PROPAGATING UPDATES BETWEEN LINKED DATASETS OF DIFFERENT SCALES

Jan-Henrik Haunert, Monika Sester

Institute of Cartography and Geoinformatics
University of Hannover
Appelstraße 9a, 30167 Hannover, Germany
jan.haunert@ikg.uni-hannover.de
monika.sester@ikg.uni-hannover.de

ABSTRACT

This work shows an incremental approach for the propagation of updates from a topographic source dataset to a generalised dataset. It aims at the result that is equal to the complete generalisation of the source dataset. For this the information from links that express correspondences of features, the knowledge of the generalisation rules and the topology of the features are considered. The problem is discussed extensively for the case of the aggregation of areas, which is based on rules that take thematic and geometrical criteria into account. Results for the generalisation of an existing dataset are shown. For the incremental method it is considered, that the update of a single feature can have different impacts on the target dataset when following the rules. An algorithm is described which is based on the derivation of the influence area on which the update is restricted.

1   INTRODUCTION

The comprehensive consistent administration of geographic data of different scales has been under scientific investigation for several years. Often links between datasets have been established, that express correspondences of features. A database which keeps layers with datasets of different scales and links between corresponding features is called Multiple Representation Database (MRDB). Starting from the dataset with the smallest scale, such a database can be built up by a successive generation of the layer with the next larger scale by automatic generalisation. The links are turned out as a by-product from the generalisation. Their storage in the database documents the generalisation and supports complex analyses.

The problem, which is addressed in this paper, concerns the update of an MRDB. Once a change has happened to the real world and is added to the source layer, the consistency between the layers is lost. Theoretically, it can be restored by a complete new generalisation of the dependent layers. Indeed, this solution of the problem is achieved with brute force and is not applicable in practice, since it requires an enormous amount of processions and is very time consuming. An update method that excludes features from the procession that are not influenced by an update will be much more efficient. This approach is often referred to as incremental generalisation. A major task for this is to find the features which are influenced by an update. For this task the analysis of links can be beneficial.

In this paper the problem will be discussed for an area partitioning and its generalisation. For this problem the influence of an update can be spatially restricted to a region which contains the updated features and their surrounding area. An exact definition will be introduced in this paper.

This work is motivated by the need for updating methods for an existing MRDB, which has been designed to keep the four digital topographic landscape models which are provided by the mapping agencies of Germany and its federal states. These models exist for the scales 1:25.000 (ATKIS Basis DLM), 1:50.000 (ATKIS DLM 50), 1:250.000 (ATKIS DLM 250) and 1:1.000.000 (ATKIS DLM 1000). The methods for the collections of these datasets differ among each other and in different federal states. A derivation of a dataset from the dataset with the lower scale by generalisation is only carried out in some federal states and for certain scales. Normally, the datasets are collected independently and show big discrepancies. However, the application of automatic generalisation methods is generally aimed for the future, to reduce the existing inconsistencies.

At the moment for the development of an updating method it is important to differ between two cases of an MRDB. The first case is a direct dependency of a generalised dataset from a source dataset. The generalisation rules which are applied to the complete generalisation are known. The second case is that two independently collected datasets with different scales have been linked and are kept in the database. Strategies for update propagation methods might be

fundamentally different for these different cases. For the incremental update method which is presented in this paper the first case is assumed.

After an overview on related work and a brief explanation of the applied aggregation method this incremental approach will be described in detail. The paper ends with a brief conclusion of the presented work and an outlook on future research.

2 RELATED WORK

Since the paper deals for the most part with an incremental method for the update of an aggregated dataset with categorical data, both the related work on incremental generalisation and the related work on the generalisation of an area partitioning will be reviewed in this section.

2.1 INCREMENTAL GENERALISATION

The concept of incremental generalisation was introduced by Kilpeläinen & Sarjakoski (1995). The basic idea is that updates which are inserted to a source dataset can be propagated to a generalised target dataset without repeating the generalisation of the entire map. Instead of this only those objects of the source dataset are considered that are possibly influenced by the change. To exclude other objects from the process it is utilized, that the generalisation process is divisible into modules, if it is composed of actions that have to be executed in a defined order. This applies for example if water areas, road networks and buildings are generalised consecutively. In this case an update to a building will not result in a change of a road in the target dataset. Hence, only buildings need to be generalised again.

Here the concept of modules comprises only groups of objects from different categories. With this restriction it would be necessary to generalise all buildings on the map again, even though the influence of an update is also spatially limited. An incremental generalisation method will be more efficient, if the generalisation task is also separated into spatially distinct modules. In the example these could be defined by grouping buildings within a mesh of the road network.

Galanda (2003) defines different types of object groups for an agent based generalisation method as group agents, that collectively aim at certain stages. In his work "category agents" correspond to the modules introduced by Kilpeläinen & Sarjakoski. The concept of the spatial partitioning by roads is equivalent to a "geographic partition agent". These analogies show, that the developed concepts by Galanda have a high relevance to incremental generalisation, since his approach is consequently build up on modularization which is also needed for an incremental method.

Harrie & Hellström (1999) implemented a system for the propagation of updates which is based on rules that are defined for special update actions like the insertion of a new road to the master dataset. After the generalisation of the objects, remaining spatial conflicts are solved. The method is a promising approach, for the case that the generalisation rules, which were used for the original complete generalisation, are not at hand. For the case that the rules for the complete generalisation are known, it is more favourable to develop an incremental generalisation method which uses the same rules than to define new rules for updating actions. Otherwise it will be very difficult to obtain consistency.

A more generic approach to the problem of automated incremental generalisation was presented by Skogan & Skagestein (2005). Here the rules of the complete generalisation are applied. The method is based on productions that are stored additionally to the datasets in a production log. A production is composed of an identifier of the generalisation rule, that is used to create a target object, the identifiers of objects in the source dataset that are examined by this rule and the identifier of the object in the target dataset that is created. By this the history of the complete generalisation is extensively documented. Influences of updates can be found by analyzing the productions in which changed objects are involved.

The method applies to systems in which simple rules define direct relationships between the features of both datasets. Often, the complete generalisation is achieved by a more complex chain of actions that are triggered by successively evaluated rules. Also, a rule can be applied iteratively many times until a certain criterion is reached. For these cases further developments are needed. The second case can be found in the example of an area aggregation, which will be discussed in this paper.

A disadvantage of the method by Skogan & Skagestein is, that a high amount of additional data has to be maintained, because every feature that is analyzed by a rule is recorded in the production. This could be reduced by using the information which is implicitly given with the rules. Often the rule which is used for the generalisation of a feature evaluates its direct neighbourhood. Here the information about the dependencies can be derived from the topology and does not need to be stored explicitly.

The ideal for an incremental method as it is described by Kilpeläinen & Sarjakoski and Skogan & Skagestein is to generate the same result as a complete generalisation with a lower effort by excluding objects that are not influenced by an update. A method that satisfies this criterion does not produce results which are dependent on the order of the updates that are added and it is always possible to restore the target dataset by a complete generalisation.

2.2   GENERALISATION OF AN AREA PARTITIONING

The generalisation of an area partitioning is a classical map generalisation problem. It aims at a partition with less areas in which the most important characteristics are preserved and certain constraints are fulfilled. An approach to this problem would be to search sub-graphs in the adjacency graph that globally represent the best approximation of the finer tessellation under the given restrictions. Indulska & Orlowska (2002) deal with the computational issue of this approach. It can be concluded that a global optimization is not functional, since it requires an enormous computation time. In many cases it will result in problems that are NP-complete.

Because of this, methods have been developed and realized, which lead only to local optimizations. This is normally done by iteratively aggregating adjacent features. The algorithm which is applied in this work for the complete generalisation of the partitioning is founded on the algorithm which has been used by van Oosterom (1995) to construct a Generalised Area Partitioning-tree (GAP-tree).

The tree represents a hierarchy of the partitioning and can be constructed as follows:

For each area feature the importance can be measured. The feature a with the lowest importance $I(a)$ is selected and merged with the neighbour b, for which the collapse function *Collapse(a,b)* is maximal. This is repeated until the minimal importance of a feature reaches a defined threshold.

In the work of van Oosterom, the importance of a feature is defined as the product of its area and a weight factor for its type. The relevant properties of the features for the Collapse function are seen in the length of their common boundary, the compatibility of their types and the weight factor for the type of b.

Bobzien (2001) compares two algorithms with a different definition of the order in which the features are selected. In the first algorithm always the feature with the smallest importance (here equal to its area) is merged. In the second algorithm a feature is merged with adjacent features until the minimal area is reached before the algorithm continues to select the smallest feature. Also it is practiced to merge the features in a defined order of classes as in van Smaalen (2003). Measures which can be utilized for the collapse function are also discussed by van Smaalen (2003) and by Peter (2001). Sester et al. (1998) show an approach for an automatic learning of aggregation rules from given examples. The automatic aggregation is applied in practice as described by Podrenek (2002). Here the Collapse function is defined by a table with ranks of feature type changes.

With an extension of the GAP-tree, which is based on a skeleton of a polygon, it is possible to resolve an area into fragments and perform merges of the resulting areas to different neighbours of the original polygon (Ai & van Oosterom, 2002). A suitable instrument for this task is the straight skeleton of a polygon, which can be applied for many generalisation cases (Haunert & Sester, 2004). It is especially suitable for the geometry type change from a two-dimensional to a linear representation.

3   RESULTS FROM A COMPLETE GENERALISATION OF AN AREA PARTITIONING

For the incremental aggregation method which is presented in this paper it is assumed that the complete generalisation is defined by an importance function and a collapse function as explained in section 2.2. It is assumed that the order in which the features are selected to become merged is defined by the importance function. Modifications of the method for other sequences are possible but will not be discussed in this paper.

To produce the assumed initial situation an existing dataset has been generalised completely. Figure 2 shows areas from the ATKIS DLM 50, which constitute the source dataset. Figure 3 shows the results in the target scale 1:250.000, after the aggregation algorithm was applied. The applied collapse function was adapted for the most part from a table with ranks of feature type changes, that is used for the derivation of the ATKIS DLM 50 from the ATKIS Basis DLM (Podrenek, 2002). The importance of a feature is defined by its area. The stop criterion for the algorithm was taken from the specifications of the ATKIS DLM 250 (AdV, 2005). Figure 1 summarizes the statistics of both datasets. The number of features and the number of categories is reduced. In some cases small elongated structures survived due to junctions to larger areas. Probably, a better result could have been achieved by considering the length of the common boundary as it is proposed by van Oosterom (1995) or shape describing parameters. Generally, the accuracy of the area boundaries is too high for the target scale, since no line simplification was applied. However, the result suffices as basis for the development of an incremental method.

	Number of features	Number of categories
ATKIS DLM 50	2137	35
generalised dataset	162	9

Figure 1: Statistics of source and target dataset

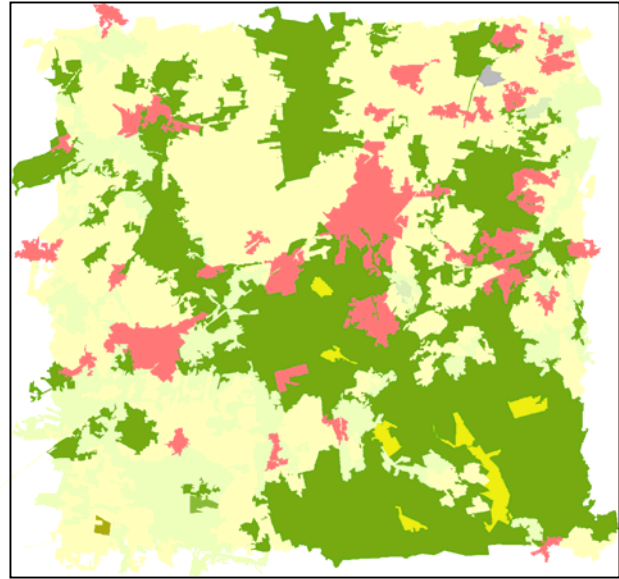Figure 2: Source dataset, areas from
ATKIS DLM 50

Figure 3: Target dataset, areas aggregated
according to specifications of ATKIS DLM 250

4 INCREMENTAL UPDATE OF A GENERALISED AREA PARTITIONING

The solution for the complete generalisation of an area partitioning relies on a very simple algorithm. However, it is very time consuming to repeat the complete generalisation every time an update has been inserted. Often, this is due to complex interlaced rules which have to be evaluated each time the collapse function is called.

Because of this, it is valuable to restrict the processing to the area that is actually influenced by the update. Another motivation is that a spatial database is often edited by more than one operator at the same time. In this case a transaction management is needed, that prevents the generation of inconsistencies. For this it is valuable to know about possible influences of updates, so that the locking of the database can be limited to restrict other users only as much as necessary.

The updates of a geographic dataset can normally be classified into insert, remove and change actions (Anders & Bobrich, 2004). In the case of an area partitioning, updates that produce gaps or overlapping areas are not allowed. This means that inserting or deleting a single feature or changing the geometry of a single feature is not feasible. With the exception of an attribute change of an area, the consistent update of the partitioning always involves more than one feature of the source dataset directly. In the following the distinction of different update actions will not be addressed, because all updates can be treated the same way. A feature will simply be called updated if any of the three actions has been applied on it.

Generally, it is aimed to produce the same results as the complete generalisation. In this case the results are reproducible and consistent. In the next section examples for dependencies between features of both datasets are discussed which can be found by applying the algorithm for the complete generalisation and comparing the resulting trees before and after the update. These examples help to understand which cases need to be considered in an incremental method.

4.1 INFLUENCES OF UPDATES WITHIN THE TREE STRUCTURE

Figure 4 shows the tree which represents the iterative aggregation of areas before and after the attribute of a feature was changed. For the simplification of the illustration the features are arranged in a one-dimensional array. However, the concepts in this section apply also for a two-dimensional area partitioning. The importance of the features is represented by the sizes of their areas. The collapse function is defined by a comparison of the integers which represent attributes of the features. The smaller the difference of these values is, the higher the value of the collapse function is. The actions of area merges are symbolized by circles and performed in the order a, b, c. After constructing the tree which is shown in figure 4a), the feature which is marked dark in figure 4b) is updated. In the given example this change can be propagated upwards through the tree and results in a change of an attribute of the corresponding feature in the target dataset. The path on which the update is triggered is marked with a bold line. The generation of the adjacent feature of the target dataset with the attribute value 4 is not influenced.

Generally, it can not be assumed, that the structure of the tree persists an update. Also it is possible, that the braches which emerge from adjacent features become influenced by an update, as the example given in figure 5 shows. Here it is not sufficient to propagate the changes upwards through the branch that contains the updated features.

The influence might spread horizontally to an adjacent branch in two cases. The first case is that, due to a change of a feature that is selected during the generalisation process, it is not assigned to the same adjacent feature as before. This happened in the example given in figure 5. The second case concerns the other way round: During the construction of the tree a selected feature is not assigned to the same neighbour as before, because a change has happened to one of its neighbours.

In both cases it is possible that the influence continues to spread to branches of adjacent features. At the worst, this might result in a chain reaction, which overthrows the whole tessellation of the target dataset.

However, a great deal of the generalisation can be saved if the original tree structure is completely given. A branch represents a sequence of area merges. In this sequence, the position of the merge which caused the horizontal spread of the influence to this branch is determined. Only those area merges that follow on this merge possibly differ from the original sequence. In other words, it is not necessary to confirm the actions in a branch that were performed earlier than the action which caused the horizontal spread of the influence to this branch. Thus, the branch needs to be reconstructed only in the upper part that follows on this event.
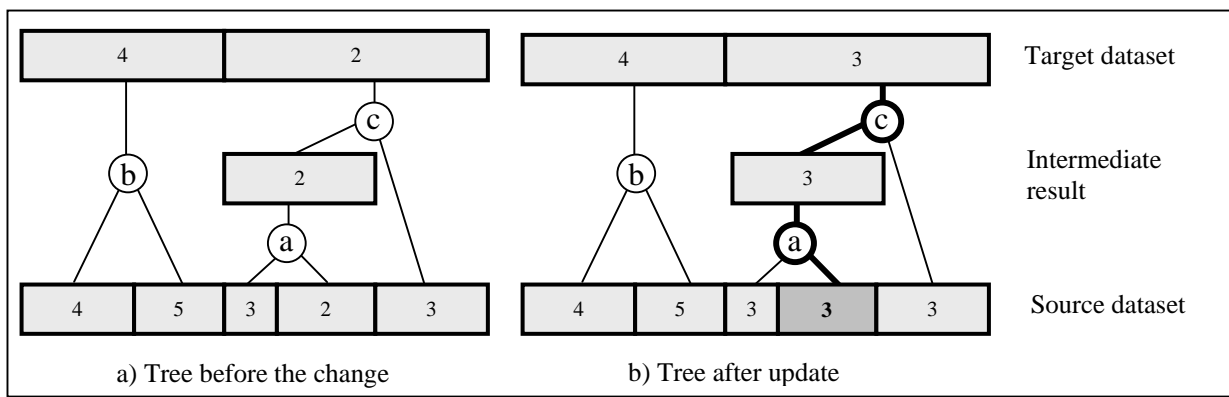

Figure 4: Propagation of an update through an area partitioning tree without a structural change of the tree.
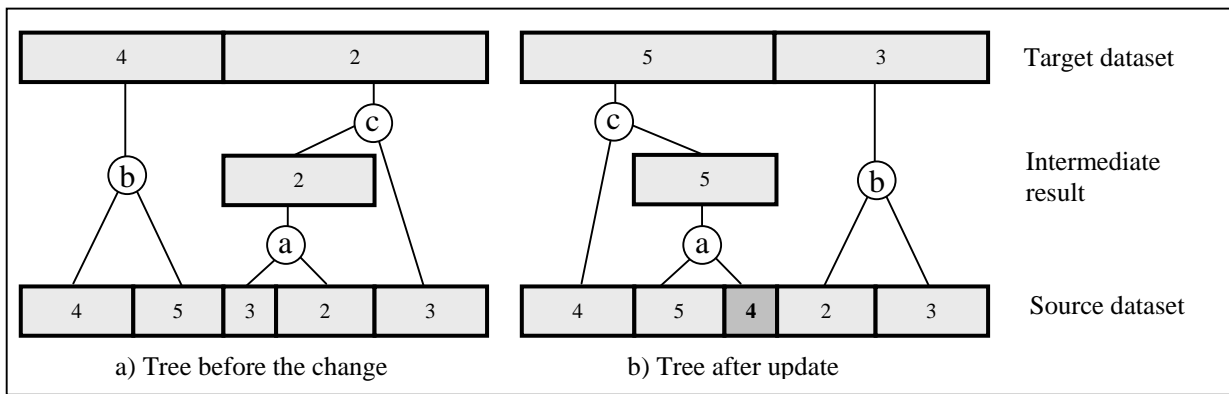

Figure 5: Generalisation after an update that results in a structural change of the tree.

4.2   SPATIAL SPREAD OF INFLUENCED REGION

As described in section 4, updates involve normally more than one feature of the source dataset at the same time. In practice, editing regions of the source dataset is more common for an operator than inserting or modifying a small group of features. Typical examples are a development area of a town or a new motorway which divides many areas. Thus, it is more natural to observe the spread of an update's influence spatially in terms of regions than to describe the propagation within the tree. This is also necessary, if the tree which documents the sequence of area merges is not at hand. As described in section 1 the underlying database to this work does only permit direct links between the two datasets. A tree structure as it is shown in section 4.1 is not provided.

The algorithm which will be shown in this section is designed for the case of direct links between features. Other differences to the concepts which were discussed before do not exist, but the problem will be observed from a more geographic perspective here in terms of regions. The following definitions help to develop an algorithm:

a) The region of features in the target dataset, which covers all updated features in the source dataset is defined as "interior region".

b) The region which is covered by all features in the target data that are adjacent to the interior region is defined as "boundary region".

The links from the MRDB can be used to find the interior region efficiently. It simply results from the features that are linked to updated features. For the area partitioning, this applies also if objects are inserted and have not been linked yet. In this case the links from the removed or deformed objects, that occupied the same area before the update happened, point to the corresponding features in the target dataset.

The following incremental approach guarantees the same result as a complete generalisation:

Apply the algorithm for the aggregation on all features in the source dataset, that are within the interior region or the boundary region. Use the collapse and importance function, which have been used for the complete generalisation.

As long as no change happens to the boundary region, this restriction of the investigation area is valid. Such a change can be caused by the two cases that were discussed in section 4.1. In both cases the change will be noticeable by a merge of a feature from the interior region with a feature from the boundary region. As mentioned, this can cause further changes to adjacent objects. Hence, the area under examination has to be extended.

If the separation of the interior region and the boundary region is violated by an area merge:
- *stop the algorithm .*
- *extend the interior region on the features in the target dataset, which cover the area which was returned by this merge.*
- *extend the boundary region on the adjacent features.*
- *perform the area merges within the region which was added to the boundary region, that are defined to happen earlier than the merge which violated the separation of the interior region and boundary region.*
- *continue with the aggregation of features, that are within the interior region or the boundary region.*
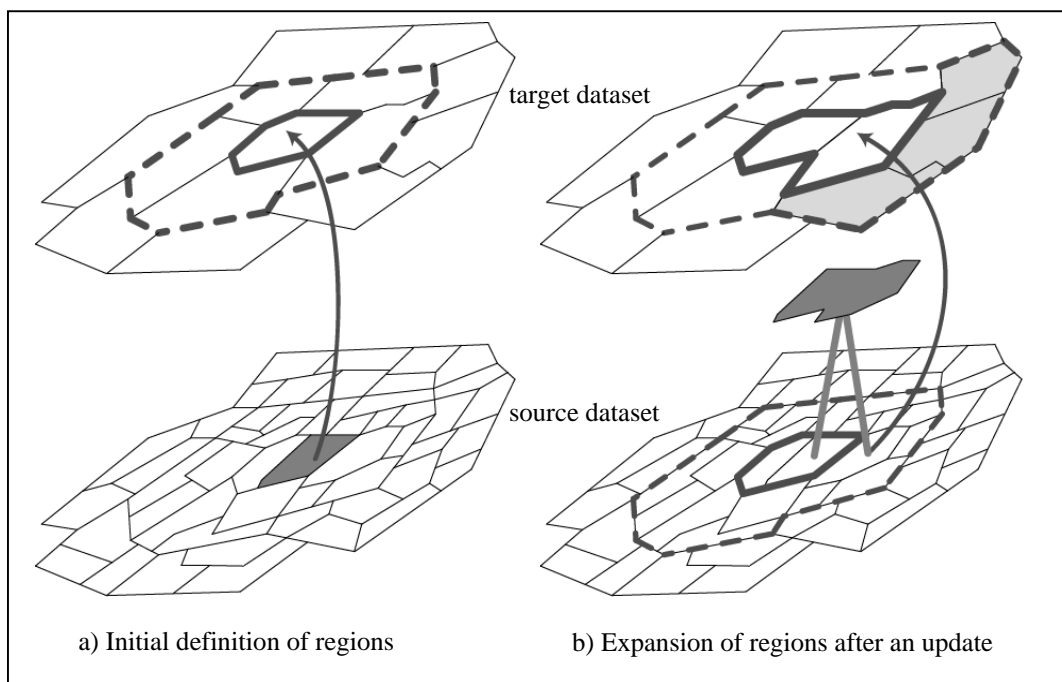

Figure 6: Two steps of the algorithm.

Figure 6 shows two steps of the algorithm. To improve the clearness of the illustration here a two dimensional area partitioning is shown, which does not imply a conceptual difference to the figures in section 4.1. In Figure 6a) the dark

grey feature is updated in the source dataset. Initially, the area of the feature in the target dataset which is connected with a link (arrow) is defined as interior region (bold line). The boundary region is defined by its neighbours (dashed line). The algorithm begins with the aggregation of all features from the source dataset that are within this region. In figure 6b) an area merge happens which connects the interior and the boundary region. This triggers an expansion of the regions. Before the algorithm continues to aggregate all features within the expanded regions, those merges of the added features (light grey) need to be performed, that are defined to happen earlier than the area merge which triggered the expansion.

With this algorithm it is achieved to restrict the generalisation to those features, which are possibly influenced by the update. For this the interior region with all updated features is defined. Also, the generalisation of a feature is performed again, if an updated feature was evaluated for its original generalisation. These features can be found in the defined boundary region, because the presented algorithm relies on the evaluation of adjacent features. An expansion of these regions needs to be considered, because an influenced feature probably has influence on other features. Although an infinite expansion of the regions is theoretically possible, the worst case is a complete new generalisation of the dataset. The links of the MRDB can be used for an efficient access to the possibly influenced features.

5   CONCLUSION AND FUTURE WORK

In an increasing degree topographic databases are designed to contain datasets of different scales. For these databases methods of update propagations are needed, which enable consistency between different datasets. The presented method satisfies this demand for the case of a generalised area partitioning. The algorithm performs an incremental update after the source dataset is modified.

For the design of the method it was assumed, that the aggregation is carried out by the algorithm which was described by van Oosterom (1995). This algorithm appears in different variations and is applied in practise by national mapping agencies for the generalisation of topographic datasets. Hence, the method can be easily applied in practice.

An interesting problem that needs further research is the adaptation of the presented approach to other generalisation problems. An important question is, weather the developed incremental method can be applied or extended on other generalisation cases than the aggregation of areas. In a prior work methods for geometry type changes and collapses of areas have been implemented (Haunert & Sester, 2004). For these problems the straight skeleton of a polygon was applied. The next step will be to extend the method for the incremental generalisation on these developed generalisation functions. The assumption is, that this will be a minor problem, because the major characteristics of the generalisation methodology will not change: Only the neighbours of a selected feature are analysed and the generalisation actions happen in a well defined order.

An interesting question which will be investigated in the future is, how the described concepts help to develop a system for a more holistic incremental generalisation of many different object types. The idea is to formulate the generalisation as a directed graph of modules which have inputs from the source data set or from other modules and deliver certain results, which finally define the target dataset. The tree shown in figure 4 and 5 can be interpreted as an instance of such a graph. Generally, it will be necessary to define the order in which the modules are executed which was done by an importance function in the given example.

The major task for the development of such a system is the modularization of the generalisation process. In some cases this will not be possible, because the complete generalisation is sometimes defined by a global optimization. However the successes by Galanda (2003) show, that an encapsulation of object groups and the definition of group hierarchies can be achieved to a high degree and is suited for generalisation tasks.

The development of an incremental generalisation system does not only support a more efficient updating of a generalised target dataset. With the availability of a system, that is capable of finding the influence region of an update and propagating the update to the target dataset, it will be also possible to observe the results of typical updating actions. Most likely, results of empirical tests will help to develop methods for the propagation of updates for the more difficult case that the target dataset has not been generated with known rules from the source dataset. As mentioned in the introduction, also for these cases update propagation methods are searched.

Also, it will be possible to analyse the sensitivity of the target dataset to changes in the source dataset. For the applied aggregation algorithm, it can happen that a small feature is the decisive weight for the generalisation and its variation causes a tremendous change. Detecting these features is important for a quality assessment of the generalisation. These aspects will be addressed with following research.

6   ACKNOWLEDGEMENTS

bundle-project entitled Abstraction of Geographic Information within the Multi-Scale Acquisition, Administration, Analysis and Visualisation.

7   REFERENCES

AdV, 2005: ATKIS – Objektartenkatalog. www.atkis.de, accessed April 2005.

Ai, T. & van Oosterom, P., 2002: GAP-Tree Extensions Based on Skeletons. – Proc. of ISPRS Symposium on Geospatial Theory, Processing and Applications, Ottawa.

Anders, K.-H. & Bobrich, J., 2004: MRDB Approach for Automatic Incremental Update. ICA Workshop on Generalisation and Multiple Representation, Leicester.

Bobzien, M., 2001: Flächenzusammenfassungen in der Modellgeneralisierung. Mitteilungen des Bundesamtes für Kartographie und Geodäsie, Band 20, Frankfurt/Main, S. 19-29.

Galanda, M., 2003: Automated Polygon Generalization in a Multi Agent System, PhD Thesis, Department of Geography, University of Zurich, Switzerland.

Harrie, L. & Hellström, A.-K., 1999: A Prototype System for Propagating Updates between Cartographic Data Sets. – The Cartographic Journal, 36(2), pp. 133-140.

Haunert, J.-H. & Sester, M., 2004: Using the Straight Skeleton for Generalisation in a Multiple Representation Environment. ICA Workshop on Generalisation and Multiple Representation, Leicester.

Indulska, M. & Orlowska, M. E., 2002: On Aggregation Issues in Spatial Data Management. – Proceedings of Thirteenth Australasian Database Conference (ADC2002), Melbourne, Australia.

Kilpeläinen, T. & Sarjakoski, T., 1995: Incremental generalization for multiple representations of geographic objects. In: Muller, J. C., Lagrange, J. P. & Weibel, R. (editors) GIS and Generalization: Methodology and Practise. Taylor & Francis, London, pp. 209-218.

van Oosterom, P., 1995: The GAP-tree, an approach to 'on the fly' map generalization of an area partitioning. In: Muller, J. C., Lagrange, J. P. & Weibel, R. (editors) GIS and Generalization: Methodology and Practise. Taylor & Francis, London, pp. 120-132.

Peter, B., 2001: Measures for the Generalization of Polygonal Maps with Categorical Data. Fourth ICA Workshop on Progress in Automated Map Generalization, Beijing, 2-4 August 2001.

Podrenek, M., 2002: Aufbau des DLM50 aus dem Basis-DLM und Ableitung der DTK50 – Lösungsansatz in Niedersachsen. In: KS, Band 6, Kartographie als Baustein moderner Kommunikation, S.126-130, Bonn.

Skogan, D. & Skagestein, G., 2005: An Automated Incremental Generalization Method. In: Skogan, D. (author) Multiple-Resolution Geographic Data and Consistency, PhD Thesis, Department of Informatics, University of Oslo, Norway.

van Smaalen, J.W.N., 2003: Automated Aggregation of Geographic Objects. A new Approach to the Conceptual Generalisation of Geographic Databases. Doctoral Dissertation, Wageningen University, The Netherlands.

Sester, M., Anders, K.-H. & Walter, V., 1998: Linking Objects of Different Spatial Data Sets by Integration and Aggregation. – GeoInformatica 2(4), pp. 335-358.

BIOGRAPHY OF THE AUTHORS

Jan-Henrik Haunert was born in 1978 and studied Surveying at the University of Hannover. He graduated in 2003 and obtained the Master's degree (Dipl.-Ing.). Since January 2004 he is a scientific assistant of the institute of cartography and geoinformatics at the University of Hannover. The focus of his research lies on different representations of geographic features at different scales. In a project funded by the German Research Foundation he develops methods for the automatic updating process of a Multiple Representation Database.

Monika Sester, born in 1961, studied Geodesy at the Technical University of Karlsruhe and obtained the Master's degree (Dipl.-Ing.) in 1986. Until October 2000 she was a staff member of the Institute for Photogrammetry of the University of Stuttgart, Germany, where she obtained her PhD in 1995. Since November 2000 she is Professor and head of the Institute of Cartography and Geoinformatics at the University of Hanover. Her primary research interests lie in multi-scale approaches in GIS and image analysis, data generalisation and aggregation, data interpretation, and integration of data of heterogeneous origin and type.