GIS/NED Route Mileage Verification

Hubo Cai, Ph.D.
GIS Analyst
URS
Morrisville, NC 27560
Phone: (919) 461-1210
email: Hubo_Cai@URSCorp.com

William Rasdorf, Ph.D., P.E., F. 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

Chris Tilley
North Carolina Department of Transportation
Geographic Information System Unit
Statewide Planning Branch
3401 Carl Sandburg Court
Raleigh, NC 27610
phone: (919) 212-6040
fax: (919) 212-5999
email: ctilley@dot.state.nc.us

L.C. Smith
North Carolina Department of Transportation
Geographic Information System Unit
Statewide Planning Branch
3401 Carl Sandburg Court
Raleigh, NC 27610
phone: (919) 212-6002
fax: (919) 212-5999
email: lcsmith@dot.state.nc.us

Forrest Robson, P.E.
North Carolina Department of Transportation
Geographic Information System Unit
Statewide Planning Branch
3401 Carl Sandburg Court
Raleigh, NC 27610
phone: (919) 212-6001
fax: (919) 212-5999
email: frobson@dot.state.nc.us

KeyWords: DEM
Geographic Information System
Digital Elevation Model
National Elevation Dataset
Route Distance
Route Measurements
Route Mileages
Transportation Routes

OUTLINE

ABSTRACT

Maintaining the nations roads and highways is a critical national infrastructure activity. To do so requires that accurate inventories of road way features as well as the roadway network itself be maintained. This paper addresses the accuracy of road network mileages, that is, the lengths of various road segments and routes. A study was conducted to determine the accuracy of using National Elevation Data to calculate these lengths. The advantage of doing so is that, if successful, costly and time consuming field studies could be eliminated or reduced. This savings could positively impact all State DOT's. The scope of the study includes interstate roads as well as all other road types including secondary roads. The primary contribution of the paper is its finding that using NED data provides significant advantages over field studies and provides a very acceptable accuracy level. It also identifies areas of questionable accuracy and indicates why. Additionally, the paper provides some insight into secondary road data which is an often neglected area.

INTRODUCTION

Geographic Information Systems (GISs) are being widely used in Transportation Engineering and are expecting more potentials in this field. Among the functionalities of a geographic information system (GIS) is its ability to derive 3D surface lengths for linear features based on elevation data.

A series of studies were conducted by the North Carolina Department of Transportation (NCDOT) to determine the appropriateness of using a GIS/NED approach to calculate centerline surface lengths for state-maintained road segments in North Carolina. A detailed description and results of the first portion of the study, which focused on interstate highways, has been published [Rasdorf, et al. 2003a]. The current paper describes the results and findings for the studies involving all road types. Additionally, it addresses some of the unique problems associated with the highway network that somewhat limit the generality of results.

State Road Inventories

State DOTs are responsible for the management of state-maintained roads. To do so requires large volume of road inventory data, including such things as road conditions, pavement data, traffic data, accident data, etc. These data are attached to a model of the roadway network. This model can be stored in a database or in a GIS. The model itself is referred to as a linear referencing system (LRS).

LRS

Linear referencing systems (LRSs) are presently being developed and used to manage and promote the analysis of spatial data elements and road inventories on transportation networks [Geo Decisions 1997, Adams, et al. 2002]. An LRS describes transportation data in terms of events that are located along roads [Rasdorf, et al. 2002].

There are three critical types of data: 1) topology, 2) geometry, and 3) attributes. According to the Environmental Systems Research Institute (ESRI) GIS glossary [ESRI 2003], topology refers to the spatial relationships between two neighboring features. For example, with a transportation network, a road segment might be connected to another road segment (connectivity) or to a point such as an intersection. The focus is on the connectivity.

Geometry, on the other hand, refers to the measures and properties of points, lines, and surfaces that enable us to represent geographic features spatially. For example, the location of a point, the length measure of a line, and the area and perimeter measures of a polygon. Attributes refer to the characteristics of a geographic feature described by numbers, characters, images, and CAD drawings, typically stored in tabular format and linked to the feature by a user-assigned identifier (e.g., the attributes of a road segment might include lane number, road conditions, traffic counts, etc.). This paper deals with the second of these data types – geometry. Even more specifically the paper focuses on the length measurements of transportation linear objects.

Using a GIS to Implement an LRS

While a GIS is not a requirement in implementing an LRS, many states are using GIS to manage road inventory data. GIS may be defined as a computer-based tool set for collecting, storing, retrieving, transforming and displaying spatial data from the real world for a particular set of purposes [Burrough and McDonnell 1998]. Spatial data represents phenomena from the real world in terms of (1) geometry, (2) attributes that are unrelated to position (examples include thickness, cost, incidence of accidents, etc.) and, (3) their spatial interrelations with each other which describe how they are linked together (this is known as topology and describes space and spatial properties such as connectivity) [Burrough and McDonnell 1998]. In other words, all three types of road inventory data have their corresponding data types in GIS and, therefore, GIS is well suited to wide use by transportation agencies to store, manage, and analyze road inventory data [Hummer et al. 2000].

Road Length Measurement

NCDOT is updating its roadway inventory database. In doing so one key item that needs to be updated is the road length measurement [Rasdorf, et al. 2003b]. The NCDOT is responsible for managing about 86,000 miles of state-maintained roads. State and federal databases store numerous data on these roads and much depends on accuracy in these databases. Prominent in importance is length measurement – roadway mileages. Keeping exact road lengths in the road inventory database is critical because it is the foundation of the geometry of the roadway network. This, in turn, is the foundation for any transportation applications that involve the use of this roadway network, for example, truck routing.

Many existing technologies are capable of obtaining length measurement with different accuracies, levels of effort, and cost. For example, length measurements can be derived from legacy design drawings. However, this approach suffers from the problems of currency and completeness. These drawings are only kept for a certain number of years. There might be many adjustments and modifications that are not represented on such drawings. The resulting conversion process is generally unsatisfactory.

Ground surveying is another technology that can be used to obtain distance information along linear objects with high accuracy when measuring the length of straight linear objects. However, this is not the case in road distance measurements (consider curves and turnovers). In addition, surveying fieldwork interrupts traffic excessively and poses safety risks. Finally, the manpower required for a full field survey would be overwhelming.

GPS is a very promising technology for measuring linear objects. However, the most notable difficulties in using this technology stem from signal blockage in certain areas caused by tree canopies or bridges as well as multipath problems caused by signal deflection from high-rise buildings. Other issues include cost. As noted above this is again a technology that mandates a physical deployment of personnel and equipment and entails all of the associated aspects of equipment training, and work process integration. These are costly. Other issues relate to GPS concerns of data quality [Karimi, et. al. 2004].

The preceding paragraphs discussed technologies which do not measure curves without approximation. The distance measurement instrument (DMI) and the inertial navigation system (INS) are other tools to measure distance directly along linear objects, even if they are curved. An INS uses accelerometers and gyroscopes that provide pitch, roll, heading information to derive a relative geo-reference for a point. According to Hummer an INS has high accuracy over small distances and is usually more accurate than the global positioning system (GPS) but much more costly [Hummer, et al. 2000]. This is true for short wave length data but INS does exhibit considerable drift over time by itself.

DMI is an inexpensive technology consisting of a mechanical device attached to one or more of the wheels of a van or other vehicles and connected to an in-vehicle recorder [Hummer, et al. 2000]. It is currently the most accurate technology available to DOTs. However, a DMI requires frequent calibrations to minimize errors that accumulate the result of travel. In North Carolina, there are about 86,000 miles of state-maintained roads. For such mileage, it is very time-consuming and labor intensive to use a DMI approach to measure the lengths for all state-maintained roads. For such mileage, the substitution of a computational GIS/NED approach in place of physical DMI measurements could significantly reduce manpower, time, and cost and increase accuracy [Rasdorf, et al. 2003a]. If we could simply generate mileages overnight using good algorithms we could achieve significant savings. However, the question of the accuracy of this approach arose and led to the initiation of a series of accuracy assessment studies, the results of which are reported herein.

Study Objectives

The objective of the work that has been completed is to determine the level of accuracy obtainable when applying a GIS/NED approach to calculate road centerline lengths for state-maintained roads and, in doing so, to answer the following questions:

(1). How accurate are the results of the GIS/NED approach compared with DMI measured lengths?

(2). What are the problems in applying the GIS/NED approach in deriving three dimensional distances for road centerlines and how could they be dealt with?

By answering these questions one can determine whether or not a GIS/NED approach is appropriate to obtain road surface lengths and to update transportation databases [Rasdorf, et al. 2003a]. In addition, the prerequisites for applying this approach to obtain road length measurement are identified.

METHODOLOGY

The approach that was used included the following steps:

1) A GIS/NED methodology was developed for deriving a three dimensional model for road centerlines, which in turn calculates the three dimensional distances for road centerlines.
2) Input data for the three dimensional model were collected and preprocessed.
3) The three dimensional model was implemented using the input data and the three dimensional distances for road centerlines were derived.

4) The results (three dimensional distances) of the GIS/NED approach were compared to the DMI length measurements to complete an accuracy assessment.
5) An analysis of discrepancies was conducted.

GIS/NED

NED is a raster product assembled by the USGS in which terrain elevations for ground positions are sampled at either regular or irregular horizontal intervals [*Anderson 1998, USGS 2001*]. The terrain (e.g. drainage basin, water supply, etc.) surface is represented by using an array with X, Y, and Z values [*Yanalak 2003*]. DEDs are readily convertible into a grid file using most commercial GIS software.

What is meant by a GIS/NED approach is the use of GIS software and NED elevation data to model and generate three-dimensional road centerline distances. The planar alignment of the roadway is obtained from a GIS road coverage, which was originally derived from overlays on orthorectified aerial photos. NED is essential a digital elevation model (DEM) that stores elevations for the ground surface. Combining NED elevations with planar alignment data enables us to derive a three-dimensional model.

This three-dimensional model consists of a series of three-dimensional points with X/Y/Z-coordinates along the road centerlines to represent the road centerline in a three-dimensional space. The X/Y-coordinates of the three–dimensional points are determined by the planar alignment of the roadway. The Z-coordinates of the three-dimensional points are determined using NED elevation data and are based on the X/Y-coordinates of the three-dimensional points on the planar plate [Rasdorf, et al. 2003a, b]. Since the elevation data are in the format of gridded cells, the elevation of a specific point is derived via a bilinear interpolation [Rasdorf, et al. 2003b]. This three-dimensional model then enables us to obtain a length measurement of the roadway centerlines by calculating the three-dimensional distance between two neighboring three-dimensional points and adding all such distances for a road segment. The accuracy of this measurement depends first on the accuracy of the original planar alignment and second, on the accuracy of the NED elevation data. The purpose of these studies is to quantify the NED elevation data accuracy for the DOT road centerline measurement application.

Limitations of Linework Digitized from Orthorectified Photos

The most critical part in implementing the GIS/NED approach is that the planimetric roadway must be correct in a spatial context. In conducting this study, the errors associated with the road centerline data were identified and one quality control mechanism was implemented by overlaying road centerlines with orthorectified aerial photos.

It was noticed that some road centerlines were not in the correct place when overlaid with these photos. Wherever there existed such a disparity, the road centerline was digitized from the orthorectified aerial photos. This process assures that all road centerlines are in the correct location with respect to the air photos. The road centerlines, as digitized from orthorectified photos, have two major error sources, the errors in the photography process and digitizing errors. Additionally, there are interpretation errors. It is these which are explained below and which were addressed in this study. These consist of dead ends and multiple lines.

Dead ends refer to the ends of roads. In the case of state-maintained roads, dead ends are mostly logical dead ends rather than physical. In other words, a dead end may simply mean the end of a state-maintained road section. In reality the road may actually continue on as a local street, for example. Since it is not a discernable physical transformation, it is impossible to accurately identify where it happens on the aerial photo. A potential quality control mechanism is to conduct a field survey.

The problem of multiple lines refer to the situation where there is more than one line connecting two neighboring nodes and all these lines have the same identity. This problem is caused by snapping algorithms in CAD software. In other words, there should be only one line connecting the two neighboring points. Instead, these multiple lines share one identity and are so close to each other that at a certain scale it is hardly noticeable and they cannot be visually deleted.

When the GIS/NED approach is used to derive a three-dimensional length, the three-dimensional lengths of these lines are all derived and added together as the GIS/NED length for the road segment between the two points. Since the distance is actually doubled (or even tripled depending on the number of lines) errors due to this problem are very significant. But they are easily identified because the resulting calculated mileages are so erroneous that the error is obvious. The quality control approach that was taken for these lines was twofold. First, we examined the road network data in detail to remove multiple lines before the GIS/NED algorithm was executed to derive the three-dimensional lengths for road segments. Second, we used the obviously different mileages to identify other such situations.

DMI

A DMI is essentially an electronic receiver that works in a way similar to an odometer. It works together with sensors and an electronic amplifier. The sensors are either wheel sensors or transmission sensors depending on where they are located. When the vehicle moves, the sensors detect the vehicle movement and generate electronic pulses. These electronic pulses are preprocessed and amplified by the electronic amplifier into the suitable working rate. The pulses from the electronic interface amplifier are sent to the DMI. Based on the received pulses, the DMI counts the revolutions of the vehicle transmission (when transmission sensors are used) or the revolutions of the wheel (when the wheel sensors are used) and derives the distance the vehicle has traveled.

There are reasons for errors to be inherent in DMI lengths. These include vehicle tire wear over time, high speed, and most significantly, which lane was driven to acquire DMI lengths. When properly calibrated, a DMI measures the actual length of a linear object such as a road segment at an accuracy level of ± 1 foot per mile (repeatability) or a 0.02% error as specified in the manufacturers' specifications. In other words, when measuring a road segment with a 2-mile length, the expected error would be ± 2 feet. As stated earlier, presently DMI measurements are the most accurate length measurements available to NCDOT and thus they were used as the basis (reference data) for comparison and accuracy assessment in this study. DMI data are available for all interstate highways in North Carolina. They were obtained during a field measurement effort in the summer of 2000.

PREVIOUS WORK

The first case study investigated this approach with respect to all interstate highways in North Carolina. Interstate highways were mandated because of the availability of DMI data and access to photo-revised roadway network data for them at that time.

In the first study, a GIS application program, written in ARC Macro Language (AML), used elevation data (USGS NED in this case) and a linear data layer (interstate highways) to build the three-dimensional model and calculate three-dimensional lengths for all interstate highway links. The calculated GIS results were compared with DMI measurements to assess their accuracy.

TABLE 1 Frequency Analysis for Groups Based on Slope [Rasdorf, et al. 2003a]

Groups			Group 1	Group 2	Group 3	Group 4	Group 5	All
Total # of Links			426	764	353	96	63	1702
DIFFERENCE	<=0.0161 km (<=0.01 mile)	# of links	426	762	351	92	57	1688
		%	100%	100%	99%	96%	90%	99%
	<=0.0322 km (<=0.02 mile)	# of links	426	764	353	96	62	1701
		%	100%	100%	100%	100%	98%	100%
	<=0.0483 km (<=0.03 mile)	# of links	426	764	353	96	63	1702
		%	100%	100%	100%	100%	100%	100%
DPERMILE	<=0.01 km/km (<=0.01 mile/mile)	# of links	372	602	268	62	45	1349
		%	87%	79%	76%	65%	71%	79%
	<=0.02 km/km (<=0.02 mile/mile)	# of links	398	677	302	75	48	1500
		%	93%	89%	86%	78%	76%	88%
	<=0.03 km/km (<=0.03 mile/mile)	# of links	409	704	320	75	52	1560
		%	96%	92%	91%	78%	83%	92%

TABLE 2 Frequency Analysis for Groups Based on Length [Rasdorf, et al. 2003a]

Groups			A	B	C	D	E
Total # of Links			145	455	476	507	119
DIFFERENCE	<=0.0161 km (<=0.01 mile)	# of links	145	455	472	503	113
		%	100%	100%	99%	99%	95%
	<=0.0322 km (<=0.02 mile)	# of links	145	455	475	507	119
		%	100%	100%	100%	100%	100%
	<=0.0483 km (<=0.03 mile)	# of links	145	455	476	507	119
		%	100%	100%	100%	100%	100%
DPERMILE	<=0.01 km/km (<=0.01 mile/mile)	# of links	95	233	397	505	119
		%	66%	51%	84%	100%	100%
	<=0.02 km/km (<=0.02 mile/mile)	# of links	95	307	472	507	119
		%	66%	67%	99%	100%	100%
	<=0.03 km/km	# of links	95	364	475	507	119
		%	66%	80%	100%	100%	100%

Three filters were applied to remove suspect road segments from the original dataset prior to processing to assure the validity of the assessment. All remaining valid segments were grouped by slope and length to evaluate the impact of these two attributes on the results. Three statistical methods were applied to assess how accurate the GIS/NED approach is. Tables 1 and 2 give the results for one of these frequency analyses. For a detailed description of the principle of the GIS/NED approach, the analysis procedure, and the results, refer to [Rasdorf, et al. 2003a].

The findings of the first case study revealed the following.

(1). The developed AML program is capable of obtaining surface lengths for road segments based on elevation data and road position data.
(2). The proposed method of using GIS to calculate surface lengths using currently available elevation data from USGS for road links provides acceptable accurate results.
(3). When there are discrepancies, these occur primarily with the links that either have high slope or short distance.
(4). The prerequisite for applying this approach is that GIS road layers must contain accurately placed linework.

It was concluded that there is much to be gained by using a GIS/NED approach to calculate roadway mileages rather than having to use DMI to get them. The accuracy of doing so is acceptable for the case of interstate highways.

But the results of the first case study still lead to several unresolved questions:

(1). What would be the accuracy of applying this approach to state-maintained roads other than interstate highways?
(2). 27% (634 out of 2336 links) of all interstate highway links were identified as suspect links. What would be the impacts of these links in a complete accuracy assessment?
(3). What will be the accuracy assessment result when a less accurate reference data is used?
(4). What happened to those links that have erroneous length measurements?

These questions lead to a new study that is described in the next section.

WAKE COUNTY TEST CASE

The original study effectively verified the accuracy of GIS/NED lengths when compared to newly driven (and thus, newly acquired) DMI lengths. However, the scope of that study was limited to interstate highways. It was obvious that additional study was needed. In the Wake County Test Case introduced here the study scope was expanded to include all state-maintained roads, while limiting the size of the study area to a single county. Doing so made the sample data more representative of all state-maintained roads when compared with the study scope of the previous study.

Overview

Wake County, NC was selected as a test county to compare computer-calculated lengths with length data from the NCDOT Pavement Management Unit (PMU) database. The use of the PMU database mileage as the baseline for comparison is questionable, though. The PMU database acquired road centerline length information over a long period of time. It came from numerous sources including DMI measurements, map scaling, and field reporting. It is obvious that the accuracy of the length information in the PMU database is not consistent. Those lengths which are DMI measured lengths might be very accurate, while others might be quite inaccurate. Sources of PMU database errors and inaccuracies are summarized as follows.

(1). Lack of measurement precision (collecting DMI length to tenth of a mile, for example);
(2). Field recording error;
(3). Transcription error; and
(4). Data input error.

In the previous interstate highway case study DMI lengths were used as the baseline for accuracy assessment of GIS/NED length calculation because they are the most accurate measurements that NCDOT currently has (though DMI measurements themselves have quantifiable errors). In the Wake County study, a less certain reference and comparison dataset was available as noted above.

Even though these mileages were not as uniformly consistent as DMI mileages they were still used. For any given road segment, if the GIS/NED calculated mileage matched the PMU database mileage (or was close), then they were consistent with each other (this assumption needs to be explained further in a subsequent study). If they significantly differed then we assumed that one of them was not correct. These disparities are quantified in the next section where we find many links matched and how close these matches were. We also identify those links that were significantly deviant.

Frequency Analysis Results

In this section, the GIS/NED calculated road lengths were compared with the length from the PMU database for road links of state-maintained roads in Wake County, North Carolina. Road links were grouped into interstate highway links, US route links, NC route links, and secondary route links. Also, links on which there are bridges, that cross railroads or county boundaries, and that are dead ends, were identified. Separate frequency analyses were performed on links excluding all or part of these links. The main concern is that links on which there are bridges, that cross railroads or county boundaries, or that are dead ends, are known to have more uncertainties than other links based on experience as well as the previous study [Rasdorf, et al. 2003a] [Rasdorf, et al. 2003b].

Table 3 shows the result of frequency analysis for all links. Table 4 shows the result for links that do not have bridges, railroads, dead ends, or county boundaries. Table 5 shows the result for links that do not have dead ends.

Result Analysis and Observations

Frequency analysis provides an in-depth view of the results, which can be represented in an understandable way as illustrated in Tables 3, 4, and 5. Observations from these tables include:

(1). Accuracy is good for all link groups. Overall, 90.3% of the links are within 0.03 mile error range, which was specified by NCDOT as acceptable. The group with the worst accuracy is the group of US route links. However, this group still has 73.2% of its 470 links within this error range.

TABLE 3 Frequency Analysis for All Links in Wake County

Groups			Interstate	US Routes	NC Routes	Secondary Routes	All
Total # of Links			196	470	295	7324	8285
DIFFERENCE km (mile)	(= 0.0000 km)	# of links	77	111	111	3621	3920
	(= 0.00 mile)	%	39.3%	23.6%	37.6%	49.4%	47.3%
	(<= 0.0161 km)	# of links	139	260	207	5927	6533
	(<=0.01 mile)	%	70.9%	55.3%	70.2%	80.9%	78.9%
	(<= 0.0322 km)	# of links	158	309	240	6485	7192
	(<= 0.02 mile)	%	80.6%	65.7%	81.4%	88.5%	86.8%
	(<= 0.0483 km)	# of links	170	344	254	6716	7484
	(<= 0.03 mile)	%	86.7%	73.2%	86.1%	91.7%	90.3%
	(<= 0.1609 km)	# of links	186	431	282	7142	8041
	(<= 0.10 mile)	%	94.9%	91.7%	95.6%	97.5%	97.1%
	(<= 0.3219 km)	# of links	193	444	291	7251	8179
	(<= 0.20 mile)	%	98.5%	94.5%	98.6%	99.0%	98.7%
	(> 0.3219 km)	# of links	3	26	4	73	106
	(> 0.20 mile)	%	1.5%	5.5%	1.4%	1.0%	1.3%

TABLE 4 Frequency Analysis for Links that Do Not Have Bridges, Railroads, Dead Ends, or County Boundary in Wake County

Groups			Interstate	US Routes	NC Routes	Secondary Routes	All
Total # of Links			2	239	215	4403	4859
DIFFERENCE km (mile)	(= 0.0000 km)	# of links	0	73	86	2277	2436
	(= 0.00 mile)	%	0.0%	30.5%	40.0%	51.7%	50.1%
	(<= 0.0161 km)	# of links	2	163	158	3698	4021
	(<=0.01 mile)	%	100.0%	68.2%	73.5%	84.0%	82.8%
	(<= 0.0322 km)	# of links	2	180	176	3993	4351
	(<= 0.02 mile)	%	100.0%	75.3%	81.9%	90.7%	89.5%
	(<= 0.0483 km)	# of links	2	193	188	4095	4478
	(<= 0.03 mile)	%	100.0%	80.8%	87.4%	93.0%	92.2%
	(<= 0.1609 km)	# of links	2	227	208	4308	4745
	(<= 0.10 mile)	%	100.0%	95.0%	96.7%	97.8%	97.7%
	(<= 0.3219 km)	# of links	2	232	213	4371	4818
	(<= 0.20 mile)	%	100.0%	97.1%	99.1%	99.3%	99.2%
	(> 0.3219 km)	# of links	0	7	2	32	41
	(> 0.20 mile)	%	0.0%	2.9%	0.9%	0.7%	0.8%

TABLE 5 Frequency Analysis for Links that Do Not Have Dead Ends in Wake County

Groups			Interstate	US Routes	NC Routes	Secondary Routes	All
Total # of Links			195	466	293	5213	6167
DIFFERENCE km (mile)	(= 0.0000 km)	# of links	77	109	109	2474	2769
	(= 0.00 mile)	%	39.5%	23.4%	37.2%	47.5%	44.9%
	(<= 0.0161 km)	# of links	139	258	205	4171	4773
	(<=0.01 mile)	%	71.3%	55.4%	70.0%	80.0%	77.4%
	(<= 0.0322 km)	# of links	157	307	238	4582	5284
	(<= 0.02 mile)	%	80.5%	65.9%	81.2%	87.9%	85.7%
	(<= 0.0483 km)	# of links	169	341	252	4758	5520
	(<= 0.03 mile)	%	86.7%	73.2%	86.0%	91.3%	89.5%
	(<= 0.1609 km)	# of links	185	428	280	5069	5962
	(<= 0.10 mile)	%	94.9%	91.8%	95.6%	97.2%	96.7%
	(<= 0.3219 km)	# of links	192	441	289	5154	6076
	(<= 0.20 mile)	%	98.5%	94.6%	98.6%	98.9%	98.5%
	(> 0.3219 km)	# of links	3	25	4	59	91
	(> 0.20 mile)	%	1.5%	5.4%	1.4%	1.1%	1.5%

(2). Most interstate highway links have a bridge, cross a railroad or county boundary, or are dead ends. This reduces the number of remaining links from 196 to 2. This sample size is far from enough to make sense of the result of this group in Table 4.

(3). Based on the comparison of Tables 3 and 4, it is observed that by excluding links that either have a bridge, cross a railroad or county boundary, or are dead ends, accuracy improves. Based on the 0.03 mile error range, the overall accuracy for links is improved from 90.3% to 92.2%. All groups have their accuracies improved. The biggest improvement happened to be the US route link group (the interstate group was not in consideration due to the small number of links in Table 4), which was identified to be the worst accuracy group in observation (1). The percentage of links whose length was less than or equal to the 0.03 mile acceptable error range increased from 73.2% to 80.8%.

(4). Based on a comparison of Tables 3 and 5, it was observed that by excluding dead end links, the accuracy remains almost the same (or even a little bit worse if the 0.03 mile error range is considered) as having them.

(5). Observations (3) and (4) tells us that dead ends are not the problem that we initially thought they were with respect to length measurement. The problem lies with links that either have bridges, cross county boundaries, or cross railroads. In other words, these are not accurately positioned. They have not correctly been located in either the database or the GIS.

Deviations

The question arises as to what happened in the case of the severe deviations between the GIS/NED results and the PMU database. The first assumption is that the reference data are erroneous. In other words, the length information in the PMU database is in question. Furthermore, observation (5) suggests that if a link has a bridge(s), or crosses county boundaries or railroads, its length is questionable when the GIS/NED approach is used. The following three sections describe the analyses and results used to address these issues.

Evaluation of the PMU Dataset

To explore the possibility that the PMU dataset might be erroneous the NCDOT conducted a field study for portions of the road sections that showed a large difference between GIS/NED length and PMU database length information. Personnel from the Pavement Management Unit were dispatched to recollect DMI lengths to obtain a true baseline mileage. The selection of links to be re-driven was arbitrary. The number of roads driven was 30, including 4 US routes and 26 NC routes. The number of links driven was 65. The mileage of the roads driven was 31.45 mile.

As mentioned earlier, the reason for doing the field study is the recognition that the PMU reference dataset itself has inherent errors. Without this step, it makes no sense to assess the accuracy of a method by comparing its results to an error-prone dataset without further investigation.

After the field data collection efforts, frequency analysis was conducted to reassess the accuracy of GIS/NED approach for these re-driven links. Table 6 shows the difference between the GIS/NED calculation and the newly obtained DMI lengths, and the difference between the PMU lengths and the newly obtained DMI lengths together for these newly driven road segments. The first row of Table 6 shows the error ranges. The second and the third rows show the number and the percentage of these newly driven links that have their differences (between GIS/NED length and DMI measurements) within the corresponding error ranges. The forth and fifth rows show the same information about the differences between the PMU dataset and DMI measurements.

TABLE 6 Frequency Analysis for Re-driven Links

Error Range (km)		= 0.0000	<= 0.0161	<= 0.0322	<= 0.0483	<= 0.1609	<= 0.3219	> 0.3219
Error Range (mile)		= 0.00	<= 0.01	<= 0.02	<= 0.03	<= 0.10	<= 0.20	> 0.20
GIS/NED vs. DMI	# of Link	25	33	35	36	53	57	5
	%	40.3%	53.2%	56.5%	58.1%	85.5%	91.9%	8.1%
PMU vs. DMI	# of Link	2	4	5	5	14	25	37
	%	3.2%	6.5%	7.7%	7.7%	22.6%	40.3%	59.7%

One thing that needs to be pointed out is that in re-driving these links, it happened three times that instead of reading a DMI mileage for each individual link, the total length of two adjacent links was acquired because for each pair of links, one link is very short. In the result table, these total of 6 links were presented as 3 link combinations and each link combination is counted once.

It is clear that the majority of these links have a difference between the GIS/NED calculation and DMI length that lies within the 0.03 error range. This tells us that the GIS/NED results are "in the ballpark." Alternatively most (92.3%) links have a difference between PMU length information and DMI length exceeding the 0.03 error range. Also, a majority of links have the latter difference bigger than 0.2 mile. These differences point to the lack of validity in the stored length measurements of these links in the PMU database.

As stated earlier, one assumption in this test case is that if there is a big difference between GIS/NED length and PMU length for a link, one of them is incorrect. From Table 6, it was observed that this incorrectness belongs to the PMU dataset. But regardless of its origin we have identified a bad link and it is these which are removed from the dataset.

However, revisiting Table 6 does not show a satisfactory accuracy for the GIS/NED length calculation itself when compared to the newly obtained DMI measurements. Again, taking a 0.03 error range as the threshold, only 58.1% of the links are acceptable when the more accurate reference data (DMI lengths) is used. The question as to what happened to these links is discussed in the next subsection.

Evaluation of the Effects of Bridges, County Boundaries, Railroad Crossings, and Dead Ends

As stated earlier, the accuracy of the GIS/NED approach is not satisfactory when considering a 0.03 mile error range. Field observation revealed that some of the re-driven links had bridges or crossed railroads. The common problem related to any links that are dead ends, have bridges, or cross county boundaries or railroads is to unambiguously pinpoint the start and/or end nodes on the orthorectified aerial photos. Thus, it is very easy for those nodes to be misplaced. Figure 1 illustrates the typical situation of these links.

In the case of a link that crosses county boundaries, the link is supposed to have either its start node or end node on the county boundary (this is one of the rules that was followed to generate links). However, due to an inability to exactly pinpoint the boundary because it is not a physical line, there is often disagreement (between field personnel who obtained the DMI measurements and the GIS personnel who digitized the links) about its precise physical location.

This situation also applies to links that cross railroads. Even though the field personnel should have no confusion about the physical position of railroads, it is still possible that the GIS personnel disagree regarding the exact start nodes or end nodes for such links because a railroad occupies certain width and there is also a distance between two parallel but close to each railroad lines.

In the case of a link (Link B in Figure 1) that has a bridge, the field personnel obtaining DMI measurements might treat this bridge as part of Link A instead of Link B. That is, because of the length of the bridge, the GIS personnel may place the center of the bridge at a different location than field personnel.

In all above cases, if the two adjoining links are combined and treated as one link, the accuracy of the overall distance should not be affected significantly. It is a matter of how the overall distance is proportioned to these two adjoining links that are combined together. As long as the overall length is correct it is not of importance to this study that the feature cannot be correctly located. We are not interested in the accuracy of their inventory. Rather, we are interested in the accuracy of a rapid way to calculate mileages, thus reducing the work that it takes to obtain them.

Dead ends, however, pose a unique and different situation. In the case of state maintained roads, most dead ends are not physical dead ends. In other words, a road will still continue out from the end node of a dead end. The dead end is simply a point that separates a state maintained road segment from a local street or other road. Using only an aerial photo there can be no assurance regarding the physical position of the end node of a dead end.

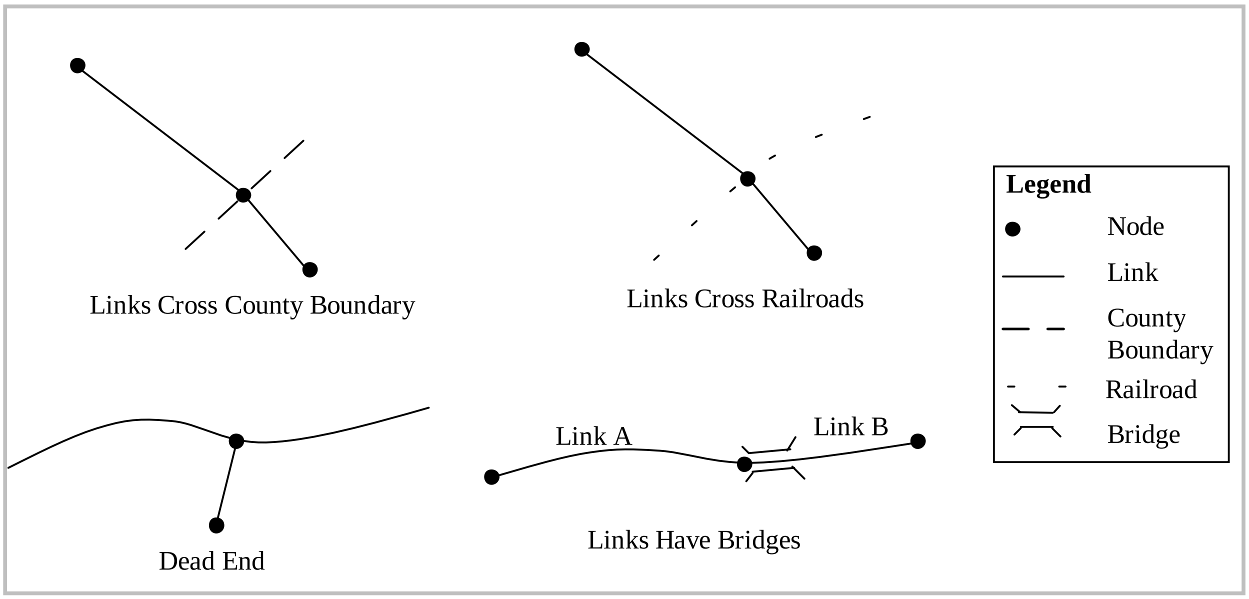

Figure 1 Illustration of Links, Nodes, Dead Ends, and the Relationship of Links and County Boundaries, Links and Railroads, and Links and Bridges

Combined Links

An important observation can be made when the DMI lengths and GIS/NED lengths are compared for combined links instead of individual links. The combining procedure was stated as: for adjoining links, if there is a bridge or railroad crossing between them (which leads to the inconsistency and incorrectness in identifying start and end nodes for these links) they are treated as a combined link with the sum of their lengths being used as the length for the combined link. This same procedure would also apply to links with county boundaries. However, since this study had its scope limited to one county, there was no way to correct the length measurements for the county boundary group of links. Thus they were omitted from the study and focus was placed on the bridge and railroad crossings.

Frequency analysis result for these newly derived combined road segments (49 total) are shown in Table 7, which is presented in the same format as Table 6 was.

It is observed from Table 7 that by combining adjoining links that either have bridges or railroad crossings, the percentage of links that are within the 0.03 mile error range increased significantly from 58.1% to 77.6%, when comparing GIS/NED results with DMI measurements. In addition, by examining the original comparison table (which shows detailed information for each individual link, and is not included in this paper due to the table size), it is also obvious that many links have bridges, or cross railroad or county boundaries [Rasdorf and Cai, 2001]. Unfortunately, some of these problematic links cannot be corrected using this procedure because the adjoining links were not measured using DMI. But these observations lead to the conclusion that for these problematic links, the problem usually originates from the failure of correctly and consistently identifying start and end nodes for them, not from the GIS/NED mileage generation process that is being assessed. These links are the links that need more effort and attention simply to improve the DOT inventory database. That is a separate issue from the mileage validation we are attempting to do.

TABLE 7 Frequency Analysis for Segments Derived from Re-driven Links

Error Range (km)		= 0.0000	<= 0.0161	<= 0.0322	<= 0.0483	<= 0.1609	<= 0.3219	> 0.3219
Error Range (mile)		= 0.00	<= 0.01	<= 0.02	<= 0.03	<= 0.10	<= 0.20	> 0.20
GIS/NED vs. DMI	Link#	22	30	35	38	45	46	3
	%	44.9%	61.2%	71.4%	77.6%	91.8%	93.9%	6.1%
PMU vs. DMI	Link#	2	5	5	6	15	23	26
	%	4.1%	10.2%	10.2%	12.2%	30.6%	46.9%	53.1%

STATEWIDE TEST CASE

The singular purpose of the statewide test was to find out how significant the difference is between the results from the GIS/NED approach and the existing length measurements of all state-maintained roads in North Carolina.

Table 8 Statewide Mileage Difference

Road Type	Original Length km (mile)	New Length km (mile)	Difference km (mile)	Percentage 1	Absolute Difference km (mile)	Percentage 2
Interstates	3670.54 (2280.83)	3674.82 (2283.49)	4.28 (2.66)	0.12%	4.28 (2.66)	0.12%
US Routes	11739.78 (7294.96)	11736.95 (7293.2)	-2.83 (-1.76)	-0.02%	2.83 (1.76)	0.02%
NC Routes	14071.01 (8743.56)	14016.50 (8709.69)	-54.51 (-33.87)	-0.39%	54.51 (33.87)	0.39%
SRs	108941.71 (67695.09)	108717.08 (67555.51)	-224.63 (-139.58)	-0.21%	224.63 (139.58)	0.21%
Totals	**138423.04 (86014.44)**	**138145.35 (85841.89)**	-277.68 (-172.55)	-0.20%	286.25 (177.87)	0.21%
***Primary Roads Only**	29481.33 (18319.35)	29428.27 (18286.38)	-53.06 (-32.97)	-0.18%	61.62 (38.29)	0.21%

In the test case, the GIS/NED mileage generation approach was applied to all state maintained roads in North Carolina. The differences of total mileages between the PMU dataset and GIS/NED computed distances for interstate highways, US routes, NC routes, secondary routes (SRs) are shown in Table 8. The *Original Length* refers to the length data from PMU database. The *New Length* refers to the GIS/NED computed distance. Both are in the units of miles. The

Difference is obtained by subtracting the *Original Length* from the *New Length* (which could be positive or negative). The *Percentage 1* is based on the *Difference* and the *Original Length* (PMU length). The *Absolute Difference* refers to the amount of *Difference* (only positive values). The *Percentage 2* is based on the *Absolute Difference* and the *Original Length*.

Observations obtained based on Table 8 include:

(1) Based on the difference in miles, the total difference between the PMU dataset and the GIS/NED generated distances for all state maintained roads in North Carolina is 172.55 miles out of about 18,300 miles, which is quite significant.

(2) Based on the difference percentage, the overall difference is only 0.18%. However, due to the large mileage of state maintained roads in North Carolina, even with this small percentage, the difference is significant as stated in Observation (1).

(3) When the absolute differences and their corresponding percentages are considered, the discrepancy is almost the same with a total of 177.87 miles and a percentage of 0.21%.

(4) According to the both difference percentages, NC routes are the group that shows the most significant discrepancy between the PMU dataset and the GIS/NED computed distances. The group of secondary roads follow the group of NC routes immediately with 139.58 miles and a percentage of 0.21%.

CONCLUSIONS

The study described in this paper resulted in two important groups of observations. First, it confirms the observations we found and conclusions we reached in the original interstate highway case study by verifying (based on Tables 1, 2, and 3) that:

(1) A GIS/NED approach generates sufficiently accurate surface lengths for all state maintained road sections if the road lines are photo-revised and they are correctly geographically positioned.

(2) A GIS/NED approach will significantly reduce mileage data collection costs, manpower needs, and time.

(3) A GIS/NED approach will establish a uniform level of accuracy across all links, route, and road mileages whereas now some mileages are very inaccurate while others are very accurate.

There are additional important observations from studies conducted and reported herein. It was observed that most errors occur in the links that either have bridges, that cross railroads or county boundaries, or are dead ends. It was concluded that these are the errors that belong to the Linework rather than to the GIS/NED approach. They could be accounted for and were removed from the accuracy assessment, thus establishing its validity. It was also very clear that without lines in their correct positions, the GIS/NED approach, or any other automated approach, is not at all an appropriate way to calculate distances. The limitations are thus clear.

This study also showed us that state road inventory databases are typically old and inaccurate for a variety of reasons. The approach used in this study can correct all the mileage inaccuracies in such databases to the level of accuracy reported herein. Again, the higher and uniform degree of accuracy makes this a very attractive approach for system-wide upgrade. Thus, DOTs should consider replacing all of their archival mileages with mileages generated using an approach like that advocated and verified herein, unless they have the resources to drive all the roads and physically measure their lengths (using DMI, GPS, or some other technology).

Our final comment is that extra attention should be paid to implementation strategies regarding how to use the mileages from the GIS/NED approach to update transportation spatial databases. A whole new set of problems emerge where different DOT divisions use different base linear referencing systems. Conversion problems between these referencing systems are significant.

The implications of this work are tremendous. It enables accurate road mileages to be obtained without physical distance measurement, thus saving significant dollars and labor hours. The direction application of this work is in transportation -- but there are broader implications for an any route-based application as well.

REFERENCES

Adams, M. T., Malaikrisanachalee, S., Blazquez, C., Lueck, S. and Vonderohe, A. (2002). "Eneterprise-Wide Data Integration and Analysis for Oversize/Overweight Permitting." *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, Volume 16, Number 1, pages 11-22, January.

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, March 5-8.

Anderson, J. M. and Mikhail E. M. (1998). *Surveying Theory and Practice*, 7th Edition, WCB/McGraw-Hill, Chapter 13, pages 742-779.

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 System*, Oxford University Press.

Environmental Systems Research Institute (ESRI). (2003). "Glossary of GIS Terms." http://www.esri.com/library/glossary/glossary.html, last updated May 7, 2003, visited May 7, 2003.

Geo Decisions. (1997). "Linear Referencing Systems, Review and Recommendation for GIS and Database Management Systems." unpublished material, Geo Decisions, a division of Gannett Fleming, Inc.

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.

Hummer, J., Karimi, H. A., and Khattak, A. (2000). "Collection and Presentation of Roadway and Inventory Data." *NCHRP Report 437*, Transportation Research Board, National Academy Press, Washington D.C.

Rasdorf, W., Cai, H., Tilley, C., Brun, S., and Robson, F. (2003). "Transportation Distance Measurement Data Quality." *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, Volume 17, Number 2, pages 75-87 April.

Rasdorf, W., Cai, H., Tilley, C., Brun, S., and Robson, F. (2004). "Accuracy Assessment of Interstate Highway Length Measurement Using DEM." *Journal of Surveying Engineering*, American Society of Civil Engineers. Volume 130, Number 3, pages 142-150, August.

Rasdorf, W., Janisch, A., Robson, F., and Tilley C. (2002). “Testing and Evaluating Alternative Algorithms for Location Referencing System Route Generation.” ***Journal of Computing in Civil Engineering***, American Society of Civil Engineers, Volume 16, Number 3, pages 194-205, July.

Rasdorf, W. and Cai, H. (2001). “A GIS Data Quality Case Study: Accuracy Comparison of Length Measurements Using GIS/NED and DMI.” ***Technical Report***, North Carolina State University, August.

USGS National Elevation Dataset. (2001). May 9, 2001. http://gisdata.usgs.net/ned/. Accessed May 26.

Yanalak, M. (2002). “Effect of Gridding Method on Digital Terrain Model Profile Data Based on Scattered Data.” ***Journal of Computing in Civil Engineering***, American Society of Civil Engineers, Volume 17, No. 1, pages 58-67, January.