An Integrated Methodology for Automated Generalization of Geological Maps

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

Der Universität Zürich

von

Azimjon Sayidov

aus

Usbekistan

Promotionskommission

Prof. Dr. Robert Weibel (Vorsitz)

Prof. Dr. Ross Purves

Prof. Dr. Stefan Leyk

Zürich, 2021

Abstract

Geological maps are one of the most complex thematic maps. They represent spatial and temporal interaction between rocks, earth materials, and landforms at the surface of the Earth. These interactions constitute the essence of geological maps. Geological maps mainly consist of interconnected polygonal features in various shapes distributed in space, representing surface rocks and deposits. It is essential to preserve these relationships between polygons during map generalization.

Most of the research on automated map generalization has been dedicated to topographical map generalization. Moreover, existing approaches proposed to deal with the generalization of geological maps often overlook peculiarities, such as the size, shape, and orientation of the features on the map.

The aim of this thesis is to develop an integrated methodology for the automation of geological map generalization. The approach aims to ensure the readability of the map during its generalization while preserving the typical traits of the map. In response to limitations of the existing approaches for the generalization of geological maps, a three-step-process is proposed that deals with polygonal features on the map. These three steps are logically sequential and complement each other.

In the first step, polygonal features are separated into the so-called foreground and background polygons. Foreground polygons (also referred to as island polygons) are small and free-standing polygons that represent geological units that are small in area but often important due e.g. to their content in valuable minerals. Background polygons represent large and bulky polygons that represent the predominant (regarding area) geological units of the map. Moreover, in this step, the so-called size constraints that deal with island polygons are identified. The size constraints are dedicated to handling an individual polygon, or pairs of map features with the minimum area and distance relations. These constraints are used to control the generalization process.

The second step is devoted to the recognition of patterns and identification of groups from the foreground polygons in the geological map. Identified polygon groups allow to ensure the preservation of patterns, such as clusters or alignments of island polygons, during the generalization process. Patterns are recognized by first constructing the Delaunay Triangulation, followed by removing global and local long edges. The network is then further refined to form the final groups based on similarities in polygon size, shape, and orientation, and the category of the geological unit.

Abstract

The third step combines the advantages of the two previous steps by utilizing the size constraints to generalize the polygon groups identified in the second step. Moreover, two generalization operators, aggregation and typification, are employed to generalize the polygon groups, while preserving their main traits. Quantitative evaluation is used to assess the quality of the generalization.

The proposed methodology for automation of geological maps maintains the readability of the map during their generalization process. Moreover, the relationships between polygons are preserved, as the pattern recognition and group identification step informs the actual generalization process. Also, the approach minimizes alteration of polygon size, shape and orientation. The methodology was developed with geological maps in mind, but has the potential to generalize other categorical maps, such as soil and vegetation maps, with small fine-tuning in the constraint definition and goal value modification.

Based on the results and open issues of this work, several areas of future research on geological map generalization can be identified. Further research is needed to formalize a comprehensive list of constraints for the complete generalization process of geological maps. Moreover, the alignment of polygons and their density must be included in pattern recognition and group identification. The current research is primarily focused on the generalization of polygonal features. However, more research is needed to generalize geological maps, including all the map layers, such as points, linear objects, symbols, etc.

Acknowledgments

Life in Switzerland would not have been exciting, and researching at the University of Zurich would not have been enthusiastic without the support and encouragement of innumerous people. I want to express my gratitude for encouraging and supporting me in the past years.

To the greatest extent, I am heartfully grateful to my main supervisor, Prof. Dr. Robert Weibel, whose considerate and wholehearted approach I felt even before coming to Switzerland. His kind assistance and wise counseling helped me and my family adapt to life in Switzerland comfortably. I want to thank him for providing me with freedom in my research yet patiently supporting and encouraging me when I got stuck, slowed down, and deviated off-topic. I cannot thank you enough, but I still say ‘Danke schön, Roby!

I would like to thank Prof. Dr. Stefan Leyk, for his assistance in breaking down my extensive research tasks into smaller and manageable parts. Moreover, I would like to thank him for his patient English editing and reviewing my paper. I want to express my gratitude to Prof. Dr. Ross Purves for helping me move forward with his advice and constructive feedback.

I am grateful to Dr. William Mackaness for acting as external reviewer of this thesis.

I want to thank ‘Dr.’ Meysam Aliakbarian for meeting me at Zurich Airport when I first arrived in Switzerland and being my ‘personal’ guide to Zurich city while explaining the basic necessities of living in Zurich. Thanks for being a travel buddy all over Europe and sharing memorable time, especially travel to Italy. Thanks, Meysamjon, for time over tea, engaging, and funny discussions on and off the research, especially in the ‘Gentlemen's Club.’

My gratitude goes to Dr. Peter Ranacher, who always appears in a supportive mode and finds arguments to carry on in a better way. I enjoyed his open-minded and humorous attitude in everyday discussions. Thanks for being the best ‘Österreichischer Alter’ and introducing me to colorful stuff like floating in a ‘Gummy’ Boat, playing tennis, and even yoga. Thanks for sharing beautiful moments and unforgettable experiences with me.

I was blessed with two delightful and engaging officemates, Hoda and Oliver. Thanks for all the discussions and jokes. You are the best officemates!

Thanks to all my friends and colleagues in the GIUZ for your wonderful and involving moments, and thanks for making the working environment enjoyable at the Irchel Campus – Annica, Katya, Olga, Michelle, Pia, Beni, Alex, Diego, Shivangi, Haosheng, Raha, Max, Elise, Curdin, Cheng, E-K, Kiran, Flurina, Nico, Zhiyong, George, Ali, Peter, Sasha, Annina, Armand, Bingjie, Sara.

Acknowledgments

Special thanks to my colleagues in Uzbekistan, Prof. Safarov, Prof. Muborakov, Prof. Mirzaliev, Asom aka, Ilkom aka for supporting me throughout my studies. Thanks to Olimjon aka, Shavkat aka, Dilshod and other colleagues for always being there for me. My thanks go to my colleague and friend Juliev Muhiddin for constant support and recommendations on geological mapping.

Football buddies, Henki, Alex, Daniel, Christian, Christophe, Samuel, Josh, Andrew, Orif, and others, have also deserved a special place in my heart to help me get rid of my stress after work with friendly yet competitive games.

I want to express my heartful gratitude to my loving parents for upbringing me and supporting me in all my ups and downs while waiting for me to pursue my doctorate. I would like to thank my siblings Gulbahor, Rahim, Dilora, for encouraging me all the time, and all my nephews and nieces, Mohinur, Abdurauf, Donoxon, Yusuf, and Miri.

Finally, and most of all, my wholehearted and warmest gratitude belongs to my beloved Nargiz Asaltoyim, who is there for me all the time with anything I need. You are the best thing that ever happened to me. Thanks for giving me Do'mboq Bachchajomin, Kamronjon, who always puts a smile on my face, pushes me to do 'cool stuff' like playing football indoors and outdoors, winning me on the car racing game on the PS, biking, hiking, tennis, table football, etc. Love you so much.

Contents

Contents ---------- vii
Abstract ---------- iii
Acknowledgments ---------- v
Abbreviations ---------- ix
1. Chapter: Introduction ---------- 3
1.1. Motivation ---------- 3
1.1.1. Map Generalization and its Automation ---------- 3
1.1.2. Problem Definition ---------- 5
1.2. Theoretical Background ---------- 6
1.2.1. Map Generalization ---------- 6
Definition of Generalization ---------- 6
Causes of Generalization ---------- 7
Generalization Operators ---------- 8
1.2.2. Geological Maps and Geological Mapping ---------- 11
Geological Maps ---------- 11
Types and Scales of Geological Maps ---------- 11
Geological Mapping ---------- 12
1.3. General Objective, Methodology, and Research Objectives ---------- 14
1.3.1. General Objective and Methodology ---------- 14
1.3.2. Research Objectives ---------- 17
1.3.3. Structure of the Thesis ---------- 17
1.4. State of the Art in Geological Map Generalization ---------- 19
1.4.1. Constraint-Based Map Generalization ---------- 19
1.4.2. Pattern Recognition in Map Generalization ---------- 20
1.4.3. Geological Map Generalization ---------- 22
Raster and Vector Formats ---------- 22
Raster-Based Generalization in Categorical Maps ---------- 22
Vector-Based Generalization in Categorical Maps ---------- 23
1.4.4. Research Challenges Addressed in this Thesis ---------- 24
2. Chapter: Generalization by Size Constraints ---------- 29
3. Chapter: Recognition of Polygon Groups ---------- 63
4. Chapter: Aggregation and Typification of Polygon Groups ---------- 87
5. Chapter: Synthesis ---------- 129
5.1. Discussion of Research Objectives ---------- 129
5.1.1. Research Objective 1: Separate and Generalize Island Polygons in Geological Maps by Incorporating Size Constraints ---------- 129

Contributions ----------129
Findings ----------130
Limitations----------130

5.1.2. Research Objective 2: Identify Groups of Small Polygons Based on Similarity Criteria, Forming Important Spatial Structures, as a Basis for Implementing Contextual Generalization Operators ---------- 131
Contributions ----------131
Findings ----------131
Limitations----------131

5.1.3. Research Objective 3: Develop Contextual Generalization Algorithms and Evaluation of Generalization Results of Polygon Groups in Geological Maps ---------- 132
Contributions ----------132
Findings ----------133
Limitations----------138

5.2. Adaptability of the Approach to other Categorical Maps ---------- 140
5.3. Future Directions and Outlook ---------- 141

Bibliography ---------- 145
Curriculum Vitae ---------- 157
Publications ---------- 159
Appindix ---------- 161

Abbreviations

OS – object separation

MA – minimum area

DT – Delaunay triangulation

NN – nearest neighbor

NSW – New South Wales

ICA – International Cartographic Association

DEM – digital elevation model

BGS – British Geological Survey

CA – cellular automata

IPQ – ipsometric quotient

OM – original map

FG – fine-grained

CM – compromise

CG – coarse-grained

MST – minimum spanning tree

IQR – inter-quartile range

UQ – upper quartile

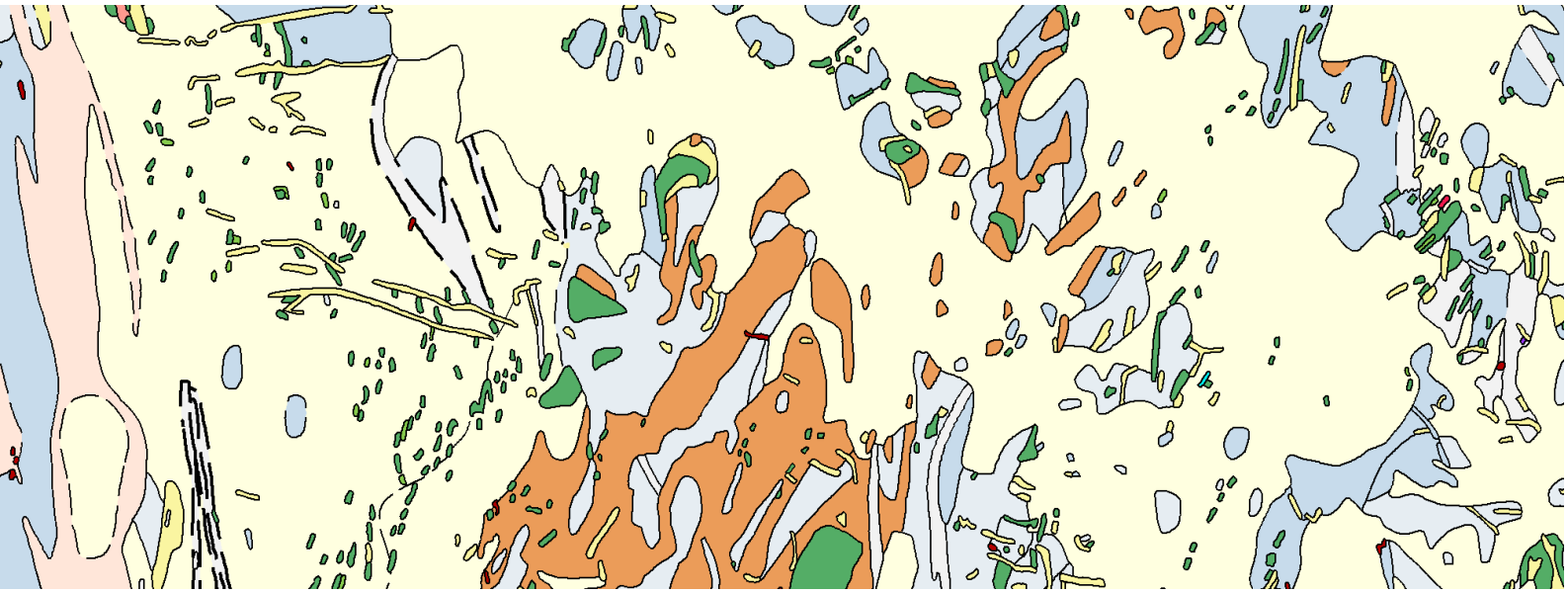
LQ – lower quartile

IUF – inner upper fence

GM – global mean

LM – local mean

CV – coefficient of variation



1

Chapter

I wisely started with a map.

J. R. R. Tolkien

1. Introduction

1.1. Motivation

1.1.1. Map Generalization and its Automation

Map generalization (short: generalization) is a crucial element of cartography. Traditionally, map generalization is responsible for reducing complexity in a map during the process of scale reduction, which is induced by the fact that through scale reduction, the available map space is also reduced. Generalization emphasizes the essential while suppressing the unimportant; it aims to maintain logical and unambiguous relations between map features and preserve the aesthetic quality of the map. The main objective of generalization is to create maps of high graphical clarity to perceive the map image easily, such that the message the map intends to deliver can be readily comprehended (Mackaness & Chaudhry, 2008).

Map generalization and its automation have been under investigation for decades (Brassel & Weibel, 1988; McMaster & Shea, 1992; Baella & Pla, 2005; Baella et al., 2007; Mackaness et al., 2007; Lecordix et al., 2007; Lecordix & Lemarié, 2007; Foerster et al., 2010; Revell et al., 2011; Mackaness et al., 2011; Duchêne et al., 2014; Stoter et al., 2014; Dumont et al., 2015; Mackaness et al., 2016; Burghardt et al., 2016; Šuba, 2018; Karsznia et al., 2020). However, most research to date focused, and continues to focus, on topographic maps, the most common type of map. Topographic maps include national maps such as those produced by the national mapping agencies (NMA) such as swisstopo (the Swiss NMA), BKG (the Federal Agency for Cartography and Geodesy, Germany), or IGN (National Geographic Institute, France), but also online maps such as Google Maps or OpenStreetMap. The generalization of specialized, thematic maps, such as so-called categorical maps (e.g., soil maps, vegetation maps, landuse maps etc.) has been significantly less in the focus of map generalization research, especially that of geological maps (Galanda, 2003; Downs & Mackaness, 2002; Steiniger & Weibel, 2005).

Categorical maps are a space exhaustive tessellation of space with discrete boundaries (Burrough & Frank, 1996). Although they are visualized by discrete boundaries, in reality, the boundaries are often better represented as fuzzy, and their shape depends on the classification used (Edwardes & Mackaness, 2000). Geological maps consist of polygonal subdivisions that illustrate the spatial and temporal interactions of rocks and landforms at the Earth's surface. These rocks and formations bear a unique natural history of the depicted area and provide valuable information on mineral resources, such as the concentration of coal, oil, uranium, or gold. As geological maps illustrate natural

phenomena, they contain various arbitrary shapes and structures, including small and compact, bulky, concave or convex, long and narrow, straight, or twisted polygons. Figure 1.1 shows some examples of shapes found in the geological map of the Euriowie Block, New South Wales, Australia. Usually, bulky polygons represent more recent sedimentary geological formations. They cover a larger portion of the surface than the mineral-rich geological units, and organic particles accumulate over millions of years and cover much older rocks underneath. The large background polygons in Figure 1.1 (light yellow) indicate metasedimentary rocks. These rocks are formed through the solidification of sediments under high pressure and temperature (Vernon & Clarke, 2008). Long and narrow, detached polygons (light red color in Figure 1.1) represent exposed layers from underneath caused by erosion or tectonic plate shift. These formations are detached because they break over time due to the pressure and compression of plates (Burton, 1998, 2001; Katz & Tuckwell, 1979; Willis et al., 1983). Twisted, convex, and concave shapes (light green and green colors) are deformed by clashes and compression of plates. They have a twisted or bent shape because, rather than breaking under pressure, they bend due to the more flexible nature of the particular rock unit. These are metamorphic rocks, such as amphibolite and basic granulates, formed under high pressure and temperature caused by tectonic movements (Brown et al., 1992; Cox et al., 1979). The small polygons (green color) are ‘amphibolite’ rocks exposed to the Earth’s surface by erosion (Vernon & Clarke, 2008). While these rock units seem to be randomly distributed on the surface and have various shapes, they are related to each other and share common properties in their formation. It is thus essential to preserve these relationships and properties during the generalization process.

The remainder of this section gives an overview of the objectives of geological map generalization and defines and motivates the research problem. Section 1.2 then addresses the necessary background on map generalization, geological maps, and geological mapping in more detail. A summary of the methodology and research objectives are provided in Section 1.3, along with the thesis structure. Section 1.4 outlines the state of the art in geological map generalization and summarizes the thesis.

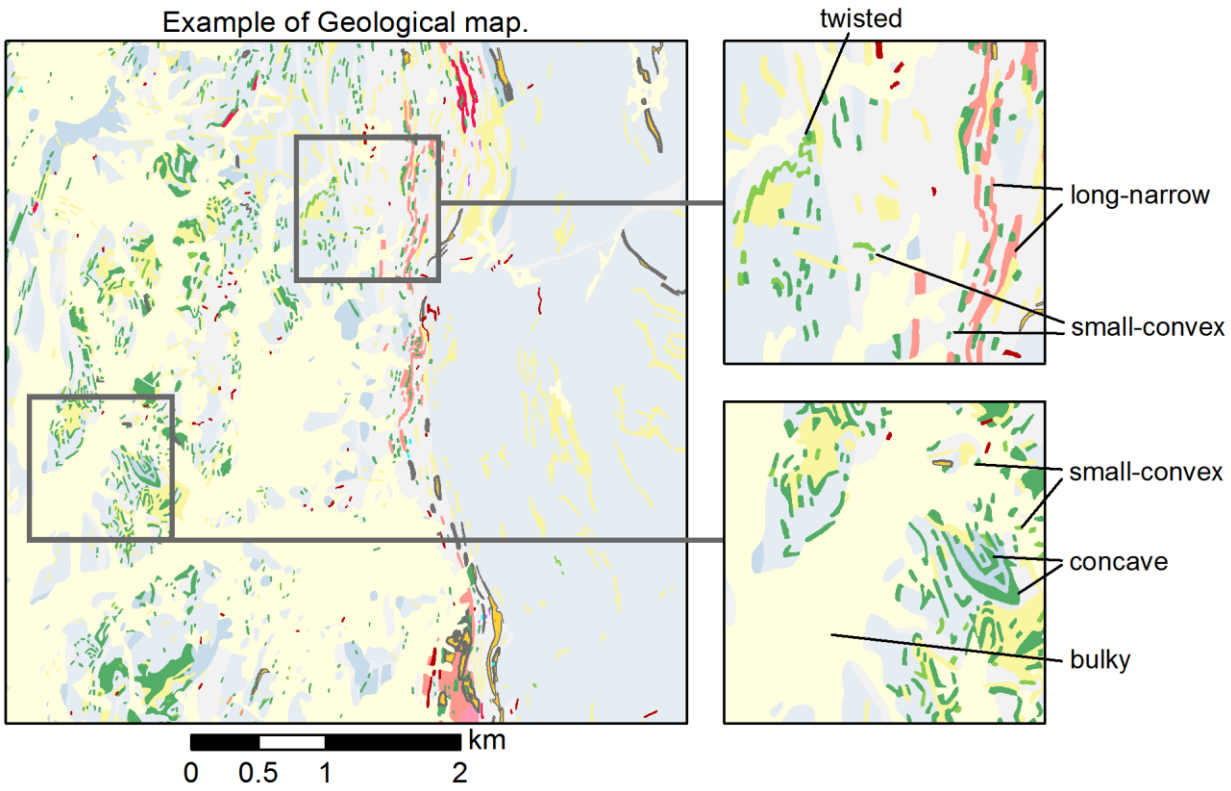


Figure 1.1. Euriowie 1:25 000 Geological Map – part of the Euriowie Block, NSW, Australia and two inset maps. (Stevens et al., 1998, 2008). https://search.geoscience.nsw.gov.au/product/251.

1.1.2. Problem Definition

Geological maps are packed with unusual shapes and structures that — unlike the anthropogenic shapes, such as building footprints found in topographic maps — hardly resemble simple geometrical shapes such as circles, rectangles, or triangles. Some examples are highlighted in Figure 1.1. Because of their complex nature and the less frequent use, the automation of the generalization of geological maps has largely been overlooked by research to date. Existing approaches do not adequately address the properties mentioned above and their peculiarities.

Map generalization must maintain readability at all scales without losing relevant content and altering the central message of the mapped geological events. Thus, the initial step to automate map generalization should consider essential cartographic requirements. Specifically, it must address constraint identification, modeling, and implementation (e.g., of legibility limits). Moreover, the polygons present on geological maps may share similarities, for instance in shape, size, orientation, and similar geological properties. Ignoring similarities during generalization may alter the main structures of the map. Methods for measuring and identifying these similarities and grouping map objects accordingly must be developed and integrated into the generalization of geological maps.

Furthermore, adequately treating the identified groups of map objects call for competent decision making and orchestration of generalization operators and algorithms.

As many authors have pointed out (Bader & Weibel, 1997; Regnauld et al., 1999; Kilpeläinen, 2000; Regnauld, 2001; Galanda, 2003; Steiniger & Meier, 2004; Duchêne et al., 2005; Harrie & Weibel, 2007; Song et al., 2009), not using structural knowledge, such as knowledge about groups of similar polygons, or alignments of polygons, can lead to improper generalization and unintended loss of content and patterns. Also, the generalization of groups of similar map objects requires structural knowledge that can adapt to local variations of patterns, facilitating the adaptability and reusability of generalization algorithms. Increasing flexibility and ensuring reusability is possible with a modular approach, which combines different, suitable generalization algorithms. For instance, the user should always be able always replace an algorithm with a more efficient or better suited one when required. In Section 1.3, we propose an integrated and modular methodology that considers the challenges mentioned above.

1.2. Theoretical Background

1.2.1. Map Generalization

Every single map has one common property: it represents a generalized, abstracted view of reality (Ormeling, 2013; Robinson, 2020). Maps can never express the full detail of the real world. They need to abstract, approximate, and generalize spatial information, preserving as much detail as possible while keeping the representation readable (and therefore removing unnecessary detail) (Keates, 2014; Monmonier, 2018). This dilemma is at the very foundation of ‘generalization’ and has been a persistent issue since the beginning of cartography (Dorling & Fairbairn, 2013). The production of maps is driven by various principles of generalization, requirements, and controls (McMaster & Shea, 1988; Beard, 1991; Ruas, 1999; Bard, 2004; Barrault et al., 2001; Ware et al., 2003; Burghardt & Neun, 2006; Stern & Sester, 2012; Stoter et al., 2014). Below, we will highlight the main properties, types, and causes of map generalization and define commonly used generalization operators.

Definition of Map Generalization

The International Cartographic Association (ICA) defines map generalization as “the selection and simplified representation of detail appropriate to the scale and/or purpose of the map” (ICA 1973). In its core, map generalization is the process of abstraction, conveying patterns and associations among a set of geographic phenomena, and aims at communicating a particular geography of the world effectively (Mackaness et al., 2016).

Introduction

The term ‘generalization’ has its roots in the Latin word ‘generalis’, meaning general or primary. The purpose of generalization is to transmit information about relevant spatial features and characteristics and show their relationship. Already in large-scale field mapping, the topographer makes an initial selection, including only certain desirable aspects of reality (e.g., relief, roads, etc.) and ignoring unimportant ones. From these, cartographers then develop medium and small-scale maps, continuously and increasingly omitting details. As the scale changes linearly, the area of the map representation changes with the square of the distance, leading to an aggravating competition for space among map objects.

Causes of Map Generalization

There are multiple reasons for generalization, such as the scale, purpose, topic (or type) of the map, or the method of visualization, all of which define the conditions and possible approaches used in the generalization process (Shea & McMaster, 1989; Müller, 1990; Kilpeläinen, 1997, 2000; Weibel & Dutton, 1998, 1999).

Each map is bound to a particular scale, which dictates the amount of information that can be shown on the map. The transition from large-scale to small-scale maps shrinks the map area; thus, selecting essential map objects and omitting minor detail is inevitable (Steiniger, 2007; Ai & van Oosterom, 2002; Neun et al., 2004; Stanislawski & Savino, 2011). The importance of map objects is linked to scale. Essential objects in a large-scale map (i.e., local landmarks such as shops, stations, etc.) may become relatively unimportant in a small-scale map and be omitted (McMaster & Shea, 1989; Sester, 2000).

As scale dictates the amount of information that can be displayed in the target map, the map’s purpose decides which objects are relevant and need to be included and which ones can be omitted. Relevant features contribute to the map’s purpose while unimportant ones do not, and if included, will hinder map comprehension (McMaster & Shea, 1989). For instance, the purpose of a political-administrative wall map for schools is to demonstrate essential features, such as main cities, main roads, and boundaries, which are indicated with larger symbols. However, if the same political-administrative map has the purpose of being a reference map, then it will contain as much information as possible, such as administrative divisions, settlements, and communication systems.

The topic (or type) of a map (e.g., landuse map, road map, geological map) indicates which objects need to be represented in detail and which ones need less attention and can be abstracted. For example, in a geological map, we need to emphasize water bodies and river networks as these features are directly linked to the topic (i.e., the geological formations presented). However, road networks and settlements are less relevant and can be largely abstracted or even omitted.

Generalization Operators

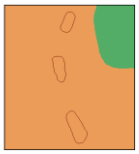
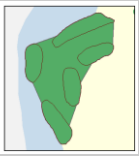
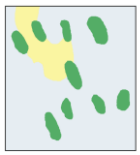
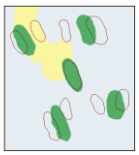
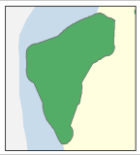
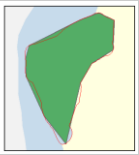
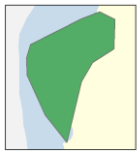
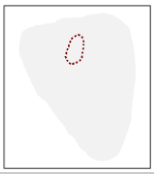
McMaster and Shea (1992) proposed a model that divides the generalization process into three phases by asking three simple questions: Why to generalize? When to generalize? How to generalize? The question ”Why to generalize?” is related to the ‘philosophical objectives’: why generalization is desirable. The Question ”When to generalize?” arises ideally due to the success (or failure) of the map product to meet its stated goal: maintaining clarity, with appropriate content, at a given scale, for a chosen map purpose and an intended audience (McMaster & Shea, 1989). The question ”How to generalize?” is related to the spatial and attribute transformations of a map object during generalization, carried out using a so-called *generalization operator*.

Generalization operators impose one or several actions in response to conflicts that occurred during scale reduction: for instance, map objects that are too small to be legible, or too close to each other to be visually distinguishable). In general, conflicts are directly linked to the goals of cartographic design, which aims at preserving as much information as possible, yet maintaining the readability and clarity of the map. Generalization operators conceptually define the action that is caused in response to a conflict (Galanda, 2003; Ruas & Plazanet, 1996; Regnauld & McMaster, 2007). In the above examples, ‘enlargement’ could be used to make small map objects large enough to be clearly visible, while ‘aggregation’ could be used to move map objects such that they can be visually distinguished. However, the conceptual definition of a generalization operator also needs a concrete implementation (Weibel, 1996; Galanda, 2003; Steiniger, 2007), and that is accomplished by a *generalization algorithm*. Since generalization operators are only conceptually — and usually rather loosely — defined, typically several algorithms exist to implement a particular generalization operator. For instance, there are usually different ways to aggregate a set of map objects, such as a group of polygons in a geological map, resulting in different outputs.

Two fundamental aspects need to be considered when applying generalization operators: ***operator selection*** and ***operator sequencing***. ***Operator selection*** is a core part of the map generalization process (Ruas & Plazanet, 1996). It may be based on the map’s purpose, the importance of the individual feature on a map, and the relationships between of map objects (e.g., whether they form a cluster of polygons that might lend themselves for aggregation). ***Operator sequencing*** is the logical ordering of generalization operators in sequences, such as ‘selection’ → ‘enlargement’ → ‘displacement’. The selection operator forms a logical start for any generalization process, controlling the amount of detail represented on the target map, and creating space on the map for the following operators (Ruas & Plazanet, 1996; Galanda, 2003; Steiniger, 2007). The enlargement operator then ensures that all map objects are clearly visible. After enlargement, however, some map objects might be too close to each other to be visually discernable, necessitating displacement.

Generalization operators vary based on the type of map. In this thesis, I will mainly focus on operators for geological map generalization, which are a representative of polygonal maps (geometrically speaking) and categorical maps (thematically speaking). The list of commonly used operators for categorical maps proposed by Galanda (2003) is presented in Table 1.1, together with their type, application domain, and definition. Based on structural transformation that they perform, operators are divided into two categories: *Spatial operators* transform spatial properties of map features and modify the graphical representation of features to fit the map's scale-specific constraints (e.g., enlarging a polygon to meet the minimum area constraint and thus make the polygon large enough to be neatly visible). *Semantic operators* transform semantic attributes altering the map's underlying statistical properties (Galanda, 2003; Cecconi, 2003; Steiniger, 2007). The reclassification operator exploits classification hierarchies that often exist in categorical data, and specifically in geological maps classification systems. For instance, the five rock types 'Hornblende + plagioclase + quartz amphibolite,' 'Garnet amphibolite', 'Epidote bearing amphibolite', 'Orthopyroxene bearing amphibolite', and 'Quartz rich amphibolite' can be merged to the group 'Amphibolite and Basic Granulite' at the next higher level of the classification hierarchy.

Table 1.1. List of operators implemented in categorical mapping (Galanda, 2003)

Type of operators	Application domain	Name of operators	Definition	Implementations: Before	Implementations: After
Spatial operators	Polygons	Elimination	Removes polygons from the map by dissolving them into the background polygons		
Spatial operators	Polygons/group of polygons	Enlargement	Enlarges small polygons to make them readable on the target map		
Spatial operators	Polygons/group of polygons	Displacement	Moves polygons away from each other to make them distinguishable		
Spatial operators	Group of polygons	Aggregation	Merges a group of polygons with the same or similar categorical properties		
Spatial operators	Group of polygons	Typification	Reduces the number of polygons and, thus, the density of a cluster of polygons by removing, enlarging, displacing objects while maintaining their typical arrangement		
Spatial operators	Outline of polygon	Simplification	Reduces the granularity of the outline of a polygon		
Spatial operators	Outline of polygon	Smoothing	Improves the aesthetic appearance of polygonal features		
Semantic operators	Map	Reclassification	Changes the category of a polygonal feature to other polygons with a similar category		

1.2.2. Geological Maps and Geological Mapping

Geological Maps

A geological map is a scaled-down interpretation of the geological structure of a selected area of the upper part of the Earth's crust, usually drawn on a topographic base map (Maltman, 2012) (sometimes also on a Digital Elevation Model; DEM). Using various colors and symbols, geological maps show the rock formations and boundaries between them that would be seen on the Earth's surface if all soils were removed (Roberts, 2013; Lisle, 2020). Geological maps are used to

- explore natural resources (raw material, underground water);
- locate rocks of a particular age, lithology, or structure;
- reconstruct the geological history of an area;
- estimate the composition and character of the soil and;
- identify natural hazards.

Types and Scales of Geological Maps

There are two types of geological maps: general and specialized maps. General geological maps consist of base geological maps and maps of mineral resources. Specialized maps focus on a particular branch of geology, i.e., quaternary formations; lithology; petrography; metamorphic rocks and tectonics; geological structure; geodynamics; geomorphology (Barnes & Lisle, 2013).

Based on the scale, geological maps are divided into five types (Lisle et al., 2011; Barnes & Lisle, 2013; Roberts, 2013; Lisle, 2020).

Detailed geological maps, with a scale of 1:10,00 and larger, are compiled onto a topographic base map. They reflect the geology of small regions with mineral deposits. Detailed geological maps are used to construct industrial structures and mines, and these maps are accurately recorded using topographic-geological surveying.

Large-scale maps, with a scale between 1:25,000 and 1:50,000, are either directly compiled onto topographical base maps or derived by generalization from detailed maps and visualize detailed geological structures and mineral resources of smaller regions. These maps are also used to solve engineering tasks, such as constructing hydropower plants or city planning. Usually, these maps are accompanied by geological cross-sections along with an explanatory note.

Medium-scale maps, with a scale between 1:100,000 and 1:200,000, are derived by generalization from detailed or large-scale geological maps and displayed on simplified topographic base maps.

They depict characteristics of a region's geological structure, the distribution of mineral resources, and their forecasting assessment.

Small-scale maps, with a scale between 1:500,000 and 1:1,000,000, are derived by generalization from maps of larger scales and displayed on a simplified topographic base map. They represent the general geological structures and the distribution of minerals in a region or a country.

Overview maps, with a scale smaller than 1:1,000,000, are derived by generalization from small-scale maps and displayed onto significantly simplified topographic base maps. They represent general geological formations across regions, countries, continents, or the entire Earth.

Geological Mapping

Geological surveying aims to identify and track geological boundaries resulting from direct observation of geological formations. Geological surveying usually results in detailed geological maps typically compiled at scales of 1:1,000 to 1:10,000 (see above), which are then used to derive smaller scale maps employing scale reduction and cartographic generalization (Lisle et al., 2011; Barnes & Lisle, 2013; Roberts, 2013; Lisle, 2020; Spencer, 2017).

Remote sensing images are also used to derive geological maps. Image processing, enhancement, and classification methods can be applied to extract geological boundaries from the image data. These processes are complex and need expert input as well as validation by geologists to confirm that the information derived from remote sensing imagery is accurate (Spencer, 2017).

Generalization of geological maps is the process of obtaining geological maps at smaller scales through the abstraction of the events represented on the map, maintaining the characteristics of geological structures in accordance with the map's purpose and scale (Spencer, 2017). Similarly to other types of maps, generalization aims to highlight relevant features, which emphasize the main structures of the geological formations, and omit non-relevant content that makes the map difficult to read and interpret (Loudon, 2000; Downs & Mackaness, 2002). Specifically, for the case of geological maps, the following requirements apply:

- Geological maps represent polygonal subdivisions. Hence, generalization must take care not create overlapping polygons or holes.
- As always, readability serves as a driving force of generalization. Since geological maps typically contain many small island polygons, this applies particularly of those small shapes, or narrow polygons (Fig 1.1), where minimal readability dimensions must be ensured.
- Since geological maps are categorical maps, the generalization operators of Table 1.1. apply. Importantly, these also include attribute transformations, i.e., reclassification.

Any algorithms used with geological maps need to be able to deal with a wide variety of possible shapes (Fig. 1.1). Algorithms developed specifically for regular shapes, such as building footprints in topographic maps, are less well suited.

1.3. General Objective, Methodology, and Research Objectives

1.3.1. General Objective and Methodology

The ***key objective*** of this dissertation is to develop an integrated methodology for the generalization of geological maps. The methodology specifically deals with small island polygons, which represent one of the major challenges in the generalization of geological maps, as explained above. Small island polygons are free-standing polygons (also referred to as area patches in the literature – Muller &Wang, (1992), which are frequently occurring features in geological maps. Also, small island polygons are often associated with older geological units, which tend to have more mineral deposits and hence more economic value than rock types of a relatively younger age, which often form larger background polygons.

The ***methodology*** proposed in this thesis is termed 'integrated' as it consists of three logically sequential and complementary parts, which represent three different approaches, as shown in Figure 1.2: a constraint-based approach, a network-based approach, and an integrated approach.

In the first part, the proposed methodology identifies and separates the island polygons, a process referred to as separation of the so-called *foreground polygons* from the *background polygons*. Foreground polygons are small island polygons that need generalization due to their small areas. Background polygons are large and bulky polygons that form the background of the map. Foreground polygons are separated from the background polygons using Tukey's outlier detection approach (Seo et al., 2006). Figure 1.3 illustrates the separation of polygons into foreground and background polygons. The foreground polygons are later (in Part 2) further divided into polygons that form groups and 'other', isolated polygons.

Small island polygons (i.e., foreground polygons) are prone to legibility issues due to scale reduction. These polygons, however, form the main message of the map in this thesis, as they represent valuable geological information in the applications mainly targeted by this work, mineral exploration and mining. Size constraints, such as minimum area and object separation, are simple yet flexible in addressing problems linked to small polygons. Thus, in the remainder of Part 1, minimum area constraints trigger different generalization operators, *removing* relatively insignificant polygons, and *enlarging* important ones. Enlarged polygons grow closer to each other, possibly violating the constraint of object separation and even resulting in overlapping features. Such issues trigger *aggregation* or *displacement* operators to resolve conflicts.

While the constraint-based approach of Part 1 can resolve cartographic conflicts of island polygons due to violation of simple size constraints, it does not have the capacity of considering the wider spatial context of these small polygons. Enabling such contextual generalization requires

identifying structures and patterns in a process called 'structure recognition' (Brassel & Weibel, 1988) or 'cartometric evaluation' (McMaster & Shea, 1992). This process is an essential prerequisite to the application of map contextual generalization operators. All modern computational processes of map generalization include structure recognition and evaluation of the source map to trigger and inform the appropriate generalization actions (Harrie & Weibel, 2007). Patterns in geological maps are, among others, created by groups of small polygons, which are essential for understanding geological formations.

The second part of the thesis is thus dedicated to recognizing patterns and identifying groups in geological maps. More specifically, patterns are recognized and groups of polygons identified from group-forming polygons (Figure 1.3). Pattern recognition is achieved by building and successively refining a network between the centroids of island polygons and grouping them based on proximity and similarity in properties such as area, orientation, shape, and category.

Part 3 of the thesis builds on the second part and is dedicated to the generalization of identified homogenous groups maintaining specific properties of a group of polygons, such as area, shape, and orientation of the group member polygons. The process of generalization is implemented by two generalization operators, typification and aggregation, for which specific algorithms are proposed. The choice between these two operators is driven by their known characteristics of altering the three key properties of the groups: area, shape, and orientation. Statistical analysis of these properties before and after the generalization process is used to evaluate the generalization results and optimize the parameterization of the operators and the operator selection process.

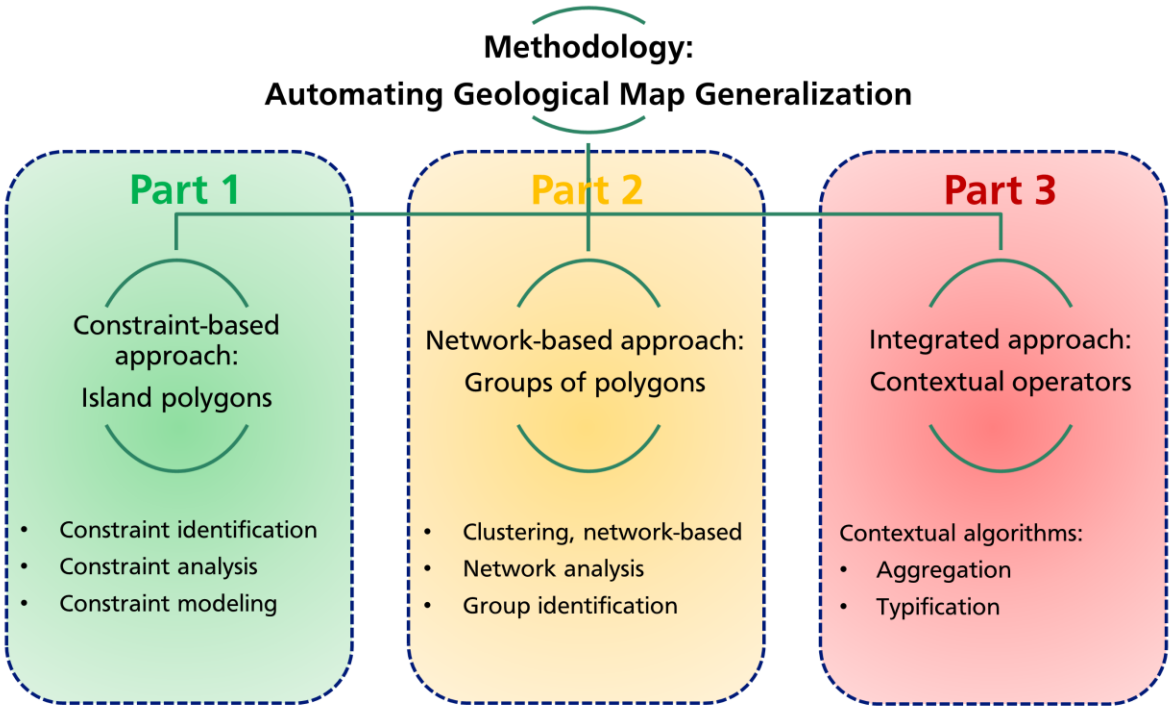


Figure 1.2. The three parts of the methodology to enable the automation of geological map generalization.

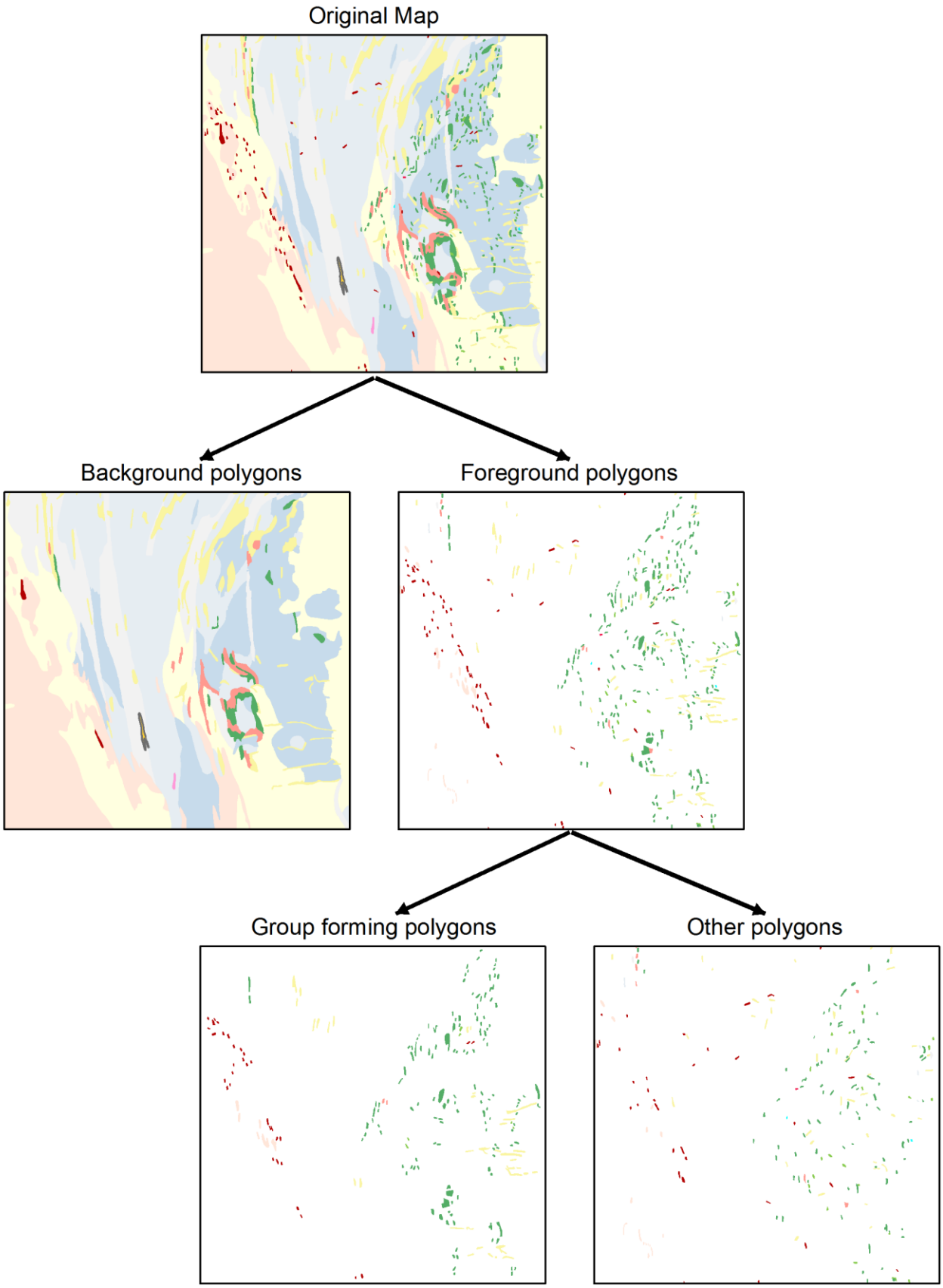


Figure 1.3. The methodology first divides polygons in the geological map into background and foreground polygons (in Part 1) and then further subdivides the foreground polygons into group-forming polygons and other, isolated polygons (Part 2).

1.3.2. Research Objectives

As stated in the preceding section, the key objective of this thesis is to develop an integrated methodology for the generalization of geological maps, focusing particularly on the treatment of small island polygons. More specifically, the three parts of the methodology of this thesis each give rise to one of the following ***specific research objectives***:

Research objective of Part 1

- Separate and generalize island polygons in geological maps by incorporating size constraints.

Research objective of Part 2

- Identify groups of small polygons based on similarity criteria, forming important spatial structures, as a basis for implementing contextual generalization operators.

Research objective of Part 3

- Develop contextual generalization algorithms and evaluation of generalization results of polygon groups in geological maps.

1.3.3. Structure of the Thesis

This thesis consists of five chapters, which can be subdivided into three main parts. The first part consists of Chapter 1 and introduces the topic, addresses the motivation and the theoretical background, describes the research objectives and approach, and reviews the state of the art of the relevant literature. The second part consists of Chapters 2 to 4, which contain the main substance and contributions of the thesis. Each of these three chapters addresses one part of the tripartite methodology summarized in Section 1.3.1, and is equivalent to one research paper (each addressing one of the above specific research objectives). The third part consists of Chapter 5 and discusses the research papers, their main findings, and provides an outlook on future research.

Of the three research papers included in this thesis, two papers have been published in international peer-reviewed journals, and one has been submitted, as shown below.

Research Paper 1 (Chapter 2)	Sayidov, A., Aliakbarian, M., & Weibel, R. (2020). Geological Map Generalization Driven by Size Constraints. ISPRS International Journal of Geo-Information. 9(4):284. https://doi.org/10.3390/ijgi9040284.

Research Paper 2 ***(Chapter 3)***	Sayidov, A., Weibel, R., & Leyk S. (2020). Recognition of group patterns in geological maps by building similarity networks. Geocarto International. https://doi.org/10.1080/10106049.2020.1730449.

Research Paper 3 ***(Chapter 4)***	Sayidov, A., Leyk, S., & Weibel R. (submitted). Integrating Aggregation and Typification to Generalize Polygon Group Patterns in Geological Maps. Cartography and Geographic Information Science.

Figure 1.4 seeks to position the three research papers with respect to the workflow of the integrated methodology for automated generalization of geological maps that forms the core of this thesis. Paper 1 proposes an approach that separates foreground and background polygons, but otherwise relies solely on size constraints. Conversely, the other papers rely on meaningful polygon group structures, where Paper 2 proposes a method to detect such polygon groups, while Paper 3 proposes new algorithms for aggregation and typification that utilize these group structures.

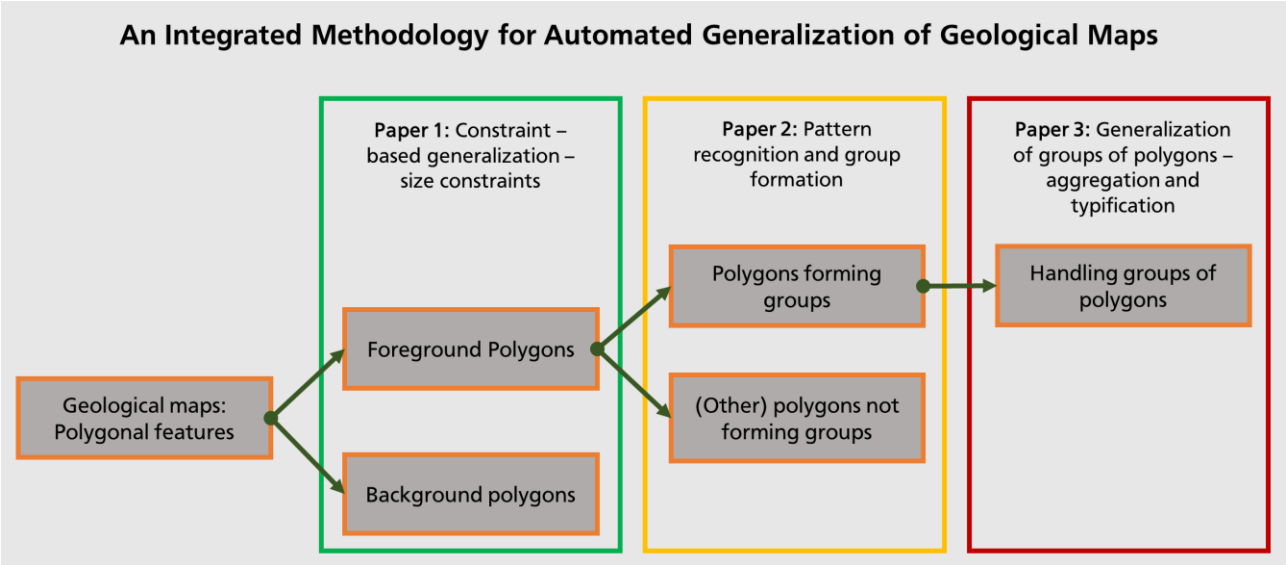


Figure 1.4. Positioning the research papers of this thesis in the workflow of the proposed integrated methodology.

1.4. State of the Art in Geological Map Generalization

This section reviews the literature on map generalization, focusing on constraint-based approaches, pattern recognition, and geological map generalization.

1.4.1. Constraint-Based Map Generalization

A review of the literature reveals that the concept of constraints is very broadly used by diverse disciplines, such as computer science, mathematics, classical mechanics, and also map generalization. Below, we focus on how the concept of constraints is used in map generalization, with a focus on the objectives of our research.

The notion of generalization constraints (Beard, 1991) and constraint-based generalization originated from discussions about designing and developing algorithms that can meet the intended purpose, which requires a careful and complete definition of requirements and criteria that govern the generalization problem (Weibel & Dutton, 1998).

The idea of constraints was taken from computer science and implemented in the automation of map generalization by several researchers, such as Beard (1991), Weibel (1996), Ruas (1999), Barrault et al. (2004), Galanda (2003), Ware et al. (2003), Burghardt and Neun (2006), Burghardt et al. (2007), Mackaness and Ruas (2007), Schmid (2008), Stoter et al. (2010), Stern and Sester (2012), Zhang, et al. (2013), Touya et al. (2019). Constraints in the context of generalization can be defined as a “design specification to which the solutions to a generalization problem should adhere” (Weibel & Dutton, 1998). A generalized map should fulfill several conditions, and these conditions control the generalization process by employing constraints. In constraint-based modeling, the generalization process is ruled by a synthesis of conditions (Ruas & Plazanet, 1996).

Weibel and Dutton (1998) proposed constraints for thematic maps consisting of polygonal subdivisions (also called categorical maps), such as geological or soil maps. Harrie (2003) put forward a typology for graphical (size) constraints, and Ruas (1999) and Ruas and Duchêne (2007) use constraints to automatically control and evaluate the generalization process based on agent-based technologies. Galanda (2003) used the same agent-based approach, which defining constraints for polygonal map generalization. Petzold et al. (2006) focus on constraint-based map generalization, controlled interactively with a workflow system.

The literature suggests that the constraint-based approach has been successfully utilized in map generalization. However, further improvement is needed to increase the efficiency and efficacy of the constraint-based approach. Harrie and Weibel (2007) highlight the limitations of the current usage of the constraint-based approach:

"*The main limitation of constraint-based techniques is the limitations of the constraints themselves. Important cartographic constraints still remain to be defined, others are poorly defined. More work is required in formulating constraints and associated measures for groups of objects. Work on pattern recognition and on constraints for preserving cartographic patterns will be particularly important…*"

While this statement dates already more than a decade back, its main message still largely drives the current ongoing research on constraint-based modeling, necessitating the identification, formalization, and parametrization of constraints, and exploring the influence of constraints on pattern recognition.

1.4.2. Pattern Recognition in Map Generalization

In automated map generalization, pattern recognition has gained increasing attention (Steiniger, 2007; Duda, et al., 2012). Accordingly, in the following paragraphs, we will take a closer look at the literature on pattern recognition for vector data in topographic and thematic maps.

Pattern recognition is a knowledge enrichment and decision-making process based on the relationships inherent in a set of spatial objects (Heinzle & Anders, 2007). It has been implemented in various domains such as engineering, astronomy, medical science, remote sensing, and many others (Fu, 2019). Pattern recognition aims to precisely identify a set of objects and distinguish them from their surroundings based on unique properties such as color, size, shape, orientation, or texture. A significant amount of research has been done on pattern recognition in remote sensing (Chen & Ho, 2008; Aksoy et al., 2010; Brilakis et al., 2011; Albus et al., 2012; Younan et al., 2012; Brereton, 2015; Chen, 2015; Lambers & Traviglia, 2016; Li & Du, 2016; Anwer et al., 2018; Fu, 2019), mainly dedicated to image analysis for automatic feature extraction and land classification (Ghamisi et al., 2014; Tuia et al., 2014; Crommelinck et al., 2016; Elmahdy & Mohamed, 2016; Lv et al., 2017).

Mackaness and Edwards (2002) define a pattern as having qualities such as shape, orientation, color, connectivity, density, distribution, and regular repetitions, which the human eye immediately interprets and associates as similar. Features on topographical maps, such as roads and buildings, have formed the main focus of pattern recognition for map generalization. Most approaches are graph-based and rely on a network of nodes and edges connecting map objects, such as roads (Heinzle & Anders, 2007; Heinzle et al., 2005; Marshall, 2005; Tian et al., 2012). These approaches recognize significant road structures, such as grid-shaped road networks and star-like or ring-shaped configurations, by obtaining so-called strokes (i.e., paths in the network with minimal deflection angle) from the network and analyzing them. Cui et al. (2019) and Yang et al. (2019) used a

density-based approach to extract patterns from road networks. Graph-based approaches have also been used for grouping buildings in urban blocks (Zhang et al., 2010; Deng et al., 2011; Zhang et al., 2013; Cetinkaya et al., 2015; Wang & Burghardt, 2017; He et al., 2018). Regnauld (2001) used a so-called divide and conquer principle to segment an initial set of buildings into homogenous groups. Characteristics of each group are used to build a new representation of the group depending on the target scale. A Minimum Spanning Tree (MST) technique is used to generate groups from centroids of buildings. The edges are then removed based on proximity, homogeneity, and the number of buildings in a group. Li et al. (2004) create a Delaunay triangulation (DT) for buildings. . Buildings whose vertices belong to the same triangle are labeled as 'neighbors.' Direct alignments between neighboring buildings are found by employing Gestalt theory (Wertheimer 1938; Steiniger & Weibel 2007). Zhang et al. (2013) produce an MST from a constrained DT of building centroids to detect building clusters, similar to Regnauld (2001), followed by an analysis of building alignments.

Similarly, Anders et al. (1999) applied graph-based clustering techniques for the automated analysis of settlement structures represented by building centroids. Anders (2003) demonstrated the successful application of different proximity graphs to find 'natural' object groups without defining any parameter, such as the number of points and distance between points. Deng et al. (2011) proposed an adaptive spatial clustering approach by removing so-called global and local long edges from the DT, which discovers patterns in the network based on various global and local criteria. This approach builds on a proximity analysis of neighboring network points, whereas attribute properties, such as size, orientation, shape, and category of the map feature, are not considered (Deng et al., 2011). Anderson-Tarver et al. (2011) proposed and implemented an approach that applies fuzzy logic to identify small polygons representing ponds in swamps and to maintain proportionate textures at the target scale.

Proximity is the primary criterion used for grouping objects in maps (Regnauld, 2001; Deng et al., 2011). The objects in a map are linked with their nearest neighbor, forming a network (or graph, to use the mathematical expression). Linked polygons can be analyzed according to their similarities and a group of homogenous objects can then be generated. This approach is called a proximity-based grouping. Proximity-based grouping methods have been widely used to recognize features in topographic maps, particularly to recognize patterns and groups of buildings, where the shapes of polygons are rather regular and the orientations often co-aligned. Geological maps are full of proximally related polygons with more different and arbitrary sizes, shapes, and orientations belonging to the same or different rock units. Preservation of the relationships between polygons is possible by recognizing patterns and identifying groups of these polygons. However, pattern

recognition and group identification using proximity-based network analysis have not been tested for generalizing geological maps, despite the potential these approaches may have. Thus, we see a need to develop a methodology for recognizing group patterns tailored to the peculiarities of a geological map. For instance, shapes of polygons in a geological map are typically less compact than building shapes and exhibit a greater degree of shape variation, which poses challenges to the formation of graphs, the analysis of the proximity of neighboring polygons as well as the similarity assessment of different polygons.

1.4.3. Geological Map Generalization

As mentioned at the outset of this thesis, the research on automated map generalization has mainly focused on topographical maps and much less on categorical maps. The existing methods for categorical map generalization can be grouped into raster and vector-based approaches. Below, the most important studies on categorical map generalization are briefly reviewed.

Raster and Vector Formats

To represent any spatial data in GIS, two main different data structures are used: vector and raster representations. Raster data represents the world divided up into cells, which have an attribute value associated with them. A digital photograph is a simple example of raster data, where each cell corresponds to a particular color value. Raster layers can also represent non-visual attribute values, such as temperature or elevation. In our case, however, it means a category belonging to a geological unit in the geological map, depicted by a color value on the map. In a GIS classified raster data must be accompanied by information detailing where the raster cells are located and how much area they cover, that is, by georeferencing information.

Vector data consist of points, lines, and polygons, where lines and polygons are made up of individual points, called vertices. Each vertex consists of a pair of coordinates, linking the position on the map to the real-world location on the Earth. The vertices are joined in sequential order to form lines or polygons. Vector data is convenient in representing discrete boundaries, such as the footprints of buildings, boundaries, and transportation links.

Raster-Based Generalization in Categorical Maps

The most commonly used raster-based generalization approach uses operators such as region growing/shrinking filters (Schylberg, 1993; Su et al., 1997, 1998; Cámara, 2000; Raposo & Samsonov, 2014; Raposo et al., 2016). Many of these approaches use operators from so-called mathematical morphology (Najman & Talbot, 2013). Schylberg (1993) uses a simple grow-and-shrink algorithm to aggregate polygons. The approach is relatively fast. However, it is inadequate

when spatial variability of categories is high, as it cannot adapt to local variation (Peter & Weibel, 1999).

A relatively recent raster-based approach designed specifically for geological map generalization is based on cellular automata (Smirnoff et al., 2008, 2012). The method achieves rather promising results for maps containing mainly large patches of geological units, but performs less convincingly in areas with more varied patch structures, as it does not address individual and local properties, such as size, shape and the orientation of geological features, which are critical for maintaining the essential structures and patterns of the map during generalization.

In general, raster-based generalization approaches are fast and efficient. However, most geological maps are compiled and stored in vector format. Converting them back and forth from one medium to another will always lead to a loss of positional accuracy.

Vector-Based Generalization in Categorical Maps

Representation of categorical data by vector data leads to so-called polygonal subdivisions or polygonal maps (Galanda, 2003). Research on polygonal map generalization can be roughly divided into two types of approaches: constraint-based generalization approaches (Peter & Weibel, 1999; Edwardes & Mackaness, 2000; Peter, 2001; Galanda, 2003; Sayidov et al., 2020a; Sayidov et al., 2020b) and database generalization (for multiple representations (Brodaric & Patera, 1997, 1998, 2001; Yaolin et al., 2001, 2002; Downs & Mackaness, 2002; Patera, 2006; McCabe, 2008). Below, we briefly review these two approaches.

The ***constraint-based generalization*** approach was developed in the 1990s (Beard, 1991; Ruas & Plazanet, 1996; Weibel, 1996; Weibel & Dutton, 1998) and implemented by a number of researchers (Ruas, 1999; Barrault et al., 2001; Ruas & Duchêne, 2007) to tackle topographic map generalization, partly also covering asepcts of categorical map generalization (e.g., Weibel & Dutton, 1998). On the conceptual level (for defining, designing, and testing constraints), constraints for polygon generalization are discussed by Peter and Weibel (1999) and Edwardes and Mackaness (2000). Galanda (2003) developed a series of constraints for categorical maps, which is employed in an agent-based approach to automate the generalization of categorical maps.

Another approach applied to the generalization of categorical maps is referred to as ***database generalization***. The concept of database generalization is based on creating a single detailed database that would allow data representation at multiple map scales and for multiple purposes (Buttenfield, 1995; Sarjakoski, 2007; He et al., 2018). Patera (2006) proposes an approach that transfers large-scale geological maps in a database to a smaller scale. The approach extracts information from the database implementing a rule-based system that utilizes SQL queries. Yaolin

et al. (2002) implemented a similar approach and proposed a spatial data structure and semantic evaluation model for database generalization, which combined classification and aggregation in a hierarchical way. Downs and Mackaness (2002) combined several generalization operators into a rule-based approach to obtain a generalized geological map. The procedure is relatively simple and flexible, yet the method incorporates potentially subjective decisions as it also allows human interaction (Downs & Mackaness, 2002).

Research on both constraint-based and database generalization yields rather promising results. However, there are some drawbacks to the approaches presented above as they do not treat categorical maps as forming patches and patterns linked together in space and time. Ignoring these properties in map generalization may miss important properties, shapes and links between patches on the map, potentially altering the essential structures and patterns implicitly contained in the map. Database generalization in particular is limited by the fact that by and large, it is restricted to the use of reclassification and aggregation operators, severely constraining the options available for adapting to local variation of map patterns.

1.4.4. Research Challenges Addressed in this Thesis

The above review of the literature has revealed several challenges in the automation of categorical map generalization. This thesis thus seeks to address the following four challenges:

1) Automated generalization of geological and categorical maps — There is an unequal distribution of research dedicated to categorical map generalization (with only rather few notable studies) and topographic map generalization (the majority of the studies). Despite differences in the character of the map objects being dealt with in topographic vs. categorical maps, some of the topographic map generalization methods can potentially be implemented for categorical map generalization. For instance, building footprints may have regular shapes, but they also often form groups whose general structure should be maintained in map generalization, similar to those of groups of small island polygons in categorical maps. However, categorical mapping approaches need to be more flexible such that the rules used for generalizing topographic maps can be adapted to different semantic content. We seek to establish a list of constraints that will handle island polygons, recognize patterns, and identify groups of similar polygons during generalization to overcome some of these challenges.

2. *Introduce constraint-based methods to geological map generalization* — The previous section highlighted several approaches dedicated to constraint-based map generalization, devoted mostly to topographic mapping. As requirements for categorical maps are different from those used in topographic maps, merely reusing approaches and processes will not provide adequate solutions

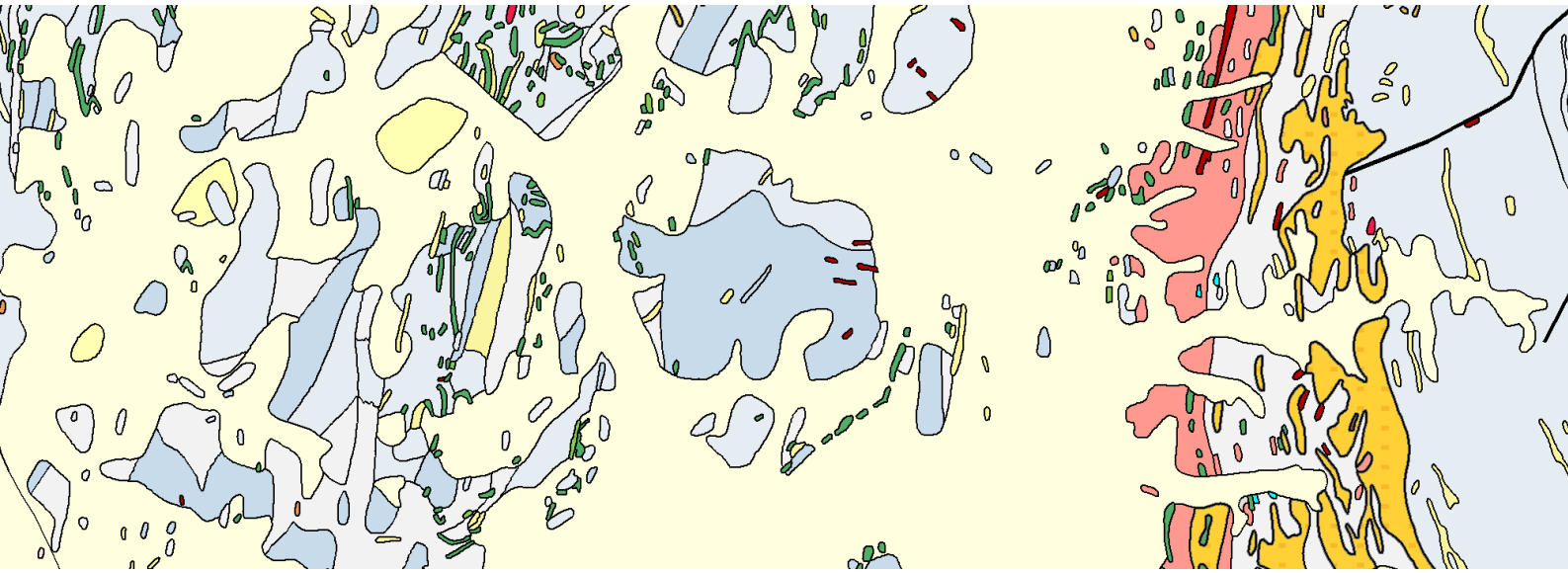
(Sayidov & Weibel, 2016, 2017). Instead, we seek to establish a list of so-called size constraints (i.e., constraints related to visual perceptibility limits), their measures, and goal values. The list should facilitate a better quantitative basis for the generalization process.

Furthermore, we seek to prioritize the logical treatment of constraints. The prioritization dictates the sequence of generalization operators that handles constraint violations. Research Paper 1 will address geological map generalization driven by size constraints.

3. *Pattern recognition in support of geological map generalization* — Several studies are dedicated to the identification of perceptual patterns in topographical maps (Regnauld 2001; Boffet & Rocca-Serra 2001; Anders 2003; Zhang et al. 2010; Deng et al. 2011; Zhang et al. 2013; Cetinkaya et al. 2015; Wang & Burghardt 2017; He et al. 2018), with a large part focusing on the detection of buildings based on spatial proximity. Thus, the first challenge is to establish whether spatial proximity principles can also recognize patterns in geological maps, as geological maps consist of irregular shapes and structures. The second challenge is to use the properties of each polygon to form groups from patterns recognized in the previous step. The second Research Paper 2 focuses on recognizing patterns and forming groups from them, utilizing similarities of polygons contained in geological maps.

4. *Adaptive operator selection to support the generalization of polygon groups* — Previous research on the generalization of groups of polygons has focused primarily on proximity (e.g., Regnauld, 2001; Deng et al., 2011) and has paid little attention to issues of preservation of orientation, shape, and size but has mostly focused on reducing the number of polygons in a group. As a result, no comprehensive approach exists that allows to generalize groups of polygons while maintaining their properties. The challenge is how to decide which generalization operators to select while preserving size, shape, and orientation of group members. In the third research paper, we present an approach that maintains the inherent properties of a polygon group during generalization using an adaptive selection of generalization operators.

In this thesis, we addressed four research challenges pertaining to particular apsects of geological map generalization. A more comprehensive list of research issues on automation of map generalization can be found in Weibel and Dutton (1999), Harrie and Weibel (2007), Stoter (2005), Stoter et al. (2010), Stoter et al. (2014), Šuba, (2018), and Sester (2020).



2

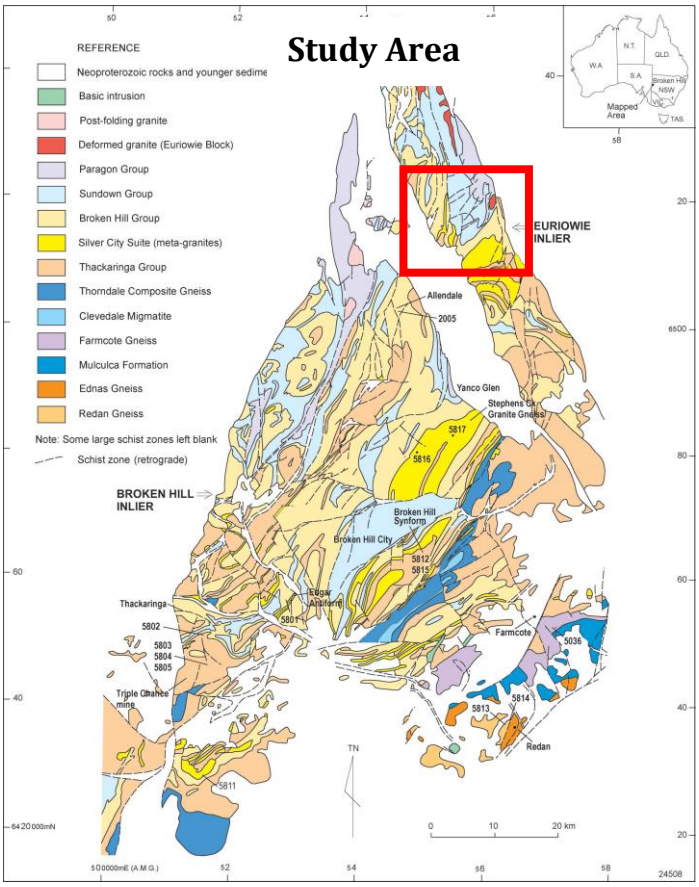
Chapter

Creativity loves constraints.

Eric Schmidt

Research Paper 1

Sayidov, A., Aliakbarian, M., & Weibel, R. (2020). Geological Map Generalization Driven by Size Constraints. ISPRS International Journal of Geo-Information. 9(4):284. https://doi.org/10.3390/ijgi9040284.



From Burton (2001). Red frame indicates the Euriowie Inlier representing the study area of this thesis.

Contribution of the Ph.D. candidate: The Ph.D. candidate developed the conception and design of the work; developed the conceptual model for the constraint-driven generalization of geological maps; scripted codes for algorithms in Python to perform analysis and processing of the geological map; drafted the article.

Authors' Contributions: Conceptualization, Data analysis, Visualization and Validation, Azimjon Sayidov; Methodology, Azimjon Sayidov, Meysam Aliakbarian and Robert Weibel; Software, Azimjon Sayidov and Meysam Aliakbarian; Writing original draft, Azimjon Sayidov; Writing review and editing, Azimjon Sayidov and Robert Weibel.

International Journal of
Geo-Information

MDPI

Article

Geological Map Generalization Driven by Size Constraints

Azimjon Sayidov *, Meysam Aliakbarian and Robert Weibel

Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; meysam.aliakbarian@geo.uzh.ch (M.A.); robert.weibel@geo.uzh.ch (R.W.)
* Correspondence: azimjon.sayidov@geo.uzh.ch; Tel.: +41-44-635-51-98

Received: 14 February 2020; Accepted: 22 April 2020; Published: 24 April 2020

Abstract: Geological maps are an important information source used in the support of activities relating to mining, earth resources, hazards, and environmental studies. Owing to the complexity of this particular map type, the process of geological map generalization has not been comprehensively addressed, and thus a complete automated system for geological map generalization is not yet available. In particular, while in other areas of map generalization constraint-based techniques have become the prevailing approach in the past two decades, generalization methods for geological maps have rarely adopted this approach. This paper seeks to fill this gap by presenting a methodology for the automation of geological map generalization that builds on size constraints (i.e., constraints that deal with the minimum area and distance relations in individual or pairs of map features). The methodology starts by modeling relevant size constraints and then uses a workflow consisting of generalization operators that respond to violations of size constraints (elimination/selection, enlargement, aggregation, and displacement) as well as algorithms to implement these operators. We show that the automation of geological map generalization is possible using constraint-based modeling, leading to improved process control compared to current approaches. However, we also show the limitations of an approach that is solely based on size constraints and identify extensions for a more complete workflow.

Keywords: geological maps; map generalization; constraint-based methods; size constraints; generalization operators

1. Introduction

The majority of the research conducted on the automation of map generalization has thus far been dedicated to topographic maps, with a focus on the generalization of roads, boundaries, river networks, and buildings; that is, with a focus on linear objects or small area objects. The generalization of thematic map types, including categorical maps (e.g., geological, soil, or land use maps), has received less attention, perhaps since categorical maps contain polygons of potentially arbitrary shapes and sizes, rendering them more complex than typical shapes found, for instance, for buildings on a topographic map. Nevertheless, categorical maps, with geological maps as a prominent representative, are a frequent map type and require specific methods with which to automate generalization. For instance, in topographic maps, buildings are usually of a regular shape and are often arranged in a regular fashion (e.g., in linear alignments), while in categorical maps, polygon features can be of an arbitrary shape, occurring in arbitrary spatial arrangements. Merely reusing the approach and processes of topographic map generalization for categorical mapping will not provide a proper solution [1,2], as requirements for categorical and thus geological map generalization are different from those used for topographic mapping, as detailed in Section 2.

Over the past two decades, constraint-based techniques have evolved as the leading paradigm used to model and automate the map generalization process [3–6] and to evaluate generalization results [7]. However, while use of the constraint-based paradigm is wide-spread in topographic mapping, generalization methods for categorical and geological maps have rarely adopted this approach (see [8] for one exception). This is somewhat disappointing, as the complexity of categorical maps could favor an approach to allow differentiated and adaptive modeling and monitoring of the conditions that govern the map generalization process.

The objective of this paper is to present a methodology using constraints, more specifically so-called size constraints, as a basis for the geological map generalization. Size constraints deal with the minimum area and distance relations in individual or pairs of map features. They have also been termed metric or graphical constraints [4] and express the natural limits of human perception in measurable, minimal dimensions [8]. The proposed approach first identifies a list of size constraints, their goal values, and measures; it also prioritizes the logical treatment of constraints, which in turn dictates the sequence of generalization operators and algorithms to be used in case a constraint is violated. The main driver for the proposed methodology is the "minimum area" (MA) constraint, which influences other constraints and is coupled with descriptive attributes of the polygon features that are being generalized. In an adaptive workflow, the MA and related constraints are successively tested and, if violated, trigger appropriate generalization operators, including object elimination (i.e., selection), enlargement, aggregation, and displacement.

The use of the constraint-based approach in map generalization has predominantly been linked to the agent-based paradigm [3,5,6,8,9]. Since building an agent engine, however, is by no means a trivial task, the uptake of the agent-based approach, and with it the constraint-based approach, in practice has been limited. Hence, the main contribution of this research is the demonstration of the usage of constraints, with a focus on size constraints, for the automated generalization of geological maps, as this approach has so far not been explicitly applied to the automated generalization of geological maps. Furthermore, we use a workflow-based approach consisting of a sequence of several generalization operators, as this has a better potential to be adopted in practice. Although designed with geological maps in mind, the approach is also applicable to other categorical maps: maps that are entirely covered with polygonal features (i.e., so-called polygonal subdivisions) such as soil maps, vegetation maps, or land-use and land-cover maps. In experiments, we show that the proposed methodology, despite its relative simplicity, is capable of more appropriately generalizing geological maps, with better local control over the generalization operations that are applied as compared to results generated by state-of-the-art solutions that do not use a constraint-based approach. Finally, we are aware that with its focus on size constraints only, our methodology has limitations, which we point out in Section 6, providing leads for further research.

In Section 2, the characteristics of geological maps are discussed. Section 3 provides a review of related work on geological map generalization. Section 4 represents the core of the article, introducing the elements of the size constraint-based methodology one by one. Section 5 presents the results of a series of experiments, firstly to illustrate the effect of the individual generalization operators, and then for different parameterizations of the combined workflow. Finally, Section 6 provides a discussion of the experimental results, while Section 7 ends the paper with conclusions and an outlook.

2. Geological Maps: Purpose and Peculiarities

Geological maps are among the most complex thematic maps, with various elaborate shapes and structures. This renders the generalization process more demanding, and thus an in-depth analysis of these structures is required before the generalization process [2,10]. Creating geological maps requires not only cartographic expertise but also general geological knowledge. Thus, for instance, knowledge on the formation and structure of rock types occurring in a study area will be crucial when it comes to identifying the relative significance of map features. However, the importance of certain map features may vary depending on the purpose and type of geological map that is to be made.

Bedrocks assist geologists in portraying the natural history of a study area and identifying associated rock formations. They also carry essential mineral resources such as coal, oil, and uranium, which are in the focus of the mining industry. Thus, a map for mining purposes highlights ancient bedrocks that may carry particular minerals, while neglecting sedimentary rocks as they are more recent [11]. Geophysicists, in turn, place more emphasis on the intrinsic characteristics of features such as porosity and permeability in rock and sediments [12].

The goal of a geological map is "to interpretively portray the spatial and temporal interactions between rocks, earth materials, and landforms at the Earth's surface" [13]. Geological materials are the igneous, metamorphic, and sedimentary rock and surficial sediments that form the landscape around us. Most geological maps use colors and labels to show areas of different geological materials, called geological units. Geological structures are the breaks and beds in the geological structures resulting from the slow but strong forces that form the world [14]. "Geological maps show the location of these structures with different types of lines. Because the Earth is complex, no two maps show the same materials and structures, and so the meaning of the colors, labels, and lines is explained on each map" [15].

Geological maps consist of diverse patterns formed by a fabric of polygons, plus additional linear objects (e.g., fault lines) and point objects (e.g., wells) which however are not of concern here. The polygons can be described by spatial, structural, and semantic properties to evaluate similarities or differences and thus infer spatial patterns relating to the perceptual structure or arrangement of polygon features on the map [16]. Figure 1 shows some sample excerpts of geological maps, ordered by increasing geometric and graphical complexity. In the simple map extract of Figure 1a, few geological units are involved, with relatively simple shapes, which could be generalized by merely using simplification operators. The next level of complexity comprises many small polygons of the same or similar geological units (Figure 1b), which may be generalized by removing or alternatively aggregating units or sub-units into a single group. Another level consists of a series of elongated polygons of the same unit embedded in, and possibly crossing, other units (Figure 1c), where a cartographer may recommend merging neighboring units while trying to maintain their overall arrangement (i.e., using the so-called typification operator). Another complicated form found in geological maps are tree-like, dendritic forms, which were created at a later stage of the quaternary period by rivers and streams carrying sediments and other minerals. This type also defines the position of a river system (Figure 1d). In this case, a possible solution is to replace several branches with a smaller number of simplified tree branches. Figure 1e shows that various kinds of units consisting of small and large/long and narrow polygons and tree-like structures may also co-exist, rendering the generalization process even more challenging. The generalization of such complex fabrics requires making multiple, interrelated (and possibly conflicting) generalization decisions.

The few examples in Figure 1 illustrate that the polygonal layer of geological maps consists of a far greater variability in category, size, shape, boundary sinuosity, and topological (in particular containment) relations of the concerned polygons than can typically be found on topographic maps. The spatial arrangement of polygons in geological maps can take many different forms, as is noticeable in Figure 1, although even in the complex map of Figure 1e we can perceive alignments and clusters of polygons. The key feature classes in topographic maps are predominantly anthropogenic and hence tend to have more regular shapes and a lesser degree of variability, and they are often arranged in regular alignments (e.g., grid street networks or straight alignments of buildings). "Natural" feature classes, such as land cover, in topographic maps usually are restricted to few categories (e.g., woodlands, waterbodies, built-up areas) and are of secondary priority. Hence, we would argue that while the same operators (elimination, simplification, aggregation, typification displacement, etc.) are valid in both domains, different generalization algorithms, or at least different combinations of algorithms, have to be used to adapt to the peculiarities of geological maps or more generally categorical maps.

The next section offers a review of various generalization methods that have been explicitly proposed to deal with the peculiarities of the geological maps listed above.

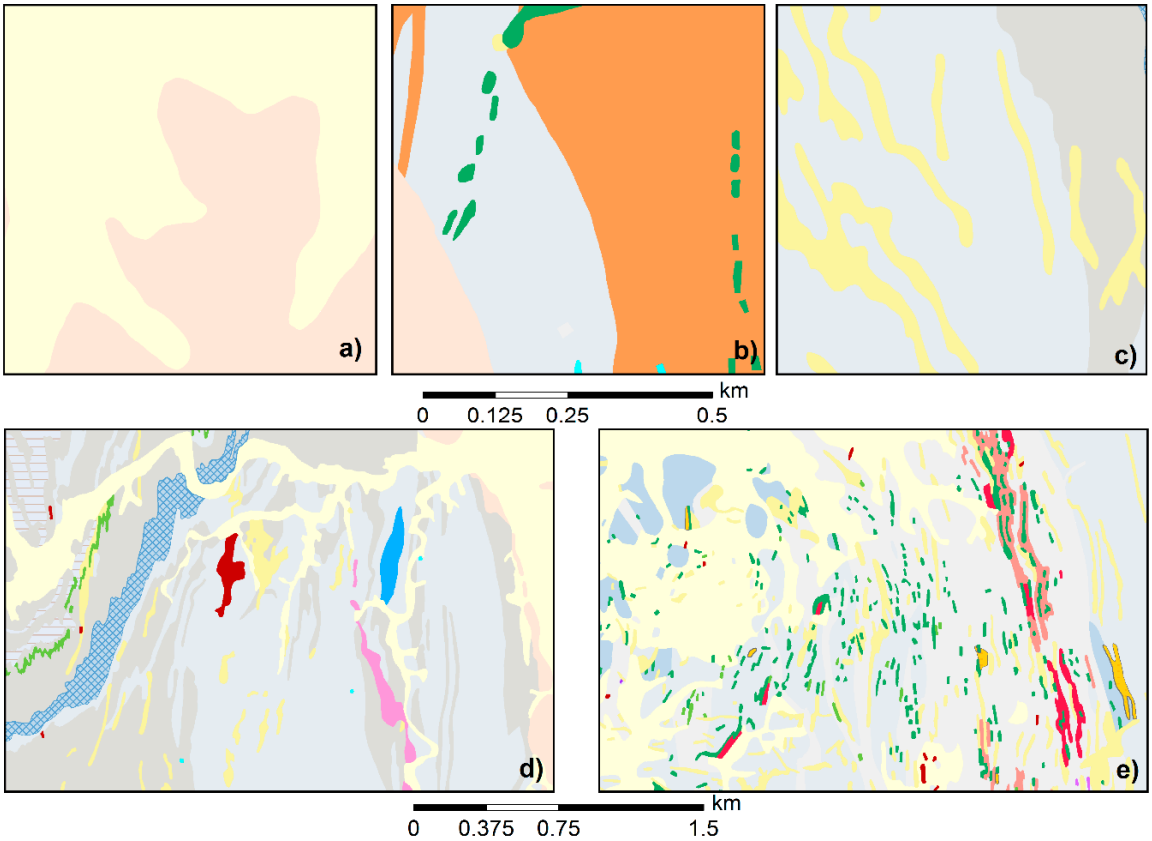


Figure 1. Some excerpts from geological maps, ordered by the complexity of the shapes and structures of geological units. (**a**) A few large polygons with simple shapes. (**b**) Several small polygons of the same geological unit and with similar geometrical properties. (**c**) Elongated polygons of the same unit with similar geometrical properties. (**d**) A complex type with tree-like dendritic shapes. (**e**) A combination of all of the above-mentioned shapes and structures.

3. Related Work

The earliest significant attempt at specifically generalizing geological maps was made by the authors of [17], who tried to automate the generalization of a 1:250,000 scale geological map from a 1:50,000 scale source map (a product of the British Geological Survey (BGS)), using the conceptual model previously suggested in [18]. While the results obtained were encouraging, the BGS concluded that the strategy still required intervention by the human operator, rendering it less flexible and more subjective. Importantly, however, this early work highlighted the significance of basing the generalization process upon an understanding of the essential structures and patterns inherent to the source map, a task that is of course not unique to the case of geological maps [18,19].

The authors of [20] presented a conceptual workflow model dedicated to the semi-automated generalization of thematic maps with three main phases: structural analysis, generalization, and visualization. Structural analysis (or "structure recognition" according to [18]) was deemed especially crucial, as once all relevant structures present in a map are made known, this information can support the decision of "when and how to generalize" [21]. The second step of their conceptual model consisted of constraint modeling in a multi-agent system, aiming to build an objective and flexible workflow.

Inspired by the aforementioned conceptual model [20], the author of [22] developed a generalization workflow based on ArcGIS tools. A sample geological map at a 1:24,000 scale was generalized to three target scales, namely 1:50,000, 1:100,000, and 1:250,000. The study results were compared with the corresponding U.S. Geological Survey (USGS) geological maps and helped

to summarize the strengths and limitations of the tools available for generalization at the time within ArcGIS.

Another experiment on the effectiveness of ArcGIS was conducted by Smirnoff et al. [23], where a cellular automata (CA) approach was developed specifically for geological map generalization using ArcGIS tools as a basis. When comparing their results with those of a process using the generalization tools directly available in ArcGIS, they concluded that the cell-based, or cellular, automata model had essential advantages for the automated generalization of geological maps. In more recent research [24], the ArcGIS toolbox called "GeoScaler" was tested on surficial and bedrock maps. The results were evaluated and found adequate. Also, repeatable results were obtained by maintaining some amount of human intervention in the process. Very importantly, since the GeoScaler toolbox was made freely available, this methodology can be used in the practice of geological mapping and has thus defined the state of the art of generalization tools for geological map generalization, which persists today. However, the approach does not consider the individual, local properties of geological features such as the size, shape, and orientation of the polygons, or the distance between them, which is crucial for carrying over the critical patterns of the source map to the derived map. Moreover, as most geological map data exist in vector format, conversion from vector to raster format (and possibly back to vector format again) causes additional, uncontrolled loss of data accuracy.

Hence, a vector-based approach that uses a combination of generalization operators that can be adaptively applied depending on the local situation seems more appropriate. In this vein, the authors of [25] proposed algorithms for several operators of polygon generalization, including elimination, enlargement, aggregation, and displacement, based on a rolling ball principle, conforming Delaunay triangulation, and skeleton approaches. The authors of [8,26] proposed a list of constraints and an agent-based process for the generalization of polygonal maps, extending earlier works by the authors of [4,5,9].

Probably the most comprehensive work regarding constraint modeling for geological maps was presented by Edwardes and Mackaness [27], who developed and illustrated a rich set of constraints and proposed ways to represent these in formal language. While conceptually intriguing, the proposed set of constraints was unfortunately not linked to particular generalization algorithms and was not implemented. The study also demonstrated the vast complexity that is involved when trying to model the constraints governing geological maps comprehensively and cast these into a computer-based process.

Müller and Wang [28] proposed an automated methodology for area patch generalization that encompassed elimination, enlargement/contraction, merging, and displacement operators. They addressed a problem similar to that dealt with in this paper, and they also used a similar set of operators. However, their approach considered all polygons as semantically identical. In contrast, geological maps easily consist of over 20 rock types, demanding simultaneous consideration of geometrical as well as attribute properties of the polygons. Their approach also leaves rather little flexibility for modifying the control parameters, as it lacks the capability of testing different solutions for a given conflict.

As argued in Section 1, the best way to detect cartographic conflicts and formalize and control generalization algorithms for resolving such situations is by using constraints, as has been shown in topographic map generalization. In this paper, we seek to demonstrate the application of the constraint-based paradigm to the case of geological maps. As shown in Section 2 and another study [27], there is an almost infinite complexity involved when trying to solve geological map generalization comprehensively. Thus, we focus on a particular generalization problem: small, free-standing polygons (area patches in the terminology of [28]), which represent a frequent case in geological maps (Figure 1b,d,e) and other types of categorical maps, such as soil maps. Since such small polygons are prone to legibility problems in scale reduction but also often represent geological units of superior economic value (Section 4.4.2), we focus on size constraints, which are simple yet allow many of the problems tied to small polygons to be addressed. A limited set of rather simple constraints alleviates tracing the effects of using a constraint-based approach in polygonal map generalization.

4. Methodology

4.1. Overall Workflow

An overview of the proposed methodology is shown in Figure 2. The overall workflow starts from a large-scale, detailed map stored in a geological database (at scales of 1:10,000 to 1:25,000) to obtain generalized maps at medium scales (1:50,000 to 1:100,000), subsequently storing them again in the database. The actual generalization workflow consists of three main stages: identification of constraints, modeling of constraints, and generalization execution. While this part of the workflow shares some similarity with the model presented in [29], the focus here is simply on representing the sequence of the main stages. In the initial stage (Section 4.2), the relevant cartographic constraints governing geological maps are identified, in our case size constraints related to the generalization of small area patches. This is followed by the stage of constraint modeling (Section 4.3), which aims to formalize the said constraints for geological map generalization. The third and main stage is the actual generalization process (Section 4.4), which takes the previously modeled constraints and uses them to orchestrate the application of different generalization operators—elimination, enlargement, aggregation, displacement—that can be used to resolve potential constraint violations. Lastly, the obtained map is evaluated based on the goal values of each representative constraint. If the map is regarded as not satisfactory by the set goal values, control is returned to the constraint modeling stage to explore alternative options, optimizing the importance and priority values of the constraints. Once the resulting map is considered satisfactory, it is stored in the database.

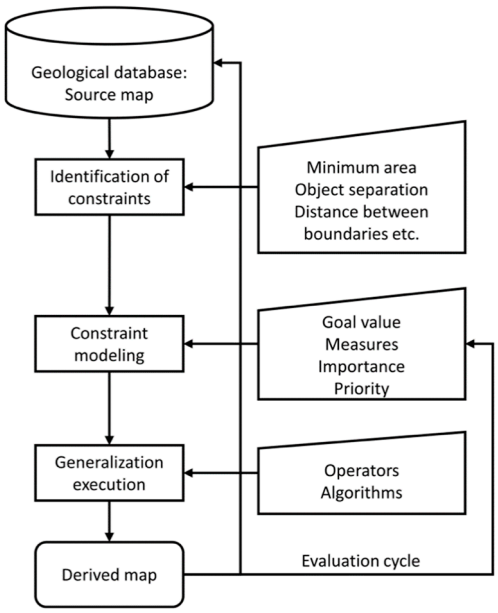


Figure 2. The overall workflow of the proposed methodology.

4.2. Identification of Constraints

As we focus on the generalization of small area patches in categorical maps, the relevant set of constraints is restricted to size constraints. Size constraints form the very starting point of the generalization process, as they deal with minimum area and distance relations in individual map features or pairs of map features. Therefore, while they may be considered to be more straightforward than, for instance, constraints dealing with groups of multiple polygons (expressing spatial pattern configurations), they are nonetheless of crucial importance as they form the foundation of all other constraints. The primary trigger of these constraints is the illegibility of map objects due to scale reduction when the limits of human visual perception are reached and polygons become too small

to be visible, or are too close, causing visual coalescence. Thus, two constraints form the core of size constraints: "minimum area" (MA), and "object separation" (OS). Based on the study of relevant literature [8,30,31], we identified the following five size constraints for our generalization of geological maps:

- Minimum area
- Object separation
- Distance between boundaries
- Consecutive vertices
- Outline granularity

4.3. Constraint Modeling

Once the constraints have been identified, they need to be further defined. Following the approach proposed in [8], which follows the agent-based paradigm, a complete constraint definition consists of the constraint itself, one or more goal values (minimum dimension to be attained on the target map), one or more measures that can quantify whether the constraint is met, possible plans (i.e., generalization operators triggered in case of constraint violation), and the importance or priority of generalization operators.

Table 1 summarizes the five size constraints, the cause that may lead to their violation, goal values, measures, possible generalization operators, and the impact that will be caused by generalization. Figure 3 illustrates the size constraints visually.

Table 1. List of size constraints and associated elements. Goal values are taken from [30] but could be modified (cf. Section 5.6).

Constraint	Cause	Goal Value	Measures	Plans/Possible Operators	Impact
Minimum area	Scale reduction	0.5 × 0.5 mm	Area	Elimination	Area loss and gain (area constraint), change of the overall configuration of polygons
				Enlargement	Change of minimum distance between features
				Aggregation	Loss of overall spatial pattern, shape distortion
Object separation	Scale reduction, enlargement	0.4 mm	Shortest distance	Displacement	Minimum distance between features, positional accuracy
				Enlargement, exaggeration	Shape distortion, ratio between features
				Aggregation	Loss of overall spatial pattern, shape distortion
				Typification	Loss of overall spatial pattern, shape distortion
Distance between boundaries	Scale reduction	0.6 mm	Internal buffer	Enlargement	Minimum distance between features
Consecutive vertices	Scale reduction	0.1 mm	Shortest distance between vertices	Elimination	Shape distortion polygon outline
Outline granularity—width	Scale reduction	0.6 mm	Shortest distance	Simplification, smoothing	Shape distortion
Outline granularity—height	Scale reduction	0.4 mm	Shortest distance	Simplification, smoothing	Shape distortion

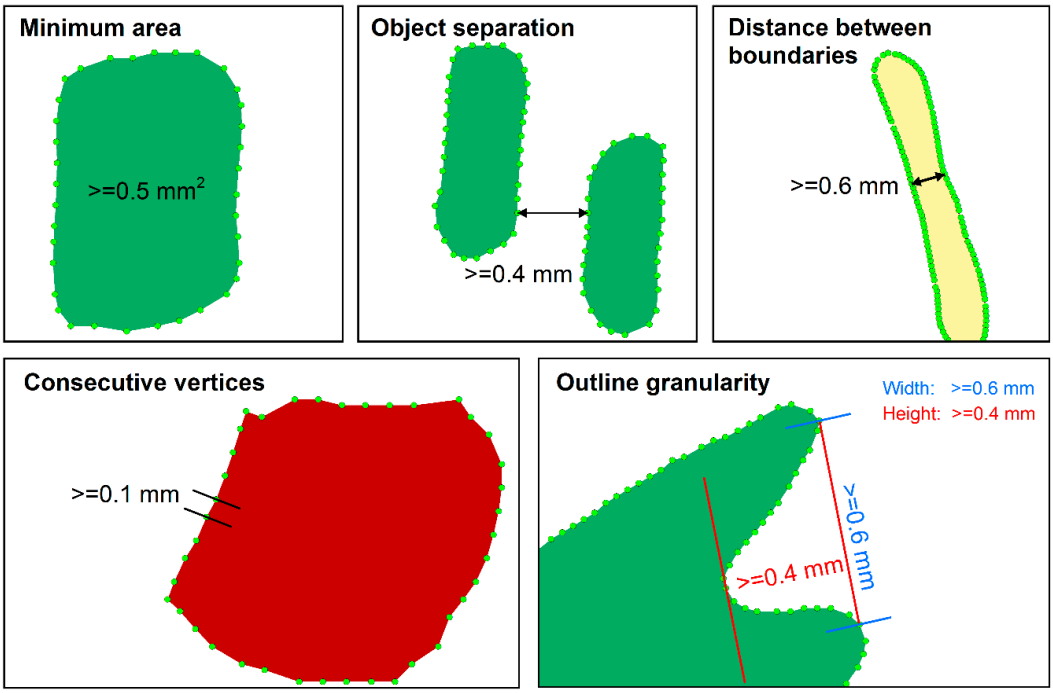


Figure 3. Graphical representation of the size constraints.

4.4. Generalization Execution

4.4.1. Generalization Workflow

The minimum area constraint is fundamental to the generalization of polygonal maps, as every type of categorical map is related to the size of polygonal features. Hence, it serves as the initial trigger and decision point for the generalization process, which is modeled after a typical process of map generalization in traditional cartography [32]. The workflow in Figure 4 is represented in two main branches, but it is essentially sequential. First, the right-hand branch deals with polygons that are too small. If they do not belong to an essential geological unit, they are eliminated. Else, they are enlarged, using a different enlargement algorithm depending on their shape. After having dealt with the elimination and enlargement operators, the workflow returns to check if now every polygon is large enough. If the answer is yes, the left-hand branch is triggered, which deals with polygons larger than the minimum area limit. Here, the next size constraint is tested. When the minimum object separation distance to its neighbors is not met, if the polygons concerned are of the same category they are aggregated or else displaced to meet the object separation constraint. Else, the polygon can be output to the target map or database. This final task also includes further line simplification and smoothing operations to take care of the "consecutive vertices" and "outline granularity" constraints.

In the remainder of this section, we will introduce the different generalization operators and algorithms one by one. Note that we assume that any reclassification operations merging sub-categories into superordinate categories take place before our workflow.

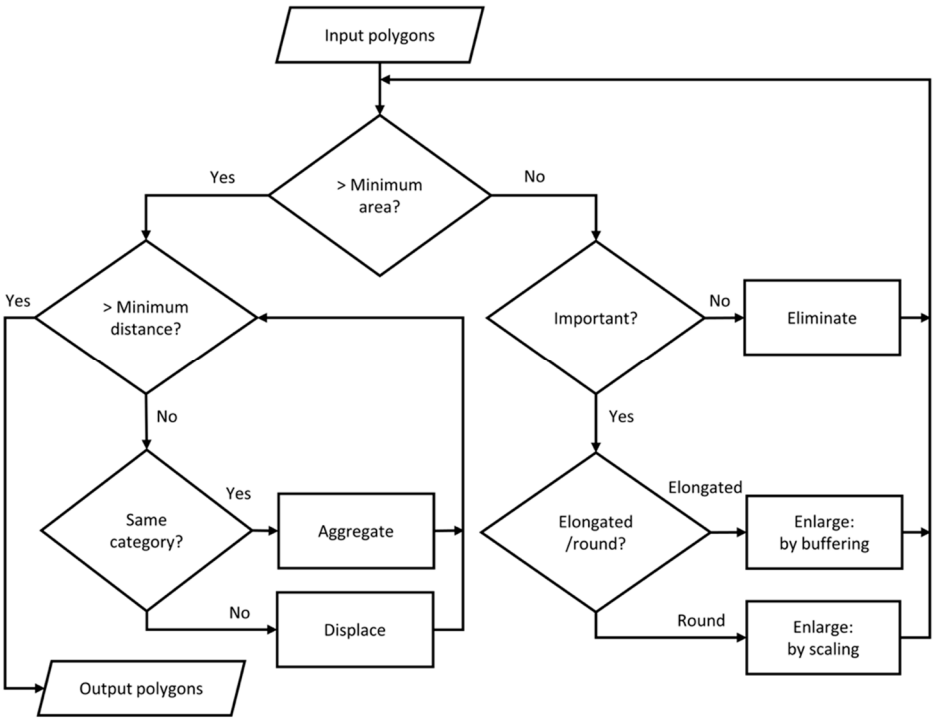


Figure 4. The workflow of the polygon generalization stage, governed by size constraints.

4.4.2. Elimination

The elimination operator dissolves small polygons into background polygons. It is triggered after a polygon fails two tests: it does not meet the MA constraint, and it is considered not sufficiently important to be nevertheless retained. However, the two tests are not applied in an unbounded way on the polygons. Still, a global target value is first determined, giving the number of polygons to be eliminated (n_{elim}) from the source map (or conversely selected for the target map). Three different selection methods are devised, all factoring in both the MA constraint and the importance of polygons given by their category (i.e., the geological unit in our case).

The first selection method, the "Radical Law" [33], is used as a baseline, as it is often used as a reference in map generalization. Consequently, the number of polygons to be eliminated (n_{elim}) is determined using the Radical Law. A detailed example of how n_{elim} is determined is given in Section 5.2. Once n_{elim} has been established, the polygons to be removed must be identified. The polygons falling below the MA threshold are sorted in ascending order according to the sum of their normalized area and their normalized geological importance value. The importance of a polygon is a function of its geological unit, which reflects geological age. Older units, such as "amphibolite" and "pegmatite", will have more mineral deposits and hence more economic value than newly generated rocks [11]; younger rock types, such as "quaternary", are thus considered less critical than ancient "Jurassic" layers. After the sorting order has been established, the n_{elim} smallest and least important polygons are then eliminated.

In the second selection method, called "area loss–gain selection", the polygons are first sorted by category in ascending order of their area, and polygon removal then also proceeds by category (i.e., by geological unit). In each category, as many small polygons are removed as are needed to balance the area gain of the remaining same-class, small polygons treated by the enlargement operator (next subsection). Thus, the total area of the removed polygons per category is always balanced with the area gained by the enlargement operator. For that purpose, this selection operator for each polygon pre-computes the area gain that would be added if the polygon was kept and enlarged.

The third approach, termed as "category-wise selection", applies the Radical Law separately for each category, as indicated by its name. For instance, if a map has 5 categories and according to the Radical Law 30% of polygons should be removed, then in each category 30% of the smallest polygons are eliminated. Note that this method does not take into account the area loss and gain of the polygons, yet it better balances the removal of polygons from each category than the pure Radical Law does, protecting categories consisting only of small polygons from vanishing.

4.4.3. Enlargement

After removing small and unimportant polygons in the elimination step, some polygons will remain that are smaller than the MA limit, but owing to their importance are flagged for retention. Therefore, the next step is to enlarge them until they reach the MA limit, and adequate readability by the human eye is ensured. Two algorithms are devised to accomplish enlargement, buffering, and scaling. Figure 5 illustrates the effect of these two algorithms: enlargement by buffering generates round polygon shapes while scaling enlarges the polygon area and maintains the initial polygon shape.

Polygon buffering is a commonly used operation in GIS and is straightforward to accomplish. However, its tendency to generate round shapes may lead to the loss of the polygon's original shape, mainly if very small polygons are involved and/or the buffer distance is significant. Hence, we recommend using buffer enlargement when only a smaller increase is necessary, and when polygons have a more elongated shape, where shape distortion is less noticeable.

The scaling algorithm can handle round polygons and extend small polygons without distortion of the shape. The algorithm loops over every vertex and calculates its distance to the point of reference, which is the polygon centroid by default. The scaled polygon is then produced by extending a line from the reference point across each initial vertex at the scaled length, with the endpoint becoming the scaled vertex. One disadvantage of the basic scaling algorithm is that when the polygon centroid is used as a reference point, topological errors may be created for non-convex polygons (Figure 5c), necessitating displacement to restore the topology. The enlarged polygon in Figure 5c is shifted from its original position to the right-hand side and below due to the scaling process.

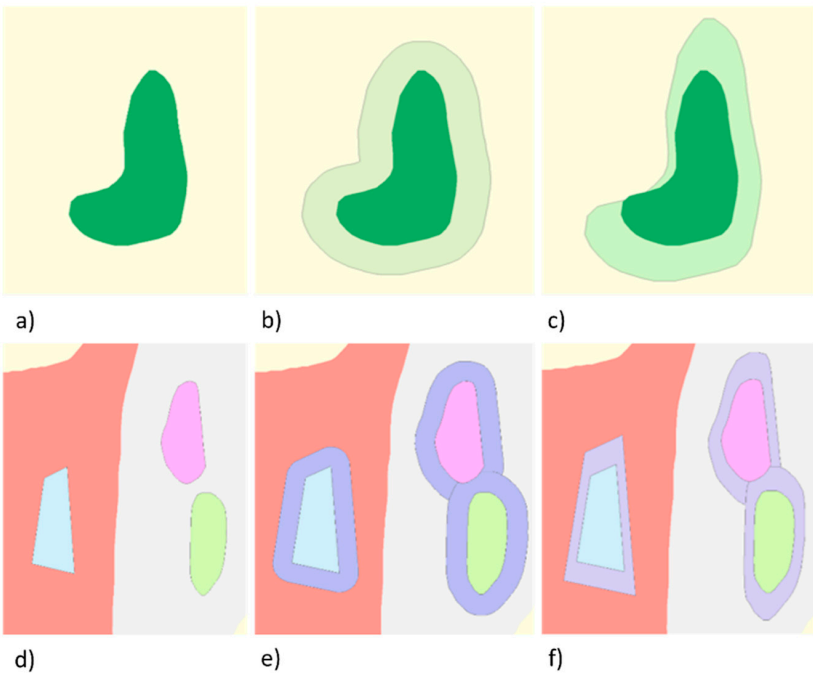


Figure 5. Enlargement. (**a**,**d**) The original polygons. (**b**,**e**) Those enlarged by the buffering algorithm. (**c**,**f**) Those enlarged by the scaling algorithm.

The choice between the two enlargement algorithms for each polygon is guided by an analysis of their shape using the "ipsometric quotient" (*IPQ*) shape index [34], used as a measure of compactness in [35]:

$$IPQ = 4\pi \times \frac{A}{P^2} \tag{1}$$

where A denotes the area and P the perimeter of the polygon. The IPQ takes values in the interval (0,1), where values approaching 0 indicate elongated shapes, while a value of 1 indicates a circle. We used a threshold of 0.5 in our experiments to distinguish between elongated shapes (for buffering) and round shapes (for scaling), as this threshold marks the midpoint between the two extremes.

The buffer width K_b that needs to be added to the polygon is calculated as a function of the target area A' to be reached at the target scale (e.g., 625 m^2 at 1:50,000), the current area A, and current perimeter P

$$K_b = \frac{A' - A}{P} \tag{2}$$

The scaling factor K_s that allows the algorithm to enlarge a polygon to the target area A' (e.g., 625 m^2 at 1:50,000) equals the square root from the target area A' over the current area A.

$$K_s = \sqrt{\frac{A'}{A}} \tag{3}$$

4.4.4. Aggregation

After execution of the enlargement operator, all remaining polygons should be sufficiently large for the target scale and thus adequately readable. However, enlargement also may give rise to another conflict: the violation of the object separation constraint, which is dealt with the aggregation and displacement operator, respectively, in this work. Both are triggered by a violation of object separation, but aggregation resolves this problem by merging polygons that are too close but also similar, while displacement moves dissimilar polygons until the minimum separability distance is again met. A conflict between features is identified by calculating the distance between any two polygons, where the polygons are closest to each other.

The similarity of polygons may be determined using several properties, such as the shape, size, orientation, and category of involved polygons [10], but in this paper, we use only the semantic similarity, that is, same category, to denote two polygons as being similar. The semantic similarity is of overriding importance: it makes only sense to aggregate polygons of same categories (or possibly of shared subcategories, if they exist).

Various algorithms have been proposed for the aggregation and typification of buildings [36,37], but for our purposes, we were more interested in an algorithm that would fill in voids between polygons that are found to be too close, akin to the algorithm in [25]. The aggregation is performed by generating a concave hull from the outline vertices of the polygons to be aggregated using the algorithm by [38]. The inset maps of Figure 12 highlight sample results of this procedure. The algorithm starts by selecting the first vertex as the one with the smallest Y value. Next, k points nearest to the current vertex are selected as candidates to be the next vertex for the output polygon. The next vertex is decided based on the biggest right-hand turn from the horizontal axis. These two steps are repeated until the selected candidate is the first vertex. Finally, the consecutive vertices found are connected by line segments to form the resulting concave polygon. Adjusting the parameter k allows for control of the degree of concavity of the resulting aggregated polygon.

4.4.5. Displacement

There are two reasons why the displacement operator may be activated. First, as a direct consequence of scale reduction, visually differentiating between map objects becomes harder and at some stage, impossible (an effect often termed congestion; [18]). Second, as a consequence of polygon

enlargement, the object separation constraint may be violated, as enlarged polygons require more space and may start to coalesce or even overlap. In both cases, map objects should be either aggregated or displaced to the minimum separability range. In instances where polygons are densely clustered, alternative operators such as typification (which is a combination of elimination, aggregation, and displacement) may be preferred.

Figure 6 illustrates the effect of the displacement operator, triggered by a violation of the object separation constraint following polygon enlargement. After the violation has been detected, polygons are moved to the minimum separability distance from each other.

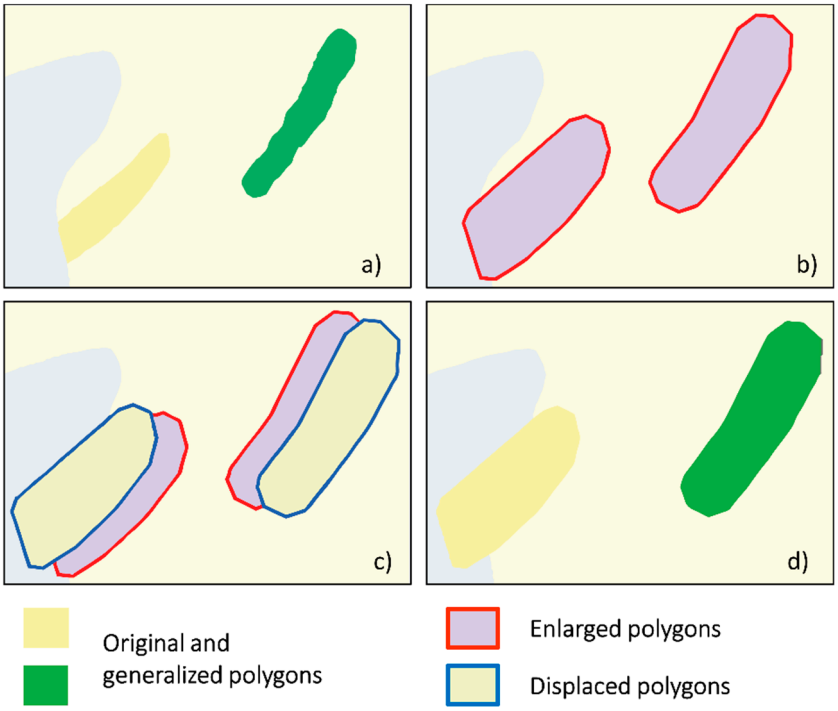


Figure 6. Displacement triggered by polygon enlargement. (**a**) Original polygons that need to be enlarged. (**b**) Enlarged polygons, in conflict with object separation. (**c**) Overlay of enlarged and displaced polygons. (**d**) Displaced polygons.

We devised two algorithms to implement the displacement operator: a simple pairwise displacement algorithm used when only two polygons are in conflict, and a Voronoi-based algorithm for cases involving multiple polygons. Note, however, that these could be replaced by other, more sophisticated algorithms featuring different displacement characteristics if so desired [39,40].

The pairwise displacement algorithm was inspired by [41] and considers a pair of polygons that are too close to each other to be distinguishable and must, therefore, be pushed back to the object separation distance. The algorithm finds the nearest points between the two polygons, calculates the movement vectors for each polygon, and displaces the polygons from each other (Figure 7). The size of the polygons controls the degree of displacement of each polygon: the smaller the polygons, the more they move, while larger polygons move less. In the example in Figure 7, the polygon with an area of 8666 m^2 moves only 5 meters, while the polygon with an area of 2471 m^2 is displaced 18 meters from its initial position.

If multiple polygons are in conflict, the pairwise displacement algorithm cannot guarantee that the minimum separability constraint is met everywhere, as it is unable to safeguard against possible knock-on effects when, e.g., a polygon is pushed into another, neighboring polygon that would then also need to be displaced. Therefore, conflict zones involving three or more polygons are handled

and are achieved by a Voronoi-based displacement approach inspired by algorithms using Voronoi or Voronoi-like cells to control the displacement [42,43].

The Voronoi-based method is divided into three steps. First, polygons are identified that violate the object separation limit (Figure 8a). In the second step, polygons are collapsed to their centroid points, which serve as seed points for the Voronoi diagram (Figure 8b). From the perpendicular bisectors of the Voronoi polygons (i.e., the edges of the Delaunay triangulation), the displacement vectors are identified for each polygon pair. Finally, the polygons are pushed to the minimum separation distance while forcing the polygons to remain within their Voronoi polygon (Figure 8c).

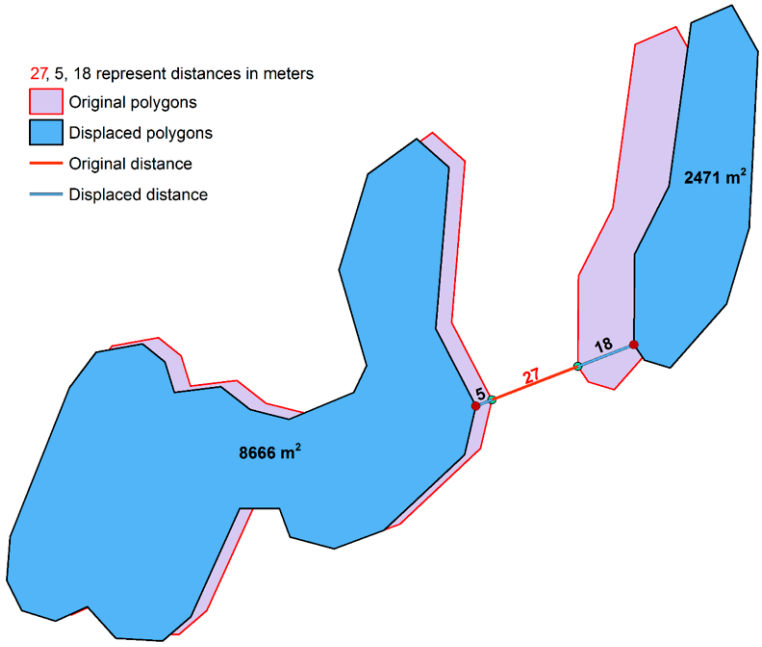


Figure 7. Pairwise displacement: Finding the closest points and calculating the displacement vector based on the proportion the polygons contribute to propagating the displacement.

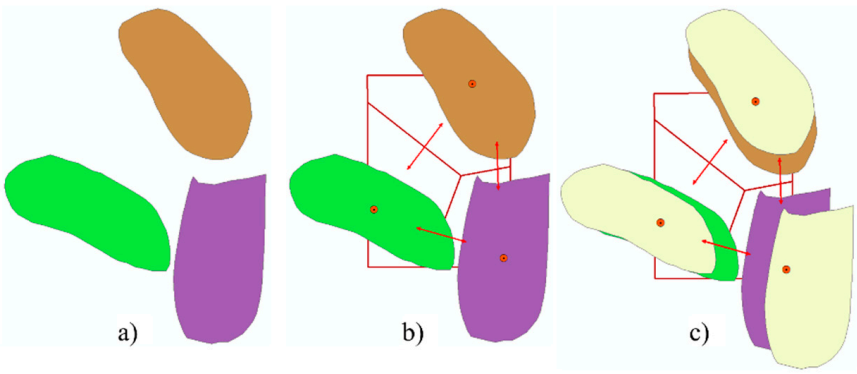


Figure 8. Voronoi-based algorithm. (**a**) Original polygons. (**b**) Generation of Voronoi polygons used to propagate displacement. (**c**) An overlay of the original and displaced polygons.

Figure 9 features two examples of the application of the Voronoi-based displacement algorithm with both the original and displaced polygon positions shown.

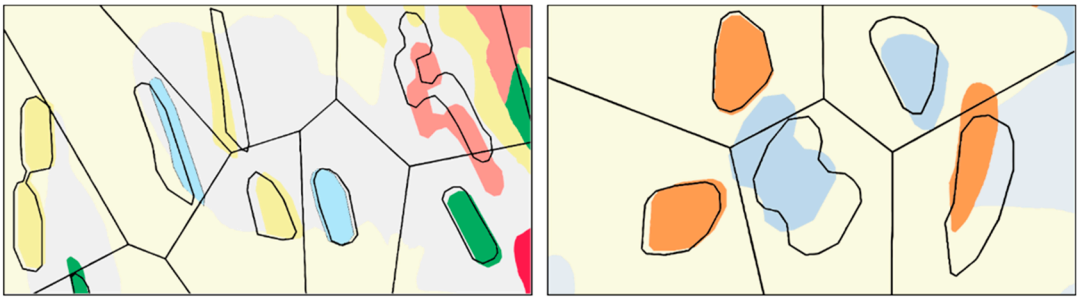


Figure 9. Two examples of the Voronoi-based approach: polygons conflicting with the minimum distance constraint are displaced to the minimum separability distance.

The completion of the displacement operator also marks the end of the process laid out in Figure 4. This process was initially triggered by the minimum area constraint and eventually ensured that this constraint as well as the object separation constraint were fulfilled. All that is left now are the finishing touches by simplification and smoothing of the polygon outlines to ensure the "consecutive vertices" and "outline granularity" constraints are also met ("Output polygons" step in Figure 4). This step, however, uses standard algorithms with conservative parameter settings and thus does not need to be described in detail here.

4.5. Implementation

The proposed workflow uses GeoPandas 0.6.0, an open-source project that handles geospatial data in the Python programming language. Geopandas methods such as dissolving, buffering, centroid identification, convex hull, rotation, scaling, and translation are used to remove, enlarge, aggregate, and displace polygons. Moreover, the ArcPy package from Esri, Inc., is used, implementing ArcGIS toolbox capabilities such as calculating area or finding near and nearby objects. Map visualization is undertaken using Esri ArcGIS 10.6 and Matplotlib, a plotting library for the Python programming language.

5. Experiments and Results

5.1. Data

The data used for our experiments were taken from the "Euriowie Block (including part of Campbells Creek)" 1:25,000 scale geological map published in the year 2000 by the Geological Survey of New South Wales (NSW), Department of Mineral Resources, Australia (https://www.resourcesandgeoscience.nsw.gov.au). A portion of this map is shown in Figure 10, left. The "Euriowie Block" is part of the bigger geological block called the "Broken Hill Block", located near the mining city of Broken Hill, NSW. It contains significant mineral resources, particularly abundant and rich lead–silver–zinc (Pb–Ag–Zn) deposits. The Broken Hill Block's geology is highly complex, and historically there has always been extensive exploration activity in the area [44,45], which gave us grounds for choosing this particular map. Moreover, the map is openly available for public and research purposes and thus promotes the reproducibility of the results.

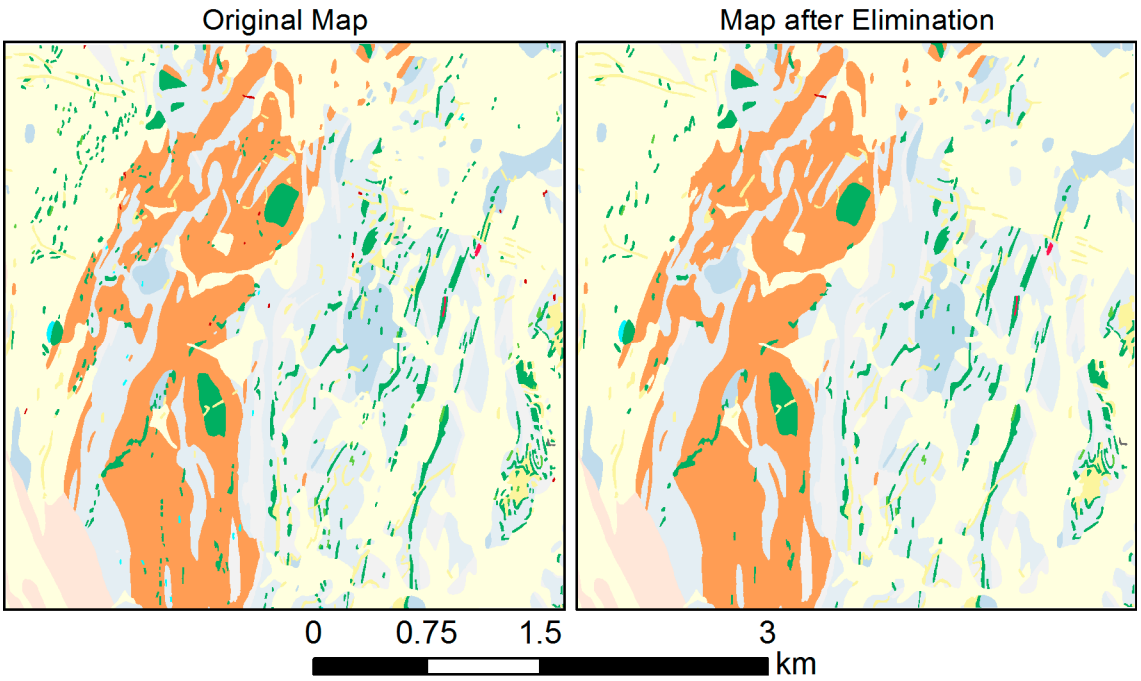


Figure 10. Part of the geological map of the Euriowie Block, NSW, Australia. The original map, left, was produced at the scale of 1:25,000, but is shown here at the scale of 1:50,000, as are all the maps in Figures 10–13. On the right, a map after elimination of less important polygons using the "Radical Law selection" approach is shown.

The map's objective is to present a broad distribution of rock types, especially those with mineralization or stratigraphic significance. The data are of moderate to high quality, reliability, and internal consistency. It has an accuracy that is appropriate for a map of a scale of 1:25,000.

The geological map consists of 215 sub-categories of rock types or units, which in turn are part of 24 higher categories. Small-sized minerals such as the "hornblende" and "feldspar" groups make up approximately 60% of terrestrial rocks by weight. These mineral groups are mostly related to plutonic (also called intrusive) igneous rocks such as pegmatites (shown in yellow in Figure 10), which form deep under the Earth's surface and are created by the solidification of magma [46]. They are also present in many types of metamorphic rocks, such as amphibolite (shown in green in Figure 10). Following erosion, they become exposed at the Earth's surface, often in small patches, represented as small polygons on geological maps. Since the amphibolite and pegmatite units have more mineral deposits and are thus of higher economic value, they will be retained longer in the scale reduction process. Since they often form small polygons, these rock types are good candidates for dealing with minimum size constraints, which are the focus of this paper.

5.2. Elimination

In the sequence of experiments reported in Sections 5.2–5.5, we use goal values for the MA constraint of 0.5 × 0.5 mm and for the object separation constraint the value of of 0.4 mm given in [30]. The whole Euriowie map database consists of n_s = 1877 polygons at the source scale of 1:25,000. With a target scale of 1:50,000, the number of polygons n_{MA} that fall below the MA limit is 1135, given a minimum area of 0.5 × 0.5 mm on the map or 625 m^2 on the ground; (Table 1). This means that almost 60% of all polygons are smaller than what may be defined as the legibility limit. Using the "Radical Law selection" strategy for the elimination operator (Section 4.4.2), the number of polygons that should

be retained, n_t, can be computed as follows (where N_s is the denominator of the source scale and N_t the denominator of the target scale):

$$n_t = n_s \sqrt{N_s/N_t} = 1877 * \sqrt{25,000/50,000} = 1327 \quad (4)$$

That is, according to the Radical Law [33], n_t = 1327 polygons should be retained, or conversely, $n_{elim} = n_s - n_t$ = 550 polygons should be eliminated. Finally, the number of polygons falling below the MA constraint that should be kept, n_{keep}, can be obtained as $n_{keep} = n_{MA} - n_{elim}$ = 585.

Thus, with the Radical Law selection strategy, n_{elim} = 550 polygons with an area below the MA threshold should be eliminated from the map, while n_{keep} = 585 polygons are also smaller than the MA threshold, but are deemed sufficiently important to be still kept on the map, and thus forwarded to the subsequent generalization operators such as enlargement.

To determine which polygons are eliminated, their importance value is established based on their area as well as their position in the "geological hierarchy" (Table 2). The area values occurring among the n_{MA} polygons smaller than the MA limit range from 7 m^2 to 625 m^2. The geological hierarchy follows the order of geological units according to their age: the older the rocks, the more important they are considered, as they likely have more mineral content. Overall, in the Euriowie Block, 15 geological units can be found, which are assigned a value of 15 for the oldest unit and 1 for the youngest unit, respectively (Table 2). For each polygon, its area and its importance value, respectively, are then normalized and summed to obtain the integrated importance value per polygon. After sorting in ascending order, the n_{elim} (550) least important polygons are then terminally eliminated using the geopandas.dissolve() algorithm of GeoPandas 0.6.1, which dissolves selected polygons into those neighboring polygons, with whom they share the largest border.

Figure 10, right, shows an excerpt of the Euriowie map after the elimination operation using the Radical Law selection approach compared to the original map, on the left. Examples using the other two selection strategies of the elimination operator will be presented in Section 5.6.1.

Table 2. Assigning importance values to polygons.

#	Property	Value Measurement	Normalized Importance
1	Area of polygons	7–625 m^2	0.0–1.0
2	Geological hierarchy	1–15	0.0–1.0

5.3. Enlargement

Those polygons that are smaller than the MA threshold but sufficiently important to be kept on the target map (n_{keep} = 585) need to be enlarged until they reach the minimum area (Section 4.4.3). Depending on the IPQ shape index of polygons, two separate algorithms, buffering and scaling, are used for enlargement.

Overall, 234 polygons (roughly 40%) out of 585 polygons had IPQ values lower than 0.5, indicating rather elongated shapes, and were thus enlarged by buffering. Accordingly, 351 polygons (around 60%) had IPQ values above 0.5, indicating rather round shapes, and were thus enlarged by scaling. The result of the enlargement step is shown in Figure 11.

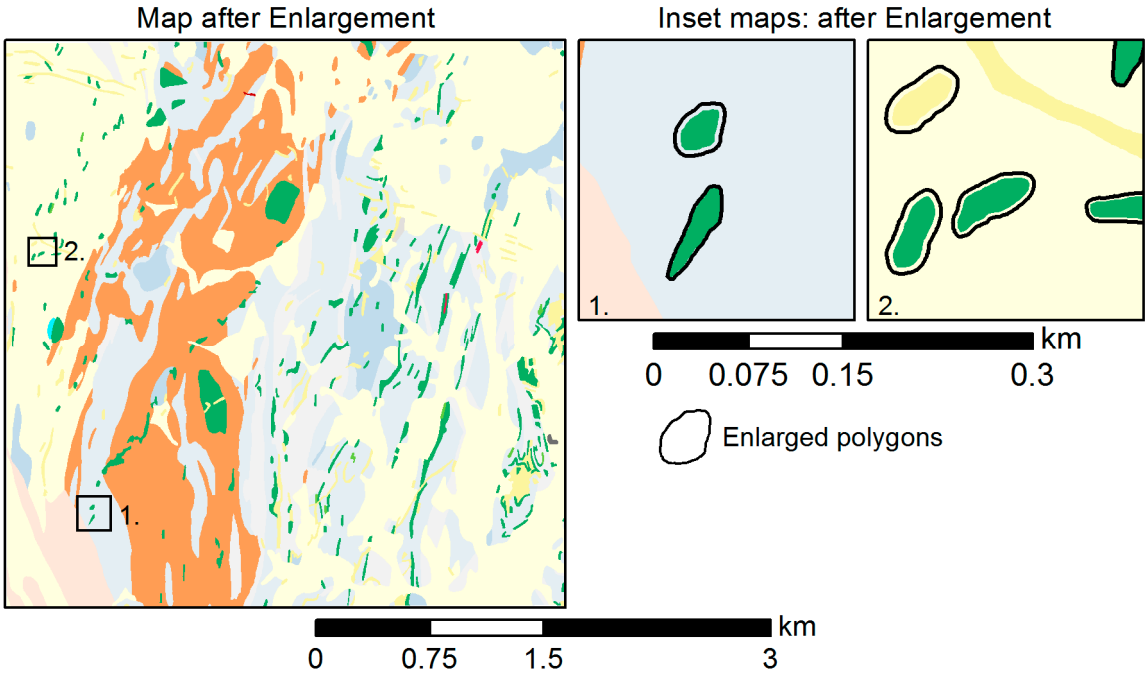


Figure 11. Part of the map after the enlargement of the number of polygons falling below the minimum area (MA) constraint that should be kept (n_{keep}), including two detailed examples illustrating enlarged polygons.

5.4. Aggregation

As the next operation, aggregation combines polygons of the same category (i.e., same geological unit) that fall within the object separation (OS) distance by applying the concave hull algorithm (Section 4.4.4). We used OS = 0.4 mm (Table 1) corresponding to 20 meters on the ground at scale 1:50,000 and k = 3 in the concave hull algorithm to produce the result depicted in Figure 12. Overall, of the n_{keep} = 585 polygons, 78 were aggregated, leaving 507.

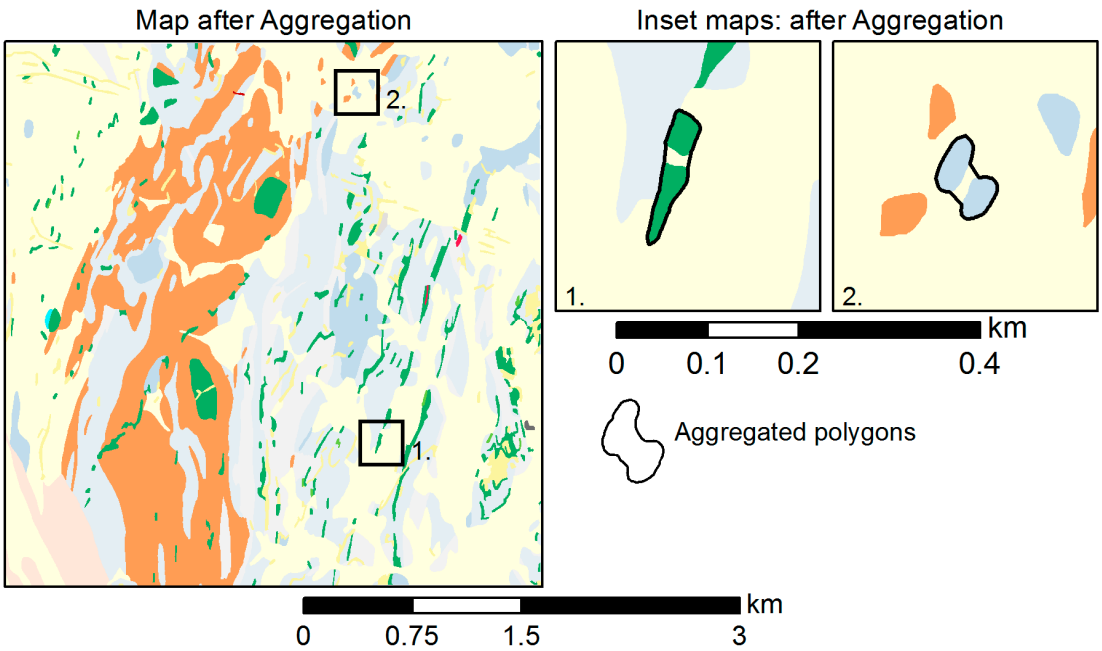


Figure 12. Part of the map after aggregation of polygons closer than the object separation (OS) threshold, including two detailed examples illustrating aggregated polygons.

5.5. Displacement

As a final step, the displacement operator is executed to relocate polygons that are too close to each other (i.e., not meeting the OS limit) from their original position to a minimum separability distance. An excerpt from the result of this operation, and thus the final generalization product, is shown in Figure 13 in comparison to the source map. Overall, 35 polygons were displaced, 24 using the pairwise displacement algorithm and 11 using the Voronoi-based approach.

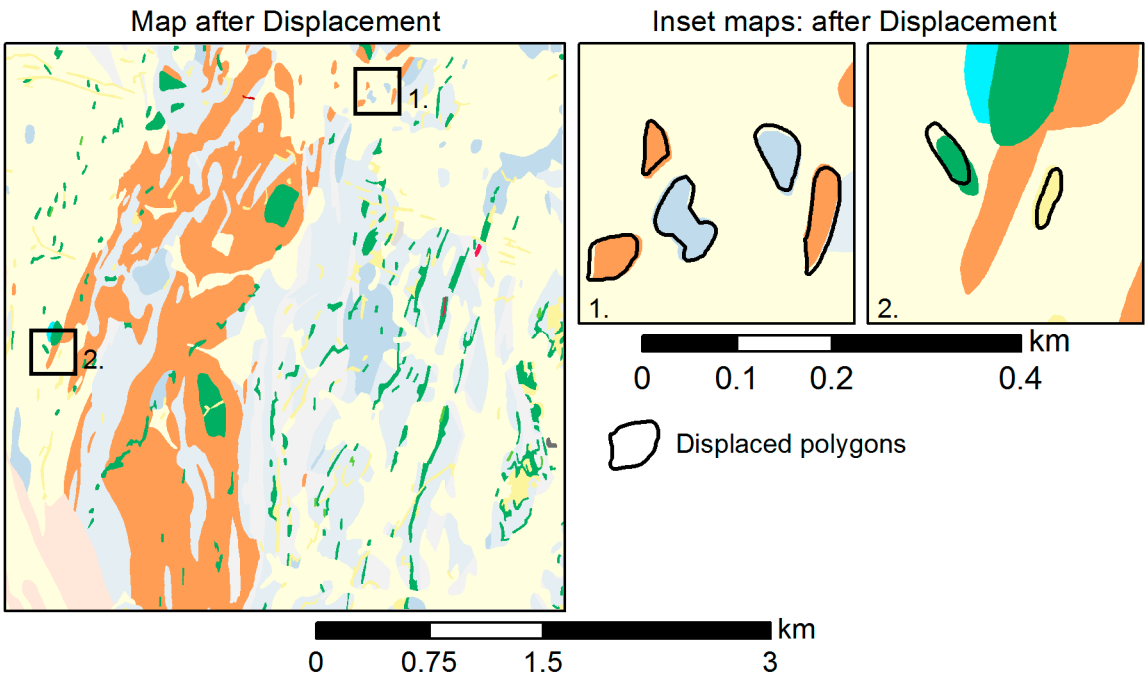


Figure 13. The generalized map after displacement. To the right, two detailed examples illustrating displaced polygons are shown.

5.6. Sensitivity to Parameter Settings

The operators and algorithms of the proposed methodology are controlled by a variety of parameters, all of which can take different values depending on the purpose and application domain of the map in question. Even for the various size constraints, although we defined goal values in Table 1, no canonical or "right" values exist. Hence, in three test cases, we illustrate the sensitivity of the generalization results in variations of the control parameters. Note that in all test cases, no simplification and smoothing was applied. These operators would normally be applied using conservative parameter settings and thus have little effect on the final result (Section 4.4.1).

5.6.1. Test Case 1

In the first test case, the influence of using the three different selection methods for the elimination operator was tested on 1877 polygons contained in the source map at 1:25,000 scale as shown in Figure 14, in the transition to a target scale of 1:50,000. This experiment used the same goal values for the MA and OS constraint as in Sections 5.2–5.5. In the "Radical Law selection", 29.29% or 563 of the source polygons were removed (Table 3). In "area loss–gain selection", only 139 were eliminated. Finally, "category-wise selection" resulted in the removal of 417 polygons. Figure 14 allows the visual comparison of the selection methods with each other and the original map. The figure also includes four inset maps for a close-up comparison of the results.

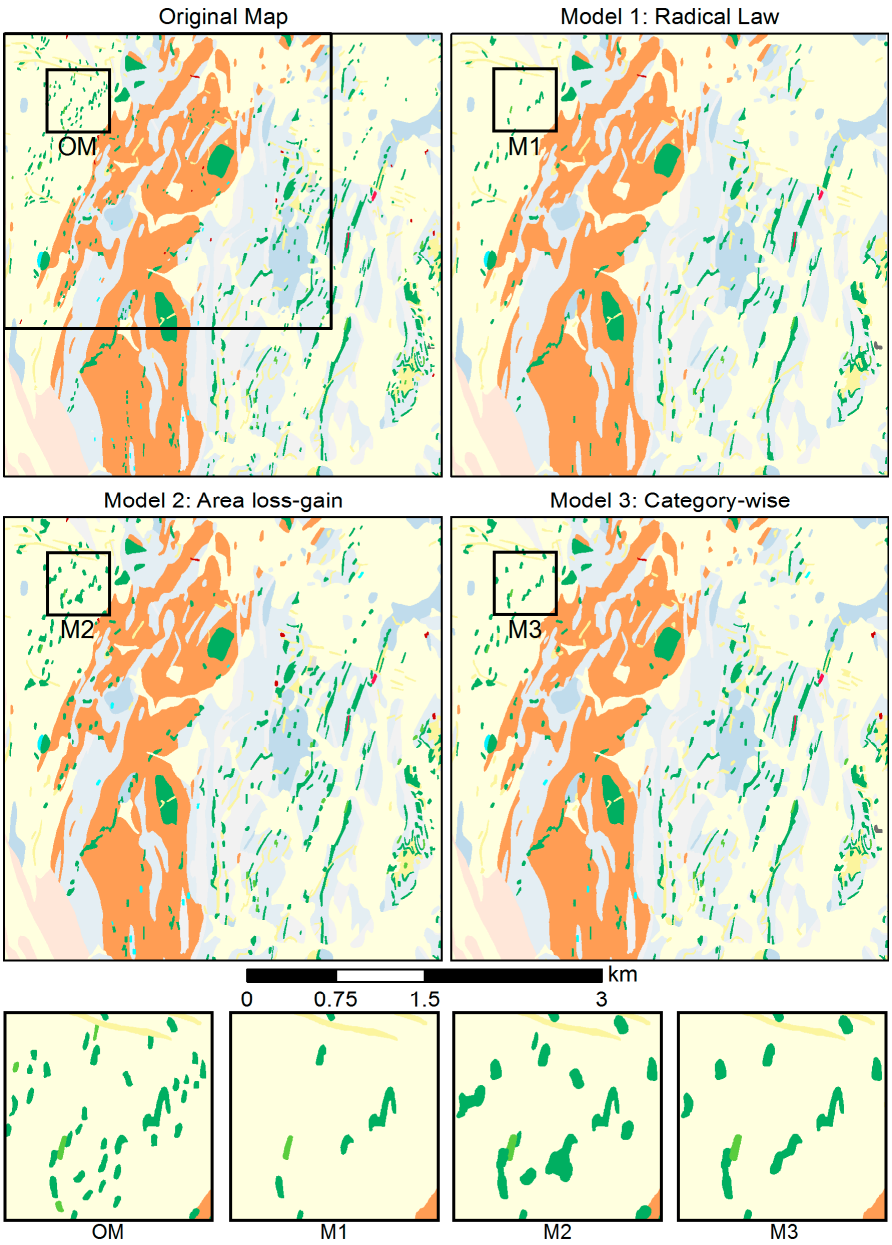


Figure 14. Effect of the selection method in the elimination operator. The large black rectangle in the original map denotes the map window used in Figure 16. The display scale is 1:50,000. The bottom row shows four magnified examples: OM (original map), M1 (Model 1), M2 (Model 2), M3 (Model 3).

Table 3. Comparison of methods of selection in the elimination operator.

Selection Method	Number of Polygons		
	Source Map (25,000)	Target Map (50,000)	Removed
Radical Law selection	1877	1314	563
Area loss–gain selection	1877	1738	139
Category-wise selection	1877	1460	417

5.6.2. Test Case 2

In the second test, we compared three different sets of thresholds for the size constraints, called "Fine-grained" (FG), "Compromise' " (CM), and "Coarse-grained" (CG) (Table 4). The first threshold set was inspired by [30]. While their fine-grained goal values were defined for regularly shaped polygons

on topographic maps, the authors noted that larger values should be used for tinted and irregularly shaped polygons, such as those typical of categorical maps. The Coarse-grained constraint set was positioned at the other end of the granularity scale and was obtained from [31], a source devoted to the symbolization of geological maps, defining the minimum size of tinted polygons to be around 2 mm^2. The Compromise constraint set tries to balance between the Fine-grained and the Coarse-grained sets. The visual comparison of the generalization results produced using the three constraint sets is given in Figure 15. The figure also presents a close-up comparison of the three approaches and the original map.

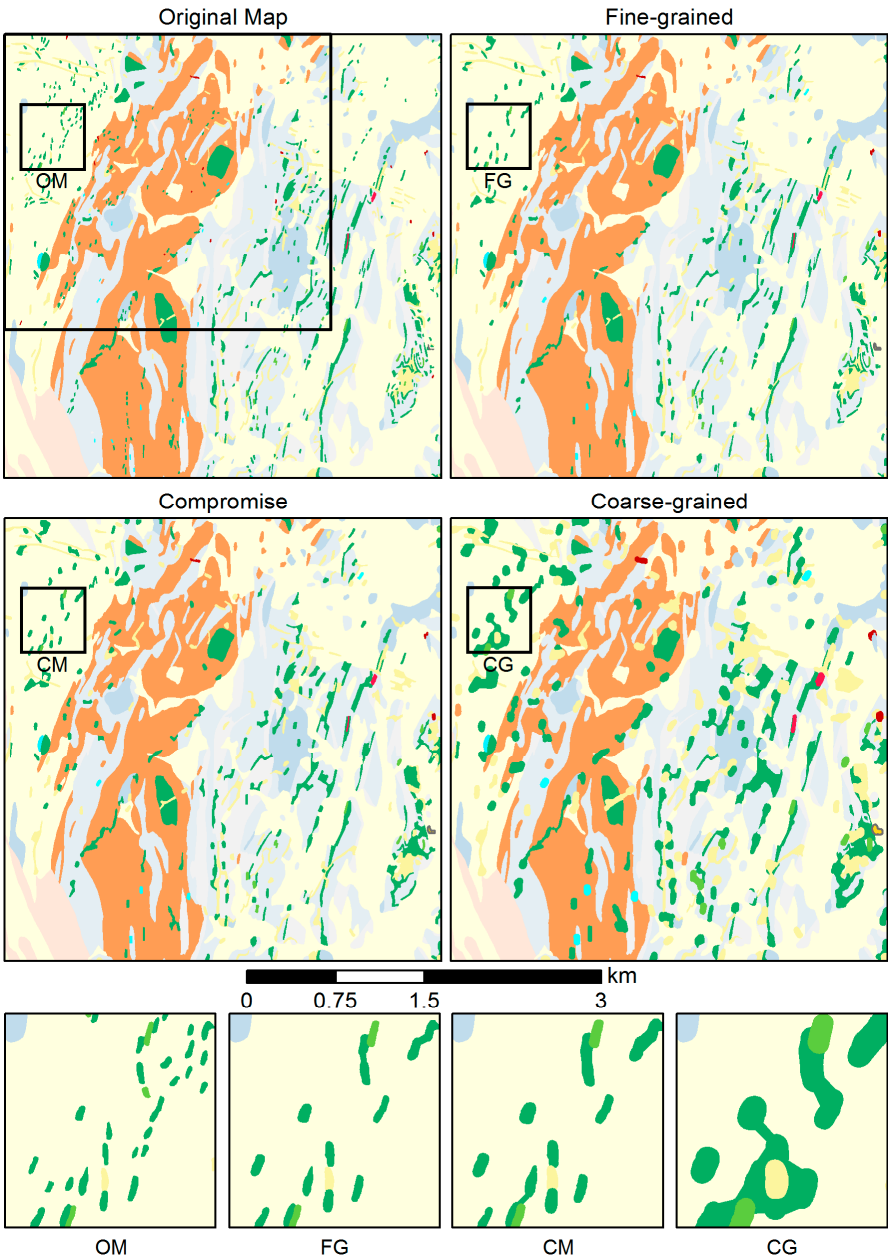


Figure 15. Effect of varying the thresholds for the size constraints using the three sets of goal values from Table 4. The large black rectangle in the original map denotes the map window used in Figure 16. The display scale is 1:50,000. The bottom row depicts four magnified examples: OM (original map), FG (fine-grained), CM (Compromise), CG (coarse-grained).

Table 4. Goal value sets for size constraints used in Test Case 2. MA: minimum area; OS: object separation, #poly = number of polygons.

Threshold Set	MA	Avg. Area	#Poly	OS	Source
Fine-grained (FG)	0.5 × 0.5 mm	19,792 m^2	1265	0.4 mm	[30]
Compromise (CM)	0.75 × 0.75 mm	20,572 m^2	1217	0.6 mm	compromise
Coarse-grained (CG)	2 × 2 mm	26,244 m^2	954	1.0 mm	[31]

5.6.3. Test Case 3

As a final test, we used the Compromise goal values set and the Category-wise selection method, which had both been found to perform best in the above tests, to produce a series of maps at the scales of 1:50,000, 1:100,000, and 1:200,000, generalized from the original 1:25,000 scale map. The interest of this test was in exploring the scale range over which the proposed methodology could be usefully applied. The visual comparison of the resulting generalization series with the original map is presented in Figure 16. Table 5 shows the evolution of the number of polygons and their combined area of four selected geological units present in Figure 16: amphibolite, pegmatite, late Proterozoic, and the soil–sand–gravel–clay unit. Amphibolite (green polygons) and pegmatite (yellow polygons) were chosen because they represent two classes that occur both mainly as small, frequent polygons, and are of high importance due to their age. Late Proterozoic (orange-brown) and soil–sand–gravel–clay (light yellow) units, on the other hand, are young and of the youngest age, respectively, and populate the study area as few but large polygons.

Table 5. Effect of scale reduction using the Compromise set of goal values for the MA and OS constraints (Table 4). The numbers are given for the map window of Figure 16.

Scales	Geological Units	Total Area in m^2 (%)	Polygons
Original (1:25,000)	All units	25,032,549 (100%)	1877
	Amphibolite	1,177,891 (4.71%)	838
	Pegmatite	915,132 (3.66%)	393
	Late Proterozoic	713,420 (2.85%)	10
	Soil–sand–gravel–clay	12,371,617 (49.42%)	18
1:50,000	All units	25,032,549 (100%)	1349
	Amphibolite	1,382,525 (5.52%)	587
MA—1406.25 m^2	Pegmatite	888,572 (3.55%)	276
OS—30 m	Late Proterozoic	713,020 (2.85%)	9
	Soil–sand–gravel–clay	12,333,353 (49.24%)	17
1:100,000	All units	25,032,549 (100%)	953
	Amphibolite	1,584,974 (6.32%)	419
MA—5625 m^2	Pegmatite	1,317,895 (5.26%)	197
OS—60 m	Late Proterozoic	697,109 (2.78%)	5
	Soil–sand–gravel–clay	11,857,308 (47.31%)	14
1:200,000	All units	25,032,549 (100%)	667
	Amphibolite	1,766,809 (7.05%)	294
MA—22,500 m^2	Pegmatite	3,126,179 (12.48%)	138
OS—120 m	Late Proterozoic	673,570 (2.69%)	5
	Soil–sand–gravel–clay	10,120,376 (40.40%)	7

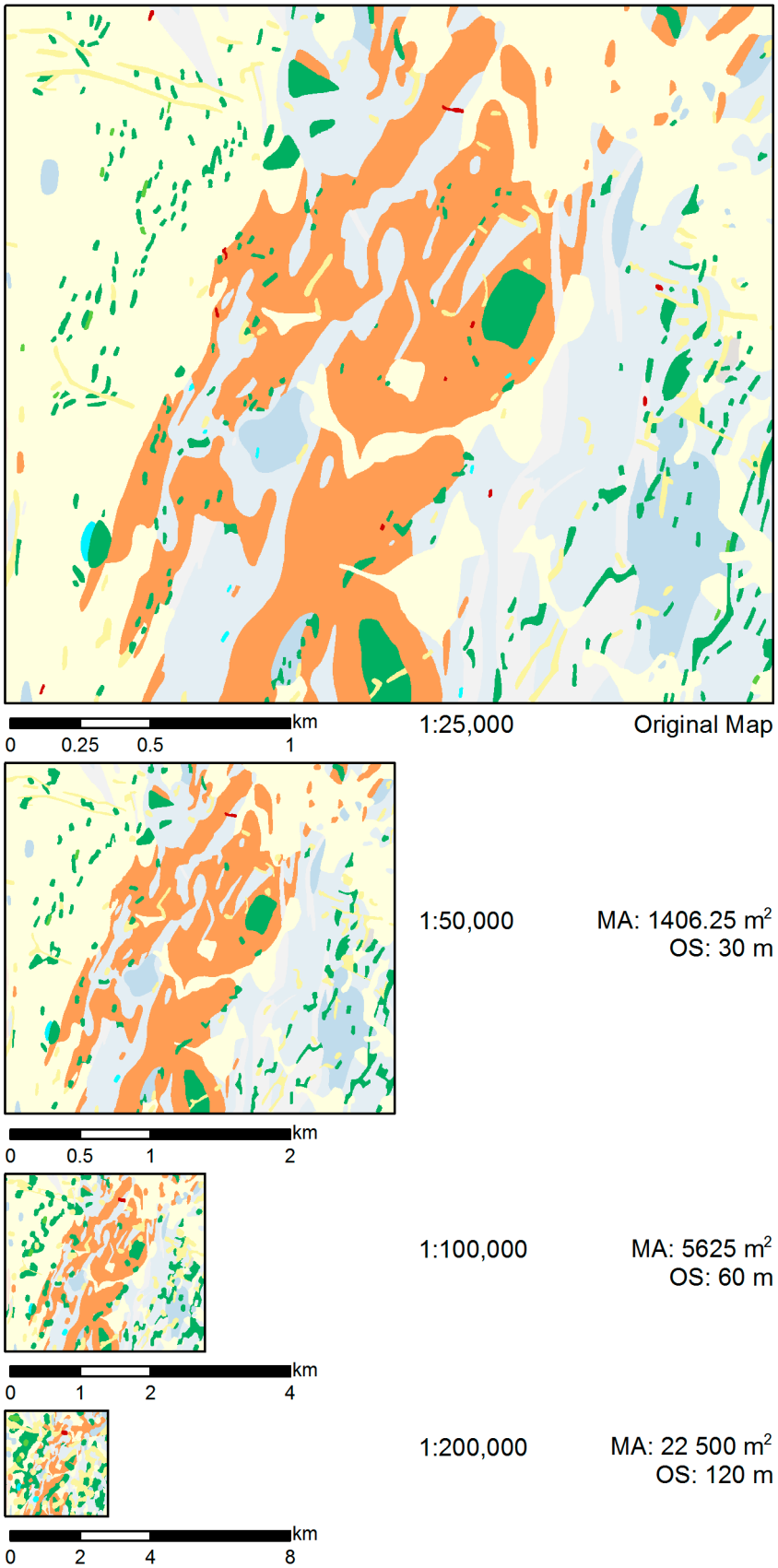


Figure 16. Scale series produced using the Compromised set of goal values for the MA and OS constraints.

5.6.4. Comparison with Cellular Automata Approach

In this section, we compare the results that can be obtained by the proposed methodology with those of an approach based on cellular automata (CA) [23,24], as implemented in the GeoScaler toolbox [47]. As mentioned in Section 2, this approach can probably be considered as the state of the art of geological map generalization available in practice, which is why we used it as a benchmark for comparison. The result of our constraint-based approach was generated using the CM goal values and the category-wise selection method (Figure 17, left). This result is equivalent to the map shown in Figure 15 (bottom left). Since GeoScaler is based on a completely different approach (raster-based, using a CA) as compared to our methodology (which works on individual polygons), it is hard to exactly match the parameter settings for the two methods. GeoScaler also offers a variety of postprocessing tools to further enhance the initial result of the CA algorithm. In order to be able to compare the results of the two methods on a level playing field, we decided not to make use of the GeoScaler postprocessing options and limit the parameter settings to those that find a correspondence between the two methods (i.e., parameters that relate to size constraints). GeoScaler [47] first asks the user for the source and target scale (1:25,000 and 1:50,000 in this case), which then automatically sets the scale reduction factor and size-related parameters. Furthermore, we rasterized the polygon map using a cell-size of 1 meter, the highest possible, to keep the map quality as high as possible. The CA was then run over the rasterized map with a Moore's neighborhood of radius 3 and 2 iterations of final CA generation. Following the CA processing, the raster map was converted back to a vector map, yielding the result shown in Figure 17 (right).

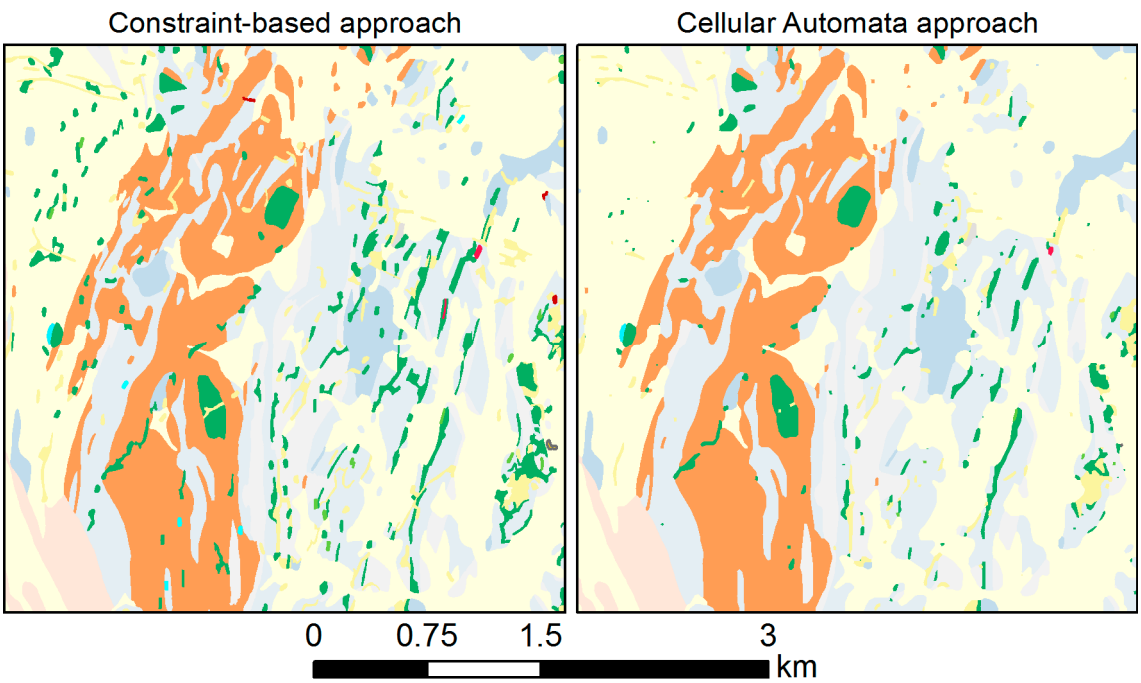


Figure 17. Comparison of two approaches: Constraint-based approach (left) and the cellular automata solution generated in GeoScaler (right).

6. Discussion

This paper presented a methodology for geological map generalization using a constraint-based approach. The main objective of this paper was to demonstrate the usefulness of the constraint-based approach in map generalization when applied to geological maps, not to solve geological map generalization comprehensively (more feature classes are addressed in [23,24]). This was done against the background that the constraint-based approach has so far not been explicitly applied to the automated generalization of geological maps. Hence, the proposed methodology focused on a

sub-problem of geological map generalization, though nevertheless an important one. In particular, it focused on size constraints as the primary driving force of maintaining map legibility through map generalization. Furthermore, we focused on the case of small, free-standing polygons (or area patches; [28]), as they represent a frequent case not only in geological maps but also other types of categorical maps such as soil maps or land cover maps.

Based on the experiments reported in Section 5, we can make several general observations:

1. The proposed methodology resolves the main legibility problems associated with small polygons in a step-by-step manner. The resulting map is more readable, and map features remain distinguishable after generalization (Figure 13).
2. Although the goal values of the size constraints are defined globally, each polygon is treated individually to its own, specific properties rather than by a global process such as cellular automata. Hence, by consideration of the semantics of individual polygons, important polygons can be protected by enlargement; by consideration of shape properties, the shape characteristics of the individual polygons are largely maintained.
3. The approach, by being based on testing whether the goal values of constraints are met, allows self-evaluation, and can guarantee that the legibility limits are met—but not more than that. Again, this differs from global approaches such as cellular automata [23,24].
4. Few parameters are required to control the generalization methodology. The process is initially triggered by the MA constraint and further assisted by few additional size constraints (most importantly, the OS constraint). Once the goal values of the size constraints and additional algorithm-specific parameters have been set, the methodology operates automatically, without further human intervention.
5. The constraints' goal values are, first of all, a function of the map legibility at the target scale, and hence allow adapting to the desired scale transition. Furthermore, the goal values also allow for controlling the overall granularity of the output map (Table 4 and Figure 15), depending on the map purpose.
6. Despite the rather low number of constraints and control parameters, the methodology is modular and features several generalization operators, thus achieving considerable flexibility.

Because the proposed methodology builds on the universally valid concept of constraints relating to map legibility, the potential exists to apply the same approach to the generalization of other categorical maps, such as vegetation or soil maps. Furthermore, since the workflow is modular, any of the algorithms used in our methodology could be replaced by another algorithm that is perhaps better suited for the peculiarities of the given generalization problem. Thus, for instance, the aggregation or displacement algorithms used in this paper could be replaced by more elaborate algorithms if so desired.

Using a workflow approach, we showed how the proposed methodology could be implemented using the libraries of a general-purpose GIS, achieving a behavior similar that of an agent-based system, but with less implementation effort and less effort for parameter tuning.

Comparing Figures 10–13, which present the results of the individual generalization operators, it becomes obvious that elimination and aggregation have the most far-reaching effects. Elimination may retain, or alternatively destroy, both the patterns of spatial arrangements of polygons, as well as the balance of area proportions of the different categories appearing on the map. Below, we discuss the available elimination algorithms separately through Test Case 1. Aggregation is the other operator that can lead to significant alteration of the map image, in this case however by area gain and shape modification of the polygons concerned. As Figure 12 suggests, the aggregation algorithm that was implemented in our workflow can sometimes lead to undesirable shape and area changes, and would be better replaced by a typification algorithm [36,37], which however would require preceding detection of group patterns amenable to typification [10].

Elimination is the first but also the most radical step of the methodology, as some polygons are permanently removed from the target map. Selective elimination of small island polygons supported by importance values gives way to preserve polygons that are important relative to others, thus reflecting geological importance or economic value. However, in the elimination process, the connectivity between neighboring polygons is not considered, which may lead to the removal or uncontrolled dissolution of certain group patterns (e.g., clusters, alignments) of polygons. Examples of this effect can be observed comparing Figure 10, left (the original map), and Figure 10, right (the map after elimination).

To address this issue, we ran Test Case 1 on three selection methods (Section 5.6.1): Radical Law selection, area loss–gain selection, and category-wise selection. As Figure 14 and Table 3 show for the scale transition to 1:50,000, area loss–gain selection preserved the number and structure of polygons best. In contrast, Radical Law selection had the most destructive effect on polygon groups. However, when the scale transition is larger, e.g., with a target scale of 1:100,000 or 200,000, more polygons will have to be removed to balance the area gain induced by enlargement. Thus, category-wise selection seems to be the better approach for the elimination operation in greater scale reductions, as it distributes the removal of polygons evenly across each category. In general, Radical Law selection is not recommended for use in categorical map generalization unless more strongly generalized ("overgeneralized") maps are desired. One of the essential requirements of categorical map generalization is to preserve areas as much as possible across scale ranges, which can be achieved using the area loss–gain balance selection method. Based on Test Case 1, however, the category-wise selection method ultimately seems more appropriate as it evens out the elimination of polygons across categories, and also represents a good compromise. Nevertheless, regardless of which of the three selection methods is used, none can safeguard against the inadvertent destruction of group patterns, especially not at greater scale reductions. The actual generalization stage should thus be preceded by a stage devoted to the recognition of essential collective patterns present in the source map [16,18], which led us to develop a process for the recognition of polygon group patterns [10].

In Test Case 2, documented in Figure 15 and Table 4, we compared three different sets of goal values to explore their effect on the granularity of the resulting map. The most fine-grained of the three sets (FG) was originally developed for topographic mapping [30]. As Figure 15 shows, it is however reaching its limits for the generalization of geological maps that are usually populated by tinted, irregularly shaped polygons that often have reduced visual contrast (as is for instance particularly the case with the yellow pegmatite polygons). For such maps, the goal values of the FG set seem too detailed. Thus, in [31] the use of considerably higher goal values for geological maps was proposed, which produces a rather coarse-grained result with more polygons being enlarged and aggregated. As Figure 15 suggests, this Coarse-grained (CG) set of goal values may lead to overgeneralization and excessive loss of detail, particularly when the focus is on small polygons, as is the case in this paper. This is confirmed by the average polygon area and the number of polygons (Table 4), which are markedly different from those of the other two sets of goal values. The Compromise (CM) set of goal values allows for a more differentiated picture to be maintained, while ensuring that legibility is maintained. Note that in this set, the values were chosen to lie closer to those of the FG set than those of the CG set, with the intention of retaining as much detail as possible, thus also allowing a greater range of scale reduction. This is also reflected by the results reported in Table 4 for the average polygon area and the number of polygons, which are very similar to those obtained with the FG set.

In Test Case 3, we used the Compromise goal values and the Category-wise selection strategy to produce a series of generalized maps, shown in Figure 16 at the corresponding target scales. While the resulting map at 1:50,000 overall looks convincing, the map at 1:100,000 starts showing signs of visual imbalance. At the scale of 1:200,000, the map breaks down completely. The visual impression is supported by Table 5, which shows that the relative area per class becomes heavily imbalanced after 1:100,000. In the original map, the Amphibolite and Pegmatite categories are the most frequent ones regarding the number of polygons, which is due to the fact that they mostly occur as small polygons.

Hence, they are candidates for removal due to their generally small size, and their number decreases significantly in the transition from 1:100,000 and 1:200,000 (Table 5). However, as they are so numerous, these two categories are also most affected by enlargement and particularly by aggregation; hence, the area gained in the generalization process is disproportional. Again, this effect suggests that beyond a scale of 1:100,000 the results are no longer pleasing. Once again, this points to the necessity of an approach that is based on the recognition of local group structures which can then inform improved contextual generalization operators [10].

Figure 17 compares the proposed constraint-based methodology and the CA approach of GeoScaler [23,24], and both methods indeed yielded comparable results in that the legibility was improved. However, the CA approach, as it is based on a moving-window operator that uses the same, essentially majority-based principle across the entire map, has a tendency towards dilating larger polygons while eroding small polygons. Hence, more small polygons disappeared than in the result of the proposed methodology, and some polygons became hardly visible. GeoScaler includes post-processing operations, such as enlargement of too small polygons, which can be applied to ensure the legibility of all polygons. However, even that postprocessing operation will not bring back those small polygons that have been removed. Thus, overall the proposed constraint-based methodology seems to outperform the CA-based approach regarding the adequate maintenance of small polygons. Additionally, due to the raster-based majority filtering of the CA approach, characteristic polygon shapes also become more uniform than in our proposed methodology.

We have already mentioned the lack of explicit polygon group recognition in the elimination operator as the first shortcoming of the proposed methodology. Here, we see the second weakness: because the methodology focuses on small polygons and favors small polygons of important geological units by enlarging them, those small polygons grow disproportionately when the scale reduction factor is large. Once again, this points to the necessity of treating groups of polygons rather than individual polygons. While the methodology of this paper includes contextual generalization operators (aggregation, displacement), "context" is merely understood as the immediate, first-order neighborhood defined by the OS constraint. If polygons are separated by a distance slightly exceeding the OS limit, no link is detected. Likewise, there is no facility for detecting higher-order neighbors and hence forming larger, contiguous groups of polygons. Hence, for a full treatment of the small polygons forming the focus of this paper, group pattern detection procedures are needed, as well as contextual generalization operators that can make use of such group patterns. In related work [10], we therefore propose a methodology that considers proximity as well as geometrical similarities (shape, size, and orientation of map features) and attribute similarities to find, refine, and form groups that can be used to subsequently inform contextual generalization operators, such as aggregation and typification, to overcome persistent limitations in generalization approaches such as those shown in this paper.

7. Conclusions

Taking the example of size constraints (minimum area, object separation, etc.), this paper demonstrated how constraint-based map generalization can be realized for the case of small island polygons, which are mainly dominated by size constraints. Size constraints are rather simple but very fundamental; everything else depends on them. Thus, starting with the most fundamental constraint—the minimum area constraint—we were able to demonstrate how a constraint-based approach can be built, triggering different generalization operators depending on the local analysis of the individual situation found. From a practical perspective, we were able to show how such an approach can be implemented using the libraries of a general-purpose GIS, achieving a behavior similar to that of an agent-based system but with less implementation effort and less effort for parameter tuning.

As our experiments also showed, an approach focusing solely on size constraints and local analysis and operations is not sufficient if greater scale reductions are to be achieved. Thus, future research will have to explore the detection of groups of polygons defined by proximity and similarity (in

attribute, size, shape, orientation, etc.), as well as contextual generalization operators for aggregation and typification that can make use of those group patterns.

Author Contributions: Conceptualization, Azimjon Sayidov and Robert Weibel; Data curation, Azimjon Sayidov; Formal analysis, Azimjon Sayidov; Investigation, Azimjon Sayidov; Methodology, Azimjon Sayidov, Meysam Aliakbarian and Robert Weibel; Project administration, Azimjon Sayidov and Robert Weibel; Software, Azimjon Sayidov and Meysam Aliakbarian; Supervision, Robert Weibel; Validation, Azimjon Sayidov; Visualization, Azimjon Sayidov and Meysam Aliakbarian; Writing—original draft, Azimjon Sayidov; Writing—review and editing, Azimjon Sayidov and Robert Weibel. All authors have read and agreed to the published version of the manuscript.

Funding: The Ph.D. research was supported by a scholarship of the Swiss Federal Commission for Scholarships (FCS) under the Grant 2015.0750.

Acknowledgments: We are grateful for the support provided to the first author by the Swiss Federal Commission for Scholarships (FCS). The authors would like to acknowledge their use of the "Euriowie (including part of Campbells Creek)" 1:25,000 geological map provided by the Geological Survey of New South Wales, Department of Mineral Resources, NSW, Australia, which formed the basis for the experiments reported in this paper.

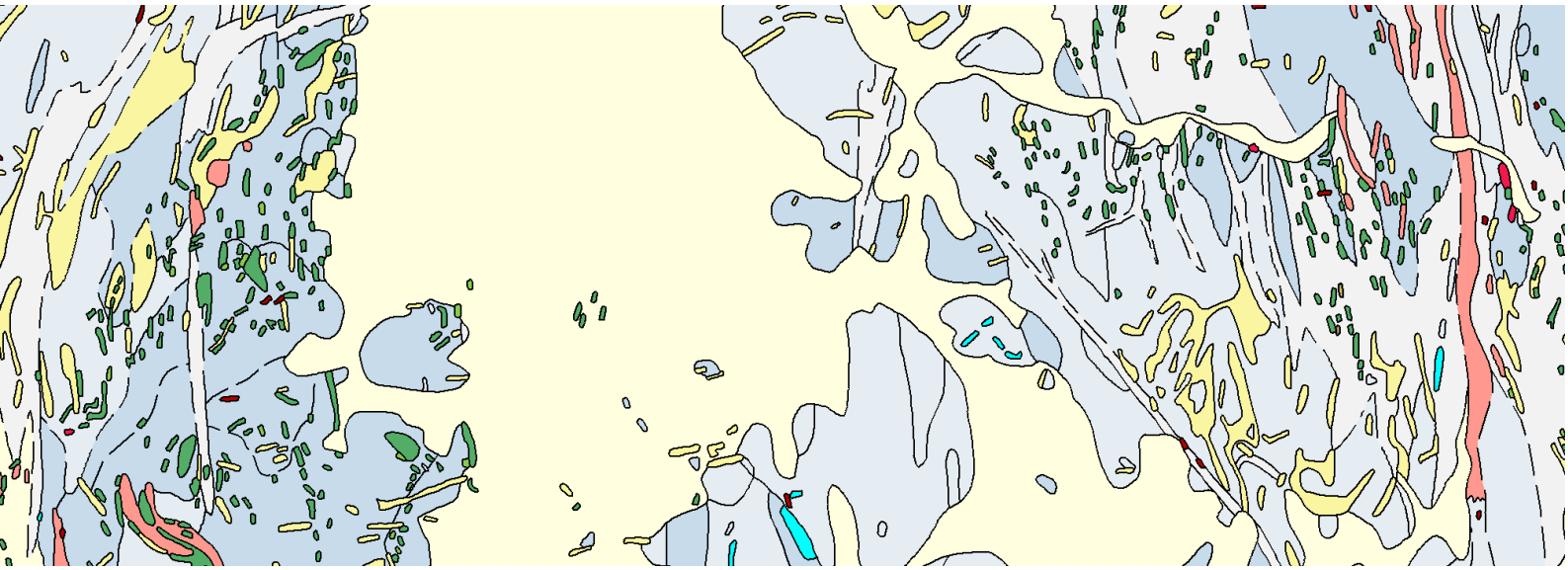
Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sayidov, A.; Weibel, R. Constraint-based approach in geological map generalization. In Proceedings of the 19th ICA Workshop on Generalisation and Multiple Representation, Helsinki, Finland, 14 June 2016.
2. Sayidov, A.; Weibel, R. Automating geological mapping: A constraint-based approach. In Proceedings of the CEUR (Central Europe) Workshop Proceedings, Leeds, UK, 30 October–2 November 2017.
3. Ruas, A.; Duchêne, C. A prototype generalisation system based on the multi-agent system paradigm. In *Generalisation of Geographic Information: Cartographic Modelling and Applications*; Elsevier: Amsterdam, The Netherlands, 2007; pp. 269–283.
4. Weibel, R.; Dutton, G.H. Constraint-based automated map generalization. In *8th International Symposium on Spatial Data Handling*; Taylor & Francis: Columbia, UK, 1998; pp. 214–224.
5. Barrault, M.; Regnauld, N.; Duchene, C.; Haire, K.; Baeijs, C.; Demazeau, Y.; Hardy, P.; Mackaness, W.; Ruas, A.; Weibel, R. Integrating multi-agent, object-oriented and algorithmic techniques for improved automated map generalization. In Proceedings of the 20th International Cartographic Conference, Beijing, China, 6–10 August 2001; Volume 3, pp. 2110–2116.
6. Harrie, L.; Weibel, R. Modelling the overall process of generalisation. In *Generalisation of Geographic Information: Cartographic Modelling and Applications*; Elsevier: Oxford, UK, 2007; pp. 67–87.
7. Stoter, J.; Burghardt, D.; Duchêne, C.; Baella, B.; Bakker, N.; Blok, C.; Pla, M.; Regnauld, N.; Touya, G.; Schmid, S. Methodology for evaluating automated map generalization in commercial software. *Comput. Environ. Urban Syst.* **2009**, *33*, 311–324. [CrossRef]
8. Galanda, M. Automated Polygon Generalization in a Multi-Agent System. Ph.D. Thesis, University of Zurich, Zurich, Switzerland, 2003.
9. Ruas, A. *Modèle de Généralisation de Données Géographiques à Base de Contraintes et D'autonomie*; Université de Marne-la-Vallée: Champs-sur-Marne, France, 1999.
10. Sayidov, A.; Weibel, R.; Leyk, S. Recognition of group patterns in geological maps by building similarity networks. *Geocarto Int.* **2020**, 1–20. [CrossRef]
11. Juliev, M.; Mergili, M.; Mondal, I.; Nurtaev, B.; Pulatov, A.; Hübl, J. Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan. *Sci. Total Environ.* **2019**, *653*, 801–814. [CrossRef] [PubMed]
12. Maine.gov. Maine Geological Survey: Reading Detailed Bedrock Geology Maps. Available online: https://www.maine.gov/dacf/mgs/pubs/mapuse/bedrock/bed-read.htm (accessed on 18 July 2019).
13. Jirsa, M.A.; Boerboom, T.J. Components of Geologic Maps. Available online: http://docslide.us/documents/components-of-geologic-maps-by-mark-a-jirsa-and-terrence-j-boerboom-2003.html (accessed on 14 January 2020).
14. Dawes, R.L.; Dawes, C.D. *Basics-Geologic Structures*; Geology of the Pacific Northwest: Wenatchee, WA, USA, 2013.
15. Graymer, R.W.; Moring, B.C.; Saucedo, G.J.; Wentworth, C.M.; Brabb, E.E.; Knudsen, K.L. *Geologic Map of the San Francisco Bay Region*; US Department of the Interior, US Geological Survey: Reston, VA, USA, 2006.

16. Steiniger, S.; Weibel, R. Relations among map objects in cartographic generalization. *Cartogr. Geogr. Inf. Sci.* **2007**, *34*, 175–197. [CrossRef]
17. Downs, T.C.; Mackaness, W. An integrated approach to the generalization of geological maps. *Cart. J.* **2002**, *39*, 137–152. [CrossRef]
18. Brassel, K.E.; Weibel, R. A review and conceptual framework of automated map generalization. *Int. J. Geogr. Inf. Syst.* **1988**, *2*, 229–244. [CrossRef]
19. Mackaness, W.A.; Edwards, G. The importance of modelling pattern and structure in automated map generalization. In Proceedings of the Joint ISPRS/ICA Workshop on Multi-scale Representations of Spatial Data, Ottawa, ON, Canada, 7–8 July 2002.
20. Steiniger, S.; Weibel, R. A conceptual framework for automated generalization and its application to geologic and soil maps. In Proceedings of the 22nd International Cartographic Conference, La Coruña, Spain, 11–16 July 2005.
21. Shea, K.S.; McMaster, R.B. *Cartographic Generalization in a Digital Environment: When and How to Generalize*; AUTO-CARTO 9: Baltimore, MD, USA, 1989.
22. McCabe, C.A. Vector Approaches to Generalizing Faults and Polygons in Santa Rosa, California, Case Study. 2007. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.643.17&rep=rep1&type=pdf. (accessed on 14 January 2020).
23. Smirnoff, A.; Paradis, S.J.; Boivin, R. Generalizing surficial geological maps for scale change: ArcGIS tools vs. cellular automata model. *Comput. Geosci.* **2008**, *34*, 1550–1568. [CrossRef]
24. Smirnoff, A.; Huot-Vézina, G.; Paradis, S.J.; Boivin, R. Generalizing geological maps with the GeoScaler software: The case study approach. *Comput. Geosci.* **2012**, *40*, 66–86. [CrossRef]
25. Bader, M.; Weibel, R. Detecting and resolving size and proximity conflicts in the generalization of polygonal maps. In Proceedings of the 18th International Cartographic Conference, Stockholm, Sweden, 23–27 June 1997; Volume 23, p. 27.
26. Galanda, M.; Weibel, R. An agent-based framework for polygonal subdivision generalization. In *Advances in Spatial Data Handling, 10th International Symposium on Spatial Data Handling*; Richardson, D., van Oosterom, P., Eds.; Springer: Berlin, Germany, 2002; pp. 121–136.
27. Edwardes, A.; Mackaness, W.A. Modelling knowledge for automated generalisation of categorical maps—A constraint based approach. In *Innovations in GIS 7. GIS and Geocomputation*; Atkinson, P., Martin, D., Eds.; Taylor & Francis: London, UK, 2000; pp. 161–174.
28. Müller, J.C.; Wang, Z. Area-patch generalisation: A competitive approach. *Cartogr. J.* **1992**, *29*, 137–144. [CrossRef]
29. Ruas, A.; Plazanet, C. Strategies for automated generalization. In *Advances in GIS Research II, Proceedings of the 7th International Symposium on Spatial Data Handling*; Kraak, M.J., Molenaar, M., Eds.; Taylor & Francis: London, UK, 1996; pp. 319–336.
30. Spiess, E.; Baumgartner, U.; Arn, S.; Vez, C. Topographic maps map graphics and generalisation. *Cartogr. Publ. Ser.* **2005**, *17*, 121.
31. Subcommittee, G.D. *U.S. Geological Survey FGDC Digital Cartographic Standard for Geologic Map Symbolization (FGDC Document Number FGDC-STD-013-2006)*; U.S. Geological Survey for the Federal Geographic Data Committee: Reston, VA, USA, 2006.
32. Lichtner, W. Computer-assisted processes of cartographic generalization in topographic maps. *Geo Process.* **1979**, *1*, 183–199.
33. Töpfer, F.; Pillewizer, W. The principles of selection. *Cartogr. J.* **1966**, *3*, 10–16. [CrossRef]
34. Osserman, B.Y.R. The isoperimetric inequality. *Bull. Amer. Math. Soc.* **1978**, *6*, 1182–1238. [CrossRef]
35. Touya, G. A road network selection process based on data enrichment and structure detection. *Trans. GIS* **2010**, *14*, 595–614. [CrossRef]
36. Regnauld, N. Contextual building typification in automated map generalization. *Algorithmica* **2001**, *30*, 312–333. [CrossRef]
37. Burghardt, D.; Cecconi, A. Mesh simplification for building typification. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 283–298. [CrossRef]
38. Moreira, A.; Santos, M.Y. Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points. In Proceedings of the Second International Conference on Computer Graphics Theory and Applications, Algarve, Portugal, 5–7 March 2011; pp. 61–68.

39. Galanda, M.; Weibel, R. Using an energy minimization technique for polygon generalization. *Cartogr. Geogr. Inf. Sci.* **2003**, *30*, 263–279. [CrossRef]
40. Bader, M.; Barrault, M.; Weibel, R. Building displacement over a ductile truss. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 915–936. [CrossRef]
41. Mackaness, W.; Purves, R. Automated displacement for large numbers of discrete map objects. *Algorithmica* **2001**, *30*, 302–311. [CrossRef]
42. Ai, T.; Zhang, X.; Zhou, Q.; Yang, M. A vector field model to handle the displacement of multiple conflicts in building generalization. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 1310–1331. [CrossRef]
43. Basaraner, M. A zone-based iterative building displacement method through the collective use of Voronoi Tessellation, spatial analysis and multicriteria decision making. *Bol. Ciências Geodésicas* **2011**, *17*, 161–187. [CrossRef]
44. Barnes, R.G. *Metallogenic Studies of the Broken Hill and Euriowie Blocks. 1. Styles of Mineralisation in the Broken Hill Block. 2. Mineral Deposits of the Southwestern Broken Hill Block*; Geological Survey of New South Wales: New South Wales, Australia, 1988.
45. Peljo, M. *Broken Hill Exploration Initiative: Abstracts of Papers Presented at the July 2003 Conference in Broken Hill*; Geoscience Australia, Ed.; Geoscience Australia: Canberra, Australia, 2003; ISBN 978-0-642-46771-3.
46. Wilson, J.R. *Minerals and Rocks*; BookBoon: London, UK, 2010.
47. Huot-Vézina, G.; Boivin, R.; Smirnoff, A.; Paradis, S.J. GeoScaler: Generalization tool (with a supplementary user guide in French). *Geol. Surv. Can.* **2012**, *6231*, 82. [CrossRef]



3

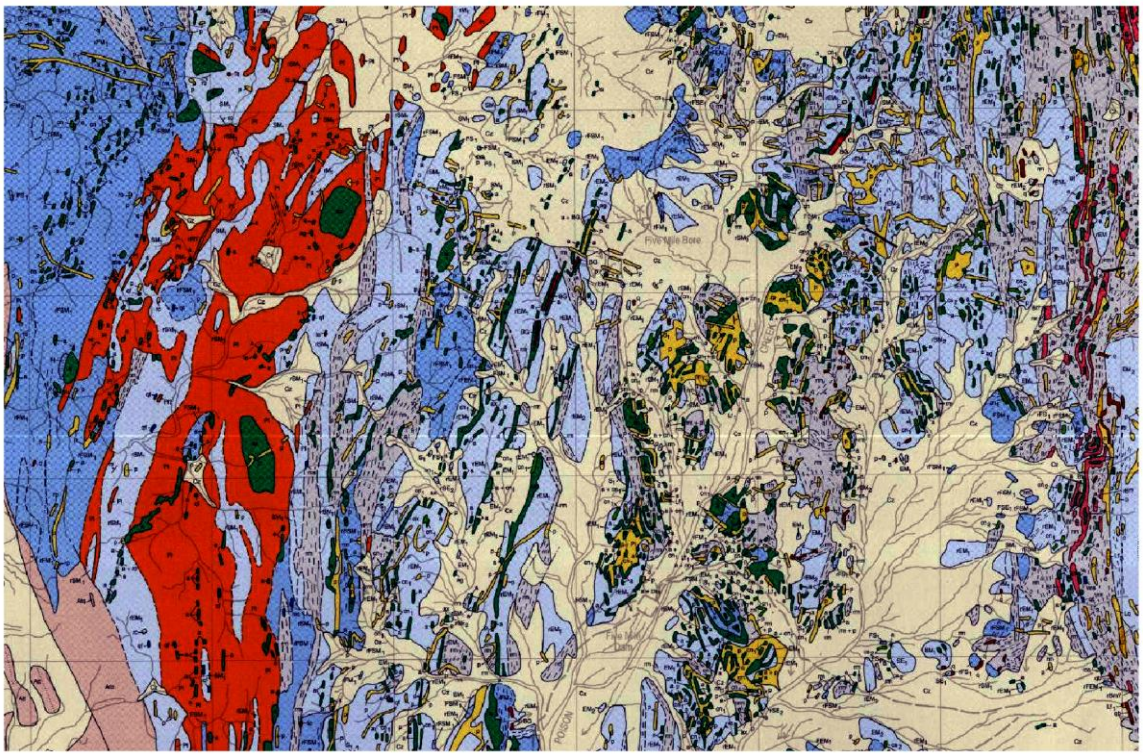
Chapter

For in and out, above, about, below,
'Tis nothing but a Magic Shadow-show,
Played in a Box whose Candle is the Sun,
Round which we Phantom Figures come and go.

Omar Khayyam

Research Paper 2

Sayidov, A., Weibel, R., & Leyk S. (2020). Recognition of group patterns in geological maps by building similarity networks. Geocarto International. https://doi.org/10.1080/10106049.2020.1730449.



Part of the Geological map of the Euriowie Inlier (scanned from a paper map – Stevens et al., 2008)

Contribution of the Ph.D. candidate:

The Ph.D. candidate designed the conceptual model to recognize and identify groups of polygons in geological maps; managed and processed the data, wrote the codes for network analysis and group identification in Python; and drafted the article.

As part of the research reported in this paper, an experiment was conducted to gain insight into how participants perceptually organize polygons into groups that they consider meaningful. While this experiment was not reported in the paper, Appendix A of thesis summarizes it.

Authors' Contributions: Conceptualization, Data analysis, Visualization and Validation, Azimjon Sayidov; Methodology, Azimjon Sayidov, Robert Weibel and Stefan Leyk; Software, Azimjon Sayidov; Writing original draft, Azimjon Sayidov; Writing review and editing, Azimjon Sayidov, Robert Weibel and Stefan Leyk.

GEOCARTO INTERNATIONAL
https://doi.org/10.1080/10106049.2020.1730449

Taylor & Francis
Taylor & Francis Group

Recognition of group patterns in geological maps by building similarity networks

A. Sayidov[a], R. Weibel[a] and S. Leyk[b]

[a]Department of Geography, University of Zurich, Zurich, Switzerland; [b]Department of Geography, University of Colorado Boulder, Boulder, CO, USA

ABSTRACT
The recognition of structures is fundamental to map generalization, furnishing structural information that assists in choosing and parameterizing generalization operators. We specifically focus on the process of recognizing groups of small polygons in geological maps as a prerequisite to subsequent aggregation or typification operators. Proximity between polygons represents an essential criterion in identifying neighboring map objects. Here, network-based analysis is used for effective definition and refinement of candidate group members, applying criteria such as the distance between polygons, polygon size, shape, orientation, and feature attributes such as rock type. Starting off from the Delaunay triangulation of the polygon centroids, the global and local long edges, which initially define the network, are removed. The modified network is loaded with additional criteria, and edges are kept or removed based on the local similarities of the polygons they connect. This approach leads to more homogeneous, meaningful groups of polygon features in geological maps.

ARTICLE HISTORY
Received 26 July 2019
Accepted 26 January 2020

KEYWORDS
Cartographic pattern detection; map generalization; geological mapping; graph-based clustering; similarity network

1. Introduction

The automation of the generalization of categorical maps has been studied for more than two decades by researchers in cartography and geographic information science (Schylberg 1993; Su et al. 1997; 1998; Peter and Weibel 1999; Downs and Mackaness 2002; Galanda 2003; Steiniger and Weibel 2005; McCabe 2008; Smirnoff et al. 2008; 2012). In particular, geological maps, as a representative of categorical maps, have been the focus of several studies (Peter and Weibel 1999; Downs and Mackaness 2002; Steiniger and Weibel 2005; McCabe 2008; Smirnoff et al. 2008; 2012), but research activity in this domain on both vector and raster data has been limited in recent years. Raster-based methods, which are considered fast and more straightforward, include techniques such as region growing/shrinking filters (Schylberg 1993; Su et al. 1997; 1998) or grid cell-based cellular automata (Smirnoff et al. 2008; 2012). However, despite those advantages, raster-based approaches tend to excessively distort the shapes of map features, due to a lack of adaptivity to local shape and structure variations depending on the underlying data resolution. In addition, raster data can represent only one attribute per raster layer that is used as a raster value resulting in lower degrees of flexibility for attribute-related data processing.

CONTACT A. Sayidov azimjon.sayidov@geo.uzh.ch; peterazim@gmail.com

Conversely, methods for generalization of geological maps in vector form have not yet been developed to the same level of sophistication as their counterparts for topographic maps, which often do not apply to the peculiarities of geological maps (Galanda and Weibel, 2002; Galanda, 2003). Overall, while promising, there is a lack of a holistic solution to geological map generalization using vector data (McCabe 2008). Geological maps consist of diverse patterns formed by polygons, which can be described by spatial, structural and semantic criteria to evaluate similarities or differences. A spatial map pattern can be described as a perceptual structure, or an arrangement of features on the map (Steiniger and Weibel, 2007). Depending on the scale of the target map, original patterns may be trimmed, merged or abstracted through the application of a range of generalization operators (McMaster and Shea, 1992). Map generalization focuses on retaining and emphasizing the essential structures and patterns of the source map, while suppressing the unimportant map features, thus preserving the main characteristics and the main message of the source map. As has already been established in the early literature on map generalization (Brassel and Weibel, 1988; McMaster and Shea, 1992), the effectiveness of the generalization process largely depends on how well existing structures and patterns can be identified in the source map. This stage of identifying structures and patterns, which has been called 'structure recognition' (Brassel and Weibel, 1988) and 'cartometric evaluation' (McMaster and Shea, 1992), is an essential prerequisite to the application of map generalization operators. All modern computational processes of map generalization include some form of pattern recognition and evaluation of the source map to then trigger and inform the appropriate generalization actions (Harrie and Weibel, 2007). In the case of geological map generalization, the pattern recognition stage relates to the identification of structures and patterns apparent among polygon features in the source map. An essential, characteristic kind of pattern in geological maps are groups of small polygons, compositions that are essential for the understanding of geological formations, and often will have to be subjected to aggregation and/or typification operations during the process of map generalization.

This paper focuses on developing a methodology that effectively recognizes important group patterns in geological maps using a network-based approach. The methodology considers proximity, as well as geometrical similarities (shape, size, and orientation of map features) and attribute similarities (rock types in the geological map) in order to find, refine and form groups that can be used to subsequently inform contextual generalization operators in order to overcome persistent limitations in existing generalization approaches.

The proposed methodology aims at small polygons, as these form the vast majority on most large-scale geological maps and need to be treated once they fall below the perceptual limit of the human visual system. Generalization of large polygons commonly only requires simplification and smoothing operators as well as local displacement (Galanda and Weibel 2003). Small polygons, however, require more complex processing and a rich repertoire of generalization operators, including selection, enlargement, aggregation, displacement, and typification, all of which require the analysis of the local spatial context, which is delivered by the proposed group detection methodology.

2. Literature review

An important measure to detect patterns of polygon groups in a map is the distance between polygons in order to determine immediate neighbours that are located (too) close to each other in the source map and thus may be candidates for aggregation or

typification operations. To date, two types of measures, geometrical and contextual, have been used to recognize groups of polygons in cartographic generalization (Bobzien et al. 2008; Burghardt and Schmid 2010).

Geometrical measures, such as size constraints including the minimum area or the minimum distance between polygons, can be useful in identifying a conflict between features and trigger appropriate generalization actions (Anderson-Tarver et al. 2011; Anonymous 2016).

In contextual cartographic generalization (Cetinkaya et al. 2015), the assumption is that objects in close proximity have a greater potential of graphic conflict (e.g., the violation of the given minimum distance constraint) at the target scale, for example, due to feature enlargement. As a consequence, such features should be treated as a group to solve this conflict (Basaraner and Selcuk 2008; Yan et al. 2008). In addition, features that are close to each other on the map are likely related to each other and may belong to the same category. Thus, an intuitive strategy is to use the object separability distance (i.e., the minimum distance required between two cartographic objects to remain visually separable) to inform the process of pattern recognition with an analysis of the proximity of features of the same and different categories.

Proximity analysis with the purpose of grouping has often been realized by graph-based clustering or grouping, techniques widely used in spatial data mining, aiming to classify spatial data into clusters of arbitrary shapes that do not require previous knowledge of the data. Graph-based grouping techniques have, for instance, been used in the grouping of buildings in urban blocks (Zhang et al. 2010; Deng et al. 2011; Zhang et al. 2013; Cetinkaya et al. 2015; Wang and Burghardt 2017; He et al. 2018). Regnauld (2001) uses Minimum Spanning Tree (MST) techniques to generate groups from centroids of buildings and remove edges based on proximity as well as homogeneity measures and the number of buildings in a group. Li et al. (2004) create a Delaunay triangulation (DT) using the vertices of buildings. Buildings whose vertices belong to the same triangle are labelled as 'neighbour'. Direct alignments between neighbouring buildings are found by employing Gestalt theory (Wertheimer 1938; Steiniger and Weibel 2007). Zhang et al. (2013) produce an MST from a constrained DT of building centroids to detect building clusters, similar to Regnauld (2001), followed by the analysis and recognition of building alignments, which is the actual purpose of their article.

Similarly, Anders et al. (1999) apply graph-based clustering techniques for the automated analysis of settlement structures represented by building centroids. Anders (2003) demonstrates the successful application of different types of proximity graphs in finding 'natural' object groups, without the need for defining any parameter. Deng et al. (2011) propose an adaptive spatial clustering approach by removing global and local long edges from the DT, which, based on various global and local criteria, discovers patterns in the network. This approach is based on the proximity analysis of neighbouring network points. However, attribute properties, such as size, orientation, shape, and category of the map feature are not considered.

Note that in the literature, both the terms 'graph' and 'network' are being used to denote the same concept; we thus use these terms interchangeably in this paper. While 'graph' is the proper mathematical term, 'network' is often used in building generalization (i.e., in the work discussed above), but has not been used in geological map generalization, as no network-based approaches exist to date.

In summary, proximity-based grouping methods have been rather widely used in the recognition of features in topographic mapping, and in particular, in the recognition of building patterns and groups, but have not been tested in the generalization of geological

maps or in the recognition of patterns in categorical maps in general. We thus see a need for developing a methodology for recognizing group patterns that is tailored to the peculiarities of categorical, polygonal map data. For instance, shapes of polygons in a geological map are typically less compact than building shapes and also exhibit a greater degree of shape variation, which poses challenges to the formation of graphs, the analysis of the proximity of neighbouring polygons, as well as similarity assessment of different polygons. In the next section, we define the problem that we are trying to solve in more detail.

3. Problem statement

3.1. Patterns and pattern descriptors in geological maps

While geological maps contain polygon features of unique shapes and structures, these features often form patterns such that individual, similar features can be grouped and spatially aggregated (Mark and Csillag 1989; Galanda, 2003). Such patterns can indicate underlying geological formations and structures that are physically (though not necessarily visibly at the surface) linked with each other (Barnes 1988). Thus, automating the generalization process of such linked features depends on the effective detection of complex polygon feature groups using criteria of geometric similarity and semantic resemblance. Once such groupings have been established, generalization operators such as aggregation, typification, and simplification can be applied, effectively.

3.2. Why group recognition is needed

The generalization process is guided by the readability and distinguishability of map features. Thus, representative constraints, minimum area, object separation, etc., trigger individual operators to solve existing conflicts. However, as all the features on the map cannot be kept throughout the generalization process, a certain proportion of them has to be removed to accommodate visually sound generalization outcomes. The Radical Law (Töpfer and Pillewizer 1966) states that the number of features on the map will vary as a function of geometrical scale progression. Although the Radical Law gives the approximate amount of detail to keep in the target map, it does not highlight which features to keep. Following the logic of removing small, thus less important features may work in some instances but will not produce the desired generalization outcome if groups and compositions of features are crucial for the generalization process. Figure 1 shows the removal of polygon features from the geological map based on the Radical Law and removing small, less-important polygons, without identifying groups. We can see that with decreasing scale of the target maps (from 1:50,000 to 1:200,000) there is an increasing loss of structural relations and thus the generalization results fail in providing a representative spatial distribution of the source map. In the inset map on the bottom left of Figure 1 (labelled 'Detail of the original map'), we compare the results of simple, number-based selection to the outcome of the group detection process proposed in this paper to highlight the advantages of forming groups of proximal, similar features that need to be preserved at the target scale. The identified groups are connected by green edges. Once groups have been formed, more informed decisions can be made and contextual generalization operators can be used, without removing all small polygons. For instance, small polygons could survive the generalization process by virtue of forming larger placeholder polygons that retain the shape and spatial arrangement of the original, too small polygons.

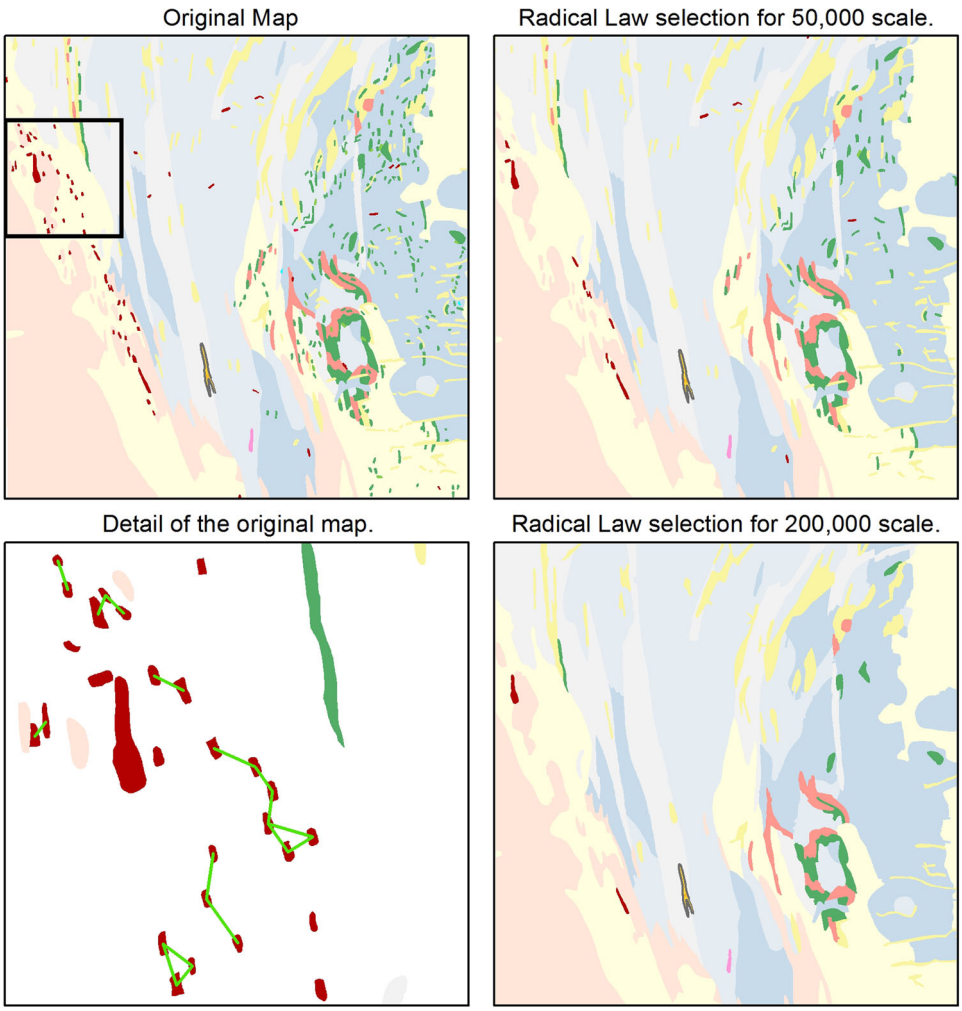


Figure 1. Using the Radical Law as a basis for generalization, implemented in 1:50,000 and 1:200,000. A detail of the original map is shown as an inset map on the bottom left.
© State of New South Wales and Department of Planning and Environment [1998].

4. Methods

Our overall process of recognizing group patterns in geological maps consists of three steps: 1) initial processing to select foreground polygons; 2) defining a network (Delaunay triangulation) of polygon centroids, and forming initial candidate clusters; and 3) refining the network based on similarity and distance criteria for final grouping (Figure 2).

4.1. Data

In this study, we use a geological map of the vicinity of the Wendelpha area in New South Wales (NSW), Australia, provided by the Geological Survey of NSW, Department of Primary Industries-Mineral Resources, available online at https://www.resourcesandgeoscience.nsw.gov.au. This geological map contains immense mineral wealth, especially a massive Pb-Zn-Ag (lead-zinc-silver) deposit (Barnes 1988). In the examples that follow,

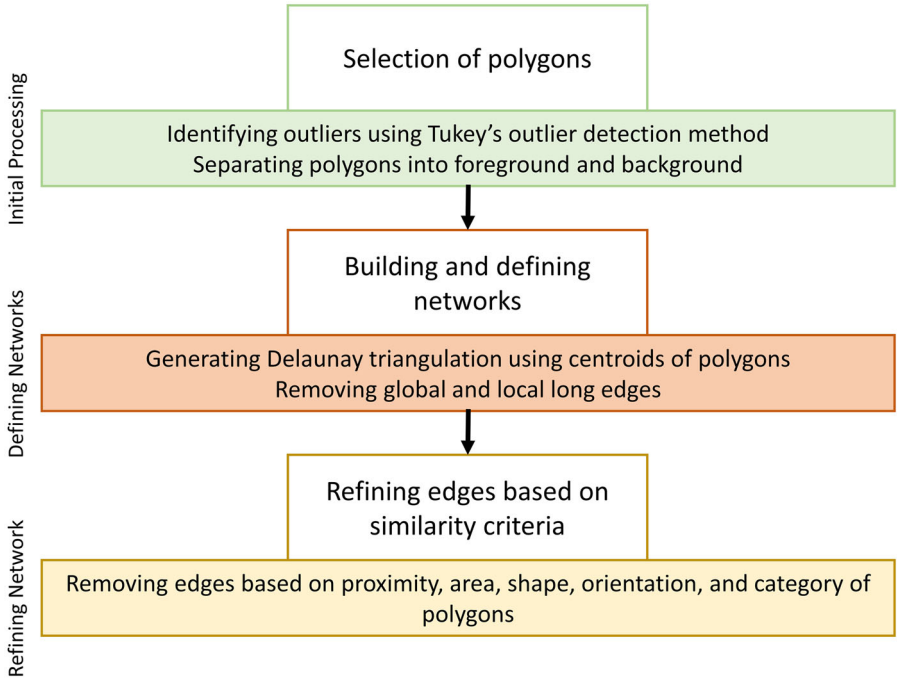


Figure 2. Proposed workflow for the recognition of polygon group patterns in geological maps.

different colours are used to denote different rock types; colour similarity indicates similar rock types. We will, however, not specifically deal with the semantics of the different rock types, as the focus of this paper is on the recognition of group patterns rather than geological mapping.

4.2. Selection of foreground polygons

Geological maps consist of polygons exhibiting a wide range of sizes. In the data set used in this study, areas range from a few m^2 to more than 34 million m^2. The minimum size of a polygon that can be legibly displayed on the target map represents the most essential constraint and receives the highest priority in selecting polygon candidates (Galanda 2003).

The majority of polygons in our geological sample map have small areas, leading to a heavily skewed long-right-tail distribution of size values i.e., a wide range of areal extents with small values, and few polygons with large areal extent. As mentioned, our grouping approach focuses on the smaller polygons and thus needs to detect outliers among all existing polygons.

We use the terms background and foreground polygons to denote the process of separating large polygons from relatively small ones. Background polygons cover vast areas that do not require a considerable amount of effort to generalize them, typically through polishing operators such as simplification and smoothing. Also, as they cover large areas, they typically have no importance for the group identification process as they will always be retained given their dominant size.

Foreground polygons are polygons that tend to cover small areas and compose the groups we are trying to identify. They do require a more complex generalization approach

Table 1. Basic statistics of the areas of the geological map.

Number of polygons in the map	7 019
Mean value	11,127 m^2
Standard Deviation	122,331
Lower Quartile (LQ) – 25 %	565 m^2
Middle Quartile (MQ) – 50 %	1,123 m^2
Upper Quartile (UQ) – 75 %	3,596 m^2
Max value	6,945,050 m^2

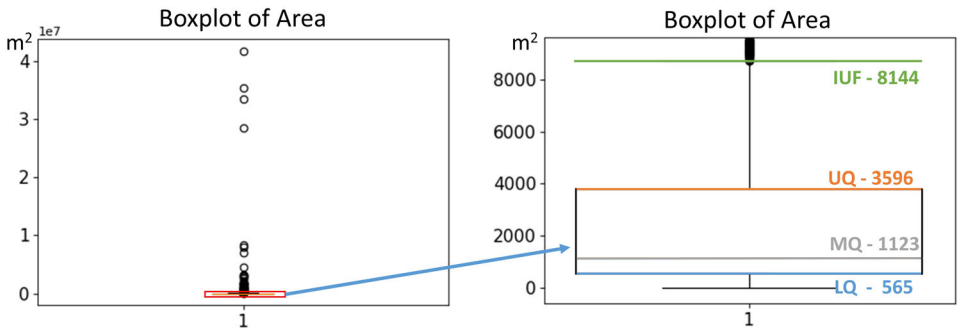


Figure 3. Boxplot of polygon areas in the geological sample map. Left, all the polygons. Right, zoomed detail of the left-skewed distribution.

using contextual operators including selection, enlargement, aggregation, displacement, and typification.

In a first step, we applied Tukey's Outlier Detection test (Seo et al. 2006). This test is independent of the mean and standard deviation and thus not influenced by extreme values in the dataset. Thus, it is useful for defining thresholds that can be used to differentiate between foreground and background features. That is, large polygons are separated from smaller polygons in such a way that a sufficiently large number of smaller foreground polygons is still retained as input for the subsequent group detection process. Many of the foreground polygons are larger than the size (i.e., visibility) threshold that would be used to determine whether or not a particular polygon should be removed at the target scale. However, they are still relatively small and therefore candidates for the grouping process.

In order to identify outliers, we first calculate the inter-quartile range (IQR = UQ-LQ, where UQ = upper quartile, and LQ = lower quartile (IQR = 3,596 – 565 = 3,032 in Table 1). Next, we calculate the inner upper fence (IUF) as UQ + 1.5 * IQR (which is 3,596 + 1.5 * (3,032) = 8,144 in Table 1). Thus, all polygons with an area greater than or equal to 8,144 m^2 are considered outliers (Figure 3). Polygons with an area less than 8,144 m^2 are considered foreground features and selected for further analysis. Approximately 80% of the polygons of the map of interest fell into this category (equal to 5,536 polygons).

Figure 4 illustrates two examples used from the geological map and also highlights three sub-regions, A, B and C, later used as close-up illustrations in Figures 6 and 7.

4.3. Building and defining initial candidate clusters by network pruning

First, the selected foreground polygons are converted to points representing centroids (i.e., the center of gravity of each polygon). A Delaunay triangulation is created using those polygon centroids, which is effective in linking immediate neighbouring polygons

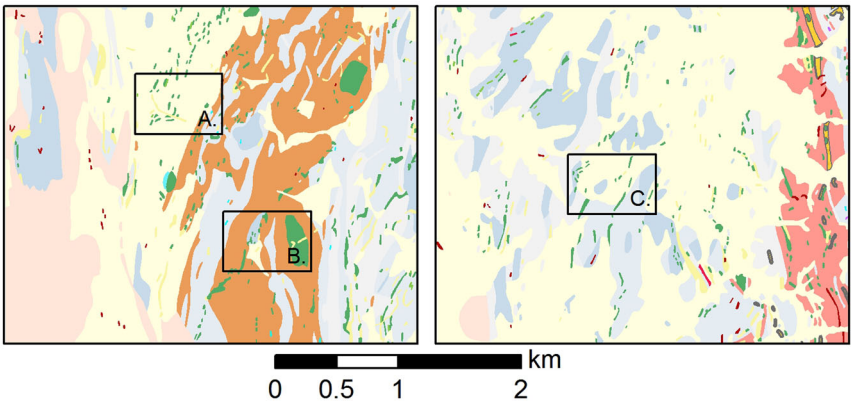


Figure 4. Two subsections of the geological map used in the research.

Refine edges

```
select → edges
    while distance <= min_distance
do
    select edge
     if  category  similar  then
        if  shape  similar  then
            if  orientation  similar  then
                Select Edge
    else
    remove edge
     end
end
```

Figure 5. The algorithm used to refine network edges based on similarity criteria.

and has been widely used in proximity-based analysis (Liu et al. 2008, Lu and Thill, 2008). Once the network has been built, global and local long edges are identified based on the approach described by Deng et al. (2011) and removed from the network. The 'global long edges' are removed based on Equation (1) (Deng et al. 2011):

$$if \begin{cases} edge > GM_{DT} + \dfrac{GM_{DT}}{LM_{DT}} \times SD_{DT} \ \ remove \\ otherwise \ \ keep \end{cases} \tag{1}$$

where GM_{DT}, LM_{DT} and SD_{DT} are the global mean, the local mean, and the standard deviation of the lengths of the edges in the Delaunay triangulation, respectively.

The removal of the global long edges disconnects the points that are linked by long edges, thus leaving candidate features that may be meaningful at the global level.

Next, the 'local long edges' are removed using Equation (2) (Deng et al. 2011):

$$if \begin{cases} edge > LM^2_{DT} + \beta \times SD^2_{DT} \ \ remove \\ otherwise \ \ keep \end{cases} \tag{2}$$

LM^2_{DT} is the mean length of the edges formed by the points that are second-order neighbours of a given point, SD^2_{DT} is the standard deviation of the edge lengths that connect to the same point, and β is used to control the sensitivity of identifying the edges to be removed. The lower the value of β, the more sensitive the cutting value. This approach is

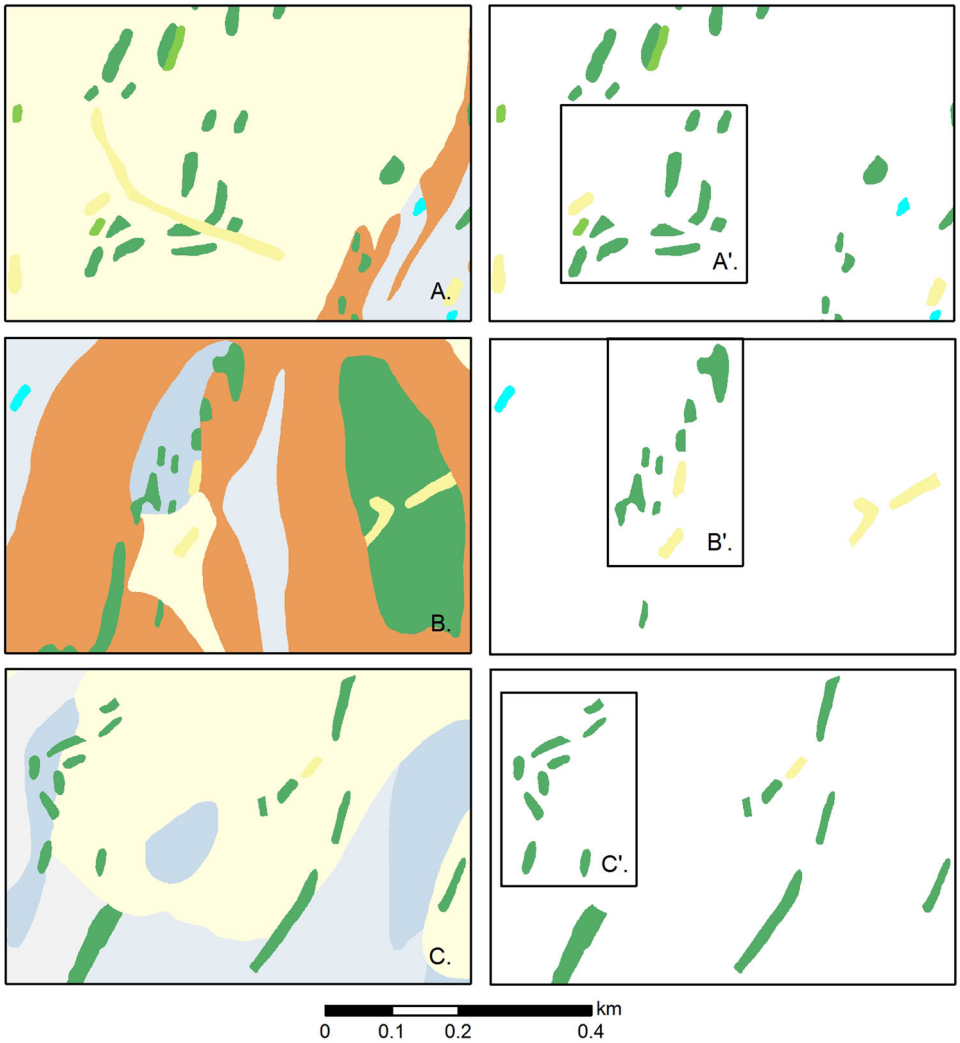


Figure 6. Three sections of the geological map representing the situation before (left) and after (right) removing background polygons from the map. The panels A, B and C are highlighted subsections in Figure 4; A′, B′, C′ are close-ups, further investigated in Figures 8, 9 and 10.

capable of identifying clusters of arbitrary shape and density of features, and allows us to remove clusters connected by 'bridge' edges, distinct edges connecting two separate polygon groups.

4.4. Refining network edges based on similarities

In the next step, the edges of the pruned triangulation are evaluated based on the similarity between the polygons participating in candidate clusters, with regard to their proximity, categories, shape, and orientation. In order to keep the existing edges connected, all these criteria must be satisfied, in a nested, prioritized manner, as shown in Figure 5. Otherwise, the edges are removed from the network.

In the following, we explain the criteria used to determine polygon similarities, one by one.

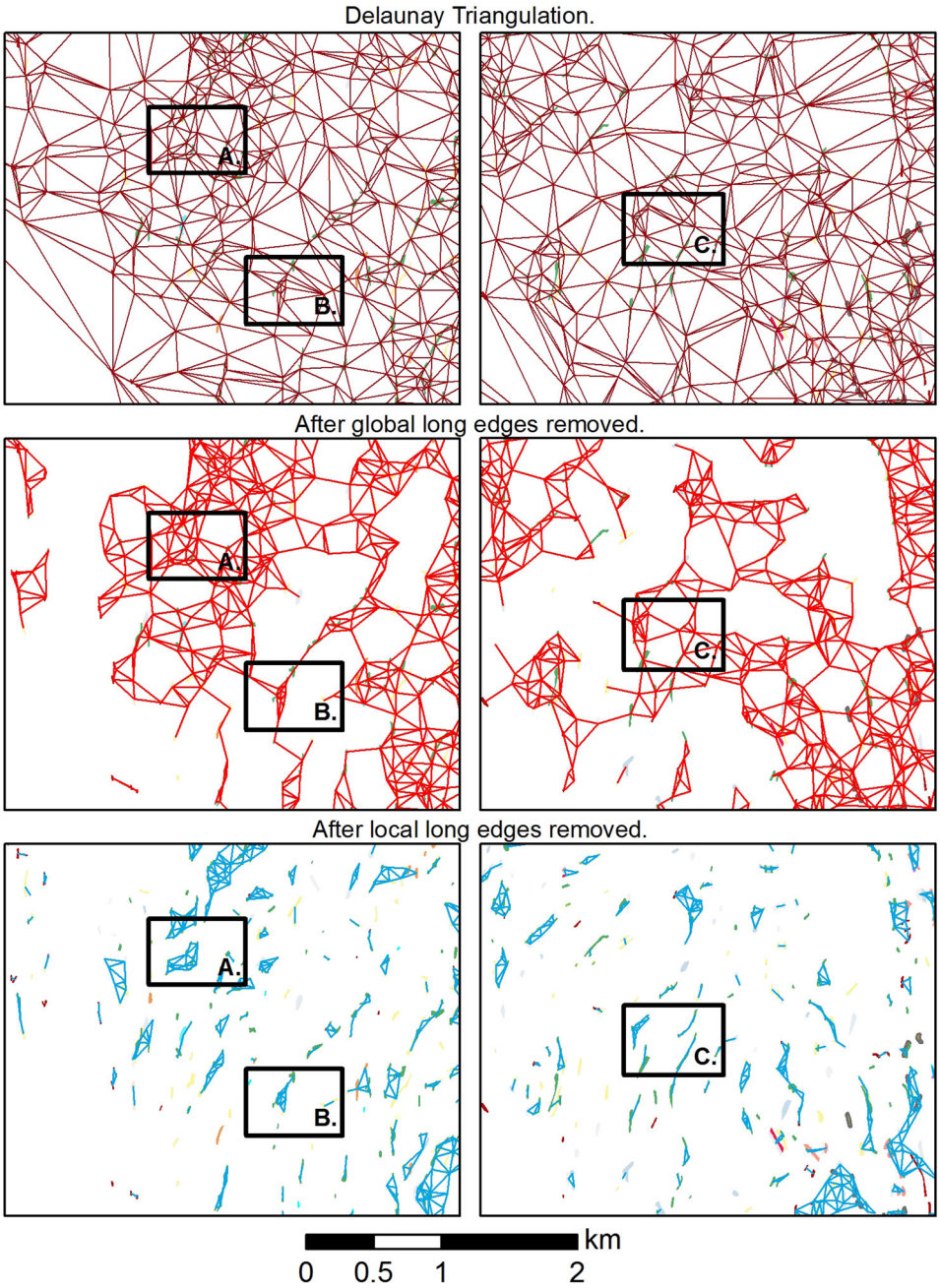


Figure 7. Overview of the partial process to obtain initial candidate clusters. Top: the DT generated from centroids; Middle: the network after removing global long edges; Bottom: after removing local long edges. The subsections A, B and C are highlighted subsections in Figure 4.

4.4.1. Size

Size refers to the areal extent of the polygons represented in the geological map. The process of selecting foreground polygons described in Section 4.2 helps to separate the background from foreground polygons. As stated above, all the polygons that have been identified as foreground polygons are input to the subsequent steps.

4.4.2. Proximity

The algorithm searches for polygons in the proximity of a given polygon feature in order to find immediate neighbors. The minimum separation distance between map objects to be distinguishable visually has been defined to be more than 0.2 mm for objects on topographic maps, with high-quality printing in black on white paper (Spiess, 2002; Spiess et al. 2005). We have chosen a minimum separation distance of 1 mm, which translates to 50 meters on the ground for a map at a scale of 1:50,000. This is a rather generous value but takes into account recommendations to increase the minimum distance for map objects or symbols printed in coloured tints (i.e., with reduced visual contrast), or for potential output to computer screens (Spiess, 2002; Spiess et al. 2005). Also, minimum distances and areas for map symbols in geological maps are generally chosen larger than in topographic maps at the same scale (e.g., Federal Geographic Data Committee 2006). Any feature located less than the minimum separation distance *(min_distance)* from a given polygon (in our case, 50 meters), is considered a candidate for grouping. Figure 5 shows the overall workflow of the algorithm.

4.4.3. Category

The category of a rock formation is represented by a unique CODE. In our example, the geological map has 26 types of major rock types (reclassified from an original fine-grained classification of 221 subcategories). Polygons with the same CODE are assumed to be of the same category and thus can form a group if all other criteria are met.

4.4.4. Shape

Using Equation (3), we calculate the shape properties of each polygon:

$$IPQ = 4 \times \pi \left(\frac{A}{P^2} \right) \tag{3}$$

where IPQ stands for Ipsometric Quotient (Osserman, 1978), A is the area, and P the perimeter of the polygon. This formula returns a normalized value between 0 and 1. Values close to 1 indicate round, compact shapes; values close to 0 indicate elongated polygons. The IPQ values are assigned to 4 equidistant classes ($0.0 - 0.25$; $0.25 - 0.5$; $0.5 - 0.75$; $0.75 - 1.0$). Polygons, whose IPQ values fall into the same class are considered similar.

4.4.5. Orientation

The orientation of a polygon is calculated by creating centrelines of polygons using the skeleton algorithm by (Haunert and Sester, 2008). The most extensive collection of subsequent segments that have similar orientation is considered the primary orientation of the feature. The orientation angles are assigned to 4 equidistant classes ($-90° - -45°$; $-45° - 0°$; $0° - 45°$; $45° - 90°$), and polygons with equal orientation class are considered similar.

4.5. Implementation

The proposed methodology was implemented in the Python 2.7 programming language, using the ArcPy package from Esri, Inc., giving access to the analysis capabilities of the Esri ArcGIS 10.6 toolbox, such as calculating area, proximity, etc. To build, process, and analyse the networks the Python package NetworkX 2.2 was used (https://networkx.github.io).

5. Results

5.1. Selection of polygons

Summaries of the selected and removed (i.e., foreground and background) polygons are given in Table 2. Figure 6 shows different parts of the geological map (highlighted in Figure 4) before and after the removal of background polygons.

78.9% of the polygons are considered foreground polygons, jointly covering only 2.5% of the total area of the sample map. The remaining 21.1% of polygons (i.e., the background polygons) cover 97.5% of the total area.

5.2. Building and defining initial candidate clusters

The trimmed Delaunay Triangulation is shown after removing global (Figure 7) and local (Figure 7) long edges. The three examples shown in Figure 8 illustrate the steps of successfully building and pruning the triangulation based on the edges connecting the centroids of foreground polygons.

This process results in the recognition of initial candidate compositions of polygon features that could be grouped, subsequently.

5.3. Refining network edges based on similarities

Figures 8 and 9 illustrate how the criteria of proximity, the similarity of shape, orientation, and rock type were exploited to identify feature groups. Applying these criteria to the initial candidate clusters found in the preceding step resulted in the final polygon groups. Figure 9 (right-hand column) shows close-up views of three particular examples that will be analysed further in the next section.

5.4. Evaluation

The polygon groups formed by the proposed three-step process (Figure 2) are evaluated by calculating the similarity regarding each criterion, often also referred to as 'homogeneity' (Boffet and Serra 2001; Christophe and Ruas 2002; Ruas and Holzapfel 2003; Zhang et al. 2010; Zhang et al. 2013), which is often based on the standard deviation (STD) of the criteria, or other measures of spread or variation. Here, homogeneity is calculated as the coefficient of variation (CV) to normalize for the different data ranges of the individual criteria, rendering them comparable. It should be noted that this is a purely numerical evaluation that does not include elements of a qualitative evaluation by expert cartographers.

We calculated the similarities among features of each final group of polygons based on proximity, size, orientation, and shape. Figure 10 illustrates three examples of groups (connected by light blue lines in the background) that are subdivided into one or more

Table 2. Number of polygons and area covered by foreground and background polygons, respectively, in the sample geological map.

	Number of polygons	Percent polygons	Area [m^2]	Area [%]
Total polygons in the map	7,019	100.0	274,712,612	100.0
Selected foreground polygons	5,536	78.9	6,982,629	2.5
Removed background polygons	1,483	21.1	267,729,983	97.5

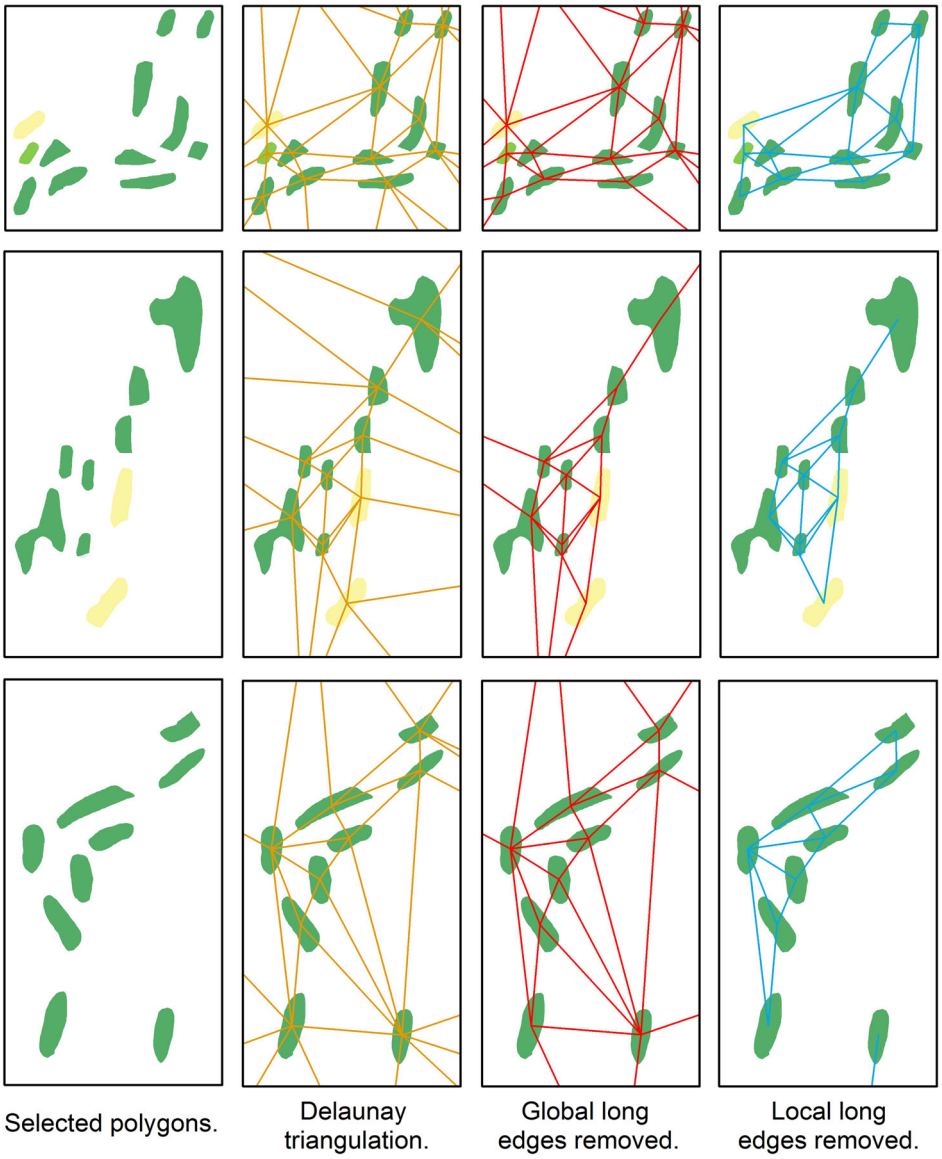


Figure 8. Close-up views of examples showing the effect of the partial process to identify initial candidate groups: Foreground polygons (left) are connected by the triangulated network (second from left); global long edges are removed (second from right); local long edges are removed (right). The different rows show the subsections A′, B′ and C′ highlighted in Figure 6.

subgroups (connected by red lines in the foreground), which are highlighted by the black boxes labelled A1, A2, etc.

Mean, STD, and homogeneity (CV) are calculated for each criterion in each group and summarized in Table 3. The CV, which is based on the mean and STD, characterizes the relative regularity of patterns linked within one group, such that smaller values indicate more features that are regular or similar. For instance, if two groups have the same STD of size, the one with the larger mean has more similarity related to the size criterion (Zhang et al. 2010; 2013).

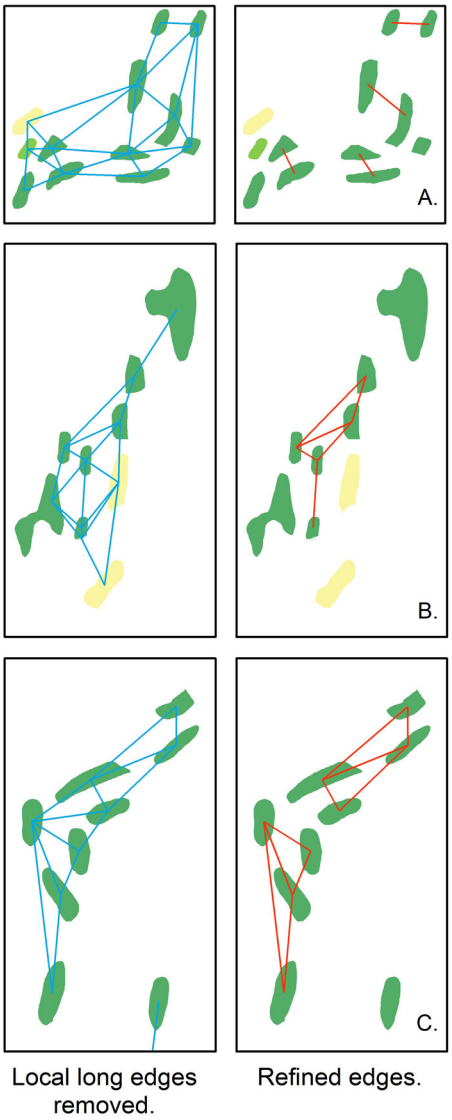


Figure 9. Refining edges based on similarity criteria after local long edges have been removed. The different rows show the subsections A′, B′ and C′ highlighted in Figure 6.

6. Discussion

6.1. Comparison to related work

We have proposed and demonstrated a three-step process for the recognition of groups of small polygons in geological (or more generally, categorical) maps. The proposed process is robust in finding groups, taking into account intrinsic properties of map features, something that has been neglected in prior research on generalization of categorical or geological maps (Schylberg 1993; Su et al. 1997; 1998; Downs and Mackaness 2002; McCabe 2008; Smirnoff et al. 2008; 2012).

Graph-based approaches to detecting group patterns have been used before in cartography, as documented in Section 1 (e.g., Regnauld 2001; Deng et al. 2011; Zhang et al.

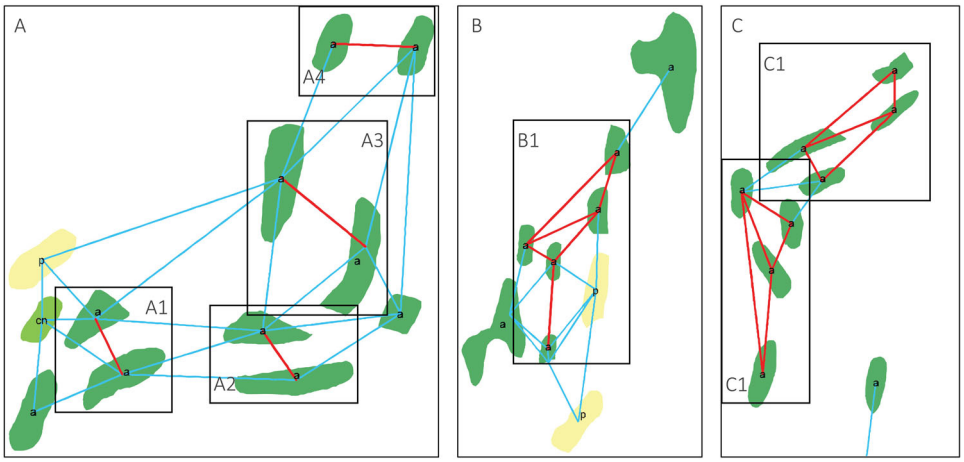


Figure 10. Examples of groups used for the evaluation. The different panels A, B and C show the subsections A′, B′ and C′ highlighted in Figure 6 and denote the initial candidate clusters (light blue lines); boxes A1, A2, A3, A4, B1, C1, and C2 highlight the final groups after refinement (connected by red lines).

2013; Cetinkaya et al. 2015; Wang and Burghardt 2017; He et al. 2018), but those methods focused on recognizing groups and alignments in the buildings feature class of topographic maps, which tend to be rather regularly shaped and sized. We have extended these graph-based methods to accommodate the peculiarities of categorical maps, in particular, geological maps, where a much wider variation of polygon sizes, shapes and compositions is typical. Hence, importantly, the network refinement step of our group recognition process extends graph-based approaches from the topographic domain to include actual polygon proximity rather than proximity of polygon centroids, and it further includes the similarity of polygons in size, shape, and orientation.

6.2. Discussion of results

Looking further into the results, we can observe that generally, the groups that have been generated are meaningful in light of subsequent contextual generalization operations. For instance, in the case of Figure 9A, small groups of two polygons have been formed. Typically, these polygon pairs would be aggregated to a placeholder polygon that takes the combined area, same shape and orientation, and median position of the two original polygons. In this way, the prevalent pattern of the arrangement of the 10 dark green polygons of Figure 9A could be maintained, in essence, achieving a typification. The 5 polygons participating in the group of Figure 9B could be typified to 2 to 3 placeholder polygons, depending on the scale reduction. In this example, we see the effect of the dissimilarity in size and shape, which splits off the 2 larger, concave polygons from the initial cluster resulting from the previous edge-removal step. This holds similarly for Figure 9C, where we see the effect of a difference in the orientation that has split the initial cluster into 2 groups, each with 4 participating polygons. These two groups could again be typified or aggregated.

The visual impression of the three examples of Figure 9 is further supported by Figure 10 and the quantitative results reported in Table 3. We can see that all values for the CV clearly drop in the transition from the initial candidate clusters (labelled A, B, C) to the final groups (A1, A2, A3, A4; B1; C1, C2), with two small exceptions for

Table 3. Summary of mean, STD and CV values (homogeneity) of initial candidate clusters and final groups, as shown in Figure 10. Boxes A, B and C denote the initial candidate clusters, while boxes A1, A2, A3, A4, B1, C1, and C2 highlight the final groups generated.

Criteria	Proximity	Size	Orientation	Shape
A				
Mean	40.2	714	132.25	0.67
STD	31.1	257	32.3	0.12
CV	0.77	0.36	0.26	0.18
A1				
Mean	17	730	120.5	0.64
STD	0	161	0.7	0.03
CV	0	0.22	0.006	0.05
A2				
Mean	15	770	90	0.53
STD	0	157	0	0.08
CV	0	0.2	0	0.15
A3				
Mean	36	1144	173.5	0.54
STD	0	59	4.5	0.05
CV	0	0.05	0.03	0.09
A4				
Mean	29	512.5	159	0.79
STD	0	16.3	5.7	0.03
CV	0	0.03	0.04	0.04
B				
Mean	25.2	721	166.3	0.625
STD	16.1	702	26	0.14
CV	0.64	0.97	0.16	0.22
B1				
Mean	31.2	363	176.2	0.74
STD	9	140	4.2	0.03
CV	0.29	0.39	0.03	0.04
C				
Mean	28	535	153.9	0.65
STD	23.1	132	38.4	0.12
CV	0.825	0.25	0.25	0.18
C1				
Mean	36	587.5	188.75	0.715
STD	33	92.4	21.5	0.075
CV	0.92	0.16	0.11	0.10
C2				
Mean	26	483	121.5	0.59
STD	15.7	158.5	13.8	0.13
CV	0.6	0.33	0.11	0.22

the CV of proximity in C1 and of shape in C2. The decreasing CV values show that the homogeneity of patterns is overwhelmingly increased in the refinement step.

The examples of Figure 9 also demonstrate that there is a clear tendency of the refinement step of breaking up the initial candidate clusters formed by the edge removal step into smaller groups. This is both due to the general principle of the refinement step, which only relies on the removal but not the insertion of edges, and the fact that the refinement algorithm uses the strict condition that all similarity criteria must be met for an edge to be retained. In the case of Figure 9A, this leads to fragmentation into polygon pairs, the smallest group that can be formed. In Figure 9C, where only the orientation differs, the fragmentation remains minimal, yet, one might perhaps argue that the initial candidate cluster already represented a cartographically meaningful group.

Generally, because they form larger and rather homogeneous groups, the initial candidate clusters, shown to the left of Figure 9, do already represent meaningful groups of sorts, but they are meaningful particularly when the scale reduction is more substantial than if the

refined, smaller groups are used for subsequent aggregation and typification operations. Basically, in our case, the refined groups relate to a scale transition to 1:50,000, while the initial candidate clusters would relate to a scale reduction to perhaps 1:100,000.

6.3. Alternatives, extensions

The three main steps of the proposed polygon group recognition process consist of several individual algorithms and decision points, some of which are parameter-free, such as the Delaunay triangulation, while others involve several interacting thresholds and parameters, such as the network refinement step. While we believe that overall, our proposed methodology has shown useful results, there is, of course, room for alternative solutions for the individual algorithms involved, and for fine-tuning of parameters and thresholds. Hence, we would like to discuss these possible alternatives for each of the three main steps of the proposed methodology.

6.3.1. Selection of foreground polygons

Using Tukey's outlier detection method had the advantage of a parameter-free, entirely data-driven procedure for establishing a threshold, splitting the polygons into the foreground and background polygons. In cartography, the Radical Law (Töpfer and Pillewizer 1966) is often used in object selection. However, this method works best for objects of similar nature, size, and importance, such as buildings in a topographic map, and typically fails in heavily skewed, long-tail distributions such as in our case. In our sample map, the Radical Law would have selected only 3,139 polygons, that is, far too few (Figure 1). Many polygons that would usefully be aggregated or typified would thus have been missed in the subsequent contextual generalization step. An alternative would have been to use the threshold of the minimum area constraint. For coloured polygon maps, this has typically been set in the order of 4 mm^2, translating to 10,000 m^2 on a 1:50,000 scale map (SSC 1977). Using this area threshold, 6,059 or 86.3% of polygons would have been designated as foreground polygons. That is, in our sample map this would have typically generated more and larger polygon groups.

6.3.2. Building and defining initial candidate clusters

The primary purpose of this step was to form initial candidate clusters of foreground polygons that could subsequently be refined using similarity criteria. This purpose has been largely met by the algorithm of Deng et al. (2011) that was used. However, since this algorithm uses the edges connecting polygon centroids for network pruning by long edges, rather than the exact distance between polygons, edges can be mistakenly removed and are lacking in the following network refinement step. This could be countered by integrating the actual proximity values in the network pruning step.

6.3.3. Refining network edges based on similarities

Of all the steps of the proposed methodology, this final step consists of the most significant number of individual algorithms and decision points. Thus, not surprisingly, this step has the most significant potential for modifications. For instance, further algorithms could be used to capture additional shape properties (e.g., convexity/concavity). More fine-grained classes could be used for the similarity assessment of shape and orientation measures. Alternatively, rather than using discrete classes on these measures, a continuous deviation tolerance threshold might be used, possibly even using fuzzy set methods (Anderson-Tarver et al. 2011). Also, how the individual similarity criteria are evaluated

and weighted could be adjusted. Currently, we use a nested intersection of the various criteria, which implies a priority ordering among the criteria, but does not assign different weights. Hence, different weighting strategies could be tested. However, a weighted evaluation scheme would also imply an even more significant number of parameters, and thus even more need for fine-tuning.

7. Conclusions

The paper has presented a network-based approach of identifying cartographically meaningful groups of polygons in geological maps using three main steps: the selection of foreground polygons; the formation of initial candidate clusters based on a Delaunay triangulation of the polygon centroids, followed by the removal of global and local long edges; and finally the refinement of the candidate clusters by applying similarity criteria such as the proximity between polygons in the map, their size, shape, orientation, and the rock type of participating polygons. An experimental evaluation of the results demonstrated adequate levels of similarity among the polygons participating in the groups that have been detected but also suggested possible extensions and alternatives for future work and experimentation. The proposed approach, inspired by the peculiarities of geological maps, is sufficiently modular and flexible to hold potential to be also adapted to the generalization of other types of categorical maps with similar polygon arrangements and patterns, such as soil maps or land cover maps, as well as small-area polygons (e.g., buildings) in topographic maps.

The next step in this research will be to use the identified groups for contextual generalization. Thus, future efforts will focus on questions related to the abstraction of the groups, involving contextual removal, aggregation, and typification operators, which will require a more detailed exploration of the group properties that have been generated in the network refinement step, leading to a semantically enriched database.

Acknowledgments

We acknowledge the support provided to the first author by the Swiss Federal Commission for Scholarships (FCS). The authors like to acknowledge their use of the Euriowie (including part of Campbells Creek) 1:25,000 Geological Map provided by the Department of Primary Industries - Mineral Resources, Geological Survey of New South Wales, NSW, Australia, which formed the basis for the experiments reported in this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by a Ph.D. scholarship of the Swiss Federal Commission for Scholarships (FCS) under the Grant 2015.0750.

ORCID

A. Sayidov http://orcid.org/0000-0001-8283-7979
R. Weibel https://orcid.org/0000-0002-2425-0077
S. Leyk https://orcid.org/0000-0001-9180-4853

References

Anders KH. 2003. A hierarchical graph-clustering approach to find groups of objects. Paper presented at: 5th ICA Workshop on Progress in Automated Map Generalization; Paris, France.

Anders KH, Sester M, Fritsch D. 1999. Analysis of settlement structures by graph-based clustering. In: Forstner W Liedtke, Ce Buckner J. Proceedings of the Semantic Modelling for the Acquisition of Topographic Information from Images and Maps Conference (SMATI'99). Munich, Germany: Deutsche Forschungsgemeinschaft, p. 41–49.

Anderson-Tarver C, Leyk S, Buttenfield B. 2011. Fuzzy modeling of geometric textures for identifying archipelagos in area-patch generalization. Cartography and Geographic Information Science. 38 (2): 137–145.

Anonymous 2016. Title withheld for anonymization purposes.

Barnes RG. 1988. Metallogenic studies of the Broken Hill and Euriowie Blocks (Bulletin № 32). New South Wales. 1. Styles of mineralisation in the Broken Hill Block. 2. Mineral deposits of the Southwestern Broken Hill Block. Sydney: Geological Survey of New South Wales.

Basaraner M, Selcuk M. 2008. A structure recognition technique in contextual generalisation of buildings and built-up areas. Cartographic Journal. 45(4):274–285.

Bobzien M, Burghardt D, Petzold I, Neun M, Weibel R. 2008. Multi-representation databases with explicitly modelled horizontal, vertical, and update relations. Cartographic Geographic Information Science. 35(1):3–16.

Boffet A, R, Serra S. 2001. Identification of spatial structures within urban block for town qualification. In: ICC, Vol. 3, Beijing, 1974–1983.

Brassel KE, Weibel R. 1988. A review and framework of automated map generalization. Int J Geograph Inform Syst. 2(3):229–244..

Burghardt D, Schmid S. 2010. Constraint-based evaluation of automated and manual generalised topographic maps. In: Gartner G, Ortag F, editors. Cartography in Central and Eastern Europe: Selected Papers of the 1st ICA Symposium on Cartography in Central and Eastern Europe. Lecture notes in geoinformation and cartography. Berlin: Springer; p. 147–162.

Cetinkaya S, Basaraner M, Burghardt D. 2015. Proximity-based grouping of buildings in urban blocks: a comparison of four algorithms. Geocarto Int. 30(6):618–632..

Christophe S, Ruas A. 2002. Detecting building alignments for generalisation purposes. In: Richardson DE, van Oosterom P, editors. Advances in Spatial Data Handling (SDH 2002). Berlin: Springer Verlag; p. 419–432.

Deng M, Liu Q, Cheng T, Shi Y. 2011. An adaptive spatial clustering algorithm based on Delaunay triangulation. Comput Environ Urban Syst. 35 (4):320–332.

Downs C, Mackaness W. 2002. An Integrated Approach to the Generalization of Geological Maps. Carto J. 39(2):137–152..

Federal Geographic Data Committee [prepared for the Federal Geographic Data Committee by the U.S. Geological Survey] 2006. FGDC Digital Cartographic Standard for Geologic Map Symbolization: Reston, Va., Federal Geographic Data Committee Document Number FGDC-STD-013-2006, 290pp., 2 plates.

Galanda M, Weibel R. 2002. An agent-based framework for polygonal subdivision generalization. In: Richardson D, van Oosterom P, editors. Advances in spatial data handling, 10th International symposium on spatial data handling. Berlin: Springer; p. 121–136.

Galanda M, Weibel R. 2003. Using an energy minimization technique for polygon generalization. Cartography and Geographic Information Science. 30 (3):263–275.

Galanda M. 2003. Automated Polygon Generalization in a Multi Agent System [PhD thesis] Department of Geography, University of Zurich. 188pp. Retrieved from https://www.geo.uzh.ch/dam/jcr:77c11613-a0f5-4b10-8dd9-491cf17527d3/thesis_MartinGalanda_2003.pdf.

Harrie L, Weibel R. 2007. Modelling the Overall Process of Generalisation. In: Mackaness WA, Ruas A, Sarjakoski LT, editors. Generalisation of geographic information: cartographic modelling and applications. Amsterdam: Elsevier Science; p. 67–87.

Haunert JH, Sester M. 2008. Area collapse and road centerlines based on straight skeletons. Geoinformatica. 12(2):169–191..

He X, Zhang X, Xin Q. 2018. Recognition of building group patterns in topographic maps based on graph partitioning and random forest. ISPRS J Photogramm Remote Sens. 136:26–40..

Li Z, Yan H, Ai T, Chen J. 2004. Automated building generalization based on urban morphology and Gestalt theory. Int J Geograph Inform Sci. 18(5):513–534.

Liu D, Nosovskiy G, Sourina O. 2008. Effective clustering and boundary detection algorithm based on Delaunay triangulation. Pattern Recog Lett. 29(9):1261–1273.
Lu Y, Thill JC. 2008. Cross-scale analysis of cluster correspondence using different operational neighborhoods. J Geograph Syst. 10(3):241–261.
Mark DM, Csillag F. 1989. The nature of boundaries on 'area class' maps. Cartographica. 26(1):65–77.
McCabe C. 2008. Vector approaches to generalizing faults and polygons in 1:24,000 geologic maps: Santa Rosa, California, case study. http://www.geovista.psu.edu/publications/2008/CMcCabe_GeologicMapGeneralization.pdf.
McMaster RB, Shea KS. 1992. Generalization in digital cartography. Washington, DC: Association of American Geographers.
Osserman B. 1978. The isoperimetric inequality. Bull Amer Math Soc. 84(6):1182–1238.
Peter B, Weibel R. 1999. Using vector and raster-based techniques in categorical map generalization. Third ICA Workshop on Progress in Automated Map Generalization, 14pp.
Regnauld N. 2001. Contextual building typification in automated map generalization. Algorithmica. 30(2):312–333.
Ruas A, Holzapfel F. 2003. Automatic characterisation of building alignments by means of expert knowledge. In: ICC, Durban, p. 1515–1604.
Swiss Society of Cartography [SSC]. 1977. Cartographic generalization – topographic maps. Cartographic Publication Series, vol. 2. Zurich: Swiss Society of Cartography.
Schylberg L. 1993. Computational Methods for Generalization of Cartographic Data in a Raster Environment [Ph.D. thesis]. Stockholm: Department of Geodesy and Photogrammetry, Royal Institute of Technology.
Seo S, Gary M, Marsh D. 2006. A review and comparison of methods for detecting outliers in univariate data sets. Department of Biostatistics. Graduate School of Public Health. 1–53. Retrieved from http://d-scholarship.pitt.edu/7948/.
Smirnoff A, Huot-Vézina G, Paradis J, Boivin R. 2012. Generalizing geological maps with the GeoScaler software: the case study approach. Comp Geosci. 40:66–86..
Smirnoff A, Paradis J, Boivin R. 2008. Generalizing surficial geological maps for scale change: ArcGIS tools vs. cellular automata model. Comp Geosci. 34(11):1550–1568..
Spiess E. 2002. Map Compilation. Anson R, and Ormeling FJ. 2002. Basic cartography for students and technicians, volume 2, edition 2. Published on behalf of the International Cartographic Association by Butterworth-Heinemann, Oxford, 2002.
Spiess E, Baumgartner U, Arn S, Vez C. 2005. Topographic maps, map graphics and generalisation. Zurich: Swiss Society of Cartography.
Steiniger S, Weibel R. 2005. A conceptual framework for automated generalization and its application to geologic and soil maps. Proceedings 22nd International Cartographic Conference, La Coruña (Spain), 11–16 July 2005.
Steiniger S, Weibel R. 2007. Relations among map objects in cartographic generalization. Cartography Geographic Information Science. 34(3):175–197..
Su B, Li Z, Lodwick G. 1998. Morphological models for the collapse of area features in digital map generalization. GeoInformatica. 2(4):359–382.
Su B, Li Z, Lodwick G, Müller C. 1997. Algebraic Models for the Aggregation of Area Features Based upon Morphological Operators. Cartogr Geogr Inf Sci. 11(3):233–246.
Wang X, Burghardt D. 2017. A Stroke-based approach to detect patterns of discrete buildings for generalization purposes. 20th ICA workshop on Generalization and Multiple Representation, Washington, DC, USA.
Wertheimer M. 1938. Laws of organization in perceptual forms. In: Ellis WD, editor. A source book of Gestalt psychology. New York: Harcourt, Brace; p. 71–88. Original work published 1923.
Yan H, Weibel R, Yang B. 2008. A multi-parameter approach to automated building grouping and generalization. Geoinformatica. 12(1):73–89.
Zhang X, Ai T, Stoter J. 2010. Characterization and detection of building patterns in cartographic data: two algorithms. Int Arch Photogrammet, Remote Sens Spatial Inform Sci. 38(II):261–266.
Zhang X, Stoter J, Ai T, Kraak MJ, Molenaar M. 2013. Automated evaluation of building alignments in generalized maps. Int J Geograph Inform Sci. 27(8):1550–1571.

4

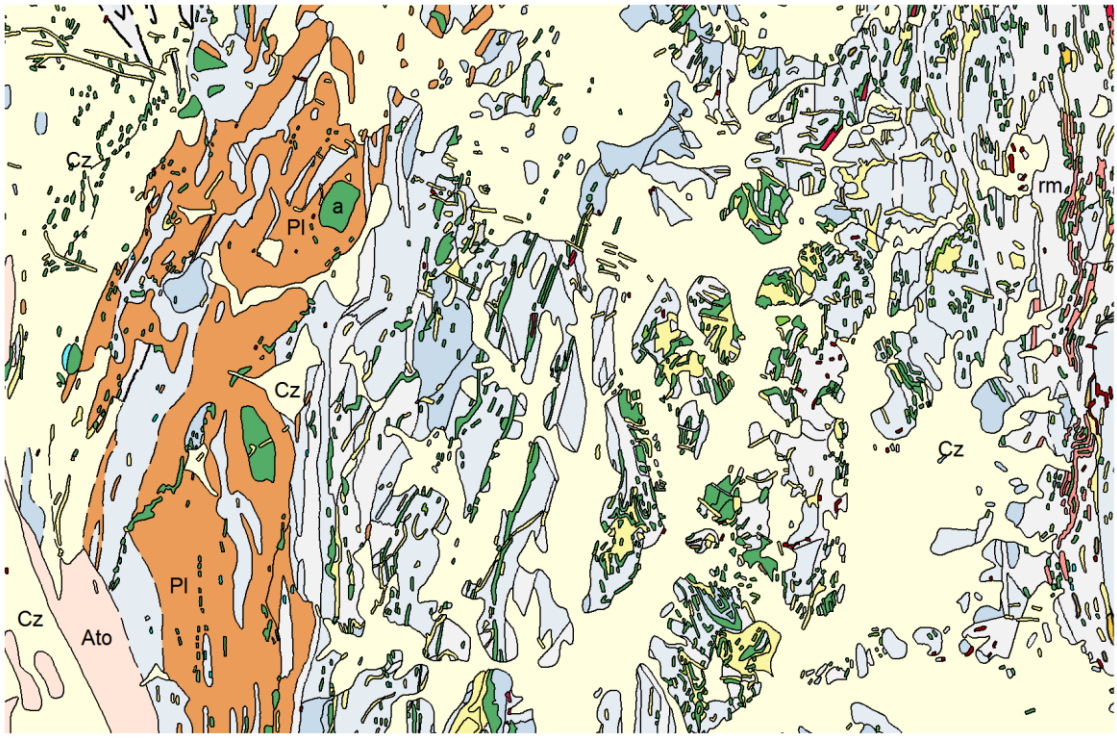
Chapter

I do believe that enduring geological features are important,
though I don't think I can be clear about exactly why.

Tracy Kidder.

Research Paper 3

Sayidov, A., Leyk, S., & Weibel R. (submitted). Integrating Aggregation and Typification to Generalize Polygon Group Patterns in Geological Maps. Cartography and Geographic Information Science.



Part of Geological map of the Euriowie Inliers (Stevens et al., 1998 - https://search.geoscience.nsw.gov.au/product/251)

Contribution of the Ph.D. candidate:

The Ph.D. candidate designed the work, processed, interpreted, and visualized the geological map; developed the conceptual model for the generalization group of polygons and its evaluation; coded the algorithms for the generalization of polygons (selection, aggregation, and typification); and drafted the article.

Authors' Contributions: Conceptualization, Data analysis, Visualization and Validation, Azimjon Sayidov; Methodology, Azimjon Sayidov, Robert Weibel and Stefan Leyk; Software, Azimjon Sayidov; Writing original draft, Azimjon Sayidov; Writing review and editing, Azimjon Sayidov, Robert Weibel and Stefan Leyk.

Integrating aggregation and typification operators to generalize polygon group patterns in geological maps

A. Sayidov[a], S. Leyk[b], and R. Weibel[c]

[ac]Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057, Zurich, Switzerland.

[b]Department of Geography, University of Colorado Boulder, UCB 0260, Boulder, 80309, CO, USA.

[a]*tel.: +4144 63 55198. Email: azimjon.sayidov@geo.uzh.ch,*

[c]*tel.: +4144 63 55253. Email: robert.weibel@geo.uzh.ch,*

[b]*tel.: +4144 63 55190. Email: stefan.leyk@colorado.edu.*

Integrating aggregation and typification operators to generalize polygon group patterns in geological maps

Abstract

Geological maps, as a representative of categorical maps, contain structures with polygons of elaborate and variable shapes, different sizes and orientations. Preserving these properties and thus the main message of the map during the generalization process is the goal of this research, with a particular focus on the treatment of small polygons. The method described here recognizes and identifies key groups of polygons in the map, and depending on the properties of these groups and potential effects due to generalization, either the typification or the aggregation operator is used during the generalization step. Changes such as the area loss and gain, shape alteration, and orientation change during the generalization are evaluated using constraint-based self-evaluation as well as global statistics. The assessment results also serve as a feedback loop to inform the operator selection process. The proposed approach demonstrates that generalizing a geological map using a grouping-based approach can preserve its main characteristics and overcomes limitations of existing approaches.

Keywords: Geological mapping; polygonal maps; categorical maps; pattern recognition; group formation; map generalization; typification; aggregation

1. Introduction

As the essence of cartography, map generalization is a complex and crucial part of the map-making process, responsible for abstracting reality and turning it into meaningful visualization at reduced scales. At its core, generalization is meant to render the main characteristics of map objects and structures, as well as their peculiarities and relationships. However, failing to preserve the map's main message during generalization may result in a document that carries ambiguous or confusing information, and misleads the map user. Thus, a successful generalization process needs to consider individual features and their properties on the map, detect and identify

relationships between them, and cautiously address these properties while applying appropriate algorithms and operators taking into account the target scale and defined constraints (Brassel & Weibel, 1988; McMaster & Shea, 1992).

Map generalization has been an active domain of research for several decades (Brassel & Weibel, 1988; Mackaness et al. 2007; Burghardt et al., 2014; Stoter et al., 2014). Nevertheless, the generalization of thematic maps, especially polygonal maps such as soil or geological maps, has received relatively little attention, with some exceptions (Downs & Mackaness, 2002; Galanda, 2003; Smirnoff et al., 2012). The role of spatial relationships between map objects for improved understanding of map contents has been acknowledged in the cartographic practice and research for some time (Brassel & Weibel, 1988; McMaster & Shea, 1992; Steiniger & Weibel, 2005, 2007; Yan, 2010; Yan, 2014; Chehreghan & Ali Abbaspour, 2017). Since geological maps are among the most complex thematic maps, there is an urgent need to carry out a rigorous analysis of the structures and inter-relationships between map objects prior to any generalization. However, existing solutions fail at combining the analysis of individual polygon properties with the assessment of structural relationships between polygon objects, and remain limited in their ability to locally adapt the generalization process (e.g., using cellular automata in Smirnoff et al., 2012).

Geological maps are a representative of categorical maps, and consist, apart from linear objects such as fault lines, or point objects such as wells, of a subdivision of polygons that vary significantly in size, shape, orientation, and category. Despite the complexity and great variability of the content, the spatial arrangements of polygons are typically highly structured — as a result of the geological processes that shaped he geological structures portrayed by polygons in the map — and show patterns of

similarities of polygon objects that are proximate, of the same category, and comparable in size, shape, and orientation (Steiniger et al., 2006; Sayidov et al., 2020a).

In this paper, we focus mainly on small polygons that fall below, or are not much larger, than the minimum area limit required for map objects to still be legible on the target map. These polygons risk being eliminated in the generalization process, but uncontrolled removal of polygons based on size only is likely to lead to loss of salient structures formed by collections of similar, small polygons (Sayidov et al., 2020b), and thus loss of representation of prevailing geological processes. Hence, we present a methodology that first identifies patterns of similar polygons as a *group* (as described in Sayidov et al., 2020a), to then support locally adaptive generalization decisions and thus overcome the limitations of existing methods. The decisions made are based on changes in group-level properties, such as mean area, shape and orientation of polygons, as well as the category of the geological units. Based on these changes, one of two generalization operators, aggregation or typification, is employed, for which we present corresponding algorithms. Since the proposed algorithms build on properties of polygons that can be derived from any polygon (i.e., size, shape, orientation, and category), the proposed methodology holds potential to also be adapted for the generalization of other categorical maps such as soil, land use, or land cover maps, with little additional tuning.

2. Related work

Most of the research and practice in map generalization to date has been directed at topographical maps and the automation of existing approaches (e.g., Baella et al., 2007; Lecordix & Lemarié, 2007; Foerster et al., 2010; Regnauld, 2011). The rather limited research on the generalization of categorical maps can be divided into raster- and

vector-based generalization. Below, we provide short summaries of both areas of study and short descriptions of generalization operators relevant for this paper.

2.1. Raster-based generalization

Filters of mathematical morphology that grow or shrink regions are commonly used in raster-based generalization approaches (e.g., Schylberg, 1993; Su et al., 1997, 1998). Such filter approaches are relatively fast and have generalization effect but do not perform well when spatial variability (e.g., of raster patch shapes and sizes) is high and small patches are involved (Peter & Weibel, 1999). Smirnoff et al. (2008, 2012) proposed a grid-cell based cellular automata approach which probably represents the most comprehensive solution of geological map generalization to date. It provides acceptable results for maps that have rather low levels of variation in their feature characteristics. However, owing to the raster-based representation, this approach cannot address individual and local properties, such as size, shape, and orientation of geological objects and cannot be used for grouping-related approaches, accordingly.

2.2. Vector-based generalization

In most cases, vector-based generalization is applied to process categorical maps either in a constraint-based approach (Peter & Weibel, 1999; Edwardes & Mackaness, 2000; Peter, 2001; Galanda, 2003; Sayidov et al., 2020b) or database generalization for multiple representations (Brodaric & Patera, 2001; Yaolin et al., 2002; Downs & Mackaness, 2002; Patera, 2006; McCabe, 2008). Both approaches are used in the domain, and some of the prominent examples are outlined below.

2.2.1. Constraint-based generalization

Peter and Weibel (1999) and Edwardes and Mackaness (2000) discussed constraints for

the generalization of polygon objects. A series of constraints were identified and modeled for the automated generalization of categorical maps by Galanda (2003) in an agent-based framework. Sayidov et al. (2020b) proposed the use of two size constraints, ‘minimum area’ and ‘object separation’, for the generalization of free-standing polygons in a geological map, using four different generalization operators: enlargement, elimination, aggregation, and displacement. While the method preserves the readability of the generalized map, a visual imbalance and distortion of the main characteristics in the generalized map was noted for larger scale transitions. It was concluded that generalization by simple size constraints alone was insufficient and that a more sophisticated process using recognition of the main map patterns prior to generalization would be required to achieve large scale transitions.

2.2.2. Database generalization

Few studies have implemented a database generalization approach for geological maps. For example, Brodaric and Patera (2001) use stratigraphic information to link stratigraphic hierarchies to class aggregation. Patera (2006) classifies geological objects that contain the lithostratigraphic sequence as attributes; these attributes are queried to generalize the geological database and obtain multiple representations. The combination of several generalization operators based on a rule-based approach has been proposed by Downs and Mackaness (2002) for the generalization of a geological map. While this approach is flexible and simple, it lacks in objectivity due to the human interaction in the process.

The major drawback of these approaches is a lack of consideration of the categorical map as a display of polygons and patterns representing geological

formations bound together in space and time, detectable by structural similarities that can inform the generalization process, as mentioned in Section 1.

2.3. Generalization operators and algorithms

The generalization of a group of small, free-standing polygons requires several operators to reduce the number and rearrange the size and position of polygons (Regnauld & McMaster, 2007). This may involve the *enlargement* of polygons falling below the minimum area limit, the removal of particular polygons by *elimination* (Töpfer & Pillewizer, 1966; Ruas, 1998), *displacement* of polygons that are too close to be visually discernible (Mackaness & Purves, 2001; Galanda & Weibel, 2003; Bader et al., 2005; Basaraner, 2011; Ai et al., 2015), fusing groups of nearby polygons by *aggregation* (also called amalgamation in the literature; Regnauld, 2003; Li et al., 2004; Regnauld & McMaster, 2007), or the reduction of the number of polygons while maintaining their distributional pattern by *typification* (Regnauld, 2001; Burghardt & Ceccoci, 2007). Of these five possible operators, enlargement is a simple operator; elimination in polygon groups invariably necessitates rearrangement of the remaining polygons, hence essentially leading to typification; and if displacement takes place within polygon groups it can also be seen as being part of the typification operator. Consequently, we restrict our review to existing algorithms for the aggregation and the typification operator, respectively.

2.3.1. Aggregation

Aggregation denotes the merging of free-standing, nearby polygonal objects during the process of scale reduction. There are three main aggregation approaches proposed in the literature. The first algorithm anticipates aggregation by buffering and merging polygons located in close proximity (Perikleous, 2006; Mackaness et al., 2008; Liu et

al., 2008). The second approach relies on building a constrained triangulation from polygon vertices of neighboring polygons and, consequently, merging the connected triangles (Bader & Weibel, 1997; Galanda, 2003; Regnauld, 2005; Li et al., 2017). A third method builds a bounding geometry (i.e., convex hull, concave hull, etc.) connecting the vertices of the representative polygon boundaries and forms a merged polygon geometry (Moreira & Santos, 2007). Similar to typification, the majority of existing algorithms has focused on the aggregation of buildings in topographic maps, including algorithms approximating building polygon groups by triangulation and rectangular shapes (Regnauld & Revell, 2007) and graph partitioning (He et al., 2018). These algorithms are not relevant for our study, as we focus on the aggregation of arbitrarily shaped polygons.

2.3.2. Typification

Typification reduces the number of polygons in a group while maintaining its typical spatial structure and arrangement. As a result, a set of polygons is represented by a subset of placeholders that have been repositioned to reflect the spatial arrangement of the source map. An approach for detecting a group of buildings based on Gestalt theory was proposed by Regnauld (2001), which included steps of partitioning building sets into groups, followed by typification as a combination of enlargement, removal, and displacement of objects in a group. A grouping method based on the detection of building alignments, which are used to define typification algorithms, was proposed by Ruas and Holzapfel (2003). A mesh simplification approach, which gradually reduces the number of polygons and repositions them, is used by Burghardt and Cecconi (2007). The overall process consists of two steps: a partitioning step, which identifies the number and position of objects based on the Delaunay triangulation, and representation,

which calculates the size and orientation of the replacement buildings.

Most of the research on typification is dedicated to building generalization, utilizing particular properties of buildings such as shape, orientation, alignment. Since buildings are represented by small (though regular) polygons, these methods have the potential of being utilized in the generalization of groups of small polygons in categorical maps.

3. Methodology

3.1. Overall methodology

The proposed approach considers polygons in geological maps as individual patches that collectively form groups of patches to facilitate their generalization. The methodology consists of three stages (Figure 1): group analysis, the generalization of groups, and evaluation of results. The group analysis stage recognizes patterns and identifies meaningful polygonal groups from those patterns based on closeness, size, shape, orientation, and geological category. In the generalization stage, the identified polygon groups are generalized by selectively implementing two generalization operators, aggregation and typification. The selection of operators is linked to differences in area, shape, and orientation of polygon objects before and after trial generalization. The evaluation of results includes a constraint-based self-evaluation, and a statistical assessment of changes in object properties such as size, shape, and orientation comparing different generalization methods, different operators and different target scales. These three stages are described in detail in the remainder of this section.

3.2. Group analysis

Group analysis consists of three steps: selection of polygons, building and defining initial networks, and refining the network edges based on similarity criteria. This stage of the overall methodology uses a procedure that has been described in detail in a separate paper (Sayidov et al., 2020a). Therefore, we will give only a summary of this procedure in Sections 3.2.1. to 3.2.3.

3.2.1. Selection of polygons

Geological maps are complex, consisting of polygons that vary in size. Typically, however, the majority of polygons has a relatively small area. These small polygons are the main focus of the group identification as they are the primary candidates to violate minimum-size constraints and thus often require generalization solutions (Sayidov et al., 2020a).

Sayidov et al (2020b) used the term ‘foreground polygons’ for small polygons, while large and bulky polygons were termed ‘background polygons’. To separate foreground from background polygons, Tukey’s outlier detection method (Seo et al., 2006) is implemented. It is independent of the standard deviation which makes it an appropriate method to differentiate foreground and background polygons in heavily skewed data distributions (with many small and few large polygons). As such, foreground polygons are seen as primary data, while background polygons are interpreted as the noise in the data to be detected as outliers.

3.2.2. Building and defining initial networks

During this step of group analysis initial candidate patches are identified based on proximity relations. An effective way of connecting immediate neighboring polygons is

by building a Delaunay Triangulation (DT) using centroids of the respective polygons (Liu et al., 2008; Lu & Thill, 2008). Once the DT is built, the so-called global and local long edges of the network are identified and removed following the approach by Deng et al. (2011). Removing global long edges from the DT forms meaningful clusters globally. Similarly, when the local long edges are removed, larger clusters are broken down into smaller clusters, forming initial candidates for the detection of meaningful polygon groups. Figure 2 illustrates the steps of building the DT, as well as identifying and removing global and local long edges. Earlier work has demonstrated that this approach can expose clusters with arbitrary shape and density (Sayidov et al., 2020a).

At the end of this step, after removal of local long edges, clusters forming initial candidates are ready to be passed to the next step for additional refinement and identification of meaningful polygon groups.

3.2.3. Refining network edges based on similarity criteria

The third and final step of group analysis is devoted to refining initial candidate networks with the purpose of identifying meaningful groups. The final groups of polygons are selected using similarity measures based on the size, shape, orientation, and category of polygons as well as proximity. To form a polygon group, all candidate polygons have to satisfy the similarity criteria (Sayidov et al., 2020a), have to be within a maximum distance, and have to be of the same category. Figure 3 illustrates the group identification process. The left panel shows the network of initial candidates in blue after removal of global and local long edges. The is initial network connects 9 candidate polygons based on proximity, which however are still different in size, shape, orientation, and category. The right panel of Figure 3 shows the final, refined network in red, connecting similar polygons. During the refinement step, the network is trimmed

down by removing edges that connect polygons that differ in relation to size, shape, orientation, or category (with reasons given in Figure 3, right).

As a result of this first stage the methodology — group analysis — we have rich information that can be utilized in the following generalization and evaluation stages, respectively: besides the Delaunay triangulation, the global long and local long edges, each identified polygon group is associated with three kinds of attributes derived from its member polygons, including size, shape, and orientation (the category is the same for all member polygons).

3.3. Generalization

During the generalization stage the essential polygon groups detected in the group analysis stage are further processed using generalization operators suitable for groups of polygons. As mentioned in Section 2.3, for small polygons this includes five possible operators: enlargement, elimination, displacement, aggregation, and typification (Regnauld & McMaster, 2007). Recent work by Sayidov et al. (2020b) provided a solution for polygon generalization that relies on size constraints and implements enlargement, elimination, aggregation, and displacement operators, however without a preceding group detection stage. Hence, aggregation and displacement, that is, the operators that would typically consider the entire spatial context of a group of map objects, were restricted to pairwise operations, while typification was simply not possible using this approach. Thus, although the results were encouraging when used for minor to moderate scale reductions, the methodology broke down when more significant scale reductions were required. Thus, in Section 3.3.1 to 3.3.3 we propose alternative algorithms for aggregation and typification, as well as a process for selecting the appropriate operator in Section 3.3.4, capitalizing on the information extracted in the

group analysis stage. Note that since the polygon groups generated by the group analysis must be of the same category, aggregation and typification will also only be performed between polygons of same category.

3.3.1. Aggregation

Aggregation involves the fusion of a set of individual, proximal polygons representing a series of lakes, islands, or rock types, due to scale reduction. The common goal of such a fusion of a group of polygons that share similar properties is to remove congestion of polygons and make them readable at a reduced map scale.

Two basic algorithms are used in this study to carry out the aggregation. The first algorithm relies on the buffering and merging of polygons using the approach described in Perikleous (2006) and Mackaness et al. (2008). The buffering algorithm is parameterized according to the 'object separation' (OS) constraint α, which defines the minimum visual separability distance between two map objects (Sayidov et al., 2020b), and takes two steps: In a first step, polygons separated by a distance less than α are expanded by a buffer distance of $\alpha/2$ and interior boundaries dissolved to form a single polygon; in a second step, the expanded polygon is contracted by an inward (or negative) buffer using the same buffer distance, giving the final, aggregated polygon. More details on the algorithm can be obtained from Perikleous (2006) and Mackaness et al. (2008).

The second algorithm builds a bounding geometry as a concave hull based on an approach developed by Moreira and Santos (2007). This algorithm was also used by Sayidov et al. (2020b). The vertex with the smallest y value is selected as a starting vertex by the algorithm, and k points nearest to the current vertex are examined as potential candidates to be the next vertex of the concave polygon's boundary. The point

with the largest right-hand (clock-wise) turn from the previously found boundary segment is selected as the next vertex to form the next segment of the polygon boundary. This process is repeated until the selected point is the first point, concluding the search process with a closed polygon shape. k serves as a control parameter to adjust the degree of concavity in the aggregated polygon. For details, see the paper by Moreira and Santos (2007).

Figure 4 illustrates the results of aggregating a small polygon group using the two aggregation algorithms. The buffering algorithm (Fig. 4b, 4d) has a tendency to form smoother shapes compared to the concave hull approach (Fig. 4c, 4e). However, after some experimentation, we found that the buffering algorithm can generate 'hourglass' shapes if two polygons are positioned at a distance approaching α (this effect is also starting to show in Fig. 4b). We therefore use the concave hull algorithm for all subsequent experiments.

3.3.2. Typification of small polygon groups

Typification is a generalization operator that replaces a set of objects by a subset of representative placeholders. An essential requirement of this generalization operator is that the typical spatial structure and arrangement of the original set of map objects are preserved (Anders & Sester, 2000; Regnauld, 2001; Burghardt & Cecconi, 2007; Wang & Burghardt, 2019).

We have implemented two typification approaches, which are triggered depending on the number of polygons included in the group to be typified. For small polygon groups of up to five polygons, we use the 'select-best' approach described in this subsection. For larger polygon groups an iterative approach is used that will be explained in the next subsection.

The select-best approach (partially inspired by Burghardt & Cecconi, 2007), performed on small polygon groups, is shown in Figure 5. It selects the most representative polygon of a particular group based on three criteria: size (i.e., area), shape, and orientation of polygons. As a measure of shape, the ipsometric quotient (Osserman, 1978) is used, which takes values between 0 and 1, where values close to 1 indicate round, compact shapes, while values close to 0 indicate elongated polygons. The orientation of a polygon is calculated by creating centerlines of polygons using the skeleton algorithm by (Haunert and Sester, 2008). Mean values of these three properties are calculated for each identified polygon group, and compared to those of the individual polygons in that group. The polygon whose properties are closest to the group mean values is selected for typification.

The selected (best) polygon is then enlarged using the approach described in Sayidov et al. (2020b) to fulfill the 'Minimum Area' (MA) constraint, rotated to the mean orientation, and shifted to the centroid point (i.e., the center of mass) of the group of polygons. The centroid of a group of polygons is calculated by generating the bounding geometry around all member polygons (using the concave hull algorithm of the aggregation operator), then calculating its geometric center.

3.3.3. Typification of large polygon groups

To typify larger groups of polygons (exceeding five polygons), we propose a divide-and-conquer strategy, which first divides large groups into small and manageable subsets (called subgroups here) and then applies the typification and aggregation algorithms on each of the subgroups, as described above. Figure 6 illustrates the processing of a group of 15 polygons that results in 6 subgroups. The subdivision process uses the same methodology for group analysis as summarized in Section 3.2,

but this time with parameters for global and local long edges adjusted to the polygon group shown in Figure 6a that is being processed.

Once the subdivision process has been achieved, the resulting subgroups are either typified, aggregated, or both (Figure 7) based on constraints for area loss and gain, shape distortion, and orientation alteration. The decision-making as to which operator has to be selected is described in Section 3.3.4.

3.3.4. Operator selection

As part of the generalization stage, the operator selection step is closely connected with the evaluation stage (Section 3.4) and thus linked to the general purpose of the map generalization, that is, to minimize the alteration of a map at reduced scale, while maintaining its integrity and readability. Each operator described above has certain properties and implications such as area loss and gain (i.e., a particular category may lose or gain area in the course of the generalization process), change of shape (alteration of the shape of modified polygons as a consequence of generalization), and orientation change (alteration of the main angle of polygons due to generalization). These traits — and particularly the minimization of alterations — are used as criteria to inform the selection of the operators.

The example of Figure 7 illustrates how the aggregation and typification operator, respectively, affect the shape, orientation, and area, respectively, of the generalized polygons (Fig. 7b and 7c). Based on these properties the more appropriate operator is selected to generalize the individual groups of polygons (Fig. 7d).

The aggregation operator maintains the relative orientation and shape of the group if its member polygons are aligned linearly in front of each other, as is the case in subgroups 5 and 6. Aggregating these polygons retains the relative shape and

orientation of the groups. Also, in this linear alignment the area gain is minimized. Conversely, subgroups 1 to 4 all have a parallel spatial configuration, which would lead to shape and orientation distortion, as well as excessive area gain, if the aggregation operator was applied (Fig. 7b). For such groups, which are linearly aligned but not in front of each other (subgroups 1, 2, 4) or in more arbitrary configurations (subgroup 3), the typification operator is used, which better preserves the relative shape and orientation of polygons, but loses area due to the removal of some member polygons (Fig. 7c).

In order to identify linearly aligned and polygons located in front of each other we introduce the 'view angle', which determines whether a polygon is 'visible' from another one within a certain angle deviation. The view angle θ is calculated as the angle between the main axis of a given polygon and the line connecting its centroid with the centroid of another polygon whose 'visibility' is tested (Figure 8). The view angle computation starts with the polygon whose *y*-coordinate of its centroid is the lowest in a (sub)group (polygon A in Fig. 8), testing the polygon whose centroid has the next higher *y*-coordinate (polygon B). If the higher polygon is positioned within the acute (inner) sector θ_{in} of the lower polygon, or in other words θ deviates less than $\pm{}^{\theta_{in}}/_{2}{}^{\circ}$ from the main axis of the lower polygon, they are deemed similar in orientation, and they are marked for generalization by the aggregation operator. The reason being that with such a small difference in the orientation of the polygons, the orientation and shape of the resulting aggregated polygon can be expected to experience alteration within tolerable bounds. However, if the differences in orientation between the polygons in the same group are greater than the inner sector, the typification operator is the preferred generalization approach. In that case, the orientations and positions of the polygons are too different to warrant acceptable results from an aggregation operator. If the group

contains more than two polygons, the view angle test is repeated with the next higher pair of polygons (B and C in this case), until all polygons in the group have been tested. Eventually, all polygons that pass the view angle test are aggregated; those that fail the test are typified.

So, in a first instance, the selection of the two operators is controlled by the view angle testing, as described above. Following this, a statistical evaluation of the area loss and gain, shape distortion, and orientation alteration is carried out to assess the validity of the generalization operations performed, as described in Section 3.4.2. If the distortions in area, shape, and/or orientation are deemed excessive, the angle θ_{in} of the inner sector can be adjusted: It can either be made smaller, thus favoring typification over aggregation, or it can be made larger (up to the angle θ_{out} of the outer sector), increasing the likelihood of the aggregation operator being selected.

3.4. Evaluation

3.4.1. Constraint-based self-evaluation

The primary purpose and driving force of the generalization is the readability of map objects at the target scale, constrained by the natural limits of human visual perception. The readability of map objects is bounded by two size constraints, the minimum area (MA) and the object separation (OS) constraint, respectively. These two constraints trigger polygons to be eliminated, enlarged, aggregated, or displaced, based on conflict situations detected by a violation of size constraints (Sayidov et al., 2020b). This means that all polygons at the target scale that are smaller than the MA value are enlarged to meet the MA constraint, and if the enlargement of polygons causes knock-on conflicts, such as overlap or congestion, these polygons are then aggregated or typified (as shown above) and possibly displaced (if they are individual polygons not participating in a

group) to the meet the OS constraint. As this evaluation capacity is intrinsically embedded in the overall generalization process and does not involve a further, external review, we call it a constraint-based self-evaluation (Sayidov et al., 2020b).

3.4.2. Statistical evaluation: area, shape, orientation

The aggregation operator has a tendency to cause area gain and shape distortion of the polygons, while typification may lead to area loss. For each polygon, therefore, three characteristics — area, shape, and orientation — are calculated before and after generalization (using the measures described in Section 3.3.2) to assess how these properties change. The results of this statistical report are used in a feedback loop to evaluate the appropriateness of the operators selected and their effect on the generalization results. As described in Section 3.3.4., initially, polygons are either typified or aggregated depending on the view angles of polygons of the same group. However, based on the statistical report of area loss and gain, shape distortion, or orientation alteration, the view angle tolerances, in particular the angle θ_{in} of the inner sector (Figure 8), may be adjusted, and thus the selected generalization operator may change for some polygon groups.

3.5. Generalization of polygons that are not part of a group

In Section 3.2.1, the selection of foreground polygons was described. For a complete generalization of a geological map both layers, that is, the foreground and background polygons, have to be treated. Also, it may happen that in the selection of foreground polygons, some polygons may be left without any group membership. Those ‘other polygons’ are typically isolated island polygons that are too far from other polygons to be included in a group. While in this study the focus is on the foreground polygons that are part of groups, we also process island and background polygons using an approach

described in Sayidov et al. (2020b).

In short, island polygons are processed in four steps: selection, enlargement, aggregation, and displacement, using a set of constraint values described in Sayidov et al. (2020b).

The generalization of large, bulky background polygons is rather straightforward, as there is no need for them to be eliminated, enlarged, aggregated, typified or displaced. Hence, the process focuses on outline generalization by refining the consecutive vertex distance and polishing the outline granularity (Galanda, 2003; Sayidov & Weibel, 2016). These two constraints are satisfied using refinement operators such as simplification and smoothing (Galanda & Weibel 2003), without the need for group detection or combinations of generalization operators to treat them. Point removal algorithms (Douglas & Peucker, 1973) are used to simplify lines by removal of redundant vertices, with conservative tolerance values. Prior research provides recommended tolerance values for maps of different scales. For example, Gary et al. (2009) applied a tolerance value of 500 meters for a map scale of 1:1,000,000; Wilmer & Brewer (2010) used 100 meters for a map scale of 1:100,000; and Stanislawski & Buttenfield (2010) proposed a 200-meter tolerance applied to a map of scale 1:200,000. Based on the above studies, a tolerance value of roughly 50 meters can be recommended for generalization of polygon boundaries in a 1:50,000 scale map.

3.6. Implementation

A range of open source and commercial software, projects, and libraries has been utilized in this study. For network generation, processing, and analysis, the Python package NetworkX 2.2 was used (https://networkx.github.io). GeoPandas 0.6.0, an open source project that handles geospatial data in the Python programming language, was

used to implement geometrical algorithms in generalization operators, such as dissolving, buffering, centroid identification, convex hull, rotation, scaling, and translation. Moreover, the ArcPy package by Esri, Inc. was used for standard geoprocessing operations such as finding proximal objects. Map visualization was carried out using Esri ArcGIS 10.6 and Matplotlib, a plotting library for the Python programming language.

4. Experiments and Results

4.1. Experimental setup

We use the work of Sayidov et al. (2020b) as a baseline for our experiments, for three main reasons. First, their methodology implements the fundamental readability constraints (which they call size constraints), including minimum area, object separation, outline granularity, and consecutive vertex distance, which can be considered as the main drivers of cartographic generalization. Second, their methodology performs reasonably well for minor to moderate scale reductions, and according to the results reported in their paper, seems to outperform the state-of-the-art method by Smirnoff et al. (2008, 2012) for the task of polygonal map generalization. And third, and most importantly, the authors note limitations of their methodology in larger scale reductions, and attribute these problems to the lack of preceding group detection procedure. And that is exactly what we would like to demonstrate: that the inclusion of the group analysis stage allows larger scale reductions and improves the generalization quality.

Thus, we adopt the experimental setting of Sayidov et al. (2020b):

- The data is taken from the 1:25,000 scale geological map “Euriowie Block (including part of Campbells Creek)” published in the year 2000 by the Geological

Survey of New South Wales (NSW), Department of Mineral Resources, Australia (available from https://www.resourcesandgeoscience.nsw.gov.au). A part of this map is shown in Figure 9. For lack of space, only a small portion is shown. Note, however, that the generalization process was executed over the entire map, and hence all numbers regarding polygons given below refer to the entire map, rather than merely the portion shown in Figure 9.

- The scale is reduced to 1:50,000, 1:100,000, and 1:200,000.
- We compare against the best-performing method used in Sayidov et al. (2020b), called 'Category-wise selection' by the authors.
- We use implementations of their enlargement and displacement algorithms.
- We use their best-performing set of goal values for the MA constraint (0.75 mm2) and the OS constraint (0.6 mm), called 'Compromise' set, representing a compromise of different goal values published in the literature. The MA constraint drives the enlargement, aggregation and typification operators; the OS constraint is used to control the group analysis and for displacement of individual polygons after aggregation and typification.

Furthermore, we use the following parameters for the proposed algorithms used here: In the concave hill algorithm for aggregation, k was set to 18; in the view angle test, the tolerances were initially set to θ_{in} = 45° and θ_{out} = 90°.

4.2. Group analysis

First, we present the results for the group analysis stage. The entire data set includes 7019 polygons, of which 5536 were identified as foreground polygons and 1483 as background polygons, respectively. Table 1 reports the group size distributions of the polygon groups identified for the three target map scales, with a goal value for the OS

constraint of 0.6 mm. The group analysis resulted in 797 groups in the 1:50,000 scale map, with group size ranging from 2 to 9 polygons (645 groups with 2 polygons, etc.). In the 1:100,000 scale map, 660 groups were identified, with 2 to 15 members. Finally, in the 1:200,000 scale map, 568 groups were found, between 2 and 33 members strong. Generally, we see that the smaller the scale, the smaller the total number of groups found, and the larger the group size.

There were 224 foreground polygons identified as not being part of a group (Section 3.5). Of these non-group-forming polygons, 14 polygons were eliminated, 160 were enlarged to the MA constraint, 50 polygons were aggregated, and 10 were displaced to meet the OS constraint. The large background polygons were subjected to simple aesthetical refinement operations, as described in Section 3.5.

4.3. Generalization of polygon groups

Figure 10 shows the effects of the proposed methodology relying on the identification of cartographically meaningful polygon groups, followed by generalization using aggregation and typification algorithms. The top row depicts the original map (left), as well as the generalized map (right), both at the scale of 1:50,000. In the top-left panel, which represents a reduction of the original 1:25,000 map to 50 % without generalization, we can clearly see that some of the small polygons are no longer clearly legible. The top-right panel shows the result of the proposed generalization methodology with parameter settings for the target scale of 1:50,000. The inset maps allow to more clearly assess the changes induced by the generalization process.

At the scale of 1:50,000, most groups were generalized using the typification operator (677 groups or 85 % of all groups), while only a small fraction was aggregated (120 groups or 15 % of all groups).

4.4. Scale series

We have created a series of maps at different scales to demonstrate the behavior of the generalization approach over larger scale reductions, using the parameter settings described in Section 4.1. Figure 11 shows the original map at scale 1:25,000 and the generalization results at three different target map scales: 1:50,000, 1:100,000, and 1:200,000.

4.5. Comparing to the baseline approach

As mentioned in Section 4.1, we also want to compare to a different approach (Sayidov et al. 2020b) that considers size constraints and has been shown to outperform the state-of-the-art, cellular automaton-based method by Smirnoff et al. (2008, 2012), but does not include the capacity of identifying polygon groups to inform contextual generalization operators such as aggregation and typification. Figure 12 allows to compare the results of the 'Size Constraints' approach by Sayidov et al. (2020b) to those of the proposed methodology (labeled "Grouping Approach") at three different scales.

4.6. Effects of generalization approach and operators on map properties

Finally, we want to examine the effects of the generalization approach and operators used on key properties of the target maps, including area loss and gain, overall shape distortion, and orientation alteration as a consequence of generalization. We focus on the transition to the target scale of 1:50,000. All results are reported for the foreground polygons only.

4.6.1. Effects of generalization approach

First, we compare the changes in area, shape, and orientation of polygons (computed

using the methods described in Section 3.3.2) after applying two different generalization approaches, the ‘size constraints’ method of Sayidov et al. (2020b) and the grouping-based methodology proposed here. Figure 13a shows that even for this rather moderate scale reduction to 1:50,000 the mean area of polygons already more than doubled when applying size constraints generalization (from 1648.56 to 4508.57 m^2), while the grouping approach resulted only in a moderate increase in polygon area (2532.18 m^2). The change in shape after size constraints generalization was applied was also considerable (from 0.60 to 0.52), indicating increasingly elongated polygon shapes. The grouping approach, however, resulted in very similar shape measures (0.58) and thus preserved this property in the generalized map. Finally, size constraints generalization resulted also in more pronounced changes in orientation of polygons (from 83.45° to 74.54°), while the changes for the grouping approach were very small (84.54°).

4.6.2. Effects of generalization operator

Second, we focus on the grouping approach and compare the same polygon properties — area, shape, and orientation — after applying only one of the generalization operators, aggregation or typification (i.e., producing results such as in Fig. 7b or 7c). The exclusive application of the aggregation operator resulted in larger polygon areas (but still smaller than for the size constraints approach) and more elongated polygon geometries compared to the exclusive typification operator (which led to more compact shapes on average), while changes in orientation were relatively minor for both operators (Figure 13b). These results demonstrate the necessity of including the statistical evaluation step described in Section 3.4.2 in the proposed methodology.

5. Discussion

In this study, we have presented a methodology for the generalization of small polygons

in geological maps by integrating the recognition of polygon group patterns (Sayidov et al., 2020a), constraint-driven polygonal map generalization (Sayidov et al., 2020b), and the aggregation and typification of groups of polygons. The generalization of isolated, non-group-forming polygons was tackled using the selection, enlargement, aggregation, and displacement operators proposed by Sayidov et al. (2020b).

The approach proposed in Sayidov et al (2020b) was used in the group analysis stage to generate groups of polygons sharing similar properties within the range of the object separation (OS) distance. As shown in Table 1, the number of groups identified decreased and the size of groups increased with decreasing map scale. This is plausible because as the scale decreases, the OS distance, and thus the search neighborhood for nearby polygons increases, thus resulting in larger groups. Table 1 also shows that the group size distribution is highly skewed towards small groups. At 1:50,000 scale, 80.93 % of groups consist of only two members, and 99.25 % consist of five or less members, that is, groups small enough to not be further subdivided in the divide-and-conquer process of Section 3.3.3. Even at the scale of 1:200,000, still 62.68 % of groups consist of only two members, and still 90.67 % contain only five members. That is, the vast majority of polygon groups can be dealt with using the simplified process.

Looking more closely into the selection of operators, we see again a skewed picture. For instance, of the 797 polygon groups identified in the group analysis stage for the target scale of 1:50,000 using the OS distance of the ‘compromise’ constraint set, 677 or 85 % were treated with the typification operator, while only 15 % were aggregated. Aggregation is only performed if the polygons of the same group are linearly aligned within the view angle bounds (Section 3.3.4), as otherwise it has a tendency of significantly altering the area, shape, and orientation of the original polygons being aggregated. Typification, by virtue of choosing polygons as

placeholders that are most representative of the polygon group, induces less alteration. Figure 10 illustrates how this strategy of mainly relying on typification, while choosing aggregation only cautiously for linearly aligned polygons is successful in maintaining the overall spatial arrangement and structure of the original map in the generalization process. The inset maps of Figure 10 also show how isolated polygons have been dealt with by enlargement and displacement.

The capacity of the proposed methodology of successfully preserving the main characteristics of the map while ensuring its readability also extends over larger scale reductions, as shown in Figure 11, where even at 1:200,000 the main directional, shape and density characteristics are largely maintained. The proposed methodology also compares favorably to the 'size constraints' approach of Sayidov et al. (2020b), as shown in Figure 12. While the size constraints methodology performs still reasonably well in the transition to 1:50,000, it falls apart at 1:100,000 and 1:200,000, which is mainly due to two reasons: The size constraints approach does not include a group recognition facility, and it does not have a typification operator. Thus, it has to solely rely on the aggregation operator for contextual generalization, and aggregation leads to excessive area gain of those polygon categories that are considered most important.

The visual impression of Figure 12 is confirmed by the quantitative evaluation presented in Figure 13a: Because all polygons falling below the minimum area (MA) threshold have to be enlarged or removed, both the size constraints approach and the grouping-based approach lead to an increase in area; but the former leads to significantly more area gain, caused by the exclusive use of the aggregation operator in a winner-takes-all strategy (important categories are favored over less important ones). Similarly, the mean shape dropped to from 0.60 to 0.52, leading to generally more elongated polygons, while the grouping approach showed only minor changes and thus

preserves the shapes relatively well. Similarly, the orientation of polygons was altered more in the size constraints approach, while grouping kept the orientation of polygons relatively constant. Overall, it appears that the proposed grouping-based methodology is able to mitigate some of the shortcomings of the size constraints approach.

The quantitative evaluation comparing the aggregation and typification operators provided important insights on their observed effects. As expected, the area increase is much higher when only the aggregation operator is applied, compared to typification. However, applying aggregation to the grouped polygons performed better than applying it on the individual polygons (as in the size constraints approach). Shapes become more elongated when applying only aggregation and relatively more compact when only typification is utilized. Again, that effect meets expectations, since aggregation is only applied if polygons are linearly aligned in a group. Furthermore, performing only aggregation in the grouping approach preserved shape better than the size constraints approach. The orientation of polygons was only slightly affected by both operators. Overall, the typification leads to less alteration of area, shape, and orientation properties, with a result that is very similar to that of the complete proposed methodology using both operators. Hence, using typification alone could be used as a simplified strategy to generalize small polygons in categorical, polygonal maps.

Despite encouraging results, several improvements are possible. First of all, the proposed methodology involves several algorithms with multiple parameters; thus, parameter tuning is an issue if the proposed approach was to be used in practice, possibly also involving machine learning (Karsznia and Weibel 2018). Further improvements are also needed to process larger groups of polygons. As mentioned in Section 3.3.3., larger groups are subdivided by removing global and local long edges from the initial group candidates. While this edge removal accurately subdivides large

groups into sub-groups containing a smaller number of polygons based on proximity characteristics, the integration of density-based analysis in forming sub-groups should be explored to better control the number of polygons to be preserved in the generalization process. Furthermore, there is a need to integrate polygonal generalization — the focus of this paper — in the comprehensive geological map generalization process, which also includes other essential content such as linear features (e.g., contact, fault, and fold lines), making mapping results more useful for practical applications (Smirnoff et al. 2008, 2012).

6. Conclusions

A new approach to geological map generalization has been proposed which integrates pattern recognition and polygon group identification overcoming persistent limitations in the generalization of groups of small polygons in complex map documents. Group identification presents an effective way of deriving geological map properties necessary to be preserved by identifying closely-located polygons with similar characteristics, such as size, shape, orientation, and the category of the geological units involved. The proposed methodology relies on a set of generalization operators, mostly importantly aggregation and typification, for which algorithms and an operator selection strategy have been introduced. In experiments, the grouping-based methodology and associated generalization operators helped maintain essential patterns of the original map and avoid significant alterations of cartographic properties, such as area, shape, and orientation; and it outperformed a state-of-the-art methodology that uses a constraint-based approach, albeit without the capacity of group identification. The presented approach has the potential of solving similar problems in other categorical maps (e.g., soil, vegetation, or landuse maps) where the grouping of similar polygons also

represents a key requirement for generalization.

Future work may address several enhancements, including parameter tuning of the existing methodology, an improved formation and subdivision of large polygon groups using density characteristics, and integration into a comprehensive strategy of geological map generalization including also the generalization of linear and point objects.

Acknowledgments

The authors acknowledge their use of the Euriowie 1:25,000 Geological Map provided by the Department of Primary Industries -Mineral Resources, Geological Survey of New South Wales, NSW, Australia, which is used in this paper.

Funding

The research was supported by a scholarship of the Swiss Federal Commission for Scholarships (FCS) under the Grant 2015.0750.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data Availability Statement

The data that support the experiments of this study are openly available from Mining, Exploration and Geoscience NSW, Australia (www.resourcesandgeoscience.nsw.gov.au) at https://search.geoscience.nsw.gov.au/product/251.

References

Ai, T., Zhang, X., Zhou, Q., & Yang, M. (2015). A vector field model to handle the displacement of multiple conflicts in building generalization. International Journal of Geographical Information Science, 29(8): 1310-1331.

Anders, K. H., & Sester, M. (2000). Parameter-free cluster detection in spatial databases and its application to typification. International Archives of Photogrammetry and Remote Sensing, 33(B4/1; PART 4), 75–83.

Bader, M. & Weibel, R. (1997). Detecting and Resolving Size and Proximity Conflicts in the Generalization of Polygonal Maps. Proceedings 18th International Cartographic Conference (ICC), Stockholm, Sweden, pp. 1525-1532.

Bader, M., Barrault, M., & Weibel, R. (2005). Building displacement over a ductile truss, International Journal of Geographical Information Science, 19(8-9): 915-936.

Baella, B., Palomar, Vázquez, J. M., Pardo, Pascual, J. E., & Pla, M. (2007). Spot heights generalization: deriving the relief of the Topographic Database of Catalonia at 1:25,000 from the master database. In: Proceedings of the 11th ICA workshop on generalisation and multiple representation, Moscow, 2007. http://generalisation.icaci.org/images/files/workshop/workshop2007/Blanca_ICAWorkshop2.pdf.

Basaraner, M. (2011). A zone-based iterative building displacement method through the collective use of Voronoi Tessellation, spatial analysis and multicriteria decision making. Boletim de Ciências Geodésicas, 17(2): 161-187.

Brassel, K.E., & Weibel, R. (1988). A review and framework of automated map generalization. International Journal of Geographic Information Systems. 2(3): 229–244.

Brodaric, B., & Patera, A. (2001). Generalizing geological map units using GIS and tabular data. Proceedings of the IAMG - International Association for Mathematical Geology, 7th Annual Conference, Cancun, Mexico, p.1 (Abstract).

Burghardt, D., & Cecconi, A. (2007). Mesh simplification for building typification. International Journal of Geographical Information Science, 21(3): 283–298.

Burghardt, D., Duchêne, C., & Mackaness ,W. (2014) Conclusion: Major Achievements and Research Challenges in Generalisation. In: Burghardt, D., Duchêne, C.,

Mackaness, W. (eds) Abstracting Geographic Information in a Data Rich World. Lecture Notes in Geoinformation and Cartography. Springer, Cham. pp 393-403.

Chehreghan, A., & Ali Abbaspour, R. (2017). An assessment of spatial similarity degree between polylines on multi-scale, multi-source maps, Geocarto International, 32(5): 471-487.

Deng, M., Liu, Q., Cheng, T., & Shi, Y. (2011). An adaptive spatial clustering algorithm based on Delaunay triangulation. Computers, Environment and Urban Systems, 35(4): 320–332.

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. The Canadian Cartographer, 10(2): 112–122.

Downs, T., & Mackaness, W. (2002). An Integrated Approach to the Generalization of Geological Maps. The Cartographic Journal, 39: 137-152.

Edwardes, A., Mackaness, W. A. (2000). Modelling knowledge for automated generalisation of categorical maps – A constraint based approach. In Innovations in GIS 7. GIS and Geocomputation; Atkinson, P., Martin, D., Eds. Taylor & Francis: London, UK, 2000; pp. 161–174.

Foerster, T., Stoter, J. E., & Kraak, M. (2010). Challenges for automated generalisation at European mapping agencies: a qualitative and quantitative analysis. The Cartographic Journal, 47(1): 41–54.

Galanda, M. (2003): Automated polygon generalization in a multi agent system. Doctoral dissertation, University of Zurich. https://doi.org/10.5167/uzh-163108.

Galanda, M., & Weibel, R. (2002). An agent-based framework for polygonal subdivision generalization. In: Richardson D, van Oosterom P, editors. Advances in spatial data handling, 10th International symposium on spatial data handling. Berlin: Springer; p. 121–136.

Galanda, M., & Weibel, R. (2003). Using an energy minimization technique for polygon generalization. Cartography and Geographic Information Science, 30(3): 263–275.

Gary, R. H., Wilson, Z. D, Archuleta, C-A. M., Thompson, F. E., & Vrabel, J. (2010). Production of a national 1:1,000,000-scale hydrography dataset for the United States—feature selection, simplification, and refinement. Scientific Investigation Report 2009-5202, U.S. Department of Interior, U.S. Geological Survey.

Haunert, J.H., & Sester, M. (2008). Area collapse and road centerlines based on straight skeletons. Geoinformatica, 12(2): 169–191.

He, X., Zhang, X., & Xin, Q. (2018). Recognition of building group patterns in topographic maps based on graph partitioning and random forest. ISPRS Journal of Photogrammetry and Remote Sensing, 136: 26-40.

Karsznia, I., & Weibel, R. (2018): Improving settlement selection for small-scale maps using data enrichment and machine learning. Cartography and Geographic Information Science, 45(2): 111-127.

Lecordix, F., Lemarié, C. (2007). Managing generalisation updates in IGN map production. In: Mackaness WA, Ruas A, Sarjakoski LT (eds) Generalisation of geographic information: models and applications, Chapter 15. Elsevier, pp. 285–300.

Li, C., Liu, X., & Guo, P. (2017). Building Aggregation Method Combining Constrained Delaunay Triangulation and a Filtering Process, 2nd International Conference on Manufacturing Science and Information Engineering (ICMSIE 2017), 15–21.

Li, Z., Yan, H., Ai, T., & Chen, J. (2004). Automated building generalization based on urban morphology and gestalt theory. International Journal of Geographical Information Science, 18(5): 513-534.

Liu, D., Nosovskiy, G., & Sourina, O. (2008). Effective clustering and boundary detection algorithm based on Delaunay triangulation. Pattern Recognition Letters, 29(9): 1261– 273.

Lu, Y., & Thill, J. C. (2008). Cross-scale analysis of cluster correspondence using different operational neighborhoods. Journal of Geographical Systems, 10(3): 241–261.

Mackaness, W.; Purves, R. Automated Displacement for Large Numbers of Discrete Map Objects. Algorithmica, 30: 302–311.

Mackaness, W. A., Perikleous, S., & Chaudhry, O. Z. (2008). Representing Forested Regions at Small Scales: Automatic Derivation from Very Large Scale Data. The Cartographic Journal, 45(1): 6–17.

Mackaness, W. A., Ruas, A., & Sarjakoski, L. T. (2007). Generalization of geographic information: Cartographic modelling and applications. Series of international cartographic association, Elsevier. ISBN 978-0-08-045374-3.

McCabe, C. 2008. Vector approaches to generalizing faults and polygons in 1:24,000 geologic maps: Santa Rosa, California, case study. http://www.geovista.psu.edu/publications/2008/CMcCabe_GeologicMapGeneralization.pdf.

McMaster, R. B., Shea, K. S. (1992). Generalization in digital cartography. Washington, DC: Association of American Geographers.

Moreira, A., & Santos, M. Y. (2007). Concave hull: A k-nearest neighbors approach for the computation of the region occupied by a set of points. In Proceedings of the Second International Conference on Computer Graphics Theory and Applications, Algarve, Portugal, 5–7 March 2011; pp. 61–68.

Osserman, B. (1978). The isoperimetric inequality. Bulletin of the American Mathematical Society, 84(6): 1182–1238.

Patera, A. (2006). Advanced Techniques for the Management of Geological Mapping. Ph.D. Thesis. The Open University, Milton Keynes, UK, 195 pp.

Perikleous, S. (2006). Automatic Derivation of Forest Patches for Representation at 1:250.000 directly From OS MasterMap. MSc Thesis, University of Edinburgh.

Peter, B. (2001). Measures for the Generalization of Polygonal Maps with Categorical Data. Fourth ICA Workshop on Progress in Automated Map Generalization. Beijing, pp. 1-21. https://kartographie.geo.tu-dresden.de/downloads/ica-gen/workshop2001/Peter_v2.pdf

Peter, B., & Weibel, R. (1999). Using Vector and Raster-Based Techniques in Categorical Map Generalization. Third ICA Workshop on Progress in Automated Map Generalization. Ottawa.

Regnauld, N. (2001). Contextual Building Typification in Automated Map Generalization. Algorithmica, 30: 312–333.

Regnauld, N. (2003). Algorithms for the amalgamation of topographic data. In: Proceedings of the 21st International Cartographic Conference (ICC), Durban, South Africa, pp. 222-234.

Regnauld, N. (2005). Spatial structures to support automatic generalization. In: Proceedings of the 22nd International Cartographic Conference (ICC), La Coruña, Spain, pp. 2-3. La Coruna.

Regnauld, N. (2011). OS VectorMap District: automated generalisation, text placement and conflation in support of making pubic data public. In: Proceedings of the 25th International Cartographic Conference (ICC), pp.-2-5, Paris, 3–8 July 2011.

Regnauld, N., & McMaster, R. B. (2007). A synoptic view of generalisation operators. In: Mackaness WA, Ruas A, Sarjakoski LT (eds) Generalisation of geographic information. Elsevier Science B.V, Amsterdam, pp 37–66.

Regnauld, N., Revell, P. (2007). Automatic Amalgamation of Buildings for Producing Ordnance Survey® 1:50 000 Scale Maps, The Cartographic Journal, 44(3): 239-250.

Ruas A, Holzapfel F. 2003. Automatic characterisation of building alignments by means of expert knowledge. In: Proceedings of the 21st International Cartographic Conference (ICC), Durban, South Africa, pp. 1515–1604.

Ruas, A. (1998). A method for building displacement in automated map generalization, International Journal of Geographical Information Science, 12(8): 789-803.

Sayidov, A., & Weibel, R. (2016). Constraint-based approach in geological map generalization. In: 19th ICA Workshop on Generalization and Multiple Representation, Helsinki (Finnland), 14 June 2016 - 14 June 2016, online.

Sayidov, A., Aliakbarian, M., & Weibel R. (2020b). Geological Map Generalization Driven by Size Constraints. ISPRS International Journal of Geo-Information. 2020(9) 284.

Sayidov, A., Weibel, R., & Leyk, S. (2020a). Recognition of group patterns in geological maps by building similarity networks. Geocarto International, online first, 1–20.

Schylberg, L. (1993). Computational Methods for Generalization of Cartographic Data in a Raster Environment [Ph.D. thesis]. Stockholm: Department of Geodesy and Photogrammetry, Royal Institute of Technology.

Seo, S., Gary, M., & Marsh, D. (2006). A review and comparison of methods for detecting outliers in univariate data sets. Department of Biostatistics. Graduate School of Public Health. 1–53. Retrieved from http://d-scholarship.pitt.edu/7948/.

Smirnoff, A., Huot-Vezina, G., Paradis, J., & Boivin R. (2012). Generalizing geological maps with the GeoScaler software: the case study approach. Computers and Geoscience, 40: 66–86.

Smirnoff, A., Paradis, J., & Boivin, R. (2008). Generalizing surficial geological maps for scale change: ArcGIS tools vs. cellular automata model. Computers and Geoscience. 34(11): 1550–1568.

Stanislawski, L. V., & Buttenfield, B. P. (2010). Hydrographic feature generalization in dry mountainous terrain. In Proceedings of the Special Joint Symposium of ISPRS Technical Commission IV & AutoCarto in Conjunction with ASPRS/CaGIS 2010 Fall Specialty Conference, Orlando, FL, USA (pp. 15-19).

Steiniger, S., & Weibel, R. (2005). A conceptual framework for automated generalization and its application to geologic and soil maps. International Cartographic Conference (ICC), La Coruña, Spain 4-5. pp. 11-16.

Steiniger, S., Burghardt, D., & Weibel, R. (2006). Recognition of island structures for map generalization. In Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems (GIS '06). Association for Computing Machinery, New York, NY, USA, 67–74.

Steiniger S., & Weibel R. (2007). Relations Among Map Objects in Cartographic Generalization. Cartography and Geographic Information Science, 34(3): 175-197.

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.

Su, B., Li, Z., & Lodwick, G. (1998). Morphological models for the collapse of area features in digital map generalization. GeoInformatica, 2(4): 359–382.

Su, B., Li, Z., Lodwick, G., & Muller, C. (1997). Algebraic Models for the Aggregation of Area Features Based upon Morphological Operators. Cartography and Geographic Information Science, 11(3): 233–246.

Töpfer, F., & Pillewizer, W. (1966). The principles of selection. The Cartographic Journal, 3: 10–16.

Vernon, R.H. and Clarke, G.L. (2008). Principles of metamorphic petrology. Cambridge University Press, 460pp

Wang, X., & Burghardt, D. (2019). A Mesh-Based Typification Method for Building Groups with Grid Patterns. ISPRS International Journal of Geo-Information, 8, 168.

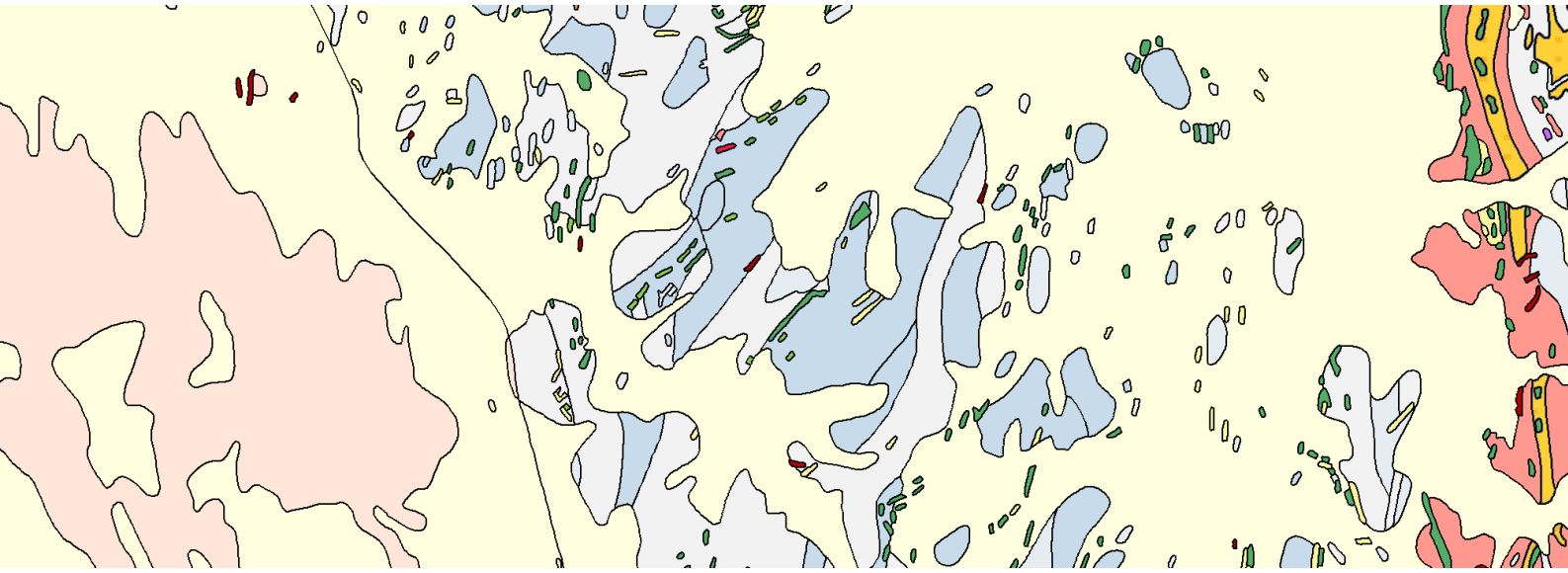
Ware, J. M., Jones, C. B., & Thomas, N. (2003). Automated map generalization with multiple operators: A simulated annealing approach. International Journal of Geographical Information Science, 17(8): 743–769.

Wilmer, J. M., & Brewer, C. A. (2010). Application of the radical law in generalization of national hydrography data for multiscale mapping. Paper presented at the Special Joint Symposium of ISPRS Technical Commission IV & AutoCarto in Conjunction with ASPRS/CaGIS 2010 Fall Specialty Conference, Orlando, FL, USA. pp.-1-3.

Yan, H. (2010). Fundamental theories of spatial similarity relations in multi-scale map spaces. Chinese Geographical Science 20: 18–22.

Yan, H. (2014). Theory of Spatial Similarity Relations and Its Applications in Automated Map Generalization. PhD Thesis, Department of Geography, University of Waterloo, http://hdl.handle.net/10012/8317.

Yaolin, L., Molenaar, M., & Kraak, M. (2002). Semantic similarity evaluation model in categorical database generalization. In: International Archives of Photogrammetry and Remote Sensing, 34: 1-4.



5

Chapter

You can't go back and change the beginning, but you can start where you are and change the ending.

C. S. Lewis

5. Synthesis

This thesis presented an integrated methodology for geological map generalization, with a focus on small island polygons, consisting of three parts. First, we implemented constraint-driven generalization to generalize island polygons in the geological map. Then, we addressed pattern recognition and identification of polygon groups in the geological map. Lastly, we utilized an adaptive operator selection to achieve the generalization of groups of polygons.

Section 5.1 summarizes the main results of the thesis relative to its research objectives. The adaptability of the approach to the generalization of other categorical map types is discussed in Section 5.2. Finally, Section 5.3 presents an outlook on directions for future research.

5.1. Discussion of Research Objectives

In this section, we summarize, for each of the research objectives laid out in Section 1.3.2, the contributions and main findings, as well as the limitations.

5.1.1. Research Objective 1: Separate and Generalize Island Polygons in Geological Maps by Incorporating Size Constraints.

Contributions

In Chapter 2 (Research Paper 1), we proposed a methodology for constraint-driven generalization that uses size constraints (i.e., constraints dealing with minimum area and distance relations in individual or pairs of map features) to generalize small island polygons in geological maps. We provided definitions for four constraints: minimum area, object separation, outline granularity, and consecutive vertices. We also proposed a set of goal values for these constraints, which strike a compromise between published goal values that appear to be too fine-grained for the purpose of geological maps (Spiess et al., 2005) or conversely too coarse-grained (FGDC, 2006).

We introduced a procedure to separate foreground (i.e., small island) polygons from background polygons. We then demonstrated how two constraints, minimum area and object separation, can trigger appropriate generalization operators — elimination/selection, enlargement, aggregation, and displacement — to resolve constraint violations among the foreground polygons, and we proposed algorithms for those operators. We thus demonstrated that size constraint-based generalization can resolve the main legibility problems of small polygons in a step-by-step manner. Responding to the weaknesses of traditional selection by the Radical Law (Töpfer & Pillewizer, 1966), we developed two new selection algorithms, termed ‘area loss-gain selection’ and ‘category-wise selection’,

which better maintain the balance of area proportions among the different categories present in the map.

Findings

- The methodology based on size constraints resolves the main legibility problems associated with small polygons and produces rather convincing results up to a scale reduction of approximately 50% (i.e., from 1:25,000 to 1:50,000), where it outperformed the cellular automata approach by Smirnoff (2008, 2012), which can be seen to represent the state-of-the-art of solutions for geological map generalization (Chapter 2, Figure 17). However, it breaks down for larger scale transitions (Chapter 2, Figure 16).
- Of the three proposed selection methods 'category-wise selection' clearly performed best. However, none of the methods can guarantee the preservation of contextual patterns of polygons, such as polygon groups, especially for larger-scale transitions.
- The 'Radical Law selection' (Töpfer & Pillewizer, 1966) has the most disfiguring effect on polygonal groups. Thus, we do not recommend using it for polygonal map generalization unless the purpose of the map is to portray an overgeneralized view of mapped events.
- While elimination generally can destroy the patterns of the spatial arrangement of polygons on the map, extensively removing a particular category of small polygons can cause a disbalance of area proportions of different categories.
- The use of the aggregation operator during generalization can lead to considerable alteration of the map image, particularly when the scale reduction is significant. Aggregation not only tends to modify the shape of polygons and but even more importantly can contribute to disproportional gain of area for each aggregated rock unit.

Limitations

When applying the category-wise selection, the generalization produced legible maps for smaller-scale shifts. However, when the scale shift increases, visual imbalances occur on the map. Especially for a scale of 1:200.000 (i.e., an 8-fold scale reduction), the relative area per class becomes significantly imbalanced. This is mainly due to the removal of small-sized polygons in large quantities belonging to categories assigned lower importance. Moreover, the area gained after enlargement and aggregation operators resulted in disproportional area gain by categories representing geological units that were given high importance. Conclusively, these drawbacks — the imbalanced elimination of polygons and the disproportional area gain of particular categories — suggest identifying polygons groups in the geological map before generalization. Recognizing

groups will help preserve the map's main character, such as the distributions of area, shape, and orientation of the polygons involved.

5.1.2. Research Objective 2: Identify Groups of Small Polygons Based on Similarity Criteria, Forming Important Spatial Structures, as a basis for Implementing Contextual Generalization Operators.

Contributions

In Chapter 3 (Research Paper 2), we proposed an approach that identifies polygon groups among foreground polygons based on similarities, in order to preserve the geological map's general structural characteristics. This process results in homogenous groups (with respect to the member polygons' similarity in category, size, shape and orientation) that can be subsequently be used to inform contextual generalization operators (e.g., aggregation, typification). The main criterion for detecting candidate group patterns is the proximity between polygons, which is analyzed on the basis of an initial Delaunay triangulation of the centroids of foreground polygons, followed by the removal of global and local long edges (Deng et al. 2011). These candidate groups are then further refined using more specific criteria such as the size, shape, orientation, and category of polygons to obtain homogenous polygon groups. The identified groups contain a wealth of information that can be exploited in contextual generalization algorithms, including the attributes of the member polygons (category, area, shape, orientation), neighborhood relations (encoded in the connecting network edges), and proximity (i.e., separating distance) between neighboring polygons.

Findings

- Applying Tukey's outlier detection approach, we successfully separated foreground polygons from background ones in the geological map considering areas of polygons.
- Our approach robustly identifies meaningful groups of polygons, considering the polygons' intrinsic properties.
- A statistical analysis of the identified groups revealed that the homogeneity of group patterns has significantly improved after the refinement step is accomplished compared to the initial candidate group formation step (after removal of global and local long edges).

Limitations

Overall, our methodology successfully identified relevant groups of polygons in geological maps. However, there is still room for improvement. First, in utilizing alternative options for the individual algorithms (e.g., for measuring shape or orientation) involved and in fine-tuning parameters and thresholds. Second, while the proposed methodology encodes a lot of information at

the level of individual polygons and network edges connecting group members, the information stored at the level of the polygon group is currently limited to simple metrics such as averages of shape, size, or orientation, while other cartographically meaningful properties such as density (and density variation) could also be extracted.

Tukey's outlier detection process robustly delineates foreground from background polygons. However, it might not always be desirable to separate all features in a map, which may affect the spatial organization between small and large polygons, although the size of polygons may differ largely.

5.1.3. Research Objective 3: Develop Contextual Generalization Algorithms and Evaluation of Generalization Results of Polygon Groups in Geological Maps.

Contributions

Chapter 4 (Research Paper 3) proposed a methodology that integrates aggregation and typification to generalize polygon group patterns in geological maps, formed by foreground polygons. Knowledge gained in the preceding group identification procedure (Chapter 3) is used to assist the operator selection process and generalize representative groups of foreground polygons. For foreground polygons participating in groups, the methodology reuses the aggregation algorithm proposed in Chapter 2, but proposes a new algorithm for typification. Based on the concept of the 'view angle' the selection between aggregation and typification is achieved to best maintain a polygon group's characteristics during generalization. As the two operators have complementary effects — aggregation leads to area gain, typification slightly loses area — they are used in combination to achieve best generalization results. For isolated foreground polygons that are not part of polygon groups, the size constraints generalization process introduced in Chapter 2 is used. We statistically evaluated our results and compared the statistics of shape, size, and orientation distributions of the polygons in the source and target maps at different scales and noted that the group-based generalization methodology clearly outperforms the approach described in Chapter 2.

As explained in Section 1.2.2, there are different types of geological maps, serving different purposes, and hence using different scale ranges. The proposed methodology can be used to produce medium-scale — i.e., *regional* scale — geological maps by generalization from detailed or large-scale geological maps. In other words, for making the transition from a scale range of about 1:5,000 to 1:25,000 used to compile the geological source maps, to a scale range of about 1:50,000 to 1:250,000. The exact scale of, and thus area covered by, a regional geological map depends very much on the intended usage, such as regional geological surveys, mineral exploration, mining, or large construction projects.

Although Section 1.2.2 has mentioned five types of geological maps in different scale ranges, the main geological operation targeted by this work — mineral exploration and mining —mainly utilizes maps in two scale ranges (Haldar, 2018): at the regional scale (~ 1:50,000 – 1:250,000) and at the detailed scale (~ 1:1,000 – 1:10,000), corresponding to the scales covered by the experiments of Chapters 2 to 4 (i.e., 1:25,000 – 1:200,000). The purpose of geological maps at the regional scale is to gain an overview of a larger area that is under investigation, researching issues such as the prospectivity of that region, or regional scale structures, such as a fold belt (Haldar, 2018). This will then allow to narrow down the area of interest to locations that seem particularly promising, which will then again be explored on maps of more detailed scales (e.g., 1:5,000, 1:10,000), where the map contains maximum information, including the extent of the mapped mineral deposit and special geological features of interest for that particular deposit (Haldar, 2018; Marjoribanks, 2010).

In other words, while the detailed scale maps usually already exist at the data compilation scale, facilities are required to produce regional geological maps by map generalization, and that is what Chapter 4 (Research Paper 3) has contributed to.

Findings

- Group identification has a significant impact on the generalization results (i.e., comparing the results of Chapters 2 and 4, respectively). Identifying group patterns can preserve the individual properties of generalized polygons and make them readable during generalization, while better preserving the statistics of the size, shape and orientation distributions of retained polygons, as well as maintaining an improved balance of area loss and gain across the different categories displayed on the generalized map (as opposed to the approach presented in Chapter 2).
- The preceding grouping identification approach (Chapter 3) allows to preserve the overall spatial arrangement of polygons on the map after generalization, by enabling truly contextual generalization operators, including aggregation and typification, while the size constraints approach of Chapter 2 only allowed pairwise operations.
- Aggregation and typification operators can be used to generalize groups of polygons while largely maintaining the properties of groups.
- Of the two operators, aggregation causes considerably more alteration of size and shape than typification does. Aggregation should only be used for polygon that are linearly aligned (i.e., within a small ‘view angle’ deviation from each other). Hence, in our experiments, only 15 % of cases were treated by aggregation.
- If only using a single operator, typification is clearly to be preferred.

Looking more specifically into the workings of the two main generalization operators proposed in Chapter 4 — aggregation and typification — Figure 5.1. and Figure 5.2 visualize two cases that highlight how each generalization operator performs and changes the size, shape and orientation of polygon groups. Figure 5.1a shows an initial situation and three main properties of the original polygons: their size in m^2, orientation in degrees, and the shape index (IPQ). Figure 5.1b then depicts the result after aggregation (red polygon outline), with the original polygons retained (in green) for comparison. The aggregated polygon has an area of 1,344 m^2, much larger than the sum of the areas of the original polygons (which comes to 459 + 358 = 807 m^2). This area gain of 66 % is primarily an effect of the side-by-side arrangement of the two original polygons, such that the space filled in by the aggregation is maximal (i.e., dominated by the long sides of the original polygons). The orientation of the aggregated polygon changes even more dramatically, from 6° and 8°, respectively, for the two original polygons, to a value of 118°, that is, basically a rotation to a perpendicular orientation. Finally, the shape of the original polygons (IPQ = 0.68, IPQ = 0.735) also changed considerably to a value of IPQ = 0.903, which is approaching the IPQ value of 1.0 for a circle. Clearly, the aggregation operator does not perform well with side-by-side arrangements of polygons, which is why the concept of the ‘view angle’ was introduced in the operator selection step, in order to avoid aggregation in such cases.

In contrast, as Figure 5.1c shows, typification preserved the relative orientation of polygons for this side-by-side arrangement: 7° for the typified polygon, cpmpared to 6° and 8° for the two original polygons. Likewise, the shape is preserved in the typification (IPQ = 0.68), since one of the original polygons is used as a placeholder. The polygon area, however, has been reduced in this example, because the typified placeholder polygon was enlarged only to a size value corresponding to the goal value for the minimum area (MA) constraint. Thus, the area of the typified polygon is now 625 m^2, which is smaller than the sum of the areas of the original polygons, 807 m^2. Of course, this effect of area loss could be countered by enlarging the typified polygon to the size that corresponds to the summed area of the original polygons. The solution of enlarging to the MA goal value, however, was chosen in order to minimize the overall area gain of foreground polygon categories.

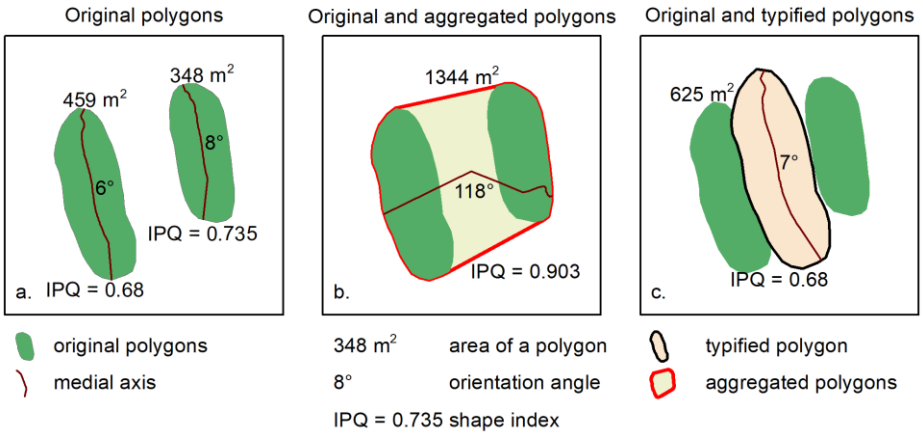


Figure 5.1. Effects of the two main generalization operators, aggregation and typification, on groups of polygons: Case 1 (side-by-side arrangement).

Figure 5.2 depicts a different case of polygons aligned in a linear arrangement. This arrangement represents the typical case in which the aggregation operator is selected, as the original polygons are positioned within the cone of the view angle from each other. As Figure 5.2b shows, the aggregation operator leads to an area gain of only 44 % (to 1054 m^2), while the orientation is preserved very well. On the other hand, the shape index does change rather considerably towards a more elongated shape (0.411 for the typified polygon, compared to 0.747 and 0.757 for the original polygons), owing to the 'concatenation' of the two linearly arranged polygons. Since no rotation takes place, however, the 'stretching' of the polygon shape is certainly tolerable.

As shown in Figure 5.2c, the area of the typified polygon comes to 625 m^2, which is smaller than the cumulative area of the two original polygons (733 m^2) would be. As explained above, this is due to the fact that the placeholder polygon is always only scaled to the goal value of the MA constraints, in order to avoid excessive area gain of the foreground polygon categories. Nevertheless, both the orientation and the shape are perfectly preserved. This may lead to the conclusion that it would be sufficient, or even preferable, to use exclusively the typification operator. This finding is further supported by the results shown in Figure 13 of Chapter 4. However, since the current typification algorithm always uses one of the original polygons of a group to represent the group as a placeholder, using only typification might lead to a more monotonous appearance of the generalized polygons, despite the better statistical performance compared to the aggregation operator. As the experiments of Chapter 4 have shown, the aggregation operator is only selected in a small minority (i.e., 15%) of cases, and hence contributes to slightly greater variation of sizes, orientations, and shapes of the generalized polygons.

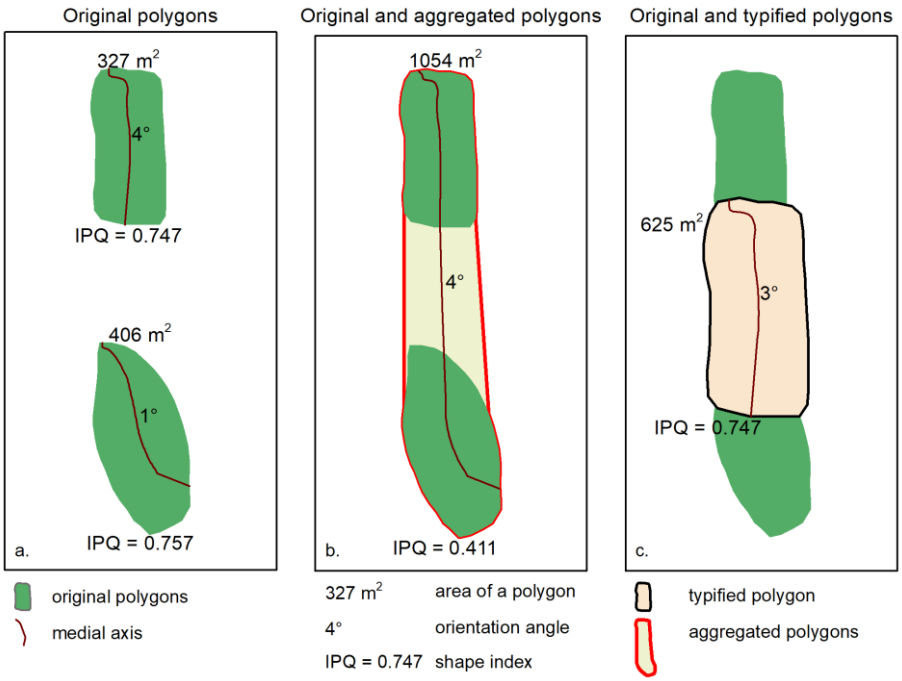


Figure 5.2. Effects of the two generalization main operators, aggregation and typification, on groups of polygons: Case 2 (linear arrangement).

Re-using Figure 10 from Chapter 4, Figure 5.3 highlights some specific examples of the performance of the proposed methodology. The successful examples (Cases 1, 2, 3, and 6) are labeled in black font, while the failures are labeled in red font and will be discussed further below. In general, combining the advantages of the two generalization operators minimizes the alterations induced by the typification and aggregation operators. For example, Cases 1 and 3 show how both simple polygon pairs (Case 3) as well as groups with several members (Case 1) can be successfully typified, while the size relations as well as orientations and shapes of the original polygons were largely preserved. Likewise, Case 2 depicts two examples of a successful aggregation of linearly arranged polygon pairs, while Case 6 demonstrates that this even possible of an alignment of four polygons. Case 6 in particular also illustrates that it would have been hard to obtain the same shape by using the typification operator.

Figure 5.4. attempts to illustrate how well the general structures of the map were maintained during the scale transitions from 1:25,000 to 1:200,000. To that end, on the right hand side of the figure lines were manually drawn to link polygons that extend over larger linear sequences across the map. Furthermore, outlines were drawn denoting the apprimate boundaries of larger complexes formed by such linear structures. While the proposed methodology only works at the level of locally formed polygon groups, we can actually see that some larger structures, such as alignments extending over many polygons not belonging to the same polygon groups, are ‘emerging’ as a result of the generalization of local polygon groups. So, overall, these larger structures are maintained rather well in the generalization process, at least to the scale of about 1:100,000.

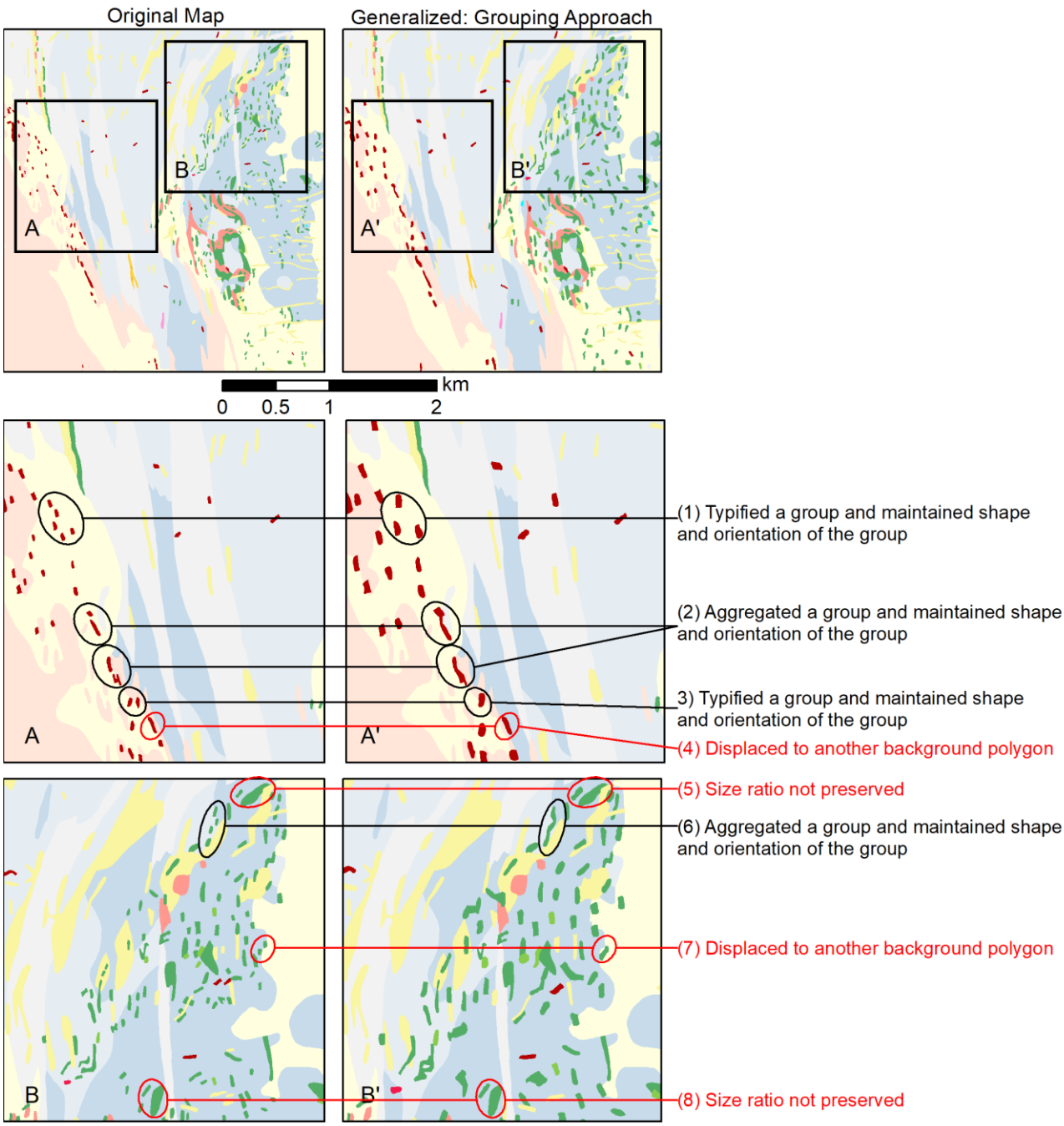


Figure 5.3. Examples of cases where the proposed methodology performed well (black font) and cases where it failed (red font).

Limitations

Overall, we showed that aggregation and typification, if based on group identification, can preserve the map's key structures and distribution characteristics during generalization. However, there are several shortcomings that still need to be addressed. As reported in Chapter 4, the size of the identified groups can vary significantly, in our example from 2 to 33 polygons, especially when the scale transition is large (1:200,000 in our case). The approach further divides large groups into smaller subdivisions based on proximity. However, when a group contains many polygons, its boundary may overlap with that of other groups, which may cause further conflicts that need to be addressed. Moreover, small polygon groups can be entirely contained by a large group and may be lost in the generalization process. Likewise, since the proposed aggregation and typification algorithms operate independently of their background, the resulting generalized foreground polygons may end up in different background polygons with different geological units, leading to topology violations.

Figure 5.3 shows some of the problematic cases, labeled in red font, where the methodology of Chapter 4 needs further refinement, such as in Cases 5 and 8, where the local size ratio of polygons is not preserved, as a consequence of the fact that polygons larger than the MA goal value are not further enlarged, while smaller polygons may be enlarged quite considerably, relatively speaking. Future work will have to distribute the area enlargement more continuously across the different polygons (in particular the small and medium-sized ones), rather than using the current dichotomous approach of enlarging small polygons while not touching all others. Cases 4 and 7 show topology violations: foreground polygons have been shifted to different background polygons, as a consequence of the typification operator currently not considering the containment relation of foreground polygons in background polygons, when displacing the placeholder polygons to a new location. Future work will have to use the boundaries of background polygons as constraints to the placement of placeholder polygons.

Furthermore, while the concept of the 'view angle' seems to be effective in preserving alignments in aggregation operations, it proceeds by a pairwise 'concatenation'. As mentioned above, we can see in Figure 5.4 that most larger structures emerge as a consequence of such local aggregation and typification operations, but at smaller scales (1:100,000 or 1:200,000), a more holistic approach of finding alignments (e.g., Wang & Burghardt, 2017) would be required. And finally, there is still plenty of room for better addressing the generalization of large background polygons, for instance, by implementing the constraints defined in Galanda (2003), who proposed more advanced algorithms for large polygons, such as the algorithm for partial displacement and exaggeration (Galanda and Weibel, 2003).

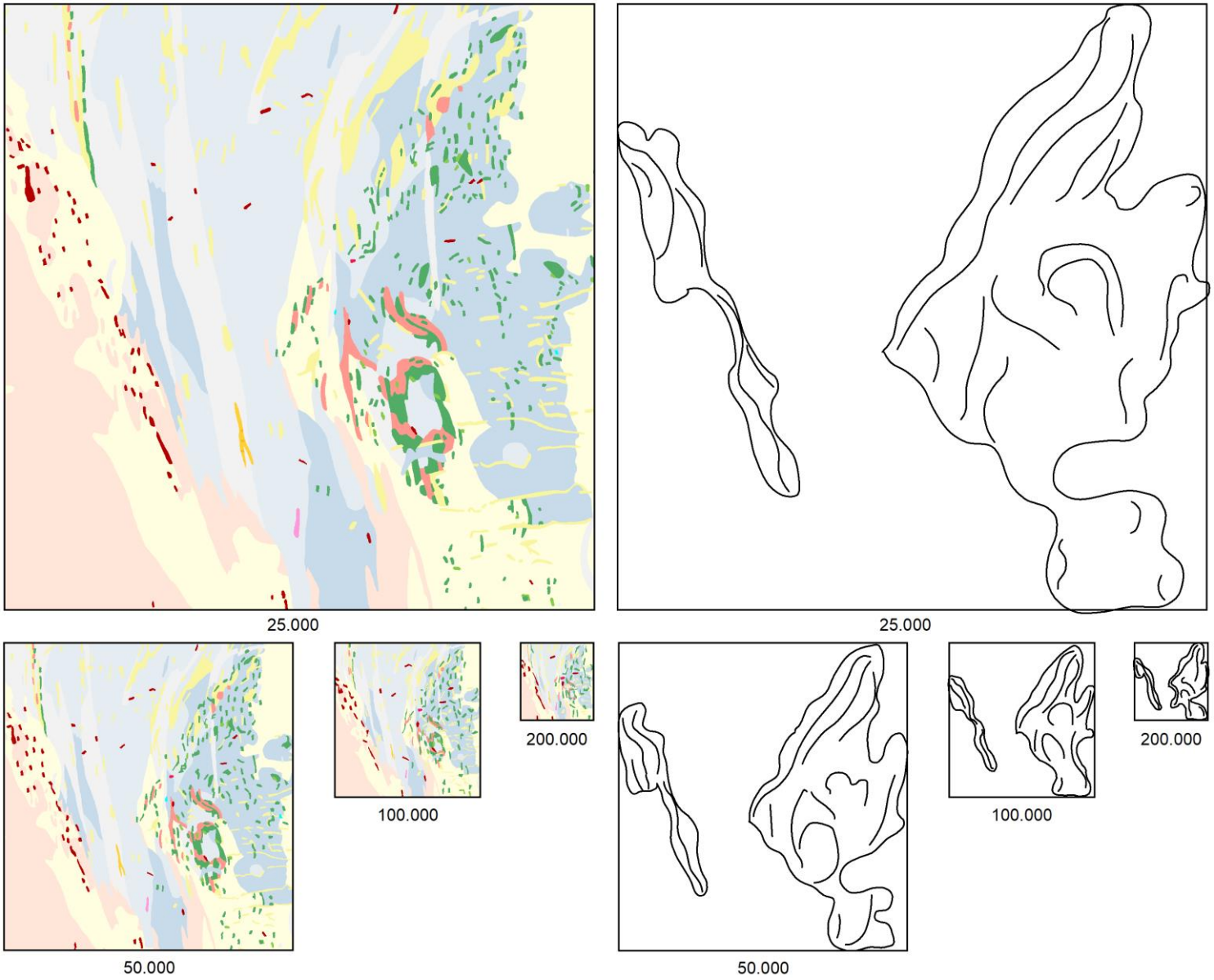


Figure 5.4. Assessing maintenance of general picture of the generalized geological maps, outlining and connecting linked polygonal patterns in the geological maps.

5.2. Adaptability of the Approach to other Categorical Maps

Geological maps belong to the group of categorical maps, that is, maps entirely covered by polygonal features with no overlaps and holes. Categorical maps show the variation of a single, categorical (i.e., nominal scale) variable by a finite number of discrete categories, such as geological units, soil types, vegetation communities, land use, or land cover types (Jaakkola, 1997; Galanda, 2003). In a geometrical sense, the unique characteristic of categorical maps is their diversity in shapes and structures, which may often vary in arbitrary ways, and are thus difficult to describe analytically and challenging to formalize during generalization.

The overall, integrated methodology put forward in this thesis 1) defines generalization constraints related to minimum areas and distances to keep the map objects and thus the entire map readable; 2) identifies meaningful groups to preserve the main structures and spatial arrangements of a map in the generalization process; and 3) implements an adaptive selection of contextual generalization operators based on the properties of identified groups and their geometric alignment. The implementation of the overall methodology is relatively easy using standard geometry and GIS libraries. It is robust in group pattern identification. And it can preserve the properties of the original polygon groups in the generalization process.

Since it is modular, the proposed integrated methodology is flexible (e.g., algorithms could be exchanged, constraints could be changed or added) and thus has the potential of being adapted to generalize other types of categorical maps, such as soil or vegetation maps, as they share common properties. Figure 5.5. shows four examples of categorical maps: a geological map, a soil map, a vegetation map, and a land use / land cover (LULC) map. The foreground polygons extracted using the method of Chapter 2 are highlighted, hence revealing similarities and differences between these map types. For instance, the foreground polygons of the geological, soil, and vegetation map are generally small and of similar sizes and shapes across these three map types, though the vegetation map contains only few island polygons. Thus, the generalization of these maps could be approached using similar steps as for geological maps: we could define and apply constraints to maintain readability, identify groups of similar polygons and other essential patterns, and implement operator selection and sequencing based on group properties. Of course, some algorithms and parameters would need to be adjusted to implement our approach. Defining suitable size constraints would be required because different maps are controlled by different requirements, and most probably, additional topological constraints would have to be implemented since containment relations (i.e., island polygons) are less frequent than in geological maps. Finally, we would need to rank

categories by their importance to establish an importance hierarchy, which will require knowledge about the particular application domain.

In contrast, LULC maps have a more regular distribution of small polygons, which conceptually cannot form a group of similar polygons. Also, adjacency relations are important. Because polygons in LULC maps are located adjacent to each other, the extraction of foreground polygons results in mostly connected and continuous polygons (Figure 5.5). This will not allow recognizing a group in this collection of polygons using the approach of Chapter 3, as most polygons are already connected. Thus, our approach cannot be applied to generalize these maps. Instead, methods based on optimal hierarchical aggregation could be used (Haunert & Wolff, 2010; van Oosterom et al., 2014; Meijers et al., 2016).

5.3. Future Directions and Outlook

In Section 5.1, a variety of limitations were identified regarding the methods proposed in the three research papers of this PhD thesis, prompting corresponding directions for future research.

In the first research paper, we proposed an approach that implements size constraints to control and manage geological map generalization. However, we did not focus on the large and bulky background polygons. We did not define constraints for the generalization of background polygons, such as the extended constraint set by Galanda (2003). Moreover, we did not consider maintaining topological relations between foreground and background polygons, such as containment relations. Since the readability of large polygons is not an issue the background polygons can be largely generalized using outline simplification and smoothing operators. However, further operators and algorithms are required. For instance, since large polygons may also exhibit narrow parts, they might need to be widened by an exaggeration operator (Galanda & Weibel, 2003).

The second research paper identified groups in a geological map based on criteria including proximity, size, shape, orientation, and the similarity of the rock units of polygons. However, because the main limiting criterion in the initial stage of the group pattern detection is proximity, the proposed approach maintains a rather local perspective, incapable of detecting patterns that extend over larger sequences or alignments of ‘linked’ polygons, possibly extending across intertwined layers of rock from different geological periods. Hence, future work needs to consider the inclusion of polygon alignments in the group identification process. Geological units are distributed over the two-dimensional map. However, almost all polygons representing the geological units are part of a larger underground body and can potentially be linked together. Thus,

the inclusion of polygon alignments as a criterion could preserve the links between polygons, further improving map generalization quality.

The third research paper focused on the generalization of the identified groups. Our approach either aggregates or typifies polygons based on their orientation relative to each other. In our experiments, the maximum group size was 15 polygons for a scale of 1:50.000, 22 polygons for a scale of 1:100.000, and 33 polygons for a scale of 1:200.000. We proposed breaking down large groups into smaller ones using a proximity-based approach. This approach could be enhanced with a density analysis. The density analysis would indicate the locations where the likelihood of conflicts and violation of constraints is highest. Also, the current aggregation and typification algorithms ignore topological and metric relations with the background polygons, such as containment relations or where an island polygon may be positioned in a background polygon, that is, more towards its center or rather towards its periphery (Weibel, 1996).

This research does not provide a complete solution for geological map generalization. It mainly focuses on the map foreground, the small island polygons, and aims to preserve them during generalization. However, geological maps contain several layers of data that need to be considered beyond polygons, such as linear features representing geological boundaries, faults, and folds. Moreover, geological maps contain other map symbols, such as mine works and cultural and topographic features. These layers also have to be considered when developing a comprehensive automated methodology for geological map generalization.

In this thesis, the results were only evaluated by quantitative measures and by comparison against baseline methods. However, geological maps belong to a specific type of thematic maps, which require particular domain knowledge. It is thus recommended to evaluate the generalization results involving expert geologists as well as expert cartographers to ensure that the quality of resulting maps corresponds to geological mapping standards.

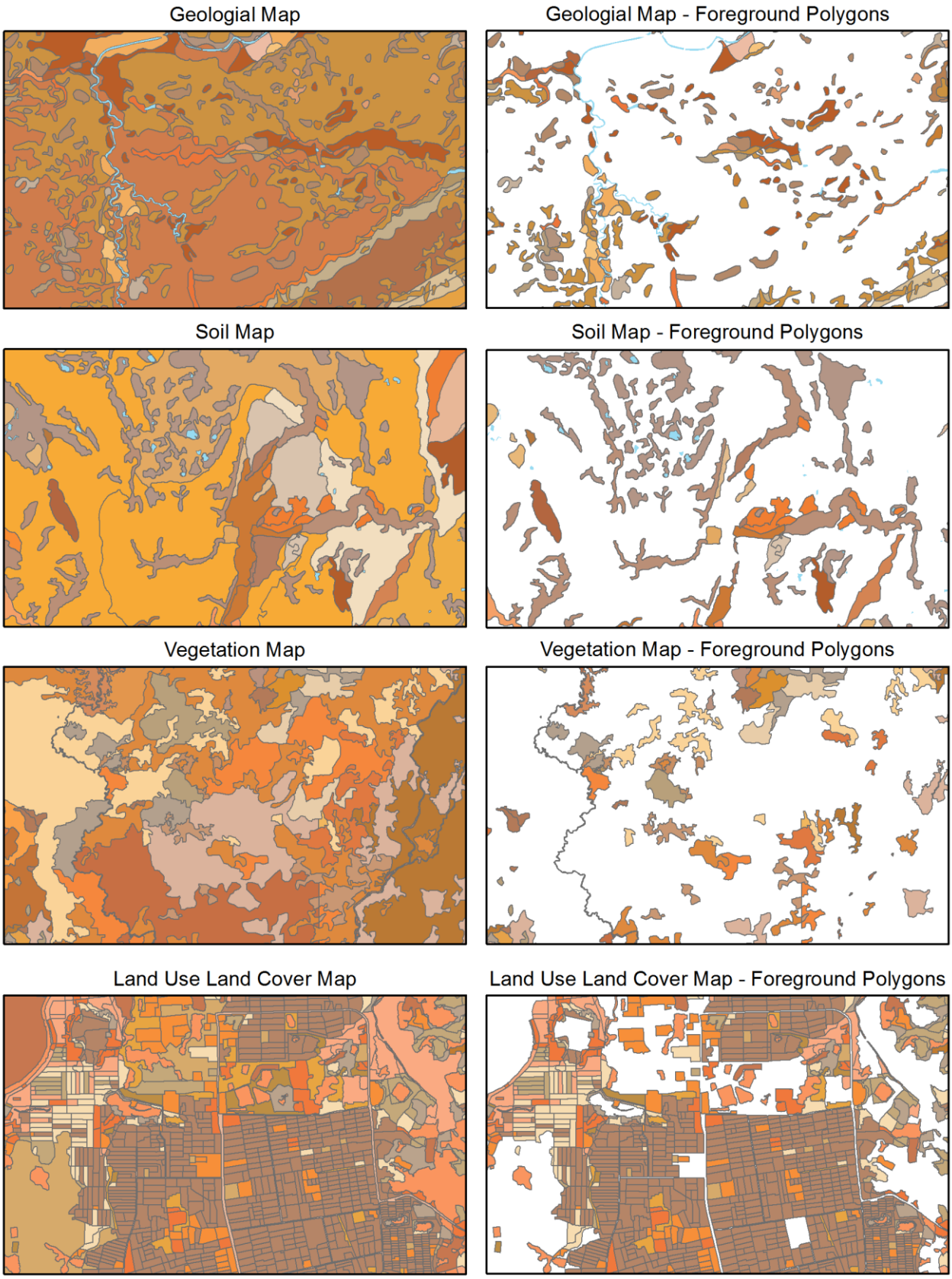


Figure 5.5. Selection of foreground polygons in categorical maps.

Bibliography

Ai, T., & van Oosterom, P. (2002). A displacement method based on field analysis. In: Symposium on Integrated Systems for Spatial Data Production, Custodian and Decision Support, Vol. XXXIV of International Archives of Photogrammetry and Remote Sensing, Xi'an.

Aksoy, S., Younan, N. H., & Bruzzone, L. (2010). Patern recognition in remote sensing. Pattern Recognition Letters, 31(10), 1069-1070.

Albus, J. E., Anderson, R. H., Brayer, J. M., DeMori, R., Feng, H. Y., Horowitz, S. L., Moayer, B., Pavlidis, T., Stallings, W. W., Swain, P. H. & Vamos, T. (2012). Syntactic pattern recognition, applications (Vol. 14). Springer Science & Business Media.

Anders, K. H. (2003). A hierarchical graph-clustering approach to find groups of objects. Paper presented at: 5th ICA Workshop on Progress in Automated Map Generalization; Paris, France, pp 4-7.

Anders, K. H., Sester, M., & Fritsch, D. (1999). Analysis of settlement structures by graph-based clustering. In: Forstner, W. Liedtke., Ce Buckner, J. Proceedings of the Semantic Modelling for the Acquisition of Topographic Information from Images and Maps Conference (SMATI'99). Munich, Germany: Deutsche Forschungsgemeinschaft, pp. 41–49.

Anderson-Tarver, C., Leyk, S., & Buttenfield, B. (2011). Fuzzy modeling of geometric textures for identifying archipelagos in area-patch generalization. Cartography and Geographic Information Science. 38 (2): 137–145.

Anwer, R. M., Khan, F. S., van de Weijer, J., Molinier, M., & Laaksonen, J. (2018). Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification. ISPRS journal of photogrammetry and remote sensing, 138, 74-85.

Bader, M., & Weibel, R. (1997). Detecting and Resolving Size and Proximity Conflicts in the Generalization of Polygonal Maps. Proceedings 18th International Cartographic Conference (ICC), Stockholm, Sweden, pp. 1525-1532.

Baella, B., & M. Pla, M. (2005). Reorganizing the Topographic Databases of the Institut Cartographic de Catalunya Applying Generalization. In Proceedings of the 8th ICA Workshop on Generalisation and Multiple Representation. A Coruña: ICA Commission on Generalisation and Multiple Representation.

Baella, B., Palomar-Vázquez, J., Pardo-Pascual, J. E., & Pla, M. (2007). Spot heights generalization: deriving the relief of the Topographic Database of Catalonia at 1: 25,000 from the master database. In 7th ICA workshop on Progress in Automated Map Generalization, Moscow.

Bard, S. (2004). Quality Assessment of Cartographic Generalisation. Transactions in GIS 8: 63–81.

Barnes, J. W., & Lisle, R. J. (2013). Basic geological mapping. John Wiley & Sons.

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 generalisation. In Proceedings of the 20th international cartographic conference (ICC 2001) (pp. 2110–2116), 6–10 August 2001, Beijing, China. CD-ROM.

Beard, K. (1991). Constraints on rule formation. In: Buttenfield B, McMaster R (eds) Map Generalization, pp. 121-135, Longman.

Boffet, A., & Serra, S. R. (2001). Identification of spatial structures within urban blocks for town characterization. In Proceedings 20th International Cartographic Conference (pp. 1974-1983).

Brassel, K. E., & Weibel, R. (1988). A Review and Framework of Automated Map Generalization. Int. Journal of Geographical Information Systems, 2(3): 229-244.

Brereton, R. G. (2015). Pattern recognition in chemometrics. Chemometrics and Intelligent Laboratory Systems, 149, 90-96.

Brilakis, I., German, S., & Zhu, Z. (2011). Visual pattern recognition models for remote sensing of civil infrastructure. Journal of Computing in Civil Engineering, 25(5), 388-393.

Brodaric, B., & Patera, A. (1997). A set-theoretic approach to managing hierarchies with relational databases Proceedings of the Joint Annual Meeting: Geological Association of Canada and Mineralogical Association of Canada, Ottawa, Canada, p. A18.

Brodaric, B., & Patera, A. (1998). New tools to aid regional geological map compilation: geological map generalization using classification hierarchies in the relational data model. Proceedings of the International conference on GIS for Earth science applications, Ljubljana, Slovenia, p. 1, (Abstract).

Brodaric, B., & Patera, A. (2001). Generalizing geological map units using GIS and tabular data Proceedings of the IAMG - International Association for Mathematical Geology, 7th Annual Conference, Cancun, Mexico, p. 1, (Abstract).

Brown, R. E., Stevens, B. P. J., & Stroud, W. J. (1992). The Early Proterozoic Willyama Supergroup in the Euriowie Block, New South Wales. Geological Survey of New South Wales, Quarterly Notes 88.

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.), Geographic information science. 4th international conference (pp. 41–46), GIScience 2006, IfGIprints 28. ISBN 9-783936-616255.

Burghardt, D., Duchêne, C., & Mackaness, W. (2016). Abstracting Geographic Information in a Data Rich World. Springer International.

Burghardt, D., Schmid, S., & Stoter, J. (2007). Investigations on Cartographic constraint formalisation. In 10th ICA Workshop on Generalisation and Multiple Representation, Moscow, Russia.

Burrough, P. A., & Frank, A. U. 1996. Geographic objects with indeterminate boundaries: GISDATA; 2: London, Taylor & Francis.

Burton, G. (2001). Characterisation of leucocratic granites of the northern Euriowie Block, Broken Hill area. Geological Survey of New South Wales, Report GS2001/060 (unpublished).

Burton, G. R. (1998). Metallogenic study and mineral deposit data sheets of the northern Euriowie Block, Poolamacca Inlier and northernmost Broken Hill Block. Geological Survey of New South Wales, Report GS 1998/226 (unpublished).

Buttenfield, B. P. (1995). Object-oriented map generalization: modeling and cartographic considerations. Müller, J. C., Lagrange, J. P., & Weibel, R. GIS and generalization: methodology and practice. GISDATA I, Serie editors. London: Taylor & Francis, 91-105.

Cámara, M., & López, F. (2000). Mathematical morphology applied to raster generalization of urban city block maps. Cartographica. 1, 33–48

Cecconi, A. (2003). Integration of cartographic generalization and multi-scale databases for enhanced web mapping (Doctoral dissertation, University of Zurich).

Cetinkaya, S., Basaraner, M., & Burghardt, D. (2015). Proximity-based grouping of buildings in urban blocks: a comparison of four algorithms. Geocarto International. 30(6):618–632.

Chen, C. H. (2015). Handbook of pattern recognition and computer vision. World Scientific.

Chen, C. H., & Ho, P. G. P. (2008). Statistical pattern recognition in remote sensing. Pattern Recognition, 41(9), 2731-2741.

Cox, K. G., Bell, J. D., & Pankhurst, R. J. (1979). The Interpretation of Igneous Rocks. George, Allen and Unwin, London.

Crommelinck, S., Bennett, R., Gerke, M., Nex, F., Yang, M. Y., & Vosselman, G. (2016). Review of automatic feature extraction from high-resolution optical sensor data for UAV-based cadastral mapping. Remote Sensing, 8(8), 689.

Cui, X., Wang, J., Wu, F., Li, J., Gong, X., Zhao, Y., & Zhu, R. (2019). Extracting Main Center Pattern from Road Networks Using Density-Based Clustering with Fuzzy Neighborhood. ISPRS International Journal of Geo-Information, 8(5), p.238.

Deng, M., Liu, Q., Cheng, T., & Shi, Y. (2011). An adaptive spatial clustering algorithm based on Delaunay triangulation. Computers, Environment and Urban Systems. 35 (4):320–332.

Dorling, D., & Fairbairn, D. (2013). Mapping: Ways of representing the world. Routledge.

Downs, T. C., & Mackaness, W. A. (2002). An integrated approach to the generalization of geological maps. The Cartographic Journal, Vol. 39, No. 2, pp. 137-152.

Duchêne, C., Baella, B., Brewer, C. A., Burghardt, D., Buttenfield, B. P., Gaffuri, J., Käuferle, D., Lecordix, F., Maugeais, E., Nijhuis, R. & Pla, M. (2014). Generalisation in practice within national mapping agencies. In Abstracting geographic information in a data rich world (pp. 329-391). Springer, Cham.

Duchêne, C., Dadou, M., & Ruas, A. (2005). Helping the capture and analysis of expert knowledge to support generalisation. In: The 8th ICA Workshop on Generalisation and Multiple Representation, La Coruna Http://ica.ign.fr.

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). Pattern classification. John Wiley & Sons.

Dumont, M., Touya, G., & Duchêne, C. (2015). Automated generalisation of intermediate levels in a multi-scale pyramid. In 18th ICA Workshop on Map Generalisation and Multiple Representation.

Edwardes, A., & Mackaness, W. A. (2000). Modelling knowledge for automated generalisation of categorical maps – A constraint based approach. In Innovations in GIS 7. GIS and Geocomputation; Atkinson, P., Martin, D., Eds. Taylor & Francis: London, UK, 2000; pp. 161–174.

Elmahdy, S. I., & Mohamed, M. M. (2016). Automatic Feature Extraction Module for Change Detection in Al Ain, UAE: Analysis by Means of Multi-temporal Remote Sensing Data. Journal of the Indian Society of Remote Sensing, 44(1), 1-10.

Federal Geographic Data Committee. (2006). Subcommittee, G.D. U.S. Geological Survey FGDC Digital Cartographic Standard for Geologic Map Symbolization (FGDC Document Number FGDC-STD-013-2006); U.S. Geological Survey for the Federal Geographic Data Committee: Reston, VA, USA.

Foerster, T., Stoter, J. E., & M. Kraak, M. (2010). Challenges for Automated Generalisation at European Mapping Agencies: A Qualitative and Quantitative Analysis. The Cartographic Journal 47 (1): 41–54.

Fu, K. S. (2019). Applications of pattern recognition. CRC press.

Galanda, M. (2003). Automated Polygon Generalization in a Multi Agent System. Ph.D. thesis, Department of Geography, University of Zurich.

Ghamisi, P., Benediktsson, J. A., Cavallaro, G., & Plaza, A. (2014). Automatic framework for spectral–spatial classification based on supervised feature extraction and morphological attribute profiles. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(6), 2147-2160.

Haldar, S. K. (2018). Mineral exploration: principles and applications. Elsevier Inc. pp 334.

Harrie, L. (2003). Weight-Setting and Quality Assessment in Simultaneous Graphic Generalisation, The Cartographic Journal, 40 (3), 221-233.

Harrie, L., & Weibel, R. (2007). Modelling the overall process of generalisation. In Generalisation of Geographic Information, pp. 67-87. Elsevier Science BV.

Haunert, J. H., & Wolff, A. (2010). Area aggregation in map generalisation by mixed-integer programming. International Journal of Geographical Information Science, 24(12), 1871-1897.

He, X., Zhang, X., & Xin, Q. (2018). Recognition of building group patterns in topographic maps based on graph partitioning and random forest. ISPRS J Photogramm Remote Sens. 136:26–40.

Heinzle, F., & Anders, K. H. (2007). Characterising space via pattern recognition techniques: identifying patterns information: cartographic modelling and applications. In Generalisation of Geographic Information, pp. 233–253. Elsevier Science BV.

Heinzle, F., Anders, K. H., & Sester, M. (2005). Graph based approaches for recognition of patterns and implicit information in road networks, in the Proceedings of the 22th International Cartographic Conference, La Coruña, July 10–16, 2005.

Jaakkola, O. (1997). Quality and Automatic Generalization of Land Cover Data, Ph.D. thesis, Department of Geography, University of Helsinki.

Karsznia, I., Przychodzeń, M., & Sielicka, K. (2020). Methodology of the automatic generalization of buildings, road networks, forests and surface waters: a case study based on the Topographic Objects Database in Poland. Geocarto International, 35(7), 735-758.

Katz, M. B., & Tuckwell, K. D. (1979). Controls of tin-bearing pegmatites and granites in the Precambrian of Broken Hill, Australia. Geological Society of Malaysia, Bulletin 11, 253-265.

Keates, J. S. (2014). Understanding maps. Routledge.

Kilpeläinen, T. (1997). Multiple Representation and Generalization of Geo-Databases for Topographic Maps. PhD thesis, Finnish Geodetic Institute, Kirkkonummi.

Kilpeläinen, T. (2000). Knowledge Acquisition for Generalization Rules, Cartography and Geographic Information Science, 27:1, 41-50, DOI: 10.1559/152304000783547993.

Lambers, K., & Traviglia, A. (2016). Automated detection in remote sensing archaeology: a reading list. AARGnews-The newsletter of the Aerial Archaeology Research Group, 53, 25-29.

Lecordix, F., & Lemarié, C. (2007). Managing generalisation updates in IGN map production. Generalisation of Geographic Information, 285-300.

Lecordix, F., Gallic, J. L., Gondol, L., & Braun., A. (2007). Development of a New Generalization Flowline for Topographic Maps. In 10th ICA workshop on Generalisation and Multiple Representation, Moscow, Russia.

Li, W., & Du, Q. (2016). A survey on representation-based classification and detection in hyperspectral remote sensing imagery. Pattern Recognition Letters, 83, 115-123.

Li, Z., Yan, H., Ai, T., & Chen, J. (2004). Automated building generalization based on urban morphology and Gestalt theory. International Journal of Geographic Information Science. 18(5):513–534.

Lisle, R. J., Brabham, P., & Barnes, J. W. (2011). Basic geological mapping (Vol. 42). John Wiley & Sons.

Lisle, R., J. (2020). Geological structures and maps: A practical guide. Butterworth-Heinemann.

Loudon, T. V. (2000). Geoscience after IT. Part B. Benefits for geoscience from information technology, and an example from geological mapping the need for a broad view. Computers and Geosciences, 26, A5-Al3.

Lv, Z., Zhang, P., & Atli Benediktsson, J. (2017). Automatic object-oriented, spectral-spatial feature extraction driven by tobler's first law of geography for very high resolution aerial imagery classification. Remote Sensing, 9(3), 285.

Mackaness, W., & Chaudhry, O. (2008). Generalization and Symbolization. In: Shekhar S., Xiong H. (eds) Encyclopedia of GIS. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-35973-1_449.

Mackaness, W., & Edwards, G. (2002). The importance of modelling pattern and structure in automated map generalisation. In Proceedings of the joint ISPRS/ICA workshop on multi-scale representations of spatial data (pp. 7-8).

Mackaness, W., & Ruas, A. (2007). Evaluation in the map generalisation process. In Generalisation of Geographic Information, pp. 89–112. Elsevier Science BV.

Mackaness, W., Burghardt, D., & Duchêne, C. (2016). Map generalization. International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology, 1-16.

Mackaness, W., Ruas, A., & Sarjakoski, L. T. (2007). In Generalisation of Geographic Information, pp. 315–323. Elsevier Science BV.

Mackaness, W., Ruas, A., & Sarjakoski, L. T. (2011). Generalisation of geographic information: cartographic modelling and applications. Elsevier.

Maltman, A. (2012). Geological maps: an introduction. Springer Science & Business Media.

Marjoribanks, R. (2010). Geological Methods in Mineral Exploration and Mining. Springer-Verlag Berlin, Heidelberg, 2nd edition. XV pp 238.

Marshall, S. (2005). Streets and Patterns. New York, Spon Press, Taylor and Francis.

McCabe, C. (2008). Vector approaches to generalizing faults and polygons in 1: 24,000 geologic maps: Santa Rosa, California, case study.

McMaster, R., & Shea, S. (1988). Cartographic generalization in digital environment: A framework for implementation in a GIS. In: GIS/LIS'88, pp. 240-249.

McMaster, R., & Shea, S. (1992). Generalisation in digital cartography. Washington DC, Association of American Geographers.

Meijers, M., Savino, S., & Van Oosterom, P. (2016). SPLITAREA: an algorithm for weighted splitting of faces in the context of a planar partition. International Journal of Geographical Information Science, 30(8), 1522-1551.

Monmonier, M. (2018). How to lie with maps. University of Chicago Press.

Müller, J. C. & Wang, Z. (1992). Area-Patch Generalisation: A Competitive Approach. The Cartographic Journal, 29(2): 137-144.

Müller, J. C. (1990). Rule based generalization: potentials and impediments. In: Proceedings of the 4th International Symposium on Spatial Data Handling, Zurich, pp. 317–334.

Najman, L. and Talbot, H. (2013). Mathematical morphology: from theory to applications. John Wiley & Sons.

Neun, M., Weibel, R. & Burghardt, D. (2004). Data enrichment for adaptive generalisation. In: The 7th ICA Workshop on Generalisation and Multiple Representation, Leicester.

Ormeling, F. (2013). Cartography as a Window to the World. The Cartographic Journal, 50(2), 175-181.

Patera, A. (2006). Advanced Techniques for the Management of Geological Mapping. Ph.D. Thesis, The Open University, Milton Keynes, UK, 195 pp.

Peter, B. (2001). Measures for the generalization of polygonal maps with categorical data. In Fourth ICA Workshop on Progress in Automated Map Generalization, Beijing (pp. 2-4).

Peter, B., & Weibel, R. (1999). Using vector and raster-based techniques in categorical map generalization. In Third ICA workshop on progress in automated map generalization (Vol. 14). Ottawa.

Petzold, I., Burghardt, D., & Bobzien, M. (2006). Workflow management and generalisation services. In: The 9th ICA Workshop on Generalization and Multiple Representation, Portland, OR Http://ica.ign.fr.

Raposo, P., Brewer, C., A. Sparks, K. (2016). An impressionistic cartographic solution for base map land cover with coarse pixel data. Cartographic Perspectives. 83, 5–21.

Raposo, P., Samsonov, T. (2014). Towards general theory of raster data generalization. In Proceedings of the 17th ICAWorkshop on Generalisation and Multiple Representation, Vienna, Austria, 23 September. p. 10.

Regnauld, N. (2001). Contextual Building Typification in Automated Map Generalization. Algorithmica, 30: 312–333.

Regnauld, N., & McMaster, R. B. (2007). A synoptic view of generalisation operators. In Generalisation of geographic information (pp. 37-66). Elsevier Science BV.

Regnauld, N., Edwardes, A., & Barrault, M. (1999). Strategies in building generalisation: modelling the sequence, constraining the choice. In ICA Workshop on Progress in Automated Map Generalization, Ottawa, Canada.

Revell, P., Regnauld, N., & Bulbrooke, G.Regnauld, N. (2011). OS Vector map District: Automated Generalisation, Text Placement and Conflation in Support of Making Pubic Data Public. In 25th International Cartographic Conference, July 2011, Paris. http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/D3-Generalisation/CO-358.pdf.

Roberts, J., L. (2013). Introduction to Geological Maps and Structures: Pergamon International Library of Science, Technology, Engineering and Social Studies. Elsevier.

Robinson, R. (2020). Georgia O'Keeffe: A Life (new edition). Bloomsbury Publishing.

Ruas, A., & Duchêne, C. (2007). A prototype generalisation system based on the multi-agent system paradigm. In Generalisation of geographic information (pp. 269-284). Elsevier Science BV.

Ruas, A., & Plazanet, C. (1996). Strategies for Automated Generalization. Proceedings of the 7th Spatial Data Handling Symposium, Delft, the Netherlands, pp. 319-336.

Ruas, A. (1999). Modèle de généralisation de données géographiques à base de contraintes et d'autonomie. Doctoral Thesis, Université de Marne-la-Vallée.

Sarjakoski, L. T. (2007). Conceptual models of generalisation and multiple representation. In Generalisation of Geographic Information, pp. 11–35.

Sayidov, A., & Weibel, R. (2016). Constraint-based approach in geological map generalization. In Proceedings of the 19th ICA Workshop on Generalisation and Multiple Representation, Helsinki, Finland, 14 June.

Sayidov, A., & Weibel, R. (2017). Automating geological mapping: A constraint-based approach. In Proceedings of the CEUR (Central Europe)Workshop Proceedings, Leeds, UK, 30 October–2 November.

Sayidov, A., & Weibel, R., & Leyk, S. (2020a). Recognition of group patterns in geological maps by building similarity networks. Geocarto International, 1-20.

Sayidov, A., Aliakbarian, M., & Weibel, R. (2020b). Geological map generalization driven by size constraints. ISPRS International Journal of Geo-Information, 9(4), 284.

Schmid, S. (2008). Automated constraint-based evaluation of cartographic generalization solutions. Master thesis, Department of Geography, University of Zurich.

Schylberg, L. (1993). Cartographic amalgamation of area objects. International Archives of Photogrammetry and Remote Sensing, 29, 135-135.

Seo, S., Gary, M., & Marsh, D. (2006). A review and comparison of methods for detecting outliers in univariate data sets. Department of Biostatistics. Graduate School of Public Health. 1–53. Retrieved from http://d-scholarship.pitt.edu/7948/.

Sester, M. (2020). Cartographic generalization. Journal of Spatial Information Science, 2020(21), 5-11.

Shea, K. S., & McMaster R. B. (1989). Cartographic generalisation in a digital environment: When and how to generalise. Proceedings of Auto Carto 9, Baltimore, Maryland, USA, 56–67.

Smirnoff, A., Huot-Vézina, G., Paradis, S. J., & Boivin, R. (2012). Generalizing geological maps with the GeoScaler software: The case study approach. Computers and. Geosciences. 2012, 40, 66–86.

Smirnoff, A., Paradis, S. J., & Boivin, R. (2008). Generalizing surficial geological maps for scale change: ArcGIS tools vs. cellular automata model. Computers and. Geosciences. 2008, 34, 1550–1568.

Song, Y. Y., & Yu, D., & Shen, Ch. (2009). Knowledge acquisition model of map generalization based on granular computing. Proceedings of SPIE - The International Society for Optical Engineering. 7492. 10.1117/12.838333.

Spencer, E., W. (2017). Geologic Maps: A Practical Guide to Preparation and Interpretation. Waveland Press.

Spiess, E., Baumgartner, U., Arn, S., & Vez, C. (2005). Topographic maps map graphics and generalisation. Cartographic. Publication. Series. 2005, Published by Swiss Society of Cartography, No 2. 17, 121.

Stanislawski, L., & S. Savino. (2011). Pruning of Hydrographic Networks: A Comparison of Two Approaches. In Proceedings of the 14thICA Workshop on Generalisation and Multiple Representation, Paris, 2011.

Steiniger, S. (2007). Enabling pattern-aware automated map generalization. PhD thesis. University of Zurich.

Steiniger, S., & Meier, S. (2004). Snakes: a technique for line smoothig and displacement in map generalisation. In: The 7th ICA Workshop on Generalisation and Multiple Representation, Leicester Http://ica.ign.fr.

Steiniger, S., & Weibel, R. (2005). A conceptual framework for automated generalization and its application to geologic and soil maps. In: Proceedings of the 22nd International Cartographic Conference, 11–16 July, La Coruna, Spain, 10 pp.

Steiniger, S., & Weibel, R. (2007). Relations among map objects in cartographic generalization. Cartography Geographic Information Science. 34(3):175–197.

Stern, C., & M. Sester, M. (2012). Towards Extraction of Constraints for Integrating Environmental Spatial Data in Digital Landscape Models of Lower Resolution – A Work In Progress. In 14th Workshop of the ICA Commission on Generalisation and Multiple Representation, Paris, June 2011.

Stevens, B. P. J., Page, R. W., & Crooks, A. (2008). Geochronology of Willyama Supergroup metavolcanics, meta sediments and contemporaneous intrusions, Broken Hill, Australia. Australian Journal of Earth Sciences, 55:3, 301-330, DOI: 10.1080/08120090701769456.

Stevens, B. P. J., Willis, I. L., & Brown R. E. (1998). Euriowie 1:25 000 Geological Sheet, 7235-III-S & 7135-II-S. Geological Survey of New South Wales, Sydney.

Stoter, J. E. (2005). Generalisation within NMA's in the 21st century, in Proceedings of the International Cartographic Conference.

Stoter, J. E., B. Baella, B., C. Blok, C., D. Burghardt, D., C. Duchêne, C., M. Pla, M., N. Regnauld, N., & Touya, G. (2010). State-of-the Art of Automated Generalisation in Commercial Software, 231. Amsterdam: EuroSDR, March. http://www.eurosdr.net/projects/generalisation/eurosdr_gen_final_report_mar2010.pdf.

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.

Su, B., Li, Z. L. & Lodwick, G. (1997). Morphological transformation for the elimination of area features in digital map generalization. Cartography, 26(2), 23–30.

Su, B., Li, Z. L. & Lodwick, G. (1998). Algebraic models for collapse operation in digital map generalization using morphological operators. Geoinformatica, 2(4), 359–382.

Šuba, R. (2018). State of the art in automated map generalization. A+ BE| Architecture and the Built Environment, (18), 9-18.

Tian, J., Ai, T., & Jia, X. (2012). Graph based recognition of grid pattern in street networks. In Advances in Spatial Data Handling and GIS (pp. 129-143). Springer, Berlin, Heidelberg.

Töpfer, F., & Pillewizer, W. (1966). The principles of selection. The Cartographic Journal, 3(1), 10-16.

Touya, G., Lokhat, I., & Duchêne, C. (2019). CartAGen: an Open Source Research Platform for Map Generalization. In International Cartographic Conference (Vol. 2, pp. 1-9).

Tuia, D., Volpi, M., Dalla Mura, M., Rakotomamonjy, A., & Flamary, R. (2014). Automatic feature learning for spatio-spectral image classification with sparse SVM. IEEE Transactions on Geoscience and Remote Sensing, 52(10), 6062-6074.

van Oosterom, P., Meijers, M., Stoter, J., & Šuba, R. (2014). Data structures for continuous generalisation: tGAP and SSC. In Abstracting geographic information in a data rich world, pp. 83-117. Springer, Cham.

Vernon, R. H., & Clarke, G. L. (2008). Principles of Metamorphic Petrology. Cambridge University Press, pp. 446. Leake, B. E. (2008). Mineralogical Magazine vol. 72 issue 5. pp: 1145-1146. Published by Cambridge University Press.

Wang, X., & Burghardt, D. (2017). A Stroke-based approach to detect patterns of discrete buildings for generalization purposes. 20th ICA workshop on Generalization and Multiple Representation, Washington, DC, USA.

Ware, J. M., Jones, C. B., & & Thomas, N. (2003). Automated map generalisation with multiple operators: A simulated annealing approach. International Journal of Geographical Information Science, 17(8), 743–769.

Weibel, R. (1996). Generalization of spatial data: Principles and selected algorithms. In Advanced School on the Algorithmic Foundations of Geographic Information Systems (pp. 99-152). Springer, Berlin, Heidelberg.

Weibel, R., & Dutton, G. (1998). Constraint-based automated map generalization. In T. K. Poiker & N. Chrisman, editors, Proceedings 8th International Symposium on Spatial Data Handling, pages 214–224. IGU - Geographic Information Science Study Group, Vancouver.

Weibel, R., & Dutton, G. (1999). Generalising spatial data and dealing with multiple representations. Geographical information systems, 1, 125-155.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In: Ellis WD, editor. A source book of Gestalt psychology. New York: Harcourt, Brace; p. 71–88. Original work published 1923.

Willis, I. L., Brown, R. E., Stroud W. J., & Stevens, B. P. J. (1983). The Early Proterozoic Willyama Supergroup: stratigraphic subdivision and interpretation of high to low-grade metamorphic rocks in the Broken Hill Block, New South Wales. Journal of the Geological Society of Australia 30, 195-224.

Yang, J., Zhu, J., Sun, Y., & Zhao, J. (2019). Delimitating urban commercial central districts by combining kernel density estimation and road intersections: a case study in Nanjing city, china. ISPRS International Journal of Geo-Information, 8(2), p.93.

Yaolin, L., Moenaar, M., & Tinghua, A. (2001). Frameworks for generalization constraints and operations based on object oriented data structure in database generalization. Proceedings of the 20th International Cartography Conference, August 6 -10, Beijing, China.

Yaolin, L., Molenaar, M., & Kraak, M. J. (2002). Categorical database generalization aided by data structure. In Wise, S., Brindley, P., Kim, Y. H., & Openshaw, C., editors, Proceeding GIS research UK 10th Annual Conference, pages 345–348. University of Sheffield, Sheffield.

Younan, N. H., Aksoy, S., & King, R. L. (2012). Foreword to the special issue on pattern recognition in remote sensing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(5), 1331-1334.

Zhang, X., Ai, T., & Stoter, J. (2010). Characterization and detection of building patterns in cartographic data: two algorithms. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science. 38(II):261–266.

Zhang, X., Stoter, J., Ai, T., Kraak, M., & J, Molenaar, M. (2013). Automated evaluation of building alignments in generalized maps. International Journal of Geographic Information Science. 27(8):1550–1571.

AZIMJON SAYIDOV

Born on 3rd of September 1984, in Bukhara, Uzbekistan

EDUCATION

- 2015 – 2020

Ph.D., GIS Unit, Department of Geography, University of Zurich, Zurich, Switzerland.

Thesis: "An Integrated Methodology for Automated of Generalization of Geological Maps."

- 2012 – 2013

Postgraduate Degree in Remote Sensing and GIS, Indian Institute of Remote Sensing, Dehradun, India.

- 2007 – 2009

Master's Degree in Cartography, Department of Geodesy, Cartography and Cadaster, Faculty of Geography, National University of Uzbekistan, Tashkent, Uzbekistan.

- 2003 – 2007

Bachelor's Degree in Geodesy, Cartography, and Cadaster, Department of Geodesy, Cartography and Cadaster, Faculty of Geography, National University of Uzbekistan, Tashkent, Uzbekistan

PROFESSIONAL EXPERIENCE

- 2015 – 2020

Research Assistant, GIS Unit, Department of Geography, University of Zurich, Zurich, Switzerland.

- 2013 – 2015

Cartographer, The Center for Remote Sensing and Geographic Information Systems, under the Geological Committee of Uzbekistan, Tashkent, Uzbekistan.

- 2011 – 2015

Researcher/Cartographer, Department of Geodesy, Cartography and Cadaster, Faculty of Geography, National University of Uzbekistan, Tashkent, Uzbekistan.

- 2009 – 2010

Topographer/Surveyor, Uzbekistan Research Institute in Geoinformation and Civil Building Cadaster, Tashkent, Uzbekistan.

- 2007 – 2009

Topographer/Surveyor, Tashkent Research Institute in Geoinformation and Civil Building Cadaster, Tashkent, Uzbekistan.

Publications

Sayidov A.K. (2012). Application of Generalization in GIS Programs. Journal of Uzbekistan Geographic Society, N-39, Tashkent-2012, 166-168 pp.

Sayidov A.K. (2012). About Generalization of River Network. “Глобализация и география” Материалы международной научной конференции, посвященная 85 – тилетию профессора М. А. Мусеибова, Баку – 2012, 512-515 ст.

Sayidov A.K. (2013). Based on GIS Developing Generalization of Road Networks, “Journal of Uzbekistan Land Resources, Geodesy, Cartography, and State Cadastre,” Tashkent.

Sayidov A.K. (2013). Algorithms of Generalization of Linear Objects (Road Network example). Journal of Uzbekistan Geographic Society, N-42, Tashkent.

Sayidov A.K. (2013). Comparative Study of Knowledge Based Algorithms for Automated Road Network Generalization, Report, CSSTEAP, Geoinformatics Department, IIRS.

Sayidov A.K. (2015). Investigation of algorithms of automated generalization of linear features, Scientific journal Izvestia vuzov «Geodesy and aerophotography», Moscow State University of Geodesy and aerophotography.

Sayidov A.K., & Weibel R. (2016). The constraint-based approach in geological map generalization, 19th ICA Workshop, Generalization and Multiple Representation, Automated generalization for on-demand mapping, Helsinki, Finland, the 14th of June 2016.

Sayidov A., Alliakbarian M., & Weibel R. (2020). Geological Map Generalization Driven by Size Constraints. ISPRS International Journal of Geographic Information. 2020, 9, 284.

Sayidov A., Weibel R., & Leyk S. (2020) Recognition of group patterns in geological maps by building similarity networks, Geocarto International, DOI: 10.1080/10106049.2020.1730449.

Conferences and Symposiums

National Symposium on Space Technology for Food and Environment Security and Annual Conventions of Indian Society of Remote Sensing and Indian Society of Geomatics, December 5-7, 2012, New Delhi, India.

19th ICA Workshop, Generalization, and Multiple Representation, Automated generalization for on-demand mapping, Helsinki, Finland, the 14th of June 2016.

Appendix A. Experiments on Perceptual Polygon Grouping

1. Introduction

Based on the literature review, it was known that pattern recognition and perceptual grouping of polygons by map readers are mostly dependent on the proximity and category of the closely located map objects. However, we hypothesized that these are probably not the only criteria used to delineate polygon groups in geological maps, and thus, more criteria would need to be considered to obtain more detailed and meaningful polygon grouping results.

In order to obtain these criteria and the associated contextual information, in preparation of the work on Chapter 3 we conducted an experiment, which aimed to gain insight from individuals who have specific expertise and exposure to maps as well as the map-making processes. Specifically, we wanted to know how they would recognize patterns and perceptually form groups on a given set of foreground polygons obtained from a geological map. The foreground polygons are relatively small in size, and thus they are the main target for the map generalization process (the notion of foreground polygons is better described in Chapter 3 of this thesis).

The experiment consisted of two parts: The participants were asked to 1) draw groups/patterns that they perceived on a sample of foreground polygons on a geological map, and 2) describe the criteria that led them in their group identification process. The second part of the experiment, describing the criteria, also contained a feedback and recommendation section.

To gain background information on the participants, they were also asked to provide information on their expertise and exposure to maps and the map-making process, and some demographic information. The experiment aimed at obtaining knowledge about group identification and formation from people who work with maps or have specific mapping expertise.

The knowledge obtained in the second part of the experiment was later translated into criteria used to recognize and identify groups in the research presented in Chapter 3 (*Recognition of group patterns in geological maps by building similarity networks*). The identified polygon groups and associated information, in turn, were used as input to inform the generalization methodology of Chapter 4 (*Integrating Aggregation and Typification to Generalize Polygon Group Patterns in Geological Maps*), which uses contextual generalization operators s— most importantly aggregation and typification — to generalize polygon groups sharing similar properties. These operators are able to maintain relevant geometric properties of the identified polygon groups and

thus ensure the preservation of the key distributional characteristics of the map in the generalization process. In the following sections, we summarize the results of the main parts of the experiment.

2. Experiment

2.1. Task 1. Drawing the Polygon Groups/Clusters Perceived in Sample Maps

In the first task, participants were asked to draw patterns of polygon groups which they perceived on given geological maps. The participants were provided with two samples of geological maps at the scale of 1:25:000 (Figure A2) that only contain foreground polygons. A brief explanation of the purpose of the experiment was given in the accompanying text, along with a short description of what the given polygons stand for (Figure A5). The participants were then asked to draw patterns they see on the map that could be grouped for the purpose of map generalization. There was no time limit for this task. However, given that the sample maps are relatively compact, it took participants on average 15 to 25 minutes to complete this task.

Snapshots of the resulting drawings of perceived polygon groups of all 15 participants are shown in Figure A1. Figure A2 shows the result of a single participant.

2.2. Task 2. Describe What Led You to Identify the Polygon Groups

The second part of the experiment was dedicated to obtaining insight about how the participants decided on drawing the boundaries delineating polygon groups. In this task, participants were asked to write down the reasoning behind their perception of patterns and decision on the boundaries of the groups which they has formed. Moreover, the users were given the option to provide their feedback and recommendation on the experiment and the assigned tasks. Figure 3A summarizes the criteria that the participants mentioned that they used during the pattern recognition and group identification process. Figure A6 shows an example description of the reasoning of the group identification of a selected user.

2.3. Participant Characteristics

In a brief questionnaire, the participants were asked to indicate their gender; age; whether they had taken any courses in cartography, GIS, or map-making; and how often they are exposed to map-making processes. This information was used to ensure that all participants had sufficient expertise. The summary of the participant characteristics is given in Figure A4; Figure A7 shows the example of a selected participant.

3. Results

Analyzing the results of the user experiment allowed to provide the necessary knowledge that was integrated in the research carried out in this thesis (Chapters 3 and 4, respectively). As a result of the experiment we identified 9 criteria for pattern recognition and grouping of small polygons in geological maps, 5 of which were used in the polygon grouping methodology described in Chapter 3. The results of the experiment are summarized below.

3.1. Participant Characteristics

A total of 15 participants conducted the experiment. The participants were almost equally from both genders, eight males and seven females. The age of the participants ranged from 26 to 40. However, most of them, 11 to be precise, are older than 30 years old. Most participants (13) had a background in either cartography or GIS. Moreover, 11 participants have experience with maps and map-making on a daily basis, 4 of them once a week, one of them once a month (Figure A4). Information provided by a selected participant is shown in Figure A7.

3.2. Task 1

Every participant was provided with two samples of geological maps to draw patterns on them. The first task produced mostly sensible results of pattern recognition and grouping (Figure A1 and Figure A2). However, two participants who did not have specific background in cartography, (Participants 1 and 2), generated over-generalized patterns on the map. Three other participants (Participants 4, 6, and 10; Figure A1) generated medium-sized clusters on the map but delineating relatively detailed patterns. The remaining 10 participants (Figure A1) generated highly detailed patterns and groups on the map, which could potentially be used in the generalization process described in Chapter 4. An example of a result obtained from the perceptual grouping task is also visualized in Figure A2. Moreover, the questionnaire form for Task 1 and Task 2, respectively, filled by a selected participant, is shown in Figure A5 and Figure A6, respectively.

3.3. Task 2

The results of the second task indicate that the perception and formation of polygon groups has more commonalities than differences among the participants. Unanimity was observed in the two primary criteria (Figure A3): proximity – the distance between features on the map; and category of polygons (the category – rock units on the geological maps). This unanimity makes sense, as the

primary factors of perceptual clustering lie in delineating objects of similar categories that are located near to each other. The polygons' size, orientation, and shape were also a popular choice for pattern recognition implemented by the Participants 9, 7, and 5, respectively. (Eventually, these 5 criteria — proximity, category of polygons, size, orientation, and shape — were the integrated in the group identification process of Chapter 3). Three test users indicated that they considered the group's shape, and two participants showed alignment as a group forming a prerequisite. Only one participant stated that the number of groups and the majority in the cluster should be considered in the group identification process. The criteria mentioned in Task 2 are summarized in Figure A3.

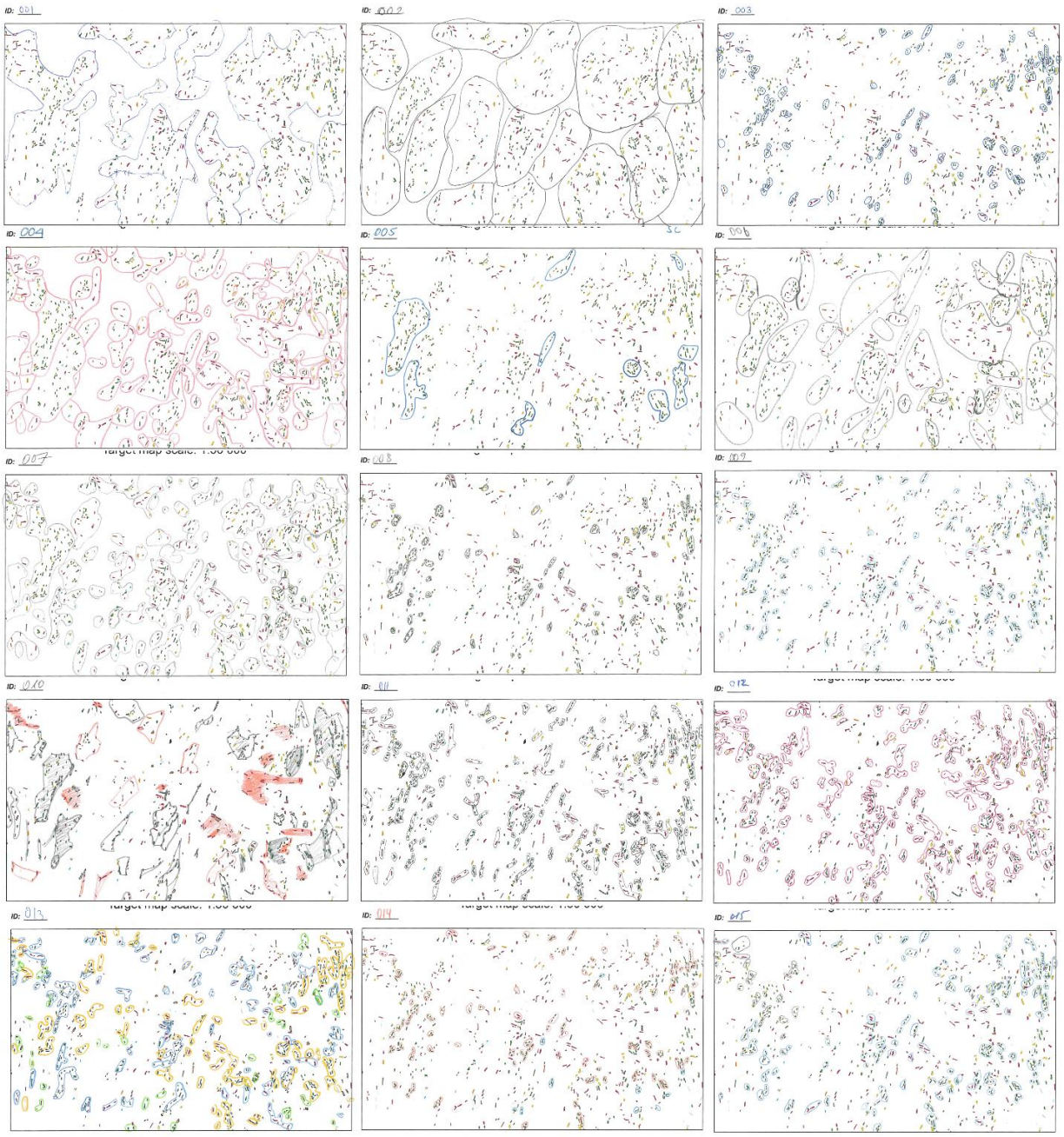


Figure A1. Results of Task 1, pattern recognition and group identification obtained from all 15 participants.

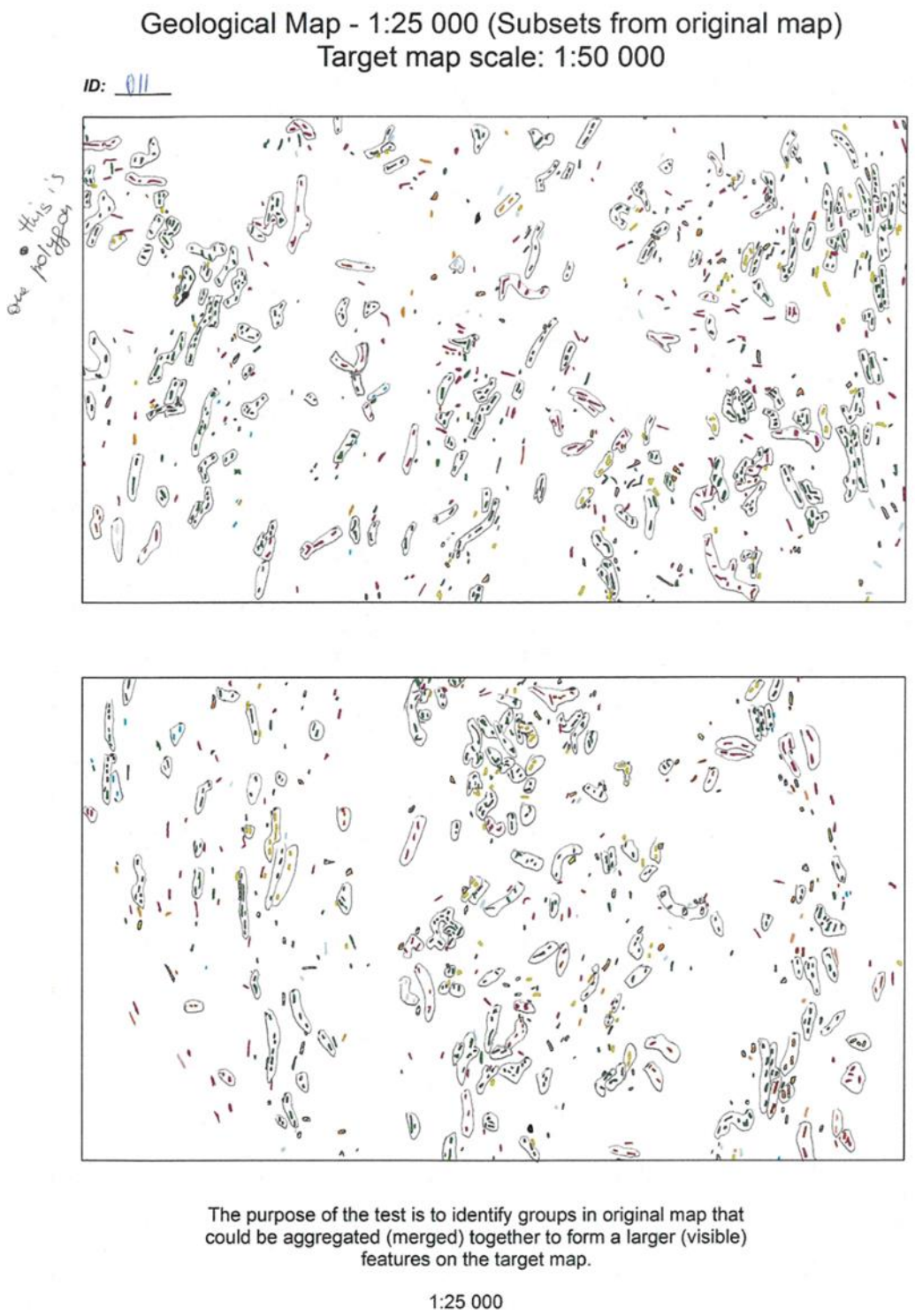


Figure A2. Result of Task 1, pattern recognition and group identification, obtained from a selected participant.

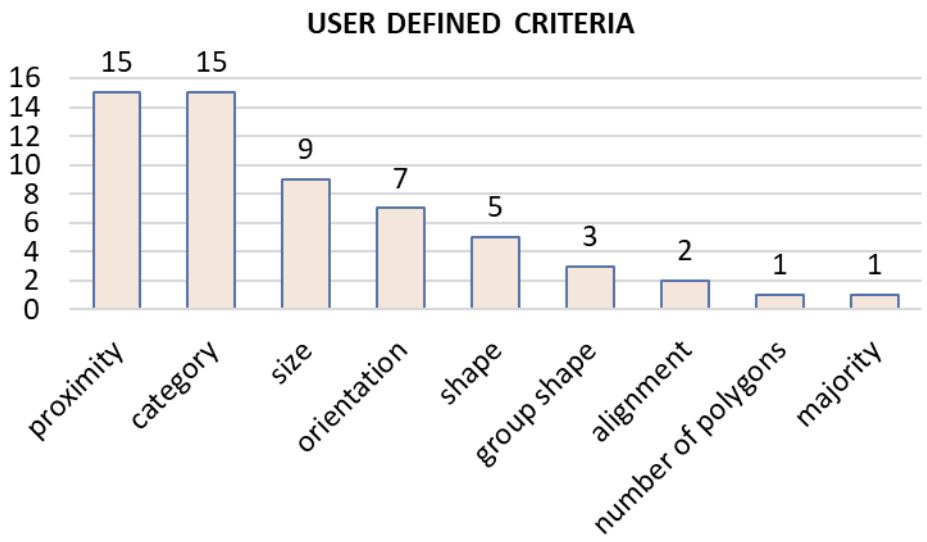


Figure A3. Criteria obtained from participants after grouping.

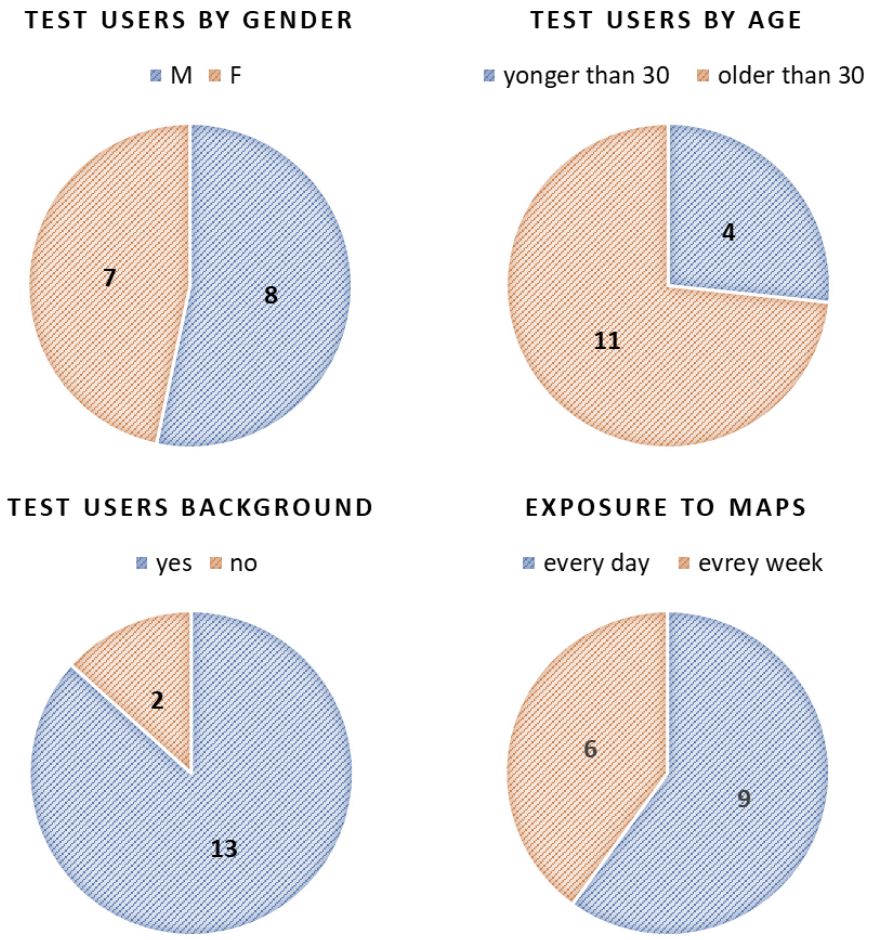


Figure A4. Participant characteristics.

ID: 011 ***Name:*** [redacted]

Test on Generalization of Geological maps.

Test name: Recognize and group polygons in geological maps.

Original (current) Map scale: 1:25 000

Target scale: 1:50 000

Auxiliary Information

The purpose of the test is to identify groups in the original map that could be aggregated (merged) together to form larger (visible) features on the target map.

Description: Grouping polygons that can be generalized. Two subsets maps from an original map scale 1:25 000 scale are provided. The aim of the test is to identify groups/clusters that can be generalized together to make map features readable.

Tasks 1:

- Draw the groups/clusters that you perceive in the attached sample maps.

Intentional Space

The polygons in the map represent rock units that are shown by color to indicate where they are exposed at the Earth surface.

Figure A5. Test example of Task 1: brief description and ancillary information on the task.

ID: 011 **Name:** [redacted]

Test on Generalization of Geological maps.

Tasks 2:

- Describe based on what (reasoning) did you find group polygons.

I've had a feeling that these are linear structures, so I've tried to connect the polygons in the parallel lines (or polygons of the same color have to be parallel).
On the second map there are also some groups of polygons which are not really linear.

Feedback\recommendations:

Making a bit clearer what the scale 1:50 000 is, I think I've made more like 1:200 000

Figure A6. Example of Task 2: user description of pattern recognition and group identification.

ID: 011 ***Name:*** [redacted]

Test user information

Please indicate you gender.

☐ Male:

☑ Female:

Please indicate your age.

29

Have you had Cartography, GIS or map-making courses before?

☑ Yes ☐ No

How many ECTS? many

How often do you deal with maps?

☑ Every day

☐ Every week

☐ Every month

☐ Every now and then

☐ Never

Figure A7. An example of participant information of a selected participant.