Scale-Dependent Point Selection Methods for Web Maps

Mathias Gröbe[1] · Dirk Burghardt[1]

Abstract

In cartographic generalization, the selection is an often-used method to adjust information density in a map. This paper deals with methods for selecting point features for a specific scale with numerical attributes, such as population, elevation, or visitors. With the Label Grid approach and the method of Functional Importance, two existing approaches are described, which have not been published in the scientific literature so far. They are explained and illustrated in the method chapter for better understanding. Furthermore, a new approach based on the Discrete Isolation measure is introduced. It combines the spatial position and the attribute's value and is defined as the minimum distance to the nearest point with a higher value. All described selection methods are implemented and made available as Plugins named "Point selection algorithms" for QGIS. Based on this implementation, the three methods are compared regarding runtime, parameterization, legibility, and generalization degree. Finally, recommendations are given on which data and use cases the approaches are suitable. We see digital maps with multiple scales as the main application of those methods. The possibilities of labeling the selected points are not considered within the scope of this work.

Keywords Generalization · Selection · Analysis · Web mapping · VGI · LBSM

Zusammenfassung

In der kartografischen Generalisierung ist die Auswahl eine häufig verwendete Methode zur Anpassung der Informationsdichte in einer Karte. Dieser Beitrag befasst sich mit Methoden zur Auswahl von Punktmerkmalen für einen konkreten Maßstab mit numerischen Attributen, wie z. B. Bevölkerung, Höhenwert oder Besucherzahl. Mit dem Label-Grid-Ansatz und der Methode der funktionellen Bedeutung werden zwei bestehende Ansätze beschrieben, die bisher nicht in der wissenschaftlichen Literatur veröffentlicht wurden. Sie werden zum besseren Verständnis im Methodenkapitel erläutert und illustriert. Weiterhin wird ein neuer Ansatz vorgestellt, der auf der Dominanz basiert. Er kombiniert die räumliche Position und den Wert des Attributs und ist definiert als der minimale Abstand zum nächstgelegenen Punkt mit einem höheren Wert. Alle beschriebenen Auswahlmethoden werden implementiert und als Plug-in „Point selection algorithms" für QGIS zur Verfügung gestellt. Basierend auf dieser Implementierung werden die drei Methoden hinsichtlich Laufzeit, Parametrisierung sowie Lesbarkeit und Generalisierungsergebnis verglichen. Abschließend werden Empfehlungen gegeben, für welche Daten und Anwendungsfälle die Ansätze geeignet sind. Als Hauptanwendungsfall für diese Methoden sehen wir digitale Karten mit mehreren Maßstäben. Die Möglichkeiten der Beschriftung der ausgewählten Punkte wird im Rahmen dieser Arbeit nicht betrachtet.

Schlüsselwörter Generalisierung · Auswahl · Analyse · Webkarten · VGI · LBSM

1 Introduction

Selection is an often utilized method in the generalization process when decreasing the scale. It entails the decision to remove or maintain objects depending on their spatial distribution and according to their attribute values. In the terminology of the fundamental generalization operator,

✉ Mathias Gröbe
mathias.groebe@tu-dresden.de

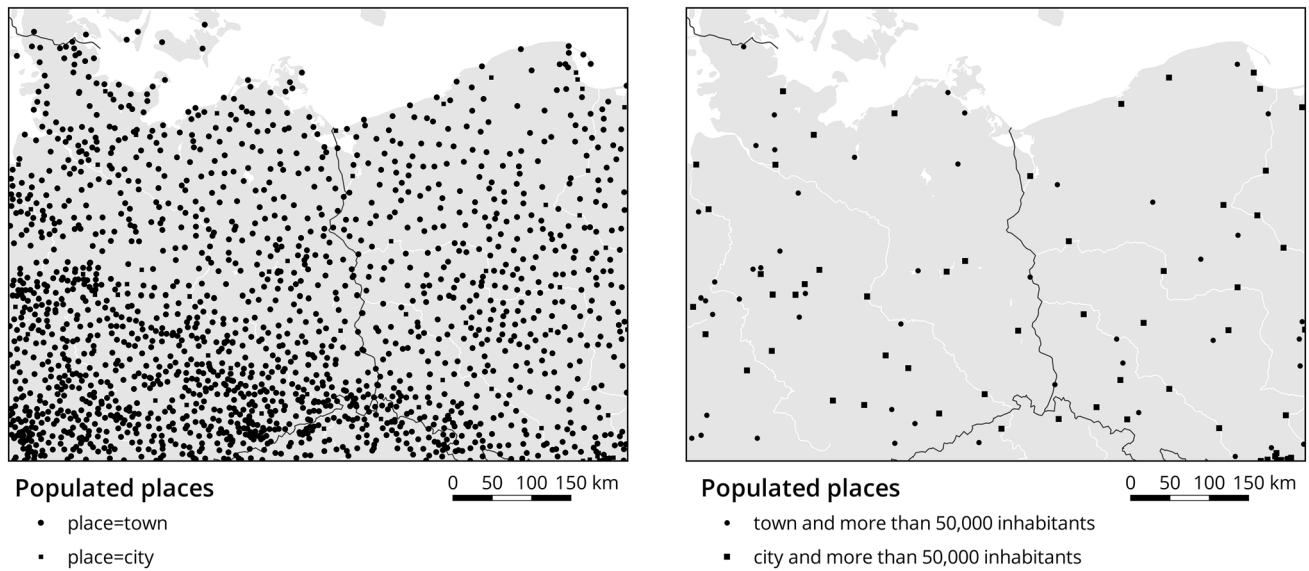[1] Institute of Cartography, TU Dresden, Dresden, Germany

Fig. 1 Example for populated places in Germany and Poland. The left map shows all settlements tagged as city or town, while the right map visualizes a selection according to the population number (Data OpenStreetMap Contributors 2021 OdbL 1.0)

the term "selection" (Kraak and Ormeling 2010; Hake et al. 2002) used here is also described as "refinement" (Slocum et al. 2009, p. 102; Stanislawski et al. 2014). The overall process aims to visualize spatial information in an easy-to-understand manner on a sheet of paper or screen. For several years, maps have been part of everyday life, especially for smartphone users: exploring the current area, finding the next bus stop or something to eat are possible tasks that can be addressed with a map. A multi-scale map app on a smartphone can present search results, and such location-based thematic information is often available as point data, e.g., restaurants, supermarkets, and tourist spots. For an excellent visualization, it is necessary to select the most important objects for the user and avoid cluttering (Ellis and Dix 2007). A selection of a category such as a restaurant's cuisine can satisfy this requirement, but often either not enough or indeed too many objects are removed. A selection according to a numerical attribute, e.g., a place's population or a restaurant's rating, is a known feasible concept that would lead to an improved result. If globally defined class borders were applied, the result would probably be disappointing because this approach could not consider the individual importance and the local spatial distribution.

How misleading a selection is that only considers the numbers and not space is shown in Fig. 1. There is a greater density of settlements in the southwest, while fewer settlements in the northeast. A selection considering only the population number leads to sparse areas on the right map. As a result, regions exist where the distance to the next shown place is greater than 100 km. Here, more places should be shown because of their local importance as a central place

for a region. For instance, they could be essential for orientation when driving through the country, necessary for the economy, or relevant for sightseeing. Nevertheless, merely the population can still be a good selection criterion in combination with space: in sparsely populated regions, a town with a smaller population is more important than in an urban area with a higher population. An absolute number does not provide appropriate results; a relative measure would be a better choice. Thus, a city's population size must be seen in relation to surrounding cities, population size, and distance.

A good solution should identify the locally most essential place (local maximum), considering the space and the chosen numerical attribute. It should be possible to integrate the approach in the production workflow of a multi-scale map. The selection algorithm might also be valuable for finding relative important points in the data as analysis tasks, e.g., points of interest (POI), recommendation platforms, and location-based social media (LBSM). A particular information request might provide many POIs within the city center with a high data density enabling the need for a clutter-free visualization.

An essential factor for the selection method in the context of multi-scale maps is the performance. Maps need to be updated, and zoomable maps usually have worldwide coverage. Updates can be done within minutes, just like the main map of OpenStreetMap demonstrates. A complex approach, e.g., using artificial intelligence (Karsznia and Sielicka 2020) with many factors, 33 ones in the example, can deliver good results but is hard to transfer on a worldwide scale, needs much preparation, and is complex to process.

1.1 Research Questions and Outline

The paper aims to describe current selection approaches, make them available to a broader audience by providing implementation, and compare the properties as well as the results by answering the following research first question:

How to select point objects with numerical attributes for visualization depending on the spatial distribution and the scale or zoom level?

In the theoretical part of the paper, existing selection methods will be described. Afterward, a detailed introduction of possibly suitable algorithms for analysis and multi-scale maps will be made. For each described method, a Python implementation is carried out and made available as a plugin for QGIS. That makes the algorithms useable for any user of the open-source GIS, and the result of the following use case in visualization and analysis reproducible. We summarize our experience with the different methods to answer the second research question:

Which selection method offers the best performance, and which selection method is suitable for multi-scale maps?

We will only consider the selection of points in the context of multi-scale maps and not the consequential labeling challenge concerning label placement as well as the length and size of the names. Based on our implementation, we compare the complexity of the selection procedures and transfer it into a web mapping working environments, such as PostgreSQL and its spatial extension PostGIS.

2 Related Work

2.1 The Traditional Selection of Points

2.1.1 Radical Law

The "Radical Law" and the derived "Principle of Selection" (Töpfer and Pillewizer 1966; Töpfer 1974) was one of the first attempts of automated methods in the context of generalization for selection or evaluation of generalization results. However, it does not offer help when selecting points, but it predicates the number of objects in the following scale and helps to identify a suitable number of selections. After all, the selection is conducted depending on the cartographer's skills and does not consider variations in map objects' density (Sarjakoski 2007). The context of multi-scale maps offers a test environment to evaluate the different approaches and check if they remain consistent over several zoom levels.

2.1.2 Selection Criteria

In former days, the selection was a matter of the knowledge and experience of the cartographer driven by recommendations, e.g., Arnberger proposes grading settlements according to the number of inhabitants, function, and importance (Arnberger 1977). A practical textbook for cartographers (Laubert et al. 1988, p. 90) recommends the following criteria: the character of the area, settlement density, area, importance, and settlement type. There are approaches to reproduce such a complex selection process by machine learning (Karsznia and Sielicka 2020; Karsznia and Weibel 2018). The authors explore former day's cartographer's decision criteria with these methods and reach similar results. For achieving this level, a high number of criteria, such as administrative, cultural, or economic factors, were collected and used.

2.2 Algorithmic Selection

There are early attempts for selection approaches existing considering the frequency and distribution of points (Srnka 1970) and different measures for selecting settlements by the population, a radius derived from population, and utilizing an R-tree (Shea 1988, p. 50). Another short overview of known cartographic selection methods is provided by Li, who also explains some algorithms similar to Sheas report (Li 2007, pp. 81–84). In particular, the circle growth algorithm (Kreveld et al. 1997) and settlement-spacing ratio algorithm (Langran 1986) are explained and illustrated. No open-source multi-scale web map style uses one of those algorithms for selection yet. The Label Grid approach is widely used. It was first implemented by Mapbox engineers and made available on GitHub.[1] Application examples can be found at the following mapping platforms: Mapbox, TopPlusOpen[2] (Kunz 2018; Kunz and Bobrich 2019) and OpenMapTiles.[3] Through a presentation, we became aware of the Functional Importance method and found a website[4] about it. It seems that there is some knowledge about selection approaches, which are rarely known and not published.

---

[1] https://github.com/mapbox/postgis-vt-util (accessed March 2021).

[2] https://www.bkg.bund.de/SharedDocs/Produktinformationen/BKG/EN/P-2018/180608-TopPlusOpen.html (accessed March 2021).

[3] https://github.com/openmaptiles/openmaptiles-tools/blob/master/sql/LabelGrid.sql (accessed March 2021).

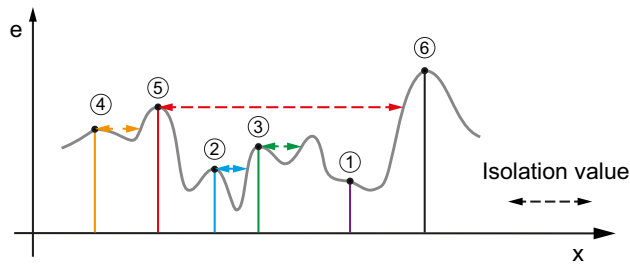[4] http://imagico.de/map/osm_populated_en.php (accessed March 2021).

Fig. 2 Principle of the topographic isolation: the grey line forms the relief, and the summits 1–6 are peaks with an elevation (e), which is used to calculate the isolation distance to the nearest point on the relief with a higher elevation

Fig. 3 Schematic profile for the Label Grid approach: selection selection of the point with the highest p-value in each grid cell. For example, p can be the population, while x describes the space

2.2.1 Selection Using Topographic Isolation

A special case in the context of algorithmic selection is the topographic isolation: it is a measure used in Geography for classifying peaks but also suitable for selections in the context of generalization. It is defined as the distance from a peak to the nearest point with the same or higher elevation, which does not need to be a peak itself. This method is in use at the OpenTopoMap[5] (OpenTopoMap/users: Germany/ OpenStreetMap Forum 2020) for the selection of peaks. As a result, at lower zoom levels, peaks with higher topographic isolation appear first. When zooming in, more peaks appear because the peaks with lower isolation are shown. Figure 2 shows the principle: the topographic isolation is the minimum horizontal distance to a higher elevation on the relief formed by the grey line with the value (e). Summit (1) has a very low or no isolation, while the isolation for the summit (6) is infinite because there is no higher point in scope.

3 Methods

In the following sections, three selection approaches are described and illustrated, each with a schematic profile and a map example. The data comes from OpenStreetMap using populated places tagged as "place = town" or" place = city" as well as the population number in the same area as the example in the introduction. Our implementation as a QGIS Plugin "Point selection algorithms"[6] allowed us to compare the three methods and made them available for further usage.

3.1 Label Grid

The first method introduced here is based on a grid approach: for an area defined by a grid cell, the points inside are ranked by a numerical attribute in descending order. This solution is feasible for all points with a numerical value. The grid width can be adjusted to the density of points on the map. It is implemented as an SQL query for PostgreSQL/ PostGIS from Mapbox.[7] The complexity of the approach is $O(n * \log n)$. It depends mainly on the sorting algorithm after the intersection of points with the grid or polygon.

The reference function from Mapbox is based on squares, but it is also possible to work with diamonds, hexagons, or any other polygon in our implementation. The result is a ranking by the point's values in the specific grid cell or polygon.

In the first step, the method requires the points to be assigned to a grid cell. Second, the ranking of the points in each grid cell follows. The result can be stored as an integer attribute with the point. Figure 3 shows an example of the Label Grid method applied to populated places, whereby only the place with the highest population in each grid cell (normally invisible) with a height and width of 156 km is shown. In the case of a typical web map, this would be the size of one tile in zoom level 8, a scale of 1:2,000,000 at the equator. The result is mainly influenced by the grid's origin and the cell size, as illustrated in Fig. 4. Moving the grid will lead to different results, so it is essential to use always the same grid.

3.2 Functional Importance

A bell-shaped curve is applied for the Functional Importance method, lowering values by distance and modeling each point's local importance (see Fig. 5). The difference

---

[5] https://opentopomap.org/.

[6] An installation is directly from the QGIS extension manager possible. More information can be found on this page: https://plugins.qgis. org/plugins/point_selection/.
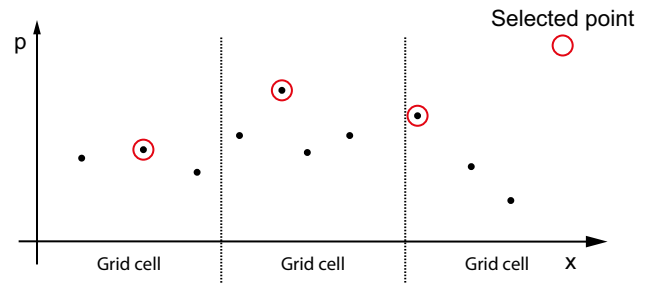
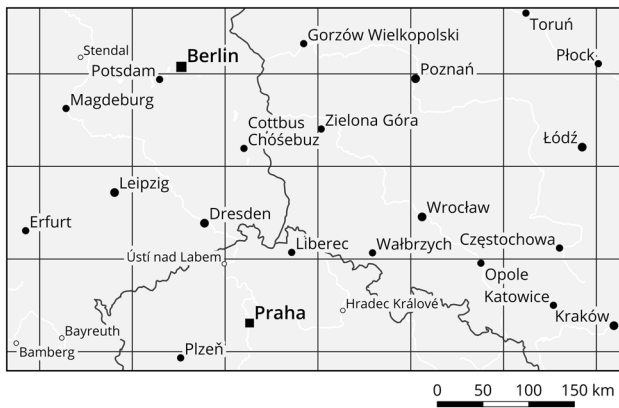[7] https://github.com/mapbox/postgis-vt-util (accessed March 2021).

Fig. 4 Example map for the Label Grid with data from NaturalEarth and OpenStreetMap contributors 2020; only the place with the highest population in each grid cell is shown
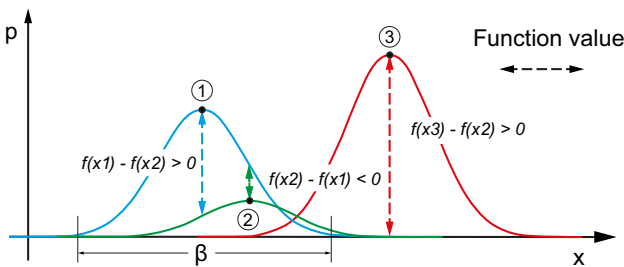

Fig. 5 Visualization of the Functional Importance method. Each place gets its function formed by the parameter population (*p*) and *β*. In case of a positive difference between function values, the place should be shown on the map. In this example, points (1) and (3) would be selected. The *x*-axis *p* shows the spatial distance

between the function value of different points can be stored and used to select the points. This way of modeling interactions between objects is not new but less in usage in the field of cartography (Horman 2021). The following formula is adapted from a description on a website (Horman 2020; imagico.de - Populated places in OpenStreetMap 2021):

$$f(d) = p * e^{-\frac{d^2}{\beta}}$$

where *d* is the distance between the point geometries, *p* is the numerical attribute value, e.g., population, and *β* is the variable to adjust how quickly the distance value is lowering and keep distance between near points.

Figure 5 illustrates the formula above and shows the influence of the parameters. By adjusting the variable *β,* it is possible to control the information load and the minimal distance between points to be shown on the map. The population value has an influence on the selection or non-selection of the point because the difference between the

function values must be greater than zero. For point (2), is the value too low to get selected. It would also be possible to use another function to model the Functional Importance—using more or fewer parameters.

For determining the highest Functional Importance value, it needs an outer loop going through all points, completed by an inner loop that calculates the distance and function values. The result is a quadratic complexity of $O(n^2)$.

Figure 6 shows an example using the Functional Importance for cities with a *β*-value of 78 km, which is the radius of the function where it nearly reaches zero and is nearly comparable with a grid cell in zoom level 8 with a size of 156 km in the web Mercator projection.

3.3 Discrete Isolation

For our approach, we transfer the principle of topographic isolation to discrete points. Usually, the isolation is the distance to the nearest higher point from a peak, often a point on a slope and not a peak. In our definition, we consider only the peaks. Therefore, there is no continuous surface, just discrete values with a known distance in between. The Discrete Isolation is the point's distance to the closest point with a higher value, as Fig. 7 shows. For computing this value, it is necessary to calculate the distance to all other points with a higher *p* value and order the results to get the point with the lowest distance and higher value, which leads to a quadratic complexity of $O(n^2)$. Figure 8 shows how this is implemented in the QGIS-Plugin. After all, the points (1)–(5) receive their Discrete Isolation distance, which is storable as an attribute. Point (6) is the highest point; thus, no distance to points with a higher value can be derived. For practical reasons, a default value can be applied, such as the earth's perimeter. An implementation of the Discrete Isolation has to calculate the distance from the point, e.g., (3), to every point with a higher value—(4), (5), (6)—and then save the lowest distance to a point with a higher value as the point (3).

The value for the Discrete Isolation is then used for the selection of the points. This procedure makes it possible to refine the selection as required. In the GIS, a graduated renderer can be used in combination with various classification methods.

As a result, selecting a lower or higher distance value for the isolation increases or decreases the number of points shown on the map. The map in Fig. 9 visualizes a selection of cities determined by the isolation method. Only places with a distance greater than 78 km to the next place with a higher population number are shown. The parameter value is the radius of a circle with a diameter of 156 km which is comparable with the tile size in zoom level 8.
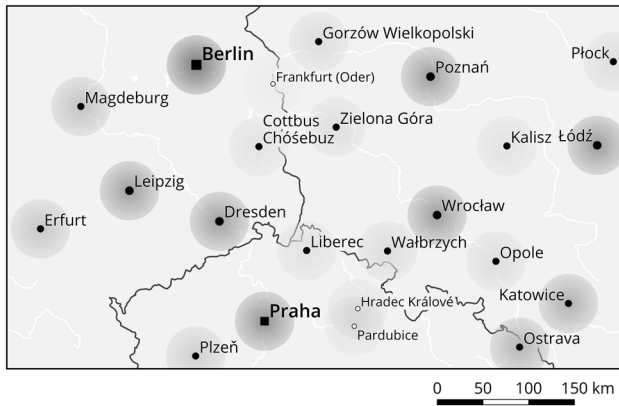
Fig. 6 Map example for the Functional Importance with data from NaturalEarth and OpenStreetMap contributors 2020. The brightness of the circles represents the Functional Importance based on population values
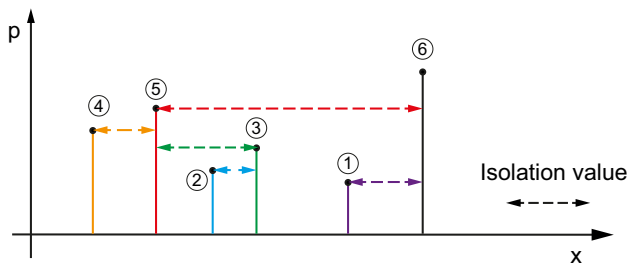

Fig. 7 Visualization of computation of the Discrete Isolation value. The points with the number 1–6 receive different isolation values because of the p value. The Discrete Isolation is the shortest distance to a point with a higher value for (p)

4 Evaluation

4.1 Runtime Comparison

For evaluating the runtime and complexity of the selection methods, we run tests with the default settings of the QGIS Plugins. Randomly distributed points were generated inside the bounding box. We started with the number of 100 points and increased the number to 2000 points with 500 and 1000 points as steps in between. A randomly numerical attribute value between 1 and 6000 was created for the test. For each number of points, the selection tool was applied six times to get average runtime values. For running the test a Microsoft Surface was utilized with 16 GB RAM, SSD hard drive, Intel Core i7 8650U, and QGIS 3.16.3.

Figure 10 and Table 1 show our test results: for 100 points, the computing time is almost identical. The fivefold increase in the number of points provides a clear differentiation of the runtime and the variance of the computing time

becomes visible. For the number of 2000 points, the runtime differs significantly between 5 s for the Label Grid, while the Discrete Isolation approach needs 80 s and 130 s for the Functional Importance. Polynomial curves were fitted into the visualization for estimating the runtime depending on the number of points. According to the result, the Label Grid approach seems to have a linear or logarithmical complexity. In contrast, the Discrete Isolation and Functional Importance have quadratic complexity depending on the number of points.

The reason for the high complexity of the Functional Importance and the Discrete Isolation is the increasing number of distances, which have to be calculated with an increasing number of points. For the Discrete Isolation, the number of calculated distances is decreased by considering only points with a higher attribute value. The Label Grid approach depends more on the number of grid cells, and in the example, the number of grid cells stays the same in our experiment. Increasing the number of grid cells would also increase the Label Grid approach's computation time; this would be necessary if the selection result is not satisfying. In the case of Discrete Isolation, this would not be needed, while for the Functional Importance, an adjustment of the parameters can be helpful for a better result.

4.2 Use Cases for the Discrete Isolation

The following section demonstrates the possibilities of the Discrete Isolation in the visualization tasks for a Cartographer or Geovisual Analyst. The data from OpenStreetMap are freely available; thus, the examples should be reproducible. In the case of the location-based social media data, we use already collected data from Twitter.
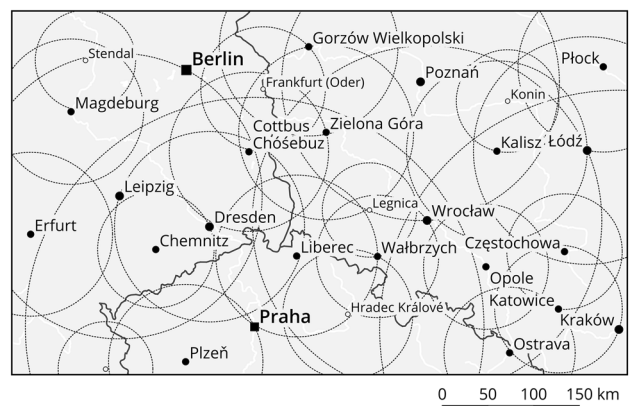

Fig. 8 Example map for the Discrete Isolation with data from NaturalEarth and OpenStreetMap contributors 2020. The circles show the distance to the nearest city with a higher population

```python
points = (...) # Set of points with numerical attribute
max_isolation = 100000

# Go over all points
for point1 in points:

    # If attribute not null
    if point1.attribute:
        point1.isolation = max_isolation
        # Loop over all points again for comparing the attribute value
        for points2 in points:
            # Compare attribute value
            if point2.attribute > point1.attribute:
                # If attribute value great calculate distance
                isolation = distance(point1, point2)
                # Compare with already calculated isolation, if smaller store it
                if isolation < point1.isolation:
                    point1.isolation = isolation

    # If no attribute aviable, then isolation have to be null
    else:
        point1.isolation = None
```

Fig. 9 Python pseudo-code to implement the Discrete Isolation similar to the implementation in our QGIS plugin

4.2.1 Selection of Peaks in a Multi-scale Environment

One possible use case of Discrete Isolation is selecting peaks similar to the topographic isolation and the example Open-TopoMap. As Fig. 11 shows, there are many peaks in OpenStreetMap. For the chosen scale, a selection is needed and performed by visualizing only peaks with Discrete Isolation higher than 3000 in Fig. 11a. It avoids clutter but does not resolve all labeling problems with a fixed label position; some black triangles are not labeled. For the demonstration of the method in a multi-scale environment from each map in the figure, the scale and the selection distance were multiplied by 1.5. This approach is intended to clarify the development of the selection while keeping the map load as constant as possible. For a consistent result, when reading the maps from (f) to (a), the respective mountains within the displayed map section must be visible, which is fulfilled in every case.

The approach has the advantage that no digital elevation model is needed for the calculation of isolation. That makes implementation straightforward and reduces the effort to chain more tools together and collect relevant data to calculate topographic isolation. The implementation is only importing the OpenStreetMap data into a PostgreSQL/PostGIS database, cleaning the "ele" tag containing the elevation values. It calculates the Discrete Isolation with an SQL implementation Discrete Isolation.[8]

4.2.2 Selection of Settlements with Additional Criteria

Cities and towns offer the population number as a suitable attribute to apply the Discrete Isolation. An improved map compared to the introduction example in Fig. 1 will be presented and explained within this use case. Figure 12 shows two maps using populated places and their population from OpenStreetMap: the upper one using a topographic isolation selection with a minimum distance of 75 km. The result is that some nearby situated cities are hidden, such as Postdam next to Berlin or Halle next to Leipzig. By ranking the importance, the selection is correct, but those cities are also significant but are in the shadow of bigger towns. Overall, the distribution of selected places is uniform.

The places are not evenly distributed; the settlements represent the population density discretely. A map should reproduce that in clutter-free visualization. The lower map in Fig. 12 shows an approach by combining the Discrete Isolation with the place type (town or city) from OpenStreetMap. A lower isolation distance is applied to cities, and nearby cities such as Halle or Potsdam get visible. The map is evenly filled with places but represents the population distribution better. A side effect of using Discrete Isolation for selection is that it is straightforward to avoid labeling problems. Choosing a suitable isolation distance can create space for labels. The selection of the isolation in these examples was made by trial and error.

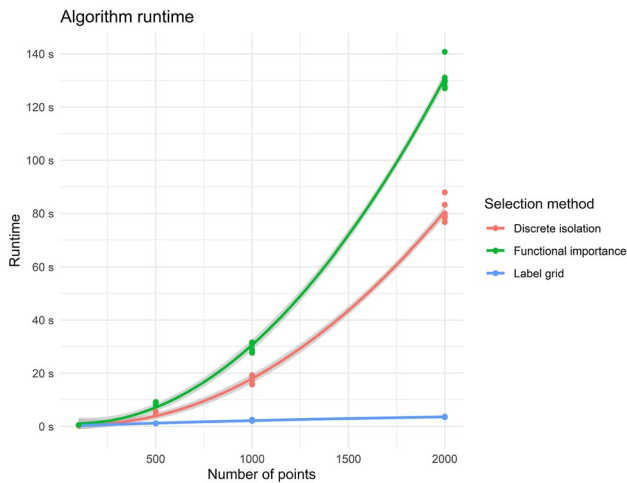[8] https://github.com/MathiasGroebe/discrete_isolation.

Fig. 10 Runtime of the tree different selection methods depending on the number of points. The result shows the complexity of the methods

Table 1 Mean runtime in seconds of the selection methods with the QGIS plugin in version 0.3, depending on the number of points in the input dataset

Number of points	Mean runtime with QGIS plugin version 0.3		
	Label grid (s)	Discrete isolation (s)	Functional importance (s)
100	0.38	0.34	0.48
500	1.14	5.32	8.57
1000	2.25	17.09	29.69
2000	3.57	80.88	131.00

4.2.3 Analyzing and Selecting Places from LBSM

The applications of Discrete Isolation are not limited to visualization. It can also be used for analysis purposes: the example shown here in Fig. 13 illustrates the method's potential in the context of Geovisual Analytics. For the determination of relevant locations for Twitter users, referenced Twitter Places[9] were counted, and the places should be visualized. The map in Fig. 13 shows a higher density of places in Dresden's city center than in rural areas. It is now possible to show the locally more important places through the Discrete Isolation, such as the tourist attractions "Frauchenkirche" or "Blaues Wunder" in Dresden's heavy crowded area. If the places would form a mountain range, and the tweet count is the elevation, the Discrete Isolation helps to find the locally highest points, which are more exciting places.

9 https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/place. Accessed April 2021.

The advantage of the approach is identifying hot spots in areas with high and low data density. For the example, we excluded points with a count of one tweet.

5 Discussion

Within the following chapter, we review our work with some example data and compare the presented methods according to several aspects. The Discrete Isolation will be discussed by the use cases and compared with the topographic isolation.

5.1 Evaluation of the Discrete Isolation

The first example for using the Discrete Isolation for selecting peaks shows that this measure can be used similarly to the topographic isolation. Using Discrete Isolation is less complicated because no digital elevation model is needed. In a multi-scale environment, it is possible to reach a similar map load and make a consistent selection where the essential points stay the same over a set of scales. A future question is how a selection distance can be transferred from one scale or zoom level to the next. The current proportional solution offers potential for improvements.

Settlement selection is another example, which shows the flexibility of the Discrete Isolation. Increasing the value for selection can simplify the labeling of places. A disadvantage can be the very unified distribution, but this can be improved using more knowledge, as shown in the example with the categories "city" and "town". The usage of the place categories from OpenStreetMap impressively shows this and the advantage that the Discrete Isolation is combinable with more attributes. A comparison with the same region from the introduction shows the great advantage of the method.

Besides, the Discrete Isolation is also suitable for analysis proposes: it is possible to identify local hot spots in the dataset and get a better overview in regions with a high data density. A further advantage is that it is also possible to consider regions with fewer data. The result is useable for a Geovisual Analysis and can help to achieve a deeper understanding of the data and underlying phenomena.

5.2 Comparison of the Approaches

Table 2 offers an overview to compare the methods according to three main aspects: performance, efficiency, and usability in web mapping. It summarizes the explanations from the methods chapter and some experiences from working with the approaches.

The test of the efficiency and performance with our QGIS plugin has shown that the Label Grid approach has the lowest complexity. The python implementation as a PostGIS
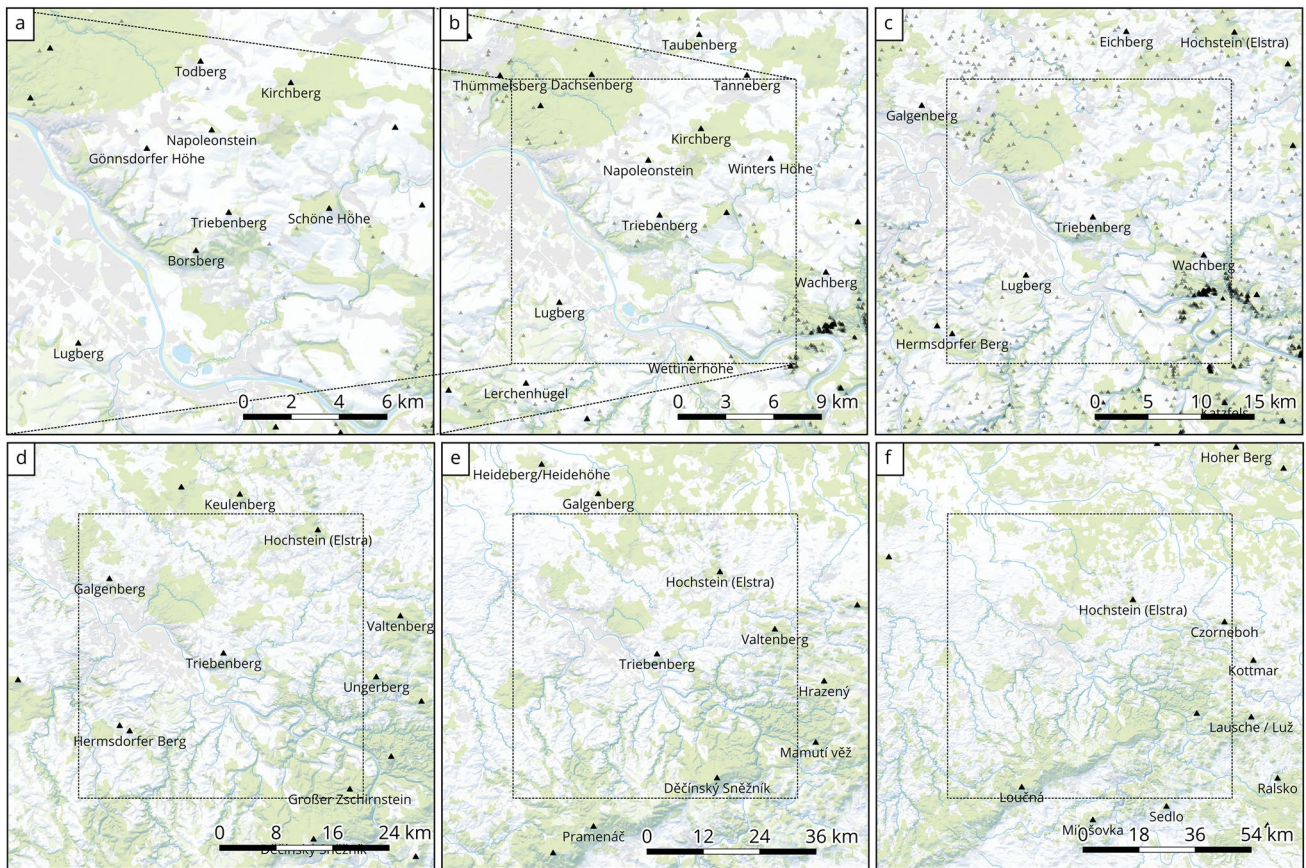
Fig. 11 Selection of peaks by the Discrete Isolation; all peaks are shown in the first row as small grey triangles; the selected peaks are black triangles and, if possible, labeled. The scale and the selection distance are multiplied by 1.5, for example to example. The dashed lines visualize the extent of the previous map. **a** Selection of peaks with a distance to the nearest higher peak than 3000 m at 1:500,000, **b** selection using a distance of 4500 m and a scale of 1:750,000. (Data OpenStreetMap Contributors 2021 OdbL 1.0, DEM GeoSN, dl-de/by-2-0 and EU DEM)

function can handle large point datasets very well. That is the reason for web mapping usage, but the result is binary: a selection or a non-selection. The Discrete Isolation performs slightly better than the Functional Importance in complexity, because the Discrete Isolation only considers points with higher values and not all Functional Importance points. The average runtime for 1000 points clearly shows the significantly higher complexity for those two approaches. In contrast to the Label Grid, a refinement of the selection is still possible because numerical values are calculated, which are used for the final selection.

Nevertheless, it needs to be considered that the Label Grid calculates a selection for a defined parameter, depending on the scale, and this is similar for the Functional Importance with its formula. From this, the necessity arises to recompute the selection for every scale or changed parameter again. In contrast, Discrete Isolation is calculated once, and only the value for showing points or not needs to be adjusted, which is a strong advantage of the method.

The Label Grid parameters are the shape type of the grid (free, square, hexagon, and diamonds) and the grid cell size. Besides, the Discrete Isolation offers the isolation distance as criteria, while the Functional Importance is very flexible: the formula can be changed completely; in our case, the radius β. That brings some flexibility into the selection and can be advantageous. The Label Grid's disadvantage is that the selection is difficult to combine with other attributes. For example, it is impossible to consider a place as an administrative center in the ranking function. Only the ranking and the function can be combined later.

If the points inside one grid cell are ranked, their spatial distribution cannot be considered by Label Grid and may cause labeling problems. In contrast, Functional Importance and Discrete Isolation with their value-based selection offer this possibility. In the Label Grid case, the shape and size of a grid depending on the map projection mainly influences the resulting selection; the other methods can be used
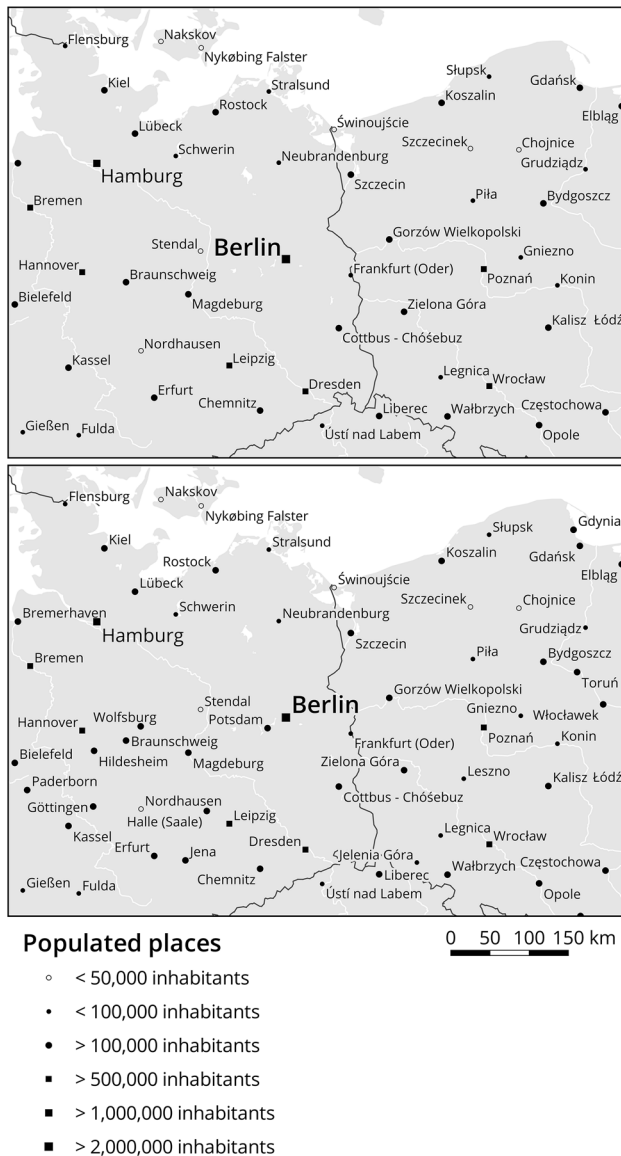
Populated places

- ◦ < 50,000 inhabitants
- · < 100,000 inhabitants
- • > 100,000 inhabitants
- ▪ > 500,000 inhabitants
- ◼ > 1,000,000 inhabitants
- ◼ > 2,000,000 inhabitants

Fig. 12 Attempts of making a better selection within the area of Fig. 1 using Discrete Isolation. The upper example uses the same Discrete Isolation of 75 km for places tagged as town or city. The lower example for cities' isolation distance is reduced to 30 km to better illustrate population density. (Data OpenStreetMap Contributors 2021 OdbL 1.0, NaturalEarth)

independently from the map projection utilizing ellipsoidal distances.

With our implementation for QGIS, the methods are accessible and can be used by everyone for free. The Label Grid was already implemented for PostgreSQL/PostGIS, and we also created an implementation for the Discrete Isolation.[10] The result of the three methods can be reviewed at

different web services; for the Label Grid, we only mention examples for which we know the usage. At the moment, only the Functional Importance seems to be in use by the originator and seems to be more or less known.

6 Conclusions

6.1 Selection of Points with Numerical Attributes

According to our knowledge, only the Label Grid approach for selecting points with a numerical attribute is in use in web mapping. We found another approach, which is called Functional Importance, and was implemented and described for comparison reasons. Besides, we present a new solution: the Discrete Isolation, which is a flexible selection measure derived from the principle of topographic isolation. There are more methods mentioned in the cartographic generalization literature, but it seems that they are not in use for current mapping applications. Table 3 offers an overview of the applications of the approaches in map production in combination with possible numerical attributes, which can be used to execute the selection algorithms.

6.2 Evaluation of Selection Methods in a Multi-scale Environment

For the Discrete Isolation, we try to demonstrate the potential of the method in some use cases. It is useable for visualization purposes such as selecting populated places, peaks, viewpoints, elevation points in the context of visualization or analysis purposes, as shown in the example with Twitter data. Comparing the algorithm's performances has shown that the Discrete Isolation is not as efficient as the currently used Label Grid but offers more flexibility and other advantages. The possibility of a flexible combination of different attributes for selection and the fact that only one computation of the Discrete Isolation is needed for different scales of zoom levels makes it handy. We offer implementations of the reviewed methods in the QGIS plugin and made them available through the built-in extension manager. It works well for small amounts of data, as our test has shown; for more than 1000 points, we recommend using the Label Grid tool only or the improved Discrete Isolation algorithm implementation for PostgreSQL/PostGIS. Otherwise, our tests have shown that the runtime of the algorithms increases rapidly. The implementation is already in use at our institutes' web map service for education, teaching, and the meinGrün project.[11] What kind of data are processed is

---

[11] https://meingruen.org/.

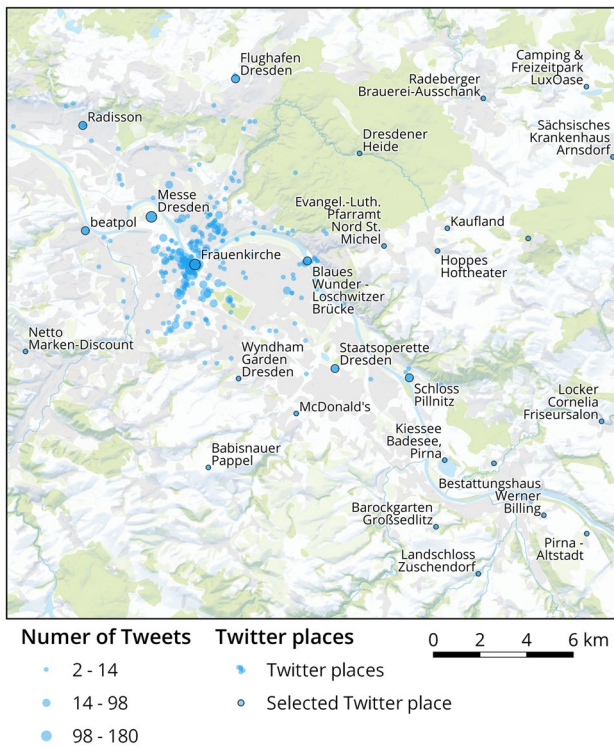[10] https://github.com/MathiasGroebe/discrete_isolation.

Fig. 13 Visualization and selection of places from Twitter by the Discrete Isolation. The places are abstracted to points, and the number of mentions in Tweets is counted as a numerical value and shown by the diameter of the points. Labeled and black outlined places have Discrete Isolation higher than 2 km (Data: OpenStreetMap Contributors 2021 OdbL 1.0, DEM GeoSN, dl-de/by-2-0, Twitter)

less critical for the choice of the selection method. Much more important is the volume of data and the desired result. The application of the Label Grid to mountains is not obvious but valuable if, for example, a uniform distribution is desired.

Overall, the Label Grid is the best choice if performance matters and can also be used with other regular or non-regular polygons. The Discrete Isolation is versatile and delivers good results but needs more computation time. Both methods work well in a multi-scale environment and seem to be suitable for efficient map production. More complicated is the Functional Importance—the results are good, but the computation time is long, and the model behind more complex. An application in multi-scale maps would require a long re-computation for each scale so that the method is unsuitable in this context.

Finally, now more GIS users can perform better selection due to the described and implemented methods. Possible improvements are a user study with the result of the different selection approaches and research about the transfer of the selection from one scale to another. Overall, there is still more work to be done in the field of generalization in the context of multi-scale maps. More research and implementation have to be carried out to produce better web maps to make them comparable with the good old manually made paper maps. Of course, possible solutions are already available, but they need to be improved to be applicable in the context of algorithmically produced worldwide maps.

Table 2 Comparison of selection methods Label Grid, Functional Importance, and Discrete Isolation

	Label grid	Functional importance	Discrete isolation
Complexity	$O(n * \log n)$	$O(n^2)$	$O(n^2)$
Average run time for 1000 points	2 s	29 s	17 s
Re-run with new parameter useful	Yes	No	Yes
Re-run for different zoom levels or scales	Yes	Yes	No
Parameters	Gird width, grid shape	Formula and their variables, function values	Isolation distance
Depending on the map projection	Yes (usage of a grid)	No	No
Combination with other attributes	No (only the final ranking)	Yes	Yes
Implementations	QGIS Plugin, PostgreSQL/PostGIS	QGIS Plugin	QGIS Plugin, PostgreSQL/ PostGIS
Usage	MapBox, TopPlusOpen, OpenMapTiles	Imagico.de	MeinGrün

Table 3 Possible application of the presented selection methods and suitable measures

Application	Settlements/places	Peaks	POIs
Possible numerical attributes	Population Area Number of visitors	Elevation Visible area	Number of visitors Rating Number of check-ins or mentions in LBSM

References

Arnberger E (1977) Thematische Kartographie: mit e. Kurzeinf. über Automation in d. themat. Kartographie, Westermann, Braunschweig, vol 231, p 1977

Ellis G, Dix A (2007) A taxonomy of clutter reduction for information visualisation. IEEE Trans vis Comput Graph 13:1216–1223. https://doi.org/10.1109/TVCG.2007.70535

Hake G, Grünreich D, Meng L (2002) Kartographie: Visualisierung raum-zeitlicher Informationen, 8., vollst. neu bearb. und erw. Aufl., De Gruyter, Berlin

Horman C (2020) Weniger ist mehr—zur Auswahl darzustellender Elemente in der digitalen Kartographie

Horman C (2021) Auswahl für Karten: funktionelle Bedeutung

imagico.de—populated places in OpenStreetMap: http://imagico.de/map/osm_populated_en.php. Accessed 22 Mar 2021

Karsznia I, Sielicka K (2020) When traditional selection fails: how to improve settlement selection for small-scale maps using machine learning. ISPRS Int J Geo-Inf 9:230. https://doi.org/10.3390/ijgi9040230

Karsznia I, Weibel R (2018) Improving settlement selection for small-scale maps using data enrichment and machine learning. Cartogr Geogr Inf Sci 45:111–127. https://doi.org/10.1080/15230406.2016.1274237

Kraak MJ, Ormeling F (2010) Cartography: visualization of geospatial data, 3rd edn. Prentice Hall, Harlow

Kreveld MV, Oostrum RV, Snoeyink J (1997) Efficient settlement selection for interactive display. In: Proceedings of Auto-Carto 13: ACSM/ASPRS annual convention technical papers, pp 287–296

Kunz P (2018) TopPlusOpen

Kunz P, Bobrich J (2019) Multi-scale cartographic visualization of harmonized datasets. Int J Cartogr 5:178–194. https://doi.org/10.1080/23729333.2019.1610931

Langran M (1986) Integration of name selection and name placement. In: Proceedings, second international symposium on spatial data handling, 1986

Laubert H, Woska E, Habel R (1988) Lehrbuch für Kartographiefacharbeiter: 2 Kartengestaltung, 2. Aufl., Haack, Geographisch-Kartographische Anst., Gotha

Li Z (2007) Algorithmic foundation of multi-scale spatial representation. CRC

OpenTopoMap/users: Germany/OpenStreetMap Forum. https://forum.openstreetmap.org/viewtopic.php?id=21802. Accessed 13 Aug 2020

Sarjakoski LT (2007) Conceptual models of generalisation and multiple representation. In: Generalisation of geographic information. Elsevier, pp 11–35. https://doi.org/10.1016/B978-008045374-3/50004-1

Shea KS (1988) Cartographic generalization. Reston

Slocum TA, McMaster RB, Kessler FC, Howard HH (2009) Thematic cartography and geovisualization, 3rd edn. Pearson Prentice Hall, Upper Saddle River, p 561

Srnka E (1970) The analytical solution of regular generalization on cartography. In: Internationales Jahrbuch für Kartographie, Kartographisches Institut Bertelsmann, Güterloh, pp 48–62

Stanislawski LV, Buttenfield BP, Bereuter P, Savino S, Brewer CA (2014) Generalisation operators. In: Burghardt D, Duchêne C, Mackaness W (eds) Abstracting geographic information in a data rich world. Springer International Publishing, Cham, pp 157–195. https://doi.org/10.1007/978-3-319-00203-3_6

Töpfer F (1974) Kartographische Generalisierung, Haack, Geograph.-Kartograph. Anst., Gotha [u.a.]

Töpfer F, Pillewizer W (1966) The principles of selection. Cartogr J 3:10–16. https://doi.org/10.1179/caj.1966.3.1.10