Polygon-based Regionalisation in a GIS Environment

Dirk TIEDE and Josef STROBL

1 Introduction

Regionalisation is a frequently used concept in the spatial sciences, applied towards numerous use cases and real world scenarios. Although underlying data describing spatial entities are mostly available in GIS-compatible frameworks, the task of defining regions is still mostly performed manually, i.e. not based upon stringent algorithms with explicit parametrisation. One reason for that is the lack of methods implemented as explicitly spatial aggregation techniques in software products. While classification methods leading to spatial types (zones) are widely available, regionalisation adding a spatial contiguity or proximity constraint to attribute-based classification is missing from even extended toolsets.

Openshaw (1996) pointed out that the analysis of spatial information aggregated to zones is one of the most important and highly relevant spatial analysis tasks which still need to be conceptualised for wide implementation. Interestingly, one of the few implementations of regionalisation based on an extension of Cluster Analysis has been implemented in the 1970ies from a multivariate statistical analysis perspective (Wishart 1987)

In order to enable spatial analysts to perform clearly defined and parameterized regionalisation, a hierarchical, aggregative, spatially constrained clustering method has been implemented by one of the authors (D. Tiede), starting from polygons as operational taxonomic units (OTU's).

In this paper an example for regionalisation in the field of renewable energy usage is presented, but possible areas of application are manifold: Depending on the potential pre-defined rules, region aggregation could be based on similarity criteria (e.g. in the ecological modelling domain) as well as on complementary criteria (e.g. to aggregate self-sustaining regions in case of resource use).

2 Method

2.1 Polygon-based regionalisation using a spatial contiguity constraint

Johnston (1976) introduced a contiguity constraint to the hierarchical clustering with centroid replacement procedures. At each hierarchical clustering step, an adjacency matrix had to be checked (manually), to see if the candidate groups of units were contiguous and if that was not the case, they were not considered for aggregation.
In contrast to Johnston the approach described in this paper starts by first selecting contiguous OTU's and then comparing their attribute values for similarity-based clustering.

To do this, the algorithm needs pre-set "seed" polygons as starting units. Next, all adjacent polygons are considered in a first aggregation step, resulting in a number of initial regions. Thus, the number of "seed" polygons determines the maximum amount of resulting regions after all polygons are clustered. A contiguity constraint as it is used herein is defined as adjacency of polygons, which means that they have at least one point of a boundary in common (cf. DeMers 2005). Nevertheless, the contiguity constraint is customizable and could also be implemented as sharing a defined length of common boundary between polygons or for example as the distance between polygon centroids.

The clustering process itself depends on customisable rules, comparing similarity or also complementarity attributes of the spatially contiguous target OTU's. The best fitting neighbouring polygon (in terms of the pre-defined rules) is merged with the initial one and new common values for the merged region are calculated according to the chosen aggregation rule. In an iterative process this is done for the entire dataset until all units are assigned to regions or another cut-off criterion is reached (e.g. maximum region size, level of dissimilarity etc).

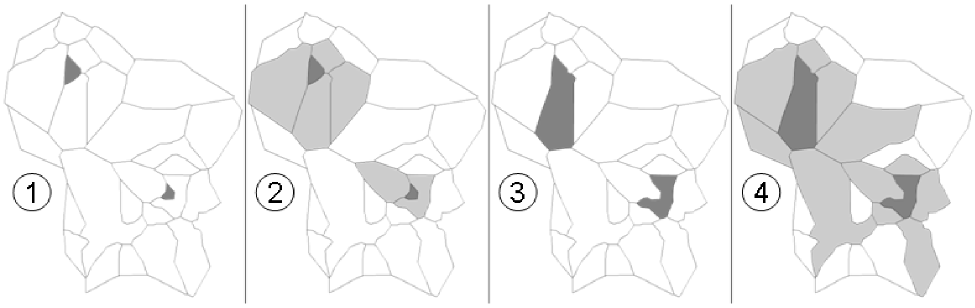

Fig. 1: Principle of the algorithm: 1. Initial "seed" polygons – 2. Selection of contiguous candidate OTU's – 3. Merging of the most similar OTU's based on pre-defined rules – 4. Next selection step

Clearly, this is a region growing algorithm starting from a fixed set of seed polygons, proceeding with a stepwise aggregative clustering strategy. Alternative approaches like divisive or non-hierarchical agglomeration techniques are not considered in this paper.

2.2 History tracking

The regionalisation process as explained above is recorded for traceability reasons. Traceability can especially be a problem with large datasets and a few hundred aggregation steps. It is also important to ensure reasonable results and to avoid a "black-box" effect, to give the user as much control as possible over the parametrisation and calculation processes. In a separate table each merged OTU gets a unique time stamp identification number (history ID). Therefore it is possible to assess and analyze the region building sequence for all original polygons = operational taxonomic units.

Additionally a tool has been developed which uses the 'history IDs' to visualize the iterative region building as animated graphics, dynamically presenting the region growing process.

2.3 Implementation

The described algorithm is implemented as an extension in the ArcGIS 9 environment. So far input and output datasets are limited to polygon shapefiles. A simple graphical user interface helps to calibrate the initial conditions and to define input and output parameters.

3 Example: Building autonomous regions based on renewable energy consumption and sources

The above described software tool is being tested for the delineation of 'energy-autonomous regions'. The objective is to balance renewable energy sources and consumption in particular areas, minimizing import of (fossile-based) energy from the outside. The data is based on a combination of the effective production potential for renewable energy sources like hydropower, wind, and photovoltaic energy, biomass conversion and geothermal energy. The power consumption per unit is subtracted, resulting in a polygon layer, representing the aggregated energy supply potential (cf. figure 2).

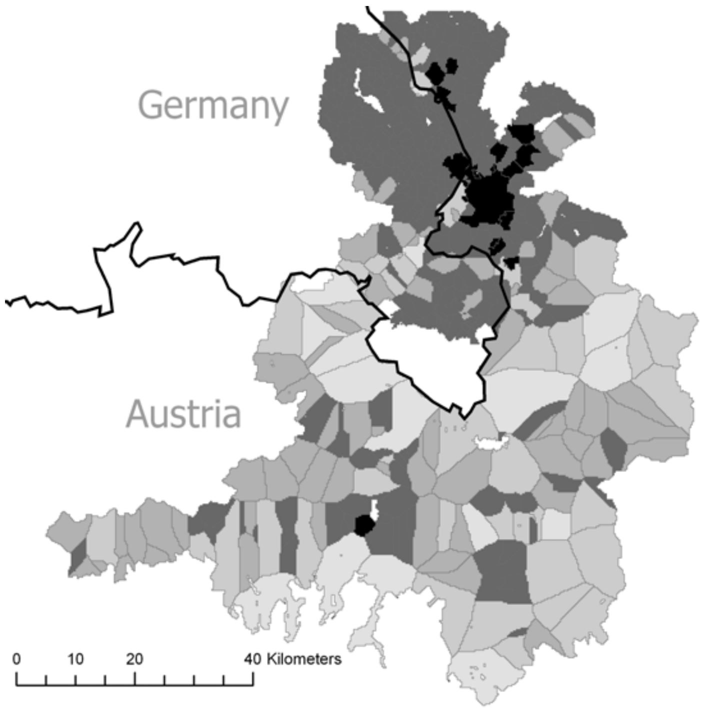

Fig. 2: Polygon layer, representing the aggregated energy potential in a transborder region between Austria and Germany. Dark values indicate higher values of energy consumption compared to locally available renewable energy potential. Brighter values indicate a net surplus, i.e. a higher renewable energy generation potential compared to average energy consumption.

Negative values indicate higher values of energy consumption compared to the renewable energy potential, while positive values are indicating a higher renewable energy potential than average energy consumption.

Renewable energy sources are primarily used to meet the demand on local level, mainly due to the low energy density and resulting high cost of transport. Exploitation on a higher than regional level is limited by economical and technical constraints as well as by resource availability and societal acceptance of transmission infrastructures. The basic conditions for the regionalisation in this example were the balance of energy sources and energy consumption inside regions to achieve a high degree of regional autonomy concerning energy supplies. Therefore the regionalisation algorithm was adapted by customizable rules to merge polygons with complementary attribute values. A polygon with an energy potential above zero (net excess potential of renewable energy) should merge with a neighbouring polygon with a value below zero (low potential of renewable energy but high energy consumption = net energy deficit). Five densely populated areas with high energy consumption rates (i.e., population centres) are chosen as “seed” polygons, building the starting points for region growing (cf. figure 3).

4 Results

Figure 3 shows on the left the initial data set with five initial “seed” polygons and on the right the resulting regions. The target specification is completely fulfilled: All resulting regions are optimized concerning their level of self sufficiency, minimizing in- and output transfers.

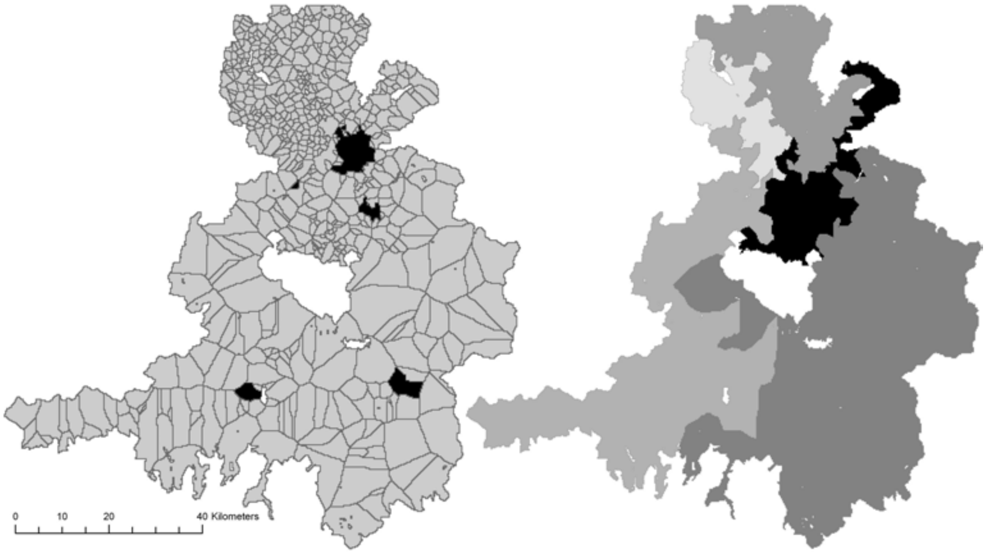

Fig. 3: Renewable energy units and five urban agglomerations as “seed” polygons (left). Resulting self-sufficient regions for renewable energy sources (right).

Clearly the result is linked to the set of initial polygons. Not only the minimum number of regions is determined by that, also the region forming is influenced. This mainly depends on the spatial restrictions given by the dataset: If a growing region is surrounded by other regions, a balanced growth is influenced or even stopped by competing cluster building.

5 Conclusions & Outlook

In this paper a regionalisation tool based on the implementation of a hierarchical, aggregative, spatially constrained clustering method has been described. An example for regionalisation in the field of self-reliance on renewable energy usage was demonstrated and promising results are shown. Possible further areas of applications are manifold: Depending on the potential pre-defined rules, region aggregation could be based on homogeneity criteria (e.g. in the ecological modelling domain) as well as on complementarity criteria. Future work will target various areas:

- Integration of further parameters and statistical rules for the clustering process.
- A weighting algorithm to bias the region building process. This is already in development status.
- Implementation of a form descriptor (e.g. shape index), to obtain more compact regions and therefore avoid elongated region shapes.
- Optimisation of processing time, especially for large datasets.

This paper addresses multiple methodology issues which are currently not fully resolved for GIS analysts using standard software:

- Regionalisation of operational taxonomic units like simple polygons by adding proximity or contiguity constraints to attribute-based classification techniques.
- Implementing regionalisation based on pre-set seed locations (e.g., population centres).
- Generalising the standard usage of similarity measures towards measures of complementarity, demonstrating a conceptual approach to identifying self-sustaining regions.

Although concepts and practical implementation are only briefly outlined in this paper, the far-reaching potential of regionalisation methods as a currently underrepresented set of spatial analytical techniques is clearly demonstrated.

6 Acknowledgements

The presented tool was developed in cooperation with the Research Studio iSPACE (a division of ARC Seibersdorf research GmbH) within a project co-funded under the programme "Energy Systems of Tomorrow" – an initiative of the Austrian Federal Ministry of Transport, Innovation and Technology.

7 References

Abler, R., J.S. Adams and P. Gould (1972): *Spatial organization: The geographer's view of the world.* London: Prentice-Hall International.

DeMers, M. N. (2005): *Fundamentals of Geographic Information Systems.* 3rd Edition, New York, John Wiley and Sons, Inc.

Johnston, R. (1976): *Classification in Geography*. CATMOG No. 6 (Concepts and Techniques in Modern Geography) Study Group in Quantitative Methods of the Institute of British Geographers, London, 1-40

Openshaw, S. (1996): *Developing a GIS-relevant zone-based spatial analysis method*. In: Longley P., and M. Batty (Eds), Spatial Analysis: Modelling in A GIS Environment, Geoinformation International, 55-73

Wishart D. (1987): *Clustan User Manual,* University of Edinburgh Press.