The automatic generalisation of building polygons with ArcGIS standard tools based on the 1:50`000 Swiss National Map Series

A. Vetter[1], M. Wigley[1], D. Käuferle[2], G. Gartner[3]

[1] Esri Switzerland Ltd., Zurich, Switzerland (a.vetter@esri.ch, m.wigley@esri.ch)
[2] Federal Office of Topography swisstopo, Wabern, Switzerland (Dominik.Kaeuferle@swisstopo.ch)
[3] Vienna University of Technology, Vienna, Austria (georg.gartner@tuwien.ac.at)

1. Introduction

For many National Mapping Agencies (NMAs) the need for automated generalisation is from significant interest (Burghardt et al., 2014, chapter 11). Through automated generalisation NMAs are able to improve their map production lines and so save valuable resources such as time and money.

Due to the fact that generalisation upon large- to midscale map production lines is known as one of the most time consuming of all generalisation tasks for NMAs (Stoter et al., 2010), today only an automatic workflow is seen as being acceptable.

The Federal Office of Topography swisstopo, the NMA of Switzerland, already uses automatic generalisation within its map production, however is constantly seeking new approaches and methods to further increase its efficiency. An exceptional generalisation challenge found within this organisation is that of the individual house representation, a characteristic for which the Swiss national maps are famous for and which is followed, with a typical Swiss precision, up to a scale of 1:100`000.

This paper supplies a possible working solution for the automatic generalisation of the individual polygon house features of the Swiss TLM (Topographic Landscape Model, scale 1:10`000) whilst retaining the individual representation for an end scale of 1:50`000 and this whilst maintaining the settlement structures.

ArcGIS for Desktop out-of-the box functionalities were used to conduct the research. The focus of this paper lies on defining the appropriate tools for building generalisation and to automate the workflow within ModelBuilder. After the workflow has been established and the data automatically generalised, the results were presented to and evaluated by various expert groups.

2. The test process for the practical implementation

The research was conducted with the tools available in the version 10.2 of ArcGIS for Desktop Advanced. As Stoter et al. (2014) says "this may not seem innovative". Stoter et al. (2010) also highlighted that there are major problems when applying existing generalisation tools in commercial software. Firstly, the tools are often difficult to parameterise and secondly, it is also often difficult to put them in the correct order. Based on Stoter et al. (2014) this research will address these two main problems. The proposed automated workflow applies the generalisation operators in the correct order with the correct parameter values. Unlike Stoter et al. (2014) this research focuses only on the generalisation of buildings whilst maintaining the existing settlement structure for the scale of 1:50`000. Furthermore, no aggregation of the urban areas is performed.

The development of the workflow was done systematically and consists of the following steps:

- Analysis of swisstopo's predefined cartographic requirements.
- Identification of necessary generalisation operators which might be of importance when generalising buildings.
- Performing model generalisation to reduce the amount of data to be visualised.
- Solving cartographic conflicts between symbolised features (e.g. Streets and Buildings) and so performing a graphic generalisation.
- Improving the generalisation process by reviewing each step and enriching the source data wherever necessary (Stoter et al., 2014).
- Verifying the workflow at different stages and adapting where necessary.
- After the processes have been optimised the evolved workflow is then joined together. The complete generalisation workflow is finally implemented within the ModelBuilder tool.

In order to develop a correct workflow it is very important to verify the results after each applied operator and compare them with the cartographic requirements. This allows the process to be improved step-by-step until an acceptable solution is found. This step is of significant importance throughout the practical implementation in order to develop a suitable workflow. Mackaness highlighted this necessity already in 1995.

"we start with some hazy thumbnail sketch of what we want, we then source the data, apply some set of generalisation operators, view the result and repeat and refine subsequence application of generalisation operators in a cycle until a satisfactory solution is found" (Mackaness, 1995)

3. Cartographic requirements

In order to be able to develop and implement the required building generalisation workflow, map specifications were defined as a set of cartographic constraints. The cartographic constraints were defined by the cartographic experts at swisstopo and were mainly based upon (Spiess et al., 2002).

The cartographic constraints relate to various generalisation considerations when dealing with buildings such as the selection, form and graphic generalisation as well as the existing buildings and settlement structure.

4. Appropriate Generalisation operators

In this research the proposed classification of operators by Foerster et al. (2007) according to Gruenreich`s model is used. This model, which distinguishes between model and cartographic generalisation, has been found to be the most suitable for the data and maps within the NMAs (Foerster et al., 2007).

Consequently all the tools which were deemed appropriate when considering building generalisation within the ArcGIS for Desktop Advanced version 10.2 were considered and categorised according to the operator classification defined by Foerster et al. (2007).

4.1. Tools in ArcGIS for model generalisation

According to Foerster et al. (2010) NMAs consider the operator’s *amalgamation* and *simplification* to be the most important when working on the scale of 1:50`000 within model generalisation. These are followed by *class selection, reclassification* and *collapse.* The operator *combine* was found however to have no significant role within building generalisation. Table 4.1 indicates the ArcGIS tools which correspond to the proposed classification of Foerster et al. (2007).

Operators by Foerster et al.:	**Corresponding tools within ArcGIS:**
Amalgamation	Aggregate Polygons Delineate Built-Up Areas
Simplification	Simplify Building
Class Selection	Select Layer By Attribute Select Layer By Location Select
Reclassification	Field calculator
Collapse	No significant role
Combine	No significant role

Table 4.1: Operators by Foerster et al. (2007) and their corresponding tools for model generalisation within ArcGIS for Desktop 10.2

Aggregate Polygons: combines polygons within a specified distance of each other into new polygons. A minimum hole size may be defined in order to retain inner courtyards for example. When considering buildings the orthogonally function is of particular advantage to help specify the characteristic for the aggregated boundary. Barrier features may also be defined to help protect features from being aggregated across streets or other line features (Esri, 2015a).

Delineate Built-Up Areas: creates polygons to represent built-up areas by delineating densely clustered arrangements of buildings on small-scale maps. This tool is useful to identify dense settlement arrangements. Buildings are clustered based upon a grouping distance (Esri, 2015a).

Simplify Building: simplifies the boundary or footprint of building polygons whilst maintaining their essential shape and size. Small details such as indentations are deleted by setting a simplification tolerance (Esri, 2015a).

Select Layer By Attribute: adds, updates or removes a selection on a layer or table view based on an attribute query.

Select Layer By Location: selects features in a layer based on a spatial relationship to features in another layer. Each feature in the input feature layer is evaluated against the features in the selecting features layer or feature class and if the specified relationship is met, the input feature is selected.

Select: extracts features from an input feature class or input feature layer, typically using a select or Structured Query Language (SQL) expression and stores them in an output feature class.

Field Calculator: performs simple and advanced calculations on all or only selected records. With allows for example a reclassification of the building hierarchy.

4.2. Tools in ArcGIS for cartographic generalisation

Foerster et al. (2010) states that the importance of cartographic generalisation operators is significantly higher at larger scales (< 1:50`000) than at smaller scales (> 1:50`000). This is because model generalisation is more important at smaller scales and therefore the number of features partaking in cartographic generalisation is higher. For the scale of 1:50`000 NMAs consider *displacement* as the most important operator followed by *enhancement*, *enlargement*, *typification* and *amalgamation*. Table 4.2 indicates which tool within ArcGIS correspond to the proposed classification by Foerster et al. (2007). At present there is a single tool for building generalisation which combines all five operators.

Operators by Foerster et al.	Corresponding tool within ArcGIS:
Displacement Typification Enhancement Amalgamation Elimination	Resolve Building Conflict

Table 4.2: Operators by Foerster et al. (2007) and their corresponding tool for cartographic generalisation within ArcGIS

Resolve Building Conflicts: assesses graphic conflicts of symbolised features under consideration of a given reference scale. Firstly, the buildings are enlarged to a specified minimum size. Next symbol conflicts within buildings and with respect to linear barrier features are then resolved by moving or hiding buildings. This ensures that the buildings do not graphically overlap or violate the minimum spacing requirements (Esri, 2015b). The cartographic operators Displacement, Typification, Enhancement, Amalgamation and Elimination are all handled by this algorithm. The operator improves the display of the buildings by adjusting the position, orientation, size and visibility whilst maintaining the representative pattern and distribution of buildings. This algorithm, which is explained in Punt and Watkins (2010) in detail, resolves symbol conflicts applying an optimization technique. The optimization approach means that each task is made up of constraints, reflexes, and actions. A constraint is for example that a building cannot be closer than a certain distance to another, a reflex might be that a building cannot be moved onto a road and an action that the building has to move away or move back. An underlying optimiser kernel seeks to improve the fulfilment of constraints by applying various actions.

5. Automated workflow for building generalisation

The input data is harvested from the building features existing in the swisstopo TLM (Topographic Landscape Model) for a scale of 1:10`000. The corresponding road network has already been generalised for the scale of 1:50‘000. The workflow consists of both model- and cartographic generalisation. The final output is the DCM50 (Digital Cartographic Model 1:50’000). Figure 5.1 illustrates this workflow.

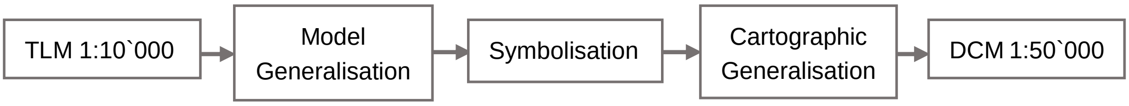

Figure 5.1: The automated workflow for the generalisation process of the 1:50`000 buildings

After extensive testing of the individual tools for the model- and cartographic generalisation, these tools were connected together with the view of creating the desired automatic workflow. This process also required extensive testing to help define the optimal order in which the individual tools should be executed.

For connecting the tools together ArcGIS provides a user friendly environment in form of the ModelBuilder. This application can deal with both simple or very complex connection matrixes.

The following sections 5.1 and 5.2 list the individual steps for the models created within ModelBuilder. Section 5.1 covers the Model generalisation and section 5.2 the cartographic generalisation.

5.1. Model generalisation

Step 1: The first step is the aggregation of all buildings with the *Aggregate Polygon* tool. This is especially important because of the way in which the buildings have been captured. In TLM the building features are captured by the individual roofs and not by their outlines. During the generalisation process it is very important that the footprint of a building is used by the operator and not the roof polygons. The intention is that overlapping polygons are aggregated together when within a distance of 1 meter, this being set as the aggregation distance. Because buildings are mainly orthogonal shaped the orthogonal optional setting available is used. Should non orthogonal buildings exist these must be excluded from this process. In order to only aggregate buildings which are not separated by other feature classes, such as roads, these are set as so-called barrier features. To illustrate the result, figure 5.2 shows the original data on the left-hand side and the aggregated buildings on the right side.

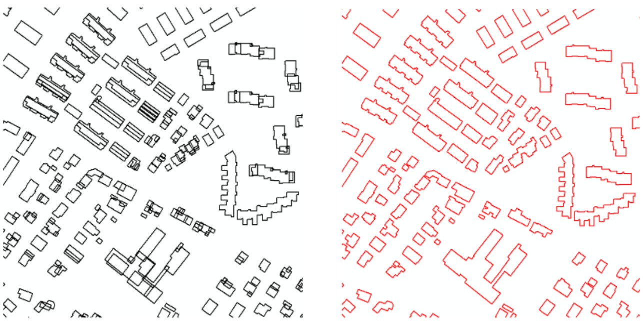

Figure 5.2: Aggregate Polygons

Step 2: The *Aggregate Polygons* tool is used a second time to perform an elimination of the inner courtyards below a minimum dimension. The same settings are used as in Step 1 complemented by setting a minimum hole size of 400 m^2 in order to eliminate the courtyards. Figure 5.3 illustrates the result.

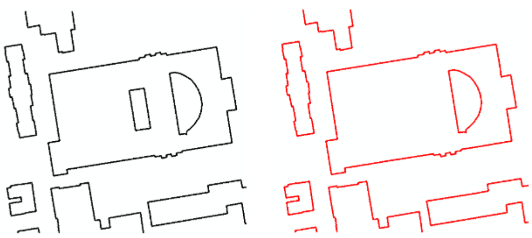

Figure 5.3: Aggregate Polygons to eliminate courtyards

Step 3: Due to the fact that all attributes are lost during aggregation there is the requirement to reattach these attributes using a *Spatial Join.*

Step 4: To apply an initial general simplification of the resulting buildings the *Simplify Building* tool is applied with a simplification tolerance of 4 meters, this value was found to produce the best results for removing details such as indentations whilst maintaining the essential shape and size relation. The value was arrived at after having tested the effect of various values and comparing there results. The decision to apply this tool after Aggregation was as a result of the extensive testing done to define the optimal execution order. It was found that processing the Simplification tool before Aggregation led to many more errors such as the overlapping of features which in turn led to the wrong buildings being aggregated. Figure 5.4 shows the results of the Simplify Building tool (on the right-hand side).

Figure 5.4: The result of the Simplify Building tool is represented on the right

Step 5: A pre-processing step is used to *add a hierarchy field* to the attribute table thus allowing different hierarchies for different building sizes to be calculated. The idea behind this is that the new hierarchy value can then be used to simplify the buildings differently.

Step 6: This step consists of *Selection* and *Classification* of the hierarchy attribute. The buildings are firstly selected according to their building sizes, this selection is based on a building size classification as defined by swisstopo. Based on this classification buildings smaller than 250 m^2 are given a hierarchy value of 3, buildings ranging from 250 – 756 m^2 a value of 2, buildings ranging from 756 – 1000 m^2 a value of 1 and buildings larger than 1000 m^2 a hierarchy value of 0.

Step 7: In this step the buildings are simplified with the *Simplify Building* tool according to the selection process defined in step 6. A different simplification tolerance is set for each of the four hierarchy values. This decision was made based on the fact that small buildings should be squared off whereas larger buildings should retain their particular footprint, hence the simplification tolerance is reduced according to the size of the building. Again after extensive testing a simplification tolerance of 12 meters for buildings with hierarchy 3, 8 meters for hierarchy 2, 7 meters for hierarchy 1 and 6 meters for hierarchy 0 was decided upon. Figure 5.5 shows the results of both methods, the simple simplification from Step 4 (left-hand side) and the simplification according to the hierarchy (right-hand side). The different colours indicate the building sizes according to hierarchies 0 to 3: blue, green, orange, yellow.

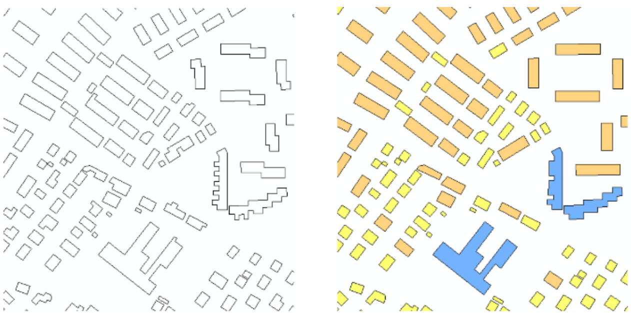

Figure 5.5: Simplify Building according to hierarchy

Step 8: In this step the smallest buildings are *selected* and *eliminated*. These small buildings where found to be mainly private garages next to the corresponding house. In order to maintain a better representation of the terraced house structure it was decided to delete these features before conducting the cartographic generalisation. Figure 5.6 highlights these small buildings and shows the result after deletion.

Figure 5.6: Selection and Elimination of small buildings

Step 9: A further requirement to fulfil was to eliminate small buildings in dense settlement areas. Therefore, features smaller than 60 m^2 are *selected* and *eliminated*. In order to achieve this there is first the need to identify what the dense areas are. As there is no clear guideline as to what defines a dense area this is done using the operator *Delineate Built-Up Areas*. With this tool it is necessary to define both the grouping distance (50m) as well as the minimum building count (4). Based on the created built-up area it is possible to select the features by location. This results in all buildings within the built-up-area being selected and then eliminated. In Figure 5.7 the grey area indicates an area of dense settlement where the buildings under the defined minimum size, here represented in black, will be deleted.

Figure 5.7: Selection and Elimination of small buildings within dense settlement areas

Step 10: Large buildings are of major importance and will notably require more space in order to be preserved whilst conducting the graphic generalisation. Therefore small buildings within a specific distance of a large building are *selected* and *eliminated*. Figure 5.8 illustrates the small buildings (represented in black) which have been selected for elimination due to their proximity to a large building.

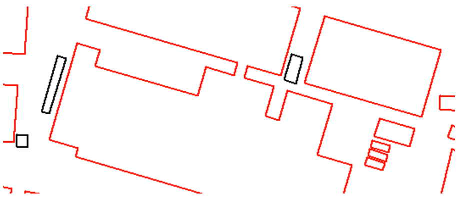

Figure 5.8: Selection/Elimination of small around large buildings

Step 11: The *Resolve Building Conflict* tool for the graphic generalisation uses the hierarchy of buildings. After running several tests and receiving cartographic feedback it was clear that the *classification* of a buildings hierarchy which run along a street needed to be modified as too many buildings were being deleted and the structure of the settlement was being lost. A pre-processing step was necessary. The solution proved to be to generate a general buffer around the generalised street

Figure 5.9: The right image shows the result of the Reclassification of buildings along a street

data which covered even the widest road. Step one was then to select all buildings with a hierarchy of two within the buffer and *reclassify* these to have hierarchy of one. In step two, all buildings within the buffer and having a hierarchy of three were selected and *reclassified* to have a hierarchy of two. Figure 5.9 shows the differences between the classifications along the streets (The colours indicate the building sizes according to the reclassified hierarchies 0 to 3: blue, green, orange and yellow).

5.2. Cartographic generalisation

The graphic generalisation process for buildings consists of a single tool within ArcGIS – the *Resolve Building Conflict* tool. In order to achieve the best results, which were not only acceptable to the cartographers involved in the evaluation but also to fulfil the predefined requirements of swisstopo a number of pre-processing steps proved necessary. The pre-processing steps necessary are step 12, step 13 and step 14.

Step 12: The first pre-processing step is the addition of two extra attribute fields, this is necessary in order to run the *Resolve Building Conflict* tool. The fields added are an invisibility and the resolve building conflict size field. These fields will be populated with values when the tool is executed.

Step 13: All the building features are *symbolised* according to swisstopo's symbol catalog and were defined as cartographic representations. Cartographic representations are a property of a ArcMap feature class that are stored in system tables inside the geodatabase and in the feature class itself. This method of symbolization was a predefinition from swisstopo.

Step 14: One of the possibilities of the *Resolve Building Conflict* tool is that of being able to define so-called conflict barrier layers. This allows for a set gap to be defined for any buildings which orient themselves along these barriers. For swisstopo it is a requirement that the house edge is overlapped by the road network signature with 3 meters. To accomplish this the original streets symbol width is reduced by the value of minus 0.06 millimetre and used as the barrier layer. This was conducted because the tool automatically snaps the buildings to the defined barrier features.

Figure 5.10: Example result using the Resolve Road Conflicts tool

Step 15: The *Resolve Building Conflict* tool separates buildings from each other and from any defined barriers whilst retaining the relative density and pattern. The tools contains the parameters Minimum Allowable Building Size and Building Gap. Swisstopo predefined the minimal dimensions for a single house as being 400 m^2 and also stated that a minimum distance of 11 meters had to be preserved between buildings. By defining the parameter minimum allowed building size as 20 meters, which is 400 m^2, the size can be enforced. It is also possible to adjust a features visibility as well as the spacing between buildings. The gap size is defined as 11 meters. Another possibility is that of managing the distance and orientation from and to the barrier features. The orient value is Boolean, specifying whether buildings should be oriented to the barrier layer. After extensive testing it was decided not to orient the buildings to the road, this was because not only the buildings along the road were affected but also those buildings within the residential areas, resulting in the existing pattern being lost. As an orientation property exists within the initial TLM data it was decided to try using this and the results were very pleasing, preserving any exiting patterns and therefore it was decided not to make use of the orientation. A hierarchy value can be optionally assigned which was done in this research. Figure 5.10 shows before (left-hand side) and after running this operator.

6. Results and Conclusion

This paper supplies a possible working solution for the automatic generalisation of the individual polygon house features of the Swiss TLM for the scale of 1:50`000. It has been proven that the generalisation tools in ArcGIS are very suitable for the generalisation of buildings with this test data as well as surprisingly fulfilling the rather special requirement of swisstopo for retaining the individual house representation and this whilst maintaining the existing settlement structure. Indeed swisstopo confirmed that this workflow achieved an acceptance level of approximately 80 percent.

The most important goal *to retain the settlement structure* was handled very satisfactory. For instance the black white ratio and the overall look and feel of the density pattern, dense areas still appearing dense, while the sparse areas remain sparse, have in most cases been handled very well indeed. Further important points to consider are the preservation of the relative building sizes and the orientation of buildings, either when considered building to building or building to road. A conclusion can also be drawn that these parameters have also been very successfully handled by this workflow. The preservation of both regular and irregular building forms can be said to have been successful. A further successful result can be seen in that the settlements extents have been generalized to a level corresponding to that of the final map scale.

Another important goal to achieve during building generalisation is *to generalize the shape correctly* such as to preserve the special character of the settlement. Building patterns and settlement forms have been maintained with varying degrees of success. The terraced house pattern and linear structures have been very successfully maintained. Also the settlement structures village, scattered settlements, urban and industrial quarters as well as individual houses where resolved in a successful way. When one considers the generalisation of the individual characteristic buildings one can see that both the main shape and the form of large building complexes as well as individual houses have been very well maintained.

Finally the *graphic generalisation* plays a key role in the generalisation process and helps guarantee the legibility of the map. Here the minimum dimensions of the buildings could be maintained in all except the densest of areas such as that of an old historic town center. The minimum distance constraint was also handled well.

Figure 6.2 shows various sections of the 1:50`000 map, displaying the predefined generalised road network as well as the automatically generalised buildings as created using the workflow described in the previous chapter. By comparison, the original TLM building is depicted in figure 6.1.

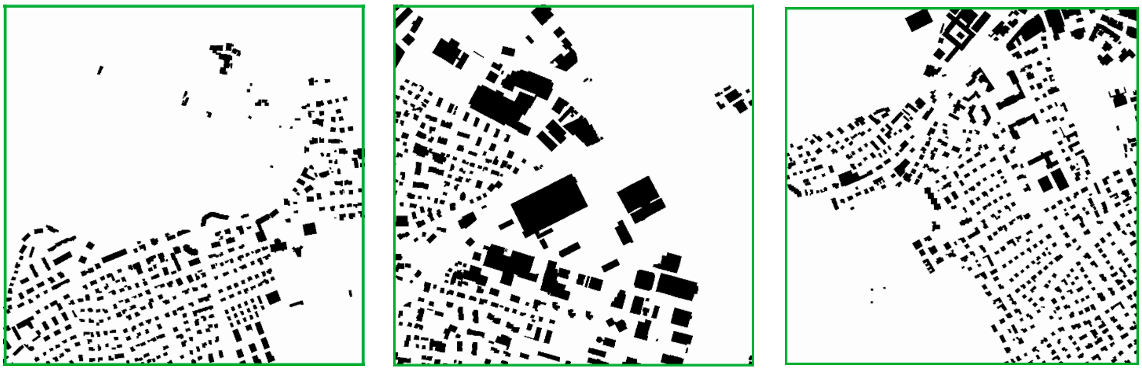

Figure 6.1: Extract TLM Data (1:10`000)

Figure 6.2: Result of the fully automatic obtained buildings (1:50`000)

Despite the fact that the overall results were very well received and offer a promising basis for further development a number of problem areas have also been identified where further research will be necessary to improve upon these results.

Regarding the *settlement structure* there are a view areas where the density was considered as problematic, especially in the area in and around the historic old town. One possible workaround would be to execute a second differentiation of a dense settlement area and to eliminate buildings within this area based on the adapted parameters.

For *generalising the shape* there is also room for further improvement. The biggest issue in this respect was that the character of the historic old town is completely lost and hand editing would be unavoidable. A possible reason for this is that the buildings within the old town centre have a highly complex form, some being even circular in shape. As there is no way to adapt the Simplify Building tool and the Resolve Building Conflict tool it might be best to exclude these highly complex area from the complete generalisation process.

A further issue is that the tools are developed to deal with the average square house shape and runs into difficulties when dealing with other shapes such as circular, resulting that these are often generalized into a square form, once more this could be a case for exclusion.

Large Building boundaries could be further simplified. Small extrusions or recesses have also not been removed or filled to further generalise the buildings. A further issue is that sometimes the small buildings have not been generalised enough resulting in a restless and poorly interpretable picture. All of these issues can be easily resolved by adjustment of parameters within the Simplify building tool until the desired results have been reached.

Regarding the *graphic generalisation* neither the minimum dimensions nor the minimum distances were maintained in the historic old town. This might be resolved by using a different approach when generalising the shape, this will however need more extensive testing.

There are also very important aspects which have not been tackled at all in this research but would be definitely worth for further exploration in the future. In this study the buildings were considered only as one single feature type because the aggregation would have become too complex. Another important issue which was not touched on at all is that of preserving feature links. As this dataset originates from a TLM it might be considered adventurous to preserve the feature links back to the mother database. The study of how these links can be preserved would probably fill a complete thesis by itself.

To summarize it can be stated that some problem areas, as stated above, can be definitely refined by further investigation and in the adjustment of parameters which is thankfully easily possible within the ModelBuilder environment. However there are also tasks where creativity will be needed to find a possible workaround, such as those of the historic old town.

References

Burghardt D, Duchêne C, Mackaness W (2014). Abstracting Geographic Information in a Data Rich World. Lecture Notes in Geoinformation and Cartography. Springer International Publishing, Cham.

Esri (2015a). Resources: Generalization toolset. http://resources.arcgis.com/en/help/main/10.2/index.html#//00700000000s000000. Accessed 25 March 2015

Esri (2015b). Resources: Graphic Conflicts toolset. URL http://resources.arcgis.com/en/help/main/10.2/index.html#//00700000000s000000. Accessed 25 March 2015

Foerster T, Stoter J, Kobben B (2007). Towards a formal classification of generalization operators. In Proceedings of the 23rd International Cartographic Conference, Moscow.

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

Mackaness W (1995). A constraint based approach to human computer interaction in automated cartography. In Proceedings of the 17th International Cartographic Conference, Barcelona, pages 1423–1432.

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

Regnauld N, McMaster R.B (2007). A Synoptic View of Generalisation Operators. In Generalisation of Geographic Information: Cartographic Modelling and Applications. Chapter 3, 37 – 66. Elsevier, Amsterdam

Spiess E, Baumgartner U, Arn S, Vez C (2002). Topographic Maps – Map Graphics and Generalisation. Swiss Society of Cartography

Stoter J, Baella B, Blok C, Burghardt D, Duchêne C, Pla M, Regnauld N, and Touya G. (2010) State-of-the-art of automated generalisation in commercial software. Technical report, EuroSDR, Amsterdam.

Stoter J, Post M, Altena V. van, 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.