JTLU

A multiscale classification of urban morphology

Patrick M. Schirmer

Institute for Transport Planning & Systems, ETH Zurich

patrick.schirmer@ivt.baug.ethz.ch

Kay W. Axhausen

Institute for Transport Planning & Systems, ETH Zurich

axhausen@ivt.baug.ethz.ch

Abstract: Various studies in the field of urban planning and design have given recommendations for "good urban forms," suggesting that specific spatial characteristics inform the quality of an urban landscape and the way people perceive and behave in them. When modeling spatial behavior in the form of location choice models or hedonic prices, we should reflect these spatial characteristics through the integration of quantitative attributes such as model variables, which is currently only done in a very limited way.

The increasing availability of disaggregated geodata enlarges the options to characterize urban morphology in the form of such attributes. The question for the researcher is which attributes are most useful to reflect characteristics of urban morphology and how can they be processed from the given data.

In this paper, we want to address this issue and give an overview of quantitative descriptions of urban morphology. We base our work on a data model that is simple enough to allow for reproducibility in any study area. These attributes are classified in multiple scales to reflect different perceptions of urban morphology. In a case study on the canton of Zurich, we furthermore prove how these characteristics allow for the definition of
urban typologies at different scales.

Keywords: Urban Form; Typology; Morphology; Cluster Analysis; Urban Simulation

1   Introduction and state of the art

Urban morphology is the study of the form of human settlements at different scales in order to understand the spatial structure and character of these as well as the process of their development. Analyses are often undertaken using cartographic data or geodata that represent the physical structure of the built environment, which can be reduced to the buildings, open space, and connecting networks of streets. This environment defines the outline of its possible usage and the options for possible activities—it is the initial setting in which enterprises, individuals, and points of interests can locate.

One might expect this setting to be important for the character of a certain area and useful for understanding behavior in space. However, at the present time, spatial models of behavior such as location choice and hedonic price models often integrate only very limited aspects of urban morphology. We expect that this is due to missing best practices on how to quantify the characteristics of urban morphology.

As Clifton et al. (2008) stress, the characteristics of the urban environment are multiple and depend on the point of view, such as the discipline and purpose of a study. Many aspects of this system are subjective perceptions that are difficult to quantify, while some others allow for quantitative and deterministic descriptions.

Throughout the 20th century, we find an emerging interest in analyzing the urban fabric, defining models of the ideal city, and experimenting with its form and layout; among others, Ebenezer Howards' Garden City, the Charta of Athens, and the Plan Voisin of Corbusier are probably the best-known examples.

However, some of these newly developed ideas created places that are now described as being unpleasant or having a low quality of life, often linked to social problems. Researchers began to link life in the city with its shape and the distribution of uses and gave recommendations on the developments of cities (Jacobs 1961; Katz 1993; Lynch 1984). Furthermore, rather young disciplines focused on the planning and development of cities were established in the form of urban planning and urban design. The focus of these disciplines lies in the functional and morphologic composition of urban space at different scales.

Meanwhile, other research fields such as economics and transportation developed an interest in representing the processes inside the urban landscape through mathematical models. Von Thünen (1826) formulated one of the first models with a simple representation of a monocentric city explaining the distribution of prices and land use around a city. Later studies advanced this rather simple model to form a variety of economic models, transport models, and land-use models. Good overviews are given by Wegener (2004), Batty (2008), and Maguire et al. (2005). The introduction of discrete choice models allowed for the modeling of urban processes through the behavior of its actors, such as its residents or enterprises (McFadden *et al.* 1978; Train 2009). As discussed above, this is expected to be closely related to urban morphology. Furthermore, we would expect that the integration of morphologic characteristics as variables in such models will allow for the evaluation of design proposals in the related disciplines of architecture, urban design, and urban planning. This forms an essential step for the practical use of these models as planning support tools.

The representation of the urban environment using independent variables in location choice models has mainly been simplified to the socioeconomic distributions available in census data (Schirmer et al. 2014). Reasons for this included the unavailability of spatial data, limited computational power, or limited geoprocessing capabilities in these research fields. However, throughout the last decade we find an emerging number of behavioral models that include attributes regarding the built environment. The wide range of attributes indicates limited knowledge of how to represent urban morphology through quantitative attributes, which is clearly addressed by Clifton et al. (2008): "It would advance urban form research to have some standardization in operational definitions and measurement protocols." A standardization of these attribution processes might help to simplify their integration into behavioral models, even if the researcher has limited knowledge of geoprocessing tools.

While Gil et al. (2012) claim "the use of computer technology to support urban morphological studies and to build bridges to contemporary urban design processes," we find that this will require a standard set of attributes that characterize urban morphology. These attributes need to be based on a data model and a definition of variables that allow the representation of urban characteristics independent of the study area. Some approaches to such morphological characterization have been undertaken in the past. Alexander et al. (1977) proposes patterns for cities, neighborhoods, and buildings as answers to design problem, but they do not define algorithms that evaluate these patterns or their underlying characteristics on real-world data. Hillier and Hanson (1984) present a network-based algorithm

to analyze the configuration of a network and finds a connection to the distribution of functions and uses (Bafna 2003; Hillier 1996). Marshall (2004) extends these network analyses to a complete set of network characteristics and street typologies using configuration, composition, and complexity of streets. Sevtsuk (2010) also uses network analysis to include connectivity measures in the location choice of retail.

Most of these studies are based on network analysis, but one would expect the urban form to not only be characterized by its network but also by its buildings and open space. In recent years, we find the first studies in this direction. Meinel et al. (2008) classify buildings and blocks using attributes such as the shape of the footprint and the block to define urban density measures.

Gil et al. (2012) use cluster analysis of blocks and streets to identify morphologic urban typologies. Dillenburger (2010) presents a tool that sketches a building outline and searches for morphologically similar buildings. Ewing and Handy (2009) get even more detailed and ask people to rate images of streets. They identify factors that lead to the perception of eight qualities of urban design: imageability, enclosure, human scale, transparency, linkage, complexity, coherence, and tidiness. Using regression analyses, they can link these to objective characteristics of the street, façade, and buildings.

Planning disciplines are also undertaking morphological studies: The Award of the American Society of Landscape Architects in 2010 honored an interdisciplinary study by landscape architects, architects, and planners that compares urban spaces based on morphological and functional characteristics of North American case studies (Cherry 2009).

All these studies indicate that form, quality, and use of a location are connected to each other and suggest that urban morphology is of high importance to the modeling of behavior and prices. However, as Gil et al. (2012) state, "the analytical process is laborious and not entirely objective" and "there is a need to integrate different morphological approaches to obtain a more complete and complex set of urban environmental attributes."

Within this paper, we will address this limitation and present a set of attributes that we believe to represent the built environment at different scales. These attributes can then be used as variables in quantitative modeling processes. We classify location characteristics at different scales that we believe represent the different disciplines of architecture, urban design, and urban planning in order to support the future development of planning support systems for design proposals.

2 Data model

2.1 Data models

The amount of geo-located data available to researchers has exponentially risen throughout the last decade, and we can expect that this tendency will be reinforced through open-data initiatives (**?**). While there is a tendency for harmonization across national boundaries—for example, the projects INSPIRE, SRTM, and CORINE (European Environment Agency 1985; INSPIRE 2013; SRTM 2000)—the structures of most geodata currently differ significantly between study areas. This reduces usability in spatial studies that cross administrative borders and reduces the comparability of studies in different areas.

2.2 Global data

Google Earth and Google Maps (Google Inc. 2013a,b) were the first well-known public platforms where this limitation did not exist. They opened the door to what could be called the first global dataset of spatial objects. Diverse Web 2.0 developments and national initiatives, such as OpenStreetMap, Wikipedia, or the Shuttle Radar Topography Mission of NASA, followed that line and collected data that became available independent of administrative boundaries. Automated image processing methods allow for the extraction of spatial information for large areas from satellite images. Light detection and ranging data (LIDAR) offer precise volumetric representations of spatial objects from overflights

and allow for the automatic creation of detailed volumetric models of whole cities. We can thus expect the availability of spatial information to become more and more independent from official sources and see a development away from local data and toward global data.

However, at the present time, these data are either not available to researchers because they are owned by private companies or they are incomplete, such as those being collected by Web 2.0 projects. In this paper, we assume a data model that reduces the impact of such limitations by being flexible enough to allow for the integration of most datasets, whether or not it is official data, Web 2.0 data, or a design proposal.

2.3   An object-driven data model

Our data model relies on objects representing the built environment in the form of built structures, such as geometries and volumes of buildings, blocks, and the connecting street networks, which are available in most official geodata and on OpenStreetMap. The transformation of these objects in the real world requires construction activity. They thus form long-lasting structural components of the built environment. We limit the use of further attributes describing these objects. We also avoid including objects missing a standardized classification, since these would reduce the general usability of our data model. Examples of excluded objects are recreational space, agricultural land, natural areas, soil coverage, and open space. Our data model uses the feature types of polygons, lines, and points as representation for real-world objects (see Figure 1) with the attributes shown in Table 1. These are linked to specific methods of analyses.

2.3.1   Blocks

Blocks are defined by enclosing streets and have only the attributes given by their geometry. Meinel et al. (2008) use this block for their morphologic analyses based on raster maps. A block is clearly defined by the border of the street and can be precisely described when cadastral information on streets or parcels is given. Alternatively, it can be derived in sufficient precision when streets are given in the form of an axial network with information or assumptions on their width.

It should be stressed that with this definition a block does not reflect urbanity and should not be confused with the building typology of perimeter block development. Every building refers to exactly one block. At the border of the settlement, these blocks might be very large and consist mainly of unbuilt land. This specific situation is reflected in Section 3.2.1.

If available, the parcel could form an alternative to the block for most of the analyses presented here. This would be advantageous since parcels reflect land ownership and legal regulations on construction. However, our previous projects in various European countries showed that the parcel is cadastral information that is rarely available on a regional or national scale. These data are created and hosted by municipal administrations and only available with high costs. With a focus on the location-independent usability of our data model, we thus decided to use the block as the morphological entity instead.

2.3.2   Buildings

The building composite is a common representation of buildings in many geodata: Buildings touching each other are grouped as one object. Examples for such data are BDTopo in France, ATKIS in Germany, Vector25 or TLM in Switzerland, and many buildings in OpenStreetMap. We choose the building composite as the representation of buildings in our data model as well. However, the number of buildings should be kept track of if available, since it defines the granularity and diversity of a building compound and its facade.

Buildings are expected to have geometry regarding their footprint and information on their height (2.5D). Height information can be available as absolute height or the number of stories: either allows

Figure 1: Initial objects represented in data model: buildings, blocks, and streets. Source: ©2010 Amt fuer Raumordnung und Vermessung (GV2113) ©2011 Swisstopo (JA100120)

us to derive the other value. Absolute height is generally available through digital elevation models in combination with a digital terrain model, but in many cases the height of buildings is also given as a number of stories.

2.3.3 Streets

Streets form a connecting network, and their configuration and capacity form an important characteristic of the built environment. As such, it is common to represent them in an abstract form through their centerlines.

While capacities of streets are rarely available, we often find classifications of streets, which are closely related to movement and access. Clifton et al. (2008) differentiate streets based on capacities and speeds. In accordance with the Federal Highway Administration they distinguish local roads, collectors, arterials and expressways or freeways.

We use a comparable classification in our data model for simplified routing. We identify categories for motorways, arterials, collectors, or distributors and local streets as motorway, primary, secondary, and tertiary roads, which is the definition as used in OpenStreetMap. Furthermore, we recommend differentiating bridges and tunnels, since they simplify the cleaning of the network topology. As many datasets do not include real three-dimensional objects, we assume that blocks and streets are situated on a flat plane.

3 Classification

In their review on quantitative analysis of urban form Clifton et al. (2008) differentiate the scales of block face, neighborhood, submetropolitan, metropolitan, and regional, which they link to different data and research fields. The classification proposed here uses a predefined data model to define a set of attributes characterizing urban morphology. This slightly changes the scales of analysis and includes alternative approaches to characterize space from that reported by Clifton et al. (2008).

The scales introduced by Clifton et al. (2008) represent levels of aggregation. In location choice models, we consider locations and their characteristics as alternatives. It thus seems more convenient

Table 1: Proposed datamodel of the built environment

Feature	Columns	Type	Examples
buildings	building_id	primary key	
	stories	integer	1,2,10,35
	height	float	24.5
	postcode	integer	271, 8048
	the_geom	geometry	
blocks	block_id	primary key	
	postcode	integer	271, 8048
	the_geom	geometry	
streets	street_id	primary key	
	type	string	primary, secondary
	bridge	boolean	TRUE/FALSE
	tunnel	boolean	TRUE/FALSE
	maxspeed	integer	50
	hierarchy	integer	1,2,3,4,5
	the_geom	geometry	

to represent the scales of perception relating it to the actual location. As a consequence, we differentiate six levels or scales of perception that are useful for the description of the objects in our data model: object, composition, neighborhood, district, municipality, and region. Small-scale analyses are specific to the form and appearance of objects. A composition is characterized by the interaction of individual objects with their direct neighbors and the space these create. A neighborhood is defined by the characteristics of the walkable surroundings. A district includes further transport modes than the neighborhood and characterizes the area that is accessible within a short driving distance. The municipality is comprised of the extent of the municipal border and includes information on a polycentric or monocentric structure. Meanwhile, the region evaluates the interaction of various municipalities and their catchment areas. We believe this classification to reflect the planning guidelines and designs in the fields of architecture, urban design, and urban planning, although this paper and the implied data model cannot cover all characteristics and details of these disciplines. Further details on classification will be described later.

3.1 Object

The appearance and spatial impact of an object is defined by its outer geometry and volume. These characteristics define the architectural type of a building. They are commonly communicated through models and plans; it is rather uncommon to extract these characteristics in the form of attributes. However, as Meinel et al. (2008) show, these attributes can be useful to characterize and classify objects, and they are common in other applications such as image processing. We differentiate the attribute from the form of an object according to the initial feature type.

3.1.1 Line

In the proposed data model, the street network consists of lines representing the streets' axes. Its shape only includes the information on the orientation and length of a street segment. In a three-dimensional

(3D) model, we get additional information on the slope and 3D length, but as mentioned earlier we assume a flat plane in our data model (see Section 2).

3.1.2 Polygon

Polygon characteristics Buildings and blocks are represented as volumes in 2.5D—i.e., through polygons and an attribute on its height. The geometry of a polygon includes information on the area and the perimeter as well as its compactness. This allows us to distinguish linear housing from closed forms or compact, detached housing. We also find the amount and size of inner polygons, such as courtyards in a building.

Meinel et al. (2008) extract, among others, the bounding box and enclosing circle of buildings and derive their area, perimeter, and compactness. Furthermore, the characteristics of the enclosing objects and the initial objects are evaluated against each other in the form of ratios.

Volumetric information Volumetric information is not used by Meinel et al. (2008). We propose building the ratio of volume to area of the facade and using the height to estimate floorspace, assuming a general floor height of 3 meters, which reflects most residential buildings but is too low for commercial and retail buildings. If given, the use of the number of stories is advantageous.

Topological skeleton The topological skeleton is a representation of polygons often used for shape recognition and shape analyses (image processing, for example). A skeleton is a line string that represents a thinned version of the shape's boundary but keeps the topological information—i.e., where each point of the line string lies within the polygon and has a maximum distance to the closest boundary (equidistant to boundary). It can be extracted using thinning techniques in image processing (Lam et al. 1992) or through vector-based thinning techniques (Lee 1982; Ogniewicz and Kübler 1995). The skeleton of a building allows us to extract typological information such as the length, width, and orientation of a building and also to characterize its shape by evaluating the number of unconnected lines ("deadends") and the number of connected lines. Figure 2 shows how these attributes can help to define building types. However, it should be mentioned that due to the variance in existing construction and their cartographic representation, these ideal types cover only a part of the building data and alternative classification techniques are needed to classify all buildings.

3.2 Composition

By relating blocks, networks, and buildings to each other we characterize the immediate environment, or composition, of a location. While the building form mainly represents architectural characteristics, composition reflects the spatial configuration of buildings and their perception—elements of urban design. These types of attributes have rarely been used in quantitative models, so we have to rely on expert knowledge in urban design to define them. We distinguish the configuration of a building from other buildings, blocks, and streets.

3.2.1 Buildings

Influence zone The interaction of buildings structures the space and defines its perception. The distance to the next building allows us to differentiate loosely spread buildings from grouped buildings. By forming a topological skeleton on the unbuilt space, we can determine the "morphological influence zone" of a building, which represents the area of attraction that a building has if no other barriers such as fences, vegetation, or roads exist. This influence zone forms a polygon that allows us to measure the same characteristics as given in Section 3.1.2, including the ratio to the area covered by the building and the volume of the building. For unrealistically large influence zones near open landscapes, their extent is limited to 300 meters, which reflects the walkable surrounding of a location.
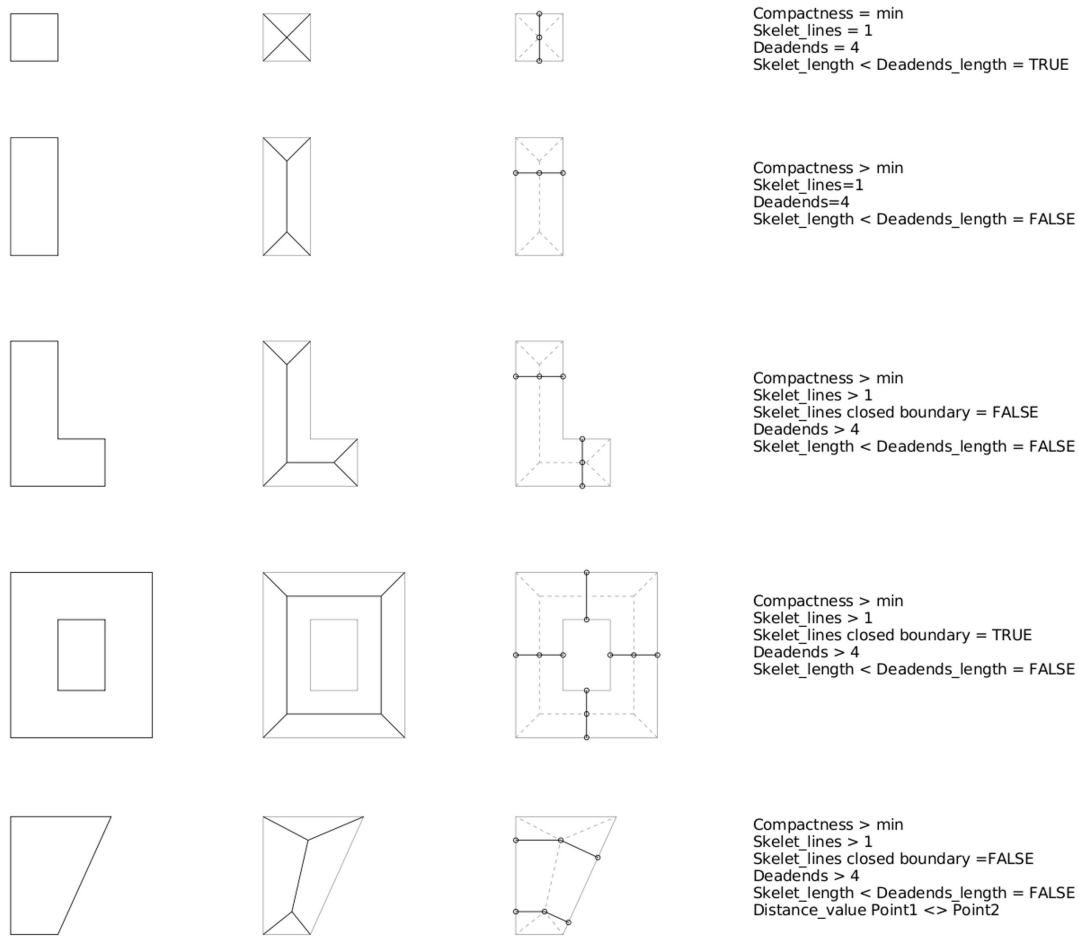
Compactness = min
Skelet_lines = 1
Deadends = 4
Skelet_length < Deadends_length = TRUE

Compactness > min
Skelet_lines=1
Deadends=4
Skelet_length < Deadends_length = FALSE

Compactness > min
Skelet_lines > 1
Skelet_lines closed boundary = FALSE
Deadends > 4
Skelet_length < Deadends_length = FALSE

Compactness > min
Skelet_lines > 1
Skelet_lines closed boundary = TRUE
Deadends > 4
Skelet_length < Deadends_length = FALSE

Compactness > min
Skelet_lines > 1
Skelet_lines closed boundary =FALSE
Deadends > 4
Skelet_length < Deadends_length = FALSE
Distance_value Point1 <> Point2

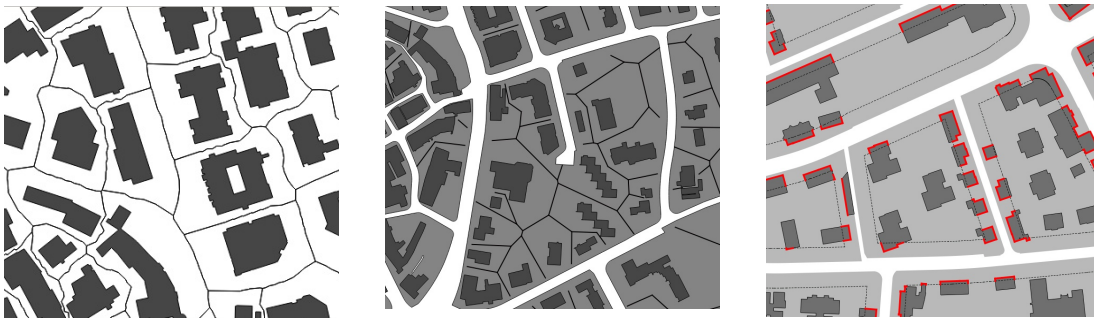**Figure 2:** The shape and morphologic skeleton help distinguish building typologies

Figure 3: Influence zone of buildings (left), permeability of block (middle), and closeness of block (right). Based on ©2010 Amt fuer Raumordnung und Vermessung (GV2113).

Courtyards   Courtyards form a semipublic space that can substantially benefit a location. While we evaluated inner polygons forming a courtyards in Section 3.1.2, a courtyard can also be formed by different buildings. Depending on buildings' footprints and heights, the space between them can form a courtyard, even when they do not touch each other, as long as the area's access is perceived as "closed." This perception strongly depends on the width of the access in relation to the height of nearby buildings. Assuming a courtyard is perceived as closed when the access is smaller than the building height,

we can evaluate the number of courtyards based on their perception as semi-private spaces as well as by their size.

Urban places   The configuration of buildings furthermore allows us to extract and classify urban places. Krier (1979) classifies urban places using their morphological characteristics. Following his theory an urban place is characterized by enclosing buildings similar to a courtyard, but will additionally have access to the network of streets and pedestrian links. The street connectivity, buildings height and the size of the place can thus help to distinguish types of urban places. However, if applied on non-urban areas, large parts would be recognized as urban place, so that further constrains on the compactness of an urban place will need to be defined if applied on those.

3.2.2   Street and block

The positioning of a building in relation to nearby streets defines its perception. This can be essential for the success of a retail store, for example, where buildings with street-aligning facades are preferred for shop windows. We differentiate the distance from the block border, the number of enclosing streets, the percentage of a facade that is visible from the street, and the orientation towards the street. The last can be derived from the building skeleton, while the facade that can be seen from the street and positioning at the corner of a block depend on the outer polygon shape.

Similar to courtyards, the character of a street also depends on its section. Therefore, not only the width of the street is relevant, but also its ratio to the height of the aligning buildings.

While the block is defined by streets, it has individual characteristics through its polygon shape. A Voronoi diagram of buildings reflects the visual permeability of a block and allows us to measure the number of axial looks through the block.

Furthermore, we define the closeness of a block as the ratio of the building facades that are perceived from the street to the actual perimeter of a block. Urban blocks can be expected to have a higher closeness than areas consisting of many detached buildings, with a maximum closeness of 1.0.

Figure 3 illustrates the closeness, permeability, and influence zone as proposed here.

3.3   Neighborhood and district

In addition to the composition, we need to reflect the wider context of a location as well. In their residential location choice model, Guo and Bhat (2007) differentiate between fixed neighborhood boundaries, based on census tracts and blocks, and alternative neighborhood definitions using Euclidean distance and network distances. Their results show a variation of spatial effects between the attributes describing the built environment and the socioeconomic variables, which justifies a differentiation of the surroundings into multiple dimensions: the neighborhood representing the walkable surrounding and the district as the accessible space using various transport modes. While the neighborhood mainly reflects local densities, the district should reflect structural configurations of the urban landscape. We believe both of these scales to be mainly linked to the disciplines of urban planning, while partly still reflecting issues of urban design.

However, our data model of the built environment does not allow for the inclusion of socioeconomic characteristics; it just reflects structural characteristics. As a consequence, we will group these scales and group the same characteristics for different extents—i.e., using the Euclidean distance of 300 meters as an approximation for walkable surrounding and the network distance of one minute of free-flow travel time as the extent for a district (see Figure 4).
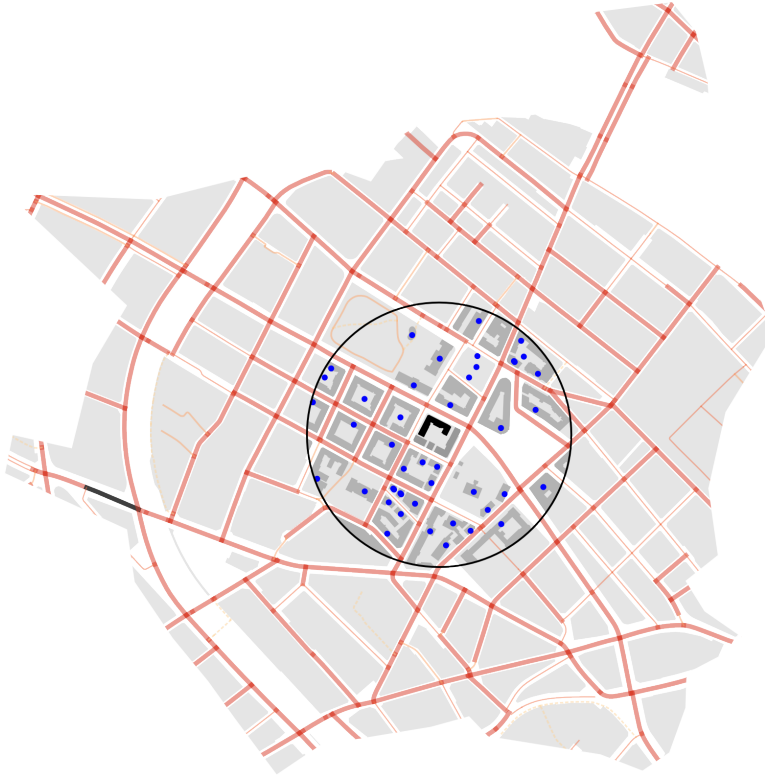
Figure 4: The neighborhood of a location. Neighborhood characteristics capture the densities of the built environment and the network structure in the close surrounding area of a location. Based on ©2010 Amt fuer Raumordnung und Vermessung (GV2113) and ©2011 Swisstopo (JA100120).

3.3.1   Densities

The density of building footprints or of floorspace is a common measure in community design as proposed by urban planners. They normally measure density in reference to another polygonal object, such as the parcel or a traffic analysis zone.

However, this does not necessarily reflect the perception of space. Especially at the border of the reference objects, where the space might be strongly influenced by neighboring buildings, we can expect inconsistencies. A simple and common way to avoid these is to use the sum within a predefined Euclidean distance instead—within a walkable neighborhood of 300 meters, for example.

The Euclidean density might be appropriate for the direct neighborhood of buildings, but it has significant limitations for evaluations of larger areas since it ignores barriers that reduce the accessibility of the neighborhood. In that context, a network distance-based density is a better approach. It is sometimes labeled as accessibility since it represents a cumulative opportunities measure—the number of objects accessible in a certain time (Bhat et al. 2000; Vickerman 1974; Wachs and Kumagai 1973). The definition of a district in this paper reflects these considerations: we measure densities as the amount of opportunities in the area that are accessible within two minutes of driving. However, we prefer the term cumulative opportunities or network density, so as not to confound it with the weighted distance accessibility measure presented below.

3.3.2 Homogeneity

We expect not only a mix of use but also the mix of building types to form an important characteristic of space. Building types can be defined by the form of an object, as described in Section 3.1. On one hand, we can count the variety of building types within a certain Euclidean distance. On the other, we can form a zone around similar buildings and gain information on the extent and form of this area.

3.3.3 Network structure

If streets are represented as axial lines, we can create a graph consisting of nodes and edges. Graph theory has been employed in various disciplines such as sociology, economics, or biology (Newman 2003) and various network analyses on graphs have been developed and well discussed (Newman 2010). Marshall (2004) gives a good overview of these with a focus on street networks. We can differentiate attributes characterizing the network structure and those relating the network to destinations along it as discussed in Section 3.4.2.

Dill (2004) reviews the literature and distinguishes connectivity measures based on intersection density, street density, percent of four-way intersections, connected intersection ratio, the percentage of grids, and pedestrian route directness. We base our network analyses on these previous studies and evaluate the number of intersections, links, length, and number of dead-ends in the district of a location (within two minutes of free-flow travel time).

3.3.4 Local access

Clifton et al. (2008) highlight that the accessibility measures at the level of a neighborhood are more diverse than on a metropolitan scale, with reference to Holtzclaw (1994) and Song and Knaap (2003), who count among others specific commercial and retail locations within walking distance. Randall and Baetz (2001) evaluate pedestrian connectivity; Dill (2004) extends this to bicycle connectivity.

Sevtsuk (2010) introduces further network-based characteristics that form extensions of the gravity-based model and are used in his retail location choice models. Besides the ones already discussed, he introduces remoteness as the cumulative distance, cumulative number of turns, or cumulative number of intersections required to reach each of the available destinations in a given network radius along shortest paths. He defined these destinations as the built volume, residents, jobs, transit stations, and different types of retailers. Remoteness thus becomes an index that measures how compact a network is in combination with the density of destinations.

We do not include information on retail or commercial locations or on pedestrian networks and cycling networks in our data model. These types of characteristics thus cannot be evaluated and will need to use dedicated points of interest, but the connectivity measures as proposed by Sevtsuk (2010) can be implemented for floorspace (see also Section 3.4.2).

3.4 Municipality and region

While the neighborhood around a building defines the urban quality at a human scale, its relation to the municipality and region defines the uniqueness of location and its role within a larger context. On a municipal level, we search for centers and the monocentric or polycentric structure of a settlement, while we evaluate the impact of these on the regional scale. These characteristics are mainly operationalized through regional planning disciplines.

Christaller (1933) studies the urban settlements in Southern Germany in order to understand how urban settlements evolve and are located in relation to each other. He finds a system of lower-order, middle-order, and higher-order centers with a clearly defined spatial distribution, as argued in his Central Place Theory. Even if these can be expected to be at least partly the outcome of planning decisions, the weight of a center and its distance from similar centers have a high impact on available opportunities, such as retail and administrative facilities as well as the diversity of job types.
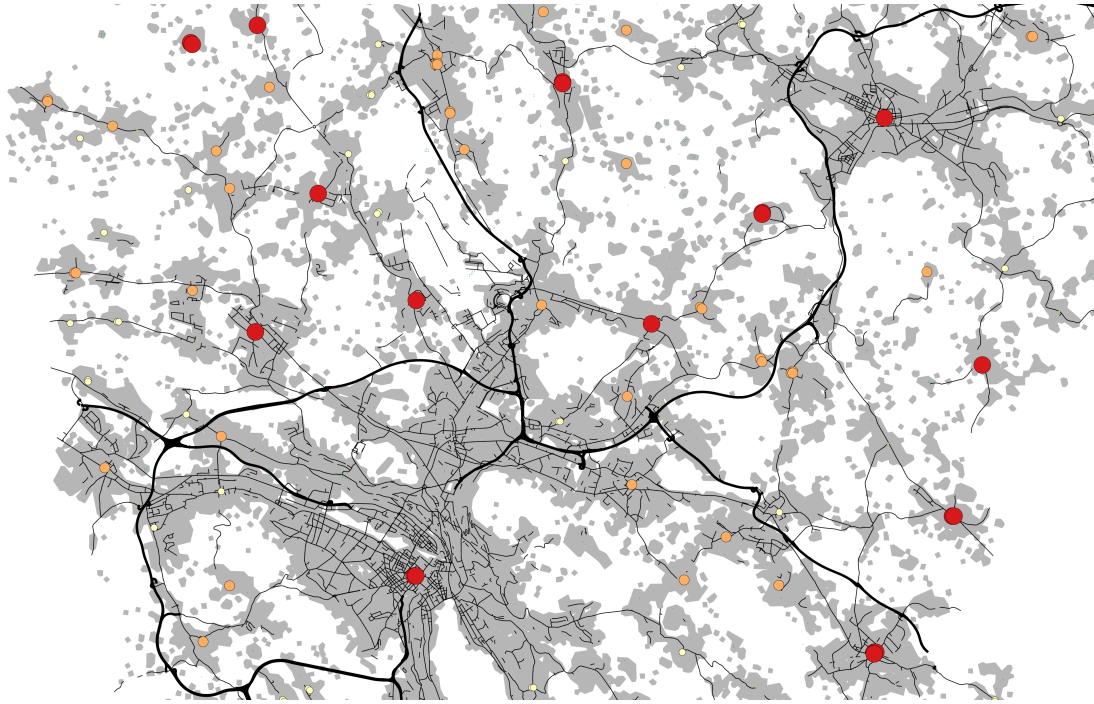
Figure 5: Municipalities can be characterized by their settlement area (grey), networks (black), and global or local centers (red and orange). Based on ©2010 Amt fuer Raumordnung und Vermessung (GV2113) and ©2011 Swisstopo (JA100120).

3.4.1   Settlement area

Municipalities have administrative borders that do not necessarily represent morphological characteristics and reduce their comparability. Those administrative boundaries form political regions, but the current extent of the settled area often differs substantially from these. By dissolving the buffer around buildings, we can define the morphologic settlement area; small areas need to be filtered to represent only conglomerations of buildings.

This shape is a good basis for evaluating and comparing density measures and center structures (e.g., polycentric, monocentric) of different municipalities. Their form and extent can then be described using the attributes already mentioned in Section 3.1. The position of a location inside the settlement area is described through its distance to the boundary.

We can also relate the settlement area to other the objects in our data, as described in Section 3.3. The ratio of the block area to the settlement area forms an index related to the space occupied by roads and railways.

We can use the ratio of the settled area's size to the size of the administrative boundaries of a municipality and gain information on the unused space or natural reserve and the potential for growth. The last can be expected to reflect the pressure on the housing market in growing cities.

3.4.2   Centrality and accessibility

While object-specific network measures have already been discussed in relation to the district (see Section 3.3.3), the topological information can be used to quantify the structural characteristics of cities and regions. Centrality measures—namely closeness, betweenness, and degree (Newman 2010)—are used to study the relevance of a node within a network. These analyses break down the network into nodes and links to evaluate the connection of the nodes to each other and to consider the cost for each link. The degree counts the number of connections each node has and allows us to distinguish

intersections. The closeness is the inverse of the summed up distances to all other nodes, while the betweeness quantifies the number of times a node is part of the shortest path between other nodes. As reported above, Sevtsuk (2010) calculates these characteristics on a predefined network radius instead.

Different weights or costs can be applied to the road segments in the calculation of centrality measures. We propose three cost functions: a constant of one per link, the length of a link, and the free-flow travel time per link. The last can be estimated by assuming a certain speed per street type. An alternative cost function would be the change of orientation between links, as used in Space Syntax analysis (Bafna 2003).

Accessibility is a measure of the spatial distribution of activities around a point, adjusted for the ability and desire of people and firms to overcome this spatial separation or "the potential of opportunities of interaction" (Hansen 1959). Accessibility calculations are used to represent local and regional differences within the urban landscape. Accessibility as distance-weighted opportunities or a gravity-type measure can use the free-flow travel time in the distance decay function to weight the impact and number of opportunities around a location. We propose using floorspace in a building as such an opportunity measure.

3.4.3 Urban centers

Density gradients as first presented by Clark (1951)(for a review see McDonald (1989)) are a common way of comparing cities in urban economics and have a proven correlation with the travel behavior and distribution of income (Clifton et al. 2008). However, this assumes and only allows comparisons of monocentric cities.

Motivated by these studies, we often find the distance to the central business district or other supply centers in choice models and regression analyses. These lack reproducibility and reduce the comparability of study areas since the location of centers is being defined manually, while an automatic evaluation would be advantageous. Since we only integrate the built structure in our data model, the distribution of functions cannot be evaluated here. However, assuming a higher density for urban areas, we can analyze the built structure and define the location of centers based on these.

The spatial distribution of built densities can be represented by means of a kernel density function or the interpolated accessibility measures of buildings. Both approaches form GIS-raster analyses and result in a map that represents a smoothed surface. This data can be analyzed and described using methods of terrain analysis (Wilson and Gallant 2000). An algorithmic definition of centers needs to define the location of urban centers in the form of points and also express their extent and impact within the urban landscape.

A comparable problem can be found in the geomorphologic classification of mountains, which differentiate major peaks from minor peaks according to their topographic prominence. This topographic or orometric prominence defines the height of a peak, measured in relation to the lowest contour line encircling it and no higher summit (see Figure 6). Using this method, we can not only distinguish the location of a center—i.e., its peak—but also its local weight (peak height), global weight (height above sea level), and catchment area (area of lowest contour line). These values allow us to quantify the impact inside a settlement area as well as the regional relevance. To allow for the comparison of local centers, we propose scaling the range to a level of 0-255, which represents the range of a monochrome raster with a bit depth of 8.

A location is linked to its closest center through an evaluation of the distance to this center and the aforementioned values of the center. On a regional scale, we can evaluate the catchment area of a city's largest center and the distance to the next center with a higher global value (see Figure 5).

3.5 Defining scale-dependent types

These morphological attributes allow for the scale-dependent classification of locations by using attributes of the individual scales. In our data, these single points of observations are the individual
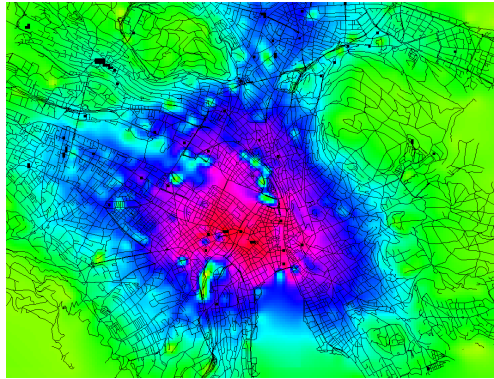
Interpolated accessibilities
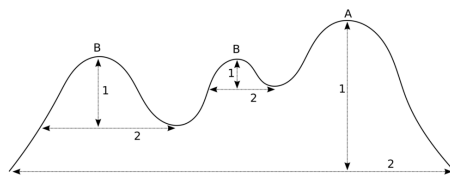
Interpolated accessibilities

Figure 6: Weight and location of centers. The location can be evaluated through prominence analyses on the interpolated accessibilities. The section allows for the differentiation of peaks as global centers (A) and local centers (B) based on their local impact (1) and catchment area (2).

buildings, which form the highest level of disaggregation. However, through the aggregation of these characteristics, the given attributes would also allow us to describe and classify alternative aggregation areas, such as census tracts.

Hecht et al. (2009) introduce a rule-based classification using the morphological characteristics of buildings, blocks, and the interactions between them. They predefine the types of buildings through weighting impact and range of the morphologic attributes per building type. This method involves the training of the data on a reference sample. Gil et al. (2012) take an alternative approach and use a k-means cluster analysis to distinguish what they call urban typologies—each cluster represents one typology.

The advantage of the earlier approach is the comparability of the typologies across study areas: the definition of typologies is static and we can observe variations in behavior, price, and urban structure in relation to each typology. The drawback of such an approach is the lack of usability in significantly different areas and without prior knowledge of the types. If the predefined typologies do not cover all observations, we fail in assigning locations to individual types.

In cluster analysis, the number and content of clusters are not predefined but data-driven. In consequence, the number and characteristics of attributes used to define a cluster are less rigid as well. On the other hand, the resulting clusters can vary across studies and reduce the comparability in relation to the typologies. This comparability can then only be achieved through a comparison of the initial attributes used for clustering.

In this paper, we propose the use of morphologic attributes that represent different scales of perception for use in behavioral models. As we are not aware of the number existing types at these scales, it seems more convenient to use cluster analysis, which we present in Section 4.4.

4 Case study

4.1 Study area

After having described the morphologic characteristics at different scales, we will evaluate these characteristics on a case study of the canton of Zurich. The data we use covers a large part of the case study area and has been described in detail by Schirmer et al. (2011). Buildings and blocks are from cadastral plans and have been enriched with additional information on use, value, residents, and jobs. The street network comes from the TLM Swisstopo (2013).

4.2 Software

Geoprocessing requires the functionality of a geographical information system (GIS), enabling the querying of data through attributes as well as through their spatial location. Since desktop GIS has limited performance and flexibility in terms of large data processing, so we decided to run this processing on the open source database PostgreSQL 9.1 (PostgreSQL Global Development Group 2013) with the spatial extension PostGIS 2.0 (PostGIS Development Team 2013). The latter allows for the creation and transformation of geometric objects inside the PostgreSQL database. For routing algorithms, we use the PostgreSQL-extension PGRouting 2.0 (pgRouting Global Development Group 2013), which includes functionality for the transformation of lines into a topological network. The geometric features in these tools are rapidly evolving, but at the present time not all functions needed are available. For example, the routines for Voronoi creation skeletonization are missing. Individual functions and routines have been created for such processes.

While raster processing is possible with the latest version of PostGIS, it still showed limited functionality at the present time. We thus relied on the raster algorithms of GRASS GIS 6.4.3 (GRASS Development Team 2013), an OpenSource GIS that can be linked to PostgreSQL. GRASS GIS uses a specific topological format that requires the import of the vector geometries, but the attributes can be kept and processed directly with PostgreSQL. By importing locations as vector points, we matched these to the raster-sets inside GRASS. Their values are directly written to PostgreSQL.

The evaluation of our data in the form of plots and descriptive statistics, as well as cluster analyses, have been performed with the open source software package R (R Core Team 2013).

The whole proceeding has been scripted as shell-scripts and runs in an iterative process, allowing for simultaneous execution on different cores or computers (multi-threading). For this case study, the processing was performed on a dual core laptop with 8GB RAM under Ubuntu 13.04. The computation of attributes for the study area needs 15 hours, and the clustering described below needs an additional 20 hours since different clustering algorithms are tested and the number of clusters is estimated.

4.3 Clustering

The cluster analysis has been performed for all scales described in Section 3 using k-means cluster analysis with the Hartigans algorithm (Hartigan and Wong 1979) as integrated in the open-source software package R (R Core Team 2013). However, in the following we will only report the details for the clustering of individual objects.

Since the attributes described in Tables 1, 2, 3, and 4 represent different units and vary in range, we standardize them prior to clustering. Many attributes proved to be strongly correlated—for example, perimeter, width, and size of the building footprint. A principal component analysis (PCA) on the standardized variables is used to remove these correlations. In order to be less sensitive to outliers, we apply the robust principal component analysis, which is available as *PcaHubert* in *R*. Following the kaiser-criterium, we included only those principal components that have an Eigen value higher than 1, which resulted in five principal components (PC) on the level of objects, one PC on the level of the composition, four PC for the neighborhood, and four PCs for the scale of municipality. Since one

PC for the scale of composition was not enough for the clustering, we also used the PC with an Eigen value smaller than one, which resulted on a total of five PC.

4.3.1   Number of clusters

A key to k-means-cluster analysis is the definition of the number of clusters. Hecht et al. (2009) defines 4 basic types and 15 more detailed morphologic types in their rule-based approach, which uses morphological aspects of buildings, blocks, and height information. Based on this classification, we would expect to be able to distinguish between 4 and 15 types of buildings based on object form and volume. However, we can expect this to be specific to the German or eventually European context. The evaluation of typologies in other regions and at other scales needs individual research or a data-driven definition of the ideal number of clusters as proposed in this paper.

Following Rousseeuw (1987) the silhouette of a k-means clustering can be used to control the quality of a clustering. The silhouette evaluates the tightness of a cluster and the separation of its data points from its neighboring clusters. A silhouette coefficient around 1 indicates that an object lies inside a cluster; a coefficient around 0 indicates that an object lies between clusters; and a negative co-efficient value shows that objects are closer to the neighboring cluster than to their own cluster center. The average silhouette width forms a quality measure of a clustering and allows us to compare different cluster outcomes in terms of optimal separation of data points: the highest coefficient value represents the optimal number of clusters.

We ran the clustering for 1 to 50 clusters and compared the average silhouette width to choose the optimal number of clusters based on this one as *k1*. Furthermore, we chose the second best as *k2* and the third best proposal as *k3* to control for alternatives (see Table 2).

The k-means algorithm chooses random centroids and generates their Voronoi to define the initial cluster, which are then optimized based in Euclidean distances. This makes it rather sensible to the initial starting points. An alternative approach is the use of medoids instead of centroids. The k-medoids algorithm chooses data points as centers (medoids or exemplars) and works with a matrix of distances between these. According to its documentation, it is more robust regarding noise and outliers than k-means because it uses the "mean (equivalent to the sum) of the dissimilarities of the observations to their closest medoid" instead of squared Euclidean distances (R Core Team 2014).

The function *clara* of the software *R* uses k-medoids and is optimized for large datasets. The function *pamk* makes use of *clara* to estimate the ideal number of clusters through the maximum silhouette width. These functions proved to be very stable in the estimation of outcome and highly efficient in terms of computational performance. Following the k-means methods described above, we run the algorithm for the optimal silhouette width (*c1*) and also for the second-best (*c2*) and third-best result (*c3*) in order to have variety in the observations and be able to argue on the effect of the chosen number of clusters.

As shown in Table 2, the proposed number of clusters is identical for the optimal silhouette width in all scales, but it differs for the less-optimal ones. Both algorithms have a tendency to choose very few clusters as the ideal number of clusters—all scales propose two clusters as the best solution. In consequence, the clusters with a smaller silhouette width represent more detail of the urban structure and seem more suitable for our purpose. Further studies will have to elaborate in detail on how to control the ideal number of clusters.

The k-means method shows significantly less variance in the silhouette plot (Figure 7), which leads to higher variety for the second and third optimal average silhouette width when using the k-medoids algorithm. The effect and relevance of this observation will be discussed later.
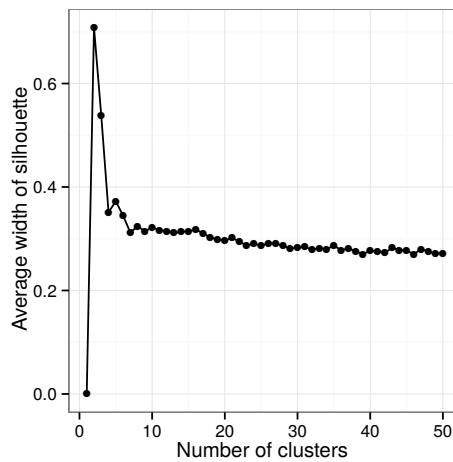
4.4   Typologies

In this section, we give an overview of scale-dependent clustering and compare the results of k-means and k-medoid. The attributes used for the scale-dependent cluster analysis are given in Tables 1, 2, 3,

Figure 7: Average silhouette width for different number of clusters in the scale of objects.
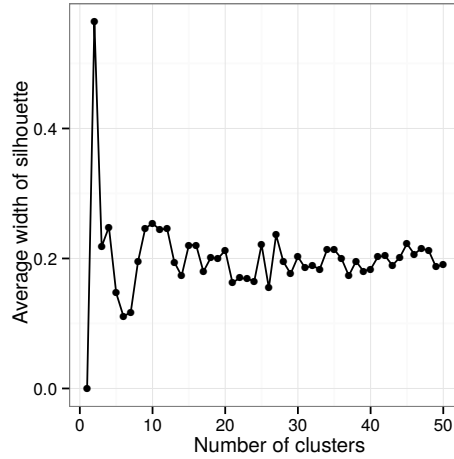
K-medoid results

K-means results

Table 2: Estimated number of clusters per scale

Algorithm for estimation	Object	Composition	Neighborhood	Municipality
kmeans centroid				
Silhoutte (best)	2	2	2	2
Silhoutte (second)	3	3	3	3
Silhoutte (third)	4	4	4	4
kmeans medoid				
Silhoutte (best)	2	2	2	2
Silhoutte (second)	4	35	3	3
Silhoutte (third)	19	6	4	4

and 4. We review the differences and variances of the clusters based on the individual boxplot for each of these variables. However, for interpretation and the practical use of this approach, we present these results through maps showing the clusters of the different scales using the building geometry.

Object The high number of clusters proposed in *c3* leads to a finer differentiation than for *k3*. Both have the tendency to define buildings based on their number of wings. The four clusters of k-means

represent punctual building developments, linear building forms, structures consisting of more than one wing, and group large-scale buildings. These clusters can easily be understood and interpreted in the maps and follow our expectation (see Figure 8).

Also *c3* distinguishes punctual housing, linear housing, and buildings consisting of several wings. Large-scale, mid-scale and small-scale objects are differentiated, and very long buildings get separated from short buildings that still have a defined orientation. Buildings having two wings (forming an "L") are found in different clusters than those with three wings (forming a "U") or those with multiple wings forming irregular or closed structures.

This last group shows the drawback of using cadastral information as the starting point: buildings often have small variations in shape (e.g., to represent entries). While these have no effect on the general form of the building, they affect the skeletonization and the forming of different groups. The use of a dataset with simplified polygons for the buildings, such as is present in the Swiss TLM, might be more optimal in such cases.

Figure 8: Selection of building typologies
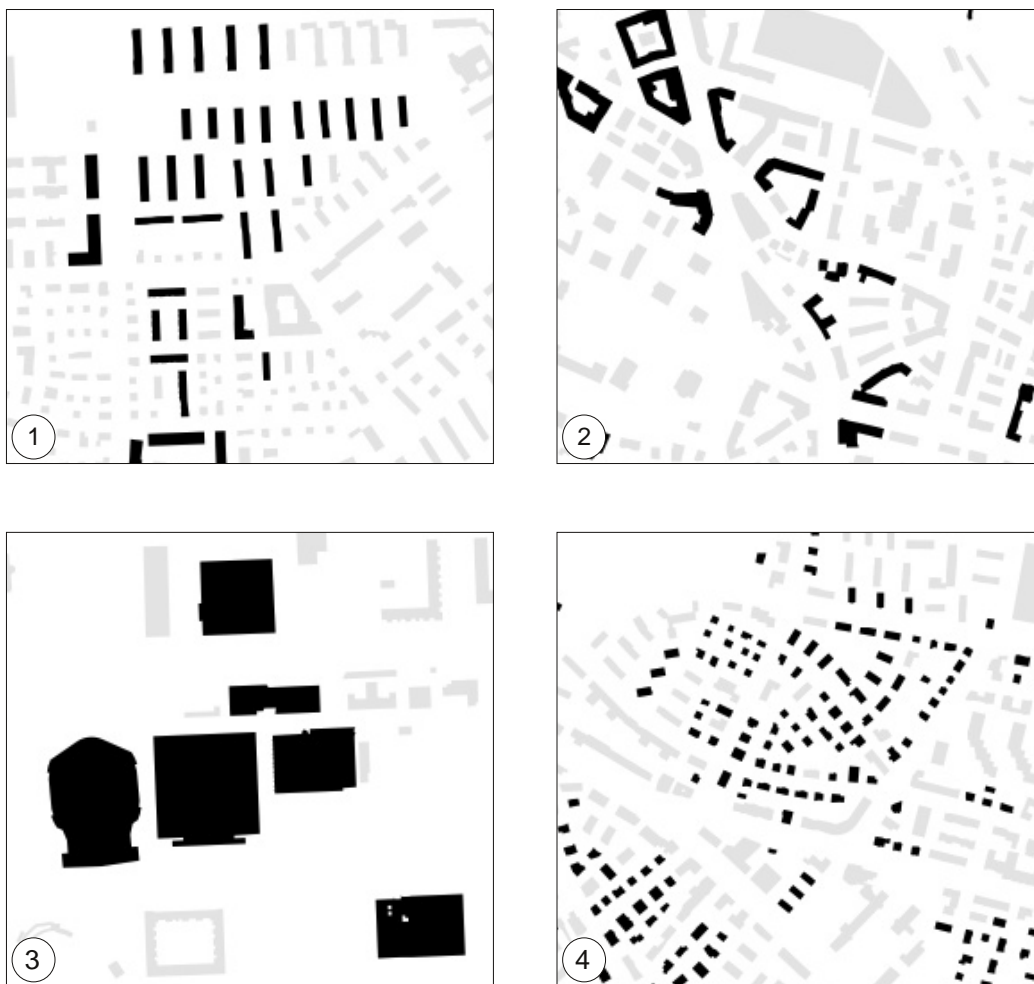
Source: based on ©2010 Amt fuer Raumordnung und Vermessung (GV2113)

These images show examples of the clustering results of the kmedoid-clustering on the scale of the object. The focus lies on the similarity of the buildings. We can see a differentiation of linear housing (1), large scale buildings (3), punctual building development (4) and buildings with multiple wings (3).

Composition   Only one principal component has an Eigen value larger than 1 for the scale of composition, which does not lead to any cluster solution. As a consequence, we include all principal components for this scale, but further studies will have to elaborate on whether this scale needs to be merged with the scale of objects.

The two cluster algorithms lead to partly different outcomes. A selection of the k-medoids results are shown in Figure 9. We find that *c3* groups buildings that are street aligning, thus having a high portion of street-facing facade, and defines locations facing unbuilt space as their own clusters. The latter generally have large influence zones (IZ) combined with large block sizes and low soil coverage in their IZ. This typology is also found in the results of the k-means clustering (*k3*). However, we find large-scale solitaries with irregular structures, such as found for campuses or industrial sites, to have very similar characteristics, often with more extreme values: they have a very large IZ, big distances to neighbors and block borders, and low soil coverage in their IZ.

Urban residential areas form a type of composition that is characterized by high density and a closed block structure. They show high soil coverage in their IZ and have a high percentage of construction along the block border. Residential areas of punctual or linear construction have smaller soil coverage in their IZ.

The last typology consists of freestanding objects with a low amount of street front. These are also characterized through a large distance to the block border, relatively high soil coverage in their IZ, and a rather low percentage of construction along the block border.

K-means also identifies large-scale solitary objects and locations facing the unbuilt space. Furthermore, locations with high urban density get grouped: they have a high soil coverage ratio and a lot of street front, thus occupying a large percentage of the block border. All other buildings fall into one group.

Neighborhood   Opposed to the building and composition clusters, the neighborhood clusters define areas of spatial homogeneity—they form spatial entities with comparable characteristics of density and networks. The main effect we observe for all clustering algorithms is the differentiation of inner urban areas and the border of settlement areas, which we expect to result from the lower amount of floorspace in a certain radius of travel time. When comparing k-means (*k3*) to k-medoid (*c3*), we find k-means creates more compact areas—i.e., it defines strict borders between the clusters. The results of k-means are shown in Figure 10.

Both algorithms represent disconnected areas, the borders of settlement, and urban cores very well. K-means defines an additional cluster for locations of medium density, which represent the cores of small settlements very well. K-medoid defines a dedicated group for areas that have a high percentage of dead-ends in the close surroundings instead. This leads to the expectation that using more clusters on this scale could reveal structural differences of the networks. Otherwise the clusters mainly rely on the density of buildings.

Municipality   The clusters on the scale of the municipality also define areas of similar spatial characteristics. The k-means and k-medoid algorithm propose a maximum of four clusters based on the average silhouette width. Their clustering results are consistent and comparable. They differentiate urban cores, the border of settlements, and the city of Zurich from the outskirts.

Locations in the city of Zurich form an individual cluster due to the unique and large catchment area of the global center and its high values of accessibility and connectivity. Urban centers outside of Zurich form their own typology; furthermore, we find highly accessible areas outside of Zurich and areas of low access in separated clusters.

The proposed set of variable shows a consistent behavior for the clustering and proves to be of value. However, when we "force" the clustering algorithms to use more than four clusters, we get a differentiation of locations along main streets from less-accessible locations (see Figure 11). The limitation to four clusters thus reduces the amount of detail in the clusters and might have to be revised.

Figure 9: Selection of composition typologies
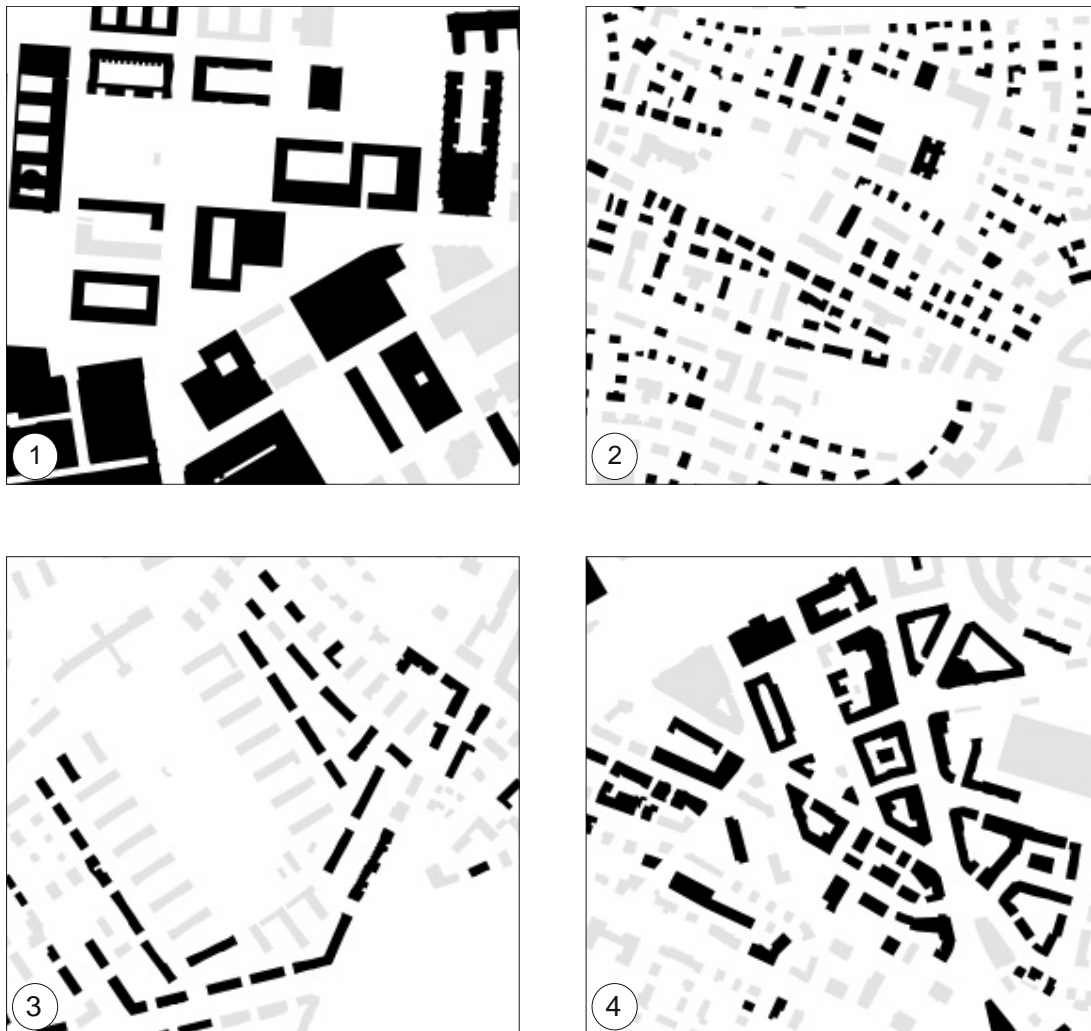
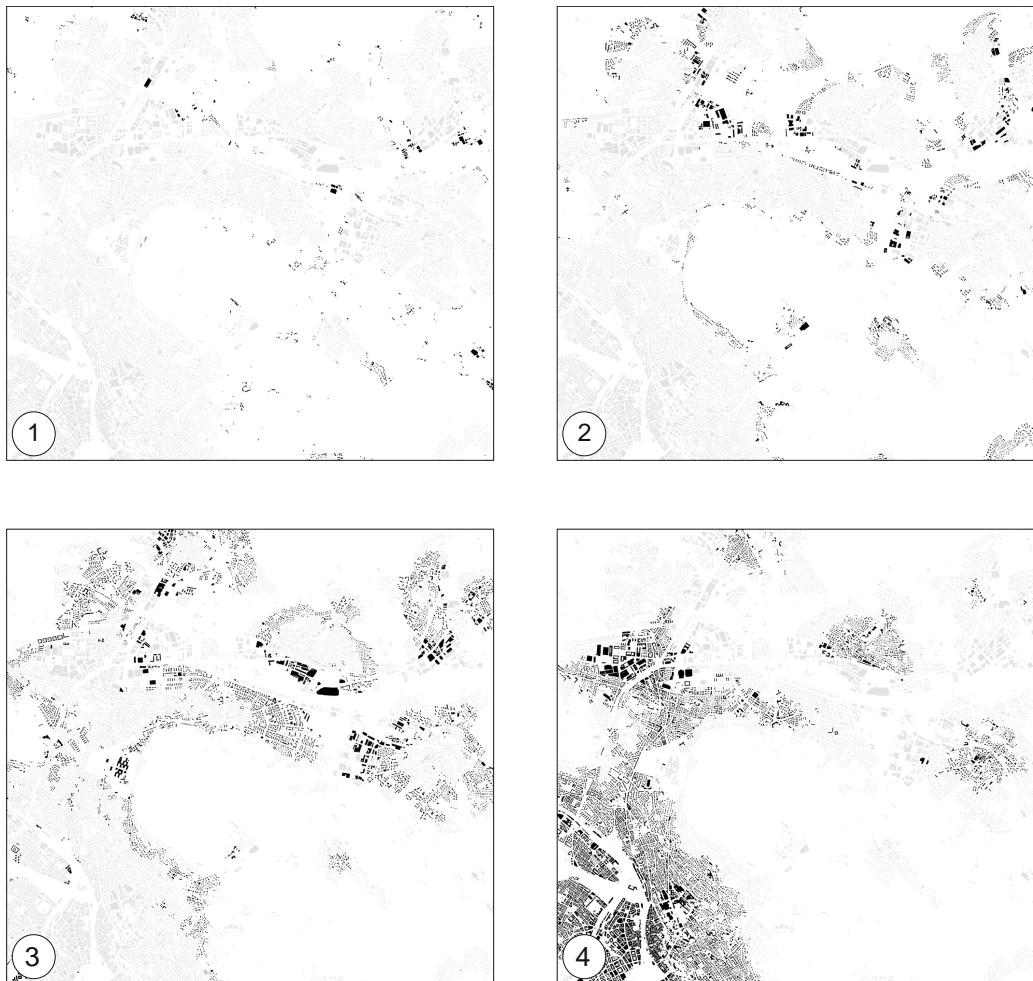Source: based on ©2010 Amt fuer Raumordnung und Vermessung (GV2113)
These images show examples of the clustering results of the kmedoid-clustering on the scale of the composition. These types refer to similarity in the relation to neighboring objects. We find street-aligning buildings (3), block-defining residential (4), block-filling solitaries of large scale (3), and detached punctual forms that form a block (2).

5   Summary and conclusion

In this paper, we present attributes describing the urban morphology at the scales of objects, compositions, neighborhoods, and municipalities. We introduce a set of attributes suitable for quantifying characteristics of these scales for modeling purposes (see Tables 1, 2, 3, and 4) and run a case study on the canton of Zurich to use these for the scale-dependent definition of typologies. The attributes are derived from geographic data with an underlying data model being simple enough to allow implementation in other study areas: it includes only form- and volume-based information buildings, networks, and the street blocks.

    We show the use of the proposed attributes in a k-means cluster analyses and a k-medoid cluster analysis to identify typologies of the proposed scales. With this approach, we classify the built environment into typological classes that can be directly used in location models or hedonic models. We found these clusters to form different views of urban morphology and believe they can help to com-

Figure 10: Selection of neighborhood typologies

Source: based on ©2010 Amt fuer Raumordnung und Vermessung (GV2113)

These images show examples of the clustering results of the k-means clustering on the scale of the neighborhood. These types refer to similarity in the density of buildings and comparability of network structures. In this example, the algorithms mainly classify based on densities. We find detached areas (1), the border of settlements as low density (2), mid density at the border of Zurich and in the core of small settlements (3), and the city of Zurich and smaller centers as high-density sites (4)

pare and understand spatial structures. While the scale of objects and composition differ between individual buildings, we find the typologies of neighborhood and municipalities to define boundaries of coherent structures. The results for individual buildings, the neighborhood, and the municipality are especially promising and follow our expectations.

However, we also found limitations of the given approach. The processing had to be performed per municipality for computational reasons, so a regional interaction across the municipal borders could not be represented. This especially has an impact on the accessibility calculations, where small municipalities directly near the city of Zurich are shown to have low accessibility values and become similar to disconnected rural areas. The definition of center structure is based on the density of floorspace of buildings since no other data is given in the data model. As a result, some retail centers as well as large-scale buildings are interpreted as centers by the given algorithm. Further studies should evaluate the alternative use of functions and points of interests to define such centers. The study area includes many small-scale municipalities, where a differentiation of the network structure seems awkward: a

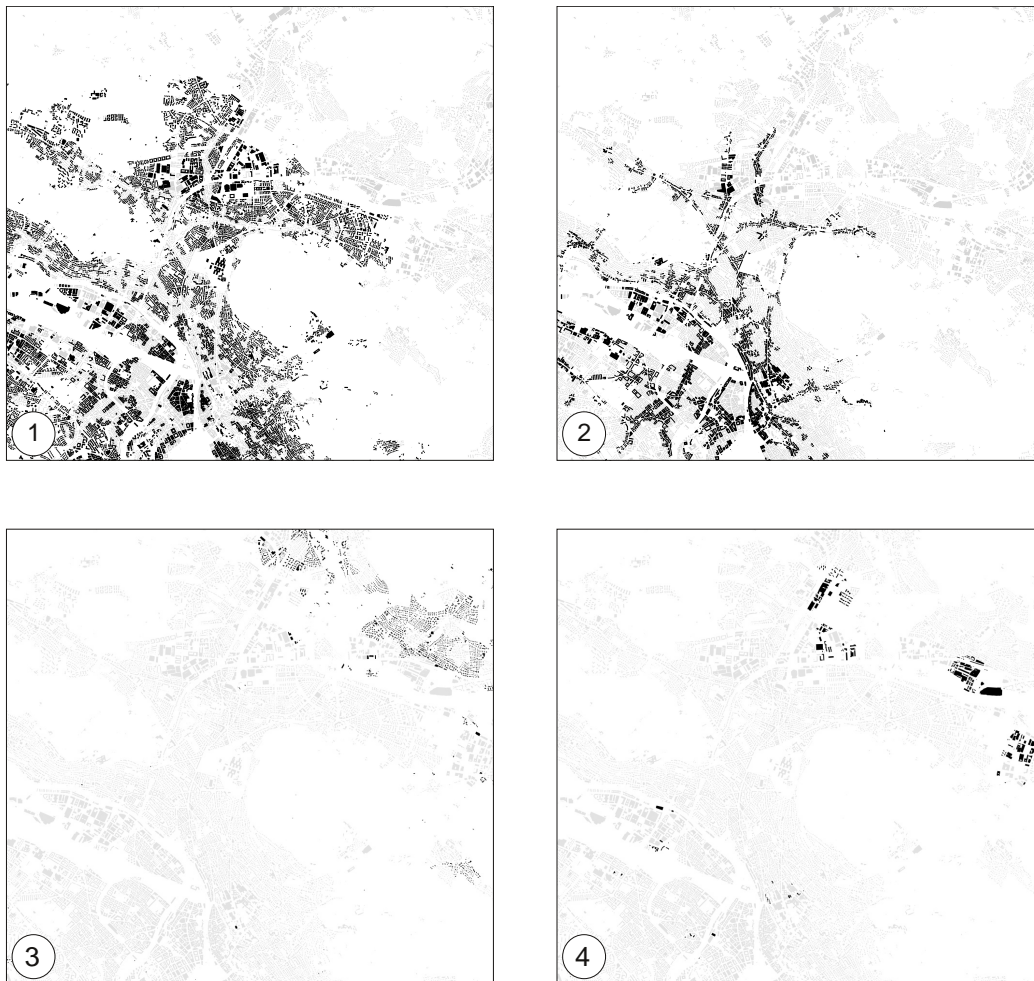**Figure 11:** Selection of municipal typologies

Source: based on ©2010 Amt fuer Raumordnung und Vermessung (GV2113)
These images show examples of the clustering results of the k-medoids clustering on the scale of the municipality. These types refer to similarity in the accessibility and connectivity of locations. The upper two pictures distinguish high accessible sites along the arterials (2) from locations of lower accessibility (1) in the city of Zurich, the lower pictures distinguish residential areas of low accessibility (3) from centers (4) in the outskirts of Zurich

free-flow travel time of only a few minutes includes the whole settlement area and the centrality measures on the small network structure do not result in such large variations as found in the city of Zurich. As a result, we find that the city of Zurich is grouped into one typology for the scale of neighborhood and municipalities.

This paper forms a first attempt to define a set of attributes on urban morphology that allow for implementation in different study areas. However, the simple data model limits the characteristics of the built environment and furthermore ignores a large set of aspects. Examples are the pedestrian networks, public transport network, open space, natural resources, the topography, points of interests, and land use, which will need similar standard attributes. While this work reviews various other studies and extends their findings, the validity of the proposed characteristics in the behavioral models or regression models is outstanding. This will be part of our ongoing work and will be presented in subsequent publications.

References

Alexander, C., S. Ishikawa, M. Silverstein, M. Jacobson, I. Fiksdahl-King, and S. Angel. 1977. *A pattern language*. Oxford: Oxford University Press.

Bafna, S. 2003. Space syntax a brief introduction to its logic and analytical techniques. *Environment and Behavior*, 35(1):17–29.

Batty, M. 2008. Planning support systems: progress, predictions, and speculations on the shape of things to come. In R. K. Brail and R. E. Klosterman, eds., *Planning Support Systems for Cities and Regions.*, pp. 1–30. Cambridge: Lincoln Institute of Land Policy.

Bhat, C. R., K. Kockelman, Q. Chen, S. Handy, M. Hani, and L. Weston. 2000. Urban accessibility index: Literature review. Research Report TX_01/7-4938-1, University of Texas, Austin.

Cherry, N. 2009. *Grid/ Street/ Place: Essential Elements of Sustainable Urban Districts*. Chicago, IL: APA Planners Press.

Christaller, W. 1933. *Die zentralen Orte in Sueddeutschland: eine oekonomisch-geographische Untersuchung ueber die Gesetzmaessigkeit der Verbreitung und Entwicklung der Siedlungen mit staedtischen Funktionen*. Jena: Fischer Verlag.

Clark, C. 1951. Urban population densities. *Journal of the Royal Statistical Society. Series A (General)*, 114(4):490–496.

Clifton, K., R. Ewing, G.-J. Knaap, and Y. Song. 2008. Quantitative analysis of urban form: a multidisciplinary review. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 1(1):17–45.

Dill, J. 2004. Measuring network connectivity for bicycling and walking. In *83th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Dillenburger, B. 2010. Space index - a retrieval-system for building-plots. In G. Schmitt, L. Hovestadt, L. Van Gool, F. Bosché, R. Burkhard, S. Colemann, J. Halatsch, M. Hansmeyer, S. Konsorski-Lang, and A. Kunze, eds., *Future cities: 28th eCAADe Conference Proceedings*, pp. 893–899. Zurich: eCAADe and ETH Zurich.

European Environment Agency. 1985. CORINE land cover. URL http://www.eea.europa.eu/publications/COR0-landcover.

Ewing, R. and S. Handy. 2009. Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban Design*, 14(1):65–84.

Gil, J., J. N. Beirão, N. Montenegro, and J. P. Duarte. 2012. On the discovery of urban typologies: data mining the many dimensions of urban form. *Urban Morphology*, 16(1):27–40.

Google Inc. 2013a. Google earth. URL http://www.google.de/intl/de/earth/.

Google Inc. 2013b. Google maps. URL maps.google.com.

GRASS Development Team. 2013. Geographic resources analysis support system (GRASS) software. URL http://grass.osgeo.org.

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

Hansen, W. 1959. How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, 25(2):73–76.

Hartigan, J. A. and M. A. Wong. 1979. A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.

Hecht, R., H. Herold, and G. Meinel. 2009. Automated Analysis of Settlement Development using Topographic Maps. In *GI2009 - 9. Sächsisches GIS-Forum*. Dresden.

Hillier, B. 1996. *Space is the machine: a configurational theory of architecture*. Press Syndicate of the University of Cambridge.

Hillier, B. and J. Hanson. 1984. *The social logic of space*, volume 1. Cambridge University Press Cambridge.

Holtzclaw, J. 1994. *Using residential patterns and transit to decrease auto dependence and costs*. San Francisco: Natural Resources Defense Council.

INSPIRE. 2013. Infrastructure for spatial information in european community. URL http://inspire.jrc.ec.europa.eu/.

Jacobs, J. 1961. *The Death and Life of Great American Cities*. New York: Random House Inc.

Katz, P. 1993. *The New Urbanism: Toward an Architecture of Community*. New York: McGraw-Hill Education.

Krier, R. 1979. *Urban Space*. London: Academy Editions.

Lam, L., S.-W. Lee, and C. Y. Suen. 1992. Thinning methodologies-a comprehensive survey. *IEEE Transactions on pattern analysis and machine intelligence*, 14(9):869–885.

Lee, D. T. 1982. Medial axis transformation of a planar shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(4):363–369.

Lynch, K. 1984. *Good City Form*. Cambridge: MIT Press.

Maguire, D. J., M. Batty, and M. F. Goodchild. 2005. *GIS, spatial analysis and modelling*. Redlands: ESRI Presss.

Marshall, S. 2004. *Streets and patterns*. London: Routledge.

McDonald, J. F. 1989. Econometric studies of urban population density: a survey. *Journal of Urban Economics*, 26(3):361–385.

McFadden, D. et al. 1978. Modelling the choice of residential location. In A. Karlsqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, pp. 75–96. Amsterdam: North Holland.

Meinel, G., R. Hecht, H. Herold, and G. Schiller. 2008. Automatische Ableitung von stadtstrukturellen Grundlagendaten und Integration in einem Geographischen Informationssystem. Technical Report Heft 134, Bundesamt für Bauwesen und Raumordnung, Bonn.

Newman, M. E. 2003. The structure and function of complex networks. *SIAM review*, 45(2):167–256.

Newman, M. E. J. 2010. *Networks: an Introduction*. Oxford: Oxford University Press.

Ogniewicz, R. L. and O. Kübler. 1995. Hierarchic voronoi skeletons. *Pattern Recognition*, 28(3):343–359.

pgRouting Global Development Group. 2013. pgRouting. URL http://pgrouting.org/.

PostGIS Development Team. 2013. PostGIS. URL http://postgis.net/.

PostgreSQL Global Development Group. 2013. PostgreSQL. URL http://www.postgresql.org.

R Core Team. 2013. R: A language and environment for statistical computing. URL http://www.R-project.org.

R Core Team. 2014. R: Clustering large applications. URL http://stat.ethz.ch/R-manual/R-patched/library/cluster/html/clara.html.

Randall, T. A. and B. W. Baetz. 2001. Evaluating pedestrian connectivity for suburban sustainability. *Journal of Urban Planning and Development*, 127(1):1–15.

Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Schirmer, P. M., M. A. B. v. Eggermond, and K. W. Axhausen. 2014. The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use*, 7(2):3–21.

Schirmer, P. M., C. Zöllig, K. Müller, B. R. Bodenmann, and K. W. Axhausen. 2011. The zurich case study of UrbanSim. In *51st European Congress of the Regional Science Association*. Barcelona.

Sevtsuk, A. 2010. *Path and Place: A Study of Urban Geometry and Retail Activity in Cambridge and Somerville, MA*. PhD thesis, Massachusetts Institute of Technology, Cambridge.

Song, Y. and G.-J. Knaap. 2003. New urbanism and housing values: a disaggregate assessment. *Journal of Urban Economics*, 54(2):218–238.

SRTM. 2000. Shuttle radar topography mission. URL http://www2.jpl.nasa.gov/srtm/.

Swisstopo. 2013. Topographic landscape model TLM. URL http://www.swisstopo.admin.ch/internet/swisstopo/en/home/topics/geodata/TLM.html.

Train, K. 2009. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Vickerman, R. W. 1974. Accessibility, attraction, and potential: a review of some concepts and their use in determining mobility. *Environment and Planning A*, 6(6):675–691.

Von Thünen, J. H. 1826. *Der isoli[e]rte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Hamburg: Perthes.

Wachs, M. and T. Kumagai. 1973. Physical accessibility as a social indicator. *Socio-Economic Planning Sciences*, 7(5):437–456.

Wegener, M. 2004. Overview of land use transport models. In D. Hensher, K. J. Button, E. H. Haynes, and P. R. Stopher, eds., *Handbook of Transport Geography and Spatial Systems*, volume 5, pp. 127–146. Oxford: Elsevier.

Wilson, J. P. and J. C. Gallant. 2000. *Terrain Analysis: Principles and Applications*. John Wiley & Sons.

A APPENDIX

Table 1: Attributes describing the building

ATTRIBUTE		DEFINITION	MEAN	MED.	s.d.
Building					
floorsqm	[m²]	Sum of floorspace indiv. buildings)	3696.50	1546.00	7120.08
footprint	[m²]	Sum of footprints covering the soil	787.24	409.00	1553.18
height	[m]	Height of buildings, ignoring the roof	12.73	12.00	6.12
stories	-	Avg number of stories in composite	*4.12*	4.00	1.82
var_floorsqm	-	Variance of area in composite	1993277.94	0.00	34265933.70
var_footprint	-	Variance of footprints in composite	110919.20	0.00	1905676.55
var_height	-	Variance of height in composite	8.96	0.00	53.36
var_stories	-	Variance of stories in composite	0.50	0.00	2.20
Composite					
box_area	[m²]	Area of bounding box	1734.59	773.00	3198.15
box_length	[m]	Length of bounding box	35.34	27.00	26.58
box_perim	[m]	Perimeter of bounding box	142.61	114.00	95.79
box_width	[m]	Width of bounding box	35.96	28.00	25.75
box_width2length	-	Ratio of width to length in bounding box	0.74	0.79	0.20
circ_area	[m²]	Area of enclosing circle	2346.14	977.00	4496.09
circ_perim	[m]	Perimeter of enclosing circle	140.45	111.00	98.88
circ_radius	[m]	Radius of enclosing circle	22.35	18.00	15.76
conv_area	[m²]	Area of enclosing convex area	909.53	444.00	1626.45
conv_perim	[m]	Perimeter of enclosing convex area	111.92	90.00	74.21
nr_bldg	-	Number of buildings in a composite	2.71	2.00	3.50
obj_area	[m]	Area of shape, i.e. the footprint	677.67	397.00	1135.20
obj_centnr	-	Number of centerlines	2.94	2.00	3.68
obj_compact	-	Compactness of the building defined as $C = obj\_peri2/(4*box\_area*\pi)$	2.50	2.03	1.49
o_courtnum	[m²]	Number of courtyards	0.07	0.00	0.69
o_courtsiz	[m²]	Size of inner courtyard	1145.73	876.00	1258.12
obj_deadnr	-	Number of lines in skeleton that are only connected to one side	5.78	5.00	3.15
obj_dist2cent	[m]	Avg.distance of building corners to centroid	16.70	14.00	10.35
obj_EW	[m]	Length of centerline w.orientation E-W	8.76	0.00	23.72
obj_floorsqm	[m²]	Floorspace, estimated on avg stories	3317.35	1442.00	6364.10
obj_length	[m]	Total length of centerline	40.52	23.00	60.90
obj_lengthmain	[m]	Longest centerline of unique orient. (<45°)	20.24	15.00	18.45
obj_maxwidth	[m]	Maximum width	12.46	11.21	6.32
obj_minwidth	[m]	Minimum width	9.35	9.36	2.97
obj_NESW	[m]	Length of centerline w.orientation NE-SW	8.98	0.00	22.60
obj_NS	[m]	Length of centerline w.orientation NS	10.20	0.00	22.28
obj_orient	-	Main orientation of centerline	-	-	-
obj_orientnr	-	Number of orientations for centerline	1.69	1.00	0.91
obj_perim	[m]	Perimeter of shape	136.56	99.00	122.68
obj_SENW	[m]	Length of centerline w.orientation SE-NW	12.48	0.00	26.30
obj_volskin	-	Ratio of volume to area of facade	3.24	3.11	0.90
obj_width	[m]	Average width along centerline	11.03	10.46	4.19
obj_wings	-	Number of wings	2.15	1.00	2.79
r_boxarea	-	Ratio of area to bounding box area	0.51	0.49	0.16
r_boxperim	-	Ratio of perimeter to bounding box perim.	0.90	0.88	0.16
r_circarea	-	Ratio of area to enclosing circle area	0.41	0.41	0.13
r_convarea	-	Ratio of area to enclosing convex area	0.86	0.89	0.13
r_convperi	-	Ratio of area to enclosing convex area	1.15	1.10	0.18
r_courtsqm	-	Ratio of courtyard to building footprint	0.37	0.42	0.34

Table 2: Attributes describing the composition

ATTRIBUTE		DEFINITION	MEAN	MED.	s.d.
Street					
obj_streetfr	[m]	Length of facade that is perceived from street	37.93	20.00	53.32
obj_streetor	-	Facing the street or aligning the street	*dummy*	-	-
Building					
dist_bldg	[m]	Distance to closest neighboring builinding	10.89	10.26	6.08
dist_bldgbl	[m]	Distance to next building in neighboring block, as alias for street-width	36.30	25.08	30.38
obj_courtyper	-	Number of perceived courtyards: Allows for gab between buildings according to height (1/2*height)	0.12	0.00	0.38
obj_inflzone	[m²]	Area if influence zone around building	3494.93	2358.00	3480.99
obj_inflscr	-	Soil coverage ratio of building footprint to influence zone	0.20	0.19	0.10
Block					
dist_block	[m]	Distance of building to block border: is alias for distance to street	9.25	5.62	11.55
r_blockscr	-	Ratio of building footprint to block area	0.03	0.01	0.07
bl_area	[m²]	Area of block	216143.59	31336.00	1011385.80
bl_corner	-	Building is at corner of block, i.e. at streets with change of orientation (> 45°)	*dummy*	-	-
bl_permind	-	Permeability: number of views between buildings in ratio to perimeter of block	0.02	0.02	0.01
bl_closind	-	Closeness of block: ratio of street facing facades to length of perimeter of block	0.45	0.43	0.25

Table 3: Attributes describing the neighborhood

ATTRIBUTE		DEFINITION	MEAN	MED.	s.d.
Building					
dens_area100	[m²]	Sum of sqm from footprints in 100m radius	10716.54	8374.00	8549.01
dens_area300	[m²]	Sum of sqm from footprints in 300m radius	55139.72	48780.00	31891.91
dens_area500	[m²]	Sum of sqm from footprints in 500m radius	124248.93	107257.00	71949.85
dens_bldg100	-	Number of buildings in 100m radius	26.49	26.00	10.71
dens_bldg300	-	Number of buildings in 300m radius	148.58	148.00	59.91
dens_bldg500	-	Number of buildings in 500m radius	331.42	326.00	135.04
dens_sqm100	[m²]	Sum of floorspace sqm in 100m radius	44715.62	26056.00	51208.07
dens_sqm300	[m²]	Sum of floorspace sqm in 300m radius	224104.17	145908.00	209793.94
dens_sqm500	[m²]	Sum of floorspace sqm in 500m radius	506427.27	313623.00	480519.09
Streets					
net_dead100	-	Number of deadends within 100m	1.71	1.00	1.01
net_dead300	-	Number of deadends within 300m	7.48	7.00	4.44
net_dead500	-	Number of deadends within 500m	18.71	17.00	9.33
net_inter100	-	Number of intersections with 100m	4.25	4.00	2.49
net_inter300	-	Number of intersections with 300m	32.37	31.00	13.84
net_inter500	-	Nr of intersections with 500m	80.74	79.00	33.37
net_length100	[m]	Total length of network within 100m	1665.77	1556.00	711.62
net_length300	[m]	Total length of network within 300m	10965.91	10671.00	3064.01
net_length500	[m]	Total length of network within 500m	27006.88	27166.00	7215.57
Streets & Buildings					
net_prox_bldg	-	Buildings in 2min driving distance	349.69	321.00	182.91
net_prox_sqm	[m²]	Floorspace in 2 min driving distance	517091.33	310493.00	531606.89
net_prox_area	[m²]	Catchment area of 2 min driving distance	797542.26	803723.00	331298.91
net_prox_length	[m]	Network in 2 min driving distance	12669.59	11967.50	6611.57
net_prox_dead	-	Dead ends in 2 min driving distance	21.09	19.00	12.09
net_prox_inter	-	Intersections in 2 min driving distance	101.40	93.00	58.30

Table 4: Attributes describing the settlement area

ATTRIBUTE		DEFINITION	MEAN	MED.	s.d.
Streets					
net_betw_length	-	Betweeness of road-network using length of segment as cost	251726.75	36512.00	510824.22
net_betw_one	-	Betweeness of road-network using constant per segment as cost	641827.12	42436.00	1703168.58
net_betw_speed	-	Betweeness of road-network using traveltime of segment as cost	287518.95	39579.00	590446.09
net_close_length	–	Closeness of road-network using length of segment as cost	3179.27	2656.21	1745.70
net_close_one	-	Closeness of road-network using constant per segment as cost	37.49	28.63	20.96
net_close_speed	-	Closeness of road-network using traveltime of segment as cost	249.18	220.00	127.54
net_degr_speed	-	Degree of road-network, identifies intersections	0.00	0.00	0.00
Streets & Buildings					
net_acc_bldg	-	Accessibiltity of buildings as distance decay function	9.85	9.16	5.01
net_acc_sqm	-	Accessibiltity of floorspace as distance decay function	14755.86	9849.28	14741.02
Urban Centers					
cent_area	[m^2]	Catchment area (prominence) of next center	505259.15	90921.95	1708064.28
cent_dist	[m]	Distance to next center	200.83	187.30	120.68
cent_loc	-	Local impact of clostest center, scaled to range of 0-255	49.34	30.00	52.06
cent_max	-	Maximum value of closest center, scaled to range of 0-255	136.34	128.00	56.75
Settlement					
dist_settle	[m]	Distance to border of own settlement	303.59	263.16	176.01