ON THE GENERALIZATION ALGORITHMS APPLIED IN GIS ENVIRONMENT

Gheorghe BADEA, Prof. PhD. Eng., Technical University of Civil Engineering Bucharest, Faculty of Geodesy, Romania, badeacadastru@gmail.com, gheorghe.badea@utcb.ro
Ana Cornelia BADEA, Prof. PhD. Eng., Technical University of Civil Engineering Bucharest, Faculty of Geodesy, Romania, anacorneliabadea@gmail.com, ana.badea@utcb.ro

***Abstract**: Generalization in GIS is a very complex and important aspect, and the evolution of GIS technology has led to the creation of tools based on complex algorithms, which greatly facilitate the work of specialists. GIS software capabilities thus include spatial transformations that change geometric data and topological representation. Data processing by generalization refers to raster data and vector data. The article focus is on presenting and comparing the usability of four simplification algorithms (Douglas-Peucker (DP), Wang-Müller(WM), Zhou-Jones (ZJ), Visvalingam-Whyatt(VW)), applied for some examples on polygon layers, to emphasize the changes due to generalization.*

***Keywords:** Generalization, GIS, ArcGIS Pro*

1. Introduction

The generalization issues in GIS environment has advanced and include complex strategies that support analysis and feature recognition, exploiting spatial and semantic contexts, and preserving relationships between and among geospatial features. Functions for generalizing spatial data are of fundamental importance in GIS because of a variety of requirements for scale-changing and thematic reduction and emphasis. Many professional and academic GIS specialists consider the process of generalization is one of the most technically challenging components of making representations at different scales.

In contrast to the analogic representations, the digital ones in GIS have a wider meaning. It is a process which realises transitions between different models representing a portion of the real world at decreasing detail, while maximising information content with respect to a given application. [9]

Simplification reduces the number of points in a line, while retaining the most representative points. [12] The scale of a representation is not the only factor which affects generalization and it is needed to know the purpose of the representation, because it can be determined what is important, which objects or categories of objects to use, and the symbology of objects. [7]

In recent years, research on line simplification has focused on shape maintenance, topological consistency, geo-characteristics, and data quality and it was presented a new approach for line simplification based on shape detection of geographical objects in self-organizing maps. [2] For instance for a given boundary, to obtain the essence of the shape is the aim of the approximation and we can have as result few segments as possible. In the figure 1 are emphasied the generalization concept. This concept is answering at three questions: why, when and how to generalize. The first one The last one is about spatial and attribute transformations.

In the figure 1, the theoretical elements consist of reducing complexity, maintaining spatial accuracy, maintaining attribute accuracy, maintaining aesthetic quality, maintaining a logical hierarchy, consistently applying rules. The application-specific elements are representation purpose, appropriateness of scale, retention of clarity. The computational elements are cost effective algorithms, maximum data reduction and preferably minimum memory/ storage usage.

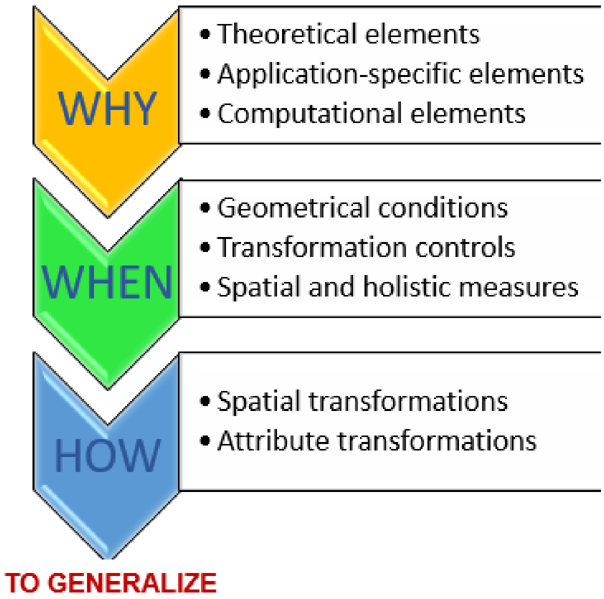

Figure 1 – Concept of Generalization (adapted from [9])

As geometrical conditions, it is needed to be taken into account: congestion, merging, conflict, complexity, inconsistency, imperceptibility. The spatial and holistic measures are linked with some specific measurements, like density, distribution, length and sinuosity, shape, distance measurements. [9] When it comes about applying generalization tools, transformation controls are needed: generalization operator, algorithm and parameter selection.

In GIS environment are tools focuse on spatial transformations: simplification, smoothing, aggregation, amalgamation, merging, collapse, refinement, exaggeration, enhancement, displacement, and also attribute transformations, like classification and symbolization. [11]

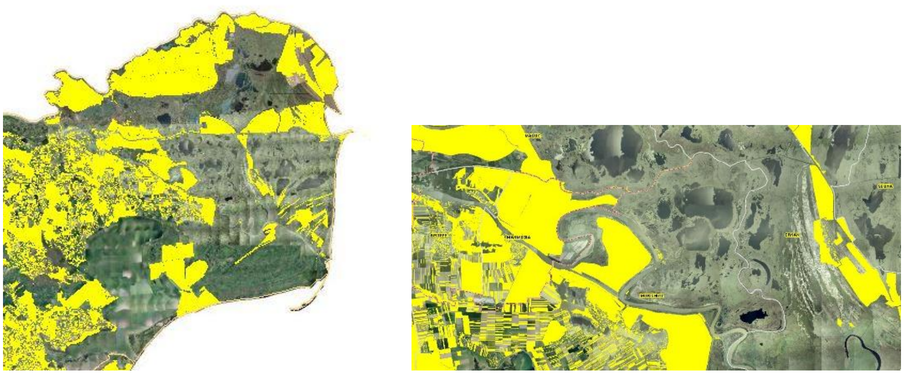

Figure 2 – Generalization Example (National Agency of Cadastre and Land Registration – NACLR Geoportal [21])

2. Simplifying Algorithms

The problem of generalization from the simpify features viewpoint is very complex, there are multiple approaches that have not been implemented in commercial GIS software products. From the specialized literature [14] it needs to be mentioned the Reumann-Witkam routine, the sleeve-fitting polyline simplification algorithm (also called the Zhao–Saalfeld algorithm), the Opheim simplification algorithm, the Lang simplification algorithm.

In ArcGIS Pro, there are useful tools to obtain data less detailed and less complex for analysis and representation in reduced scales or for other special purposes. There are implemented several methods that reduce the complexity of a polygon feature while retaining its basic shape, useful for simplifying features for display at smaller scales or standardize a dataset to a more uniform scale resolution. Beyond the simplification tool are implemented the above mentioned algorithms for different purposes.

Through applying simplifying tool, a feature changes curved segments to series of line segments and existing line segments remain unchanged. The maximum allowable set offset influence how the final shape is generalized.

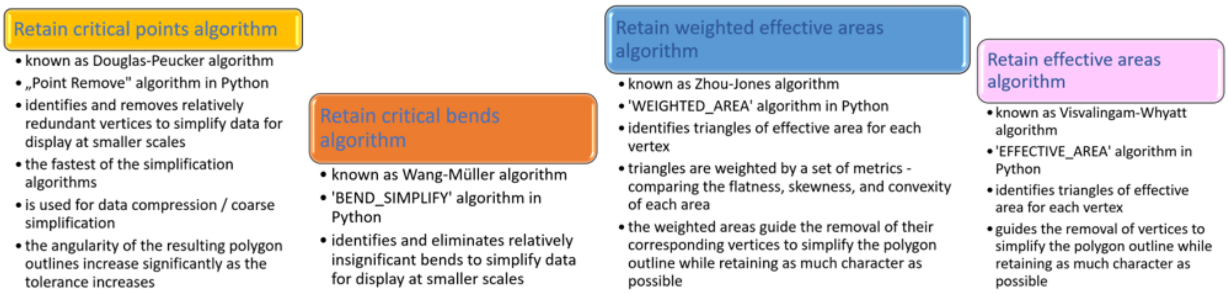

Figure 3 – Analyzed Algorithms

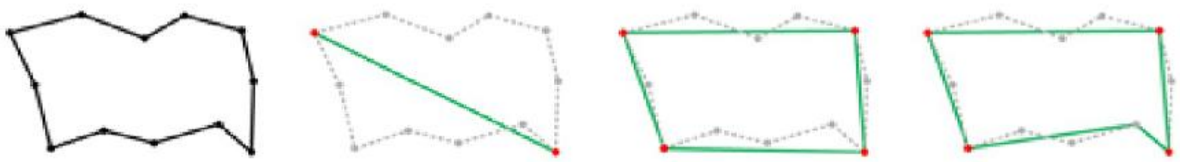

Figure 4 - Polygon simplification by modified Douglas-Peucker algorithm [13]

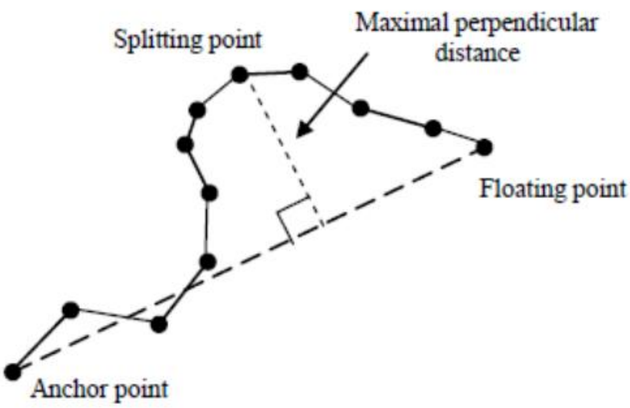

Figure 5 – Douglas-Peucker Algorithm [4]

From the mathematical viewpoint, polygon simplification by modified Douglas-Peucker algorithm (figure 5) starts with the input polygon, identifying the initial start and end nodes with maximum curvatures. It is obtained the simplified polygon after first round and finally the simplified polygon. The regions defined by the layer contours may have overlaps and gaps in between and the conflicts are solved by snapping contours.

One important parameter is the simplification tolerance parameter which has impact on the degree of simplification, because for a larger tolerance is obtained a more coarse resulting geometry, and for a smaller tolerance it is obtained a geometry more similar to the input features.

For the Douglas-Peucker algorithm, the tolerance means the maximum allowable perpendicular distance (figure 5) between each vertex and the newly created line. In case of the Wang-Müller algorithm, the tolerance is the diameter of a circle that approximates a significant bend. In case of Zhou-Jones algorithm, the square of the tolerance is the area of a significant triangle defined by three adjacent vertices. The further a triangle deviates from equilateral, the higher weight it is given, and the less likely it is to be removed. [20] In case of Visvalingam-Whyatt algorithm, the square of the tolerance is the area of a significant triangle defined by three adjacent vertices.

It exists the option that any polygons that are smaller than the minimum area parameter are removed from the output feature class. For a group of adjacent polygons that share common edges, the minimum area applies to the total area of the group. The keep collapsed points parameter is used to retain a record of removed polygons as point features. The output polygon feature class is topologically correct. Any topological errors in the input data are flagged in the output polygon feature class. Multipart polygons are simplified as individual parts.

Input barrier layers is a parameter used to identify features that must not be crossed by simplified polygons and the barrier features can be points, lines or polygons.

For the Wang-Müller algorithm, the main criterion guiding the process of generalization is line structure, which itself can be decomposed into a series of line bends. [3] This method preserves the overall structure with line bends which are mathematically defined according to size, shape, and context. According to [14], line simplification introduces positional error in the resultant map and may cause topological errors and positional error is compared with positional error resulting from the Douglas–Peucker algorithm. After comparing workflow [14] it was concluded that the Douglas–Peucker algorithm is the most effective to preserve the shape of the line and the most accurate in terms of position.

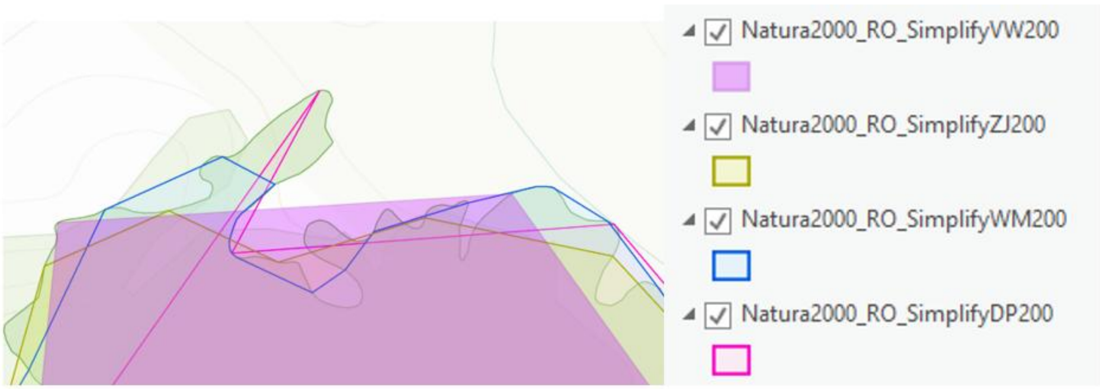

Figure 6 – Simplification applied on Protected Areas layer, v. 2019 - Bucegi Area (algorithms (Douglas-Peucker (DP), Wang-Müller(WM), Zhou-Jones (ZJ), Visvalingam-Whyatt(VW) – 200m)

Figure 7 – Simplification applied on Counties layer - Salaj Area (algorithms (Douglas-Peucker (DP), Wang-Müller(WM), Zhou-Jones (ZJ), Visvalingam-Whyatt(VW))

In ArcGIS Pro are two methods for smoothing [14]. The polynomial approximation with exponential kernel (PAEK) method smooths polygons based on a smoothing tolerance and each smoothed polygon may have more vertices than its source polygon. The Bezier interpolation method smooths polygons without using a tolerance by creating approximated Bezier curves to match the input polygons. For large datasets, the second smoothing algorithm requires more time. (figure 9)

In figure 10 is highlighted an example of building simplification based on two parameters: the minimum area and the simplification tolerance.

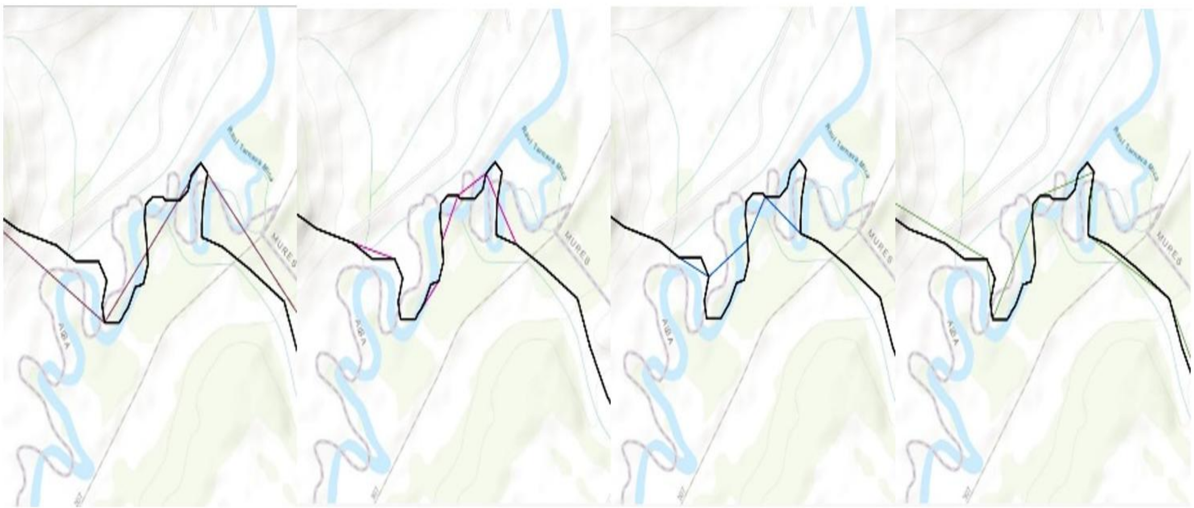

Figure 8 – Simplification applied on Counties layer – Mures and Alba County Limit (algorithms (Douglas-Peucker (DP), Wang-Müller(WM), Zhou-Jones (ZJ), Visvalingam-Whyatt(VW))

Figure 9 – Example of Smoothing Results (Polynomial Approximation with Exponential Kernel – left; Bezier - right) obtained after Applying DP Algorithm

Figure 10 – Building Simplification - min area 200sqm., with checking of spatial conflicts

3. Conclusions

First applied algorithm, retain critical points, can be succesfully applied for data compression and for eliminating redundant details. The disadvantage of the result is that the line that results may contain sharp angles and spikes. This algorithm is more adequate for small datasets reduction or compression and not for a high quality of representation. The second algorithm is respecting more the input geometry than the first, but it is needed more time to process.

NACLR (National Agency of Cadastre and Land Registration), like others NMA (National Mapping Agencies), maintains spatial databases at different levels of detail, in which are store multiple representations of the same geographic phenomena. Multi-representation databases are created by connecting objects among existing databases or by generating smaller scale representations from a single large scale database via automatic generalization.

Currently the GIS environment offers multiple generalization possibilities, based on complex algorithms. However, the principles underlying the creation of geoprocessing tools [8] must be known and analyzed in order to make the best choice to reflect the representation needs.

4. References

1. *Kolanowski, B., Augustyniak, J., Latos, D., Cartographic Line Generalization Based on Radius of Curvature Analysis, ISPRS International Journal of Geo-Information, 2018, https://www.mdpi.com/2220-9964/7/12/477/pdf*
2. *Shen, Y., Ai, T., He, Y., A New Approach to Line Simplification Based on Image Processing: A Case Study of Water Area Boundaries, ISPRS International Journal of Geo-Information, 2018 https://www.mdpi.com/2220-9964/7/2/41/pdf*
3. *Wang, Z., Muller, J.C,. Line generalization based on analysis of shape characteristics., Cartogr. Geogr. Inf. Sci. 1998, 25, 3–15, https://www.tandfonline.com/doi/abs/10.1559/152304098782441750?src=recsys*
4. *Douglas, D., Peucker, T., Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, The Canadian Cartographer 10(2), 112-122, 1973*
5. *Badea, A. C,. Badea, G., Concepte 2D, 3D si analiza GIS -Sinteze, Aplicaţii, in Planificare spaţială şi GIS pentru dezvoltare durabilă, Editura MATRIX ROM Bucureşti, 2017, ISBN vol 1: 978-606-25-0379-6, ISBN vol 2: 978-606-25-0380-2, https://www.librarie.net/p/313928/planificare-spatiala-si-gis-pentru-dezvoltare-durabila-sinteze, https://www.matrixrom.ro/produs/planificare-spatiala-si-gis-pentru-dezvoltare-durabila-aplicatii/*
6. *Badea, A. C., Badea, G., Cadastru, bănci de date și aplicații GIS în zone urbane, Editura Conspress, 2013, ISBN 978-973-100-310-8, http://www.agir.ro/carte/cadastru-banci-de-date-si-aplicatii-gis-in-zone-urbane-121878.html;*
7. *Droppová, V., The Tools of Automated Generalization and Building Generalization in an ArcGIS Environment, Slovak Journal of Civil Engineering, Vol. XIX, 2011, No. 1, 1 – 7, DOI: 10.2478/v10189-011-0001-4;*
8. *Badea, A. C., Badea, G., The Advantages of Creating Compound GIS Functions for Automated Workflow, pp. 943 – 949, INFORMATICS, GEOINFORMATICS AND REMOTE SENSING, Conference Proceedings, vol. I, ISBN 978-954-91818-9-0, DOI:10.5593/SGEM2013/BB2.V1/S11.043*
9. *Weibel, R., Dutton, G., Generalising spatial data and dealing with multiple representations, https://www.geos.ed.ac.uk/~gisteac/gis_book_abridged/files/ch10.pdf*
10. *Müller, J.-C., Lagrange, J.-P., Weibel, R., GIS And Generalisation: Methodology And Practice, Taylor and Francis, 1995*
11. *Davis, C. A. Jr., Laender, A. H. F., Multiple Representations in GIS: Materialization Through Map Generalization, Geometric, and Spatial Analysis Operations, http://www.dpi.inpe.br/gilberto/papers/davis_multiple_representations.pdf*
12. *Jamali, S., Generalization algorithms, http://web.nateko.lu.se/courses/ngen06/Lectures/L9-Generalisation.pdf*
13. *Xiong, B., Oude Elberink, S., Vosselman, G., Footprint Map Partitioning Using Airborne Laser Scanning Data, 10.5194/isprs-annals-III-3-241-2016, https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/III-3/241/2016/isprs-annals-III-3-241-2016.pdf*

14. *Shi, W., Cheung, C., Performance Evaluation of Line Simplification Algorithms for Vector Generalization, The Cartographic Journal Vol. 43 No. 1 pp. 27–44 March 2006*
15. *https://doc.arcgis.com/ (last accessed May 2020)*
16. *https://gisgeography.com/ (last accessed May 2020)*
17. *https://proceedings.esri.com/ (last accessed May 2020)*
18. *https://pro.arcgis.com/ (last accessed May 2020)*
19. *http://geoportal.ancpi.ro/geoportal/imobile/Harta.html (last accessed May 2020)*