Cartography M.Sc.

Master Thesis

Analysis of a Workflow for the Automated Generalization of Geological Maps

Ana Oliva Pinilla Pachon

Technical University of Munich
TECHNISCHE UNIVERSITÄT WIEN Vienna University of Technology
TECHNISCHE UNIVERSITÄT DRESDEN
UNIVERSITY OF TWENTE. ITC

September 9, 2022

TECHNISCHE UNIVERSITÄT DRESDEN

Statement of Authorship

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

„Analysis of a Workflow for the Automated Generalization of Geological Maps"

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Dresden, September 9, 2022 *Ana Oliva Pinilla Pachon*

TECHNISCHE UNIVERSITÄT DRESDEN

Faculty of Environmental Science Department of Geoscience Institute of Cartography

Master Thesis

Analysis of a Workflow for the Automated Generalization of Geological Maps

submitted by	**Ana Oliva Pinilla Pachon**
born	03.06.1992 in Bogotá Submitted for the academic degree of Master of Science (M.Sc.)
Submission on	September 9, 2022
Supervisors	Prof. Dipl.-Phys. Dr.-Ing. habil. Dirk Burghardt (Technische Universität Dresden) Dipl.-Geogr. Marco Schwarzak (Bayerisches Landesamt für Umwelt)
Reviewer	Dr.-Ing. Holger Kumke (Technische Universität München)

TECHNISCHE UNIVERSITÄT DRESDEN

Analysis of a Workflow for the Automated Generalization of Geological Maps

submitted for the academic degree of Master of Science (M.Sc.) conducted at the Institute of Cartography, TU Dresden

Author	**Ana Oliva Pinilla Pachon**
Study course	Cartography M.Sc.
Supervisors	Prof. Dipl.-Phys. Dr.-Ing. habil. Dirk Burghardt (TUD) Dipl.-Geogr. Marco Schwarzak (LfU)
Reviewer	Dr.-Ing. Holger Kumke (TUM)
Chair of the thesis assessment board	Prof. Dipl.-Phys. Dr.-Ing. habil. Dirk Burghardt
Submission on	September 9, 2022

Statement of Authorship

Herewith I declare that I am the sole author of the thesis named

„Analysis of a Workflow for the Automated Generalization of Geological Maps“

which has been submitted to the thesis assessment board today.

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Dresden, September 9, 2022 *Ana Oliva Pinilla Pachon*

To my parents, Germán and Oliva, my siblings, Edwin, Bety, Maria, and Marcela, and my niece and nephew, Isabella and Nikolai.

Acknowledgments

I would like to express my deepest gratitude to my supervisors Prof. Dr.-Ing. Dirk Burghardt from TU Dresden and Dipl.-Geogr. Marco Schwarzak from the LfU for their support, guidance, and invaluable patience during the completion of this thesis. I also would like to thank Dr.-Ing. Holger Kumke for his review and comments on my thesis.

I am also grateful to Katja Welzel and Dr. Melanie Meyer from LfU for their help collecting and organizing the necessary data for the study. Many thanks to the cartographers and geologists from LfU who evaluated the outputs maps. Additionally, I would like to thank Juliane Cron for her kindness and support from the beginning of this master's.

Special thanks to Peter for his constant support, calls to motivate me when I needed it the most, and company. Lastly, many thanks to my family for their unconditional support along this journey.

Abstract

This thesis attempts to analyze and evaluate an automatic procedure for the generalization of geological maps from a 1:25,000 scale (source) to 1:150,000, 1:500,000, and 1:1,000,000 scales (target) based on an existing workflow. The workflow and the source geodatabase were provided by the Bayerisches Landesamt für Umwelt (LfU). The procedure was implemented in a study area in southern Bavaria, Germany, to produce geological maps at the three target scales. Then cartographers and geologists from the LfU performed a subsequent qualitative evaluation to identify the advantages and disadvantages of the workflow. It consisted of delivering PDFs and GDBs, reviewing the outputs, and writing comments either on the PDFs or in a different document. The results showed that the workflow has the potential to be used for future generalizations of geological maps at LfU. However, limitations found need adjustments to achieve the desired outputs. Finally, based on the findings, recommendations for future improvements were proposed.

Keywords: geological maps, map generalization, cartography.

Contents

1 Introduction

1.1 Motivation

The nature of geological maps makes it difficult to adapt the amount of detail when reducing the scale. A geological map is a graphic representation of the distribution of rocks, geological structures, and mineral deposits (Downs & Mackaness, 2002, p. 1). It consists of a topographic map in which colored polygons are mainly overlapped to depict the different types of rocks (Maltman, 1990, p. 1). Some of its applications include civil engineering, land use planning, mining, and energy resources. Due to the complexity and variety of applications of geological maps, the generalization process normally requires the participation of a geologist with expertise in geological mapping and knowledge of the area (Smirnoff et al., 2012, p. 67).

Cartographic generalization is a central and complex process within mapmaking. Initially, there was a focus on individual operators (e.g., simplification, aggregation) that later changed to holistic solutions because of the challenge represented by the operators' interplay (Sester, 2020, p. 6). Thus, over the last two decades, almost purely manual production lines have started to be replaced by fully or semi-automatic production workflows. However, the research in this field has been mainly focused on topographic maps (Steiniger & Weibel, 2005), which has led to gaps in other areas such as geological mapping.

1.2 Problem Statement

This research will focus on the analysis of an existing workflow for the automatic generalization of geological maps. This will be carried out in three main stages (i) data processing, (ii) evaluation of the results, and (iii) workflow orchestration. For this purpose, the Bayerisches Landesamt für Umwelt (LfU) will provide the 1:25,000 scale database and the existing procedure.

1.3 Research Sub-Objectives and Research Questions

Main Objective

Analyze and evaluate an automatic procedure for the generalization of geological maps from a 1:25,000 scale (source) to 1:150,000, 1:500,000, and 1:1,000,000 scales (target) based on an existing workflow.

Sub-Objective and Reseacrh Question

The following sub-objectives (1, 2, and 3) will be achieved by answering the following research questions.

1. Execution of the generalization process of geological maps scaled at 1:150,000, 1:500,000, and 1:1,000,000 (target scales) from a 1:25,000 scale geological database (source scale).
   (a) Is generalization of different target scales achievable by using different parameter values, or do the workflows need to be modified?
   (b) What errors cannot be allowed as a result of the generalization process?
2. Evaluation of the results considering different target scales.
   (a) How can the evaluation process be carried out?
   (b) What are the semantic and geometric constraint parameters that lead to a successful automated generalization of a geological map?
   (c) How well do the results of this workflow satisfy the qualitative evaluation at the LfU for the different target scales?
3. Orchestration of the workflow to ensure its use at the Bayerisches Landesamt für Umwelt (LfU).
   (a) How can changing the values of the constraints influence the results of the generalization process?

1.4 Thesis Structure

In the **background and related work** chapter, I present a literature review on the characteristics of geological maps and generalization in categorical maps. In the generalization section, I focus on vector-based and raster-based generalization methodologies, operators, algorithms, constraints, levels of analysis, and evaluation in generalization.

In the **methodology** chapter, I introduce the current state of the geological cartography at the LfU, the overview of the software used, the structure and components of the workflow, and the evaluation process.

In the **implementation** chapter, I introduce the study area, explain the step-by-step workflow implementation in the study area, and outline how the evaluation was performed.

In the **results** chapter, I present the results from the evaluation grouped into different categories.

In the **discussion** chapter, I discuss the results found about the workflow itself and the main limitations.

In the **conclusion** chapter, I answer the research questions formulated for this thesis. Then, I discuss the overall findings and reflect on the shortcomings.

In the **outlook** chapter, I make some recommendations for future improvements.

2 Background and Related Work

This chapter presents a background on the characteristics of geological maps and explains the generalization process applied to categorical maps—additionally, a brief description of the Polsby-Popper (PP).

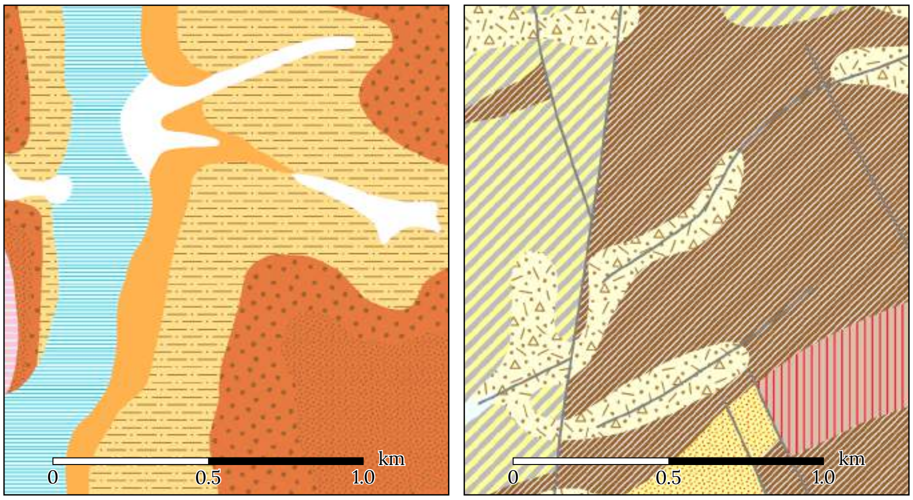
In section 2.1, a description of the particularities of geological maps can be found. Section 2.2 explores the generalization process. In the latter, generalization is first discussed from the point of view of vector and raster formats. Then the operators, algorithms, constraints, and levels of analysis (or spatial levels) are described. Finally, some approaches for the evaluation of the generalization process are mentioned. Section 2.3 explains what Polsby-Popper is and what it indicates.

2.1 Particularities of Geological Maps

A geological map contains valuable information not only for geologists themselves but also for professionals from other disciplines. For instance, in civil engineering, a geological map assists in the planning (e.g., location and construction) of engineering structures. For natural resource exploration, geological maps contain information about bedrocks that can carry raw materials and groundwater. Some other applications include land use planning, mining, and academia. Therefore, the purpose and type of geological map will determine the relative importance of certain elements in the map (Sayidov et al., 2020, p. 2).

A geological map is probably one of the most complex thematic maps that depict the composition of the earth's surface and at the depth of an area. It has colored areal objects that represent geological units. A specific type of rock and age are assigned to each polygon. Additionally, lineal elements show geological structures (e.g., folds and faults) while point elements represent features such as volcanoes, impact craters, wells, etc. depending on the scale. Given its complexity, geological map creation requires cartographic and geological knowledge (Sayidov et al., 2020, p. 2).

One of the key elements in a geological map is its polygons (Smirnoff et al., 2012, p. 67). These areal elements vary greatly both in their geometry and semantics. Figure 2.1 shows polygons with diverse shapes, sizes, boundaries, colors, and textures. On the other hand, a topographic map contains, among others, anthropogenic features which have more regular shapes (Sayidov et al., 2020, p. 3). Thus, although the generalization operators are applicable in both types of maps, the arrangement between the algorithms may differ (Sayidov et al., 2020, p. 3).



(a) (b)

Figure 2.1. Parts of a geological map that show different levels of complexity and interplay with geological structures (e.g., faults). **(a)** Large polygons. **(b)** Polygons of different sizes and shapes as well as the presence of faults. Data source: Bayerisches Landesamt für Umwelt, www.lfu.bayern.de. Adapted from the geological map of Bavaria, Germany at a scale of 1:25,000 (dGK25) (Bayerisches Landesamt für Umwelt, n.d-a).

2.2 Generalization on Geological Maps

Generalization is the process responsible for scale reduction (map generalization or cartographic generalization). It implies the reduction of the complexity to obtain a high-quality map. It could be achieved by emphasizing the essential while removing the unimportant. Hence, the generalized map should be easily perceived and the message behind understood. (Weibel, 1997, p. 101)

McMaster and Shea (1992) defined “digital generalization” as deriving cartographic data from a data source using spatial and semantic transformations. In the context of digital generalization, three types of generalization can be distinguished: object generalization, model generalization, and cartographic generalization.

- → *Object generalization* is defined by Weibel and Dutton (1999) as the process that occurs when defining and building the original database from the real world.
- → *Model generalization*, in turn, has the purpose to create a modified database taking the primary database as the source. That means that this generalization occurs in the geographic database and graphical aspects are not considered (Sarjakoski, 2007, p. 17).
- → *Cartographic generalization* encompasses the generalization of spatial data, especially for cartographic visualization (Weibel & Dutton, 1999, p. 128). It can be also understood as the needed processes to achieve the required cartographic presentation for the generalized maps (Sarjakoski, 2007, p. 16).

There are two main approaches for generalization in categorical maps: vector and raster-based (Sayidov, 2021, p. 22). Galanda (2003a, p. 8) defines polygon generalization as the generalization process in polygon subdivision, which means in the vector data model. Peter and Weibel (1999, p. 4) state that a categorical dataset in raster format consists of an array of cells in which each cell corresponds to a particular value. In a geological map, a color is assigned to a cell. Each color in this case depicts a geological unit. Some examples for each approach are presented in section 2.2.1 and section 2.2.2, respectively.

2.2.1 Polygon Generalization Approaches

Galanda (2003a) defines polygonal subdivision as the categorical coverage represented by vectors. Polygon generalization can be seen with a focus either on algorithms or holistic solutions.

On the one hand, the **algorithms** developed for generalization that are or can be adapted to polygon generalization can be divided into two groups: areal objects or exterior boundaries algorithms. An example for each can be the Delaunay triangulation (DT) and line simplification algorithms, respectively.

Delaunay triangulation consists of triangles whose circumcircles do not contain a vertex or vertexes from another triangle in the triangulation (Bader & Weibel, 1997). One derivation is the so-called constrained Delaunay triangulation (CDT) in which the boundary edges are maintained without being split (no points added) which means that some of the triangles might not meet the Delaunay condition of no points within the interior of circumcircles (Fleischmann, 2000). An implementation of this method on map generalization was made by Jones et al. (1995; as cited in Bader and Weibel, 1997). Another variation is the conforming Delaunay triangulation (RDT) in which the boundary edges are split by inserting points. In this case, all triangles meet the Delaunay condition (Fleischmann, 2000). Bader and Weibel (1997) use the RDT to eliminate small polygons. In this example, a skeleton of the polygon under consideration is built based on the RDT of the polygon interior (Figure 2.2). Then the chains of points in the adjacent polygons are connected to the closest point in the skeleton. Hence, the remaining skeleton represents the new boundary between the polygons (Figure 2.2). Algorithms based on CDT and RDT work for different generalization operators such as aggregation, collapse, displacement, elimination, enlargement, and exaggeration (Galanda, 2003a, p. 18).

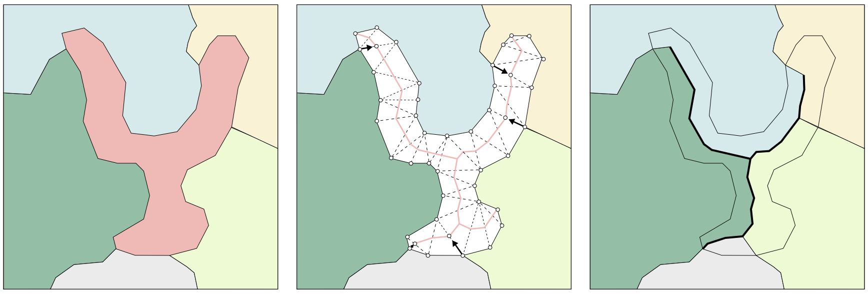


Figure 2.2. Elimination of polygons using RDT. Adapted from Bader and Weibel (1997) and Galanda (2003a, p. 18).

Another way to approach polygon generalization is by considering only the boundaries of the polygons. To do this, the use of line simplification algorithms is a good method. Two of the most well-known are Douglas-Peucker (Douglas & Peucker, 1973) and Visvalingam-Whyatt (Visvalingam & Whyatt, 1993) algorithms. Additionally, one important point to consider is that polygon generalization must guarantees both no self-intersection within a polygon outline and non-intersections with other elements (Galanda, 2003a). That means that a line simplification method must ensure that condition too (De Berg et al., 1998). De Berg et al. (1998) proposed a solution to overcome this limitation. They used a more advanced geometric algorithm applied specifically for polygonal subdivision simplification.

On the other hand, there are attempts to achieve more **holistic processes** for polygon generalization. Some of the answers to holistic approaches include constraint-based (Sarjakoski, 2007, p. 22) and database generalization.

The basic idea behind the constraint-based approach is to find a state in which a set of constraints will be met. The focus is not on individual actions but instead on how a goal can be reached (Sarjakoski, 2007, p. 22). Sayidov et al. (2020) are an example of it. Their approach used size constraints to generalize geological maps.

One of the techniques to confront all the constraints at the same time is the agent-based approach (Savino, 2011). In this approach, each object of a geographical database can be modelled as an agent. Each agent has the objective of satisfying a series of constraints that are requirements of the generalization process. An agent can be seen as capable of controlling the decision-making process (Sarjakoski, 2007, p. 23). Galanda (2003a) outlined a framework that uses the multi-agent system for the automated generalization of polygonal maps.

Database generalization, in turn, could be considered a form of model generalization (Patera, 2006). The idea of database generalization is to have a single detailed database from which other data representations at multiple scales would be derived (Sayidov, 2021). One example of it specifically applied to geological maps was approached by Downs and Mackaness (2002). They proposed a rules-based approach for semi-automated generalization to derive a map from a 1:250,000 scale from 1:50,000.

2.2.2 Raster Generalization Approaches

Raster data is a form of tessellation that parts the space into cells. As data structures, raster-based tessellation is simpler than vectors due to the tessellation of space and the implicit adjacency between the cells (Galanda, 2003a, p. 15). One early example of raster generalization is provided by Monmonier (1983) who demonstrated that raster-based generalization of categorical maps is more practical than vectors. He attributed it, among other things, to grid-based generalization algorithms' ability to maintain graphic coherence at smaller scales according to given specifications.

Due to the similarity with image processing imagery, it is notable that many raster algorithms are based on image processing operations (Galanda, 2003a, p. 15). Schylberg (1992), for example, presented the grow and shrink operator to execute amalgamation in categorical data.

Su et al. (1997) described techniques to eliminate small area features and simplify the boundaries of the remaining features. Those techniques are based on operators such as erosion and dilatation which were developed in mathematical morphology. Su et al. (1998) developed a set of morphological models based on mathematical morphology to describe the area to line collapse and partial collapse. They divided the collapse operation into two types: complete collapse and partial collapse, in which the first is subdivided into an area to point and area to line collapse.

Brown et al. (1996) used cost surfaces to describe a map generalization process especially to model when two regions can and cannot be aggregated. For that, each pixel is weighted according to its closeness to the least cost path (optimal path between source pixels) and its distance to the source cells. Thus, source cells will be merged when the area that separates them from each other has a low cost while not merged when instead that area has a high cost.

A further application used the cellular automata (CA) model to generalize geological maps (Smirnoff et al., 2008; Smirnoff et al., 2012). Cellular automata (CA) was first introduced by John von Neumann in the late 1940s (Toffoli & Margolus, 1987). It describes a model consisting of an array of cells arranged in a grid. Each cell has a state which can change over time depending on a set of rules guided by the states of the neighboring cells.

Smirnoff et al. (2012) explained the application of this model in the generalization of geological maps. Initially, the geological map is depicted as a raster image. Then each cell takes a finite number of states that are equivalent to the number of geological units which are represented by a color in Figure 2.3a. The following step is to evaluate each cell according to the rule. In this case, the rule was to check if a cell had the same state as the majority of neighboring cells, for which, if the rule was fulfilled the cell maintained the same state, but if not, the cell changed its state to the most frequent state in the neighborhood. The neighborhood, also called a Moore neighborhood, is a square composed of cells that are considered for the evaluation of the state of the cell in the center. In the end, the process finishes when no further cell alterations will happen (Figure 2.3).

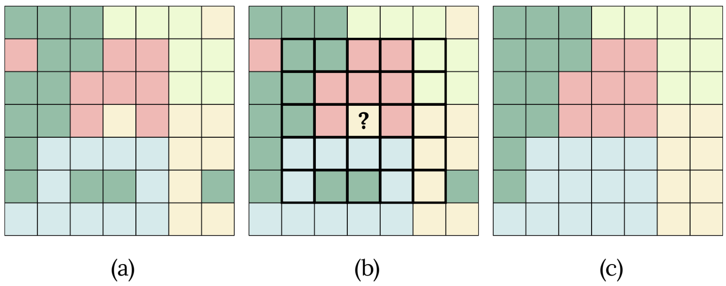


Figure 2.3. Application of cellular automata to the generalization of geological maps. **(a)** A raster image depicting a geological map. **(b)** The cell in the center will take the state of the most frequent state among its neighbors. In this case, the Moore neighborhood is defined as two cells around the cell in the center (R=2). **(c)** Result of the generalization. Adapted from Smirnoff et al. (2012, p. 69).

According to Smirnoff et al. (2008), CA showed a high generalization level and great accuracy, as well as good at meeting the set requirement for the generalization process. They also stated that CA showed important advantages in generalization for geological maps.

2.2.3 Operators to Generalize Categorical Data

McMaster and Shea (1992) proposed a conceptual model that divides the generalization process into three areas: philosophical objectives, cartometric evaluation, and spatial and attribute transformations. These components arise in response to the questions: "why to generalize?", "when to generalize", and "how to generalize?", respectively. The first question is related to why generalization is advantageous. The question when is conducted to identify situations in which generalization is necessary. These situations arise due to the success or failure of the cartographic product to achieve its stated objectives: maintaining clarity, with appropriate content, at a given scale, for a chosen map purpose, and target audience (Shea & McMaster, 1989). The third area is commonly considered as the operators that perform the generalization process (Shea & McMaster, 1989).

Generalization operators describe the spatial and semantic transformations that must be made to reduce the complexity of the categorical coverage during the generalization process (Galanda, 2003a, p. 11; Steiniger & Weibel, 2005, p. 2). In other words, the operator is a description of the type of modification to be applied (Roth et al., 2011, p. 33). Simplification, for example, is an operator to reduce the number of vertices of a polygon boundary. In this sense, the generalization operators provide one or more actions in response to conflicts during scale reduction (Sayidov, 2021, p. 8). Sometimes, a conflict is referred to as a constraint violation (Mackaness, 1994, as cited in Ruas & Plazanet, 1996).

Galanda (2003a) provides a textual and graphical overview of operators used in his work that applies to categorical map generalization (see Figure 2.4). The operators are classified by type between semantic and spatial transformers. The former changes the semantic attributes which alter the underlying statistical properties (Sayidov, 2021, p. 9). The last transform the spatial properties of objects' maps and modify their graphical representation to adapt them to the specific constraints of the map scale (Sayidov, 2021, p. 9). Then a further division is made according to their scope. That is, considering the objects that could potentially be affected by an operator (i.e., a whole map, a group of objects, an object, or an object's boundary). Sequentially, a definition for each is provided.

2.2.4 Algorithms to Generalize Categorical Data

A generalization operator is a description of the type of spatial transformation that is to be achieved when reducing the scale. On the other hand, a generalization algorithm is the implementation of this operator (Regnauld & McMaster, 2007, p. 41). There is a difference between the algorithm approaches implemented for vector and raster data. The vector-based generalization operators rely on geometry and their complexity is due to the necessity of modelling different relationships such as topology and connectivity (Regnauld & McMaster, 2007, p. 41). In general, the generalization of raster-based data is more straightforward (Regnauld & McMaster, 2007, p. 41). One example is the amalgamation operator for raster features (e.g., cost distances). It has more flexibility and its implementation is easier than methods for vector data (Peter & Weibel, 1999, p. 10).

The representation of categorical data can be made in both formats raster and vector. Given the similarity of raster maps with digital images, some of their processing tools have been adapted to

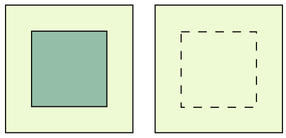
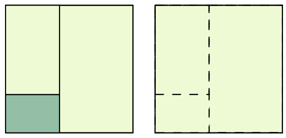
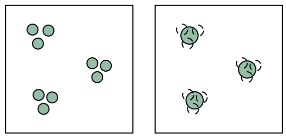
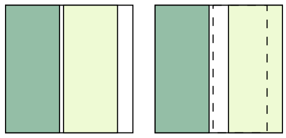
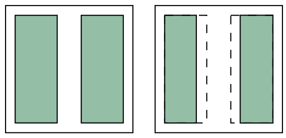
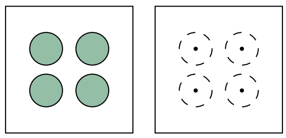
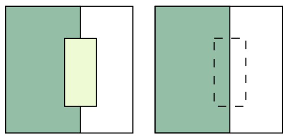
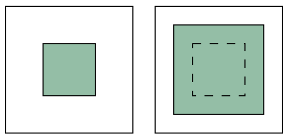
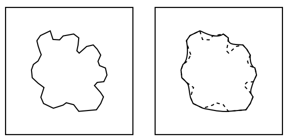
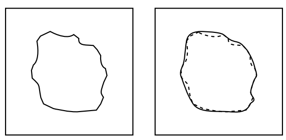
Semantic transformation	Map	**Reclassification**	Changes the category of an object and merges it with the neighboring polygons of the same class.	
Spatial transformation	Group of polygons	**Aggregation**	Merges an object with some others of a similar class into one object.	
		Typification	Decreses the complexity of a group of objects by removing, displacing, enlarging, and aggregating individual objects, but keeping the arrangement.	
	Polygon/group of polygons	**Displacement**	Means the movement of an entire object.	
		Exaggeration	Describes a local increase (decrease) of an object.	
	Polygon	**Collapse**	Collapses a polygon either to a line or to a point.	
		Elimination	Deletes an object from the dataset and the freed space is assigned to another object.	
		Enlargement	Describes a global increase (decrease) of an object.	
	Outline of polygon	**Simplification**	Reduces the granularity of an ouline.	
		Smoothing	Improves the visual appearacnce of an object's outline.	

Figure 2.4. Generalization operators used in categorical maps. Adapted from Figure 2.3 in Galanda (2003a, p. 12)

raster generalization (Regnauld & McMaster, 2007, p. 55). For example, Peter and Weibel (1999, p. 10) described the amalgamation using the cost-distance method in which certain candidate regions are promoted over low-cost regions while high-cost areas remain as barriers (see Figure 8 in Peter and Weibel, 1999).

Conversely, categorical data can be also represented by vectors. Even though the conceptual operations are like in raster format, their implementation is different (Regnauld & McMaster, 2007, p. 56). The followings are two examples of algorithms applied to polygonal map generalization. Bader and Weibel (1997) described several algorithms developed for metric conflict resolution (especially those related to size and proximity conflicts). These algorithms include the elimination of small polygons, enlargement of narrow polygons, aggregation of polygons, and displacement of polygons. Galanda and Weibel (2003) used a snakes-based algorithm for the resolution of violations associated with metric constraints (size and proximity conflicts). It consists of the translation of some or all vertices of the polylines that delimited the polygons in a polygonal subdivision. The generalization operators included in this approach are displacement, enlargement, and exaggeration.

2.2.5 Constraints

Cartographic generalization constraint is understood as the design of specifications for which a map generalization should be respected (Galanda, 2003b, p. 1). According to Galanda (2003a, p. 63), a constraint refers to a specification of a final product regarding an object's property which should be considered by a good generalization process. Unlike rules, constraints are neither linked to a single condition (Ruas & Plazanet, 1996) nor a particular action (Beard, 1991, as cited in Ruas & Plazanet, 1996). Instead, constraints are bound to a synthesis of conditions (Ruas & Plazanet, 1996).

The generalization constraints (Beard, 1991, as cited in Sayidov, 2021) and constraints-based generalization came as a result of the discussions around the design and development of algorithms to meet an intended objective (Weibel & Dutton, 1998, as cited in Sayidov, 2021). The constraint-based approach is a method used to define users' requirements as well as control the generalization process (Stoter, Post, et al., 2014, p. 3). Some examples of the implementation of the idea in the automation of map generalization are as follows: Barrault et al. (2001), Ware et al. (2003), Burghardt and Neun (2006), and Zhang et al. (2013).

There have been several classifications for constraints postulated by different authors. Following are some of them. Beard (1991; as cited in Burghardt et al., 2007) specified four categories: graphic, structural, application, and procedural. Another classification by Ruas and Plazanet (1996) divided the constraints into legibility, shape, spatial, and semantic as shown next:

- → Legibility constraints define the perceptibility thresholds of the objects on a map.
- → Shape constraints ensure the preservation of the shape of both an individual object and a set of objects.
- → Spatial constraints apply to an object (absolute position and relative displacement constraints) and a set of objects (topological and proximity constraints).
- → Semantic constraints are divided into quantity, inter-classes quantity, and functional. (Ruas & Plazanet, 1996)

A further division was made by Weibel and Dutton (1998; as cited in Galanda, 2003a) and Ruas (1999; as cited in Galanda, 2003a). They proposed a constraint classification into four categories: metric, topological, structural, and procedural. The following is a brief explanation of each:

→ Metric (or graphical) constraints define the limits of human perception as minimal dimensions on a map.
→ Topological constraints ensure the maintenance of the polygon topology or its consistency when the polygon topology is modified.
→ Structural constraints attempt to preserve the spatial and semantic structures of the data.
→ Procedural constraints influence the order of generalization operations and the selection of parameters and algorithms. (Galanda, 2003a, p. 65)

Galanda (2003a) presents an extended conceptual discussion of the mentioned constraints (metric, topological, structural, and procedural).

Later, Burghardt et al. (2007) suggested a constraint typology that divides initially into two categories: legibility and preservation of appearance constraints (see Figure 2.5). On the one hand, the legibility constraint is further divided into two types, minimal dimension and removal/emphasize:

→ The former ensures that objects are larger than the minimum size to make sure they are visible.
→ The last one is defined for objects that are less or more semantically relevant or do not meet the minimal dimension or a combination of both. (Burghardt et al., 2007)

On the other hand, the preservation constraints are subdivided into topology, position/orientation, shape, pattern, and distribution/statistics. Following is a brief explanation:

→ Topology constraints ensure the topology (e.g., adjacency and self-intersection).
→ Position/orientation constraints handle translation and rotation.
→ Shape constraints control the objects' deformation.
→ Pattern constraints deal with repetitions of objects or object parts.
→ Distribution/statistics constraints model global effects. (Burghardt et al., 2007)

The two further levels of classification are shown in Figure 2.5 as "geometry type" and "number of objects".

The constraint-based approach was developed in the 1990s and has been mostly implemented for topographic maps (Sayidov, 2021, p. 23). Some of the attempts to implement this idea include Barrault et al. (2001), Ware et al. (2003), Burghardt and Neun (2006), and Zhang et al. (2013). However, there have also been efforts to implement it in categorical maps. One of them is Peter and Weibel (1999) who specified generic cartographic constraints for the generalization process. Edwardes and Mackaness (2000) discussed a constraints-based approach for the automated generalization of categorical maps. Sayidov (2021) attempted to establish a list of size constraints, their measures, and their goal values to ease the generalization process of geological maps. Later, Galanda (2003a) proposed a series of constraints used within an agent-based approach for the automated generalization of categorical maps.

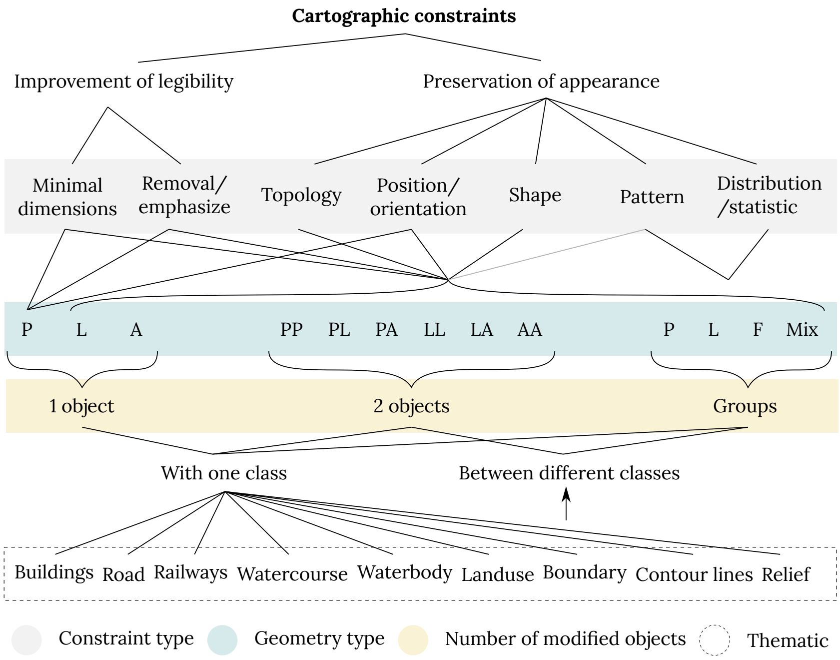


Figure 2.5. Subdivision of typology of constraints suggested by Burghardt et al. (2007). It shows four levels of classification that are from top to bottom: improvement of legibility/preservation of appearance constraints, constraints type, geometry type, and the number of modified objects. The gray line relates the pattern constraint and the objects part of one object. P = point; L = line; A = area. Adapted from Figure 2 in Burghardt et al. (2007)

2.2.6 Levels of Analysis

The map space is organized into levels of analysis or spatial levels. According to Galanda (2003a), these spatial levels are divided into micro, meso, and macro. The micro level refers to individual geographical objects (e.g., a street, a building, etc.) and deals with the generalization of individual objects. The meso level indicates a group of objects (e.g., all the streets of a town) and handles the contextual generalization. The macro level refers to a population of objects (e.g., all the streets of a dataset). This level guides and controls the generalization of populations. These levels of analysis suggest a hierarchical order of the geographic objects which considered together with spatial levels lead to achieving better generalization solutions. (Galanda, 2003a, p. 14)

2.2.7 Evaluation in Generalization

Evaluation in the context of generalization has changed due to the emergence of automatic generalization processes. Since manual generalization and visual assessment are closely related, the error resulting from automated processes did not exist in the manual generalization evaluation. As an example, some of these errors associated with automated solutions are on topology. (Ehrliholzer, 1995, p. 2244)

Generalization quality assessment is defined by Ehrliholzer (1995) as “The application of a set of selected criteria relevant to generalization on a cartographic product that intends to ascribe quantitative or qualitative attributes for which the rating scale has to be transformed in order to get a quality judgement that combines both quantitative and qualitative aspects.” (p. 2243). The criteria are defined as specifications of quality factors that are the cartographic requirements for generalization. In a constraint-based approach, the criteria could be equivalent to the cartographic constraints. (Schmid, 2008, p. 24).

Mackaness and Ruas (2007) consider three aspects of automated generalization evaluation: algorithmics, informatics, and cartographic evaluation. The first one, algorithmics, focuses on the measurement of the efficiency and robustness of the generalization algorithms. The second, informatics, deals with information handling such as the user’s specification as well as integration in the evaluation phase. The last one, cartographic evaluation, deals with the optimization of presentation forms by modelling symbolization, abstraction, and aesthetics (Mackaness & Ruas, 2007, p. 90). In turn, cartographic evaluation, covers features evaluation before, during, and after map generalization.

According to Mackaness and Ruas (2007), the evaluation process could be carried out at object, group, or population levels. Micro-evaluations are the assessments made at the object level in which one non-generalized object is compared to one generalized object. Meso-evaluations occur at the group level. Therefore, a group of non-generalized objects is compared with a group of generalized objects. As an example, a specific area of a map can be compared with respect to the arrangement of geological units before and after the generalization process. Macro-evaluations happen at the dataset level. It means that the comparison considers all the data of a specific type before and after the generalization. Therefore, macro-evaluations are used mainly to quantify the objects before and after the process. It is in terms of number, length, or area. (Mackaness & Ruas, 2007, p. 103)

Quality evaluation of generalization can be made using two approaches: manual or computational (Schmid, 2008, p. 24). The first one refers to visual evaluation. This type of evaluation is subjective and qualitative (Mackaness & Ruas, 2007, p. 98). On the other hand, computational refers to a quantitative evaluation that includes the measurements and evaluation of transformations during the generalization process (Schmid, 2008, p. 25).

Mackaness and Ruas (2007) differentiated three types of evaluation:

→ *Evaluation for tunning* that occurs before the generalization process (i.e., finding the optimal set of parameters).
→ *Evaluation for controlling* happens during the generalization.
→ *Evaluation for assessing* the quality of the generalized data. It is subdivided into:
  - *Evaluation for editing* that aims to detect errors and inconsistencies in either object or group of objects that were not correctly generalized.
  - *Descriptive evaluation* aims to characterize the generalized data. For that, it provides information regarding the objects that were removed, emphasized, or altered.
  - *Evaluation for grading* tries to grade the generalization process with a unique value that reveals the quality of the generalized data which can be useful when compared

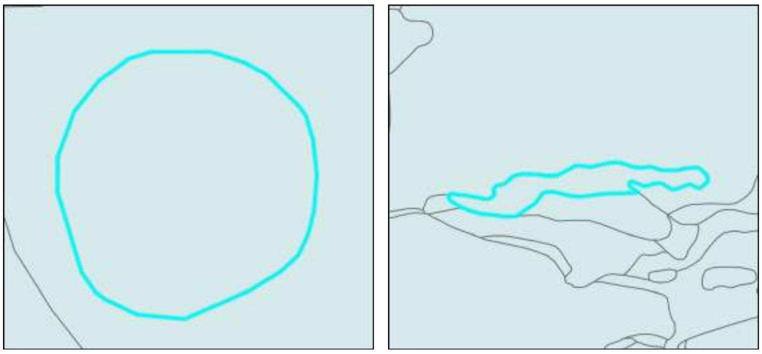
with other results of generalization with different parameters. (Mackaness & Ruas, 2007)

2.3 Polsby-Popper

The Polsby-Popper (Polsby & Popper, 1991) measures compactness. Compactness is used to assess geographic gerrymandering, which refers to the action of manipulating the boundaries of voting districts to take political advantage ("Measuring Compactness", n.d.). The Polsby-Popper (PP) score is the ratio of the area of district (A) to its perimeter (P), which is the area of a circle whose circumference is equal to the perimeter of the district ("Measuring Compactness", n.d.).

$$PP = \frac{4\pi A}{P^2}$$

The Polsby-Popper (PP) score can take a value between the range [0,1] in which a value of one indicates a circle. Figure 2.6a shows a highlighted polygon with a PP of 0.98. Figure 2.6b displays a highlighted polygon with a PP of 0.15.



(a) (b)

Figure 2.6. PP score in two different polygons. **(a)** A polygon with PP 0.98. **(b)** A polygon with PP 0.15.

3 Methodology

This chapter presents the state of geological mapping at the LfU and the methodology used for generalizing geological maps.

Section 3.1 describes the current state of the geological cartography at the LfU, hence, why an automated process to improve the production lines is needed. Section 3.2 explains the two software used for the workflow. Section 3.3 describes the step-by-step procedure used for generalization. Section 3.4 outlines how the evaluation was carried out. Finally, section 3.5 explains how the Polsby-Popper score was measured.

3.1 Current State of the Geological Mapping at the Bayerisches Landesamt für Umwelt

The LfU has geological mapping at scales 1:25,000 (GK25), 1:50,000 (GK500), 1:75,000 to 1:200,000, 1:500,000, and 1:1,500,000. However, the only maps that cover Bavaria completely are the 1:25,000, 1:500,000, and 1:1,500,000 scale maps. The 1:25,000 maps represent the current state of knowledge of the LfU Geological Survey, which can be accessed from the LfU website [1]. The 1:500,000 scale map was manually generalized in 1996 by K. Doben, G. Doppler, W. Freudenberger, H. Jerz, R. K. F. Meyer, H. Mielke, W. -D. Ott, J. Rohrmüller, H. Schmidt-Kaler, K. Schwerd, and H. J. Unger (see Figure 3.1).

3.2 Overview of the Software Used in this Workflow

3.2.1 FME

FME is a platform used for data integration. In this regard, it means bringing together data from different sources and creating new datasets. Creating FME workflow involves reading the data, transforming it by changing or restructuring it, and writing it.

There are many functions available in FME with which reading and transforming are carried out. That is why FME supports a huge number of formats. Some of them are Esri geodatabase, Esri shapefile, GeoTIFF, PostgreSQL, GeoJSON, and HTML, among others. Besides, there are also numerous transformers for multiple purposes. An example applied to spatial analysis is the SpatialFilter with which point, line, area, and text are filtered according to spatial relationships. Another function is the ContourGenerator to create contour lines from raster or points data.

[1] https://www.lfu.bayern.de/geologie/geo_karten_schriften/dgk25_uab/index.htm

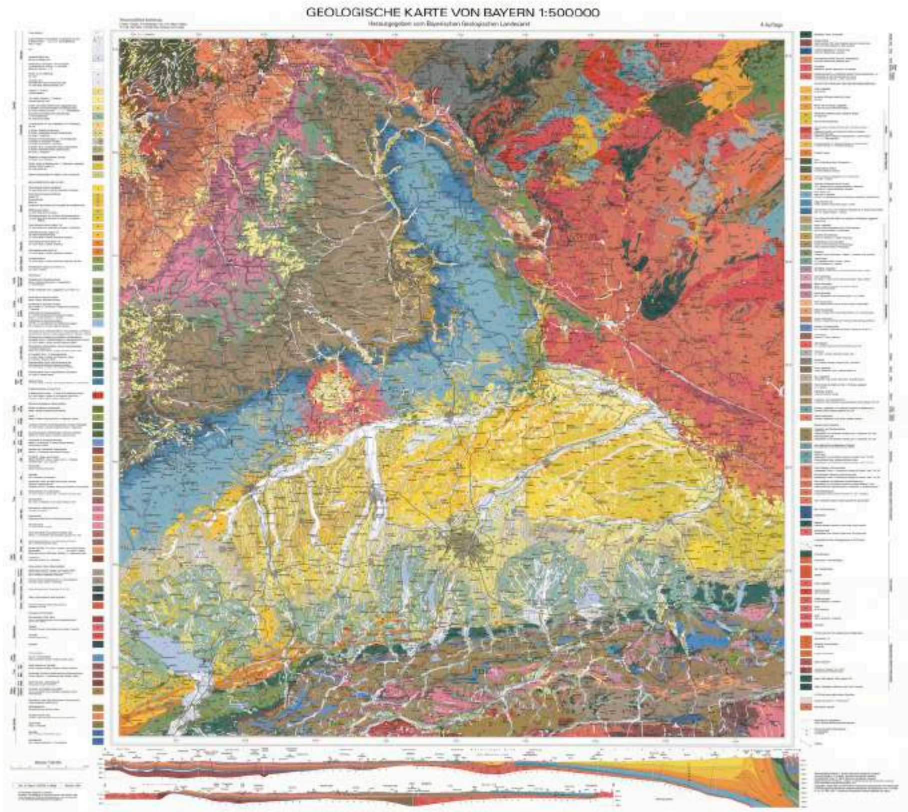


Figure 3.1. Geological map at scale 1:500,000. Taken from the Bayerisches Landesamt für Umwelt (n.d-b).

3.2.2 GeoScaler

GeoScaler is a free software developed for the generalization of surficial and bedrock geology maps at the LCNP of the Geological Survey of Canada. It was initially available as a plug-in for ArcGIS in 2009, which required ArcMap 9.3 with ArcInfo license and ArcInfo Workstation (Smirnoff et al., 2012, p. 69). Later in 2012, a second improved edition was released under ArcGIS 10.0 and 10.1 environments. Some of the requirements include an ArcInfo (Advanced) license and the Spatial Analysis extension. (Huot-Vézina et al., 2012, p. 15)

Two separate modules are available to carry out the generalization process on both surficial and bedrock data. The first one is aimed at executing the generalization of surficial formations (Quaternary) which is divided into two models: polygons and symbols. The polygon model includes data preparation, generalization, postprocessing, and polygon-to-symbol conversion (Smirnoff et al., 2012, p. 72). Along with it, surficial symbol generalization starts with separation according to symbol representation (polygon, point, and line) follows by a specific generalization procedure for each layer (Smirnoff et al., 2012, p. 73). The second module is for the generalization of bedrock in which two considerations are relevant. On the one hand, the generalization of boundaries between units has to take special attention to the presence of fault lines because it limits the freedom with which this process is carried out. On the other hand, if there are special features, they might be preserved regardless of the target scale (e.g., dykes). (Smirnoff et al., 2012)

3.3 Workflow

The LfU provided the workflow for its analysis and evaluation. This workflow was proposed at its base by Schuff (2019) and further developed by Landesamt für Geologie, Rohstoffe und Bergbau (LGRB) in cooperation with the company con terra GmbH (Münster). This workflow provides the generalization of categorical data focused on areal objects based on geometric and semantic criteria. Figure 3.2 displays a schematic overview in which the terms bedrock and surficial correspond to the definition given in the context of GeoScaler. The former focuses on bedrock geology while the latter refers to surficial formations (Quaternary) (Huot-Vézina et al., 2012). The individual parts are described one by one in the following sections.

3.3.1 Data Enrichment

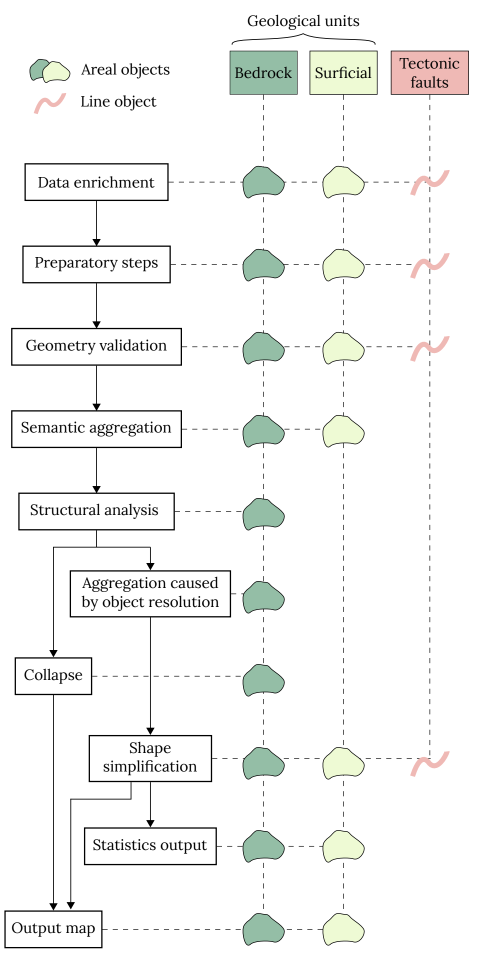


Figure 3.2. Diagram of the generalization process. To the left is the step-by-step workflow. On the right is the data, either polygons or lines, for which each step applies.

The data enrichment phase provides the necessary information to carry out the generalization process. In this step, the consistency and completeness of the geological and tectonic data, as well as the generalization parameters, are verified. The last ones are of particular importance as they will define the values for the constraints that will determine the characteristics of the output.

An example that applies to both geological units and tectonic faults is the parameter referred to here as “flag”. This “flag” parameter provides information about the relevance of a specific object for the target scale. Figure 3.3 shows an example that applies to tectonic faults. The green symbol lines (value 0) represent tectonic faults to be preserved at all generalization scales (1:25,000, 1:150,000, 1:500,000, and 1:1,000,000). Red lines (value 1) represent faults to be preserved at the 1:25,000, 1:150,000, and 1:500,000 generalization scales. The blue lines (value 2) represent those faults to be preserved at scales 1:25,000 and 1:150,000. Finally, there are the orange lines (value 3) represent faults relevant to the 1:25,000 scale. This same range of values from 0 to 3 and their corresponding target scales also apply to the geological units.

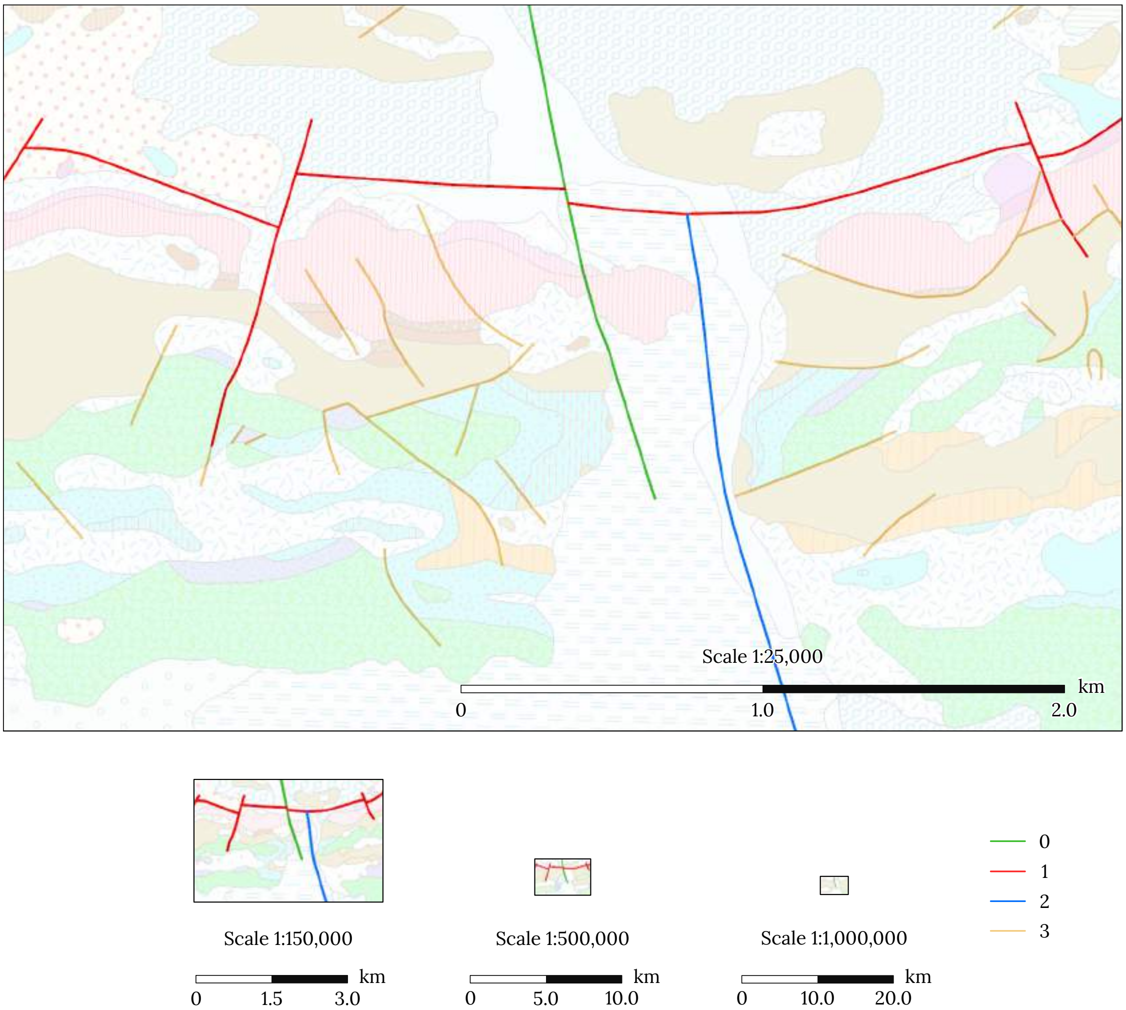


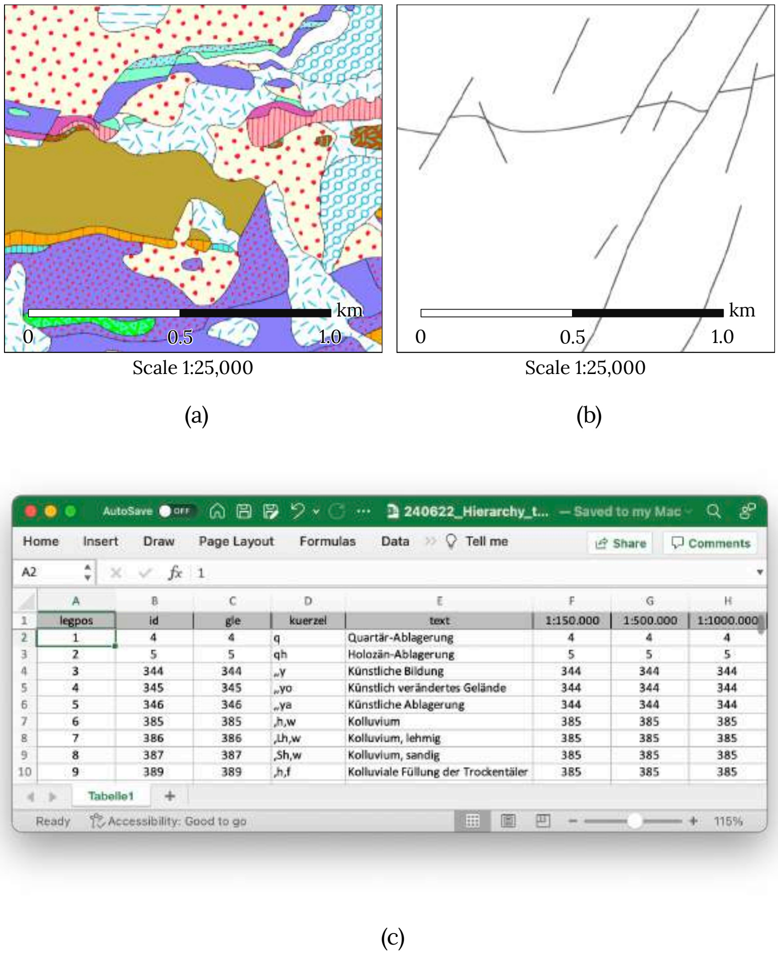
Figure 3.3. Attribute “flag” applied to tectonic faults.

3.3.2 Preparatory Steps

Preparatory steps are carried out at the beginning of the workflow. In this phase, geological and tectonic information is transferred to the workflow and filtered according to the initial parameters. The initial input data is a geodatabase (GDB) containing the geological units at 1:25,000 scale, tectonic faults, a polygon of the boundary of the study area, and the hierarchy table (see Figure 3.4). The parameters are map type and target scale (see Tables 3.1 and 3.2).

Table 3.1. Type of map parameter.

Parameter	Choices
Map type	Bedrock or surficial



	legpos	id	gle	kuerzel	text	1:150.000	1:500.000	1:1000.000
2	1	4	4	q	Quartär-Ablagerung	4	4	4
3	2	5	5	qh	Holozän-Ablagerung	5	5	5
4	3	344	344	„y	Künstliche Bildung	344	344	344
5	4	345	345	„yo	Künstlich verändertes Gelände	344	344	344
6	5	346	346	„ya	Künstliche Ablagerung	344	344	344
7	6	385	385	‚h,w	Kolluvium	385	385	385
8	7	386	386	‚Lh,w	Kolluvium, lehmig	385	385	385
9	8	387	387	‚Sh,w	Kolluvium, sandig	385	385	385
10	9	389	389	‚h,f	Kolluviale Füllung der Trockentäler	385	385	385

Figure 3.4. Required data to start the workflow. **(a)** Geological units at scale 1:25,000. **(b)** Tectonic faults at 1:25,000. **(c)** Hierarchy table.

Table 3.2. Initial scale and target scales. Each target scale represents the smallest value within the range target scale.

Initial scale	**Target scale 1** [1:70,000–1:150,000]	**Target scale 2** [1:150,000–1:500,000]	**Target scale 3** [1:500,000–1:1,000,000]
1:25,000	1:150,000	1:500,000	1:1,000,000

3.3.3 Geometry Validation

In this phase, problems in the geometry of geological units and tectonic faults are detected according to the Open Geospatial Consortium (OGC) geometry specifications. Each feature is processed individually in this phase, ensuring that faulty geometries do not abort the further processes.

3.3.4 Semantic Aggregation

This first process of polygon aggregation is performed according to the semantic characteristics of the units. The framework uses the information in the hierarchy table (Figure 3.4c) to accomplish this process. Each polygon is associated with a geological unit, equivalent to a numeric value in the hierarchy table. In this phase, each polygon converts the numerical value to the equivalent value of the unit in the target scale. In other words, if the target scale is 1:150,000, then the initial value of each polygon changes to the value that corresponds to it at the 1:150,000 scale according to the hierarchy table. Thus, at the moment of aggregation, each polygon is evaluated with its neighbors for semantic similarity. Adjacent polygons sharing the same value are then aggregated.

3.3.5 Structural Analysis

The structural analysis is divided into several parts. The first part is a geometric characterization in which the area and circularity of each polygon are calculated. After this, there is a data separation between surficial and bedrock geology. The second phase corresponds to the estimation of the number of neighbors for each polygon in the study area. In the third phase, the data is filtered according to the minimum area value (see Table 3.3), keeping the objects whose areas are less than the minimum required. In the fourth phase, the polygons are filtered by elongated and non-elongated objects. The elongation of the polygons is calculated based on the area/length ratio. Finally, the polygons adjacent to the polygons under the minimum required area that have semantic similarity are selected and kept.

3.3.6 Aggregation Caused by Object Resolution

The aggregation considers both the semantic similarity of the polygons concerning their geological age and the area of the polygons. First, it evaluates if the units correspond to the same hierarchical level. In the case of more than two neighbors, the neighbor with the largest area is chosen. In the case of only two polygons, A and B, sharing semantic characteristics, it is evaluated which is the larger polygon between A and B. Thus the smallest polygon is added to the largest one.

Subsequently, the geological data (bedrock and surficial) are joined with the geological faults in a geodatabase to process it in shape simplification.

Table 3.3. Minimum area according to each target scale.

Target scale	Minimum area (m^2)	Minimum area in map units (mm^2)
1:150,000	202,500	9
1:500,000	2,250,000	9
1:1,000,000	9,000,000	9

3.3.7 Collapse

In this phase, the polygons that meet the following specifications:

(a) are under the minimum accepted area,
(b) have at least one neighboring polygon with semantic similarity to which be aggregated, and
(c) are relevant to the target scale ("flag" parameter)

will be transformed into lines or points according to their geometry.

3.3.8 Shape Simplification

In this step, the surficial and bedrock data are separated for further processing in GeoScaler according to the steps mentioned for each one in section 3.2.2. In this phase, the shape simplification process is performed using cellular automata (CA) as explained in section 2.2.2 and according to the parameters specified in Table 3.4. Tectonic faults play an important role in this phase as they will remain invariant areas. This means that the area, which corresponds to the number of cells specified as an "unaltered radius around faults" (see Table 3.4), will not change. This is to maintain the coherence between the geological units and mainly the faults that are boundaries between units.

Table 3.4. Parameters for GeoScaler.

Parameter	Possible values
Moore's neighborhood radius	[1, 2, 3, 4, 5]
Number of iterations	[1–255]
Unaltered radius around faults	[1–5]

3.3.9 Statistics Output

The statistical output is used as an evaluation for the generalization process and as a basis for further data enrichment. This file records the existing objects before and after object resolution and shape simplification, respectively. It provides information for each geological unit about the number of objects per unit and the area covered, among others. This information is relevant to evaluate, for example, if it is worth it to preserve a geological unit that disappears during the process. If so, the "flag" parameter is used to ensure the unit collapse into a line or point. Otherwise, there would be no need to provide additional information.

3.3.10 Output Map

The result will be a geodatabase that includes the generalized polygons corresponding to the geological units. Additionally, the points and lines from the result of the collapse.

3.4 Evaluation of the Results

A quantitative evaluation tends to examine only certain characteristics and thus evaluate specific requirements, which is difficult to integrate into an automatic holistic process (Stoter, Zhang, et al., 2014) such as the one analyzed here. Given this, a qualitative visual assessment was performed after the generalization process to evaluate the quality of the generalized maps of the study area (see section 4.1). This evaluation consisted of delivering the outputs (PDFs and GDBs) (see appendix A) to LfU experts who visually examined their readability and detected errors and inconsistencies.

3.5 Polsby-Popper Measurement

To characterize the change in the polygons' compactness during the generalization process, the Polsby-Popper was calculated. The mean value of the PP score was measured for all polygons in the input and output data. For the latter, the value was estimated at three different numbers of iterations (10, 80, 150).

4 Implementation

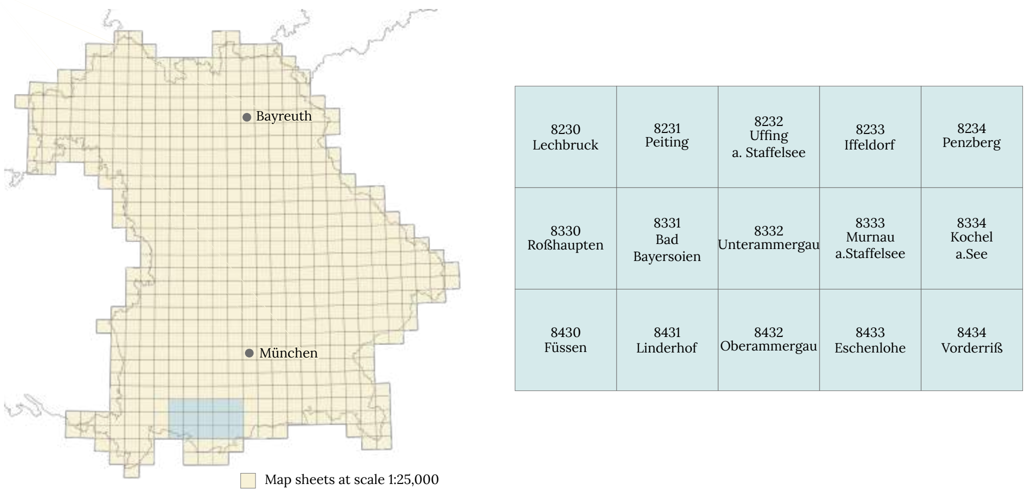
This chapter presents the workflow implementation using images and text that describe the changes per step.

Section 4.1 shows the study area and the criteria for why it was selected. Section 4.2 describes the required software specifications, while section 4.3 describes the step-by-step generalization procedure. Section 4.4 explains briefly how the evaluation was in practice. Finally, section 4.5 describes how the Polsby-Popper score was calculated.

4.1 Study Area

The study area chosen for the implementation of the workflow corresponds to the area to the south of Bavaria, Germany, shown in Figure 4.1a in blue. This area comprises a total of 15 maps at 1:25,000 which can be seen in Figure 4.1b. The criteria for choosing this region were

(i) the completeness of the input data regarding especially the parameter “flag” related to geological faults and
(ii) the geological complexity that this area represents makes it interesting to analyze.



(a) (b)

Figure 4.1. Distribution of map sheets at 1:25,000 scale in Bavaria. **(a)** The location of the study area is highlighted in blue. **(b)** Distribution of the sheets at 1:25,000 scale for the study area.

To demonstrate the implementation of the workflow, examples are taken from different zones within the study area to show the operations carried out at each step.

4.2 Software Specifications

Below is a list of the software required for workflow implementation and its specific features.

→ FME 2019.1 or higher
→ ArcGIS Desktop Professional 10.7 or higher
→ ArcGIS Desktop-Extension: Spatial Analysis
→ Python 2.7

For the correct operation of the workflow, it is necessary to have ArcMap or ArcGIS Pro open, as well as to set ArcGIS administrator to “Professional”.

4.3 Implementation of the Workflow

In the workflow in FME (Figure 4.2) several parameters determine the specifications of the output map. These parameters can be configured from a selection list at the start or incorporated as part of the input data (see Figure 4.3). The values used for the parameters in the example presented in this chapter are described in Table 4.1. These values are taken from specifications set by LfU experts (i.e., the minimum area in square meters), Schuff (2019), and Dupke et al. (2021).

Figure 4.2. Workflow in FME.

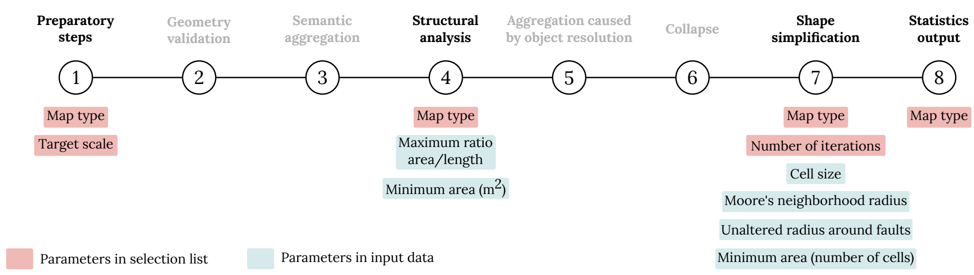


Figure 4.3. Parameters used in the workflow. Each parameter is located in the phase in which it is used. The pink boxes indicate the parameters configure as a selection list at the start. The blue boxes define the parameters previously set up and entered as input data.

Table 4.1. Values of the parameters described in Figure 4.3.

Parameter	Possible values
Map type	Bedrock
Target map	1:150,000
Number of iterations	80
Maximum ratio area/length	30
Minimum area (m^2)	202,500
Cell size	10
Moore's neighborhood radius	2
Unaltered radius around faults	2
Minimum area (number of cells)	200

4.3.1 Preparatory Steps

The first three filters affecting the input data include the type of map, the target scale, and the number of iterations for the cellular automata process. For this thesis, the data processing and analysis focused on bedrock data, so it was the map type chosen at the beginning of the processing. The target scales, 1:150,000, 1:500,000 (GK500), and 1:1,000,000, were pre-established by LfU as being of interest for the automation process of the production lines in that institution.

4.3.2 Geometry Validation

The process involved in this step is described in the section 3.3.3.

4.3.3 Semantic Aggregation

Each unit of the original scale (1:25,000) becomes its corresponding parental geological unit according to the target scale (1:150,000, 1:500,000, or 1:1,000,000). In this case, the 1:150,000 scale parental unit is taken into account. Figures 4.4a and 4.4b show the units at scales 1:25,000 and 1:150,000, respectively. As an example, units 360, 366, and 367 become 355 at a 1:150,000 scale as shown in Figure 4.4 (see also Table 4.2). The dashed line in Figure 4.4b shows the remnant or where the amalgamated unit was.

4.3.4 Structural Analysis

This phase produces two results. On the one hand, it categorizes the polygons between elongated and non-elongated. On the other hand, it selects the polygons with an area less than nine mm^2 in map units (their equivalent according to the scale is in Table 3.3). Additionally, it selects those polygons with an area less than the minimum area and that have one or more neighboring polygons with similar semantics.

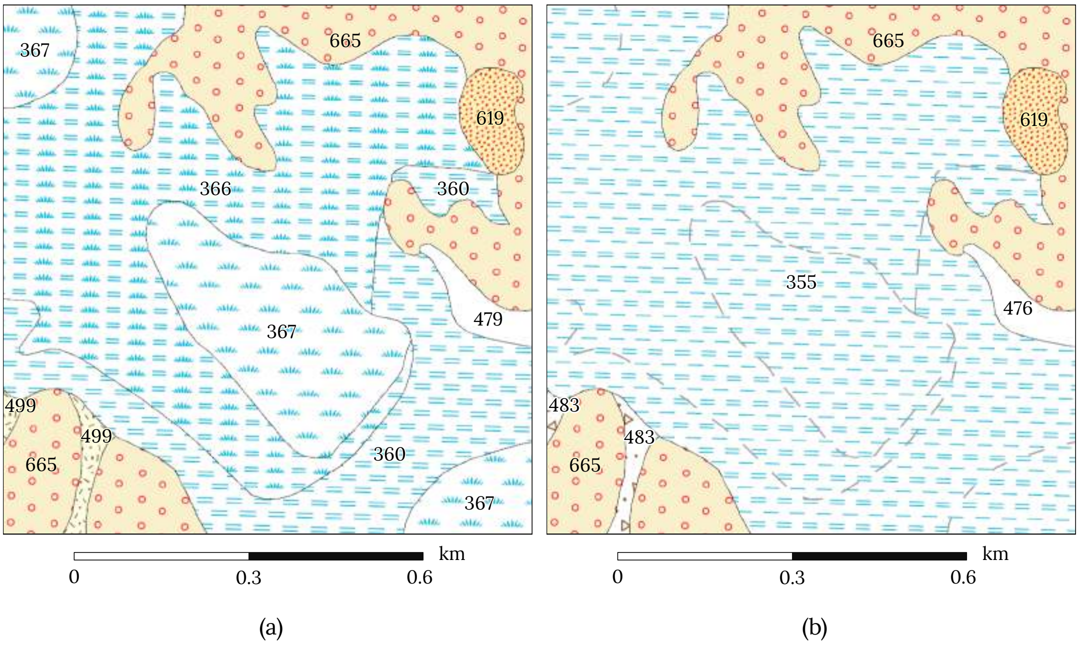


Figure 4.4. Semantic aggregation step. **(a)** Before. **(b)** After.

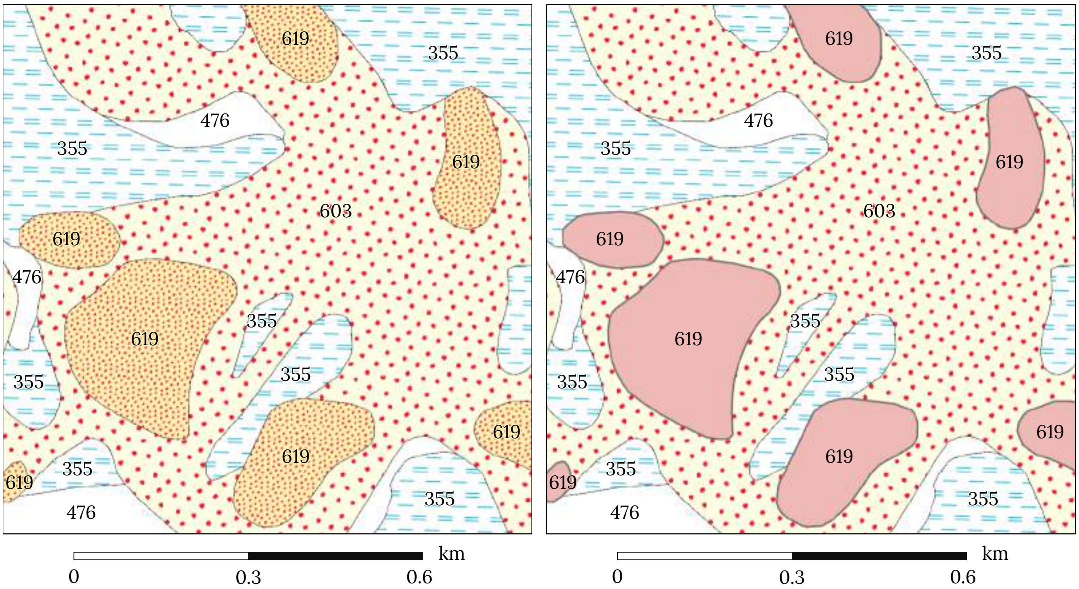
Table 4.2. Hierarchy table.

Unit at scale 1:25,000	Parental unit at scale 1:150,000	Parental unit at scale 1:500,000	Parental unit at scale 1:1,000,000
360	355	355	355
366	355	355	355
367	355	355	355
479	476	476	476
499	483	483	483
619	619	609	609
665	665	665	599

Figure 4.5a shows the map at the beginning of the process. Figure 4.5b shows in pink the units with an area less than the minimum area and that have at least one neighbor with similar semantics as shown in Figure 4.5c. For example, unit 619 (see Figures 4.5a and 4.5c) shares semantic similarity with its neighboring unit 603 since at the next generalization scale (1:500,000) both would become 609.

4.3.5 Aggregation Caused by Object Resolution

The second aggregation of polygons is made based on the information collected in the previous phase. For that purpose, two aspects are considered (i) the similarity with the next parental level in the geological hierarchy and (ii) the area of the candidate polygons to be aggregated.



(a) (b)

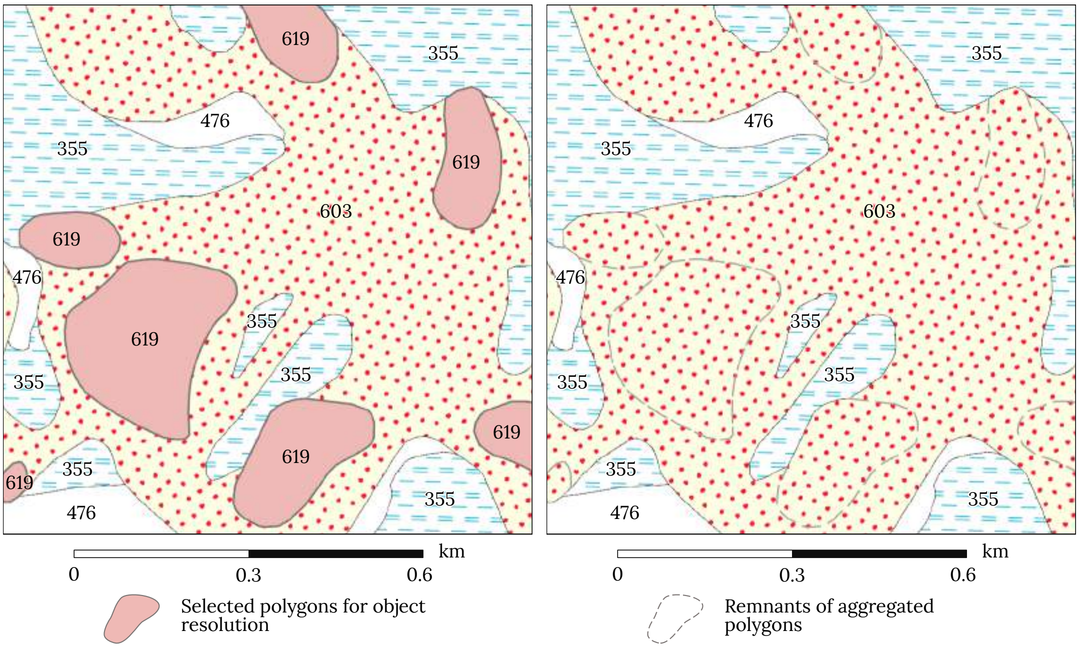
Unit at scale 1:150,000	Parental unit at scale 1:500,000
355	355
476	476
603	609
619	609

Selected polygons for object resolution

(c)

Figure 4.5. Structural analysis step. **(a)** Before. **(b)** After. **(c)** Equivalences of units according to scale.

Figure 4.6a shows the candidate polygons selected in the previous phase. Figure 4.6b displays the result of this aggregation process in which the dashed lines represent where the aggregated polygons were.



(a) (b)

Figure 4.6. Aggregation caused by object resolution step. **(a)** Before. **(b)** After.

4.3.6 Collapse

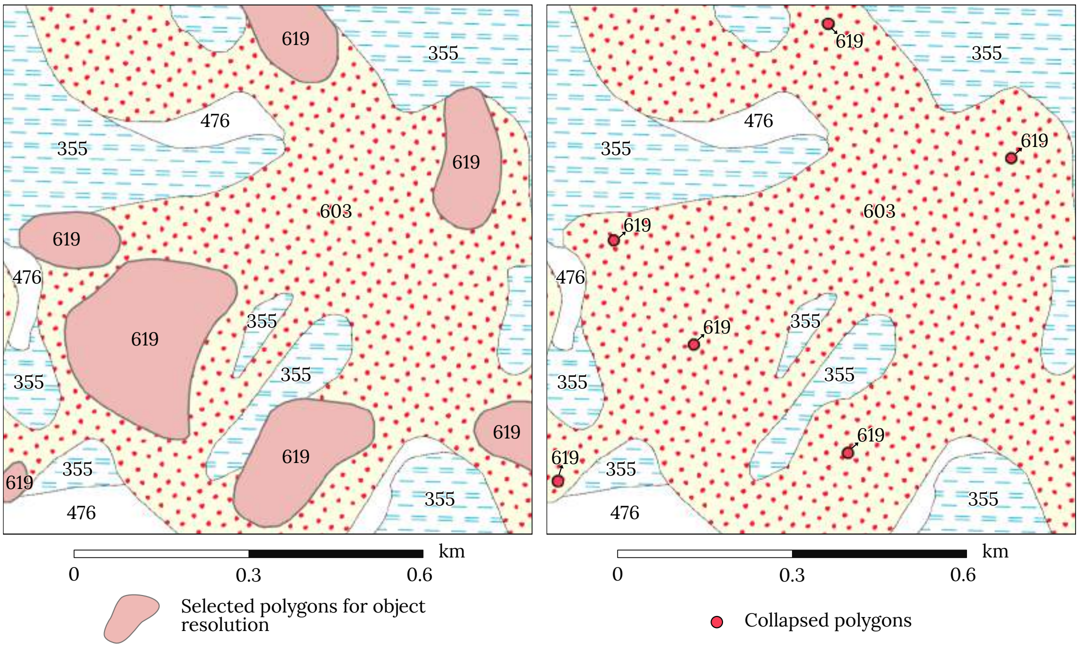
Polygons marked with the “flag” attribute for the target scale with at least one neighboring polygon with semantic similarity collapse to points or lines according to their elongation. The attribute “flag” applies similarly in polygons than in faults. The explanation given in section 3.3.1 about the “flag” attribute also applies here. Figure 4.7a shows the polygons in pink with an area less than the minimum area, attribute “flag” of 0, 1, or 2, and at least a neighboring polygon with semantic similarity. Figure 4.7b shows the collapse of polygons to points.

4.3.7 Shape Simplification

This phase is carried out in raster format in GeoScaler. Figure 4.8a shows the polygons before shape simplification. Figure 4.8b shows the polygons after shape simplification with 80 iterations as described in Table 4.1.

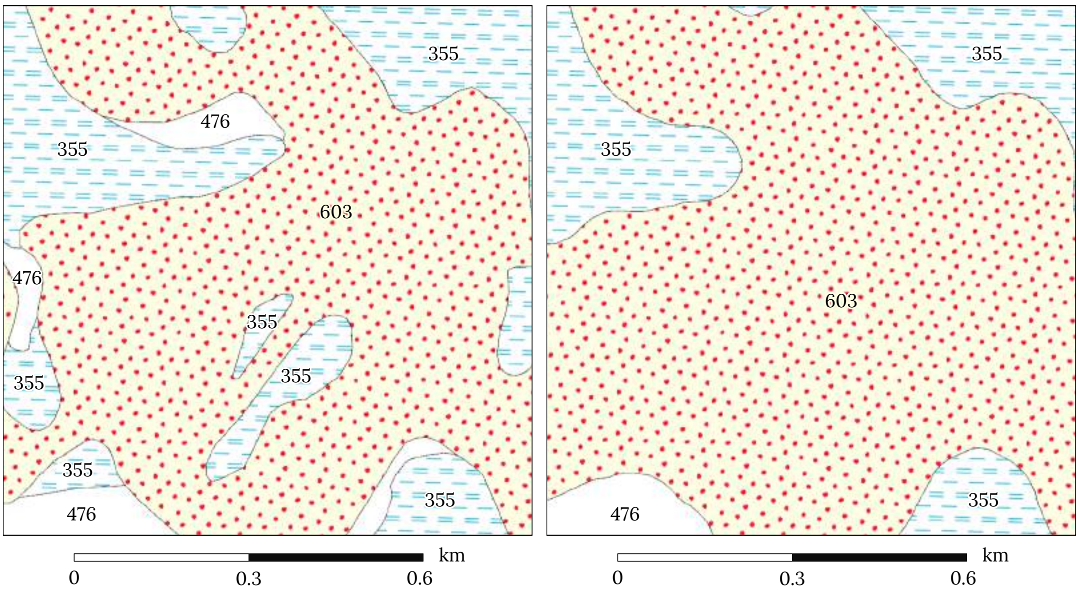
4.3.8 Statictics Output

Figure 4.9 shows the statistics obtained from the object resolution and shape simplification processes. This step serves for further data enrichment as it shows, for example, the number of polygons or the area of polygons before and after these two processes (Figure 4.9a). For example, some units in the second column in Figure 4.9b have a “None” value which means those units disappeared.



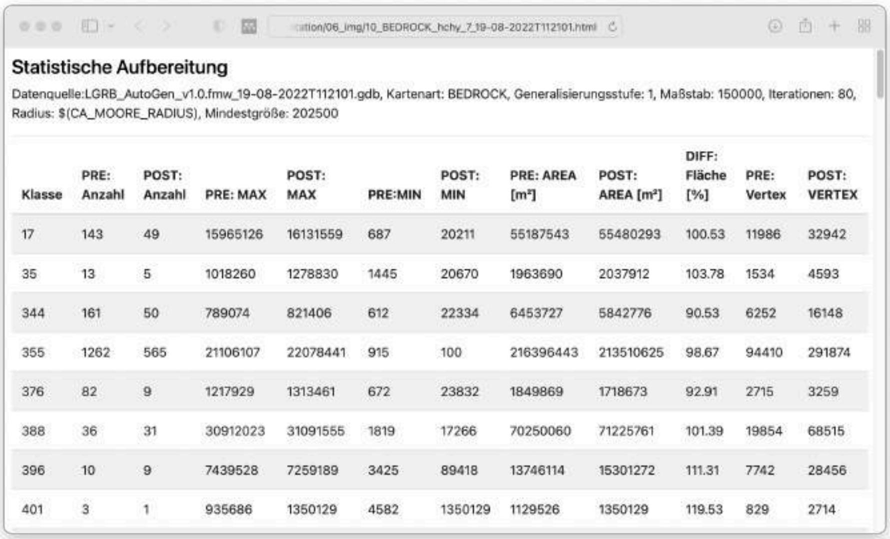
(a) (b)

Figure 4.7. Collapse step. **(a)** Before. **(b)** After.



(a) (b)

Figure 4.8. Shape simplification step. **(a)** Before. **(b)** After.

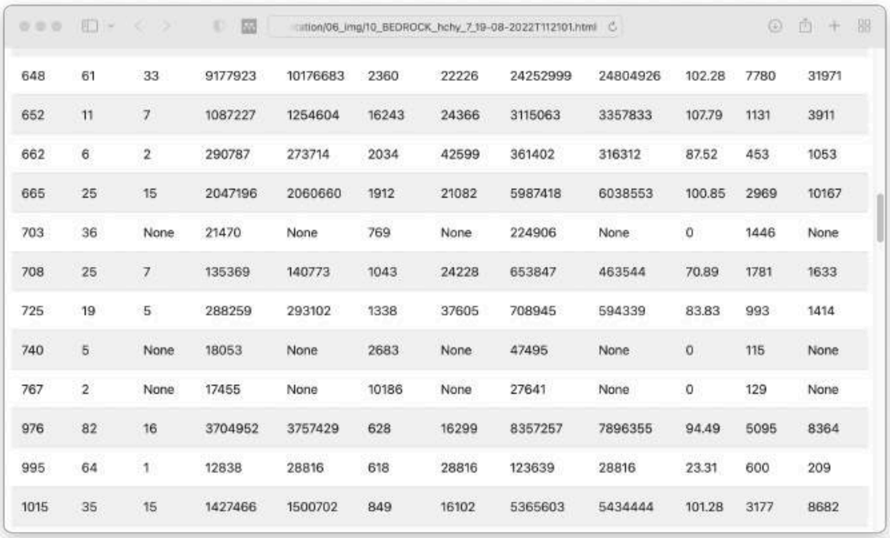


Statistische Aufbereitung

Datenquelle:LGRB_AutoGen_v1.0.fmw_19-08-2022T112101.gdb, Kartenart: BEDROCK, Generalisierungsstufe: 1, Maßstab: 150000, Iterationen: 80, Radius: $(CA_MOORE_RADIUS), Mindestgröße: 202500

Klasse	PRE: Anzahl	POST: Anzahl	PRE: MAX	POST: MAX	PRE:MIN	POST: MIN	PRE: AREA [m²]	POST: AREA [m²]	DIFF: Fläche [%]	PRE: Vertex	POST: VERTEX
17	143	49	15965126	16131559	687	20211	55187543	55480293	100.53	11986	32942
35	13	5	1018260	1278830	1445	20670	1963690	2037912	103.78	1534	4593
344	161	50	789074	821406	612	22334	6453727	5842776	90.53	6252	16148
355	1262	565	21106107	22078441	915	100	216396443	213510625	98.67	94410	291874
376	82	9	1217929	1313461	672	23832	1849869	1718673	92.91	2715	3259
388	36	31	30912023	31091555	1819	17266	70250060	71225761	101.39	19854	68515
396	10	9	7439528	7259189	3425	89418	13746114	15301272	111.31	7742	28456
401	3	1	935686	1350129	4582	1350129	1129526	1350129	119.53	829	2714

(a)



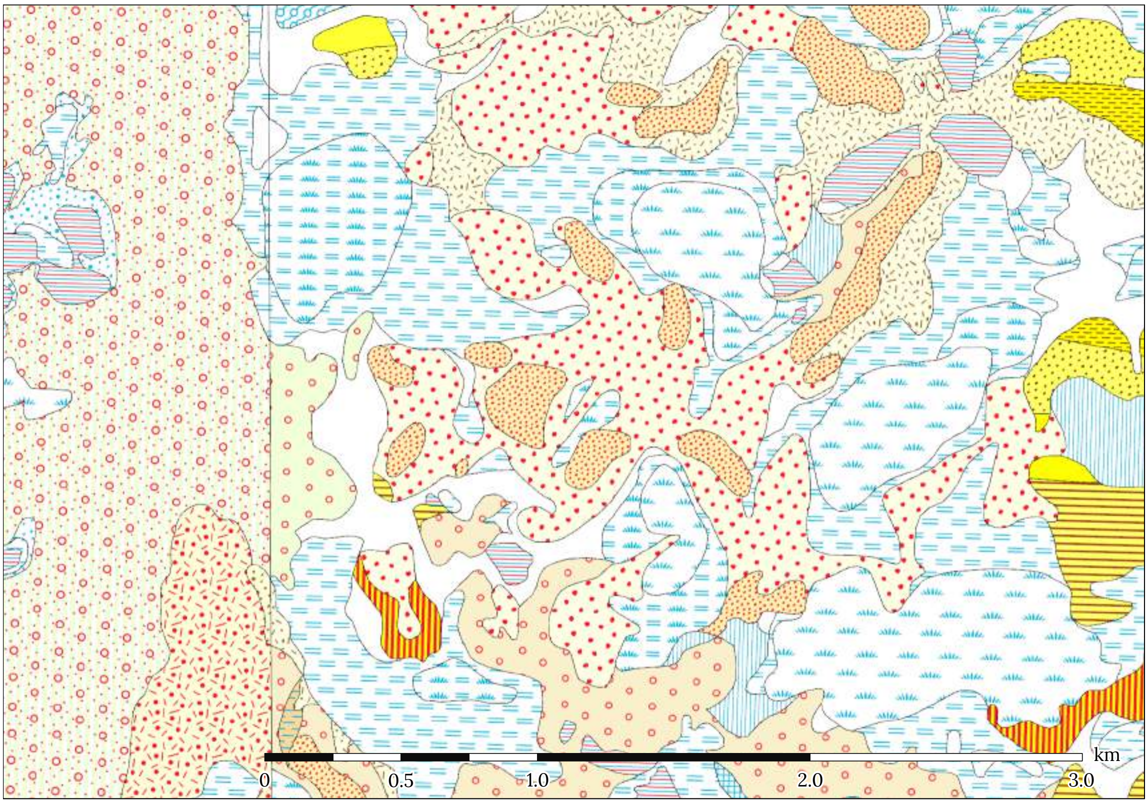
648	61	33	9177923	10176683	2360	22226	24252999	24804926	102.28	7780	31971
652	11	7	1087227	1254604	16243	24366	3115063	3357833	107.79	1131	3911
662	6	2	290787	273714	2034	42599	361402	316312	87.52	453	1053
665	25	15	2047196	2060660	1912	21082	5987418	6038553	100.85	2969	10167
703	36	None	21470	None	769	None	224906	None	0	1446	None
708	25	7	135369	140773	1043	24228	653847	463544	70.89	1781	1633
725	19	5	288259	293102	1338	37605	708945	594339	83.83	993	1414
740	5	None	18053	None	2683	None	47495	None	0	115	None
767	2	None	17455	None	10186	None	27641	None	0	129	None
976	82	16	3704952	3757429	628	16299	8357257	7896355	94.49	5095	8364
995	64	1	12838	28816	618	28816	123639	28816	23.31	600	209
1015	35	15	1427466	1500702	849	16102	5365603	5434444	101.28	3177	8682

(b)

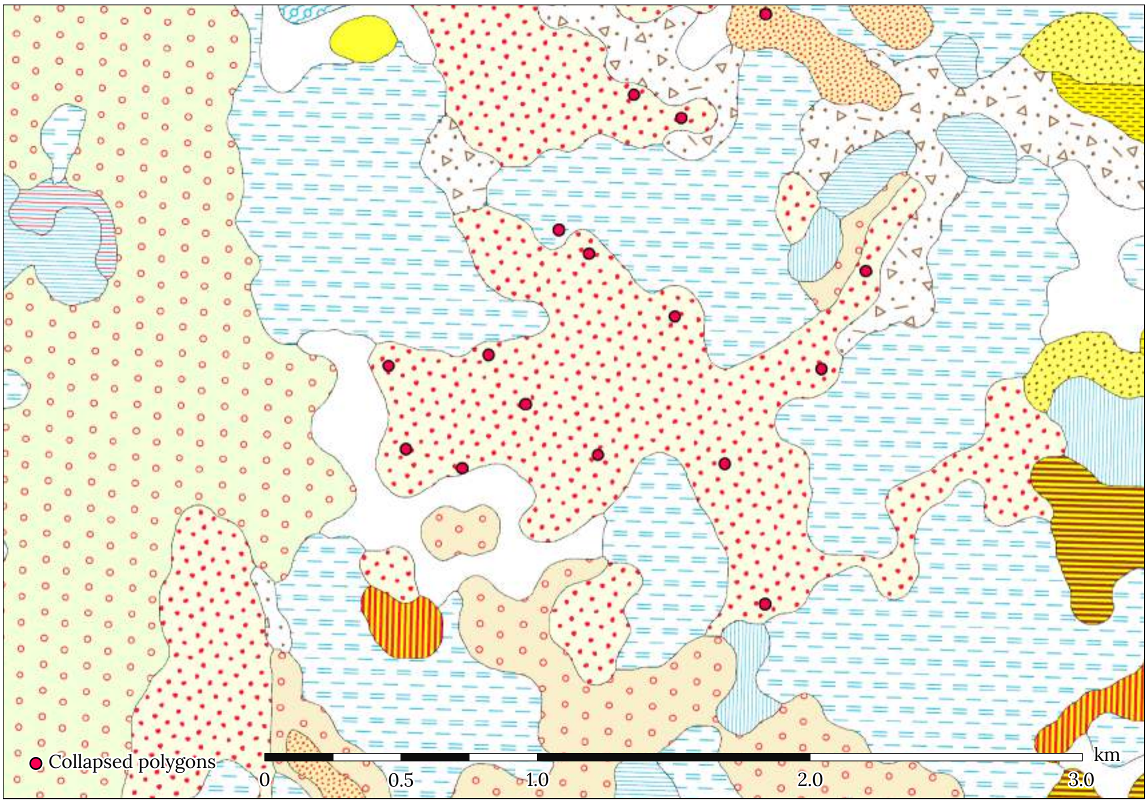
Figure 4.9. Statistics output. **(a)** Table header. **(b)** Table.

4.3.9 Output Map

Figure 4.10 shows the before and after of the generalization process. Both maps are at the same scale for a better comparison of the transformations.



(a)

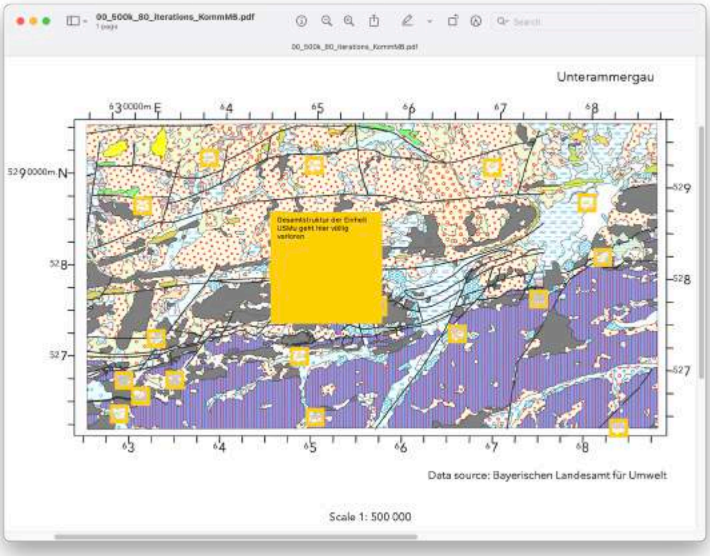


(b)

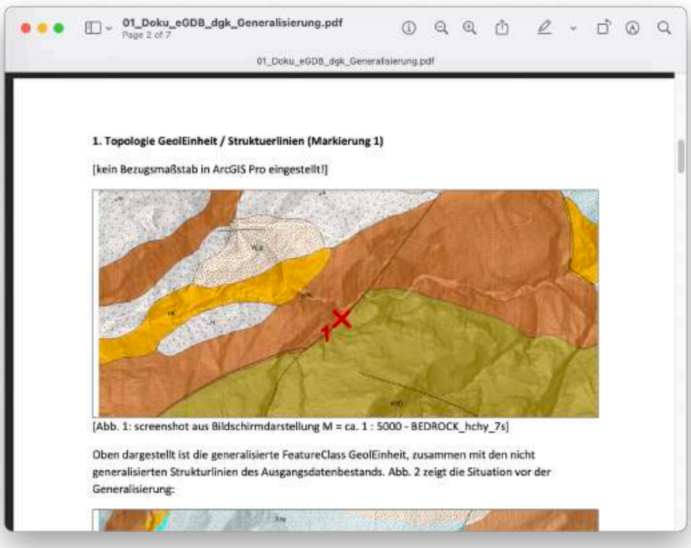
Figure 4.10. Output. **(a)** Before the generalization process. **(b)** After the generalization process.

4.4 Evaluation of the Results

For the evaluation, the experts reviewed either the PDF output maps or the GDBs. Some commented directly on the PDF maps, while others compiled their reviews in a separate document. An example of these two formats is shown in Figure 4.11. The aspects evaluated on each map were both cartographic and geological.



(a)



(b)

Figure 4.11. Comments from LfU experts. **(a)** Annotations. **(b)** A compiled PDF file with the reviews.

4.5 Polsby-Popper Score Calculation

The calculation of the PP score was performed in ArcGIS Pro using a model created in Model-Builder (Figure 4.12). The workflow has as input data a GDB containing data sets with the input and output data from the generalization process. The PP value is calculated using the tool "Calculate Field," which uses the formula described in section 2.3.

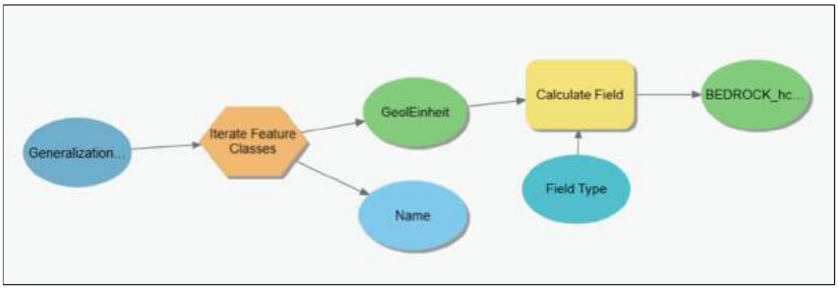


Figure 4.12. Model in ModelBuilder for PP calculation.

5 Results

This chapter presents the results from the experts' evaluation of the LfU and the analysis of these evaluations.

Section 5.1 describes the qualitative evaluations grouped into different categories. Each category explains specific examples according to the found limitations. Finally, section 5.2 describes the results from the Polsby-Popper score and its relation with the generalization process.

5.1 Qualitative Evaluation of Generalized Maps

The qualitative evaluation consisted of a visual examination of the generalized maps. This evaluation was carried out with the help of LfU geologists and cartographers to examine the readability and the geographic and geological coherence with respect to the input data (1:25,000 scale maps). The experts mainly analyzed maps with 80 iterations at three target scales: 1:150,000, 1:500,000, and 1:1,000,000.

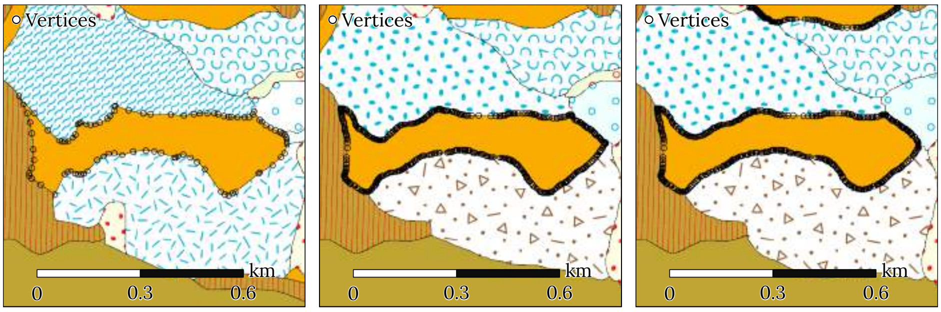
The evaluation from a macro level perspective (i.e., the whole map) can be divided into two points of view. On the one hand, some consider this generalization could be used for visualization on the Internet in the UmweltAtlas. On the other hand, some others think that due to the errors its use is limited.

Some errors found after the generalization process are explained in detail below. They are grouped according to their characteristics.

5.1.1 Shape Simplification

Concerning shape simplification, a drawback was related to the increase in the number of vertices in some generalization scales (see Figure 5.1). This increase is mainly due to the shape simplification process in the raster format. Although the shape simplification satisfies both non-self-intersection within a polygon outline and non-intersections with other elements, the increment in the number of vertices implies more storage and a subsequent shape simplification process.

Figure 5.1 shows the same polygon at 1:25,000 and 1:150,000 scales (the latter with 1 and 80 iterations). Figure 5.1a shows the polygon before the generalization process with 98 vertices. Figures 5.1b and 5.1c display the same polygon after the generalization process at the 1:150,000 scale with 417 and 403 vertices, respectively. Hence, the lower the number of iterations, the higher the number of vertices.



Scale 1:25,000 — (a)

Scale 1:150,000, 1 iteration — (b)

Scale 1:150,000, 80 iterations — (c)

Figure 5.1. The number of vertices before and after the generalization process. **(a)** Vertices at scale 1:25,000. **(b)** Vertices at scale 1:150,000 with one iteration. **(c)** Vertices at scale 1:150,000 with 80 iterations. The sizes of the figures have been adjusted for better comparability.

The disposition in which each polygon is located with respect to neighboring polygons can play for or against during the shape simplification phase (see Figure 5.2). Figures 5.2b, 5.2c, and 5.2d show the decrease in the area of the green polygon (pointed with a red arrow) that is greater to the east than to the west. This effect is because of the cellular automata process since it considers the pixels around it for the simplification process. In this case, to the west, there are four nearby polygons with different characteristics while the east-side distribution of nearby polygons is not as diverse and the polygon in contact with the green polygon covers a larger area which makes it the most predominant value when using cellular automata.

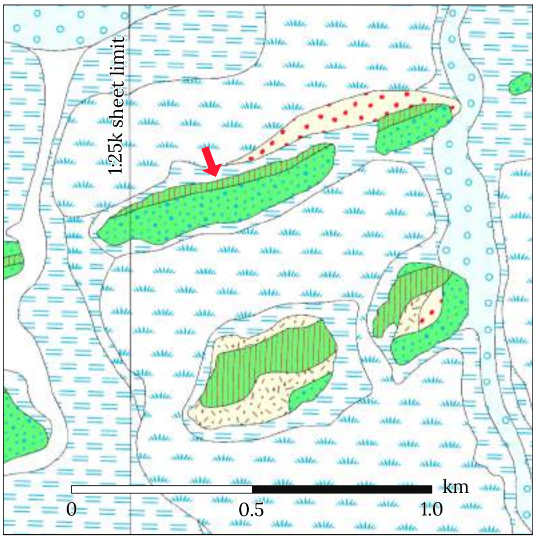
Figure 5.2a shows the green polygon marked with the red arrow at the original scale of 1:25,000. Figures 5.2b, 5.2c, and 5.2d show the same polygon generalized to the 1:150,000 scale using 10, 80, and 150 iterations, respectively. The asymmetric decrease in the area to the east and west of the green polygon is evident.

5.1.2 Aggregation Caused by Object Resolution

The process of aggregation caused by object resolution carries out as explained in section 4.3.5. However, sometimes this process gives unexpected results. An example of it is explained below.

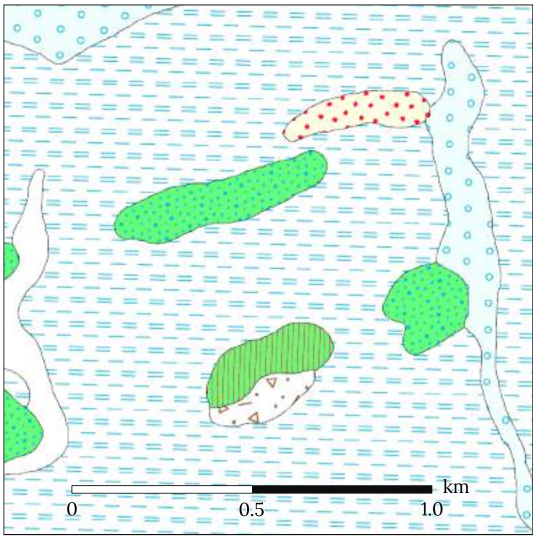
Figure 5.3a shows an area of the input map at a scale of 1:25,000. The red arrow highlights geological unit 423, which will be used as a case study for this example.

Figure 5.3b displays the first process of semantic aggregation. This phase considers the equivalence of each unit to the unit at the target scale, which in this case, corresponds to 1:150,000. Table 5.1 contains information on the correspondence of the geologic units to the three target scales. For this first step of semantic aggregation, the information in columns 1 and 3 corresponding to "unit at scale 1:25,000" and "parental unit at scale 1:150,000" is taken into account. For example, the first three units in Table 5.1 (401, 408, and 423) correspond to the same units at scale 1:150,000 (see Figures 5.3a and 5.3b). On the other hand, units 452 and 458 become unit 445 at 1:150,000.



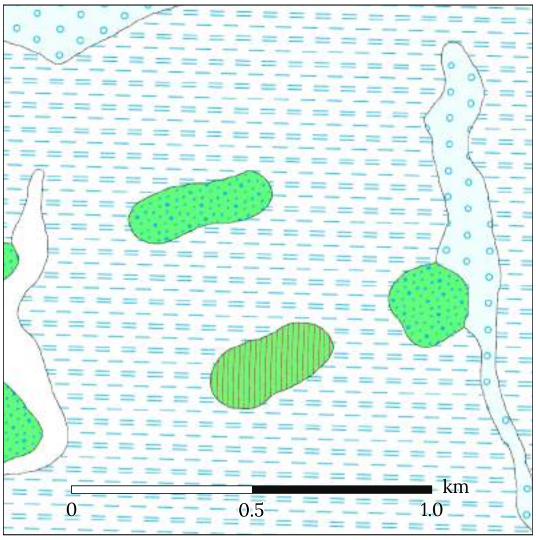
Scale 1:25,000

(a)



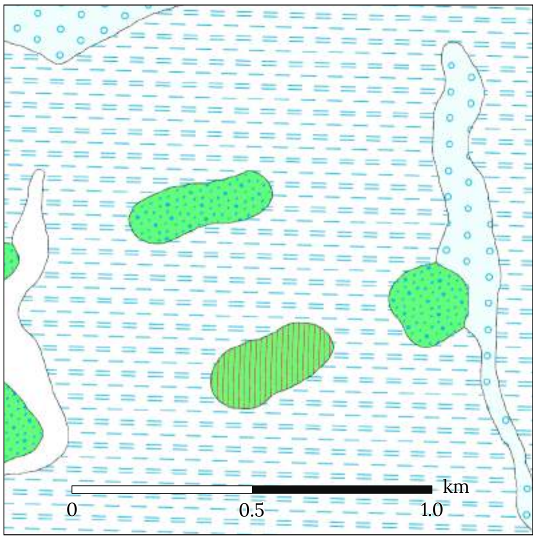
Scale 1:150,000 - 10 iterations

(b)



Scale 1:150,000 - 80 iterations

(c)



Scale 1:150,000 - 150 iterations

(d)

Figure 5.2. Shape simplification. **(a)** Input data at 1:25,000 scale. **(b)** Generalized output at 1:150,000 scale with 10 iterations. **(c)** Generalized output at 1:150,000 scale with 80 iterations. **(d)** Generalized output at 1:150,000 scale with 150 iterations. The sizes of the figures have been adjusted for better comparability.

Figure 5.3c depicts the structural analysis phase. In this step, the polygons that are under the minimum required area value according to the target scale (see Table 3.3) and that have at least one neighboring polygon with which they have semantic similarity are identified (highlighted in pink). To evaluate the semantic similarity in this step, the values that the units would take at the next level of generalization, which for this example would be 1:500,000, are taken into account.

Thus, for example, units 401 and 408 on the right would become unit 396 at a scale of 1:500,000. Therefore, unit 408 in Figure 5.3c is aggregated with unit 401 in Figure 5.3d. However, the highlighted geological unit 423 in Figures 5.3a and 5.3c became unit 408 in Figure 5.3d, which does not match the information in Table 5.1.

According to Table 5.1, unit 423 (scale 1:25,000) should be converted into units 423, 396, and 388 at scales 1:150,000, 1:500,000, and 1:1,000,000, respectively. It is why this result is unexpected. The problem is that the polygon with code 423 (highlighted with the red arrow) selects the neighboring polygon 408 in Figure 5.3c as the polygon to which it should be aggregated because of its semantic similarity. However, polygon 423 (red arrow) is not aggregated to polygon 408 on the right because the latter is aggregated to polygon 401. In this case, the correct solution would be that polygon 423 (highlighted with the red arrow) is also aggregated to the neighboring polygon 401 since units 401, 408, and 423 share semantic similarity by becoming unit 396 at 1:500,000 scale.

Table 5.1. Hierarchy table for Figure 5.3.

Unit at scale 1:25,000	Abbreviation of the unit at scale 1:25,000	Parental unit at scale 1:150,000	Parental unit at scale 1:500,000	Parental unit at scale 1:1,000,000
401	qhj3„	401	396	388
408	qhj2„	408	396	388
423	qhj1„	423	396	388
452	qhm2,G	445	439	388
458	qhm1°1°,G	445	439	388
462	qha,G	459	439	388
465	qh„l	521	521	521
476	„ta	476	476	476
494	„w	483	483	483
499	„hg	483	483	483
512	„ru	509	509	509
612	W,U,g	609	609	609
624	W,G,ge	619	609	609
650	Wh3°2°,G	648	647	599
651	Wh3°1°,G	648	647	599
976	OSM	976	976	975
1298	miOB	1298	1037	1032

5.1.3 Collapse of Features

In the current workflow setup, the collapse of polygons is controlled by the minimum area condition (see Table 3.3) and the semantics. The latter executes using the "flag" parameter (see sections 3.3.1 and 4.3.6) and the similarity of polygons. Even though collapsing polygons to points or lines was executed successfully, it was also a point of criticism (see comment 5.1).

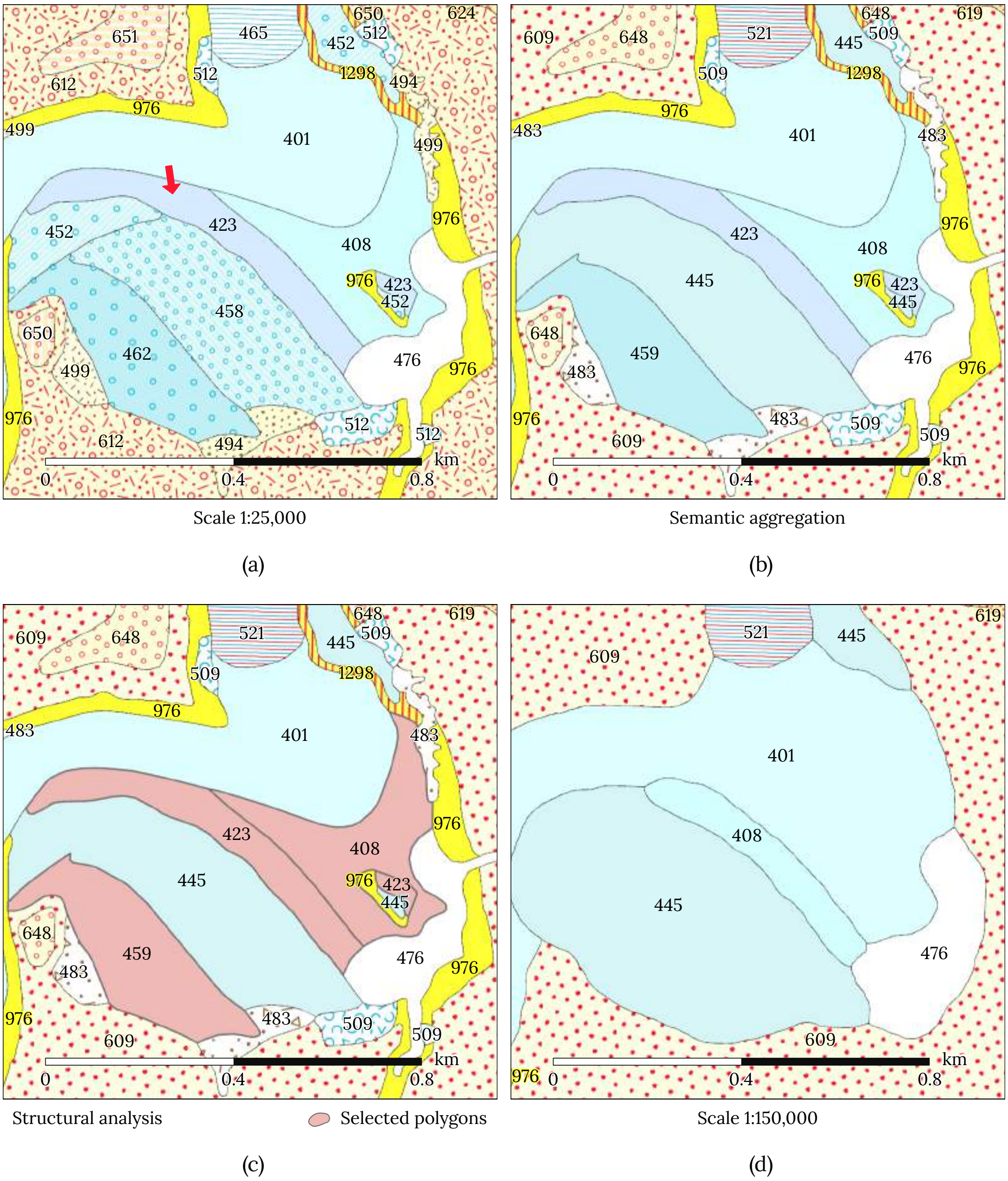
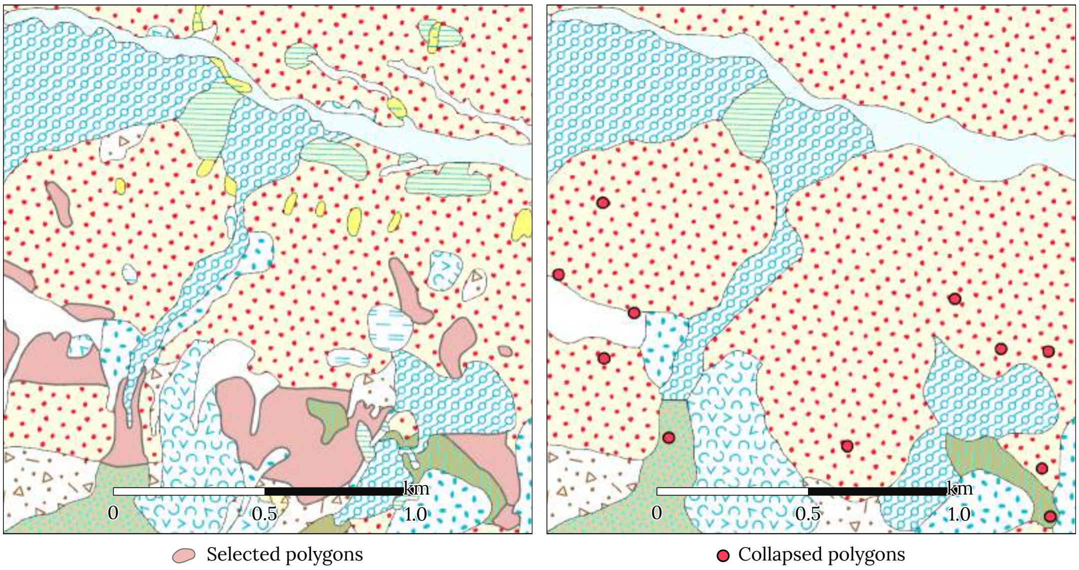


Figure 5.3. Example of error in aggregation caused by object resolution. **(a)** Input data at scale 1:25,000. **(b)** First semantic aggregation. **(c)** Structural analysis process. **(d)** Output at scale 1:150,000. The sizes of the figures have been adjusted for better comparability.

Comment 5.1: “Die Umwandlung kleiner Flächen in Punkte ergibt im Alpenraum keinen Sinn, da der stratigraphische Zusammenhang verloren geht. Die Punkte werden meist innerhalb einer Einheit angezeigt (Mittelpunkt der ursprünglichen Fläche), was meistens falsch ist. Außerdem erscheinen im Zielmaßstab sehr viele Punkte dicht gedrängt nebeneinander, so dass man es teilweise durch eine Fläche darstellen könnte.
Punkte könnten sinnvoll für kleinere, punktuelle geologische Flächen, wie z.B. Vulkane,

> kleinere Intrusionen, sein." ["The conversion of small areas into points does not make sense in the alpine region, because the stratigraphic context is lost. The points are mostly displayed within one unit (center of the original surface), which is mostly wrong. In addition, in the target scale many points appear densely packed next to each other, so that it could be partially represented by a surface. Points could be useful for smaller, punctual geologic surfaces, such as volcanoes, minor intrusions."] (U. Teipel, personal communication, August 22, 2022)

Figure 5.4a displays the polygons selected to be collapsed in pink. Not all polygons under the minimum area were chosen to be collapsed; on the contrary, only those that fulfilled this condition and, in addition, had at least one neighboring polygon with which they shared semantic similarity. Figure 5.4b shows the generalized map with the selected polygons in the left collapsed to points. The polygons under the minimum area (e.g., yellow polygons on the left) that did not have similar neighbors were removed after subsequent processes, such as shape simplification.



(a) (b)

Figure 5.4. Shortcoming caused by collapse. **(a)** Data before collapse. **(b)** Data after collapse.

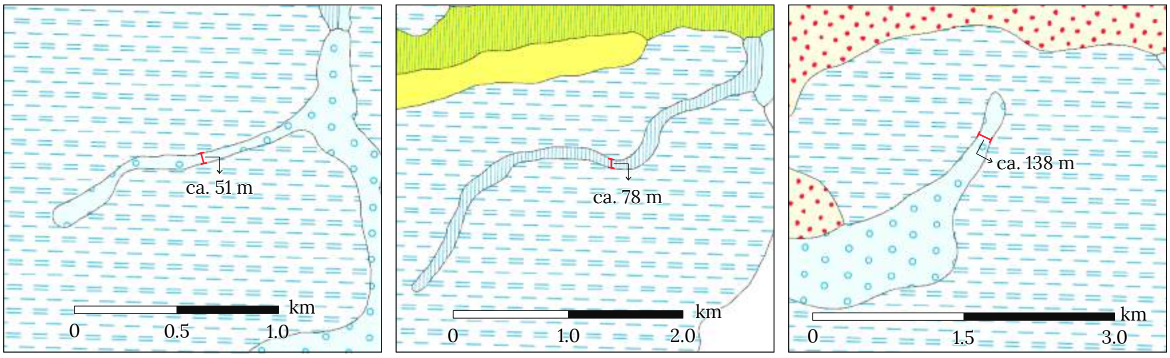
As the comment 5.1 indicates, collapsing polygons to points or lines is a helpful solution for geological features such as volcanoes or intrusions. It is also a way to inspect the data for further data enrichment. Nevertheless, the previous points also reveal the need for an alternative solution; for example, instead of collapsing the polygons marked with a flag, enlarging them to guarantee their permanence as polygons.

5.1.4 Minimum Dimensions

The generalization process analyzed here ensures that the minimum area condition is met (see Table 3.3). Hence, polygons below the minimum area are collapsed or aggregated to another polygon. However, there are polygons that, even though they are above this minimum area, the minimum distance between border points is less than three mm in map units, which makes their legibility difficult.

Although currently, no step guarantees a minimum distance between polygons between border points, it should be considered for future improvements to the workflow.

Figure 5.5 shows three examples where the minimum distances between polygon border points are less than the threshold distances established for each target scale (see Table 5.2). From left to right are the three target scales, 1:150,000, 1:500,000, and 1:1,000,000 each processed at 80 iterations.



Scale 1:150,000 - 80 iterations | Scale 1:500,000 - 80 iterations | Scale 1:1,000,000 - 80 iterations

(a) | (b) | (c)

Figure 5.5. Different examples of under minimum distances at various scales. **(a)** 1:150,000 scale. **(b)** 1:500,000 scale. **(c)** 1:1,000,000 scale. The sizes of the figures have been adjusted for better comparability.

Table 5.2. Distance between boundary points per scale.

Target scale	Minimum distance between boundary points (m)	Minimum distance between boundary points in map units (mm)
1:150,000	450	3
1:500,000	1,500	3
1:1,000,000	3,000	3

5.1.5 Geological Characteristics

A geological map represents the distribution of rocks. This distribution is complex and involves semantic relationships that are important to preserve (e.g., folds, tectonic faults, stratigraphic sequences, etc.). Even though the output data satisfy the spatial and semantic criteria set up in the current workflow, some geological features worth preserving of the study area fade as the scale is decreased. Some examples follow.

Figure 5.6 depicts the first example of large non-preserved structures with downscaling. Figure 5.6a shows a section of the 1:500,000 scale map manually generalized in 1996. This figure is used as a basis for comparison of the automatic generalizations shown in Figures 5.6b, 5.6c, and 5.6d. The purpose of this comparison is to see the level of preservation of the orange structure located in the central part which has a horseshoe shape.

Figures 5.6b, 5.6c, and 5.6d represent the automatic generalizations of the study area at scales 1:150,000, 1:500,000, and 1:1,000,000, respectively. The decrease in detail of the structure under study is evident as the scale is reduced (from Figure 5.6b to 5.6d). Comment 5.2 agreed with this effect.

> **Comment 5.2:** "Manchmal verschwinden auch größere zusammenhängende Strukturen, die als Ganzes betrachtet werden sollten, z. B. eine Art langgezogenes Hufeisen im Alpenvorraum, die bei den groben Generalisierungsstufen teils verschwinden." ["Sometimes larger coherent structures that should be considered as a whole, e.g. a kind of elongated horseshoe in the Alpine foothills, partly disappear at the coarse generalization levels."] (M. Balg, personal communication, August 19, 2022)

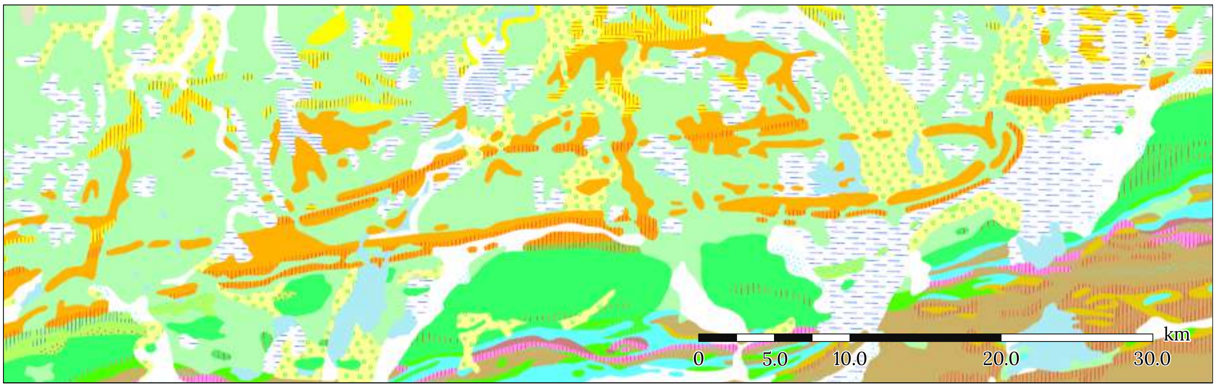
If this drawback is caused by the collapse of polygons to points or lines, a possible solution is to include a function that enlarges polygons according to their semantic importance.

The second example corresponds to the decrease or disappearance of SW-NE to W-E structures as the scale is reduced. Figure 5.7 shows an example of geological units aligned in the SW-NE direction on the left at the original scale (1:25,000) and on the right at the generalization scale 1:150,000. The increase of roughness or "nodulation" of the polygons is visible (see comment 5.3).

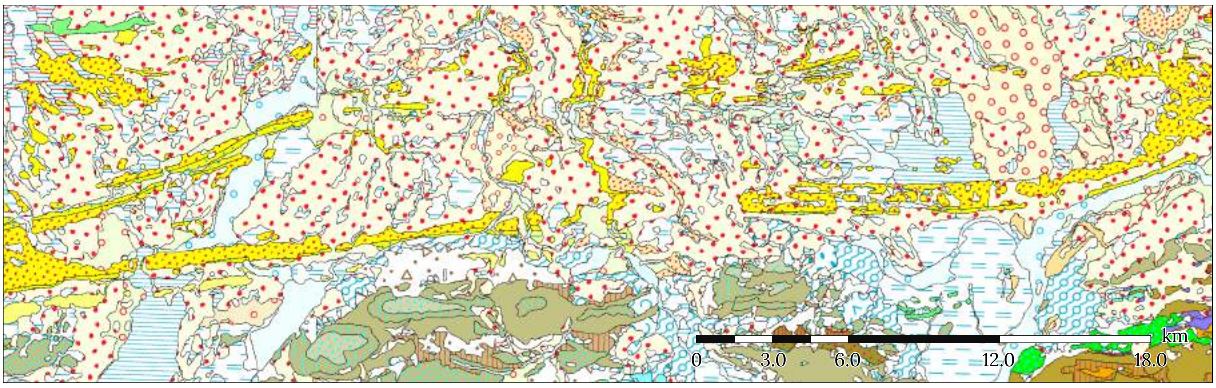
> **Comment 5.3:** "im Vergleich der beiden Bilder zeigt sich schön die „Verknödelung" der Flächen im Zuge der Generalisierung: die für die Alpen charakteristischen, langgezogenen SW-NE bis W-E streichenden Strukturen (meist Faltenkerne) werden zugunsten der größeren umgebenden Fläche verkürzt und erscheinen dann oft (mit kleineren Maßstäben zunehmend) wie Knödel." ["the comparison of the two images nicely shows the "nodulation" of the surfaces in the course of generalization: the elongated SW-NE to W-E strike structures (mostly fold cores) characteristic of the Alps are shortened in favor of the larger surrounding surface and then often appear (with smaller scales increasingly) like nodules."] (U. Teipel, personal communication, August 22, 2022)

The third example is related to coherence in stratigraphic sequences. Sequences have continuity in geologic time, so their position and relationship with adjacent units have a semantic meaning that must be respected. Figure 5.8 shows an example of an altered sequence in the generalization process. Comment 5.4 addressed this case.

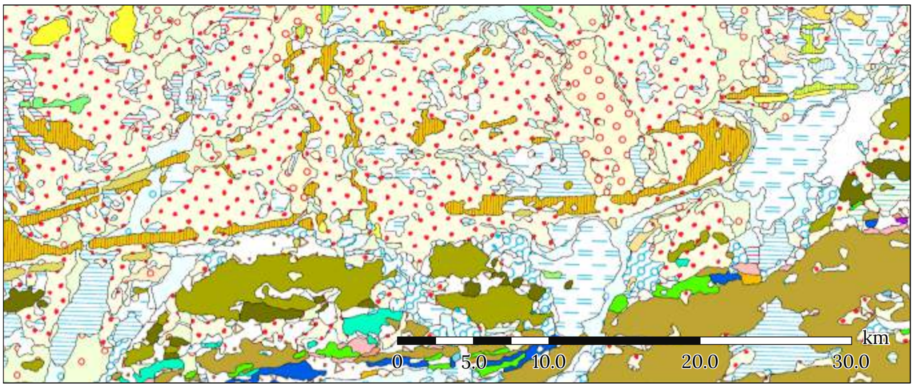
> **Comment 5.4:** "Muldenstruktur im Zentrum verschwindet: aus der Abfolge (N → S) nHD-nO-nSC-nA-nSC-nO/nK-nHD wird nHD-nSC-nA-nK-nPK/nHD – die symmetrische Abfolge durch die Mulde ist verschwunden. Die einzelnen Flächen wären (mit Ausnahme nO) groß und breit genug, um die richtige Abfolge zu generalisieren." ["Trough structure in the center disappears: the sequence (N → S) nHD-nO-nSC-nA-nSC-nO/nK-



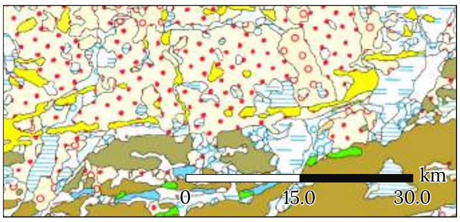
(a)



(b)

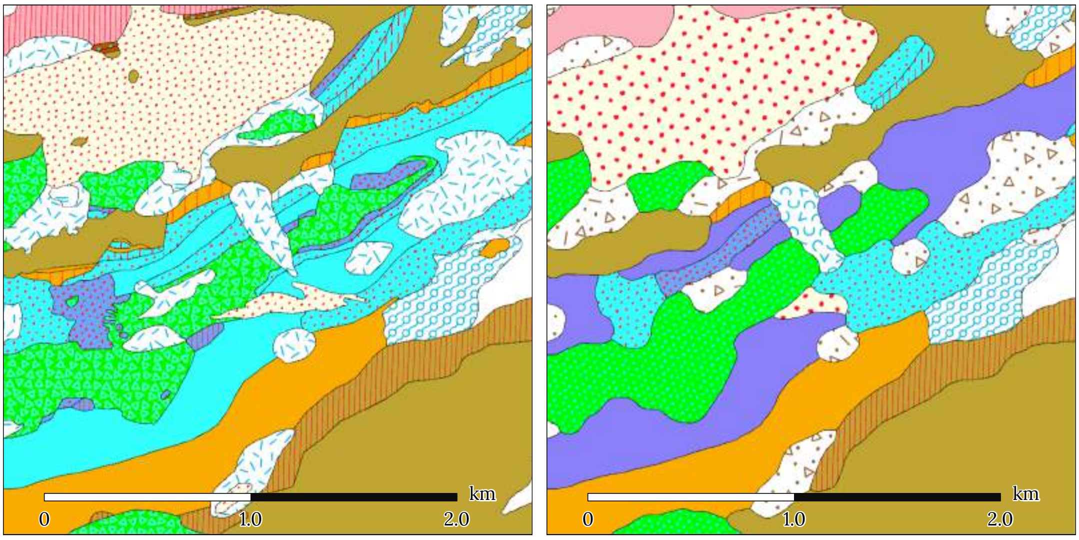


(c)



(d)

Figure 5.6. Example of large non-preserved geological structure. **(a)** The manually generalized map at the scale 1:500,000. **(b)** The automatically generalized map was 1:150,000 but reduced to a scale of 1:300,000 to fit the page size. **(c)** and **(d)** The automatically generalized maps at scale 1:500,000 and 1:1,000,000, respectively.



Scale 1:25,000

(a)

Scale 1:150,000 - 80 iterations

(b)

Figure 5.7. Change of SW-NE to W-E aligned elongated structures with generalization. **(a)** Original data at scale 1:25,000. **(b)** Generalized data at scale 1:150,000. The sizes of the figures have been adjusted for better comparability.

nHD becomes nHD-nSC-nA-nK-nPK/nHD - the symmetrical sequence formed by the trough has disappeared. The individual surfaces would be large enough and wide enough (except for nO) to generalize the correct sequence."] (U. Teipel, personal communication, August 22, 2022)

Figure 5.8a highlights with a white halo the labels of the units belonging to the sequence mentioned in the previous paragraph nHD-nO-nSC-nSC-nA-nSC-nO/nK-nHD (from north to south). Figure 5.8b shows the selection of some polygons that would be aggregated into neighboring polygons according to their semantic characteristics in the next phase of the process. Some of those polygons are part of the sequence.

The result of the aggregation phase is shown in Figure 5.8c where the nSC and nO units of the southern section of the sequence disappear. Figure 5.8c displays the inversion of the nPK and nHD units at the lower right, which in Figure 5.8b were nHD and nPK. This drawback occurs similarly to the selection and aggregation process described in section 5.1.2 and Figure 5.3. The output after the shape simplification process is seen in Figure 5.8d. Here it is evident the elimination of the nO unit during the shape simplification phase resulted in the sequence nHD-nSC-nA-nK-nPK/nHD, which no longer has the symmetry it had at the beginning of the process.

5.1.6 Harmonization

The study area is non-harmonized even though some borders of the 1:25,000 scale sheets are harmonized. Although this work does not focus on the harmonization subject, it is worth noting

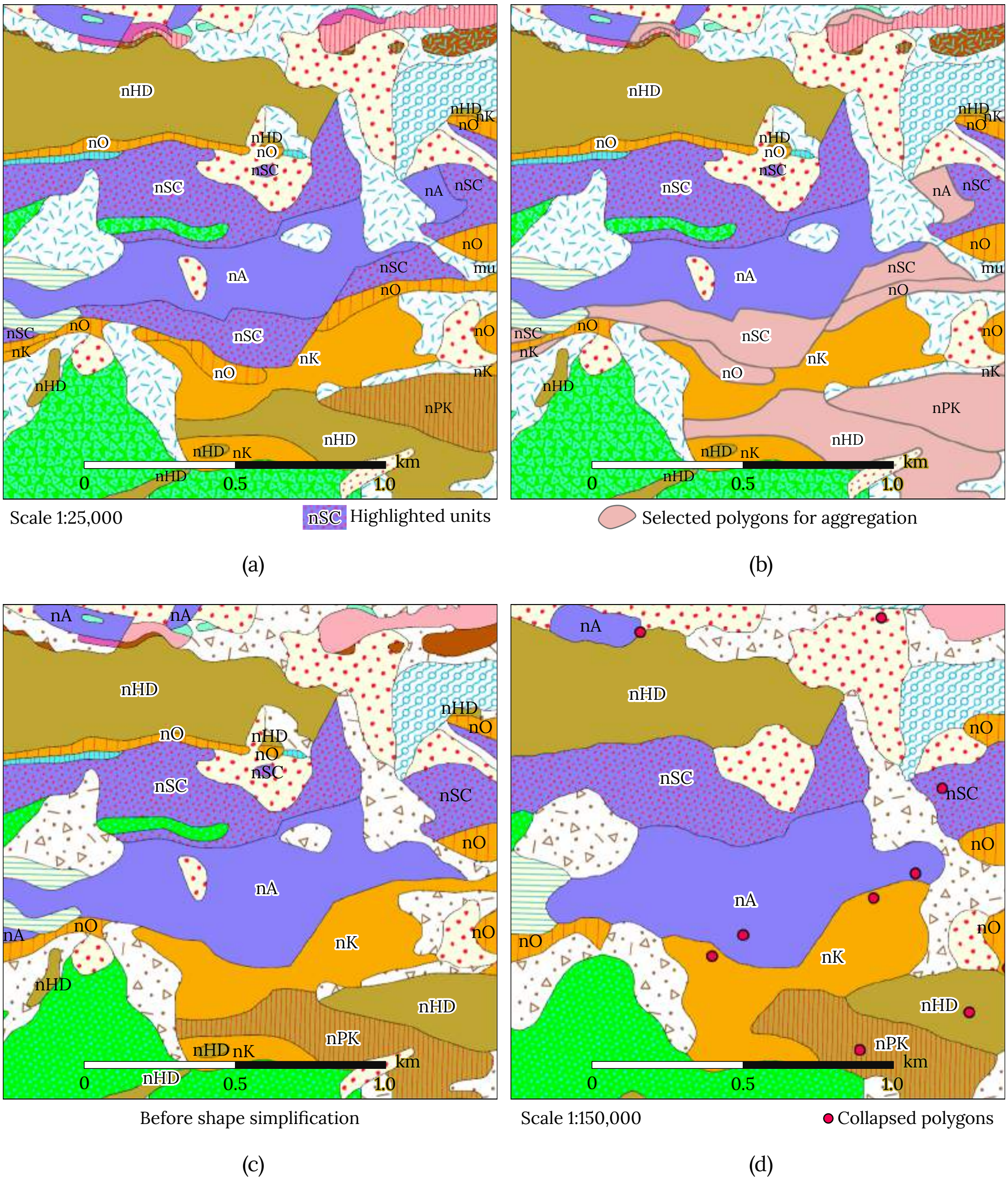


Figure 5.8. Example of stratigraphic sequence lost. **(a)** Input data at scale 1:25,000. **(b)** Structural analysis that identifies polygons to be aggregated to neighboring polygons. **(c)** Data before shape simplification process. **(d)** Output map at scale 1:150,000. The sizes of the figures have been adjusted for better comparability.

that the results obtained here show the need to harmonize the 1:25,000 scale sheets before the automatic generalization process.

Figure 5.9a represents the output of the generalization process at the 1:150,000 scale. In the upper left, marked with a red arrow, the evidence of the non-harmonization of the sheets at a 1:25,000 scale is visible. Due to the units on both sides do not match, they are not aggregated during the generalization process, and thus the boundary of the sheets at 1:25,000 remains almost intact.

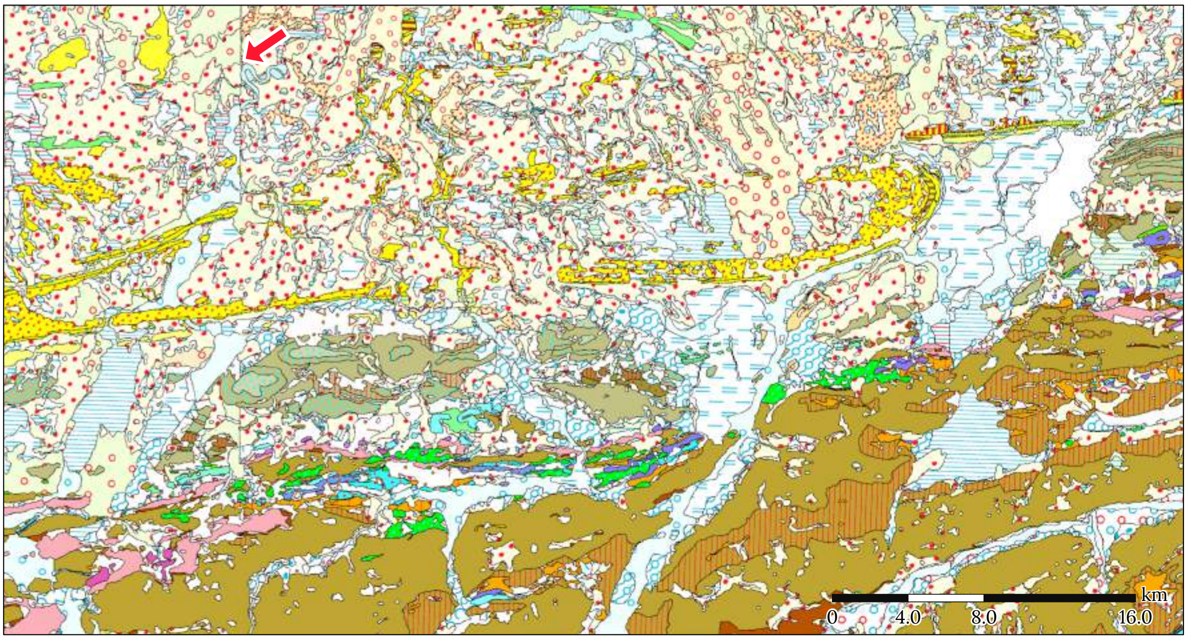
A similar result, but no longer as marked as before, is seen in the generalized 1:500,000 scale map in Figure 5.9b (marked with a red arrow). Similarly, Figure 5.9c shows the same effect but reduced (highlighted with the red arrow) on the generalized 1:1,000,000 scale map.

5.2 Polsby-Popper Score

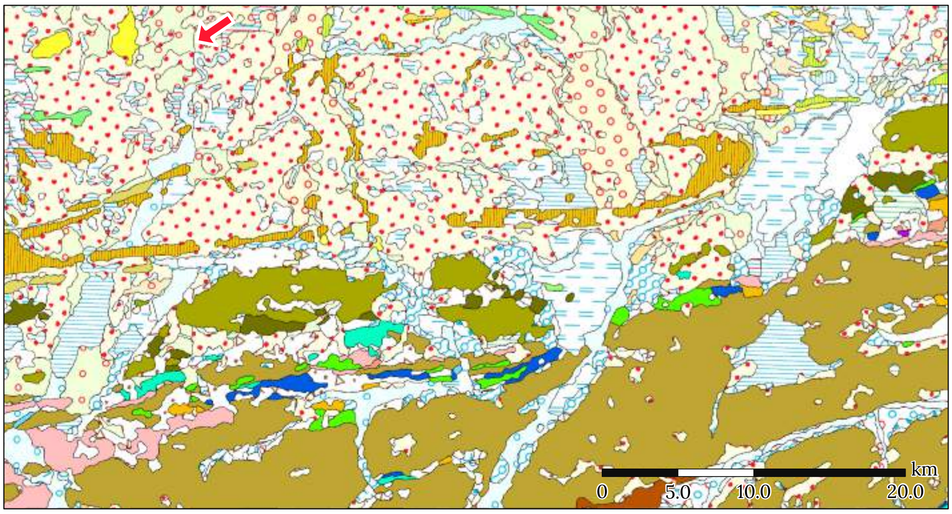
Section 5.1.5 mentioned the increase of polygons' roundness after generalization. The Polsby-Popper score also indicates that the higher the number of iterations, the more round the polygons are, which means closer to one (see Table 5.3). An example of it is shown in Figure 5.2. It displays the input and output data. The output is shown in Figures 5.2b, 5.2c, and 5.2d, which correspond to different numbers of iterations (10, 80, 150).

Table 5.3. Polsby-Popper score.

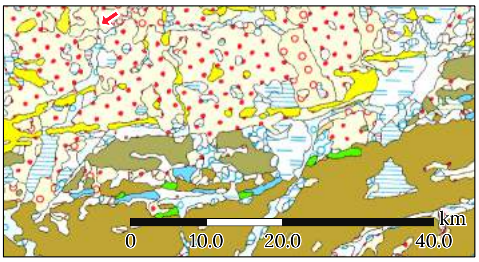
	Input data at 1:25,000	**Output data at 1:150,000**		
Iterations	0	10	80	150
Total number of polygons	18433	4701	4291	4263
Mean PP score	0.445	0.480	0.521	0.523



(a)



(b)



(c)

Figure 5.9. Error caused by non-harmonization of input data. **(a)** The generalized map was at 1:150,000 but reduced to a scale of 1:400,000 to fit the page size. **(b)** and **(c)** Outputs at scales 1:500,000 and 1:1,000,000, respectively.

6 Discussion

The main objective of this research was to analyze and evaluate an automatic procedure for the generalization of geological maps from a 1:25,000 scale (source) to 1:150,000, 1:500,000, and 1:1,000,000 scales (target) based on an existing workflow. In this chapter, I discuss the workflow and its results.

The workflow implemented in this thesis has some limitations. The first is the need for a previous learning phase on the structure and execution of the workflow. This prior knowledge is fundamental to executing the workflow, evaluating its results, and proposing solutions to drawbacks. The time this phase takes depends largely on the person's expertise, both in terms of knowledge of the subject and experience with software used in this workflow.

The second shortcoming is related to data enrichment. This process is time-consuming as it involves compiling and organizing the geological information for each unit in the study area. In this phase, the geologists with geological knowledge of the study area establish the hierarchical relationships between the "base" units at the 1:25,000 scale and their parent geological units at the target scales, for this case study 1:150,000, 1:500,000, and 1:1,000,000. The time required to carry out this stage depends, among other things, on the current state of geological mapping coverage in the study area at different scales.

The third disadvantage lies in shape simplification. Two main limitations are identified in this step. On the one hand, the evident increase in the number of vertices after the raster shape simplification process implies a subsequent simplification process. On the other hand, the disadvantage of not preserving the original shape proportions of the polygons depends on the number and distribution of the neighboring units per polygon.

The fourth limitation is found in the aggregation phase caused by object resolution. This process focuses on the polygons' aggregation under the minimum area and with neighbors with semantic similarity. When several adjacent polygons are candidates for aggregation, it sometimes results in the occurrence of an isolated polygon with its semantic category changed (i.e., the geological unit it represents).

Another disadvantage is the collapse of all polygons that do not satisfy the minimum area condition and have at least one adjacent polygon with which they share a semantic similarity. It represents a disadvantage because, although collapse is useful for features such as volcanoes or intrusions, it is not useful for polygons that are part of larger structures. Applying an operator such as enlargement to preserve these polygons is preferable in this case.

Another drawback is the absence of a function to measure minimum dimensions. Although the minimum area is a condition that is guaranteed in this workflow, other conditions that satisfy minimum dimensions (e.g., the distance between boundary points and object separation) are fundamental to ensure the readability of the resulting map.

Additionally, another shortcoming is related to the non-preservation of geological features. This non-preservation applies to large structures formed by a group of non-adjacent polygons, geometric proportions of elongated elements, and stratigraphic sequences. The first case occurs because the polygons that conform to the large structure collapse or aggregate into other units. In this case, it would be useful to associate the flag attribute with the enlargement operation rather than the collapse. The second case is related to the raster method used for shape simplification and the number of iterations of this process. The higher the number of iterations, the rounder the polygons which makes elongated thin polygons more rounded. The third case is also the result of vector-based collapse and aggregation and raster-based shape simplification. The last one counts since, in the raster-based simplification, polygons are eliminated according to the minimum number of cells.

7 Conclusion

The main objective of this thesis was to analyze and evaluate an automatic procedure for the generalization of geological maps from a 1:25,000 scale (source) to 1:150,000, 1:500,000, and 1:1,000,000 scales (target) based on an existing workflow. For this purpose, a study area south of Bavaria was defined based on the completeness of the input data and the geological complexity that this area represents.

The first sub-objective was to execute the generalization process of geological maps to the target scales from a geological database at the source scale. Along with the practical part, a theoretical step was carried out. The theoretical part consisted of the literature compilation on generalization applied to geological maps. The practical part was carried out in three phases. The first phase consisted of correcting operational problems of the workflow to be able to use it. The second step consisted of a learning process to understand the structure and functions of the workflow. Experiments in a small area were performed. In the preliminary tests, an error was found. It was about gaps in the output map due to incomplete data in the hierarchy table. However, this error was corrected before continuing the study area tests. The third step consisted of running the generalization process in the study area at the three target scales. It was found that the generalization process for each target scale is determined by parameter values established for each scale but implemented in the same workflow.

The second sub-objective was to evaluate the results at different target scales. This objective was achieved by conducting a literature review on evaluation in generalization, assessing the results with the help of LfU experts, and compiling the evaluations. The constraint parameters that led the generalization process were identified. Then, values were assigned according to LfU standards and reference literature. It led to the results obtained for the generalization evaluation presented to the LfU experts. The most appropriate evaluation method for this holistic process was a qualitative visual assessment after generalization, given the complexity of integrating a quantitative evaluation that usually examines only specific characteristics. Additionally, the results are considered promising for use in the web service of the UmweltAtlas since, due to the identified disadvantages, their use for printed maps is limited.

The third sub-objective was to orchestrate the workflow to ensure its use in the Bayerisches Landesamt für Umwelt (LfU). This objective aimed to test the constraints' values and analyze their influence on the results of the generalization process. The values of the constraints with which tests were performed were obtained from LfU's standards and reference documents of this workflow. One parameter that significantly influences the generalization process is the number of iterations, which affects the shape simplification process. It was found, for example, that the higher the number of iterations, the higher the roundness of polygons in the resulting maps.

However, it was also found that to ensure the use of this workflow within the LfU production lines, especially for printed maps, efforts should first be addressed to solve the limitations encountered rather than to establish ideal parameter values.

The generalization of geological maps is difficult, given the complexity of this type of map. The workflow analyzed and evaluated in this thesis achieves fundamental requirements for the generalization of this type of map. First, it proposes a solid strategy to handle the semantic relationships of geological units at different scales. In addition, it allows generalization to different objective scales through a series of parameter values assigned for each scale. Furthermore, it achieves seamless integration of vector- and raster-based processes. Fourth, it guarantees the resolution of conflicts related to the minimum area by collapsing or aggregating the objects involved in the conflict. Finally, it guarantees in the shape simplification process the topological correctness of the generalized polygons: no self-intersections within a polygon outline and no intersections of different objects. Based on the analysis performed here, this workflow promises, in principle, to be valid for use on the LfU website, although further evaluations are necessary.

Although this workflow represents an advance in automating the generalization process in geological maps, some limitations must be addressed to fulfill the quality requirements. These limitations are grouped into three: geometric, semantic, and functional. Geometric shortcomings deal mainly with the non-preservation of the original shape proportions of the polygons. Semantic limitations especially involve cases where unaggregated polygons are changed semantic category (i.e., geologic unit) in the process of aggregation caused by object resolution. In addition, the non-preservation of consistent stratigraphic sequences can cause, for example, hiatus or structural contacts where there are none. Finally, functional limitations have to do mainly with the collapse process and the measurement of minimum dimensions, which require an additional reconfiguration and creation of other functions.

Based on the analysis and evaluation performed here, this thesis concludes with the overall finding that the workflow has the potential for implementation as an automatic generalization of geological maps in the LfU. However, the limitations mentioned in the paragraph above need solutions to provide a resulting map that ensures legibility and maintains polygon characteristics.

8 Outlook

Based on the discussion and experience gained during the development of this thesis, some aspects that require further improvement were identified. Therefore, this section aims to highlight potential improvements to the methods presented here.

→ Morphological operators such as opening and closing can improve the shape simplification process, guaranteeing a shape simplification that keeps the original shape proportions of the polygons.
→ To avoid the appearance of isolated polygons with the changed geological unit, an iterative process could be considered to verify these "isolated" polygons and thus add them to the neighboring polygons in case the semantic characteristics are similar.
→ The collapse process should be restructured to only collapse geological features such as intrusions. In addition, a parallel process of enlargement of relevant polygons below the minimum area should be created.
→ Measurement of minimum dimensions should be incorporated to measure at least the distance between boundary points and object separation.
→ Since the workflow does not have a line simplification process for tectonic faults, it should be included in a way that considers the distribution of the generalized data.

A Appendix

Outputs for Evaluation

The generalized 1:150,000 scale map in Figure A.1 has an original size of A2 (420 × 594 mm).

The generalized 1:500,000 scale map in Figure A.2 has an original size of A4 (210 × 297 mm).

The generalized 1:1,000,000 scale map in Figure A.3 has an original size of A5 (148 × 210 mm).

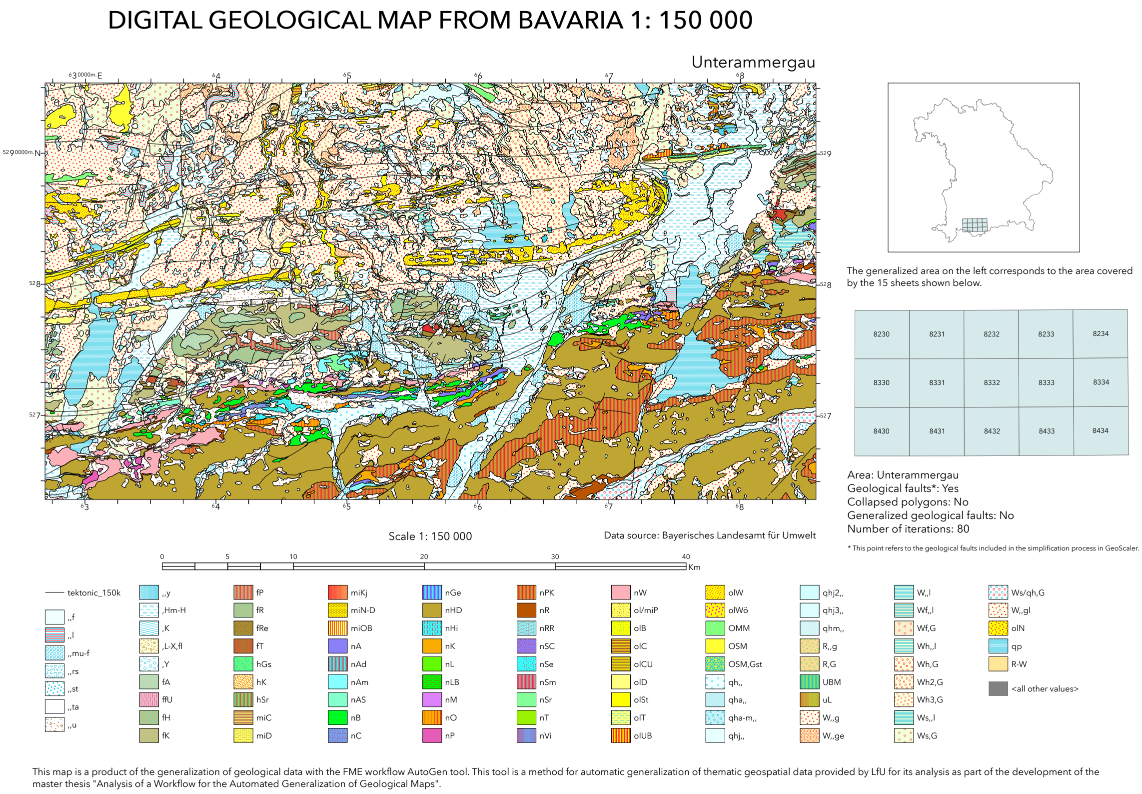


Figure A.1. Result of generalization at 1:150,000 and 80 iterations.

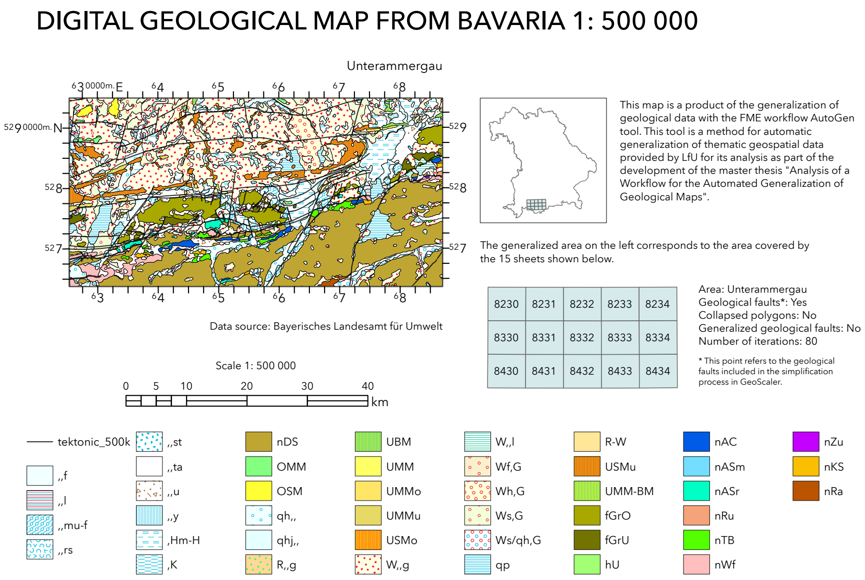


Figure A.2. Result of generalization at 1:500,000 and 80 iterations.

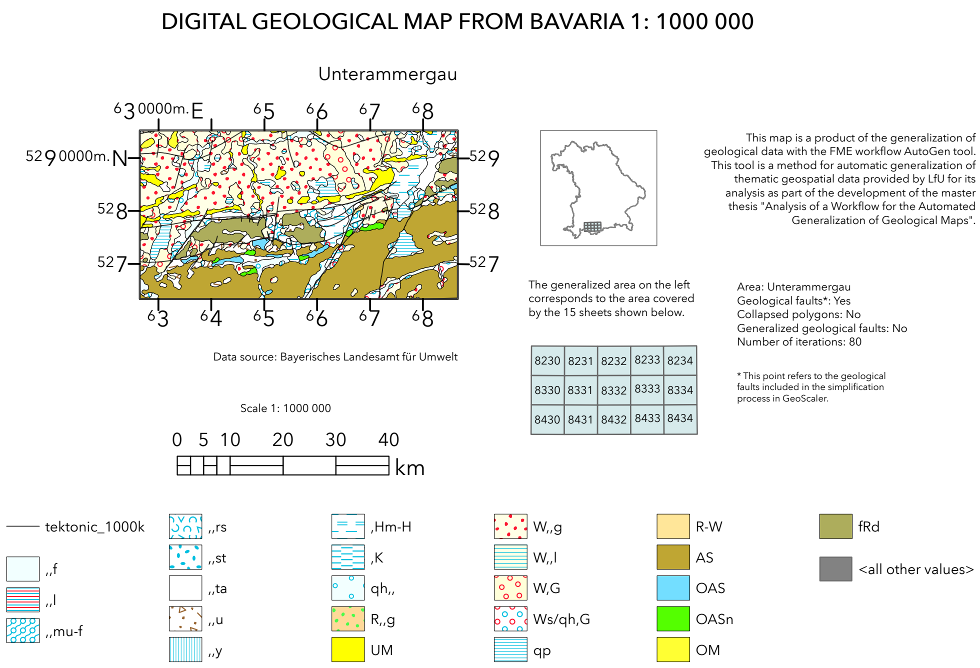


Figure A.3. Result of generalization at 1:1,000,000 and 80 iterations.

List of Abbreviations

CA	cellular automata	7
CDT	constrained Delaunay triangulation	5
DT	Delaunay triangulation	5
GDB	geodatabase	18
LfU	Bayerisches Landesamt für Umwelt	I
LGRB	Landesamt für Geologie, Rohstoffe und Bergbau	17
LCNP	Laboratoire de cartographie numérique et de photogrammétrie	
OGC	Open Geospatial Consortium	19
PP	Polsby-Popper	3
RDT	conforming Delaunay triangulation	5
TUD	Technische Universität Dresden	
TUM	Technische Universität München	

List of Figures

List of Tables

Bibliography

Bader, M., & Weibel, R. (1997). Detecting and resolving size and proximity conflicts in the generalization of polygonal maps. In *18th international cartographic conference* (pp. 1525–1532). https://icaci.org/files/documents/ICC_proceedings/ICC1997/icc1997_volume3_part2.pdf. (Cit. on pp. 5, 10)

Barrault, M., Regnauld, N., Duchêne, C., Haire, K., Baeijs, C., Demazeau, Y., Hardy, P., Mackaness, W., Ruas, A., & Weibel, R. (2001). Integrating multi-agent, object-oriented and algorithmic techniques for improved automated map generalization. In *20th international cartographic conference*. https://icaci.org/files/documents/ICC_proceedings/ICC2001/icc2001/file/f13041.pdf. (Cit. on pp. 10 sq.)

Bayerisches Landesamt für Umwelt. (n.d-a). Digital geological map 1:25.0000 (dgk25) [geological map]. *Creative Commons Attribution 4.0 International*. Retrieved May 18, 2022, from www.lfu.bayern.de. (Cit. on p. 4)

Bayerisches Landesamt für Umwelt. (n.d-b). Geologische karte von bayern 1:500.000 [geological map]. *Creative Commons Attribution 4.0 International*. Retrieved August 15, 2022, from www.lfu.bayern.de. (Cit. on p. 16)

Brown, N. J., Wright, S. M., & Fuller, R. M. (1996). A technique for the removal of outliers during a computerised map generalisation process. *The Cartographic Journal*, 33(1), 11–16. https://doi.org/10.1179/caj.1996.33.1.11 (cit. on p. 7)

Burghardt, D., & Neun, M. (2006). Automated sequencing of generalisation services based on collaborative filtering. In M. Raubal, H. J. Miller, A. U. Frank, & M. Goodchild (Eds.), *4th international conference, giscience 2006* (pp. 41–46). IfGI prints. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.85.3725&rep=rep1&type=pdf. (Cit. on pp. 10 sq.)

Burghardt, D., Schmid, S., & Stoter, J. (2007). Investigations on cartographic constraints formalisation. In *11th ica workshop on on generalisation and multiple representation*. IfGI prints. https://kartographie.geo.tu-dresden.de/downloads/ica-gen/workshop2007/Burghardt-ICAWorkshop.pdf. (Cit. on pp. 10–12)

de Berg, M., van Kreveld, M., & Schirra, S. (1998). Topologically correct subdivision simplification using the bandwidth criterion. *Cartography and Geographic Information Systems*, 25(4), 243–257. https://doi.org/10.1559/152304098782383007 (cit. on p. 6)

Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2), 112–122. https://doi.org/10.3138/FM57-6770-U75U-7727 (cit. on p. 6)

Downs, T. C., & Mackaness, W. A. (2002). An integrated approach to the generalization of geological maps. *The Cartographic Journal*, 39(2), 137–152. https://doi.org/10.1179/caj.2002.39.2.137 (cit. on pp. 1, 6)

Dupke, S., Hagemeister, A., & Schuff, J. (2021). *Fme workflow autogen*, Landesamt für Geologie, Rohstoffe und Bergbau. (Cit. on p. 24).

Edwardes, A., & Mackaness, W. (2000). Modelling knowledge for automated generalisation of categorical maps - a constraint based approach. In P. Atkinson & D. Martin (Eds.), *Innovations in gis 7: Gis and geocomputation* (pp. 161–173). Taylor & Francis. https://doi.org/https://doi.org/10.1201/9781482268263. (Cit. on p. 11)

Ehrliholzer, R. (1995). Quality assessment in generalization: Integrating quantitative and qualitative methods. In *17th international cartographic conference*. https://icaci.org/files/documents/ICC_proceedings/ICC1995/PDF/Cap431.pdf. (Cit. on pp. 12 sq.)

Fleischmann, P. (2000). *Boundary integrity*. Retrieved August 27, 2020, from https://www.iue.tuwien.ac.at/phd/fleischmann/node51.html. (Cit. on p. 5)

Galanda, M. (2003a). *Automated polygon generalization in a multi agent system* (PhD dissertation). University of Zurich. https://doi.org/10.5167/uzh-163108. (Cit. on pp. 4–6, 8–12)

Galanda, M. (2003b). Modelling constraints for polygon generalization. In *7th ica workshop on progress in automated map generalization*. https://kartographie.geo.tu-dresden.de/downloads/ica-gen/workshop2003/galanda_v1.pdf. (Cit. on p. 10)

Galanda, M., & Weibel, R. (2003). Using an energy minimization technique for polygon generalization. *Cartography and Geographic Information Science*, 30(3), 263–279. https://doi.org/10.1559/152304003100011199 (cit. on p. 10)

Huot-Vézina, G., Boivin, R., Smirnoff, A., & Paradis, S. J. (2012). *Geoscaler: Generalization tool (with a supplementary user guide in french)* (2nd ed.), Geological Survey of Canada. https://doi.org/10.4095/291993. (Cit. on pp. 16 sq.)

Mackaness, W. A., & Ruas, A. (2007). Chapter 5 - evaluation in the map generalisation process. In W. A. Mackaness, A. Ruas, & L. T. Sarjakoski (Eds.), *Generalisation of geographic information* (pp. 89–111). Elsevier Science B.V. https://doi.org/https://doi.org/10.1016/B978-008045374-3/50007-7. (Cit. on pp. 13 sq.)

Maltman, A. (1990). *Geological maps: An introduction*. Springer. https://doi.org/https://doi.org/10.1007/978-1-4684-6662-1. (Cit. on p. 1)

McMaster, R. B., & Shea, K. S. (1992). *Generalisation in digital cartography*. Association of American Geographers. (Cit. on pp. 4, 8).

Measuring compactness. (n.d.). Retrieved August 22, 2020, from https://fisherzachary.github.io/public/r-output.html. (Cit. on p. 14)

Monmonier, M. S. (1983). Raster-mode area generalization for land use and land cover maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 20(4), 65–91. https://doi.org/10.3138/X572-0327-4670-1573 (cit. on p. 6)

Patera, A. (2006). *Advanced techniques for the management of geological mapping* (PhD dissertation). The Open University. https://doi.org/10.21954/ou.ro.0000fb1a. (Cit. on p. 6)

Peter, B., & Weibel, R. (1999). Using vector and raster-based techniques in categorical map generalization. In *3rd ica workshop on progress in automated map generalization*. https://

kartographie.geo.tu-dresden.de/downloads/ica-gen/workshop1999/peter_weibel.pdf. (Cit. on pp. 5, 8–11)

Polsby, D. D., & Popper, R. D. (1991). The third criterion: Compactness as a procedural safeguard against partisan gerrymandering. *Yale Law & Policy Review*, 9(2), 301–353 (cit. on p. 14).

Regnauld, N., & McMaster, R. B. (2007). Chapter 3 - a synoptic view of generalisation operators. In W. A. Mackaness, A. Ruas, & L. T. Sarjakoski (Eds.), *Generalisation of geographic information* (pp. 37–66). Elsevier Science B.V. https://doi.org/https://doi.org/10.1016/B978-008045374-3/50005-3. (Cit. on pp. 8–10)

Roth, R. E., Brewer, C. A., & Stryker, M. S. (2011). A typology of operators for maintaining legible map designs at multiple scales. *Cartographic Perspectives*, (68), 29–64. https://doi.org/10.14714/CP68.7 (cit. on p. 8)

Ruas, A., & Plazanet, C. (1996). Strategies for automated generalization. In *7th international symposium on spatial data handling*. https://www.oicrf.org/-/strategies-for-automated-generalization. (Cit. on pp. 8, 10)

Sarjakoski, L. T. (2007). Chapter 2 - conceptual models of generalisation and multiple representation. In W. A. Mackaness, A. Ruas, & L. T. Sarjakoski (Eds.), *Generalisation of geographic information* (pp. 11–35). Elsevier Science B.V. https://doi.org/https://doi.org/10.1016/B978-008045374-3/50004-1. (Cit. on pp. 4, 6)

Savino, S. (2011). *A solution to the problem of the generalization of the italian geographical databases from large to medium scale: Approach definition, process design and operators implementation* (PhD dissertation). University of Padova. https://kartographie.geo.tu-dresden.de/downloads/ica-gen/publications/savino.pdf. (Cit. on p. 6)

Sayidov, A. (2021). *An integration methodology for automated generalization of geological maps* (PhD dissertation). University of Zurich. https://www.zora.uzh.ch/id/eprint/213697/. (Cit. on pp. 4, 6, 8, 10 sq.)

Sayidov, A., Aliakbarian, M., & Weibel, R. (2020). Geological map generalization driven by size constraints. *ISPRS International Journal of Geo-Information*, 9. https://doi.org/10.3390/ijgi9040284 (cit. on pp. 3, 6)

Schmid, S. (2008). *Automated constraint-based evaluation of cartographic generalization solutions* (Master's thesis). University of Zurich. Switzerland. https://www.geo.uzh.ch/dam/jcr:a1dfc9a8-dfc8-49d3-9dc0-f519490836c9/msc_stefan_schmid_2008.pdf. (Cit. on p. 13)

Schuff, J. (2019). *Verfahren zur automatisierten Generalisierung flächenhafter Geofachdaten* (Master's thesis). UNIGIS. http://unigis.sbg.ac.at/files/Mastertheses/Full/104640.pdf. (Cit. on pp. 17, 24)

Schylberg, L. (1992). Cartographic amalgamation of area objects. In L. W. Fritz & J. R. Lucas (Eds.), *Xviith isprs congress. technical commission iv: Cartographic and data base applications of photogrammetry and machine vision* (pp. 135–138). ISPRS Archives. https://www.isprs.org/proceedings/xxix/congress/part4/135_XXIX-part4.pdf. (Cit. on p. 6)

Sester, M. (2020). Cartographic generalization. *Journal of Spatial Information Science*, 21. https://doi.org/10.5311/josis.2020.21.716 (cit. on p. 1)

Shea, K. S., & McMaster, R. B. (1989). Cartographic generalization in a digital environment: When and how to generalize. *9th Auto-Carto: International Symposium on Computer-Assisted Cartography*. https://cartogis.org/docs/proceedings/archive/auto-carto-9/pdf/

cartographic-generalization-in-a-digital-environment-when-and-how-to-generalize.pdf (cit. on p. 8)

Smirnoff, A., Huot-Vézina, G., Paradis, S. J., & Boivin, R. (2012). Generalizing geological maps with the geoscaler software: The case study approach. *Computers & Geosciences*, 40, 66–86. https://doi.org/https://doi.org/10.1016/j.cageo.2011.07.013 (cit. on pp. 1, 3, 7, 16)

Smirnoff, A., Paradis, S. J., & Boivin, R. (2008). Generalizing surficial geological maps for scale change: Arcgis tools vs. cellular automata model. *Computers & Geosciences*, 34(11), 1550–1568. https://doi.org/https://doi.org/10.1016/j.cageo.2007.10.013 (cit. on p. 7)

Steiniger, S., & Weibel, R. (2005). A conceptual framework for automated generalization and its application to geologic and soil maps. In *22nd international cartographic conference*. https://www.researchgate.net/publication/228918417_A_conceptual_framework_for_automated_generalization_and_its_application_to_geologic_and_soil_maps. (Cit. on pp. 1, 8)

Stoter, J., Post, M., van Altena, V., Nijhuis, R., & Bruns, B. (2014). Fully automated generalization of a 1:50k map from 1:10k data. *Cartography and Geographic Information Science*, *41*(1), 1–13. https://doi.org/10.1080/15230406.2013.824637 (cit. on p. 10)

Stoter, J., Zhang, X., Stigmar, H., & Harrie, L. (2014). Evaluation in generalisation. In D. Burghardt, C. Duchêne, & W. Mackaness (Eds.), *Abstracting geographic information in a data rich world: Methodologies and applications of map generalisation* (pp. 259–297). Springer International Publishing. https://doi.org/10.1007/978-3-319-00203-3_9. (Cit. on p. 22)

Su, B., Li, Z., & Lodwick, G. (1997). Morphological transformation for the elimination of area features in digital map generalization. *Cartography*, *26*(2), 23–30. https://doi.org/10.1080/00690805.1997.9714064 (cit. on p. 7)

Su, B., Li, Z., & Lodwick, G. (1998). Morphological models for the collapse of area features in digital map generalization. *GeoInformatica*, *2*(4), 359–382. https://doi.org/10.1023/A:1009757422454 (cit. on p. 7)

Toffoli, T., & Margolus, N. (1987). *Cellular automata machines: A new environment for modeling*. MIT Press. https://doi.org/https://doi.org/10.7551/mitpress/1763.001.0001. (Cit. on p. 7)

Visvalingam, M., & Whyatt, J. D. (1993). Line generalisation by repeated elimination of points. *The Cartographic Journal*, 30(1), 46–51. https://doi.org/10.1179/000870493786962263 (cit. on p. 6)

Ware, J. M., Jones, C. B., & Thomas, N. (2003). Automated map generalization with multiple operators: A simulated annealing approach. *International Journal of Geographical Information Science*, *17*(8), 743–769. https://doi.org/10.1080/13658810310001596085 (cit. on pp. 10 sq.)

Weibel, R. (1997). Generalization of spatial data: Principles and selected algorithms. In M. van Kreveld, J. Nievergelt, T. Roos, & P. Widmayer (Eds.), *Algorithmic foundations of geographic information systems* (pp. 99–152). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-63818-0_5. (Cit. on p. 4)

Weibel, R., & Dutton, G. (1999). Generalising spatial data and dealing with multiple representations. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical information systems: Principles, techniques, management and applications* (Second, pp. 125–155).

Wiley. https://www.geos.ed.ac.uk/~gisteac/gis_book_abridged/files/ch10.pdf. (Cit. on p. 4)

Zhang, X., Stoter, J., Ai, T., Kraak, M.-J., & Molenaar, M. (2013). Automated evaluation of building alignments in generalized maps. *International Journal of Geographical Information Science*, 27(8), 1550–1571. https://doi.org/10.1080/13658816.2012.758264 (cit. on pp. 10 sq.)