# IMPLEMENTATION OF COMPREHENSIVE MODELING TECHNIQUES ON *KARTOGEN* GENERALIZATION SOFTWARE

Özlem SİMAV, Serdar ASLAN, Bülent ÇETİNKAYA, O.Nuri ÇOBANKAYA
ozlem.simav@hgk.msb.gov.tr
General Command of Mapping, Harita Genel Komutanligi, Tip Fakultesi Caddesi, TR06100, Dikimevi, Ankara, Turkey.

## ABSTRACT

This paper describes the implementation of three well-known modeling techniques namely condition-action (C-A), human interaction (HI) and constraint based (CB) model on *KartoGen* (*KG*) Generalization Software designed by the Cartography Department of General Command of Mapping (GCM). Throughout the software several generalization algorithms are developed and an overall modeling is performed containing where, when and how to use and control these algorithms. Preliminary results in text generalization reveal that CB model highly prevents the human interactivity and lead the automation of the generalization processes to a promising level. This satisfactory experience in text generalization encourages us to implement CB modeling techniques into displacement of buildings and transportation generalization.

## INTRODUCTION

Automated generalization, one of the most interesting and complex problem of cartography, is an indispensible part of computer-assisted map production systems that serves to derivation of small scale and less detailed maps from larger scale ones. Depending on the continuous developments in the digital cartographic techniques and their implementation in the production environments, the generalization approaches and algorithms are encouragingly improved to produce more consistent data at target scales. However, any generalization operation can impact the data quality which may results in unpredictable consequences related with topological and semantic accuracies (Haunert and Sester, 2008). Besides generalization requires quality checked specific digital data, skilled cartographers and automation processes composed of smart software algorithms in order to retain the legibility of map features as well as their contextual characteristics.

Through the conventional mapping production flow, the cartographers in GCM were carrying out manual generalization to derive 1:50K and 1:100K maps from the main scaled 1:25K maps following the basic generalization principles and rules until the beginnings of 2000 (Aslan et al, 2004). A project was initiated in 2002 the aim of which is to automate the generalization processes in the derivation of 1:50K and 1:100K scale standard topographic maps at an optimum production time with high standardization and automation levels as much as possible. Three years later the first version of KartoGen (*KG*) software was officially released which was developed by the GCM staffs using ESRI ArcGIS environment and its ArcObjects developing components. *KG* software includes data pre-processing, automatic and semi automatic generalization toolboxes (Figure 1). The toolboxes are still under improvement. Before the execution of main generalization programs, data pre-processing including data transformation from coverage to geodatabase, attribute exchange, adding supplementary fields into feature classes required for the next processes, deleting the unnecessary features without any condition, splitting-unsplitting lines, compressing database etc. are performed automatically. In the subsequent steps, generalization process is concluded with automatic and interactive semi automatic chain of operations.

Generalization is now rendered digitally, as programming code. But the core point is how to arrange this batch code with intelligence which is the issue of the modeling task. From simple batch processing to sophisticated methods, different approaches for modeling the automated generalization processes are developed in accordance with the advances in computer science and technology. Apart from primitive batches, Harrie and Weibel (2007) distinguished the generalization process into three modeling techniques. These are C-A (also called rule based), HI and CB modeling.

The basis of C-A model is the attributes of the object (condition) and the generalization rule which trigger the algorithm (action) according to this object. For particular generalization tasks like aggregating vegetation polygons with raster generalization operators, C-A model generates encouraging outputs. However, for the overall generalization, some hindrances are revealed. For instance, formulation of the rules embedded into the algorithm is a hard work due to the variety of map objects and their relations. These relations between map objects cause different algorithms to exist.

The limitations faced with the C-A model and the impossibility of taking all situations into consideration cause a new interactive model called HI model to reveal. HI model is contact with the concept of *amplified intelligence* (Weibel, 1991), based on sharing the workload of the generalization process between the software and human. While a simple interactive process tackles only with the map object in order to edit with limited generalization operators, an interactive process supported by amplified intelligence, which has not been used so far by any generalization system (Harrie and Weibel, 2007), produces the solution of the situation and executes related operators. Though HI model is well enough for single map objects, it is time consuming and highly confined to cartographer’s ability.

Generalization process has already chained with consecutive conditions and there exist several alternative actions to solve them, as well. Without any interactive intervention, CB model resolves conditions (constraints) with any action, contrary to the C-A model. Thus, CB model let us evaluate which action is better, but the constraints must be defined coherently to eliminate mapping conflicts.

In the current version the *KG* software, the system is mainly managed by means of C-A and HI modeling techniques. We are now updating our software by taking the advantage of CB modeling. This study focuses on the practical implementation of three well-known modeling techniques on the text generalization and text placement within the *KG* software and search for an appropriate model in terms of the requirements of GCM for the production of 1: 50K and 1:100K scale maps.

## *KG* SOFTWARE

It is acknowledged that there have been still no perfect solution exist for modeling the whole generalization process fully automatic (Harrie and Weibel, 2007). In order to acquire high quality derived maps, we have been studying to develop and improve generalization processes to satisfy the requirements of GCM for the production of 1:50K and 1:100K scale maps. In the current version of *KG* software, most of the operations are automated, but only the undecided situations are handled by the cartographers manually. Selection of roads in transportation feature class may be a good example for the undecided situation. Solution performed for this problem is tested in various sheets including different types of roads and the outputs are unsatisfied and time consuming. Therefore, for the time being the selection of roads are completely done by the cartographers manually. Studies regarding the road selection with weighting and shortest way algorithms according to road types and junctions are still going on. As soon as the assessment of the results is completed, it is going to be included into the automated process.

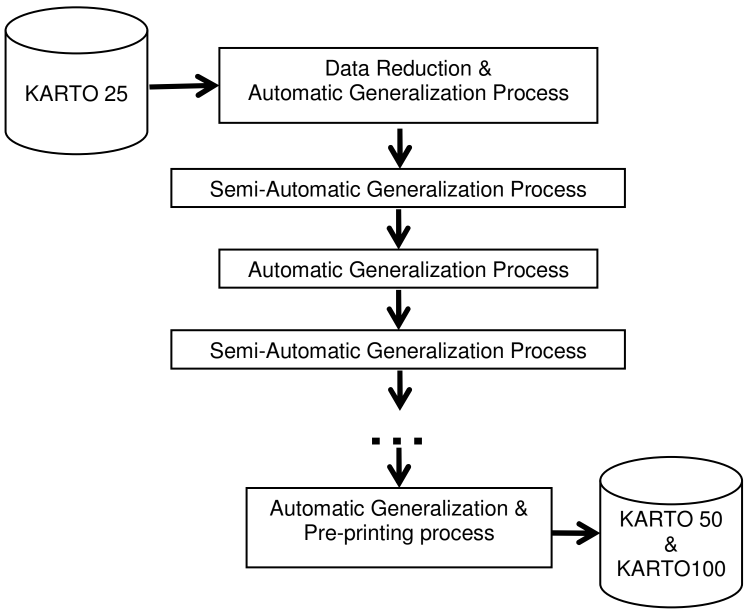



**Figure 1.** Generalization workflow of *KG* software.

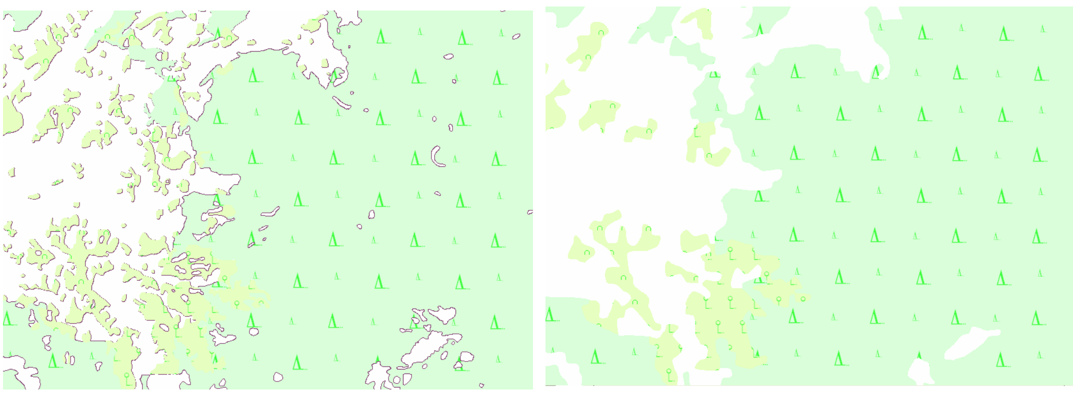
The system in the current version of the *KG* software is mainly managed by means of C-A and HI modeling techniques. The CB modeling technique has been only implemented for the text generalization and text placement so far. Table 1 summarizes the modeling techniques used in *KG* software with respect to the feature classes. In the table one can see that whilst vegetation and utility generalization are based on C-A model only (Figure 2), the text generalization uses CB model (Figure 3). The other feature classes use both C-A and HI model.

**Table1.** Modeling techniques used in *KG* software.

| | C-A Model | HI Model | CB Model |
|---|---|---|---|
| Elevation Generalization | ***A*** | ***A*** | - |
| Hydrographic Network Generalization | ***A*** | ***A*** | - |
| Transportation Network Generalization | ***A*** | ***A*** | ***R*** |
| Building Generalization | ***A*** | ***A*** | ***R*** |
| Industry Generalization | ***A*** | ***A*** | - |
| Utility Generalization | ***A*** | - | - |
| Physiography Generalization | ***A*** | ***A*** | - |
| Boundary Generalization | ***A*** | ***A*** | - |
| Vegetation Generalization | ***A*** | - | - |
| Text Generalization and Text Placement | - | - | ***A*** |

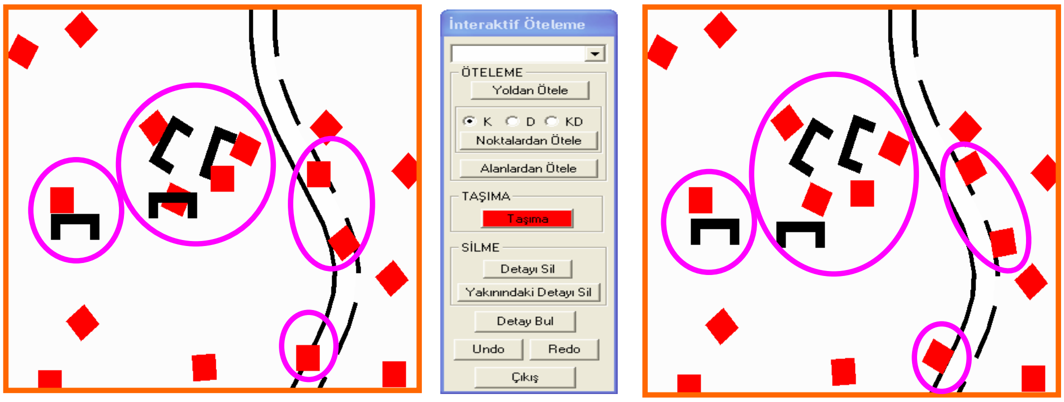
(***A*** depicts the applied model and ***R*** depicts the model in research level)

After checking and preparing of the raw data, known as *data reduction,* the *KG* software carries out the generalization in two steps. In the first step where the C-A model is adapted, the vegetation, elevation, hydrography, utility, industry, boundary, physiography and annotation feature classes are exposed to automatic generalization. Without any interactivity, this step resolves a generalization problem with only one predefined solution. The second step called *interactive generalization* that handles 9 feature classes successively. Despite the name "*interactive*", this semi automated step comprises not only the simple human interaction but also certain automated generalization operators and algorithms related to C-A or CB modeling techniques.



(a) (b)

**Figure 2.** Vegetation generalization from 1:25K (a) to 1:50K (b) using C-A model (After conversion of vector dataset into raster, certain raster neighbor operations are applied for simplifying and smoothing the vegetation polygons using statistical calculations with predefined thresholds)




(a) Before (b) After

**Figure 3.** Building displacement using both C-A and HI model (An generic interface appears for the situations that can not be decided trough the execution of automated processes.)

Regarding the C-A model, the conditions and actions embedded into the scripts of the software trigger the predefined generalization operations with the parameters. The parameters are taken from the tables of geodatabase, but the cartographers have chance to change these parameters via an interface and assess the results if necessary. In Figure 4 a generic user interface of hydrography generalization can be seen for setting up the measures of actions. These measures can be stored in the system rather than on-the-fly calculations. Some parameters of simplification, smoothing, point remove etc. are given in this interface.

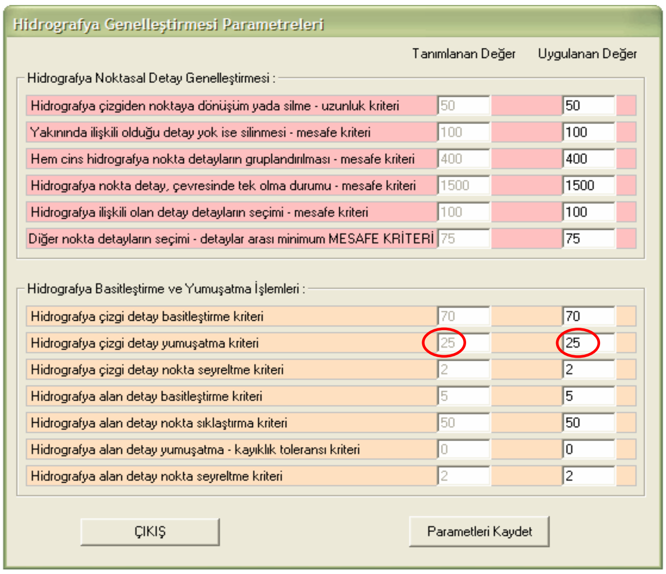



**Figure 4.** Parameters that are used in Hydrography generalization

## COMPARISION OF THREE MODELING TECHNIQUES ON THE TEXT GENERALIZATION

The C-A and HI models were adapted for the text generalization and placement processes in the former version of the *KG* software and they had been used until last year. Recently we have decided to change the C-A modeling techniques with CB model. The principle reason for this change is to benefit from the flexibility of the CB model, to save the process time by canceling the human interaction and to make these processes fully automated. The ability of flexible placement choices of the CB model according to weightings and priorities of text stacks let us control where and how the texts should be placed relative to the features. The CB model is supposed to be faster than C-A together with HI since there will be very little manual interaction.

In the older version we had some semantic problems with the text placement regarding the absence of connections between features and their corresponding text labels. In order to put the texts into a map automatically, a relation between the text and the feature need to be constructed. Since there had no concrete connection between text and the related feature, no automated operation could be carried out. A program is developed which can construct relation between the features and their corresponding annotations where the names of the settlements are retrieved from the settlement database and the other annotations from the 1:25K main database. Some features like hills and ridges can not define certain topography due to its fuzzy borders. These texts are hard to place, but the summits and ridge lines derived from slope and aspect maps let us to use ESRI Maplex extension for the placement of these texts as well.

Another drawback of the older version is about the selection of the text labels from the base database and the placement of them. Only the annotations of the settlements (city, town, village, etc.), height values of the geodetic control points, and the some important way points such as building names, graveyard names were being selected and no additional automated operation was being carried out for their placement. In other words, the selected labels were being placed without any change in their positions and their shapes as they were in the original base map.

To compare CB with C-A and HI modeling techniques we performed a case study in a 1:100K topographic map derived from 1:25K maps which have highly dense annotations. The adapted comparison criteria are the processing time, the accuracies of the text positions and the number of the missing text data according to the GCM production requirements.

Figure 5 shows a sample area in an ideal 1:100K map that satisfies the GCM production requirements which will be taken as reference for the comparison. To produce the same map we first apply C-A modeling together with HI. The output is depicted in Figure 6. It takes approximately 5 minutes to complete the whole map with C-A. The same map is also produced by CB approach with the same computer. The duration of the execution is three times more than the C-A only. However we can observe lots of mistakes in the derived map in Figure 6 regarding the missing text data and the positions. So to correct them with HI takes ten more hours. The output of the CB is almost similar with the ideal map (Figure 7) and needs almost no human interaction. These results show that, the CB is an appropriate modeling for the text generalization and placement operations.

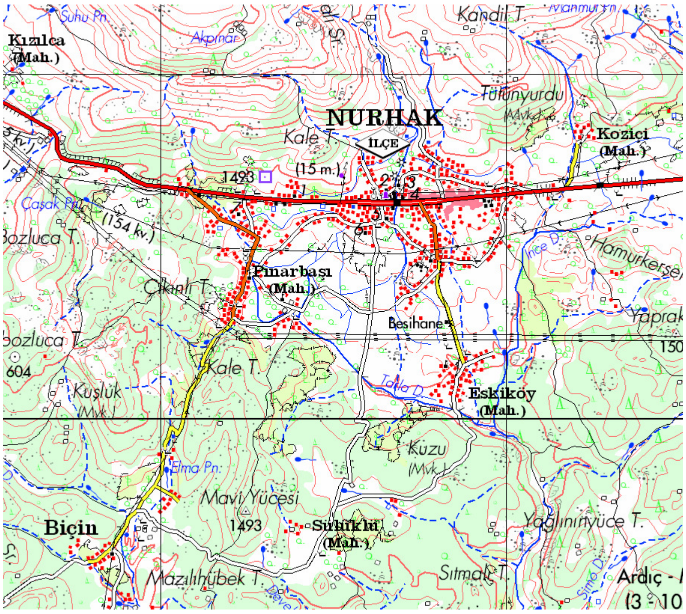



**Figure 5.** A sample area of 1:100K ideal case map

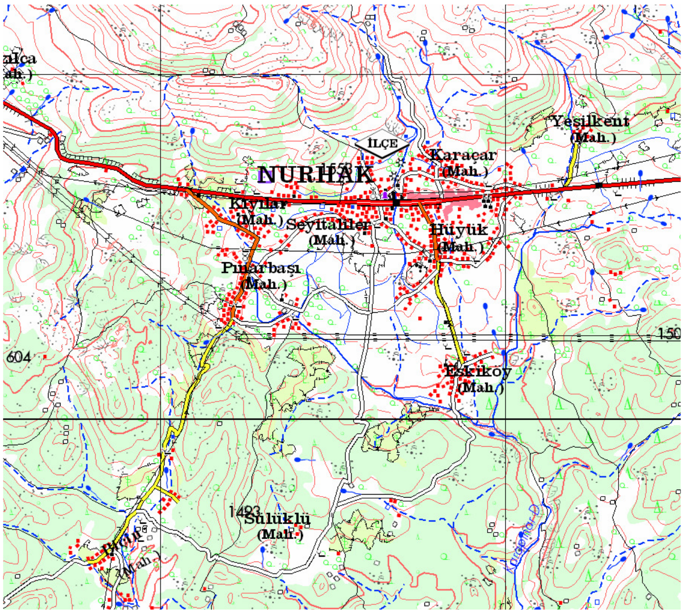



**Figure 6.** The same area in Fig.5 with C-A

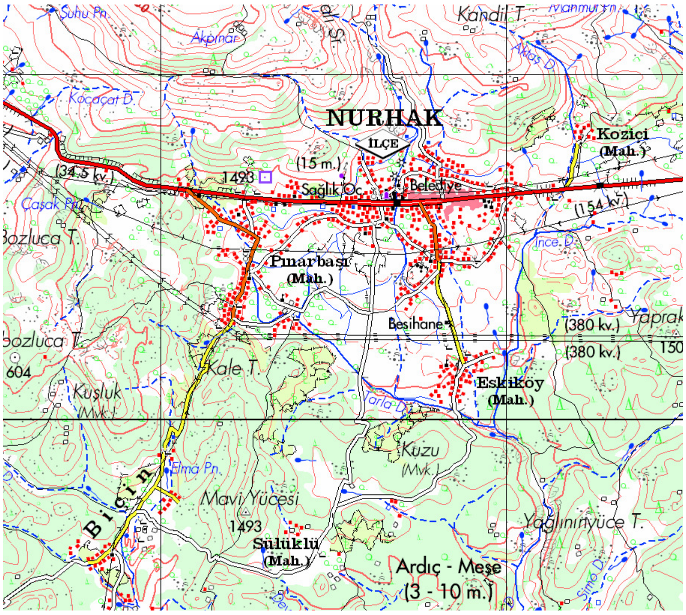



**Figure 7.** The same area in Fig.5 with CB

## CONCLUSION

Due to its complex nature, automation of generalization processes still constitutes an important issue in cartography. Different approaches including simple batch processing, C-A modeling, HI modeling and finally CB modeling techniques contribute to the automation of generalization. In this study we give a brief overview of *KG* software, an application software used in the Cartography Department of GCM for the cartographic generalization activities, and the modeling techniques integrated into the software. A case study about text generalization and placement is conducted which simply compares the three modeling techniques namely C-A, HI and CB.

Most of the cartographic generalization operations within the *KG* software are based on the C-A and HI modeling. Efforts to integrate the CB technique into the system have increased to eliminate the limitation of the condition-action approach and so far it has been only implemented for the text generalization and text placement.

The preliminary results particularly in text generalization reveal that CB model highly prevents the human interactivity and lead the automation of the generalization processes to a promising level. This experience in text generalization encourages us to implement CB modeling techniques into the displacement of buildings and transportation generalization including elimination, simplification and displacement operations and researches for its application have been recently initiated. It should be noted that whatever the approach is used, the human interactivity is still necessary.

Experiences obtained during the development and update of the *KG* software (not mentioned in detail here) show that each model be used in different parts of the system to obtain desirable and encouraging solutions. Feature classes that have not dense data such as industry, utility and boundary yield satisfactory results based on C-A model and need little interaction, whereas the densely distributed data such as text, building and transportation can be benefit from the advantage of CB model.

**REFERENCES**

Aslan, S., Çetinkaya, B., Ilgın, D.E., Yıldırım, A. 2004, Some intermediate results of KartoGen Generalization Project in HGK, ICA/EuroSDR Workshop on Generalisation and Multiple Representation, Leicester, UK, August 19-22.

Burghardt, D., Schmid, S. and Stoter, J. 2007, Investigations on cartographic constraint formalization. In: *Workshop of the ICA Commission on Generalization and Multiple Representation*, August, Moscow.

Filippovska, Y., Walter, V., Fritsch, D., 2008, Quality evaluation of generalization algorithms, ISPRS Beijing.

General Command of Mapping, 1999, Technical directions for topographic Maps, GCM Press (in Turkish)

Harrie, L., Weibel, R., 2007, Modelling the Overall Process of Generalisation; Generalisation of Geographic Information: Cartographic Modelling and Applicatons, Elsevier, p. 67-87

Haunert, J-H., Sester, M., 2008, Assuring logical consistency and semantic accuracy in map generalization , Photogrammetrie - Fernerkundung – Geoinformation (PFG) , vol. 2008 , no. 3 , p. 165-173.

Podolskaya, E.S.,, K.-H. Anders, Haunert, J-H., Sester, M., 2008, Quality Assessment for Polygon Generalization , Quality Aspects in Spatial Data Mining , CRC Press, Taylor & Francis Group , p. 211-220

Weibel, R., 1991, Amplified intelligence and rule-based systems. In Buttenfield, B. P. And McMaster, R. B., editors, *Map Generalization: Making Rules for Knowledge Representation*, pages 172–186. Longman, London.