AUTOMATIC GENERALIZATION OF LAND COVER DATA

Olli Jaakkola
Finnish Geodetic Institute
Geodeetinrinne 2
FIN-02430 Masala, Finland

Abstract

The study is related to the production of a European (previously CORINE) Land Cover map of Finland.
It presents an implementation of a new methodological concept for map production, supervised classification
and automatic generalization. The theory behind automatic generalization and some of the possible
methods for describing the quality of generalized land cover data are given. In the case study, the existing
supervised classification is supported with digital maps and attribute databases. All input data is
combined to a detailed land cover with a very small minimum feature size, and automatically generalized
to the standard European Land Cover. According to the quality tests performed, the generalization
method used here meets the present quality specifications. The method is fast, and it gives an opportunity
for multiple data products in various scales, optimized for different purposes, and for updating the land
cover database.

1 Introduction

The study includes a theoretical part describing the theory of automatic generalization and a proposal for
testing the quality of generalized land cover data. The case study is related to the conversion of an
existing land use database of Finland to a standard European Land Cover database. At present, the decided
common European methodology consists of the visual photo-interpretation of satellite images for the
production of land cover features (see [1, 2]). The present method is found to be unsuitable for
continuous change detection. In the Finnish version the conversion includes multiple input data in
different spatial data forms, combination of different input data into one coverage, and a automatic
generalization of the coverage to the European Land Cover. The paper shows that it is possible to
automatically generalize a detailed land cover to the European Land Cover database and fulfill the
requirements specified for the database.

There is a more detailed report on the feasibility study for the Finnish CORINE land cover, with greater
consideration on the theory of generalization and quality of generalized land cover, as well as a detailed
description of the contents of the case study [3]. Some of the main results are presented here.

2 The automatic generalization

2.1 Why, when and how to generalize

Automatic generalization can be defined as the process of deriving, from a detailed (large scale)
geographic data source, a less detailed (small scale) data set through the application of spatial and
attribute operations. In attribute operations the spatial contiguity is not taken into account, whereas in
spatial operations the value of the class at one point depends on spatial associations with the area
surrounding it.

Objectives of a generalization process are: to reduce the amount, type, and cartographic portrayal of the
mapped data consistent with the chosen purpose and intended audience, and to maintain clarity of
presentation at the target scale. Clearly, the objectives state that the selection of generalization operations

and parameters is dependent on the intended purpose of the generalized data.

The conditions for generalization have to be specified. The spatial measures specify the land cover features itself, such as its size, shape, distance, and the holistic measures specify the distribution or density of the land cover features. The spatial measures can be given as parameters to the generalization operators, e.g. the minimum feature size, the minimum feature width, the minimum distance between the features, the minimum inlet/outlet width etc. The parameters used specify the content of the generalized data.

The forms of spatial and attribute operations possibly needed for the generalization of the land cover may vary for different data sets. Presented here are the ones needed for the Finnish land cover. In the raster domain these terms may be somewhat fuzzy, since the terms are originally defined for vector-based transformations [4]. The approach is similar to that described as object-based raster generalization [5, pp. 109].

In the generalization, we have used one attribute operation. Reclassification is the regrouping of features into classes sharing identical or similar attribution. Four spatial operations have been used. Aggregation is the lassoing of a group of individual point features in close proximity and representing this group as one continuous area. Amalgamation is the joining of contiguous features together, either by merging the feature to the semantically closest one, or by dividing it between the neighbouring features. Amalgamation is sometimes called aggregation of areal features. Smoothing is the relocating or shifting of a boundary line to plane away small perturbations and capturing only the most significant trends. Simplification is the reducing of a boundary line complexity by removing changes smaller than a certain threshold.

Also, the execution order of the generalization operations is important. Firstly, we have to execute the operations, where the data types (point, line, area) are changed, i.e. aggregation. Secondly, we have to execute the operations, where the attribute definitions of the features are changed, i.e. reclassification. Thirdly, we have to execute the operations, where the spatial definitions of the features are changed, i.e. amalgamation (size), or smoothing and simplification (shape). The order of execution has a major effect on the generalized output data.

2.2 The quality of generalized data

Most of the spatial data quality measures are scale dependent. Actually, none of the present quality measures are suitable, as such, to describe the quality of the generalized data. The generalization reduces the complexity of the data structure and adds error to the database. The quality always deteriorates in favour of simplicity and legibility. Most of the quality assessment procedures do not take these errors into account, and therefore do not understand that the error rate in the generalized database includes both the degree of generalization and the real error, the bias in summary measures and unintended positional and attribute errors produced by generalization. The different generalization operations have different effects on the quality of the result [3, pp. 16-17]. Also, the generalization operations are dependent on one another. Thus, we need to test the quality of different generalization operations combined with each other.

For testing the case study generalization, we have modified the normal discrete multivariate analysis, i.e. the error matrix, to account for generalized land cover data. From the error matrix we have derived quality measures for the feature based attribute accuracy. Also, from the error matrix we have derived measures [3, pp. 20-21], for analysing the quality of land cover class shares after different generalization methods, namely the residual differences between the areas of the generalized and the original land cover. Visual inspection could be used for detecting the major positional (and attribute) errors.

3 Case study: the Finnish Land Cover

The purpose of the Finnish Land Cover -project is to use existing land cover classification with auxiliary data and automatic generalization to produce a standard European Land Cover database. The purpose of the European Land Cover is twofold: to provide quantitative data on land cover for statistics, and to provide maps of different scales for European environmental policy. The land cover database in the nominal scale of 1:100000 is considered accurate enough, but not too large in volume. The spatial measures, when to generalize, are given with a minimum mapping feature size of 25 hectares and a minimum feature width of 100 metres. The quality specifications in the original CORINE procedure include a positional accuracy limit of 75 metres, and a feature based attribute accuracy of 85 percent for a total classification [2]. There is no specification for the areal shares of the classes, since the original method does not include a separate generalization task.

3.1 The procedure

As input data we proposed to use existing supervised satellite data classification (forest and semi-natural areas), map masks (fields and wetlands) , statistical records with position (buildings), and statistical records with digitization (see Figure 1).

Firstly, we aggregated point features, namely the individual buildings to the areal feature, i.e. built-up area. Thus we expanded the ground areas of certain buildings in artificial surfaces with 1 pixel to also cover the surrounding concrete areas. Secondly, we combined the SLAM, and the building register to an ungeneralized land cover database (Figure 1). At the same time with combination we reclassified 64 SLAM and some other classes to CORINE classes (Figure 2 a). Thirdly, we formed the heterogeneous classes (CORINE 242, 243) with combined aggregation, reclassification, and statistical comparison (method see [6, pp. 10-11], it is not in report [3]). Fourthly, we amalgamated small areal features to larger ones. We performed separate amalgamations in the different hierarchical levels of the CORINE classes. With the hierarchical grouping we actually defined the priorities of amalgamation, first grouping and amalgamations for classes in the CORINE class level 1, then for land classes in level 1, and finally for all the classes. The amalgamations were done in several iterative steps, first to a minimum feature size of 1 hectare, then of 5 hectares, and finally of 25 hectares minimum feature size (see Figure 1 and Figure 2 b, c, and d). Fifthly, after this we smoothed the border lines, and thus reduced the narrow inlets and outlets in all features (Figure 2 e). After raster-vector conversion we simplified arcs with too many points by giving a weed-tolerance of 25 metres, and obtained the final standard European Land Cover database for Finland.

3.2 The quality of products

A general note is that the quality assessment results are from a single area of 1200 km^2 (40 * 30 km). For the quality of the CORINE land cover concepts applied to the Finnish environment see [3, appendix A3]. The generalization level for the European Land Cover is the area of generalized features divided by the whole area, and is 22 percent. It could be interpreted that generalized map features have on average about 22 percent of other classes included.

The positional accuracy was verified by overlaying the original raster map with vector borderlines of the produced European Land Cover. The borderlines of the main classes in level 1, e.g. water bodies, forests, and agricultural areas, are preserved quite nicely [3, appendices B5,B6]. The feature based attribute accuracy was tested with a feature based comparison to the original combined data. In the derived measures we have the real errors produced by generalization and the level of generalization together [3, appendices A5,B5,B6]. The generalized pixels are graphically presented as a difference map between the original and generalized land cover map [3, appendix B7]. The quality of areal shares of different classes
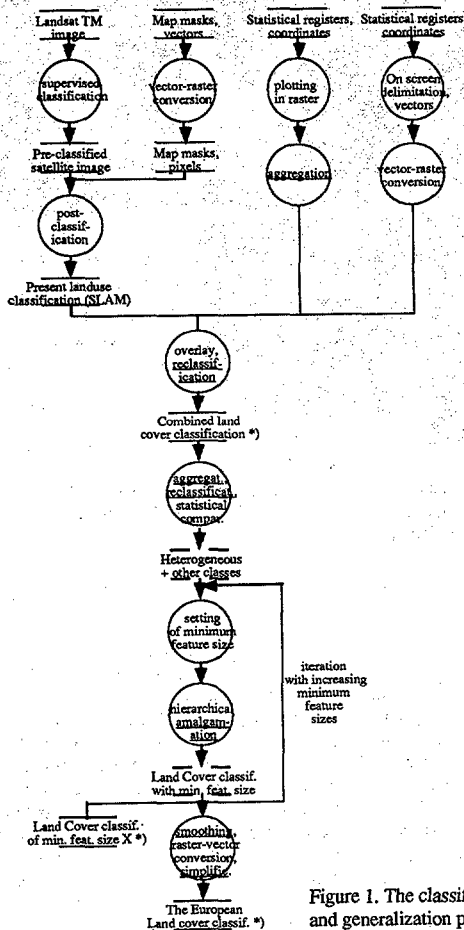
Figure 1. The classification, combination
and generalization procedure. *) see figure 2.

is important for the statistical use of the results. The areal shares changed systematically during the generalization, and the quality of shares were tested with the residual differences. Approximately, a change of 21 percent in areal shares of classes occured during the generalization, but for some classes the change is significantly larger. Note, that the residual differences include both the degree of generalization and the real systematic errors.
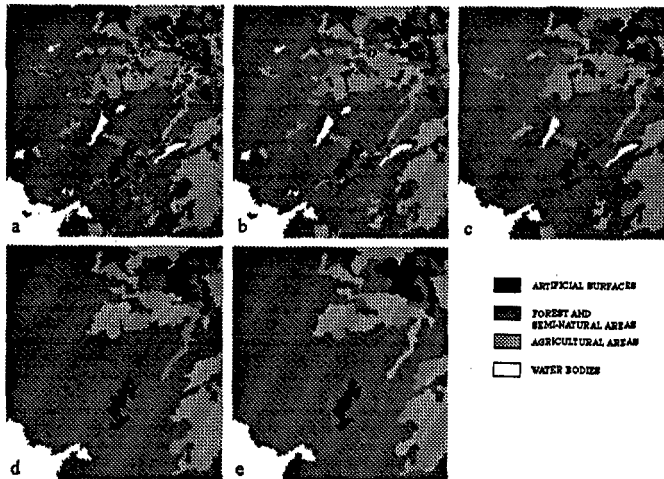
Figure 2. The result after aggregation, overlay and reclassification a), after amalgamation for minimum feature size of 1 b), 5 c), and 25 hectare d), and after smoothing the borderlines e).

4 Conclusions

The presented procedures for the automatic generalization are fully operational. The processing time for the present generalization procedure for the test area of 1200 km^2 was from 3 to 4 hours. The automatic generalization can flexibly produce multiple products in different scales. Actually, the present procedure produces four different databases: an original combination map, maps of 1, 5, and 25 hectare minimum feature size, all at the same time, as side-products of the standard European Land Cover. All maps are in digital form and can be stored on a database. The size and shape parameters can be modified for different classes, e.g. we may give a smaller minimum feature size for some important classes. Also, we can test the influence of different operations and parameters on the results, and select the most optimized ones. The procedure is areally homogeneous and objective if the supervised classification is controlled. Perhaps the most important property of supervised classification and automatic generalization is updating consistency. The proposed method gives the possibility of using exactly the same methods for two or more different classification dates from the same area. In addition, the method is open for topological constraints, i.e. the land cover features can be kept topologically correct with certain point or linear features.

Noting the quality specifications for CORINE land cover (chapter 3) and the testing results (chapter 3.2), the quality of automatic generalization is probably good enough for the European Land Cover. Further improvement in quality can be introduced with more precise objectives. We venture to say this despite not having made the comparison with the visual photo-interpretation. It is also good to bear in mind, that visual interpretation of the Finnish landscape is very difficult [7]. Thus, the advantages (+) and disadvantages (-) of the new method compared to the old method are be summarized in Table 1.

	sc + ag	vis
fast processing	+	-
direct digital product	+	-
multiproduct approach	+	-
multiple input data	+ -	+
class definition consistency	+	-
areal homogeneity	+	-
consistency in updating	+	-
revision possibility for classification	+	-
positional accuracy	+	-
attribute accuracy (of features)	+ -	+ -
attribute accuracy (of areal shares)	+ -	+ -
systematic errors controllable	+	-

Table 2. The advantages and disadvantages of the supervised classification
supported with digital data (sc) and the automatic generalization (ag)
compared to the visual photo-interpretation (vis).

Finally, we should understand that the aims of generalization can be multiple, and some procedures
produce high quality maps, some high quality statistics, and these two aims can be conflicting.

References

[1] CORINE land cover, 1992. Brochure by European Environment Agency Task-Force. 22 p.

[2] CORINE land cover - Guide technique, 1993. CECA-CEE-CEEA, Bruxelles. 144 p.

[3] Jaakkola, O., 1994. Finnish CORINE land cover - a feasibility study of automatic generalization and
data quality assessment. Reports of the Finnish Geodetic Institute, 94:4, 42 p.

[4] McMaster, R. & K. Shea, 1992. Generalization in digital cartography. Association of American
Geographers, Washington. U.S.A. 134 p.

[5] Schylberg, L., 1993. Computational methods for generalization of cartographic data in a raster
environment. Royal Institute of Technology, Department of Geodesy and Photogrammetry. Doctoral
thesis. Stockholm, Sweden. 137 p.

[6] Fuller, R.M. & N.J. Brown (1994). A CORINE map of Great Britain by automated means: a
feasibility study. Institute of Terrestrial Ecology (Natural Environment Research Council). ITE
project T02072J1. 37 p.

[7] Ahokas, E., J. Jaakkola & P. Sotkas, 1990. Interpretability of SPOT data for general mapping.
European Organization for Experimental Photogrammetric Research (OEEPE) Publ. off. No. 24. 61
p.