CO-359

A PROCESS CHAIN FOR THE AUTOMATISED CONSTRUCTION OF CHOREMATIC DIAGRAMS

REIMER A., DRANSCH D.

GFZ German Research Centre for Geosciences, POTSDAM, GERMANY

BACKGROUND AND OBJECTIVES

Chorematic maps, as introduced by French geographer and cartographer R. Brunet in the 1980ies, are a class of diagrams that allow to depict complex situations in a highly synthesised and generalized manner. In essence they are a tool for the structural and iconic representation of complex geospatial situations. They consist of terms and graphics that largely abstract from actual objects and precise cartographic symbols and have enjoyed several documented successes as effective communication tool. Until now, chorematic diagrams have been created manually. In the case of automatizing the construction of chorematic diagrams, two major research questions need to be answered: How can chorematic diagrams be described formally? Which process steps are needed to derive chorematic diagrams from geodata? Our research consists of two parts: In the first part existing chorematic diagrams found in printed sources such as books, magazines and agency documents are analysed in terms of their attributes. The analysis results are used to create a formal description of chorematic diagrams (Reimer and Fohringer 2010, Reimer 2010). In the second part of the research, a process chain for the construction of chorematic diagrams from geodata is developed on the basis of the formal description derived in part one.

The results are combined into a formal description of chorematic diagrams. Theformal description of chorematic diagrams is enriched by concrete measures for specific attributes in form of generalization constraints and a chorematic database in which existing mappings of thematic content to choreme usage are stored and classified according to a matrix of geometrical and semantic scale level structure. The results of the analysis are used to develop a prototypical process chain for the automated construction of chorematic diagrams from geodata. The projected process chain for automatic construction poses geometric generalisation problems as well as semantic generalisation problems.Two sub-problems, namely outline schematisation and choropleth to area-class diagram transformation, are presented in further detail.

A first validation of the process chain is the recreation of existing diagrams or parts thereof.

APPROACH AND METHODS

When we strive for automatised construction of chorematic diagrams, we are mostly concerned with the graphical dimension. That is to say that while chorematic diagrams have been the results of a certain kind of geographic thought, they also share a certain kind of cartographic expression of these thoughts. While simple or elegant explanations and hypotheses on the structure of geographic space can come from many sources and procedures, the cartographic expression of these geographic ideas is often lacking in its execution. The reams of chorematic diagrams that were produced in the wake of Brunet's institutional and scientific success share some cartographic properties, which we strive to understand to be able allow chorematic modes of graphical expression be used wherever they might seem appropriate. For suggestions of usage see Cheylan et al. 1997, Laurini et al. 2006 and Ligozat et al. 2007.

Taking a successful cartographic paradigm in order to use the mode of expression for other subjects or domains is not unprecedented. The most famous example would be schematic maps, inspired by the famous London tube map or metro-map. Other examples include the street-map analogy that business process modellers use to describe their need for a certain kind of generalisation of base data (Van der Aalst 2009), or the numerous attempts at mapping multidimensional data-“spaces” with spatial analogies like „data surfaces“. Whether remaining in geographic space or moving to data space, these attempts need to understand what one might call the design rules of a given type of map. In fact, the great advances that automated generalisation has made, especially with topographic maps in the custody of the respective NMAs (National Mapping Agencies), stem from the fact that their respective design rules have been documented and understood so well.

Our research into the design rules of existing chorematic diagrams has resulted in a taxonomy (Reimer 2010) and first attempts at formalising constraint-based and rules-based design principles (Reimer and Fohringer 2010), and is being refined. As discussed in Reimer 2010, we deem the cartographic results and notions about chorematic design rules of other research into chorematic diagrams to be insufficient. Most other research in that area is mainly inspired by the chorematique instead of trying to recreate or model the existing diagrams.

The basic visualisation strategies of existing chorematic diagrams are mostly those of complexity reduction. Complexity is reduced by not depicting quantitative data at all and only very seldom ordered data. The visual variables used need therefore only be able to convey qualitative information (Bertin 1973) of a given geo-object. The usage of the quantitave-enabling variable „size“ is restricted to the (relative) size of the geo-object in geographic space. The overall cartographic complexity is kept low by using only very few cartographic lines per square milimeter of drawing space, i. e. they all share a cartographic line frequency of around or below 1 [Bt] (see Reimer and Fohringer 2010; Harrie and Stigmar 2007; Bertin 1973). Furthermore, the number of point objects, as well as annotations is severely reduced. In the case of textual annotations, they are most of the time omitted or abbreviated and explained in the legend to reduce visual clutter. The point symbols also are often just circles, and only very few more complex symbols are used. Hachures and patterns are simple line patterns differentiated by colour and direction, if used at all.

The main questions that pose themselves before being able to construct a chorematic diagram in a data-driven way then are:

1.Which object types?

2.How many classes of the same object type?

3.How many instances of the classes, i.e. How many objects for which class?

4.How to visualise them?

Assuming some way to answer the questions 1.-3. has been decided upon, the visualisation question comes to the forefront. The quantitative analysis of existing chorematic diagrams and the cartographic line frequency serve as upper bound for the answers to these three questions (Reimer 2010, Reimer and Fohringer 2010). As has been argued for above, the geometric shapes of the geo-objects in their peculiar “choremised” form are the key to creating chorematic diagrams; neither quantities or orders need to be depicted, but rather configurations and patterns of pre-classified geo-objects.

PROCESS CHAIN OVERVIEW

The basic approach to a process chain for automatically creating chorematic diagrams can be described as a specific interpretation of the visualisation pipeline, where source data is pre-processed, analysed, filtered, mapped to a specific visualisation and then rendered for the viewer. In our case, the main focus lies on the „mapping“ part, where expressiveness and effectiveness of the visualisation are the goal. In the case of spatial data visualisation this „mapping“ step is structurally the same as cartographic generalisation, the difference being that the output is generalised differently than is usual, i. e. chorematic schematisation takes place. As chorematic diagrams use very simplified, smoothed, schematised or even caricatured geometries, all of the filtered data must undergo a geometric transformation. Rather than transforming all the geometries at once or in the same manner, we propose a layer-based approach (Figure 1). The first step is the generalisation of the territorial outline of the area of interest. Then, one or more thematic layers are mapped into this new geometry, for most cases with different methods than those that were used for the outline. The crucial task here is to ensure that both generalisation results can be combined again, in other words to ensure cartographic consistency via geometric anchors in regards to horizontal relations (Steiniger and Weibel, 2005). In a further step, geographic processes that have been identified in the data analysis that are implicit in the data but are not geographic entities themselves can be added. They need to be assigned some form of symbolisation from the dictionary of dynamic choremes. While most dynamic choremes are symbolised with arrows, their number, starting points and length as well as their horizontal relationships need to be defined from the particular semantics of the dataset and placed into the diagram in a consistent cartographic way. A very simple example for such an operation is an arrow in the direction of a population movement (Laurini et al. 2006). The database implicitly contains that information, but the gradient arrow itself has no geometric precursor in the dataset; the geometric and graphic properties need to be assigned via symbolization instead of being cartographically generalised.

It is important to note, that not all chorematic diagrams are alike. In fact, a strong case can be made that indeed several distinct classes with distinct cartographic expressions exist. From that it becomes clear, that any attempts at automation must use different methods for the distinct classes of chorematic diagrams. While the broad taxonomy has been presented (Reimer 2010), we here present a refinement of that taxonomy that takes into account the progress of data within the process chain.

While geometric accuracy has been used to differentiate between chorematic diagram classes, this is not precise enough for choosing methods & algorithms. Whether a chorematic diagram's territorial outline uses curves or line segments only, is crucial to the methods used for automated generalisation. The following sub-classes of chorematic territorial outline depiction have been encountered in the published evidence:

- symmetric polygons that can be defined by less than their full point list (ex: circle is defined by one point and a radius, a square by the points of the diagonal etc.); Figure A

- polygons without parametrised curves that are defined by a list of points; Figure F

- polygons with some vertices as parametrised curves; Figure B

- polygons only made from parametrised curves; Figure C

- polygons that can be fully described by a set of circular arcs (sometimes called Lombardi-style, after Mark Lombardi); Figure D

The chorematic territorial outlines furthermore can present geographic space either with the region of interest acting as a container (Inselkarte, Figures A-E and F), or as a continuum, only bounded by the edge of the map (Rahmenkarte, Figure G). From a geometric viewpoint this fact seperates methods that can be used for polygons from more sophisticated methods needed for handling polygonal subdivisions. These are needed for area symbolisation, too.

Area Symbols & Subdivisions are needed for the most expressive parts of chorematic diagrams, the geometric shapes within the region of interest. The shape depicting the major economic growth corridor of the European Union, the famous Blue Banana, is the prime example. From a generalisation viewpoint, the different ways that areas are visualised in chorematic diagrams are:

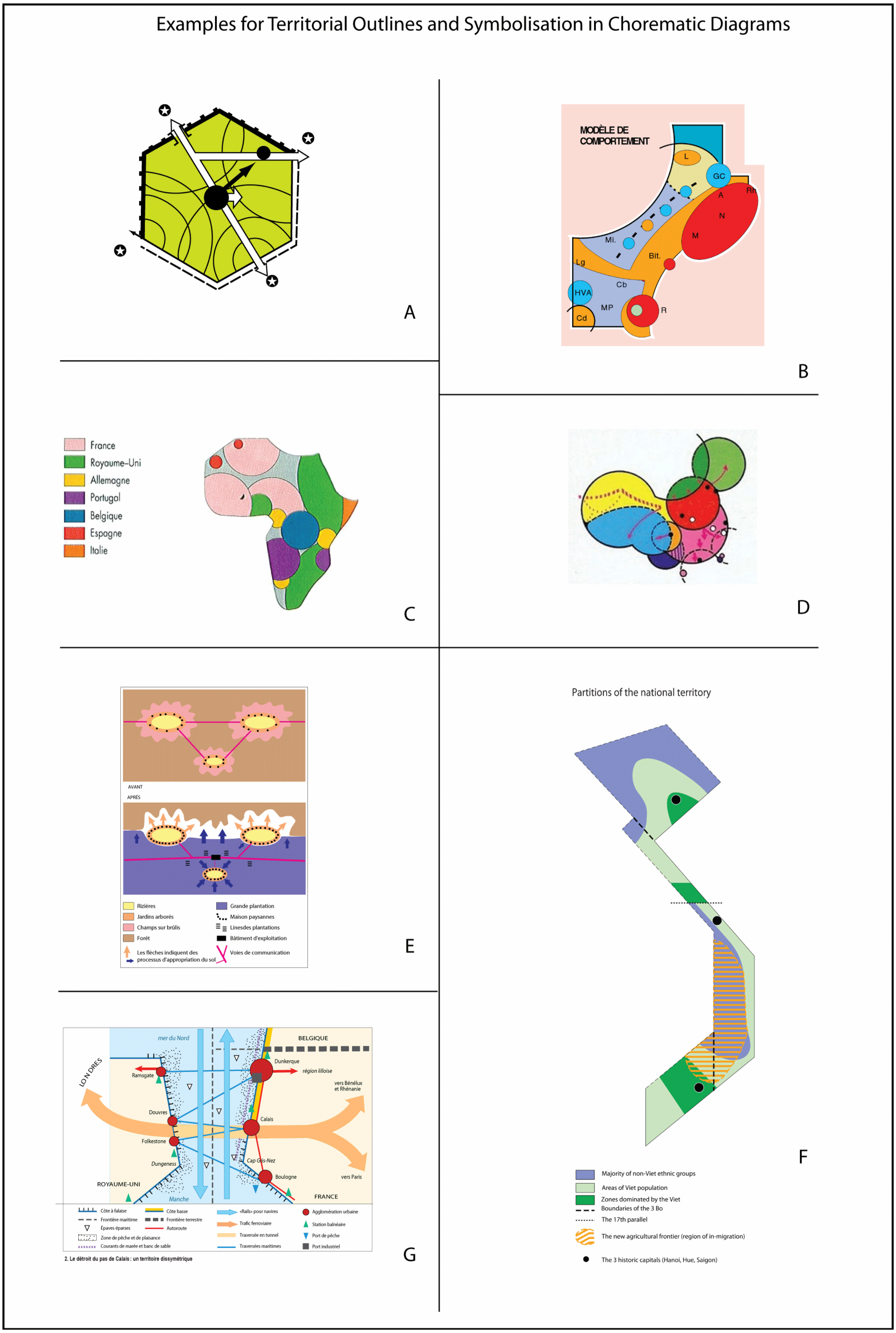
- polygonal, reduced set of angles (rectilinear/octilinear etc.; cf. Fig. 6 in Reimer 2010)

- parametrised curves anchored at the outline; Figure F

- parametrised curves within the outline i. e. isolines and clines

- circular arcs only; Figures C and A

- smoothed, non-convex hulls Figure G



This differentiation highlights the fact that different methods & algorithms can and must be used to reach the specific style.

The thematic layers in relation to each other can be understood as either being vertically interacting with or not interacting. The difference lies in the fact that thematic layers that interact with each other and are to be portrayed in just one diagram must take layer hierarchies into specific consideration. A layer will then have to take visual precedence over another in one part of the diagram, where the ordering is reversed at another place, depending on the specific interaction. This requires a modelling of the interaction scheme on a case by case basis rather than a straightforward layer hierarchy that is applied across the whole diagram. Solving such issues is still an open problem.

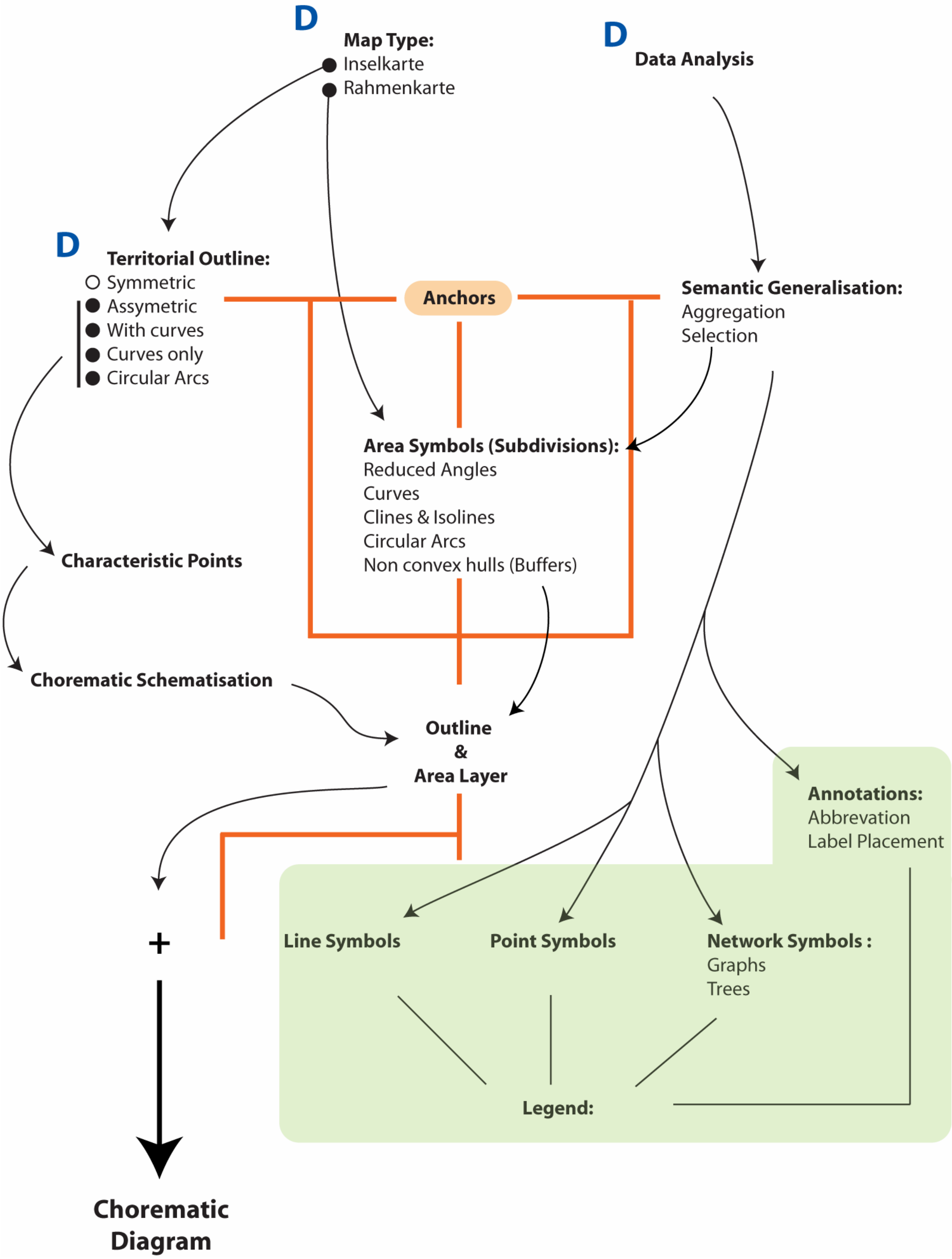


Figure 1: Detail of the cartographic sub-processes. Human decisions are needed at the steps marked with a blue "D".

RESULTS

In the above we have outlined which sub-problems have to be solved before the full process chain can be implemented. The sub-problem we have approached so far, is the generation of territorial outlines without parametrised curves. Research on other tasks is ongoing and we highlight our current proposals for tackling these.

For the generalisation of territorial outlines a simulated annealing method has been chosen.

Simulated Annealing for Territorial Outlines

Simulated Annealing is an optimisation approach, that has been used for other schematisation tasks in cartography (Agrwala and Stolte 2001; Anand et al 2007). The basic idea behind simulated annealing is to iteratively produce solutions from a given input and compute a score for the solution.

In the case of the territorial outline production, the input is the polygon: area features made up of edges, which in turn are made up of vertices. The initial state is evaluated using a scoring function S; this function assigns to the input state a score that reflects how well it measures up against a set of given constraints (Beard 1991).

To generate the input polygon, regular datasets of the given region are simplified with point removal techniques such as Imai-Iri or Douglas-Peucker algorithms. Self-parametrisation is done by selecting a a specific number of characteristic points for the output, based on chorematic class choice.

The score is compared each time with the current best score achieved so far. If the score is higher, the new solution is kept as current best one. The number of iterations is covered by a symbolic temperature, that decreases with every iteration. That temperature also governs the probability of how the new solutions is being computed. The higher the symbolic temperature, the higher the probability that a totally randomised solution is computed, whereas with decreasing temperature, the new solutions will be small modifications of the last solution. Due to that strategy, simulated annealing has the ability to avoid getting stuck in local maxima, while still settling into an acceptable refined solution after the cooling point is reached.

The scoring function is the crucial part of a given simulated annealing approach. Following our proposed for generalisation constraints for chorematic schematisation tasks; Reimer and Fohringer 2010), we based the scoring function on parallelity of line segments. The scoring function calculates all edge slopes, and compares them, assigning the normalized edge length as a score. All edges that share their slope value with another will contribute to the score. A bonus score of with the value of those parts of normalized edge lengths that are parallel to each other is assigned.

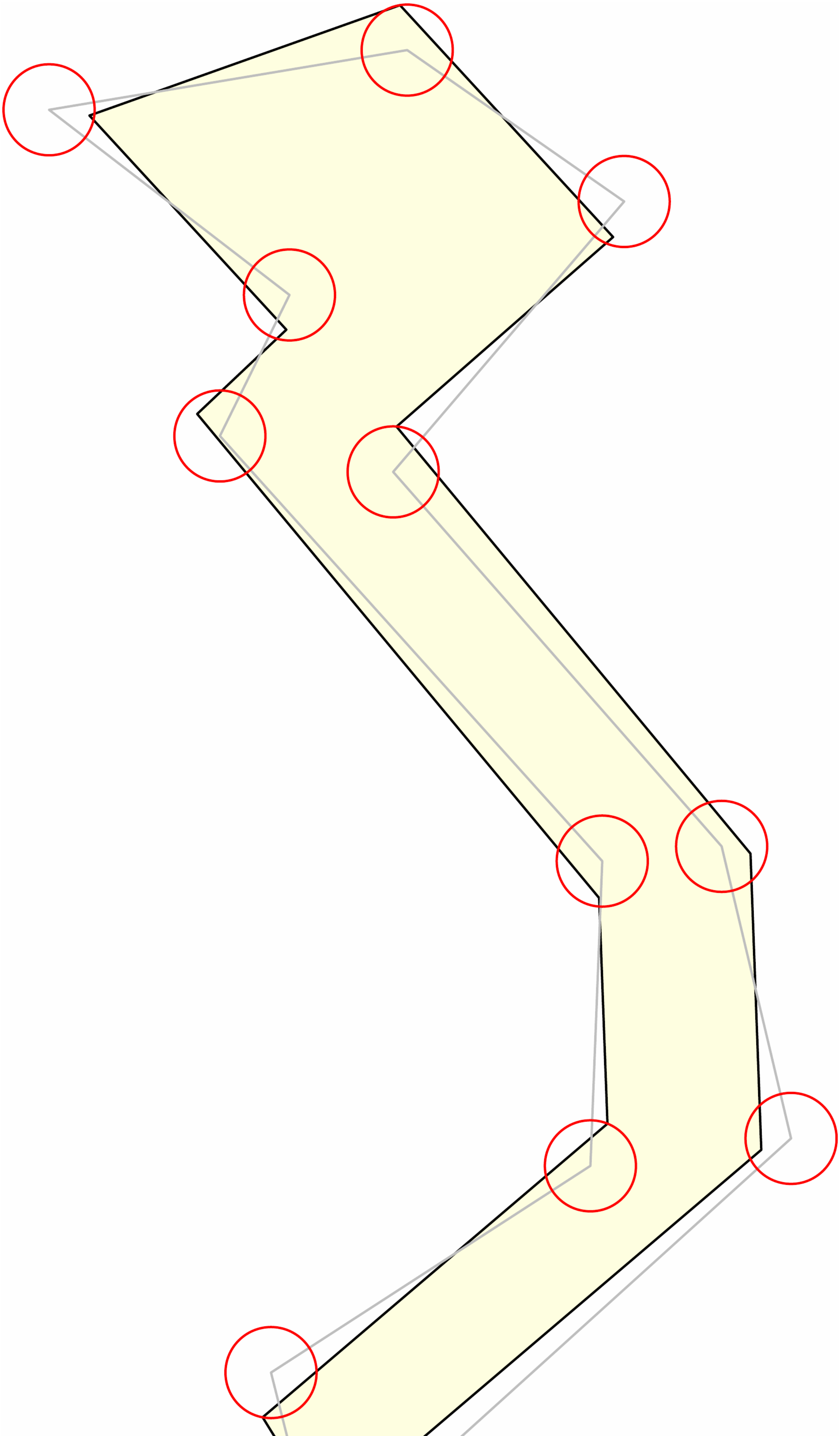


Figure 2: Recreation of the chorematic outline of Vietnam. Circles indicate the maximum allowed distance a single point may be moved, parametrised as fraction of the total circumference of the polygon.

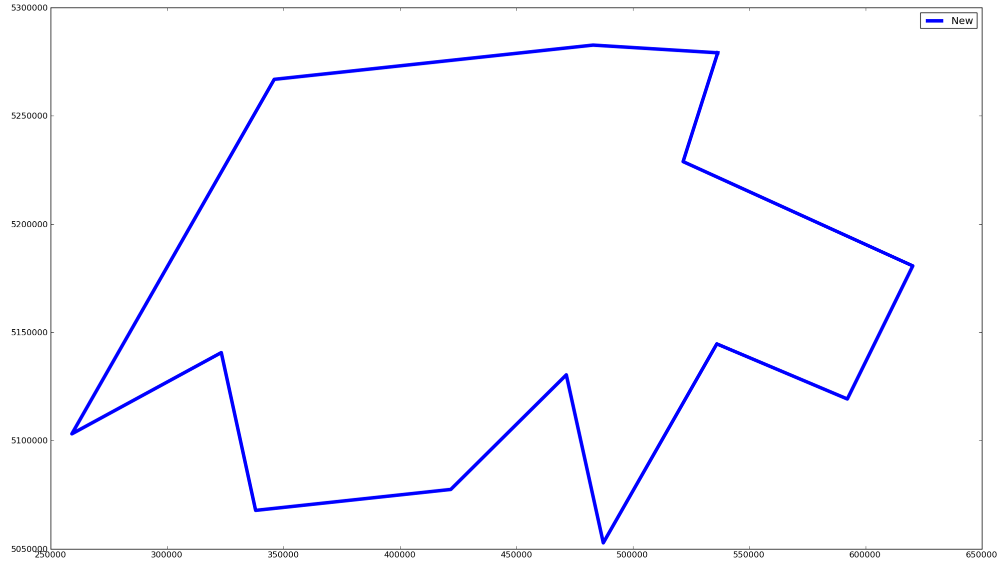


Figure 3: Generalisation result for Switzerland, using the same parameters as for Vietnam.

Bezier-curve-fitting for choropleth to area-class map conversion

One of the most common types of thematic geodata as well as maps are in the choropleth-style. As chorematic diagrams depict mainly spaces with similar qualities instead of the qualities of a given administrative region, area-class maps are needed (Mark and Scillag 1989; Bucher and Schlömer, 2006). If we assume that some form of aggregation and selection has already taken place, the new subdivisions must be drawn. Currently, we work on fitting cubic Bezier-curves to the aggregated dividing polylines, and provide a spatial false-attribution measure. While area-preserving geometric solutions are still an open problem due to the high degrees of freedom of control point placement for Bezier-curves, several approximation strategies exist (e. g. Ito and Ohno 1993; Masood and Ejaz 2010), which are being evaluated.

CONCLUSION & FUTURE PLANS

We have outlined a process chain for the construction of chorematic diagrams, and have highlighted the sub-problems that need to be solved. Our approach moves back human intervention to the start of the process, and for several cases, a projected system could either assume some of those decisions, or more practical, be tailored for recurring tasks. User decisions (marked with blue „D") are then made once at inception only. For example, if an decision support system is to be furnished with automatically generated chorematic diagrams on population concentration, the type of outline, area-symbolisation, city hierarchy and density classes would be made just once. Then the corresponding diagrams can be generated fully automated once the current region of interest is clear.

For future research, we concentrate on widening the polygonal techniques to curves and especially subdivisions, as these are the most common form of chorematic expression techniques. Another very important area remains the detection strategies for the high level dynamic choremes, such as tendency surfaces, and tissues of change (Brunet 1987). While being carto-syntactically straightforward (arrows, non-convex hulls buffers around tree-structures, point symbols etc.), they pose a serious challenge towards semantic generalisation in all but the most simple cases.

REFERENCES

Images: [A] Renerie, R. (1988): La carte-modèle: une carte pour comprendre, une carte pour agir. L'exemple du Bassin de Belly. In: Mappemonde 1/88, p. 9. [B] Brunet, R. (1991): La population du Languedoc-Roussillon en 1990 et la Croissance Recente (1), MappeMonde, 91, pp. 34–36. [C] Brunet, R. and Dollfus, O. (1990): Mondes Nouveaux, p. 300, Hachette-Reclus, Montpellier. [D] Gentelle, P. (1987): Chine, un atlas économique, Reclus. [E] Meyer, E. (1992): Types de contact village-plantation à Sri Lanka 1840-1940. In: Mappemonde 4/92, p.37. [F] Taillard, C. and tu Lâp, V. (1994). Atlas du Viêt-Nam. Reclus.

[G] Picouet, P. And Renard, J.-P. (2002): Les détroits: de nouveaux territoires? L'exemple du pas de Calais. In: Mappemonde 1/2002, p.9.

Agrawala, M.; Stolte C. (2001): Rendering effective route maps: improving usability through generalization. In: Proceedings of SIGGRAPH 2001. Los Angeles: 241-250.

Anand, S.; Avelar, S.; Ware, M.; Jackson, M. (2007): Automated schematic map production using simulated annealing and gradient descent approaches. In: 15th Annual GIS Research UK Conference.

Beard, M. (1991): Constraints on rule formation. In: Map Generalization: Making Rules for Knowledge Representation, Longman.

Bertin, J. (1973): Graphische Semiologie. Diagramme, Netze, Karten, Gruyter, Berlin.

Brunet, R. (1987): La carte: Mode d'Emploi, Fayard-Reclus, Paris.

Bucher, H.; Schlömer, C. (2006): Die neue Raumordnungsprognose des BBR. In: Raumforschung und Raumordnung 3/2006, pp. 206-212, Springer.

Cheylan, J.-P., Libourel, T. and Mende, C. (1997). 'Graphical modelling for geographic explanation', in Spatial Information Theory: A Theoretical Basis for GIS, pp. 473–483, Springer.

Harrie, L.; Stigmar, H. (2007): An evaluation of measures for quantifying complexity of a map. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences Volume XXXVI – 4/W45 covering contributions of the Joint Workshop "Visualization and Exploration of Geospatial Data" June 27-29, 2007, Stuttgart.

Itoh, K., Ohno, Y. (1993): A Curve Fitting Algorithm for Character Fonts, pp. 195–205. John Wiley and Sons Electronic Publication.

Laurini, R., Milleret-Raffort, F.; Lopez, K. (2006): A primer of geographic databases based on chorems', Lecture Notes in Computer Science, 4278, pp. 1693–1702.

Ligozat, G.; Nowak, J.; Schmitt, D. (2007): From language to pictorial representations. In: Proc. of the Language and Technology Conference (L&TC'07), Poznan, Poland, 5-7 September, 2007.

Mark, D. M.; Csillag, F. (1989): The nature of boundaries on 'area-class' maps. Cartographica 26(1), pp. 65–77.

Masood, A.; Ejaz, S. (2010): An efficient algorithm for robust curve fitting using cubic Bezier curves. In: D.-S. Huang et al. (Eds.): ICIC 2010, LNAI 6216, pp. 255–262, 2010.

Reimer, A. (2010): Understanding Chorematic Diagrams: Towards a Taxonomy. In: The Cartographic Journal, Vol. 47 Nr. 4 pp. 330-350, 2010.

Reimer, A.; Dransch, D. (2009): Information aggregation: Automatized construction of chorematic diagrams. workshop (GeoViz & The Digital City), 2009, March, Hamburg, Germany.

Reimer, A.; Fohringer, J. (2010): Towards constraint formulation for chorematic schematisation tasks". In: Geographic Information on Demand: 13th Workshop of the ICA commission on Generalisation and Multiple Representation, Zurich.

Steiniger, S.; Weibel, R. (2005): Relations and structures in categorical maps. In: 8th ICA Workshop on generalisation and Multiple Representation, A Coruna (Spain), July 7-8th, 2005.

Van der Aalst, W.M.P. (2009): Using Process Mining to Generate Accurate and Interactive Business Process Maps. In: A. Abramowicz and D. Flejter, (ed): Business Information Systems (BIS 2009) Workshops, volume 37 of Lecture Notes in Business Information Processing, pages 1-14. Springer-Verlag, Berlin.