UNIL | Université de Lausanne
Faculté des géosciences
et de l'environnement

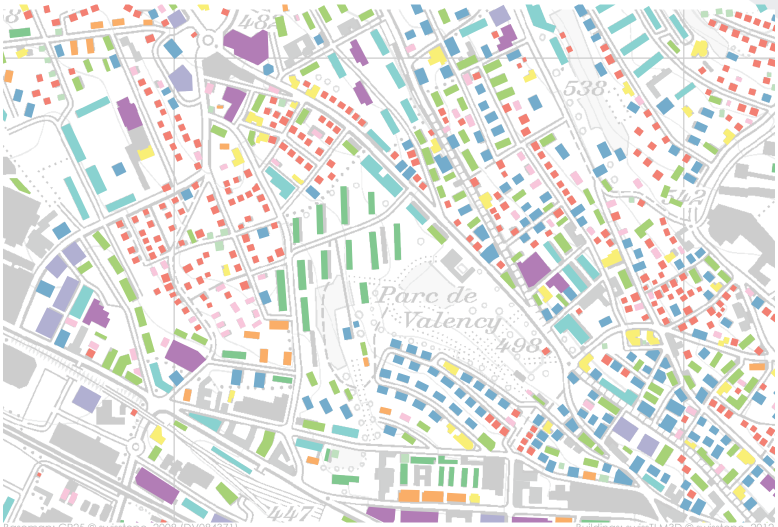
Master of Science in Geography

How does Lausanne's agglomeration shape up ?

Typologies of the urban structures at the building, composition, and neighbourhood scales

Raphaël Bubloz

Under the supervision of Dr. Christian Kaiser
Under Prof. François Bavaud's expertise



Maîtrise universitaire ès sciences en géographie | Juin - 2018

Secrétariat du master en géographie | www.unil.ch/mastergeographie

How does Lausanne's agglomeration shape up ?
Typologies of the urban structures at the building, composition, and neighbourhood scales

Raphaël Bubloz

Spring Semester 2018

Abstract

When we move in an urban environment, it is possible to notice similarities in the organisation of urban structures, despite visible architectural differences. This master's thesis attempts to distinguish the urban structures present, using various quantitative indicators associated to buildings and road networks.

Three scales were selected for the analysis: the building, the composition (the space formed by the building and its block) and the neighbourhood. The focus is therefore respectively on the shape (volume, compactness, size, etc.) of the building, the arrangement between the building and the elements in the immediate vicinity and, finally, on the reachable space around a building.

The analysis aims to bring out typologies of structures at these three levels, within Lausanne's urban agglomeration. Several definitions of the agglomeration are retained (statistical, morphological, functional) in order to highlight possible nuances in the typologies from one definition to another.

Keywords: Urban morphology, urban shape, built typology, urban agglomeration, urban fabric, clustering, Lausanne

Résumé

Lorsque nous nous déplaçons en milieu urbain, il est possible de remarquer des similarités dans l'organisation de structures urbaines, et ce, malgré les différences architecturales existantes. Le présent mémoire tente de distinguer les structures urbaines présentes à l'aide de différents indicateurs quantitatifs relatifs aux bâtiments et réseaux routiers.

Trois échelles ont été retenues pour l'analyse : le bâtiment, la composition (l'espace formé par le bâtiment et son bloc) ainsi que le voisinage. L'intérêt est donc respectivement mis sur la forme (volume, compacité, taille, etc.) du bâtiment, l'agencement entre le bâtiment et les éléments à proximité immédiate et, finalement, sur l'espace atteignable autour d'un bâtiment.

L'analyse vise à faire ressortir des typologies de structures à ces trois niveaux, au sein de l'agglomération lausannoise. Plusieurs définitions de l'agglomération sont retenues (statistique, morphologique, fonctionnelle) afin de mettre en avant d'éventuelles nuances dans les typologies d'une définition à l'autre.

Mots-clés : Morphologie urbaine, forme urbaine, typologie du bâti, agglomération urbaine, tissu urbain, clustering, Lausanne

Acknowledgements

Even if the making of this master's thesis was mainly an individual task, several people have contributed to the final outcome found in the following pages.

First of all, I would like to thank Christian Kaiser, Senior Lecturer at the University of Lausanne, for his supervision. His general suggestions have helped me with the structure and orientation of this work. The assistance received on some coding difficulties has permitted to avoid spending too much time in the worst part of the circle of frustration.

My thanks also go to François Bavaud, Associate Professor at the University of Lausanne, who has accepted to provide his expertise on this work. His accurate remarks have contributed to more precision concerning several aspects of the methodology.

My colleagues and friends, Adeline, Adrienne, Aurélie, Christian, Damian, Daniela, Jasmine, Julie, Léa, Lucas, Nicolas, Patrick, Romain, Tiffany, and Tim should also be given credit. They helped as much on some technical and practical difficulties as on the quality of the off time spent in Lausanne or elsewhere in Europe. Particular mention for Léa and her unfailing support and smile throughout our university studies.

Thanks to my family for their uninterrupted backing and to Thérèse for her eagerness to welcome and host me during the writing of this master's thesis.

Finally, my gratitude goes to the dozens of anonymous people who devote time and energy to maintain and develop quality open source and free software. The numerous tutorials, documentations, and forum answers ease the access to knowledge and we can only hope this tendency will be the norm in the near future.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Bronx, Hollywood, or the City are names of neighbourhoods known all around the globe. What makes them memorable neighbourhoods ? Culture, cinema, or bankers may be some of the elements contributing to their fame. Would it be possible to define the inner structures and limits of a neighbourhood using solely the buildings and roads?

Neighbourhoods and urban structures have long been studied in less known cities in the United Kingdom, France, or Italy by urban morphologists (Conzen, 1960; Moudon, 1994). These monographs relied on town plans and other sources of information to retrace the evolution of urban tissues. This required a lot of work and was really focused on local contexts.

With the arrival of more spatial data and a higher capability in the analytic powers, some new methods were developed to try to define the neighbourhoods across cities in a comparable manner. Recently, Schirmer and Axhausen (2015) have proposed a way to classify urban morphologies based on simple and easily available data elements.

This paper builds on their methods, with some tweaks, to see if it can reveal similar urban forms at three scales and group them in coherent sets. This is applied on the (slightly) renowned Olympic capital of Lausanne. To test those urban classifications, several definitions of Lausanne's agglomeration are retained to compare their results. They are based on statistical, morphological, and/or functional criteria.

These analyses could lead to new opportunities in city design or a change in the approaches on renovation or addition of new buildings among existing ones.

First a review of the literature on cities, geometry, and clustering is proposed, The problem is stated in the next chapter, followed by the details of the methodology used. Results are described in the eponymous chapter and then discussed and criticised. Ultimately, a conclusion with a brief summary and an overture round off this master's thesis.

Chapter 2

Literature review

This chapter summaries and discusses the main theoretical elements mobilised by the analysis. It is divided in several parts each addressing a general thematic.

The first is focused on the study of several aspects of cities. The second on geometric elements allowing characterisation of shape and the final is dedicated to methods of clustering.

2.1 Cities under scrutiny

From the emergence of the first human settlements to the current megalopolis, most of cities have not ceased to evolve, grow, and accommodate an ever expanding urban population (United Nations, 2015). This dynamism and the sheer complexity of cities make them subject of studies in a large variety of fields.

As Berry (1964) put it, cities are systems comprised in systems of cities, they are therefore seldom analysed under the scope of only one particular field. When trying to retrace the historical evolution of cities, their courses are always linked to other elements such as economics (Bairoch, 1991), art, culture, and technology (P. Hall, 1998) and many more.

Hence, the study of cities is not the prerogative of only one field. Yet still, there is a whole branch of geography dedicated solely to urban studies and their multiple facets.

2.1.1 Urban geography

It may be complicated to precisely define urban geography, but there are some common ground between urban geographers. T. Hall (2006) split those concerns in three large types: descriptive, interpretive and explanatory. The first is the identification of internal structures or association between cities and the incumbent processes. The second is the analysis of people's reactions to these structures and processes. The last one is the investigation around the causes and the local adaptations of the processes.

Approaches in urban geography have evolved and the focus on these concerns has shifted with time. The early monographs of cities' sites and situations have been replaced by more quantitative approaches at the start of last century's second part (Latham, McCormack, McNamara, & McNeill, 2008). Those paradigms, looking to find general laws describing the location of people in cities, the influence of urban environment on people, or the influence of human behaviour on spatial organisation, have been accompanied by some more cultural theories later in the century (Pacione, 2009).

This diversity in conception of cities can explain the large spectrum urban geography still has today. Even if they are working on the same concept, urban geographers can come to differing interpretations depending on the paradigm used (Hamnett, 1991). Despite the recent tendency to favour interdisciplinary studies to better integrate complexity (Baerwald, 2010), the divide between qualitative and quantitative approaches in urban geography is still visible.

Some fields rely on surveys, interviews, and sensible diagnostics to apprehend urban quality, citizen participation, or people's representations of the city. On the other side of the methods' spectrum, modelling, simulation, or statistical analyses are used (Clifford, Cope, Gillespie, & French, 2016).

As stated above, urban geography is hard to define but can yet be separated according to theories applied or methods used. However, the definition of what makes a city or urban environment is no consensus, despite it being the main subject of the branch.

Defining the city

For practical (uniformity on the whole territory) and sometimes legal (different status) reasons, cities are defined at the national scale. For example, in Switzerland, a city is a municipality with at least 10'000 inhabitants (Schuler, Dessemontet, Joye, & Perlik, 2005). In France, an urban unit (*unité urbaine*) is based on a minimum number of 2'000 inhabitants and a built continuity with no more than 200 metres between buildings (Clanché & Rascol, 2011). Le Gléau, Pumain, and Saint-Julien (1996) have studied the different definitions for some European countries. They underlined the absence of an uniform definition and the inherent difficulties to compare cities between countries of varying sizes and densities.

Comparing cities' characteristics across countries or continents is an action frequently realised in newspapers (Murphy, 2017a, 2017b) or scientific papers. In order to be able to make these comparisons on a well-defined base, several methods, based on statistics or functional features, have been proposed. These methods consider not just the city but the urban agglomeration around a center.

The Federal Statistical Office (2014) has developed a statistical method to define agglomerations in Switzerland. They use a variable (IJO) based on inhabitants, employments and an equivalence of overnight stays in hotel and health institutions. With a grid of 300 metres resolution, they check for cells having at least a 500 IJO per km^2 density and keep the contiguous (by the side) cells with a minimum total size of 15'000 IJO. A core center is found in these areas if adjacent cells have a density of no less than 2'500 IJO per square kilometres and 5'000 IJO in total. The results are transposed to the municipalities and refined with commuter data to categorise their urban types (*e.g.* core, commuting zone, multi-oriented, or without urban character). These definitions do not necessarily correspond to those used in the substantial projects of urban agglomerations development and have led to some interrogations on their coexistence (Swiss Federal Council, 2015).

In the USA, statistical analysis on cities are mainly based on the Metropolitan Statistical Areas but their complex definition (Office of Management and Budget, 2010) has led to authors proposing alternative ways to define urban areas. Rozenfeld et al. (2008) have designed an algorithm to delineate urban agglomerations based on population's distribution. This city clustering algorithm (CCA) overlays a grid, of a resolution varying with scale, on the studied area. Next, starting with a random cell, it checks if it is populated or not. If yes, it iteratively repeat this check for the neighbouring cells and add them to the same cluster. When there are no more adjacent populated cells, another populated one is randomly selected and the process starts again until all populated cells are in a cluster. This technique, inspired from the works on percolation of Stauffer and Aharony (1994), can also be used on points (Rozenfeld, Rybski, Gabaix, & Makse, 2009). Some parameters can be adjusted, for example, what is considered as populated or if next nearest neighbours should be considered. This method is nonetheless reliant on the granularity of available statistics and the grid placement and resolution.

A further method who rely entirely on physical (built) elements is based on fractal geometry. This technique has been proposed by Tannier, Thomas, Vuidel, and Frankhauser (2011) to avoid the use of a predefined threshold distance between buildings (as in the French definition for example) to define agglomerations. A series of size-increasing buffers is applied on the polygons representing the buildings to obtain a dilation curve relating the buffer size to the number of clusters at this size. A polynomial estimation of the dilation curve is then chosen using the Bayesian information criterion of Schwarz and the curvature function of this estimate is computed. The point of maximum curvature of this function is identified and reported on the estimated curve to find the distance threshold. This distance is then used to compute the urban agglomeration by applying a buffer of radius corresponding to half the distance threshold (which is a diameter).

Several other methods exist or are created on the fly to enable cities comparisons (*e.g.* Lamanna et al., 2018). With the limits of cities better defined, it is possible to study their internal organisation, evolution and some inner delimitations too.

2.1.2 Urban morphology

The study of cities inner shape and divisions is called urban morphology. This field has a rich history of research but has only recently been formalised in a more interdisciplinary perspective (Moudon, 1997). The first pre-war analyses using maps were based on town plans and realised around 1894 by Fritz, whose work had a great influence in several countries (Gauthier, 2004). Post-war, three main schools of thought advanced in the field (Gauthier, 2004; Moudon, 1994). The Italian one was composed mainly of architects interested either in the history of architectural types or the dominance of types in development projects. The English one was carried by a geographer who notably realised a monograph of a British city, studying the whole evolution process of buildings, lots, and streets (Conzen, 1960). In France, the evolution of plots and the medieval towns were the most studied.

This large history of research in the field of urban morphology has, nonetheless, a common theoretical basis. Moudon (1997) identifies three principles :

- The form is defined by three fundamental elements : buildings, plots, and streets
- The form can be analysed at different scales
- The form can only be understood historically because of transformations and replacements

On top of these principles, studies concentrate on urban tissues, groups of cohesive elements found in cities. Yet still, with the focus on the evolution of shapes and their path through time, urban morphology requires lot of data about structures and socio-economics attributes of the inhabitants in order to delimit zones in the city (Moudon, 1997).

Without the focus urban morphology puts on time in the definition of similar areas, several other methods have been proposed to find neighbourhoods in cities.

2.1.3 Delineating neighbourhoods and urban structures

Flint (2009) in his definition of neighbourhoods and communities emphases the growing importance of neighbourhoods in the urban policies. This return of governance in sub-parts of cities is made to bring decisions closer to the people it impact. Yet still, as for the outer limits of cities, several ways to define neighbourhoods exist and can all influence the shape and representations a community can have.

Saunier (1994) has retraced the evolutions of French neighbourhoods in History. At first, neighbourhoods were mainly administrative with generally one church and police station in each. They gradually evolved in a more subjective space in the minds of people who have common conceptions of where neighbourhoods start and end.

Delineating spaces using the representations of inhabitants will surely lead to neighbourhoods varying from one person to the other in function of their personal interests and motivations. To counter this variability in the collective imaginary, it is possible to use methods based on shape, connectivity, and/or socio-economics characteristics.

The city of Lausanne has defined seventeen neighbourhoods for statistical purposes (Roh, Huissoud, Schaffner, & Cunha, 1989). They have used characteristics on the built environment, socio-economic status of inhabitants, ethnicity, jobs, and densities. This allow to define statistical sectors which can be grouped in larger neighbourhoods with respect to the physical barriers (railways, rivers, etc.). These administrative neighbourhoods have been defined in order to have a sense, which is not always the case for some delineations.

To change the definition of neighbourhoods used in residential location choice, Guo and Bhat (2007) have tested other methods. The first one is the circular-unit representation where circles of varying radii are centred on each building. This is best used where there are few physical barriers and opportunities are the same in all directions. The second one is the network-band representation where the neighbourhood is the convex-hull of all reachable points along the roads within a fixed distance from the building. This better takes into account physical barriers and lead to different shapes of neighbourhoods.

Those alternative definitions have been used by Schirmer and Axhausen (2015) in their study of urban morphology in the Canton of Zurich. They proposed a set of indicators to classify urban structure at several scales, based solely on buildings and roads data. They argue that this type of data will become more available with the recent expansion of open (source) spatial data.

Schirmer and Axhausen (2015) also classify buildings according to their shape characteristics. They found similar group of buildings like linear housing, large buildings, or building with multiple wings. They worked on the composition scale as well, with more contrasted results depending on the clustering method used. They find that some buildings have similar attributes even if they are in quite different environments and question the pertinence of this scale. For the results at the bigger scales, they deplore the fact that operations have been made municipality by municipality and some interactions between neighbourhoods and regions have been lost in the process.

2.1.4 Accessibility, networks and routing

As highlighted above, the notion of reachability crosses administrative borders and can be considered as a measure of accessibility. As Bhat et al. (2000) have underlined in their comprehensive review of the notion of accessibility, there is (once again) no fixed definition. It can be understood as a measure of connectivity, opportunities, reachability, etc. In the present work, the emphasis will be on the places reachable from a certain place of origin.

To compute accessibility (or analyse some of its characteristics), the street network is considered as a graph of intersections (nodes) and streets (edges) (Marshall, 2005). The graph can be considered as directed (e.g. A and B are linked, it is possible to go from A to B but not from B to A) or not. The edges can take street characteristics such as class, width, slope, or speed limits.

It is then possible to perform routing on the graph. This operation consists of finding the ways between nodes in the graph according to specific rules. For example, it is possible to assign a cost to each edge (length, class rank, travel time, etc.) and to choose the path from A to B with the least cost. Methods can be adapted to precisely reflect the reality of the terrain or the preferences of people (Meeder, Aebi, & Weidmann, 2017).

2.2 Shape, geometry and topology

As seen in the previous section, some studies in geography describe and analyse human and natural objects or processes. From the beginning of the twentieth century, geographers have use shape mainly to describe their studied elements, such as administrative areas, urban form, or physical features. In physical geography, shapes were also used to compare and classify, notably, drainage basins (Boyce & Clark, 1964).

MacEachren (1985) has retraced the use of compactness in geography. He identified four main types of compactness measures: the perimeter to area ratios (first used in the basin drainage studies), the measures based on parameters of the related circles (inscribed, circumscribed, same area), comparison to a standard shape, and dispersion of elements of area. In addition to this classification, he also evaluated the different measures and noted the influence of some characteristics, such as border complexity, on the perimeter-area measures.

More recently, Angel, Parent, and Civco (2010) have taken the analyse of compactness a step further by enumerating ten compactness properties of the circle and how to compute them. For example, the exchange index describe the part of the shape lying in the circle of same area placed on the shape centroid. This has been used to track gerrymandering on really elongated Congressional Districts in the USA. Another method compute the difficulty to circumnavigate around shapes, when they are considered as obstacles.

Associated to shapes are other geometric objects which can be used to describe them. A bounding box is the minimal rectangular (axis-aligned) geometry which envelops the whole of a shape, thus its coordinates correspond to the shape extremities. Bounding boxes are notably used for spatial databases indexing (Nguyen, 2009).

Convex hulls are another geometry enveloping the whole shape, except it has the shortest (convex) perimeter. A commonly given example of it is to put a rubber band around a shape and see it adopt the minimum perimeter (de Berg, Cheong, van Kreveld, & Overmars, 2008).

Blum (1967) developed a way to study shapes while removing the exterior aspects of it. It keeps only the skeleton of the shape, which correspond to points in the shape being at equal distance from the border. There are different methods to obtain the skeleton of two and three dimensional shapes, which are used in image recognition or medical imagery (Palagyi, n.d.).

2.3 Clustering

In order to simplify analysis on individuals in large sets, it is possible to group those that have similar characteristics. This allow to name the groups obtained and look for the differences between classes. The multiple processes of grouping similar object are called clustering (Hartigan, 1975). The two methods presented below are in the unsupervised category, they output group without prior guidance on the destination groups.

K-Means

The K-Means method divide the observations into k clusters. The observations are assigned to the cluster with the nearest mean (centroid) using an euclidean dstance. The algorithm does so iteratively by trying to shift observations between clusters in order to reduce the distance to the centroid. When no more movement is needed, the process is halted (Hartigan, 1975). This algorithm rely on the user to define the parameter k and will include outliers in sometimes far (but nearest) clusters.

DBSCAN

The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm uses density and reachability to group similar observations. It has two parameters: ϵ (the maximal distance between points) and the minimum number of observations to form a cluster. This method can be used with several measures of distances. Observations are core points if, within ϵ, they have the minimum number of needed neighbours. Points situated at the edge of clusters are density-reachable if they can be reached via core points. Finally, observations that are not reachable or connected in too few numbers are considered as noise. Starting with a random observation, the algorithm will check if it is a core point and assign all other core points and density-reachable points to the same cluster as this point. This continues iteratively on the remaining points with no clusters (Ester, Kriegel, Sander, & Xu, 1996). This process is also dependent on the user to decide the value of the parameters, but sets aside the outliers (observations in not dense enough neighbourhoods).

Chapter 3

Problem statement

The first part of this chapter sets out the theoretical basis of the research. In the second part, the research problem is stated and the main hypotheses explained. In the last part, the various research questions are broken down.

3.1 Theoretical frame

As seen in the previous chapter, the theoretical and practical aspects of defining cities are still controversial topics. Moreover, the retained delineation of a city or agglomeration can be the cause of a bias in line with the modifiable areal unit problem (Fotheringham & Wong, 1991). In order to grasp the possible impacts these interpretations can have, three distinct definitions have been retained:

- the statistical agglomeration of Lausanne, as defined by the Federal Statistical Office (2014)
- the agglomeration of Lausanne, according to the clustered distribution of inhabitants and employments, as defined by Rozenfeld et al. (2008)
- the morphological agglomeration of Lausanne, as defined by the fractal approach proposed by Tannier et al. (2011)

This diversity could bring out other perspectives in the use of non-administrative boundaries to delimit the urban perimeter.

Regardless of their definitions, cities are such complex systems that they can (and should ?) be analysed at multiple scales. Based on the work of Schirmer and Axhausen (2015), three, out of the six levels for which they submit classifications, will be studied: building, composition, and neighbourhood.

The building level comprises all the analyses related to the form and shape of the constructions. The composition relates objects to their direct neighbours and the spatial disposition thus created. Finally, they also propose the neighbourhood as the characteristics of the space comprised in a given walking distance, and the district, comprised in a driving distance (Schirmer & Axhausen, 2015). This study will group those scales in an extended neighbourhood as well.

In order to have an outlook as complete as possible, concepts and tools from various fields discussed in the literature review are employed across the scales mentioned above. The emphasis will be on geometry and topology for the buildings, proximity for the compositions, and accessibility for the neighbourhoods.

Once those various attributes have been computed, it is necessary to classify them so as to find how the elements are distributed in groups. At this step, it is necessary to sort out indices which contribute only a little (or not) to the clustering operations.

Finally, by mapping the groups obtained, it will be possible to determine if it is conceivable, at all scales, to find coherent sets that could be used to reveal urban structures within the city.

3.2 Problem definition and general hypotheses

The main objective of this research is to **reveal similar urban structures, at different scales, and group them into coherent sets**.

In order to reveal urban structure at the chosen scale, arrays of indices must be built to reflect the levels' specific characteristics. At the building scale, indicators must help distinguish between simple and complex forms, small and big structures, compact or elongated shapes, etc. At the composition level, the blocks' main attributes are used and at the neighbourhood scale, accessibility allows to gather informations about the surroundings. This leads to the first hypothesis:

Hypothesis 1: For each scale, the selected indices contribute to the distinction of different types of urban structures.

The several indices having been selected to discern the scales, the clustering results should reveal this selection. If the results are indeed different, their coherence evaluation cannot be based on a fixed set of criteria. The building scale may generate a typology where same type buildings shouldn't necessarily be next to each other, whereas for the neighbourhood we expect results to be grouped by proximity:

Hypothesis 2: The final results of clusters differ according to scale and their respective coherence is a function of the scale

Agglomerations have distinct boundaries depending on their definitions, therefore the space considered and the buildings included will be different. This diversity in the selected elements may cause some differences between the results for each agglomeration. This is expressed in the final hypothesis:

Hypothesis 3: Clustering results differ significantly across the agglomeration's definitions

3.3 Research questions

To guide the research process, a subset of questions, each addressing a part of the bigger problem, has been defined.

- How does Lausanne take shape under each city's definition ?
- Which indices can be used to quantify shapes and forms, surroundings, and neighbourhood ?
- Which clustering method best suits the resulting data attributes?
- Which indices should be left aside in order to improve the clustering ?
- What are the differences between definitions of the city, are they significant ?

Chapter 4

Methodology

This chapter provides an overview of the data and software used, the process to define the different agglomerations, the computed indices and their specificities, and the clustering methods applied. Those parts mainly describe the process followed without digging into the particular techniques used for the analysis. An online repository with detailed procedures and code extracts is available at https://github.com/Raphbub/master-thesis.

4.1 Data sets

Several sets have been used in order to have the appropriate data in the best condition available. The main source of geodata is the Swiss Federal Office of Topography, swisstopo. As their data is not freely available, a brief summary and linked download information about the sets are in Appendix A.

For the buildings, geometries are issued from swisstopo's Topographic Landscape Model (swissTLM3D). The data is from March 2012, has a precision ranging from 0.2 to 1.5 meters and describes the footprint of every building or infrastructure construction in Switzerland. The swissBUILDINGS3D 1.0 data set was used to retrieve a height for the swissTLM3D buildings. It describes building volumes (ignoring the roof) and was lastly actualised in 2005 for the studied area. For the analysis, all buildings without height or with a height inferior or equal to zero have been removed.

For the roads, geometries are also issued from swissTLM3D (excepting for routing queries) which allows for a better alignment with buildings. In this case, the most recent changes are from 2011. Roads used have been corrected after considering the network as a topology to accommodate for little imprecisions (order of the millimetre) at a few line junctions. To help compute blocks, exterior lines of the *Léman* were taken from the swissTLM3D bodies of water data set and considered as roads circumnavigating the lake.

For the routing queries, the network is based on the roads and ways of OpenStreetMap (OSM). OSM is an open source, editable map, covering the entire world, where everybody can contribute to keep it up to date. The data is licensed under the Open Data Commons Open Database License (ODbL) and was downloaded in early April 2018.

In their assessment of OSM data quality, Girres and Touya (2010) have underlined the fast-changing mapping environment of OSM. They also drew attention to the uncertainty created by having little limitations on submitted content, and the mix of users with different profiles and interests. The author can only concur with this point having added more than a thousand driveways leading to buildings for the sake of the routing analysis. Despite these imperfections, OSM data has been privileged over the swissTLM3D set because the former has attributes relative to the maximum speed allowed on roads' sections whereas the latter only has values describing the roads' categories according to their widths.

To delineate the agglomerations, some supporting datasets where used. The FSO edited a list of all municipalities related to an urban center and of their corresponding status in the agglomeration. The associated geometries describing their territories are issued from swisstopo's swissBOUNDARIES3D. Both data sets are from 2014.

The FSO also has statistical data aggregated at the hectare level (tiles of 100 x 100 meters). This is the case for the 2016 STATPOP census and the businesses' structures statistics of 2014. All FSO data sets used are freely available online (see Appendix A).

4.2 Software

In the process of downloading, storing, manipulating, using, computing, or analysing the data sets described above, several tools have been mobilised.

Most of the data preparation, storage and computing has been done using PostgreSQL version 9.6.5. PostgreSQL (https://www.postgresql.org) is a free, open source, object-relational database system to which extensions can be added. To handle spatial data and operations, and the routing, Postgis (sfcgal) version 2.3.3 (https://www.postgis.net) and pgRouting version 2.4.1 (https://www.pgrouting.org) were used. Working with databases allows faster computation and better reproducibility. Code can be executed in the terminal or using a graphical user interface (GUI) such as pgAdmin (https://www.pgadmin.org) v. 4.2. These software come with command line tools to ease some operations, namely importing a shapefile or OSM file in the database.

Visualisation of the spatial data, control of the results and map designs were made with QGIS version 3.0.2 https://www.qgis.org/, a free, open source GIS. In its latest version a database manager is incorporated and support the selection, display, and/or modification of entities in PostgreSQL from QGIS.

R version 3.4.3, a statistical environment and programming language, was used for the statistical analyses, visualisation of data and some modifications in the database (R Core Team, 2018). More than 15'000 packages with specific purposes and functions are available for use, notably *rpostgis* and *RPostgreSQL*, *dbscan*, *tidyverse* and *ggplot2*.

For the fractal delineation of the urban agglomeration, MorphoLim v. 1.5 https://sourcesup.renater.fr/morpholim/ was used. It is a Java-based open source software, developed by members of the *Théoriser & Modéliser pour Aménager* research group from Bourgogne and Franche-Comté universities, designed to help compute the urban limits.

4.3 Case study

This section will firstly describe Lausanne's agglomeration site and situation in a general way before focusing specifically on the methods used to define the different agglomerations. A general comparison of the resulting agglomeration concludes this part.

Lausanne is located on the northern shores of the *Léman* (Lake Geneva) on a hilly landscape, mainly urbanised, with some woods north of the city. The agglomeration extends into the Plateau, the primary flat corridor extending from Geneva to Zurich where most of the Swiss inhabitants reside.

According to the Federal Statistical Office (2014), it is the fourth most populated city and the fifth agglomeration overall. On a regional scale, it is the second agglomeration in french-speaking Switzerland (Figure 1). Lausanne, chief town of the Canton of Vaud, is situated not too far from Geneva and Bern, the Swiss capital city. It is also possible to reach Paris or Milan in train in just under four hours.

The agglomeration (Figure 2) is spanning across several municipalities, from Morges to Cully. The city has only a few parks and the wooden areas are located mainly in the north. Some vineyards are present east and west of Lausanne. The highway rounds the city with only one section entering the inner city but with main arteries linking the city to several highway's entrances. Lausanne's railway station is not far from the center and the network is mainly oriented NW - SE and completed by two subways and several bus lines.

4.3.1 Statistical agglomerations

The statistical agglomeration can be divided in two entities depending on the used center - periphery relation. For the remainder of the analysis, a distinction will be made between the

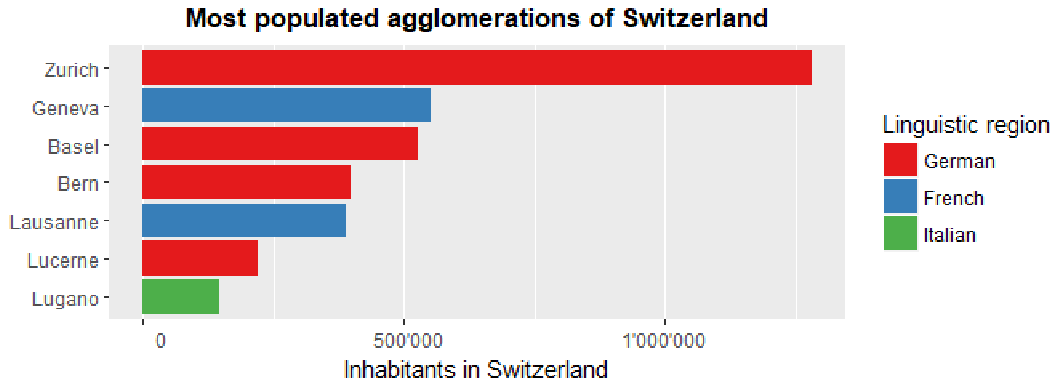


Figure 1: Population statistics of Swiss agglomerations

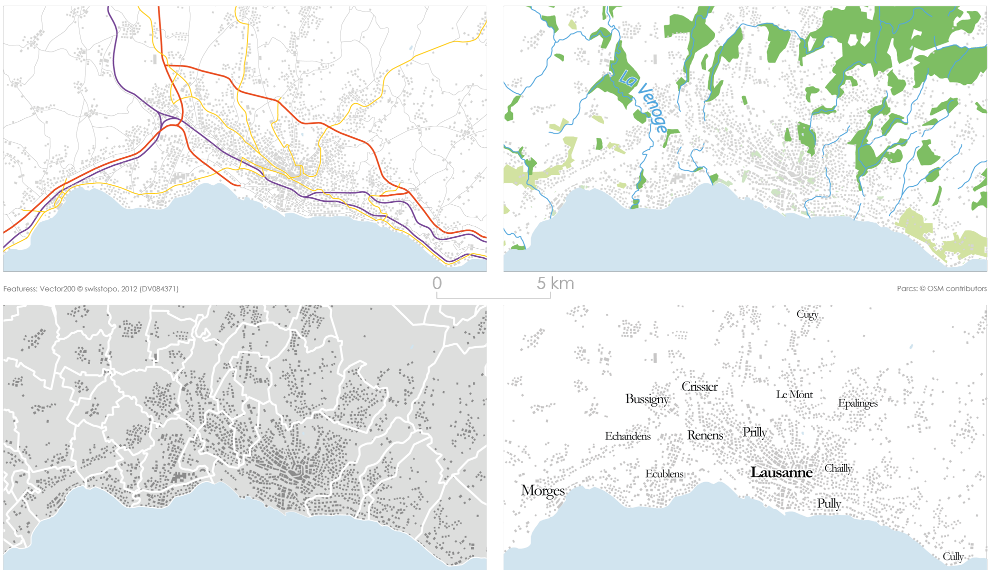


Figure 2: Transportation network, natural areas, municipalities limits, and places around Lausanne's municipality

agglomeration comprising the center-city and its surrounding main center (StatA) and the one including the municipalities from the center-city to the commuting zone (or *couronne*) (StatB).

Both agglomerations respect the municipalities limits with, for the larger StatB, four of them in the Canton of Fribourg (Figure 3). StatA covered an area of 122.96 km^2 and 24 municipalities for a total of 280'051 inhabitants in 2012. StatB was nearly six times bigger in size with 772.7 square kilometres covered for 131 municipalities and a population of 389'614.

4.3.2 Morphological agglomerations

As seen in the section 2.1.1, there are several definitions of an urban agglomeration based on more functional or morphological criteria. Here two methodologies have been retained to be able to compare results. Contrary to the statistical definition, the following agglomerations do not follow municipalities' borders.

Fractal methodology

To define the fractal agglomeration of Lausanne, the buildings' shapefile was used in MorphoLim. The polynomial degree retained for the estimation of the dilation curve was the 6th (Figure 4).

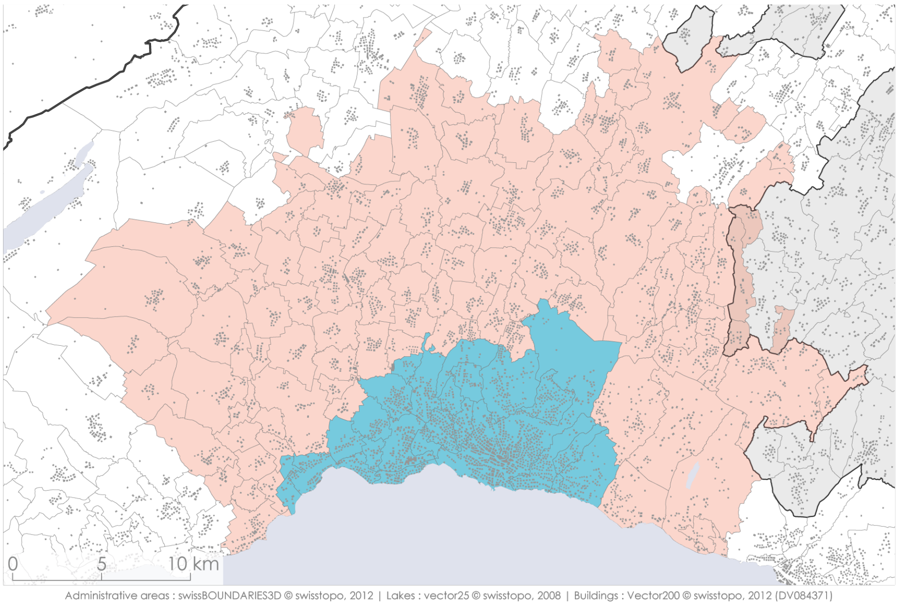


Figure 3: Statistical agglomerations of Lausanne

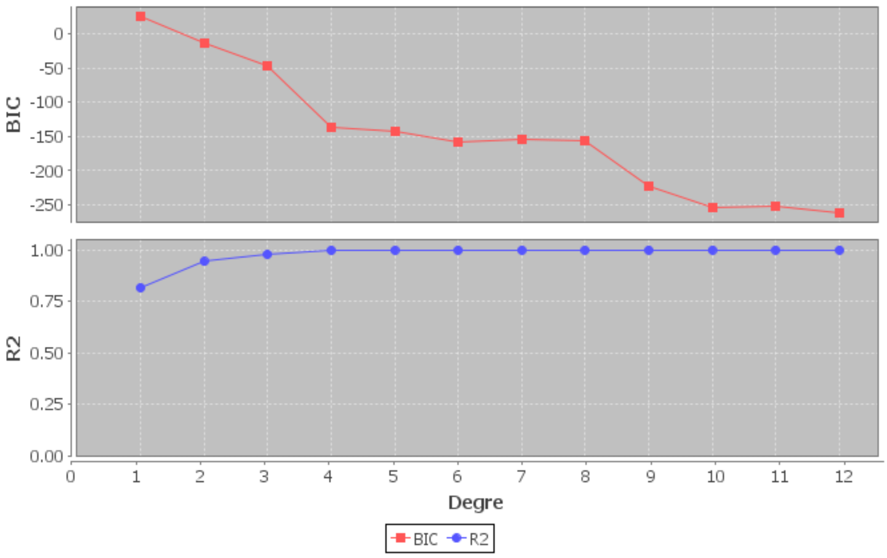


Figure 4: BIC and R2 for polynomial degrees of dilation curve approximation

Then the absolute maximum point of curvature is identified and reported on the dilation curve to find the distance threshold Figure 5. In Lausanne's case, the threshold is 178.2 metres.

The program then computes a positive and negative buffer of half the threshold on the buildings. The resulting geometry corresponds to the morphological agglomeration defined by the fractal method. As can be seen on Figure 6, the zone is smaller (53.48 km^2) than the statistical agglomerations and has some holes, mainly matching the wooded areas of the city. The west and east extremities are just included, respectively by a thin zone over the Venoge river and thanks to buildings lining the lake shore up to Cully.

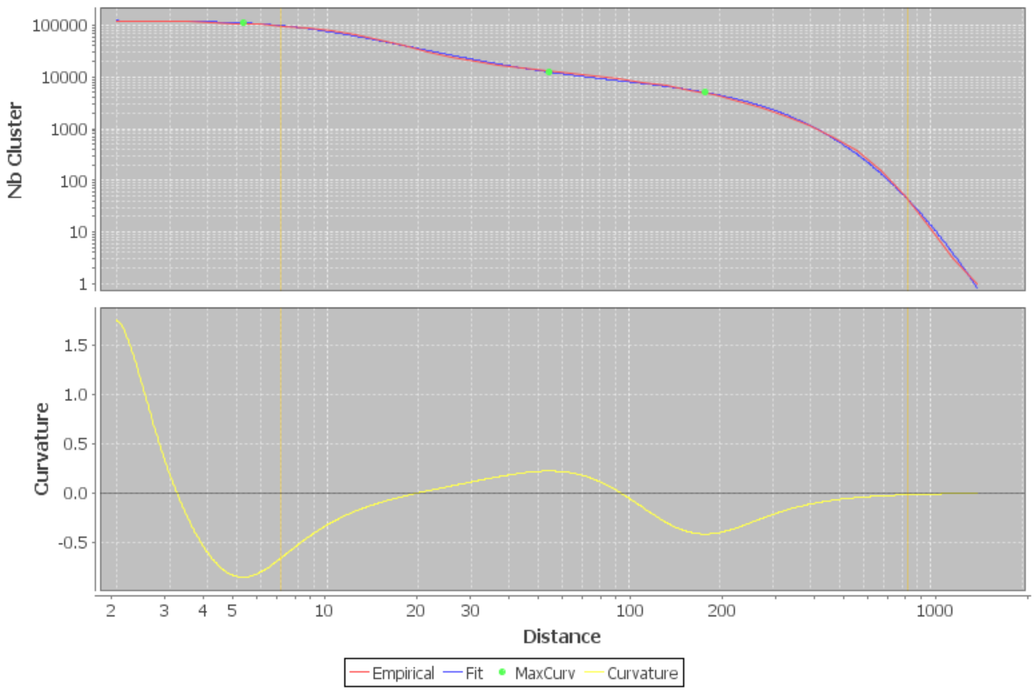


Figure 5: Polynomial approximation and curvature function

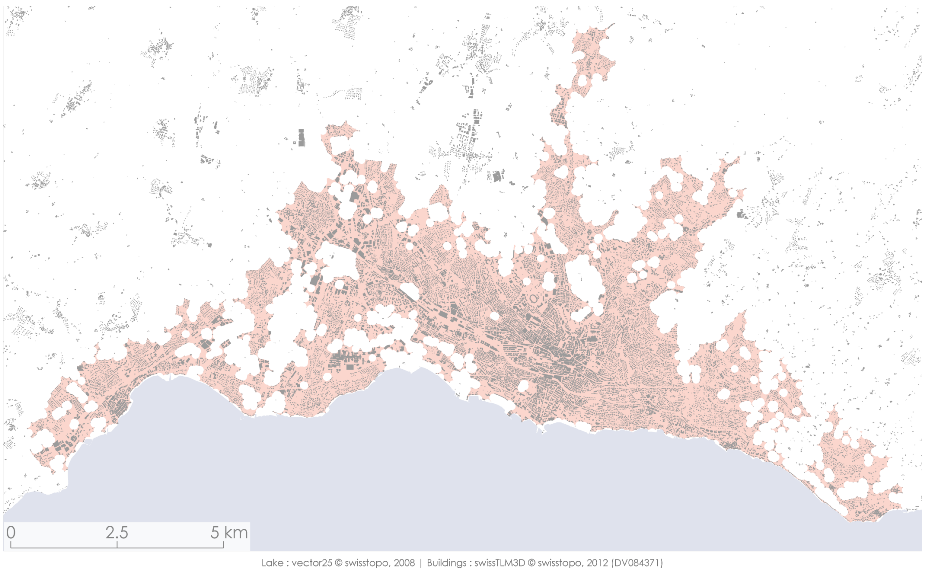


Figure 6: Morphological agglomeration of Lausanne

City clustering algorithm

The final technique used to define an urban agglomeration is the CCA. This method, based on density and spatial proximity, has been applied on a combination of census data and employment statistics based on businesses' locations. This approach is necessary to take into account economic activities located in industrial zones where people may not be living. This will ensure buildings in these areas are also included in the analysis.

The grid position and resolution (100 x 100m) are given and correspond to the ones used by the FSO for their hectares statistics. A few others parameters can be tweaked to join the hectares in a satisfying result. It is possible to define the neighbourhood rule (via a set spatial distance) and to choose a minimum density threshold for which hectares should be considered.

	Area [km^2]	Perimeter [km]	Buildings
StatA	122.96	90.6	24094
StatB	772.7	260	51704
Fractal	53.48	252.7	24037
CCA King	27.64	251.4	13990
CCA Lenient	36.56	390.6	18538

Table 1: Summary of the agglomerations' main characteristics

Those parameters have been tested and adjusted at the Swiss national level and the density considered is, at least, 20 inhabitants and/or employments combined per hectare. For the spatial distance, two different ones have been retained. The first corresponds to a king neighbourhood (142 metres - or one hectare - in all directions). This result is quite relevant at the national scale (see Appendix B) but is limiting for the case of Lausanne.

As can be seen on Figure 7, the CCA with a king neighbourhood (in blue) ignores a large part of the agglomeration's western side. This is mainly due to the gap caused by the Venoge river and its surrounding forest, which is just larger than the 142 metres threshold. To make up for this borderline case, a more lenient neighbourhood definition was used with a distance of 224 meters. This corresponds to a maximum gap of one hectare in diagonal. The result (Figure 7 red and blue) is a more disjointed area - with a few isolated patches - which better considers the densely occupied zones between Morges and Crissier.

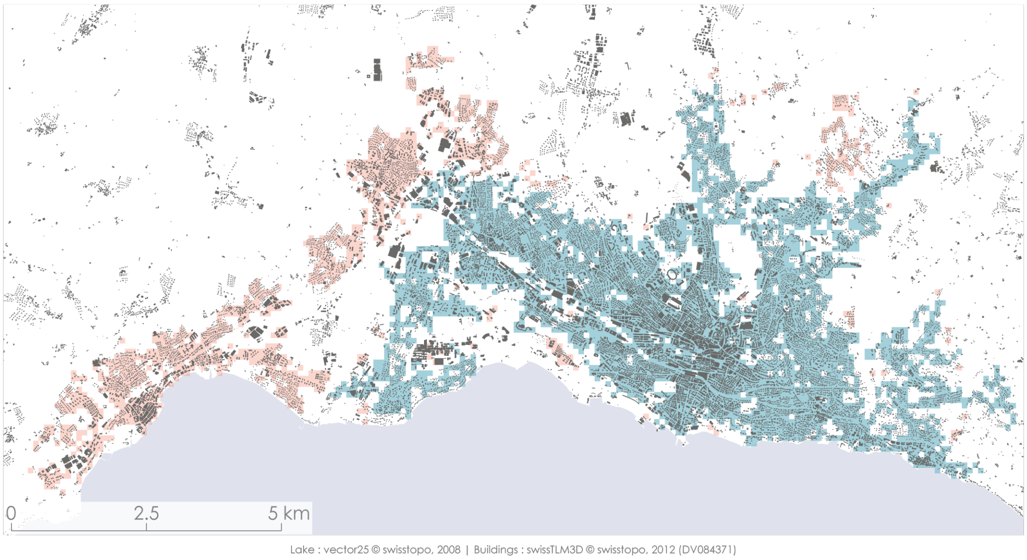


Figure 7: Agglomerations obtained using the CCA method

4.4 Indices

The following subsections address the indices used at the three different scales. When the indices are not self-explanatory a brief description, or the formula used, is provided. As stated earlier, the exact procedures can be found online and the complete indices list is in Appendix C.

To avoid the introduction of an artificial border effect, all indices have been computed on buildings in a two kilometres buffer of the biggest agglomeration.

4.4.1 Building level

At this scale, indices look to describe the form and shape of the buildings. To ensure clarity in this section, they have been arbitrarily divided into groups according to the characteristics upon which they rely.

Basics

The first group comprises what can be called basic descriptors of shape and form:

Perimeter

Area

Volume to facade — Ratio of the building volume divided by the facades' area. Corresponds to $\frac{Area}{Perimeter}$

Widths — Longest and shortest edge of the building's boundary.

Stories — Number of stories in building. It is approximated using the height (roof non included) and a generic story height of 3 meters. All buildings are considered to have at least one story.

Floorspace — Total floor area of the building, equals to $Area \cdot Stories$

Compactness

For the following compactness indices, the name of the author proposing the measure (discussed in section 2.2) is used to distinguish them:

Miller — Computed as $\frac{Area}{(.282 \cdot perimeter)^2}$

Schumm — Computed as $\frac{Diameter\,of\,circle\,with\,same\,area}{Diameter\,of\,circumbscribing\,circle}$, with diameter of same area circle equal to $2 \cdot \sqrt{Area/\pi}$

Haggett — Computed as $\frac{Diameter\,of\,inscribed\,circle}{Diameter\,of\,circumbscribing\,circle}$

Lee & Sallee — Computed as $\frac{Area\,of\,intersection}{Area\,of\,union}$ between the shape and the circle of same area aligned on the shape's centroid.

Ehrenburg — Computed as $\frac{Area\,of\,inscribed\,circle}{Area}$

Bounding box

Bb perimeter

Bb area

Length & width — Longest side as length, the other as width

Length to width — Ratio between the sides of the bounding box

Area to bb area

Perimeter to bb perimeter

Inscribed, circumscribed and other circles

The remaining descriptors linked to one or more circles associated to a shape.

Circ. circle radius

Exchange index — Computed as $\frac{Overlap}{Area}$, with *overlap* the common surface between same area circle and the shape.

Detour index — Ratio between the same area circle perimeter and the convex hull perimeter

Convex hull

Ch area

Ch perimeter

Area to ch area

Perimeter to ch perimeter

Skeleton and centerlines

Skeleton dead ends

Nb. of lines in center line

Centerline length

Nb. of orientation in center line

Centerline length by orientation Orientations are NS, EW, NESW, SENW

Centerline main orientation

Miscellaneous

Av. distance to corners

Nb. of inner courtyards

Area of inner courtyards

Inner courtyard area to area

4.4.2 Composition level

Some indicators at this scale are based on the notion of block. The retained definition for the block is the space comprised between roads (Schirmer & Axhausen, 2015). The road width is inferred from the road category attribute. To avoid streets cutting through buildings, the minimum width for the category is always taken. For example, in the analysis all roads of class *6m Strasse* have a width of 6.21 metre, even though the effective width can be between 6.21 and 10.20 metres. Every building is assigned to a unique block. If a building spans across multiple blocks, it is allotted the block with which it shares the biggest area.

Streets

Facade length Length of facade perceived from the street. A facade is considered as perceived if it is within a 3 metres distance from the roads. Only the length within this distance is counted.

Part of facade perceived Ratio between the perceived facade and the building's perimeter

Buildings

Distance to nearest building

Distance to nearest building in another block

Influence zone area Open area closest to the building. Computed with the medial axis of the block's non-built area. Akin to the Voronoi polygons of the buildings. This indicator is limited to the surface corresponding to a circle of a 300 metres radius.

Area to influence zone area

Blocks

Distance to block limit

Block area

Area to block area

Views between buildings — Number of spaces between buildings that can be seen from the streets. Computed by making the Voronoi polygons of the buildings' centroids (using the point on surface algorithm) and then counting the number of edges intersecting the streets.

Perceived facades to block perimeter

4.4.3 Neighbourhood level

At this scale, the interest is on the characteristics of the building's surroundings. The indicators are computed for several distances, using different methods.

The first one is the most basic. The considered surroundings consist of a circle which center is the building. The radii considered are 100, 300, and 500 metres. This approach is rather quick but does not take into account the anisotropy of space. Physical barriers (rivers, highways, or railways) could prevent movement in some directions and using the features in those unreachable spaces to compute neighbourhood statistics may not be relevant.

The second method uses the same distances as above (100, 300, 500m) but along the road network. All buildings are first attributed the nearest node on the network. For these routing queries, the whole graph is considered as undirected and the edges' costs correspond to their lengths.

The last method is similar to the second but considers the area reachable in a two minutes drive by car. The buildings are attributed the nearest node located on the end of a drivable edge (*e.g.* no pedestrian roads, cycle lanes, nor stairs). Only the drivable edges of the network are used and the graph is deemed as directed to account for one-way streets. The costs correspond to the time (in seconds) required to travel along the edge, using its maximum allowed speed.

For both routing methods, all the reachable points are first computed, then the α-shape of these points is taken as the catchment area of the building.

Radius & network distance

Cumulated area in x m.

Buildings in x m.

Cumulated floorspace in x m.

Dead ends in x m.

Intersections in x m.

Cumulated network length in x m.

Driving time

Catchment area

Buildings

Cumulated floorspace

Dead ends

Intersections

Cumulated network length

4.5 Clustering

Clustering has been used to group the buildings of every agglomeration into classes, based on the indicators described above. For each scale, all values have been standardised before using the DBSCAN method. This first step allows to detect outliers, buildings which are considered too dissimilar from the majority of buildings. The different indicators were given the same weight and the densities were computed using the euclidean distance. Once those noise points were set aside, the remaining cluster could be further analysed.

The second step was to perform a K-Means clustering on the remaining buildings. For each scale, a selection of the available indicators has been used. However, the clustering on these subsets was unweighed. The parameter k was determined by looking at the explained variance according to the number of group and trying to find a plateau once the explained variance levels were superior to 70%. This explained variance corresponds to the within cluster sum of squares. The k retained is the one before the plateau occurs.

Chapter 5

Results

For each studied scale, results are presented. First a brief reminder of the scale's context and the indicators used are presented. The general overview of the clustering and the different class are also detailed. Finally, if applicable, differences between agglomerations are shown.

5.1 Structures at the building scale

The building scale's indices reflects the shape and form of each building. The DBSCAN clustering method has been applied on all indices to discover the less dense data spots which are deemed as outliers. The remaining buildings have been clustered with the K-Means method.

To improve the interpretation of the results all indices related to the centerlines' lengths by orientation and the ones about the courtyards have not been considered for the clustering. In fact, keeping the orientations tended to put similar but distinctly oriented buildings in different clusters, thus limiting the possibilities to find typologies of comparable buildings. For the courtyards, the reason is that most of buildings having one or more courts are set aside by the DBSCAN.

5.1.1 General overview

Those particular buildings, considered as noise by the DBSCAN, can mainly be separated in three categories, common to the five agglomerations. The first one concerns the city-center buildings. As the analysis uses the composite of buildings, this leads to the center (Figure 8a) being made of big, irregular masses, having outstanding attributes instead of the closely-packed, side-touching housing existing in reality. The second category is made of large industrial or commercial buildings and/or warehouses (Figure 8b). Finally, the last category of really complex shapes (Figure 8c) and buildings with courtyards (Figure 8a) are also not assimilated with the DBSCAN.

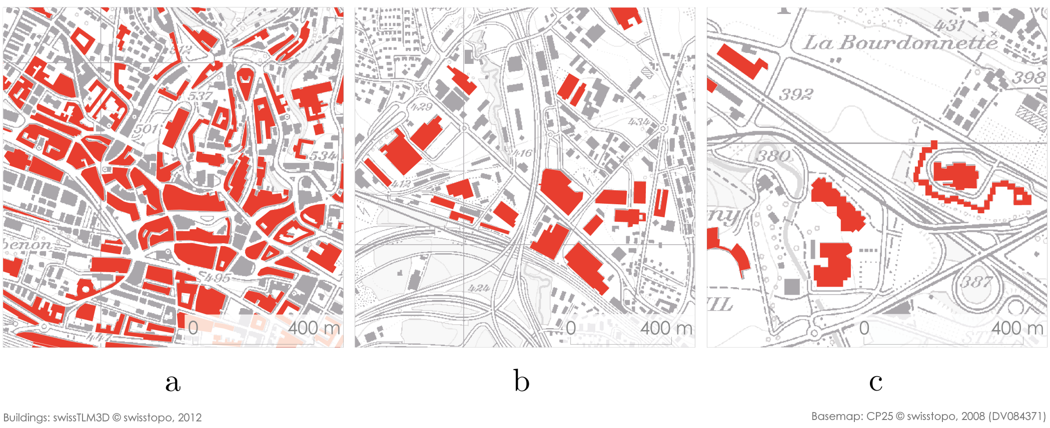


Figure 8: Structures set aside at the building scale

Results of the K-Means on the remaining buildings are quite satisfying, with an explained variance between 73% and 76% for the five agglomerations. This corresponds to 12 clusters for the agglomerations and 15 for the bigger *statB*.

From these clusters it is possible to identify common building typologies that change only a little across agglomerations (see below). These typologies can be grouped in bigger categories:

- Voluminous and elongated buildings
- Compact rectangular buildings of at least a couple stories
- Detached or semi-detached buildings, nearly square in shape, with maximum four stories
- Large structures or aggregated blocks
- Buildings with a "complex" shape, such as a *T*, *U*, or *L*

Voluminous and elongated buildings

Those buildings tend to be big to large apartment complexes or similar looking business offices. They are located in distinct neighbourhoods, in a rather orderly alignment. Even though they indicate a higher density, they are not dominant in the agglomeration's core center. There are three distinct clusters in this group.

The first one is made of the most elongated and highest buildings (Figure 10a). They have the smallest value of compactness for three of the five definitions used (Schumm, Haggett, and Ehrenberg).

The buildings in the second cluster are also elongated but with a smaller width and less storeys. Their bounding box's attributes are also different. A small change in the orientation of buildings causes the bounding box to increase in size, thus influencing all the related indices (Figure 9). The buildings of this cluster tend to be orientated SN or EW (Figure 10b) where the ones in the first cluster span along other directions (Figure 10a).

a b

Figure 9: Difference of bounding boxes for the same building

Finally, the last cluster of this group includes the rectangular buildings but less elongated and of a medium height (Figure 10c).

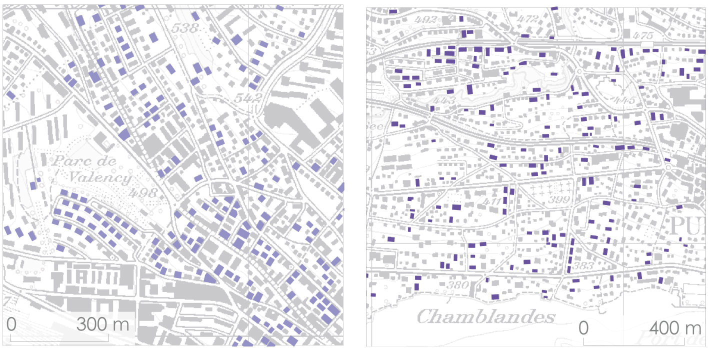
a b c

Buildings: swissTLM3D © swisstopo, 2012 Basemap: CP25 © swisstopo, 2008 (DV084371)

Figure 10: Clusters in the elongated buildings group

Compact rectangular buildings of at least a couple stories

This regroup buildings which are quite similar to the previous group but with some more compact shapes. They tend to be three to seven-storey blocks of apartments. There are two distinct clusters in this group (Figure 11). They are mainly differentiated by their (non-) alignment along the north and south axis used to compute the bounding boxes.



a b

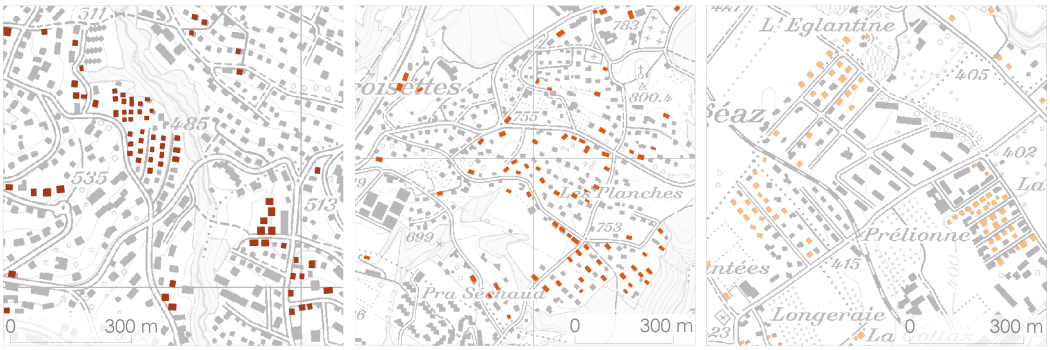
Buildings: swissTLM3D © swisstopo, 2012 Basemap: CP25 © swisstopo, 2008 (DV084371)

Figure 11: Clusters of medium-sized buildings

Detached or semi-detached buildings

This group represent more than half of the agglomeration's buildings. In this one, the buildings are smaller and mainly composed of detached or semi-detached houses. They are scattered across the agglomeration, at the exceptions of the core center and between the aligned bigger structures. On the three clusters of this group, two are, once again, separated according to their orientations (Figure 12a and b).

The last cluster is made of buildings which shapes near a square, whereas the two previous clusters have buildings with a longer side (Figure 12c). This allows to distinguish either bigger villas or semi-detached houses.



a b c

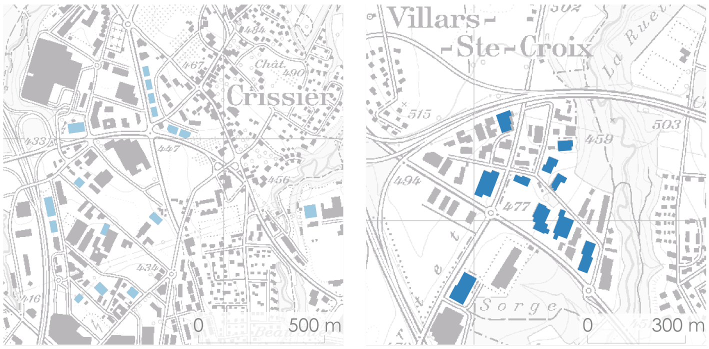
Buildings: swissTLM3D © swisstopo, 2012 Basemap: CP25 © swisstopo, 2008 (DV084371)

Figure 12: Clusters of the smaller buildings

Large structures or aggregated blocks

The buildings in this group are large structures (important area and edges' values) corresponding to industrial warehouses but still smaller than the outliers described above. There also are some blocks located in the city center and that are not too complex to be considered as outliers.

The main difference between these clusters is in the compactness and complexity of the shape. The first cluster is composed of buildings with a mainly regular shape (Figure 13a). The second has more complex geometries (Figure 13b).



a b

Buildings: swissTLM3D © swisstopo, 2012 Basemap: CP25 © swisstopo, 2008 (DV084371)

Figure 13: Clusters of large structures

Buildings with a "complex" shape

Based on the compactness of shapes, it is difficult to link the buildings in this group to particular buildings in an urban environment. Thus, they are found all over the agglomerations. The two clusters in this group are separated according to the size of the buildings. Some of the "complex" shapes are houses (Figure 14a) and the other are larger buildings (Figure 14b).

a b

Buildings: swissTLM3D © swisstopo, 2012 Basemap: CP25 © swisstopo, 2008 (DV084371)

Figure 14: Clusters of more complex structures

5.1.2 Differences between agglomerations

The statistical agglomeration B excluded, differences across the definitions are not on the general typology but more on the distribution of buildings in it. Some structures may change from one cluster to the other but they are a minority. This shows that the common core of these agglomerations is the same and the inclusion or exclusion of some parts of the city does not change the typology. Therefore, buildings in those parts are not different (or typical) enough to have a new class emerge in the typology.

For *StatB*, there are more clusters, 15 in total. Even with this higher number, the general categories found for the other definitions are still visible. Some differences are made within the categories, with, for example, a higher diversity in the "complex" buildings clusters (Figure 15). The difference is not only on the size of the structure, but on some other attributes such as the number of lines in the skeleton.

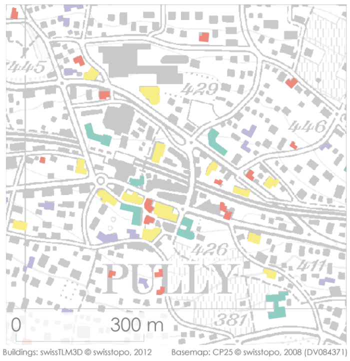


Figure 15: Buildings in the "complex" group for StatB

As said previously, those slight modifications excepted, structures at the building level are quite similar from one agglomeration to the other. The classes resulting from this analysis can be linked to buildings with a function typology (houses, apartment complexes, industrial warehouses, etc.) but some may explain only the shape of structures.

5.2 Structures at the composition scale

At this scale, the indicators relate the building to its immediate environment, mainly the block, streets and neighbouring buildings. The DBSCAN and K-Means clustering have been performed using all the defined indicators for this scale.

5.2.1 General overview

The DBSCAN method set aside a number of buildings with really high values for some of the indicators. Once again buildings in the center are discarded, mainly because they occupy the entirety of their blocks (Figure 16b), thus scoring high on the ratios between the area and the influence zone and/or the block areas. Buildings located in large blocks (Figure 16a) or the ones far from other buildings - whether in the same block or not - (Figure 16c)are also not considered for the K-Means clustering.

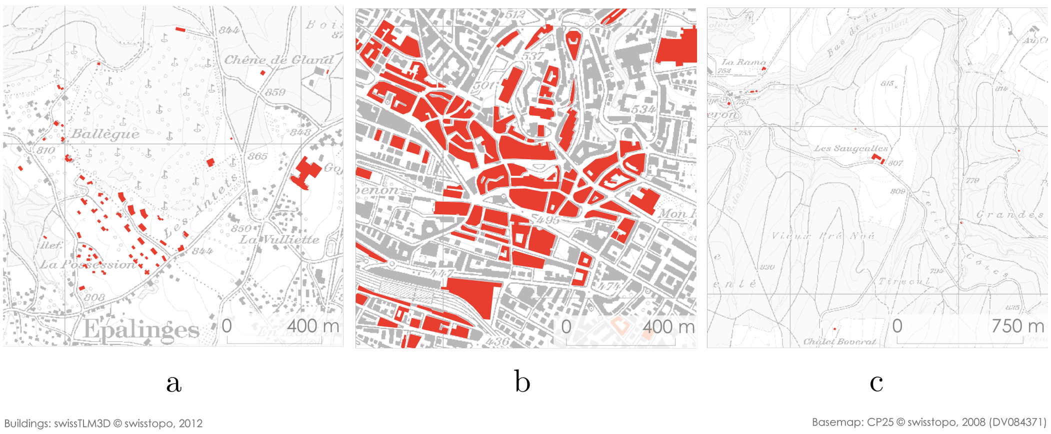


Figure 16: Buildings set aside at the composition scale

On the statistical agglomeration B, the DBSCAN also yields the above-mentioned noise points but distinguishes two more clusters of 10 and 18 buildings. The former is composed of very isolated structures in large blocks and the latter correspond to the buildings located in the biggest occupied block of the agglomeration (nearly 4 km^2).

Contrary to the building level, the clustering at the composition scale is quite inconclusive. There is an important trade off to be made between the number of clusters and the part of explained variance.

For the functional definitions of the agglomeration, the number of clusters needed to explain at least two third of the variance is around fifteen. For the same number of clusters, the explained variance is a bit over 70% in the statistical definitions.

Moreover, the number of indicators is smaller than the one in the building scale and some are common to the buildings of the same block. This causes some difficulties when trying to analyse the content of each cluster.

Some of the clusters found regroup buildings from the biggest blocks, buildings located directly next to roads, or buildings occupying a large part of their blocks. Those are rather simple interpretations to make, but the other clusters are a lot harder to explain (Figure 17), without forgetting that the explained variance is not sufficient to ascertain their meaning. Some neighbouring buildings are sometimes in different clusters only because one of them is closer to a building in another block, which can still be several dozens metres away.

Figure 17: Clusters at the composition scale

For all these reasons it is complicated to make a robust typology emerge and to compare the results across the agglomeration's definitions.

5.3 Structures at the neighbourhood scale

For this last scale, the focus in on the reachable area surrounding a building and its characteristics. The indicators used reflects on buildings and streets attribute with some varying methods to define the reachable area.

5.3.1 General overview

As for the previous scale, the DBSCAN has been performed on all indicators to make the first sort. Once again the majority of buildings in the center have been considered as noise (Figure 18a). This can, in part, be explained by the same reasons as for the buildings scale, that is they have big area and floorspace values.

Other elements are eliminated at this step. The buildings near those big structures are influenced by their attributes and set aside. Smaller urban core (Figure 18b) are also eliminated due to their (pedestrian) street network and higher densities, which lead to bigger values in regard of the number of buildings and intersections.

Finally, the last group of buildings which do not make the cut is more heterogeneous. It consists of buildings located near special features in the road network, such as those buildings located next to a cemetery which alleys are included in the street network (Figure 18c). This yields very high values on some attribute, thus leading to the removal of the buildings.

To further analyse the remaining buildings, several combinations of the indicators have been tested for the clustering. K-Means on all attribute led to some really poor results with just over half the variance explained for a reasonable number of clusters used.

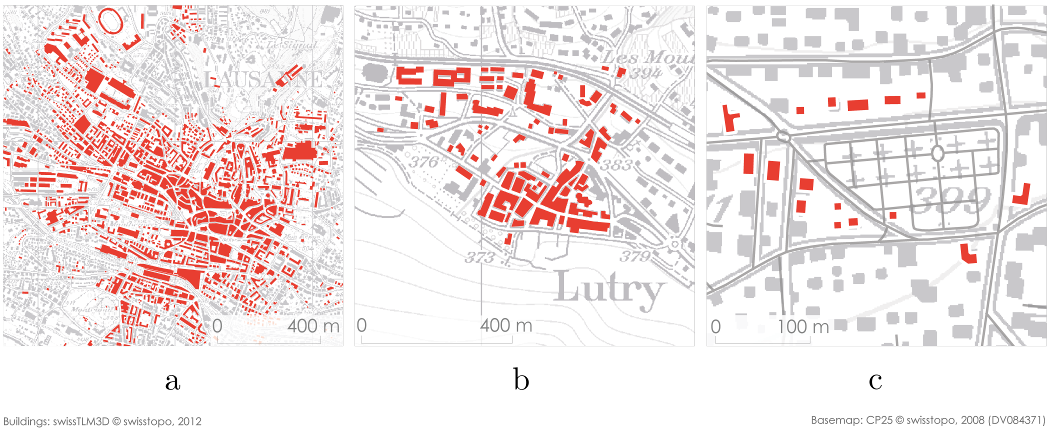


Figure 18: Buildings set aside at the neighbourhood scale

Amongst the tried associations, two have been retained : all indicators related to the driving distance and all the ones related to the 500 metres radius and walking distance. Clustering on the former explain around 85% of the variance for the agglomerations but uses only 6 attributes. Applied on the 12 indicators related to the 500 meters threshold, K-Means results explain between 71% and 82% of the variance.

Neighbourhoods according to driving time

Generally speaking, clusters tend to group buildings situated next to each other, thus forming large sets. Those sets are not unique in the agglomeration and can be found at several places.

Five of the clusters are located around the main center and have high values for the reachable number of buildings and floorspace (Figure 19). The two clusters in the vicinity of the center are differentiated by the size of their catchment area and the length of the road network lying in it. The three others are a bit further away from the core center, with one including buildings next to main and secondary roads while the other has the buildings located in the middle of neighbourhoods or next to natural barriers. The final clusters has higher values in number of dead ends, intersections and length of network which may indicate places with more alleyways.

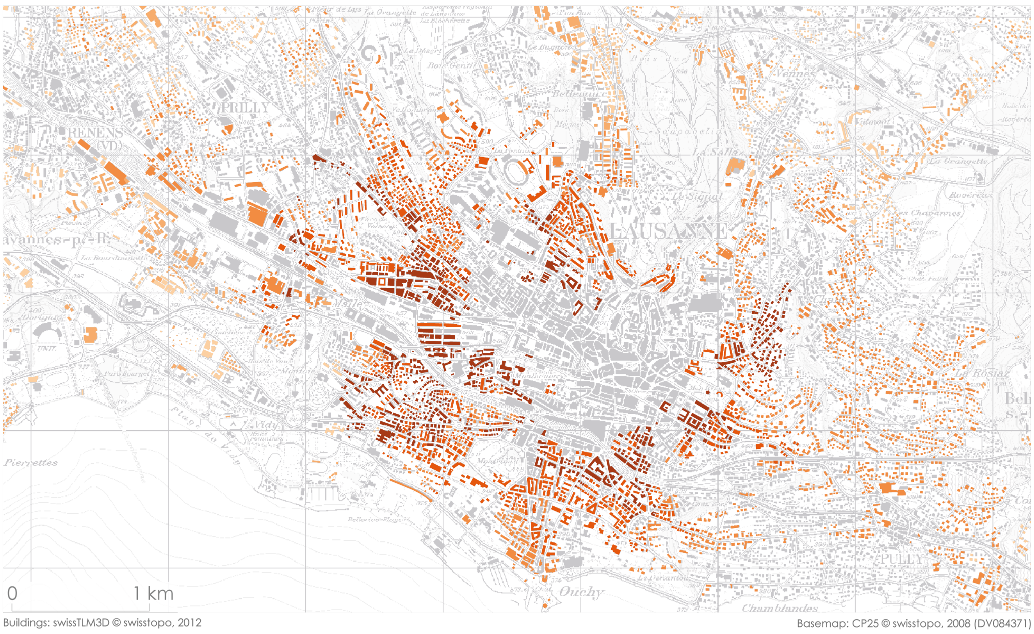


Figure 19: Five clusters around and near the core center

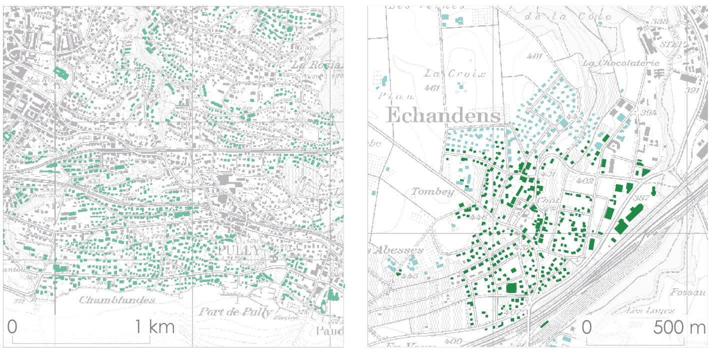
Two other clusters also have a really high number of reachable buildings. The first one is found in the secondary centres of Pully, Prilly, Renens, and Chailly. Contrary to clusters in the main centre's proximity, the buildings here have smaller values of floorspace thus indicating tinier (in size and/or height) structures in the reachable areas (Figure 20a). The second one has values well higher than average for the six attributes. This case can be explained by the location of buildings in this cluster. They are located near main arteries and highways which, thanks to the maximal authorised speed, give them a big catchment area (Figure 20b). Therefore, all indicators have higher values due to the size effect induced by these particular locations.

a b

Buildings: swissTLM3D © swisstopo, 2012 Basemap: CP25 © swisstopo, 2008 (DV084371)

Figure 20: Clusters in secondary centres and near main roads

Three more clusters are located either at the fringe of the agglomeration or near physical barriers. The buildings are distributed in the separate clusters mainly based on the size of their catchment area. A cluster regroup buildings next to physical barriers but still in quite dense areas (Figure 21a). The other two are in less densely built areas and are distinguished by the proximity to the border / physical barrier (Figure 21b). Sometimes, the difference can be between buildings along main roads and the ones in more remote neighbourhoods.



a b

Buildings: swissTLM3D © swisstopo, 2012 Basemap: CP25 © swisstopo, 2008 (DV084371)

Figure 21: Clusters near physical barriers or in the fringe

Finally, the two remaining clusters are quite different from the rest. The first one regroup buildings with a lot of dead ends in their surroundings (Figure 22a). Buildings in this cluster are anywhere but in the center of the agglomeration. The final one regroup buildings with really low values in all attributes. They can be found in neighbourhoods depending on a long road to connect to the principal arteries or in the middle of other clusters where there are some one way roads drastically reducing their catchment area (Figure 22b).

On the whole, it is quite difficult to name those clusters, maybe with the exception of secondary centres. However, while looking at the mapped results, the pattern are easier to see, even if some exceptions remain.

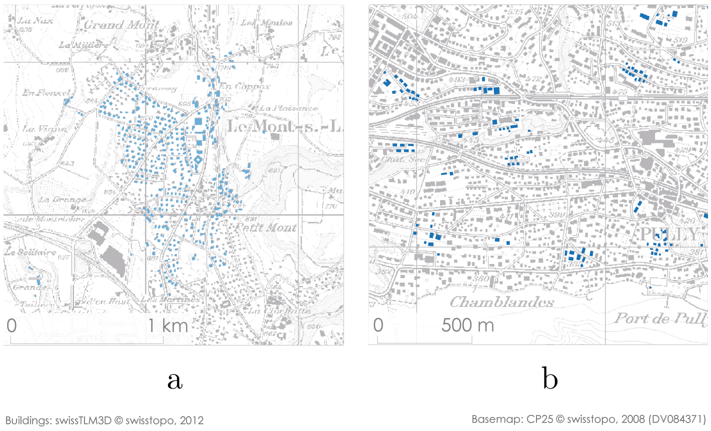


Figure 22: Atypical clusters

Neighbourhoods in a 500 metres distance

Those 500 metres are taken both along the road network and as the crow flies. In general, this distance gives smaller catchment areas than the ones computed with the driving distance. This has an influence on the results, with more compact-looking neighbourhoods and finer differences near the limits of the agglomerations or physical barriers.

Compared to the classification based on the driving distance, there also is a large number of clusters based on the densities. Contrary to the previous classification, the secondary centres also have a radial distinction in the densities (Figure 23) and are better delimited.

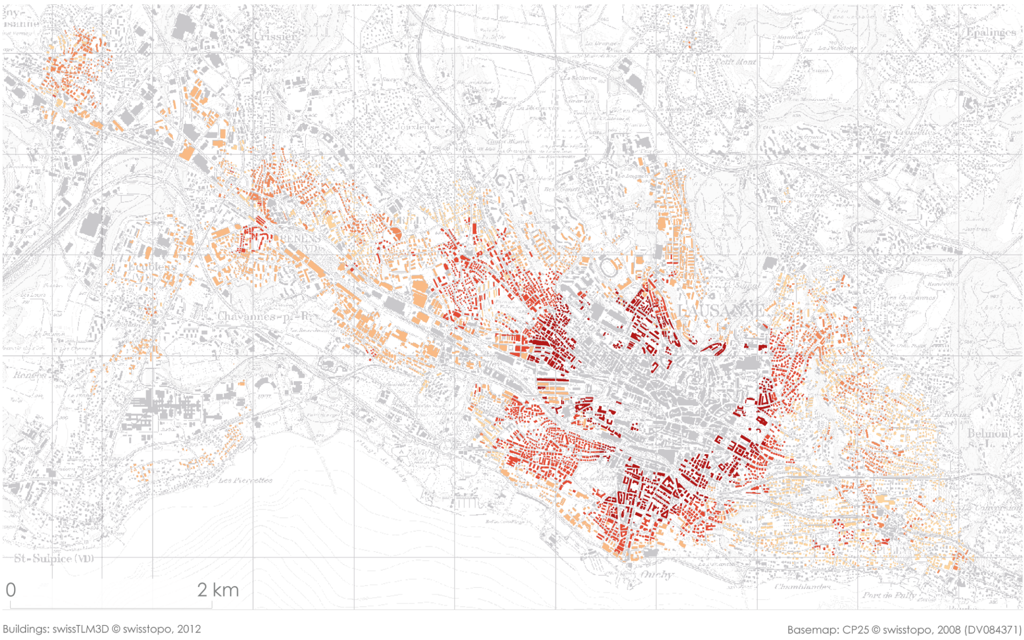


Figure 23: Clusters relying on densities for the walking distance

In less densely built areas, the same type of radial structures can be found. On Figure 24, the usage of the 500 metres radius circles can be seen quite well. This shows a radial structure (slightly mitigated by the network distance) with an important influence of the borders. This type of organisation in neighbourhoods can be found throughout the agglomeration where densities are smaller, for example in residential areas or near open areas.

The two remaining clusters are separated due to their road network characteristics. The first one has neighbourhoods in places with a long network of roads and intersections (Figure 25a). The other is the contrary, it group buildings in places with few roads and thus low values for

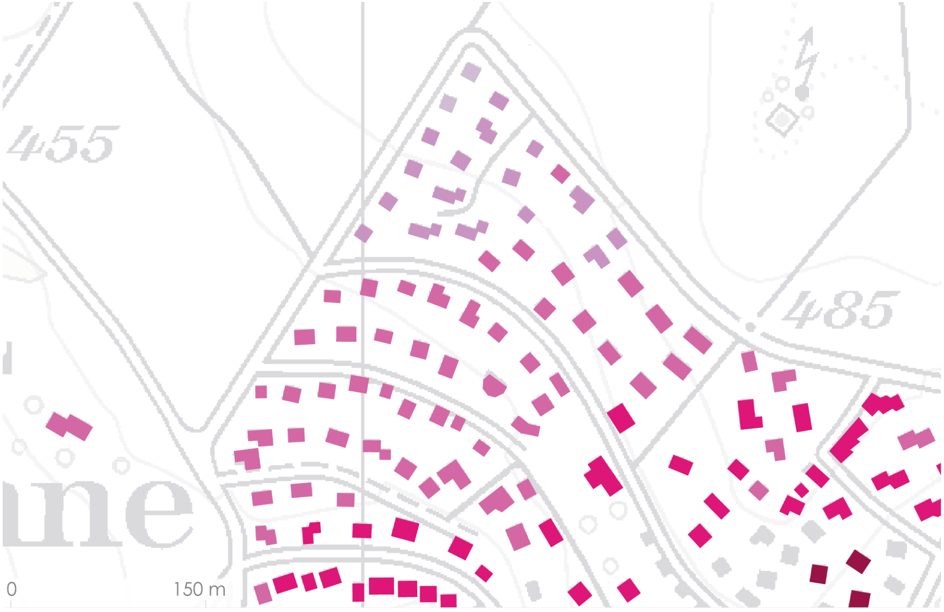
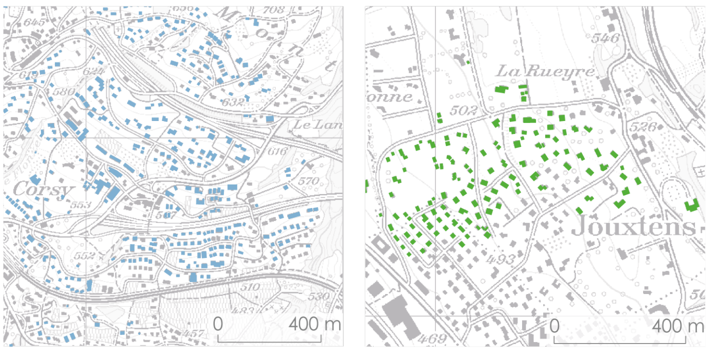


Figure 24: Radial structures near the borders

these attributes (Figure 25). However, for reasons addressed in the section 6.2, it would be too adventurous to consider these as significantly different from the other low-density clusters.

On the whole, the results obtained are somewhat different depending on how the notion of neighbourhood is considered. The results obtained for the driving distance are more varied than those of the walking and circle distance. The homogeneity showed in the latter does not necessarily reflect better results to determine coherent urban structures at this scale, as Figure 24 demonstrates.



a b

Buildings: swissTLM3D © swisstopo, 2012 Basemap: CP25 © swisstopo, 2008 (DV084371)

Figure 25: Clusters in neighbourhoods with and without lots of intersections and roads

5.3.2 Differences between agglomerations

Once again, results by agglomeration do not differ significantly. This time, the statistical agglomeration B is not the exception that proves the rule. The buildings in smaller villages tend to be grouped like the houses in residential areas of the smaller agglomerations (Figure 26).

The main differences are in the CCA agglomerations where some more distinctions are found in the low-density areas near physical barriers. Those fluctuations do not improve the classification as a whole and tend to divide the zones near obstacles in even smaller chunks of neighbourhoods (Figure 27).

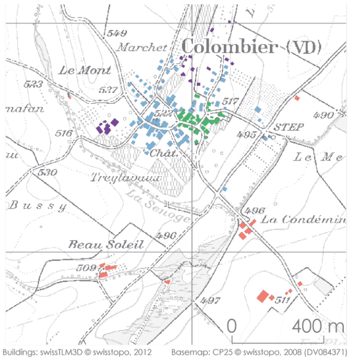


Figure 26: Small village divided like a city neighbourhood

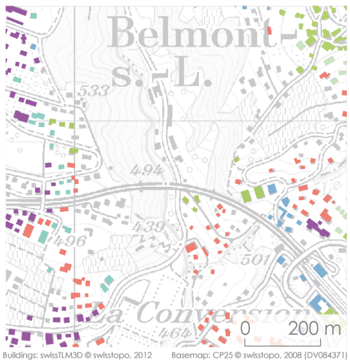


Figure 27: Divisions near physical barrier for the CCA agglomerations

Chapter 6

Discussion

In the first part of this chapter, the results previously detailed are used to answer the hypotheses formulated in the chapter 3. The second part details the encountered limits that may temper the strength of the analysis.

6.1 Results in regard with the hypotheses

Knowing the results obtained for the three scales and the differences across the agglomerations it is possible to provide answers to the hypothesis.

The first one stated that the selected indicators contributed to the distinction of types in urban structures. For the building scale, indicators relative to the orientation and bounding boxes of elements have only added subtle differentiations between buildings of the same type, but oriented otherwise.

For the composition scale, all indicators have been used for the clustering.

Finally, for the neighbourhood level, several tries on the whole set of indicators have been inconclusive. The best results were obtained while considering indicators in the biggest surroundings, namely the 500 metres (walking or radius) threshold and the two minute driving distance. Unfortunately, this means that it is not possible to distinguish the close environment (100 - 300m) of buildings using the computed indicators. This show the need to consider the neighbourhood in its larger definition, at least for Lausanne.

Therefore, this first hypothesis can only be considered as partially true, with some indicators having played little to no part in the distinction of urban structures at the three scales.

The second hypothesis assumed that the resulting clusters would differ across the scales and that their coherence is a function of the level. At the building scale, results identified the shape of buildings and some of them could be linked to the type of the buildings (houses, apartment complexes, industrial warehouse, etc.). This possibility to attach this shape typology to existing types of building shows a certain rationality in the group obtained. The pertinence of differentiating so-called complex structures, which can be any type of building, remain up to discussion.

Results obtained at the composition scale were not conclusive. This may be due to the selection of indicators or the important variance in blocks size and structure in the region. Schirmer and Axhausen (2015) questioned the relevance of this scale, (non) results obtained in this work abound in this way. Instead of simply abandoning this scale, it could be interesting to find other indicators or maybe try them in cities with chequered road.

In the search of quantified neighbourhoods, results were mitigated. Clusters obtained using the driving distance highlighted the differences of accessibility around the center, along the main arteries, and in the suburbs. The identified groups were quite difficult to name. While they showed coherent structures in their delimitations, they could hardly be considered as neighbourhoods as we know it. The walkable neighbourhood was marginally different with more compact neighbourhoods and finer distinctions in the fringes.

As results were different across the scales, the first part of this hypothesis is accepted. The second part is more nuanced as results were obtained in two of the three scales. For these ones, obtained clusters were in line with the indicators used but did not necessarily correspond to expected outcomes (*e.g.* entire neighbourhood in the same cluster, same type of buildings together, etc.). In hindsights, using the clusters from the building scale to help define neighbourhoods could have contributed to neighbourhoods grouping same type buildings. The use of socio-economic indicators could also be tried to refine the limits of the clusters (if similar people live close to each other).

The last hypothesis stipulated that the results would be different across the agglomeration's definitions. At the building scale, outside of slight changes in the distribution in the clusters, only the statistical agglomeration B was different in the number of clusters. Nonetheless, those changes were mainly some more nuances in two of the identified groups.

As results at the composition scale were inconclusive, it was not relevant to compare the results across the agglomerations.

As for the building scale, results at the neighbourhood level were only lightly changed for the CCA agglomerations. This led to more divisions in already cluttered neighbourhoods.

Given these observations, this hypothesis is rejected. Even if the task of comparing results for different definitions was appealing, the shared core of the agglomeration is so important that no particular areas tend to stand out. This can also mean that more rural areas are not that different from neighbourhoods on the edge of the city.

6.2 Limits

Answers to the hypotheses presented above should be nuanced with the limits arising from the defined methodology and the quality of data. As we try to get close to reality with models, choices made to approximate it at best have some repercussions on other aspects of the work. These limits are detailed below according to the scale they affect.

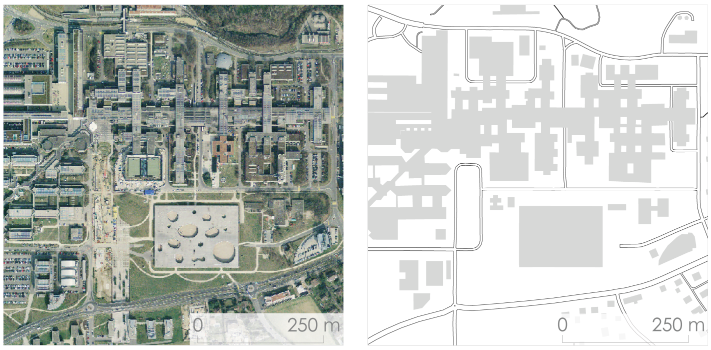
Building scale

Indices at this scale were solely based on the shape of the buildings and several factors have influenced these indicators. The coordinates of the buildings in the swissTLM3D dataset have a precision of 5 digits even if the announced error margin is between 1.5 and 0.2 metres. This has an influence of some skeletons of shapes which have more complicated geometries because of the locations of some points. Moreover, this excessive precision is quite strange in regard of the digitisation of some buildings.

Examples of the differences in the precision of building can be found everywhere in the agglomeration and are quite hard to spot. Sometimes identical buildings (identified as such with aerial photography) do not have the same footprint in the dataset. This can lead to identical buildings in reality being classified in different groups. Figure 28 is another example of some disparities in the quality of data. On the first structure, holes are present and numerous whereas for the second, there are only some gaps in the perimeter and no holes at all.

Buildings used for the analysis are the composites, which joins all touching buildings into one structure. This way of considering building is quite appropriated when considering dense dwellings in the center of cities. It may have less expected outcomes in a few other places. Some composite regroup detached houses and their garage or separate buildings of different heights with roofs superposed. These approximations affect not only all indices related to the shape but also the volumetric ones. As datasets were crossed to retrieve the heights of buildings, it is possible that whole structures are assigned the value the smallest element in it. This can lead to detached houses with adjoining garages having heights of around 2 metres.

Those problems of height also influence the floorspace values which are computed using the number of story, derived from the height. Some buildings can therefore have smaller than expected values. The contrary is also true by design. For example, large warehouses have really high values of floorspace when, in fact, they mainly consist of only one tall story to accommodate for all the machinery or vehicles.



a b

Aerial photo: SWISSIMAGE 25cm © swisstopo, 2015 Buildings: swissTLM3D © swisstopo, 2012

Figure 28: Digitalisation differences

Composition scale

As seen in the previous chapter, results at this scale are not conclusive, some of the following elements may explain a part of this outcome.

The retained definition of the block - the space encompassed by roads - has raised some issues when creating the relevant geometries. First of all, the width of the roads are always the minimum possible to avoid buildings being on roads. This precaution implies that roads can be up to 4 metres smaller than their true width, which has quite an impact on the size of blocks. Secondly, only some type of roads have been considered: roads without particular structures (tunnel, underpass), bridges, and steps. They were considered as "block breaking" ways. Finally, some buildings span across multiple blocks. There are (pedestrian) ways at ground level going under higher parts of a building, thus closing the block.

Perception of buildings from the streets is a theoretical measure of buildings being less than 3 metres from the block border. As mentioned above, road width is quite conservative and no difference is made on the kind of space between the road and the building. A shopping building with windows next to the pavement is more perceptible than a house separated of the road by a fence, a hedgerow, and some trees. This measure may be too conceptual to actually reveal relevant tendencies.

Neighbourhood scale

This level is subject to limits on two fronts: computation of the reachable area and computation of the attributes in the area.

The routing configuration was thought to find the best compromise between accuracy, data availability, and computation time. As with all compromises, this led to some drawbacks detailed hereafter. All buildings have been assigned a starting node on the road graph. It corresponds to the node nearest to the building. If this works well when driveways are mapped up to buildings, it is quite flawed in some numerous other cases. Buildings located along a road without intersections will be assigned one of the two points at the extremities of the segment. Some buildings may be attributed a starting point directly on a highway or on the other side of a property. All theses approximations make a notable difference because some routing queries start precisely at the doorstep of buildings whereas some have a head start of several dozens metres or a speed advantage when starting from a highway.

A distinction was made to use only the drivable network for car routing. As old city-centres are mainly pedestrian zones the starting points tend to be away from the buildings themselves. Worse, some of the roads in the middle of these zones are mapped as normal roads Figure 29. If the starting point is on this road, the maximal reachable area will consist of only the road because cars are not allowed in pedestrian areas. For walking routing, the usage of the whole graph could be discussed because it means that pedestrians are allowed on highways or other streets without pavements. This can be nuanced by the fact that on short distances it has close to little impacts.

Figure 29: Drivable street isolated in a pedestrian area

Another difficulty was to find a reasonable time-distance for the driving routing. As no obstacles are taken into account, the trip is calculated at the maximal authorised speed for each segment. These unrealistic rides sometimes give buildings with starting points on the main axis (limited to 80 and 120 km/h) exaggerated areas because of the elongated α-shape obtained. It would be possible to refine the definitive shape but, depending on the set of reachable points, it can be a costly computational operation. Therefore, reachable areas better be considered as a little bit exaggerated estimation rather than a precise and definitive catchment area.

Finally, regarding the quality of data, the swissTLM3D sets for the road is simpler than the available OSM data. This has had an influence regarding the network attributes used at this scale. The driveways and other smaller roads in neighbourhoods tend not to be mapped, which in turn lead to less intersections, dead ends, and a smaller network length in those areas. Figure 30 shows what a neighbourhood looks like in reality (a), how the roads are mapped in the swissTLM3D data set (b), and the network mapped by OSM users (c). It is easily conceivable that, with fully mapped ways and roads, some neighbourhoods would be classified otherwise. For example, neighbourhoods with a lot of detached houses would have driveways leading to the houses and therefore a high number of dead ends and intersections. Another flaw introduced by the selection of only certain type of roads is the creation of artificial dead ends where there are tunnels, underpasses, or other structures not initially considered. For example, the highway entrance/exit (Figure 31b) yields a lot of dead ends because not all roads are connected (Figure 31a).

Figure 30: Differences in the street network compared to the reality

Globally the quality in the details of swisstopo's data set was below par in comparison to OSM's data. Unfortunately the vast majority of these limits was discovered during the analysis

Figure 31: Missing links in the network

and could not be attenuated. For future works, it may be interesting to compare the available data and to try and reduce the limits, even if it means to concentrate on only one scale.

General limits

The indicators used have been chosen based on the work of Schirmer and Axhausen (2015) with some added (notably several compactness indicators) and a few removed. This feature selection was made to have comparable results and lacks a critical view on the sets of proposed indicators. For example, the five compactness measures are quite correlated and generally tend to distinguish the same buildings. There is also a certain redundancy in the comparisons to the perimeter and area of several forms associated to the building's shape.

As the clustering was unweighed, all indicators were considered to have the same quality to help differentiate buildings. This is not the case as results showed some classes separated by differences in indicators which pertinence were questionable. A further analysis could be done on the extraction of the relevant features to emphasize some major aspects which may not have been given a sufficient weight in this work.

Chapter 7

Conclusion

This paper set out to try to find urban structures at different scales. Three sets of indicators have been computed to quantify attributes. They were based solely on the buildings' shapes and the roads and ways network. Buildings were then partitioned into groups using a couple of clustering algorithm.

To see if the extent of the considered urban environment has an influence on the outcomes, several definitions of Lausanne's agglomeration were retained. These definitions are based on varying criteria emphasising the density, the built continuity, or the proximity to determine what is urban or not. Results were compared across the agglomerations.

At the building scale, some of the indicators were outed as non relevant. They were characterising the orientation of the structures. Some of the obtained clusters can be linked to concrete type of buildings, such as apartment complexes or detached house. Some other focus on the complexity of the shape. Overall, results at this scale were quite coherent even if the level of details was not consistent in the original data.

At the composition scale, results were inconclusive. This level had the fewest number of indicators and the lowest explained variance. Only a few of the retrieved clusters could be explained clearly. This absence of a coherent result is in line with previous studies at this level.

The neighbourhood level was based on the accessibility around buildings. Several ways to compute the accessibility have been used but only two were relevant. Smaller distances of 100 and 300 metres did not permit to certainly distinguish between buildings. Using a 500 metres distance yielded clusters who could be separated in three large groups: high-density areas, low-density areas, and areas with peculiar network features. The results in neighbourhoods near physical barriers were more detailed than for the clusters computed in the driving time. The latter returned some different results, with secondary centres put in evidence. Both of those results were quite coherent but did not really fit in the general representation of a neighbourhood one can have. Unfortunately, several elements related to the quality of the data and the choices made in the methodology have influenced some of the outcomes.

Surprisingly, with some little exceptions, results were substantially similar for all the different agglomeration's definitions studied.

The results obtained show the potential of a quantitative classification as Schirmer and Axhausen (2015) have proposed. The sets of indicators suggested in their paper may be revised to yield some better results. The integration of different definitions of *the urban* was not rewarded with some persuasive differences but it may indicate that the common built tissue in these definitions is quite important and diversified.

Finally, this quantification of the urban structures has a potential to grow and be improved to fill in the limits detailed at the previous chapter. Some new indicators may appear with the increasing availability of spatial data. The quality of the data is quite important as too few details can give a botched classification and too much details can lead to over specific clusters.

Once the classification based on physical elements is perfected, the introduction of socio-economic factors could open the way to finer definitions of structures in cities. Eventually, the integration of the time factor could show the developments of the urban arrangements and, who knows, predict their future evolutions.

References

Angel, S., Parent, J., & Civco, D. L. (2010). Ten compactness properties of circles: measuring shape in geography. *Canadian Geographer / Le Géographe canadien*, *54*(4), 441–461. doi:10.1111/j.1541-0064.2009.00304.x

Baerwald, T. J. (2010, June 25). Prospects for Geography as an Interdisciplinary Discipline. *Annals of the Association of American Geographers*, *100*(3), 493–501. doi:10.1080/00045608.2010.485443

Bairoch, P. (1991). *Cities and economic development: from the dawn of history to the present* (C. Braider, Trans.). Chicago: Univ. of Chicago Press.

Berry, B. (1964). Cities as systems within systems of cities. *Papers in Regional Science*, *13*(1), 147–163.

Bhat, C. R., Kockelman, K., Chen, Q., Handy, S., Mahmassnani, H., & Weston, L. (2000). *Urban accessibility index: Literature review* (No. TX-01/7-4938-1). Center for Transportation Research, University of Texas at Austin. Austin.

Blum, H. (1967). A Transformation for Extracting New Descriptors of Shape. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form: Proceedings of a symposium*. Cambridge, Ma: MIT Press.

Boyce, R. R., & Clark, W. A. V. (1964). The Concept of Shape in Geography. *Geographical Review*, *54*(4), 561–572. doi:10.2307/212982. JSTOR: 212982?origin=crossref

Clanché, F., & Rascol, O. (2011). Le découpage en unités urbaines de 2010. L'espace urbain augmente de 19 % en une décennie. *INSEE Première*, *1364*, 1–4. Retrieved from https://www.insee.fr/fr/statistiques/1280970

Clifford, N. J., Cope, M., Gillespie, T. W., & French, S. (Eds.). (2016). *Key methods in geography*. London: SAGE.

Conzen, M. R. G. (1960). Alnwick, Northumberland: A Study in Town-Plan Analysis. *Transactions and Papers (Institute of British Geographers)*, (27), 1–122. doi:10.2307/621094

de Berg, M., Cheong, O., van Kreveld, M., & Overmars, M. (2008). Computational Geometry : Introduction. In *Computational geometry: algorithms and applications* (3rd ed, pp. 1–18). Berlin: Springer.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Kdd*, *96*(34), 226–231.

Federal Statistical Office. (2014). *Areas with urban character 2012* (No. 1477-1200). Federal Statistical Office. Neuchâtel. Retrieved from https://www.bfs.admin.ch/bfs/fr/home/statistiques/themes-transversaux/analyses-spatiales/niveaux-geographiques.assetdetail.349566.html

Flint, J. (2009). Neighborhoods and Community. In *International Encyclopedia of Human Geography* (pp. 354–359). doi:10.1016/B978-008044910-4.01065-8

Fotheringham, A. S., & Wong, D. W. S. (1991). The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A: Economy and Space*, *23*(7), 1025–1044. doi:10.1068/a231025

Gauthier, B. (2004). The history of urban morphology. *Urban Morphology*, *8*(2), 71–89.

Girres, J.-F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, *14*(4), 435–459.

Guo, J. Y., & Bhat, C. R. (2007). Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography*, *15*(1), 31–45. doi:10.1016/j.jtrangeo.2005.11.001

Hall, P. (1998). *Cities in civilization*. New York: Pantheon Books.

Hall, T. (2006). *Urban geography* (3rd ed). Routledge contemporary human geography series. London: Routledge.

Hamnett, C. (1991). The blind men and the elephant: the explanation of gentrification. *Transactions of the Institute of British Geographers*, 173–189.

Hartigan, J. A. (1975). *Clustering algorithms*. New York: Wiley.

Lamanna, F., Lenormand, M., Salas-Olmedo, M. H., Romanillos, G., Gonçalves, B., & Ramasco, J. J. (2018). Immigrant community integration in world cities. *PLOS ONE*, *13*(3), e0191612. doi:10.1371/journal.pone.0191612

Latham, A., McCormack, D., McNamara, K., & McNeill, D. (2008). *Key concepts in urban geography*. Thousand Oaks, CA: SAGE Publications.

Le Gléau, J.-P., Pumain, D., & Saint-Julien, T. (1996). Villes d'Europe : à chaque pays sa définition. *Economie et statistique*, *294*(1), 9–23. doi:10.3406/estat.1996.6079

MacEachren, A. M. (1985). Compactness of Geographic Shape: Comparison and Evaluation of Measures. *Geografiska Annaler. Series B, Human Geography*, *67*(B), 53–67. doi:10.2307/490799. JSTOR: 490799?origin=crossref

Marshall, S. (2005). *Streets & patterns*. London: Spon.

Meeder, M., Aebi, T., & Weidmann, U. (2017). The influence of slope on walking activity and the pedestrian modal share. *Transportation Research Procedia*, *27*, 141–147. doi:10.1016/j.trpro.2017.12.095

Moudon, A. V. (1994). Getting to know the built landscape: typomorphology. In K. A. Franck & L. H. Schneekloth (Eds.), *Ordering space: types in architecture and design* (pp. 289–311). New York: Van Nostrand Reinhold.

Moudon, A. V. (1997). Urban morphology as an emerging interdisciplinary field. *Urban Morphology*, *1*, 3–10.

Murphy, D. (2017a, May 11). Where is the world's densest city? *The Guardian: Cities*. Retrieved May 5, 2018, from http://www.theguardian.com/cities/2017/may/11/where-world-most-densely-populated-city

Murphy, D. (2017b, April 19). Where is the world's most sprawling city? *The Guardian: Cities*. Retrieved May 5, 2018, from http://www.theguardian.com/cities/2017/apr/19/where-world-most-sprawling-city-los-angeles

Nguyen, T. T. (2009). Indexing PostGIS databases and spatial query performance evaluations. *International Journal of Geoinformatics*, *5*(3), 1–9.

Office of Management and Budget. (2010). *2010 Standards for Delineating Metropolitan and Micropolitan Statistical Areas*. Federal Register 75, N.123, pp. 37246-37252.

Pacione, M. (2009). *Urban geography: a global perspective* (3rd ed). London ; New York: Routledge.

Palagyi, K. (N.d.). Skeletonization. Retrieved February 7, 2018, from http://www.inf.u-szeged.hu/~palagyi/skel/skel.html

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Vienna. Retrieved from https://www.R-project.org/

Roh, C., Huissoud, T., Schaffner, C., & Cunha, A. (1989). *Rapport sur la délimitation des quartiers et secteurs statistiques*. Office d'études socio-économiques et statistiques. Lausanne.

Rozenfeld, H. D., Rybski, D., Andrade, J. S., Batty, M., Stanley, H. E., & Makse, H. A. (2008). Laws of population growth. *Proceedings of the National Academy of Sciences*, *105*(48), 18702–18707.

Rozenfeld, H. D., Rybski, D., Gabaix, X., & Makse, H. A. (2009). *The area and population of cities: new insights from a different perspective on cities* (Working Paper 15409). Cambridge, Ma: National Bureau of Economic Research.

Saunier, P.-Y. (1994). La ville en quartiers : découpages de la ville en histoire urbaine. *Genèses*, *15*(1), 103–114. doi:10.3406/genes.1994.1234

Schirmer, P. M., & Axhausen, K. W. (2015). A multiscale classification of urban morphology. *Journal of Transport and Land Use*. doi:10.5198/jtlu.2015.667

Schuler, M., Dessemontet, P., Joye, D., & Perlik, M. (2005). *Les niveaux géographiques de la Suisse*. Neuchâtel: Federal Statistical Office.

Stauffer, D., & Aharony, A. (1994). *Introduction to percolation theory*. London: Routledge.

Swiss Federal Council. (2015). *Politique des agglomérations 2016+ de la Confédération* (No. 18 février 2015). Swiss Federal Council. Bern.

Tannier, C., Thomas, I., Vuidel, G., & Frankhauser, P. (2011). A Fractal Approach to Identifying Urban Boundaries. *Geographical Analysis*, *43*(2), 211–227. doi:10.1111/j.1538-4632.2011.00814.x

United Nations. (2015). *World Urbanization Prospects. The 2014 Revision* (No. ST/ESA/SER.A/366). UN, Department of Economic and Social Affairs, Population Division. New York. Retrieved from https://esa.un.org/unpd/wup/Publications/Files/WUP2014-Report.pdf

Appendix A

Data sets

The present work uses proprietary and open access data to perform the analyses. Those can be replicated using the procedures described in the online repository, which also contains links to the data sets.

Swisstopo

The TLM3D data set (https://shop.swisstopo.admin.ch/fr/products/landscape/tlm3D) has multiple types of features arranged in several categories. Three of those categories have been used in this analysis: the roads and ways, the buildings, and the lakes' contours. Samples of the data can be found at the address mentioned above.

The swissBUILDINGS3D 1.0 data set (https://shop.swisstopo.admin.ch/fr/products/landscape/build3D) was used to attribute a height to the TLM3D buildings. It is no longer updated but samples can still be found at this address.

The swissBOUNDARIES3D data set is freely available on the website for the swiss open data (https://opendata.swiss/fr/dataset/swissboundaries3d-landesgrenzen). It was used in the analysis to determine which municipalities were in the statistical agglomerations.

OpenStreetMap

Data added in OpenStreetMap can be downloaded directly from the website or via third parties such as Geofabrik or OverPass Turbo. As the map is updated nightly, the more recent the download, the more precise the data (technically...). The OSM extract used for the analysis was downloaded on April 9 2018.

Federal statistical office

Two data sets of the national statistical office were used to delineate the CCA agglomerations

- Census statistics concerning the population at the hectare scale (https://www.bfs.admin.ch/bfs/fr/home/services/geostat/geodonnees-statistique-federale/btiments-logements-menages-personnes/resultats-recensement-depuis-2010.assetdetail.3543467.html)
- Business census statistics concerning the enterprises and available at the hectare level (https://www.bfs.admin.ch/bfs/fr/home/services/geostat/geodonnees-statistique-federale/etablissements-emplois/statistique-structurel-entreprises-statent-depuis-2011.assetdetail.3303058.html)

The final data set is about the urban character of space as defined in 2012 by the FSO. It was used to delineate the statistical agglomerations (https://www.bfs.admin.ch/bfs/en/home/statistics/catalogues-databases/press-releases.assetdetail.188853.html).

Appendix B

CCA at the Swiss scale

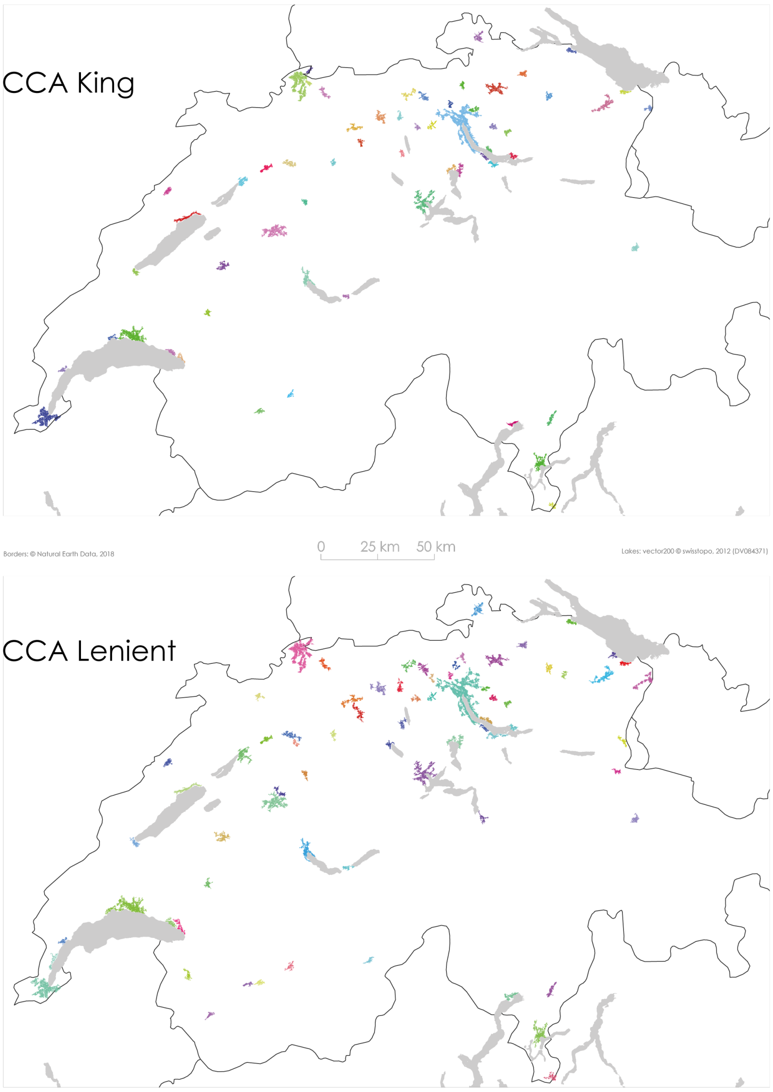


Figure 32: Results of the CCA at the Swiss scale

Appendix C

Indicators used

Short name	[Unit]	Description
b_perim	[m]	Perimeter of building
b_area	[m^2]	Area of building
b_r_vol_fac	-	Volume to facade ratio
b_maxEdge	[m]	Longest edge of building
b_minEdge	[m]	Shortest edge of building
b_stories	-	Number of stories
b_floorsqm	[m^2]	Floorspace area
c_Miller	-	Miller's compactness
c_Schumm	-	Schumm's compactness
c_Haggett	-	Haggett's compactness
c_LeeSallee	-	Lee & Sallee' s compactness
c_Ehrenburg	-	Ehrenburg's compactness
bb_perim	[m]	Perimeter of the bounding box
bb_area	[m^2]	Area of the bounding box
bb_length	[m]	Longest edge of the bounding box
bb_width	[m]	Shortest edge of the bounding box
bb_r_lw	-	Bounding box's edges ratio
bb_r_area	-	Bounding box area to building area ratio
bb_r_perim	-	Bounding box perimeter to building perim.
cc_rad	m	Radius of the circumbscribed circle
cc_exch	-	Value of the exchange index
cc_detour	-	Value of the detour index
ch_area	[m^2]	Area of the convex hull
ch_perim	[m]	Perimeter of the convex hull
ch_r_area	-	Convex hull area to building area ratio
ch_r_perim	-	Convex hull perimeter to building perim. ratio
s_deadend	-	Number of dead ends in the skeleton
sc_lines	-	Number of lines in the centerline
sc_length	[m]	Length of centerline
sc_orien	-	Number of orientation in the centerline
sc_l_sn	[m]	Total length of centerline with orientation S-N
sc_l_ew	[m]	Total length of centerline with orientation E-W
sc_l_nesw	[m]	Total length of centerline with orientation NE-SW
sc_l_senw	[m]	Total length of centerline with orientation SE-NW
sc_m_orient	-	Main orientation of centerline
m_corndis	[m]	Average distance from corners to centroid
m_court	-	Number of inner courtyards
m_court_area	[m^2]	Area of courtyard
m_court_rel_a	-	Area of courtyard to area of building ratio

Table 2: Indicators for the building scale

Short name	[Unit]	Description
cp_fac_lgt	[m]	Length of facade perceived from the street
cp_fac_lgt_rel	-	Part of the building's perimeter perceived from street
cp_d_nxtbuil	[m]	Distance to nearest building
cp_d_builnxtblk	[m]	Distance to nearest building in another block
cp_infzn_area	[m^2]	Area of influence zone
cp_r_ftp_infzon	-	Area of building to influence zone area ratio
cp_d_blklim	[m]	Distance to the block limit
cp_blk_area	[m^2]	Area of the block
cp_r_ftp_blk	-	Area of building to block area ratio
cp_permeab	-	Number of views between buildings
cp_closeness	-	Sum of street facing facades to the perimeter of the block ratio

Table 3: Indicators at the composition level

Short name	[Unit]	Description
Indicators for a 100m, 300m, 500m radius and walking distance		
de_ar	[m^2]	Sum of buildings' areas in reachable area
de_bd	-	Number of buildings in reachable area
de_sq	[m^2]	Sum of building's floorspace in reachable area
net_de	-	Number of dead ends in reachable area
net_int	-	Number of intersections in reachable area
net_lgt	[m]	Total length of network in reachable area
Indicators for a 2 minute driving distance		
net_dd_carea	[m^2]	Surface of catchment area
net_dd_bd	-	Number of buildings in reachable area
net_dd_sq	[m^2]	Sum of building's floorspace in reachable area
net_dd_de	-	Number of dead ends in reachable area
net_dd_int	-	Number of intersections in reachable area
net_dd_length	[m]	Total length of network in reachable area

Table 4: Indicators at the neighbourhood scale