Acknowledgement

At this point I would like to thank everybody who supported me at my work for this thesis. I'm indebted to Prof. Dr. Alfred Ultsch (Philipps-University of Marburg) as he provided his helpful support within regular discussion meetings which resulted in valuable input for my work. Sincere thanks goes to Prof. Dr. Josef Strobl for his advice during my whole study. The basic data for completing this work was delivered by the Federal Statistical Office (=BFS, Bundesamt für Statistik) and the Federal Office for Spatial Development (=ARE, Bundesamt für Raumentwicklung). Special thanks go to all members of UNIGIS Msc for guiding me through my whole study period at the Paris-Lodron-University of Salzburg. Finally I would like to thank my whole family for their continuous motivation.

Hint

It is to emphasize that the work is mainly based on an already existing program library (DBT). It has been developed by the Databionics research group for scientific purposes. The author wants to thank the whole group for the possibility to use this program library. Please contact Prof. Dr. Alfred Ultsch for further information (Department of Mathematics and Computer Science, Philipps-University of Marburg)

Declaration of Own Work

I hereby declare that the work submitted is my own and that all passages and ideas that are not mine have been fully and properly acknowledged. I confirm that I have referenced and put in inverted commas all quoted text (from books, web, etc). I have given the sources of all pictures, data etc. that are not my own. I have not made any use of the essay(s) of any other student(s) either past or present. I understand that any false claim will be penalized in accordance with the University regulations.

gez. M. Behnisch

Martin Behnisch - Zurich, 30. November 2009

Abstract

Most of the large databases currently available have a strong spatiotemporal component and potentially contain information that might be of value. Spatial analysis is far from adequate handling the huge volumes of data and the growing complexity. Based on Data Mining techniques and Knowledge Discovery the population development of the 2896 Swiss communities is examined by time intervals. The time intervals orientate on the census days of the Swiss federal population census, which was carried out every 10 years since 1850 (=15 decades). The question is how many patterns will occur. The patterns are described concerning their characteristics (size and properties by decade). The patterns with similar properties are grouped into classes. To explain these classes, the discovered classification is compared with already existing classifications of Switzerland (height zones, community types, urban rural types and NUTS2-Regions). The classes are presented in localized way and proofed in mind of the spatial analyst. By using cartograms the communities and their patterns are presented in proportion to their middle population during the 15 decades.

Kurzfassung

Durch den schnellen Fortschritt in der Informationstechnologie und das rapide Anwachsen der Datenmengen mit raumzeitlicher Komponente steigen die Anforderungen, aus diesen Daten Wissen zu extrahieren und darzustellen. Auf der Grundlage von Techniken des Data Mining und der Knowledge Discovery wird die Bevölkerungsentwicklung der 2896 Schweizer Gemeinden anhand ausgewählter Zeitschnitte untersucht. Die Zeitschnitte orientieren sich an den Zählungsstichtagen der Eidgenössischen Volkszählung, die seit 1850 alle 10 Jahre durchgeführt wurde (=15 Dekaden). Es stellt sich die Frage wie viele Entwicklungsmuster in der Schweiz existieren. Die Aufgabe besteht darin, relevante Entwicklungsmuster zu identifizieren. Die Entwicklungsmuster werden nach Art und Ausmaß präzisiert. Die Entwicklungsmuster mit ähnlichen Eigenschaften werden in sogenannte Entwicklungsklassen gruppiert. Um die Entwicklungsklassen zu erklären, wird die gefundene Klassifikation über Kontingenztabellen mit bereits bestehenden Klassifikationen in der Schweiz verglichen (Höhenstufen, Gemeindetypen, Stadt-Land-Typen sowie NUTS2-Regionen). Die Eigenschaften der Entwicklungsmuster werden lokalisiert dargestellt. Mit Hilfe von Kartogrammen werden die Gemeinden und ihre Entwicklungsmuster in Proportion zu ihrer Einwohnerzahl über die 15 Dekaden abgebildet.

Contents

List of Figures

List of Tables

List of Abbreviations

BFS	Swiss Federal Office of Statistics
GIS	Geographical Information System
ESOM	Emergent Self Organizing Map
PDE	Pareto Density Estimation
GWR	Geographically Weighted Regression
SAR	Spatial Auto-Regression
LISA	Local Spatial Autocorrelation Analysis

1 Introduction

In general, mankind's demographic history is characterized by periods of more or less dramatic growth, but also by times of stagnation and decline. The demographic growth bears witness to the processes of interaction between populations and their environment, as well as to the relations between individuals and between subsets of societies.

The comprehensive description of the long-term changes in population in a continuous area such as Switzerland demands a uniform data collection over a longer period. Switzerland offers such a memory indicating the spatial, social and economic development of the Swiss Confederation over the last 150 years. A Census was held in Switzerland every 10 years since 1850. The data collection was initiated by the great efforts of the Federal Councilor Stefano Franscini (1796-1857). Today it is therefore possible to decode the population dynamics on the level of communities.

Although the Swiss heartland and the great Alpine valleys have been well populated since the first millennium before Common Era, the total population of Switzerland has always remained modest in size compared with the neighboring countries (Rothenbacher, 2002).

The territory within the present national borders had less than half a million inhabitants in the year 1000. It was only in about 1600 that the one million mark was passed. Like Europe as a whole, Switzerland experienced a marked growth from the 18th century onwards. In 1848, when the Federal State was established in 1848 the number of inhabitants in the country reached 2.4 million. The development in the analyzed time period is characterized by a continuous population growth but with different intensities. Periods with strong growth can be observed around 1900, the time after the Second World War until 1970 and the last two decades until 2000. During the last decades Switzerland had one of the largest growth rates in Western Europe (Haug, 2002).

The population has tripled since 1850, amounting to 7.3 million in the year 2000.

At present, stabilization in the population can be observed, as is typical of the Western world. Swiss population accounts for 0.1% of the world's population (Watkins, 2007). According to the most plausible estimates (BFS, 2006), it should only continue to rise slightly in the course of the next decades, reaching a ceiling of some 8.2 million towards the year 2030, after which it will probably decline. Other prognoses points at an early decline in the year 2015 and a population of 6.5 million in 2050.

1.1 Motivation

Urbanized areas are a major component of the modern environment. For the first time, more half of the world's population will be living in urbanized areas by the end of this decade (United Nations, 2009). Switzerland is a highly urbanized country. Urban, suburban and rural areas are closely linked by dense flows of people, goods, materials and information. It is widely acknowledged that the contemporary physical space presents a complex structure; research on the nature of this structure and the pattern of its growth has remained indispensable. For many years it was very difficult to start a long-term analysis in Switzerland and the alpine regions in general because of heterogeneous spatial and statistical definitions and a general lack of uniform data.

Against this background the motivation of this thesis is to brighten the knowledge about the long-term population development of Swiss communities in terms of patterns and localized properties. The initial idea of the thesis is the analysis of the established census data of the year 2000 (Schuler et al., 2002). Actually urban planners and politicians have several impressions about the recent problems of Swiss population losses in peripheral alpine regions (source, year) as well as about the urban sprawl in the Midland (Tschopp et al., 2003; Oswalt/Baccini, 2003). The long-term development of all Swiss communities is often not quantified and therefore more or less nebulous in context of actual planning and decision processes (Bätzing, 2001). But the long-term aspects should be taken into account to avoid dramatic losses of economic, social and cultural capitals in the coming decades.

1.2 Assumptions and Questions

The investigation of Swiss communities will focus on the long-term behavior of population between 1850 and 2000 and is based on several assumptions. These assumptions lead to specific research questions and future perspectives:

1. Spatial data analysis (in conjunction with the techniques of data mining) is an appropriate way to quantify the long-term development of communities. The summary of a large amount of communities (2896 objects) to a smaller amount of meaningful patterns leads to a better understanding of the processes generating the attribute values. Thus it is possible to formulate hypothesis on the general development of communities in the past.

   Questions: How many different (long-term) patterns exist in Switzerland?

2. Several development directions are characterizing the 2896 Swiss communities. The variety of existing long-term patterns is extremely wide and differing. The expected patterns are therefore not just influenced by an increase or decrease but probably by a multitude of opposing and recurring trends. Clustering will support the interpretation and characterization of the observed population episodes in a transparent and systematic way.

   Questions: What are the relevant patterns of population development? How many clusters of patterns do exist? What are the characteristics of clustered long-term patterns? Where do these patterns occur?

3. The amount and change of population might be influenced both by other spatial and non-spatial characteristics. It is expected that the communities are not independent of each other. Attribute values in nearby places tend to be more similar than attribute values drawn from locations far away from each other.

   Questions: Are there official classifications (spatial or non-spatial) that might be valuable for the explanation of patterns? Are there analytical techniques that might be of interest for the explanation of such results?

Future Perspectives: What are the possibilities for further spatial investigations?

1.3 Modus Operandi

The description of the long-term development in a continuous area such as Switzerland requires a systematic analysis. The approach of this thesis refers to a cyclical data mining approach (see for deeper information into the lecture notes of Prof. Ultsch, 2009). A central issue of data mining is the transition from data to knowledge. An important goal of knowledge discovery is the search for patterns in data that can help explain the underlying process that generated the data. Techniques of data mining are therefore of high relevance to reveal logical or mathematical and partly complex descriptions of patterns and regularities inside the set of 2896 Swiss communities.

- Inspection of Data and Modeling: A preliminary data observation is taken into account as it helps to ensure the integrity of the data set before proceeding further with the analysis of the long-term population development. Typical Indexes of population change will be discussed in view of their properties. Hypothesis about the data distribution will be formulated. A data model will be elaborated and later on proofed.

- *Properties of population change:* At the beginning of this thesis it is assumed by the author that Swiss communities allow a grouping process based on processes of population decline, stagnation or increase. It might be possible to identify specific development profiles by using characteristics of all 15 decades. Such characteristics are explored to trigger discussions in the application domain and to reveal insights about spatiotemporal phenomena. Since communities with similar characteristics are occurring a clear definition and description of their properties should be elaborated.
- *Clustering:* Clustering is an appropriate method in context of the comparison of the population development over time. Such method is often applied to group objects such that objects in the same group are similar and objects in different groups are unlike each other. A great deal of variability in the range and distribution of variables is a problem for cluster algorithms which involves distance measurements. Since the data does not follow a normal distribution, other techniques and transformations should be taken into account to achieve normality or symmetry. In this case other techniques should be integrated to ensure the clustering and classification.
- *Cluster Explanation:* Since a partition of classes is realized it might be good to foster the understanding of their spatial characteristics. In view of already existing techniques in the field of data mining a comparison with other common spatial classifications should be developed. The explanation provides the mentioned transition from data to knowledge and generates several hypotheses for further investigations. The partition of classes and the machine-generated explanations should be validated mindful of the spatial analyst.
- *Geovisualization:* The process of knowledge conversion and communication is supplemented by spatial reasoning. The style of the presentation and the technical realization is task of the population cartography as a sub domain of the thematic cartography. Maps dealing with population usually provide information of distribution, density, structure or the spatial long-lasting/temporary change of population and stratification (Witt, 1971). Against this background several results of the long-term population analysis will be presented in map and diagram form. Such maps can be used as basis for further spatial analysis and content based interpretation in a future perspective.

1.4 Restrictions

Due to the lack of several long-term data dimensions the initial idea of this thesis is to use the development of population between 1850 and 2000 as a kind of overall indicator for the observable (cumulated) situation of Swiss communities.

The level of communities is often discussed as an appropriate scale to analyze the population development (see section 3.2). However, the author wants to point out that the whole study depends on the quality and accessibility of Swiss population data. At this stage the different size of communities (area) cannot be taken into account due to official data restrictions and limited amount of time. In the future it might be possible to get access to other population data that is not directly influenced by statistical aggregation and official territories.

1.5 Audience

This thesis aims to be a contribution to everyone interested in the development of communities. In particular the long-term dynamic of population in Switzerland will be examined and presented. By discovering different patterns it might be possible to think in new spatial relations and neighborhoods (e.g. comparative strengths, interregional communication and cooperation), thus, communities obtain a new condition in a long-term perspective.

It is supposed that the integration of geographical information systems and in addition the application of Data-Mining techniques will sharpen the planner's present view to the past development of communities. As this work integrates techniques of Data Mining it may also be of interest in the field of Geographic Information Science. Practitioners in the field of Spatial Planning and Historians are requested to validate the detected long-term patterns and dependencies.

The long-term population analysis of this thesis might be useful to establish a more general framework for deeper investigations and enlarged explanations using other dimensions such as economic, social or cultural data. Quantitative spatial investigations in general might lead to advanced strategic instruments such as semi or fully automated urban monitoring systems or a benchmark system for regional policy. In the future it might be possible that politicians and planners might intensify the effort to integrate long-term analysis into the planning and design process. A comprehensive, dynamic understanding of the past evolution of communities is however an essential condition for formulating comprehensive and reliable long-term visions (e.g. 2030/2050).

1.6 Structure of the thesis

The thesis is separated into nine chapters. One of them provides fundamentals (chapter 2) to explain some theoretical aspects that are relevant for the later developed approach. A survey of literature in the subject area is called related work (chapter 3). The long-term analysis of the Swiss population between 1850 and 2000 is then divided in to 4 specifiable work packages:

- The description of data and the subsequent inspection and modeling process are the basis for the computed results (see chapter 4).
- Essentially, the quality of analysis is a function of the quality of the data. Due to the detailed investigation of data distributions several subtasks lead to the identification of so called patterns (see chapter 5).
- These patterns are analyzed by their frequency and also described by relevance. Such measurement provides later on clustering and classification of patterns (chapter 6). The localization and spatial reasoning supports the understanding of population change over time (15 decades) and the spatial distribution of patterns and related communities.
- The explanation of classified patterns is realized in chapter 7. Existing spatial typologies are compared to the classification results of this thesis.
- The discussion implies a critical drawback of the presented work, the methodological procedure and a short comparison of expected and observed results (chapter 8).
- Conclusion and some perspectives for future steps are to find in chapter 9.

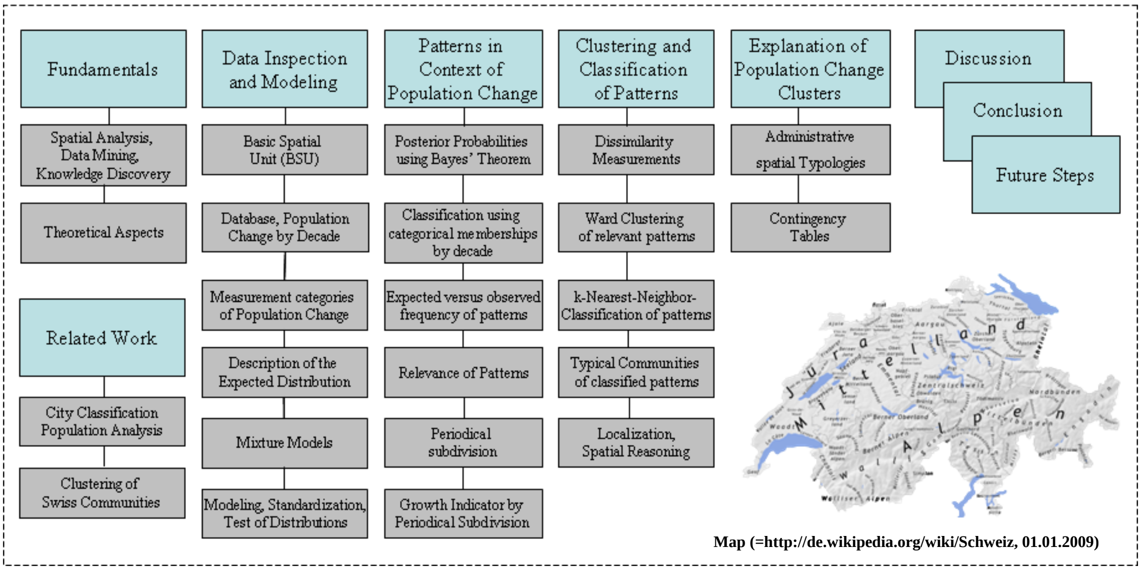


Figure 1: Structure of the thesis

2 Fundamentals

2.1 Spatial Analysis, Data Mining and Knowledge Discovery

Most of the large databases currently available have a strong spatiotemporal component and potentially contain information that might be of value. Miller and Han (2009) are quoted as follows: “Due to the growth and wide availability of geo-referenced data in recent years, traditional spatial analysis tools are far from adequate at handling the huge volumes of data and the growing complexity of spatial analysis task. Geographic data mining and knowledge discovery represent important directions in the development of a new generation of spatial analysis tools in data-rich environment”. Urban Data Mining (Behnisch, 2009) describes in similar manner such methodological approach to reveal logical or mathematical and partly complex descriptions of patterns and regularities inside a set of multidimensional geospatial data. Data Mining is commonly defined as the inspection of data. Mining implies a laborious process of searching for hidden information in a large amount of data (Han and Kamber, 2006). The ultimate goal of Data mining is to provide evidence-based insight through a deeper understanding of data (in the mind of the analyst) and to produce results that can be utilized at policy and strategy levels. Important requirements for ‘knowledge discovery’ are interpretability, novelty and the usefulness of results. Since the use of the term ‘data-mining’ is quite diverse, a short but more general definition of data-mining and knowledge discovery will be presented (Ultsch, 1987). Data mining means the inspection of a large data set with the aim of knowledge discovery. Knowledge discovery is the discovery of new patterns in the data, i.e. knowledge that is unknown in this form so far. This knowledge has to be presented symbolically and should be understandable for human beings as well as useful in knowledge-based systems. An important goal of knowledge discovery is the search for patterns in data that can help explain the underlying process that generated the data. A central issue of Data Mining is the transition from data to knowledge. The conversion of sub-symbolic patterns and trends in data to a symbolic form is seen as the most difficult and most critical part of data analysis (Ultsch and Korus, 1995). Symbolically represented knowledge – as sought by data-mining – is a representation of facts in a formal language such that an interpreter with competence to process symbols can utilize this knowledge. In particular, human beings must be able to read, understand and evaluate this knowledge. The knowledge should be useful for analysis, diagnosis, simulation and/or prognosis of the process that generated the data.

A cyclical data mining procedure was developed by the mentor of this thesis (Figure 2) in former times and later on successful applied. Applications of the methods are reported e.g. for medicine, meteorology, biology, pharmacy, stock prediction, customer relation management or spatial pattern detection and explanation. "Urban Data Mining" is interested in methods and approaches for community examination (Behnisch, 2009).

When analyzing population data in this thesis important steps of such cyclical methodology procedure are taken into account. The cyclical methodology procedure is characterized by six main tasks (Figure 2) following the initial step of data collection. The main tasks on the far right of Figure 2 contain several aspects within the circle and are roughly explained below. It should be considered that the analysis certainly starts with a relevant problem or specific research question. According to the presented steps within the circle there are often several combinations and processes necessary to find an appropriate solution or probably surprising answer. In particular, the cyclical approach provides the ability to identify hidden relationships and unusual patterns within a large amount of data. But human interaction is important during the mining process to analyze and validate partial results as early as possible and to guide further steps.

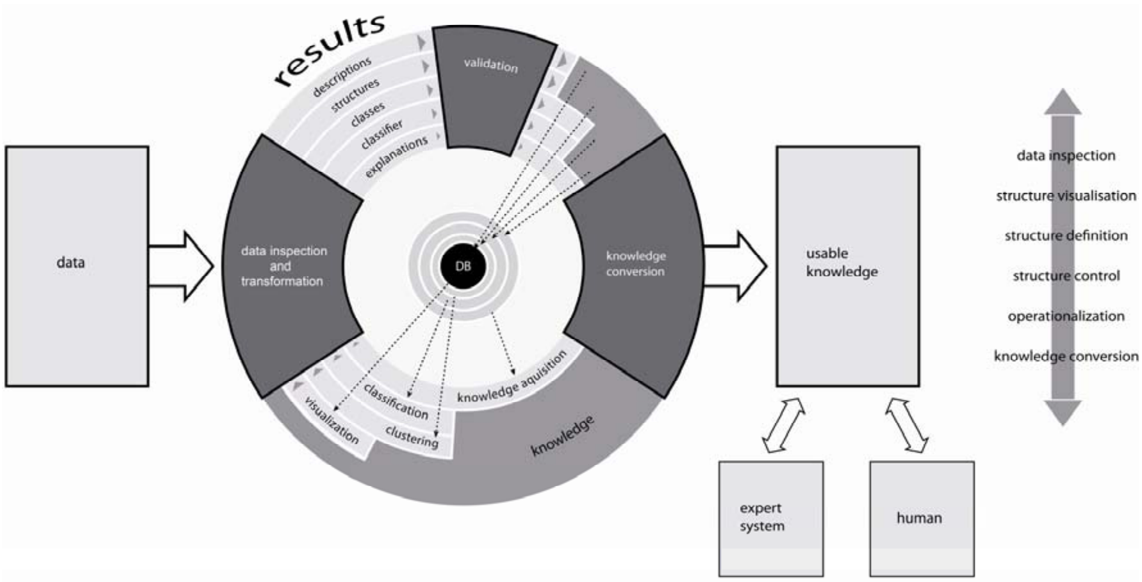


Figure 2: Cyclical Data Mining Approach, Source: Ultsch (2009), applied to community data in "Urban Data Mining" (Behnisch, 2009)

Data inspection

Examination of the variables to gain insight into the data and the relations between data reformulate variables to make them compatible and comparable. Data inspection is crucial for a successful outcome of the analysis. The inspection of data is commonly realized by visualization in form of histograms, Quantile-Quantile-plots, PDE-plots

(Ultsch, 2005b) and Box-Plots. If the data are not cleansed and normalized, there is a danger of obtaining spurious and meaningless results. For many similarity measures, e.g. the commonly used Euclidean distance, normalization of data needs to be considered to avoid undesired emphasis of features with large ranges and variances. The application of transformation measurements such as ladder of power is often recommended to take into account restrictions of statistics (Hand et al., 2001). Furthermore the set of variables is usually proofed by correlation coefficients and scatter plots to discover relations or unforeseen dependencies in the set of variables.

Structure visualization

Many methods offer a two dimensional projection with respect to some quality measure. Most commonly principal component analysis (PCA) preserving total variance and multidimensional scaling (MDS) preserving distances as good as possible are used. The output of these methods are merely coordinates in a two dimensional plane. Since there are not clearly separated clusters in a dataset it will be hard to recognize groups for examples. More visualization capabilities than simple low dimensional projections are offered by the Emergent SOM (ESOM). The original high dimensional distances can be visualized with the canonical U-Matrix (Ultsch, 1992). The projection leads to sharpen cluster boundaries. The visualization can be interpreted as height values on top of the usually two dimensional grid of the ESOM, leading to an intuitive landscape. The data space can be displayed in form of topographical maps, intuitively understandable also by users without scientific education. Clearly defined borders between clusters, where large distances in data space are present, are visualized in the form of high mountains. Smaller intra cluster distances or borders of overlapping clusters form smaller hills. Homogeneous regions of data space are placed in flat valleys. Toroid maps should be used to avoid border effects. The U-Map is a non-redundant view of the U-Matrix of such a border-less ESOM (A. Ultsch., 2003b; Ultsch, 2005c; Ultsch/Mörchen, 2005d).

Structure definition

Clustering (i.e. unsupervised classification) is the process of finding intrinsic groups (classes) in a set of data without knowing a priori which data set belongs to which class. Classification is the task of assigning class labels to a data set according to a model where classes are known. Results can suggest a general typology and lead to the development of prediction models using subgroups instead of the total population (amount of objects, e.g. communities). Clustering can be applied to nonspatial variables, spatial variables (e.g. shape), and proximity of the objects or events in space, time and space-time. Each cluster should be as homogeneous as possible and distinct from other clusters. For example a cluster can be defined based on distance (e.g. agglomerative (Ward, 1963), divisive), density (e.g. DBSCAN (Sander et al., 1998)), partitioning (e.g. EM algorithm (Bilmes, 1997)) or grid structure (e.g. STING (Wang et al., 1997)).

Structure control and explanation

The openness of the formation of clusters needs an additional validation and explanation of results. Regression is the task of explicitly modeling variable dependencies to predict a subset of the variables from others (Hastie et al., 2009). Regression can also be used to replace missing values. Discriminant analysis is applicable to determine the class of an observation based on a set of variables. The Explanation of clusters can be also realized for example by a classification and regression tree (Breiman et al., 1984) or by contingency tables. The structure control supports the explanation and description of a classification result.

Operationalization

New objects can be associated to existing classes by classifiers representing a model in the form of rules or decision trees. A classifier is based on learning, testing and validation of data sets. It is expressed in a sub-symbolic or symbolic form whereas a symbolic classifier (e.g. Sig* (Ultsch, 2008)) assists human skills of comprehension.

Knowledge conversion

The most important step is the generation of useful, new and unsuspected knowledge. It is required to be representable in a linguistic form that is understandable to humans and automatically usable by knowledge-based systems. With extracted knowledge it is possible to diagnose unknown examples. Geographic Visualization (MacEachren, 1994) supports the interpretation of results. Geographic Visualization is commonly defined as the integration of cartography, GIS, and scientific visualization to explore geographic data and communicate geographic information to private or public audiences (MacEachren and Kraak, 1997). Spatial analysis provides a synoptic view of observed spatial patterns. Maps are essential for visualizing such patterns. Important tasks are the spatial feature identification, spatial feature comparison and in particular spatial feature interpretation. Identification allows to spot the emergence of spatiotemporal patterns at different levels of spatial aggregation and to explore boundaries between spatial classes. Spatial feature identification and comparison can guide spatial query formulation. Spatial feature interpretation can help to build geographic domain knowledge.

2.2 Dissimilarity Measurements

It is to point out that there are specific forms of dissimilarity. Dissimilarity usually fulfills three criteria (Izenman, 2008):

1. $d(p_i, p_j) \geq 0 \rightarrow$ positivity
2. $d(p_i, p_j) = 0 \Leftrightarrow p_i = p_j \rightarrow$ separate identity
3. $d(p_j, p_i) = d(p_i, p_j) \rightarrow$ symmetry

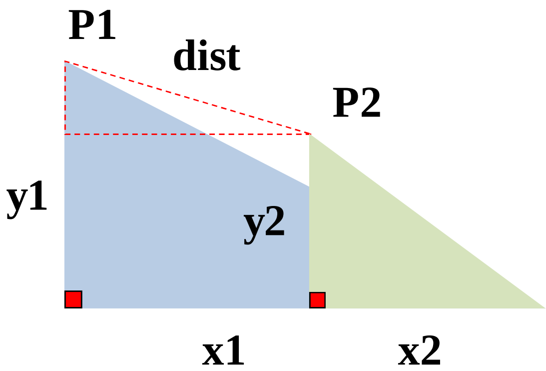
Metric dissimilarity satisfies the fourth property:

4. $d(p_i, p_j) \leq d(p_i, p_k) + d(p_k, p_j) \rightarrow$ triangle inequality

Ultrametric dissimilarities can be displayed graphically (see dendrogram in section 2.3) and satisfies the fifth property:

5. $d(p_i, p_j) \leq max\{d(p_i, p_k), d(p_j, p_k)\}$

Several reasons for the use of the Euclidean distance are later on discussed in the chapter about clustering. Generally Euclidean Distance is one of the most popular distance measurements. The Euclidean distance E is based on the Pythagoras theorem. It corresponds to the geometric distance into the multidimensional feature space and is not limited to any orthogonal dimension.



$$E_{2D} = \sqrt{(x1 - x2)^2 + (y1 - y2)^2}$$

$$E_{3D} = \sqrt{(x1 - x2)^2 + (y1 - y2)^2 + (z1 - z2)^2}$$

...

$$E_{nD}(x_i, x_j) = \sqrt{\sum_n (x_{in} - x_{jn})^2}$$

Figure 3: Sketch of the Euclidean Distance between points in the feature space

2.3 Ward Clustering

Clustering is a well-known example of unsupervised learning and is often used to arrange large quantities of high dimensional data into natural cluster (groups). Hierarchical Clustering is an appropriate approach and is subdivided into agglomerative methods, which proceed by series of fusions of objects into cluster (bottom-up), and divisive methods, which separate a given partition of objects successively into finer partitions (top-down). The later on applied clustering approach is based on a technique that is one of the most used in practice: WARD algorithm (Ward, 1963) is a typical hierarchical (agglomerative) algorithm. Such algorithm seeking to form the partitions $C_c, C_{c-1}, \dots, C_1$.

Ward algorithm begins with a partition that treats each object as its own cluster. At each agglomerative step, the union of every possible partition pair is considered and the two partitions whose fusion results in minimum increase in "information loss" are combined. At each agglomerative step the number of distinct partitions is reduced by 1. Any particular partition c is characterized by a sum of square measure of variation(V).

$$V = \sum_{c=1}^{C} \sum_{n=1}^{3} \left(x_{pnc} - \bar{x}_{nc}\right)^2$$

with x_{pnc} = values of period n of object p (for all objects p$= 1, \dots, P$) in partition c

and $\bar{x}_{nc}$ = mean value or the mean vector (centroid) of period n in partition c.

Each partition is further characterized by a specific value R^2 (=mean squared error) describing the mentioned "information loss". Let $V(1)$ be the value when all objects are aggregated into a single (very heterogeneous) partition. $V(1)$ plays the role of the "total sum of squares". Let $V(C)$ be the value of V for the partition into C partitions.

A proportion of variability $\frac{V(c)}{V(1)}$ is measured for each partition c, R^2 describes the information loss e.g. characteristics of objects that are unexplained by the new partition:

$$mean\ squared\ error = R^2 = \left(1 - \left(\frac{V(c)}{V(1)}\right)\right)^2$$

The total number of objects should be considered to be partitioned into several similar ones. At the start there is no information loss ($V(C)$ is zero and R^2 is 1). At $V(1)$, R^2 is 0. At any intermediate step, R^2 measures the proportion of variability explained by the current partition c. The analysis of R^2 at intermediate agglomerative steps leads to the identification of a decrease in similarity within partitions as the number of partitions drops from C to 1.

The aim of Ward algorithm is to unify cluster such that the variation inside these cluster is not increasing dramatically. In contrast to other agglomerative algorithms (e.g. Single Linkage) this algorithm does not put together cluster with smaller distance, but it joins cluster that do not excessively increase the information loss.

Clusters are compact if all of objects within them are relatively homogeneous together (high similarity) as compared with objects in different clusters (high dissimilarity). However, like variance, mean squared error has a disadvantage of heavily weighting outliers. This is a result of the squaring of each term, which effectively weights large errors more heavily than small ones. Furthermore the algorithm tends to join clusters with a small number of observations, and it is strongly biased toward producing clusters with the same shape and with roughly the same number of observations. Generally it is to emphasize that no provision can be made for a relocation of objects that may have been "incorrectly" partitioned at an early stage. When clustering in this thesis the data is already inspected. The clearly bounded and symmetrical range provides the whole clustering process (see suggestion of Prof. Ultsch in section 5.2). Against this background unforeseen circumstances which may affect the results are minimized.

The clusters are defined by the partition at the point Wards algorithm is stopped. But how to find the point between C and 1 that leads to a clear and distinguishable structure and a good representation of content. All hierarchical clustering methods can be displayed in a dendrogram. Such tree-like diagram can depict the mergers or divisions which have been made at successive level. The dendrogram may be drawn in a horizontal or vertical form. It visualizes the height of the linkage of objects. That means the difference in height defines how close objects are to each other. Objects are similar to each other at low heights whereas objects are more dissimilar are combined higher up the dendrogram. By cutting the dendrogram at an appropriate height a partition of objects into a specified number of groups can be obtained. If a line is drawn on the dendrogram at a given height, then the marked branches of the tree constitute a cluster.

2.4 k-Nearest-Neighbor-Classification

The k-Nearest-Neighbor-Classifier is a sub-symbolic one. That means the classifier does not require any deeper understanding of the class. The k-Nearest-Neighbor classifier supports the labeling by finding a labeled object that is the nearest neighbor of an unlabeled object.

The k-Nearest-Neighbor classifier was firstly introduced to the beginning of the 1950s (Fix / Hodges, 1951 and 1952) as a method of non-parametrical classification. The learning data will be arranged in ascending order in a chosen metric IR^d to a given observation $p \in IR^d$. Thus the following equation is satisfied:

$$\|p - P_{R1n}\| \leq \|p - P_{R2n}\| \leq \cdots \leq \|p - P_{Rnn}\|$$

Whereas $(R_{1:n}, \dots, R_{n:n})$ is defined as a randomly permutation of the tuple $(1, \dots, n)$. It is possible that different points of the learning series have the same distance to p. A k-Nearest-Neighbor classifier determines the class that is most frequently under k-nearest-neighbors of p. It should be mentioned that a classifier with $k = 1$ lead to a construction of a Voronoi diagram.

2.5 Class explanation using Contingency Tables

A contingency table, often referred to as cross-classifications or cross-tabulations, usually shows frequencies for particular combinations of values of two discrete random attributes X and Y. Each cell in the table represents a mutually exclusive combination of X and Y values. Contingency tables contain row attributes across the horizontal axis and column attributes down the vertical. Cell entries give the number of cases (e.g. communities, patterns, or other unit of analysis) that occur in each cell. The cells themselves are formed by combining one category from each of the row and column attributes. Marginal totals (or marginals) give the total number of cases found in each category of the attributes — in other words they are the row and column totals. The mentioned elements are shown in tabular form below (Table 1).

Table 1: Structure of a Contingency Table (absolute / relative frequency)

$(\boldsymbol{k\ x\ m})$ $\boldsymbol{contingency\ table}$

	c_1	...	$\boldsymbol{C_m}$
$\boldsymbol{r_1}$	f_{11}	...	f_{1m}
$\boldsymbol{r_2}$	f_{21}	...	f_{2m}
⋮	⋮		⋮
$\boldsymbol{R_k}$	f_{k1}	...	f_{km}

$\boldsymbol{i = 1, \dots, k}$
$\boldsymbol{j = 1, \dots, m}$
$\boldsymbol{c\ =\ column}$
$\boldsymbol{r\ =\ row}$
$\boldsymbol{f_{ij}\ =\ frequency}$
$\boldsymbol{p_{ij}\ =\ f_{ij}/n_{km}}$

$\boldsymbol{Y(y_m)} \longrightarrow$

		c_1	...	$\boldsymbol{C_m}$	$f_{i.}\ =\ f_{i1} + \ldots + f_{im}$
$\boldsymbol{X(x_k)}$	$\boldsymbol{r_1}$	f_{11}	...	f_{1m}	
	$\boldsymbol{r_2}$	f_{21}	...	f_{2m}	$\boldsymbol{f_{2.}}\ =\ Total\ of\ \boldsymbol{X}$
$\downarrow$	⋮	⋮		⋮	
	$\boldsymbol{R_k}$	f_{k1}	...	f_{km}	
$f_{.j}\ =\ f_{1j} + \ldots + f_{kj}$		$\boldsymbol{f_{.1}}\ =\ Total\ of\ \boldsymbol{Y}$			$\boldsymbol{n_{km} = Total\ of\ all\ Frequencies}$

$\boldsymbol{Y(y_m)} \longrightarrow$

		c_1	...	$\boldsymbol{C_m}$	$\boldsymbol{p_{i.}}\ =\ \sum_{j=1}^{m} p_{ij} = f_{i.}/n_{km}$
$\boldsymbol{X(x_k)}$	$\boldsymbol{r_1}$	p_{11}	...	p_{1m}	
	$\boldsymbol{r_2}$	p_{21}	...	p_{2m}	$\boldsymbol{p_{2.}}\ =\ Total\ of\ \boldsymbol{X}$
$\downarrow$	⋮	⋮		⋮	
	$\boldsymbol{R_k}$	p_{k1}	...	p_{km}	
$\boldsymbol{p_{.j}}\ =\ \sum_{i=1}^{k} p_{ij} = f_{.j}/n_{km}$		$\boldsymbol{p_{.1}}\ =\ Total\ of\ Y$			**1**

Normally, cell entries are expressed as either row or column percentages (depending on the point of analysis).

Table 2: Contingency Table, Frequency as vertical and horizontal Percentages

$Y(y_m) \longrightarrow$

$X(x_k) \downarrow$		c_1	…	C_m
	r_1	v_{11}	…	v_{1m}
	r_2	v_{21}	…	v_{2m}
	⋮	⋮		⋮
	R_k	v_{k1}	…	v_{km}
$Total\ (X \mid Y = r_j)$		**1**	**1**	**1**

$v_{ij} = f_{ij}/f_{.j}$

$i = 1, \ldots, k$

$j = 1, \ldots, m$

$Y(y_m) \longrightarrow$

$h_{ij} = f_{ij}/f_{i.}$

$X(x_k) \downarrow$		c_1	…	C_m	$Total\ (Y \mid X = c_i)$
	r_1	h_{11}	…	h_{1m}	**1**
	r_2	h_{21}	…	h_{2m}	**1**
	⋮	⋮		⋮	
	R_k	h_{k1}	…	h_{km}	**1**

The expected frequency for each cell ($f_{E_{ij}}$) is computed by multiplying the marginal frequencies for the row and column (row and column totals) of the desired cell and then dividing by the total number of observations. The formula of the expected frequency can be represented as follows:

$$f_{E_{ij}} = \frac{(\text{Row Total } * \text{ Column Total})}{\boldsymbol{n_{km}}} = \frac{f_{i.} \ * f_{.j.}}{\boldsymbol{n_{km}}}$$

One important question for the interpretation of contingency tables is as follows: "Is the proportion of observed values significantly higher or lower than would be expected?" Under the assumption of statistical independency of classes and a constant probability it is possible to model the frequency distribution of communities with the binomial inverse cumulative distribution function (CDF). It provides the proof whether the given number of communities of a class i differs significantly from the expected number for class j. The prior probability is already known due to the computation of the expected value $f_{E_{ij}}$. It is assumed that there is a 5 error. Then the binomial inverse CDF provides the identification of limits for the decision about significance. That means the identification of a positive (values are to high) or negative significant (values are to low) number of communities.

A significant result means that the cells of a contingency table should be interpreted. A non-significant test means that no effects were discovered and chance could explain the observed differences in the cells. In this case, an interpretation of the cell frequencies is not useful. The deviation for each cell is computed by the difference of observed and expected values. These values are helpful during the interpretation process of significant cell entries.

Table 3: Deviation for each cell (observed - expected)

$i = 1, \dots, k$ $j = 1, \dots, m$	$\boldsymbol{Y(y_m)} \longrightarrow$ c_1	...	$\boldsymbol{C_m}$	$\boldsymbol{Total}\ (\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{c_i})$
$\boldsymbol{X(x_k)}$ $\boldsymbol{r_1}$	$f_{11} - f_{E_{11}}$	...	$f_{1m} - f_{E_{1m}}$	0
$\boldsymbol{r_2}$	$f_{21} - f_{E_{21}}$	...	$f_{2m} - f_{E_{2m}}$	0
$\downarrow$ ⋮	⋮		⋮	
$\boldsymbol{R_k}$	$f_{k1} - f_{E_{k1}}$	...	$f_{km} - f_{E_{km}}$	0
$\boldsymbol{Total}\ (\boldsymbol{X} \mid \boldsymbol{Y} = \boldsymbol{r_j})$	0		0	Sum of the (Observed – Expected) for both the rows and columns equals zero.

With regard to the publications of section 4.3.2 (Ultsch, 2005) the equation for determining the relative difference, could be also used in case of the presentation of the difference of expected and observed values:

$$RelDiff\left(f_{ij}, f_{E_{ij}}\right) = \frac{difference}{mean} = \frac{f_{ij} - f_{E_{ij}}}{\frac{1}{2}\left(f_{E_{ij}} + f_{ij}\right)} = \frac{f_{ij} - f_{E_{ij}}}{f_{E_{ij}} + f_{ij}} * 2$$

The value of relative difference is adjusted by multiplying by 100% to reduce rounding errors. The values of $RelDiff\left(f_{ij}, f_{E_{ij}}\right)$ is displayed in each contingency tables of the thesis and supports the interpretation.

Another procedure should be mentioned that is often used to test the significance of contingency tables. It is called the chi-square statistic. The Chi-squared statistic is based on the postulate of empirical independence. This test assumes a sample with a sufficiently large size. If a chi square test is applied on a sample with a smaller size, then the chi square test will yield an inaccurate inference. Using the chi-square for each cell the observed frequency is compared with the expected frequency $f_{E_{ij}}$.

Chi-square statistic is represented by the formula below:

$$x^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\left(f_{ij} - f_{E_{ij}}\right)^2}{f_{E_{ij}}}$$

In general, the larger the difference between the observed and expected values, the greater is x^2. The chi-square test of significance is also useful as a tool to determine whether or not it is worth the researcher's effort to interpret a contingency table.

For this purpose the degree of freedom is computed by multiplying one minus the number of rows, times one minus the number of columns:

$$degree\ of\ freedom = df\ = (Row\ Total\ -\ 1\ )\ \cdot\ (Column\ Total\ - 1)$$

To provide the interpretation it is therefore necessary to compute the contingence coefficient K as presented by the formula below:

$$K = \sqrt{\frac{x^2}{n + x^2}}$$

The contingence coefficient K is defined in the range $[0, K_{max}]$. The value K_{max} is the upper limit and is computed as a function based on the table dimension. That means related to the number of columns and rows. Based on the formula below M is equal to the smaller value of k or m.

$$K_{max} = \sqrt{\frac{M-1}{M}}\quad with\ \ M = min\{k, m\}$$

The contingence coefficient K is normalized to $K^* \in [0,1]$. Thus the coefficient is not any more dependent of the table dimension.

$$K^* = \frac{K}{K_{max}}$$

Using K^* it is possible to decide about the relation of two different categories. Values of K^* near by zero indicate that there is probably no relation of both values. Values of K^* near by 1 indicate that there is probably a clear relation.

3 Related Work – City Classification, Analysis of Population

3.1 City Classification and Urban Portraits

There are several trials to build up content-based classifications in the focus of research about geospatial objects (e.g. buildings, building stocks, cities and regions) and their similarities. For example, Harris started in 1943 (Harris, 1943). He was a pioneer in city classification and ranked US cities according to industrial specialization data. Later on in the 1970s studies were geared to measure socio-economic properties and shifted more towards the goals of public policy. In recent years the evaluation of the performance of different cities is becoming increasingly important for sustainable development (Arlt et al., 2001). The patterns of demographic and economic changes in Germany are also part of several investigations (Siedentop et al., 2003; Gatzweiler et al., 2003). Critical properties of geospatial objects are discussed and analyzed by Demsar (2006). Methods of data mining are applied to analyze Swedish communities.

The demographical analysis of the geographical Institute of the University of Bern have shown that the level of communities is essential to identify growing or declining areas in countries of the alpine region (Bätzing, 1993 and Bätzing et al, 1996). Thus it is also well known that the higher level of NUTS-3 (Nomenclature des unités territoriales statistiques: Bezirke, departments, provincies) is not appropriate to detect spatial disparities (as a negative example see ABIS 1999).

Another study on the level of communities was interested in the classification of characteristic agrarian structure regions in the Alps (Tappeiner et al., 2003). The clustering process deals with 43 variables (30 static and 13 dynamic) and 5.558 alpine communities. Such study has demonstrated the risk to fail when handling such a large amount of variables. The important step of data inspection including the investigation of data distributions is not discussed and presented in detail. The study applies the CLARA algorithm, the k-means algorithm and the k-medoid algorithm. The authors claim that the k-mediod algorithm has proper characteristics not to overweight extreme objects and is presented by the authors as a suitable method to classify communities. But the authors do not take into account that it might be better to form clusters with just a subset of relevant variables. Furthermore a cluster explanation with other variables of the 43 variables is not considered. The results are presented in just one map and the integration of spatial analysis or GIS in general is sparsely presented.

In recent years another study has tried to define a consistent definition of the mountain region in Europe based on GIS and several non-spatial community indicators including demographical data (Hill et al, 2004). Such project was further interested to harmonize statistical data for deeper combined analysis. About 115.000 communities in Europe are therefore characterized and it is obvious that the demand of a multidimensional analysis (e.g. Geographical Data Mining, Knowledge Discovery) will be of rising interest.
It is to emphasize that several former classification studies are calculated by hierarchical clustering algorithms (e.g. WARD, k-means). Especially in the field of urban and spatial planning as well as regional science, data are usually multidimensional, spatially correlated and heterogeneous. These properties make some of the former approaches often inappropriate for the data, as their basic assumptions cease to be valid (e.g. identically distributed). For example a great deal of variability in the range and distribution of variables may pose a problem for cluster algorithms which involves distance measurements. Furthermore, several cluster algorithms are limited to find clusters of specific shape (e.g. spherical, ellipsoid). Extracting knowledge from geospatial data requires therefore an intensive data understanding and inspection as it helps the researcher to become familiar with the nature of the data.
In context of the Swiss spatial organization and demography the interval 1850 to 2000 is already investigated on national scale and in parts on the level of communities (e.g. Tschop et al., 2002). Some general statements to population and settlement patterns are given. However the characterization is briefly and the localization of patterns on the level of communities is missing. A scatter plot allows the comparison of two time points (1920 and 1990). The concentration of population is displayed in a Lorenz curve. The relative development of population is displayed along transit axes (e.g. west-east). Thereby it is possible to start a visual comparison of districts. It should be mentioned that many other studies of spatial properties (e.g. urbanization) are more descriptive and metaphorically. Classification of communities and GIS are not in the scope of interest. They often refer to urban sprawl or uncontrolled development resulting in concepts like the “Zwischenstadt” (Sieverts, 1999). Swiss approaches centered on the Netzstadt concept related the urban and regional development to the urban metabolism (Baccini, Oswald 2003). Another famous study addresses a qualitative portrait of the Swiss urban conditions (Diener et al., 2005). There is a great demand for the classification of communities. To tap the full potential of spatial interpretation and analysis the author of this thesis would suggest spatial analysis, Data Mining and Knowledge Discovery.

3.2 Clustering of Swiss Communities concerning the development of Population in a long-term Perspective

Many former demographic studies have been made by using cross-section analysis (=comparison of two time points). In this section one study should be discussed in a deeper way. The here presented study was the starting point and motivation to tackle such a spatiotemporal clustering problem in this master thesis.

The study is dealing with a longitudinal section analysis (=long-term analysis) of the population development in European alpine communities (Dickhörner, 2000, Bätzing/Dickhörner, 2001). The target was the typing of all alpine communities concerning the development of their population between 1870 and 1990. Basis for that was the absolute population of communities in 1870, 1950, 1960, 1970, 1980 and 1990 from the Alpine Database (developed by Werner Bätzing and Manfred Perlik). A special variable is the first time interval from 1870 to 1950, which leads over 80 years. This big step summarizes the changing from an agrarian society in 1870 to an industrial society in 1950. At the same time the demographic results of both world wars have been smoothed. This time interval can be recognized as the beginning, which makes it easier to assess the development after 1950, which is of greater interest.

To avoid weighting, the relative growth rate per year R_I for the time interval $I\ (t_1\ ,\ t_2)$ has been used. The authors argue that the five metric variables of a dedicated number have the same quantitative meaning on all (harmonized) feature scales. By using of the variable ‚relative growth rate per year' the attribute's space is by the opinion of the authors extensively standardized. So in their opinion an important condition for the „regional taxonomic method" and also „all other multivariate numeric methods" is fulfilled, which depends on „direct numeric comparisons".

The mentioned asymmetric range of values $R_I \in [-100, \infty[$ has a critical influence in particular on the calculation of distances and similarity patterns. Normalizing by the empirical variance is a big problem, when values are not normally distributed and the results are mainly influenced by extreme cases. Therefore in the opinion of the authors a Gaussian transformation for example should be carried out with these values. From their point of view this would lead to results which are difficult to be interpreted and would be less understandable because of more abstracted values. In the present case less data inspection and improvement is not worth to be legitimate from their point of view.

The authors discuss the linear independence of variables as another requirement of cluster analysis. That means that the same information will not be used more than once with different variables in an analysis. The variables have been proved for correlation and only one minor significant correlation has been found.

In the work of the authors the Euclidian distance is used in spite of a measure for dissimilarity. All communities with a relative growth rate over +10%/a respectively below –10%/a in a specified interval have been eliminated during the clustering process. Ward algorithm is used for clustering and results are later on optimized by the non-hierarchical Quick-Cluster-Analysis. This procedure was realized for outlier treatment and general optimization. The communities are partitioned into 17 clusters. 11 of them are characterized by different growing processes. These clusters consist of 51.2% of the alpine communities with 75.1% of the alpine population in 1990 (51.3% in 1870) living on 48.8% of the alpine area. One big cluster is characterized by stagnation (17.4% of communities, 16.5% of alpine population in 1990 and 23.4% in 1870, 23.8% of alpine area). Five clusters are characterized by different declining processes (31.4% of alpine communities, 8.4% of alpine population in 1990, 25.2% in 1870 and 27.4% of the alpine area).

Besides the changing of population further variables are added in the already mentioned Alpine Database. To detect relations between them and the development of community population data relations have been proved. The distribution of all alpine communities (total amount) to the seventeen clusters served as a reference. A correlation analysis was done with the following variables: nationality, geographic height, community area, population, and urbanization zones. In contrast to this cluster explanation the author of this thesis suggests to use of contingency tables.

The longitudinal study concludes that the distribution of Swiss Alpine communities to the clusters is mainly characteristic in accordance to the Alpine average. So in the Swiss Alps not only all development types are present, but also their quantitative weighting is comparable near the alpine average. Therefore analyses in this area have a high grade of representativeness for the whole Alp area. This study has tried to offer an insight to the demographic development in European Alpine regions. But this study was mainly interested in the last five decades.

Against this background the here presented thesis aims to examine the population change by 15 decades (1850 to 2000). The author is mainly interested in a consistent approach.

4 Population Data – Data Inspection and Modeling

4.1 Basic Spatial Unit (BSU)

The analysis of population development requires initially the selection of an appropriate administrative level (i.e. community) or rather the definition of a comparable spatial unit of interest. In view of growth and decline it is very important to analyze such processes on an appropriate administrative level or scale. Furthermore the selection depends on the availability and amount of official statistics. The basic spatial unit (BSU) denotes the smallest type of areal unit for which data are available. Traditionally in Switzerland most of the statistical data focus on the cantonal or community level. In particular census data is typically established on the level of communities. Against this background statistics of population are often aggregated and published on this level. In Switzerland data is actually available in the time frame 1850 to 2000. Such valuable and long range data (population per community) has a great influence on the final selection of spatial units.

It is to point out that the actual system of Swiss communities derived from different historical events and institutional decisions and factors (Meyer, 1978). Thus the size of communities ranges from 31 to 28221 ha and the number of population ranges from 22 to 363273. The specific interpretation of one community demands a critical handling with respect to local properties such as growing settlement areas in the valleys and several shrinking settlement areas on the hillside. For example such community characteristics are to find in Ticino or Valais. It is to remark that the analysis of communities is already influenced by generalization processes but it is also to state that this official statistical level is the only lower one with countrywide significance. At least it is possible to observe regional linkages and characteristics as well as suburban specifics. At present there are no other alternatives allowing a deeper spatial investigation of the long-term development of Swiss population. This thesis is therefore based on the level of communities as a spatial reference system.

For the future it should be necessary to get access to other official data and to extend the scope of spatial interest (Manley et al., 2006). For example population data needs some spatial disaggregation to reach the level of individual houses or precisely defined and comparable settlement areas (official geo data). Such approaches lead to highly resolute data (Thinh, 2004, p. 52 ff.). In the future it seems to be promising to compare results of different spatial levels (e.g. settlement area vs. community vs. canton).

4.2 Database

The database of this thesis is characterized by data of official statistics and in particular of the Swiss Federal Population Census (German: Eidgenössische Volkszählung, French: Recensement fédéral de la population, Italian: Censimento federale della popolazione). The geometry of all 2.896 communities is available in scale of 1:2.000.000. Such generalized geo-data (GEOSTAT) is just as well established by the Federal Office of Statistics. All variables computed results are stored in a relational database that is joined to the matched spatial units in a GIS.

The population of a spatial unit is typically defined (Bähr, 2004) as a certain amount that consists of different distinguishable elements (e.g. individuals, persons) belonging to it permanently (e.g. legally established resident or de jure inhabitants.). Population covers both nationals and aliens, native and foreign-born persons, intemees, refugees and any other group physically present within the borders of a country at a specified time. A population census comprises the total process of collecting, compiling, evaluating, analyzing and publishing or otherwise disseminating demographic, economic and social data pertaining, at a specified time, to all persons in a country or in a well delimited part of a country. The Swiss Federal Population Census has been realized every 10 years starting in 1850 (Schuler, 2002). Exceptions of the regular cycle are to mention for the years 1888 and 1941. In 1980 there is one community “Vellerat” hindering the census (Missing Census Data: “NaN”, undisclosed 69 persons in 1980). However, the Swiss Federal Population Census follows criteria of a common modern census (Witthauer, 1969): completeness (≠double count or omissions), concurrence (reference date: e.g. 05.12.2000), individuals (≠groups or families), delimitation of areas (Swiss federal territory), scientific review (published, reported) and periodicity (10 years, by decade). In 2000, statistical data was optimized and temporally harmonized in a comparable way by the Federal Statistical Office. Therefore it is possible to analyze Swiss communities in a long-term perspective – this means that population data by decade is available since 1850. The harmonization process has followed modifications of the territorial area (e.g. community fusion, community separation, line of the border), and characteristics of political and statistical definitions (e.g. factual and legal position). Collected data include population data (citizenship, place of residence, place of birth, position in household, number of children etc.), household data (number of individuals living in the household, etc.), accommodation data (surface area, amount of rent paid, etc.) and building data (geo-coordinates, time of construction, number of floors, etc.).

4.3 Population Change by Decade

Actually several indexes are in use to measure the change in population (Schwarz, 1970; Woods, 1979; Bähr, 2004). Typical indexes aim to compare two values of different temporal states without knowing everything in-between. Changes in populations (P) are measured to describe either a gain or a loss between different temporal states (t and $t-1$). The change found is an increase if the new value $P_t > previous\ value\ P_{t-1}$; the change found is a decrease if new value $P_t < previous\ value\ P_{t-1}$.

The absolute quantity is a measure of the absolute occurrence of the population. The relative quantity is a measure of the absolute occurrence of population in relation to a specific output quantity.

In view of the Swiss Federal Population Census it is possible to realize a community comparison on the basis of 15 variables representing a specific change by decade (e.g. 1850 to 1860, 1860 to 1870 etc.). In context of such a classification task examining relative quantities is usually more informative than absolute quantities (Vogel, 1975). The index relative percent change (RPC) of population is well known in the field of geography (Husa and Wohlschlägl, 2007). Relative percent change allows a comparison of communities with a different absolute population (e.g. small/large communities). However, there are some crucial properties in context of classification discussed by Ultsch (2009) that will be similarly presented in the following sub-sections. As an alternative to relative change calculation, relative differences (RelDiff) are therefore also proposed. This index results from the research in the field of Data Mining and Knowledge Discovery (Ultsch, 2003 a). It is already demonstrated as a superior index for classification and in particular for the search of similarity patterns.

4.3.1 Relative Percent Change (RPC)

The relative change (RC) of population is defined as the absolute change divided by the previous value. Absolute change means the difference of population from a previous value (P_{t-1}, $P_{t-1} \geq 0$) to a new value (P_t, $P_t \geq 0$).

$$RC(P_t, P_{t-1}) = \frac{absolute\ change}{previous\ value} = \frac{new\ value - previous\ value}{previous\ value} = \left(\frac{P_t - P_{t-1}}{P_{t-1}}\right)$$

The relative change (RC) of population can be expressed as percent change (RPC).

$$RPC(P_t, P_{t-1}) = RC * 100\% = \left(\frac{P_t - P_{t-1}}{P_{t-1}}\right) * 100\%$$

In a theoretical sense the percent change is problematic for very small values of the previous population P_{t-1} ($P_{t-1} \cong 0$).

Another special case will appear if the community population is lost completely within 10 years ($P_t = 0$).

$$RPC(P_t, P_{t-1}) = \left(\frac{0 - P_{t-1}}{P_{t-1}}\right) * 100 = -1 * 100 = -100\ \%$$

The range of values is therefore bounded to the bottom. The minimum of percent change is -100 %. Within ten years the population of one community might increase tenfold or more. That is why there is no upper bound (infinite endpoint) and the range is asymmetric: $RPC(P_t, P_{t-1}) \in [-100, \infty[$. It is expected that the majority of Swiss communities seeks to show values between -50 % and $+100$ %.

Limits: The definition problem of the denominator $P_{t-1} \to 0$ (previous value) and the mentioned assymetric range of values ($RPC_t \in [-100, \infty[$) have a critical influence in particular on the calculation of distances and similarity patterns. Typical measures are Euclidian distances, correlation measures, Mahalanobis distances and others. Since the population data of several communities needs to be compared, the variance of the data have to be taken into account. To normalize by the empirical variance is a big problem, when values are not normally distributed and the results are mainly influenced by extreme cases. Typically comprehensive transformations (Mathematical functions) are necessary to achieve normality or symmetry in data distributions. But this procedure is sophisticated and statistical and visual tests are prerequisite to obtain reliable models.

4.3.2 Relative Difference (RelDiff)

Generally percent difference is the numerical interpretation of comparing two values with one another (Abramowitz and Stegun, 1972). It is often used as a quantitative index of quality control for repeated measurements where the outcome is expected to be the same. Percent difference is similar to percent error, which is applied when one determines an experimental value and is comparing it to the accepted or actual value.

The here presented index relative difference (RelDiff) results from the research of the Databionics Research Laboratory, Department of Computer Science, University of Marburg. Applications are already to find in context of DNA array experiments (Ultsch, 2005a) or return measurements in stock investments (Ultsch, 2008).

With regard to the mentioned above publications the equation for determining the relative population difference, could be similarly defined by decade as follows:

$$RelDiff(P_t, P_{t-1}) = \frac{difference}{mean(decade)} = \frac{P_t - P_{t-1}}{\frac{1}{2}(P_{t-1} + P_t)} = \frac{P_t - P_{t-1}}{P_{t-1} + P_t} * 2$$

The equation is based on the understanding of relative change (RC) as a general semantic. The ratio is now adapted using the mean of both values in the denominator.

$$RC = \frac{absolute\ change}{previous\ value} = \left(\frac{P_t - P_{t-1}}{P_{t-1}}\right) \Rightarrow RPD = \frac{absolute\ change}{mean(decade)} = \frac{P_t - P_{t-1}}{\frac{1}{2}(P_{t-1} + P_t)}$$

Relative population difference is used where a new value P_t and a previous value P_{t-1} correspond to the idea of change found. Both values must contain the same units in order to be compared correctly with one another. The numerator is similarly defined as the difference of population (new value P_t and a previous value P_{t-1}). The denominator represents a so called basic population (universal set per decade) that is compareable in a wider sense to the previous equation of relative change (RC).

The general requirement for selecting two values to be compared is that the user of this technique expects the two values (assuming both $P_{t-1} \geq 0$ and $P_t \geq 0$ are positive) to be numerically equivalent. A relative difference of 0% explains that the two values are exactly the same. In this case the population seeks to be stable within one decade. The relative difference is greater 0 (less then 0) if the new value of population P_t is greater then the previous value P_{t-1} (P_t is smaller than P_{t-1}).

A total loss of population is uncritical for the definition of relative difference:

$$RellDiff(P_t, P_{t-1}) = \frac{0 - P_{t-1}}{P_{t-1} + 0} * 2 = -2$$

Zero values of population do not lead to undefined values. Extreme values are generally fixed to ± 2. In a theoretical sense the upper bound represents an exorbitant population increase. The range of relative difference is bounded and symmetric ($[-2{,}2]$).

Finally the value of relative difference is adjusted by multiplying by 100% to reduce rounding errors:

$$RelDiff(P_t, P_{t-1}) = \frac{difference}{mean(decade)} * 2 * 100 = \frac{P_t - P_{t-1}}{P_{t-1} + P_t} * 200$$

Advantages: Generally a great deal of variability in the range and distribution may pose a problem for cluster algorithms which involve distance measurements. For example atypical scores in a distribution (outlier) can wildly determine the Euclidian Distance. It is obvious that the influence of outlier (extrem values) on relative percent difference is clearly alliviated due to the symmetric and limited range. Furthermore the numerical stability for relative percent difference is much better than for relative percent change. Relative percent difference is particularly suitable for normalisation and standardization. Typically those measurements require statistical operations to be applied to each individual value using the global parameter of each variable such as its minimum value, mean or variance. The above mentioned properties of relative difference are here clearly superior to the common relative percent change. Finally the values of relative difference ensure a direct interpretation.

4.4 Description of the Expected Distribution

At the beginning of the long-term analysis it is assumed that the population has been increased in all 2896 communities since the initial date of 1850. However it seems to be necessary to gain insight into the population change by decade. For this purpose expected characteristics of population change by decade are summarized:

I. **The first assumption deals with the state of population change by decade.**
   Three states of population change by decade are expected: "Loser", "Typical" and "Winner". Such typical communities represent the general trend of population change in Switzerland within 10 years. The mean μ of the Swiss population change by decade is described by typical communities. The mean value of population change is a first indicator to formulate hypothesis about the general situation in Switzerland. The two other groups ("Loser", "Winner") are characterized by a development that is above or below the detected trend by decade. The "Winner" group of communities is often represented by large positive values. The "Loser" group of communities is often characterized by negative large values. But usually it is to consider that these two developments are just better or rather worse than the typical development. General characteristics are sketched below:

"Loser"	**"Typical"**	**"Winner"**
$<$	$- \ldots 0 \ldots \emptyset \ldots +$	$>$
$below$	$Trend\ in\ Switzerland$	$above$

II. **The second assumption deals with the size of community classes.**

The major group of communities is formed by typical communities. In view of several decades the minor group "Winner" is sometimes bigger than the group "Loser" and vice versa. The mentioned characteristics of all groups are illustrated as follows:

"Loser"	**"Typical"**	**"Winner"**
$minority$	$majority$	$minority$

III. **The third assumption deals with the distribution of population change**

a) *Which distribution represents the "typical" state of population change?*

Since communities are characterized by a **normal distribution** it is to state that the sum of many unobserved random population members is acting independently of one another. The central limit theorem (CLT) offers general conditions under which the mean of a sufficiently large number of independent random variables (in this case "communities with population"), each with finite mean and variance will be approximately normally distributed. The location x of the peak of the distribution represents the mean μ of the Swiss population change by decade (see Figure 4: 15 points in time since 1850 $(d(t), d(t-1))$. For normal distributed data, the interval $\mu \pm \sigma$ covers a probability of 68.3%, while $\mu \pm 2\sigma$ covers 95.5%.

b) *Which distribution represents the "winner" state of population change?*

Since communities are characterized by a **log-normal distribution** it is to state that the product of many unobserved random population member is acting independently of one another. Communities with an increase of population follow a skewed distribution to the left characterized by just a few communities with very large positive values. This asymmetric distribution is probably described by a median smaller than the mean value, large variance and clearly positive values. A major difference to the normal distributed data of typical communities is that the effect of population is multiplicative. The force of multiplication is represented by booming processes (growth, concentration). All communities play major roles in a developing situation whereas processes are intensified by adjacencies and other spatial conditions.

c) *Which distribution represents the "loser" state of population change?*

Since communities are characterized by a **log-normal distribution** it is to state that the product of many unobserved random population member is acting independently of one another. Communities with a decreasing population follow a skewed distribution to the right characterized by just a few communities with very large negative values. This asymmetric distribution is probably described by a median bigger than the mean value, large variance and clear negative values. The effect of population is multiplicative. The force of multiplication is represented by intensive declining processes (concentration, abandoned areas). All communities play major roles in a negative situation whereas processes are intensified by adjacencies and other spatial conditions. Due to the mentioned properties data of decreasing communities might be usually log-normal distributed.

"Loser"	**"Typical"**	**"Winner"**
$lognormal$	$normal$	$lognormal$

4.5 Mixture Models

The assumed distribution of population change is a composite distribution. A mixture model will be elaborated which is based on such a composite (log-normal,normal,log-normal) distribution.

Mixture modelling has the advantage of being able to model distributions of continuous variables. Such process is based on density estimation of the data (Ultsch, 2005). The probability density function is thought of as the density according to which the data (population change by decade) is generally distributed. The mixture density is a probability density function which is expressed as a convex combination of several probability density functions. A convex combination of probability distributions is a weighted sum of its component probability density functions $P(X|C)$, with probability density function:

$$P(X) = \sum_{C=1}^{3} P(C) \cdot P(X|C)$$

$0 \leq P(C) \leq 1,$ where $P(C=1) + P(C=2) + P(C=3) = 1$

$C = 1 = "Loser";\ C = 2 = "Typical";\ C = 3 = "Winner"$

The components are described by either normal or log-normal probability distributions. These are two-parameter distributions (Crow and Shimizu, 1988, p. 1) and without loss of generality the parameters can be taken to the mean M and standard deviation S.

Let $M(C|X)$ represent the mean and $S(C|X)$ the standard deviation of the component density function $P(C|X)$. The objective of fitting the mixture to the data is to estimate the following parameter: frequency $P(C = Loser)$, $P(C = Typical)$, $P(C = Winner)$; mean M: $M(C = Loser)$, $M(C = Typical)$, $M(C = Winner)$ and standard deviation S: $S(C = Loser)$, $S(C = Typical)$, $S(C = Winner)$.

The Expectation Maximization (EM-) algorithm proposed by Dempster, Laird, and Rubin (1977), is appropriate for solving parameter estimation problems for a Mixture of distributions (Bilmes 1997). When there is a need to learn parameters of a mixture, the EM algorithm starts with initial values for all parameters and they are re-estimated iteratively. It is crucial to start with 'good' initial parameters as the algorithm only finds a local, and not a global optimum. Therefore the solution (to where the algorithm converges) strongly depends on the initial parameters and needs several recalculations.

The Mixtures are verified by Pareto Density Estimation (PDE) and probability density functions (PDF). Density estimation using hyperspheres with a global radius are a simple and efficient way to estimate data density (Ultsch, 2005b). Quantile-Quantile (Q-Q) plots are used as a graphical technique to compare the distribution of population change to a theoretical model. It is a plot of the quantiles of the first data set against the quantiles of the second data set. If the distributions are similar, the points in the Q-Q plot will approximately lie on a line. Q-Q plots are used to compare the shapes of distributions, providing a graphical view of how properties such as location, scale, and skewness are similar or different in the two distributions.

4.6 Modeling, Standardization and Test of Distributions

The modelling process takes into account that different mixture distributions by decade need to be comparable. The idea is therefore to use a two-stage modelling process regarding a standardization procedure.

Initially the aim is to model the distribution of typical communities as a Gaussian (normal distribution). The detection of the mean μ (Figure 4) is here meaningful to characterize the typical Swiss population change by decade (15 points in time since 1850, $d(t), d(t-1)$). For example in 1950 it is 5 % by decade.

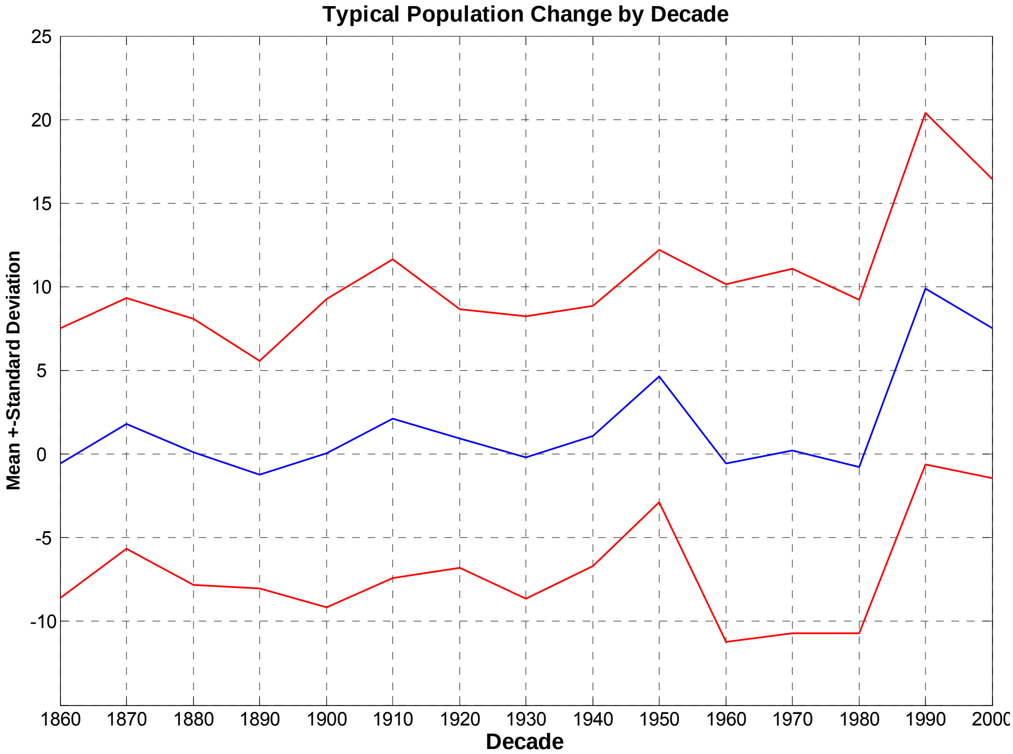


Figure 4: Typical Communities: Mean of the population change by decade

The typical population change is a kind of "clinical thermometer" over time. Table 4 gives a rough overview of the history in Switzerland (1850-2000). Spatial and functional linkages between demographic and urbanization (urban/rural) processes are of interest for the explanation. This relation is crucial to control spatial developments. However a precise examination of such communities is necessary.

Table 4: Short Timeline of Switzerland's History (1850-2000)

1848	Federal State	The Principles of the constitution are still valid today.
1863	Thomas Cook	organizes tours (all included) to Switzerland: start of mass tourism
1882/83	Emigration	13,500 persons leave Switzerland: USA,83%; Argentina, 11%; Canada,4%; Brasil, 2%
1898	State Railways	Companies parliament and electorate decide to nationalize the major railway lines
1914 - 1918	World War I.	Armed neutrality works when surrounded by war faring nations.
1918 - 1933	Economic Crisis	Inner conflicts, general strike and world economic crisis hit this industrialized country.
1933 - 1939	Spiritual Defense	Hitler in Germany is soon seen as a danger to Switzerland's independence. Thousands of German refugees (jews, intellectuals) are accepted.
1939 - 1945	World War II.	Neutral Switzerland surrounded by fascist troops or collaborating regimes.
Since 1945	Prosperity	Recent history is characterized by political stability, economic progress, increased social security and a new openness and tolerance.
1973-1983	Oil Crisis	Affected by the hike of oil prices which resulted in a decrease of energy consumption
1950-2000	Suburbanization	A first period of fast growing town centres is followed by suburban growth dispersion

The typical population change is characterized by a normal distribution (Figure 1). PDE and Q-Q-Plot verify the assumptions of data distribution (see example of population change by decade in 1900).

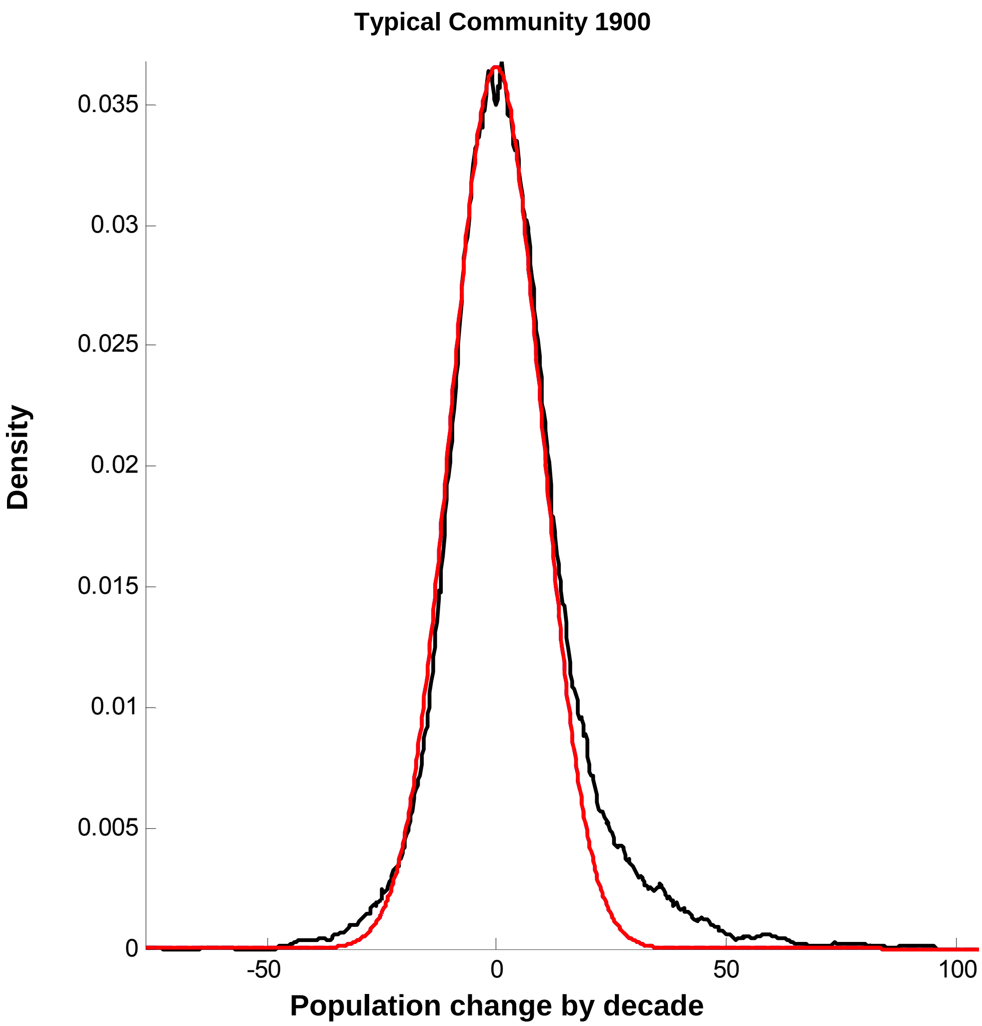


Figure 5: Proof of Data Distribution using Gaussian Model and Q-Q-Plot

Gaussian= red line, PDE= fine line

Variable: $RelDiff$ 1900 ($t = 1900$, $t-1 = 1890$)

Generally z-transformation is a famous measure to standardize or auto-scale the statistical data. Mean and standard deviation of typical population change are used as a precondition for the standardization of the whole mixture.

$$z_i = \frac{X_i - M}{S}$$

z_i = z-transformed values and X_i = original values of population change by decade

M = Mean and S = standard deviation of the typical population change by decade

Computed z-scores become comparable by measuring the observations in multiples of the standard deviation. If the original distribution is a normal one, the z-transformed data belongs to a standard normal distribution. The mean of z-transformed data is always zero.

The second modelling stage deals with the whole mixture model (Log-Normal-Normal-Log-Normal). Each component of the standardized mixture model is described precisely (e.g. Figure 6). Mixture modeling is used for density estimation as a reasonable approximation of population change by decade. In particular the elaborated mixture model represents a generic model, combining three categories (“Loser”, “Typical”, “Winner”) by decade. Each subpopulation has its own characteristic parameters: Mean, Standard Deviation and Weight. Since the original distribution of typical communities is a normal one, the z-transformed data belongs to a standard normal distribution ($M = 0; S = 1$). The mean and standard deviation of typical communities (z-transformed data of the normal distribution) are therefore used to control the modelling process of the whole mixture model. The Q-Q-Plot confirms the assumed distribution. Figure 7 shows the points of compared distributions (quantiles of the first and second data set) that lie on a clear line.

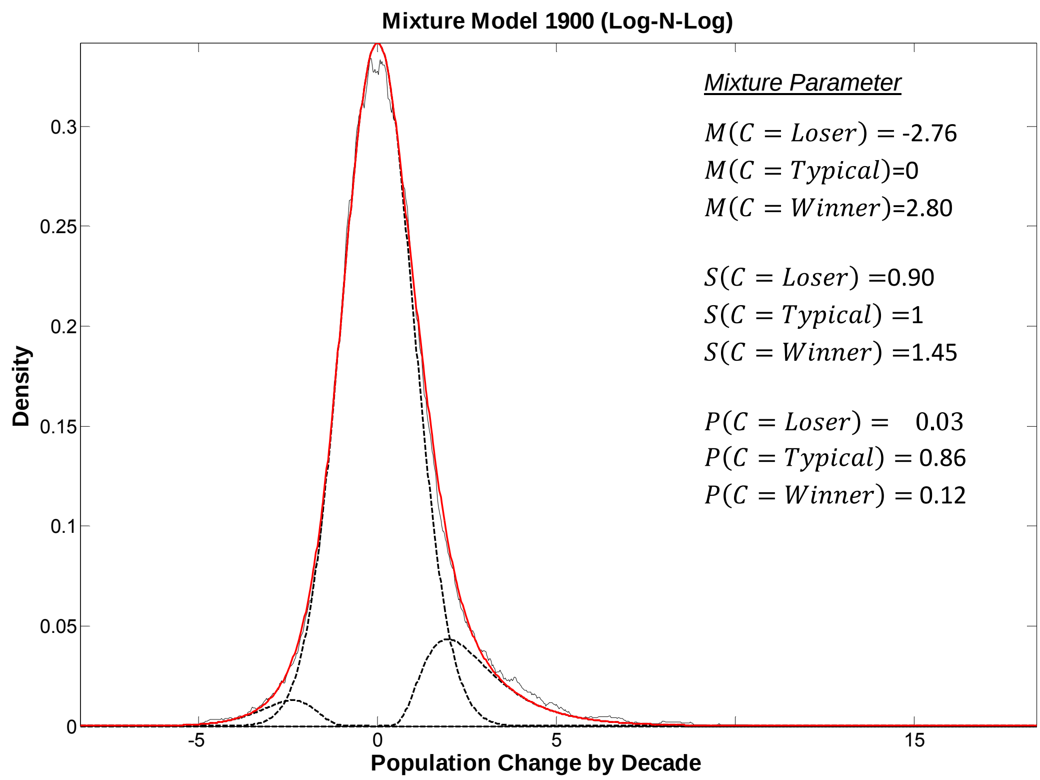


Figure 6: Mixture Model (Log-Normal-Log) of population change by decade
Mixture= red line, Component Density Function= flat line, PDE= fine line
Variable: $Population\ Change\ by\ Decade$ ($t = 1900, t - 1 = 1890$)

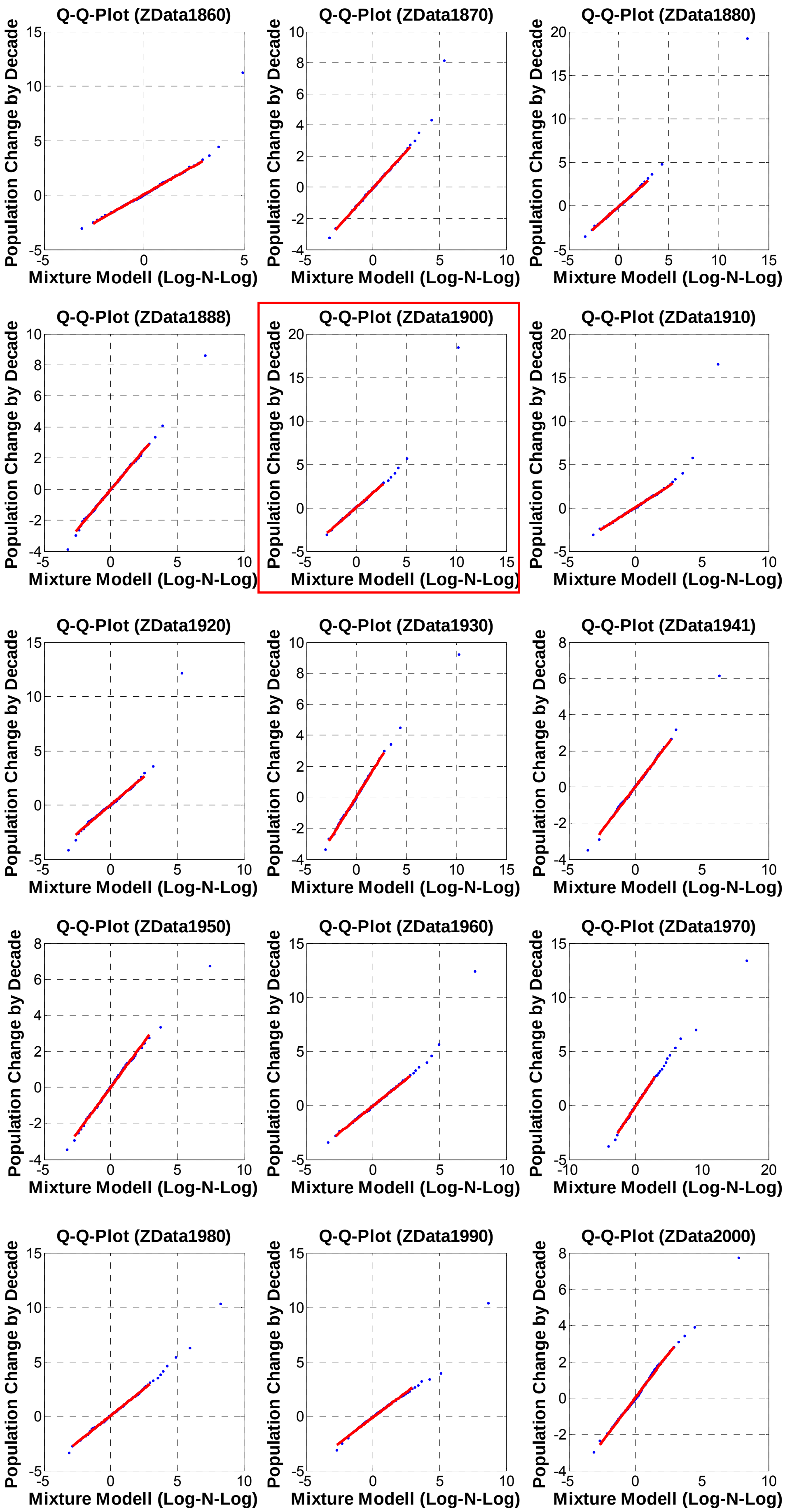


Figure 7: Q-Q-Plot (population change by decade vs. Mixture Model), 1860-2000

5 Patterns in Context of Population Change

5.1 Posterior Probabilities using Bayes' Theorem

Based on the value of population change by decade and the elaborated category C ("loser", "typical", "winner") it is possible to calculate posterior probabilities defining the degree of membership to a specific category. Each category is already characterized by a specific prior probability $P(C)$ and a component probability density function $P(C|X)$. The Bayes' theorem supports the computation of the posterior probability that a specific category C is observed for the given feature X (Han/Kamber, 2006, p. 311):

$$P(X|C) = \frac{P(C) \cdot P(C|X)}{P(X)}$$

The posterior probability of X occurring given that category C has occurred is equal to the prior probability of category C occurring, multiplied by the probability of component $P(C|X)$ and divided by the probability of $P(X)$. This famous ratio poses the first definition of the conditional probability in the 18th century and was created by Thomas Bayes (1702-1761). Bayes' theorem is used to compute the posterior probabilities (see Figure 8). Each community is described by three values related to the three categories of the mixture model. A specific category C ($posterior \geq 0.5$) is predominant observed for a given value of population change.

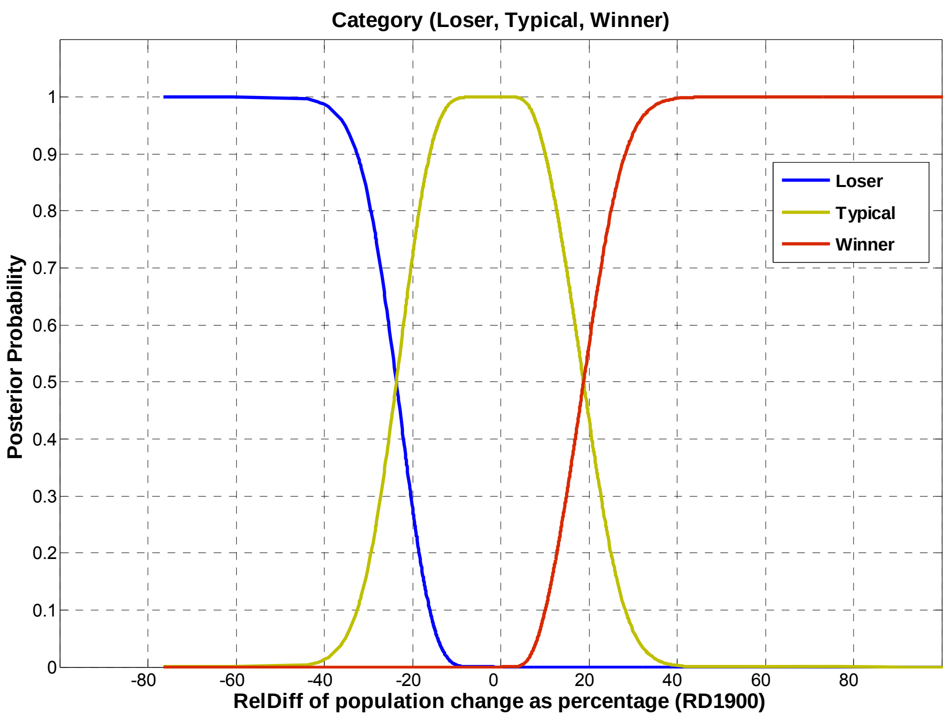


Figure 8: Posterior probabilities based on population change and categories

5.2 Classification using categorical memberships by decade

In mathematics any sequence of numbers that may be modeled by a mathematical function is considered as a pattern. Hardy (1992, p.84) is quoted as follows: “A mathematician, like a painter or a poet, is a maker of patterns. If his patterns are more permanent than theirs, it is because they are made with ideas”.

For the purpose of classification a specific function was suggested by Prof. Ultsch. His advice ensures the classification procedure in an efficient and appropriate way. This procedure is generally based on Bayes’ decision boundaries and the definition of classes. Based on the above mentioned posterior probabilities and category C LTW is described as follows:

$LTW : [-2,2] \rightarrow [-1,1]$, (LTW=Loser, Typical, Winner)

$LTW(X) = P(X|Winner) - P(X|Loser)$ e.g. $LTW(X) = 0.65 - 0 = +0.65$
$LTW(X) = 0 - 0.49 = -0.49$
$LTW(X) = 0 - 0.8 = -0.8$

In case of $P(X|C) \neq 1$ $LTW(x)$ is equal proportionally to scaled Bayes’ posterior probability. Scaling means the translation of values into the positive and negative range using the equation above. In case of integer values LTW has the following properties:

$LTW(X) = -1 \quad <=> P(X|Loser) = 1$

$LTW(X) = 0 \quad <=> P(X|Typical) = 1$

$LTW(X) = +1 \quad <=> P(X|Winner) = 1$

The above mentioned properties of LTW are valuable to detect directly a specific category (Loser, Typical, Winner). That means it forms categories based on decision rules.

- “Winner” will arise when $LTW(X) \geq +0.5$. The category by decade is $+1$.
- “Loser” will arise when $LTW(X) \leq -0.5$. The category by decade is -1.
- “Typical” will arise when the following decision criteria is satisfied: $-0{,}5 > LTW(X) < 0.5$. The category is equal to zero.

Each category by decade is element of a pattern. The term pattern is used in this thesis to describe unique profiles of categories over time (15 decades). In summary a pattern consists of a set of 15 categorial values (-1, 0, 1) related to the membership label (“Loser”, “Typical”, “Winner”).

LTW is particularly suitable due to the symmetric and limited range. In order of a clustering approach (Clustering and Classification of Patterns, in section 6) and related proximity measurements several advantages are further discussed. The above presented decision rules have the following properties:

- The classification of patterns is simple and easy to realize.
- The rules are as compact as possible to handle the complexity.
- It is possible to represent meaningful properties of the patterns to be classified.
- The categorical features ease the selection of relevant patterns.
- It is possible to classify new patterns and to extend the classification by increasing its dimension (e.g. another decade).
- The classification is able to highlight the relationship between patterns.

It is possible to discuss the expected results of such a classification process. When the number of decades is 15 and the number of categorical features ("Loser", "Typical", "Winner") by decade is 3 the possible number of different patterns is 3 to the power of 15. By the author of this thesis it is assumed that a number of approximately 1000 different patterns will arise. One reason for this number might be the possibility that there are about three or five patterns consisting of a large group of communities in relation to the total amount of 2896 communities. For example a individual pattern size between 300 and 500 communities is conceivable.
Based on the data distribution of population change and the knowledge that most communities show a typical development by decade over time there are probably several patterns characterized in most decades by a "Typical". In consideration of different patterns it is further assumed that there are many patterns consisting of only one or two communities. Perhaps such group of more or less outliers is defined by an amount of 300 communities (10 % of the total amount of communities). In view of the Swiss spatial planning policy (e.g. desired spatial development, balancing interests, structure plan = "Richtplan", plan directeur in French) the author of this thesis does not expect any patterns characterized just by "Losers" or just by "Winners".

Figure 9 shows the result of pattern classification sorted by frequency. There are 880 different patterns observed. Each pattern is characterized by a unique set of categories ($LTW(d1, \dots, d15)$). The observed pattern matrix (2896 Swiss communities $\times$ 15 decades) gives an overview of the frequency and variety of patterns.

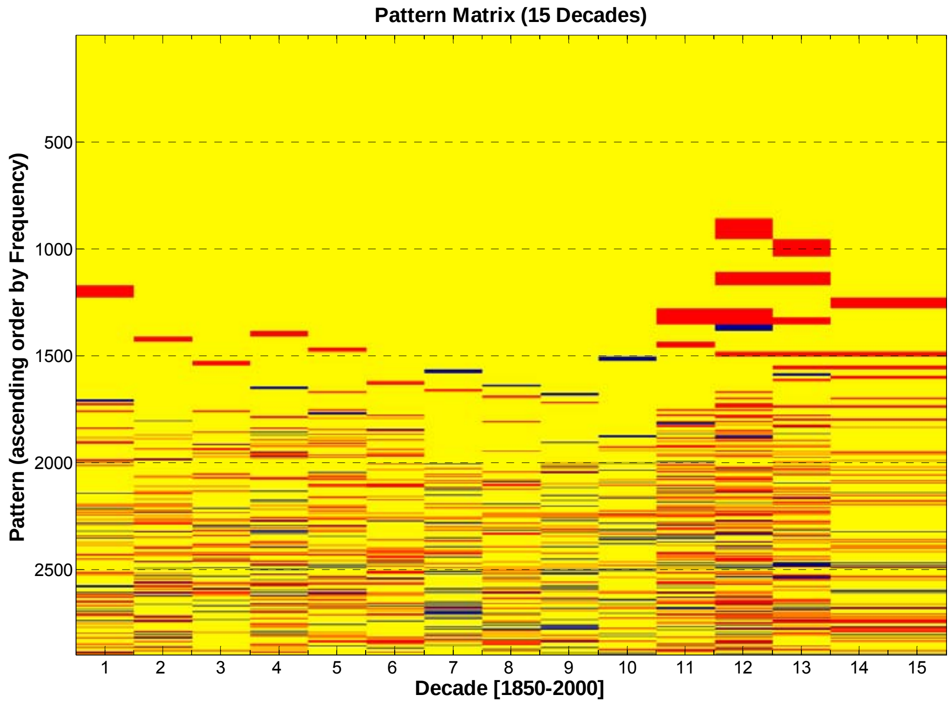


Figure 9: Frequency of patterns, blue="Loser", yellow="Typical", red="Winner"

5.3 Expected versus observed frequency of patterns

Based on the classification and categories by decade it is possible to discuss the joint distribution of their occurrence and the results of patterns in general. In view of all 15 decades ($decade\ d = 1, \dots, 15$) the observed probability $P_{T,d} = P(C = Typical)$ does not vary over time. The mean value is 85 % and the standard deviation of the probability $P_{T,d}$ is just **6%**. Figure 11 shows the observed probability $P_{T,d}$ over time.

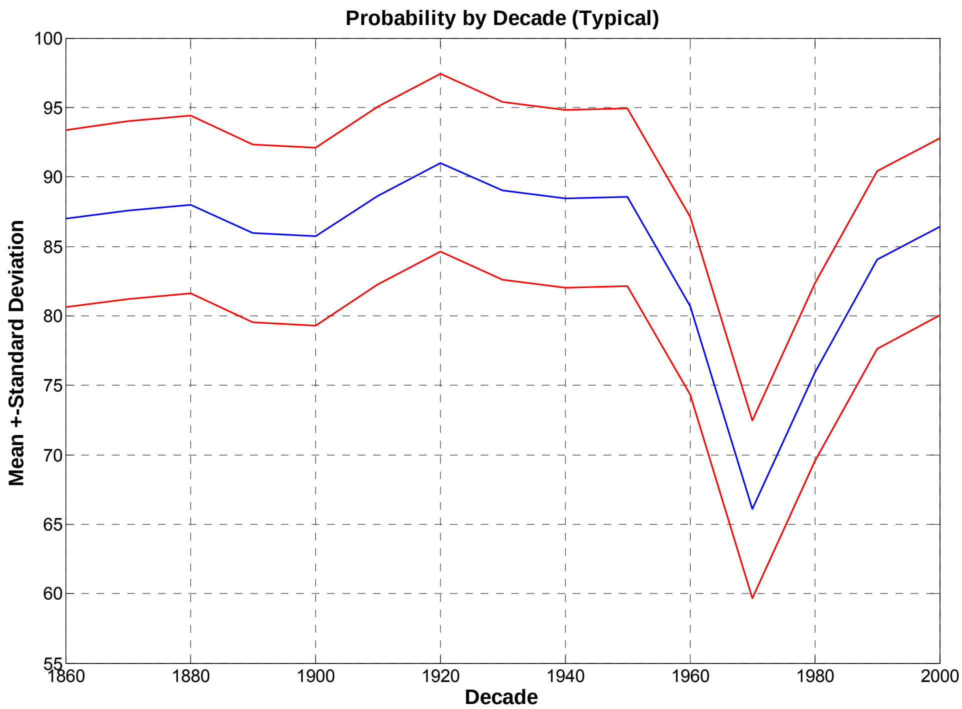


Figure 10: Probability $P_{T,d}$ observed in 15 decades

Under the assumption of statistical independency of decades and a constant probability $P_{T,d}$ it is possible to model the frequency distribution of patterns as a binomial distribution (Bortz, 2005, p.65). Such model is realized in this form: If P_T is given as the probability that a community will be described as "Typical" the probability P_{NT} is also given that a community is a "Non-Typical": $P_{NT} = 1 - P_T$. That means a "Winner" or a "Loser". In addition u is the amount of "Non-Typical" features in a pattern. The probability p_{bin} of a pattern with $u\ in$ [0,15] is then computed as follows:

$$p_{bin}(u) = \binom{15}{u} \cdot P_{NT}^{u} \cdot P_{T}^{(15-u)}$$

The expected frequency of a pattern is obtained using the probability p_{bin} in relation to the total amount of communities (=2.896). The comparison of expected pattern frequency by the model to the frequency by observation identifies differences. Figure 11 shows the observed and expected frequencies of patterns in relation to the number of "Non-Typical" features per pattern.

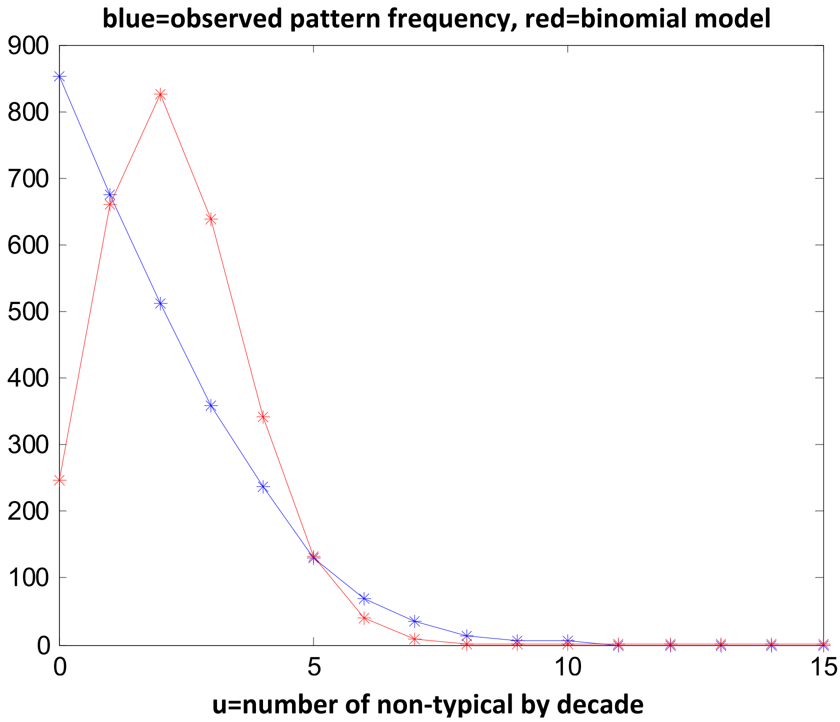


Figure 11: Size of classified patterns (total amount=880)

Due to the visualization of pattern frequencies (observed versus expected) it is obvious that the observed patterns which do not consist of any "Non-Typical" feature are clearly above the expected frequency (by the factor 4). The inspection of distribution has already indicated such big amount of "Typical" communities by decade. This pattern is essential for the description of the Swiss population change.

In contrast to this major deviation of frequencies the observed patterns with two or three "Non-Typical" features are below the expected frequency (by the factor 2). Those patterns should be inspected more precisely when later on deciding about relevance of patterns. In focus of any further deviations the observed frequency of patterns with six or seven "Non-Typical" features is also above the expected frequency.
The expected and observed frequencies are further examined by using the cumulative distribution function. According to the observed patterns with number $u = 0$ or u=1 (number of "Non-Typical") it is remarkable that more than 50 percent of all Swiss communities are characterized by these patterns. Furthermore the frequency of observed patterns is clearly disproportional to the expected frequency of the model. Nearly seventy percent of communities are described by patterns with zero, one or two "Non-Typical". The observed total number of "Non-Typical" per pattern is 10.

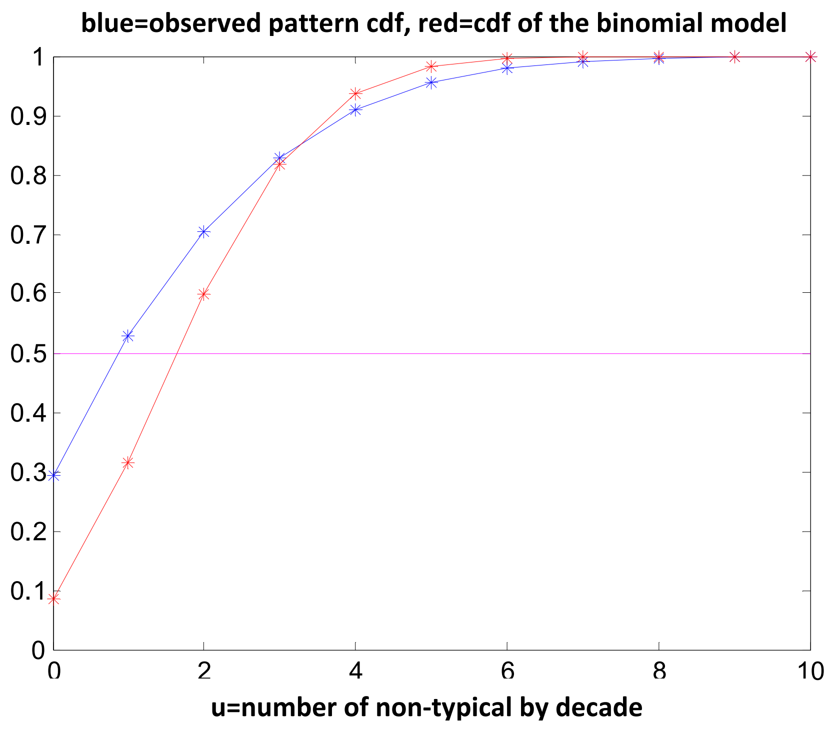


Figure 12: Size of classified patterns (total amount=880)

In conclusion such comparison of frequencies has led to the detection of patterns that are of special interest in view of feature characteristics that are emphasized by the observed data, e.g. one "Non-Typical" in 15 decades. Clustering should be based on patterns showing on the one hand such a disproportional frequency (clear deviation of expected and observed frequencies) and on the other hand a meaningful amount of population per pattern. Each pattern has a specific impact of population that is discussed in detail in the following section.

5.4 Relevance of Patterns

The goal of this section is the final selection of patterns that are relevant and interesting enough for the clustering approach (similar/dissimilar). Two main tasks are to mention for this purpose. The first one deals with the definition of a specific criterion of relevance (population impact). The second one supports the pattern selection using in addition a procedure of information optimization (Pareto Principle). It is assumed that the final amount of relevant patterns varies from 80 to 150.

5.4.1 Long-term Impact of Population

In view of strategic spatial planning it is a crucial task to measure the relevance of each pattern. Each pattern is already characterized by the amount of communities. There are few patterns comprising a large amount of communities and there are many patterns comprising a relatively small amount of communities. Since the Swiss community system is very heterogeneous (e.g. in population, size etc.) it is by far not sufficient to use the number of communities for the detection of relevant patterns.

It is a pragmatic planning approach to have a deeper look to the size of population. The question is how many people have an impact on one pattern? For this issue it is possible to use population data of one specific year in a short-term perspective or summarized population data in a long-term perspective. The mean of population by decade was already mentioned in Chapter 4.3.2. This value represents the denominator of *Relative Difference* and describes all 2896 communities by decade. In consideration of the long-term description of population change (1850 to 2000) it seems conceivable to compute the Mean of all 15 mean values by decade. Furthermore it appears complete to sum up all these “long-term Means” of communities related to one pattern. Finally there is one number (LPI) describing the long-term impact of population on each pattern:

$$LPI = Longterm\ Population\ Impact = \sum_{1}^{CP} Longterm\ Mean = LM_1 + \cdots + LM_{CP}$$

with:

$$Longterm\ Mean = LM = \frac{1}{15} \cdot \sum_{D=1}^{15} Mean_D = \frac{1}{15} \cdot (M_1 + \cdots + M_{15})$$

$Mean\ Decade = Mean_D = \frac{1}{2}(P_{t-1} + P_t)$, new value P_t, $previous\ value\ P_{t-1}$

$CP = Number\ of\ communities\ \in pattern$

5.4.2 Optimum of Patterns

The amount of 880 unique patterns will be now scrutinized in order of relevance. Thus several observed criteria are used such as the disproportional frequency of patterns (expected versus observed), the number of patterns, the number of “Non-Typical” per pattern and in particular the population impact.

The first step is to find relevant patterns by ordering all patterns by frequency. The most frequent pattern is described by 15 “Typical” over time. It is of course relevant due to the representation of the Swiss “Typical” development of population change over time (total=852, 30 % of Swiss communities, LPI=959,636).

For the selection of other relevant patterns it is helpful to have deeper look to the population impact on each pattern. The order of patterns leads to the selection of six other relevant patterns. The population impact on each pattern is here always above 100000. Since the next following pattern is described by just 76948 such value is reasonable as a clear boundary. Three of these 6 further patterns are described by just one community. In view of population it is not surprising that these singular patterns are the famous and largest Swiss cities such as Zürich (=7 “Non-Typical”), Bern (=4 “Non-Typical”) and Basel (=4 “Non-Typical”). Geneva is to find in an additional pattern of these 6 patterns. It consists of 4 communities (= 2 “Non-Typical). All these patterns are nearly describing 45 % of the whole Swiss population (=Long-term Population Impact).

The question is how to identify an optimal number of other relevant patterns in the amount of remaining patterns? The patterns and computed values of the long-term population impact are now used for information optimization. The same idea is used as for a theoretical foundation of the Pareto 80/20-law (Ultsch, 2001). It is assumed by the author of this thesis that the minimal value describing a relevant population impact on a pattern is within the range of 2000 to 10,000.

A Lorenz curve presents the association of the number of patterns and the computed long-term population impact (see Figure 13). From the ideal point 0% of long-term population impact and 100% of knowledge of the patterns the distance to the real situations on the Lorenz curve is measured. The identification mark ’a’ in Figure 13 shows the shortest of such distances. From this it is concluded that in order to gain different patterns only about 14% of the patterns, the 14% relevant ones, should be examined in deep detail. This is consistent approximately with the well known Pareto 80/20-law.

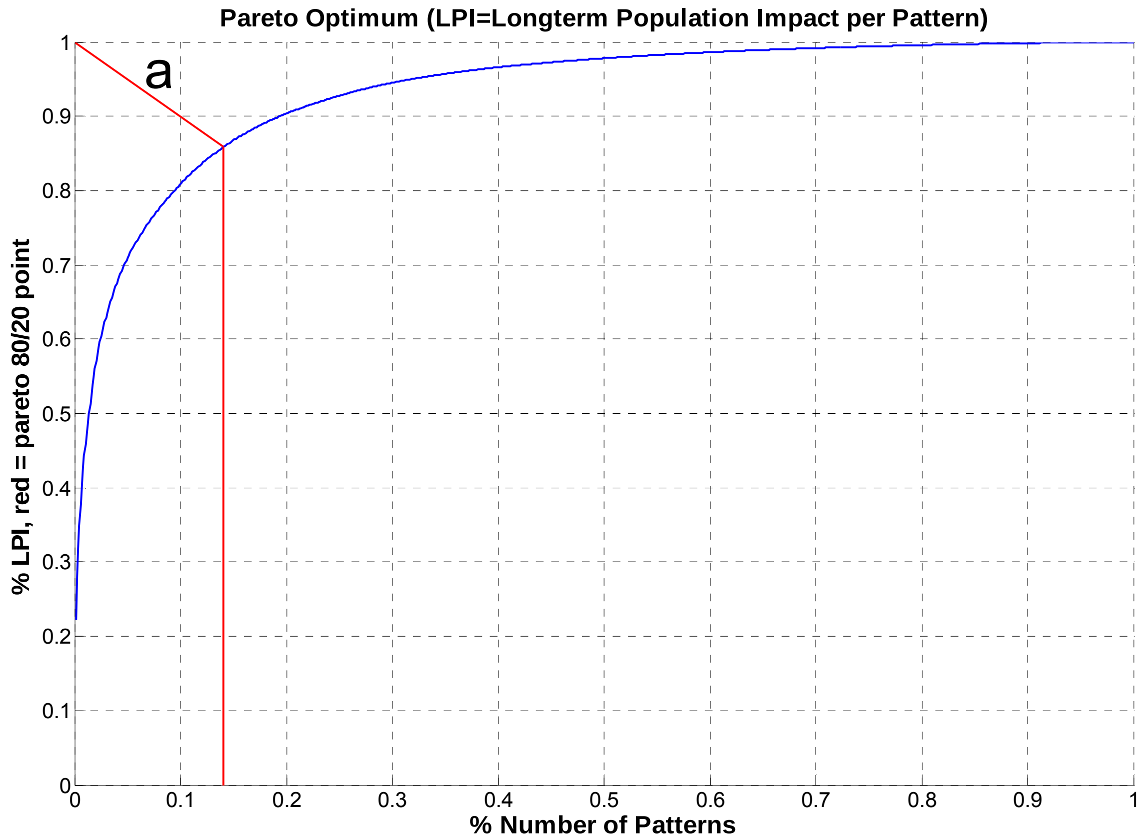


Figure 13: Information Optimization in View of relevant Patterns in Switzerland

A long-term population impact of 5000 per pattern is the observed boundary. 122 relevant patterns are selected (~14 percent of 880). All relevant patterns are characterizing about 1899 Swiss communities (65% of 2896 communities). All these relevant patterns will be later used for the clustering procedure.

5.5 Periodical subdivision

For the purpose of interpretation, comparison and in particular for clustering it is useful to define a periodical subdivision. Such periodical subdivision is based on knowledge about the general population growth and population development in Switzerland. Furthermore historical events (e.g. negative impact of World War I and II) and urbanization processes (e.g. Schaeffer, 1992; Tschopp, 2002) are taken into account.

The already observed patterns and localized results (see appendix) are the basis to define reasonable intervals. Figure 15 shows in addition the pattern matrix (122 patterns × 15 decades) allowing a visual comparison of patterns (development of population by decade). The view to the patterns by decade leads to the formulation of three characteristic periods. The first one comprises the first six decades (1-6) and is equal to the period 1850 to 1910. The second one takes 4 decades (7-10) and is equal to the period 1910 to 1950. The last period covers 5 decades (11-15) and is equal to the period 1950 to 2000. All three periods are roughly characterized in view of the Swiss population as a whole and general urbanization process.

Period 1850-1910 (Industrialization and urbanization): This period is characterized by a strong population and urban growth. Two phases of growth generally occur (1850 to 1880 and 1888 to 1910). Furthermore a short period of stagnation (1880-1888) is discussed in the literature (Rey, 2003; Schuler et al., 2002). Characteristics are rapid economical and social changes, tourism and furthermore railway construction.

Period 1910-1950 (World War I, II): The period of both World Wars and the related meantime is characterized by stagnation in view of the whole Swiss population and growth rate. Switzerland is subject to separation and negative impact of both wars.

Period 1950-2000 (Urbanization/Suburbanization): A period of population growth and economic boom faces Switzerland. In the 1950s a suburbanization trend was starting in Switzerland because core urban centers were growing slower than smaller urban and rural areas. Later on counter-urbanization (Mulugeta/Schaeffer, 2009) is occurring (e.g. 1970 to 2000). In general counter-urbanization occurs when population growth in areas with small populations exceeds that in large population centers (Dean et al., 1984). Counter-urbanization is also defined as the reversal of the long trend towards more and larger urban settlements. Such trend implies a process of settlement system change.

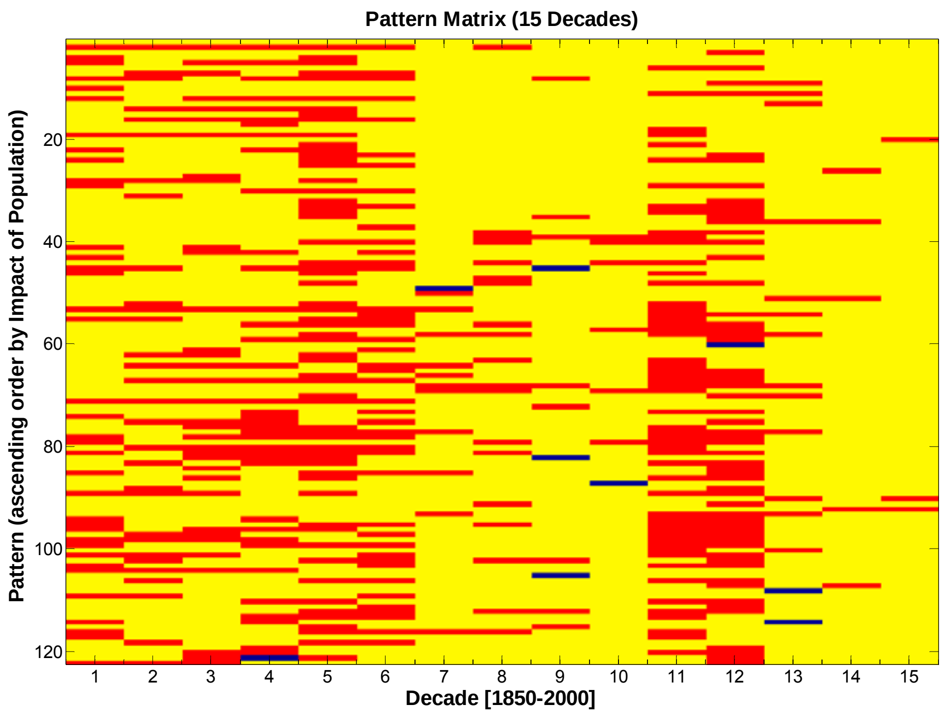


Figure 14: Relevant Patterns (blue="Loser", yellow="Typical", red="Winner")

5.6 Growth Indicator by Periodical Subdivision

For a unique and comprehensive pattern description of each period it is necessary to define a specific indicator. Such indicator measures the population development per period and sums up the observed patterns by decade:

period 1 (1850-1860, 6 decades): $GI_1: \{-1, 0,1\} \rightarrow [-6, +6]$

period 2 (1910-1950, 4 decades): $GI_2: \{-1, 0,1\} \rightarrow [-4, +4]$

period 3 (1960-2000, 5 decades): $GI_3: \{-1, 0,1\} \rightarrow [-5, +5]$

The growth indicator GI_n is computed as follows:

$$growth\ indicator = GI_n = \sum_{D=a}^{b} p_D = p_a + p_{a+1} + \cdots + p_b$$

with $p = pattern\ by\ decade, D = decade\ , n = period,$

$a = first\ decade\ by\ period$, b= $last\ decade\ by\ period$

The growth indicator GI_n is characterized for each period $n \in \{1,2,3\}$ by a different number of decades D. Such decades are in interval $[a, b]$ and represent a different number of patterns by decade.

GI_n has the following properties: Large negative values characterize a “Loser”. Large positive values indicate a “Winner”. Values by Zero represent a “Typical”. In case of only one positive and only one negative value in a longer period such characteristics are cancelling each other. Thus the pattern is finally characterized as a “Typical”. In view of the pattern matrix of the following section such specific situation is not really occurring.

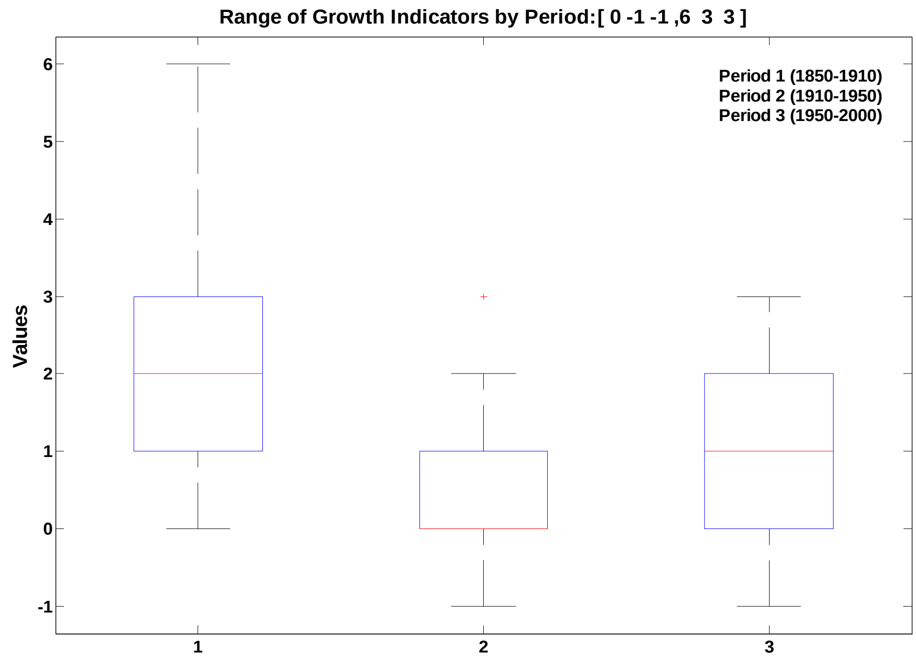


Figure 15: Growth Indicators of Relevant Patterns

6 Clustering and Classification of Patterns

In this thesis clustering is used as a measure to detect dissimilarity of one pattern p (see definition in section 5.2) to one another on the basis of the Euclidean Distance.

6.1 Dissimilarity of Patterns

Decisions about dissimilarity are often based on distance measurements and are used in this thesis to quantify the proximity of patterns in the multidimensional feature space (3 periods $n \in \{1,2,3\}$ and specific additive values of pattern features). Thus let patterns $p_i, p_j \in R^n$. Based on a number of patterns P a symmetric PxP matrix is realized. For patterns with n dimensions (number of periods $n = 1, \ldots, 3$), the Euclidean distance $E(p_i, p_j)$ between two patterns p_i and p_j is defined as follows:

$$E(p_i, p_j) = \sqrt{\sum_n (p_{in} - p_{jn})^2}$$

Where p_{in} and p_{jn} represent the n^{th} dimension values of p_i and p_j. Matrix E is symmetric. Euclidean Distance is used to represent relatedness in pattern content, such that semantically similar patterns are placed closer to one another in feature space than less similar ones. Such measurement is comparable in a wider sense (feature space) to the famous first law in geography: „Everything is related to everything else, but near things are more related than distant things“ (Tobler, 1970).

<u>Why is Euclidean distance an appropriate measurement for clustering of patterns?</u>

- Since data of patterns will be compared, the variance of the data has to be taken into account. Generally a great deal of variability in the range and distribution may pose a problem for cluster algorithms which involve distance measurements (e.g. Ward). For example atypical scores in a distribution (outlier) can wildly determine the Euclidian distance. The values of GI_n are here advantegous. The influence of outlier (extrem values) on Euclidean Distance is clearly alliviated due to the symmetric and limited range of the growth indicator: (*period 1*: $[-6,6]$, *period 2*: $[-4,4]$, *period 3*: $[-5,5]$). When dealing with dissimilar patterns the distance is in minimum 1. The pattern properties and related integer values of the growth indicator have the advantage of a precise distinction of patterns. Such measurement provides the clustering by distance.

- It is often the case that components of data feature vectors are not immediately comparable. When variables are on different measurement scales (e.g. km, area, persons), standardization is necessary to balance the contribution of the variables in the computation of distance. GI_n values of patterns are on a similar scale.
- The Euclidean distance matters on peak heights, so two patterns of similar shape but different mean absordance would be treated as much different. When comparing patterns based on $measured\ values$ such criteria is not necessary due to meaningful characterisics by period ("Loser", "Typical", "Winner").
- Euclidean distance is not appropriate for variables that are correlated. This property was already proofed and is not significant.

6.2 Ward Clustering

When analyzing patterns the goal is to partition the **122** patterns into cluster, where the number of cluster is unknown and has to be determined from the data of patterns (see section 5.5). A cluster is determined by using 3 periodical growth indicators (=each period is defined by additive pattern information). The advantage is the measurement of dissimilarity in terms of integer values by 1. Using Ward algorithm takes such property into account because the algorithm unifies cluster such that the variation inside these cluster is not increasing dramatically. In contrast to other agglomerative algorithms (e.g. Single Linkage) this algorithm does not put together cluster with smaller distance, but it joins cluster that do not excessively increase the information loss. The clustering depends upon how similar (dissimilar) the patterns are to each other. Similar patterns are treated as homogeneous cluster, whereas dissimilar patterns form additional cluster. As a result clusters should be far enough apart that cluster are easily identifiable. Each cluster is later be replaced by an integer representing each cluster. When analyzing patterns the requested result is a clear and compact cluster structure. Before clustering it is assumed by the author that there are about 6 to 12 clusters which fulfill such requirements (3 periods and a specific value of GI_n lead to discriminable cluster descriptions). The interpretation of the dendrogram leads finally to an eight cluster solution. Cutting at the marked point provides well separated clusters. Dissimilar cluster are therefore combined higher up to the diagram. A deeper look to the distance leads to the understanding of a clear and compact structure. A finer solution would lead to an elusive and marginal distinguishable amount of clusters. Furthermore the eight cluster solution seems reasonable in view of the total amount of 122 clustered patterns.

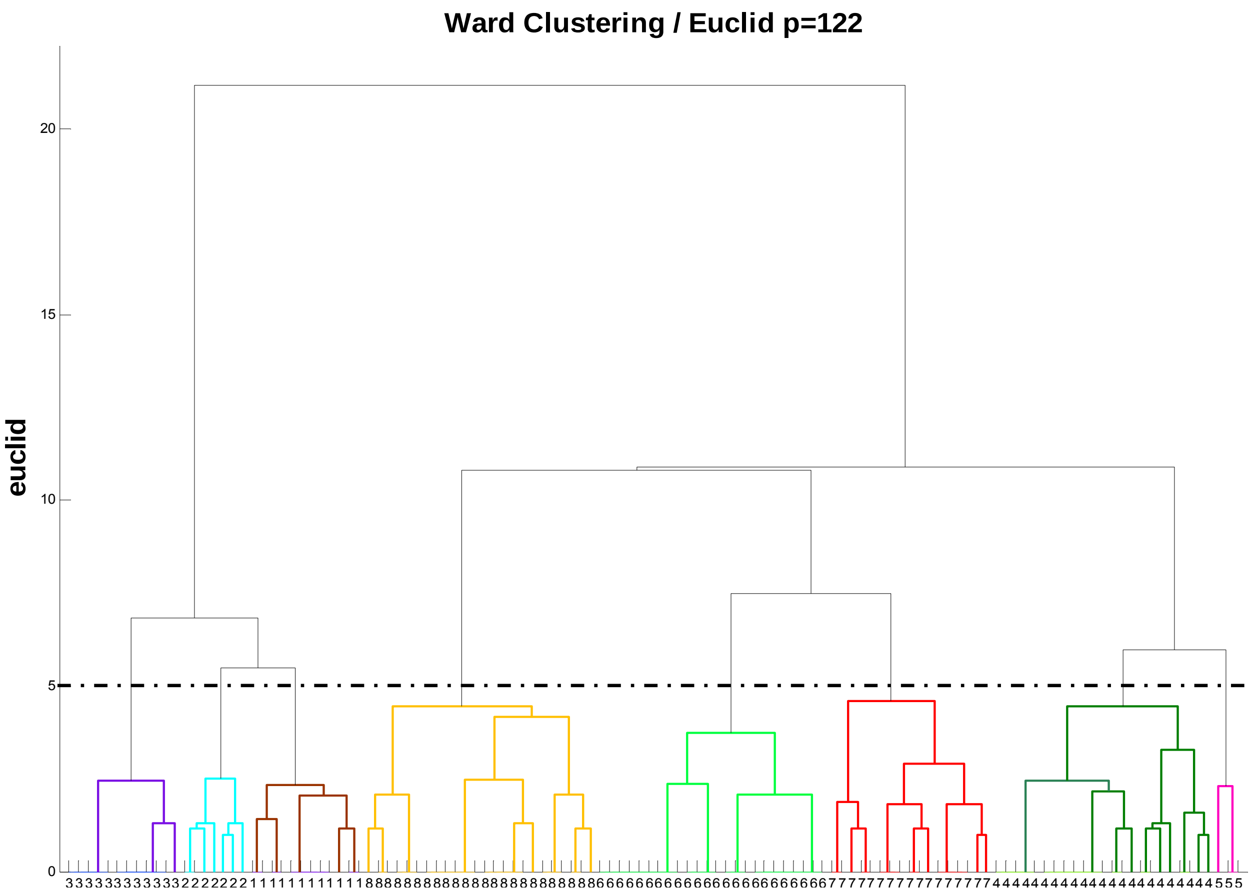


Figure 16: Result of WARD Clustering (dendrogram)

6.3 k-Nearest-Neighbor-Classification of Patterns

In the following the author will use the term class instead of a cluster in order to avoid misunderstanding. The clustering of 122 relevant patterns has led to 8 classes with different size observed. Each class has specifiable properties described by three growth indicators. These indicators are related to the three defined periods (1850-1910, 1910-1950, 1950-2000). The classes are already characterizing 1899 Swiss communities. However, there are 997 Swiss communities and related patterns without a class membership. The patterns of these communities should be allocated to the 8 classes.

The aim is therefore to find an allocation procedure which allows identifying the class membership of all 2896 communities respectively their patterns. For this purpose a *k*-nearest neighbor classifier will be realized (see section 2.4). The first step is to construct a classifier based on the classified patterns. The second step is to start the allocation of the rest of patterns using the constructed classifier. Figure 17 shows the principle using training and test data in general form.

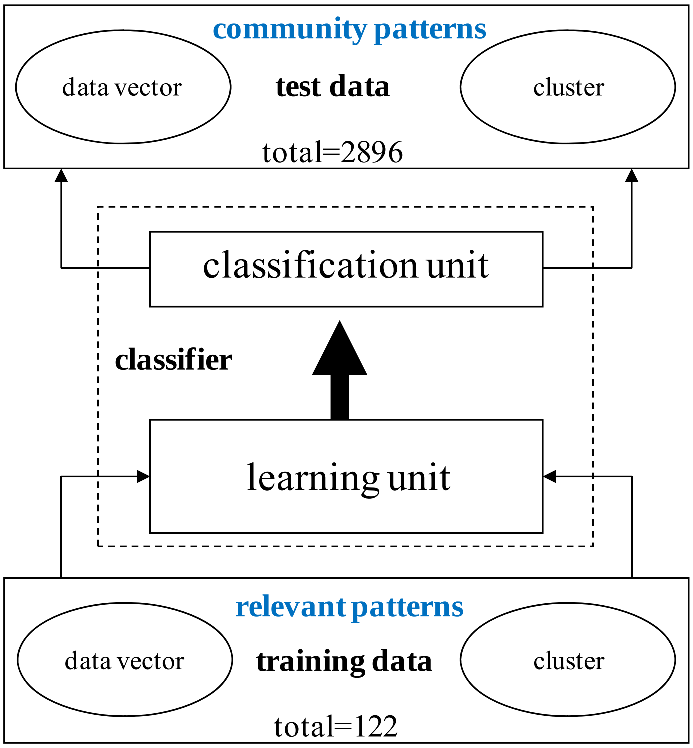


Figure 17: Construction of a classifier with training and test data

In case of the pattern classification the k-Nearest-Neighbor classifier will be initially constructed (trained) on the basis of all labeled patterns. Such procedure is necessary to assign the quality of the classifier. It is tested how far the observed classes will be rediscovered.

The number of neighbors is $k = 1$ and the growth indicators are used to identify the accuracy of allocation to the given classes (overall accuracy). Such procedural step leads to an accuracy of 100% (expected 100%). It is confirmed that the accuracy is good enough to classify the other patterns. This set of 122 relevant patterns is now a kind of reference and is the basis to start the comparison with the other patterns without a class membership (test data). For that purpose a value of $k = 1$ is chosen (Figure 18).

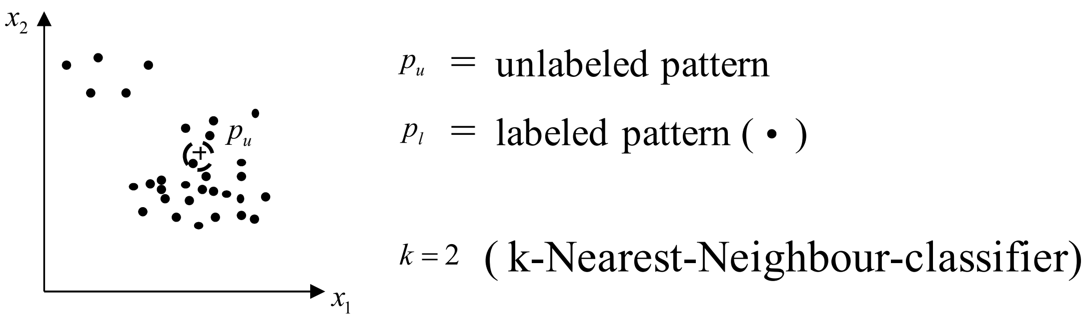


Figure 18: Allocation problem (labeled versus unlabeled pattern)

The result of classification is displayed in view of the size of the classes and the amount of population per class (see Figure 19, Figure 20).

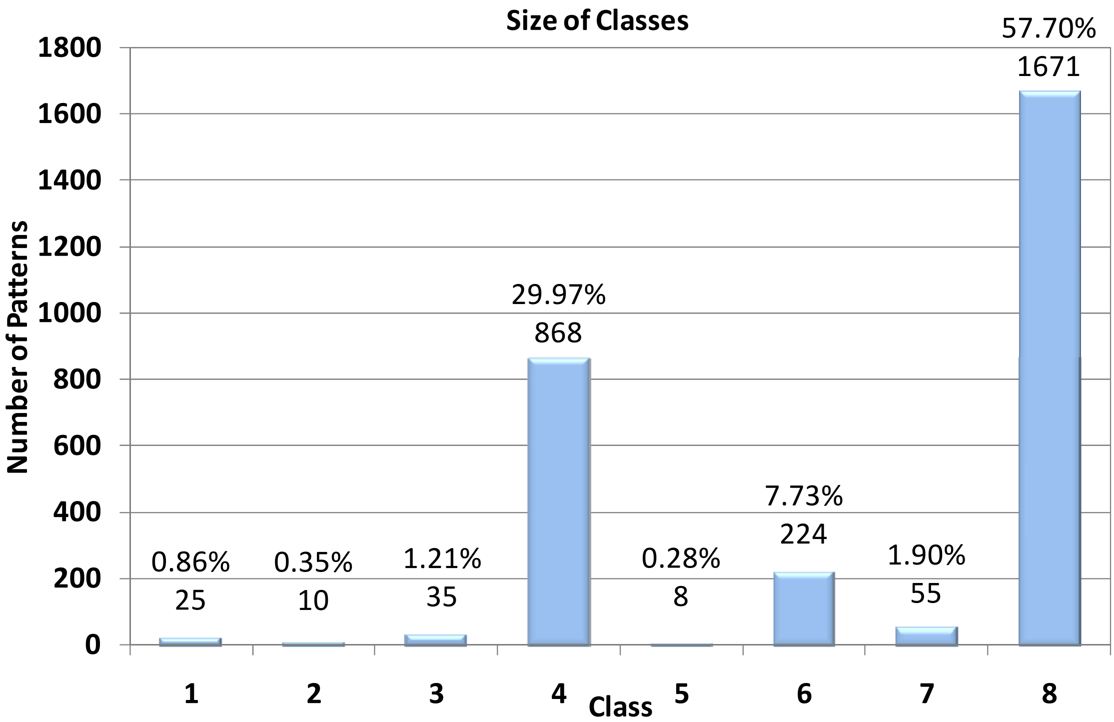


Figure 19: Size of classes

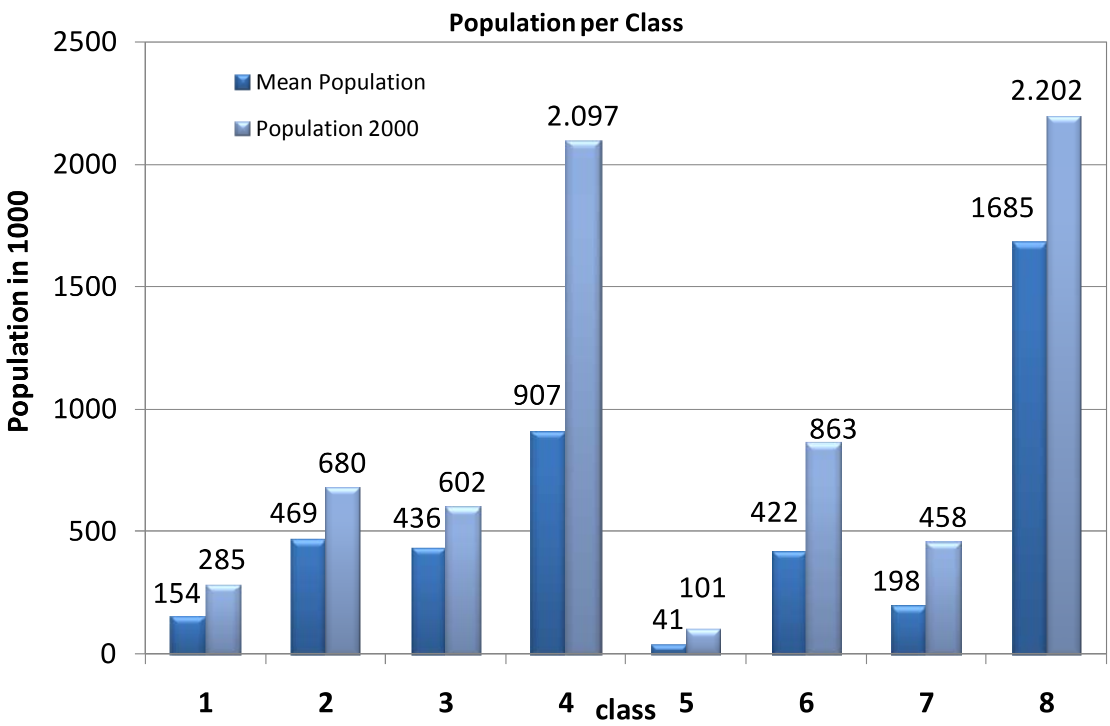


Figure 20: Mean Population and Population of the year 2000 per class

6.4 Typical Communities of classified Patterns

Each class should be further described by a typical community. The identification of such a community is realized in a two stage procedure. At the beginning each class is characterized by the mean value of each growth indicator by period. The median of each growth indicator is also computed for the purpose of comparison of both values. The Median is the middle value below which 50% of the cases fall. The three presented values allow the selection and description of a characteristic (mean) pattern by class c:

$$Period\ 1: \overline{GI}_{1c} \qquad Period\ 2: \overline{GI}_{2c} \qquad Period\ 3: \overline{GI}_{3c}$$

Secondly the typical community of each class is identified on the basis of the mean population of 15 decades. As mentioned before the impact of population is reasonable when dealing with population development. All communities of the selected pattern are ordered in descending order by population. The community with the maximum is described by the author as a typical community. A class can be characterized using the label of the typical community: “is like St. Gallen” or “is like Solothurn”. Table 5 shows the eight typical communities and their properties by period. The value of the growth indicator by period can be used to find a label for each period. The author decides to characterize a community as a “Winner” by period when the positive relative frequency is equal or above 25%. That means for example that a value of 1 in period 2 is equal to an overall winner in that period: $\boldsymbol{GI_2 = \frac{1}{4} * 100 = 0,25}$%. A winner can be further characterized by different intensities per period.

The classification of patterns and the identification of typical communities lead to another important aspect. The author wants to emphasize that Solothurn is the **typical** Swiss community in context of the long-term population development (1850-2000).

Table 5: Typical communities of 8 classes and their properties

c	Typical community	GI_{1c}	1850-1910	GI_{2c}	1910-1950	GI_{3c}	1950-2000
1	“Gossau”	+4	“Winner”	0	“Typical””	+1	“Typical”
2	“St. Gallen”	+5	“Winner”	0	“Typical”	0	“Typical”
3	“La Chaux-de-Fonds”	+3	“Winner”	0	“Typical”	0	“Typical”
4	“Uster”	0	“Typical”	0	“Typical”	+2	“Winner”
5	“Ascona”	0	“Typical”	+3	“Winner”	+1	“Typical””
6	“Zug”	+1	“Typical”	0	“Typical”	+1	“Typical””
7	“Dietikon”	+3	“Winner”	+1	“Winner”	+2	“Winner”
8	“Solothurn”	0	“Typical”	0	“Typical”	0	“Typical”

6.5 Localization, Spatial Reasoning

The aim is to obtain a view to each class and to formulate some spatial abstractions. The interpretation leads to explanations and triggers some hypothesis that might be valuable for further explanations and subsequent analysis.

The communities of class 1 are mostly to find in the Swiss agglomeration zone. The 25 communities are just representing 1 percent of Swiss communities. Those communities are mainly showing an increase of population in the first period (1850-1910). This increase is above the general Swiss trend and therefore there are many winning communities in this period observed.

Class 2 is described by many big cities respectively cores of an agglomeration. In total there are 10 communities allocated to this class. Zurich, St. Gallen, Luzern, Lausanne and Montreux belong for example to this class. The urbanization process in the first period has a big influence on these communities. In the first period communities are in average characterized by a “winner” in 5 of six decades. Later on many of these communities follow the typical population development in Switzerland. It is further interesting that only Zurich (1910-1950: 1 “Winner”), Lausanne (1910-1950: 1 “Winner”) and Olten (1910-1950: 1 “Winner”, 1950-2000: 1 “Winner”) are characterized by some “Winners” in the subsequent periods.

Communities of the Class 3 are often to find in the Midlands (19/35) and thus there are many to find also in the lower regions. The communities of this class are also characterized by many “winning” communities in the first period (1850-1910). In the subsequent periods the class is in average following the typical population development.

There are 30 % of Swiss communities allocated to Class 4. There is a widespread of this class in the Swiss area. It is characterized by winning communities in the last 5 decades. It is assumed that these communities did clearly benefit from the economical and social changes in Switzerland in the recent years. Furthermore the changes in transport systems did probably support such a development of communities.

The communities of class 4 are frequently to find in the agglomeration zone. Communities are often suburban or peri-urban. It is to state that about 2 million of the Swiss population is living in these communities. However there are not so many high populated communities belonging to this class. The class 4 is often localized in the canton Zurich, Vaud, Aargau, Basel (hinterland), Ticino and Geneva. The typical community was “Uster” nearby Zurich.

There are 8 communities allocated to class 5. In comparison to other Swiss communities these communities are characterized by a “winning” period of population between 1910 and 1950. In average these communities are “winning” in three of 4 decades. These communities are located in the canton of Zurich, Valais, Geneva, Basel, Bern and Ticino and the area of these communities is small. Examples of this class are Ascona, Riehen nearby Basel, Montana.

The class 6 is also often located nearby big cities and cores of agglomeration. These communities are mostly to find in the lower regions. The majority of communities are located in the suburban area of Switzerland. This class is in third position in view of the total number of 224 Communities. These communities are showing in average “Typical” properties by period. In comparison to class 8 the majority of communities can be characterized at least by one “Non-Typical” (“Winner”/”Loser”) in the first and third period.

The class 7 contains 55 communities. There are many to find in the canton of Zurich and Valais. The communities are often to find in the agglomeration zone and represent small city centers. Bellinzona, Emmen, Grenchen, Wettingen, Thalwil and Dietikon (typical community) are just some examples. The communities are often “Winners” in the first and in particular in the third period. It is assumed that communities of class 4, 6 and class 4 express recent developments of urban sprawl.

Class 8 describes the typical population development in Switzerland. All periods are therefore described by a development that represents the Swiss trend of population change. There are 60 % of all 2896 Swiss communities that represent such a development. Furthermore this class comprises a large amount of population both in view of the long-term mean of population and the population in the year 2000. The midlands and Eastern Switzerland are influenced by typical communities. The typical communities are often to find in the canton of Bern (287), Vaud (191), Graubünden (156), Fribourg (147), Valais (114), Ticino (117), Aargau (104), Jura (75), Luzern (71) and Solothurn (68). Typical communities are to find both in the lower and higher regions. About 1400 communities are located below 1200 meters but many of them are to find in rural regions of Switzerland. Solothurn is the typical Swiss community in view of population change by 15 decades.

Figure 21 determines the eight class solution and displays the spatial structure of classified communities.

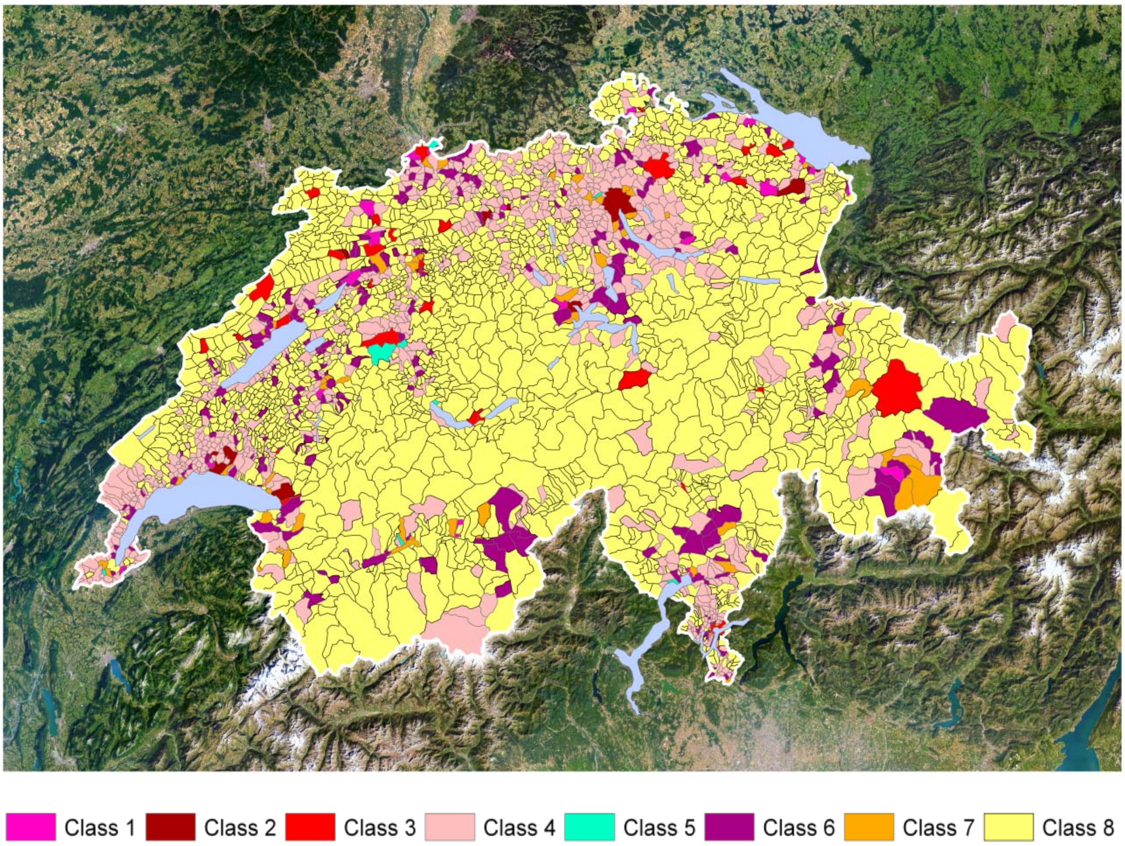


Figure 21: Localization of the classification result (8 classes)

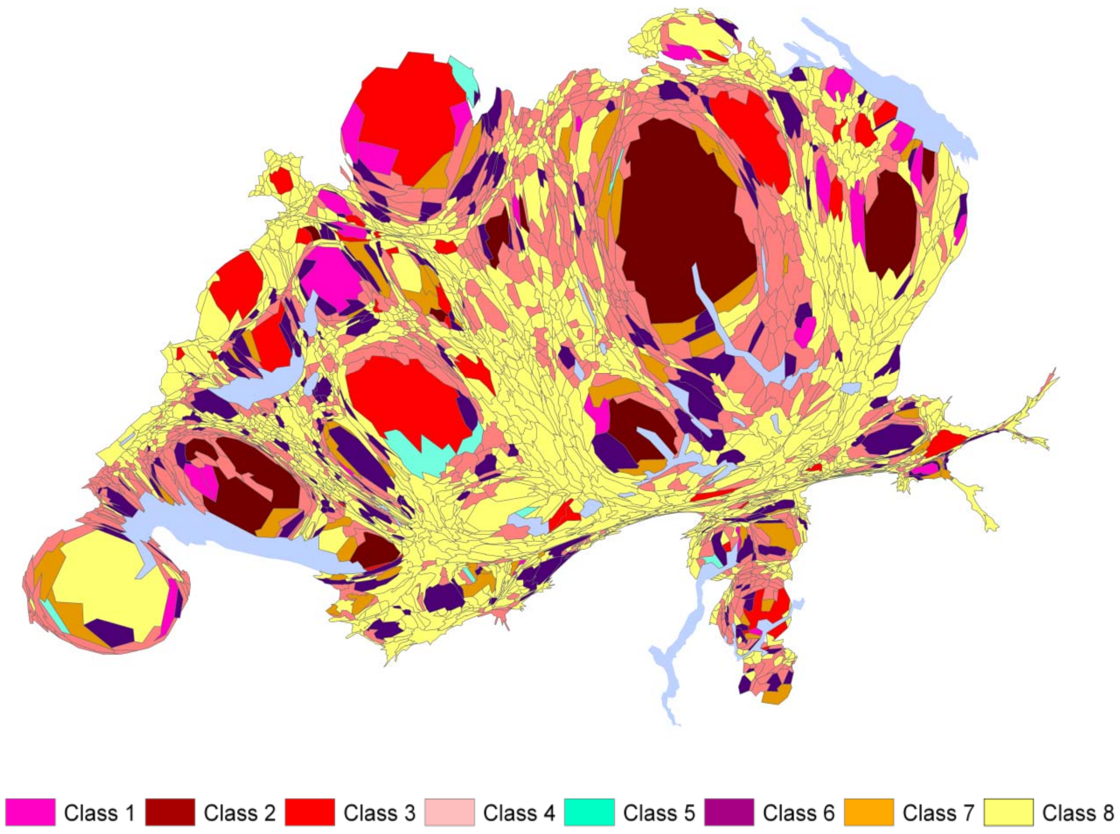


Figure 22: Localization of the classification result in proportion to the population

For many years map makers have searched for a way to construct cartograms. These are maps in which the sizes of geographic regions such as communities appear in proportion to their population or some other analogous property. Such maps are used in this thesis for the representation of patterns by decade (see appendix). Furthermore Figure 22 displays the classification results using the long-term mean of population in proportion to the area of a community. The method of Gastner and Newman (2004) is the basis for the realization of such maps. It is a method for constructing density-equalizing cartograms. The method supports to choose the balance between good density equalization and low distortion of map regions. The implementation in GIS software was used straightforward. (number of cells in width and height is 2048, see Burgdorf, 2008). The cartogram provides to emphasize the cores and other communities of agglomerations in visual manner. But the proportion of the suburban areas (e.g. rings) is narrowed in comparison to the center. The low populated regions and in particular alpine regions are less important in this display.

In addition the observed patterns are now presented in view of the property “Typical” and “Non-Typical”. Figure 23 and Figure 24 show all localized patterns. Under the assumption of statistical independency of decades and a constant probability it was possible to model the frequency distribution of these two types (patterns) as a binomial distribution (see section 5.4). As a result it was detected that the observed patterns of “Typical” feature are clearly above the expected frequency (by the factor 4). The geographical regions Jura, Central Plateau and Alps are characterized by the “Typical” pattern in different frequency and spatial coherence. When thinking about the own expectation to the localization of typical communities it is interesting for the author that the canton of Bern, Luzern, St. Gallen are certainly characterized of this pattern. Higher and in particular alpine regions seem to be characterized in most cases by other patterns. Further spatial analysis and techniques of data mining should be taken into account to find several reasons for this special frequency and spatial distribution. This pattern characterized by 852 communities is interesting in view of a discrete sampling. It is further to see that the applied clustering and classification procedure (Figure 21) has led to a deeper insight into the patterns with “Non-Typical”. The above mentioned clustering and classification provides the identification of similar developments over time. Several classes can be labeled e.g. typical communities, suburban communities, small city centers, cores of agglomeration or rural and remote communities. These spatial abstractions need a further explanation with other nonspatial or spatial attributes.

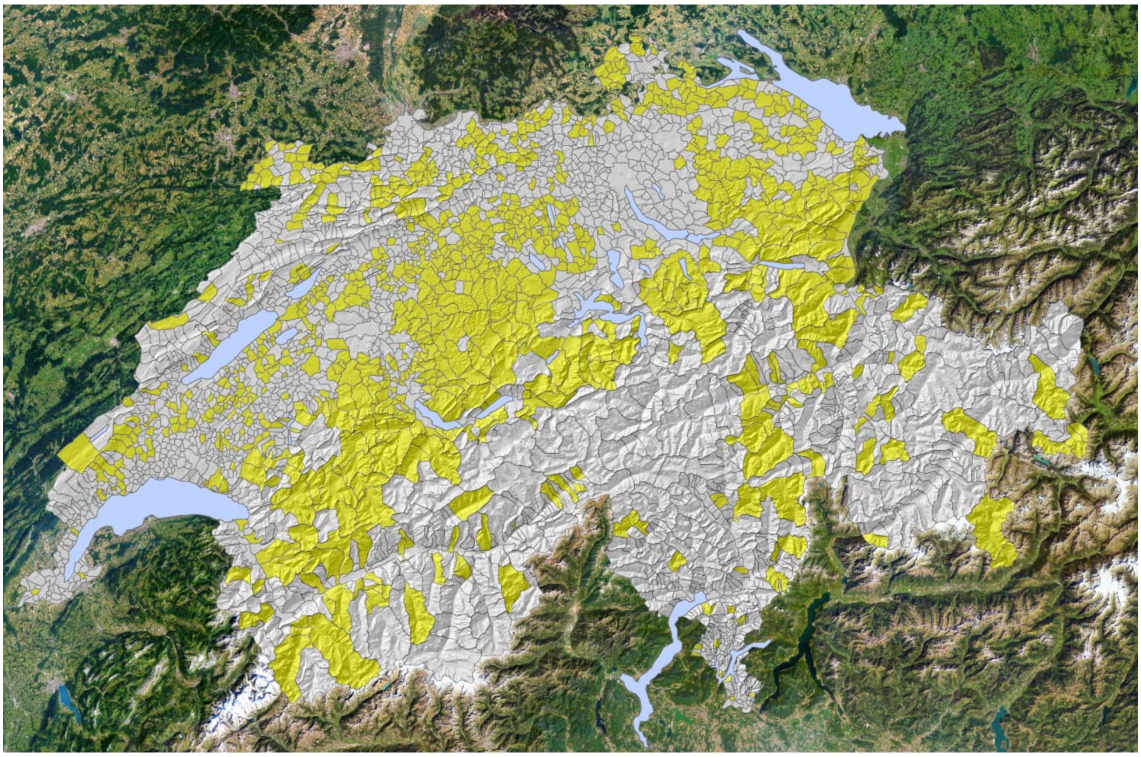
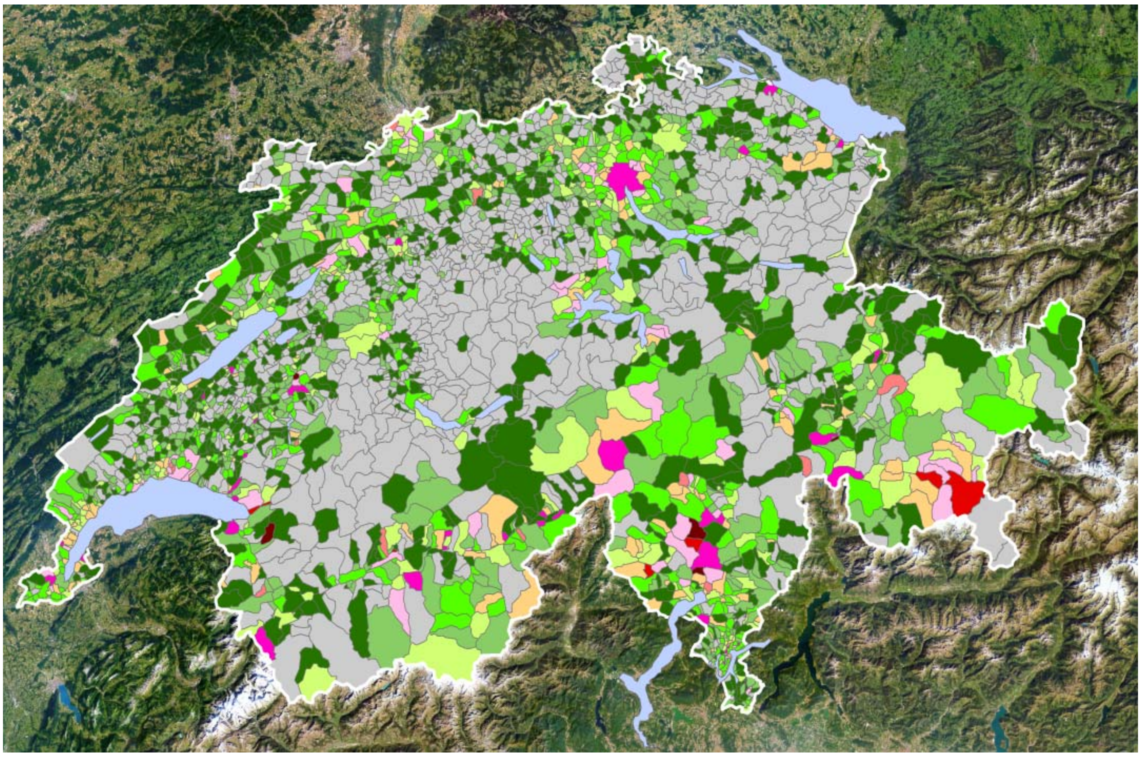


Figure 23: Map of 15 observed "Typical" per Pattern in 15 Decades (total = 852)



non-typical:	0	1	2	3	4	5	6	7	8	9	10
frequency:	852	675	511	359	238	131	70	36	13	6	5 = 2896
cumulative:	29.42	52.73	70.37	82.77	90.99	95.51	97.93	99.17	99.62	99.83	=100%

Figure 24: Map of observed "Non-Typical" per pattern in 15 decades (total = 2044)

7 Explanation of Classified Patterns

It is possible to integrate contingency tables to understand the observed classes in a complementary form. Such procedure supports the finding of class explanations based on well-known typologies. At the beginning a short description of relevant typologies provides the general understanding of typologies and spatial content. Later on a comparison is made on contingency tables in order to decide whether or not dependencies are present.

7.1 Administrative Spatial Typologies

Swiss spatial representations and administrative analysis are usually based on different spatial units and structures, so that different phenomena depending on subject and frequency may be presented in adequate manner. Spatial structures can be distinct into regionalization and typologies. Regionalization implies structuring the national territory in spatial coherent subareas (with the exception of enclaves), whereas typologies are summarizations of spatial discontinuous, but similar structured subareas. Spatial structures and typologies, which are currently most used in statistics of Switzerland, are the following (Schuler et al., 2005):

- Institutional structures: communities, counties, cantons.
- Regional political structures: areas of spatial planning, IHG-Regions, areas of economical regeneration.
- Regions of analysis: greater regions, language areas, MS-Regions, agglomerations and metropolitan areas.
- Spatial Typologies: type of community (according to Centre-Peripheral Model), typology of the MS-Regions (MS = Mobilité spatiale).
- Typology of height (e.g. Schmidt, 1969, Walter/Breckle, 1991).

The about 3000 communities of Switzerland have been classified according to a Centre-Peripheral-Concept and diverse criteria into 22 types of communities, which even have been combined in 9 main types (Table 6, Table 7). The spatial typology of communities has been developed about 20 years ago by a team of researchers on the ETH Lausanne ordered by BFS. The updating after the confederate census in 1990 and 2000 generated adjustments to the changed economical and social reality, but also confirmed the chosen method. The typology uses variables in conjunction with employment (commuter activities, ratio of employees, economic sector) development, tax revenue (income of the direct federal tax), tourism (room nights), structure of population as well as central functions. The assignment of a community to a determined type in 1980 also was dependent of the kind of MS-Region to which it belonged.

Table 6: Spatial Typology of Communities (Centre-Peripheral-Concept, 22 types)

Key	Label	Main Criteria	Communities	
1	big Centres (Grosszentren)	main town of a MS-Region, population >300.000	5	0.17%
2	medium Centres (Mittelzentren)	main town of a MS-Region, population >14.000	22	0.76%
3	small Centres Kleinzentren)	main town of a MS-Region, population >7.000	42	1.45%
4	peripheral Centres (Peripheriezentren)	main town, MS-Region, population <7.000, not located in agglomeration	27	0.93%
5	communities with high-income (Einkommensstarke Gemeinden)	located in an agglomeration zone, tax income >800 SFR/inhabitant	88	3.04%
6	touristic communities (touristische Gemeinden)	not located in an agglomeration zone and dependent on population and accommodations/inhabitant	53	1.83%
7	semi-touristic communities (semitouristische Gemeinden)	not located in an agglomeration zone and dependent on population and accommodations/inhabitant	111	3.83%
8	homes and institutions (Gemeinden mit Heimen/ Institutionen")	collective households > 10%	39	1.35%
9	workplaces, of metropolitan region (Arbeitsplatzgemeinden metropolitaner Regionen)	dependent on population and percentage of jobs/employees	114	3.94%
10	suburban communities of metropolitan regions (Suburbane Gemeinden metropolitaner Regionen)	dependent on population and percentage of apartment houses	72	2.49%
11	peri-urban communities of metropolitan region (Periurbane Gemeinden metropolitaner Regionen)	located in an agglomeration zone	245	8.46%
12	workplace communities of non-metropolitan region (Arbeitsplatzgemeinden nicht-metropolitaner Regionen)	dependent on population and percentage of jobs/employees	93	3.21%
13	suburban communities of non-metropolitan region (Suburbane Gemeinden nicht-metropolitaner Regionen)	dependent on population and percentage of apartment houses	53	1.83%
14	peri-urban communities of non-metropolitan region (Periurbane Gemeinde nicht-metropolitaner Region)	located in an agglomeration zone	219	7.56%
15	commuters and high immigration (Wegpendlergemeinden mit hoher Zuwanderung)	born people in that community <35.5%	367	12.67%
16	commuters and low immigration (Wegpendlergemeinde mit geringer Zuwanderung)	born people in that community >35.5%	265	9.15%
17	industrial tertiary communities (Industriell-tertiäre Gemeinden)	primary sector <9%, secondary sector <38%	177	6.11%
18	industrial communities (Industrielle Gemeinden)	primary sector <9%, secondary sector >26	106	3.66%
19	agrarian-industrial communities (Agrar-industrielle Gemeinden)	importance of primary sector is middle	216	7.46%
20	agrarian-tertiary communities (Agrar-tertiäre Gemeinden)	primary sector: 9-23.5%, tertiary double as secondary sector	278	9.60%
21	agrarian communities (Agrarische Gemeinden)	primary sector >23.5%	233	8.05%
22	communities with decline in population (Gemeinden mit starkem Bevölkerungsrückgang)	elderly people >28% or Population development 1970-2000: + 60%	71	2.45%

Table 7: Spatial Typology of Communities (Centre-Peripheral-Concept, 9 types)

Key	Label	Main Criteria	Communities	
1	Centres (1, 2, 3) (Zentren)	main town of a MS-Region, population >7.000	69	2.38%
2	Suburban Communities (9,10,12,13) (Suburbane Gemeinden)	dependent on population and percentage of apartment houses	332	11.46%
3	Communities with high-income (5) (Einkommensstarke Gemeinden)	located in an agglomeration zone, tax income >800 SFR/inhabitant	88	3.04%
4	Peri-urban Communities (11+14) (Periuburbane Gemeinden)	located in an agglomeration zone	464	16.02%
5	Touristic Communities (6+7) (touristische Gemeinden)	not located in an agglomeration zone and dependent on population and accommodations/inhabitant	164	5.66%
6	Industrial and tertiary Communities (4,8,17,18) (Industrielle und tertiäre Gemeinden)	dependent on percentage of primary and secondary sector	349	12.05%
7	Rural Communities with commuters (15+16) (Ländliche Pendlergemeinden	dependent on percentage of workplace, population and commuters	632	21.8%
8	Agrarian-mixed Communities (19+20) (Agrar-gemischte Gemeinden)	dependent on percentage of primary and secondary sector	494	17.06%
9	Agrarian Communities (21+22) (Agrarische Gemeinden)	primary sector >23.5%	304	10.50%

In Switzerland exist currently 50 agglomeration areas with their allocated core (=Kernstadt) and other communities which belong to the agglomeration. An agglomeration area is defined by a population of minimal 20.000. There are 5 communities defined as cities, which don't belong to an agglomeration area, these are called 'isolated cities' and are mainly located in dead end valleys (Table 8). The highest and the lowest point of Switzerland are located between more than 4000 m vertical height. Depending on altitude Switzerland is divided into 5 classes (Table 9).

Table 8: Typology of urban/rural regions depending on state of urbanization

Key	Class Label	German Translation	Communities	
1	core of agglomeration	„Kernstadt einer Agglomeration“	64	2.21%
2	(other) communities in agglomeration zone	„Andere Agglomerationsgemeinde“	910	31.42%
3	isolated city	„Isolierte Stadt“	5	0.17%
4	rural community	„Ländliche Gemeinde“	1917	66.19%

Table 9: Typology of communities depending on height

Key	Class Label	German Translation	Height (NN)	Communities	
1	hill zone	Hügelstufe	bis 600 m	1713	59.15%
2	mountain zone	Bergstufe	bis 1200 m	1012	34.94%
3	subalpine zone	Untere Alpenstufe	bis 1800 m	164	5.66%
4	alpine zone	Obere Alpenstufe	bis 2500 m	7	0.24%
5	snow zone	Schneestufe	ab 2500 m	0	0.00%

7.2 Class Explanation using Contingency tables

The data of the spatial typologies were examined according to significant properties using contingency tables. If the observed value is considerable greater than the expected we can notice a positive significance. Otherwise if it is considerable less we state a negative significance. The discovered significance has been marked in the tables with green as positive and red as negative.

Class 1 (suburban communities)

This class is significantly characterized by suburban (16 of 25) communities. Many of them are workplace communities (7 of 25). It is significant that the most are located in the hill zone (21 of 25).

Class 2 (core communities)

This class represents communities in the core of an agglomeration zone. It is positive significant that the communities of this class are in the core of an agglomeration. There are 7 of 10 communities with this property. These core communities can be further characterized as big or medium centres as indicated by significance.

Class 3 (communities in agglomeration zone, Medium centres)

The communities of class 3 are mainly located in an agglomeration zone. A significance is observed for the medium centres. Another significant feature for some communities is the characteristic as peripheral centres. There are significantly less rural communities in this class than expected. Most of the communities are located in the Midland (19 of 35).

Class 4 (peri-urban communities)

In class 4 we find the most communities in the suburban and peri-urban zone of an agglomeration. These communities are often so called workplace communities. Another significant feature of the communities of this class is high income. The class is significantly characterized by commuters and immigrants. Many rural but non agrarian communities belong to this class, but the number is significant less than expected. It is significant that 667 of 888 communities are to find in the hill zone. It is also observed that most of the communities are located in the western part of Switzerland, which includes the Région Lémanique, the Midland and northwest of Switzerland.

Class 5 (rich communities)

The communities of this class belong to an agglomeration zone (7 of 8). A significant feature of these communities is high income. Some can also be characterized as touristic. It can also be stated that there is a negative significance for rural communities.

Class 6 (suburban communities)

Some communities of class 6 are located in the core zone and can be identified as small centres. The most communities (135 of 225) are located in the suburban and peri-urban zone. This is double of the expected value. It is significant that some communities are characterized as workplaces. One third of the communities are rural communities, but this feature has a negative significance. Most of the communities are located in the hill zone (166 of 225). In the mountain zone are fewer communities than expected.

Class 7 (suburban workplace communities)

Most of the communities (45 of 55) are located in the suburban zone of an agglomeration. Some of them (5 of 55) can be identified as touristic communities. There is a high significance for workplace communities (21 of 55). The expected value of commuters could not be confirmed. Also was not confirmed that the expected value of rural communities could be found in this class. It is significant that most of the communities are located in the hill zone (43 of 55). In the mountain zone there are less than half of the expected communities to find. The most communities are located in the Région Lémanique, in the Midland and significantly in the Zurich area.

Class 8 (typical Swiss communities)

Most of the communities (1481 of 1671) can be identified as rural communities. It is surprising that the communities are not mainly located in the hill zone as expected. But it is obvious, that the communities in this class have a positive significance in higher areas of the Midland and Eastern Switzerland compared to all other classes. Most of the communities a significance in the field of industrial-tertiary, agrarian-mixed and agrarian economics can be identified.

The following tables present the results of the analysis by contingency tables. Maps of Switzerland are shown additionally to deepen the understanding of the spatial distribution and localization of administrative or other already established community classifications in Switzerland.

Table 10: Typology of urban/rural regions compared to the 8 classes

1= core of agglomeration, 2= communities in agglomeration zone, 3= isolated city, 4= rural community

	1	2	3	4	Row	%
1	7	16	0	2	25	0,86
expected	0,55	7,86	0,04	16,55		
% change	171%	68%	-200%	-157%		
2	7	2	0	1	10	0,35
expected	0,22	3,14	0,02	6,62		
% change	188%	-44%	-200%	-148%		
3	10	11	1	13	35	1,21
expected	0,77	11,00	0,06	23,17		
% change	171%	0%	177%	-56%		
4	7	518	1	342	868	29,97
expected	19,18	272,75	1,50	574,57		
% change	-93%	62%	-40%	-51%		
5	1	7	0	0	8	0,28
expected	0,18	2,51	0,01	5,30		
% change	140%	94%	-200%	-200%		
6	14	135	2	73	224	7,73
expected	4,95	70,39	0,39	148,28		
% change	96%	63%	135%	-68%		
7	5	45	0	5	55	1,90
expected	1,22	17,28	0,09	36,41		
% change	122%	89%	-200%	-152%		
8	13	176	1	1481	1671	57,70
expected	36,93	525,08	2,88	1106,12		
% change	-96%	-100%	-97%	29%		
Column	64	910	5	1917	2896	100,00
%	2,21	31,42	0,17	66,20	100,00	

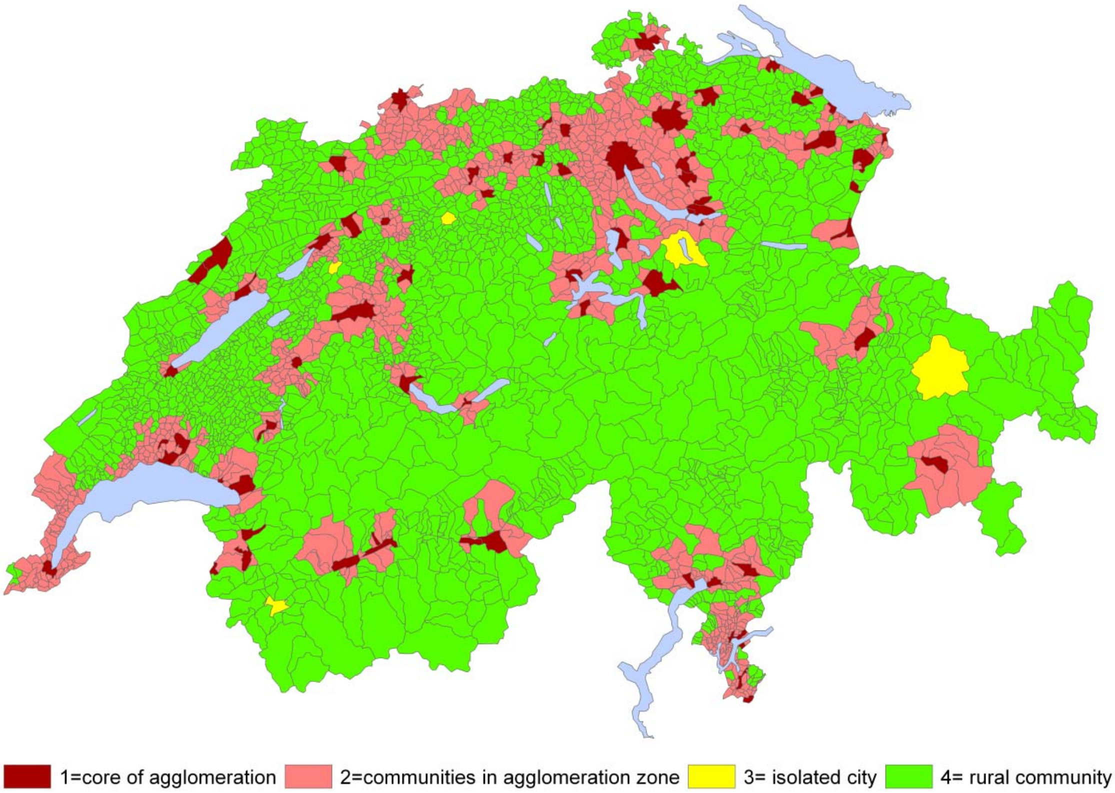


Figure 25: Localization of urban/rural regions in Switzerland

Table 11: Typology of height compared to the 8 classes

1= hill zone, 2= mountain zone, 3= subalpine zone, 4= alpine zone

	1	2	3	4	Row	%
1	21	3	0	1	25	0,86
expected	14,79	8,74	1,42	0,06		
% change	35%	-98%	-200%	177%		
2	8	2	0	0	10	0,35
expected	5,92	3,49	0,57	0,02		
% change	30%	-54%	-200%	-200%		
3	25	9	1	0	35	1,21
expected	20,70	12,23	1,98	0,08		
% change	19%	-30%	-66%	-200%		
4	667	173	28	0	868	29,97
expected	513,43	303,32	49,15	2,10		
% change	26%	-55%	-55%	-200%		
5	7	0	1	0	8	0,28
expected	4,73	2,80	0,45	0,02		
% change	39%	-200%	75%	-200%		
6	166	48	8	2	224	7,73
expected	132,50	78,28	12,69	0,54		
% change	22%	-48%	-45%	115%		
7	43	7	4	1	55	1,90
expected	32,53	19,22	3,11	0,13		
% change	28%	-93%	25%	153%		
8	776	770	122	3	1671	57,70
expected	988,41	583,93	94,63	4,04		
% change	-24%	27%	25%	-30%		
Column	1713	1012	164	7	2896	
%	59,15	34,95	5,66	0,24		100,00

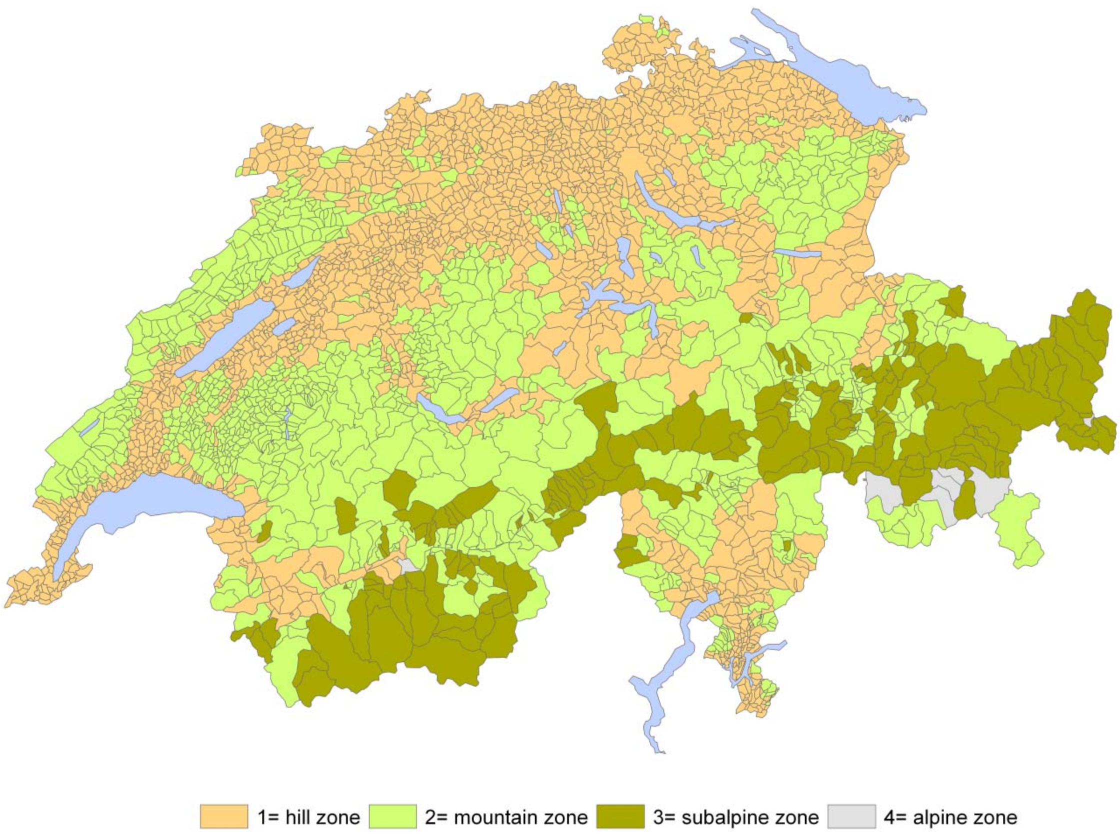


Figure 26: Localization of communities by height in Switzerland

Table 12: 9er Typology of Centre-Peripheral-Concept compared to the 8 classes

1= centre, 2= suburban, 3= community with high income, 4= peri-urban, 5=tourism, 6=industrial/tertian, 7=commuters, 8=agrarian mixed, 9=agrarian

	1=CEN	2=SUB	3=RE	4=PERI	5=TOUR	6=IND	7=PEND	8=MIX	9=AGR	Row	%
1	6	12	2	1	3	1	0	0	0	25	0,86
expected	0,60	2,87	0,76	4,01	1,42	3,01	5,46	4,26	2,62		
% change	164%	123%	90%	-120%	72%	-100%	-200%	-200%	-200%		
2	7	2	0	0	0	1	0	0	0	10	0,35
expected	0,24	1,15	0,30	1,60	0,57	1,21	2,18	1,71	1,05		
% change	187%	54%	-200%	-200%	-200%	-19%	-200%	-200%	-200%		
3	8	8	0	3	3	8	5	0	0	35	1,21
expected	0,83	4,01	1,06	5,61	1,98	4,22	7,64	5,97	3,67		
% change	162%	66%	-200%	-61%	41%	62%	-42%	-200%	-200%		
4	6	177	56	282	31	58	191	56	11	868	29,97
expected	20,68	99,51	26,38	139,07	49,15	104,60	189,42	148,06	91,11		
% change	-110%	56%	72%	68%	-45%	-57%	1%	-90%	-157%		
5	0	3	3	0	2	0	0	0	0	8	0,28
expected	0,19	0,92	0,24	1,28	0,45	0,96	1,75	1,36	0,84		
% change	-200%	106%	170%	-200%	126%	-200%	-200%	-200%	-200%		
6	16	63	14	52	8	24	37	9	1	224	7,73
expected	5,34	25,68	6,81	35,89	12,69	26,99	48,88	38,21	23,51		
% change	100%	84%	69%	37%	-45%	-12%	-28%	-124%	-184%		
7	5	30	2	10	5	2	1	0	0	55	1,90
expected	1,31	6,31	1,67	8,81	3,11	6,63	12,00	9,38	5,77		
% change	117%	131%	18%	13%	46%	-107%	-169%	-200%	-200%		
8	21	37	11	116	112	255	398	429	292	1671	57,70
expected	39,81	191,56	50,78	267,73	94,63	201,37	364,66	285,04	175,40		
% change	-62%	-135%	-129%	-79%	17%	24%	9%	40%	50%		
Column	69	332	88	464	164	349	632	494	304	**2896**	
%	2,38	11,46	3,04	16,02	5,66	12,05	21,82	17,06	10,50		**100,00**

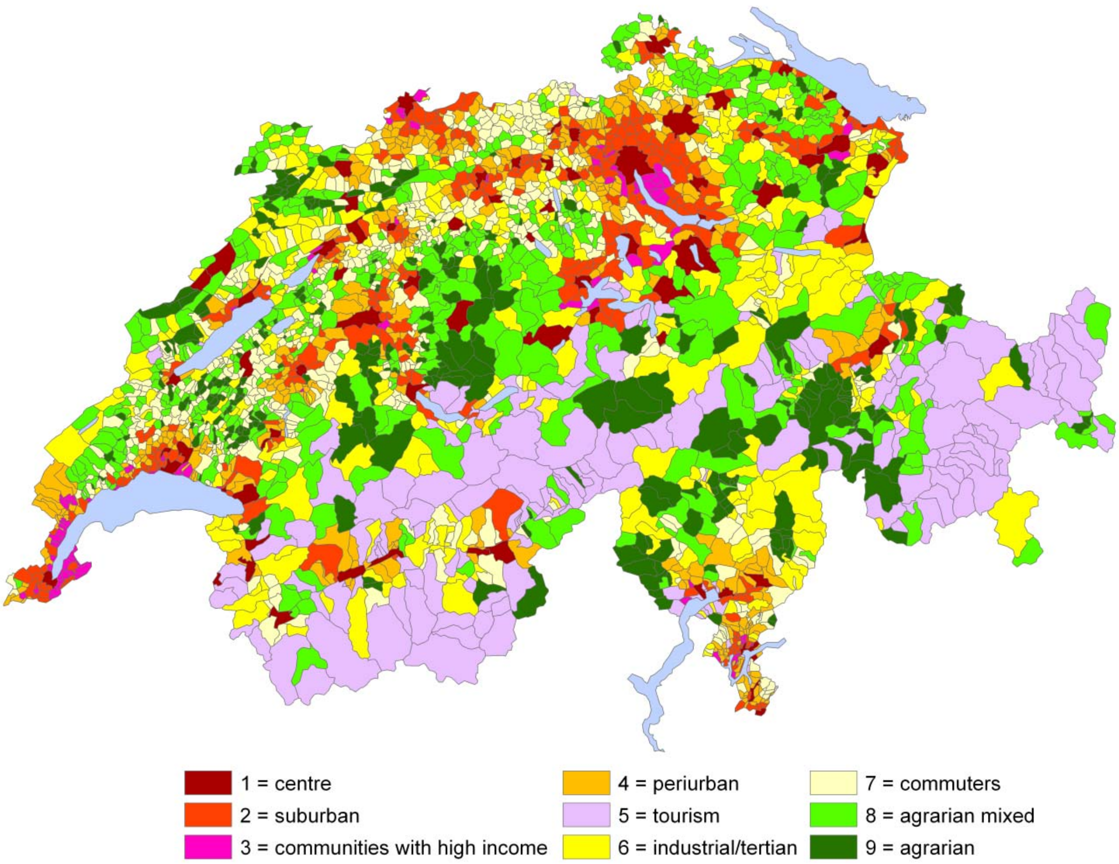


Figure 27: Localization of types of the Swiss Centre-Peripheral-Concept

Table 13: Nuts 2-Typology of 7 greater regions compared to the 8 classes

1= Région lémanique, 2= Espace Mittelland, 3= Nordwestschweiz,

4= Zürich, 5= Ostschweiz, 6= Zentralschweiz, 7= Ticino

	1	2	3	4	5	6	7	Row	%
1	4	5	4	1	8	1	2	25	0,86
expected	5,08	7,88	2,77	1,48	4,07	1,61	2,11		
% change	-24%	-45%	36%	-38%	65%	-46%	-6%		
2	2	4	0	1	2	1	0	10	0,35
expected	2,03	3,15	1,11	0,59	1,63	0,64	0,85		
% change	-2%	24%	-200%	51%	21%	44%	-200%		
3	1	19	1	2	6	1	5	35	1,21
expected	7,12	11,03	3,88	2,07	5,69	2,25	2,96		
% change	-151%	53%	-118%	-3%	5%	-77%	51%		
4	208	180	156	99	87	52	86	868	29,97
expected	176,53	273,65	96,21	51,25	141,17	55,75	73,43		
% change	16%	-41%	47%	64%	-47%	-7%	16%		
5	2	3	1	1	0	0	1	8	0,28
expected	1,63	2,52	0,89	0,47	1,30	0,51	0,68		
% change	21%	17%	12%	72%	-200%	-200%	39%		
6	48	83	20	12	26	8	27	224	7,73
expected	45,56	70,62	24,83	13,23	36,43	14,39	18,95		
% change	5%	16%	-22%	-10%	-33%	-57%	35%		
7	14	12	5	8	6	3	7	55	1,90
expected	11,19	17,34	6,10	3,25	8,95	3,53	4,65		
% change	22%	-36%	-20%	85%	-39%	-16%	40%		
8	310	607	134	47	336	120	117	1671	57,70
expected	339,85	526,80	185,21	98,67	271,77	107,32	141,36		
% change	-9%	14%	-32%	-71%	21%	11%	-19%		
Column	589	913	321	171	471	186	245	2896	
%	20,34	31,53	11,08	5,90	16,26	6,42	8,46		100,00

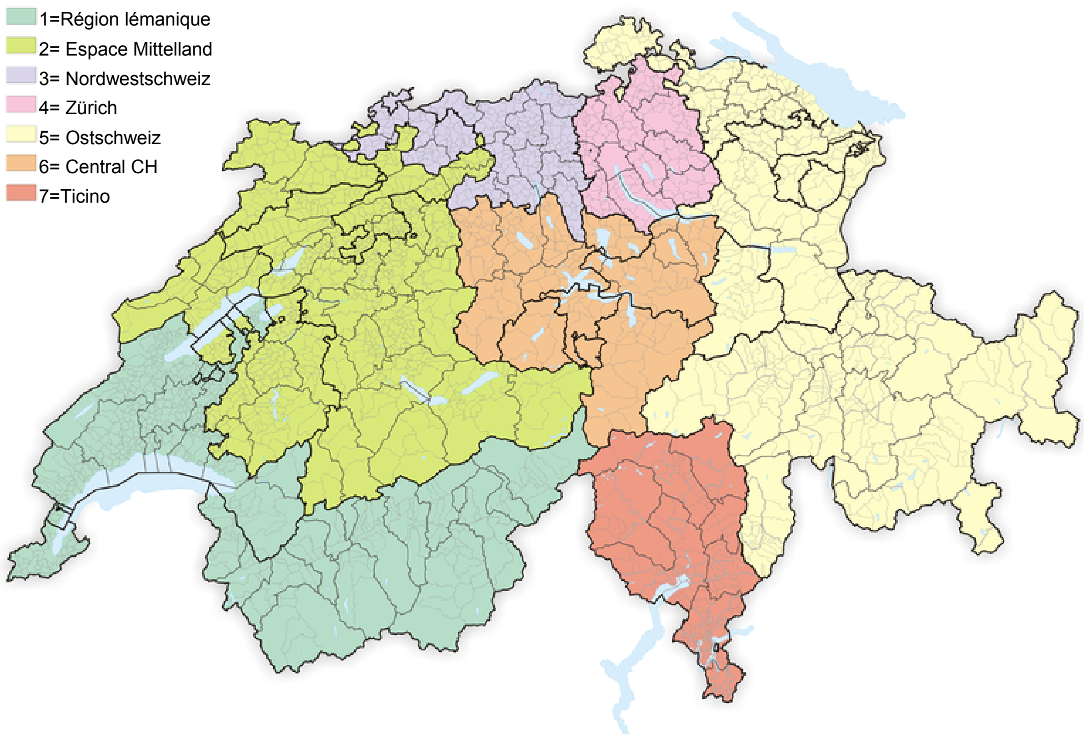


Figure 28: Greater Regions (Nuts 2 Regions) in Switzerland

1 – big centre
2 – medium centre
3 – small centre
4 – peripheral centre
5 – high income
6 – touristic
7 – semi-touristic
8 – homes /institutions
9 – workplaces of a metropolitan region
10 – suburban of a metropolitan region
11 – peri-urban of a metropolitan region
12 – workplaces of non metropolitan region
13 – suburban of non metropolitan region
14 – peri-urban of non metropolitan region
15 – commuters, high immigration
16 – commuters, low immigration
17 – industrial tertiary
18 – industrial
19 – agrarian industrial
20 – agrarian tertiary
21 – agrarian
22 – decline in population

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Row	%
1	0	2	4	0	2	3	0	0	7	0	0	3	2	1	0	0	0	1	0	0	0	0	25	0,86
expected	0,04	0,19	0,36	0,23	0,76	0,46	0,96	0,34	0,98	0,62	2,11	0,80	0,46	1,89	3,17	2,29	1,53	0,92	1,86	2,40	2,01	0,61		
% change	-200%	165%	167%	-200%	90%	147%	-200%	-200%	151%	-200%	-200%	116%	126%	-62%	-200%	-200%	-200%	9%	-200%	-200%	-200%	-200%		
2	2	4	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	10	0,35
expected	0,02	0,08	0,15	0,09	0,30	0,18	0,38	0,13	0,39	0,25	0,85	0,32	0,18	0,76	1,27	0,92	0,61	0,37	0,75	0,96	0,80	0,25		
% change	197%	193%	149%	-200%	-200%	-200%	-200%	-200%	-200%	-200%	-200%	103%	138%	-200%	-200%	-200%	-200%	93%	-200%	-200%	-200%	-200%		
3	2	4	2	3	0	3	0	0	0	2	0	4	2	3	5	0	2	3	0	0	0	0	35	1,21
expected	0,06	0,27	0,51	0,33	1,06	0,64	1,34	0,47	1,38	0,87	2,96	1,12	0,64	2,65	4,44	3,20	2,14	1,28	2,61	3,36	2,82	0,86		
% change	188%	175%	119%	161%	-200%	130%	-200%	-200%	-200%	79%	-200%	112%	103%	13%	12%	-200%	-7%	80%	-200%	-200%	-200%	-200%		
4	0	0	6	4	56	11	20	3	65	53	181	37	22	101	152	39	38	13	11	45	9	2	868	29,97
expected	1,50	6,59	12,59	8,09	26,38	15,89	33,27	11,69	34,17	21,58	73,43	27,87	15,89	65,64	110,00	79,43	53,05	31,77	64,74	83,32	69,84	21,28		
% change	-200%	-200%	-71%	-68%	72%	-36%	-50%	-118%	62%	84%	85%	28%	32%	42%	32%	-68%	-33%	-84%	-142%	-60%	-154%	-166%		
5	0	0	0	0	3	2	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	8	0,28
expected	0,01	0,06	0,12	0,07	0,24	0,15	0,31	0,11	0,31	0,20	0,68	0,26	0,15	0,60	1,01	0,73	0,49	0,29	0,60	0,77	0,64	0,20		
% change	-200%	-200%	-200%	-200%	170%	173%	-200%	-200%	104%	134%	-200%	-200%	149%	-200%	-200%	-200%	-200%	-200%	-200%	-200%	-200%	-200%		
6	0	5	11	3	14	4	4	5	25	9	18	15	14	34	31	6	10	6	0	9	1	0	224	7,73
expected	0,39	1,70	3,25	2,09	6,81	4,10	8,59	3,02	8,82	5,57	18,95	7,19	4,10	16,94	28,39	20,50	13,69	8,20	16,71	21,50	18,02	5,49		
% change	-200%	98%	109%	36%	69%	-2%	-73%	49%	96%	47%	-5%	70%	109%	67%	9%	-109%	-31%	-31%	-200%	-82%	-179%	-200%		
7	0	1	4	0	2	5	0	0	11	5	2	10	4	8	1	0	1	1	0	0	0	0	55	1,90
expected	0,09	0,42	0,80	0,51	1,67	1,01	2,11	0,74	2,17	1,37	4,65	1,77	1,01	4,16	6,97	5,03	3,36	2,01	4,10	5,28	4,43	1,35		
% change	-200%	82%	133%	-200%	18%	133%	-200%	-200%	134%	114%	-80%	140%	120%	63%	-150%	-200%	-108%	-67%	-200%	-200%	-200%	-200%		
8	1	6	14	17	11	25	87	31	5	2	44	23	7	72	178	220	126	81	205	224	223	69	1671	57,70
expected	2,88	12,69	24,23	15,58	50,78	30,58	64,05	22,50	65,78	41,54	141,36	53,66	30,58	126,36	211,77	152,91	102,13	61,16	124,63	160,41	134,44	40,97		
% change	-97%	-72%	-54%	9%	-129%	-20%	30%	32%	-172%	-182%	-105%	-80%	-125%	-55%	-17%	36%	21%	28%	49%	33%	50%	51%		
Column	5	22	42	27	88	53	111	39	114	72	245	93	53	219	367	265	177	106	216	278	233	71	2896	
%	0,17	0,76	1,45	0,93	3,04	1,83	3,83	1,35	3,94	2,49	8,46	3,21	1,83	7,56	12,67	9,15	6,11	3,66	7,46	9,60	8,05	2,45		100,00

8 Discussion

In the presented long-term analysis, a classification and clustering procedure was applied to community data (for 2896 communities). The author expects that in the future the concept of Data Mining in connection with Knowledge-Discovery techniques will become increasing important for geographical information science (GISc), urban research and future planning processes. The issue for this long-term analysis is to apply typical Data Mining techniques and methods to validate the process and outcomes.

The author has assumed at the beginning that Swiss communities allow a grouping process based on processes of population decline, stagnation or increase. The idea was to identify specific development profiles by using characteristics of all 15 decades. The presented long-term analysis provides the ability to identify patterns within a large amount of Swiss community data.

First, several indexes are discussed to decide about their properties. Typical indexes aim to compare two values of different temporal states without knowing everything in-between. As an alternative to relative change calculation the author has used relative differences (RelDiff). This index is already in use in the field of Data Mining and Knowledge-Discovery (Ultsch, 2003 a). The thesis has shown that relative difference is superior to common relative percent change because the influence of outlier (extrem values) on relative percent difference is alliviated due to an symmetric and limited range. Other indexes are critically in consideration of normalizing by the empirical variance when values are not normal distributed. The results then are mainly influenced by extreme cases when distances (e.g. Euclidean Distance) are calculated.

It was then started to gain insight into the distribution of population change by decade. For this purpose expected characteristics of population change by decade are assumed and discussed in this thesis. This step leads to a general assumption of the distribution of population change. A composite distribution is assumed when taking into account three characteristic developments: Losing Communities (e.g. multiplicative process → log-normal), Typical Communities (e.g. population is acting independently → CLT Theorem), Winning Communities (e.g. multiplicative process, growth → log-normal). Due to the assumed distribution of typical communities the Swiss trend of population development over time is observable. A mixture model is realized based on the distribution assumption as composite of a log-normal, normal, log-normal distribution.

The Expectation Maximization (EM-) algorithm is used in this thesis for parameter computation of the mixture distributions (Bilmes 1997). It should be mentioned that 'good' initial parameters are important as the algorithm only finds a local and not a global optimum. Several re-calculations are necessary to proof the intermediate results. The modeling process uses also Pareto Density Estimation (PDE; Ultsch, 2005) and probability density functions (PDF) to verify the Mixture. In addition to other common mixture models such technique ensures the modeling process. Q-Q-plots are finally used to proof the modeled distribution of "Typical" to a theoretical model.

The modeling process takes a standardization procedure into account by using a two-stage modeling procedure. This process realizes that mixture distributions by decade are different and need to be comparable. Thus the distribution of typical communities is first modeled as a Gaussian (normal) distribution. The detection of the mean and standard deviation are helpful to characterize the typical Swiss population change by decade (clinical thermometer). The proof by Pareto Density Estimation and Q-Q-Plot leads to the verification of the distribution assumption (Normal Distribution). The second modeling stage deals with the whole mixture model (Lognormal-Normal-Lognormal). It is to emphasize that the whole model is now based on standardized data. The mean and standard deviation of the first modeling process are the basis for the z-Transformation. The standardization provides the clustering processes and comparisons of decades. In consideration of the distribution of typical communities as a normal one, it is further known that the z-transformed data belongs to a standard normal distribution. The mean and standard deviation of typical communities provide to control the modeling process of the whole mixture model (M=0; S=1). The whole model is also proofed by Q-Q-plots. It is therefore verified that the distribution assumptions and the elaborated model are reasonable. It can be concluded that population change is described precisely using the mentioned composite distribution.

In comparison to other clustering approaches dealing with population data the pool of data is examined in depth in this thesis. In particular the variance and the properties of distance measurements in view of clustering are considered. The importance for the investigation of distributions is demonstrated. It has led to intermediate results that three categories by decade ("Loser", "Typical", and "Winner") characterize Swiss communities in one decade. The modeled distribution and determined parameters (Mean, Standard Deviation and amount of communities by distribution) provide the identification of the degree of membership to a specific category.

The posterior probability and component probability density function are the basis to compute the posterior probabilities using Bayesian theorem. This theorem offers advantages through its ability to formally incorporate prior knowledge into model specification via prior distributions and allows considering the variability. A specific category C is then predominant observed for a given value of population change (posterior ≥ 0.5) in a community. Each category by decade can be therefore used as an element of unique profiles over time (15 decades). Due to the comprehensive modeling process it is now possible to get a first insight into each decade and the specific amount of communities by category. The approach supports the discovery of probably multiple and partly unsuspected profiles over time. For a better understanding and characterization of communities the term pattern is defined in this thesis to describe such profiles consisting of unique categories by decade. In comparison to typical clustering approaches of population data such approach allows a deeper understanding in advance of communities in one decade and also in a combined view to several decades.

Due to his long and intensive experience in Data Mining and Knowledge-Discovery the mentor of this thesis (Prof. Dr. A. Ultsch) suggests a function that supports the classification of patterns using decision rules (LTW: $[-2,2] \rightarrow [-1,1]$, (LTW=Loser, Typical, Winner). The function is valuable to detect directly a specific category. That means it forms categories related to the membership label ("Loser", "Typical", "Winner"). Based on results of Bayesian posterior probabilities and the decision rules a set of 15 categorical values (-1, 0, 1) allow a description of patterns. A classification process leads to the identification of the total number of patterns. At the beginning the author had the expectation that a number of approximately 1000 different patterns will arise. It was conceivable that in view of all 2896 communities and the characteristics of the distribution there will arise three or five large groups of a single pattern. It was further assumed that many patterns are describing only one or two communities. The size of this group was assumed by 300 communities and labeled in advance as outlier. With regard to the classification result such assumptions are mostly confirmed. In total 880 patterns are observed in Switzerland and this value is nearby the expectation of the author. According to the pattern frequency it is also true that a larger group of patterns is dominating. There is one pattern characterizing 852 communities and it shows always typical categorical values (15x "Typical"). This pattern is surprising because it is clearly representing the "Typical" Swiss population development over time.

However, there are also about 863 communities allocated to 775 patterns that are described by only 1 or 2 communities. The group of observed specific developments (“outlier”) is much bigger than expected. The author uses a pattern matrix to display all 15 characteristics in a combined view. It shows the frequency and variety of patterns in graphical form. Such measurement is possible due to the classification of patterns. In comparison to other approaches dealing with population data the description of decades based on patterns is helpful to get insight into existing developments. The author is interested to understand the data and patterns more precisely before clustering. Such approach is not just interested in the application of a so called “black-box” clustering procedure.

For this reason the author takes another comparison of expected and observed patterns into account. At first the author describes each pattern by the number of “Typical” and “Non-Typical”. Under the assumption of statistical independency of decades and a constant prior probability it is possible to model the frequency distribution of these patterns as a binomial distribution. The mixture model supports the determination of the prior probability and results by decade are now very useful for this procedure. The comparison of the expected pattern frequency by the model to the observed frequency identifies several differences. For example the observed mentioned “Typical” pattern (=852 communities) is above the expected frequency by factor 4. Further spatial analysis and techniques of Data Mining should be taken into account in the future to find explanations for this unexpected frequency and specific distribution in the Swiss space. This pattern might be valuable for other investigations of population as well as be interesting to deepen the investigation by using randomly subsets of this pattern. Another interesting aspect deals with patterns with one or zero “Non-Typical”. They are characterizing more than 50 percent of all Swiss communities. In context of geographical information science it should be further mentioned that there are also some other interesting techniques which are based on the assumption that some spatial areas having higher or lower values than being expected alone by chance. These techniques lead to the identification of local clusters where the values are above or below those of a random distribution in space (Anselin, 1995).

According to a clustering procedure the author aims to select relevant patterns for clustering. A pragmatic planning approach is to have a deeper look to the size of population. The question is therefore how many people have an impact on one pattern?

But in consideration of the long-term description of population change (1850 to 2000) it seems to be more informative to compute the overall mean of all 15 mean values by decade. This value is labeled in this thesis as the long-term impact per pattern and is also an interesting value in general for the description of other communities and regions over time. For the purpose of pattern selection another already developed technique of Data Mining is applied (Ultsch, 2001). It is a procedure of information optimization based on the famous Pareto principle. The number of patterns and the values of the long-term population impact are used in this manner. From this it is concluded that in order to gain different patterns only about 14% of the patterns, the 14% relevant ones, should be examined in deep detail. The assumption of the author was that the minimal value of the population impact on a pattern is within the range of 2000 to 10,000. The observed value is 5000 per pattern. The author has concluded that clustering should be based on patterns showing either a disproportional frequency (clear deviation of observed and expected frequencies; binomial model) or a meaningful amount of population per pattern. Finally 122 patterns are selected and declared as a relevant pattern in Switzerland. It is to remark that 65% of all 2896 communities belong to these relevant patterns and about 85 percent of population. The procedure of information optimization is generally a good technique in context of further spatial investigations. Such technique allows deciding about a subset of data that might be particularly relevant for the description of a specific property or phenomena. However in this thesis it should be mentioned that the criteria that might be useful for such procedure of information optimization are multitude and thus the author decided in a first pragmatic way to choose the amount of population as a criterion of relevance. The author wants to suggest the use of other criteria (e.g. number of buildings, percentage of urbanized area per community) . Furthermore it might be helpful to compare the results and the number of relevant patterns.

For the purpose of interpretation and in particular for clustering of relevant patterns the author has defined three periodical subdivisions. These periods are identified based on knowledge about the general population growth and population development in Switzerland. The periods are as follows: period 1 (1850-1910, industrialization and urban growth), period 2 (1910-1950, World War I, II, subject of separation) and period 3 (1950-2000, urbanization, suburbanization, economic boom faces Switzerland). The use of a pattern matrix (122 patterns × 15 decades) allows a visual comparison and confirms such division in a visual way.

Nevertheless the author would suggest to detect some more reasons and explanation for this subdivision. The verification should be based on other multidimensional descriptions and on the general historical and societal Swiss context.

For clustering of relevant patterns the author has defined growth indicators for each period. The clustering is therefore based on three indicators per pattern. These indicators sum up the observed patterns by decade for each period. Large positive values indicate a Winner by period, values by zero indicate a typical development by period and large negative values by period indicate a Loser. At this point it should be mentioned that the data modeling of population change has led to the awareness about the quantities of patterns by decade. Thus it is already discovered that there are more “Winner” than “Loser” and the large amount of communities is described by “Typicals” by decade. Against this background it was not assumed that there are many growth indicators by period showing large negative values.

The clustering has required a decision about dissimilarity of patterns. Distance measurements are used in this thesis to quantify the proximity of patterns in a multidimensional feature space (3 periods $n \in \{1, 2, 3\}$ and specific additive values of patterns). A precise distinction of patterns is realized due to the properties of the growth indicator and related integer values. Such measurement provides the clustering by distance. The thesis has shown that other approaches are often influenced by large variances when using Euclidean distance. The influence of outliers (extreme values) on Euclidean Distance is alleviated due to the symmetric and limited range of the growth indicator. One advantage is that the values of the growth indicator are on similar scale and comparable. The similarity of patterns is handled by integer values. This indicator benefits from the intensive data inspection. It is based on the identified patterns and the translation to a semantic by decade. That means pattern are different from each other by a value of 1. Another advantage is furthermore that standardization is not necessary.

The use of Ward algorithm takes such developed properties better into account than other hierarchical algorithms. Generally it is an appropriate method for clustering. In this specific case the algorithm forms clusters according to the values of the growth indicator. The information loss refers to the inner and outer cluster differences and is supported by the scaled values of the growth indicator. When clustering the 122 relevant patterns the aimed expected result was a clear and compact cluster structure whereas a solution of 6 to 12 clusters were expected. It is conceivable when clustering 122 patterns based on information about population change that each cluster will contain about 10 patterns in average. The interpretation of the dendrogram has confirmed the expectation and has led to an eight cluster solution. These clusters are describing 1899 communities.

The aim of this thesis was to describe all 2896 communities according to their population development over time. The use of a k-Nearest Neighbor Classifier is based on the assumption that the 122 relevant patterns are point of reference for the whole Swiss population development. Thus the patterns can be used as training data. A k-Nearest Neighbor classifier is constructed to determine a class membership in particular for the rest of 997 unlabeled Swiss communities. The classifier provides the identification of the nearest neighbor that has already class information to allocate an unlabeled community.

It might be possible to compare the results of the k-Nearest Neighbor classification to other techniques in the field of Geographical Information Science. A comparison of quality of several spatial techniques is certainly of theoretical interest. For example it is interesting to compare the classification results of this thesis (2896 communities with a unique class membership) to results of spatial interpolation. The two spatial data sets (1899 communities / 997 communities) can be used for the integration of methods and allow future research.

In comparison to several former approaches in clustering and city classification the author of this thesis is additionally interested in class explanation. Knowledge conversion provides the transition from data to knowledge and generates several hypotheses (e.g. class descriptions and explanations) for further investigations. 2896 classified communities allow now a deeper explanation. The size of classes is described at first and also the impact of population per class. By using a weighted mean of the three growth indicators it was further possible to identify a typical community for each class. Based on the analyzed population data by decade (1850-2000) it can be stated that Solothurn is the typical community in Switzerland.

The localization of classes is used for the spatial verification and spatial reasoning. Several classes can be therefore characterized by additional labels such as suburban communities, small city centers, cores of agglomeration or rural and remote communities. The method of Gastner and Newman (2004) is the basis for the realization of scaled maps. The size of communities is therefore scaled in proportion to the amount of population. The high populated regions and classes are thereby emphasized.

A comparison of scaled and common maps allows a deeper description in the future. The maps in the appendix are a cartographical basis for the interpretation of periods and decades. It might be possible to discuss these results with other spatial planners and geographers to find further reasons for the development of communities.

Structure interpretation and spatial reasoning were also realized in this thesis by using contingency tables with the aim of identifying a significant number of explanatory characteristics. Such procedure is based on well-known typologies in Switzerland. The comparison is made on contingency tables in order to decide whether or not dependencies are significant.

At the end of this discussion it should be emphasized that specific classes should be investigated in detail by other structural and temporal parameters (e.g. age of the population, infrastructure, buildings, etc.).

9 Conclusion and perspectives for future steps

9.1 Achieved Targets

At the beginning of this thesis several questions were formulated and the achieved targets are shortly summarized.
The first question dealt with the number of patterns probably arising when analyzing the long-term development of population (1850-2000). Due to an intensive inspection and modeling of Swiss population data it was possible to identify 880 patterns according to their properties by 15 decades. Mixture modeling and the Bayesian theorem are basis for the classification of patterns by using defined decision rules. A pattern is described as unique profiles of categories over time. A set of 15 categorical values (-1, 0, 1) related to the category “Loser”, “Typical”, “Winner” provides the further interpretation of 2896 Swiss communities. A “Typical” represents the Swiss trend of population change by decade. All intermediate results are presented by decades in localized form.

The second question was interested in the number of relevant patterns of population development. To answer this question a two stage procedure was realized. At first the frequency of observed and expected values was analyzed. Under the assumption of statistical independency of decades and a constant probability it has been shown that it is possible to model the frequency distribution of patterns (“Typical”/”Non-Typical”) as a binomial distribution. The comparison of expected and observed values has led to the selection of patterns with a disproportional frequency. Secondly the patterns and their (long-term) mean value of population by 15 decades are used for information optimization. Such technique relates to the theoretical foundation of the Pareto 80/20-law (Ultsch, 2001). As a result 122 patterns are selected and described by 3 growth indicators (1850-1910, 1910-1950, and 1950-2000).

The third question asked for the number of clusters. By using the Euclidean Distance and Ward algorithm a clustering was realized. The result was a compact structure of eight clusters. The specific characteristics of Euclidean Distance are taken into account by using scaled growth indicators (symmetric and limited range, proximity of different patterns is 1). Later on a k-Nearest Neighbor classifier is used to allocate all Swiss communities to the partition of 8 classes. In view of the population change by 15 decades it can be further stated that the typical community is observed by “Solothurn”.

The localization provides the verification of results in a spatial abstraction and leads to a finer description of classes. Classification results are also presented using the long-term mean of population in proportion to the area of a community (Gastner and Newman, 2004). The explanation of the observed classes is realized on the basis of other already existing spatial typologies. Several significant relations are extracted on the basis of contingency tables.

In conclusion the application of methods from Data Mining and Knowledge Discovery are probably more and more important to develop a new generation of spatial analysis tools in a data-rich environment. Quantitative spatial investigations in general might lead to advanced strategic instruments such as semi or fully automated urban monitoring systems or a benchmark system for regional Swiss and European policy.

9.2 Perspectives for future Steps

Patterns of population change are different for the countryside near major cities, for metropolitan villages and for remote rural villages. In consideration of different patterns, it is possible to provide a data-based approach on spatial relations and neighborhoods (e.g. comparative qualities, interregional communication and cooperation). Swiss urbanized landscapes are highly dynamic, complex and multifunctional. Detailed inventories of landscape conditions and monitoring of change are urgently needed.

It is important to work with extracted knowledge when formulating strategies for the future development of Swiss communities. Therefore, a need for adjusted planning tools exists. A good base for the implementation of such tools is the spatiotemporal data exploration in a long-term perspective leading to specific details and explanations. The exploration of communities and other spatial lower scaled geospatial objects triggers discussions in the application domain and reveals insights about spatiotemporal phenomena and long-term processes. Furthermore, field investigation in selected areas should be conducted to obtain more reliable statistical data in space and time. In particular, long-term time courses serves as a basis for making decisions, as well as to control decisions that have been taken. To foster the understanding of the here presented classification other techniques are certainly of interest for validation and further explanation. Spatial outliers (Shekhar et al., 2003) might be of specific interest for further research. The modifiable unit problem (Openshaw, 1984) should be also taken into account to optimize the analysis of Swiss communities. At the end of this thesis some techniques and methods are briefly presented and discussed in view of future research.

9.2.1 Verification by Structure Visualization

The goal of clustering is to determine the intrinsic grouping in the set of data. But how to decide what constitutes a good clustering? Against this background the technique of an Emergent Self Organizing Map (Ultsch, 1999) is briefly presented in context of future perspectives in spatial analysis and geographical information sciences. Such technique would be an appropriate technique to verify the cluster structure of this thesis (computed by WARD algorithm). The aim is then to get visual insight into set of data. The power of self-organization allows the emergence of structure in data and supports its visualization, clustering and labeling concerning a combined distance and density-based approach. To visualize high-dimensional data, a projection from the high dimensional space onto two dimensions is needed (=planar map). This projection onto a grid of neurons is called a self-organizing map (SOM). There are two different SOM usages. The first are SOM, introduced by Kohonen (1982). Neurons are identified with clusters in the data space (k-means SOM) and there are very few neurons. The second are SOM where the map space is regarded as a tool for the visualization of the otherwise high-dimensional data space. These SOM consist of thousands or tens of thousand neurons. Such SOM allow the emergence of intrinsic structural features of the data space and therefore they are called emergent SOM (ESOM). The ESOM preserves the neighborhood relationships of the high-dimensional data and the weight vectors of the neurons are thought as a sampling point of the data. The U-Matrix has become the canonical tool for displaying the distance structures of the input data on ESOM. The P-Matrix takes density information into account. The combination of a U-Matrix and a P-Matrix leads to the U*-Matrix. On this U*-Matrix a structure in the data set can be detected directly. Figure 29 allows a comparison of both methods using the same data to see in whether there are cluster structures.

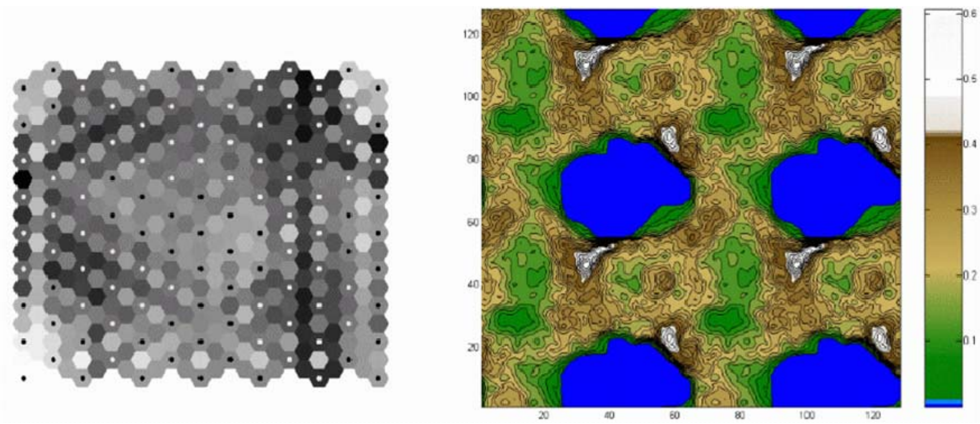


Figure 29: k-means SOM by Kaski et al. (2002) (left) and U*-Matrix (right)

The often-used finite grid as map has the disadvantage that neurons at the rim of the map have very different mapping qualities compared with neurons in the centre versus the border. This is important during the learning phase and structures the projection. In many applications important structures appear in the corner of such a planar map. Using ESOM has the advantage of a non-linear disentanglement of complex structures. The clustering of the ESOM can be performed at two different levels. The 'best match' visualization can be used to mark data points that represent a neuron with a defined characteristic (in this thesis clustering result by WARD). On the U*-Matrix the cluster structure in a set of data can be proofed and detected directly. Such visualization is used in tiled form to avoid border effects. Afterwards, a so-called island view is realized by mask to reduce redundancies which means each neuron is nearly visible at once. The corresponding U*-Map (island view) delivers a geographical landscape of the input data on a projected map (imaginary axis). The cluster boundaries are expressed by mountains, which means the value of height defines the distance between different patterns, which are displayed on the z-axis. A valley describes similar objects (e.g. communities or other spatial objects and according to this thesis patterns), characterized by small U-heights on the U*-Map. Objects found in coherent regions are assigned to one cluster. All local regions lying in the same cluster have nearly the same properties. The here presented U*-Map is an example for the possibilities of to integrate such method into the common geospatial analysis. It offers a proof by visualization of a given structure (e.g. hierarchical clustering results) and fosters a spatial abstraction.

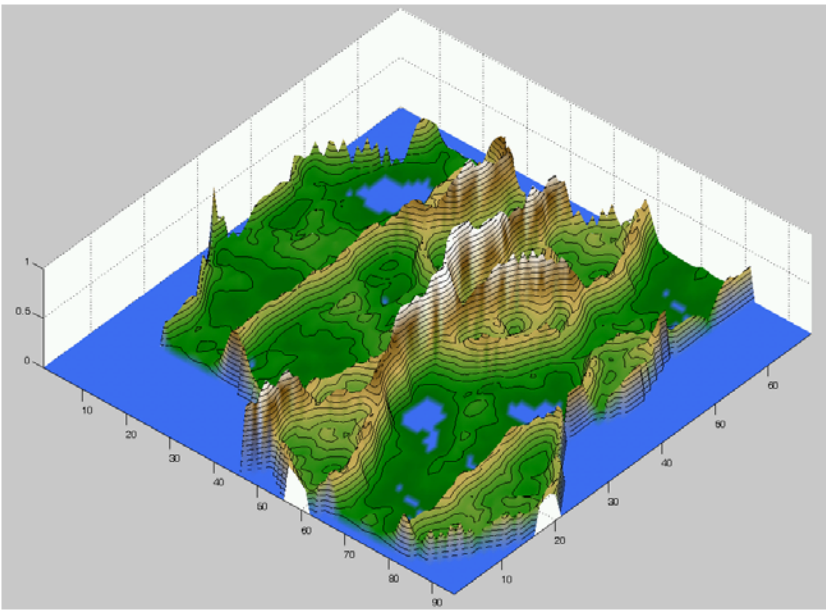


Figure 30: Example of an island view (U* -Map, see also Ultsch, 2005c).

9.2.2 Verification by Spatial Explanation

In spatial datasets „dependence is present in all directions and becomes weaker as data locations become more and more dispersed" (Cressie, 1993). Furthermore Tobler's 'First Law of Geography' is to keep in mind when analyzing spatial data (Tobler, 1979): „Everything is related to everything else, but near things are more related than distant things." Against this background it might be valuable to deepen the explanations of the elaborated classification of this thesis by other non-spatial and probably spatial data.

Other techniques might be useful for the ongoing explanation of the classification (longterm development of population 1850-2000).

The first option is to use indicators of spatial association. These are statistics that evaluate the existence of clusters in the spatial arrangement of a given variable. In case of extended data local clusters in the values mean that there are areas that have higher or lower values than is to be expected by chance alone; that is, the values occurring are above or below those of a random distribution in space. Local spatial autocorrelation analysis is based on the Local Moran LISA statistics (Anselin, 1995). This yields a measure of spatial autocorrelation for each individual location. A cluster map represents a special choropleth map showing those locations with a significant Local Moran statistics classified by type of spatial correlation. Conclusions depend on the significance level, and thus provides an informal mechanism to deal with multiple comparisons.

The second option is to develop regression models. Spatial Auto-Regression models (SAR) are an extension of the classical regression model for incorporating spatial dependence. They are popular for prediction and classification of spatial data. For example spatial contextual classification and prediction models for mining geospatial data should be taken into account (Shekhar, 2002). Geographically weighted regression (GWR) is testing for significance of spatial versus non-spatial effects (see Fotheringham et al.). It can combine spatial and non-spatial variables and can test the relative plausibility of models. Furthermore it can map various statistics (e.g., y-intercept, slope coefficients, standard errors, t-values, residuals, diagnostic test results) and provides the possibility for mapping variation within district sub-units of a larger space.

"Never concerned that the answer may prove disappointing, with pleasure and confidence we turn over each new stone to find unimagined strangeness leading on to more wonderful questions and mysteries – certainly a grand adventure." Richard P. Feynman (1988, S.243)

References

Abramowitz, Milton and Stegun, Irene A. (Eds.) (1972) Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. 9th printing, New York.

Arlt, G., Gössel, J., Heber, B., Hennersdorf, J., Lehmann, I. and Thinh, N.X. (2001) Auswirkungen städtischer Nutzungsstrukturen auf Bodenversiegelung und Bodenpreis. IÖR Schriften No. 34, Leibniz-Institut für ökologische Raumentwicklung e.V., Dresden.

Bähr, Jürgen (2004) Bevölkerungsgeographie. UTB, Stuttgart.

Bätzing, Werner and Dickhörner, Yven (2001) Die Typisierungen der Alpengemeinden nach 'Entwicklungsverlaufsklassen' für den Zeitraum 1870–1990. Mitteilungen der Fränkischen Geographischen Gesellschaft, 48, 273–303.

Behnisch, M. (2009) Urban Data Mining, Universitätsverlag, Karlsruhe.

Bilmes J (1997) A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report (http://crow.ee.washington.edu/people/bulyko/papers/em.pdf, 15.10.2009), University of Berkeley, ICSI-TR-97-021.

Bishop, Christopher (2006) Pattern Recognition and Machine Learning. New York.

Bortz, Jürgen (2005) Statistik für Human und Sozialwissenschaftler, Berlin.

Breiman, L.; Friedman, J.H.; Olshen, R.A. and Stone, C.J. (1984) Classification and Regression Trees, CRC Press, Boca Raton, FL.

Bundesamt für Statistik, BFS (Ed.) (2006) Szenarien zur Bevölkerungsentwicklung der Schweiz 2005-2050. Neuchâtel.

Burgdorf, Markus (2008) Verzerrungen von Raum und Wirklichkeit in der Bevölkerungsgeographie. In: Kartographische Nachrichten, Heft 5 /08, Bonn, pp. 234-242

Crow, Edwin L.; Shimizu, Kunio (1988) Lognormal Distributions: Theory and Applications, In: *Series Statistics: A Series of Textbooks and Monographs*, Vol. 88, New York.

Dean, K.G.; Brown, B.J.H.; Perry, R.W.and Shaw, D.P. (1984) The Conceptualisation of Counterurbanisation. In: *Area* Vol. 16 No. 1, pp 9-14.

Dempster, A.P.; Laird, N.M., and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. In: *Journal oft the Royal Statistical Society, Series B (Methodological)*, Vol. 39, No. 1, pp. 1-38.

Demsar, U. (2006) Data Mining of Geospatial Data: Combining Visual and Automatic Methods, Urban Planning Department, KTH Stockholm.

Dickhörner, Yven (2000) Typisierung der Bevölkerungsentwicklung in den Alpen auf Gemeindeebene. Magisterarbeit, Erlangen (Mentor: Prof. Dr. Werner Bätzing).

Diener, Roger; Herzog, Jaques; Meili, Marcel; de Meuron, Pierre; Schmid, Christian (2005) Die Schweiz - Ein städtebauliches Portrait. Zürich.

Feynman, Richard P. (1988): What do you care, What Other People Think? Norton & Company, New York.

Fortheringham, A. Steward (2002) Geographically Weighted Regression. The analysis of spatially varying relationships. Chichester.

Gastner, M.T. and Newman, M.E.J. (2004) Diffusion-based method for producing density equalizing maps. In: Proceedings of the NAS, Vol. 101, No. 20, pp. 7499-7504.

Gatzweiler, H.P., Meyer, K. and Milbert, A. (2003) Schrumpfende Städte in Deutschland? Fakten und Trends, in BBR (ed.): Informationen zur Raumentwicklung, Selbstverlag, Bonn, 10/11, pp. 557–574.

Han, Jiawei; Kamber, Micheline (2006) Data Mining – Concepts and Techniques. San Francisco CA.

Hand, D., Mannila, H. and Smyth, P. (2001) Principles of Data Mining, MIT Press, Cambridge, MA.

Hand, David J.; Yu, Keming (2001) Idiot's Bayes - not so stupid after all?, In: *International Statistical Review*. Vol. 69, No. 3, pp. 385-399.

Hardy, G. H. (1992) A Mathematician's Apology. Cambridge.

Harris, C.D. (1943) A functional classification of cities in the United States. Geographical Review, 33(1), 86–99.

Haug, W. (2002) Räumliche und strukturelle Bevölkerungsdynamik der Schweiz 1990-2000. Bundesamt für Statistik (Ed.), Neuchâtel.

Husa, Karl; Wohlschlägl, Helmut (2007) Proseminar Grundzüge der Bevölkerungsgeographie (Skriptum). Institut für Geographie und Regionalforschung (Hg.), Universität Wien.

Izenman, Alan Julian (2008) Modern Multivariate Statistical Techniques – Regression, Classification, and Manifold Learning. Berlin.

Kahsaia, Mulugeta, S.; Peter V. Schaeffer (2009) Deconcentration, Counter-urbanization, or Trend Reversal? The Population Distribution of Switzerland, Revisited Research Paper (published in: *Socio-Economic Planning Sciences*, Vol. 26, No. 2, pp. 89-102.

Kaski, S., Nikkilä , J., Törönen, P., Venna, J., Castrén, E. and Wong, G. (2002) Analysis and visualization of gene expression data using self-organizing maps. Neural Networks, Vol. 15 No. 8–9, 953–966.

Kohonen, T. (1982) Self-organizing formation of topologically correct feature maps. Biological Cybernetics, 43(1), 59–69.

MacEachren, A.M. (1994) Visualization in modern cartography: setting the agenda, in A.M. MacEachren, D.R.F. Taylor (eds): Visualization in Modern Cartography, Pergamon, Oxford, pp. 1–12.

MacEachren, A.M. and Kraak, M.-J. (1997) Exploratory cartographic visualization: Advancing the agenda", In: *Computes and Geosciences,* Vol. 23, pp. 335-343.

Manley, David; Flowerdew, Robin; Steel, David (2006) Scales, levels and processes: Studying spatial patterns of British census variables. Computers, Environment and Urban Systems, Volume 30, Nr. 2, pp. 143–160.

Meyer, Hannes (1978) Wandlungen im Bestande der Gemeinden. Dissertation, Universität Zürich.

Openshaw, S. (1984) The Modifiable Areal Unit Problem. Geo Books, Norwich.

Oswald, Franz; Baccini, Peter (2003): Netzstadt. Einführung in das Stadtentwerfen. Birkhäuser, Basel.

Pearl, J. (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufman Publishers.

Pearl, Judea (2000) Models: Reasoning, and Inference, Cambridge University Press.

Redner R, Walker H (1984) Mixture Densities, Maximum Likehood and the EM Algorithm, In: *SIAM Review*, Vol. 26, No. 2 (April 1984).

Rey, Urs (2003) Demografische Strukturveränderungen und Binnenwanderung in der Schweiz 1850-1950. Dissertation, Universität Zürich.

Romesburg, C.H. (1984) Cluster analysis for researchers. Wadsworth, London.

Rothenbacher, Franz (2002) The European Population 1850-1945, in the Series Societies of Europe. Macmillan Press, London.

Sander, J.; Ester, H.-P.; Xu, X. (1998) Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. In: *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, pp. 169-194.

Schaeffer, Peter, V. (1992) Deconcentration, Counter-Urbanization, or Trend Reversal? The population Distribution of Switzerland, 1900-1980. In: *Socio-Economic Planning Sciences*, Vol. 26, No. 2, pp. 89-102.

Schmidt, G. (1969) Vegetationsgeographie auf ökologisch-soziologischer Grundlage. Leipzig.

Schuler, Martin; Dessemontet, Pierre; Joye, Dominique (2005) Eidgenössische Volkszählung 2000 - Die Raumgliederungen der Schweiz. Bundesamt für Statistik (BfS), Neuchâtel.

Schuler, Martin; Ullmann, Dominik; Haug, Werner (Ed.) (2002) Eidgenössische Volkszählung 2000 - Bevölkerungsentwicklung der Gemeinden 1850-2000. Bundesamt für Statistik (released: 01.11.2005), Neuchâtel.

Schwarz, K. (1970) Maßzahlen zur Beurteilung der räumlichen Verteilung der Bevölkerung im Bundesgebiet, in *Wirtschaft und Statistik*, Heft 7, S. 337–342.

Shekhar, S., Lu, C.-T. and Zhang, P. (2003) A unified approach to detecting spatial outliers. GeoInformatica, Vol. 7, No. 2, pp. 139–166.

Shekhar, S.; Schrater, P.; Vatsavai, R.; Wu, W. and Chawla, S. (2002) Spatial Contextual Classification and Prediction Models for Mining Geospatial Data , IEEE Transactions on Multimedia (special issue on Multimedia Databases)

Siedentop, St., Kausch, St., Einig, K. and Gössel, J. (2003) Siedlungsstrukturelle Veränderungen im Umland der Agglomerationsräume, Heft 114, Bundesamt für Bauwesen und Raumordnung, Bonn.

Sieverts, Thomas (1997) Zwischenstadt zwischen Ort und Welt, Raum und Zeit, Stadt und Land. Basel.

Tappeiner, Ulrike; Tappeiner, Gottfried; Hilber, Andreas; Mattanovich, Ernst (2003) The EU Agricultural Policy and the Environment – Evaluation of the Alpine Region. European Academy EURAC (eds.), Oxford.

Thinh, Xuan, Nguyen (2004) Entwicklung von mathematisch-geoinformatischen Methoden und Modellen zur Analyse, Bewertung, Simulation und Entscheidungsunterstützung in Städebau und Stadtökologie. Habilitationsschrift der Universität Rostock, Fakultät der Agrar- und Umweltwissenschaften.

Tobler, Waldo (1970) A computer movie simulating urban growth in the Detroit region. In: *Economic Geography*, Vol. 46 Nr. 2, pages 234-240.

Tschopp, M., R. Sieber und K. W. Axhausen (2002) Demographie und Raum in der Schweiz: Ein historischer Abriss, Arbeitsberichte Verkehrs- und Raumplanung, 134, IVT, ETH Zürich.

Tschopp, Martin; Sieber, René; Keller, Peter; Axhausen, Kay W. (2003) Demografie und Raum in der Schweiz. Ein historischer Abriss, in *DISP 153*, pp. 25-32.

Ultsch, A. (2005c) U*C self-organized clustering with emergent feature map, in *Proceedings of Lernen, Wissensentdeckung und Adaptivität* (LWA/FGML 2005), Saarbrücken, Germany, pp. 240–246.

Ultsch, A.; Mörchen, F. (2005d) ESOM-Maps: tools for clustering, visualization, and classification with emergent som. Technical Report 46, Dept. of Mathematics and Computer Science, University of Marburg, Germany, 2005.

Ultsch, Alfred (1992). Self-organizing neural networks for visualization and classification. In: Opitz, Otto; Lausen, Berthold; Klar, Rüdiger (eds.): Information and Classification - Concepts, Methods, and Applications. Proceedings of the 16th Annual Conference of the GfKl, University of Dortmund, 1992.

Ultsch, Alfred (2003) Is log ratio a good value for identifying differential expressed genes in microarray experiments?, Technical Report No. 35, Dept. of Mathematics and Computer Science, University of Marburg, Germany.

Ultsch, Alfred (2003b) Maps for the visualization of high dimensional data spaces. In T. Yamakawa, editor, Proceedings of the 4th Workshop on Self-Organizing Maps (*WSOM'03*), pages 225–230.

Ultsch, Alfred (2005a) Improving the Identification of Differentially Expressed Genes in cDNA Microarray Experiments, In: Claus Weihs and Wolfgang Gaul: Classification − the Ubiquitous Challenge (Proceedings of the 28th Annual Conference of the Gesellschaft für Klassifikation e.V. University of Dortmund, March 9–11, 2004), Berlin, Heidelberg.

Ultsch, Alfred (2005b) Pareto Density Estimation: A Density Estimation for Knowledge Discovery, In: Daniel Baier and Klaus Dieter Wernecke: Innovations in Classification, Data Science, and Information Systems (Proceedings of the 27th Annual Conference of the Gesellschaft für Klassifikation e.V. Brandenburg University of Technology, Cottbus, March 12–14, 2003), Berlin, Heidelberg.

Ultsch, Alfred (2008) Is log ratio a good value for measuring return in stock investments?, in: Andreas Fink, Berthold Lausen, Wilfried Seidel and Alfred Ultsch (eds.) Advances in Data Analysis, Data Handling and Business Intelligence (Proceedings 32nd Annual Conference of the German Classification Society, Helmut-Schmidt-University, Hamburg, July 16-18, 2008), Berlin, Heidelberg.

Ultsch, Alfred (2009) Lecture Notes: Databionic − Knowledge Discovery, Dept. of Mathematics and Computer Science, University of Marburg, Germany.

United Nations (2009) World Population Prospects. Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, New

Vogel, Friedrich (1975) Probleme und Verfahren der numerischen Klassifikation: unter bes. Berücks. von Alternativmerkmalen. Göttingen.

Walter, Heinrich; Breckle, Siegmar-W, Eds. (1991) Ökologie der Erde, Band 1 - Ökologische Grundlagen in globaler Sicht. Stuttgart.

Wang, W.; Yang, J.; Muntz, R. (1997) A statistical information grid approach to spatial data mining. In: Proceedings 1997 International Conference. Very Large Data Bases (VLDB, 1997), Athens, Greece, August 1997, pp. 186-195.

Ward, Joe H. (1963) Hierarchical Grouping to optimize an objective function. In: *Journal of American Statistical Association*, Vol. 58, pp. 236-244.

Watkins, Kevin (2007) Human Development Report 2007/2008. United Nations Development Program, UNDP (ed.), New York.

Witthauer, K. (1969) Verteilung und Dynamik der Weltbevölkerung, in Petermanns Geographische Mitteilungen, Sonderheft 272, Gotha/Leipzig.

Woods, R. (1979) Population Analysis in Geography. London, New York.

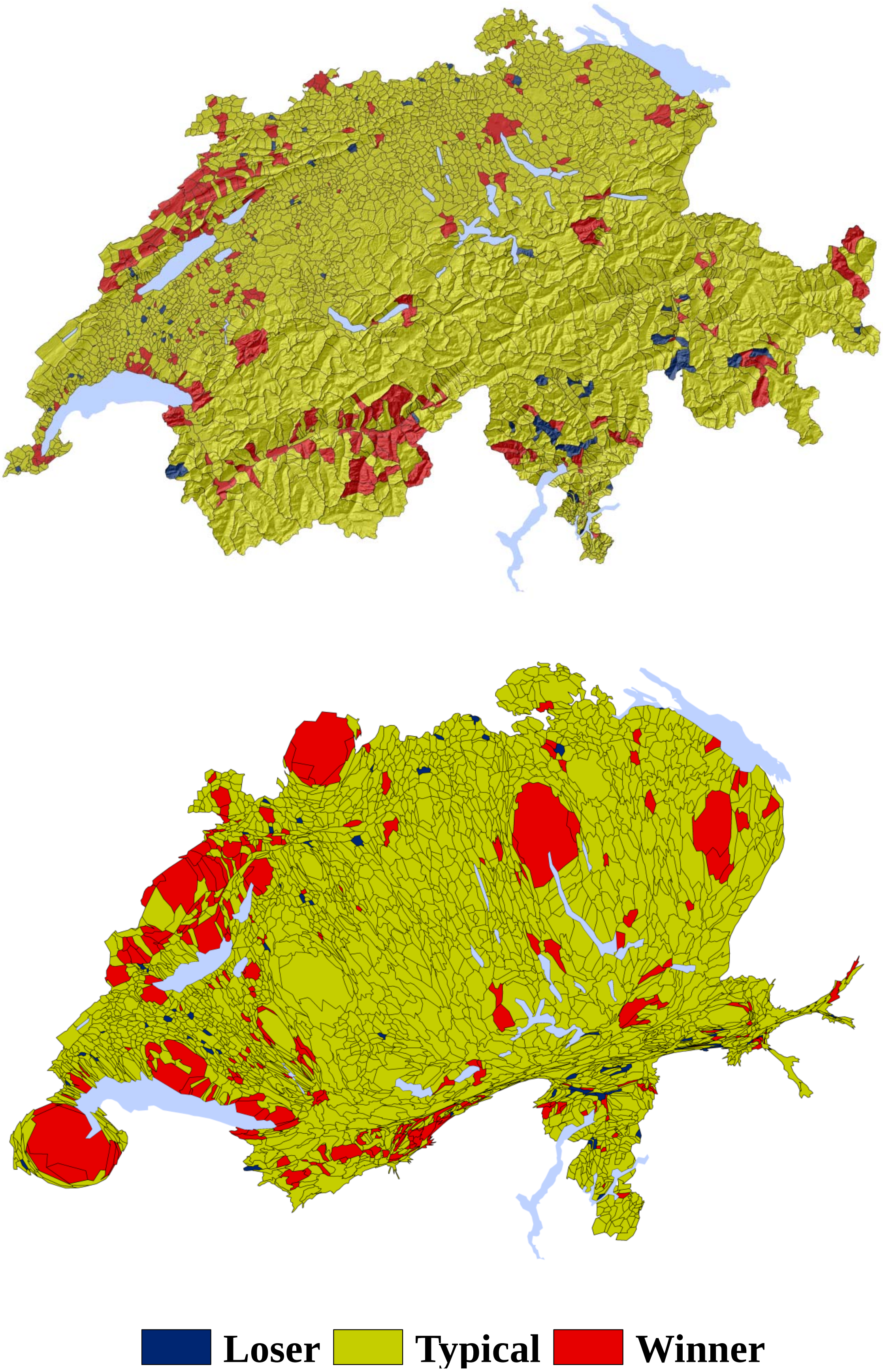
York, 2008 Online: http://esa.un.org/unup/ (20.05.2009).

Zhang, Harry (2005) Exploring Conditions for the Optimality of Naïve Bayes', In: *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI),* Vol. 19, No. 2, pp. 183-198.

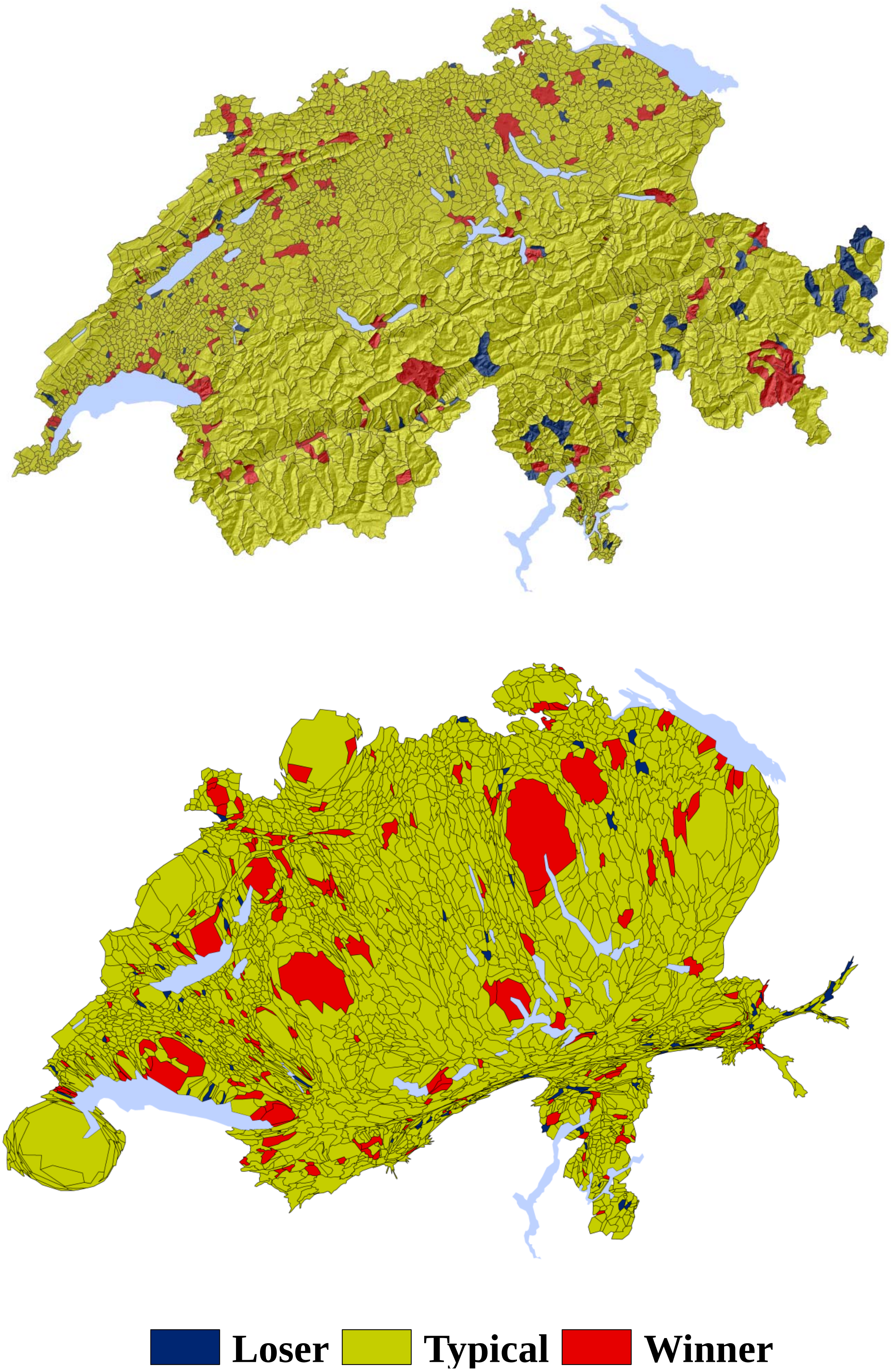
Appendix

Appendix 1: Cantons in Switzerland

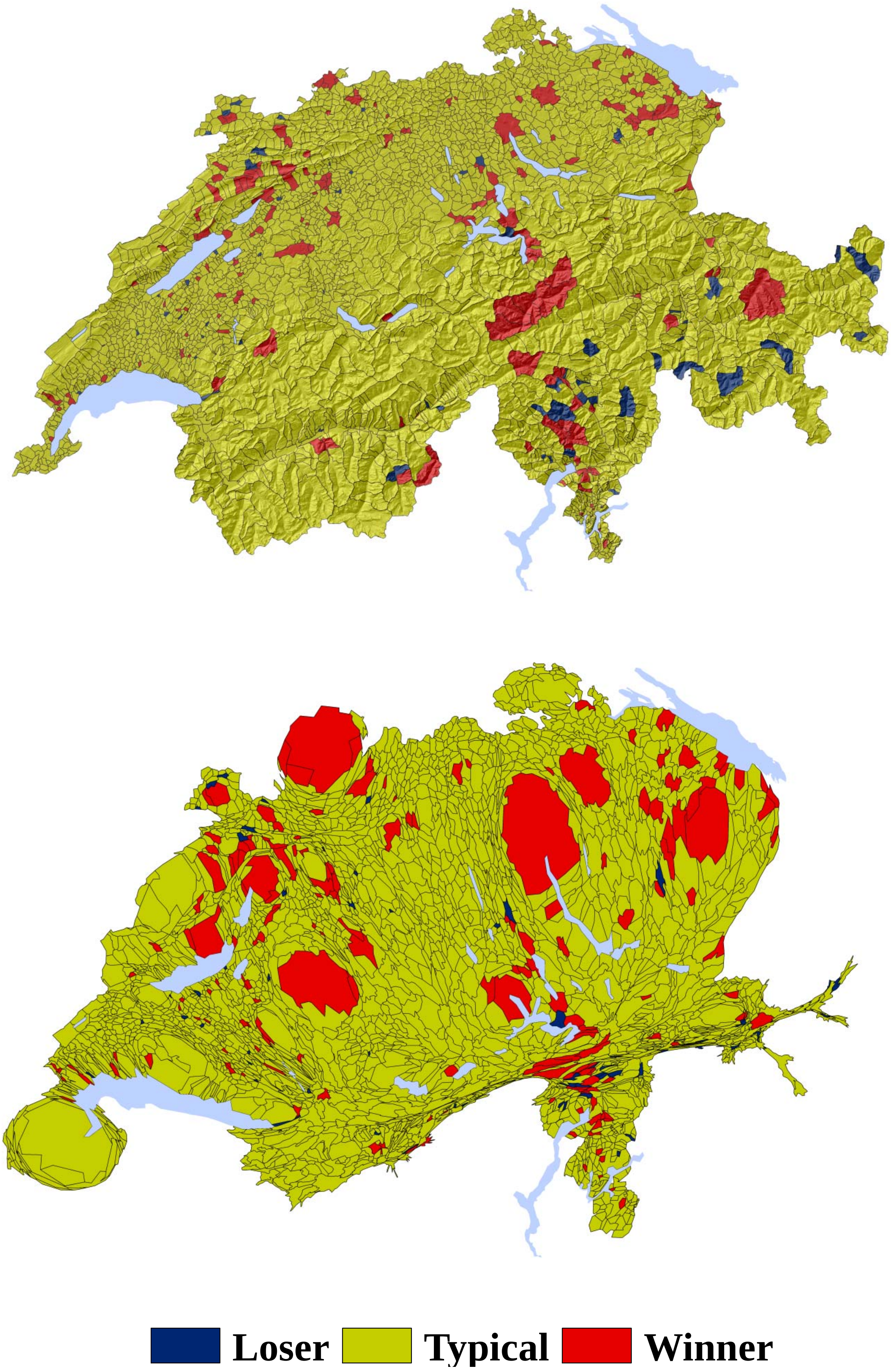
Appendix 2: Regions in Switzerland (=http://de.wikipedia.org/wiki/Schweiz, 01.01.2009)



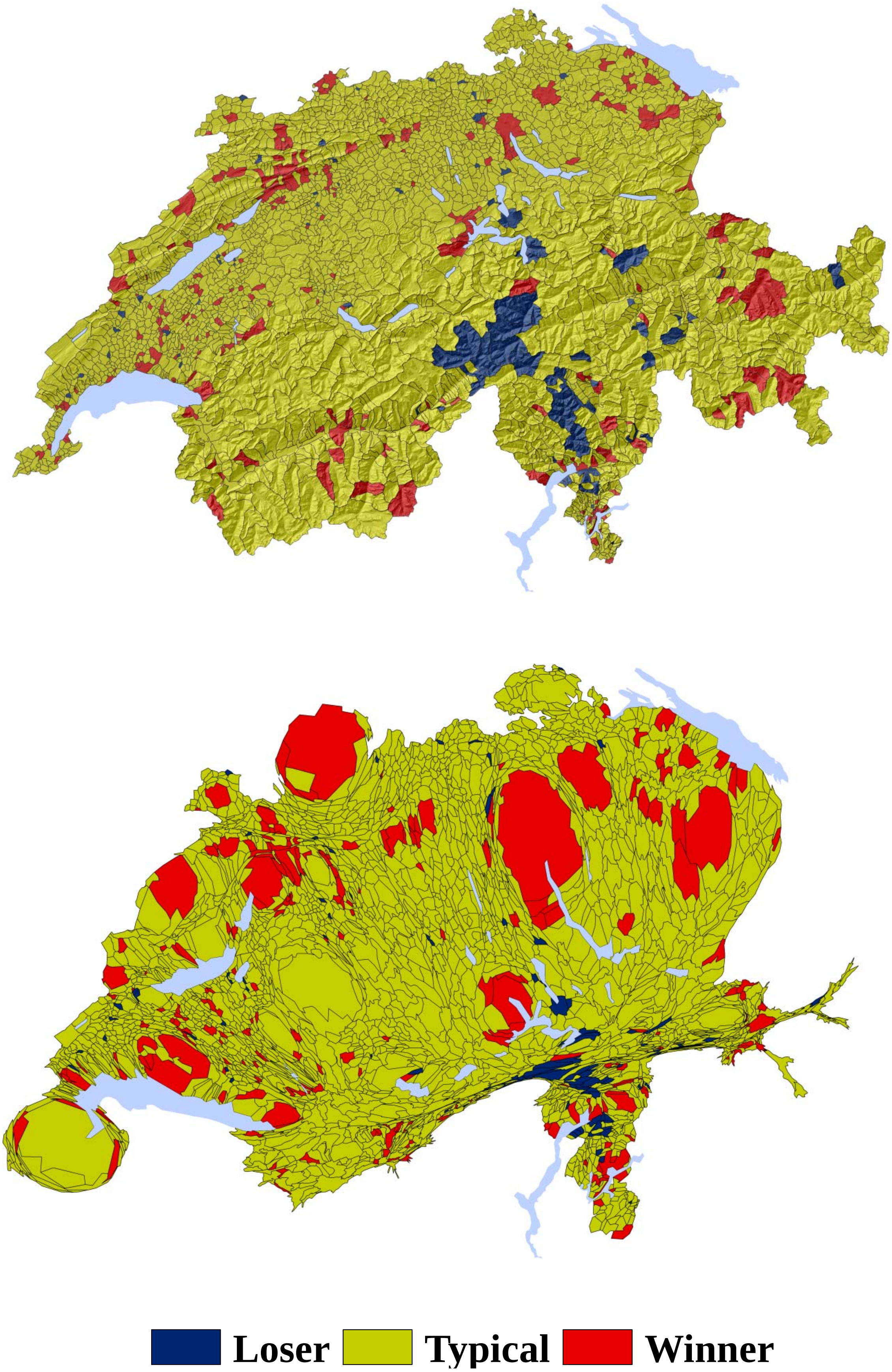
Appendix 3: Decade 1860: “Loser”, 61; “Typical”, 2578; “Winner”, 257 communities



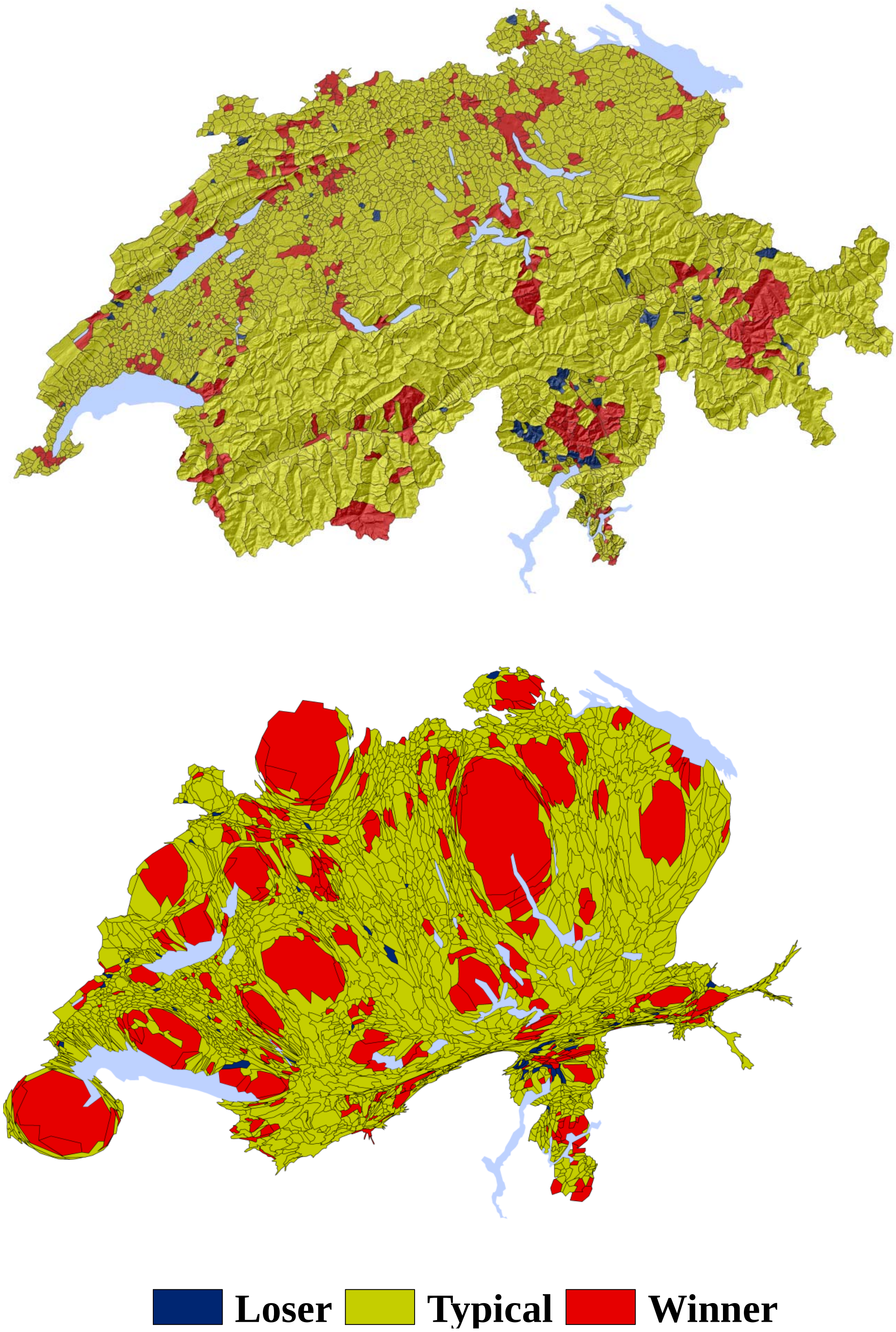
Appendix 4: Decade 1870: “Loser”, 71; “Typical”, 2628; “Winner”, 197 communities



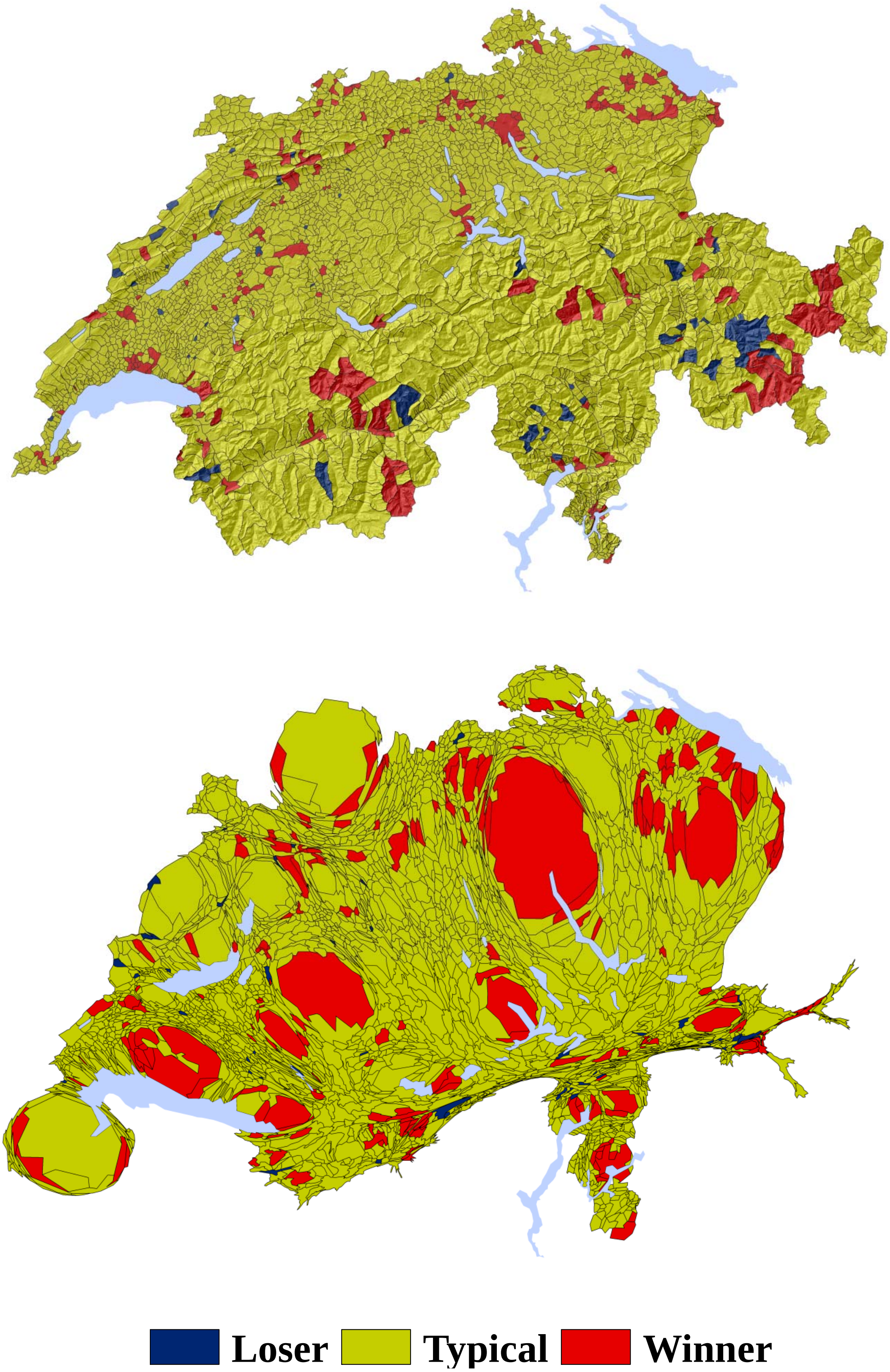
Appendix 5: Decade 1880: “Loser”, 63; “Typical”, 2650; “Winner”, 183 communities



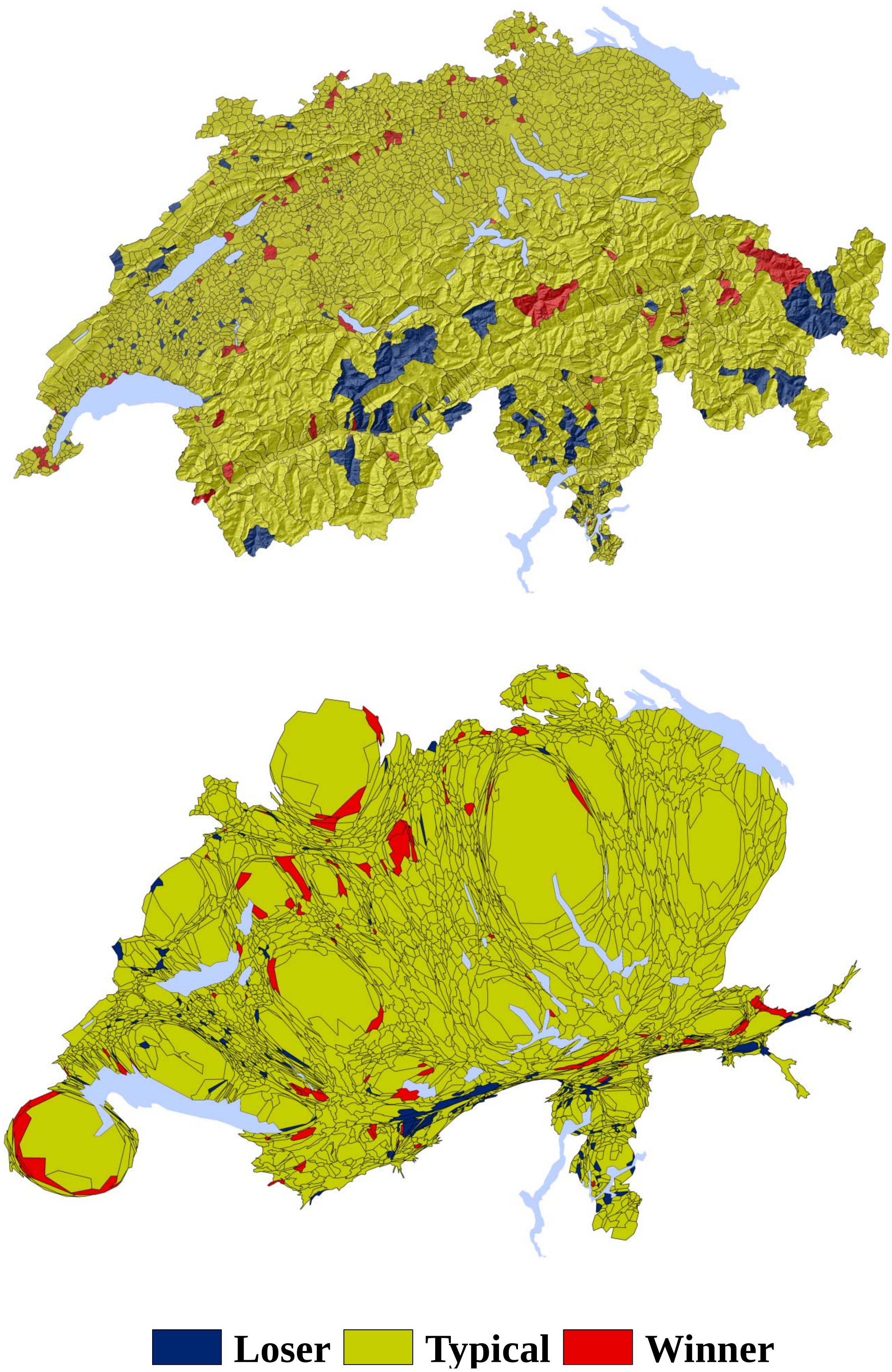
Appendix 6: Decade 1888: “Loser”, 99; “Typical”, 2549; “Winner”, 248 communities



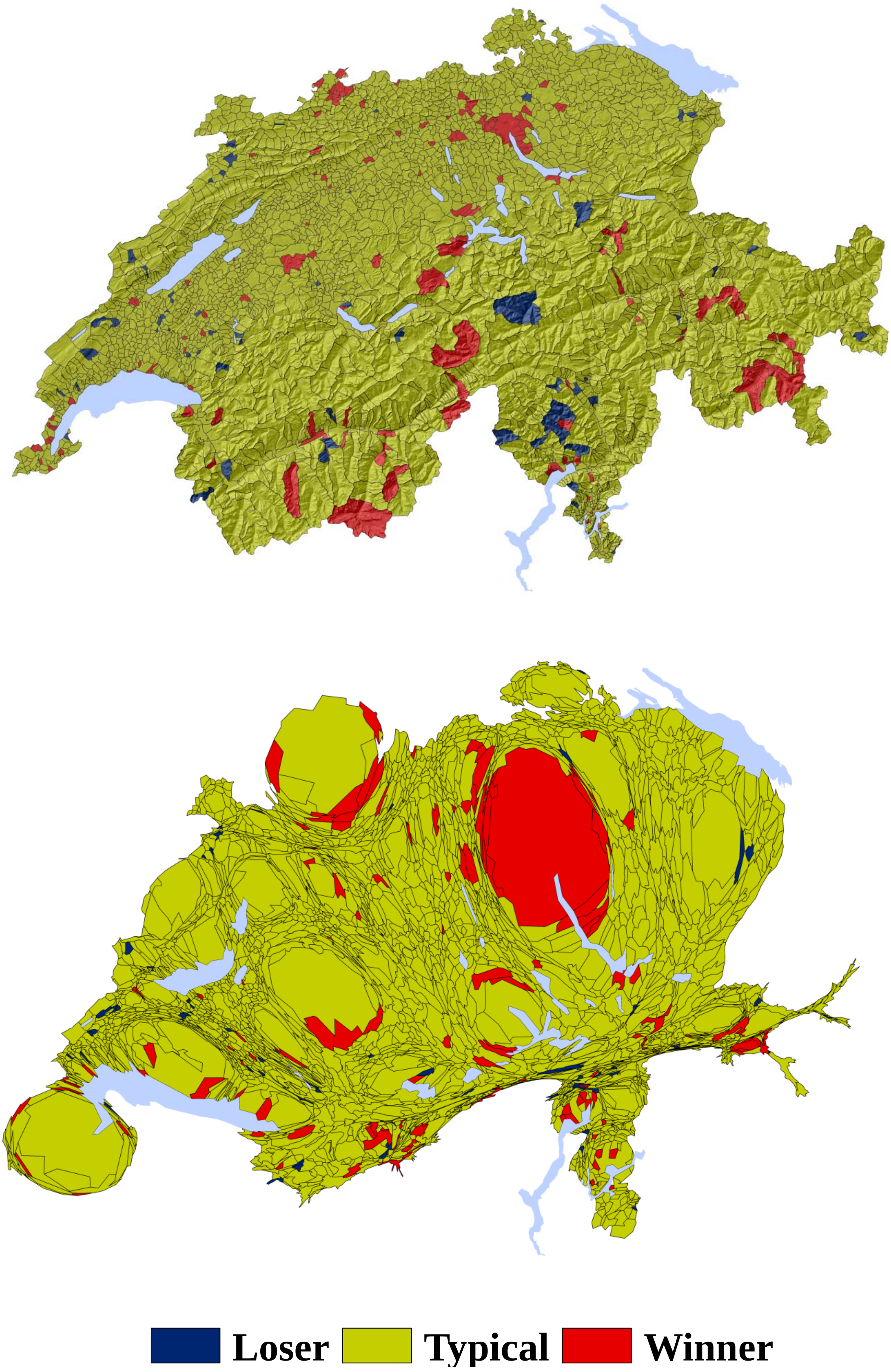
Appendix 7: Decade 1900: “Loser”, 54; “Typical”, 2544; “Winner”, 298 communities



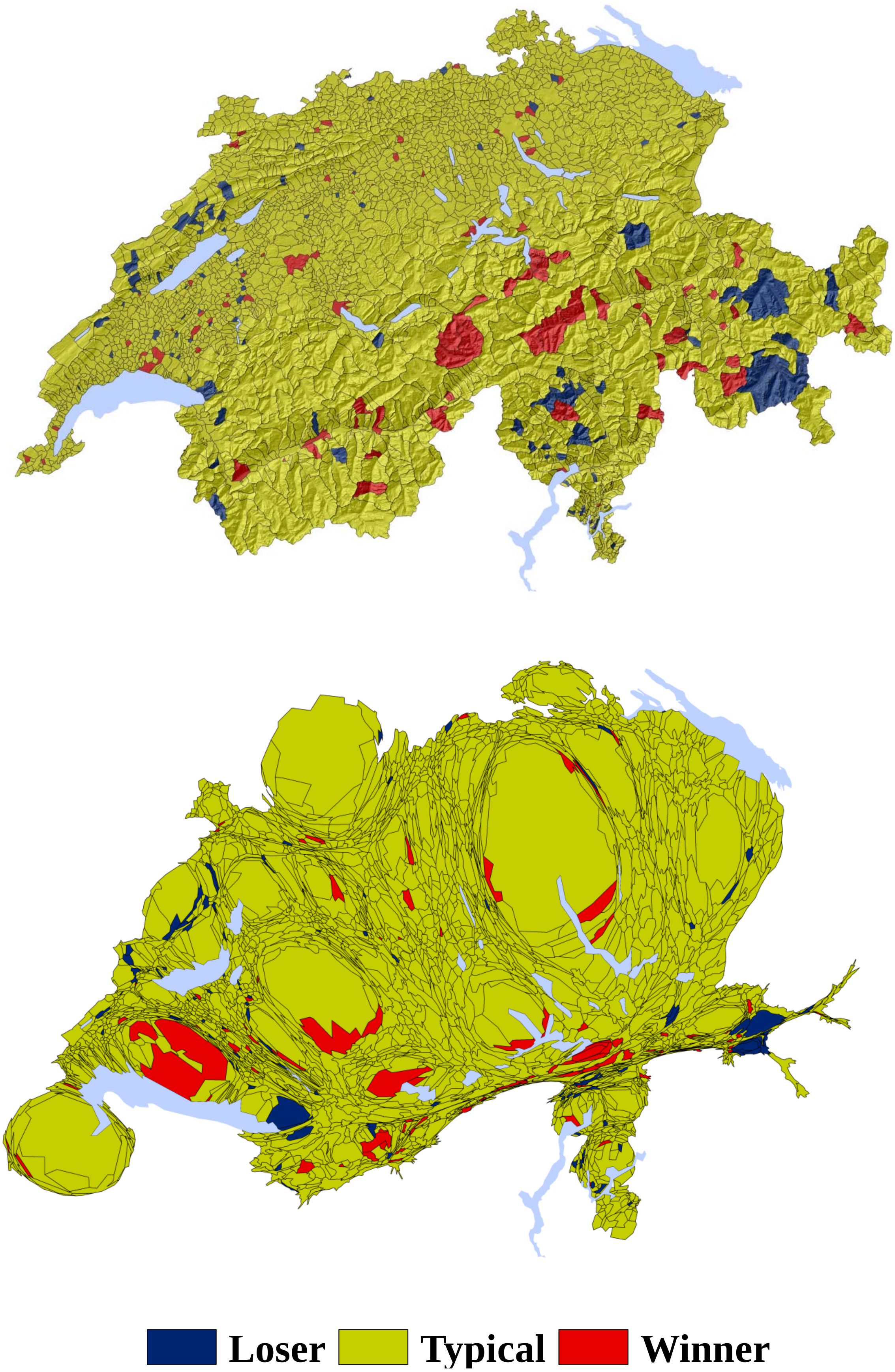
Appendix 8: Decade 1910: “Loser”, 47; “Typical”, 2622; “Winner”, 227 communities



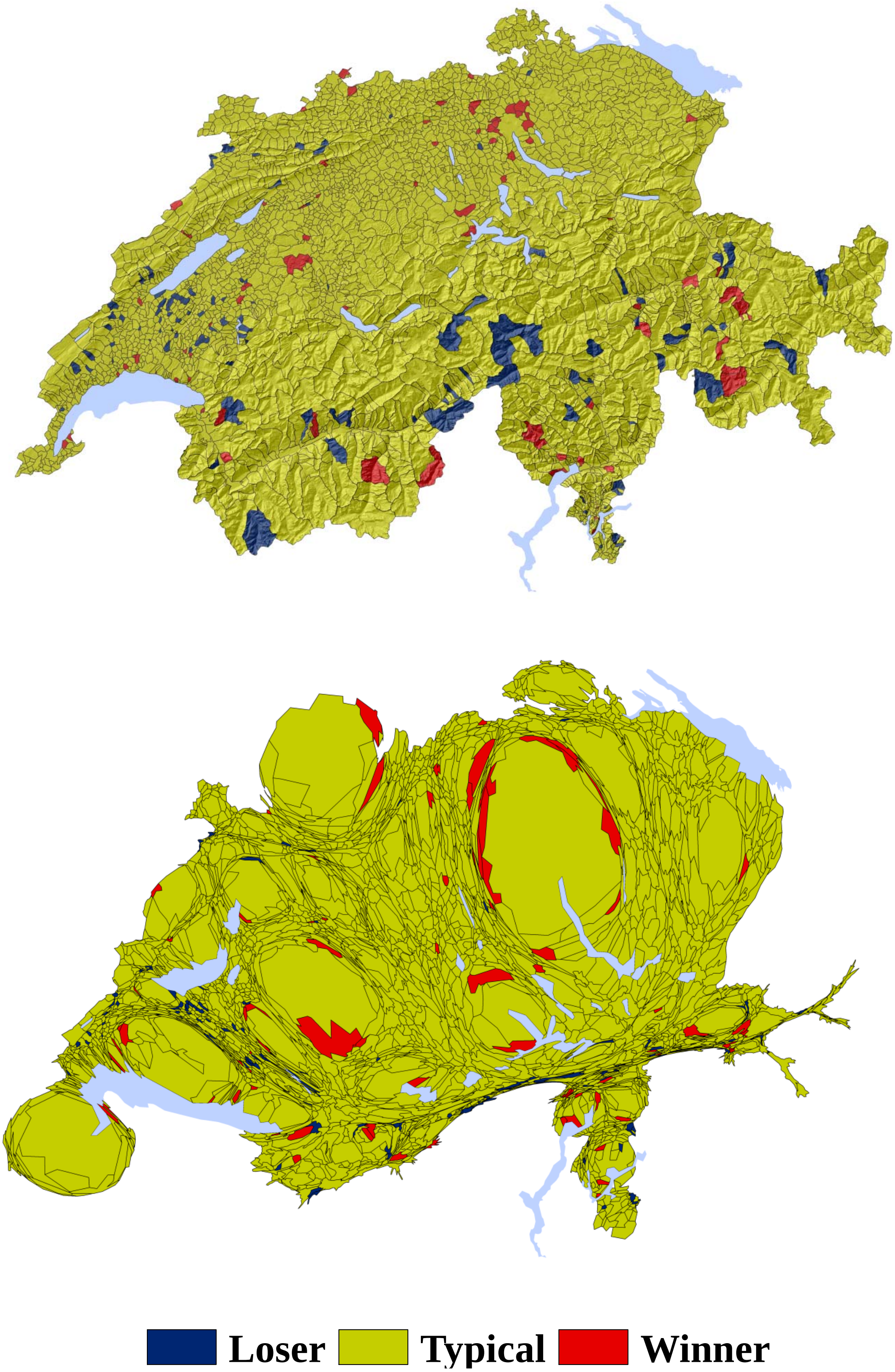
Appendix 9: Decade 1920: “Loser”, 130; “Typical”, 2667; “Winner”, 99 communities



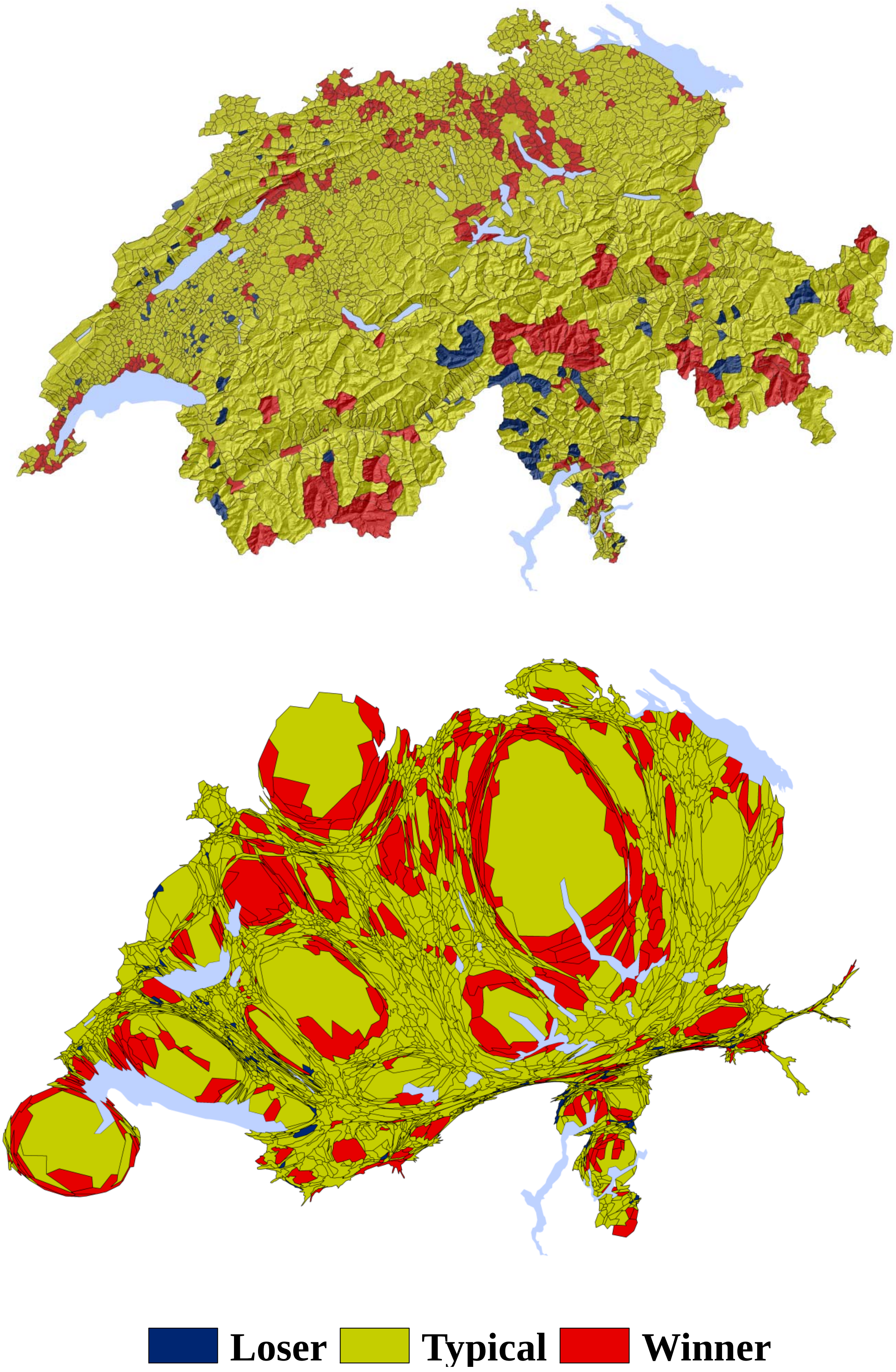
Appendix 10: Decade 1930: “Loser”, 68; “Typical”, 2684; “Winner”, 144 communities



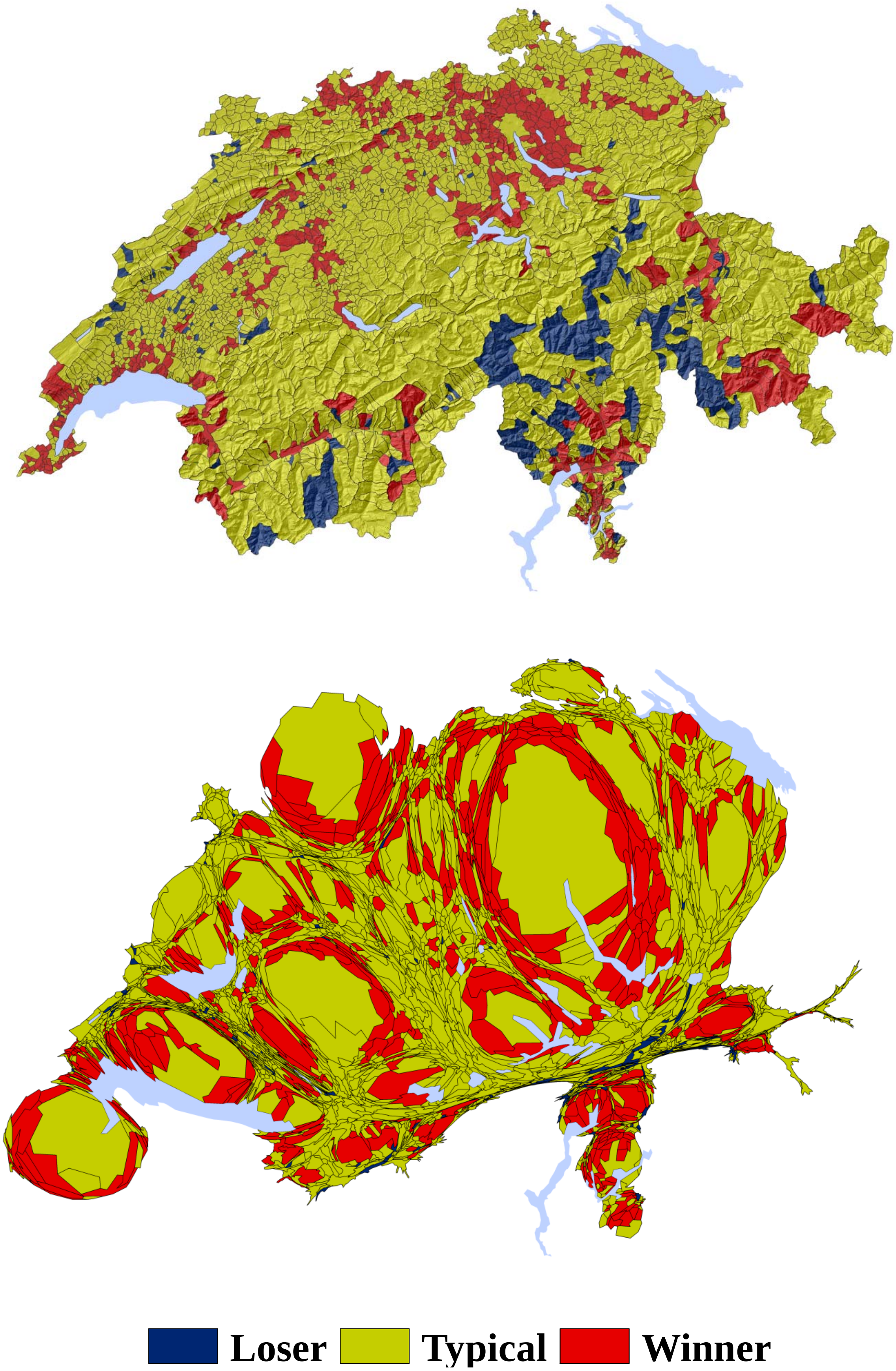
Appendix 11: Decade 1941: “Loser”, 109; “Typical”, 2693; “Winner”, 94 communities



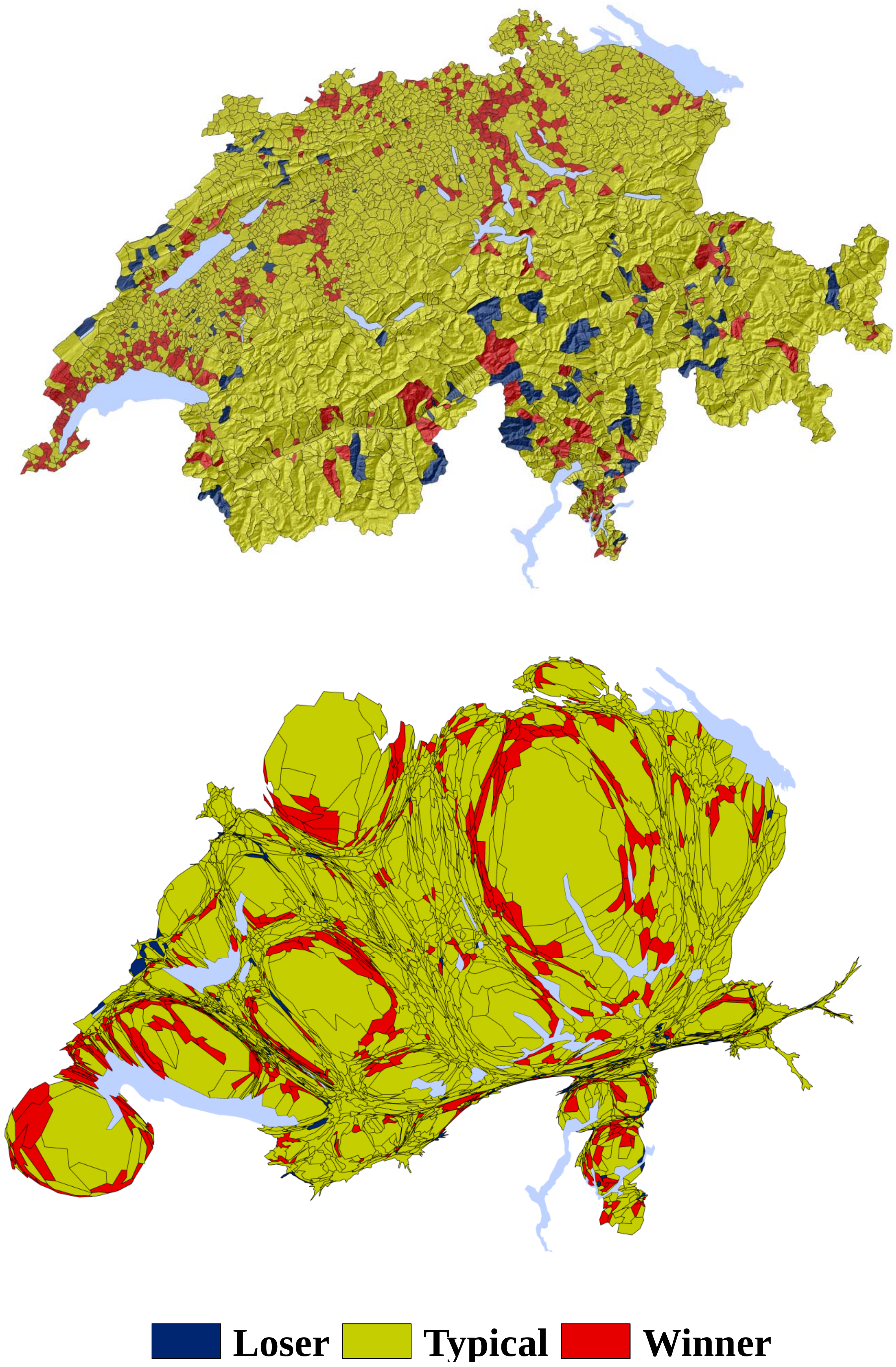
Appendix 12: Decade 1950: “Loser”, 127; “Typical”, 2686; “Winner”, 83 communities



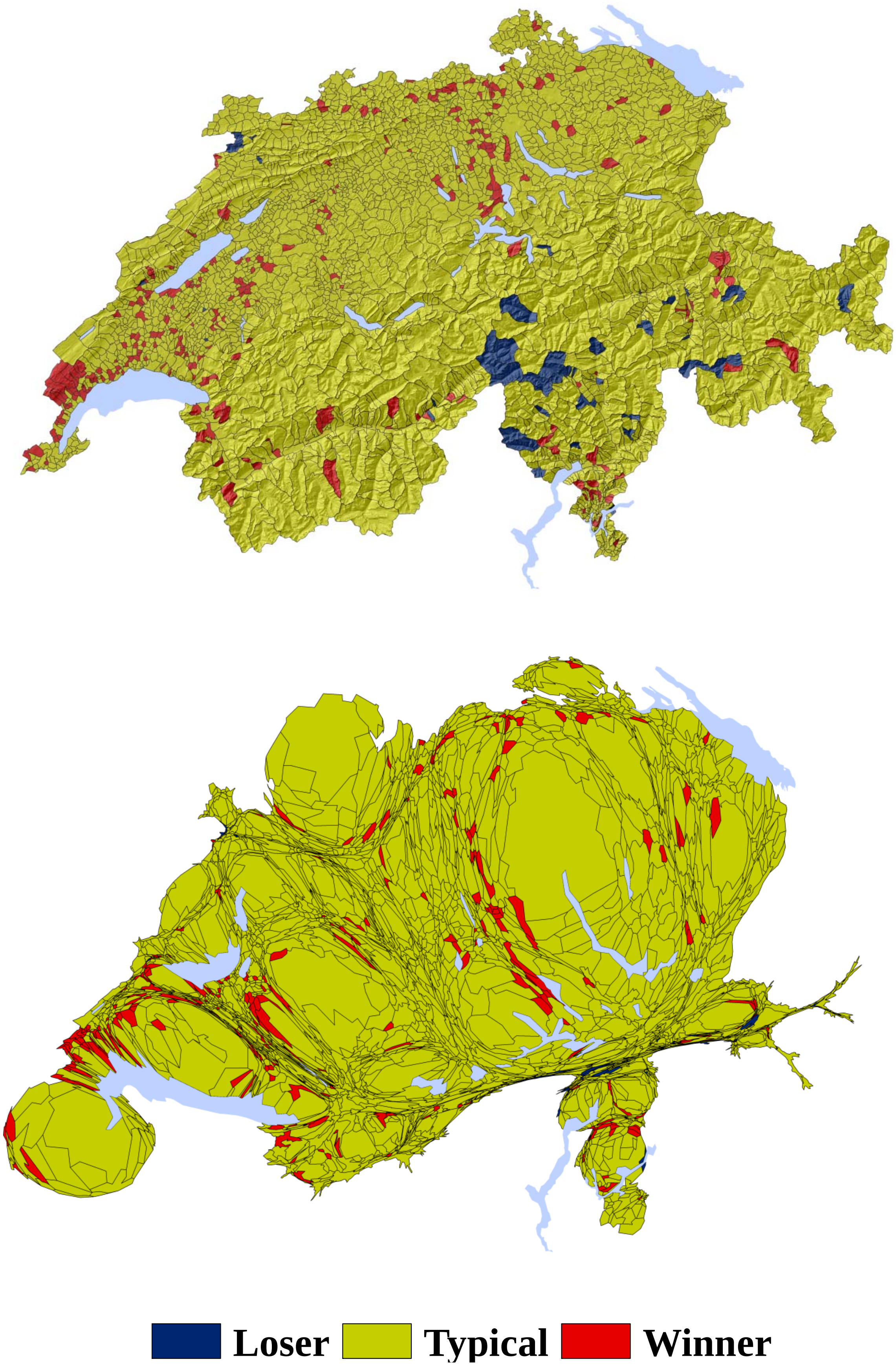
Appendix 13: Decade 1960: “Loser”, 99; “Typical”, 2387; “Winner”, 410 communities



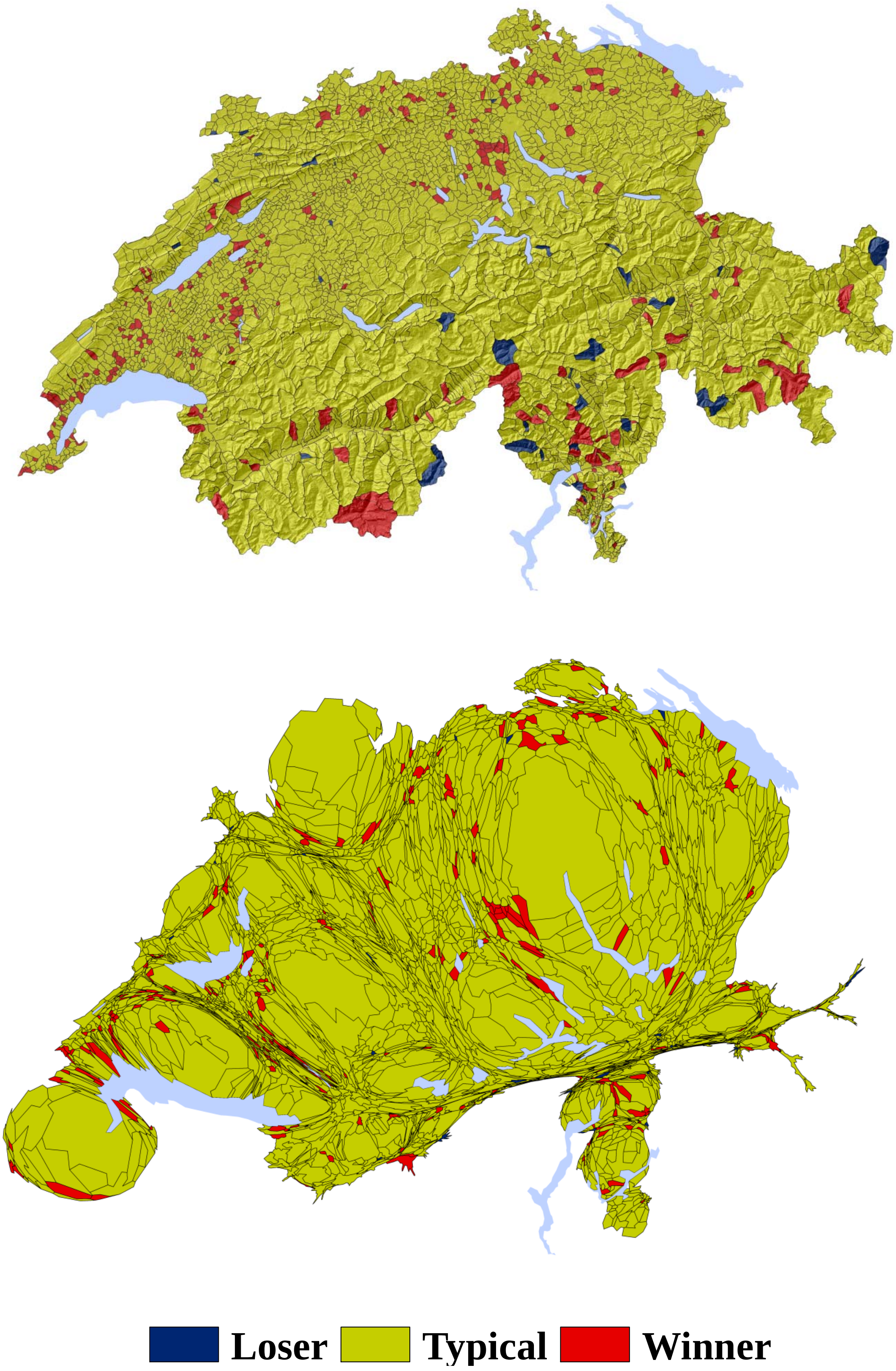
Appendix 14: Decade 1970: “Loser”, 156; “Typical”, 2067; “Winner”, 673 communities



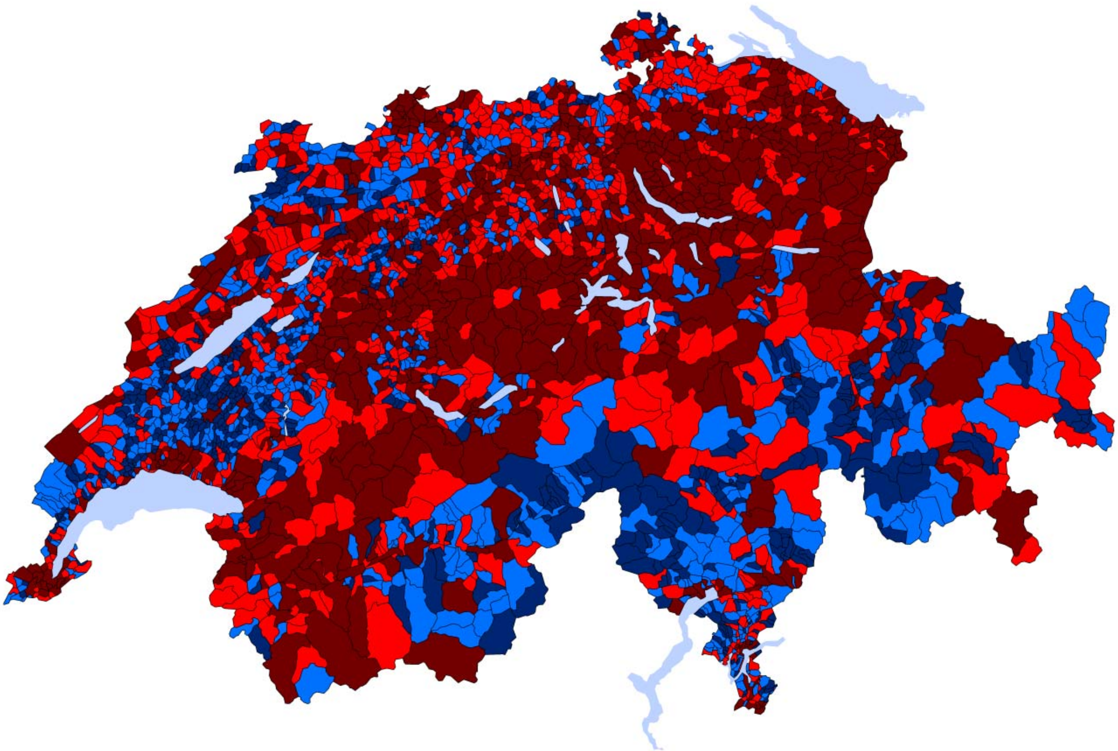
Appendix 15: Decade 1980: “Loser”, 114; “Typical”, 2272; “Winner”, 510 communities

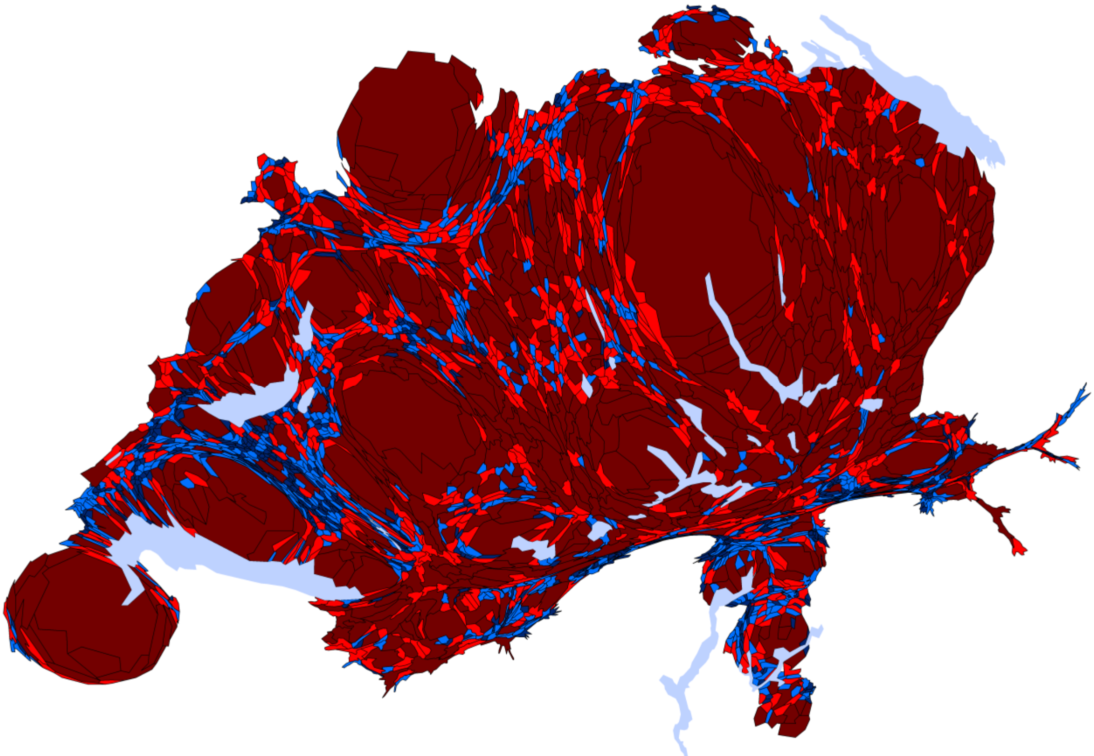


Appendix 16: Decade 1990: “Loser”, 44; “Typical”, 2560; “Winner”, 292 communities



Appendix 17: Decade 2000: “Loser”, 39; “Typical”, 2592; “Winner”, 265 communities





31 - 296 297 - 585 586 - 1324 1325 - 249567

Appendix 18: Long-term Mean of Population (1850-2000), Mean=1489, Median=585