Alternate definitions of linear and areal displacement between polylines – The plot thickens

Jiaxin Deng [a], Barry J Kronenfeld [a, *], Barbara Buttenfield [b], Larry Stanislawski [c],

[a] *Eastern Illinois University, jdeng@eiu.edu, bjkronenfeld@eiu.edu*
[b] *University of Colorado – Boulder, babs@colorado.edu*
[c] *U.S. Geological Survey, Center of Excellence for Geospatial Information Science, lstan@usgs.gov*

* Corresponding author

Keywords: cartographic generalization, line simplification, Hausdorff distance, Fréchet distance, shift displacement

Abstract:

Line simplification is an important component of cartographic generalization that aims to reduce the complexity of polylines while maintaining fidelity of location, shape and other characteristics. Line simplification has been implemented by various algorithms in previous studies. In recent years, line simplification has remained an active area of research in generalization, owing to its complexity and the diverse characteristics of linear geographic features, and to the recognition by the generalization community that adaptive strategies must be considered. Quite simply, there is no 'perfect' algorithm that is universally effective. And as cartographers experiment with different algorithms tailored to different situations, they must assess the quality of the results according to both qualitative and quantitative standards. In this study, we review standardized error metrics to evaluate and compare the quantitative geometric error associated with different simplification algorithms.

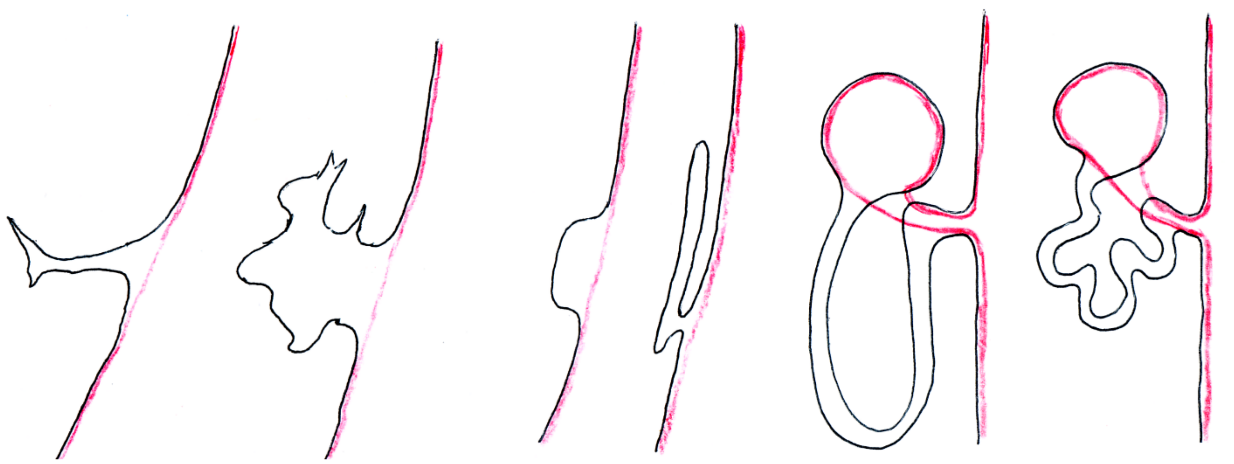
In particular, we review two natural classes of error metrics used to assess error in line simplification, namely linear displacement and areal displacement. We briefly review common semantic definitions of each metric class, identifying five distinct measures of linear displacement and three distinct measures of areal displacement. We highlight ambiguities that have not been noted in previous cartographic studies. We perform preliminary empirical analysis to clarify the differences among these measures, to determine their significance and to provide tentative guidelines for cartographic researchers.

Linear displacement metrics typically measure the maximum value of the shortest distance from each point on one line to the nearest point on the other. Two broad metric definitions are common: Hausdorff and Fréchet distances. Hausdorff distance is defined as the maximum separation between two sets of points in a metric space. The so-called metric space is also a point set, in which the distance between arbitrary elements can be defined. In practice, Hausdorff distance can be measured in at least three ways: point-to-point, point-to-segment, and segment-to-segment. The latter is the true Hausdorff distance between polylines, but it is difficult to calculate and is not implemented in at least some libraries containing functions that purport to calculate the Hausdorff distance (e.g. Shapely, https://pypi.org/project/Shapely).

Another measure of linear displacement is the Fréchet distance, which is used widely in the mathematical research field. Fréchet distance is measured by sweeping continuously forward along two polylines simultaneously, without backtracking, and measuring the furthest separation in this ordered sweep. Unlike Hausdorff distance, this method also considers the path sequence, and therefore is more suitable for measuring the similarity of polylines than the Hausdorff distance. However, calculation of the Fréchet distance is complicated and can be computationally intensive for large datasets. The discrete Fréchet distance is an approximation of the continuous Fréchet distance and can be computed more easily. When a sufficient number of discrete points are selected along the polyline, the discrete Fréchet distance is approximately equal to the continuous Fréchet distance.

Areal displacement is another natural class of standardized error metrics. Elsewhere, we show that areal displacement can be defined in at least three distinct ways. *Enclosure* displacement may be defined as the sum of absolute areas of all non-overlapping sub-regions that are enclosed by two segmental polylines. We introduce a novel definition referred to as *shift displacement*, which is the sum of the areas of regions that "shift sides", i.e. displacement regions that are to the right or left side of one polyline but to the left or right of the other polyline, respectively. Finally, *minimum homotopy area* is defined as the minimum area that must be swept across in a continuous transformation of one polyline into the other. The term *homotopy* comes from the field of mathematical topology and defines a continuous deformation between curves. In this sense, the minimum homotopy area is related to the Fréchet distance which may be considered a minimum homotopy distance.

The various measures of linear and areal displacement are distinguished by semantic and practical computational differences, and both factors can affect the cartographer's choice of metric. Semantically, three important distinctions can be made. First, linear displacement metrics are distinct from areal displacement metrics in that they measure distance (as opposed to area), and in that they measure the maximum (as opposed to average) error. This distinction is illustrated in Figure 1a. Conceptually, areal displacement may be thought of as the product of average linear displacement and the length of the polyline, although in practice both components of this product are semantically ambiguous.



(a) (b) (c)

Figure 1: Pairs of scenarios illustrating differences in measured displacement depending on the metric used. Original polylines shown in black, simplified polylines shown in red. (a) approximately equal linear displacement but very different areal displacement, (b) approximately equal Hausdorff distance but very different Fréchet distance, (c) approximately equal shift in areas covered by adjacent polygon features but very different positional displacement of line features.

Second, both linear and areal displacement measures can be divided into those that treat the polyline as an unordered set of points (Hausdorff distance, enclosure displacement) vs. those that treat the polyline as an ordered set (Fréchet distance, shift displacement, minimum homotopy area). This distinction is illustrated in Figure 1b. The latter conceptualization captures the process of deformation that occurs during simplification.

Third, measures of areal displacement can be distinguished according to whether the original polylines are considered to be linear features or implicit boundaries of areal features. In particular, minimum homotopy area measures the positional displacement of a polyline as a linear feature, whereas shift displacement measures the change in area covered by adjacent polygons. This difference is illustrated in Figure 1c.

In terms of practical computation, computation of metrics for both linear and areal displacement is more complicated than most cartographers have credited. For example, Hausdorff distance cannot always be determined by measuring distances between pairs of vertices, as it sometimes requires measuring vertex-to-segment or even segment-to-segment distances. Furthermore, functions in existing code libraries are not always well documented, so that users cannot be certain what metric is being computed. In the presentation, we will discuss computational issues with each metric and survey existing tools and functions available to perform computation of each metric. We will also illustrate a set of fairly simple data examples that can be used to determine what metric is actually being calculated by a given code function.

So far we have documented a number of potential semantic and computational issues in measuring displacement of polylines, but are the effects of these issues substantial in real-world cartographic generalization scenarios? To answer this question, we compare two linear and two areal displacement metrics as computed for a diverse sample of hydrographic lake boundary data taken from the USGS National Hydrography Dataset (NHD) and simplified using common line simplification algorithms. Our dataset includes 10 lakes from different geomorphic settings, taken from the high-resolution (HR) National Hydrography Dataset (NHD). Each lake is simplified to four different target scales using four different line simplification algorithms, resulting in 160 pairs of original and simplified features. For each feature pair, we measure two linear displacement metrics: the vertex-to-vertex Hausdorff distance, an easily computed metric that is available in many programming libraries, and the modified (vertex-to-line) Hausdorff distance that we have developed. We also assess two areal displacement metrics: enclosure displacement and shift displacement. The frequency with which metrics of the same type (linear or areal) have non-identical values and the mean and maximum magnitude of these differences are computed.

In sum, we identify at least five different definitions of linear displacement and three different definitions of areal displacement, and provide a comparative analysis of two linear and two areal displacement metrics. The questions we address are, how do these differ from each other, how can they be computed, and which metric(s) should be used by cartographers for given tasks? The results of our systematic analysis can provide guidance for cartographers seeking appropriate error metrics for use in assessing line simplification algorithms.