Cartography and Geographic Information Science, 2014
Vol. 41, No. 1, 1–13, http://dx.doi.org/10.1080/15230406.2013.824637

Taylor & Francis
Taylor & Francis Group

Fully automated generalization of a 1:50k map from 1:10k data

Jantien Stoter[a,b]*, Marc Post[b], Vincent van Altena[b], Ron Nijhuis[b] and Ben Bruns[b]

[a]*OTB, Delft University of Technology, Delft, The Netherlands;* [b]*Kadaster, Apeldoorn, The Netherlands*

(*Received 20 June 2012; accepted 26 April 2013*)

This article presents research that implements a fully automated workflow to generalize a 1:50k map from 1:10k data. This is the first time that a complete topographic map has been generalized without any human interaction. More noteworthy is that the resulting map is good enough to replace the existing map. Specifications for the automated process were established as part of this research.

Replication of the existing map was not the aim, because feasibility of automated generalization is better when compliance with traditional generalizations rules is loosened and alternate approaches are acceptable. Indeed, users valued the currency and relevancy of geographical information more than complying with all existing cartographic guidelines. The development of the workflow thus started with the creation of a test map with automated generalization operations. The reason for the test map was to show what is technologically possible and to refine the results based on iterative users' evaluation. The generalization operations (200 in total) containing the relevant algorithms and parameter values were developed and implemented in one model. Particular effort was made to enrich the source data in order to improve the results. The model is context aware which means it is able to apply different algorithms or adjust parameter values in accordance with a specific area. The result of the research is a fully automated generalization workflow that produces a countrywide map at scale 1:50k from 1:10k data in 50 hours.

A fully automated workflow may be the only way to produce flexible and on-demand products; consequently, the results were implemented as a new production line in 2013. Issues for further research have been identified.

Keywords: automated generalization; cartography; multi-scale topographic data

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

Introduction

Many National Mapping Agencies (NMAs) have introduced or are introducing automated generalization to improve parts of their map production lines. This article presents research enabling the implementation of a fully automated workflow to generalize a complete 1:50k topographic map (covering all classes and the complete country) from 1:10k object-oriented data.

The research was prompted by problems encountered through the Netherlands' Kadaster, the organization responsible for topographic mapping in The Netherlands. The Kadaster is responsible for producing topographic vector data and raster maps at the scales of 1:10k, 1:50k, 1:100k, 1:250k, 1:500k, and 1:1000k and updating them every two years at least. To meet this legal obligation the Kadaster has been converting its vectorized maps into object-oriented databases since 2007. The manual updates of these products by cartographers take too much time and is, consequently, too costly to realize the required update cycle within budget. For smaller-scale maps this problem is even bigger since every scale is generalized from the previous larger (i.e., more detailed) scale in a ladder approach. Therefore, updates in 1:10k data are propagated much later in maps at scale smaller than 1:50k. The time-consuming update process also prevents the creation of flexible map production lines for on-demand products (i.e., different products for different purposes). The need for such products is the reason behind the Kadaster's great interest in implementing automated generalization.

Many researches have studied fully automated generalization of topographic maps and mapping agencies in many countries have introduced automated generalization. Examples are Great Britain (Regnauld 2011), Turkey, USA, Norway, China, France (Lecordix et al. 2007), Denmark (Foerster, Stoter, and Kraak 2010), Catalonia (Baella and Pla 2005), Germany, and Switzerland.

As already mentioned, this article focuses on research enabling the implementation of a fully automated generalization workflow. Although Regnauld (2011) also implemented a fully automated generalization workflow, his study differs from ours in some aspects. Regnauld's aim was to produce a background map at the district level for Great Britain which would focus on a few themes (meaning gaps in the map are allowed) and which would complement the existing multi-scale products. In contrast, our study focused on automatically producing a complete topographic map that replaces the existing map products (i.e., it is not an additional product). Further, since Regnauld's generalization involved a map at the scale of 1:15k–1:35k, no displacement was necessary, in contrast to our research. Finally, whereas Regnauld (2011) applied generalization to 10 × 10 km tiles and features were

*Corresponding author. Email: j.e.stoter@tudelft.nl

© 2013 Cartography and Geographic Information Society

chopped at tile edges in order to be able to process data for the entire country, our generalization was complete – no gaps and no overlap, and resulted in a needless map. As shown later in this article, 100% automation was possible because map specifications were adjusted to meet technological possibilities. As a result of the automated workflow, data for the entire area of The Netherlands were generalized within 50 hours, which is a three-weeks' turnaround including pre-processing, generalization, visualization, and printing of the maps.

The research builds largely on existing (and recently released) tools. This may not seem innovative. However, as concluded by Stoter et al. (2010), the main problems of applying existing generalization tools to real-world map production are that the tools are difficult to parameterize and that automatic detection tools for discriminating between specific contexts are missing. This research addressed these problems and built an automated workflow that implements generalization operators in the right order with the right parameter values while discriminating between different contexts. Also, new is the introduction of partitions which do not require features to be chopped at artificial (map sheet) borders. This enables to generalize the whole country while maintaining road and water networks.

Although involving the users in establishing the new map specifications was crucial, the focus in this article is on the data processing in the automated workflow. The evaluation that was performed at several stages of the research and by different types of users is detailed in other publications. This article will only summarize the users' evaluations results when relevant or necessary.

It should be noted that our research focuses on generalization of topographic maps representing landscapes and urban areas. Generalization of bathymetric data and isolines or monothematic data (such as geological maps) faces different challenges and was not attempted by this research.

This article is organized as follows. The following section defines the scope of automated generalization in our research and presents the approach in relation to previous research on the topic. The next section presents the automated generalization workflow developed and tests carried out in a few areas to validate it. The results and findings are presented in the next section. The article closes with conclusions and future work in the last section.

Scope and methodologies

Scope

The main objective of the research was to accomplish fully automated generalization of a topographic map (at 1:50k scale) from large-scale object-oriented data (scale 1:10k). Replication of the existing map was not the aim, because feasibility of automated generalization is better when compliance with traditional generalizations rules is loosened and alternate approaches are acceptable.

This is because topographic products (many of which originated more than 60 years ago) may overemphasize cartographic principles and ignore the new requirements of multi-scale topographic information, such as supporting differing applications in a wide array of user contexts as well as user preference for up-to-date maps that may not meet all traditional cartographic principles. In addition, automating a previously interactive process designed for a different technical and organizational context is extremely complicated (Foerster, Stoter, and Kraak 2010; Stoter, Burghardt, et al. 2009). Another reason is that due to the time saved by automation, it is possible to generalize various products meeting different demands. Therefore, the specifications do not longer have to define one product that should serve all needs and uses.

A number of aspects further refine the scope of our research:

- The focus is on producing a map. Therefore, disruptions of the geometry to meet cartographic requirements do not have to be controlled apart from ensuring their consistency in the resulting map, i.e., roads and water should still form networks after generalization. To accomplish keeping the networks, the workflow performs generalization using 'smart' partitioning, instead of performing generalization in map sheets.
- The first focus is on generalizing a 1:50k map from 1:10k data (see Figure 1). Other scales will follow in the (near) future. Both the source and the target map cover a planar partition (i.e., no gaps or overlaps are allowed).
- Generalization without any interaction is the best guarantee for efficiency and consistency and the only way to produce on-demand products. Therefore, the output is generalized fully automatically and it is not allowed to improve the results afterwards by human interaction.
- The most straightforward way for updates in a fully automated generalization workflow is to completely replace the old version map, in line with Regnauld (2011). This also makes it possible to adjust a next version of the map to developing technologies and data models. Therefore, we do not maintain links between the objects at the two scale levels currently employed. If these links are required they will be part of a subsequent study.
- The scope is limited by static visualizations of the map on screen or on paper. Dynamic generalization for Internet use (real time and not *per se* related to the predefined scales) has different requirements but also different solutions and is therefore outside the scope of this research.

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

Figure 1. Source data (1:10k) and target map (1:50k) shows how much generalization needs to be applied.

Methodology to implement an automated generalization process

Our research employed an empirical approach. We generated an initial 1:50k map in a semi-automatic manner by extending the work of Stoter, van Smaalen, et al. (2009, 2012) to show how much automation can be achieved with current technology. This work implemented existing guidelines for interactive generalization as a set of automated steps. Comparing these initial results with existing maps provided insights into how cartographers interpret the guidelines. These insights were then used to improve the process and to determine the optimal approach, algorithms, and desired parameter values. If needed, the steps were prototyped with different alternatives to identify the best results. In addition, the intermediate results also helped to identify the optimal order of steps and operators and to discriminate between different contexts.

The implementation relied on a set of standard ArcGIS tools, self-developed tools within Python, and several FME tools. ArcGIS version 10 (Punt and Watkins 2010; ESRI, Redlands, CA, USA) contains some specialized generalization tools for collapsing two lanes of a road into a single road line, displacing symbolized geometries, simplifying (symbolized) buildings, and thinning out of networks. Other commercially available generalization software systems are Change/Push/Typify (University of Hannover, Hannover, Germany), Axpand (Axes Systems, Switzerland), and Clarity (Cambridge, UK; see also Stoter et al. 2010).

The complete generalization workflow is implemented within the Model builder tool of ArcGIS. The resulting workflow consists of three main models, each consisting of about 200 sub-models that are responsible for solving each specific generalization problem in the process.

The remainder of this section explains some other characteristics of the methodology, including enrichment of source data, the role of constraints and redefining map specifications for automated generalization with the help of users.

Enrichment of source data

Since the aim is full automation, “touching” the output map afterwards is disallowed. In our research, a fully automated generalization process was realized by improving the process step-by-step, or, if that did not work, by improving and enriching the source data. Many others have considered enriching source data and making implicit information explicit as important step for automated generalization. Examples are Bobzien et al. (2008), Chaudhry and Mackaness (2006a), López and Balboa (2008), Anderson-Tarver et al. (2011, 2012), Neun, Burghardt, and Weibel (2008), Steiniger et al. (2008), and Zhang et al. (2013a, 2013b).

Our approach enriches the source data in two ways: 1. external data sources are used to obtain the knowledge or 2. the required knowledge is made explicit by computation (for more information, see further).

Constraints

We applied constraints in a different manner than is common in constraint-based generalizations. The constraint-based approach is a shared method, used to express user requirements and to control and evaluate the automated generalization process. See, for example, McMaster and Shea (1988), Beard (1991), Ruas (1999), Bard (2004), Barrault et al. (2001), Ware, Jones, and Thomas (2003), Burghardt and Neun (2006), and Stern and Sester (2012).

We use constraints in our research, but we do not use them to automatically control and evaluate the automated generalization process as is done with agent-based technologies (Ruas 1999; Ruas and Duchene 2007). Instead, the constraints are formulated in terms of new map specifications that need to be addressed by the workflow, while iterative controlling and evaluation of the process was used to obtain the best generalization workflow. This process made it possible to define and adjust the map specifications (i.e., constraints) as part of the process.

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

Defining map specifications for automated generalization

One of the main challenges was defining new specifications for automated generalization taking existing guidelines as starting point, while ensuring both that user requirements are met and the new specifications can be implemented in a fully automated workflow.

Various researchers have studied (new and changing) specifications for automated map generalization. Müller and Mouwes (1990) examined existing map series and concluded that "superficial" generalization knowledge exists in the form of map specifications written down for interactive generalization. Rieger and Coulson (1993) carried out a survey among a group of cartographers performing interactive generalization and concluded that a common view on the classification of generalization operators does not exist. Nickerson (1991) and Kilpelaïnen (2000) acquired knowledge from experts to define rules for knowledge-based map generalization. Various studies used reverse engineering to collect generalization knowledge by comparing map objects across scales (Buttenfield 1991; Leitner and Buttenfield 1995; Weibel 1995). Other studies describe methods to generate rules from interactive generalization carried out by a cartographic expert (Weibel 1991, 1995; McMaster 1995; Reichenbacher 1995). Several studies applied machine learning techniques to convert expert knowledge into map specifications for automated generalization, e.g., Weibel (1995), Plazanet, Bigolin, and Ruas (1998), Mustiere (2001, 2005), and Hubert and Ruas (2003). Brewer and Buttenfield (2007) ran map exercises with students on different data sets at various scales to provide guidelines for generalization processes. Taillandier and Gaffuri (2011) propose a human-machine dialog to evaluate automated generalization samples and to improve specifications accordingly.

In our research, users were involved in several ways. Initial results were sent to a selection of main customers of the current 1:50k map (who are formally organized in a users group) in order to test the main principles and assumptions. The insights obtained were used to improve the generalization process. The evaluation and improvement process was then repeated by asking more (potential) users to evaluate the resulting map in more detail and for different types of areas, i.e., "Relief and dense road pattern," "Complicated crossings and dense parcel boundaries," "Dense water network," and "Urban and industrial area."

Interestingly, the evaluations showed that users appreciated the "same appearance of the map" less than "more frequent update cycles". Indeed, they expressed to be very pleased that the 1:50k maps will be consistent with the 1:10k source data. This level of consistency is achievable because it takes approximately three weeks to obtain a complete 1:50k map from updated 1:10k data. Therefore, a new version of a 1:50k map will be released shortly after a new release of 1:10k data. This will avoid the 1:10k data containing updates which are not yet visible in the 1:50k map.

Another interesting observation made on the basis of the users' evaluation was that some results of the automated generalization were appreciated more than the results of interactive generalization. For example, the automatically thinned road network appeared to be more appropriate for navigation than the interactively thinned road network. In addition, several respondents expressed satisfaction with the improved uniformity of the entire map.

Besides validation by users, an internal validation was performed to ensure that the target map still meets minimal conditions despite the fact that the resulting map is a different product. These quality aspects which include keeping consistency (i.e., objects should keep their relative positions) and preservation of the main characteristics of the initial data (i.e., the target data should still reflect reality to some extent) cannot all be evaluated by users. Therefore, this evaluation task was performed internally by cartographers who pointed at errors as well as situations that they would have solved differently but are not *per se* an error. Cartographic experts also helped the technical team with the explanation and interpretation of the cartographic rules that were used in the manual process and served as guidelines for the automatic generalization.

After two years of (prototype) implementation, iterative testing, and user evaluations, the research has resulted in a fully automated generalization process that went into production in 2013.

Automated workflow

This section describes the resulting workflow that implements the 100% fully automated generalization process employed to produce a countrywide 1:50k map from 1:10k data (i.e., TOP10NL, the nationwide object-oriented data set at the scale 1:10k). The workflow consists of the following four steps (see Figure 2):

- Pre-processing.
- Model generalization aiming at reducing the data that has to be visualized.
- Symbolization of the data.
- Graphic generalization to solve cartographic conflicts of symbolized objects.

The partitioning that we applied to be able to process the complete nationwide data set of the Netherlands in one run is described in a later section.

Pre-processing the data

Our experience indicates that the more rich, consistent, correct etc. the input data, the better results can be

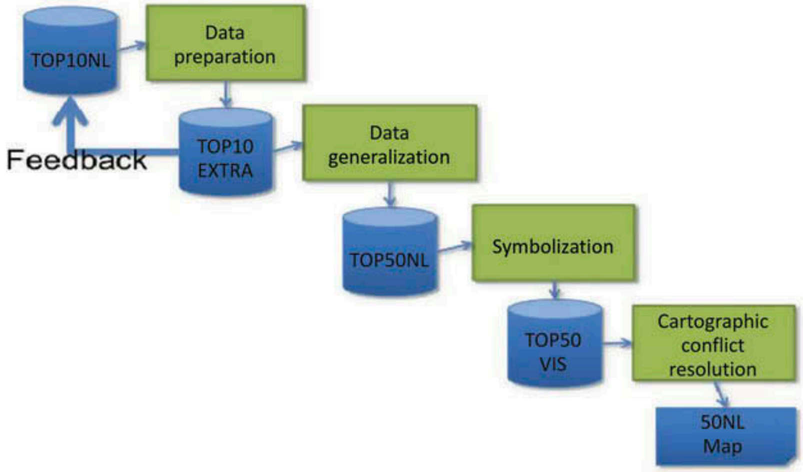


Figure 2. Automated workflow for generalization process 1:50k map (can be repeated for maps at smaller scales).

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

obtained. Therefore, we gave considerable attention to enhancing the input data by correcting (hidden) errors and the enrichment of the source data. The latter is done in two ways. Either external data sources are used or the required knowledge is made explicit by computation. In addition to this enrichment process that results in newly encoded information, dynamic enrichments are also obtained, i.e., calculated on-the-fly each time the process is run to automatically detect specific situations.

Examples of enrichments of the input data that resulted in newly encoded information are as follows:

- Urban extents are identified by calculating areas with high density of buildings (i.e., more than 15%).
- Industrial areas are obtained from the statistic bureau of the Netherlands, and added.
- TOP10NL road segments have been attributed with information on exits to better control the process that generalizes the road network from the TOP10NL road polygons. Since the process of generalizing a good road network was not trivial and because the generalized network is beneficial for all mid-scale and smaller-scale maps as well, it was decided to enrich the source data with this new information.
- Pruning of the (artificial) water networks. This appeared to be easier with drainage information attached to the water objects. Therefore, drainage information was added to TOP10NL data obtained from water management organizations.
- The proper placement of buildings in a limited amount of space is improved by adding extra information to building objects. For example, by assigning lower importance rates to small (calculated) and unimportant buildings such as sheds (derived from the function of the building).

The resulting enhanced database (TOP10Extra shown in Figure 2) is the input for the automated generalization process of several smaller scales.

Model generalization

The model generalization process consists of a set of steps the optimal order and implementation of which was determined by iterative testing.

Step 1: Reclassify all TOP10NL classes, attributes, and attribute values into TOP50 classes, attributes, and attribute values. This also involves aggregation, such as combining neighboring deciduous and pine forest in one object of class "forest".

Step 2: Generalize TOP10NL road centerlines into centerlines appropriate for scale 1:50k. Road centerlines in TOP10NL (additional to the polygon geometries that are used for the map) represent single lanes. Road centerlines at scale 1:50k cover complete roads (see Figure 3).

The TOP50 centerlines are generated with an algorithm, available in ArcGIS, which merges two lanes of a single

Figure 3. Centerlines in TOP10NL (left) and TOP50 (right), both projected on TOP10NL polygons.

road into one centerline. Merging is also done when a verge divides the lanes (Figure 4). Other approaches were considered, such as generating centerlines from the road polygons, but these gave unsatisfying results, specifically at crossings.

Step 3: Extend terrain parcels to road centerlines by an algorithm developed in Stoter, van Smaalen, et al. (2009). Due to replacing road polygons with road centerlines, gaps exist in a small buffer around the road centerlines. The original road areas need to be assigned to the neighboring area objects to ensure complete coverage of topographic classes. This is done by extending terrain parcels to the new centerlines (see Figure 5).

Step 4: Convert areas with many buildings to built-up areas and remove the original building objects in those areas. Buildings are only converted to built-up area if they are located within urban areas. In addition, important buildings such as schools, hospitals, and churches are kept as separate point symbols.

Step 5: Prune the road network, see also Chaudhry and Mackaness (2006b), Thom (2007), and Thompson and Richardson (1999). This is not straightforward, specifically because TOP10NL data does not contain many attributes for pruning. The pruning of the road network consists of several steps (see Figure 6 for results).

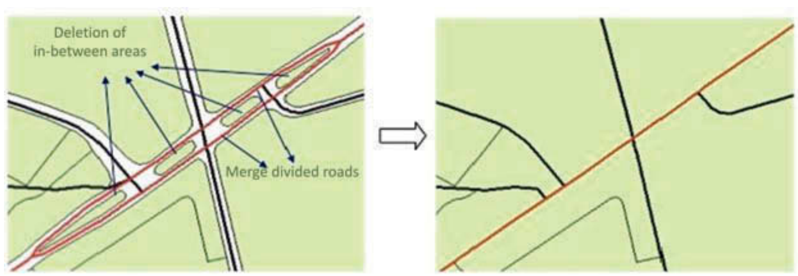


Figure 4. Merging divided roads that enclose verges.

Figure 5. Extending parcel boundaries to the new road centerlines. (a) Original 1:10k polygons; (b) gaps in the data due to replacing road polygons with lines; and (c) extending parcel boundaries to fill the gaps.

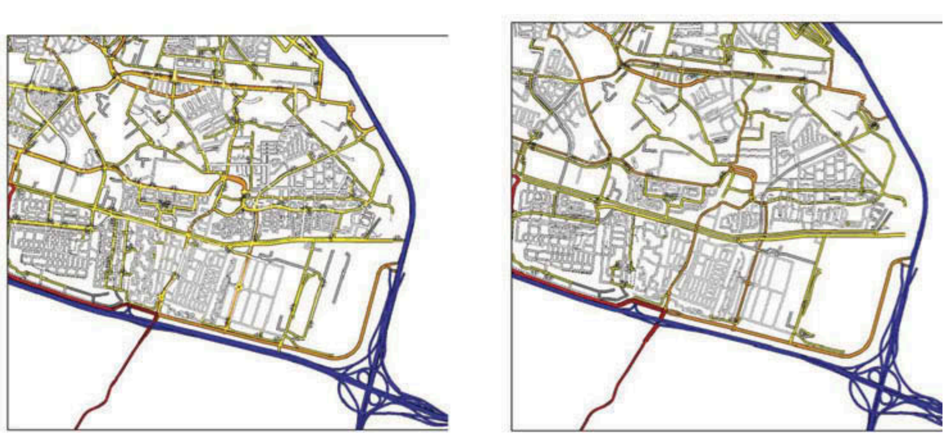


Figure 6. Thinning of the road network before (left) and after (right) generalization.

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

(1) Cycle paths parallel to (i.e., "touching") roads are deleted; free cycle paths are maintained.
(2) Access roads to buildings in rural areas are detected and selected. Automated detection finds access roads "if buildings are located within 200 meters of the end of the road AND no other roads are located in the neighborhood (within 200 meters) of the building."
(3) The remaining road network is pruned by the "thin road network algorithm" available in ArcGIS. This algorithm retains connectivity and the general character of the network while using a hierarchy for the relative importance of a road and its minimum length. We did not further use a self-developed algorithm based on road continuity (see Stoter, van Smaalen, et al. 2009).

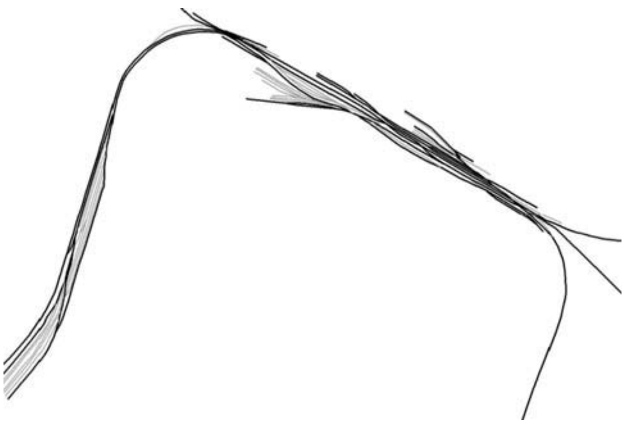


Figure 7. Simplyfing the railway network (orginal data in gray; simplified network in black).

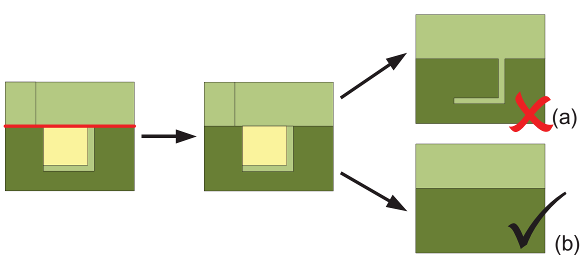


Figure 8. Narrow areas (although sufficient in size) are detected and deleted or amalgamated.

Step 6: Prune water network, first by removing small water parallel to (i.e., "touching") roads and then by applying the thin road network algorithm on the linear waterways. Lake shorelines and other water area boundaries are added in this process to maintain connectivity. We use the thin road network algorithm on the waterways because the Dutch water network is almost completely man-made and its structure resembles a road network rather than a natural water network. Therefore, algorithms available for pruning natural hydrographic networks, as studied, for example, in Stanislawski and Savino (2011), are less appropriate.

Step 7: Detect the main railway network by a routing algorithm and thinning the railway tracks at yards by keeping the outer ones (see Figure 7).

Step 8: Generalize small areas of forest, water, and other terrain types. These objects can either be too small or too narrow (see Figure 8). Narrow areas are determined by calculating the ratio area/perimeter.

If these small and narrow areas are isolated, these are deleted; if not, then these are amalgamated. The amalgamation takes into account that areas should not be amalgamated over physical linear boundaries as water, road etc., i.e., only amalgamate if the areas are in the same partition (see Figure 9).

Step 9: Reduction of data by filtering the vertices of all linear geometries (linear objects and polygon boundaries) using the Douglas-Peucker algorithm (parameter value 1 m).

Following model generalization, symbols are assigned to all geometries. Symbolization sometimes results in objects that appear larger on the map than they are in reality. Symbolization therefore steers the next graphic generalization process that solves the cartographic conflicts. The basis symbols used in this process correspond exactly to shape and

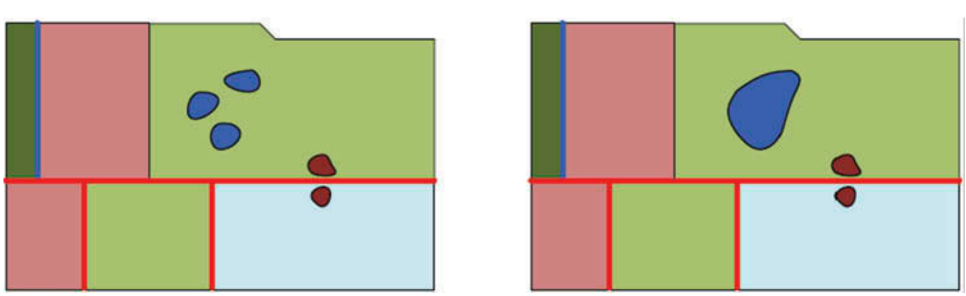


Figure 9. Amalgamation of small water areas (blue), not over hard topography (red roads).

outline of portrayed features, but they lack all cartographic refinement. Sophisticated symbolization and cartographic enhancement (such as aligning symbols to other features to enhance legibility) are dealt with at a later stage in the process.

Graphic generalization

The graphic generalization process consists of the following three steps.

Step 1: Generalize remaining buildings in order to meet minimum building size and avoid overlap. At first, small details of buildings are deleted with an ArcGIS algorithm. This simplifies the boundary or footprint of building polygons while maintaining their essential shape and size (see Figure 10(b)). Second, an algorithm (also available in ArcGIS) is applied to remove graphic conflicts of symbolized buildings. This helps meeting the minimum distance between buildings and between buildings and other surrounding objects (see Figure 10(c)). The algorithm, explained in Punt and Watkins (2010), resolves symbol conflicts applying an optimization technique. The algorithm combines object removal and emphasizing operations. The position, orientation, and size of the objects are adjusted in order to avoid overlap or violation of spacing requirements, but the representative pattern and distribution are maintained.

Step 2: Displace linear objects (railways, roads, water), boundaries of symbolized water and terrain objects, as well as all other point and linear objects (e.g., administrative boundaries, height contours, engineering constructs) with an algorithm that displaces symbolized objects and reshapes them in order to avoid overlap (see Figure 11).

Displacing the boundaries of polygons and rebuilding the polygons afterwards (see next step) showed better results for a planar partition (used in our case) than displacing the polygons themselves. See Smaalen, Nijhuis, and Stoter (2011) for more details on these experiments. The displacement algorithm (available in ArcGIS 10.0) is implemented via an optimization method using a hierarchy of object types. In this process, main roads and highways may never be displaced, since these are used to generate partitions (see next section).

Step 3: Rebuild terrain and water polygon objects from the displaced boundaries and assign former codes to the new areas by using left/right information of the boundaries.

Partitioning and updating

To obtain a generalized map for the whole of The Netherlands with the above workflow requires partitioning. We use the main road network to partition the country to perform countrywide generalization. The main roads are often physical boundaries in the real world and therefore it

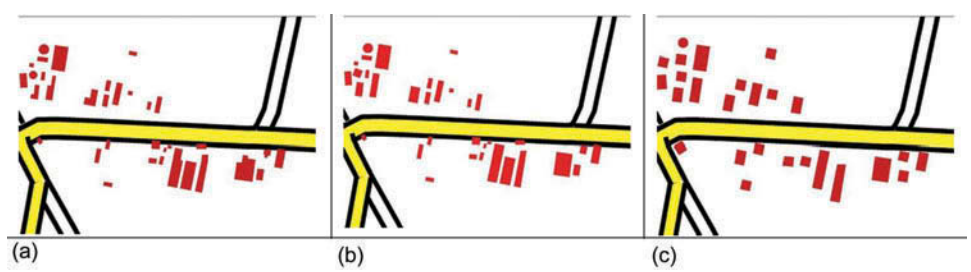


Figure 10. Simplification and displacements of buildings. (a) Source data, displayed at target scale; (b) simplification of buildings; and (c) displacement.

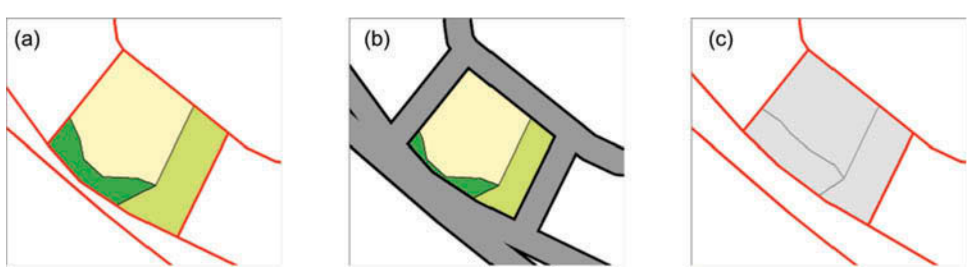


Figure 11. Displacement to avoid overlap of symbolized linear objects. (a) Input lines and polygons; (b) symbolized polygon boundaries (enforces minimal size of polygons) and lines; and (c) displacement of lines and polygon boundaries.

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

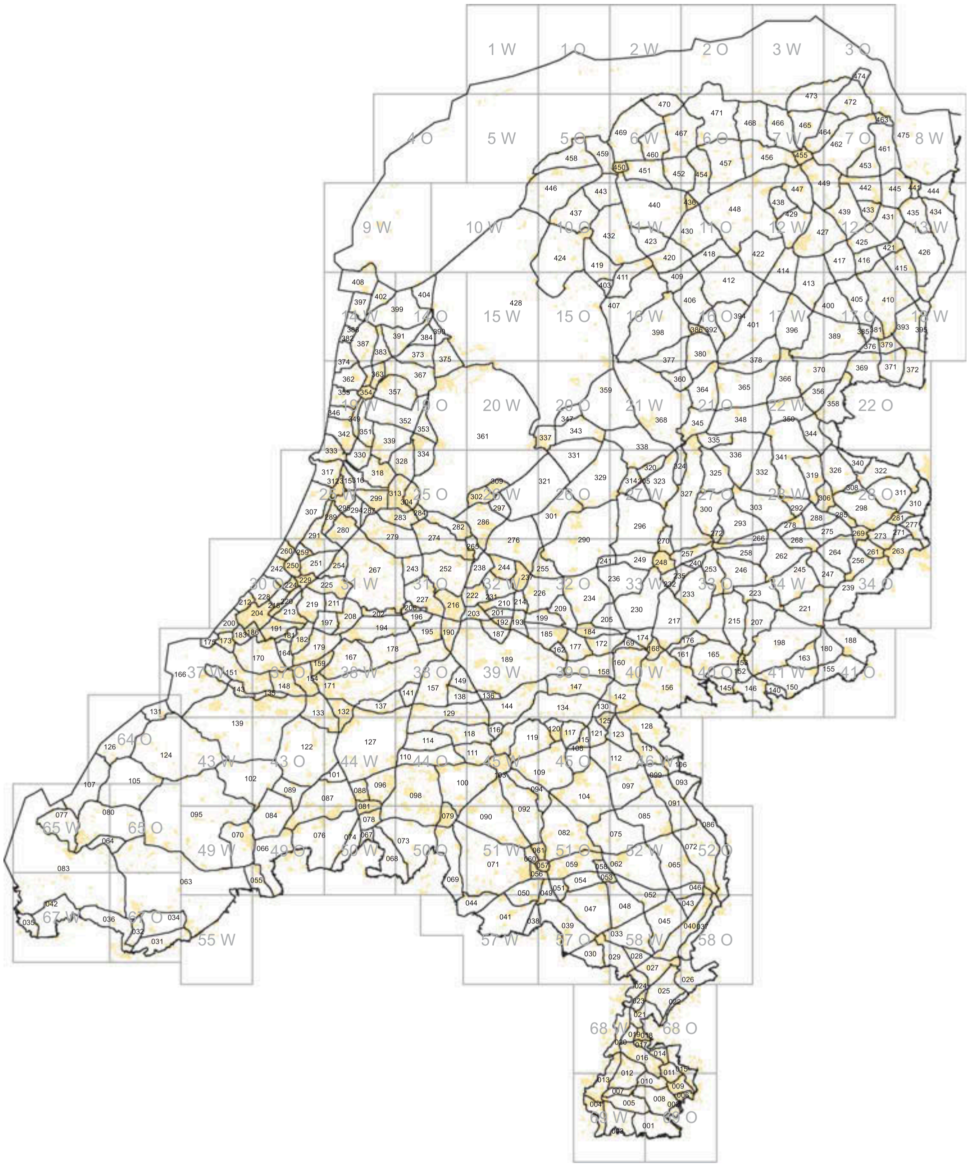


Figure 12. Partitioning of the Netherlands to enable automated generalization of a seamless map.

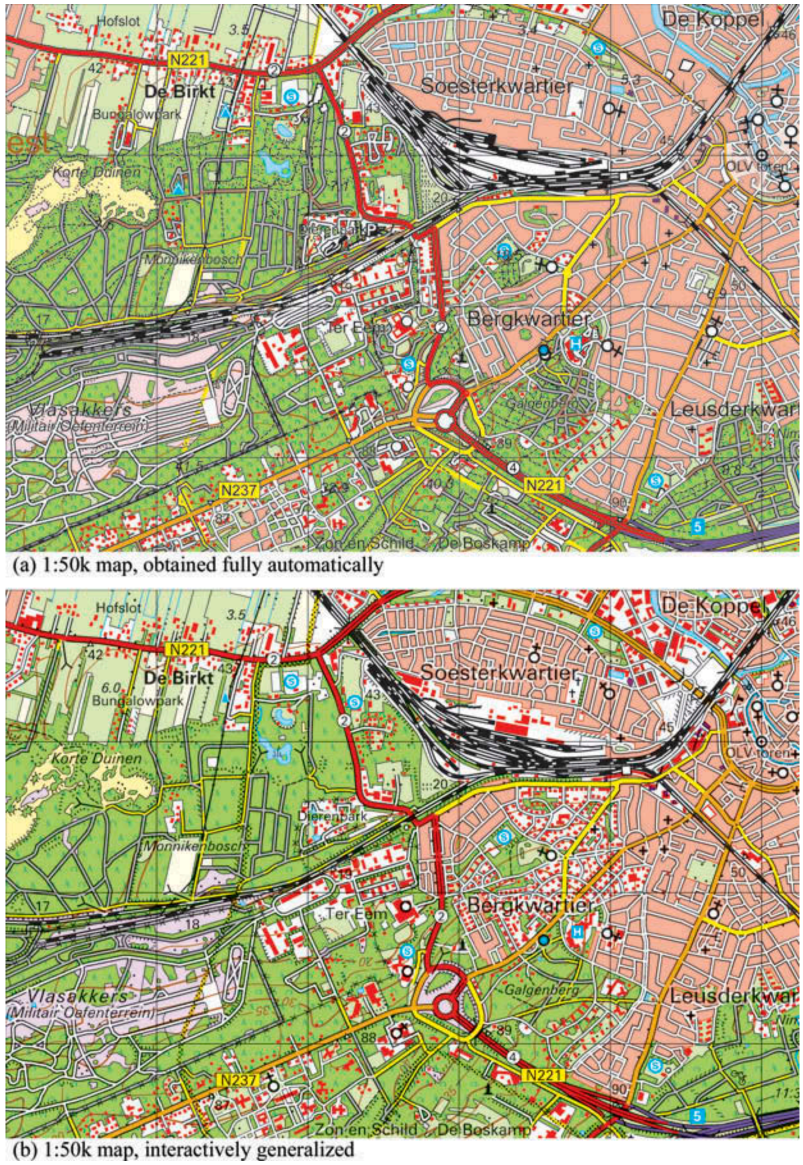
is unlikely that these linear boundaries would go through (divide) any object. Using these real-world boundaries as partition boundaries makes it less complicated to keep the topology at partition boundaries.

Near the coast, where roads are missing to demarcate the areas, artificial partitions have been made. The resulting 460 areas are shown in Figure 12. Despite some global operations that were applied to the country as a whole (such as creating and simplifying the road network), the workflow was applied per partition, and partitions were connected afterwards. Because our process prohibited to move vertices of objects at and near partition boundaries in the displacement process, the objects at neighboring partitions still fit after generalization.

Processing the generalization for the whole country in a reasonable time is achieved by multiprocessing capabilities within Python. It processes six partitions in parallel on each of the six available systems, i.e., 36 partitions at the same time.

Results and findings

Figure 13(a) shows the 1:50k map that was generalized automatically from the data shown in Figure 1(a) with the workflow described in the previous section. Figure 13(b) shows the interactively generalized map of the same area (same area shown in Figure 1(b)). The two maps in Figure 13 should be compared with caution, since the automated generalization process is not meant to replicate



(a) 1:50k map, obtained fully automatically

(b) 1:50k map, interactively generalized

Figure 13. A map generalized by an automated workflow process (a) compared with an interactively generalized map (b) of the same area.

the existing maps. Another reason for caution is that names are not generalized in Figure 13(a).

However, differences also exist because of the changed map specifications. For example, the interactive generalization guidelines prescribe that detached houses may never be converted into built-up areas. It was impossible to always meet this rule, because detached houses are often enlarged to meet the minimum size (15×15 m). If building blocks are at least 40 m, it would be possible to accommodate enlarged detached houses at a minimum distance of 10 m (required for readability). But this is rarely the case. At the same time, streets in TOP50 are symbolized with line widths of 20 m, which is wider than these streets are in reality. Because our users did not consider the conversion of detached houses into built-up areas as problematic, we decided to always convert detached houses in built-up areas when the building density threshold was exceeded (see Figure 14). The same applies to buildings in industrial areas; in contrast to the interactively generalized map, buildings in industrial areas were converted into built-up areas by the fully automated generalized process because it was too difficult to treat the buildings as individual objects.

Another issue that was solved differently by the fully automated process is that ditches were treated as ordinary terrain boundaries because of their limited importance and because they can be visualized as terrain boundaries. In addition, all water with linear

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

(a) Automatically generalized map (b) Interactively generalized map

Figure 14. The difference between the interactive (b) and automated generalized map (a) for converting single houses into built-up areas.

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

geometries in TOP10NL (i.e., smaller than two meters) parallel to roads was eliminated in the new map to save space.

Finally, unlike the rule in the interactively generalized map that prohibits dikes to be displaced, dikes were displaced in the automatically generalized map. This is because damming dikes are not available as single objects in TOP10NL; they are represented by hatches on the map. Since the users' consult showed the unimportance of this rule in combination with the difficulty to identify dikes, it was decided to ignore this guideline. The 1:50k map does show dikes (with hatches), but they may have been moved to make space for other objects.

A main question is how does the integrated workflow respect the context of the data, which is important in automated generalization. Context-aware generalization is implemented in our workflow, even though the same workflow is used for the entire area. The context is taken into account in the several sub-models. For example, buildings in urban, rural, and industrial areas are generalized differently; water and cycle paths parallel to roads are detected and generalized differently from the other waters and cycle paths; dead-end roads that lead to buildings are generalized in a specific manner; land-use areas that are divided by a road are not amalgamated etc.

Conclusions

This article presents findings about a fully automated generalization process of 1:10k data into a 1:50k map. The primary reason for using fully automated generalization is to meet a fast update cycle of source data in a cost-effective manner. Ability to deliver more flexible on-demand products is another important reason. Early in the process it was concluded that 100% automation is only possible if current products are reconsidered and changed to meet new requirements with available technologies. Therefore, users of multi-scale data and maps were closely involved in assessing the intermediate results of the generalization process and to formulate usage criteria for the new 1:50k map.

Although other Mapping Agencies have developed automated processes and even though we use existing tools, the significant contribution of this research is that fully automated generalization could be achieved because map specifications were adjusted to meet technological possibilities. Automated generalization by most other NMAs still require interaction in some parts of the process and, therefore, the time saved of our workflow is significant.

Based on the results and favorable users' feedback, the Netherlands Kadaster decided that a fully automated generalization workflow that produces a renewed map is the most appropriate workflow for producing up-to-date and on-demand map products. Consequently, automatically generalized maps replace existing maps from 2013.

As mentioned, the thirty-six parallel processes in Python are able to generalize the complete Netherlands in about 50 hours (www.python.org). Including pre-processing, generalization, visualization, and printing, the whole turnaround is 3 weeks for the whole country. Therefore, a 1:50k update is foreseen with every new delivery of TOP10NL (every two months).

On-going and future work will extend the automated generalization approach we detail in this article to the 1:100k map and to on-demand products. Since some basic concepts have been tackled in the source data set (i.e., merging of divided roads; calculation of urban areas and railway tracks; adding extra attributes for water, roads, and buildings), these can be reused for other multi-scale products.

A few issues require further attention to make the workflow work at any scale and for other multi-scale products, apart from improving specific algorithms to obtain better results.

The road-thinning algorithm is sufficient to generalize water for 1:50k representations. However, at a scale of 1:100k, narrow waterways (under 40 m) may need to be merged into lines (i.e., collapsed), which brings extra challenges for keeping a good water network, apart from the lack of a thorough centerline tool.

The handling of updates needs to be studied further. Some of the issues that could be addressed are as follows: what would be the impact of replacing an entire area during an update and what the users' perspective on this might be; is it possible to keep unaltered areas in the map as part of the automated process?

In conclusion, implementing an automated generalization workflow brings a revolutionary change in the products and the map production workflow itself. This research shows automated map production is a development, which will have a significant impact on the provision of geographical information in the future.

Acknowledgments

This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs (project code: 11300).

References

Anderson-Tarver, C. A., B. Buttenfield, L. Stanislawski, and J. Koontz. 2011. "Automated Delineation of Stream Centerlines for the USGS National Hydrography Dataset." In *Advances in Cartography and Geographic Information Science*, edited by A. Ruas, Lecture Notes in Geoinformation and Cartography, Vol. 1, 409–423. Berlin: Springer-Verlag.

Anderson-Tarver, C. A., M. Gleason, B. Buttenfield, and L. Stanislawski. 2012. "Automated Centerline Delineation to Enrich the National Hydrography Dataset." In *Geographic Information Science,* edited by N. Xiao, M.-P. Kwan, M. F. Goodchild, and S. Shekhar, Lecture Notes in Computer Science, Vol. 7478, 15–28. Berlin: Springer-Verlag.

Baella, B., and M. Pla. 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.

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

Barrault, M., N. Regnauld, C. Duchêne, K. Haire, C. Baeijs, Y. Demazeau, P. Hardy, W. Mackaness, A. Ruas, and R. Weibel. 2001. "Integrating Multi-Agent, Object-Oriented, and Algorithmic Techniques for Improved Automated Map Generalisation." In *Proceedings of the 20th International Cartographic Conference (ICC 2001)*, Beijing, August 6–10, 2110–2116. International Cartographic Association, CD-ROM.

Beard, M. K. 1991. "Constraints on Rule Formation." In *Map Generalisation: Making Rules for Knowledge Representation*, edited by B. P. Buttenfield and R. B. McMaster, 121–135. London: Longman Group. ISBN: 0–582–08062–2.

Bobzien, M., D. Burghardt, I. Petzold, M. Neun, and R. Weibel. 2008. "Multi-Representation Databases with Explicitly Modeled Horizontal, Vertical, and Update Relations." *Cartography and Geographic Information Science* 35 (1): 3–16.

Brewer, C. A., and B. Buttenfield. 2007. "Framing Guidelines for Multi-Scale Map Design Using Databases at Multiple Resolutions." *Cartography and GIS* 34 (1): 3–15.

Burghardt, D., and M. Neun. 2006. "Automated Sequencing of Generalisation Services Based on Collaborative Filtering." In *Geographic Information Science,* edited by M. Raubal, H. J. Miller, A. U. Frank, and M. Goodchild, 4th International Conference, GIScience 2006, IfGIprints 28, 41–46. Munster: GIS.

Buttenfield, B. P. 1991. "A Rule for Describing Line Feature Geometry." In *Map Generalisation: Making Rules for Knowledge Representation*, edited by B. P. Buttenfield and R. B. McMaster, 150–171. Essex: Longman.

Chaudhry, O. Z., and W. A. Mackaness. 2006a. "Modelling Geographic Phenomena at Multiple Levels of Detail." In AutoCarto, Vancouver, WA, June 26–28.

Chaudhry, O. Z., and W. A. Mackaness. 2006b. "Rural and Urban Road Network Generalization Deriving 1:250,000 from 1:1250." In *International Cartographic Conference*, 10 p., La Coruña , July 10–16.

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

Hubert, F., and A. Ruas. 2003. "A Method Based on Samples to Capture User Needs for Generalisation." In *5th ICA Workshop on Progress in Automated Map Generalisation*, Paris, 2003. http://generalisation.icaci.org/images/files/workshop/workshop2003/hubert_et_al_v0.pdf

Kilpelaïnen, T. 2000. "Knowledge Acquisition for Generalisation Rules." *Cartography and Geographic Information Science* 27 (1): 41–50.

Lecordix, F., J. L. Gallic, L. Gondol, and A. Braun. 2007. "Development of a New Generalization Flowline for Topographic Maps." In *10th ICA workshop on Generalisation and Multiple Representation*. Moscow, August 2–3.

Leitner, M., and B. Buttenfield. 1995. "Acquisition of Procedural Cartographic Knowledge by Reverse Engineering." *Cartography and Geographic Information Systems* 22 (3): 232–241.

López, F. J. A., and J. L. G. Balboa. 2008. "Generalization-Oriented Road Line Segmentation by Means of an Artificial Neural Network Applied over a Moving Window." *Pattern Recognition* 41 (5): 1593–1609.

McMaster, R. B. 1995. "Knowledge Acquisition for Cartographic Generalisation." In *GIS and Generalisation: Methodology and Practice*, edited by J. C. Mueller, J. P. Lagrange, and R. Weibel, 161–179. London: Taylor & Francis.

McMaster, R. B., and K. S. Shea. 1988. "Cartographic Generalisation in a Digital Environment: A Framework for Implementation in a Geographic Information System." In *GIS/LIS Proceedings*, 240–249. San Antonio, TX: American Society for Photogrammetry and Remote Sensing.

Müller, J. C., and P. J. Mouwes. 1990. "Knowledge Acquisition and Representation for Rule Based Map Generalisation: An Example from the Netherlands." In *GIS/LIS Proceedings 90*, Vol. 1, 58–67. Anaheim, CA: American Society for Photogrammetry and Remote Sensing.

Mustière, S. 2001. "Apprentissage Supervisé Pour La Généralisation Cartographique." Thèse de doctorat, Université Paris VI.

Mustière, S. 2005. "Cartographic Generalisation of Roads in a Local and Adaptive Approach: A Knowledge Acquisition Problem." *International Journal of Geographical Information Science* 19 (8–9): 937–955.

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

Downloaded by [Bibliotheek TU Delft] at 08:26 04 March 2015

Neun, M., D. Burghardt, and R. Weibel. 2008. "Web Service Approaches for Providing Structural Cartographic Knowledge to Generalisation Operators." *International Journal of Geographical Information Science* 22 (2): 133–165.

Nickerson, B. G. 1991. "Knowledge Engineering for Generalisation." In *Map Generalisation: Making Rules for Knowledge Representation*, edited by B. Buttenfield and R. B. McMaster, 40–55. London: Longman.

Plazanet, C., N. Bigolin, and A. Ruas. 1998. "Experiments with Learning Techniques for Spatial Model Enrichment and Line Generalisation." *Geoinformatica* 2 (4): 315–333.

Punt, E., and D. Watkins. 2010. "User-directed Generalization of Roads and Buildings for Multi-Scale Topography." In *13th ICA Workshop on Generalisation and Multiple Representation*, 2010. Zurich.

Regnauld, N. 2011. "OS Vectormap 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

Reichenbacher, T. 1995. "Knowledge Acquisition in Map Generalisation Using Interactive Systems and Machine Learning." In *Proceedings of the 17th International Cartographic Conference*, 2221–2230. Barcelona: International Cartographic Association.

Rieger, M. K., and M. R. C. Coulson. 1993. "Consensus or Confusion: Cartographers' Knowledge of Generalisation." *Cartographica* 30 (2 and 3): 69–80.

Ruas, A. 1999. "Modèle De Généralisation De Données Géographiques À Base De Contraintes Et D'Autonomie." Doctoral thesis, Université de Marne-la-Vallée.

Ruas, A., and C. Duchêne. 2007. "A Prototype Generalisation System Based on Multi-Agent System Paradigm." In *Generalisation of Geographic Information: Cartographic Modelling and Applications*, edited by W. Mackaness, A. Ruas, and L. T. Sarjakoski, Kidlington: Elsevier.

Smaalen, J., R. Nijhuis, and J. E. Stoter. 2011. "Automated Generalisation of Land Cover Data in a Planar Topographic Map." In 14th Workshop of the ICA Commission on Generalisation and Multiple Representation, Paris, June 2011.

Stanislawski, L., and S. Savino. 2011. "Pruning of Hydrographic Networks: A Comparison of Two Approaches." In *14th Workshop of the ICA Commission on Generalisation and Multiple Representation*, Paris, June 2011. http://generalisation.icaci.org/images/files/workshop/workshop2011/genemr2011_Stanislawski.pdf

Steiniger, S., T. Lange, D. Burghardt, and R. Weibel. 2008. "An Approach for the Classification of Urban Building Structures based on Discriminant Analysis Techniques." *Transactions in GIS* 12 (1): 31–59.

Stern, C., and M. Sester. 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.

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

Stoter, J. E., D. Burghardt, C. Duchêne, B. Baella, N. Bakker, C. Blok, M. Pla, N. Regnauld, G. Touya, and S. Schmid. 2009. "Methodology for Evaluating Automated Map Generalization in Commercial Software." *Computers, Environment and Urban Systems* 33 (5): 311–324.

Stoter, J. E., J. van Smaalen, N. Bakker, and P. Hardy. 2009. "Specifying Map Requirements for Automated Generalisation of Topographic Data." *The Cartographic Journal* 46 (3): 214–227.

Stoter, J. E., J. van Smaalen, R. Nijhuis, A. Dortland, J. Bulder, and B. Bruns. 2012. "Fully Automated Generalisation of Topographic Data in Current Geo-Information Environments." In *Urban and Regional Data Management – UDMS Annual 2012*, edited by S. Zlatanova, H. Ledoux, E. M. Fendel, and M. Rumor, 111–121. Leiden: CRC Press, Taylor & Francis Group.

Taillandier, P., and J. Gaffuri. 2011. "Using Human-Machine Dialogue to Refine Generalisation Evaluation Function." In *Conference of the International Cartographic Association*, Paris, July. http://hal.upmc.fr/docs/00/68/84/25/PDF/ICC-2011_Taillandier-Gaffuri.pdf.

Thom, S. 2007. "Automatic Resolution of Road Network Conflicts Using Displacement Algorithms Orchestrated by Software Agents." In *10th ICA Workshop on Generalization and Multiple Representation*. Moscow, August 2–3.

Thompson, R. C., and D. E. Richardson. 1999. "The Good Continuation Principle of Perceptual Organization Applied to the Generalization of Road Networks." In *Proceedings of the ICA 19th International Cartographic Conference*, Ottawa, August 14–21, 1215–1223. International Cartographic Association.

Ware, J. M., C. B. Jones, and N. Thomas. 2003. "Automated Map Generalisation with Multiple Operators: A Simulated Annealing Approach." *International Journal of Geographical Information Science* 17 (8): 743–769.

Weibel, R. 1991. "Amplified Intelligence and Rule-Based Systems." In *Map Generalisation: Making Rules for Knowledge Representation*, edited by B. P. Buttenfield, and R. B. McMaster, 172–186. London: Longman.

Weibel, R. 1995. "Three Essential Building Blocks for Automated Generalisation." In *GIS and Generalisation: Methodology and Practice*, edited by J. Mueller, J. P. Lagrange, and R. Weibel, 56–70. London: Taylor & Francis.

Zhang, X. T., A. J. Stoter, M. J. Kraak, and M. Molenaar. 2013a. "Building Pattern Recognition in Topographic Data: Examples on Collinear and Curvilinear Alignments." *Geoinformatica* 17: 1–33. http://dx.doi.org/10.1007/s10707-011-0146-3.

Zhang, X., J. E. Stoter, A. Tingha, M. Molenaar, and M. J. Kraak. 2013b. "Automated Evaluation of Building Alignments in Generalized Maps." *International Journal of Geographical Information Science*. doi:10.1080/13658816.2012.758264.