for point set generalization. Li (2007) has distinguished two groups of algorithms: selective omission and structural simplification. Within the algorithms for selective omission the following algorithms can be distinguished: settlement-spacing ratio, distribution-coefficient, gravity-modeling, set-segmentation and quadrat-reduction proposed by Langran and Poiker (1986) as well as one proposed by van Kreveld *et al.* (1997) – circle-growth algorithm. As pointed out by van Kreveld *et al.* (1997) the set-segmentation and

quadrat-reduction algorithm require a great deal of human intervention thus are not appropriate for automation. The other methods do not directly give a clear hierarchy of the base set of objects, what is more the objects with very high importance value have too much influence on the selection process. In the end it may result in not preserving the local objects density, which might not be suitable especially at medium and small scale generalization process.

The second group of algorithms simplifies a set of points by removing some of them based on various characteristics. One of the most popular solutions is to use Voronoi-diagrams, which helps to rank the points according to their Voronoi areas and additional thematic importance. Then the points with the lowest ranks can be removed iteratively (Samsonov, Krivosheina 2012).

Apart from the settlements selection, a very important factor which strongly influences on a way of its generalization is also a shape. The ways how settlements are represented and generalised, and especially how the outlines of towns or villages are generalised, has evolved as a result of many years of practice, based above all on experience and intuition. Generalisation of settlements' outlines is a challenging process owing to its complex and multifaceted nature. This is because settlements are those elements of map content which undergo the biggest changes following the process of generalisation from large-scale topographic maps to small-scale reference (general) maps. Therefore, generalisation of settlements should be subordinated to the principles and constraints of visual perception on the one hand, and on the other this process should take into account the characteristic spatial, physiognomic and functional features of settlements.

Mathematical morphology distinguishes a set of operations (especially erosion and dilation) allowing to incorporate the contextual shape characteristics, which makes generalization process more effective and correct.

Mathematical morphology is called a science considering a form and structure. It has developed on the basis of Hermann Minkowski's works from the early years of the 20th century. Other precursors of wider applications of mathematical morphology were also Georges Matheron (1975) and Jean Serra (1982).

The key concept in mathematical morphology is the notion of the structuring element. It is a subset of an artificially generated section of the image with one distinguishing point, the so-called *central point*, which is the centre of the system of coordinates for

this particular subset. As a result of the interaction between the digital image and the structuring element, morphological transformation is possible. The shape and size of the structuring element determines the scope of the morphological function. The purpose of this function is to modify the object so as to reveal its structure.

Erosion as a mathematical operation is defined as the process of removing the surface layer of the object. It can be conceptualised as a process whereby the external layer of the object is cut off, and the width of the affected area depends on the size and shape of the structuring element. Dilation is defined as an opposite process: roughly, as an addition of an external "belt" with a width depending on the size and shape of the structuring element.

Owing to such characteristics as contextual aggregation of parts of objects or emphasising their typical features, morphological operations can be used as excellent tools in the process of cartographic generalisation, particularly generalisation of built-up areas at topographic scales as well as settlements' outlines at medium and small scales.

Studies published so far on the generalisation of built-up areas as well as possible applications of mathematical morphology operations are associated with generalisation both vector and raster data at topographic scales.

The topic of the work by Su *et al.* (1997) is the elimination of small surface areas (such as building outlines, which is necessary in the process of generalisation as the map scale is reduced. The authors also point out that, with a scale reduction, there is a problem with sharp, irregular shapes which are produced by following the generalisation. That is why the process of the elimination of smaller areas must be preceded by the simplification of their contours. The paper presents techniques aimed at eliminating smaller objects and simplifying the remaining areas. These techniques have their origin in mathematical morphology and are associated with erosion and dilation operations. In this context, the generalisation has three stages:

- Elimination of smaller-surface areas using erosion operators;
- Return to the original object shapes;
- Simplification of the object outlines.

Examples of the applications of mathematical morphology operations in the process of generalisation of built-up areas in topographic scales can also be found in the works by Urena and Ariza (2000) and Boffet (2001). A combination of the traditional

Douglas-Peucker simplification algorithm based on erosion and dilation operations is also used by Ordnance Survey in the UK during the generalisation of built-up areas (Regnauld, Revell 2007; Revell 2008).

2. Proposed solutions

The focus of the presented research is the selection parameters for settlements generalization on medium and small scale maps as well as the investigation of contextual algorithms of settlements' outlines aggregation and simplification.

2.1. Selection of the settlements presented by signatures

During the settlement selection process at medium and small scales the population of the settlement is usually the leading factor. However, as stated before the selection process based only on the population values may result in an improper presentation of a size structure of settlements. In particular, the problem exists on the areas characterized by a significant ratio of villages with large number of population. In this case, near the administrative problem also the criteria of a size of settlements should be taken into account.

To keep a characteristic spatial order of settlements it seems to be very appropriate to take into account both a density as well as a structure of settlements' sizes. From that point of view such approach can be treated as supporting criteria. For that reason there was defined an indicator (called "density index" and expressed as a number of localities on 1 dm²) which could have been used in the automated generalization. Based on analysis of maps at similar scales it has been noticed that for the urbanized area of Poland density index should be in between 50 and 150 localities per 1 dm². In order to reflect the structure of settlements' sizes in the selection process, the assumption has been made that differentiation of settlement density should reflect differences of the population density.

To verify the density index two test areas with different ratio of settlements density were chosen. In following attempts there have been chosen localities over 100, 200, 300 and 400 population values. After that it has been checked which population choice reflects the most optimal settlement density and also density of population at this area. Furthermore, the values of settlements' density indexes were calculated. This made it possible to define the border value from which the numbers of citizens should have been displayed to

obtain an optimal density of localities. However, it had been difficult to find out what was the optimal density in range of the whole region (the whole Province area) due to the fact that it was dependent on the level of concentration and localities' size. As the proper solution it was considered an application of regional differentiation of the generalized area respectively to density and size of localities together with a local selection of density parameters. Nearby the density of localities it was introduced an additional supporting parameter – a density of population. As it has been highlighted the assumption was that the density of settlements should reflect the density of population. According to this fact as well as the decision that the regional selection parameter differentiation should take place within particular districts (the smaller administrative areas within two selected provinces), another analysis has been conducted. Firstly, only the rural districts were taken into account, whereas the urban districts containing only one large city were excluded from the analysis. What is more, within the remaining districts (in both provinces) two other groups of districts were identified as specific ones. The first group consisted of the districts from which the large city (in terms of its population) has been delimited and based on that, an urban district was created. It results in a situation where a general district population has been "artificially" undervalued. Another group of districts there were these where a large city was not excluded. As a result, a density of settlements in this district was not high, however, at the same time a real population density was significant. In order to provide comparability among all districts within a given province it was decided to introduce a concept of an "average" district city. A theoretical population of the "average" district city has been calculated by summing the population of all district within given province (excluding districts of "overstated" or "undervalued" population density), and then dividing the result by the number of concerned districts. In case of the districts of undervalued population density the "average" district city population has been added and the population has been recalculated. It resulted with obtaining the theoretical population density values. Respectively, for the districts with an "overstated" population density a city with the largest population value has been replaced with the "average" district city. Then, theoretical population densities were calculated for the considered districts. The relation between the density of settlements and the population density within the particular districts has been shown at Figures 1 and 2 (respectively for two

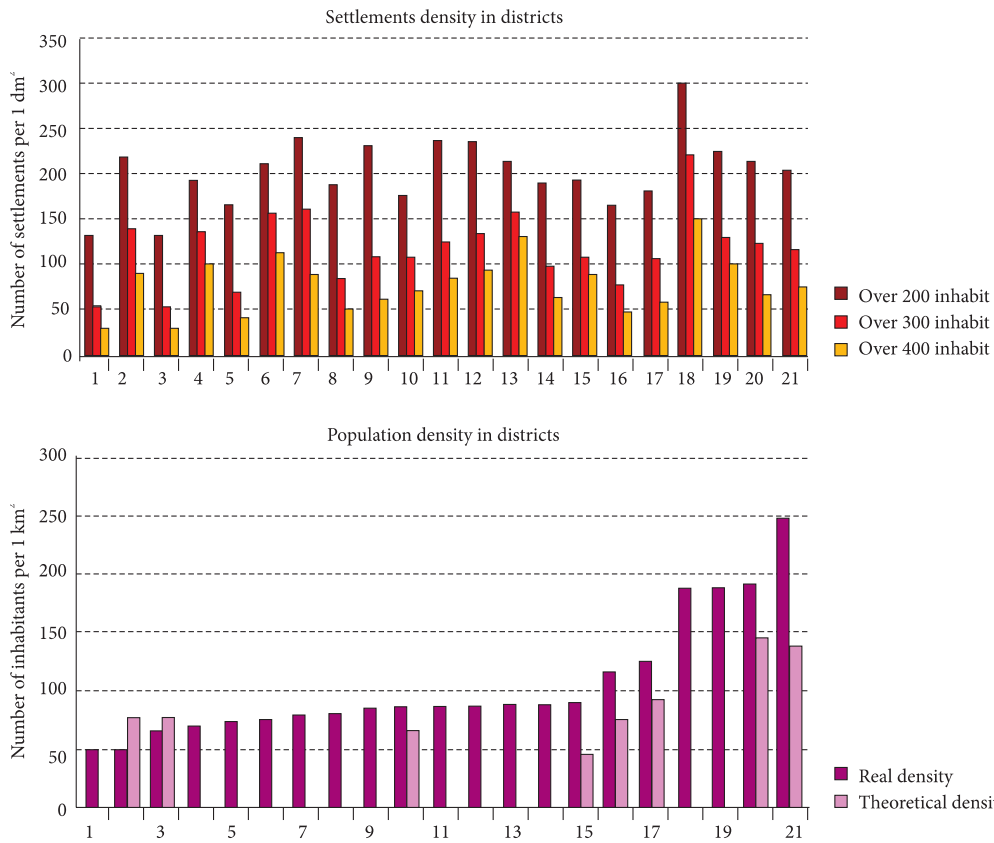


Fig. 1. The relation between settlements density and population density in districts within the Lodz Province

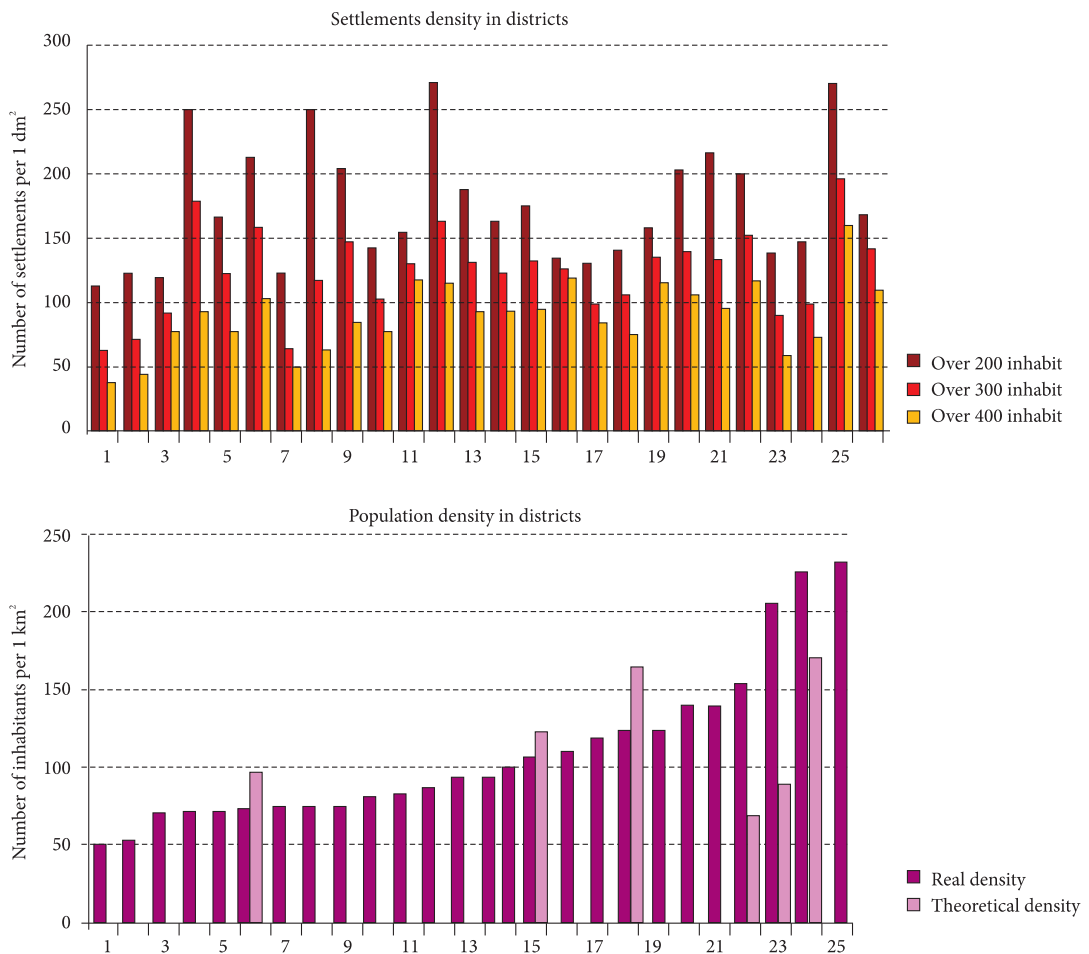


Fig. 2. The relation between settlements density and population density in districts within the Lower Silesia Province

provinces). The performed analyze considering this parameter made it possible to find out which of those two density factors reflect a better way for a population density in districts. Based upon that, the regional differentiation of settlements' density within particular districts was executed (Figs 3, 4).

2.2. Generalization of the settlements presented by outlines

The second aspect of the presented research is the investigation of contextual algorithms of settlements' outlines aggregation and simplification. The author proposes to use erosion and dilation operations in

general scales in the process of settlements' outlines generalisation of the General Geographic Database.

As part of the research, a *cluster settlements* algorithm was developed. This is an implementation of a sequence of erosion and dilation operations, which makes it possible to combine different outlines representing parts of towns or villages situated at a specific distance into one town or village. To achieve that, the algorithm was enriched with the functionality to identify, count, and finally amalgamate, parts of towns/villages having the same name.

In order to obtain optimised settlements' outlines shape the *action polygon erode* algorithm (supplied by

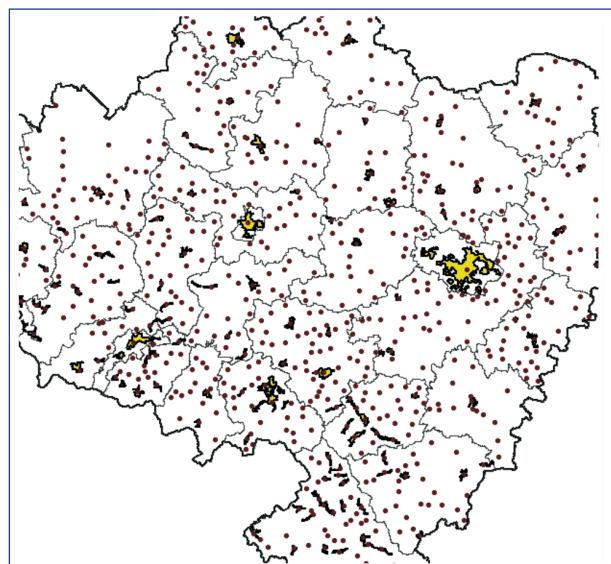
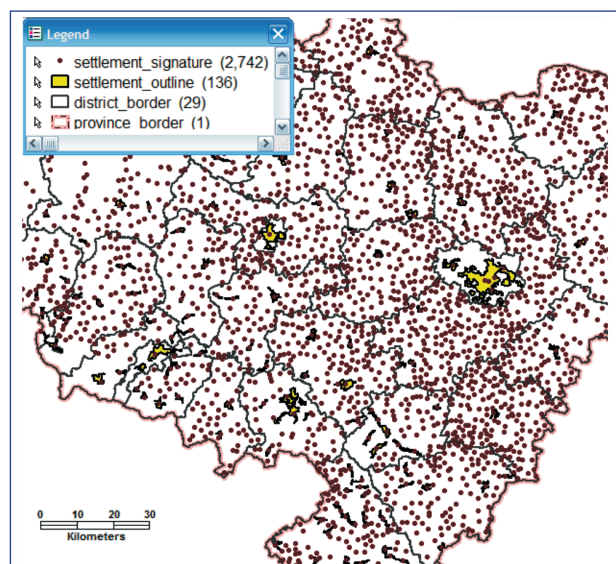


Fig. 3. The results of the regional differentiation of settlements' density within particular districts in Lower Silesia Province

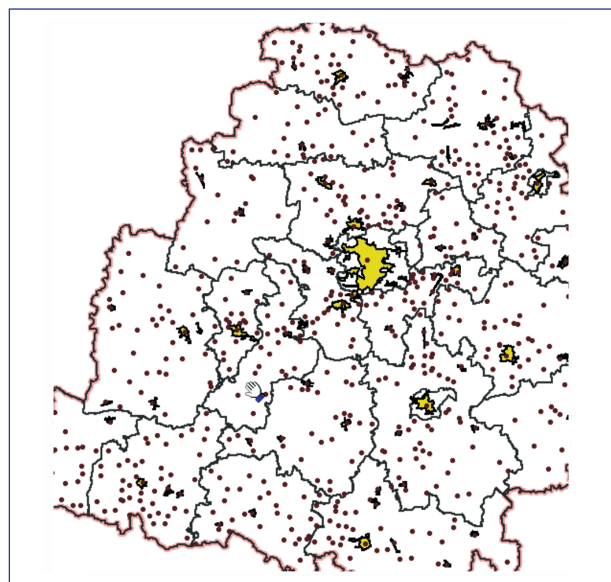
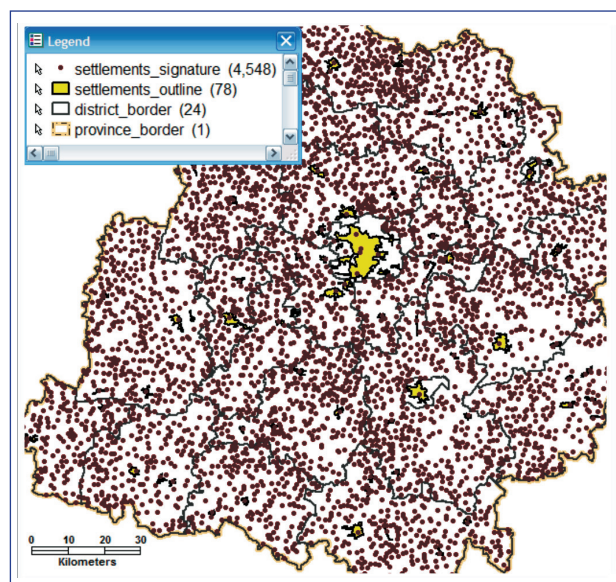


Fig. 4. The results of the regional differentiation of settlements' density within particular districts in Lodz Province

the Ordnance Survey) was used. This algorithm makes it possible to return to the original shape of the object following the buffering process. As the final step the settlements' outlines were respectively simplified according to the Douglas Peucker simplification algorithm.

The application and modification of the listed tools allowed for contextual aggregation of parts of the outlines of towns and villages, i.e. linking their outlines based on their names and the distance between them. In addition, putting together the Douglas Peucker simplification algorithm and the mathematical morphology operations discussed in this paper made it possible

to retain the characteristic elements relating to the shape of the generalised settlements' outlines. The results of the settlements' outlines aggregation and simplification process can be seen at the Figures 5 and 6.

Conclusions

The focus of the presented research was to highlight and present the main aspects and difficulties in the settlements' generalization process at medium and small scales. In order to investigate the generalization process of the settlements at medium and small scales the settlements generalization, both presented in a form of the signatures as well as outlines were

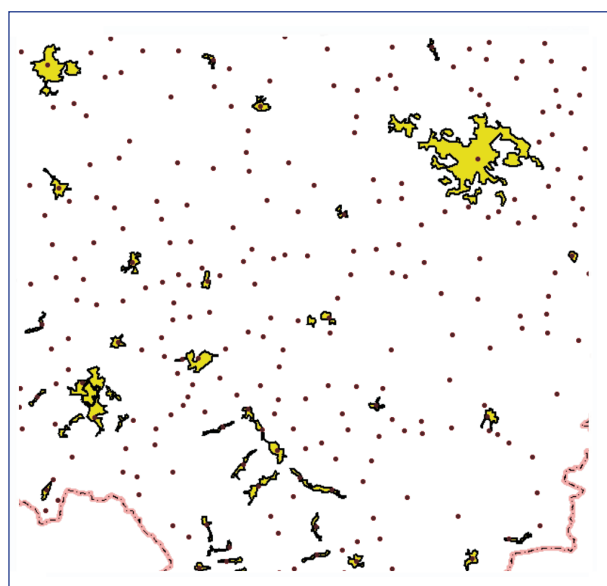
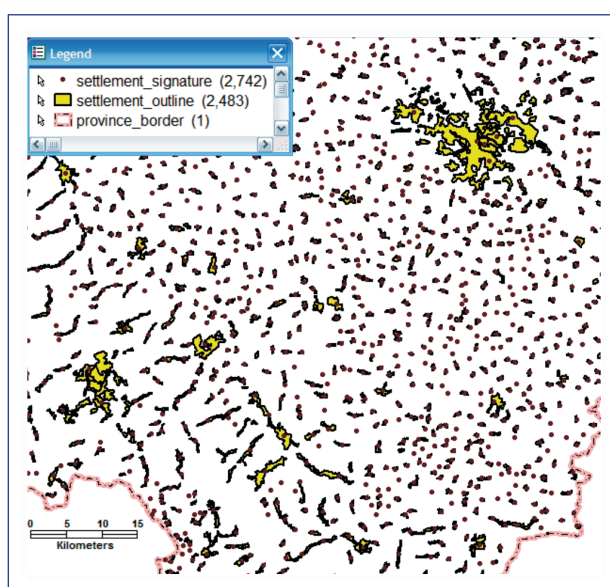


Fig. 5. The examples of the settlements' outlines aggregation and simplification process in the Lower Silesia Province

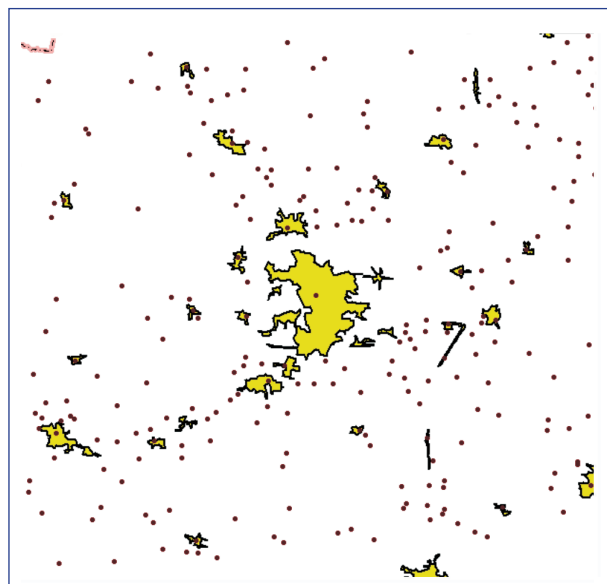
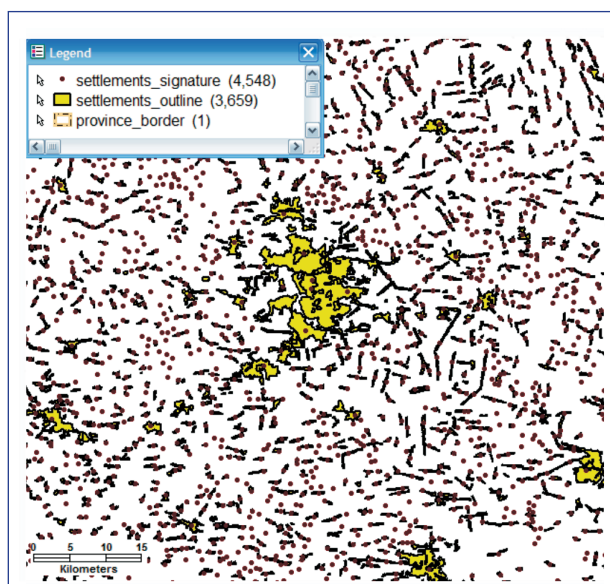


Fig. 6. The examples of the settlements' outlines aggregation and simplification process in the Lodz Province

investigated. The conducted experiments have been prepared based on the data collected within the Polish General Geographic Database.

According to the settlements presented by signatures, the regional differentiation of settlements' selection parameters within particular districts has been proposed. This solution makes it possible to take into account the settlements specifics, like settlements' size structure and settlements' density. It also helps to reflect the differences of the population density within the whole province.

In relation to the settlements' outline generalization it should be noted that the specific nature of the generalisation of general maps in many cases requires that different solutions to one problem should be found, depending on the context and surroundings of the objects. One significant restriction of the tools and systems supporting the automation of the generalisation process currently in use is the lack of contextual algorithms, which allow for the implementation of generalisation operations at a high conceptual level. An example here is the contextual aggregation of parts of towns or villages bearing the same names and situated at a given distance from one another, an operation which cannot be carried out using the out-of-the-box version of popular GIS systems.

On the other hand, the above-mentioned characteristics of morphological operations, such as inclusion of the objects' context and retaining (or even highlighting) their characteristic features, could offer excellent tools for the process of cartographic generalisation, and generalisation of built-up areas as well as settlements in particular. We can therefore say that the properties of morphological operations encapsulate the very essence of generalisation viewed as an intentional process involving the selection and highlighting of the basic geographic attributes of objects while taking into account a perception potential of a recipient, adapted to the map's function, topic and scale (Tyner 2010).

An automation of the generalization process of the settlements at medium and small-scale still lacks coherent and formal methodology. According to this situation the results of this research can contribute to understanding and formalizing a generalization process of medium and small scales.

References

- Boffet, A. 2001. *Methode de creation d'informations multi-niveau pour la generalisation cartographique de l'urbain*. Ph.D. thesis, Universite Marne La Valle, Paris.
- Burghardt, D.; Neun, M.; Weibel, R. 2005. Generalization services on the web – classification and an initial prototype implementation, *Cartography and Geographic Information Science* 32(4): 257–268.
<http://dx.doi.org/10.1559/152304005775194665>
- Flewelling, D. M.; Egenhofer, M. J. 1993. Formalizing importance: parameters for settlement selection from a geographic database, in *Proceedings of Auto-Carto 11* [online], [cited 04 May 2013]. Minneapolis: ACSM-ASPRS, 167–175. Available from Internet: <http://mapcontext.com/autocarto/proceedings/auto-carto-11/pdf/formalizing-importance-parameters-for-settlement-selection-from-a-geographic-database.pdf>
- Kadmon, N. 1972. Automated selection of settlements in map generation, *The Cartographic Journal* 9(2): 93–98.
<http://dx.doi.org/10.1179/000870472787352442>
- Langran, G. E.; Poiker, T. K. 1986. Integration of name selection and name placemen, in *Proceedings of 2nd International Symposium on Spatial Data Handling*. Seattle, WA, 50–64.
- Li, Z. 2007. *Algorithmic foundation of multi-scale spatial representation*. Boca Raton: CRC Press.
- Matheron, G. 1975. *Random sets and integral geometry*. New York: Wiley.
- Neun, M.; Burghardt, D.; Weibel, R. 2009. Automated processing for map generalization with web services, *GeoInformatica* 13(4): 425–452.
<http://dx.doi.org/10.1007/s10707-008-0054-3>
- Ratajski, L. 1967. Phenomenes des points de generalisation, *International Yearbook of Cartography* 7: 143–151.
- Regnauld, N.; Revell, P. 2007. Automatic amalgamation of built-up areas for producing Ordnance Survey 1:50 000 Scale Maps, *The Cartographic Journal* 44(3): 239–250.
- Revell, P. 2008. A review on the Clarity generalization platform and the customizations developed at Ordnance Survey research, in *ICA Workshop on generalisation and multiple representation*, 20–21 June, Montpellier, France.
- Richardson, D. E.; Muller, J.-C. 1991. Rule selection for small-scale map generalization, in B. Buttenfield, R. McMaster (Eds.). *Map generalization: making rules for knowledge representation*. London: Longman, 136–149.
- Samsonov, T. E.; Krivosheina, A. M. 2012. Joint generalization of city points and road network for small-scale mapping, in *Proceedings of GIScience, Seventh International Conference on Geographic Information Science*, Columbus, Ohio.
- Serra, J. 1982. *Image analysis and mathematical morphology*, vol. 1. London: Ac. Press.
- Stoter, J.; Burghardt, D.; Duchêne, C.; Baella, B.; Bakker, N.; Blok, C.; Pla, M.; Regnauld, N.; Touya, G.; Schmidt, S. 2009. Methodology for evaluating automated map generalization in commercial software, *Computers, Environment and Urban Systems* 33(5): 311–324.
<http://dx.doi.org/10.1016/j.compenvurbsys.2009.06.002>
- Su, B.; Li, Z.; Lodwick, G. 1997. Morphological transformation for the elimination of area features in digital map generalization, *Cartography* 26(2): 23–30.
<http://dx.doi.org/10.1080/00690805.1997.9714064>
- Topfer, F.; Pillewizer, W. 1966. The principles of selection: a means of cartographic Generalization, *The Cartographic Journal* 3(1): 10–16.
- Tyner, J. A. 2010. *Principles of map design*. New York: The Guilford Press.

- Urena, C. M. A.; Ariza, F. J. A. 2000. Mathematical morphology applied to raster generalization of urban city block maps, *Cartographica* 1(37): 33–48.
- Van Kreveld, M.; Van Oostrum, R.; Snoeyink, J. 1997. Efficient settlement selection for interactive display, in *Proceedings of Auto-Carto 13: ACSM/ASPRS Annual Convention Technical Papers*, Seattle, 287–296 [online], [cited 05 May 2013]. Available from Internet: <http://mapcontext.com/autocarto/proceedings/auto-carto-13/pdf/efficient-settlement-selection-for-interactive-display.pdf>
- Ware, J. M.; Jones, C. B. 2005. Map generalization in the web age, *International Journal of Geographical Information Science* 17(8): 743–769.
<http://dx.doi.org/10.1080/13658810310001596085>

Izabela KARSZNIA, Dr Eng., works as an Assistant Professor at the Department of Cartography, Faculty of Geography and Regional Studies, Warsaw University in Poland. In her research interests she concentrates on spatial data generalization, geographic information systems as well as geographic information visualisation aspects.