Accuracy Evaluation and Sensitivity Analysis of Estimating 3D Road Centerline Length using Lidar and NED

Hubo Cai and William Rasdorf

Abstract
Highway networks are represented by linear spatial objects (road segments). Having accurate length information of road centerlines is critical in transportation. This paper presents a geographic information system (GIS)-based approach that overlays planimetric road centerlines and elevation data to model road centerlines in a 3D space and estimate their lengths. Elevation sources included light detection and ranging (lidar) and the National Elevation Dataset (NED). The estimated distances were compared to distance measurement instrument (DMI)-measured distances to evaluate the accuracy. The effects of elevation datasets with varying vertical accuracies were assessed. The relationship between road geometric properties and the accuracy of distance estimates was examined. We found that (a) the proposed 3D approach is efficient in estimating 3D road centerline distances, (b) using lidar point data improves the accuracy by 28 percent over the use of NED, and (c) certain road geometric properties have direct relationship with the accuracy of distance estimates.

Introduction
Highway networks are represented by a rich set of linear spatial objects (road segments). These geo-spatial transportation segments are connected to form topological networks that provide the basis for location referencing systems to locate features such as bridges, signs, pavement conditions, and traffic accidents, and can be used to more accurately measure over-the-road travel distances between geographic locations (FGDC, 1994). Furthermore, these topological networks can be used to find shortest paths, to determine efficient routes, and to estimate traffic volumes when combined with the variety of network analysis tools.

Transportation data is usually referenced to road networks by using a one-dimensional (1D) linear referencing model (Quirroga, 1999). With this model, objects along a network are located using a set of known points on the network and distances and directions from the known points to the objects along the linear objects. All linear referencing methods are based on this concept (Baker and Blessing, 1994) which makes distance critical in transportation spatial databases.

Examining various existing transportation spatial databases reveals that there is no correct and consistent distance information (Rasdorf et al., 2003). The integration of linear referencing data with different linear referencing methods and the integration of linear referencing data with other datasets (global positioning system (GPS) data, for example) reveal that errors vary widely across databases. There is a need to obtain accurate and consistent distance information and to use this information to update existing transportation spatial databases. This is not a trivial task taking into consideration the number of roads and the mileages of these roads in the United States (Rasdorf et al., 2004).

A number of techniques are available to obtain distances along linear objects. Distance can be extracted from construction and design drawings. This approach suffers from the problems of currency and completeness. A distance measurement instrument (DMI) and an inertial navigation system (INS) might be used to measure the distance directly along lines. The DMI itself is an inexpensive technology consisting of a mechanical device attached to one or more of the wheels of a van and connected to an in-vehicle recorder (Karimi et al., 2000). However, a DMI requires frequent calibration to minimize errors that accumulate as the result of travel. An INS uses accelerometers and gyroscopes to provide pitch, roll, and heading information to derive a relative geo-referencing for a point. An INS has high accuracy over small distances and is usually more accurate than GPS (depending on types of GPS) but much more costly (Karimi et al., 2000). Ground surveying is another technology that can be used to obtain distances along linear objects with a high accuracy when working with straight lines only. GPS is a very promising technique for measuring linear objects; however, the most notable difficulties in using this technology stem from signal blockage in certain areas caused by thick tree canopies, bridges, or multi-path problems caused by signal deflection by high-rise buildings.

In addition to the difficulties and limitations mentioned above, all these approaches suffer from a common disadvantage: they are very time-consuming to use to obtain distance information for a large number of roads.

Hubo Cai is with the Department of Civil and Construction Engineering, Western Michigan University, 1903 West Michigan Avenue, Kalamazoo, MI 49009 (hubo.cai@wmich.edu).

William Rasdorf is with the Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC 27695 (rasdorf@eos.ncsu.edu).
On the other hand, the application of the 3D spatial modeling approach proposed herein for determining distances provides an efficient way to obtain distance information without field efforts (Quiroga, 1999; Rasdorf et al., 2003; Rasdorf et al., 2004). Such an approach could be used to represent road centerlines in a more accurate way, to obtain road centerline distances, and to correct and update distance information in existing transportation spatial databases. The concern is accuracy. Before this 3D technique can be successfully applied in modeling linear objects in a three-dimensional space and determining their lengths, its accuracy in overlaying planimetric road centerline data with a three-dimensional space is needed to locate a point on a road segment, leading to a one-dimensional model of three-dimensional linear objects.

This paper presents the results of a research study that focused on estimating 3D road centerline lengths using lidar data and evaluating their accuracy by comparing them to reference data, i.e., DMI data in a case study. Four elevation datasets were used: a lidar point cloud, a lidar 20-foot digital elevation model (DEM), a lidar 50-foot DEM, and an NED dataset. The sensitivity of the accuracy to the use of these different elevation datasets with varying vertical accuracies was also analyzed. In addition, the effects on the accuracy from geometric properties of road centerlines were evaluated and reported herein.

Methodology
This section briefly describes the 3D point models used in this research. Based on the chosen model, 3D distance determination for lines is proposed. In addition, model construction when working with elevation datasets in different formats, is discussed.

3D Point Model
The key of 3D spatial modeling is the treatment of the third dimension, i.e., elevation in a similar way as the other two planimetric dimensions (Li, 1994). From the 2D aspect, a line object is constructed by connecting two or more vertices. The line segment connecting two neighboring vertices could be a straight line, a circular line, or a polyline depending on the actual characteristics of the line object being constructed and the relative relationship among vertices. Together, the point locations and the connecting line segments compose the geometry of the line under construction.

A natural extension to this principle is to represent a 3D linear object by connecting two or more 3D vertices as illustrated in Figure 1; 3D vertices are points with coordinates specified in a 3D space. The line segment connecting two neighboring 3D vertices could be a straight line, a circular line, or a polyline.

In Figure 1, line segments connecting these vertices can be described using mathematical functions in the format of \( Z = f(X, Y) \) and \( Y = f(X) \), where \( X, Y, \) and \( Z \) are the three dimensions in 3D GIS and the combination of the two functions together specifies a 3D line, not a 3D surface in a three-dimensional space.

In a 2D GIS, the location of a point is specified in the context of a coordinate system. In Linear Referencing Systems (LRS), the location of a point along a road segment is specified by a distance along the segment referring to the start point. In other words, only one measure (the distance) is needed to locate a point on a road segment, leading to a one-dimensional model of three-dimensional linear objects.

Taking a similar approach, a 3D vertex of a road centerline can be located in a three-dimensional space using the distance measure (planimetric distance) and the elevation. This leads to a variation of the point model to represent road segments in a 3D space as illustrated in Figure 2.

With the variant point model proposed above, the lines connecting neighboring 3D vertices could be described by a function in the format of \( Z = f(D) \), where \( Z \) is elevation, and \( D \) is 2D distance along the line referring to the start point. This variant can be simplified by assuming that all line segments connecting neighboring vertices are straight lines. No mathematical function is necessary to describe the connecting line segments, even though it can be easily derived. This variation requires sufficient number of vertices in order to achieve a high accuracy (Cai, 2009). Generally, more 3D points indicate a closer approximation to the original linear object.

3D Distance Determination
The variation point model described in the previous section may not be desirable in visualization, but it is very convenient for specific 3D analyses such as distance determination. The distance determination method in this research was based on this simplified variant, which specifies that a 3D line consists of 3D vertices. In addition to its \( X, Y, Z \) coordinates, each 3D vertex has an associated distance that is the planimetric distance along the line from the start point of the line to this vertex. The distance of the straight line segment connecting two neighboring vertices was calculated as the square root of the sum of the square of the 2D distances between these two vertices and the square of the elevation difference of these two vertices. The 2D distance used in this calculation is the difference of the planimetric distances associated with these two vertices. Adding distances of all straight line segments connecting neighboring vertices of a 3D line leads to the 3D distance of this 3D line.

![Figure 1. Line construction in a 3D GIS.](image1.png)

![Figure 2. An LRS-based variation of the 3D Point model (Modified after Cai and Rasdorf, 2008).](image2.png)
Compared to calculating the planimetric distance between two neighboring vertices using their X-Y coordinates directly, using the difference in the planimetric distances along the line from two neighboring vertices to the start point as the planimetric distance between two neighboring vertices is more accurate. This is because it takes into consideration the planimetric curvature of the linear objects and is more accurate in 3D distance determination.

Model Construction
The critical part in constructing the proposed 3D point model for lines and carrying out 3D distance determination is the introduction of the third dimension, or the elevation. More specifically, the concern herein is how to analyze 3D vertices along road centerlines. Two different approaches were developed in this research to work with elevation datasets in different formats: a snapping approach for working with lidar point clouds and a uniform interval and bilinear interpolation approach for working with digital elevation models (DEMs).

Buffering and Snapping Approach
The buffering and snapping approach for working with point elevation data was developed based on road geometry and point density (Cai and Rasdorf, 2008). Point elevation datasets, such as lidar point clouds, have relatively high density. It indicates that many of points would be on the road pavement surface, even though not many points would lie exactly on the road centerline. Roads are man-made objects with a relatively flat surface (small cross-sectional slopes) and consequently, these points on the pavement surface have their elevations very close to the elevations of nearby points along road centerlines and could be used to obtain elevations for 3D points along road centerlines.

For a given road centerline, a buffer with a size of a half of the road surface width or smaller can be applied to identify elevation points that are on the road surface. Each identified elevation point can then be snapped to the nearest point on the road centerline. These two steps lead to a series of 3D points along the road centerline.

Uniform Interval and Bilinear Interpolation Approach
The uniform interval and bilinear interpolation approach used in this research study was designed for obtaining 3D vertices along lines when working with DEMs (Rasdorf et al., 2004). The uniform interval part identifies the planimetric positions of 3D vertices on the line while the bilinear interpolation approach introduces elevations to these vertices based on their planimetric positions and the DEMs.

The uniform interval to be used was determined based on the resolution of the DEM data being used. As a general rule, the optimum interval should be the same as the resolution which is often represented by one dimension of the cell (for example, 30 meters) of the DEM in order to achieve a reasonable accuracy and efficiency.

Case Study
A transportation case study was designed to implement the developed 3D approach to estimate 3D distances for road centerlines, using elevation datasets in different formats with varying vertical accuracies, more specifically, a lidar point cloud, lidar DEMs, and NED. The accuracy of the estimated 3D lengths was evaluated by comparing them to DMI data. The sensitivity of the accuracy of distance prediction to varying vertical accuracies of elevation datasets was evaluated. In addition, the effects on the accuracy from road centerline geometric properties calculated from the resulting 3D vertices were assessed.

Study Scope
The scope of this case study was determined based on the availability of lidar data in North Carolina at the time of conducting this research, while considering sample size and variety. The scope of this case study included portions of Interstate highways in nine counties (Durham, Granville, Halifax, Johnston, Nash, Orange, Vance, Wake, and Wilson), and US and NC routes in Johnston County.

Data Sources
This case study involved three major datasets: planimetric road centerline data, elevation data for 3D distance determination (point cloud and DEM), and DMI data for accuracy evaluation.

The planimetric road centerline data were provided by the GIS unit of North Carolina Department of Transportation (NCDOT). It was digitized from 1993 grayscale digital orthophoto quarter-quadrangles (DOQQs) and 1998 color-infrared (CIR) DOQQs with a ground resolution of one-meter, which were georeferenced to State Plane coordinates, using a horizontal datum of North American Datum of 1983 (NAD83).

The elevation data in this case study included lidar data and NED. Lidar data is further categorized into lidar point clouds, lidar DEMs with a 20-foot resolution, and lidar DEMs with a 50-foot resolution. They were obtained from the North Carolina Floodplain Mapping Program (www.ncfloodmaps.com). The obtained lidar data is bare earth data free from vegetation, buildings, and other manmade structures, in the format of ASCII files (NCCTSMP, 2003). The horizontal datum is NAD83 and the projection is the Lambert Conformal Conic projection. The coordinate system is State Plane (North Carolina). The vertical datum is North American Vertical Datum of 1988 (NAVD88). North Carolina specified a root mean square error (RMSE) of 20 cm for coastal counties and a RMSE of 25 cm for inland counties (NCCTSMP, 2001).

The NED is produced by the U.S. Geological Survey (USGS) by merging the highest-resolution, best-quality elevation data available across the United States into a seamless raster format (USGS, 2003a). It is the result of the USGS effort to provide a 1:24 000 scale Digital Elevation Model (DEM) data for the conterminous US and 1:63 360 scale DEM data for Alaska (USGS, 2003b). In this case study, NED data was obtained from North Carolina State University Spatial Information Lab (http://www.precisionag.ncsu.edu) in the format of interchange files (.e00 files). These interchange files were converted into raster grid files. The horizontal datum is NAD83, and the vertical datum is NAVD88. The projection is the Lambert Conformal Conic projection. The coordinate system is State Plane (North Carolina). The grid spacing of the converted raster data is one-arc-second (approximately 30 meters (98 feet)).

In this case study, the reference data, or DMI data, was collected by researchers for all road segments in the study scope through field work. The DMI data used to collect measurements is NITESTAR® Model NS-60 from Nu-Metrics, Inc. This model has a precision of one foot/mile (repeatability) in measuring distances (Nu-Metrics 1998). The calibration was done by NCDOT personnel working in the Road Inventory Section of the GIS Unit of NCDOT.

Data Processing
Based on the previous discussion, 3D distances were determined for each road segment. When working with lidar points, the buffering and snapping approach was used with a buffer at the size of the typical lane width of a particular road segment to ensure that only lidar points on road surface were included. When working with a particular DEM dataset, two intervals were taken: one at the full cell size
of the DEM and the other at the half cell size of the DEM. Figure 3 illustrates a constructed 3D road centerline model using lidar point data, in which the vertical scale was exaggerated to show elevation changes. In addition, the DMI distance for each road segment was obtained.

As a result, each road segment is associated with eight distances, one from using lidar point data, one from using lidar 20-foot DEMs with a 10-foot interval, one from using lidar 20-foot DEMs with a 20-foot interval, one from using lidar 50-foot DEMs with a 25-foot interval, one from using lidar 50-foot DEMs with a 50-foot interval, one from using NED with a 15-meter interval, one from using NED with a 30-meter interval, and the reference distance (DMI measured distance). All combinations of road classes and road centerline distances from different sources were investigated in this study.

Accuracy Evaluation and Sensitivity Analysis
This section presents the methods used to evaluate the accuracy of the estimated 3D distance and the sensitivity of the accuracy to elevation datasets with varying format, resolution, and accuracy. Evaluation results are also presented and discussed herein.

General Information
As stated earlier, each road segment in this case study was associated with seven datasets 3D distances and one reference distance (DMI distance). For a particular road segment, error was defined as the difference between an estimated 3D distance and the reference distance. Proportional error was defined as the error divided by the reference distance. Road segments were categorized based on road types, resulting into four groups: All Road, Interstate, US routes, and NC routes road segments. The group of All Road included Interstate, US routes, and NC routes road segments. There were a total of 255 samples of errors (differences) and a total of 28 samples of proportional errors (proportional differences).

To deal with the problem of misalignment between field county boundary signs and the digital county lines, neighboring road segments touching the same county boundary were combined into one road segment and were counted once in accuracy evaluation.

Accuracy Evaluation and Sensitivity Analysis Results
RMSE is a commonly-used method for spatial accuracy assessment (Daniel and Tennant, 2001). The most commonly used RMSEs are 100% RMSEs calculated from all checkpoints and 95% RMSEs calculated from the “best” 95% of the checkpoints that have errors smaller than the errors of the remaining 5%. Equation 1 illustrates how RMSE can be calculated:

$$ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2}; $$

where $e_1, e_2, e_3, \ldots, e_n$ are errors.

Table 1 summarizes 100% and 95% RMSEs in this case study. The difference has a unit of foot while the proportional difference has a unit of foot/1,000 feet. A careful review of the discarded 5% “worst” road segments revealed that the majority were road segments that had either underpass or overpass point(s) as the end(s). This indicates that when adopting the proposed 3D approach to estimate 3D road centerline distances, the practitioners must pay extra attention to road segments falling into the scenario of having underpasses or overpasses point(s) as end(s). The sensitivity of the accuracy to elevation datasets with varying formats, resolutions, and vertical accuracies was evaluated by comparing these RMSEs. The trend patterns observed from 95% RMSEs and 100% RMSEs were similar and consequently, only graphical comparisons of 95% RMSEs from the aspects of the difference and proportional difference are presented in Figure 4 and Figure 5, respectively. The shaded area of Table 1 comprises the graphs of Figure 4. The latitudinal axis of Figure 4 lists the seven data sets and intervals under consideration. In both Figures 4 and 5, straight lines are used to link discrete data points to illustrate trends, if they exist. Since the horizontal axis represents the use of different elevation data sets, it is meaningless to try to interpolate between these discrete data points.

Based on Figure 4, it is observed that for all road segments (diamond), the sample using the lidar point data (LP) has the lowest RMSE (the highest accuracy). The samples using lidar DEMs are higher than the RMSE of the sample using lidar point data, but lower than the RMSEs of the samples using NED. Thus, it would simply appear that lidar point cloud data inputs greater accuracy than lidar DEMs and they in turn input greater accuracy than NED data sets.

For Interstate road segments and US road segments, the trend lines are close to the trend line of all road segments. The exception is the trend line for NC road segments. The RMSE of the sample using lidar point data is higher than the others. However, these RMSEs are too close to each other to reach a meaningful observation of the trend.

Based on Figure 5, it is observed that the trend lines for All Road segments, Interstate Road segments, US Road segments, and NC Road segments are almost parallel to each other. The RMSEs of the samples of a particular road type (but using different elevation datasets and intervals) are close to each other. When all roads are considered, the sample using lidar point data has the lowest RMSE. The lowest RMSEs correspond to the samples using lidar point data for Interstate Road segments and US Road segments and the samples using lidar 20-foot DEMs for NC Road segments.

In all cases, the RMSEs of the samples using NED are imperceptibly higher than the corresponding RMSEs of the samples using lidar point data. It is worth pointing out herein that by removing the “worst” 5% of the data, the improvements on the RMSEs for proportional differences are significant and more significant than those for differences.

Overall, it is observed that when using the same DEM with different interval sizes, the difference between the RMSE values are very small, indicating that the effect from using different intervals on the accuracy of the estimated 3D distance could be ignored, given these intervals were no larger than the cell
Table 1. Summary of RMSEs

<table>
<thead>
<tr>
<th>RMSE</th>
<th>Error Format</th>
<th>Road Type</th>
<th>Lidar Point Data</th>
<th>Lidar 20-ft DEM</th>
<th>Lidar 50-ft DEM</th>
<th>NED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All Road segments</td>
<td></td>
<td>10-ft Interval</td>
<td>20-ft Interval</td>
<td>25-ft Interval</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
<td></td>
<td>25.65</td>
<td>33.30</td>
<td>33.73</td>
</tr>
<tr>
<td>100%</td>
<td>Difference</td>
<td>Interstate Road segments</td>
<td></td>
<td>24.76</td>
<td>33.38</td>
<td>33.99</td>
</tr>
<tr>
<td>RMSE</td>
<td>Difference</td>
<td>US Road segments</td>
<td></td>
<td>28.26</td>
<td>37.39</td>
<td>37.62</td>
</tr>
<tr>
<td>Proportional Difference</td>
<td>Difference</td>
<td>All Road segments</td>
<td></td>
<td>50.44</td>
<td>51.28</td>
<td>51.26</td>
</tr>
<tr>
<td>95%</td>
<td>Proportional Difference</td>
<td>Interstate Road segments</td>
<td></td>
<td>32.06</td>
<td>33.16</td>
<td>33.12</td>
</tr>
<tr>
<td>RMSE</td>
<td>Proportional Difference</td>
<td>US Road segments</td>
<td></td>
<td>63.72</td>
<td>64.99</td>
<td>64.98</td>
</tr>
<tr>
<td>Proportional Difference</td>
<td>Proportional Difference</td>
<td>NC Road segments</td>
<td></td>
<td>78.80</td>
<td>78.74</td>
<td>78.74</td>
</tr>
<tr>
<td>95%</td>
<td>Proportional Difference</td>
<td>All Road segments</td>
<td></td>
<td>22.48</td>
<td>27.65</td>
<td>28.01</td>
</tr>
<tr>
<td>Proportional Difference</td>
<td>Proportional Difference</td>
<td>Interstate Road segments</td>
<td></td>
<td>21.39</td>
<td>27.95</td>
<td>28.53</td>
</tr>
<tr>
<td>Difference</td>
<td>Proportional Difference</td>
<td>US Road segments</td>
<td></td>
<td>25.28</td>
<td>31.31</td>
<td>31.42</td>
</tr>
<tr>
<td>95%</td>
<td>Proportional Difference</td>
<td>NC Road segments</td>
<td></td>
<td>21.32</td>
<td>18.56</td>
<td>18.31</td>
</tr>
<tr>
<td>RMSE</td>
<td>Proportional Difference</td>
<td>All Road segments</td>
<td></td>
<td>24.90</td>
<td>26.36</td>
<td>26.33</td>
</tr>
<tr>
<td>Proportional Difference</td>
<td>Proportional Difference</td>
<td>Interstate Road segments</td>
<td></td>
<td>18.19</td>
<td>19.76</td>
<td>19.75</td>
</tr>
<tr>
<td>Difference</td>
<td>Proportional Difference</td>
<td>US Road segments</td>
<td></td>
<td>35.47</td>
<td>37.73</td>
<td>37.72</td>
</tr>
<tr>
<td>95%</td>
<td>Proportional Difference</td>
<td>NC Road segments</td>
<td></td>
<td>51.93</td>
<td>51.82</td>
<td>51.82</td>
</tr>
</tbody>
</table>

Resolution. This observation is confirmed by comparing other statistics including means, medians, standard deviations, absolute means (based on error magnitude) in this case study (not included in this paper) (Cai, 2003).

Significance Analysis

The purpose of significance analysis is to evaluate the effects on the accuracy of the estimated 3D distance from certain geometric properties such as distance, vertical complexity, etc. This section describes the methods being used, the geometric properties under consideration, and evaluation results with observations.

Analysis Methods

Two methods were used to evaluate the effects from certain road centerline geometric properties on the accuracy of the estimated distance. The first method was the calculation of sample correlation coefficient of two variables; one was the geometric property under consideration and the other was the error/proportional error in the estimated 3D distances. The value of the sample correlation coefficient ranges from $-1$ to $1$. It could be understood that the closer to $1$ the value of the sample correlation coefficient is, the stronger the linear association is. If the value of the sample correlation coefficient is 1, variables $X$ and $Y$ are said to be perfectly correlated (Rao, 1998). The sign of the sample correlation coefficient determines if there is a positive or a negative linear association between two variables.

The second method was the grouping and comparison method. With this method, rather than grouping roads into different road types, roads were grouped based on the value

Figure 4. Comparisons of 95% RMSEs from the aspect of the difference.

Figure 5. Comparisons of 95% RMSEs from the aspect of proportional difference.
of a geometric property under evaluation. RMSEs for these groups were obtained and compared. Given a particular set of estimated 3D distances determined by the use of elevation dataset and the model parameter, the variation of the RMSEs among the groups could be evaluated and the effects of the factor under consideration on the accuracy of the estimated 3D distances could be determined.

**Geometric Properties under Evaluation**

Geometric properties under evaluation in this research included the actual distance, the average slope and slope change, and the weighted slope and slope change. The distance was defined as the DMI measured distance. The geometric properties related to slopes and slope changes describe the vertical complexity. In this research, most linear objects are relatively long and therefore, it is unreasonable to simply take the start and end points for slope calculation. Rather, two slopes were obtained for each line: the average slope and the weighted average slope. The weight is 3D distance.

Figure 6 illustrates the calculations of average and weighted slopes and slope changes with a simplified example.

Assuming there is a line that consists of three vertices connected by two straight line segments LN1 and LN2, with slopes S1 and S2, respectively:

1. The average slope is calculated as \((S_1 + S_2)/2\).
2. The weighted slope is calculated as \((S_1 * D_1 + S_2 * D_2)/(D_1 + D_2)\), where \(D_1\) is the 3D distance of LN1, and \(D_2\) is the 3D distance of LN2.
3. The average slope change is calculated as \(((S_1 - 0) + (S_2 - S_1))/2\).
4. The weighted slope change is calculated as \(((S_1 - 0) * D_1 + (S_2 - S_1) * D_2)/(D_1 + D_2)\).

**Significance Analysis Results**

This subsection presents the results from significance analysis by using two methods: the sample correlation coefficient approach and the grouping and comparison approach.

1. **Sample Correlation Coefficients**

Table 2 summarizes the results of the sample correlation coefficients based on the difference, the absolute difference (error magnitude), the proportional difference, and the absolute proportional difference (the magnitude of the proportional error) for the geometric properties under evaluation in this research. In Table 2, road types are not under consideration, i.e., only the results for the all road segments are presented. In the original study, road types were under consideration. The observations about Interstate, US, and NC road segments are similar to those of all road segments, and therefore, are not repeated herein (Cai, 2003).

It is observed that from the aspect of the proportional difference, there is no linear association between a geometric property under consideration and the accuracy of the estimated 3D distance, as indicated by the close to 0 sample correlation coefficients. However, when the error magnitudes (the absolute difference and the absolute proportional

<table>
<thead>
<tr>
<th>Geometric Property</th>
<th>Error Format</th>
<th>Lidar Point Data</th>
<th>Lidar 20-ft DEM 10-ft Interval</th>
<th>Lidar 20-ft DEM 20-ft Interval</th>
<th>Lidar 20-ft DEM 25-ft Interval</th>
<th>Lidar 20-ft DEM 50-ft Interval</th>
<th>Lidar 20-ft DEM 15-m Interval</th>
<th>Lidar 20-ft DEM 30-m Interval</th>
<th>NED 15-m Interval</th>
<th>NED 30-m Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference</td>
<td>-0.07</td>
<td>-0.34</td>
<td>-0.36</td>
<td>0.16</td>
<td>0.06</td>
<td>-0.35</td>
<td>-0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Slope</td>
<td>-0.10</td>
<td>0.36</td>
<td>0.37</td>
<td>0.35</td>
<td>0.34</td>
<td>0.40</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted Slope</td>
<td>-0.22</td>
<td>-0.33</td>
<td>-0.33</td>
<td>-0.15</td>
<td>-0.20</td>
<td>-0.34</td>
<td>-0.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Slope</td>
<td>0.24</td>
<td>0.36</td>
<td>0.36</td>
<td>0.19</td>
<td>0.25</td>
<td>0.36</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted Slope</td>
<td>0.00</td>
<td>0.23</td>
<td>0.23</td>
<td>-0.23</td>
<td>-0.22</td>
<td>-0.22</td>
<td>-0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Slope</td>
<td>-0.25</td>
<td>-0.31</td>
<td>-0.31</td>
<td>-0.14</td>
<td>-0.19</td>
<td>-0.32</td>
<td>-0.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted Slope</td>
<td>0.19</td>
<td>0.33</td>
<td>0.33</td>
<td>0.17</td>
<td>0.23</td>
<td>0.34</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Slope</td>
<td>-0.21</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.06</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted Slope</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>-0.20</td>
<td>-0.19</td>
<td>-0.19</td>
<td>-0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Slope</td>
<td>-0.24</td>
<td>-0.36</td>
<td>-0.36</td>
<td>-0.15</td>
<td>-0.21</td>
<td>-0.37</td>
<td>-0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted Slope</td>
<td>0.26</td>
<td>0.39</td>
<td>0.39</td>
<td>0.22</td>
<td>0.29</td>
<td>0.40</td>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Slope</td>
<td>0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted Slope</td>
<td>-0.30</td>
<td>-0.29</td>
<td>-0.29</td>
<td>-0.28</td>
<td>-0.28</td>
<td>-0.27</td>
<td>-0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Slope</td>
<td>-0.25</td>
<td>-0.39</td>
<td>-0.39</td>
<td>-0.17</td>
<td>-0.23</td>
<td>-0.39</td>
<td>-0.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted Slope</td>
<td>0.27</td>
<td>0.41</td>
<td>0.41</td>
<td>0.23</td>
<td>0.29</td>
<td>0.42</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Slope</td>
<td>0.00</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted Slope</td>
<td>-0.30</td>
<td>-0.28</td>
<td>-0.28</td>
<td>-0.28</td>
<td>-0.28</td>
<td>-0.27</td>
<td>-0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
mediate RMSEs.

having intermediate distances are the groups having inter-

and 8 graphically illustrate the comparisons of these RMSEs

and the proportional difference of these groups. Figures 7

indicated that groups having longer distances are associated

together. The general trend for these seven trend lines

elevation datasets and model parameters are clustered

corresponding to the use of seven different combinations of

estimated 3D distance. In other words, the steeper the

negative linear association with the proportional error of the

slope has a positive linear association with the error but a

of the estimated 3D distance but a negative linear association

from the aspect of the proportional error.

The exact meaning of the positive or negative linear

association is demonstrated with an example: the average

slope has a positive linear association with the error but a

negative linear association with the proportional error of the

estimated 3D distance. In other words, the steeper the

average slope of the road segment is, the larger the error it

has and the smaller the proportional error it has in its

estimated 3D distance.

2. Grouping and Comparison

Table 3 illustrates the designed groups based on the distance. Group 1 was used to include all very short Road segments. Group 7 was used to include all very long Road segments. The symbol “(” means “greater than”. The symbol “)” means “less than”. The symbol “]” means “less than or equal to”.

Table 4 summarizes the RMSEs based on the difference and the proportional difference of these groups. Figures 7 and 8 graphically illustrate the comparisons of these RMSEs from the aspect of the difference and the proportional difference, respectively.

It is observed from Figure 7 that the general trend of the RMSEs indicates that the longer the distance, the larger the RMSE, when using a particular combination of an elevation dataset and an interval size. The two groups having the shortest distances are the two groups having the smallest RMSEs. The two groups having the longest distances are the two groups having the largest RMSEs. The other three groups having intermediate distances are the groups having intermediate RMSEs.

From Figure 8, it is obvious that the seven trend lines corresponding to the use of seven different combinations of elevation datasets and model parameters are clustered together. The general trend for these seven trend lines indicated that groups having longer distances are associated with smaller RMSEs of the proportional difference while groups having shorter distances are associated with larger RMSEs of the proportional difference.

These observations confirm the linear associations observed based on sample correlation coefficients. The same analysis was conducted for the factors of average slope, weighted average slope, average slope change, and weighted slope change. Similar observations were obtained, which confirmed the linear associations observed from the sample correlation coefficients. In other words, there is a general positive linear association between a factor under consideration and the error and there is a general negative linear association between a factor under consideration and the proportional error. With the aid of these observations, road segments with particular geometric properties could be identified as those that request additional attention when adopting the proposed 3D approach to estimate 3D distances for road centerlines.

Conclusions and Discussion

This paper observes that the accuracy of the calculated 3D distances varies with the use of different elevation datasets. From the aspect of differences (error magnitude) using the 100% RMSE as the measure of the accuracy, the use of lidar point data in this case study improved the accuracy by about 28 percent compared to the use of NED data. The use of lidar DEMs improved the accuracy by about 6 percent compared to the use of NED data. Based upon the 95% RMSE, the use of lidar point data in this case study improved the accuracy by about 25 percent compared to the use of NED data. The use of lidar DEMs improved the accuracy by about 9 percent compared to the use of NED data.

Table 4. Summary of RMSEs for Groups Based on the Distance

<table>
<thead>
<tr>
<th>Error Format</th>
<th>Group Name</th>
<th>Distance Range (ft)</th>
<th>Number of Road Segments</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference</td>
<td>Group 1</td>
<td>(0, 100)</td>
<td>46</td>
<td>17.36%</td>
</tr>
<tr>
<td></td>
<td>Group 2</td>
<td>(100, 1,000)</td>
<td>26</td>
<td>9.81%</td>
</tr>
<tr>
<td></td>
<td>Group 3</td>
<td>(1,000, 5,000)</td>
<td>52</td>
<td>19.62%</td>
</tr>
<tr>
<td></td>
<td>Group 4</td>
<td>(5,000, 10,000)</td>
<td>28</td>
<td>10.57%</td>
</tr>
<tr>
<td></td>
<td>Group 5</td>
<td>(10,000, 20,000)</td>
<td>38</td>
<td>14.34%</td>
</tr>
<tr>
<td></td>
<td>Group 6</td>
<td>(20,000, 30,000)</td>
<td>32</td>
<td>12.08%</td>
</tr>
<tr>
<td></td>
<td>Group 7</td>
<td>(30,000, +∞)</td>
<td>43</td>
<td>16.23%</td>
</tr>
<tr>
<td>Total</td>
<td>–</td>
<td>–</td>
<td>265</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 3. Illustration of Groups Based on the Distance

<table>
<thead>
<tr>
<th>Group Name</th>
<th>Distance Range (ft)</th>
<th>Number of Road Segments</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>(0, 100)</td>
<td>46</td>
<td>17.36%</td>
</tr>
<tr>
<td>Group 2</td>
<td>(100, 1,000)</td>
<td>26</td>
<td>9.81%</td>
</tr>
<tr>
<td>Group 3</td>
<td>(1,000, 5,000)</td>
<td>52</td>
<td>19.62%</td>
</tr>
<tr>
<td>Group 4</td>
<td>(5,000, 10,000)</td>
<td>28</td>
<td>10.57%</td>
</tr>
<tr>
<td>Group 5</td>
<td>(10,000, 20,000)</td>
<td>38</td>
<td>14.34%</td>
</tr>
<tr>
<td>Group 6</td>
<td>(20,000, 30,000)</td>
<td>32</td>
<td>12.08%</td>
</tr>
<tr>
<td>Group 7</td>
<td>(30,000, +∞)</td>
<td>43</td>
<td>16.23%</td>
</tr>
<tr>
<td>Total</td>
<td>–</td>
<td>265</td>
<td>100%</td>
</tr>
</tbody>
</table>
From the aspect of proportional differences (proportional error magnitude), the improvements due to the use of higher accurate elevation datasets were not as significant as those shown for differences above. This was the case because 43 percent of the road segments in the case study were longer than 10,000 feet, 53 percent were longer than 5,000 feet, and about 73 percent were longer than 1,000 feet, and longer road segments have a higher proportional accuracy. The 3D distances calculated using lidar point data with the snapping approach in this research had the highest accuracy when compared to DMI measured distances.

The observations and discussions offered above provide guidance for practitioners on selecting appropriate elevation data sets to carry out road distance estimating and updating to meet accuracy needs. The RMSE values obtained in this study also identify the accuracy range of estimated road distances using the proposed 3D approach (with different elevation data sets each having varying accuracy and precision). Thus, practitioners can quickly determine whether the use of readily available NED data meets their application accuracy needs and whether it is desirable or necessary to replace NED data with more expensive lidar data.

While the accuracy of the estimated 3D distances is dependent on the elevation dataset being used, if the same elevation dataset is used, but two different intervals are taken (one equivalent to half of the cell size and the other...
equivalent to the full cell size of the elevation data), the accuracies of the estimated 3D distances are almost the same. In other words, the accuracy of the estimated 3D distances is not dependent on the interval sizes if they are less than or equal to the cell size of the elevation data. The accuracy of the estimated 3D distances also varies with road types without discernible pattern. This leads to the recommendation that when adopting the proposed 3D approach to estimate road centerline distance using DEMs, the optimum interval ought to be taken at a value that is equivalent to the DEM cell size (resolution) to achieve the greatest accuracy while minimizing the computing time and resource consumption.

It is also concluded that the geometric properties under consideration in this study (the distance, the average and weighted slope, and the average and weighted slope change) have close relationships to the accuracy of the estimated 3D distance. The distance factors capture the vertical curvature sinusously. There is a positive association between a geometric property and the error but a negative association from the aspect of the proportional error. Thus, these geometric properties become “accuracy indicators” of the estimated 3D road centerline distance. These accuracy-indicating geometric properties provide a quick accuracy estimate of a particular estimated road distance. They also help in identifying highly vulnerable road segments (groups of road segments with a great potential for large estimate errors). For these appropriate quality control measures can be applied to improve the accuracy in road distance estimating.

Some key findings are emphasized herein:

1. The 3D modeling approach developed in this research provides an efficient way to estimate 3D distances for road centerlines. The designed case study quantified the impacts from elevation data with varying accuracy to the accuracy of road distance estimates. The case study also identified a particular group of road segments that occupy certain characteristics of road geometries and require extra caution when adopting the proposed approach to build a more accurate road centerline model.

2. The preference order of the elevation datasets to be used in modeling road centerlines as 3D lines and estimating their distances is lidar point data, lidar DEMs, and NED (from higher preference to lower preference). Ladar point data are the most preferred elevation dataset to be used.

3. Using this approach, all distances are calculated by computer. For most practical applications no DTM, ground surveying, or GPS measurement are needed to determine road centerline distance. This indicates significant savings in equipment, time, and personnel. The implications of the work, however, transcend the transportation application investigated herein. In fact, the results apply to any linear entity that can be modeled using GIS, particularly infrastructure applications including water, sewer, other forms of piping, communication and power lines, etc.

Acknowledgments

The authors are grateful to the National Science Foundation (NSF) for funding portions of this work through NSF Grant No. 9978592. The authors would also like to acknowledge the many contributions of the NCDOT to this work, particularly the personnel of the Information Technology Branch and the GIS Unit.

References


