Transactions in GIS, 2008, 12(1): 103–130

Research Article

Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National Standard for Spatial Data Accuracy

Paul A Zandbergen
Department of Geography
University of New Mexico

Abstract
Spatial data quality is a paramount concern in all GIS applications. Existing spatial data accuracy standards, including the National Standard for Spatial Data Accuracy (NSSDA) used in the United States, commonly assume the positional error of spatial data is normally distributed. This research has characterized the distribution of the positional error in four types of spatial data: GPS locations, street geocoding, TIGER roads, and LIDAR elevation data. The positional error in GPS locations can be approximated with a Rayleigh distribution, the positional error in street geocoding and TIGER roads can be approximated with a log-normal distribution, and the positional error in LIDAR elevation data can be approximated with a normal distribution of the original vertical error values after removal of a small number of outliers. For all four data types considered, however, these solutions are only approximations, and some evidence of non-stationary behavior resulting in lack of normality was observed in all four datasets. Monte-Carlo simulation of the robustness of accuracy statistics revealed that the conventional 100% Root Mean Square Error (RMSE) statistic is not reliable for non-normal distributions. Some degree of data trimming is recommended through the use of 90% and 95% RMSE statistics. Percentiles, however, are not very robust as single positional accuracy statistics. The non-normal distribution of positional errors in spatial data has implications for spatial data accuracy standards and error propagation modeling. Specific recommendations are formulated for revisions of the NSSDA.

Address for correspondence: Paul A Zandbergen, Department of Geography, University of New Mexico, Albuquerque, NM 87131, USA. E-mail: zandberg@unm.edu

1 Introduction

Data quality is a key component of any spatial data used in a Geographic Information System (GIS). In the context of spatial data, a number of components of data quality can be identified, including accuracy, precision, consistency and completeness as well as a number of dimensions, including space, time and theme (e.g. Veregin 2005). The combination of components and dimensions leads to more specific characterizations of spatial data quality; for example, in the case of accuracy a distinction can be made between spatial accuracy, temporal accuracy and thematic accuracy.

Concerns for data quality issues are clearly expressed in the development of standards for spatial data acquisition and dissemination. In the United States, the most commonly used standards are the Spatial Data Transfer Standard (SDTS) and the Content Standards for Digital Geospatial Metadata developed by the Federal Geographic Data Committee (FGDC). The FGDC was established to promote coordinated development and dissemination of geospatial data. The FGDC has been involved in several activities related to geospatial data quality, including the development of metadata content standards and spatial data accuracy standards.

In the following the emphasis will be on positional accuracy (or spatial accuracy) of vector data as one example of spatial data quality, although the accuracy of a continuous surface in the form of LIDAR elevation data will also be discussed. Accuracy in this context is defined as the absence of error, and therefore techniques to characterize accuracy rely on developing quantitative estimates of error. Positional accuracy of vector data depends on dimensionality; metrics are relatively well defined for point entities, but widely accepted metrics for lines and areas have not been established. For points, error is usually defined as the discrepancy (normally Euclidean distance) between the encoded location and the reference location derived from a data set of known and very high positional accuracy. Error can be measured in any of the three dimensions of spatial direction and in any combination of the three; the most common measures are horizontal error (distance measured in X and Y simultaneously) and vertical error (distance measured in Z).

Various metrics have been developed to summarize positional error for a set of points. The first one is mean error, which tends to zero for a single dimension when bias is absent. Bias refers to a systematic pattern of error – when bias is absent error is said to be random. Another common metric is Root Mean Square Error (RMSE), which is computed as the square root of the mean of the squared errors. RMSE is a measure of the magnitude of error and does incorporate bias in the X, Y and Z domains (Maling 1989). Under the assumption that the positional error follows a statistical distribution (like the normal distribution), statistical inference tests can be performed and confidence limits derived for point locations.

For lines and areas the situation is more complex since there is no simple statistical measure of error that can be adopted. Errors in lines arise from the errors in the points that define those lines. However, as these points are not randomly selected, the errors present at points cannot be regarded as typical of errors present in the line (Veregin 2005). Several approaches for characterizing the positional error in lines and areas have been proposed, from relatively simple buffering techniques (Goodchild and Hunter 1997, Tveite and Langaas 1999) to more complex stochastic simulation techniques (Leung and Yan 1998, Zhang and Kirby 2000, Shi and Liu 2002) and fractal geometry (Duckham and Drummond 2000). In practice most spatial data quality standards have circumvented the complexities of characterizing positional errors in lines and areas by specifying the use of "well-defined points" as part of the standards, requiring the selection of a sample of points from line and area vector data.

The most established map accuracy standard in the United States is the National Map Accuracy Standard (NMAS) (USGS 1999). NMAS was developed in 1947 (US Bureau of the Budget 1947) specifically for printed maps. It has lost some of its meaning in recent years for digital spatial data but it remains a widely employed standard. NMAS contains both a horizontal and vertical component. For horizontal accuracy, NMAS reads as follows:

> "For maps on publication scales larger than 1:20,000, not more than 10 percent of points tested shall be in error by more than 1/30th of an inch, measured on the publication scale. For maps on publication scales of 1:20,000 or smaller, 1/50th of an inch. These limits of accuracy shall apply in all cases to positions of well-defined points" (US Bureau of the Budget 1947).

The 1/30th and 1/50th of an inch are referred to as the Circular Map Accuracy Standard (CMAS); the underlying assumption is that the errors in the X and Y directions are of similar magnitude.

Vertical accuracy in NMAS is specified in terms of the contour interval at the 90th percent confidence interval as follows:

> "Vertical accuracy, as applied to contours maps on all publication scales, shall be such that not more than 10 percent of the elevations tested shall be in error more than one-half the contour interval. In checking elevations taken from the map, the apparent vertical error may be decreased by assuming a horizontal displacement within the permissible horizontal error for a map of that scale" (US Bureau of the Budget 1947).

This one-half contour interval is referred to as the Vertical Map Accuracy Standard (VMAS).

Despite the widespread use of NMAS, limited guidelines were developed on how to determine exactly if the vertical and horizontal standards for a particular publication scale were met. For example, with the exception of digital elevation data, there is no specific minimum number of points to be used in the evaluation nor are there specific guidelines for how to select the sampling locations and how to determine the positional error at those locations. NMAS also leaves every map producer in charge of the decision whether to test or not, and in fact most map products are not tested, even when claimed they comply with NMAS (Chrisman 1991). Most US agencies infer that a particular map product would pass the test based on compliance with certain specified procedures and equipment.

The American Society for Photogrammetry and Remote Sensing (ASPRS) also developed map accuracy standards (ASPRS Specifications and Standards Committee 1990), which provide accuracy tolerances for maps at a scale of 1:20,000 and larger. These guidelines are similar to NMAS in that they specify standards that have to be met at a particular map scale. The statistic employed, however, is the Root Mean Square Error (RMSE) and the error components are considered separately in the X, Y and Z directions.

With the advent of digital spatial data, the NMAS and the ASPRS guidelines became technically obsolete, since within a computerized environment the display scale is independent of the scale for which the data was created. Nevertheless, many users of geospatial data still commonly think about spatial data in NMAS terms, in part because that is how much of the data in use today were created. The ASPRS guidelines are also still in widespread use, since many digital photogrammetry products are often still treated as paper-like products when it comes to reporting accuracy specifications.

In 1998 the FGDC published the National Standard for Spatial Data Accuracy (NSSDA) (FGDC 1998), which superseded NMAS for digital mapping products. The ASPRS guidelines have also undergone revision following the publication of the NSSDA, in particular for vertical accuracy reporting (ASPRS LIDAR Committee 2004). The NSSDA implements a testing methodology for determining the positional accuracy of locations on maps and in digital spatial data relative to clearly defined georeferenced positions of higher accuracy. NSSDA also provides specific guidelines for what type of reference data to use (i.e. data of known and higher accuracy, or ground control points derived using surveying or GPS), the minimum number of points to be used (20), as well as their spatial distribution (a minimum of 20% in each quadrant of the study area, and no points closer together than 1/10th of the length of the diagonal of study area) (FGDC 1998).

One key assumption of the NSSDA is that the data do not contain any systematic errors and that the positional errors follow a normal distribution. Based on this assumption the NSSDA specifies a 95% confidence interval between test locations and reference locations. Horizontal accuracy is defined as a radial error, and the X and Y components of the error are not evaluated separately.

Horizontal accuracy in the NSSDA is defined as the "radius of a circle of uncertainty, such that the true or theoretical location of the point falls within that circle 95% of the time" (FGDC 1998). The horizontal accuracy statistic is determined as 1.7308 X RMSE. Horizontal accuracy is defined as a circular (or radial) error and therefore the annotation $RMSE_r$ is used. Vertical accuracy in the NSSDA is defined as "the linear uncertainty value, such that the true or theoretical location of the point falls within of that linear uncertainty value 95% of the time" (FGDC 1998). The vertical accuracy statistic is determined as 1.9600 X RMSE. Vertical accuracy is defined as a linear error in the Z direction and therefore the annotation $RMSE_z$ is used.

It is important to note that the NSSDA uses a 95% confidence interval, but that this is derived from the calculation of the RMSE; the values of 1.7308 and 1.9600 for horizontal and vertical accuracy, respectively, are directly derived from observations on the normal distribution as described by Greenwalt and Schultz (1968). This assumes there are no systematic errors and no major outliers, and that the distributions of both vertical and horizontal errors are independent and follow a normal distribution. However, these assumptions have not undergone much testing and are not elaborated upon in the original FGDC documents on NSSDA. The NSSDA does provide an alternative characterization of horizontal error if the X and Y components of the RMSE are different (FGDC 1998).

The Greenwalt and Schutz (1968) approximations used in the NSSDA have been criticized by MCollum (2003) as inappropriate, in particular when the values for $RMSE_x$ and $RMSE_y$ are very different, or when the error distribution in the X and Y directions are not independent. McCollum (2003) has suggested the use of circular error tables (Harter 1960, Beyer 1966, Folks 1981) as an alternative to the Greenwalt and Schutz (1968) estimators but maintains the assumption that the errors are normally distributed.

One of the key differences with NMAS is that the NSSDA does not specify thresholds for positional accuracy; instead, it provides a framework for determining and reporting the positional accuracy of spatial data, and agencies are free to set thresholds of tolerances for their product specifications.

The NSSDA has been widely endorsed by major federal and state agencies, although its adoption as an actual reporting standard in the metadata for spatial data has not been very rapid. One of the major strengths of the NSSDA is that is provides very clear guidance on how to select reference locations and calculate the RMSE values. In addition to the original FDGC documents, there is also a very easy-to-follow Positional Accuracy Handbook (Minnesota Planning 1999), complete with worked examples and ready-to-use spreadsheets.

The NSSDA has received widespread adoption in other spatial data accuracy standards and guidelines. For example, the Guidelines for Digital Elevation Data (NDEP 2004) include very specific references to the NSSDA and its protocol using the RMSE. The vertical accuracy test of the NDEP guidelines are directly copied from the NSSDA documents: "fundamental vertical accuracy is calculated at the 95% confidence level as a function of vertical RMSE" (NDEP 2004, p. 31). In addition, however, the NDEP guidelines provide some discussion of the fact that large errors are known to occur in the evaluation of elevation data and that therefore a normal distribution cannot be assumed. As a result, the 95th percentile statistic is suggested as a supplemental accuracy test, but not required. The experience of NDEP in the evaluation of many different elevation datasets has resulted in a proposal to the FGDC to revise the NSSDA (NDEP 2003). The proposal in effect suggests that the NDEP guidelines should be incorporated into the NSSDA. The recommendation is to maintain the current procedure to utilize the vertical RMSE statistic, but to limit its use to check points in open terrain. For areas with substantial ground cover, where the error distribution is more likely to deviate from the normal distribution, the 95th percentile is suggested as an alternative statistic. NDEP also proposes that the NSSDA include the recommendation to report vertical accuracy separately for different ground cover types

The NSSDA is also being adopted by the GPS community and is starting to replace other existing accuracy measures for GPS-collected data such as Circular Error Probable (CEP) and Spherical Error Probable (SEP). The use of the RMSE statistic and the 95th percentile is already very common in GPS, so the adoption of NSSDA does not present a very major conceptual change. Nevertheless, the GPS community in the United States has not had a single unified standard to report accuracy, so several agencies have taken the opportunity to develop GPS data standards that comply with the NSSDA, including guidelines for test procedures (e.g. USFS 2003).

The discussion of NMAS and NSSDA so far has been limited to positional errors in point locations; for line and area data the standards specify that a sample of "well-defined points" is to be used. The use of such "well-defined points", however, is somewhat open to interpretation, and can lead to bias and error (Van Niel and McVicar 2002). More sophisticated approaches to characterize positional errors in line and area features have been developed (Harvey et al. 1998, Zhang and Kirby 2000, Shi and Liu 2000), but these have not been implemented as standards. In addition, these approaches also assume that the errors in the composite points follow multi-dimensional normal distributions.

There have been several empirical descriptions of the distribution of positional error in different types of spatial data. Earlier work by Vonderohe and Chrisman (1985) on the positional error of USGS DLG data found evidence of non-normality. The work by Bolstad et al. (1990) on the accuracy of manually digitized map data also found small but statistically significant differences between the observed error distribution and a normal distribution. Positional accuracy of lines and areas has also been extensively studied by the Institut Géographique National (IGN), the French mapping agency, including studies by Vauglin (1997) and Bel Hadj Ali (2002). Their empirical validation using the Hausdorff distance (Munkres 1999) suggests that the error distribution is largely normal. More recent empirical studies have used small sample sizes which do not allow for proper distribution testing (e.g. Van Niel and McVicar 2002).

One application where error distributions have received much attention is the characterization of vertical error in Digital Elevation Models. One recent example is the study by Oksanen and Sarjakoski (2006) which determined the vertical error in a high resolution DEM. Results indicate that the error distribution was non-normal in nature and could not be characterized with a single estimator of dispersion. Similar evidence of non-normality of the error distribution has been found by other studies (Fisher 1998, López 2000, Bonin and Rousseaux 2005). This research on vertical error in DEMs suggests a number of explanations for the non-normality of the error distribution: (1) the frequent occurrence of gross errors (or blunders), in particular when data is interpolated from contours; (2) large positive spatial autocorrelation of the vertical error in DEMs; and (3) non-stationary processes underlying the occurrence of vertical errors in DEMs. One critical source of non-stationary behavior in DEMs obtained through Light Detection and Ranging (LIDAR) is the variability of vertical error with land cover (Hodgson and Bresnahan 2004, Hodgson et al. 2005).

A second application where the distribution of positional error has also received some attention is street geocoding. The positional accuracy of street geocoding appears to follow a log-normal distribution (Cayo and Talbot 2003, Karimi and Durcik 2004, Whitsel et al. 2004) but limited distribution testing has been performed on this type of positional error. Despite this (limited) evidence of non-normal behavior, most existing research on geocoding quality continues to assume a normal-distribution that can be characterized with statistics such as the mean, standard deviation or RMSE (Dearwent et al. 2001, Ratcliffe 2001, Bonner et al. 2003, Ward et al. 2005).

The error distribution of spatial data collected with Global Positioning System (GPS) has also received some attention. GPS equipment specifications commonly use error statistics such as Root Mean Square Error (RMSE) or CEP (Circular Probable Error of 50% of the error distribution, identical to median). Equipment specifications also often assume the RMSE is equivalent to the 63rd percentile of the error distribution and that two times the RMSE value (referred to as 2dRMSE) is equivalent to the 95th percentile of the error distribution. Much of the published literature on the empirical validation of positional errors in GPS locations uses the mean value or standard deviation (e.g. Bolstad et al. 2005, Wing et al. 2005). Very few studies have tried to properly characterize the error distribution of GPS observations. An exception is Wilson (2006) who argues that the distribution can be approximated by a bivariate normal distribution with no correlation between the two variables; only a slice in any direction will be a normal distribution. If the variance is assumed to be the same in each direction (empirical evidence by Wilson (2006) suggests this is not completely true), then the error distribution can be described by the Weibul distribution with shape factor $\beta = 2$, or the Rayleigh distribution. The Rayleigh distribution usually occurs when a two dimensional vector has its two orthogonal components normally and independently distributed (Papoulis 1984).

The assumptions of the Rayleigh distribution are not perfectly met for GPS errors since the error in the easting and northing are not exactly identical, in part due to the slight variability of GPS error with latitude resulting from poor satellite visibility at higher latitudes (Parkinson 1996). However, the Raleigh distribution represents the best known distribution to approximate the GPS error distribution (Wilson 2006). This has received very limited attention in the GPS literature and no mention is made of this in the spatial data standards discussed previously.

The positional error in TIGER road networks has also received some attention. The initial TIGER database was created by the US Census Bureau from a variety of sources, including the US Geological Survey's 1:100,000 scale maps series. The occurrence of relatively large positional errors in TIGER data has been widely recognized (e.g. O'Grady and Goodwin 2000, Trainor 2003, Wu et al. 2005) and a major effort to enhance the quality of the TIGER data was initiated in 2002. The MAF/TIGER Enhancement Program is expected to result in substantial improvements in the positional accuracy of the TIGER data, including road networks (US Census Bureau 2006).

In many other applications rigorous characterization of the error distribution has not received a lot of attention, and the assumption that the error is driven by a stationary random process is widespread. The non-normal distribution of positional error observed in several applications presents a serious challenge to current map accuracy standards which rely on assumptions of normality and utilize simple statistics to characterize its distribution, such as RMSE. The objective of this study, therefore, is threefold:

1. To develop a more rigorous characterization of the distribution of positional errors in spatial data, using a range of applications. These applications include Global Positioning Systems (GPS), street address geocoding, TIGER road networks and LIDAR elevation data.
2. To test the reliability of the RMSE statistic to characterize the positional error distribution as well as a comparison of alternative descriptors. These alternatives include: 90th and 95th percentile, and 90% and 95% RMSE. The underlying hypothesis is that the positional error of most spatial data is not normally distributed, and that the RMSE statistic as employed in the NSSDA protocol is unreliable to characterize the error distribution.
3. To develop recommendations for the revision of spatial data accuracy standards with specific reference to the NSSDA.

2 Methods

2.1 Data Collection

GPS. GPS observations were recorded using a mapping grade receiver (Trimble GeoXM). The unit was placed on a tripod on top of a surveyed bench-mark of 1st order horizontal accuracy located on the campus of the University of South Florida. GPS positions were logged every second for an 8-hour interval during which the unit was not moved. The coordinate system employed was UTM Zone 17N NAD 1983. No real-time differential correction was employed and no post-processing corrections were applied, since this could introduce a potential bias related to base-station characteristics. The raw GPS locations were plotted in GIS software and a random selection of 1,000 points was created for further analysis. The location of the surveyed benchmark was also plotted in the same UTM coordinate system, and the Euclidean (horizontal) distance between each GPS location and the surveyed benchmark was determined, as well as the X (easting) and Y (northing) error components.

One of the most critical components in achieving accurate GPS locations is the satellite geometry at the time the position fix is determined, which is captured in the values for Position Dilution of Precision (PDOP). Low PDOP values (i.e. < 6) suggest that GPS observations will be accurate to within the equipment specifications (Trimble 2002). The average PDOP value for the 1,000 positions was 2.06, with a minimum of 1.5, a maximum of 4.30 and a standard deviation of 0.37. The number of satellites used for position fixes varied between 6 and 8 (GPS receiver was limited to 8 channels) with an average of 7.7 and a standard deviation of 0.5. These values suggest that variability of satellite geometry was limited and not expected to be a major factor in influencing the error distribution. Multivariate regression analysis using PDOP and number of satellites as independent variables and positional error as dependent variables did not result in significant relationships. The sample was also split into quintiles based on PDOP values; the error distributions with the highest and lowest PDOP values were not statistically different based on the Kolmogorov-Smirnov (K-S) two-sample test. Under more variable conditions the effect of satellite geometry on positional accuracy would be considerable, but in this particular test design conditions were nearly ideal during the entire data collection period.

Geocoding. Student enrollment records for 2005 were obtained from the Orange County School Board for all public schools in Orange County, Florida. These records contain the home residence of each student, including street number, street name and postal codes (in the form of 5-digit United States ZIP code). These 163,886 addresses were street geocoded using a 1:5,000 street centerline network from Orange County for 2005 and parcel geocoded using a 1:2,000 parcel database of the Property Appraisers Office of Orange County for 2005, both using ArcGIS 9®. A small offset of 10 m was used in the placement of the street geocoded locations to indicate the side of the street the address was located on. The coordinate system employed was UTM Zone 17N NAD 1983. Only those records which could reliably be geocoded using both techniques were used in further analysis. Duplicate locations resulting from siblings residing at the same physical address were also removed, resulting in a total of 62,142 records. A random selection of 1,000 points was created for further analysis. The positional accuracy of the street level geocoded locations was determined by measuring the Euclidean distance between the street level geocoded point and the centroid of the associated parcel.

TIGER Roads. A TIGER 2000 road network was obtained for Orange County, Florida from the US Census Bureau. A topological data structure was enforced, creating a dataset of all endpoints and intersections in the road network. The TIGER road network was compared to 1 m digital color orthophotography for 2005 and a 1:5,000 street centerline network from Orange County for 2005. The coordinate system employed was UTM Zone 17N NAD 1983. The orthophotography was originally created to meet NMAS for 1:12,000 maps, i.e. the 90th percentile of the error distribution is 10 m or less, but no formal accuracy testing for this particular area has been performed. While the error in the orthophotography and the street centerlines presents a confounding factor, the error in the TIGER data is approximately an order of magnitude higher, suggesting that improved accuracy in the determination of intersections would produce similar results. Intersections were randomly selected from the TIGER road network and a determination was made whether a meaningful comparison to the street centerlines could be identified. (Note: the representation of the road network in TIGER data is sometimes so poor that it does not approximate the network as represented by the orthophography, making it often impossible to determine which intersections it is supposed to represent). Where a logical match could be made, the Euclidean (horizontal) distance between the two intersections was determined as a measure of the positional error in the TIGER data. This process was repeated until 1,000 reliable comparisons were identified to produce the final dataset for further analysis.

LIDAR. LIDAR accuracy assessment data was obtained from the North Carolina Flood Mapping Program for the combined Neuse and Tar-Pamlico river basins. The raw LIDAR data for this area was collected in 2002 and processing of the data was completed in 2004. A 20 feet bare earth DEM was created by the North Carolina Flood Mapping Program. The coordinate system employed was State Plane North Carolina NAD 1983 in US survey feet. Details on the collection, processing and accuracy assessment of the LIDAR data are provided in a series of Issue Papers produced by the North Carolina Flood Mapping Program (2006); a brief summary follows. The original LIDAR data was collected with a ground spacing of sampling points of approximately 3 m. To produce the bare earth DEM a combination of manual and automated cleaning techniques were employed. These post-processing techniques included the use of automated procedures to detect elevation changes that appeared unnatural to remove buildings, as well as the use of last returns to remove vegetation canopy.

As part of the data collection, the North Carolina Flood Mapping Program also completed a County-by-County accuracy assessment of the LIDAR data using independent survey contractors, which employed a combination of traditional surveying and Real-Time Kinematic GPS to achieve a target vertical survey accuracy of 5 cm (North Carolina Flood Mapping Program 2005). The survey reports for all Counties within the Neuse and Tar-Pamlico river basins were obtained; these reports include the accuracy assessment of the LIDAR data derived through a comparison with the elevation at each surveyed location with a Triangulated Irregular Network (TIN) elevation model based on the bare-earth LIDAR elevation points. Vertical accuracy was reported in meters as a positive or negative value with three significant digits. The locations of all survey points were plotted. River basin boundaries were used to select only those survey points falling exactly within the basin; 1,000 random locations within the basin were selected for further analysis. Both the original error values and the absolute error values were used to account for the fact that the RMSE statistic does not distinguish between positive and negative error values.

2.2 Positional Error Characterization

The positional error distributions were characterized in a number of ways to determine the degree to which they follow a normal distribution. First, basic descriptive parameters of the distributions were derived, including mean, median, minimum, maximum, standard deviation, and inter-quartile range. Second, the degree of normality of the distribution was determined using skewness and kurtosis, in addition to standard normality tests, including the Lilliefors test (Lilliefors 1967) and the Shapiro-Wilk test (Shapiro and Wilk 1995). Third, histograms and normal Q-Q plots were created and compared to the normal curve. Since initial findings and previous studies suggested that some error distributions appear to follow a log-normal distribution, error distributions were $\log_{10}$ transformed, and the characterization steps above were repeated. The $\log_{10}$

transformation was not performed on the original values of the LIDAR errors because of the presence of many negative values; only the absolute values were used in the transformation. In addition, the GPS error distribution was tested against the Raleigh distribution, based on previous work by Wilson (2006) suggesting that this distribution is commonly observed in GPS errors. This also included characterizing the GPS error distribution separately in the X and Y directions. Testing against the Raleigh distribution and characterizing the error separately in the X and Y directions was also performed for the errors in geocoding and TIGER roads, but did not provide a more meaningful description than using the Euclidean horizontal distance alone and are therefore not reported here.

2.3 Evaluating Error Statistics

For each of the four different error distributions the following error statistics were determined using the original 1,000 points in each sample: (1) 90th percentile – the maximum error value of 900 points, disregarding 100 points with the highest error values; (2) 95th percentile – the maximum error value of 950 points, disregarding 50 points with the highest error values; (3) 90% RMSE – Root Mean Square Error using only 900 points, disregarding 100 points with the highest error values; (4) 95% RMSE – Root Mean Square Error using only 950 points, disregarding 50 points with the highest error values; and (5) 100% RMSE – Root Mean Square Error using all 1,000 points.

For each of the four spatial data types used in this study, only the untransformed positional error distributions were employed in the evaluation of the accuracy statistics. For the GPS data the horizontal error was used and not X and Y separately. For the LIDAR data the absolute value of the vertical error was used. This approach most closely reflects the implementation of current spatial data accuracy standards, including the NSSDA.

The original error statistics were based on the complete sample of 1,000 locations. To determine the robustness of these error statistics, a Monte Carlo simulation technique was employed for each of the four error distributions which is briefly described below. The NSSDA protocol for testing the positional accuracy of spatial data requires a minimum of 20 points. The Monte Carlo simulation employed in this study carries out the NSSDA protocol and then repeats it 100 times by using a conditional stratified random selection of 20 points from the sample of 1,000. A stratified random selection is necessary since the NSSDA protocol imposes additional limitations on the selection of sample points: no less than 20% of sample points have to be selected within each of the four quadrants, and individual sampling locations can be no closer together than 1/10th the diameter of the study area. To accomplish these requirements, each sampling location was assigned a quadrant (NW, NE, SE or SW) and 25% (or 5 out of 20) locations were randomly chosen from each quadrant. A $1{,}000 \times 1{,}000$ proximity matrix was constructed using GIS for each of the error distributions to limit the selection of points within close proximity of each other. If a selected point was found to be within the minimum allowed distance of another selected point in the sample, another point was selected at random within the same quadrant. This process was repeated until the minimum distance condition was met for the complete sample of 20 points. This stratified-random sampling process was not employed for the GPS data, since all 1,000 GPS positions are estimates of the same location. Sampling restrictions by quadrant and/or proximity would therefore introduce undesirable bias towards selecting points at greater distances from the reference location. For each sample of 20 points the same error statistics were determined: 90th percentile, 95th percentile, 90% RMSE, 95% RMSE and 100% RMSE. The Monte Carlo simulation was implemented using a Macro in Microsoft Excel®.

For each spatial data type, the average values based on 100 estimates for each error statistic were compared to the original values based on the complete sample of 1,000 points. The variability in the 100 estimates for the error statistics was also determined by comparing the standard deviation of the estimates for each error statistic to the average value. An error statistic is considered robust in this study if the average value approximates the original value and if the relative variability is small.

The chosen approach to testing the robustness of the error statistics has a very practical application: if the NSSDA protocol were applied as specified (a single sample of a minimum of 20 well-defined and evenly distributed points) what is the likelihood that results would be very different from the original error statistic? The hypothesis is that RMSE statistics in general are more robust than percentiles since RMSE uses the entire distribution. However, for non-normally distributed positional errors the 100% RMSE is not expected to be very robust relative to statistics that "trim" a small percentage of outliers.

3 Results

3.1 Description of Positional Error Distributions

Table 1 provides descriptive statistics of the positional errors in the four types of spatial data considered. For the GPS data the horizontal error, the error in the X direction and the error in the Y direction are reported separately. For the LIDAR data, both the original error value and the absolute values are reported. Each of the four types of error will be described below, accompanied by a diagram showing the locations of the 1,000 observations.

Figure 1 shows a map of the 1,000 GPS positions, including the location of the surveyed benchmark. The mean error is 2.470 m and the largest observed horizontal error is 7.800 m. These values are typical for uncorrected GPS position fixes using a mapping-grade receiver (Bolstad et al. 2005). The median error of 2.215 m is a bit

Table 1 Summary descriptive statistics of positional errors (in meters) in four types of spatial data (n = 1,000)

Statistic	GPS	GPS-X	GPS-Y	Geocoding	Roads	LIDAR Orig.	LIDAR Abs.
Mean	2.470	0.019	0.002	30.420	38.467	–0.045	0.146
Median	2.215	–0.081	0.101	17.668	29.869	–0.045	0.109
Standard Deviation	1.419	2.008	2.022	40.731	37.232	0.214	0.162
Minimum	0.126	–6.104	–6.821	0.725	1.327	–1.921	0.000
Maximum	7.800	6.040	6.778	600.345	628.515	2.335	2.335
Interquartile range	1.847	2.580	2.550	34.412	29.639	0.215	0.141

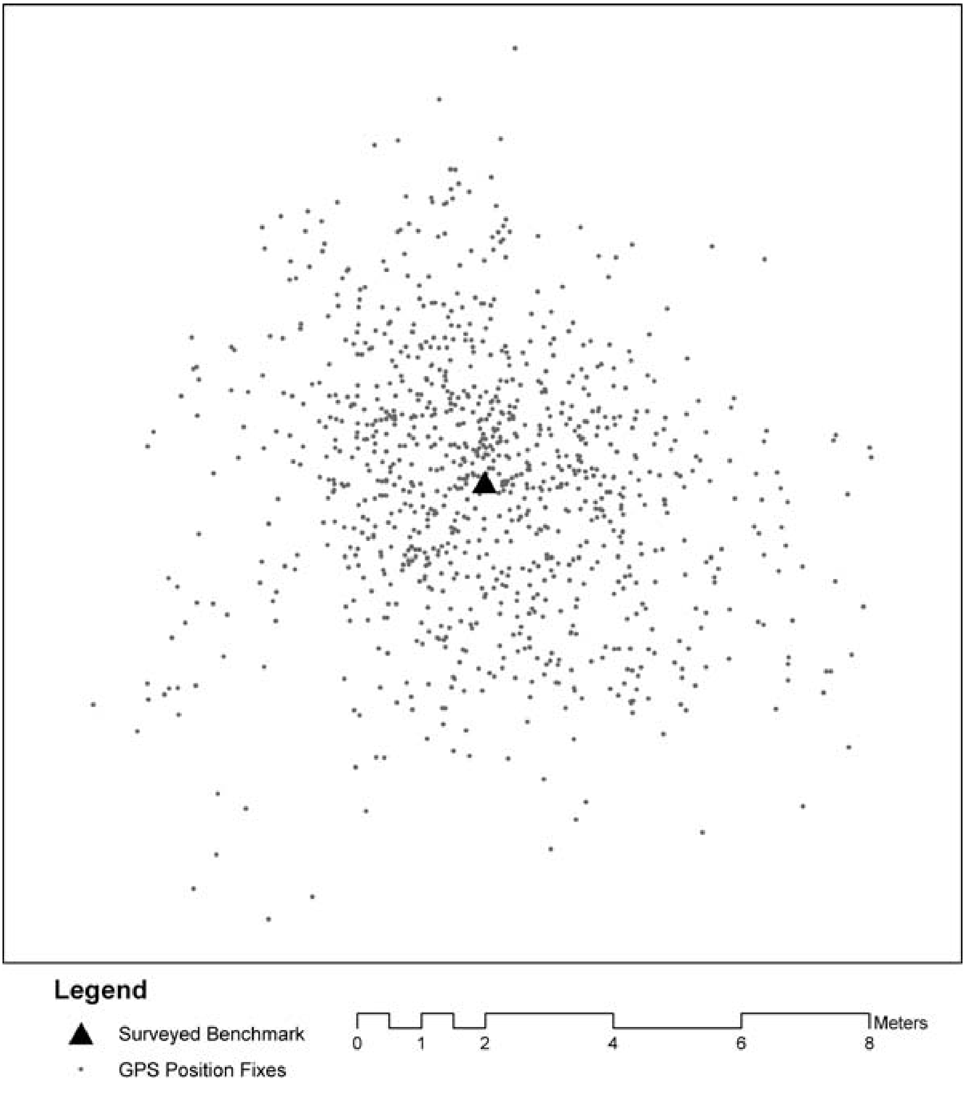

Figure 1 Uncorrected GPS positions collected over an 8-hour period (n = 1,000)

smaller than the mean, providing a first indication the distribution is not normal. Table 1 also reveals that the X and Y components of the GPS error are very similar, but not identical. The mean and median values for the X and Y components are very similar, suggesting both distributions could be normal.

Figure 2 shows the locations and magnitude of the positional error in street geocoding results within Orange County, Florida. The spatial distribution reflects the overall population density within the study area. The spatial pattern in the magnitude of the errors does not reveal any spatial bias, i.e. errors of varying magnitude occur in all parts of the study area, and there are no neighborhoods with only small or only large errors. The mean error is 30.420 m, which is similar to those reported in previous studies on

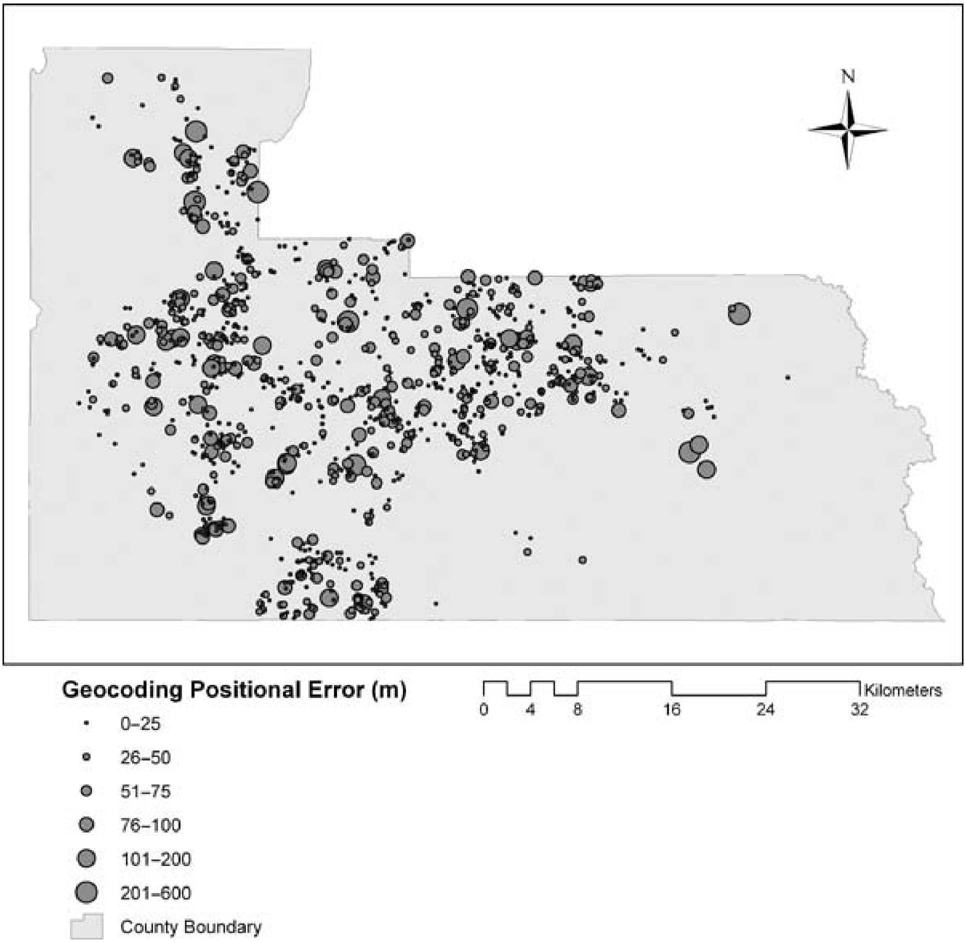

Figure 2 Location of parcel centroids and error magnitude in street geocoding of associated addresses (n = 1,000)

the positional error in street geocoding (Dearwent et al. 2001, Bonner et al. 2003, Cayo and Talbot 2003, Karimi and Durcik 2004). The maximum is around 600 m, representative of a small number of large outliers. The median error of 17.668 m is much smaller than the mean, suggesting the distribution is not normal.

Figure 3 shows the locations and magnitude of the positional error in the TIGER road network within Orange County, Florida. The spatial distribution is more dispersed than the results of street geocoding, reflecting the presence of major arterial roads in areas of low population density. The spatial pattern in the magnitude of the errors reveals that some of the larger errors occur in the lower density suburban and rural areas of the County, in particular in the southeastern portion. These locations reflect intersections of arterial roads with local roads in rural areas or in newly developed suburban communities. The mean error is 38.467 m, while the maximum of more than 600 m represents a small number of large outliers. The median error of 29.869 m is smaller than the mean, providing a strong indication the distribution is not normal.

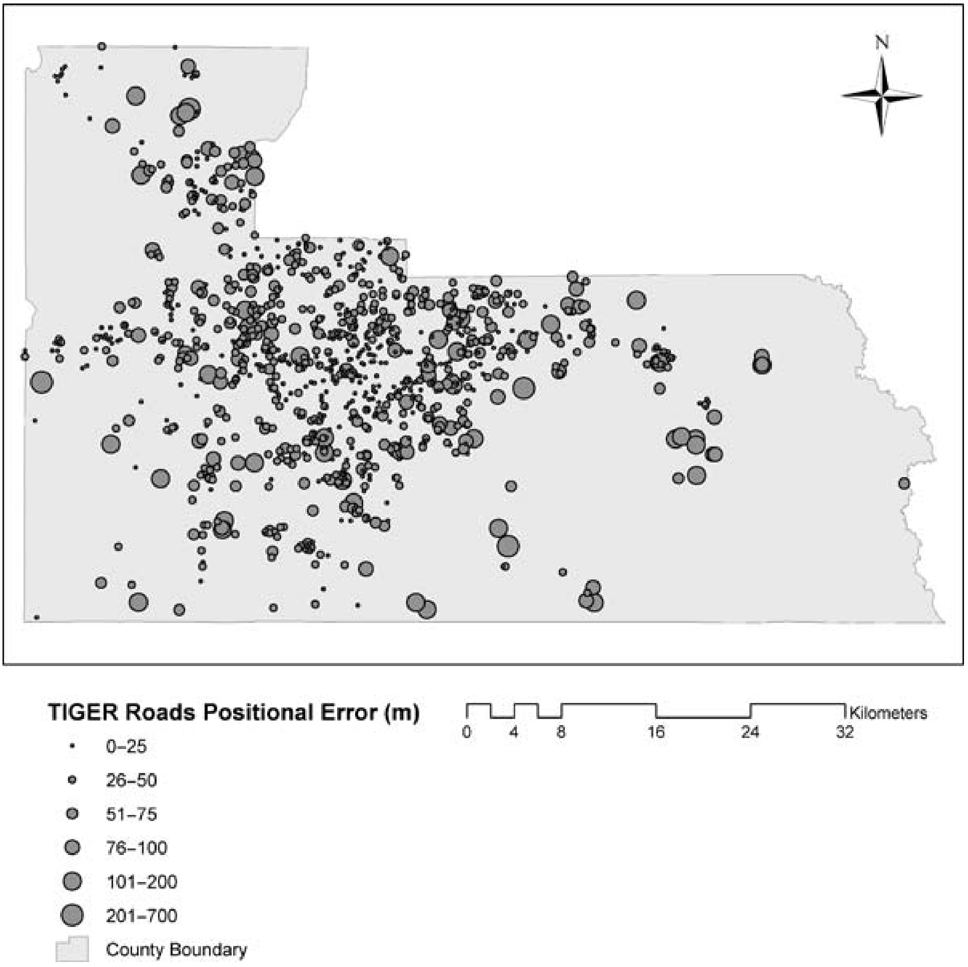

Figure 3 Location of TIGER road network intersections and error magnitude compared to 1:5,000 street centerlines (n = 1,000)

Figure 4 shows the locations and magnitude of the vertical error in the LIDAR DEM. The survey points are located throughout the entire river basin, but some degree of local clustering can be observed. This reflects the nature of the collection of the survey data: within a County a number of study areas of several square miles are identified, and several dozen samples in relative close proximity are taken within each study area across several land cover types. Individual survey locations are typically not closer together than several hundred meters, but still appear as clustered at the scale shown in Figure 4. For the original error values, both the mean and the median are –0.045 m, suggesting a very symmetrical distribution around zero. The minimum value is –1.921 m and the maximum value is 2.335 m; these observations clearly do not meet the typical accuracy expectations for LIDAR data, but reflect a very small number of outliers. When considering the absolute error values, the mean (0.146) and median (0.109) are different, and the standard deviation and range have logically been reduced.

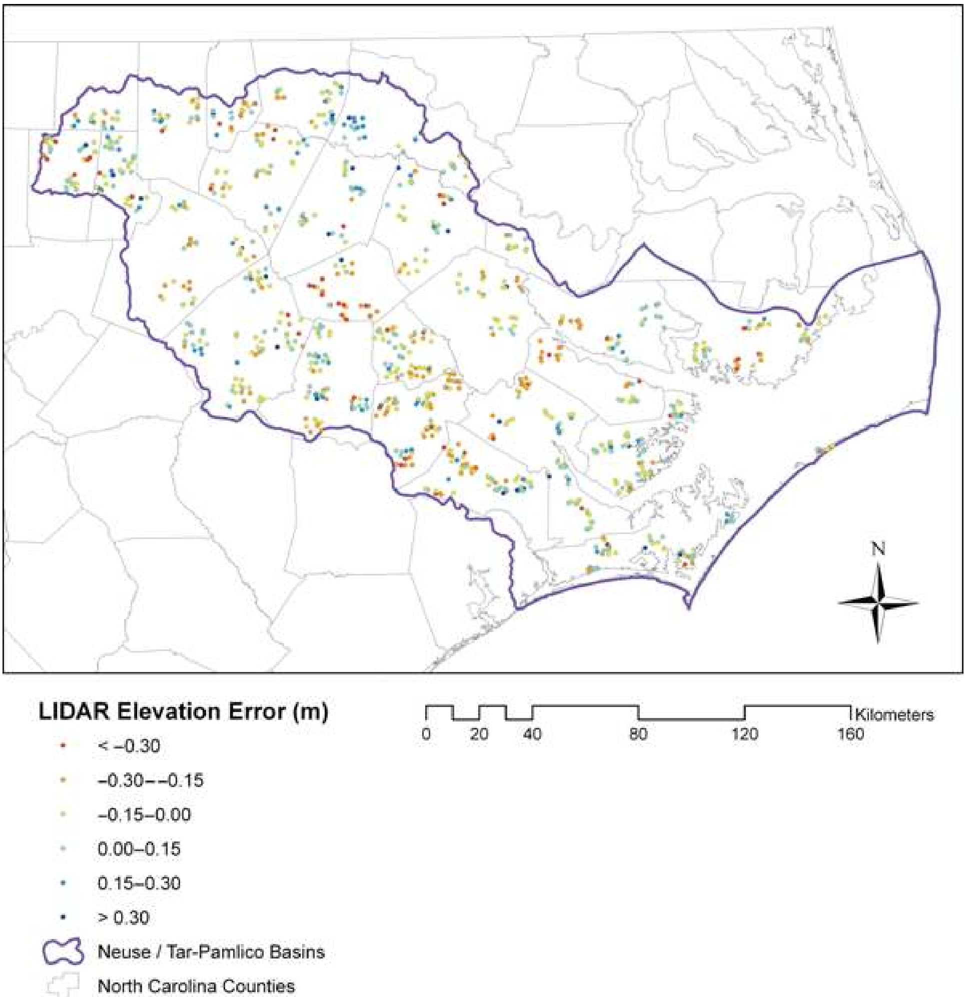

Figure 4 Location of survey points and error magnitude of LIDAR elevation data (n = 1,000)

3.2 Normality Testing

The descriptive summaries so far have provided some initial indication that several of the error distributions are not normal. This is more fully described using measures of normality (skewness and kurtosis), normality testing (Lilliefors and Shapiro-Wilk) and distribution plots (histogram and Q-Q plots). Table 2 reports the results for the measures of normality and normality testing, and Figures 5 through 9 show the distribution plots for each of the four spatial data types considered.

GPS. The distribution plots of the positional error of GPS locations are shown in Figure 5. The distribution of the horizontal error is somewhat skewed with a tail towards the right. As reported in Table 2, values for skewness and kurtosis exceed the standard error for these parameters and confirm the lack of symmetry and strong clustering. Normality tests confirm the distribution is not normal. Log-transformation of this

Table 2 Results of normality testing of positional error distribution for four types of spatial data (n = 1,000)

	GPS				Geocoding		Roads		LIDAR Orig.		LIDAR Abs.	
	Orig.	X	Y	Log10	Orig.	Log10	Orig.	Log10	Orig.	Log10	Orig.	Log10
Normal distribution												
Skewness[1]	0.812	0.157	–0.075	–0.882	5.48	–0.081	5.823	–0.290	–0.134	–	5.995	–0.952
Kurtosis[2]	0.411	0.365	0.134	1.042	53.264	–0.111	69.955	0.757	27.271	–	61.786	1.577
Tests of normality												
Lilliefors[3]	0.083	0.040	0.038*	0.063	0.228	0.058	0.170	0.035	0.072	–	0.185	0.080
Shapiro-Wilk[3]	0.953	0.993	0.997*	0.957	0.594	0.987	0.642	0.991	0.833	–	0.612	0.948

[1] Standard Error for Skewness: 0.077
[2] Standard Error for Kurtosis: 0.155
[3] All tests of normality significant at $P < 0.001$ (rejecting the hypothesis of a normal distribution), with the exception of the marked results for GPS-Y
* $P > 0.001$

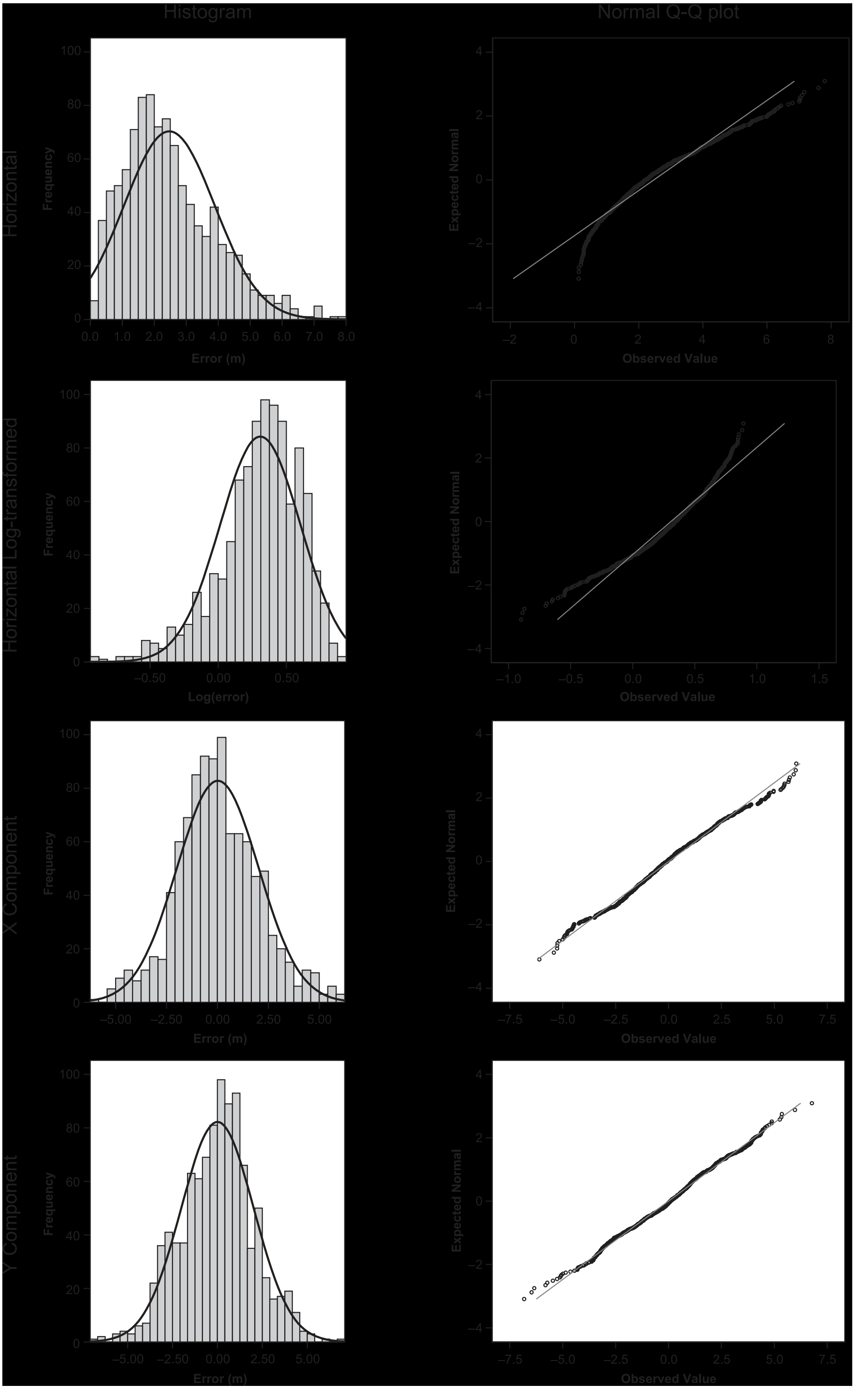

Figure 5 Distribution of positional errors in GPS locations (n = 1,000)

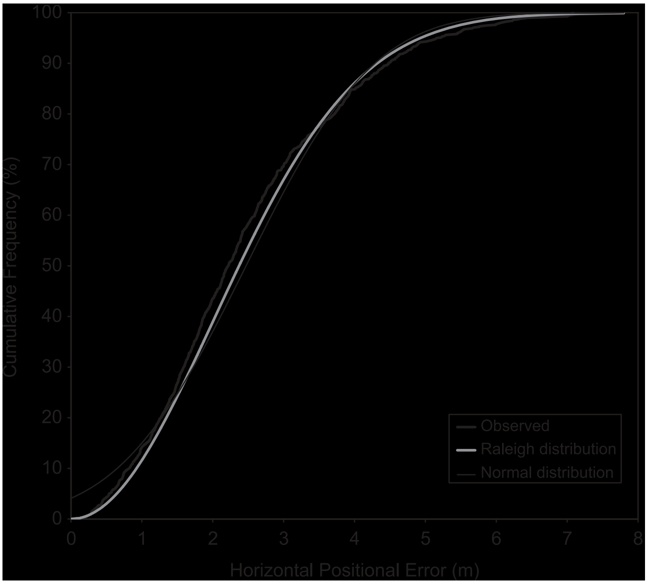

Figure 6 Cumulative distribution function of positional errors in GPS locations (n = 1,000) compared to normal and Raleigh distributions

distribution does not result in any improvement: values for skewness and kurtosis are further from normal than before. Distribution plots in Figure 5 confirm that log-transformations do not provide a more meaningful characterization.

When considering the X and Y components of GPS errors separately, both distributions are very close to normal. The distributions are very symmetrical, with low values for skewness and kurtosis, although only the values for the Y component are lower than the standard error. The histograms and Q-Q plots for both X and Y components visually suggest the distributions are normal, but formal testing only confirms that the distribution of the Y component is normal. Comparing the histograms of the X and Y components with the horizontal error (X and Y combined) reveals an interesting pattern. While the distributions for the X and Y component are symmetrical with a peak around zero, the distribution of the horizontal error is skewed with a peak around 2 m. The implication of this is that the true position is much more likely to be 1 to 2 m away from a GPS position fix than it is to be 0 to 1 m away. The reason for this is that although the probability of a position fix being within any unit area falls off with distance from the true position, the circumference at that distance gets larger (meaning there is more area at that distance), which increases the probability of the true position being at that distance. These opposite effects on the probability play against each other and yield the observed histogram.

Since normal distributions for the X and Y component were expected, the horizontal error (X and Y combined) was expected to follow the Rayleigh distribution. This is explored in Figure 6, which plots the cumulative distribution function of the observed error compared to normal and Rayleigh distributions with the same standard deviation. The Rayleigh distribution indeed approximates the observed distribution much closer, in particular for low positional errors of < 1 m. The differences between the observed and Rayleigh distributions can be attributed to the (small) difference in the distributions of the X and Y components and the (small) deviation from normality of the X component.

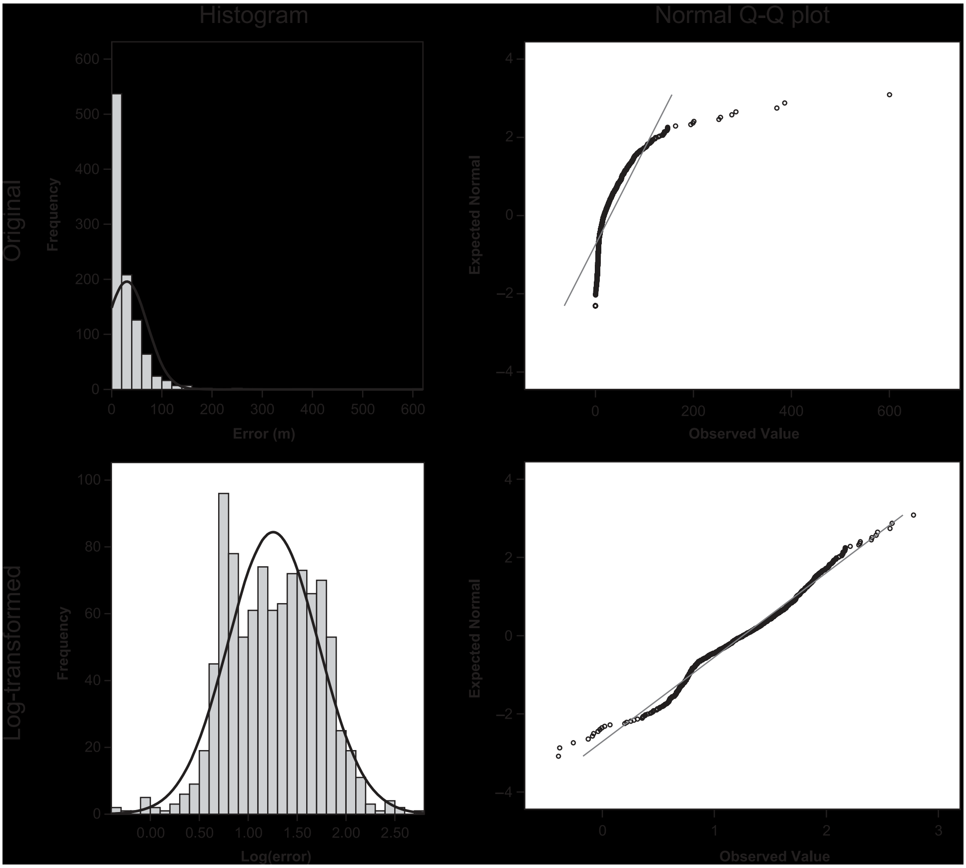

Figure 7 Distribution of positional errors in steet geocoding (n = 1,000)

Geocoding. The distribution plots of the positional error of street geocoding are shown in Figure 7. The distribution of the errors is highly skewed with a long tail towards the right. As reported in Table 2, very high values for skewness and kurtosis confirm the lack of symmetry and strong clustering. Normality tests confirm the distribution is not normal. Log-transformation of this distribution results in a substantial improvement: values for skewness and kurtosis are slightly negative and very close to the standard error for these parameters. Distribution plots in Figure 7 confirm that log-transformation results in a distribution that is much closer to normal, but some evidence of non-normal behavior is observed, such as the occurrence of a major peak to the left of the mean. Normality tests confirm the error distribution of street geocoding is not log-normal, although the general shape of the distribution approximates a log-normal distribution.

TIGER roads. The distribution plots of the positional error of the TIGER roads are shown in Figure 8. The distribution of the error is highly skewed with a long tail towards the right. As reported in Table 2, very high values for skewness and kurtosis confirm the lack of symmetry and strong clustering. Normality tests confirm the distribution is not normal. Log-transformation of this distribution results in a substantial

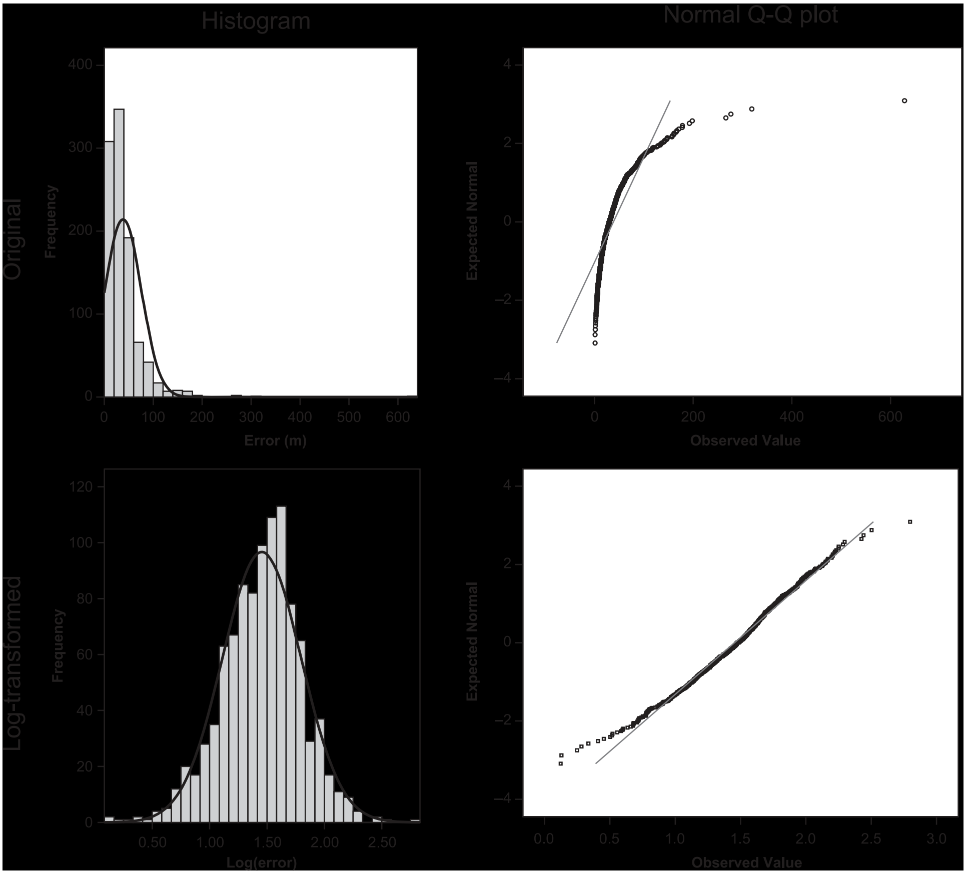

Figure 8 Distribution of positional errors in TIGER roads (n = 1,000)

improvement: values for skewness and kurtosis are smaller but are still not close to the standard error for these parameters. Distribution plots in Figure 8 confirm that log-transformation results in a distribution that is much closer to normal, but some evidence of non-normal behavior is observed, such as a slightly asymmetrical peak and a tail at the left side of the distribution. Normality tests confirm the error distribution of street geocoding is not log-normal, although the general shape of the distribution approximates a log-normal distribution.

LIDAR. The distribution plots of the positional error of LIDAR data are shown in Figure 9. The distribution of the original error values is very symmetrical with a low value for skewness. Kurtosis, however, is very high as a result of very long tails at both ends. Normality tests confirm the distribution is not normal. The distribution plots reveal an interesting pattern, which is most clearly observed in the Q-Q plot; the distribution follows the normal curve very closely with the exception of a very small number of outliers on both ends of the distribution. This confirms the occurrence of non-stationary behavior in elevation data reported by other studies (Bonin and Roussseaux 2005, Oksanen and Sarjakoski 2005).

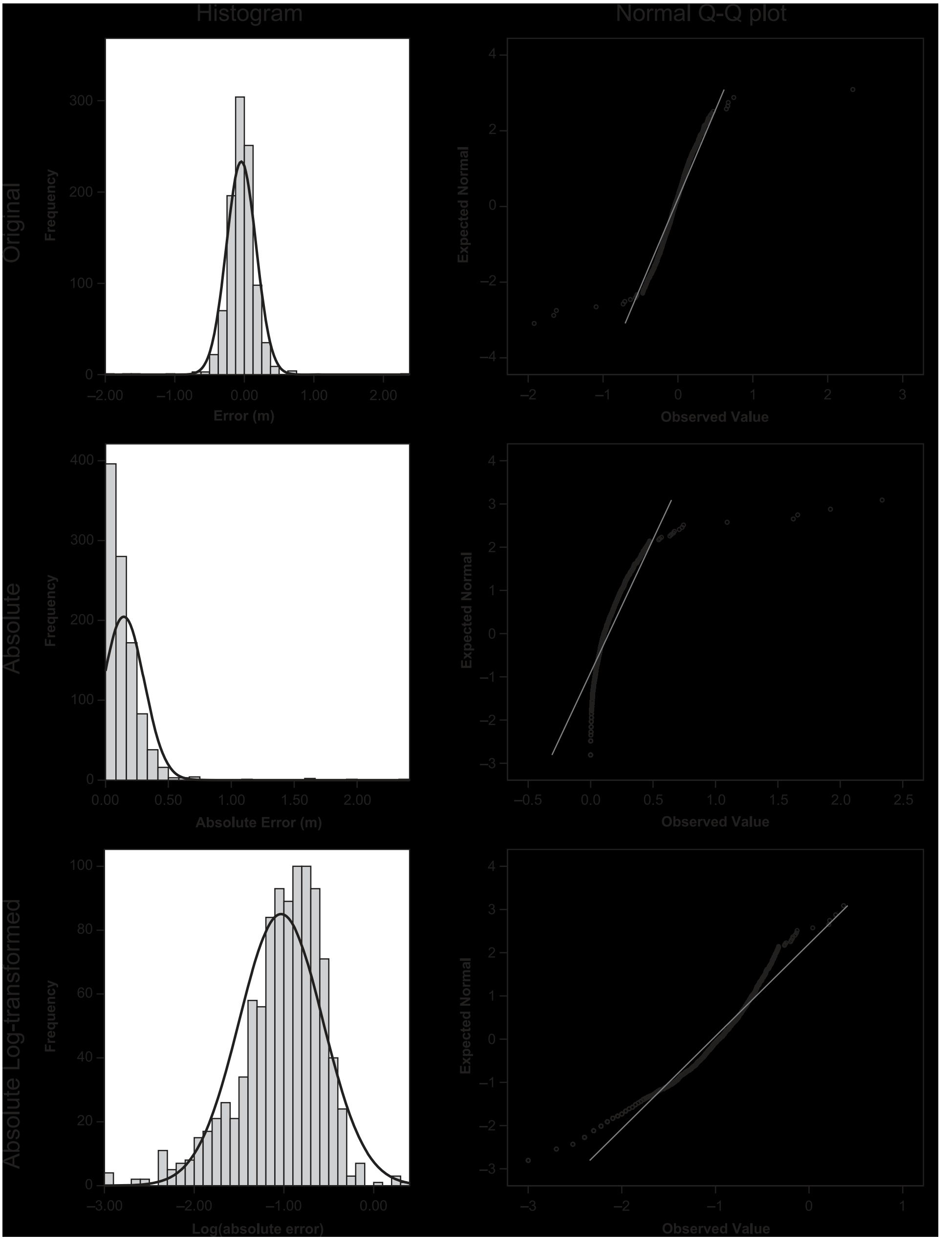

Figure 9 Distribution of positional errors in LIDAR data (n = 1,000)

It should be noted that the sampling design of the original LIDAR accuracy assessment was not random. Sampling locations were clustered in selected areas to facilitate the field work and stratified across land cover types: 20% bare earth and low grass, 20% high grass, weeds and crops, 20% brush lands and low trees, 20% urban areas and 40% forested. Separate analysis (not reported here) has revealed non-normal behavior for all land cover types.

Table 3 Positional accuracy statistics for entire sample of 1,000 observations for four spatial data types

Error statistic	GPS (m)	Geocoding (m)	Roads (m)	LIDAR (m)
90th Percentile	4.478	68.145	74.734	0.293
95th Percentile	5.158	90.337	95.623	0.372
99th Percentile	6.439	194.596	167.045	0.663
90% RMSE	2.395	27.245	33.986	0.135
95% RMSE	2.574	31.925	38.370	0.151
100% RMSE	2.848	50.821	53.521	0.218
NSSDA	4.930	87.961	92.635	0.428

When considering the absolute error values, the distribution becomes highly asymmetrical with a long tail on the right. Values for skewness and kurtosis reported in Table 2 are much higher than for the original values and normality tests confirm the distribution is not normal. The distribution plots in Figure 9 reveal strong deviation from normal behavior.

Log-transformation of this distribution results in some improvement: values for skewness and kurtosis are smaller but are not close to the standard error for these parameters. Distribution plots in Figure 9 confirm that log-transformation results in a distribution that is closer to normal, but the distribution is clearly asymmetrical. Normality tests confirm the error distribution of the absolute values of LIDAR errors is not log-normal.

3.3 Evaluating Error Statistics

Table 3 reports the positional accuracy statistics for the four spatial data types. These statistics are based on the complete sample of 1,000 observations for each data type, employing the NSSDA methodology, i.e. using horizontal error for GPS, geocoding and TIGER roads (not X and Y components separately) and using absolute vertical error for LIDAR elevation data. The percentiles and RMSE values are determined from the error distributions, while the NSSDA accuracy statistic is determined by multiplying the horizontal or vertical 100% RMSE values with 1.7308 or 1.9600, respectively.

A comparison of the accuracy statistics reveals a number of interesting patterns. First, the percentiles are higher than their corresponding RMSE values for every one of the four distributions, i.e. the 90th percentile is higher than the 90% RMSE value and the 95th percentile is higher than the 95% RMSE value. This reflects the fact that the RMSE statistic considers all data values within the portion of the distribution considered, not just the maximum value. What is most noteworthy is that the difference between the percentiles and RMSE value is much larger for the skewed distributions. For example, the 95th percentile for the GPS error distribution is roughly twice the 95% RMSE value, while the 95th percentiles for the geocoding and TIGER roads are roughly three times the 95% RMSE values. For the skewed distributions the relative difference

Table 4 Robustness of positional error statistics based on Monte Carlo simulation of 100 samples of 20 observations for four spatial data types

	Avg/Original Ratio	Rank	SDev/Avg Ratio	Rank
GPS				
90th Percentile	0.928	4	0.151	4
95th Percentile	0.927	5	0.156	5
90% RMSE	0.992	2	0.133	3
95% RMSE	0.996	1	0.126	2
100% RMSE	0.983	3	0.120	1
Geocoding				
90th Percentile	0.880	3	0.284	3
95th Percentile	0.857	5	0.344	4
90% RMSE	0.992	2	0.253	1
95% RMSE	1.002	1	0.254	2
100% RMSE	0.870	4	0.433	5
Roads				
90th Percentile	0.881	4	0.258	3
95th Percentile	0.882	5	0.342	5
90% RMSE	1.013	1	0.167	1
95% RMSE	1.014	2	0.197	2
100% RMSE	0.926	3	0.307	4
LIDAR				
90th Percentile	0.905	3	0.242	3
95th Percentile	0.860	5	0.285	4
90% RMSE	0.989	2	0.216	2
95% RMSE	0.991	1	0.215	1
100% RMSE	0.900	4	0.474	5

between the 90th and 95th percentiles is also much larger than between the 90% and 95% RMSE values. This result is significant since it shows that a consistent relationship between RMSE and percentiles cannot be assumed unless the distribution has been empirically tested.

A second interesting comparison can be made between the calculated values for the NSSDA statistic and the 95th percentiles. The NSSDA statistics is determined from the 100% RMSE value and is assumed to be identical to the 95th percentile. Table 3 shows this is clearly not the case for the four datasets considered. For GPS, geocoding and TIGER roads, the NSSDA statistic underestimates the 95th percentile by 4.4, 2.6 and 3.1%, respectively. For LIDAR, the NSSDA statistic overestimates the 95th percentile by 15.0%.

The results of testing the robustness of the accuracy statistics using Monte Carlo simulation are reported in Table 4. For each of the five accuracy statistics, the average and standard deviation for 100 simulations of 20 samples are determined. The average is compared to the original values reported in Table 3 and reported as the ratio between average and original value. The closer this ratio is to a value of 1, the more robust the accuracy statistic. The standard deviation is compared to the average and reported as the ratio between standard deviation and average. The lower the value of this ratio is, the more robust the accuracy statistic. The values of the two ratios were ranked from 1 to 5 within each of the four spatial data types considered.

Results indicate a number of consistent patterns. First, for the GPS errors, which most closely approximate a normal distribution of the four spatial data types considered, the three different RMSE statistics are similarly robust. The 95% RMSE is most robust in terms of the average/original ratio and the 100% RMSE is most robust in terms of the standard deviation/average ratio, but overall the ratios are very similar. Second, for the three other spatial data types, the 90% and 95% RMSE statistics emerge as the most robust, consistently ranking either 1st or 2nd. The 100% RMSE statistic does not perform well for these distributions, ranking 3rd only once, and 4th or 5th in all other cases. Third, the 90th and 95th percentiles do not perform well for any of the distributions; the highest rank for the 90th percentile is 3rd (5 out of 8 comparisons) and the highest rank for the 95th percentile is 4th (2 out of 8 comparisons).

Overall, the results presented in Table 4 suggest that the traditional 100% RMSE used in the NSSDA protocol is robust only if the positional error in spatial data is (close to) normally distributed. In this case, "data trimming" by removing the largest 5 to 10% of the data is not necessary. If data is not normally distributed, the 100% RMSE does not perform well, and some amount of data trimming is recommended to produce more robust accuracy statistics such as the 90% and 95% RMSE. Percentiles, while useful to describe the distribution, are not very robust as single descriptors.

4 Discussion and Conclusions

This research has presented strong evidence that the positional error distributions of several different spatial data types are not normally distributed. Characterizing positional errors, therefore, may require different approaches for different types of spatial data, based on an understanding of the underlying processes which cause positional errors. A single statistic, such as the horizontal or vertical RMSE value, appears to be insufficient to properly characterize the nature of the positional error of many spatial datasets.

The positional error in GPS locations can be approximated with a Rayleigh distribution, which assumes that the X and Y components are normally distributed and independent. The positional error in street geocoding and TIGER roads can be approximated with a log-normal distribution. The positional error in LIDAR elevation data can be approximated with a normal distribution of the original (positive and negative) vertical error values after removal of a small number of outliers. For all four data types considered, however, these solutions are only approximations, and some evidence of non-stationary behavior resulting in lack of normality was observed in all four datasets.

The characterization of positional errors requires a more thorough understanding of the processes causing them. For example, in the case of GPS, accuracy is affected by factors such as poor satellite geometry, receiver noise, ionospheric disturbance, obstacles, canopy, and multi-path errors, all of which can contribute to non-stationary behavior in the results. In the case of street geocoding, non-stationary behavior can be introduced by variability across urban-rural gradients, differences in parcel sizes, and accuracy of the street reference network. For TIGER roads, non-stationary behavior can be introduced by the merging of datasets from various sources and time periods, and the completion of updates by different contractors. For LIDAR elevation data, non-stationary behavior can be introduced by variability in accuracy by land cover and terrain complexity, as well as effects of signal processing and data cleaning.

One of the limitations of this research is the effect of the accuracy of the reference data. While in all four cases the positional accuracy of the reference data is at least one order of magnitude better than the data being evaluated, improved accuracy of reference data may provide more robust results. A second limitation is the effect of non-random sampling. In the case of the LIDAR accuracy assessment the reference locations are stratified by land cover type, and a closer investigation on non-normal behavior by land cover type is warranted. In the case of the TIGER roads accuracy assessment the reference locations were determined by looking at intersections and this may not represent the accuracy of the entire line segments.

The results of this research reveal some very significant recommendations for the use of spatial data accuracy standards, in particular the NSSDA:

1. The NSSDA should consider using alternative approaches to characterize the positional error in spatial data; the emphasis on the use of a single accuracy statistic seems overly simplistic considering the variability and complexity of error distributions of common spatial data types.
2. The assumption of normality of positional error should be reconsidered since there is strong evidence to suggest that many spatial data types are not normally distributed. This would also include revisiting the assumption that the 95th percentile can be reliably determined from the 100% RMSE.
3. A broader view on the use of accuracy statistics should be embraced which allows for proper characterization of a range of different distributions. This should include the use of separate X and Y components for horizontal accuracy and the use of the original (positive and negative) values for the vertical accuracy of elevation data.
4. A minimum sample size of 20 locations seems insufficient for data types for which the underlying distribution is not well established considering the variability and complexity of many positional error distributions.
5. Revised techniques need to be developed to characterize commonly observed distributions, such as the Rayleigh and log-normal distributions.
6. When using a single accuracy statistic for comparative purposes, some amount of data trimming is recommend to improve the robustness of the statistic: measures like 90% and 95% RMSE are preferred over the traditional 100% RMSE, while percentiles are the least robust when used as single descriptors.
7. Accuracy testing procedures for line and area features need to be implemented.

The non-normal distribution of positional errors in spatial data has implications beyond spatial data accuracy standards since most error propagation techniques for spatial data are also based on an assumption of normality. For example, the modeled error in numerical error propagation modeling of DEMs is usually a random error based on the expected standard deviation of the vertical error in the DEM. The error is modified using either an exponential (e.g. Holmes et al. 2000) or Gaussian (e.g.

Goovaerts 1997) spatial autocorrelation model. This type of error propagation modeling has become widely employed for a range of applications (Ehlschlaeger 1998, Lee et al. 1992, Veregin 1997, Lindsay and Creed 2005, Oksanen and Sarjakoski 2005) but assumes that the vertical error in a DEM follows a distribution whose dispersion can be described meaningfully with a single statistic, such as the RMSE. Given the occurrence of major errors in DEMs, alternative approaches to error propagation modeling will need to be developed.

In addition to the practical implications for spatial data accuracy standards and error propagation modeling, a better understanding of the distributions of positional error can provide insights into the underlying processes which explain the occurrence of errors in spatial data.

References

American Society for Photogrammetry and Remote Sensing (ASPRS) Specifications and Standards Committee 1990 ASPRS Accuracy standards for large-scale maps. *Photogrammetric Engineering and Remote Sensing* 56: 1068–70

American Society for Photogrammetry and Remote Sensing (ASPRS) LIDAR Committee 2004 *ASPRS Guidelines: Vertical Accuracy Reporting for LIDAR Data*. Bethesda, MD, American Society for Photogrammetry and Remote Sensing

Bel Hadj Ali A 2002 Moment representation of polygons for the assessment of their shape quality. *Journal of Geographical Systems* 4: 209–32

Beyer W 1966 *CRC Handbook of Tables for Probability and Statistics*. Cleveland, OH, The Chemical Rubber Co

Bolstad P V, Gessler P, and Lillesand T M 1990 Positional uncertainty in manually digitized map data. *International Journal of Geographic Information Systems* 4: 399–412

Bolstad P, Jenks A, Berkin J, and Horne H 2005 A comparison of autonomous, WAAS real-time and post-processed Global Positioning Systems (GPS) accuracies in northern forests. *Northern Journal of Applied Forestry* 22: 5–11

Bonin O and Rousseaux F 2005 Digital terrain model computation from contour lines: How to derive quality information from artifact analysis. *GeoInformatica* 9: 253–68

Bonner M R, Han D, Nie J, Rogerson P, Vena J E, and Freudenheim J L 2003 Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology* 14: 408–12

Cayo M R and Talbot T O Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics* 2: 10

Chrisman N R 1991 The error component in spatial data. In Maguire D J, Goodchild M F, and Rhind R W (eds) *Geographical Information Systems: Principals and Applications*. New York, John Wiley and Sons: 165–74

Dearwent S M, Jacobs R J, and Halbert J B 2001 Locational uncertainty in georeferencing public health datasets. *Journal of Exposure Analysis and Environmental Epidemiology* 11: 329–34

Duckham M and Drummond J 2000 Assessment of error in digital vector data using fractal geometry. *International Journal of Geographical Information Science* 14: 67–84

Ehlschlaeger C R 1998 The Stochastic Simulation Approach: Tools for Representing Spatial Application Uncertainty. Unpublished Ph.D. Dissertation, University of California, Santa Barbara

Federal Geographic Data Committee 1998 *Geospatial Positioning Accuracy Standards: Part 3, National Standard for Spatial Data Accuracy*. Washington, D.C., Federal Geographic Data Committee Report No STD-007.3-1998

Fisher P 1998 Improving modeling of elevation error with geostatistics. *GeoInformatica* 2: 215–33

Folks J 1981 *Ideas of Statistics*. New York, John Wiley and Sons

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

Goovaerts P 1997 *Geostatistics for Natural Resources Evaluation*. New York, Oxford University Press

Greenwalt C and Shultz M 1968 *Principles of Error Theory and Cartographic Applications*. St. Louis, MO, Aeronautical Chart and Information Center

Harter L 1960 Circular error probabilities. *Journal of the American Statistical Association* 55: 723–31

Hodgson M E and Bresnahan P 2004 Accuracy of airborne LIDAR-derived elevation: Empirical assessment and error budget. *Photogrammetric Engineering and Remote Sensing* 70: 331–9

Hodgson M E, Jensen J, Raber G, Tulils J, Davis B A, Thompson G, and Schuckman K 2005 An evaluation of LIDAR-derived elevations and terrain slope in leaf-off conditions. *Photogrammetric Engineering and Remote Sensing* 71: 817–23

Holmes K W, Chadwick O A, and Kyriakidis P C 2000 Error in USGS 30-meter digital elevation model and its impact on terrain modeling. *Journal of Hydrology* 233: 154–73

Karimi H A and Durcik M 2004 Evaluation of uncertainties associated with geocoding techniques. *Computer-Aided Civil and Infrastructure Engineering* 19: 170–85

Lee J, Snyder P K, and Fisher P E 1992 Modeling the effect of data error on feature extraction from digital elevation models. *Photogrammetric Engineering and Remote Sensing* 58: 1461–7

Leung Y and Yan J 1998 A locational error model for spatial features. *International Journal of Geographical Information Science* 12: 607–20

Lindsay J B and Creed I F 2005 Sensitivity of digital landscapes to artifact depressions in remotely-sensed DEMs. *Photogrammetric Engineering and Remote Sensing* 71: 1029–36

Lilliefors H 1967 On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association* 62: 399–402

López C 2000 Improving the elevation accuracy of digital elevation models: A comparison of some error detection procedures. *Transactions in GIS* 4: 43–64

Maling D H 1989 *Measurements from Maps: Principles and Methods of Cartometry*. New York, Pergamon

McCollum J M 2003 Map error and root mean square. In *Proceedings of the Towson University GIS Symposium*, Baltimore, Maryland

Minnesota Planning 1999 *Positional Accuracy Handbook: Using the National Standard for Spatial Data Accuracy to Measure and Report Geographic Data Quality*. St. Paul, MN, Minnesota Planning Land Management Information Center

Munkres J 1999 *Topology* (Second Edition). Englewood Cliffs, NJ, Prentice Hall

NDEP 2003 *Proposed Changes to the Geospatial Positioning Accuracy Standards: Part 3, National Standard for Spatial Data Accuracy*. Reston, VA, United States Geological Survey, National Digital Elevation Program Proposal (submitted to the Federal Geographic Data Committee)

NDEP 2004 *Guidelines for Digital Elevation Data (Version 1.0)*. Reston, VA, United States Geological Survey, National Digital Elevation Program

North Carolina Flood Mapping Program 2005 *Quality Control of Light Detection and Ranging (LIDAR) elevation data in North Carolina for Phase III of NCFMP*. Raleigh, NC, North Carolina Division of Emergency Management, North Carolina Flood Mapping Program

North Carolina Flood Mapping Program 2006 *Issues Papers of the North Carolina Cooperating Technical State Flood Mapping Program*. Raleigh, NC, North Carolina Division of Emergency Management, North Carolina Flood Mapping Program

O'Grady K and Goodwin L 2000 *The Positional Accuracy of MAF/TIGER*. Washington, D.C., U.S. Census Bureau, Geography Division

Oksanen J and Sarjakoski T 2005 Error propagation analysis of DEM-based drainage basin delineation. *International Journal of Remote Sensing* 26: 3085–102

Oksanen J and Sarjakoski T 2006 Uncovering the statistical and spatial characteristics of fine toposcale DEM error. *International Journal of Geographic Information Science* 20: 345–69

Papoulis A 1984 *Probability, Random Variables, and Stochastic Processes* (Second Edition). New York, McGraw-Hill

Parkinson B W 1996 GPS error analysis. In Parkinson B W and Spiler Jr J J (eds) *Global Positioning Systems: Theory and Applications* (Volume 1). Washington, D.C., American Institute of Aeronautics

Ratcliffe J H 2001 On the accuracy of TIGER-type geocoded address data in relation to cadastral and census areal units. *International Journal of Geographic Information Science* 15: 473–85

Shapiro S S and Wilk M B 1965 An analysis of variance test for normality (complete samples). *Biometrika* 52: 591–9

Shi W and Liu W 2000 A stochastic process-based model for the positional error of line segments in GIS. *International Journal of Geographical Information Science* 14: 51–66

Trainor T 2003 US Census Bureau geographic support: A response to changing technology and improved data. *Cartography and Geographic Information Science* 30: 217–23

Trimble 2002 *GeoExplorer CE Series Equipment Specifications*. Sunnyvale, CA, Trimble Navigation

Tveite H and Langaas S 1999 An accuracy assessment method for geographical line data sets based on buffering. *International Journal of Geographical Information Science* 13: 27–47

United States Bureau of the Budget 1947 *United States National Map Accuracy Standards*. Washington, D.C., U.S. Bureau of the Budget

United States Census Bureau 2006 *MAF/TIGER Accuracy Improvement Project*. Washington, D.C., U.S. Census Bureau, Geography Division

United States Forest Service 2003 *GPS Data Accuracy Standard: Draft*. Washington, D.C., U.S. Forest Service

United States Geological Survey 1999 *Map Accuracy Standards*. Washington, D.C., U.S. Geological Survey Fact Sheet FS-171-99

Van Niel T G and McVicar T R 2002 Experimental evaluation of positional accuracy estimates from a linear network using point- and line-based testing methods. *International Journal of Geographical Information Science* 16: 455–73

Vauglin F 1997 Modèles statistiques des imprécisions géométriques des objets géographiques linéaires. Unpublished Ph.D. Dissertation, University of Marne-La-Vallée, France

Veregin H V 1997 The effects of vertical error in digital elevation models on the determination of flow-path direction. *Cartography and Geographic Information Systems* 24: 67–79

Veregin H V 2005 Data quality parameters. In Longley P A, Goodchild M F, Maguire D J, and Rhind D W (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications* (Second Edition). Hoboken, NJ, John Wiley and Sons: 177–89

Vonderohe A P and Chriman N R 1985 Tests to establish the quality of digital cartographic data: Some example from the Dane County Land Records Project. In *Proceedings of Auto-Carto 7*, Washington, D.C.: 552–9

Ward M H, Nuckols J R, Giglierano J, Bonner M R, Wolter C, Airola M, Mix W, Colt J, and Hartge P 2005 Positional accuracy of two methods of geocoding. *Epidemiology* 16: 542–7

Whitsel E A, Rose K M, Wood J L, Henley A C, Liao D, and Heiss G 2004 Accuracy and repeatability of commercial geocoding. *American Journal of Epidemiology* 160: 1023–9

Wilson D 2006 GPS Horizontal Positional Accuracy. WWW document, http://users.erols.com/dlwilson/gpsacc.htm

Wing M G, Eklund A, and Kellogg L D 2005 Consumer-grade Global Positioning Systems (GPS) accuracy and reliability. *Journal of Forestry* 103: 169–73

Wu J, Funk T, Lurman F W, and Winer A M 2005 Improving spatial accuracy of roadway networks and geocoded addresses. *Transactions in GIS* 9: 585–601

Zhang J and Kirby R P 2000 A geostatistical approach to modeling positional errors in vector data. *Transactions in GIS* 4: 145–59