Inferring the Scale of OpenStreetMap Features

Guillaume Touya, Andreas Reimer

Inferring the Scale of OpenStreetMap Features

Guillaume Touya [1]

Andreas Reimer[2]

Abstract Traditionally, national mapping agencies produced datasets and map products for a low number of specified and internally consistent scales, *i.e.* at a common level of detail (LoD). With the advent of projects like OpenStreetMap, data users are increasingly confronted with the task of dealing with heterogeneously detailed and scaled geodata. Knowing the scale of geodata is very important for mapping processes such as for generalization of label placement or land-cover studies for instance. In the following chapter, we review and compare two concurrent approaches at automatically assigning scale to OSM objects. The first approach is based on a multi-criteria decision making model, with a rationalist approach for defining and parameterizing the respective criteria, yielding five broad LoD classes. The second approach attempts to identify a single metric from an analysis process, which is then used to interpolate a scale equivalence. Both approaches are combined and tested against well-known Corine data, resulting in an improvement of the scale inference process. The chapter closes with a presentation of the most pressing open problems

Keywords: Generalization, Scale, Data Quality, Level of Detail, OpenStreetMap

[1] G. Touya

COGIT,

Institut national de l'information géographique et forestière, France
Email: guillaume.touya@ign.fr

[2] A. Reimer

Department of Geography,

Heidelberg University, Germany
Email: andreas.reimer@geog.uni-heidelberg.de

1. Introduction

Studying the quality of OpenStreetMap (OSM) data has been a hot research topic in recent years, as OSM has grown and applications flourished. OSM quality can be assessed by comparisons with a reference dataset (Haklay 2010, Girres and Touya 2010, Zielstra and Zipf 2010), with intrinsic measurements (Barron *et al.* 2013), with contributor trust inference (Skarlatidou *et al*. 2011, Kessler *et al.* 2013), or in a fitness for use context, like pedestrian routing (Mondzech and Sester 2011). The latter method is the concern of this chapter, particularly the assessment of the fitness of OSM data for high-quality cartography (Sester *et al.* 2014). Besides being crucial for automated derivation of cartographic products, scale evaluation also touches issues for the appraisal of OSM data for geographic analyses, linking the issue with the bigger problem of scale in Volunteered Geographic Information (VGI) (Jokar 2014, Feick and Robertson 2014).

OSM data can be very rich in the regions with good completeness, but the level of detail of objects is very heterogeneous (Touya 2012). For instance, Figure 1 shows three very complex religious buildings: the first one is captured with many details while the others are captured coarsely. This heterogeneity is a major obstacle for automated cartography, as existing automatic processes, like map generalization processes, are parameterized for a given homogeneous input scale. As a consequence, producing smaller scale maps by automatic generalization requires the inference of the initial scale or level of detail for every object as it can be different for each. Even for the production of large-scale maps (*e.g.* city plans) where map generalization is not necessary, level of detail is important. By introducing confusion in the spatial relations between detailed and less-detailed objects, such inconsistency may mislead the map reader. Whether a building is actually in a clearing or in a patch of woodland that was imported from a generalized source or even incompletely mapped is undecidable without some information on scale (Touya and Brando 2013). Apart from cartographic communication, investigations into land cover/land use obviously face similar problems. Two concurrent methods have been proposed to infer the scale/LoD of OSM objects individually (Touya and Brando 2013, Reimer *et al.* 2014). In this chapter, we review and compare both methods to improve the quality of scale/LoD inference for OSM.

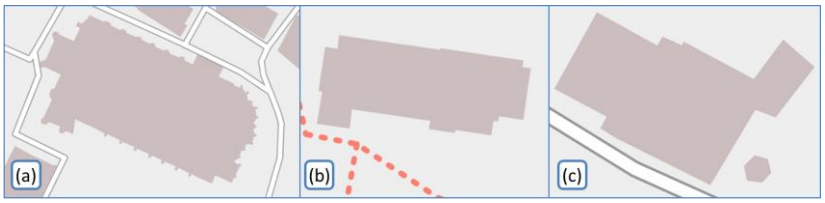

Figure 1. Different levels of detail for three very complex religious buildings with similar granularity. (a) Is captured from cadastre data while (b) and (c) are captured from Bing images.

The second section discusses the cartographic and geographic notions of scale and level of detail. The third section briefly presents both scale inference methods, while Section 4 explores how these methods can be combined to improve the scale inference of OSM data. Section 5 describes how scale inference can be used to enhance the automated mapping techniques for OSM data, to illustrate the usefulness of the research. After a discussion in the sixth section, the chapter is concluded and further research is proposed.

2. Scale and Level of Detail

All scientific modelling approaches implicitly or explicitly must deal with the question of the conceptual and numerical relationships between reality and the model being created. In the most straightforward interpretation, a numerical relationship between reality and the model is the ratio of values in reality to values in the model. For all spatial sciences and endeavors, be they model railroads or traffic simulations, informative numerical ratios as an expression of the reality–model relationship are based on geometries. For maps as graphical models of geographic phenomena, the ratio between distances as measured on the map and as measured in reality is called “scale” (International Cartographic Association 1972). As it has so far been inefficient to provide maps at an arbitrary scale, mapping agencies have prepared maps and map series at certain scale groups. Furthermore, in cartography, scale has become shorthand for usage environment for the map and conceptual organization of the subject matter depicted. Freitag (1962) writes: “The scale of a map is one of its constitutive elements; it determines information density, readability, significant contents and area of application.” Due to this historical development, the seemingly objective and quantifiable numerical scale has become attached to precepts about the conceptual relations between geographic reality and its graphical model (Freitag 1962, Bollmann and Koch 2002). This was driven by and is congruent with developments in geography (Sudgen and Hamilton 1971) and other geosciences such as ecology (Steinhardt 1999, Levin 1992,

Chave 2013). In all cases, ranges of numerical scales are grouped into conceptual hierarchies. Consequentially, each level of the hierarchy is understood to be the realm of certain phenomena and specialized research methods only applicable within the bounds of that level. Haggett *et al.* (1965) proposed normalization over the surface of the earth as the basis for a logarithmic subdivision as a "yardstick" (G-scale) for a continuous hierarchy of modelling perspectives. Without any direct connection to the geosciences or cartography, computer graphics research saw itself confronted with the problem of "enhancing performance and realism" of "computer produced pictures" for graphical models of reality (Clark 1976). It is not surprising that one of the most enduring solutions was to devise hierarchical geometric models, which came to be known as "Levels of Detail" (LoD) (Clark 1976). As such, LoD and the hierarchical models of space in the geosciences are quite similar in purpose and argumentation. Both approaches also share the need for a concretization for specific use cases.

Touya and Brando (2013) define the LoD of a geographical dataset as the conjunction of several factors, namely the conceptual schema of the data, the semantic resolution, the geometric resolution, the geometric precision, and the granularity. The *conceptual schema* component is the way ground truth is represented in the geographical database: polygonal features representing forests or point features representing individual trees are conceptual schemas that correspond to different LoDs. The *semantic resolution* is the quantity of details in the attribute data attached to geographical features. By analogy with raster resolution, the *geometric resolution* of vector features is approximately the minimum distance between two vertices of the geometry. The *geometric precision* is simply the positional shift between ground truth and the represented feature. Finally, *granularity* describes the size of the smallest shapes of features, such as the smallest protrusions of buildings or the smallest width for sharp bends in a road. Biljecki *et al*. (2014) proposed a formal framework to measure LoD, but it is applicable for 3D city models.

As CityGML standard did for 3D city models (Kolbe 2009), LoD categories can be defined for maps. Touya and Brando (2013) proposed five LoD categories that would be used in this chapter, from the more detailed to the less detailed: street, city, county, region, and country. The *street* LoD contains features represented for parcel management or street orientation. For example, the British Ordnance Survey MasterMap can be considered as a street LoD dataset. The *city* LoD contains features to describe what is visible on the ground (buildings, roads, rivers, forests, *etc*.). It is the LoD of classical 1:25k topographic maps. The *county* LoD contains features that represent a small region to allow tourist-like displacements (*e.g.*, visits, hiking, cycle rides). The *regional* LoD is related to the representation of a large region and only contains important roads and geographic features. Finally, the *country* LoD is even less detailed, for the representation of countries or very large regions.

In order to compare LoD and scale inferences, we have to match the LoD categories used in this chapter with scale values. Table 1 summarizes an attempt at such a matching. Although this matching is somewhat subjective, it was used in the remainder of the chapter to compare scale and LoD inferences.

Table 1. Matching the LoD categories from Touya and Brando (2013) with map scale ranges, *i.e.* the objects with a given LoD category can be mapped within this scale range, without significant generalization.

LoD Category	Street	City	County	Region	Country
Scale range	<1:15k	1:15k–1:50k	1:50k–1:150k	1:150k–1:750k	>1:750k

3. Two Methods for the Automatic Inference of Scale

As we have noted before, information on the scale/LoD of a given dataset is crucial for further geoprocessing operations such as automated generalization. Before the rise of datasets like OSM, information on scale/LoD was just part of the metadata, but with user generated geodata, capture or intended scale/LoD are not documented. To address these problems, we present two different data enrichment approaches that attempt to infer the scale/LoD of OSM-data. The first approach can be understood as a rationalist (Touya and Brando 2013), the second as an empiricist approach (Reimer *et al.* 2014).

3.1 Scale Inference with a Multiple Criteria Decision Technique

3.1.1 Measuring Level of Detail

The definition of LoD presented in Section 2 showed that it is more complex than geometric resolution alone, as many aspects are involved in its characterization. Assuming that LoD inference is the aim, rather than scale, several measures have to be used to properly infer LoD. Following the definition given in Section 2, measures can be used to infer LoD while covering all its aspects (Touya and Brando 2013).

Measuring the conceptual schema:

The feature type of the object, in the Open Geospatial Consortium sense, can help us to measure the schema aspect of the level of detail. For instance, buildings or points of interest have a higher LoD than built-up areas. Although the Resource Description Framework (RDF) structure of OSM data prevents deriving of the feature type from classes, it can be derived from the main tags of objects like highway, building, or amenity.

Measuring the semantic resolution:

The annotation process to add semantic information on objects improves the level of detail of OSM objects by giving specifications on the object, but OSM is sparsely tagged (Mooney and Corcoran 2012) and only a few objects contain several tags. A very simple measure that counts tags is used. Metadata-like tags, such as source, or created_by, are counted as well as property tags like name, and language specific tags, such as species:fr that gives the French name of a tree species while the generic tag species gives the Latin name.

Measuring the geometric resolution:

Geometric resolution can be measured by assessing the density of vertices in relation to the length of objects. Indeed, we assume that a higher frequency of vertices denotes a digitization of objects at a larger scale or level of detail. The vertex density measure has been empirically normalized between 0 and 1 using the maximum and minimum values found in French authoritative datasets at different LoDs (1:25k, 1:100k and 1:250k). However, such a measure penalizes curvy lines that require more vertices to digitize than a straight line. So, curved objects (*e.g.* rivers) may be artificially considered as more detailed. So, we added a measure, the median of edge lengths (Figure 2), to better capture the resolution of curved objects. It has been proved more effective than vertices density on mountain roads by Girres (2011). This measure is also empirically normalized using the same authoritative datasets.

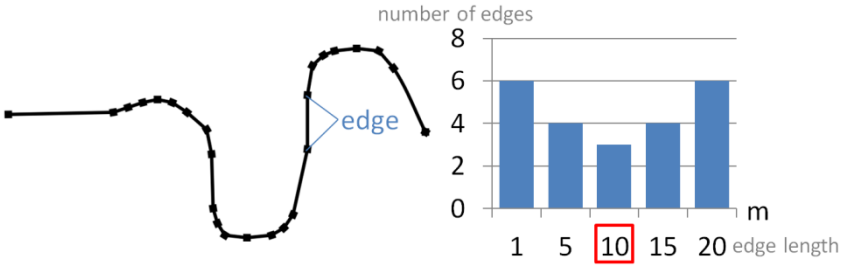

Figure 2. The median of edge lengths gives a clearer idea of geometric resolution for curved objects like mountain roads or rivers.

Measuring the geometric precision:

Intrinsically assessing the geometric precision is a complex task and previous research giving a geometric precision for OSM objects used a reference dataset (Haklay 2010, Girres and Touya 2010). However, the source tag in OSM informs on where the contribution comes from (*e.g.* GPS survey, Bing imagery, imported from open datasets). Knowledge on the rough precision of such sources can give us a vague but useful insight on geometric precision. The main possible values for the source tag have been listed from the TagInfo tool, and a rough LoD equivalency was derived when possible: a GPS survey is usually quite precise (value set to city) while digitization from Landsat creates imprecise objects (value set to region).

Measuring granularity:

Granularity captures the minimum size of details in an object shape. The detail can be the whole shape, a simple protrusion, or a bend in a linear object. To capture granularity in all three cases, three measures are proposed: size, smallest edge, and coalescence. First, small areas indicate a high granularity for small-sized objects but do not indicate anything about the granularity of large objects. The measure was calibrated with the smallest sizes found in the authoritative datasets at several LoDs used before. The size criterion is only applied to polygons. Then, the smallest edge criterion analyzes the length of the shortest edge between two vertices to infer granularity. This criterion is a classical measure to assess building granularity in cartographic generalization (Stoter *et al.* 2009). Note that when used on raw data, shortest edge values can become infinitesimally small due to nigh-double clicks and other digitization errors. A clean-up step is suggested when used on OSM data, for example. Finally, the coalescence (*i.e.* symbol overlap in a curve) criterion is used for the inference of linear features' granularity. This crite-

rion is based on the principle that if a linear feature symbol coalesces at a given symbol width, the feature cannot be displayed at a scale that requires such a width for eye perception (Girres 2011). It could also be used on polygons but it has not been tested.

To summarize, eight criteria can be used to infer LoD in OpenStreetMap:

- Feature type (conceptual schema).
- Number of tags (semantic resolution).
- Vertex density (geometric resolution).
- Median edge length (geometric resolution).
- Capture source (geometric precision).
- Size (granularity).
- Smallest edge (granularity).
- Coalescence (granularity).

3.1.2 Combining Criteria to Infer a LoD Category

The eight proposed criteria are quite diverse, and hardly comparable, so their aggregation to infer a LoD category is challenging. Multiple-criteria decision-making is a computer science domain that researches methods to that are able to cope with diverse and intuitively hard-to-compare measurements (Figueira *et al.* 2005a). Such techniques can help us to infer LoD from the eight proposed criteria. Among the large number of existing techniques, we chose the ELECTRE TRI method (Figueira *et al.* 2005b) as its properties match our needs:

- The decision is a classification into categories.
- There are more than three criteria.
- The criteria are hardly comparable (*i.e.* how to compare the vertex density measure with the capture source criterion?).
- The criteria give fuzzy results that may be insignificant taken individually.

In ELECTRE TRI, the comparison between two vectors of measures (that gather the measures for all criteria) is not made by aggregating the measures. Comparisons are made criterion by criterion, and according to this comparison, each criterion votes for or against the assertion that “vector u outranks vector v”. To make a classification, each category is assigned a lower-bound vector and an upper-bound vector to which the vectors to classify are compared. Therefore, we had to define, the lower-bound and upper-bound vectors for each category: *e.g.*, the feature type “building” is the feature type criterion value for the lower-bound vector of the category City LoD. The bounds definition is based on the authoritative datasets. The criteria are able to vote indifference when the criterion alone is

not sufficient to decide what outranks what. For instance, the criterion "feature type" is indifferent when the values up for comparison are "footpath" and "residential road". The criteria are also able to vote a veto, *i.e.* the outranking decision is based only on this criteria. For instance, the granularity criterion will put a veto if a 1-m smallest edge is compared to a 500-m smallest edge.

As the importance of any single criterion in the global method is hard to assess, a sensitivity analysis was conducted. We compared the global result (*i.e.* with all criteria) with results derived from only a subset of the criteria. The sensitivity analysis shows stability for some objects (*e.g.* building and forest), *i.e.* whatever the criteria used, the LoD category is the same. Other objects show more variability: for the tested streets or rivers, the choice of the criteria among the eight available significantly impacts the LoD category that is inferred.

As a strategy to make the method more robust against such feature type-dependent sensitivity, we compute the LoD for all the possible permutations of criteria (with at least four criteria to all criteria). The geometric mean of all the inferred categories is used as the new LoD value, robust against outliers and changing sensitivity.

The multiple criteria decision method and all the criteria have been implemented in the open source research platform GeOxygene (Grosso *et al.* 2012), which is able to load OSM datasets. Some results automatically computed on built-up areas, like in Reimer *et al.* (2014), are presented in Figure 3.

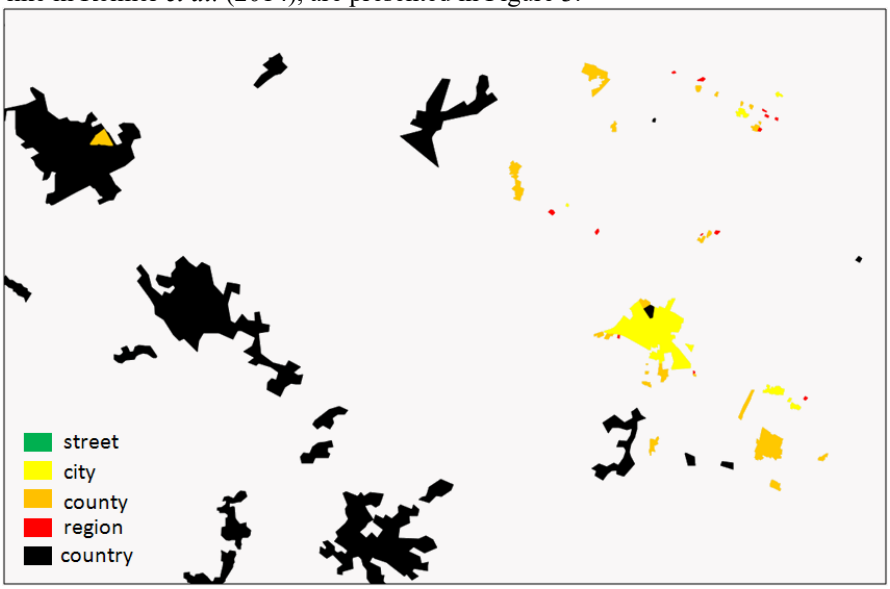

Figure 3. LoD inference results for built-up areas in France: LoD is globally low because built-up areas are imported from Corine Land Cover.

3.2 Empiric Scale Inference

The empiric approach as presented in Reimer *et al.* (2014) is based on an inversion of Töpfer's radical law (Töpfer and Pillewizer, 1966). The basic idea is to create a specific radical law applicable to a specific class of objects by empirically determining a measurable constant and inverting it, to determine the scale equivalency of a given object. The scale equivalency is then expressed as if it was a cartographic scale. This empiric scale inference method is thus divided into three phases:

1. Identification of a suitable measure.
2. Determination of the specific constant.
3. Automated application.

Phases 1 and 2 are currently not fully automated, but conceivably could be automated, assuming some further work in defining criteria for selecting suitable measures. Note that phase 2 needs to be informed by the relation of the selected measure to the object, potentially modifying the degree of power of the equation. Area-based measures will most likely be of third degree and linear measures of second-degree complexity, disregarding other factors such as a drop or increase in apparent density. For a detailed discussion of the determination of the exponent x in the equation:

$$n_F = n_A \cdot \sqrt{\left(\frac{M_A}{M_F}\right)^x},$$

where n_F is a measure in the follow-up scale, n_A is a measure at the original scale and M_A and M_F are the scale factors at original and follow-up scale respectively, please see Töpfer (1979, pp. 43).

For the case of urban area polygons, the vertex frequency was identified from many other potential measures generated for urban area polgyons for 1:250k scale map products from three different national mapping agencies. Note that vertex frequency was used for a related purpose by Dutton (1999). Other potential measures such as minimum edge length, area, vertex per area, angular resolution, minimum angle, *etc*. were tested and discarded either due to being too different across the tested maps or due to being statistically dependent on the area of the polygon. The vertex frequency tested negatively for dependence on the area of the polygon, *i.e*. there is no statistically significant correlation between frequency and polygon size. Furthermore, the measure showed very stable distributions with low standard deviations around a concrete value of 1 vertex per millimeter drawing space across a wide range of maps of the same scale produced by different organizations. We thus set up the scale equivalency S_E as:

$$S_E = 250000 \cdot \left(\frac{1.0 \frac{v}{mm}}{n_F} \right)^2 .$$

Note how this is equivalent to the 1st degree (x = 1) selection equation based on Töpfer (1979):

$$M_F = M_A \cdot \left(\frac{n_A}{n_F} \right)^2 .$$

Contrary to common experience, no significant numerical difference could be detected whether curvy rivers and coastlines were part of the urban area polygons or not. We highlight this as an example of the conceptual difference compared to the multi-criteria analysis, where the choice of criteria is made *a priori*, *i.e.* without a preceding statistical–empirical analysis. Whereas intuition and expert knowledge suggest that there is a strong influence on the artificial/natural form dichotomy of polygons on vertex frequency, these effects are not measurable for this specific object type and scale range.

4 Combining Both Methods to Improve Scale Inference

4.1 Compared Evaluation of Both Inference Methods

To compare both methods, scale was used and the improved LoD inference (the mean of category inferences when criteria vary) was interpolated to scale, using the scale ranges from Table 1, and a linear interpolation inside each category (*i.e.* a 2.5 LoD mean corresponds to a 1:50000 scale). Then, five test areas were selected with only built-up areas as the empiric scale inference is only tuned for such objects. Four of the test areas come from areas of interest highlighted in Reimer *et al.* (2014): Africa, western Ukraine, Australia, and Belgium. A test area in France, where the multiple criteria method was initially tested, was also added. Figure 4 clearly highlights the general trend on all tested datasets, that both methods do not infer scale the same way. The scale equivalency (SE) tends to infer larger scale than the multiple criteria method (SL). The results are summarized for all built-up areas of all five test datasets in Table 2.

Table 2. Comparison of scale inferences from both methods in all five datasets (17402 objects).

	LoD Mean	Scale from LoD (S_L)	Scale from equivalency (S_E)	Scale gap (S_L/S_E)
mean	3.94	1:18996375	1:81368	638.9
max	5.00	1:75375000	1:32017460	11499.7
min	1.93	1:14475	1:5	0.3
median	3.94	1:411549	1:11149	61.1

Occasionally, SL is larger than S_E, but the mean scale gap is around 40 in most datasets (the gap is only significantly bigger in the French dataset). However, both methods are quite consistent in assessing the wide diversity of scales for built-up areas across the test datasets. When such a difference is measured, it is necessary to assess which one is closer to some form of ground truth. For this purpose, we used the French dataset where almost all built-up areas are derived from an automatic import of the Corine Land Cover European dataset, which is confirmed by the *source* tag on the objects. Corine Land Cover is produced by remote sensing for a scale of 1:100k. So, the both scale inferences should tend to 1:100k on the French dataset. S_E appears to be the more accurate scale: it slightly underestimates the scale with a mean scale of 1:43k, while S_L is much further with a mean scale of 1:2.5M.

The large bias of S_L may have several explanations. First, the multiple criteria method was not intended to infer scale but LoD, being a different and fuzzier notion (see Section 2). Measurements on geometry are balanced with measures on feature type, or metadata, which significantly impacts the inferred scale. Moreover, the geometry measurements were not calibrated on built-up areas but mostly on roads and buildings (Touya and Brando 2013), and they were not changed for this comparison to see if it mattered. Finally, the interpolation between LoD categories and scales is quite fuzzy and inaccurate. The ranges used for each category could be tuned and a non-linear interpolation could be used for small scales, where the scale-range is very large compared to the scale ranges of *street* and *city* categories. On the other hand, the scale equivalency also appears to be biased, as the only consideration of vertex density tends to increase scale equivalency too much. Taking granularity into account would maybe make S_L more accurate.

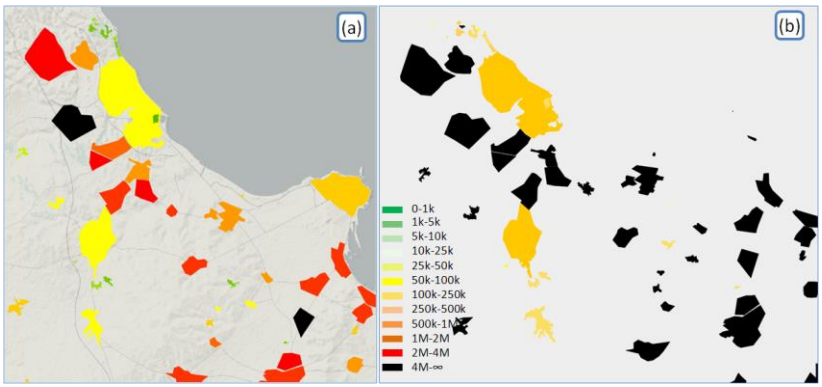

Figure 4. (a) Reimer *et al.*'s (2014) scale equivalency on built-up areas in Tunisia. (b) Interpolated scale from multiple criteria LoD inference on the same dataset.

Although the exemplary scale equivalency was calibrated with built-up areas in authoritative datasets, it was computed on other map types of objects, as well as the multiple criteria LoD inference, in order to further explore both methods. Such evaluations can help us to know if the scale equivalency really has to be calibrated several times for several types of objects, which is the baseline assumption. Tests were carried out on rivers (tag *waterway* = river), forest (tag *landuse* = forest) that are geographically far from built-up areas and industrial areas (tag *landuse* = industrial), which is a type of object close to built-up areas (Figure 5). Two areas with good completeness in France and Germany were selected.

In all three cases, the scale gap between S_L and S_E is less pronounced, particularly on rivers that are quite different objects, while forest and industrial areas are considered as land use areas. The gap decrease is partly due to the inference of smaller scales by S_L compared to the inference for built-up areas. Indeed, vertex density is correlated to the shape of objects, and curved shapes like forests and rivers tend to be composed of more vertices than built-up or industrial areas with the same length

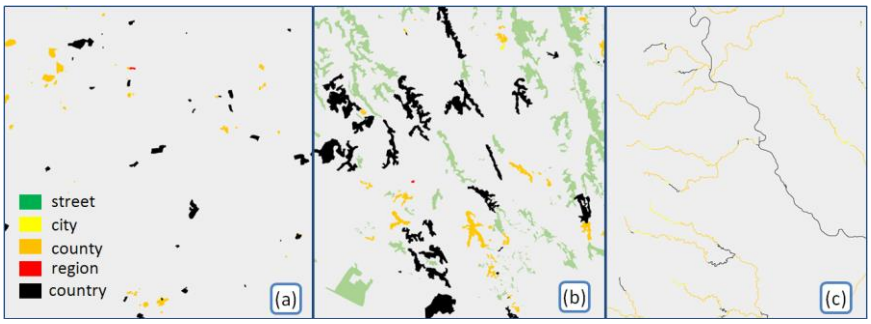

Figure 5. LoD inference on industrial areas in (a) Germany, (b) forests in France, and (c) rivers in Germany.

4.2 Mixing Both Inference Methods

The exemplary scale equivalency for polygons as a function of the vertex density measure is naturally quite similar to the vertex density criterion used by Touya and Brando (2013), normalized by benchmarking on existing datasets. Hence, we tried to compute the multiple criteria scale inference replacing the vertex density criterion by the scale equivalency measure. The previous section showed that the multiple criteria method tended to yield much smaller scales compared to Reimer *et al.*'s (2014) scale equivalency, so adding the scale equivalency as a criterion could plausibly improve inference results.

Table 3. Comparison of LoD inference with and without Reimer *et al.*'s (2014) scale equivalency as a criterion, on 4858 built-up areas from the French and African datasets (tag landuse = residential).

	Mean LoD with vertex density	Mean LoD with Scale Equivalency	LoD difference
mean	4.18	3.12	1.06
max	5.00	4.20	1.72
min	1.98	1.98	-0.03
median	4.64	3.06	0.94

The scale inference was computed with the vertex density criterion and with the scale equivalency instead on the same datasets of built-up areas (Table 3). The comparison was also computed on three buildings datasets (Table 4) in different parts of the world in order to assess the impact of the scale equivalency criterion on geographical data it was not calibrated for. The result for built-up areas shows that replacing the vertex density criterion clearly influences the LoD inference with a mean decrease of the LoD of 1 category, giving a mean scale of approximately 1:90k, which is consistent with the computation of scale equivalency on the same datasets. As most built-up areas in the tested datasets are imported from Corine Land Cover, a dataset specified for 1:100k scale, the scale inference is more accurate with Reimer *et al.*'s (2014) scale equivalency for these objects.

On the contrary, the LoD inference on building remains nearly the same with or without Reimer *et al.*'s (2014) scale equivalency, which means that the calibration of both measures for small and detailed features like buildings produce quite similar results.

Table 4. Comparison of LoD inference with and without Reimer *et al.*'s (2014) scale equivalency as a criterion, on 9153 buildings from the three chosen datasets.

	Mean LoD with vertex density	Mean LoD with Scale Equivalency	LoD difference
mean	2.05	1.99	0.05
max	4.66	3.95	1.48
min	1.64	1.61	-0.25
median	2.02	2.00	0.03

5 LoD Harmonization for Large-Scale Automatic Mapping

When OSM is an input for the automatic derivation of large-scale maps (*e.g.* 1:10k), highly detailed objects are displayed with less detailed objects. When these objects share spatial relation (*e.g.* inclusion) in reality, the LoD difference may alter this spatial relation and blur the map readability (Touya 2012). In this case, the automatic inference of scale/LoD for individual objects is very useful. It helps identifying those spatial relations that are damaged because of LoD difference (Touya and Brando 2013). The spatial relations are identified by spatial rela-

tions algorithms and stored as complex objects of the dataset. Then, the LoD inference is a required input for an automatic LoD harmonization process that could restore the spatial relation (Touya and Baley 2014); here spatial relations could be made explicit as map objects in order to assure their restoration.

For instance, OSM built-up areas are often poorly detailed (see experiments above, Section 4.1) while buildings are quite detailed in France (imported from cadastral data). So, buildings may lie just outside built-up areas. The harmonization process extends the geometries of built-up areas to include those buildings, restoring the spatial relation of buildings included in a built-up area (Figure 6).

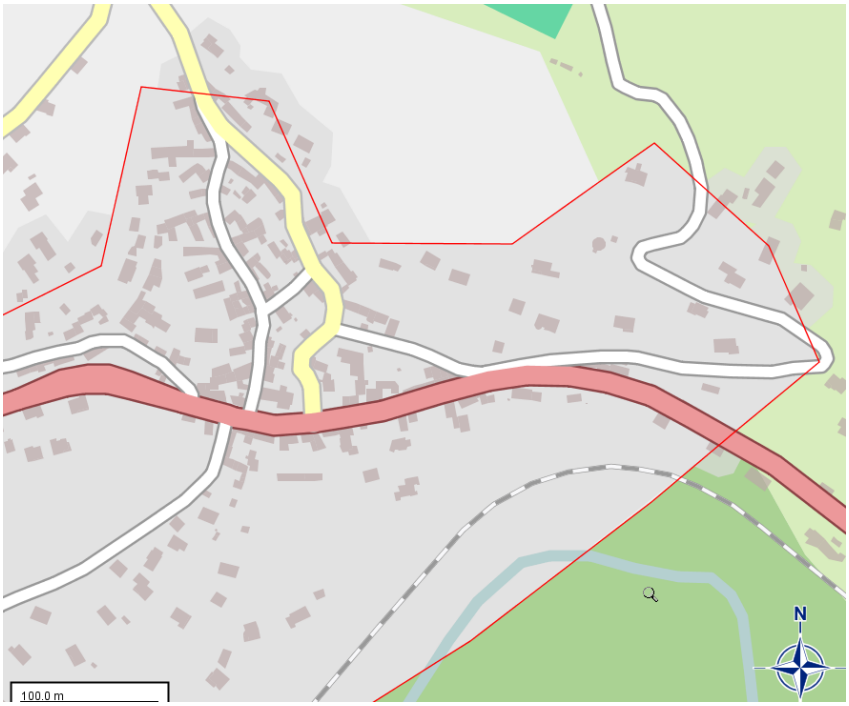

Figure 6. Automatic harmonization of a built-up area (in red), extended (in grey) to include buildings with higher LoD.

6 Open Problems

6.1 Scale Inference for Point Objects

Both methods rely on some kind of point density estimation based on the series of vertices that describe objects' geometry, which is not immediately applicable to point objects (Dutton 1999, Bereuter and Weibel 2013). So, how can we infer the scale or LoD of point objects?

In the multiple criteria method, all criteria related to geometric resolution and granularity become meaningless, and the conceptual schema, semantic resolution and geometric precision criteria remain. We tested the method with the remaining criteria on several point features in OSM: trees, aerial way pylons, bicycle_rental, bus stops, towers, peaks, and power poles. These features are diverse in terms of real extent and geographical neighborhood. A fourth criterion, called version number, was added to assess the features with a crowdsourcing approach (Goodchild and Li 2012): although the Linus' law (*i.e.* quality increases with the number of active contributors) has not been completely proved for VGI (Haklay *et al.* 2010), the number of versions of a feature gives hints on the number of times a feature has been improved by OSM contributors. This criterion assesses the objects with many versions as detailed objects.

As the source tag is seldom filled, the most differentiating criterion is the *feature type* criterion. The point objects that represent a small object like a tree, bus stop or a pylon are attributed to a detailed LoD, while objects representing a large (and fuzzy) extent like peaks or bicycle rental places are given less detailed LoD. Within a feature type, the three other criteria (source, number of tags, and version number) make the objects more or less detailed: for instance, a peak captured by GPS with elevation and name tags is given a higher LoD than a peak captured from Bing imagery (Figure 7).

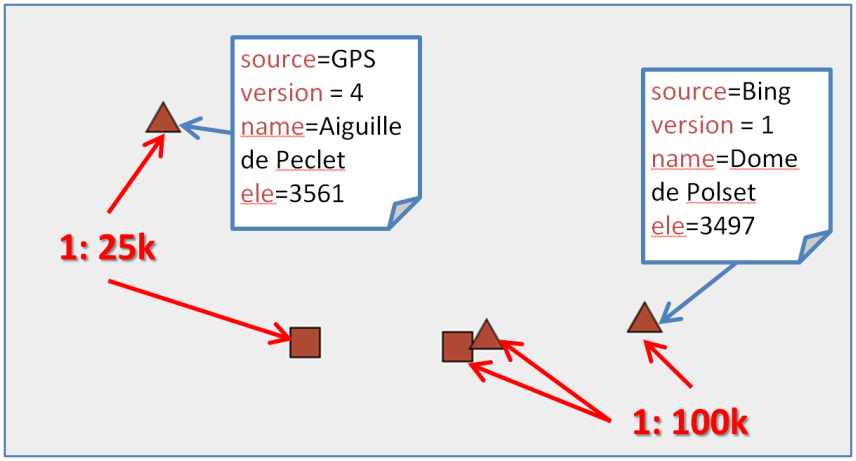

Figure 7. Peaks (triangle symbols) and passes (square symbols) extracted from OSM in the French Alps. Scale inference (computed from the multiple criteria method) varies according to the tags on the objects.

These results show that the LoD inference of point objects could be improved. First, the version number criterion only uses the version number, so there is no information about the number of contributors that improved the object, and it may be a single contributor. However, the wisdom of crowd theory assumes that better decisions can be made when there are multiple independent contributors (Surowiecki 2004). Using the full history OSM database, as proposed by Barron *et al.* (2013), could further enhance the version number criterion with the number of different contributors involved in the object improvement. Moreover, the peaks example is the ideal case for introducing what Goodchild and Li (2012) call the geographic approach of VGI quality assurance. If we could cross the peaks with digital terrain model information, it would be easy to assess the proximity of the feature and a peak in the relief as modelled in the DTM.

6.2 Does Feature Density Alter Scale Level?

Inferring the LoD of objects individually is sometimes insufficient, as the legibility of the object can be affected by the local density of objects. It is plausible to assume that certain dense patterns alter the LoD/scale of its composite object, potentially raising partonomic questions. Figure 8 shows the example of a cross-country ski resort in the Pyrenees. Individually, the paths are approximately inferred as 1:100k scale objects by the improved multiple criteria method. However,

the density of such paths prevents any legible display at this scale (Figure 8c). In this case, the scale inferred is not the scale at which the paths could be mapped, but the scale at which they could be individually mapped.

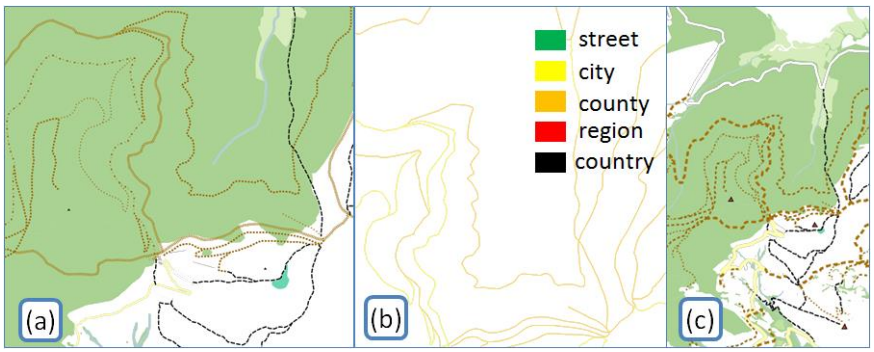

Figure 8. (a) Multiple paths in a cross-country ski resort in the Pyrenees. (b) Individual scale inference tells that paths should be mapped at 1:100k, but (c) there are too many close paths at this scale.

Regarding the multiple criteria method, two kinds of criterion could be added to integrate such density issues. First, clutter measures (Rosenholtz *et al.* 2007) at several scales could be made to assess at which scales the density of objects from one feature type causes clutter problems. Then, the principles from the coalescence criterion could be extended to the neighborhood of objects: as symbol size increases when scale decreases, when does the object symbol overlap with symbols from its neighbors?

6.3 The Scale of Objects with Simple Shapes

In a dataset with a homogeneous scale, simple objects like rectangular-shaped buildings are captured with simple geometries, but we know the scale is similar to all the other buildings that may have complex shapes. In OpenStreetMap, if an object has a very simple shape, it is hard to know if its scale is small, or if it is just a simple geometry that has been accurately captured. The scale equivalency measure will infer that such simple objects have a small scale, as the number of vertices to capture them is small (Figure 9). Within the empirical scale equivalency approach one must come up with coping strategies for individual object types separately, as was done for city blocks (Reimer *et al.* 2014). LoD inference too, provides quite different results for buildings depending on the complexity of their original shape (Touya and Brando 2013).

Figure 9. Two buildings captured from the source (French cadastre) by the same contributor with quite different scale equivalency because the upper one has a much simpler shape.

When the application is the production of small-scale maps with a generalization process, this issue will not alter the generalization result. Indeed, if the inferred scale is small, the impact will only be a slight generalization, but if the shape is simple, it does not require as much generalization as a building with a complex shape. However, this can be an issue when the application is the production of a large-scale map, with LoD consistency checking (see Section 5). Then, objects with simple shapes should be identified before consistency checking, for instance, by analysis of the neighbors of the objects: if only one has a low resolution, the shape is probably simple. This analysis can be related to the geographic approach of quality of Goodchild and Li (2012).

7 Conclusions

Until the widespread availability and success of VGI and OSM specifically, scale was an easily determined metadatum. With non-government grassroots geodata becoming more encompassing by the day, users are confronted with rich but scale-wise inhomogeneous data. This is as true for geometries as it is on a conceptual level. This chapter reviewed the first two attempts at automatized scale inference for OSM and showed their relative strengths and weaknesses and ways to improve both by combining them. Both approaches currently and purposefully work on an individual object level. The open problems such as point features and feature density could plausibly be approached with meso-regions as intermediate steps, as it has been done in agent-based automated generalization. A more fundamental problem seems to be the further automation of parameterization. The

current approaches make extensive use of tacit and explicit cartographic knowledge in selecting the measures and/or criteria. It is unclear at the moment how or if automation of these fundamental steps is attainable at all in the near future.

References

Atkinson, P. M and Tate, N. J. (2000). Spatial Scale Problems and Geostatistical Solutions: A Review. Professional Geographer, 52(4), pp. 607-623.

Barron, C., P. Neis, and A. Zipf (2013). Towards intrinsic quality analysis of OpenStreetMap datasets. In Online proceedings of the International Workshop on Action and Interaction in Volunteered Geographic Information (ACTIVITY). AGILE.

Bereuter, P., and Weibel, R., (2013). Real-time Generalization of Point Data in Mobile and Web Mapping Using Quadtrees. Cartographic and Geographic Information Science, 40(4), 271-281

Biljecki, F., H. Ledoux, J. Stoter, and J. Zhao (2014). Formalisation of the level of detail in 3D city modelling. Computers, Environment and Urban Systems 48, 1-15.

Burt, J, E. and Barber, G. M. (1996). Elementary Statistics for Geographers, second edition. The Guilford Press.

Chave, J. (2013). The problem of pattern and scale in ecology: what have we learned in 20 years?. Ecology Letters 16: pp.4-16.

Clark, J.H. (1976). Hierarchical Geometric Models for Visible Surface Algorithms. Communications of the ACM Vol. 19 Nr. 10, pp. 547-554.

Dutton, G. (1999). Scale, sinuosity and point selection in digital line generalization. Cartography and Geographic Information Science . 26(1), 33-53.

Feick, R. and Robertson, C (2014). A multi-scale approach to exploring urban places in geotagged photographs. In: Computers, Environment and Urban Systems, in press.

Figueira, J., S. Greco, and M. Ehrogott (Eds.) (2005). Multiple Criteria Decision Analysis: State of the Art Surveys, Volume 78 of International Series in Operations Research and Management Science. Springer.

Figueira, J., V. Mousseau, and B. Roy (2005). ELECTRE methods. In J. Figueira, S. Greco, and M. Ehrogott (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, pp. 133-162. Springer.

Girres, J.-F. (2011). An evaluation of the impact of cartographic generalisation on length measurement computed from linear vector databases. In Proceed-

ings of 25th International Cartographic Conference (ICC'11), Paris, France. ICA.

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

Goodchild, M. F. and L. Li (2012). Assuring the quality of volunteered geographic information. Spatial Statistics 1, 110-120.

Grosso, E., J. Perret, and M. Brasebin (2012). GEOXYGENE: an interoperable platform for geographical application development. In Innovative Software Development in Gis, Chapter 3, pp. 67-90. John Wiley and Sons.

Haggett, P., Chorley, R. J., Stoddart, D. R. (1965). Scale Standards in Geographical Research: A new Measure of Areal Magnitude. In: Nature, Vol. 205. pp. 844-847.

Haklay, M. (2010). How good is volunteered geographical information? a comparative study of OpenStreetMap and ordnance survey datasets. Environment and Planning B: Planning and Design 37 (4), 682-703.

Haklay, M., S. Basiouka, V. Antoniou, and A. Ather (2010). How many volunteers does it take to map an area well? the validity of linus's law to volunteered geographic information. The Cartographic Journal 47 (4), 315-322.

Jokar Arsanjani, J. (2014). Case Study I: VGI Platforms and Data Generalisation: In Sester *et al.* 2014

Keßler, C. and R. T. de Groot (2013). Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. In D. Vandenbroucke, B. Bucher, and J. Crompvoets (Eds.), Geographic Information Science at the Heart of Europe, Lecture Notes in Geoinformation and Cartography, pp. 21-37. Springer International Publishing.

Kolbe, T. H. (2009). Representing and exchanging 3D city models with CityGML. In J. Lee and S. Zlatanova (Eds.), 3D Geo-Information Sciences, Lecture Notes in Geoinformation and Cartography, pp. 15-31. Springer Berlin Heidelberg.

Levin, S. A. (1992). The problem of Pattern and Scale in Ecology. Ecology, 73(6), pp. 1943-1967.

Mondzech, J. and M. Sester (2011). Quality analysis of OpenStreetMap data based on application needs. Cartographica: The International Journal for Geographic Information and Geovisualization 46 (2), 115-125.

Mooney, P. and P. Corcoran (2012). The annotation process in OpenStreetMap. Transactions in GIS 16 (4), 561-579.

Reimer, A., Kempf, C., Rylov, M. and Neis, P. (2014) Assigning Scale Equivalencies to OpenStreetMap Polygons. In: AutoCarto, International Symposium on Automated Cartography 2014, Proceedings (accepted), 2014.

Rosenholtz, R., Y. Li, and L. Nakano (2007). Measuring visual clutter. Journal of Vision 7 (2).

Sester, M, Jokar Arsanjani, J, Klammer, R., Burghardt, D. and Haunert, J-H (2014). Integrating and Generalising Volunteered Geographic Information. In: Burghardt, D., Duchêne, C. and Mackaness, W. (Eds.) (2014). Abstracting Geographic Data in a Data Rich World.

Skarlatidou, A., M. Haklay, and T. Cheng (2011). Trust in web GIS: the role of the trustee attributes in the design of trustworthy web GIS applications. International Journal of Geographical Information Science 25 (12), 1913-1930.

Steinhardt, U. (1999) Die Theorie der geographischen Dimensionen in der Angewandten Landschaftsökologie. In: Schneider-Sliwa *et al.* (Eds.) (1999) Angewandte Landschaftsökologie. Springer.

Stoter, J., D. Burghardt, C. Duchêne, B. Baella, N. Bakker, C. Blok, M. Pla, N. Regnauld, G. Touya, and S. Schmid (2009). Methodology for evaluating automated map generalization in commercial software. Computers, Environment and Urban Systems 33 (5), 311-324.

Sudgen, D and Hamilton, P. (1971) Scale, systems and regional geography. Area, Vol. 3, No. 3, pp. 139-144

Surowiecki, J. (2004). The Wisdom of Crowds. Anchor.

Töpfer, F and Pillewizer, W (1966) The principle of selection. The Cartographic Journal, Vol. 3, pp. 10-16.

Töpfer, F. (1979) Kartographische Generalisierung, 2nd Edition. VEB Hermann Haack, Gotha.

Touya, G. (2012). What is the level of detail of OpenStreetMap? In Workshop on Role of Volunteered Geographic Information in Advancing Science: Quality and Credibility.

Touya, G. and C. Brando-Escobar (2013). Detecting Level-of-Detail inconsistencies in volunteered geographic information data sets. Cartographica: The International Journal for Geographic Information and Geovisualization 48 (2), 134-143.

Touya, G. and M. Baley (2014). Harmonizing Level of Details in OpenStreetMap Based Maps. Submitted to GIScience 2014, Vienna, Austria.

Zielstra, D. and A. Zipf (2010). A comparative study of proprietary geodata and volunteered geographic information for germany. In Proceedings of 13th AGILE International Conference on Geographic Information Science, Guimaraes, Portugal.