An Accuracy Assessment Method for Geographical Line Data Sets Based on Buffering

Håvard Tveite, Department of Mapping Sciences,
Agricultural University of Norway,
POBox 5034, N-1432 Ås, Norway, email: lanht@nlh.no

Sindre Langaas, UNEP/GRID-Arendal,
c/o Dept. of Systems Ecology, Stockholm University,
S-10691 Stockholm, Sweden, email: langaas@grida.no

Abstract

A method for quantitative assessments of spatial accuracy and completeness for line data sets is suggested and explored. Data sets of higher accuracy are used for the assessments. The method utilizes only generally available GIS operations. It works by establishing a number of buffers of increasing width around the lines. For each width, the maps are overlaid, and statistics are computed. By plotting the results in graphs, the usefulness of the method is demonstrated.

The method is explored using the Digital Chart of the World (DCW), with the Norwegian mapping authority's national N250 map series as reference data set.

Keywords

Accuracy, completeness, spatial, geographical, digital, data, line, buffer, overlay, DCW

Short title

Accuracy Assessments for Line Data Sets

1 Introduction

There is an increasing agreement within the spatial information society, both the academic sector and the community of practitioners, about the importance of having quality information associated with digital spatial data sets (Chrisman 1984, Goodchild and Gopal 1991, Morrison 1995). This recognition is basically driven by two forces -

1. The increasing economic and legal importance of spatial data and information in decision making processes, such as the AM/FM sector.

2. The possibility of combining multiple spatial data sets, possibly developed for different purposes, in sophisticated analyses using GIS.

There is also wide consensus about the need for different types of quality information. Although the identification of various elements of spatial data quality are still under debate and development, a considerable degree of consensus exists within major spatial data standards about certain key elements. Some of these quality elements are highly descriptive (qualitative), such as lineage, while others, such as positional accuracy, are principally quantitative by nature. Table 1 shows the main elements identified for the proposed European (CEN 1995) and US (USGS 1990) standards.

US Standard (SDTS)	European Standard (CEN/TC287)
• Lineage	• Lineage
	• Usage
	• Quality parameters:
	◇ Primary parameters:
• Positional accuracy	· Positional accuracy
• Attribute accuracy	· Thematic accuracy
• Logical consistency	· Logical consistency
• Completeness	· Completeness
	· Temporal accuracy
	◇ Secondary parameters:
	· Textual fidelity

Table 1: An overview of the data quality structure and characteristics for the US Spatial Data Transfer Standard (SDTS) and a working draft of a European Standard (CEN/TC287)

Despite the increased recognition of spatial data quality information, both qualitative and quantitative, as an integral part of spatial data sets for use in GIS, there has not been a matching increase in the elaboration of such information according to the metadata standards recently developed or under development. This can be explained by various causes, such as

- lack of conclusive, relevant and convenient measures and metrics
- lack of "toolboxes" for data quality assessments and descriptions
- lack of institutional and personal attitudes, competence, priorities and funding within the spatial information society

Scientifically, all of these causes deserve intensified research in order to reduce the gap between the awareness of the importance of spatial data quality assessments and the realization of such assessments within the spatial data community.

In this paper, we exploratively attempt to address the first point above, lack of conclusive, relevant and convenient measures and metrics. Specifically, we have indicated new measures and metrics related to the positional accuracy of line objects. However, the suggested method should, with some customisation, also be applicable to other types of geometric objects such as points, regions in 2D and 3D space and surfaces in 3D space.

The second point above, lack of "toolboxes" for data quality assessments and descriptions, is addressed by exploring a method for assessing quantitative quality metrics for spatial data sets using buffer and overlay, applying data sets of better quality as reference. Experimental results based on data from the Digital Chart of the World (DCW) and reference data from the Norwegian Mapping Authority (N250) are presented and discussed.

Quality measures for lines have also been discussed by Blakemore (1984), Skidmore & Turner (1991) and Dutton (1992). They all suggest the use of epsilon error bands in some form to model the effects of digitising errors on the positional accuracy of lines. The method presented in this paper can be used to assess the width of epsilon error bands, using buffer and overlay routines, and applying data sets of better quality as reference.

Kiiveri (1997) provides a general discussion of error models for assessing, representing and transmitting positional uncertainty in rectangular maps using a probabilistic approach. Kiiveri suggests the use of a smooth random perturbation function **P** over the unit square that maps onto the unit square. In contrast to the work of Kiiveri, this work only provides overall quality measures.

Parallel work on the use of buffering for assessing the spatial accuracy of lines has been done by Goodchild and Hunter (1997). Their research focuses strictly on spatial accuracy, and does not address completeness and miscodings for line data sets.

The research presented in this article is a result of the project "Issues of Error, Quality, and Integrity of Digital Geographical Data: The Case of the Digital Chart of the World (DCW)" (Tveite and Langaas 1994, Langaas and Tveite 1994). A preliminary presentation of this research was made at the ScanGIS'95 conference (Tveite and Langaas 1995).

The paper is structured as follows. In section 2, the characteristics of linear geographical phenomena are discussed. In section 2.2, different ways of assessing geographical line quality are presented. Our method for quantitative assessment of geographical line quality on the basis of data of higher geometric accuracy is introduced in section 3. The experimental results are presented and discussed in section 4, and section 5 rounds it all up with a short conclusion.

2 Linear geographical phenomena

The geometric line abstraction can be used to represent linear aspects of many geographical phenomena. Some examples:

- Centrelines of roads and railways
- Administrative (state, municipality) and economical (property) borders
- Utility lines (power lines, telephone lines, water and sewage tubes)
- Centerlines of rivers and streams
- Boundaries of natural phenomena (e.g. vegetation, soil)
- Shorelines

Some of these are natural phenomena and some are human "constructions" (often constrained by nature in some way).

There are many ways of providing quality measures for linear features. The choice of a quality measure depends to some extent on the type of linear feature that is under consideration.

2.1 "Scale"

Geometric accuracy is in many cases closely related to "scale". The "scale" of a line data set can to a certain extent be determined on the basis of the geometry of the line alone. Some indications on scale that can be derived from line data sets:

- The numerical resolution of the representation of the coordinates is the crudest measure of "scale" / spatial accuracy of a data set. For instance, if coordinates are stored as meters with no decimal digits, the accuracy cannot be expected to be better than about a meter. This measure is seldom reliable, and it is normally not a useful measure when the original data have been manipulated (for example transformed to a new projection), as most software do not consider accuracy in their transformations.

- Distance between neighbouring points on a line. The intended scale of the data set can normally be derived from the smallest distance between neighbouring points on a line. This is not true if, for instance, the data set has been smoothed by inserting new points on the lines using interpolation.

- Frequency of curvature change. For a curving phenomenon which changes curvature at a higher frequency than can be captured using the assumed geometric accuracy of the data set of interest, the maximum rate of curvature change can be used as an indication of the "scale" of the data set. This requires the phenomenon to exhibit fractal behaviour (Mandelbrot 1967, Mandelbrot 1982, Barnsley 1988) up to larger scales than the expected scale of the data set under consideration. Many linear features in nature seem to exhibit fractal behaviour over a large spectrum of scales. Examples of such phenomena are: shorelines, rivers/streams, vegetation boundaries and other natural boundaries.

2.1.1 Fractal behaviour of infrastructure

Infrastructure will cease to exhibit fractal behaviour at large scales. Roads will normally not change curvature more frequently than each 100 meter (>1000 meters for modern motorways, while perhaps 10-20 meters for really old roads). The same applies to railways, power lines, telephone lines and other utilities. At some point they will cease to exhibit fractal behaviour. The fractal behaviour of infrastructure is, in addition to cultural/historical issues, normally also influenced by the topography of the area (e.g. when building roads one used to try to avoid crossing steep hills, valleys and lakes).

2.2 Methods for assessing the quality of lines

In the following sections, some methods for calculating / quantifying the geometric accuracy of lines will be presented and discussed.

For our assessments, we assume that we have two independent line data sets (independent in the meaning that one is not derived from the other), X and Q, covering the same line theme and the same area (and with about the same temporal validity). The geometric accuracy of Q should be better (preferably by an order of magnitude) than the expected geometric accuracy of the data set X. If the expected difference in accuracy is smaller, it is advantageous that the data set Q has a known geometric accuracy. It is also expected that the completeness and consistency of data set Q is equal to or better than that of data set X.

The geometric accuracy of a line can be decomposed into two components:

- Positional point accuracy: Positional accuracy can easily be given for well defined points on the line (e.g. the end-points). For the rest of the line, it is difficult to say anything about positional accuracy and to quantify it.
- Shape fidelity: To be able to say something about the accuracy of a line, it is useful to talk about its shape fidelity as compared to another line. The shape fidelity should indicate to what extent the curvature of two lines are similar.

The types of spatial "errors" that can occur for linear data sets can be classified into categories. E.g.:

- Fuzziness of lines. The position of most linear phenomena get fuzzy as the scale gets larger, and it is generally impossible to determine an exact location (Burrough and Frank 1996). River centrelines and soil and vegetation boundaries are examples of fuzzy linear phenomena in nature, but also human constructions can be difficult to measure with extremely high accuracy (it is difficult to determine the centreline of an existing road with millimeter accuracy).
- Scale-dependent errors. These are errors that result from a too low sampling frequency when collecting data on the linear phenomenon of interest. Can be the result of generalisation (McMaster and Shea 1992).
  - Generalisation by sampling: A line-representation that has been generated by sampling a line of higher geometric accuracy represents a special case (e.g. using the Douglas-Peucker algorithm (Douglas and Peucker 1973)). Each point of the line is very accurately specified, but between the represented points, there can be large deviations between the interpolated line and the original position of the linear feature.
- Errors that result from erroneous sampling and data processing. Most digitising errors belong to this category.

It would be desirable to be able to separate these types of "errors" when describing the spatial accuracy of the geometric representations of linear geographical features.

2.2.1 Point measures

As long as one can determine corresponding points in X and Q, it is straightforward to calculate the geometric accuracy at these points. For single points one can measure the deviation vector ($\vec{e}$) of the point representation ($\vec{P_X}$) as compared to another representation of the same point with better (and known) geometric accuracy ($\vec{P_Q}$).

$$\begin{aligned} \vec{e} &= \vec{P_X} - \vec{P_Q} \\ &= (\vec{P_X}_x - \vec{P_Q}_x, \vec{P_X}_y - \vec{P_Q}_y, \vec{P_X}_z - \vec{P_Q}_z) \text{ for 3D space} \end{aligned} \tag{1}$$

The absolute value of this deviation vector ($|\vec{e}| = \sqrt{\vec{e}_x^2 + \vec{e}_y^2 + \vec{e}_z^2}$ for 3D space) is a useful measure for further (standard) statistical calculations.

For multiple points one has to resort to statistical measures to determine quality parameters. Standard deviation and variance can be estimated whenever the point-errors are independent and identically distributed from a Gaussian (normal) distribution. Such measures are most useful when the data sets have no spatial bias relative to one another.

The mean error vector (spatial bias) can be estimated as ($\vec{P_X}_i$ and $\vec{P_Q}_i$ are corresponding points in the two data sets):

$$\overline{\vec{e}} = \overline{\vec{P_X} - \vec{P_Q}} = \frac{1}{N}\sum_{i=1}^{N}(\vec{P_X}_i - \vec{P_Q}_i) \tag{2}$$

This is also an estimator of the statistical mean (μ) of the point error vectors. In the case of no point error bias ($\overline{\vec{e}} = 0$), estimators for the variance (σ^2) and standard deviation of the point errors ($\vec{e}$) are simply:

$$var(\vec{e}) = \frac{1}{N-1}\sum_{i=1}^{N}\vec{e_i}^2 \tag{3}$$

$$SD(\vec{e}) = +\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\vec{e_i}^2} \tag{4}$$

A combination of spatial bias and variance or standard deviation (SD) can be used as a measure of the spatial accuracy of points.

End-points The positional errors of line end-points can not be used to estimate the geometric accuracy of a line, even though they might give an indication of spatial bias. But for a large line-network data set, the positional errors at these points can be used as samples of the positional errors of the lines in the data set, and as such they can provide an indication of the positional accuracy of the line data set. End-points could be cross-roads and dead ends in a road network, river meets and lakes in a river/watercourse system or joints and end-points in a tube network.

If it is possible to identify corresponding end-points in the reference data set (Q) and the data set of unknown spatial accuracy (X), it is straightforward to find a statistical measure of the geometric accuracy of the end-points using the formulaes presented above.

An earlier effort on quantitative quality assessment on the DCW was performed using 40 evenly distributed cross-roads in the road and railway network in the area covered by ONC G18 (the south-west coast of USA), and using 1 : 100 000 scale topographical data (US DLG) as reference data sets (1 : 24 000 data were used for testing vertical accuracy). This work is described in a DMA report (ESRI 1990). The horizontal accuracy of the DCW automated from the G18 was measured to 583 meters at a confidence level of 90%. The DCW data tested were found free from bias.

Intermediate points As long as intermediate points are not well-defined features, corresponding intermediate points could be established by searching for the closest point on the other line.

A method for determining spatial accuracy of a line as compared to a line of better accuracy could then be to traverse the line, and at regular intervals (spacing ϵ) along the line take out a sample point, and for each of these points do a search for the closest point on the reference line. At each sample point, the distance vector, $\vec{e}$, to the closest point on the reference line is an indication of the spatial accuracy of the line at that point. An overall measure of the accuracy of a single line can be calculated statistically using $\vec{e}$ as in the formulaes presented above. The $\vec{e}$ will be a lower bound on the error for each sampled point on the line. The corresponding spatial accuracy measure will therefore always be too optimistic.

On the data set level, this method has to be applied for all lines that have corresponding lines in the reference data set, arriving at an overall measure of the positional accuracy of the lines in the data set.

This method can either be considered as an integration along the lines to find an average displacement, or as a sampling of point errors along the line. In the first case, the spacing ϵ should be made as small as practically possible in the numerical integration process. For point error sampling, the choice of spacing ϵ must be based on the spatial accuracy of the data sets. The lines we are interested in do not exhibit completely random behaviour, and this implies that the smaller ϵ that is chosen, the more strongly will the $\vec{e}$'s of neighbouring point samples be correlated. To get an overall statistical measure for the data set, ϵ should therefore be chosen so large that the $\vec{e}$'s of neighbouring points can be considered not correlated. The choice of ϵ will depend on the characteristics of the data set under consideration. It should at least be greater than the shortest distance between curvature changes along the lines in the data set. It would be interesting to do several calculations based on different ϵ's to give an assessment of the stability of the calculated spatial accuracy.

To determine separate measures for the line end-points and the interior of the lines, a transformation will have to be performed on each individual line prior to the traversal of the line, in such a way that the end-points of the corresponding lines match exactly.

2.2.2 Overall measures

While point measures may be easy to understand and calculate, they do not capture all aspects of line accuracy.

In earlier research on accuracy measures for lines, the epsilon error band (Blakemore 1984) has been a popular concept. Blakemore (1984) uses epsilon error bands to represent positional uncertainty of linear geographical features resulting from the geocoding and digitising process. Kraus (Kraus 1994) and Hunter & Goodchild (Hunter and Goodchild 1995) apply the error band technique to digital elevation models.

Skidmore and Turner use line intersect sampling theory to assess line accuracy by calculating the length of "generated" line that lies within the epsilon error band of the "true" line (Skidmore and Turner 1991).

In this paper we propose a method to assess overall measures of accuracy and completeness of a line data set relative to another line data set of better quality using buffering.

By applying buffering in our accuracy assessment method, it implicitly builds on the concept of epsilon error bands.

3 The buffer-overlay-statistics (BOS) method

Using the buffer-overlay-statistics (BOS) method proposed below one can find approximations for the epsilon error band (Blakemore 1984), average displacement, generalisation level and completeness for a line data set of unknown quality relative to another data set (preferably of better and known quality). The method works by establishing a number of buffers of various sizes around the lines in both data sets and comparing them (using overlay and statistics). The process is iterative. Graphs can be produced to visualize the above mentioned aspects of line data quality.

The BOS method assumes homogeneous quality for the whole area of both the reference data set and the data set of unknown quality.

3.1 Background

The elements of the BOS method are shown in figure 1. To the left are the two original lines. X is the line of unknown quality and Q is the line of know quality. The right part of the figure shows the result of buffering and overlaying the two lines. Four different types of areas result from the buffer and overlay operations: Areas that are inside both the buffer of X (XB) and the buffer of Q (QB) ($XB \cap QB$), areas that are inside XB and outside QB ($XB \cap \overline{QB}$), areas that are outside XB and inside QB ($\overline{XB} \cap QB$) and finally areas that are outside both ($\overline{XB} \cap \overline{QB}$). The last one is not used in our calculations.

If the lines are very similar, the area of ($XB \cap QB$) will dominate the other two, but as the lines become more different, the area of the other two will increase as a function of the size of the displacements. The area of ($XB \cap QB$) compared to the total area that is inside XB or QB could therefore be used as a measure for line data set accuracy.

The areas resulting from the buffer and overlay process can also be used to assess other aspects of line data set quality. Some of these will be explored further.

In order to produce useful assessments of line quality, a strategy for normalizing the statistics and aggregations had to be determined. We wanted to be

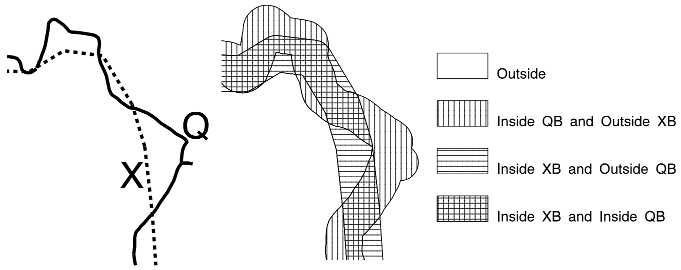

Figure 1: The elements of the BOS method.

conservative in our assessments of accuracy and quality and chose the normalization methods accordingly. For instance, since the area of XB (XB is the area resulting from buffering the data set of unknown quality) is likely to be smaller than the area of QB (QB is the area resulting from buffering the data set of known and better quality), the use of $Area(XB)$ for normalization will give the most conservative assessment of the positional accuracy of the line data set (average displacement). Using $Area(QB)$ would normally result in a lower average displacement, due to the expected difference in detail for the two line data sets.

We have chosen to buffer the lines in both the reference data set and the data set of unknown quality. Another approach would be to buffer only the dataset of unknow quality (the data set we want to assess the error band for), overlay it with the lines of the reference line data set, and calculate the percentage of the reference data set that fall outside the buffer. Trying this for many buffer sizes, on could determine the width of the error band (for instance the buffer size that includes 99% of the reference data set). Such an approach would have to assume that all the lines in the reference data set are also present in the data set of unknown quality. By studying the shape of a plot of buffersize versus inclusion percentage one could, however, get an impression of the size of the error band, even in situations of data set incompleteness. Our completeness/miscoding estimates are based on this principle and can also be used to find an epsilon error band. See figure 5 for example output.

The reason why we have buffered both of the lines in this work is that it gives a weighting of the errors, and therefore can give a measure of average displacement. Line pieces that are far away will contribute more to the error than line pieces that are close to the reference lines as long as they are not farther away than the width of the buffer. By buffering only one of the data sets, we can only determine if a line piece is inside or outside the buffer.

3.2 The reference data set

The proposed BOS method relies on a reference data set for the quality assessments. In order to give good quality estimates, the reference data set should be known to be of significantly better quality than the data set of unknown quality. But since the BOS method only gives relative measures, it will work with reference data sets of all kinds. The BOS method can therefore be used for exploring how a line data set behave with respect to another line data set, no matter what the quality of the data sets are.

3.3 Single line

In the method proposed below, buffering of lines and subsequent overlay analysis is performed to give a quantitative assessment of the geometric accuracy of a line relative to another line (of higher accuracy). The method is iterative, because it will not be possible to determine an optimal buffer size in advance (we do not yet know the spatial accuracy of the line under consideration). The size of the first buffer can be determined on the basis of the known spatial accuracy of the reference line (e.g. the standard deviation, SD, of point positions if that is available). For each iteration, the size of the buffer should be increased by a suitable delta. 4-5 iteration will probably be sufficient for optimum values of the first buffer size and the delta (for the examples we use 20 iterations), and the process should be terminated when the results seem to stabilize.

3.3.1 The iterative process

Before starting the iterative process it is useful to do some statistical calculations on the lines. The most interesting measure at this point in the process is the total length of the lines. Then we start the iterations:

For a number (n) of buffer sizes: $bs_i, i \in \{1, 2, 3, \cdots, n\}$ (bs_i is the size of the ith buffer), perform the following 3 steps:

First step - line buffering Perform a buffer operation on each of the two lines, X and Q, using the buffer size bs_i (resulting in a buffer $2 \cdot bs_i$ wide). Call the resulting polygons XB_i and QB_i, respectively.

Second step - overlay Perform an overlay of the two polygons XB_i and QB_i, the result being a new polygon data set: $XBQB_i$.

Third step - statistics Calculate statistics (depending on the method to be used: total area, number of polygons, total perimeter, perimeter/area for each polygon) on $XBQB_i$ for the following situations:

- areas inside XB_i and inside QB_i: $(Area(XB_i \cap QB_i))$
- areas inside XB_i and outside QB_i: $(Area(XB_i \cap \overline{QB_i}))$
- areas outside XB_i and inside QB_i: $(Area(\overline{XB_i} \cap QB_i))$
- areas inside XB_i or inside QB_i: $(Area(XB_i \cup QB_i))$

Example output from the iterative process (20 iterations) is represented graphically in figure 2. The statistics have been normalized using the sum of the areas inside either XB_i or QB_i ($Area(XB_i \cup QB_i) = Area(XB_i \cap QB_i) + Area(XB_i \cap \overline{QB_i}) + Area(\overline{XB_i} \cap QB_i)$).

The sum of the normalized areas of $Area(XB_i \cap \overline{QB_i})$ and $Area(\overline{XB_i} \cap QB_i)$ will approach 100% when the buffer size approaches 0, as long as the lines do not coincide for any length (they may cross). The limit value for a buffer size of 0 for the normalized $Area(XB_i \cap \overline{QB_i})$ should then be the length of the line X divided by the sum of the length of the lines X and Q. By choosing a very small first buffer size, good approximations of the relative length of the lines can easily be read from the figure. The graph will always show two crossing points. The buffer sizes for these crossing points will give an indication of the relative spatial accuracy of the two lines.

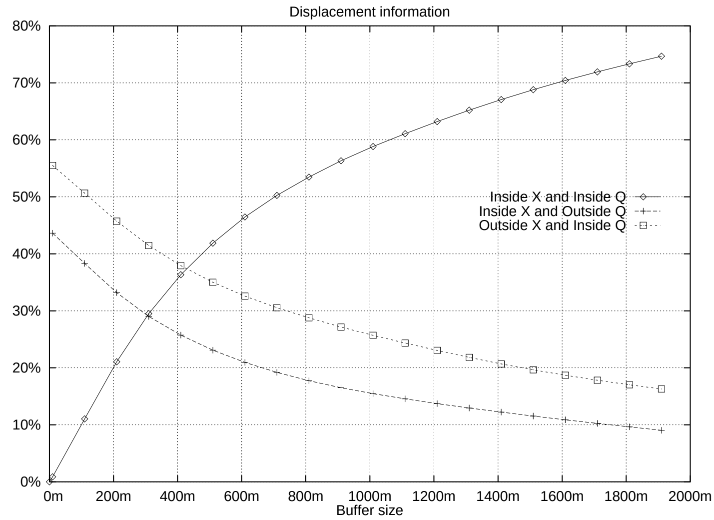

Figure 2: Example output from the iterative process ("normalized").

3.3.2 A measure for the geometric accuracy of a line

The statistics calculated in the above steps can be used to give measures of the deviation of the line X from the line Q.

Average displacement The average displacement, DE_i, for each step i is given by equation 5.

$$DE_i = \frac{\pi}{2} \cdot \frac{2 \cdot bs_i \cdot Area(\overline{XB_i} \cap QB_i)}{Area(XB_i)} = \pi \cdot bs_i \cdot \frac{Area(\overline{XB_i} \cap QB_i)}{Area(XB_i)} \quad (5)$$

The width of the buffer around the two lines is $2 \cdot bs_i$. Displacements along the line will not contribute to an increase in $Area(\overline{XB_i} \cap QB_i)$. Assuming an even distribution of line directions and errors, the underestimation due to this will be $\int_{x=0}^{\pi/2} cos(x) = 2/\pi$. This has been corrected for in equation 5 (the correction term $\pi/2$).

The computation is done relative to the area of the buffer around the line X of unknown quality (for "normalization" of the results), hence the term $Area(XB_i)$ in the denominator of equation 5. The choice of X instead of Q for normalization could be argued. Normally, the total length of Q will be greater then the total length of X (completeness), so using X for normalization will generally yield a more conservative estimate for the average displacement.

All the DE_is are used to determine the average displacement of a line Q (of better accuracy in our case) relative to a line X.

Example average displacement output from the iterative process (20 iterations) is presented graphically in figure 3. The graph must be expected to increase steadily until the buffer size reaches the average displacement of the lines, then the graph should start to flatten out. The shape of the graph will therefore give an indication of average displacement. An assessment of the average epsilon error band (Blakemore 1984) of the line X could also be made from this figure.

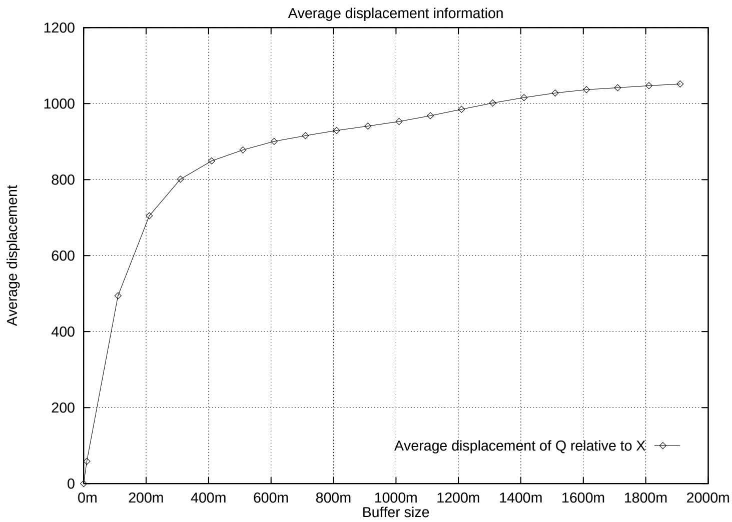

Figure 3: Example average displacement results.

Oscillation An indication of the oscillation of the lines X and Q relative to one another, O_i, for each step i is given by equation 6.

$$O_i = \frac{\#A(\overline{XB_i} \cap QB_i)}{Length(X)} \tag{6}$$

Where $\#A(...)$ is the count of areas/polygons. The length of the line of unknown quality (X) is used to normalize the result.

This measure is most useful for "randomly" oscillating phenomena, where it could be used as an indication of bias (there would probably be a bias if the oscillation, O, is low for randomly oscillating lines of different accuracy). By inspecting the behaviour of O_i for increasing buffer sizes, one can get an indication of spatial bias. An even better indication of spatial bias can be found using a modified equation:

$$O_i = \frac{\#A(\overline{XB_i} \cap \overline{QB_i})}{Length(X)} \tag{7}$$

Equation 7 should be expected to show a steeper decrease for buffer sizes of about half of the spatial bias, while for data sets without spatial bias there should be a steady decrease.

Figure 4 shows the results when equation 7 was applied on a data set with spatial bias and the same data set when spatial bias had been removed. The absolute spatial bias used for correction was in this case 559 meters.

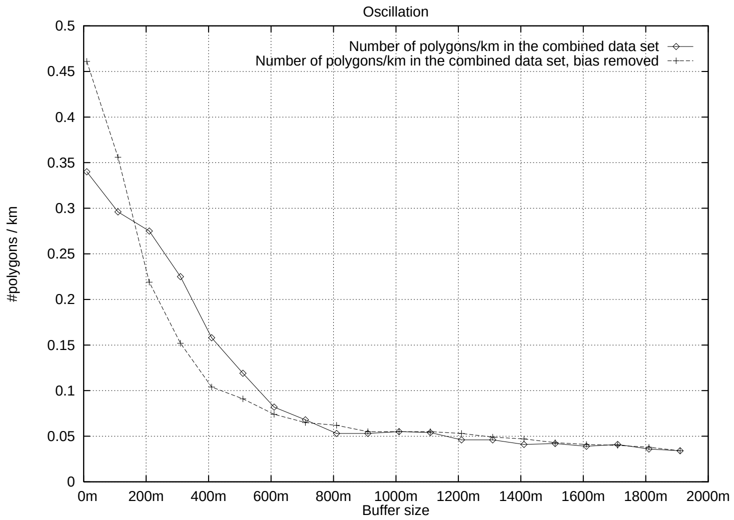

Figure 4: Oscillation, indication of spatial bias.

Another measure of oscillation could be found using X and Q directly, by counting the number of nodes introduced when overlaying the two lines. Such a method is more sensitive to bias than the one proposed above.

O is also a measure of relative scale for "randomly" (that is random appearance at the relevant scales) oscillating linear phenomena.

3.4 Line data sets

An advantage of the method described above is that the buffering method for calculating the geometric accuracy of lines can just as easily be applied to line data sets as to individual lines.

To apply the method directly on the data set level, all lines should exist in both data sets (the completeness criterion). In the real world this is seldom the case, since data sets of different scale are normally not at the same generalisation level. Lines that only are present in one of the data sets will introduce errors in the calculations.

In conjunction with spatial accuracy assessments on real linear data sets using this method, it is important that an assessment of the relative completeness of the data sets is made.

If there is known to be a spatial bias in the line data set of unknown quality, as compared to the reference data set, this should be removed before starting the calculations of average displacement.

3.4.1 Completeness and miscodings

Using an approximate measure of the geometric accuracy of a data set (X), it is possible to make an assessment of the completeness / number of miscodings of the X data set, as compared to the reference data set, Q. An approximate measure of the geometric accuracy can be obtained by applying the method presented above for single lines on the complete data sets (temporarily ignoring the lack of completeness measures).

The method outlined below use a combination of buffer, overlay and selection.

First step - buffer Perform buffering on the line data sets X and Q, using a buffer size, bs, which should be significantly larger than the geometric accuracy measure found for data set X.

It is necessary to choose the buffer size larger than the statistical measure of the spatial accuracy (could be expressed as the standard deviation, SD, that is a kind of weighted mean value of the errors). For example, when choosing a buffer distance four times as large as the SD for both line data sets, we ignore all errors within $4SD$'s of the reference lines.

The results of these bufferings are the data sets XB and QB.

Second step - overlay Do two line-polygon overlays: Overlay X with QB and XB with Q, resulting in the new mixed data sets XQB and XBQ.

Third step - statistics Statistics is run to determine completeness and miscodings according to the formulaes presented below.

Completeness Using XBQ, calculate the sum of the length of the lines from Q inside XB and compare it to the total length of lines in Q:

$$Completeness(X) = \frac{Length(Q\ inside\ XB)}{Length(Q)} \tag{8}$$

A more "exact" measure can be obtained by using the identity of the lines that are not in X, and calculate the length of the complete lines, as opposed to the part of the lines that do not fall within the buffer.

Miscodings Using XQB, calculate the sum of the length of the lines outside QB and compare it to the total length of lines in X. This is a measure of the amount of miscodings in X as compared to Q.

$$Miscodings(X) = \frac{Length(X\ outside\ QB)}{LengthX} \tag{9}$$

This can also be done in a more "exact" way using in the same method as described above.

A graphical presentation of the results of the completeness and miscodings calculations from equations 8 and 9 for a number of buffer sizes is shown in figure 5. The completeness graph must be expected to show a steady increase until the buffer size approaches the average displacement of the line data set, from there on the graph should flatten out. Conversely, the miscoding graph should decrease steadily until the buffer size approaches the average displacement.

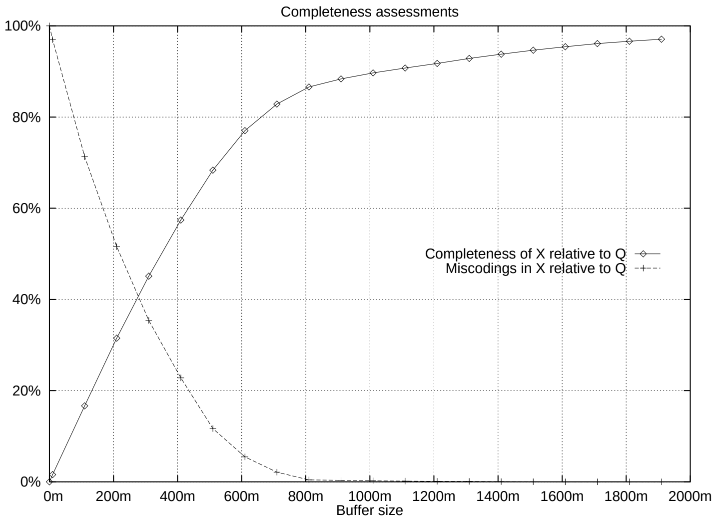

Figure 5: Example completeness and miscoding results.

3.4.2 Ensuring completeness

To prepare for the spatial accuracy assessments to come, all miscoded lines in X and all lines in Q that are not represented in X should be removed from the line data sets. The lines to be removed can be found using XBQ and XQB, as described above. The resulting data sets should be used in the rest of the process.

3.4.3 Accuracy assessments

If the data sets operated on are the original data sets, as opposed to the completeness adjusted data sets, the results could be roughly adjusted using the completeness measures determined above, as suggested in equation 10. The behaviour and usefulness of this equation has not been investigated further.

$$DE_i = \pi \cdot bs_i \cdot \frac{Area(\overline{XB_i} \cap QB_i) - Area(QB_i) \cdot (1 - Completeness(X))}{Area(XB_i) \cdot (1 - Miscoding(X))} \quad (10)$$

4 Experiments

In order to investigate the properties of the BOS method, it was applied to some of the line data sets the project had access to. The data sets that were used had some interesting differences, but in general they were quite similar. A less homogeneous collection of data sets would have been desirable, but the results of the experiments do illustrate some of the strengths and weaknesses of the method.

4.1 Data sets

The following data sets were used in the experiments:

N250 Digital geographical data from the Norwegian Mapping Authority (NMA). N250 is the 1 : 250 000 national map series of Norway. Datum: WGS84. Projection: UTM.

DCW (Digital Chart of the World) version 1, tile NK33 These data have been produced by the Defense Mapping Agency (DMA), USA, and is digitised from the DMA Operational Navigation Chart (ONC). The DCW is expected to have a scale of 1 : 1 000 000. Datum: Uncertain. Projection: Uncertain.

WVS (World Vector Shoreline) These data have been produced by the Defense Mapping Agency (DMA), USA, from existing digital and hard copy sources. It's nominal scale is 1 : 250 000. Datum: WGS84. Datum: UTM.

The N250 data set was used as the reference data set (X) in all the experiments.

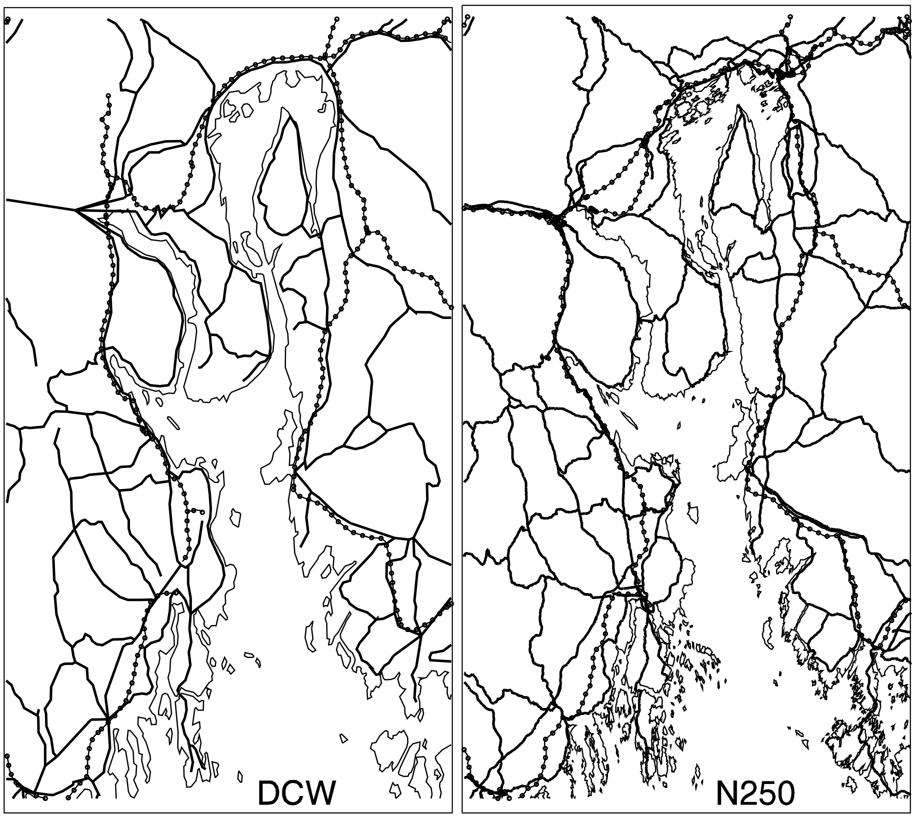

Figure 6: The X data set (DCW) and the Q data set (N250).

An overview of the themes from the clipped DCW and N250 data sets that have been used in the experiments are shown in figure 6. The coastline is shown with thin lines, the roads are shown with thick lines and the railways are shown with thin lines and dots. The DCW data sets have a spatial bias relative to the N250 data sets. This is probably due to reference system inconsistencies. The reference system assumed for the DCW data in these experiments ("WGS84", with the “Clarke 1866” projection) is not in conflict with the vague DCW specifications. The WVS data set has no significant spatial bias relative to the N250 data set.

4.2 Implementation

The method described, with some of the proposed measures and metrics, has been implemented using ARC/INFO version 7.0.4 (ESRI 1998) and its macro language AML on a Sun Ultra 1/140 running Solaris 2.5 with 96MB of internal memory. The overlay and buffer operations were performed using the “union” and “buffer” commands, while the statistics were calculated using “statistics” in “tables”. The graphs were produced automatically from the results of the statistics using GNU-plot. The “expensive” part of the processing is the buffer and overlay operations. For the original DCW road data set and the N250 adjusted road data set, the reported CPU time was about 50 minutes for 20 iterations (buffer sizes). The statistics calculations also took some minutes due to lots of disk traffic. For ease of implementation and comparison, 20 iterations / buffer sizes were used in all the experiments, starting with 10 meters and increasing with 100 meters for each iteration. The AML-file and GNU-plot files are available at the project WWW-page (Tveite and Langaas 1994).

4.3 Results

Experimentation was done using the road data, coastline data and railway data. The results seem to confirm the assumed behaviour of the method.

4.3.1 Road data

Four data sets were used in the experiments: The N250 reference data set (only main roads were used and this could mean that small roads present in the DCW data set were not in the reference data set which would result in a number of “miscodings”). The original DCW data set and a corrected version of the DCW (a spatial bias of 514 meters to the east and 264 meters to the south (578 meters absolute bias) was determined by manually measuring corresponding points in the two data sets. This bias was removed (Bordal 1995)). And finally, a completeness adjusted DCW was derived and used. Size of N250: 1721 arcs, 15171 arc segments. Size of DCW: 172 arcs, 737 arc segments.

The left part of figure 7 shows the results of running the buffer method on the DCW road network and the N250 road network in the Oslo region in Norway. The right part of figure 7 shows the results when bias had been removed from the DCW data set.

The left part of figure 8 shows the completeness/miscoding calculations (equation 8 and 9) for the raw data set, while the right part of figure 8 shows the the same calculations when bias had been removed.

The left part of figure 9 shows the average displacement information (equation 5) for the road data set. The upper curve shows the results for the original road data set, while the curve in the middle shows the results when the original road data set had been corrected for spatial bias. The lower curve shows the results when bias had been removed and the data sets had been completeness adjusted.

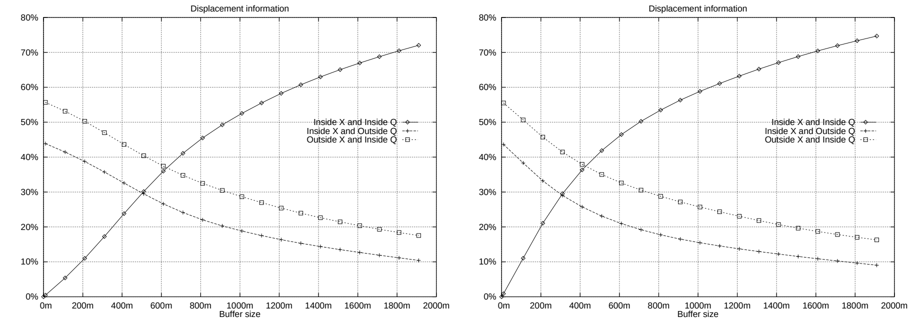

Figure 7: Accuracy results, road data (bias removed to the right).

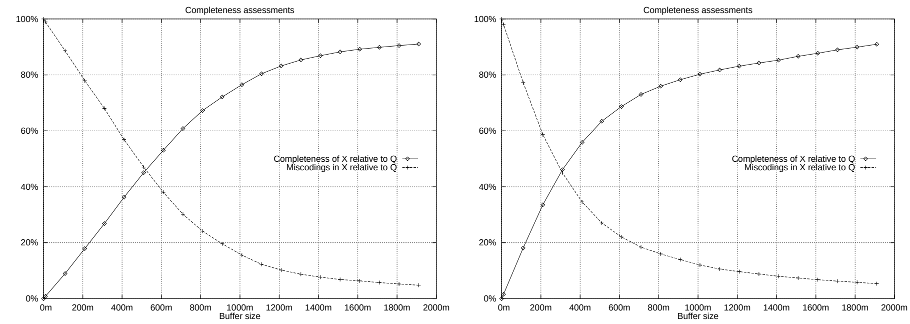

Figure 8: Completeness results, road data. Bias removed to the right.

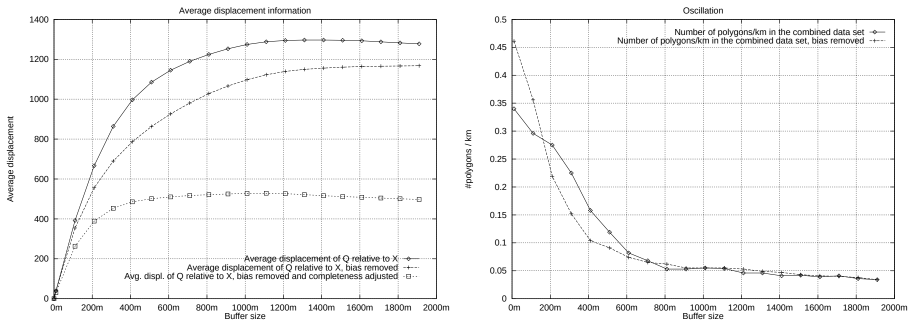

Figure 9: Average displacement (left) and oscillation (right), Roads. With and without bias and completeness adjustments.

The results of the oscillation statistics (equation 7) are shown in the right part of figure 9 for both the original data set and the bias adjusted data set. The effect of bias adjustment on the curve is evident.

4.3.2 Coastline data

Three data sets (N250, WVS, DCW) were used in the experiments. Size of N250: 709 arcs, 7470 arc segments. Size of WVS: 225 arcs, 4694 arc segments. Size of DCW: 42 arcs, 983 arc segments.

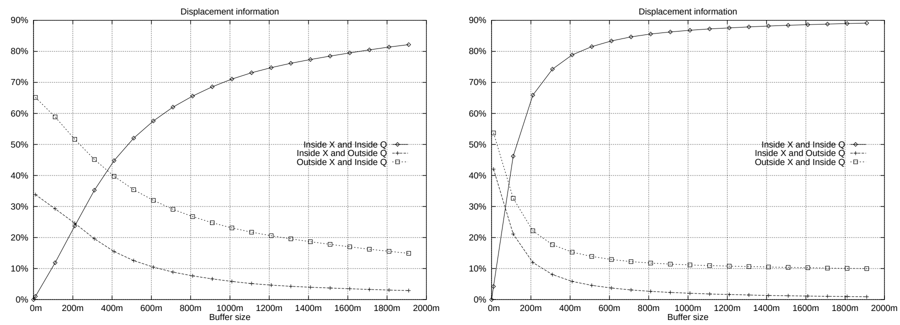

Figure 10: Accuracy results - coastline. DCW (left) and WVS (right).

The diagrams indicate a significant difference in the accuracy of DCW and WVS. The diagrams also indicate that WVS is quite accurate. The curves of the accuracy results for WVS in figure 10 cross for buffer sizes of 75 and 100 meters, whereas for DCW the crossings are at 200 and 400 meters.

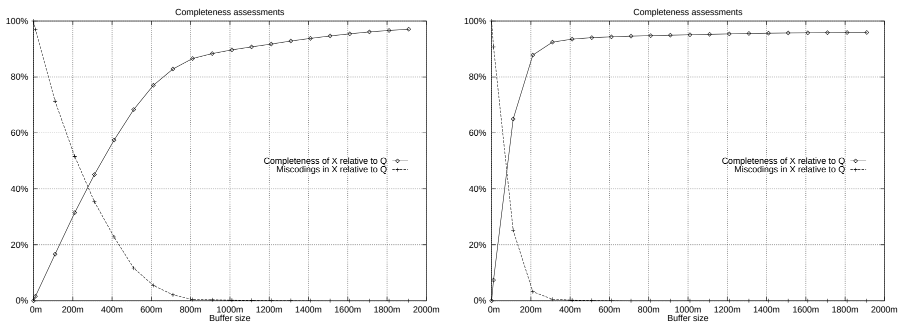

Figure 11: Completeness results - coastline. DCW (left) and WVS (right).

The completeness curves in figure 11 indicate that the completeness for WVS is probably about 90% (islands and bays), while the number of miscodings seems fairly small. Similar observations can be made for DCW (completeness about 80%). Completeness adjusted data sets would therefore probably have given better accuracy results.

The increase in the average displacement curve for WVS in figure 12 seem to flattens out before 100 meters. The continuing rise in the average displacement curve for buffer sizes above 100 meters is probably due to completeness problems.

Figure 13 shows the oscillation statistics for DCW and WVS compared to the N250 coastline. The DCW graph gives an indication of bias, having an irregularity at 100 meters, while the WVS graph does not give any such indication of bias.

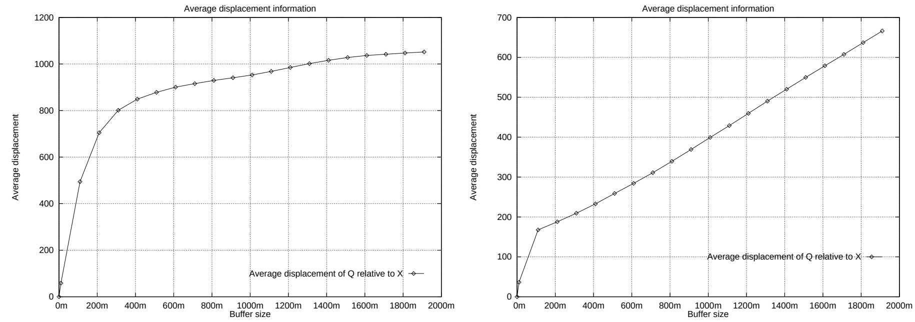

Figure 12: Average displacement - coastline. DCW (left) and WVS (right).

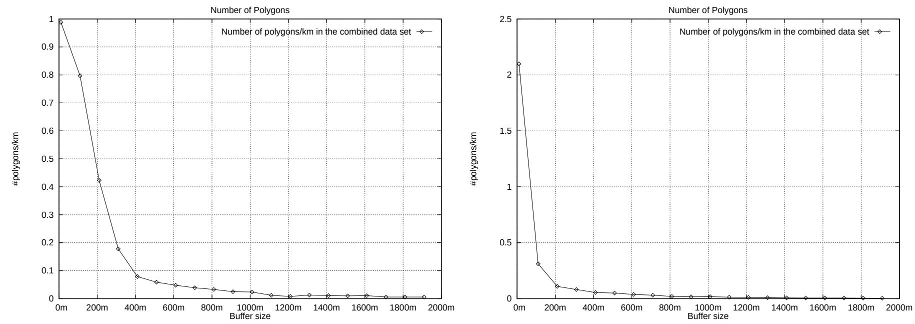

Figure 13: Oscillation - coastline (#polygons/km). DCW (left) and WVS (right).

4.3.3 Railway data

The railway data sets are rather small. Size of N250: 111 arcs, 3383 arc segments. Size of DCW: 39 arcs, 203 arc segments. Figures 14 to 15 show the same indications of bias as the unadjusted road data set, and from the completeness results in figure 14 one would expect the results to be significantly affected by the completeness and miscoding properties.

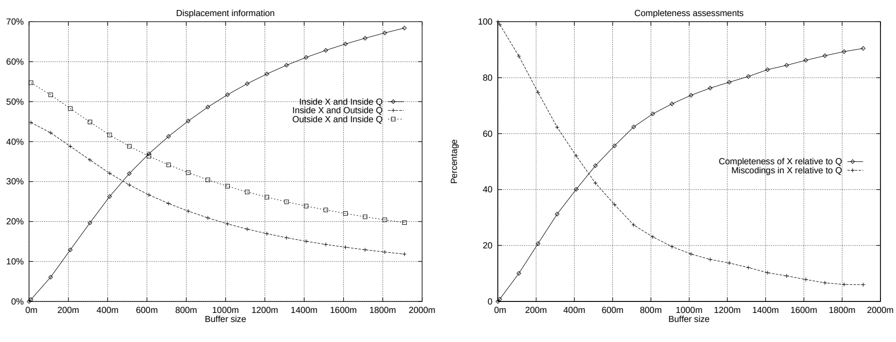

Figure 14: Accuracy (left) and completeness (right) results - railways.

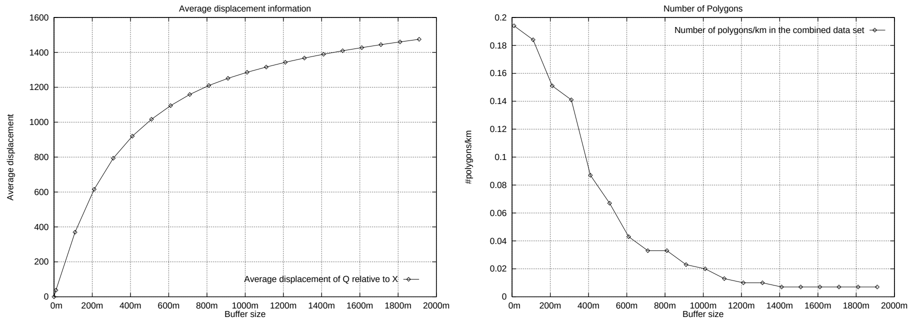

Figure 15: Average displacement (left) and oscillation (right) - railways.

5 Conclusions

The BOS method for quantitatively assessing the spatial accuracy of the representation of geographical linear features has been suggested and explored. Assessments are made relative to a reference data set. The method utilizes the standard GIS operations buffer and overlay to produce line and polygon data sets that can be analysed using simple aggregation statistics (e.g. sum and count). The method has been tested on real-world data sets and the results have been presented. The results show that the method produces reasonable measures of the geometric accuracy of line data sets.

Some points can be made as to the utility of the method and the suggested measures and metrics.

1. The method is currently tailored to line data sets, but can probably be extended to other kinds of spatial data sets.
2. The method relies on a reference data set that covers the region of interest.
3. The method utilizes the common GIS-operations *buffer* and *overlay* in addition to straightforward statistics. This means that the method should be easily adaptable to all kinds of GIS environments.
4. The method can be used to get indications of:
   - The spatial accuracy of a data set relative to another data set.
   - The completeness of a data set relative to another data set.
   - Oscillation of a data set relative to another data set.
   - The presence of bias in a data set relative to another data set.
5. The results seem to be useful when the data sets have a similar contents (reasonable completeness and miscoding characteristics).
6. The presence of spatial bias seems to affect the results in a systematic way, and can therefore probably be detected by the method.

7. The method has not been tested on data sets where the amount of data differs by more than an order of magnitude, but it is assumed that the method will not give satisfactory results under such conditions.

New quality metrics are needed for spatial data in general and line data in particular. We hope that we have provided some useful input on these issues.

We are looking forward to further testing and refinement of the BOS method.

Acknowledgements

We would like to thank Peter Fisher for providing literature references that helped put this work into context, and Michael Goodchild for suggestions that helped us improve the calculations of average displacement. We would also like to thank the three anonymous reviewers for their help in improving the article.

This work has been partially funded by the Norwegian Research Council under the Geographical Information Technology programme. Many thanks to the Norwegian mapping authority for giving us access to excerpts of the digital N250-data set for this research.

References

BARNSLEY, M., 1988, *Fractals Everywhere* (London: Academic Press).

BLAKEMORE, M., 1984, Generalisation and error in spatial data bases. *Cartographica*, **21**, 131–139.

BORDAL, J.-A., 1995, Accuracy Assessment of Geographical Line Data Sets, The Case of the Digital Chart of the World. Master's thesis, Agricultural University of Norway.

BURROUGH, P. A., AND FRANK, A. U. (editors), 1996, *Geographic Objects with Indeterminate Boundaries* (London: Taylor and Francis).

CEN, 1995, TC 287 N369: Geographic information - Data description - Quality. Technical report, European Committee for Standardization. Working Draft at stage 32.

CHRISMAN, N., 1984, The role of quality information in the long-term functioning of a geographic information system. *Cartographica*, **21**, 79–87.

DOUGLAS, D. H., AND PEUCKER, T. H., 1973, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Canadian Cartographer*, **10**, 110–122.

DUTTON, G., 1992, Handling positional uncertainty in spatial databases. In *Proceedings of the 5th International Symposium on Spatial Data Handling* (Columbia: International Geographical Union), pp. 460–469.

ESRI, 1990, Digital Chart of the World - DCW Error Analysis. Technical report, Defense Mapping Agency, Systems Center (AQE) 8613 Lee Highway Fairfax, Virginia 22031-2138. Prepared by Environmental Systems Research Institute, Inc. for Defense Mapping Agency, USA, Contract No.: DMA600-89-C-0023 CDRL C002.

ESRI, 1998. The homepage of Environmental Systems Research Institute, Inc. http://www.esri.com.

GOODCHILD, M. F., AND GOPAL, S., 1991, *The Accuracy of Spatial Databases* (London: Taylor and Francis).

GOODCHILD, M. F., AND HUNTER, G. J., 1997, A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science*, **11**, 299–306.

HUNTER, G. J., AND GOODCHILD, M. F., 1995, Dealing with error in spatial databases: a simple case study. *Photogrammetric Engineering and Remote Sensing*, **61**, 529–537.

KIIVERI, H., 1997, Assessing, representing and transmitting positional uncertainty in maps. *International Journal of Geographical Information Science*, **11**, 33–52.

KRAUS, K., 1994, Visualization of the quality of surfaces and their derivatives. *Photogrammetric Engineering and Remote Sensing*, **60**, 457–462.

LANGAAS, S., AND TVEITE, H., 1994. Project Proposal: Issues of Error, Quality, and Integrity of Digital Geographical Data: The Case of the Digital Chart of the World. http://www.nlh.no/ikf/gis/dcw/quality.ps.

MANDELBROT, B., 1967, How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science*, **156**, 636–638.

MANDELBROT, B., 1982, *The Fractal Geometry of Nature* (New York: W.H. Freeman and Co.).

MCMASTER, R. B., AND SHEA, K. S., 1992, *Generalization in Digital Cartography* (Washington: Association of American Geographers).

MORRISON, J. L., 1995, Spatial Data Quality. In *Elements of Spatial Data Quality*, edited by S. Guptill and J. L. Morrison (Oxford: Elsevier), pp. 1–12.

SKIDMORE, A., AND TURNER, B., 1991, A measure of vector map accuracy. In *Proceedings of the Symposium on Spatial Database Accuracy* (Melbourne: Department of Surveying and Land Information, The University of Melbourne), pp. 161–174.

TVEITE, H., AND LANGAAS, S., 1994. The Digital Chart of the World (DCW) & Data Quality Project. http://www.nlh.no/ikf/gis/dcw/dcw.html.

TVEITE, H., AND LANGAAS, S., 1995, Accuracy assessments of geographical line data sets, the case of the digital chart of the world. In *Proceedings of the 5th Scandinavian Research Conference on Geographical Information Systems (ScanGIS'95)* (Trondheim: Department of Surveying and Mapping, University of Trondheim), pp. 145–154.

USGS, 1990, Spatial Data Transfer Standard, version 12/90. Technical report, US Geological Survey.