Transportation Distance Measurement Data Quality

William Rasdorf, Ph.D., P.E.
Member, ASCE
North Carolina State University
Department of Civil Engineering
NCSU Campus Box 7908
Raleigh, NC 27695
phone: (919) 515-7637
fax: (919) 515-7908
email: rasdorf@eos.ncsu.edu

Hubo Cai
Student Member, ASCE
North Carolina State University
Department of Civil Engineering
NCSU Campus Box 7908
Raleigh, NC 27695
phone: (919) 515-6454
fax: (919) 515-7908
email: hcai@eos.ncsu.edu

Chris Tilley
North Carolina Department of Transportation
Geographic Information System Unit
Statewide Planning Branch
1587 Mail Service Center
Raleigh, NC 27699-1587
phone: (919) 715-9684
fax: (919) 212-3103
email: ctilley@dot.state.nc.us

Soren Brun, Ph.D.
North Carolina Department of Transportation
Geographic Information System Unit
Statewide Planning Branch
1587 Mail Service Center
Raleigh, NC 27699-1587
phone: (919) 250-4188 ext. 206
fax: (919) 212-3103
email: sebrun@dot.state.nc.us

Hassan Karimi, Ph.D.
University of Pittsburgh
School of Information Sciences
Department of Information Sciences and Telecommunications
135 N. Bellfield Avenue
Pittsburgh, PA 15260
phone: (412) 624-4449
fax: (919) 624-2788
email: hkarimi@sis.pitt.edu

Forrest Robson, P.E.
North Carolina Department of Transportation
Geographic Information System Unit
Statewide Planning Branch
1587 Mail Service Center
Raleigh, NC 27699-1587
phone: (919) 250-4188 ext. 204
fax: (919) 212-3103
email: frobson@dot.state.nc.us

KeyWords: accuracy assessment, CAD, data quality, descriptive statistics, DEM/DTM, distance measurement, DMI, frequency analysis, GIS, GPS, NED, photo-editing, RMSE, road layer, roadway mileage, spatial data accuracy, surface analysis, surface length, surface modeling, TIN, transportation model

OUTLINE*

* The outline is provided only for reviewer convenience to grasp the overall organization of the paper during the review.

Transportation Distance Measurement Data Quality

by
William Rasdorf
Hubo Cai
Chris Tilley
Soren Brun
Hassan Karimi
Forrest Robson

ABSTRACT

Data quality and spatial data accuracy issues are critical to any Geographic Information Systems (GIS) application, especially GIS applications in the transportation community. This paper addresses one specific aspect of spatial data accuracy issues – linear measurement (length measurement) through a transportation case study.

In the case study, an alternative to Distance Measurement Instruments (DMI) was proposed to determine road lengths for interstate highways in North Carolina. In the proposed alternative the road lengths were calculated by overlaying GIS roadway linework over elevation data (in this case the National Elevation Dataset, which was developed based on U. S. Geological Survey (USGS) 7.5 minute DEMs and calculating a centerline roadway slope distance. The results of this approach were collected and compared with DMI lengths to assess the accuracy of the proposed approach. Error sources were tentatively identified and control mechanisms were discussed.

Computer tools and models used to model surfaces and roadway linework are emphasized in this paper. The computer algorithms used for length calculations and accuracy assessment are described. This research concluded that by carefully controlling quality of both the roadway

linework data and the elevation data, GIS programs can be written to provide accurate length measurements to the transportation community. Furthermore, instrumentation like global positioning systems (GPS), high resolution cameras, and precise odometers can be combined to create productivity enhancing automated engineering systems.

INTRODUCTION

This section reinforces the need for a national infrastructure information system and suggests that it could be a useful tool in infrastructure management. Geographic Information Systems (GIS) are mentioned as one useful tool in such a system. Key issues in GIS applications, including data quality and spatial data accuracy, are raised as critical to the proper use of such a system.

Background

The nation's civil engineering infrastructure includes, among many other things, roads, bridges, and other transportation components. These infrastructure components are integrally linked to our society and our environment and must be efficiently maintained and managed. Increasingly, we are relying on computer tools and databases for information about infrastructure to enable us to better develop, use, and maintain it. Prominent among these tools are GISs which are optimized to input, store, and manipulate spatial data. However, even with the recent advances in the capabilities of GISs, one key problem is still data quality and accuracy issues (Rasdorf 2000).

Geographic Information Systems (GISs)

A GIS can be defined as a computer-based tool set for collecting, storing, retrieving, transforming and displaying spatial data from a discipline-specific domain for a particular set of purposes (Burrough and McDonnell 1998). A GIS has at least five components: people, data, hardware, software, and procedures.

Two key data items in any infrastructure information management system are (1) the location of an object and (2) the spatial relationships between objects. Spatial data are what GISs are most widely recognized as working with. Spatial (or geographical) data represent phenomena from a domain in terms of (1) their positions with respect to a known coordinate system and (2) their spatial interrelations with each other (which describe spatial properties such as connectivity). Spatial data are associated with attributes (such as thickness, cost, incidents and etc.) by means of identifiers (Burrough and McDonnell 1998). The availability and accuracy of spatial data directly affect the analysis results of GIS applications.

From a functional point of view, a GIS must be able to provide at least the following functionality with respect to spatial data.

(1) Data acquisition and verification,
(2) Data storage and database management,
(3) Data analysis,
(4) Data delivery, exchange, and presentation, and
(5) Data transformation.

GIS technology has benefited the field of transportation for many years because it is a domain that is rich in spatial data (Fletcher 1987, Nyerges and Dueker 1988, Dueker and Kjerne 1989, Ries 1993, Vonderohe *et. al.* 1993 and 1994, Dueker and Butler 1998). With the ongoing enhancements to GIS software that have occurred over the past few years, the transportation community has an unprecedented and promising opportunity to obtain, use, and distribute spatial data by applying GISs (Fletcher 2000). Thus, GISs have become critical in the transportation community.

Data Quality

Data quality is an important factor to any information system whose goal is to effectively and accurately convey information. The Federal Geographic Data Committee (FGDC) has been working on data quality standards for years. One of its outcomes, the *US Spatial Data Transfer Standard* (SDTS 1997), addresses this subject as follows.

> "Quality is an essential or distinguishing characteristic necessary for cartographic data to be fit for use."

The quality of data is one of the most important factors for information systems that are intended to conduct analysis and support decision-making. It is clearly recognized that the validity of analysis is based on the validity of the data used to perform the analysis. Any engineering data are an abstraction of the real-world phenomena they represent. It is not always the case that a phenomenon can be exactly visualized, understood, and modeled in an information system. In other words, errors and deviations are inherent in information management systems, which leads to the necessity for data quality control.

Five parameters, positional accuracy, attribute accuracy, data lineage, completeness, and consistency - are often used as parameters for assessing spatial data quality and providing accuracy information for metadata that describes the data itself and is known as data about data. Data lineage describes the original source of derived data, its derivation methods, and all of the transformations employed in producing the final data. As defined in the *US Spatial Data Transfer Standard*, completeness describes the objects represented and the abstract universe of all such objects (SDTS 1997). Consistency refers to the harmonious uniformity of agreement among parts of a database. It is the lack of apparent contradiction in the data.

In addition, data quality needs and levels can vary from person to person, organization to organization, or from application to application. A hiker, for example, may be satisfied to find themselves within 100 meters of a targeted destination whereas a surveyor may wish to be within 0.001 meters of their target. Currently, it is most often the responsibility of the users to decide if a data set is sufficient to meet their or their organization's quality requirements, and the standard may differ by application or use (Rasdorf 2000).

Spatial Data Accuracy

Errors and uncertainties are intrinsic in spatial data and need to be addressed properly (Burrough and McDonnell 1998). Spatial data accuracy is one data quality indicator that is composed of positional accuracy (whether objects are in their correct positions), geometric accuracy (e.g., size or length of linear objects), and interrelationship accuracy (topological) that describes the relationships between objects.

Errors in spatial data can occur at various stages of the overall data acquisition and transformation process. Errors and approximations in determining a geographical location might depend on surveying skills, the provision of hardware, and the choice of map projections, spheroids, and datum. Errors in the measurement of attributes depend on phenomenon variation, the accuracy of the measurement device, or observer bias and can occur during the recording of raw data. Errors can occur when data are stored in the computer due to space limitations or storage differences. Errors can also occur when deriving new attributes from existing data by applying logical or mathematical rules that are flawed or that use low quality data.

It is impossible and unnecessary to completely eliminate errors in spatial data. A certain error tolerance may be fully acceptable (recall the hiker). However, errors are by no means desirable. Knowledge about how errors and uncertainties occur would be very helpful in controlling and possibly reducing them. Finally, a thorough understanding of errors and error propagation can be used to improve our understanding of spatial patterns and processes (Burrough and McDonnell 1998).

In summary, a variety of different kinds of errors and uncertainties contribute to spatial data inaccuracy and these need to be well understood and carefully controlled in any GIS applications. They will occur even with the recent advances in technologies to more accurately acquire raw data and in improved computer hardware and GIS software which process and analyze data. It is an important goal to provide a quantified approach to give users, analysts, and decision-makers information about the quality and accuracy of the data being provided, analyzed, and used.

Work by Others

Spatial data accuracy and related issues have been previously studied. Some researchers tried to identify error sources and error ranges. Amrhein and Schut discussed the range of errors that can accompany any data sets, and made comprehensive statements about data quality that is needed by users (Amrhein and Schut 1990). Goodchild, with others, described the error and quality issues in the spatial world in a series of publications (Goodchild 1991a, b, Goodchild 1994, Goodchild and Jeansoulin 1998).

Discussions and descriptions of data quality control are also available in most GIS books and publications. For example, Burrough and McDonnell identified and described errors and uncertainties in GIS modeling and pointed out the stages where errors and uncertainties could creep in (Burrough and McDonnell 1998). Also, as mentioned earlier, errors and uncertainties are not always bad. They provide better understanding, which leads to better quality control and informed use (Burrough and McDonnell 1998).

Given a basic understanding of data quality issues in GISs, one must next evaluate, test, model and control errors and uncertainties. Beard and Buttenfield proposed graphical methods to detect and evaluate errors. A rationale for the use of these graphical methods was outlined, examples were provided, and points of research challenges were discussed (Beard and Buttenfield 1999).

As a result of these efforts, some researchers have proposed solutions for modeling spatial data accuracy. An error model was developed and tested in 1992 by Goodchild and others (Goodchild *et. al.* 1992). A new model was designed specifically for handling vector data uncertainties in GIS around 1995 (Hunter and Goodchild 1995, Hunter and Goodchild 1996). A comprehensive discussion of error modeling and management for GIS data was provided by Chapman and others (Chapman *et. al.* 1997).

Some researchers studied the propagation of errors in spatial modeling (Heuvelink 1999). Error propagation illustrates that errors in GIS are not only inherent, but they are also dynamic rather than being static. And finally, emerging from all the studies of data quality issues in GIS, a definition for formal spatial data standards has been proposed and discussed (Backe 1996).

The FGDC has been coordinating the efforts and contributions to data quality standards from governmental agencies, research institutes, and private vendors. One of the best-known standards is the Spatial Data Transfer Standard, or SDTS, which is a robust way of transferring earth-referenced spatial data between dissimilar computer systems with the potential for no information loss (FGDC-STD-002). Another example is the Geospatial Positioning Accuracy Standard that is purposed to provide consistency in reporting the accuracy of point geospatial data collected by different activities (FGDC-STD-007.3-1998).

However, most studies on errors are still at the research level (Fisher 1995, Goodchild and Gopal 1989) although systematic studies of spatial data quality have been published (Guptill and Morrison 1995). In addition, most studies related to data quality and accuracy issues serve general GIS purposes from a theoretical point of view. Presently, there are no available systematic models and algorithms for GIS users to determine data quality. This appears in transportation GIS applications as well. In this paper, a transportation case study from one state department of transportation is completely examined to provide an in-depth view of one aspect of spatial data quality and accuracy in a transportation GIS application.

MODELS AND TOOLS

This section introduces the concepts behind elevation, surface modeling, and the relationship between them. Equipment that would be used to acquire accurate distance measurements (e.g. DMI) is described. Finally, a purely computational alternative (GIS approach) to physical measurement is presented and discussed in detail.

What we eventually see is that there are numerous ways to model elevation. These various models can be incorporated into GIS software thus enhancing the functionality available to users like those in the transportation, environmental, and geotechnical communities who might benefit from it. The key objective is to understand the accuracy of the various models and to select models whose accuracy matches that needed in the application. Doing so truly provides productivity enhancement and an improved engineering process.

In this section a number of elevation models are introduced and described. These give us information to superimpose on a 2-D representation of the surface under consideration. That 2-D linework representation is also examined and issues linking elevation to 2-D representation are discussed. Finally instrumentation that provides data is considered and examined.

Surface and Surface Models

Five different ways of modeling a surface are points, contour lines, a TIN (triangulated irregular network), a DEM/DTM (digital elevation model/digital terrain model), and a mathematical model. The examples used in this section to illustrate the concepts are extracted from the paper "An Exercise in Digital Elevation/Terrain Models: From Point to Mathematics" (Massasati 2000).

Surfaces

In a GIS points, lines, and polygons are presented in a planar view. In modeling terrain surfaces, elevation values must somehow be stored throughout the planar surface to represent the actual surface. However, there are an infinite number of points that compose a surface. The only representational choice we have is to use a limited number of points with accompanying elevation values to approximate an entire surface. All surface models contain some inaccuracies, each of which depends on a number of factors (Shearer 1990). Thus, the challenge in modeling surfaces is to provide high accuracy elevations with low consumption of resources.

Mass Points and Breaklines

Mass points are irregularly spaced points. Each of them has x/y/z values, in which the x/y values identify its location on a planar coordinate system while the z value indicates its elevation. A breakline is a line feature that describes a change in the smoothness or continuity of a surface. A soft breakline ensures that known elevations along a linear feature are maintained in the final surface model. But it does not define interruptions in surface smoothness. A soft breakline is depicted with a series of x/y/z coordinates. On the other hand, a hard breakline defines interruptions in surface smoothness. It is often depicted as a series of x/y coordinates only, which have no elevations. Hard breaklines are usually used to define locations with abrupt surface changes, for example, streams, ridges, and etc. Breaklines and mass points provide input data to construct the final surface model.

Point Model

In a point model, all elevations from sampling points are captured by storing the position and elevation of every sampling point. Each point has three accompanying numeric values, X, Y, and Z. The X and Y values identify the point position in a planar (e.g. state plane grid coordinates) or geodetic (e.g. latitude and longitude) coordinate system while the Z value represents the elevation for that point. Elevations for any other points are interpolated from the elevations of sampling points. Figure 1 provides an example of point model that represents a pyramid. The dimensions of the pyramid are simplified and reduced entirely to values 750 feet at the base and 450 feet in height. The orientation is assumed to be north-south and east-west. The coordinate of the southwest corner is zero east, zero north, and zero elevation. The entire point model for this structure consists of the X, Y, and Z coordinates of 5 points.

Contour Lines

A contour is an imaginary line of constant elevation on the ground surface. If the locations of several closely spaced ground points of equal elevation are plotted on a drawing, the line joining these points is called a contour line. Thus, a terrain surface may be represented by series of contour lines with fixed interval, the vertical elevation difference between neighboring contours.

The contour line model is widely applied in surveying and mapping to represent surfaces on topographic maps like those produced by the USGS (USGS 2001d). Elevations for points along contour lines can be read directly from the elevation values represented by the contour line itself. The elevations of points between contour lines are obtained through interpolation. Figure 2 provides an example contour line model to represent the pyramid assuming a 100 foot vertical elevation interval which results in five contour lines. It must be noted that the very top of the pyramid is a single point.

DEM/DTM

DEM/DTM has been referred to as a "form of computer surface modeling which deals with the specific problems of numerically representing the surface of the earth" (Kennie and Petrie 1990).

The accuracy of DEM/DTM in representing surfaces depends on several factors, one important factor being grid resolution (Heesom and Lamine 2001). DEM/DTM is the most commonly used format to represent terrain surfaces in a cell-based data structure. In a cell-based data format, space is divided into either regular or irregular units called cells. The most common cell-based data structure has either rectangular or square cells, each being the same size. It is usually referred to as raster. Using a raster data structure model DEM/DTM stores a single elevation data value for every cell that is supposed to represent the elevation for the entire continuous cell surface. The elevations used for the surface are interpolated from the original sampling points.

DEM/DTM possesses the advantages belonging to a raster data structure. First, it is easy to understand and visualize. Next it is also in a form that is readily processable by spatial analysis. Figure 3 illustrates the DEM/DTM representing the surface of the pyramid using a 75×75 gridsize. The maximum elevation within a cell is used as the elevation for that cell for illustration purposes.

TIN (Triangulated Irregular Network)

The Triangulated Irregular Network (or TIN) was designed by Peuker and co-workers (Peuker *et. al.* 1978) for digital elevation modeling to accurately model surfaces while avoiding the data redundancies of raster format elevation models. A TIN is a *vector* format terrain model that uses a sheet of continuous, connected triangular facets (cells) that are based on a Delaunay triangulation of *irregularly* spaced nodes or observation points. Each facet stores coordinates and elevation for its three nodes. Elevations for all other points in the facet are readily derivable by interpolation from these three nodes. Figure 4 provides the surface of the pyramid represented by a TIN.

A TIN overcomes some disadvantages inherent in raster data structures, particularly the problem of data redundancy. It also solves another problem inherent in a raster representation's inability to represent areas with differing relief complexity without changing cell sizes. Thus both data redundancy and fixed cell sizes are the key problems with the raster model. It is worth pointing out that TINs and DEM/DTMs are mutually convertible with potential information loss, which is dependent on the converting algorithm.

An inherent component of TIN is the invisible mathematical model that depicts the relationship between an elevation and its X and Y coordinates analytically. For example, using the numerical values of Figures 1 and 4, each surface (of the pyramid) can be represented by the following mathematical model.

$$Z = a*X + b*Y + c$$

Where:

X, Y, and Z are surface point coordinates and

a, b, and c are constants that can be determined for each surface by solving the equation using the three corner points.

As an example of how the mathematical model works consider the calculation procedure for the surface of Triangle 4 from Figure 4 as shown in the following equations.

$$0 = a*0 + b*0 + c \qquad (1)$$

$$0 = a*750 + b*0 + c \qquad (2)$$

$$450 = a*375 + b*375 + c \qquad (3)$$

By solving the three equations one finds that b = 1.2, a = 0, and c = 0 and the resulting equation is:

$$Z = 1.2*Y, \qquad \text{where: } 0<Y<375 \text{ and } 0<Z<450$$

Likewise, equations for other triangular surfaces and lines bordering two triangular surfaces are acquired and represented in a like manner as shown in Figure 4.

Linework Models

There are many ways to represent linework using computers. One is computer aided design (CAD) software, which was developed initially for drawing and cartographic purposes and has been optimized to do so. Another is GIS software, in which linework and other features are captured as thematic maps (or coverages). The reason CAD files are still in use is that much useful linear spatial information was originally stored using CAD files. For GIS software to use the available spatial legacy data stored in CAD files (instead of repeating the data input tasks already completed during the development of those CAD files), GIS must be capable of using existing CAD files without losing important information.

CAD Files

CAD facilitates design work by providing fast, flexible, and easy to use computerized drawings and maps. With CAD, time consuming and often interrelated editing tasks are easy to complete by working on parts of a drawing instead of redrawing it entirely. Many commercial CAD programs are available today. The two most commonly used in transportation are AutoCAD and MicroStation. CAD data has no associated attributes.

In order to extract information from CAD files for GIS applications, the GIS must be capable of converting those files into a format or formats (e.g., shapefile or coverage) that provide full support to the GIS spatial analysis capabilities.

GIS Thematic Maps

In GIS, real-world phenomena are represented by thematic maps (or coverages). A thematic map is a map that displays selected kinds of information relating to specific themes, such as city and county boundaries, land use types, streets, population density, and so on. Within the concept of a thematic map layer, real-world objects may be represented as a cell, a number of cells, or they may be abstracted into point, line, or polygon features, depending on what data structure (raster or vector) is used. GIS thematic layers are geo-referenced, which locates spatial data with respect to a common real-world frame of reference to enable spatial analysis (Burrough and McDonnell 1998). In addition, each thematic layer is accompanied by an attribute table that stores descriptive characteristics for its features. In a vector data structure, each feature has one record in the attribute table. If necessary, additional tables can be linked with thematic layers.

Thematic maps are not the end products in GIS. They just store information that is necessary for analysis and decision-making. Maps, views, reports, displays, etc. can all be extracted from these thematic layers to meet user needs without changing the underlying thematic maps themselves. This characteristic also applies to CAD files.

Measurements

As noted earlier, real-world objects are explicitly represented in GISs by points, lines, or polygons in a vector data structure, depending on the characteristics of these objects, measurement scales, and their roles in analysis. Different spatial attributes describe different types of objects. Point objects use coordinates to specify their locations. Line objects use length as an essential measurement, while perimeter and area are considered to be important measurements for polygon objects.

Point objects are spatial phenomena, each of which occurs at only one location in space (DeMers 2000). Each point feature is discrete in that it can occupy only a given location in space at any time. Point objects are assumed to have no spatial dimension (no length or width). Thus, points are said to have “0” dimensionality. Points play a critical role in many of the models that were described in the last section. In the point model they were used exclusively to define the pyramid. DEM/DTM data is also originally point data. Points are also critical in TIN and mathematical models.

Linear or line objects are conceptualized as occupying only a single dimension in coordinate space. These “one-dimensional” objects may be any objects that are fundamentally long and narrow, such as roads, canals, rivers, power lines, pipelines, etc. (DeMers 2000). One very important feature in addition to locational coordinates for line features is length. This attribute allows us to measure the spatial extent of a line. Other than length, which is the major concern of this case study, shapes and orientations of linear objects could also be obtained from GISs, if needed.

Polygon objects represent objects that occupy “two-dimensional” space. Two important attributes these objects possess are perimeter and area. As with lines, we can describe a polygon’s shape and orientation, but in addition, we can also describe the amount of territory (area) a polygon occupies.

Finally, if we add a third dimension (height) to polygon features, we could represent the existing terrain as a surface as mentioned earlier. Similarly, another important attribute, volume, could be calculated, if needed, as a function of an area contained beneath a surface object.

Instruments and Tools

Points, lines, polygons, and surfaces are conceptual objects. What gives them strength and utility is their ability to aid in modeling the physical world. One key link between these concepts and the physical world is instrumentation – particularly instrumentation to measure point locations and lengths. Measuring lengths is discussed in the following section. Measuring point locations is now often done using global positioning instruments. These offer coordinate locations measured in familiar terms such as latitude and longitude.

Of concern and interest here is the accuracy of the values obtained using these instruments. It is arguable that not only must the instrument data be recorded, but an indication of its accuracy must be recorded and stored in such tools as database management and geographic information systems as well. Furthermore, when used by application programmers or others these measures of accuracy must be presented, along with the coordinate data, and must be accounted for in subsequent analysis. Not to do so is to pass on limited information that can have subsequent negative effects.

Distance Measurement Instrument

In the case study we were interested in a particular linear object, one that models road and highway segments. More specifically, we were interested in one particular attribute of those linear objects -- length. A DMI is a device that measures a linear path. It operates much as a standard odometer in an automobile but its accuracy is greater with calibrations. As noted above, the global positioning system (GPS) is a technology that is used to obtain positional data from which a length can theoretically be obtained by approximation. GPS can provide accurate locational coordinates for points along linear objects while DMI measures the spatial extent (length) of these linear objects. GPS can also calculate straight-line length between a pair of points and derive road lengths by accumulating those straight-line segment lengths. The integration of DMI and GPS tools provides a highly accurate distance measurement and verification capability for a highway network that measures as well as positions a road.

By driving cars equipped with DMI and GPS, surface length (or actual length between any two points) is directly read from DMI and positioned by GPS. This method is known to have high precision and a high accuracy with calibrations (±1 foot per mile as specified in the manufacturer's specifications of the DMI currently available to NCDOT). But these technologies are very large time and labor consumers and they are expensive. The cost to use DMI includes personnel to perform calibration and to drive cars on the road, gasoline, equipment, and others. In many other applications, including real-time emergency vehicle routing, automatic vehicle location (AVL), and monitoring of construction equipment, this combined approach is either not possible or not practical.

GIS Software

Examples of commercially available GIS software include ArcGIS from ESRI (Environmental Systems Research Institute, Inc.), MapInfo from MapInfo Corporation, and MGE (Modular GIS Environment) from Intergraph Corporation (Quiroga 1999). There is also commercially available GIS software to meet the needs of specific fields, for example, TransCAD from Caliper Corporation is designed specifically for transportation GIS applications. In addition, software developers are continuing to make improvements and provide extensions to existing programs and to create new GIS software.

Even with more recent advances in 3-D simulations, GIS still stores data in 2-D formats. In other words, 3D length measurements for linear objects are not directly available from GIS. GIS must use functions provided by GIS software to calculate surface lengths. The accuracy and precision of these calculations depend not only on the accuracy of the original data sets, but on the GIS algorithm involved in the calculation.

It must be noted that any GIS software has its own strengths and weaknesses. There is no single GIS program that is capable of providing all the spatial functionality any user might wish to have. When selecting GIS software to develop customized applications, two important factors must be taken into consideration, one is cost, the other is need. The user should select an appropriate program based on his/her own situation. For the present we are interested in evaluating the ability of existing GIS software to calculate surface length. In particular, we would like to determine its accuracy in calculating centerline road surface mileages.

GIS Surface Analysis

In this study, functions provided by ArcGIS were used to calculate surface lengths for road linework. These functions operate using two inputs, vector linework data and elevation data (either in raster *or* vector format).

Calculation of Elevations

With a grid file, the elevation for any arbitrary point on the grid is actually interpolated from elevations of the surrounding grid cells (evaluated as a point at the center of the cell) using a bilinear interpolator. An example is shown in Figure 5. On the left we see elevations for a set of points that represent the center points of the grid's cells (A, B, C, and D). On the right we see an enlarged illustration of the interpolation process for point E.

In order to interpolate the elevation value for point E, the GIS uses the elevation values of four surrounding cells (evaluated at the cell center). First, the elevation for two intermediate points (F and G) in the top-to-down direction on the paper page are calculated by linear interpolation. The elevation for the left intermediate point F is calculated between A and B as:

$$1035 + [(22.44 / 30.00) * (1041 - 1035)] = 1039.49$$

The elevation for the right intermediate point G is similarly calculated between C and D as:

$$1048 + [(22.44 / 30.00) * (1060 - 1048)] = 1056.98$$

And finally the elevation for point E is calculated using points F and G and, performing a linear interpolation in the horizontal direction as:

$$1039.49 + [(22.25 / 30.00) * (1056.98 - 1039.49)] = 1052.46$$

In a TIN model, as noted earlier, the terrain surface is represented using a network of irregular triangles. The three nodes of each triangle store elevations and coordinates. As demonstrated in both the earlier *TIN* Section, mathematical functions describing the relationship between the X and Y coordinates and elevation (Z value) can be derived from the three nodes for each triangular surface. When interpolating elevation values for a point (as was illustrated in Figure 5), the X and Y coordinates of that point are taken into consideration to decide which surface the point belongs to. The relationship between coordinates and elevation for that surface is then used to interpolate the elevation value for the new point.

In general, even though grid and TIN are two different data structures used by ArcGIS to represent surfaces, both of them actually store elevations for some points throughout the surface. Elevations for points other than these points are interpolated when necessary rather than stored. This concept makes it possible to simulate continuous surfaces with a limited number of available elevations.

Calculation of Surface Lengths

ESRI ArcGIS uses the *SURFACELENGTH* function to calculate surface lengths for linear features based on elevation data. The elevation data for this function could be raster elevation data (grid files) or vector elevation data (TINs).

When using grid elevation data to calculate surface lengths for linear paths, a sample distance along the linear path must be determined before performing calculations. This sample distance defines the points along the path for which we wish to obtain Z values. Suppose, for example, we wish to know the true 3-D road centerline distance from point A to point B in Figure 6. The polyline feature is divided into segments based on the sample distance (our segments being a-b, b-c, and c-d). Elevations for the start node and the end node (for example, nodes a and b for segment a-b) for each segment are first interpolated from grid elevation data by applying the bilinear interpolation explained earlier.

Next, the surface length calculation for each segment can be obtained by visualizing the curved line of Figure 6 as a straight line as shown in Figure 7 and using geometry to calculate a straight line surface length for each segment. In other words, a series of straight lines are used to approximate the actual roadway centerline surface and the overall length is obtained by summing surface lengths of all segments for each polyline feature. Figure 8 shows the method when used to calculate surface lengths with a sample distance of 20 meters, which coincides with the sample distance used in the case study. Our goal was to determine if this method actually produces a reliable real-world result. In other words, can we accurately use maps generated by GISs to obtain distances or must we drive the roads with DMI?

If we were to use TINs to provide elevations for the *SURFACELENGTH* function, no sample distance would be required. Input vertices for the polyline features and triangle edge intersection locations serve as the nodes to divide links into segments. Again, these segments are assumed to be straight lines to approximate the actual polyline features. Elevations for the start node and the end node of each segment are interpolated from the triangle nodes by applying the relationship functions explained earlier. Surface lengths for segments belonging to the same feature are added together to achieve the overall surface length for this feature. The differences between using grid and TIN models include the method used to divide the polyline feature into segments, the method used to represent the surface, and the method used to interpolate elevations for surface points. The underlying geometry used to calculate segment lengths and feature length is the same for both and is illustrated in Figure 8.

One important point to emphasize again is that the accuracy of the calculated surface lengths is based on the correct positions of the linework and the quality of the elevation data.

CASE STUDY

The transportation community is seeing increasing numbers of GIS applications due to its rich spatial data processing, analysis, and presentation capabilities. In spite of an enormous existing base of GIS applications in transportation, there are presently no systematic and quantified models and algorithms to provide transportation GIS users with information about spatial data accuracy and its effects on analysis results. For example, a spatial query might tell us that the length of a road from one intersection to another is 5.26 miles. But is it? Presently that 5.26 miles is often accepted on faith as being accurate. But is it true, and if it is, how accurate is it?

The context of the case study was interstate highways in North Carolina, and it focused on one transportation data quality issue – length measurement. In case study, a GIS program was developed to calculate 3-D road centerline lengths based on elevation data (rise) and on a GIS highway network (run), and then assessed the accuracy of the resulting distance measurement by comparing those GIS results with highly accurate DMI measurements. In addition to making an accuracy assessment, error sources were tentatively identified and control mechanics were discussed.

This section introduces the case study, the three major data sources, the procedure followed to perform the accuracy assessment, and the analysis results. The purpose of presenting the case study is to establish the context of the work. Only a summary is presented here. For a complete description of the case study, its results, and their implications see (Rasdorf *et. al.* 2001).

Case Study Overview

Maintaining accurate road length measurements is one of the key issues in road inventory databases. An alternative to a DMI measurement approach is to use GISs to combine elevation data with a GIS road layer to perform a surface analysis, leading to a computer generated road length (DeMers 2000). This is fast and economical.

The purpose of the case study was to determine if this method provided reasonable accuracy. For this case study, the accuracy requirements for North Carolina Department of Transportation (NCDOT) were determined according to an application needs analysis. The accuracy requirements are summarized as below.

- Errors less than or equal to 0.03 mile/mile (with respect to length) are considered to be acceptable.
- Errors less than or equal to 0.01 mile/mile (with respect to length) are considered to be good.

In addition to address the accuracy concerns from NCDOT, our purpose here is to take a broader view of the case study and its results to gain important information about data quality and accuracy issues that are a concern within the GIS community (Beard and Buttenfield 1999, Amrhein 1990, Heuvelink 1999). To do so, this case study presented the errors not only in the proportional format as stated in the accuracy requirements from NCDOT, but also in the absolute format. Statistical analyses were carried out for both formats to evaluate the overall accuracy and the impacts from road characteristics including slope and length.

Data Sources

Major data sets comprised of NED data, a photo-edited road layer, and DMI data, were used in this study. These are individually discussed below.

NED Data

NED data involved in this study were purchased from USGS in 1998. This data was developed based on 7.5-minute DEMs, which were derived from hypsographic data (contour lines) and/or photogrammetric methods using USGS 7.5-minute topographic quadrangle maps. NED data captures terrain surfaces with regular or irregular sampling points that have X, Y, and Z values. Elevations for positions other than those sampling points are interpolated from elevations of the sampling points to create a DEM file. The DEM data files are mostly based on a 30-meter by 30-meter grid spacing while some parts have 10-meter by 10-meter grid spacing.

Figure 8 shows a sample NED data file in text format and illustrates how NED data is converted into a grid file for further calculations. Both files essentially provide elevations in a raster data structure.

The first two rows shown in Figure 8 specify the number of columns and rows in the file (5 rows and 5 columns for this sample file). The xllcorner and yllcorner values specify the coordinates of the center point of the leftmost, lowest cell. The cell size indicates the 30 meter grid resolution. Following that is a set of numbers that specify the elevations for each cell in the sequence of rows first, top to bottom, and then columns, left to right. In other words, it follows the sequence of (row 1, column 1), (row 1, column 2), …, (row 1, column 5), then (row 2, column 1), (row 2, column 2), …, until it reaches (row 5, column 5).

GIS Road Layer

North Carolina DOT maintains the roadway network in sets of CAD data in the format of MicroStation Design Files (DGN) for state-maintained roads, which include all interstate highways, U.S. Routes, NC Routes, and state-maintained secondary roads. The CAD files are readily usable in GIS by geo-referencing. The results can be stored and viewed as a thematic layer in GIS. This type of data is called road layer data in this study. This road layer data has a 1:24,000 scale. In this case study, the road layer data were corrected by referencing to aerial photos (Rasdorf *et. al.* 2001).

DMI (Distance Measurement Instrument) Data

This type of data was acquired by driving cars with highly precise and accurate DMI along roads to measure their lengths (Karimi, *et. al.* 2000). DMI data is the most accurate length measurement (± 1 foot per mile with calibrations as specified in the manufacturer's specifications) presently available to NCDOT and thus it was used as the basis for comparison for this study. DMI data is available for all interstate highways in NC. It was obtained during a field measurement effort in the summer of 2000. NCDOT calibrated the DMI well in advance to the field measurement to assure its high accuracy.

Data Analysis

The overall analysis procedure included surface analysis and grouping and statistical analysis.

Surface Analysis

The goal of surface analysis is to calculate the actual centerline length of the links comprising the highway network. To do so a horizontal distance obtained from the GIS linework, a vertical distance obtained from NED data, and a GIS function to calculate a 3-D true slope distance were used. This calculated distance was then compared to a slope distance actually measured using DMI.

First, the original NED data was converted into a 30-meter by 30-meter grid file. Next, using GIS surface analysis, the grid file and the GIS road layer were combined to produce the actual 3-D length for all links based on the 2-D length of the road layer and the elevation information from the grid file.

Grouping and Statistical Analysis

Working with the cleaned road layer data set all links were grouped based on slope and length. The reason for doing so was to determine whether slope and length are significant factors influencing length measurement accuracy, and if so, what impact they might have.

First, all the links in the cleaned road layer were sorted based on SLOPE, which resulted in 5 sub categorizations as groups G1 to G5.

Group G1	0% =< SLOPE < 2%	511 Links
Group G2	2% =< SLOPE < 4%	900 Links
Group G3	4% =< SLOPE < 6%	415 Links
Group G4	6% =< SLOPE < 8%	119 Links
Group G5	8% =< SLOPE < 24%	89 Links

Second, all the links in the cleaned road layer were sorted again based on LENGTH, which resulted in another 5 sub categorizations as groups A to E.

Group A	0.00 mile =< DMI_MEASURED_LENGTH <= 0.10 mile	166 Links
Group B	0.10 mile < DMI_MEASURED_LENGTH <= 0.50 mile	565 Links
Group C	0.50 mile < DMI_MEASURED_LENGTH <= 1.00 mile	561 Links
Group D	1.00 mile < DMI_MEASURED_LENGTH <= 2.00 mile	598 Links
Group E	2.00 mile < DMI_MEASURED_LENGTH <= 5.00 mile	144 Links

Finally, frequency analysis, descriptive statistics, and RMSE analyses were performed for each group. The analyses and assessments were then summarized together to presents the results.

Analysis Results

Two error measurements are used in this study. One is the absolute error with name DIFFERENCE, which is the absolute difference between GIS calculated length and DMI length. The other is the proportional error with name DIFFERENCE_PER_MILE (DPERMILE in

Tables), which is derived by dividing the absolute error by the length. This case study assumed these errors were normally distributed. Examining the distributions of these errors validated this assumption, though the examinations were not shown in this paper. In our case study we were only interested in the magnitude of both the absolute and proportional errors. No concern about whether the length is longer than or shorter than (e.g. positive or negative error) was given in this case study. This concern will be addressed in another study that follows this one.

Frequency provides an in-depth way to view error distribution using percentages. It represents the accuracy for every group by counting the number and calculating the percentage of links within certain error ranges for each group. Table 1 summarizes the results of frequency analysis for groups based on slope. Table 2 summarizes the results of frequency analysis for groups based on length.

For each slope and length group, descriptive statistics were obtained for error, length, and slope attributes. Parameters include Mean, Variance, Standard Error, Median, Range, Confidence Level, and Count.

From a statistics point of view, RMSE is useful in describing differences between two datasets. In the spatial world, this statistic is widely used to describe positional accuracy, including horizontal and vertical accuracy (FGDC-STD-007.3-1998). However, with its ability to depict the difference between two datasets, RMSE was deemed to be a useful tool for us in evaluating the accuracy of our results. Thus, RMSE has been calculated for all groups in this study to reveal the difference between DMI length measurements and GIS length calculations. The following equation was used to calculate RMSE.

$$RMSE = [(e_1^2 + e_2^2 + e_3^2 + \ldots + e_n^2)/n]^{½}; \quad e_1, e_2, e_3, \ldots, e_n \text{ ---- errors}$$

In this study, RMSE was calculated for both absolute error (DIFFERENCE) and proportional error (DIFFERENCE_PER_MILE). While assuming the normal distribution of errors, a 95% confidence level was determined by multiplying RMSE by 1.96, which corresponds to $Z_{0.025}$ in a standard normal distribution. The 95% confidence level indicates the higher bound of the error range (or confidence interval) in which an occurrence has a 95% probability of falling. In this case study, the 95% confidence intervals all have zero (no error situation) as their lower bounds. For example, the higher bounds of the 95% confidence level for the absolute errors of group G2 is 0.033, which indicates the 95% confidence interval for the absolute errors of this group is from 0 to 0.033. Table 3 provides RMSE and 95% confidence level for all groups based on slope. Table 4 provides the same information for groups based on length.

Descriptive statistics illustrate the distribution of errors. Table 5 shows summarized descriptive statistics for all groups with two measurements, mean and standard error, for two variables, DIFFERENCE (the absolute error) and DPERMILE (the proportional error).

Table 6 provides average lengths for groups based on slope and average slopes for groups based on length. The purpose of this table was to see if there was an indication of the relationship between slope and length.

With our case study, errors less than or equal to 0.03 mile/mile are considered acceptable and errors less than or equal to 0.01 mile/mile are considered to be good based on an application needs analysis at NCDOT. Using this accuracy criterion, several important points were observed from the above tables:

(1) The magnitude of absolute errors is less than that of the proportional errors.

(2) Most links are relatively short. Average values of link length are less than 1 mile for 4 groups out of 5 groups based on slope. From groups based on length, groups having less than 1 mile length (A, B, and C) have a total of 1292 links, which is about 64% of all links. This observation supports the explanation of observation (1).

(3) Accuracy is acceptable for all groups except for steep slope group G5 and for short length group A. Elsewhere, we have defined good and acceptable error.

(4) The overall trend shows that accuracy decreases as the slope increases and accuracy decreases as the length decreases.

(5) Short steep links are potentially unacceptable from the length accuracy point of view.

(6) Observations (4) and (5) lead to the concern about how much of the error is contributed by the slope and how much of the error is contributed by the length. This information could possibly reveal the causes of errors. For example, human recording errors might tend to show up more for short lengths and GIS/calculation/storage errors tend to show up more for steep slopes. A further study following this will address this concern.

ERRORS AND CONTROL

This section discusses some of the control issues related to data quality errors in the case study. Three major error sources encountered are identified and control mechanisms for dealing with these errors are also discussed.

Data Quality and Errors in GISs

As mentioned earlier, data quality is always a critical issue in GIS applications. It is an important factor in the process of effectively and accurately conveying both spatial and nonspatial information (FGDC-007.3-1998). Errors are inherent in any GIS applications. Some are from datasets (errors from elevation data and road layer) that are involved in analysis and some are from analysis methods themselves (surface length calculation uses a series of straight line segments to approximate a curved line). In this study, major error sources were identified and controlled through careful data preparation and by applying screening filters during the analysis. It was recognized though that not all errors could be eliminated from the accuracy assessment in the case study. The following subsections discuss the nature of some of these errors.

Errors from DMI

DMI data were supposed to be the most accurate length measurement (± 1 foot per mile as specified in a manufacture's specifications) in this study and they were used as the base reference for accuracy assessment. The GIS length calculations were compared with that DMI data to determine how accurate the GIS data was. Recall that the GIS measurement derivations were based on NED/DEM elevation data whose quality was being evaluated for use in the distance measurement study. DMI is capable of providing accurate distance measurement between two locations by driving a car equipped with a highly precise and accurate DMI along the actual roads.

Aside from DMI instrument errors, the primary errors that occur with DMI come from the operator's interpretation of where to begin and end the measurement of physical features. With intersections there is often the operator's interpretation of where the center of the intersection is. With roads there is the question of precisely where they begin and end. These questions are related to the use of DMI rather than to its accuracy. Traveling at a speed of 50 mph a delay of ¾ of a second would introduce a 0.01-mile error in the length measurement. So it is clear that operator error in using DMI is significant.

Errors from NED

NED data involved in this study came from USGS. This set of data is the result of USGS efforts to provide seamless raster elevation data for the United States. It is essentially a DEM, which captures terrain surfaces with regular or irregular sampling points with X, Y and Z values. Elevations for positions other than those sampling points are interpolated from elevations of those sampling points.

It is clear that the elevations for all NED data sampling points have errors. The degree of accuracy for these points depends on the equipment, method, and personnel involved in measuring the elevations. Given that a mechanical transport vehicle (aircraft) is involved (which is likely GPS guided) that takes GPS position readings and then extrapolates surface-based readings from this position there is a lot of room for error. Second, elevation values for points with missing values are obtained from interpolation. Interpolation is just an approximate method that assumes that the elevation for one unknown point is related to its neighboring points. In addition, by using points to represent the continuous terrain surface and interpolating unknown intermediate elevations from known sampling points, a second order of approximation is introduced.

Accuracy of NED is still under assessment (USGS 2001e). As recommended by USGS, it is best to refer to published information on the accuracy of the source digital elevation models (DEM's) from which NED was assembled (USGS 2001e). In our case study, the source DEM has level 1 quality, which has vertical RMSE between 7 and 15 meters (USGS 1993). Please note the current version of Data Users Guide 5 contains all changes through March 15, 1994 (USGS 1993). In our case study, 3100 tiles composing the NED data from the USGS for North Carolina were mosaiced together. Further improvements on the data accuracy were also performed at NCDOT, which leads to a better, yet unknown, accuracy.

The errors mentioned here cannot be eliminated from the study, but their importance should be noted. The goal herein was not to improve elevation data. Rather, it was to determine whether or not its quality was high enough to enable the automatic generation of reasonably accurate slope distances that happen to represent roadway centerline distances.

Errors from the GIS Road Layer

The vector data representing interstate highways (the GIS road layer) was originally stored as CAD files in a DGN format. Some roads in these files are not shown in their actual locations. This is obvious when matching orthorectified aerial photos with the linework in the CAD files. Without roads being drawn in their actual locations, surface length calculations would generate incorrect lengths and the resulting accuracy assessment would have little validity.

In our case study, this set of errors was reduced by matching orthorectified aerial photos (1:24,000) with GIS road layer coverages and correctly placing all the linework directly over the roadway centerline by digitizing. Still it must be recognized that by digitizing roads to assure that they are in their correct locations, digitizing errors and scale-dependent errors, are introduced into the road layer.

CONCLUSIONS

GIS software is seeing increasing use in civil engineering applications, and appropriately so. It continues to evolve so that it possesses new capabilities that are of use to civil engineering.

GIS data quality and accuracy were addressed in the case study described herein. This was done by comparing results from a GIS measurement analysis with DMI data. Three methods (descriptive statistics, frequency analysis, and RMSE) were used to determine whether or not the GIS method provides acceptable accuracy of distance measurement when compared with DMI data.

This study leads to a number of conclusions. First, the method of using GIS plus NED is a technically feasible way to determine roadway centerline lengths using surface analysis functions provided by GIS tools. Programs may be written using ARC Macro Language that are capable of calculating surface length based on a shapefile or coverage and NED. Second, the GIS method is as accurate as DMI, which was demonstrated by the analysis results. Third, most errors occur with links that have short lengths and high slopes. For these links, alternatives to the proposed GIS/NED approach, such as DMI or GPS should be employed to acquire length information.

It was also recognized that the origins of errors are plentiful and might originate in the data sets, the method itself, or other sources. The grid file that contains elevation data is built using NED, which is acquired by sampling and interpolating. Thus, that original NED data contains both horizontal and vertical errors and the road layer also has inherent position and digitizing errors. It is these errors that we were trying to quantify and assess.

A slope distance is calculated from an accurate elevation difference between two points and an accurate 2-D planimetric distance between them. The accuracy of the slope distance is thus a function of the accuracy of the linear and elevation distances.

The results of this study show that when you have accurate planimetric distances, elevations obtained from DEM data will give a degree of slope distance accuracy that is acceptable to most NCDOT applications. This tells us that we could immediately use current DEM data and straightforward algorithms to calculate accurate road mileages if we have accurate planimetric distances.

It must be noted, however, that the use of calculated distance, as opposed to driven distance, requires that all GIS linework be photo-revised using orthorectified aerial photos. In doing so all lines will then be accurate to the same degree, and they will be so to the accuracy level of the 1:24,000 aerial photographs, thus introducing a significant amount of consistency into planimetric distance values. Presently, one distance value might be highly accurate and one might be highly inaccurate. It is difficult to have a degree of consistency in the data and in their accuracy. Using DEM, all the measurements will have primarily the same degree of accuracy, thus improving user confidence in the distances.

REFERENCES

Amrhein, C. G. and Schut, P. (1990). "Data Quality Standards and Geographic Information Systems," *Proceedings of National Conference 'GIS for the 1990's*, Canadian Institute of Surveying and Mapping, Pages 918-930.

Backe, K. (1996). "Formal Spatial Data Standards – What are They and Who Does Them," *Proceedings of ASPRS/ACSM, Volume 1, Remote Sensing and Photogrammetry*, Baltimore, Maryland, Page 111.

Beard, M. K. and Buttenfield, B. P. (1999). *Geographic Information Systems – Principles and Technical Issues – Volume 1 – Detecting and Evaluating Errors by Geographical Methods*, John Wiley & Sons, New York, Chapter 15, Pages 219-233.

Burrough, P. A. and McDonnell, R. A. (1998). *Principles of Geographic Information Systems – Errors and Quality Control*, Oxford University Press, New York, Chapter 9, Pages 220-240.

Chapman, M. A., Alesheikh, A., and Karimi, H. A. (1997). "Error Modeling and Management for Data in Geographical Information Systems," *COASTGIS'97*, Aberdeen, Scotland.

DeMers, M. N. (2000). *Fundamentals of Geographic Information Systems*, 2nd Edition, John Wiley & Sons, New York, Chapter 8, Pages 208-234, Chapter 10, Pages 256-297.

Dueker, K. J. and Butler, J. A. (1998). "GIS-T Enterprise Data Model with Suggested Implementation Choices," *URISA*, Volume 10, Number 1, pages 12-36.

Dueker, K. J. and Kjerne, D. (1989). "Multipurpose Cadastre: Terms and Definitions," *Proceedings Annual Convention ACSM-ASPRS*, Volume 5, pages 94-103.

FGDC-STD-002 (1998). "Spatial Data Transfer Standards," Federal Geographic Data Committee (a modified version was adopted as ANSI NCITS 320:1998).

FGDC-STD-007.3-1998 (1998). "Geospatial Positioning Accuracy Standards," Subcommittee for Base Cartographic Data, Federal Geographic Data Committee,. http://www.fgdc.gov/standards/status/sub1_3.html. Accessed June 16, 2001.

Fisher, P. F. (1995). "An Exploration of Probable Viewsheds in Landscape Planning," *Environment and Planning B: Planning and Design*, 22: 527-46.

Fletcher, D. R.(1987)."Modeling GIS Transportation Networks," *Proceedings Volume II, Urban and Regional Information Systems Association, 25th Annual Conference*, Fort Lauderdale, FL.

Fletcher, D. R. (2000). "Geographic Information Systems for Transportation: A Look Forward," *Transportation in the New Millennium*, Transportation Research Board, Washington, DC.

Goodchild, M. F. (1994). "Spatial Accuracy," ***Proceedings, 22nd Annual Conference,*** Australian Urban and Regional Information Systems Association, Sydney, Australia.

Goodchild, M. F. (1991a). "Issues of Quality and Uncertainty," In J. C. Muller (Ed.), *Advances in Cartography*, Elsevier, New York, Pages 113-139.

Goodchild, M. F. (1991b). "Issues of Quality and Uncertainty," *Proceedings of the State of Indiana Geographic Information System Conference*, Indianapolis, IN, Indiana University GIS Alliance, Pages 17-53.

Goodchild, M. F. and Gopal, S. (1989). *The Accuracy of Spatial Databases*, Taylor & Francis, London.

Goodchild, M. F. and Jeansoulin, R. (Eds.) (1998). *Data Quality in Geographic Information*, Editions Hermes, Paris.

Goodchild, M. F., Guoqing, S., and Shiren, Y. (1992). "Development and Test of an Error Model for Categorical Data," *International Journal of Geographic Information Systems*, Volume 6, Number 2, Pages 87-104.

Guptill, S. C. and Morrison, J., (eds.) (1995). *The Elements of Spatial Data Quality*, Elsevier, Armsterdam.

Heesom D. and Lamine M. (2001). "Effect of Grid Resolution and Terrain Characteristics on Data from DTM," *Journal of Computing in Civil Engineering*, Volume 15, Number 2, American Society of Civil Engineers, Pages 137-143.

Heuvelink, G. B. M. (1999). *Geographic Information Systems – Principles and Technical Issues - Vol. 1 – Propagation of Error in Spatial Modeling with GIS*, John Wiley & Sons, New York, Chapter 14, Pages 207-217.

Hunter, G. J. and Goodchild, M. F. (1996). "A New Model for Handling Vector Data Uncertainty in Geographic Information Systems," *Journal of the Urban and Regional Information Systems Association*, Volume 8, Number 1, Pages 51-57.

Hunter, G. J. and Goodchild, M. F. (1995). "A New Model for Handling Vector Data Uncertainty in Geographic Information Systems," *Proceedings, URISA 95,* San Antonio, TX, Volume 1, Pages 410-419.

Karimi, H. A., Khattak, A. J., and Hummer, J. E. (2000). "Evaluation of Mobile Mapping Systems for Roadway Inventory Data Collection," Journal of Computing in Civil Engineering, Volume 14, No. 3, Pages 168-173.

Kennie, T. J. M. and Petrie, G. (1990). "Introduction to Terrain Modeling – Application Fields and Terminology," *Terrain Modeling in Surveying and Civil Engineering*, T. J. M. Kennie and G. Petrie, editors, Whittles Publication, Latherwheel, Caithness, Pages 1-2.

Massasati, Ahmad S. (2000). "An Exercise in Digital Elevation/Terrain Models: From Point to Mathematics," *NCGIA Core Curriculum in GISscience*, http://www.ncgis.ucsb.edu/giscc/extra/e001/eoo1.html, accessed June 26, 2001.

Nyerges, T. L. and Dueker, K. J. (1988). *Geographic Information Systems in Transportation*, Office of Planning HPN-22, Federal Highway Administration, U. S. Department of Transportation, Washington, DC.

Peuker, T. K. *et. al.* (1978), "The Triangulated Network," *Proceedings of the Digital Terrain Models-DTM Symposium*, American Society of Photogrammetry – American Congress on Surveying and Mapping, St. Louis, Missouri, Pages 24-31.

Quiroga, C. A. (1999). "Accuracy of Linearly Referenced Data by Using Geographic Information Systems," *Transportation Research Record 1660*, Transportation Research Board, Pages 100-107.

Rasdorf, W. (2000). "Spatial Data Quality," Technical Report, Department of Civil Engineering, North Carolina State University, Raleigh, NC.

Rasdorf, W., Cai, H., Tilley, C., Brun, S., and Robson, F. (2001). "Accuracy Assessment of Road Length Measurement Using DEM," *Journal of Transportation Engineering*, American Society of Civil Engineers, Submitted.

Ries, T. (1993). "Design Requirements for Location as A Foundation for Transportation Information Systems," *Proceedings, AASHTO Symposium on Geographic Information Systems in Transportation*, Portland, OR.

SDTS, (1997). "American National Standard for Information Systems – Spatial Data Transfer Standard (SDTS) – Part 1. Logical Specifications," Draft for Review, Federal Geographic Data Committee, Published by American National Standard Institute, New York, page 5.

Shearer, J. W. (1990). "The Accuracy of Digital Terrain Model," *Terrain Modeling in Surveying and Civil Engineering*, T. J. M. Kennie and G. Petrie, editors, Whittles Publication, Pages 315-334.

USGS (2001a). "National Elevation Dataset," http://edcnts12.cr.usgs.gov/ned/, accessed May 26, 2001.

USGS (2001b). "National Elevation Dataset," http://edcnts12.cr.usgs.gov/ned/about.htm, accessed May 26, 2001.

USGS (2001c). "Digital Elevation Model," Western Geographic Science Center, USGS, http://craterlake.wr.usgs.gov/dem.shtml, accessed June 2, 2001.

USGS (2001d). "Topographic Maps," http://mcmcweb.er.usgs.gov/topomaps/, accessed August 16, 2001.

USGS (2001e). "Accuracy of the National Elevation Dataset," http://edcnts12.cr.usgs.gov/ned/Accuracy.htm, accessed May 23, 2002.

USGS (1993). "Data Users Guide 5," ftp://mapping.usgs.gov/pub/ti/DEM/demguide/dugdem.txt, accessed May 23, 2002.

Vonderohe, A. P., Travis, L., Smith, R. L., and Tsai, V. (1994). "Adapting Geographic Information Systems for Transportation," ***Transportation Research News***, Number 171, Transportation Research Board.

Vonderohe, A. P., Travis, L., Smith, R. L., and Tsai, V. (1993). "Adaption of Geographic Information Systems for Transportation," ***NCHRP Technical Report 359***, Transportation Research Board.

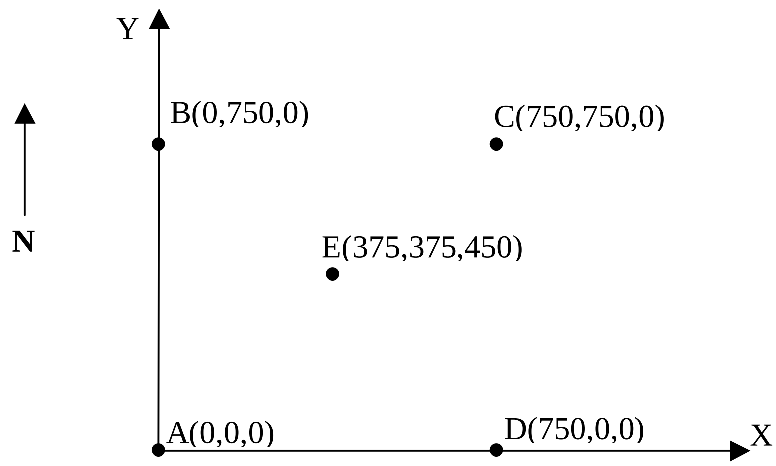

Figure 1. The Point Model for a Pyramid

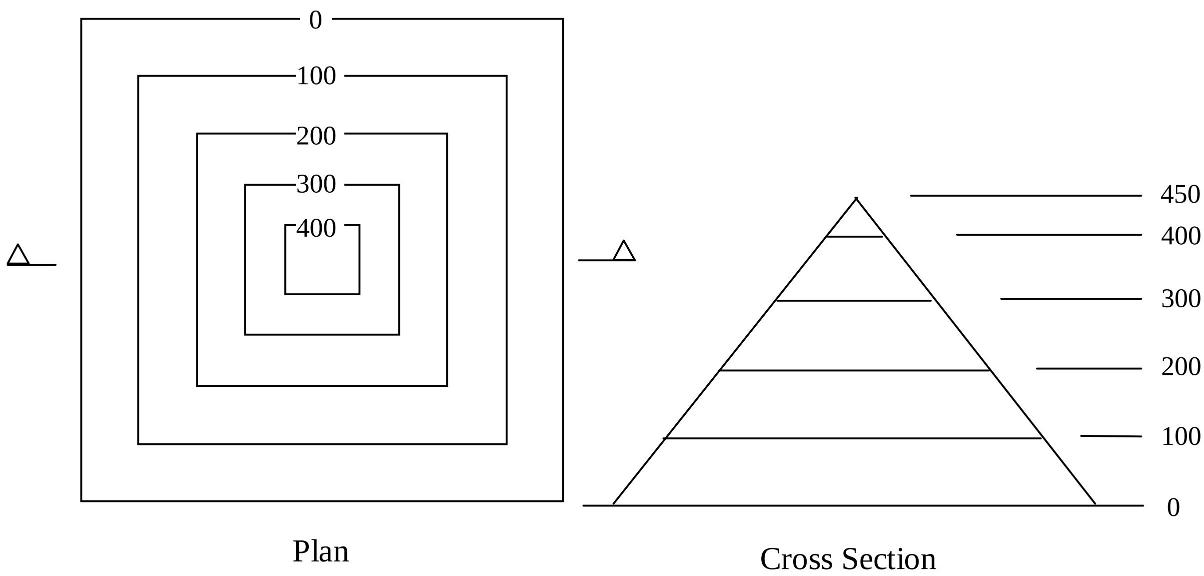

Figure 2. The Contour Line Model for a Pyramid

90	90	90	90	90	90	90	90	90	90
90	180	180	180	180	180	180	180	180	90
90	180	270	270	270	270	270	270	180	90
90	180	270	360	360	360	360	270	180	90
90	180	270	360	450	450	360	270	180	90
90	180	270	360	450	450	360	270	180	90
90	180	270	360	360	360	360	270	180	90
90	180	270	270	270	270	270	270	180	90
90	180	180	180	180	180	180	180	180	90
90	90	90	90	90	90	90	90	90	90

Figure 3. The DEM/DTM for the Pyramid

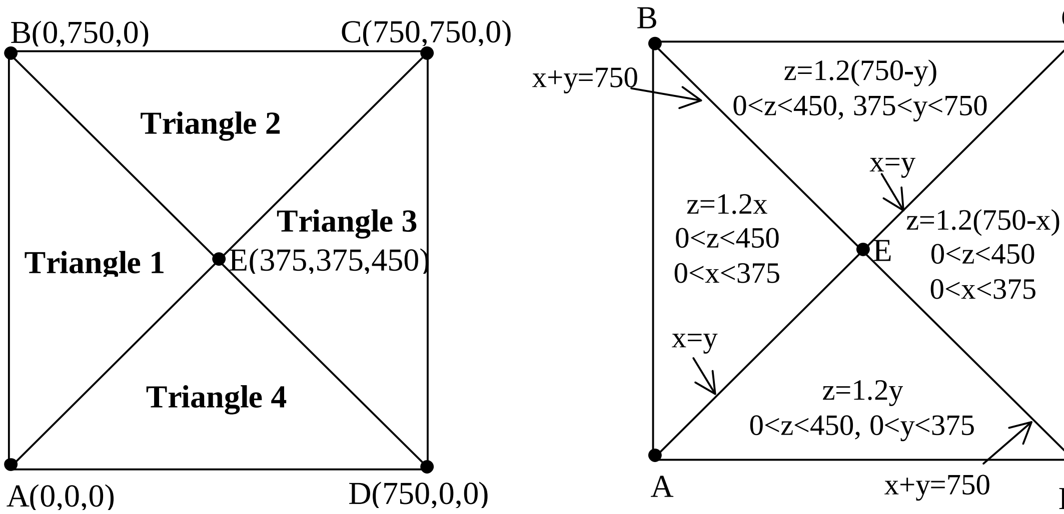

Figure 4. The TIN Model for the Pyramid

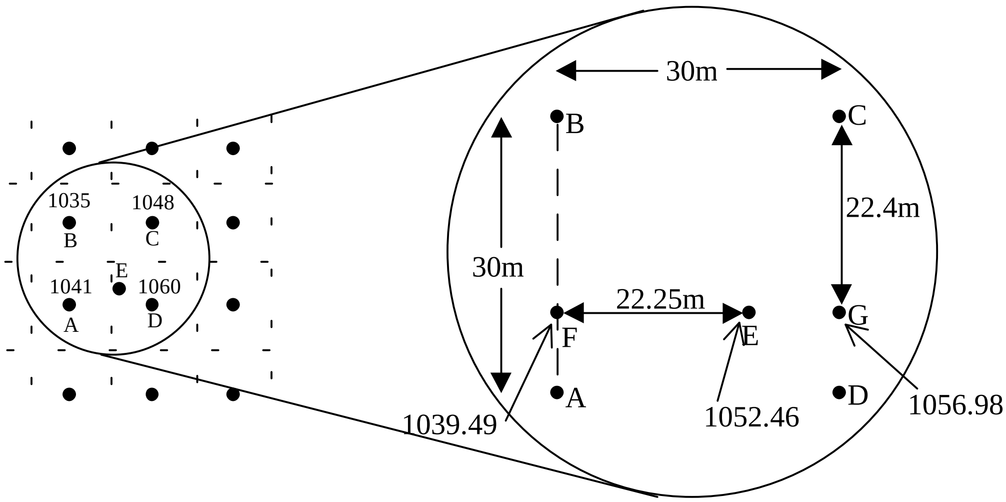

Figure 5. Interpolating an Elevation for Point E

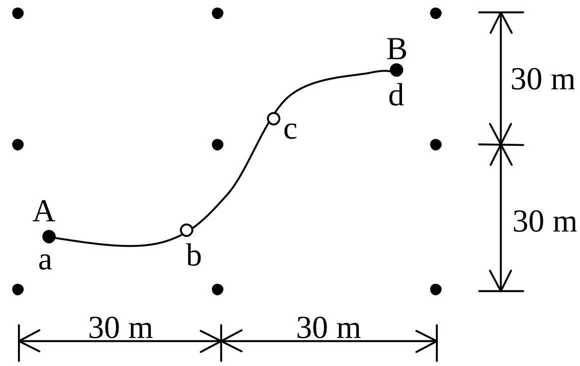

Figure 6. Line and Its Segments Defined Using A Sample Distance

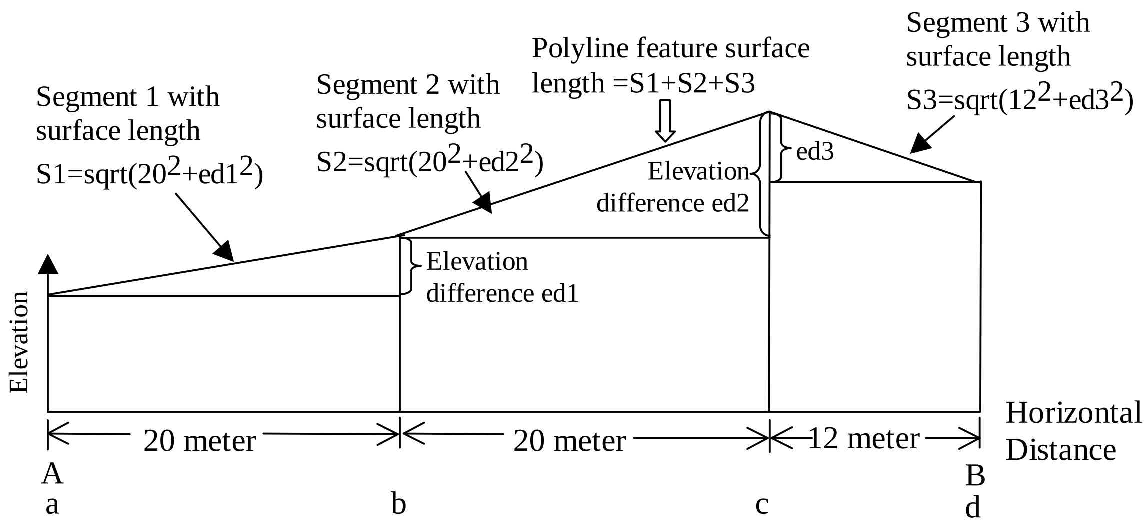

Figure 7. Calculation of Surface Length for Segments and Link Using Grid File

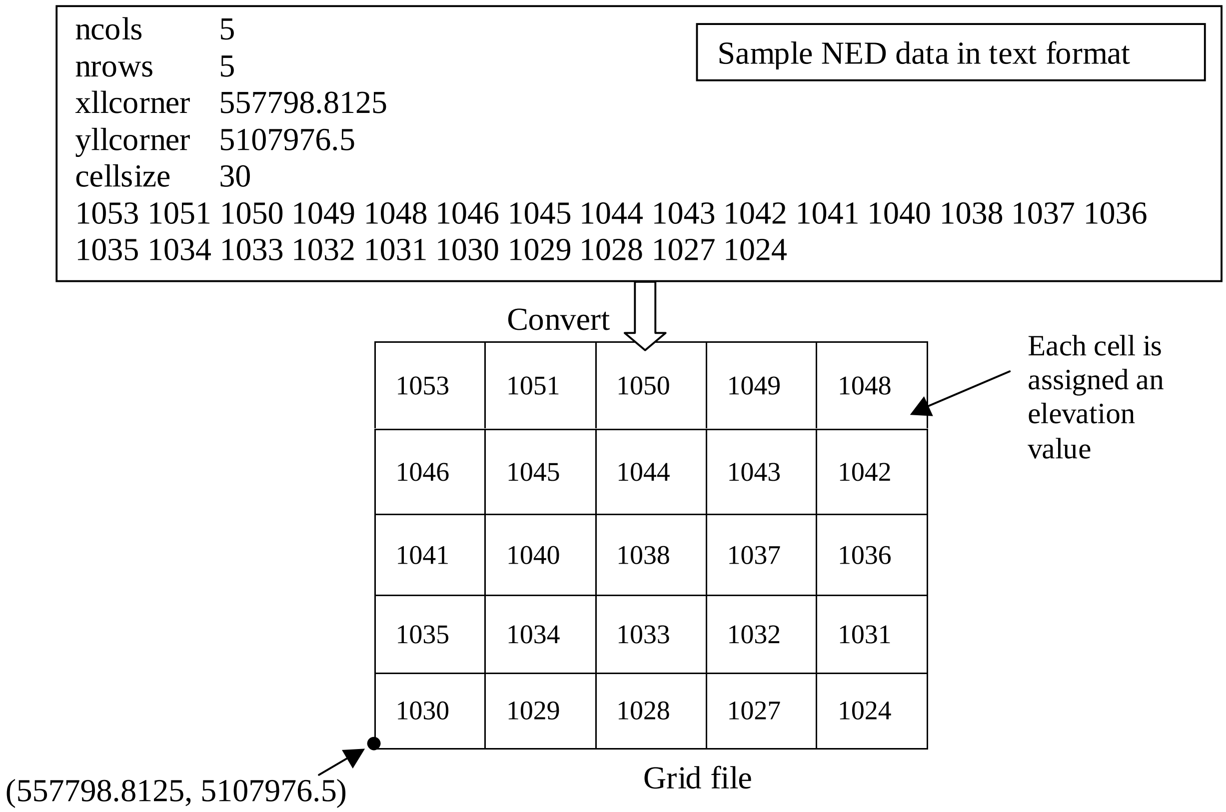

Figure 8. The Concept of Converting NEDs into Grid File

Table 1. Frequency Analysis for Groups Based on Slope

Groups			G1	G2	G3	G4	G5	All
Total # of Links			511	900	415	119	89	2034
DIFFERENCE	<=0.01 mile	# of links	427	766	352	94	64	1703
		%	84%	85%	85%	79%	72%	84%
	<=0.02 mile	# of links	475	828	382	108	72	1865
		%	93%	92%	92%	91%	81%	92%
	<=0.03 mile	# of links	486	856	394	111	78	1925
		%	95%	95%	95%	93%	88%	95%
DPERMILE	<=0.01 mile/mile	# of links	389	618	277	64	53	1401
		%	76%	69%	67%	54%	60%	69%
	<=0.02 mile/mile	# of links	431	712	322	80	60	1605
		%	84%	79%	78%	67%	67%	79%
	<=0.03 mile/mile	# of links	449	763	345	84	67	1708
		%	88%	85%	83%	71%	75%	84%

Table 2. Frequency Analysis for Groups Based on Length

Groups			A	B	C	D	E
Total # of Links			166	565	561	598	144
DIFFERENCE	<=0.01mile	# of links	145	455	476	510	117
		%	87%	81%	85%	85%	81%
	<=0.02mile	# of links	150	501	525	555	134
		%	90%	89%	94%	93%	93%
	<=0.03mile	# of links	155	522	536	574	138
		%	93%	92%	96%	96%	96%
DPERMILE	<=0.01 mile/mile	# of links	95	233	400	535	138
		%	57%	41%	71%	89%	96%
	<=0.02 mile/mile	# of links	95	307	490	570	143
		%	57%	54%	87%	95%	99%
	<=0.03 mile/mile	# of links	95	364	523	582	144
		%	57%	64%	93%	97%	100%

Table 3. RMSE and 95% Confidence Accuracy for Groups Based on Slope

		Groups	G1	G2	G3	G4	G5
SLOPE	DIFFERENCE	RMSE	0.000	0.017	0.017	0.018	0.029
		95% Confidence Level	0.000	0.033	0.033	0.035	0.057
	DPERMILE	RMSE	0.019	0.094	0.118	0.386	0.276
		95% Confidence Level	0.037	0.184	0.231	0.757	0.541

Table 4. RMSE and 95% Confidence Accuracy for Groups Based on Length

	Groups		A	B	C	D	E
LENGTH	DIFFERENCE	RMSE	0.018	0.021	0.016	0.015	0.021
		95% Confidence Level	0.035	0.041	0.031	0.029	0.041
	DPERMILE	RMSE	0.506	0.095	0.023	0.012	0.006
		95% Confidence Level	0.992	0.186	0.045	0.024	0.012

Table 5. Summarized Descriptive Statistics

Group	DIFFERENCE		DPERMILE	
	Mean	Standard Error	Mean	Standard Error
G1	0.009	0.001	0.027	0.006
G2	0.009	0.000	0.029	0.003
G3	0.009	0.001	0.037	0.006
G4	0.011	0.001	0.107	0.034
G5	0.016	0.003	0.111	0.027
A	0.008	0.001	0.238	0.035
B	0.012	0.001	0.047	0.003
C	0.009	0.001	0.012	0.001
D	0.009	0.000	0.008	0.000
E	0.010	0.001	0.002	0.000

Table 6. Average Lengths for Groups Based on Slope and Average Slopes for Groups Based on Length

Groups	G1 (Flat)	G2	G3	G4	G5 (Steep)
Average Length	1.14 mile	0.85 mile	0.77 mile	0.72 mile	0.90 mile
Groups	**A (Short)**	**B**	**C**	**D**	**E (Long)**
Average Slope	4.89%	3.73%	3.39%	3.05%	3.00%