

Abstract

CAI, HUBO Accuracy Evaluation of a 3-D Spatial Modeling Approach to Model Linear Objects and Predict Their Lengths. (Under the direction of Dr. William J. Rasdorf.)

Real world objects are three-dimensional. Numerous applications in geographic information systems (GISs) require modeling spatial objects in a 3-D space, but many current GISs only represent two-dimensional information. The GIS community has been struggling with solving complex problems dealing with 3-D objects using a 2-D approach.

This research focused on modeling linear objects in a 3-D space, predicting their 3-D distances, and evaluating the accuracy. A point model was developed, which modeled a 3-D line with a group of 3-D points (with X/Y/Z-coordinates) connected by straight lines. It required two input datasets, an elevation dataset and a planimetric line dataset. With elevation datasets in different formats (point data and digital elevation models (DEMs)), two approaches were proposed, differing in how the third dimension (elevation) was introduced. With point data, a snapping approach was developed. With DEMs, elevations for points uniformly distributed along planimetric lines were obtained via bilinear interpolations. Mathematical equations were derived to predict 3-D distances.

A case study was designed in the transportation field because of the rich source of linear objects and the criticality of 3-D distances in GIS-T and LRS. Two elevation datasets were used: LIDAR and national elevation dataset (NED). LIDAR datasets were further categorized into point data and DEMs (20-ft and 50-ft resolutions). Two intervals were taken to locate points planimetrically along lines when using DEMs (full cell size and half cell size). Consequently, each line was associated with seven calculated 3-D distances (one from LIDAR point data, two from LIDAR 20-ft DEM, two from LIDAR 50-ft DEM, and two from NED).

The accuracy of predicted 3-D distances was evaluated by comparing them to distance measurement instrument (DMI) measured distances. Errors were represented in two formats: difference and proportional difference (based on DMI measured distances) between the predicted 3-D distance and the DMI measured distance, taking

road types into consideration. Evaluation methods included descriptive statistics, error distribution histograms, hypothesis tests, frequency analysis, and root mean square of errors (RMSE). The effects from the use of different elevation datasets and intervals on the accuracy were evaluated via a sensitivity analysis. The effects from the geometric properties of linear objects on the accuracy were evaluated via significant factor analyses. Factors under consideration included distance, average slope and weighted slope, average slope change and weighted slope change, and the number and density of 3-D points. The usefulness of this research was proved by applying the resulting 3-D road centerlines to determine flooded road segments under flooding scenarios.

This research concluded that errors in the predicted 3-D distance varied with elevation datasets and road types, but not with the use of different intervals with the same elevation dataset, given the interval was less than or equal to the cell size. Using elevation datasets with higher vertical accuracies resulted in higher accuracies in predicted 3-D distances. In this research, using LIDAR point data improved the accuracy by 28% and using LIDAR DEMs improved the accuracy by 6%, compared to using NED data, with 100% RMSEs as the accuracy measure.

It was also concluded that there was a positive association between the error and any one of these factors from the aspect of the difference but a negative association from the aspect of the proportional difference. Each factor had a threshold, above which effects from the increase of the factor value were insignificant.

**ACCURACY EVALUATION OF A 3-D SPATIAL MODELING
APPROACH TO MODEL LINEAR OBJECTS AND PREDICT THEIR
LENGTHS**

by
HUBO CAI

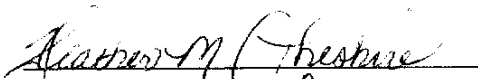
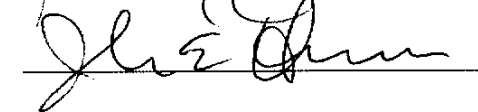
A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor in Philosophy



CIVIL ENGINEERING

Raleigh

2003

APPROVED BY:



Chair of Advisory Committee

BIOGRAPHY

Hubo Cai was born in China, on December 26, 1975. He received the degree of Bachelor of Science in Construction Management Engineering from Tongji University, Shanghai, in June 1998. Hubo came to the United States to pursue graduate studies in 1999. He spent a year at Iowa State University, as a Master student in Construction Engineering and Management. In year 2000, he transferred to North Carolina State University to continue his graduate studies, with an emphasis in Construction Engineering and Management. He obtained his Master of Civil Engineering from North Carolina State University on December 2001. After his graduation, Hubo decided to pursue a PhD degree in Civil Engineering, with his research focus moved to Computer Aided Engineering. He graduated with his PhD from North Carolina State University in May 2004. Hubo will work in the United States for a few years to accumulate working experiences and will go back to China in the near future.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude and appreciation to my advisor, Dr. William J. Rasdorf for his invaluable guidance and support throughout the research efforts and the writing of this dissertation. His sincerity and involvement in the work and his standards of excellence have been a constant source of inspiration. Sincere appreciation is extended to the members of my advisory committee, Dr. Joseph E. Hummer, Dr. Hugh A. Devine, and Dr. Heather M. Cheshire for the time and effort they have expended in assisting me by serving on the committee.

The assistance offered by the GIS unit of the North Carolina Department of Transportation is gratefully acknowledged. They provided equipment and data access to complete this research. The suggestions and help from Chris Tilley of the GIS unit of NCDOT are gratefully appreciated. Also, Mr. Worawat (former Master student in the Construction Engineering and Management) helped with collecting data for spatial measurement technologies. Amber Anderson (undergraduate student at NCSU) helped collect field data. Their time and efforts spent on this research are also gratefully appreciated.

Special thanks to my wife Jing for her patience, understanding, and support that saw me through this research.

I am pleased to dedicate this thesis to my parents and my little baby girl, Evelyn Cai, through whose prayers this work has come to fruition.

TABLE OF CONTENTS

	Page
LIST OF TABLES	xi
LIST OF FIGURES	xv
1 INTRODUCTION AND BACKGROUND	1
1.1 BACKGROUND	1
1.2 GIS	3
1.3 2-D, 2.5-D, AND 3-D MODELING	4
1.4 DATA QUALITY	7
1.5 GIS-T AND ROAD CENTERLINE	7
1.6 RESEARCH NEEDS	10
1.7 TERMS AND TERMINOLOGIES	12
1.8 ORGANIZATION OF THE DISSERTATION	20
2 RESEARCH PLAN	21
2.1 RESEARCH PROBLEM STATEMENT	21
2.2 OBJECTIVES OF THE RESEARCH	22
2.3 CASE STUDY DESCRIPTION	23
2.4 RESEARCH METHODOLOGY	25
2.5 BENEFITS AND SIGNIFICANCE OF THE RESEARCH	27
2.6 EXPECTED RESEARCH RESULTS	28
3 SPATIAL MODELING	29
3.1 DATA MODEL, DATA STRUCTURE, AND SPATIAL DATA MODEL	29
3.2 2-D AND 2.5-D SPATIAL MODELING	30
3.3 3-D SPATIAL MODELING	32
3.3.1 NATO Advanced Research Workshop in 1989	32
3.3.2 Applications of 3-D Spatial Modeling	37
3.3.3 Observations Regarding Research and Applications of 3-D Spatial Modeling around 1990	38
3.3.4 Recent Research Efforts in 3-D Spatial Modeling	39
3.3.5 Observations Regarding Recent Research Efforts in 3-D Spatial Modeling	43
3.4 SPATIAL MODELING IN TRANSPORTATION	44
3.4.1 Modeling Surfaces	44
3.4.1.1 Breaklines and Mass Points	45
3.4.1.2 Point Model	45
3.4.1.3 Contour Lines	46
3.4.1.4 DEM	47
3.4.1.5 TIN (Triangulated Irregular Network)	48
3.4.1.6 Researches in Terrain Modeling	49
3.4.2 Modeling Discrete Features	51

3.4.3	Modeling Transportation Networks	52
3.4.3.1	Linear Referencing Systems (LRS)	54
3.4.3.1.1	<i>LRS Data Model</i>	55
3.4.3.1.2	<i>Location Referencing Methods</i>	57
3.4.3.2	Current Research in LRS	59
3.4.3.3	Federal Efforts	62
3.4.4	3-D Spatial Modeling in GIS-T	64
4	DATA QUALITY	66
4.1	DEFINITION OF DATA QUALITY AND SPATIAL DATA ACCURACY	66
4.1.1	Data Quality	66
4.1.2	Errors, Accuracy, and Precision	67
4.1.3	Data Quality Parameters	69
4.1.4	Data Quality Standards	71
4.2	MODELING ERRORS	73
4.2.1	Error Sources	73
4.2.2	Error Analysis and Accuracy Evaluation	75
4.2.3	Error Propagations	83
4.3	DATA QUALITY ISSUES IN GIS-T	86
5	MEASUREMENT TECHNOLOGIES	88
5.1	INTRODUCTION	88
5.2	LENGTHS DERIVED FROM LEGACY DRAWINGS OR MAPS	88
5.3	GROUND SURVEYING	90
5.3.1	Background	90
5.3.2	Fundamentals of Total Station in Distance Measurement	91
5.3.2.1	Electronic Distance Measurement Instrument (EDMI)	91
5.3.2.2	Electronic Angle Measurement/Electronic Theodolite	93
5.3.2.3	On-board Microprocessor	93
5.3.2.4	Data Collector	94
5.3.2.5	Prisms	95
5.3.3	Data Quality Issues	95
5.3.3.1	Errors	95
5.3.3.1.1	<i>Instrumental Errors</i>	96
5.3.3.1.2	<i>Non-instrumental Errors</i>	96
5.3.3.2	Accuracy	97
5.3.4	Advantages and Disadvantages	98
5.4	GLOBAL POSITIONING SYSTEMS (GPSs)	99
5.4.1	Background	100
5.4.2	Fundamentals of GPS	100
5.4.2.1	Basic Principle	100
5.4.2.2	GPS Elements	101
5.4.2.2.1	<i>Satellites (Space Segment)</i>	101
5.4.2.2.2	<i>Receivers (User Segment)</i>	101

5.4.2.2.3	<i>Tracking Stations</i>	102
5.4.3	Data Quality Issues	103
5.4.4	Using GPS to Measure Distance along Linear Objects	105
5.4.5	Advantages and Disadvantages	106
5.5	DISTANCE MEASUREMENT INSTRUMENT (DMI)	106
5.5.1	Background	106
5.5.2	Fundamentals of DMI	107
5.5.2.1	Wheel Sensor	108
5.5.2.2	Transmission Sensor	108
5.5.2.3	Electronic Interface Amplifier	109
5.5.2.4	Calibration	109
5.5.3	Data Quality Issues	110
5.5.3.1	Errors	110
5.5.3.2	Accuracy	111
5.5.4	Advantages and Disadvantages	112
5.6	OBSERVATIONS	112
6	GLOBAL 3-D MODELING CONCEPTS FOR LINEAR OBJECTS AND LENGTH PREDICTION	114
6.1	ROAD GEOMETRY	114
6.2	ROAD CENTERLINE	115
6.3	3-D MODEL FOR ROAD CENTERLINE	116
6.3.1	Point Model	116
6.3.2	Mathematical Model	119
6.3.3	Relationship between Point Model and Mathematical Model	120
6.3.4	Model Selection	121
6.4	MODEL CONSTRUCTION	122
6.4.1	Source Data	123
6.4.1.1	USGS DEM AND NATIONAL ELEVATION DATASET (NED)	124
6.4.1.1.1	<i>USGS DEMs</i>	124
6.4.1.1.2	<i>NED</i>	137
6.4.1.2	Light Detection and Ranging (LIDAR)	140
6.4.1.2.1	<i>Background</i>	140
6.4.1.2.2	<i>Fundamentals of LIDAR</i>	140
6.4.1.2.3	<i>Data Collection and Generation of End Product</i>	143
6.4.1.2.4	<i>Advantages</i>	144
6.4.1.2.5	<i>Limitations</i>	145
6.4.1.2.6	<i>Errors and Accuracy</i>	146
6.4.1.2.7	<i>LIDAR in North Carolina</i>	148
6.4.2	Construction of 3-D Point Model	148
6.4.3	Construction of 3-D Mathematical Model	149
6.5	DISTANCE PREDICTION	150
7	COMPUTATIONAL IMPLEMENTATIONS OF 3-D MODELS	152
7.1	DEVELOPMENT ENVIRONMENT	152
7.2	OBTAINING 3-D POINTS ALONG LINEAR OBJECTS	154

7.2.1	Obtaining Planimetric Position	154
7.2.2	Obtaining Elevations	156
7.2.2.1	Obtaining Elevations from LIDAR Point Clouds	156
7.2.2.1.1	<i>Interpolation Approach</i>	157
7.2.2.1.2	<i>Approximation Approach</i>	159
7.2.2.2	Obtaining Elevations from LIDAR DEM	164
7.2.2.3	Obtaining Elevations from USGS DEMs and NED	164
7.2.3	Deriving Mathematical Function and Adjusting 3-D Points	166
7.3	PREDICTING 3-D DISTANCE	169
7.4	ALGORITHMS	169
7.4.1	Pre-Processing	170
7.4.2	Processing	170
7.4.2.1	Processing Procedure with LIDAR Point Data and Road Network Data	171
7.4.2.2	Processing Procedure with DEM or NED Data and Road Network Data	172
7.5	ACCURACY ASSESSMENT	174
8	CASE STUDY	176
8.1	STUDY DEFINITION	176
8.1.1	Study Problem Statement	176
8.1.2	Study Objectives	178
8.1.3	Study Scope	178
8.1.4	General Approach	181
8.2	CASE STUDY INFORMATION SOURCES AND DATA COLLECTION	182
8.2.1	Road Network	182
8.2.1.1	Digitizing Road Centerlines from DOQQs	182
8.2.1.2	DOQQs	184
8.2.1.3	DOQQs in North Carolina	184
8.2.1.4	Description of the Road Centerline Data	186
8.2.1.5	Errors and Accuracy	187
8.2.2	Elevation Data	191
8.2.2.1	NED	191
8.2.2.1.1	<i>Data Collection and Description</i>	191
8.2.2.1.2	<i>Errors and Accuracy</i>	192
8.2.2.2	LIDAR	193
8.2.2.2.1	<i>Data Collection and Description</i>	194
8.2.2.2.2	<i>Errors and Accuracy</i>	195
8.2.3	Reference Data (DMI Data)	198
8.2.3.1	Data Collection	198
8.2.3.2	Errors and Accuracy	200
8.3	PRE-PROCESSING	202
8.3.1	Pre-Processing Road Centerline Data	202
8.3.2	Pre-Processing LIDAR Data	205
8.3.2.1	Pre-Processing LIDAR Mass Point Data	206
8.3.2.2	Pre-Processing LIDAR DEMs	207
8.3.3	Pre-Processing NED Data	208
8.4	MODELING ROAD CENTERLINES IN 3-D	208

8.4.1	Using LIDAR Points	208
8.4.1.1	Snapping	209
8.4.1.2	Obtaining Elevations at Start/End Points	211
8.4.1.3	Quality Control	215
8.4.1.4	Improvement	217
8.4.2	Using LDIAR DEMs	219
8.4.3	Using NED	222
8.5	3-D DISTANCE PREDICTION	223
9	ACCURACY EVALUATION AND ERROR ANALYSIS	226
9.1	COMBINATION OF FTSEGS	226
9.2	STATISTICAL METHODS APPLIED TO THE CASE STUDY	228
9.2.1	Describing Samples	228
9.2.2	Statistical Inferences	230
9.2.3	ANOVA	231
9.2.4	100% RMSE and 95% RMSE	233
9.2.5	Frequency Analysis	234
9.3	ACCURACY EVALUATION RESULTS	234
9.3.1	General Information	234
9.3.2	Accuracy Assessment	236
9.3.2.1	Descriptive Statistics	237
9.3.2.2	Histograms	242
9.3.2.3	Hypothesis Tests and Confidence Intervals	251
9.3.2.4	100% and 95% RMSEs	259
9.3.2.5	Frequency Analysis	268
9.4	RESULT ANALYSIS AND SENSITIVITY ANALYSIS	275
9.4.1	Comparison of Means	277
9.4.1.1	ANOVA	277
9.4.1.2	Comparison of 100% Means	280
9.4.1.3	Comparison of 95% Means	283
9.4.2	Comparison of Medians	285
9.4.3	Comparison of Absolute Means	288
9.4.4	Comparison of RMSEs	292
9.4.4.1	Comparison of 100% RMSEs	293
9.4.4.2	Comparison of 95% RMSEs	296
9.4.5	Comparison of Frequencies	298
9.4.5.1	Comparison of Frequencies for Differences	299
9.4.5.2	Comparison of Frequencies for Proportional Differences	305
9.5	ERROR ANALYSIS AND ERROR PROPAGATION	313
9.6	CONCLUSIONS OF ACCURACY ASSESSMENT AND SENSITIVITY ANALYSIS	316
10	IDENTIFICATION OF SIGNIFICANT FACTORS	320
10.1	INTRODUCTION	320
10.2	FACTORS UNDER CONSIDERATION	320
10.3	ANALYSIS METHODS	325

10.4 SIGNIFICANCE DETERMINATION	327
10.4.1 Sample Correlation Coefficient and Sample Coefficient of Determination	327
10.4.1.1 Distance	327
10.4.1.2 Average Slope	330
10.4.1.3 Weighted Average Slope	332
10.4.1.4 Average Slope Change	335
10.4.1.5 Weighted Average Slope Change	337
10.4.1.6 Number of 3-D Points and Average Density of 3-D Points	340
10.4.2 Grouping and Comparison	343
10.4.2.1 Distance	343
10.4.2.2 Average Slope	346
10.4.2.3 Weighted Average Slope	350
10.4.2.4 Average Slope Change	353
10.4.2.5 Weighted Average Slope Change	356
10.4.2.6 Number of 3-D Points and Average Density of 3-D Points	359
10.5 CONCLUSIONS AND LIMITATIONS	362
11 USE OF THE MODEL TO ASSESS HIGHWAY FLOODING	367
11.1 INTRODUCTION	367
11.2 FLOOD EXTENT AND DEPTH DETERMINATION AND FLOODED ROAD SEGMENT IDENTIFICATION	368
11.3 MODELS AND ALGORITHMS	370
11.3.1 Conceptual Model	370
11.3.1.1 The Flood Extent Prediction Model	370
11.3.1.2 The Flooded Road Segment Identification Model	373
11.3.2 Computational Model	374
11.3.2.1 Development Environment	374
11.3.2.2 Data Sources	374
11.3.2.3 Pre-Processing	375
11.3.2.4 Algorithms for Flood Extent and Depth Prediction	375
11.3.2.5 Enforcement of Constraints	378
11.3.2.6 Algorithms for Identifying Flooded Road Segments	381
11.4 TESTING	384
11.4.1 Study Area	384
11.4.2 Results	385
11.5 LIMITATIONS AND CONCLUSIONS	389
12 SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	390
12.1 SUMMARY	390
12.2 CONCLUSIONS	395
12.3 KEY FINDINGS	401
12.4 RECOMMENDATIONS	402
12.5 FUTURE WORK	404
REFERENCES	406

APPENDICES	419
APPENDIX A PROGRAM CODE FOR EXTRACTION OF FTSEGS FROM THE LINK-NODE SYSTEM	420
APPENDIX B PROGRAM CODE FOR SNAPPING LIDAR POINTS TO LINEAR FEATURES	423
APPENDIX C PROGRAM CODE FOR WORKING WITH DEMS TO OBTAIN 3-D POINTS ALONG LINES	427
APPENDIX D PROGRAM CODE FOR CALCULATIONS OF AVERAGE SLOPE, WEIGHTED AVERAGE SLOPE, AVERAGE SLOPE CHANGE, AND WEIGHTED AVERAGE SLOPE CHANGE	433
APPENDIX E PROGRAM CODE FOR CHECKING THE ERROR TOLERANCE AND THE MAXIMUM DROP AND MAKING ADJUSTMENTS	436
APPENDIX F PROGRAM CODE FOR FLOOD EXTENT PREDICTION	440
APPENDIX G PROGRAM CODE FOR FLOODED ROAD SEGMENT IDENTIFICATION	447

LIST OF TABLES

Page

Table 1.1	The Significant Difference in Lengths between 2-D Approach and 3-D Approach	11
Table 3.1	Link Table	53
Table 3.2	Impedance Table	53
Table 4.1	A Sample Error Matrix	80
Table 5.1	Summary of GPS Error Sources and Error Budget (Worawat and Rasdorf 2003)	105
Table 5.2	Accuracy and Application (Worawat and Rasdorf 2003)	105
Table 5.3	Accuracy of DMIs (Based on Market Models)	111
Table 8.1	Distribution of State-Maintained Roads in the Study Scope	181
Table 8.2	Available Digital Elevation Data Products from the North Carolina Floodplain Mapping Program	194
Table 8.3	Accuracy Assessment Results for Durham, Johnston, and Wake	197
Table 8.4	Distribution of FTSegs in the Study Scope	205
Table 8.5	Average Density of 3-D Points before and after Further Improvement	219
Table 8.6	Sample 3-D Point Attributes (from LIDAR Point Data)	223
Table 8.7	Sample 3-D Point Attributes (from LIDAR 20-ft DEM with a 20-ft Interval)	224
Table 9.1	FTSeg Distribution before and after Combination	228
Table 9.2	ANOVA Result	232
Table 9.3	Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (LIDAR Point Data) and DMI Measurements	237
Table 9.4	Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 10-ft Interval) and DMI Measurements	238
Table 9.5	Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 20-ft Interval) and DMI Measurements	239
Table 9.6	Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 25-ft Interval) and DMI Measurements	240
Table 9.7	Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 50-ft Interval) and DMI Measurements	240
Table 9.8	Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (30-m NED with 15-m Interval) and DMI Measurements	241
Table 9.9	Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (30-m NED with 30-m Interval) and DMI Measurements	241
Table 9.10	Naming Schema	243
Table 9.11	The Criteria for α -level Hypothesis Tests (Rao 1998)	252
Table 9.12	Hypothesis Test Results (LIDAR Point Data)	253
Table 9.13	Hypothesis Test Results (LIDAR 20-ft DEM, 10-ft Interval)	254
Table 9.14	Hypothesis Test Results (LIDAR 20-ft DEM, 20-ft Interval)	255
Table 9.15	Hypothesis Test Results (LIDAR 50-ft DEM, 25-ft Interval)	256
Table 9.16	Hypothesis Test Results (LIDAR 50-ft DEM, 50-ft Interval)	257
Table 9.17	Hypothesis Test Results (NED 30-m DEM, 15-m Interval)	258
Table 9.18	Hypothesis Test Results (NED 30-m DEM, 30-m Interval)	259
Table 9.19	100% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR Point Data) and DMI Measurements	261
Table 9.20	95% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR Point Data) and DMI Measurements	261
Table 9.21	100% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 10-ft Interval) and DMI Measurements	262
Table 9.22	95% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 10-ft Interval) and DMI Measurements	262
Table 9.23	100% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 20-ft Interval) and DMI Measurements	263
Table 9.24	95% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 20-ft Interval) and DMI Measurements	263

Table 9.25	100% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 25-ft Interval) and DMI Measurements	264
Table 9.26	95% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 25-ft Interval) and DMI Measurements	264
Table 9.27	100% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 50-ft Interval) and DMI Measurements	265
Table 9.28	95% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 50-ft Interval) and DMI Measurements	265
Table 9.29	100% RMSE for Differences and Proportional Differences between 3-D Distances (30-m NED with 15-m Interval) and DMI Measurements	266
Table 9.30	95% RMSE for Differences and Proportional Differences between 3-D Distances (30-m NED with 15-m Interval) and DMI Measurements	266
Table 9.31	100% RMSE for Differences and Proportional Differences between 3-D Distances (30-m NED with 30-m Interval) and DMI Measurements	267
Table 9.32	95% RMSE for Differences and Proportional Differences between 3-D Distances (30-m NED with 30-m Interval) and DMI Measurements	267
Table 9.33	Results of Frequency Analysis (LIDAR Point Data)	269
Table 9.34	Results of Frequency Analysis (LIDAR 20-ft DEM, 10-ft Interval)	270
Table 9.35	Results of Frequency Analysis (LIDAR 20-ft DEM, 20-ft Interval)	271
Table 9.36	Results of Frequency Analysis (LIDAR 50-ft DEM, 25-ft Interval)	272
Table 9.37	Results of Frequency Analysis (LIDAR 50-ft DEM, 50-ft Interval)	273
Table 9.38	Results of Frequency Analysis (NED, 15-m Interval)	274
Table 9.39	Results of Frequency Analysis (NED, 30-m Interval)	275
Table 9.40	Result of ANOVA for 100% Means (Differences, All Samples Considered)	278
Table 9.41	Result of ANOVA for 100% Means (Proportional Differences, All Samples Considered)	278
Table 9.42	Results of ANOVA for Pairwise Comparisons of 100% Means for Differences	279
Table 9.43	Summary of 100% Sample Means of Differences	280
Table 9.44	Summary of 100% Sample Means of Proportional Differences	282
Table 9.45	Summary of 95% Sample Means of Differences	283
Table 9.46	Summary of 95% Sample Means of Proportional Differences	285
Table 9.47	Summary of Medians of Differences	286
Table 9.48	Summary of Medians of Proportional Differences	287
Table 9.49	Summary of 100% Absolute Means for Differences	288
Table 9.50	Summary of 95% Absolute Means for Differences	289
Table 9.51	Summary of 100% Absolute Means for Proportional Differences	290
Table 9.52	Summary of 95% Absolute Means for Proportional Differences	291
Table 9.53	Summary of 100% RMSEs for Differences	293
Table 9.54	Summary of 100% RMSEs for Proportional Differences	295
Table 9.55	Summary of 95% RMSEs for Differences	296
Table 9.56	Summary of 95% RMSEs for Proportional Differences	297
Table 9.57	Summary of Percentages for Differences ([-5, 5])	299
Table 9.58	Summary of Percentages for Differences ([-10, 10])	300
Table 9.59	Summary of Percentages for Differences ([-20, 20])	301
Table 9.60	Summary of Percentages for Differences ([-30, 30])	302
Table 9.61	Summary of Percentages for Differences ([-50, 50])	303
Table 9.62	Summary of Percentages for Differences($(-\infty, -50)$ and $(50, +\infty)$)	304
Table 9.63	Summary of Percentages for Proportional Differences ([-1, 1])	305
Table 9.64	Summary of Percentages for Proportional Differences ([-5, 5])	306
Table 9.65	Summary of Percentages for Proportional Differences ([-10, 10])	307
Table 9.66	Summary of Percentages for Proportional Differences ([-20, 20])	308
Table 9.67	Summary of Percentages for Proportional Differences ([-30, 30])	309
Table 9.68	Summary of Percentages for Proportional Differences ([-50, 50])	310
Table 9.69	Summary of Percentages for Proportional Differences ([-100, 100])	312
Table 9.70	Summary of Percentages for Differences($(-\infty, -100)$ and $(100, +\infty)$)	313
Table 9.71	RMSEs from Comparing 2-D Lengths with DMI Measured Distances	319

Table 10.1	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Difference with the Factor of Distance	327
Table 10.2	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Difference with the Factor of Distance	328
Table 10.3	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Proportional Difference with the Factor of Distance	329
Table 10.4	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Proportional Difference with the Factor of Distance	329
Table 10.5	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Difference with the Factor of the Average Slope	330
Table 10.6	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Difference with the Factor of the Average Slope	331
Table 10.7	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Proportional Difference with the Factor of the Average Slope	331
Table 10.8	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Proportional Difference with the Factor of the Average Slope	332
Table 10.9	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Difference with the Factor of the Weighted Average Slope	333
Table 10.10	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Difference with the Factor of the Weighted Average Slope	333
Table 10.11	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Proportional Difference with the Factor of the Weighted Average Slope	334
Table 10.12	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Proportional Difference with the Factor of the Weighted Average Slope	334
Table 10.13	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Difference with the Factor of the Average Slope Change	335
Table 10.14	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Difference with the Factor of the Average Slope Change	336
Table 10.15	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Proportional Difference with the Factor of the Average Slope Change	336
Table 10.16	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Proportional Difference with the Factor of the Average Slope Change	337
Table 10.17	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Difference with the Factor of the Weighted Average Slope Change	338
Table 10.18	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Difference with the Factor of the Weighted Average Slope Change	338
Table 10.19	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Proportional Difference with the Factor of the Weighted Average Slope Change	339
Table 10.20	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Proportional Difference with the Factor of the Weighted Average Slope Change	339
Table 10.21	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination with the Factor of the Number of 3-D Points for Errors of Using LIDAR Point Data	341
Table 10.22	Summary of Sample Correlation Coefficients and Sample Coefficients of Determination with the Factor of the Average Density of 3-D Points for Errors of Using LIDAR Point Data	342
Table 10.23	Illustration of Groups Based on the Distance	343
Table 10.24	Summary of RMSEs from the Difference for Groups Based on the Distance	343
Table 10.25	Summary of RMSEs from the Proportional Difference for Groups Based on the Distance	344
Table 10.26	Illustration of Groups Based on the Average Slope	346
Table 10.27	Summary of RMSEs from the Difference for Groups Based on the Average Slope	346
Table 10.28	Summary of RMSEs from the Proportional Difference for Groups Based on the Average Slope	348
Table 10.29	Illustration of Groups Based on the Weighted Average Slope	350

Table 10.30	Summary of RMSEs from the Difference for Groups Based on the Weighted Average Slope	350
Table 10.31	Summary of RMSEs from the Proportional Difference for Groups Based on the Weighted Average Slope	351
Table 10.32	Illustration of Groups Based on the Average Slope Change	353
Table 10.33	Summary of RMSEs from the Difference for Groups Based on the Average Slope Change	353
Table 10.34	Summary of RMSEs from the Proportional Difference for Groups Based on the Average Slope Change	354
Table 10.35	Illustration of Groups Based on the Weighted Average Slope Change	356
Table 10.36	Summary of RMSEs from the Difference for Groups Based on the Weighted Average Slope Change	356
Table 10.37	Summary of RMSEs from the Proportional Difference for Groups Based on the Weighted Average Slope Change	357
Table 10.38	Illustration of Grouping Based on the Number of 3-D Points	359
Table 10.39	Illustration of Grouping Based on the Average Density of 3-D Points	359
Table 10.40	Summary of the RMSEs for Groups Based on the Number of 3-D Points	360
Table 10.41	Summary of the RMSEs for Groups Based on the Average Density of 3-D Points	361
Table 11.1	Part of the Attribute Table of the Identified Flooded Road Segments	387

LIST OF FIGURES

Page

Figure 1.1	The Procedure of Map Projections	5
Figure 1.2	The Differences in Representing a Linear Objects in 2-D, 2.5-D, and 3-D spaces	6
Figure 1.3	GIS-T, Product of an Enhanced GIS and an Enhanced TIS (Vonderohe et al. 1993)	8
Figure 1.4	The Concepts of Horizontal Alignment and Vertical Alignment	9
Figure 1.5	Typical Cross Section in Cut with Table Drain (Khisty 1999)	9
Figure 1.6	Definition of Road Centerline	10
Figure 3.1	Representation of an Object with n=2 (Modified after Gargantini 1989)	34
Figure 3.2	Octree for Object in Figure 3.1 (Modified after Gargantini 1989)	34
Figure 3.3	Classification of 3-D Geometric Representations (Li 1994)	41
Figure 3.4	Surface Based 3-D Geometric Representations (Li 1994)	41
Figure 3.5	Volume Based 3-D Geometric Representations (Li 1994)	42
Figure 3.6	The Point Model for a Pyramid	46
Figure 3.7	The Contour Line Model for a Pyramid	46
Figure 3.8	The DEM for the Pyramid	47
Figure 3.9	The TIN Model for the Pyramid	48
Figure 3.10	A Simple Network Model	53
Figure 3.11	Conceptual Overview of the LRS Data Model (Modified after Vonderohe et al. 1995)	55
Figure 3.12	The Conceptual Model of a Linear Referencing System (Summarized after Vonderohe et al. 1995)	56
Figure 3.13	Milepost/Milepoint/Reference Post Location Reference Method (Modeled after Geo Decisions 1997)	57
Figure 3.14	Link-Node Location Reference Method (Modeled after Geo Decisions 1997)	58
Figure 3.15	X-Y Point Location Reference Method (Modeled after Geo Decisions 1997)	58
Figure 3.16	Address/Intersection Location Reference Method (Modeled after Geo Decisions 1997)	59
Figure 4.1	Data Quality Parameter Matrix	70
Figure 4.2	Static and Operational States of Data Quality Evaluation (Modeled after Wu and Buttenfield 1994)	70
Figure 4.3	Measuring Components of Positional Accuracy for Point (Modeled after Veregin 1999)	76
Figure 4.4	The Early Model of Positional Accuracy for Lines (Modeled after Goodchild 1991a)	77
Figure 4.5	A Practical Error Evaluation Method for Lines	77
Figure 4.6	Variations in Band Shape and Error Distribution (modeled after Veregin 1999)	78
Figure 4.7	Illustration of the Dilemma in Determining Positional Accuracy for Polygons	78
Figure 4.8	The Illustration of New Data Derivation	83
Figure 4.9	Example of Errors in LRS (Modeled after Miller and Shaw 2001)	87
Figure 5.1	Add-On Electronic Distance Measuring Instrument Mounted on a Theodolite (Photo Courtesy of Sokkia Corporation)	90
Figure 5.2	Current Total Station, Trimble 3600 Series (Photo Courtesy of Trimble Navigation Limited)	90
Figure 5.3	Principles of EDM Measurement (Courtesy of Kern Instruments-Leica) (Kavanagh 1997)	91
Figure 5.4	The Horizontal Plane for Angle Measurement (Schofield 1993)	93
Figure 5.5	Illustration of Distance Calculation in Total Station (Moffitt and Bossler 1998)	94
Figure 5.6	Data Collector (Courtesy of Trimble Navigation Limited)	95
Figure 5.7	Prisms (Courtesy of Topcon Instrument Corp., Paramus, NJ)	95
Figure 5.8	Plan View Showing Collimation Error (Worawat and Rasdorf 2003)	96
Figure 5.9	Front View Showing Transit Axis Error (Worawat and Rasdorf 2003)	96
Figure 5.10	Relation between Angular and Linear Error (Moffitt and Bossler 1998)	98
Figure 5.11	Approximation of 3-D Distance between Two Locations via Segmentation	99
Figure 5.12	Three Major Segments of GPS (Source: http://www.aero.org/publications/GPSPRIMER)	101
Figure 5.13	Satellite Constellation (Source: http://www.garmin.com)	101
Figure 5.14	GPS Master Control and Monitor Station Network (Source: Peter H. Dana 5/27/1995)	102
Figure 5.15	Illustration of Multipath Errors (Source: http://www.garmin.com)	104
Figure 5.16	Hand Wheel (Worawat and Rasdorf 2003)	107
Figure 5.17	DMI Attached to Car Dash Board (Courtesy of Nu-Metrics)	107
Figure 5.18	Using DMI to Calculate a Travel Distance (Worawat and Rasdorf 2003)	107

Figure 5.19	Wheel Sensor (Courtesy of Nu-Metrics)	109
Figure 5.20	Transmission Sensor (Courtesy of Nu-Metrics)	109
Figure 6.1	Line Construction in 2-D GIS	117
Figure 6.2	Line Construction in 3-D GIS	117
Figure 6.3	A Variation of 3-D Point Model with the Use of LRS	118
Figure 6.4	An Example of Mathematical Model in the Case of LRS Road Data	120
Figure 6.5	Illustration of the Basic Structure of a Grid (Modified after Maune et al. 2001)	126
Figure 6.6	Conversion of NED Data into a Grid File (Modeled after Rasdorf et al. 2003a)	139
Figure 6.7	Illustration of Electromagnetic Spectrum (Worawat and Rasdorf 2003)	141
Figure 6.8	Components of a LIDAR System	141
Figure 6.9	Operation of Airborne LIDAR in Mapping (Courtesy of NC Floodplain Mapping Program)	141
Figure 7.1	ArcMap Displaying X/Y Coordinates for a Point on a Line	154
Figure 7.2	Illustration of the Spatial Properties and Spatial Relations for Points and a Linear Object	155
Figure 7.3	Three Groups of LIDAR Points Based on the Spatial Relationships	157
Figure 7.4	Illustration of Obtaining Intersections and Determining Their Elevations	158
Figure 7.5	Cross-Sectional View of Comparing the Vertical Errors Due to Interpolation and Approximation	160
Figure 7.6	Comparison of the Interpolation Approach with the Approximation Approach	161
Figure 7.7	Illustration of the Density of LIDAR Points in Johnston County, North Carolina	163
Figure 7.8	Bilinear Interpolation of Obtaining Elevations from DEMs (Rasdorf et al. 2003b)	165
Figure 7.9	Obtaining Intermediate Points along a Linear Object	165
Figure 7.10	Defining Best-Fitting Line	167
Figure 7.11	Maximum Vertical Adjustments	169
Figure 7.12	Processing Procedure for 3-D Spatial Modeling and 3-D Distance Prediction when Working with LIDAR Point Data and Road Network Data	172
Figure 7.13	Processing Procedure for 3-D Spatial Modeling and 3-D Distance Prediction when Working with DEM Data	173
Figure 8.1	Availability of LIDAR Data in North Carolina	179
Figure 8.2	Study Scope for the Case Study	180
Figure 8.3	Detailed View of the Study Scope for the Case Study	180
Figure 8.4	Digitizing Road Centerlines from DOQQs	183
Figure 8.5	Illustration of Intersection, Link, and Node	187
Figure 8.6	Illustration of Errors in Digitizing Linear Features	188
Figure 8.7	Illustration of Dilemmas in Digitizing Road Centerlines	189
Figure 8.8	Illustration of Misalignment between Sign Locations and County Lines	190
Figure 8.9	RMSE Trends	197
Figure 8.10	Illustration of Errors due to Stopping on Shoulders at Start and End Points	200
Figure 8.11	Quality Control in Identifying Start or End Points When Collecting DMI Measurements	201
Figure 8.12	Algorithm for Deriving FTSegs from Link-Node System	205
Figure 8.13	Illustration of Reducing the Number of LIDAR Points by Applying a Buffer – Part I	206
Figure 8.14	Illustration of Reducing the Number of LIDAR Points by Applying a Buffer – Part II	207
Figure 8.15	An Illustration of the Algorithm for Snapping	209
Figure 8.16	Illustration of the Dilemma in Determining Elevations for Start and End Points	211
Figure 8.17	Illustration of Sample Results from Snapping	211
Figure 8.18	Illustration of Interpolation, Extrapolation, and Weighted Average Scenarios for End Points	213
Figure 8.19	Illustration of Connected FTSegs with Different Road Types	214
Figure 8.20	An Example of 3-D Points Simulating the Road Centerline	215
Figure 8.21	Illustration of Suspect Points	216
Figure 8.22	Illustration of a Typical Scenario Associated with Suspect Points	217
Figure 8.23	Illustration the Scenario for Further Improvement	218
Figure 8.24	Algorithm of Working with LIDAR DEMs	221
Figure 8.25	Algorithm for 3-D Distance Prediction	225
Figure 9.1	The Problem Associated with Accuracy Evaluation for FTSegs Touching County Boundaries	227
Figure 9.2	Grouping Errors into Samples	236
Figure 9.3	Distribution Histograms (LIDAR Point Data, Differences)	244

Figure 9.4	Distribution Histograms (LIDAR Point Data, Proportional Differences)	244
Figure 9.5	Distribution Histograms (LIDAR 20-ft DEM, 10-ft Interval, Differences)	245
Figure 9.6	Distribution Histograms (LIDAR 20-ft DEM, 10-ft Interval, Proportional Differences)	245
Figure 9.7	Distribution Histograms (LIDAR 20-ft DEM, 20-ft Interval, Differences)	246
Figure 9.8	Distribution Histograms (LIDAR 20-ft DEM, 20-ft Interval, Proportional Differences)	246
Figure 9.9	Distribution Histograms (LIDAR 50-ft DEM, 25-ft Interval, Differences)	247
Figure 9.10	Distribution Histograms (LIDAR 50-ft DEM, 25-ft Interval, Proportional Differences)	247
Figure 9.11	Distribution Histograms (LIDAR 50-ft DEM, 50-ft Interval, Differences)	248
Figure 9.12	Distribution Histograms (LIDAR 50-ft DEM, 50-ft Interval, Proportional Differences)	248
Figure 9.13	Distribution Histograms (NED, 15-m Interval, Differences)	249
Figure 9.14	Distribution Histograms (NED, 15-m Interval, Proportional Differences)	249
Figure 9.15	Distribution Histograms (NED, 30-m Interval, Differences)	250
Figure 9.16	Distribution Histograms (NED, 30-m Interval, Proportional Differences)	250
Figure 9.17	Comparisons of 100% Sample Means of Differences	281
Figure 9.18	Comparisons of 100% Sample Means of Proportional Differences	283
Figure 9.19	Comparisons of 95% Sample Means of Differences	284
Figure 9.20	Comparisons of 95% Sample Means of Proportional Differences	285
Figure 9.21	Comparison of Medians of Differences	286
Figure 9.22	Comparison of Medians of Proportional Differences	287
Figure 9.23	Comparison of 100% Absolute Means of Differences	289
Figure 9.24	Comparison of 95% Absolute Means of Differences	290
Figure 9.25	Comparison of 100% Absolute Means of Proportional Differences	291
Figure 9.26	Comparison of 95% Absolute Means of Proportional Differences	292
Figure 9.27	Comparison of 100% RMSEs for Differences	294
Figure 9.28	Comparison of 100% RMSEs for Proportional Differences	295
Figure 9.29	Comparison of 95% RMSEs for Differences	296
Figure 9.30	Comparison of 95% RMSEs for Proportional Differences	297
Figure 9.31	Illustration of Percentages at the Error Range of [-5, 5] for Differences	299
Figure 9.32	Illustration of Percentages at the Error Range of [-10, 10] for Differences	300
Figure 9.33	Illustration of Percentages at the Error Range of [-20, 20] for Differences	301
Figure 9.34	Illustration of Percentages at the Error Range of [-30, 30] for Differences	303
Figure 9.35	Illustration of Percentages at the Error Range of [-50, 50] for Differences	304
Figure 9.36	Illustration of Percentages at the Error Range of $(-\infty, -50)$ and $(50, +\infty)$ for Differences	305
Figure 9.37	Illustration of Percentages at the Error Range of [-1, 1] for Proportional Differences	306
Figure 9.38	Illustration of Percentages at the Error Range of [-5, 5] for Proportional Differences	307
Figure 9.39	Illustration of Percentages at the Error Range of [-10, 10] for Proportional Differences	308
Figure 9.40	Illustration of Percentages at the Error Range of [-20, 20] for Proportional Differences	309
Figure 9.41	Illustration of Percentages at the Error Range of [-30, 30] for Proportional Differences	310
Figure 9.42	Illustration of Percentages at the Error Range of [-50, 50] for Proportional Differences	311
Figure 9.43	Illustration of Percentages at the Error Range of [-100, 100] for Proportional Differences	312
Figure 9.44	Illustration of the Effects from the Point Positional Errors on the Predicted 3-D Distances	314
Figure 9.45	Illustration of the Effects of the 3-D model on the Predicted 3-D Distances	314
Figure 9.46	Illustration of the Difference between the True 3-D Line and the Constructed 3-D Line	315
Figure 10.1	Illustration of Slope and Slope Change Calculation	322
Figure 10.2	The Algorithm of Calculating Average Slopes, Weighted Average Slopes, Slope Changes, and Weighted Slope Changes	324
Figure 10.3	Comparison of RMSEs of the Difference for Groups Based on the Distance	344
Figure 10.4	Comparison of RMSEs of the Proportional Difference for Groups Based on the Distance	345
Figure 10.5	Comparison of RMSEs of the Difference for Groups Based on the Average Slope	347
Figure 10.6	Comparison of RMSEs of the Proportional Difference for Groups Based on the Average Slope	348
Figure 10.7	Comparison of RMSEs of the Difference for Groups Based on the Weighted Average Slope	351
Figure 10.8	Comparison of RMSEs of the Proportional Difference for Groups Based on the Weighted Average Slope	352
Figure 10.9	Comparison of RMSEs of the Difference for Groups Based on the Average Slope Change	354

Figure 10.10 Comparison of RMSEs of the Proportional Difference for Groups Based on the Average Slope Change	355
Figure 10.11 Comparison of RMSEs of the Difference for Groups Based on the Weighted Average Slope Change	357
Figure 10.12 Comparison of RMSEs of the Proportional Difference for Groups Based on the Weighted Average Slope Change	358
Figure 10.13 Comparison of RMSEs of the Difference and Proportional Difference for Groups Based on the Number of 3-D Points	360
Figure 10.14 Comparison of RMSEs of the Difference and Proportional Difference for Groups Based on the Average Density of 3-D Points	362
Figure 11.1 Modeling Water Bodies Using Polylines	369
Figure 11.2 Cross-sectional View of Flooding	369
Figure 11.3 Illustration of Flooded and Not Flooded Road Segments in the Flooded Area	370
Figure 11.4 Conceptual Model for Predicting Flood Extent	372
Figure 11.5 Conceptual Model for Identifying Flooded Road Segments	373
Figure 11.6 Obtaining Points along A Polyline	376
Figure 11.7 The Procedure of Predicting Flood Extent	377
Figure 11.8 Identifying Points with Target Elevation	378
Figure 11.9 Illustration of Water Flow	379
Figure 11.10 Error Tolerance and Maximum Drop	380
Figure 11.11 Flood Road Segment Identification Algorithm, Part I	382
Figure 11.12 Flood Road Segment Identification Algorithm, Part II	383
Figure 11.13 An Illustration of Identifying Flooded Road Segment Algorithm	384
Figure 11.14 Study Scope for Testing in Wilson County	385
Figure 11.15 Testing Results – Algorithm 1	386
Figure 11.16 Testing Results – Algorithm 2	386
Figure 11.17 Detailed View – Algorithm 1	386
Figure 11.18 Detailed View – Algorithm 2	386
Figure 11.19 Flooded Road Segments – Algorithm 1	388
Figure 11.20 Flooded Road Segments – Algorithm 2	388
Figure 11.21 Detailed View of the Flooded Road Segments	388

1 INTRODUCTION AND BACKGROUND

This chapter describes briefly the development and definition of geographic information systems (GISs), their applications in various fields, the basics of geographic modeling and data structure, data quality issues, and geographic information systems in transportation (GIS-T). Section 1.6 (Research Needs) states the motivation for this research. Terms and terminologies being used are also defined in this chapter. In addition, section 1.8 summarizes the organization of this dissertation to provide better understanding.

1.1 BACKGROUND

Human beings started to locate themselves on the Earth and navigate along the Earth surface in the very early years. It is known that human beings put stones on the ground to help them keep track of their routes and had relied on the stars in the sky to help locate their positions. The spatial information was so valuable that people tried to record this information even before the emergence of paper and compass. The competition for and demands on natural resources of land, air, water, and raw materials accompanying civilization brought the needs to record land use, land ownership, and transactions in a way that is not dependent on human memories (Burrough and McDonnell 1998).

From the earliest civilization to modern times, the above-mentioned spatial data have been collected by navigators, geographers, and surveyors to be recorded in a coded, pictorial form by mapmakers and cartographers (Burrough and McDonnell 1998). It is known that the land surveyors were an important part of the government in Roman times and that the results of their work may still be seen in vestigial form in the European landscapes to this day (Dilke 1971). By the seventeenth century skilled cartographers such as Mercator had demonstrated that the registration of spatial phenomena through an agreed standard provided a model of the distribution of natural phenomena and human settlements that were invaluable for navigation, route findings, and military strategies (Hodgkiss 1981). Cartographers recorded the locations and characteristics of natural phenomena using a powerful range of tools. Point and line symbols were chosen to represent the most important characteristics of the objects while areas were colored differently to represent the changes in certain properties of the surface.

Today, much geographical information concerns the location of well-defined objects in space and the interactions among them (Burrough and McDonnell 1998). Examples include trees in a forest, roads connecting cities, and houses on a street. Spatial data and spatial analyses are being widely used in a variety of fields. Civil engineers use them to plan the routes of roads and to estimate construction costs. Police departments need to know crime distribution and to determine emergency routes. The utility companies need to record the positions and locations of the utility systems. The urban planners need detailed information about the distribution of land and resources in towns and cities. Foresters are interested in the distribution of trees. All these applications reveal that using traditional paper maps to record geographical information have several severe drawbacks that make them inappropriate in these applications. These drawbacks include:

- In order to represent spatial data in paper maps, the original data had to be greatly reduced and consequently, many details were often filtered away and thus lost.
- The information for large areas with respect to map scale could only be represented by a number of map sheets. It is not uncommon that the interested areas are near the junctions of two or more maps.
- Once the paper map is produced, it is very difficult and costly to make changes or to be combined with other spatial information.
- Most importantly, paper maps are just static and qualitative documents. Spatial analyses that deal with the interactions of spatial objects are extremely difficult with these paper maps.

More recently, using aerial photographs and satellite images enables monitoring landscape changes over time and provides an efficient way to obtain useful information about the Earth's surface. However, the products of the airborne and space sensors are not maps, rather, they are images. The digital data of these products are not in the form of points, lines, or polygons as in traditional maps. There is a need for the integration of remote sensing, Earth-bound survey, and cartography. This is only made possible by the class of spatial information handling and mapping tools known as geographic information systems (GISs).

The growing demands for spatial data and for better ways to analyze them, together with the rapid developments in computer technologies led to the emergence of geographic information systems (GISs) that are

heavily computer oriented (Burrough and McDonnell 1998). GISs model spatial objects and their relationships in a consistent manner and carry out spatial analyses efficiently; therefore, GISs overcome the drawbacks of the traditional maps.

By the late 1970s there had been considerable investments in the development and application of computer-assisted cartography, particularly in North America by government and private agencies (Burrough and McDonnell 1998). With more than twenty years of technical development in GISs, they are becoming a worldwide phenomenon. In 1995 it was estimated that more than 93,000 sites worldwide had installed GISs (Burrough and McDonnell 1998). Today, GISs are being utilized in various fields including agriculture, archaeology, Earth observations, epidemiology and health, forestry, emergency services, navigation, marketing, real estate, regional/local planning, transportation, social studies, natural resource management, land use and land cover mapping, hydrology, civil engineering, mining, geology, geography, soil science, and much more.

1.2 GIS

The concept of GISs traces its roots to a handful research initiatives in the US, Canada, and Europe during the 1950s (Thill 2001). The first real GIS was the Canada Geographic Information System set up for the Canada Land Inventory.

A 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 1986). A GIS has at least five components: people, data, hardware, software, and procedures.

A GIS works with spatial data. Spatial data represent phenomena from the real world in terms of (1) their positions with respect to a known coordinate system, (2) their attributes that are unrelated to position (examples include thickness, cost, incidence of accidents, and 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).

From a functional point of view, a GIS must be able to realize at least four tasks (Rasdorf and Cai 2001).

- (1) Data input and verification,
- (2) Data storage and database management,
- (3) Data output and presentation, and
- (4) Data transformation.

It is the functional complexity of GISs that differentiates a GIS from any other systems. The geo-visualization capability differentiates a GIS from a database management engine. The analytical capability differentiates a GIS from automated mapping applications (Thill 2000). On the other hand, the database management features in GISs enable GISs to capture spatial and topological relationships between geo-referenced entities without pre-defining these relationships (Thill 2000).

1.3 2-D, 2.5-D, AND 3-D MODELING

As indicated by the name, a GIS works with geographic data, which can be understood as the information tied to some portion of the Earth (Chrisman 2002). It is obvious that we do not store real world phenomena in the computer but only representations of these phenomena based on some formalized models (Burrough and McDonnell 1998). In other words, a model can be defined as a set of rules to abstract real world phenomena while modeling can be defined as the procedure of abstracting real world phenomena so that representations are produced. In the case of GISs, modeling enables us to obtain geographic information to describe those real world phenomena. It is this information that is being input into and stored in the computer and analyzed to support decision-makings.

Geographical phenomena require two descriptors to represent the real world, what is present and where it is (Burrough and McDonnell 1998). Geographic information is commonly broken into three components: space, time, and attribute. It requires reference systems, which are defined as a set of rules for measurement and provide means to compare a particular measurement to others performed with reference to the same set of rules, for these three components.

A spatial reference system can be defined as a mechanism to situate measurements on a geometric body. It establishes a point of origin, orientation of reference axes, and geometric meaning of measurements and units of measurements (Chrisman 2002).

As outlined earlier, modern GISs are developed based on the concepts that have been well established with the emergence and development of paper mapping. The basic spatial models used in modern GISs are little different from those used 15-20 years ago (Burrough and McDonnell 1998). All real world spatial objects exist in Euclidean 3-D space (Smith and Paradis 1989). Paper mapping spatially locates three dimensional real world objects in a two-dimensional coordinate system, more specifically, an X and Y coordinate system. Figure 1.1 illustrates the procedure of map projections to convert three dimensional objects on Earth surface into two dimensional coordinate systems. This procedure is being followed by most geographic information systems on market.

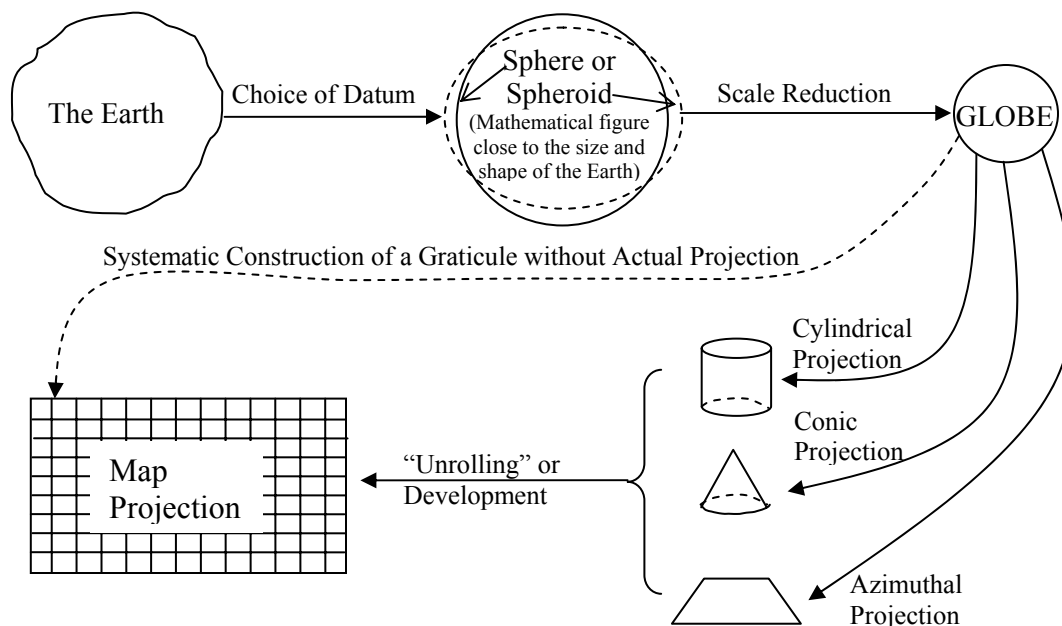


Figure 1.1 The Procedure of Map Projections

The actual shape of the Earth is too lumpy to use as a reference surface. The first step is to adopt a model for the Earth, usually in the form of a reference ellipsoid. Examples include Clarke’s 1866 and WGS 84. Then a geodetic datum is developed for the Earth using the established longitude-latitude format. The Earth’s surface

is not flat. In order to place objects on a two-dimensional plane, a procedure called projection is required. A projection transforms latitude-longitude information into planar coordinates, resulting in map construction on two-dimensional media in Cartesian coordinate systems. After projection, the origin point, axis directions, and measurement units are determined and a spatial reference system is developed.

2-D modeling captures the planimetric positions of objects on a two-dimensional plane. Most commercially available GIS products take this approach and therefore, cannot handle true 3-dimensional data, although they can handle topographic data, usually as a digital elevation model (DEM), and display isometric views, contour maps and so on (Turner 1989a, 1989b, 1989c, and 1989d). Most DEMs use either grid elevation matrices, or triangular meshes (TINs) to allow for terrain representations. This is enabled by treating the elevation, or the third dimension (Z-coordinate) as a dependent variable. This approach is defined as the quasi- three-dimensional, or 2.5-D modeling approach. In other words, the dependence of Z-coordinate on the X- and Y- coordinates in the 2.5-D approach enables deriving Z values using X and Y values and some supplementary information, and employing some interpolations or mathematical functions. In the 2.5-D approach, only one elevation value is allowed for a specific pair of X/Y-coordinates.

Different from 2-D and 2.5-D modeling approaches, 3-D modeling represents spatial objects in a true three-dimensional space. The 3-D modeling approach treats the third dimension (or Z-coordinate) as an independent variable. Figure 1.2 illustrates the differences in representing a simple linear object in 2-D, 2.5-D, and 3-D spaces. In the 3-D space, the object might not be a simple linear object as illustrated in Figure 1.2.

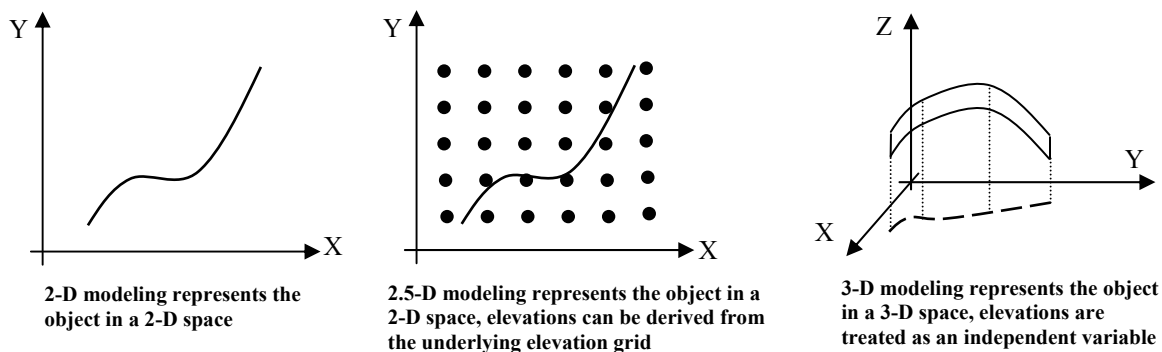


Figure 1.2 The Differences in Representing a Linear Object in 2-D, 2.5-D, and 3-D spaces

1.4 DATA QUALITY

Data quality is a critical feature of successful use of GISs, whether working in two dimensions or three (Burrough 1986, Burrough and McDonnell 1998, Rhind 1981). Data quality is an important factor to any information systems 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-makings. It is clearly recognized that the validity of analysis is based on the validity of the data used to perform the analysis (Rasdorf et al. 2001). Any geographic 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 lead to the necessity for data quality control (Rasdorf et al. 2001).

1.5 GIS-T AND ROAD CENTERLINE

GISs have benefited the field of transportation for many years due to the rich spatial data in transportation (Fletcher 1987, Nyerges and Dueker 1988, Dueker and Kjerne 1989, Ries 1993, Vonderohe et al. 1993, Dueker and Butler 1998, Fletcher 2000). In the transportation context, three classes of GIS models are relevant (Goodchild 1992, 1998).

- Field models to represent the continuous variation of a phenomenon over space. Terrain elevation uses this model.
- Discrete models to represent discrete entities (points, lines, and polygons) in space. Highway rest areas, toll barriers, urbanized areas may use this model.
- Network models to represent topologically connected linear entities such as roads.

Among these three relevant models, the network model built around the concepts of arc and node plays the most prominent role in transportation applications because single- and multi-modal infrastructure networks are vital in enabling and supporting passenger and represent data (Thill 2000). Many transportation applications only require network representation of data. Examples include pavement management systems, routing applications, traffic information systems, congestion management, accident detection, etc.

However, the network model in the conventional GISs assumes the homogeneity of each network link, which does not hold in transportation. The road parameters such the number of lanes and pavement width cannot be constrained to be constant between terminal nodes of a link. Also, traffic parameters of speed, flow and capacity cannot be expected to be constant along a link. The dynamic nature of these distributed attributes of the network precludes that the network be permanently edited to maintain the homogeneity of each link on each attribute, which leads to the concept of linear referencing (Thill 2000). In the linear referencing approach, attributes are linearly referenced and linked dynamically to the entities forming the network (Scarponcini 1999). This capability of geographic information systems in transportation (GIS-T) was identified as critical in the early researches in GIS-T (Dueker 1987, Fletcher 1987, Vonderohe et al. 1993).

To summarize, GIS in transportation is more than just one more domain of application of generic GIS functionality because GIS-T has several data modeling, data manipulation, and data analysis requirements that are not fully supported by conventional GISs (Thill 2000). Rather, it is viewed as the cross-fertilization of an enhanced GIS and an enhanced transportation information system (TIS) as illustrated in Figure 1.3 (Vonderohe et al. 1993).

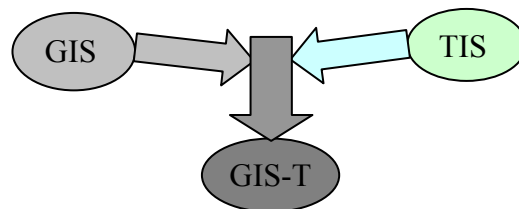


Figure 1.3 GIS-T, Product of an Enhanced GIS and an Enhanced TIS (Vonderohe et al. 1993)

GIS-T work with the physical facilities of a transportation system, particularly the roads in the context of highway transportation. Roads are three-dimensional surface objects in space. Among the seven elements in

highway geometric design, i.e. the sight distance, cross section, horizontal alignment, super elevation, vertical alignment, channelization, and pavement design, the cross section, the horizontal alignment, and the vertical alignment together depict the roads spatially as surface objects.

Cross section design refers to the profile of the facility that is perpendicular to the centerline and extends to the limits of the right-of-way within which the facility is constructed (Papacostas and Prevedouros 2001). The cross section of road includes the pavement, shoulder, clear zone, and additional R/W as required. The horizontal alignment represents the projection of the road facility on a horizontal plane. It generally consists of straight-line segments (tangents) connected by circular curves directly (simple curves) or via intermediate transition curves (Papacostas and Prevedouros 2001). The vertical alignment of highways and railways consists of grade tangents connected with parabolic vertical curves. Figure 1.4 illustrates the concepts of horizontal alignment and vertical alignment with a straight road segment connected directly to a circular segment. Figure 1.5 illustrates a typical cross section with cut in table drain.

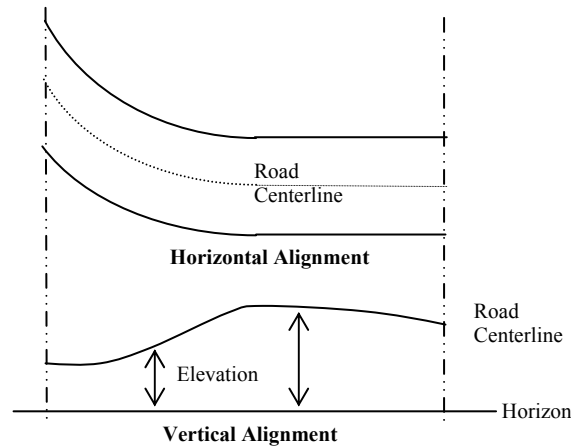


Figure 1.4 The Concepts of Horizontal Alignment and Vertical Alignment

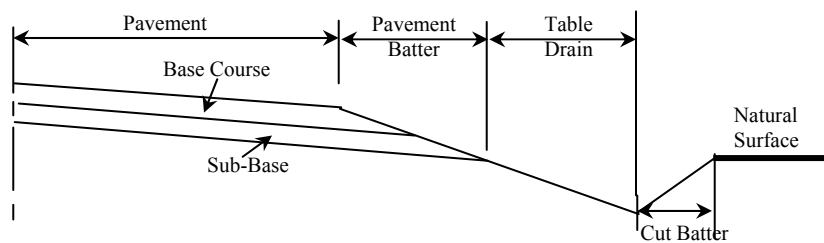


Figure 1.5 Typical Cross Section in Cut with Table Drain (Khisty 1990)

In GIS-T, roads are treated as linear objects rather than three-dimensional objects. The geometry of the road centerline, which is defined as the physical center of road pavement, is used to represent the geometry of roads. Figure 1.6 illustrates the definition of road centerlines for roads depending on the number of lanes and the way they are modeled in GIS-T. Modeling three-dimensional roads as linear objects in a 2-D space is just an example to illustrate that many of the linear objects in GIS are three-dimensional spatial objects in reality.

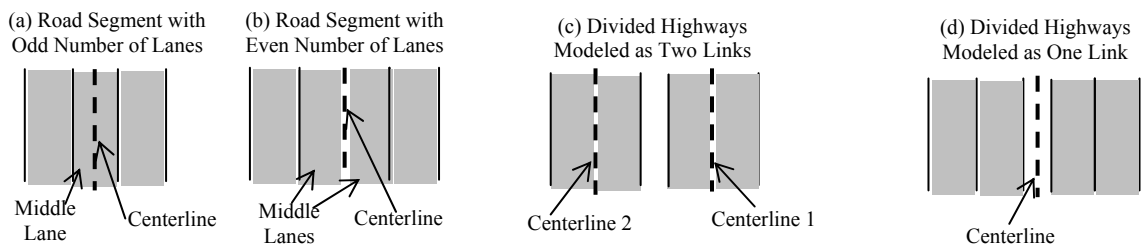


Figure 1.6 Definition of Road Centerline

1.6 RESEARCH NEEDS

Real world objects are all three-dimensional (3-D) objects. Numerous applications in GISs require modeling and representing spatial objects in a 3-D space with elevation (or height) and geometric information such as gradients, aspects, and actual lengths and areas together with their planimetric positional information. For example, terrain modeling focuses on capturing elevations for an area and models this area in a 3-D space. Triangulated irregular networks (TINs) provide a solution to terrain modeling problems by modeling and representing an area using 3-D polygons. Digital elevation models (DEMs) provide another solution by using a series of points with X, Y, and Z values to approximate a 3-D surface. More strictly, using TINs and DEMs to model terrains is not true 3-D modeling, but 2.5-D modeling because the elevation is treated as a dependent variable rather than an independent variable.

The 2.5-D modeling approach provides a surrogate to deal with 3-D objects in a 2-D space by treating elevation as a dependent variable. It is being widely used in modeling terrain surfaces. However, many phenomena in geology, geophysics, meteorology, hydrology, mining, ground water, hazardous contaminations, and the like, are 3-dimensional in nature and when trying to fit them into 2-dimensional systems it is impossible to accurately model, analyze, or display the information (Smith and Paradis 1989). Situations such as the

problems with hazardous chemicals in the ground are impossible to be modeled, analyzed, or displayed using a 2-dimensional tool (Smith and Paradis 1989). Applying a 2-dimensional tool to a 3-dimensional situation limits the scientist's work in many ways. It is not possible to accurately model the vertical relationships between two stacked 2-dimensional stacks, or to perform true 3-dimensional analytic operations (Smith and Paradis 1989).

There are other applications that require and benefit from modeling and representing spatial objects in 3-D space. Examples include hydrology (river cross-section through sediments, for example), flood response and management (locate areas that are flooded and determine the accessing routes), stream modeling and flow estimate, shoreline modeling, volume calculation in grading and excavation, geometric design in highway systems, bus routing, etc. These applications not only require representing and modeling polygon/area objects in a 3-D space, but also modeling and representing linear objects in a 3-D space with elevation and geometric information. For example, flood response and management requires determining which road sections are under flood and which road sections become dips due to flooding so that the appropriate accessing route could be determined. In bus routing, modeling roads as 3-D linear objects helps minimize the gradient changes to achieve gasoline economy. In highway geometric design, modeling roads as 3-D linear objects provides geometric information such as gradients and curves to determine the sight distance. Modeling linear objects in 3-D space also benefits transportation applications in which location or locating is an essential part because it provides more accurate locations along the linear objects.

For interest, Table 1 provides several examples to illustrate the significant difference in length measurements between 2-D approach and 3-D approach to model road centerlines (NCDOT data). The three links in Table 1.1 are all interstate highway segments that are relatively flat without steep slopes. It is reasonable to expect that the length difference between the 2-D approach and the 3-D approach will be more significant for road segments in mountainous areas with steep slopes.

Table 1.1 The Significant Difference in Lengths between 2-D Approach and 3-D Approach

Link	2-D Length	3-D Length	Difference	Percentage Difference
1783	12.69m	13.88m	1.19m	8.6%
985	45.47m	52.33m	6.86m	13.2%
1253	112.39m	126.48m	14.09m	11.1%

To summarize, there is a need to model linear spatial objects in a 3-D space and to use the 3-D modeling approach to derive geometric information for these linear objects such as the actual length and curvature. A preliminary literature review revealed that current researchers in the field of 3-D modeling focus on modeling polygon/area objects in a 3-D space while modeling linear objects in a 3-D space is mostly ignored over the last few decades. This scenario is partially due to the involvement of polygon/area objects in complicated spatial analysis tasks and the difficulties of modeling polygon/area objects in a 3-D space in GISs, which has attracted most research attention. However, for those GIS applications that deal with linear objects (for example, GIS-T applications), how to model linear objects in a 3-D space becomes critical and calls for research efforts.

1.7 TERMS AND TERMINOLOGIES

In order to establish a vocabulary for this dissertation, several terms will be defined here. Appropriate definitions found in the literature will be introduced whenever needed.

3-D Modeling: A 3-D modeling approach models 3-D objects in a three-dimensional space, in which the third dimension (elevation) is treated as an independent variable in the same way as the X and Y dimensions are treated in 2-D modeling.

Accuracy: Accuracy may be defined as the extent to which an estimated or observed value approaches the true value (Burrough and McDonnell 1998). Data accuracy is a broad term that can be divided into positional accuracy (whether objects are in their correct positions), geometric accuracy (e.g., size or length of linear objects), interrelationship accuracy (topological) that describes the relationships between objects, attribute accuracy, and temporal accuracy.

Positional accuracy is the positional closeness of an object as described in a spatial database to its real world position. Attribute accuracy is defined as the accuracy of the attributes of a spatial object. Geometric accuracy refers to the accuracy of geometric properties of spatial objects. Examples include length, direction, and orientation of linear objects, and area and perimeter of area objects. Interrelationship accuracy (topological)

refers to the accuracy of topology, which describes the spatial relationship between spatial objects including connectivity, adjacency, and proximity. Attribute or thematic accuracy is defined as the accuracy of the attributes of its data (Burrough and McDonnell 1998). Temporal accuracy is defined as that part of the data's error that arises due to the temporary nature of the data. It is affected by the interaction between the duration of the recording interval and the rate of change in the event (Veregin 1999).

ANOVA: The analysis of variance (ANOVA) is a statistical method that was designed for testing the equality of the expected responses for t treatments (Rao 1998). The ANOVA method compares a measure of the magnitude of the observed variability within the t samples with a measure of the observed variability between the means of t samples to test the equality of the expected response for t treatments (Rao 1998).

Breakline: 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 hard breakline defines interruptions in surface smoothness. Hard breaklines are usually used to define locations with abrupt surface changes, for example, streams, ridges, etc.

Data Quality: Data quality is a critical feature of successful use of GISs, whether working in two dimensions or three (Burrough 1986, Burrough and McDonnell 1998, Rhind 1981). 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.”

Data quality is a comprehensive measurement to assess the data suitability for use. 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 (SDTS 1997).

DEM: A digital elevation model (DEM) 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). DEM 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 DEM 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.

DMI: A distance measurement instrument (DMI) is a distance measure equipment that works similar to a vehicle odometer by driving along roads. A DMI measures the distance a vehicle has traveled by counting the number of pulses from the wheel sensors or the transmission sensors and converting the number of pulses into the distance traveled based on the relationship between the traveled distance and the number of pulses, which is obtained during calibrations.

DTM: A digital terrain model (DTM) is defined as a broad term that refers to any digital models that could be used to represent the terrain surface. Examples include digital contour lines, DEM, TIN, etc.

EDMI: Electronic distance measurement instrument (EDMI) is defined as an instrument that transmits a carrier signal of electromagnetic energy from its position to a receiver located at another position (Kavanagh 1997). By doing so, the distance between two positions can be determined. An EDM is an instrument that is built into the total station to measure distances in ground surveying.

Error: Error is the difference between the true and estimated values.

Error Propagation: During the GIS operations, errors in the input data accumulate with errors introduced in operations and affect the quality of the new derived data (Burrough and McDonnell 1998). This is defined as error propagation.

Floodplain Delineation: Floodplain delineation is the process of determining inundation extent and depth by comparing river water levels with ground surface elevations (Noman et al. 2001).

Frequency Analysis: Frequency analysis is an analysis method that counts the frequency of the occurrences in a given range. It is a useful tool being used in this research to evaluate the accuracy of the predicted 3-D distances. This method works in such a way that given a set of errors, first, they are ordered from “best” to “worst”; second, ranges are designed, and finally errors in a particular range are counted and the percentage is calculated.

GIS: A geographic information system (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 1986). A GIS has at least five components: people, data, hardware, software, and procedures.

From a functional point of view, a GIS must be able to realize at least four tasks (Rasdorf and Cai 2001).

- 1) Data input and verification,
- 2) Data storage and database management,
- 3) Data output and presentation, and
- 4) Data transformation.

GIS-T: GIS in transportation (GIS-T) is defined as the cross-fertilization of an enhanced GIS and an enhanced transportation information system (TIS) (Vonderohe et al. 1993). It is more than just one more domain of application of generic GIS functionality because GIS-T has several data modeling, data manipulation, and data analysis requirements that are not fully supported by conventional GISs (Thill 2000).

GPS: A global positioning system (GPS) is such a system that it locates a point by tracking the signals from satellites. A GPS consists of three parts: the space segment (satellites), the user segment (GPS receivers), and the control segment (base stations) (Kennedy 1996).

The basic working principle for GPS is such that for a specific point position on the Earth surface, first, its distance to a satellite with known spatial position is determined. Second, by obtaining two more such distances to two other satellites with known spatial positions there will be three simultaneous equations that can be solved to obtain the point position in the x , y , and z directions. Since our time measurements (and thus our distance calculations) are not perfect, distance measurements are made to four or more satellites and this enables us to eliminate the error in timing.

LIDAR: LIDAR is the acronym for Light Detection and Ranging. LIDAR technology has become the accurate, timely and economical way to capture elevation information for DEMs and DTMs (Hill et al. 2000). Elevation datasets produced from LIDAR could be mass points and DEMs with different resolutions. LIDAR systems emit rapid pulse of laser light to precisely measure distances from a sensor mounted in a port opening of an aircraft's fuselage to targets on the ground.

LIDAR is also known as airborne laser mapping. A LIDAR is similar to the more familiar radar, and can be thought of as laser radar. A LIDAR transmits and receives electromagnetic radiation, but at a higher frequency than radar. A LIDAR operates in the ultraviolet, visible, and infrared region of the electromagnetic spectrum. A modern LIDAR system consists of Airborne Global Positioning System (GPS), Inertial Navigation System/ Inertial Measurement Unit (INS/IMU), Laser Rangefinder, and supporting computer hardware and software. The GPS is used to determine a position of the sensor. Laser Rangefinder is used to determine a distance between the sensor and the target. The orientation of the aircraft is controlled and determined by the INS. Combining all together, LIDAR provides the position of the target ground feature.

LIDAR Point Data: LIDAR point data refer to the mass point data describing the bare Earth surface obtained from LIDAR. Each point has X/Y/Z-coordinates.

LIDAR DEM: A LIDAR DEM refers to the DEM produced by processing the LIDAR point data via interpolation methods.

LRS: A linear referencing system (LRS) is defined as a system that encompasses a larger set of office, field management, and integration procedures that include one or more linear referencing methods. The linear referencing system incorporates “procedures that relate all locations to each other...it includes techniques for storing, maintaining, and retrieving location information (O’Neill 1996)”. A location referencing method is “a way to identify a single location; i.e., to reference a specific position with respect to a known point (O’Neill 1996)”. The core for linear referencing is that the location of any unknown point along a linear feature can be determined by specifying the direction and distance from any known point to the unknown point (Vonderohe et al. 1995). This concept solves the problem of locating events along the network and is the basis for all linear referencing methods.

Linear referencing systems (LRS) are used to manage and promote analysis of spatial data elements on transportation networks because inherent in these data collection and analysis efforts is a requirement to describe such data in terms of location.

NED: The National Elevation Dataset (NED) is a new raster product assembled by the USGS. It is designed to provide national elevation data in a seamless form with a consistent datum, elevation unit, and projection (USGS 2001c, USGS 2003b). In the NED assembly process, data corrections were made and the elevation values were converted to decimal meters as a consistent unit of measure. NAD83 was consistently used as the horizontal datum, and all the data were recast in a geographic projection (USGS 2001c, USGS 2003b). The vertical datum is North American Vertical Datum of 1988 (NAVD88). NED has a resolution of one arc-second (approximately 30 meters) for the conterminous United States, Hawaii, and Puerto Rico and a resolution of two arc-seconds for Alaska (USGS 2001c, USGS 2003b).

The USGS NED has been developed by merging the highest-resolution, best-quality elevation data available across the United States into a seamless raster format (USGS 2003a). When a NED tile is assembled, the best available source data are selected according to the following criteria (ordered from first to last): 10-meter DEM, 30-meter level-2 DEM, 30-meter level-1 DEM, 2-arc-second DEM, 3-arc-second DEM (Osborn et al. 2001).

As a result, a NED file might have multiple resolutions (Osborn et al. 2001). NED is the result of the USGS effort to provide 1:24,000-scale Digital Elevation Model (DEM) data for the conterminous US and 1:63,360-scale DEM data for Alaska (USGS 2003a, USGS 2003b).

NED is essentially a digital elevation model (DEM) in which terrain elevations for ground positions are sampled at either a regular or irregular horizontal interval (Anderson and Mikhail 1998, USGS 2001d). The terrain surface is represented by using an array with X, Y, and Z values.

Precision: Precision is the dispersion of data from the actual data. It is estimated in terms of standard deviation of the observations over the mean. It also refers to the ability to display numbers to a certain number of decimal digits (Rasdorf 2000).

Road Centerline: A road centerline is defined as the physical center of road pavement, which is used to represent the geometry of roads in GIS-T as linear objects.

Road Geometry: In this research, road geometry refers to the geometric properties that are necessary to describe the road segments as 3-D objects in a 3-D spatial referencing system. These geometric properties include cross section, horizontal alignment, super elevation, and vertical alignment.

Sensitivity Analysis: Sensitivity analysis in this research refers to the analysis of the sensitivity of the accuracy of the predicted 3-D distances to the use of different elevation datasets with different vertical accuracy levels. It is the measure of the effects on the accuracy of the predicted 3-D distances from the use of different elevation datasets.

Significant Factor: In this research, a significant factor is defined as a factor that is related to the geometric properties of the linear object and that significantly affects the accuracy of the predicted 3-D distances using the 3-D approach proposed in this research.

Spatial Data Accuracy: The spatial data accuracy is defined as the accuracy of the spatial data. 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 (Rasdorf et al. 2003a).

Surface Length (3-D Distance): In this research, surface length or 3-D distance refers to the same thing, the actual distance of a linear object when measured in a 3-D space. It differs from 2-D length in that the effects on the length from the variations of the third dimension (elevation) are accounted for.

Terrain Modeling: Terrain modeling is one particular category of surface modeling that deals with the specific problem of representing the surface of the Earth (Kennie and Petrie, 1991). Surface modeling is a general term that is used to describe the process of representing a physical or artificially created surface by means of a mathematical expression. A surface may be defined as a three-dimensional object with its third dimension varying continuously. A terrain is a surface object with elevation as the continuously varying third dimension. There are four different ways of modeling a surface: point models, contour line models, TIN (triangulated irregular network), and DEM (digital elevation model).

TIN: The Triangulated Irregular Network (or TIN) was designed by Peucker and co-workers (Peucker *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.

1.8 ORGANIZATION OF THE DISSERTATION

This dissertation consists of 12 chapters. The content of these chapters is as follows.

1. An introduction to and the background for the research.
2. Research plan and the methodology used.
3. Literature review of spatial modeling.
4. Literature review of data quality.
5. Examination of existing technologies and methods that could be used to measure road lengths, focusing on their advantages and disadvantages in obtaining lengths and data quality issues.
6. Conceptual 3-D model developed in this research to model linear objects in a 3-D space and to predict their 3-D distances.
7. Computational implementations of 3-D models developed in this research.
8. The case study developed in the transportation context, in which the developed 3-D models and algorithms are implemented and tested in practice.
9. Accuracy evaluation of the predicted 3-D distances by applying the developed 3-D models in the case study and sensitivity analysis aiming at evaluating the effects on the accuracy from the integration of the developed models with different elevation datasets.
10. Identification of significant factors.
11. The application of 3-D modeling and the resulting 3-D road centerline data in identifying flooded highway segments for floodplain mapping and flood response and management.
12. Summary of the results, conclusions, key findings, recommendations, and guidelines for future research.

The appendices hold the program codes that were developed for this research.

2 RESEARCH PLAN

This chapter defines the research problem, the objectives of this research, the methodology or the approaches that will be used to carry out this research, the significance and benefits of conducting this research, and expected research results. Furthermore, a case study defined in the context of transportation road network is presented with brief information regarding its background and study needs.

2.1 RESEARCH PROBLEM STATEMENT

Real world objects are all three-dimensional (3-D) objects. Numerous applications in GISs require modeling and representing spatial objects in 3-D space with elevation (or height) and geometric information such as gradients, aspects, and actual lengths and areas together with their planimetric positional information. Examples include hydrology (river cross-section through sediments, for example), flood response and management (locate areas that are flooded and determine the accessing routes), stream modeling and flow estimates, shoreline modeling, volume calculation in grading and excavation, geometric design in highway systems, and bus routing.

However, many current GISs can only represent two-dimensional (2-D) information by projecting the 3-D objects onto a planar surface and referencing them to a known coordinate system (Burrough and McDonnell 1998). This is an indicator of the great influence of the traditional paper mapping on GISs. The GIS community has been struggling with solving complex problems dealing with 3-D objects in a 2-D approach (terrain modeling, for example).

Different from the traditional 2-D approach, 3-D spatial modeling represents 3-D objects in a 3-D space (Li 1994). By modeling real world objects as 3-D objects, the 3-D spatial modeling technique benefits numerous applications. As a response, three-dimensional GISs are becoming increasingly available (Burrough and McDonnell 1998).

Currently, 3-D spatial modeling techniques are being widely used in modeling three-dimensional polygon (or area) objects so that important information about gradients, aspects, and volumes is captured. They could also be applied to modeling three-dimensional linear objects and capturing geometric information about linear objects such as gradients, curves, and distances, which are very important to applications such as flood response and management, highway geometric designs, routing procedures, road inventory, etc. Despite the needs for modeling linear objects in 3-D space and the benefits it would bring, it has been largely neglected in the past few decades. Questions regarding how to model and represent linear objects as 3-D linear objects in a 3-D space and how accurate models would be in representing linear objects and predicting their lengths remain unanswered. This is the research topic that will be addressed in this research.

2.2 OBJECTIVES OF THE RESEARCH

The overall purpose of this research is to develop a 3-D spatial model to represent linear objects in a 3-D space, predict lengths for linear objects using this 3-D model, and evaluate the accuracy of these predicted distances via a case study in the context of transportation (transportation linear objects, or more specifically, road segments).

The objectives of this research can be summarized as follows:

1. To develop a three-dimensional (3-D) spatial model to represent linear objects as three-dimensional (3-D) linear objects.
2. To predict the lengths of linear objects using this 3-D spatial model.
3. To evaluate the accuracy of this 3-D spatial model and the predicted lengths via a transportation case study.
4. To identify and quantify the error sources in the input data, and apply quality control measures to improve the accuracy of this 3-D spatial modeling approach in length prediction.
5. To examine the error propagation in this 3-D spatial modeling approach in distance prediction.
6. To analyze the sensitivity of this 3-D spatial modeling approach to varying accuracy in the input data.

7. To identify properties belonging to the linear objects, which significantly affect the accuracy of this 3-D spatial modeling approach in predicting the 3-D distance.

2.3 CASE STUDY DESCRIPTION

The case study is defined in the context of transportation because transportation possesses a rich set of linear spatial objects, more specifically, the road segments in highway transportation. These geo-spatial transportation segments can be connected to form topological networks that provide the basis for several indirect location referencing systems, including street address and various linear referencing methods commonly used to locate features like bridges, signs, pavement conditions, and traffic accidents, and can be used to more accurately measure over-the-road travel distances between geographic locations (FGDC 1994). Furthermore, these topological networks can be used to find the shortest paths, to determine the most efficient route, or to estimate traffic volumes, when combined with the variety of network analysis tools.

The key to integrating and analyzing transportation spatial data, and thus to making decisions about transportation investments, is location (FGDC 1994). This applies to all transportation applications, in which location or locating is an essential part. Examples include pavement management systems, automatic vehicle locating (AVL), intelligent transportation systems (ITS), accident detection, congestion management, and routing procedures (emergency response routing, truck routing, and bus routing, for example). This also applies to applications that use transportation data. For example, in flood management, the road segments that are flooded must be located to determine accessibility and response routes.

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

However, examining the existing transportation spatial databases revealed that there is no correct and consistent distance information (Rasdorf et al. 2003a). There are lots of errors and uncertainties in this distance information. Furthermore, the integration of linear referencing data with different linear referencing methods and the integration of linear referencing data with other datasets (GPS data, for example) revealed that errors vary widely across databases. There is a need to obtain accurate and consistent distance information, and use this information to update the existing transportation spatial databases. This is not a trivial task taking into consideration the number of roads and the mileages of these roads in the United States (Rasdorf et al. 2003b).

A number of techniques are available to obtain distance along linear objects. For example, this information can be extracted from construction and design drawings. However, this approach suffers from the problems of currency and completeness. These drawings are only kept for certain years. There might be lots of adjustments and modifications, which are not represented in these drawings. The distance measurement instrument (DMI) and the inertial navigation system (INS) measure the distance directly along the linear objects. The DMI itself is an inexpensive technology consisting of a mechanical device attached to one or more of the wheels of a van and connected to an in-vehicle recorder (Karimi, Hummer, and Khattak 2000). However, a DMI requires frequent calibrations to minimize errors that accumulate as the result of travel. An INS uses accelerometers and gyroscopes to provide pitch, roll, heading information to derive a relative geo-reference for a point. An INS has high accuracy over small distances and is usually more accurate than GPS (depending on types of GPS) but much more costly (Karimi, Hummer, and Khattak 2000). The 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, which is not always the case in road distance measurements (a road has lots of curves and turnovers). In addition, fieldwork interrupts the traffic very much. GPS is a very promising technique in measuring linear objects. However, the most notable difficulties in using this technology stem from signal blockage in certain areas caused by thick tree canopies or bridges, or multipath problems caused by signal deflection by high-rise buildings.

In addition to the difficulties and limitations mentioned above, all these approaches suffer from a common disadvantage: it is very time-consuming to use any of these approaches to obtain distance information for a large number of roads (Rasdorf et al. 2003b).

The possibility of using planimetric length to approximate the actual distance has also been considered. The resulting error from using this approach may not be significant in case of road segments with small gradient such as most interstate highways. But in case of steep road segments with relatively big gradients, the resulting errors are not acceptable, especially for roads in mountainous areas.

On the other hand, the application of the 3-D spatial modeling approach in predicting distances provides an efficient way to obtain distance information without field efforts. This could be used to represent road centerlines in a more accurate way, to obtain road centerline distances, and to correct and update distance information in existing transportation spatial databases. However, the concern is the accuracy. Before this 3-D technique can be successfully applied in modeling linear objects in three-dimensional space and predicting their lengths, its accuracy must be evaluated and quantified. Therefore, this transportation case study provides an excellent opportunity to test and evaluate the accuracy and quality of the 3-D spatial modeling approach.

2.4 RESEARCH METHODOLOGY

The research objectives will be achieved via the completion of the following seven tasks.

1) Task 1 Literature Review

Literature related to 3-D spatial modeling and spatial data quality will be reviewed. The literature search will be based on, but not limited to:

- Published materials in books under the subject of GIS, LRS, spatial modeling, and spatial data qualities;
- Published materials in journals (e. g., ASCE Journal of Computing in Civil Engineering, The International Journal of Geographic Information System), conference proceedings, and magazines (e. g., GPS World, Geo World, Civil Engineering);

- Data quality standards (e. g., Spatial Data Transfer Standards, Geospatial Positioning Accuracy Standard, Metadata Content Standards);
- Agencies dealing with spatial data and spatial modeling (e. g., USGS, FGDC, NCCGIA, State DOT's) and their technical bulletins;
- Internet.

2) Task 2 Develop A 3-D Spatial Model to Represent Linear Objects and Predict Lengths

A 3-D spatial model will be developed in this phase to represent linear objects as 3-D linear objects and predict their lengths. There are two steps involved in developing this 3-D spatial model. First, a conceptual model will be developed, which identifies the procedures and the input data. Second, a computational model will be developed using commercial GIS software and the available input data.

3) Task 3 Accuracy Evaluation

The accuracy for the 3-D spatial model developed in task 2 to predict lengths will be evaluated using a case study in the context of transportation linear objects in this phase. More specifically, the 3-D model developed in phase 2 will be applied to model highways as 3-D linear objects and predict their road centerline distances. The accuracy of this distance information will be evaluated via a comparison with reference data.

4) Task 4 Error Source Identification, Error Quantification, Error Propagation, and Quality Control

The error sources in the input data will be identified. The extents of these errors will be quantified. The effects of these errors to the accuracy of the predicted distance will be examined. Quality control measures will be examined and applied.

5) Task 5 Sensitivity Analysis

The sensitivity of the accuracy of the predicted distance to the quality of input data will be analyzed in this task. Tentatively, two datasets (LIDAR data and USGS NED) with different accuracy levels will be used separately to predict the distance for road centerlines using the 3-D spatial modeling approach. Their effects on the predicted distance will be evaluated and quantified.

6) Task 6 Identification of Significant Factors

The correlation between a property of a linear object (for example, the gradient, the curve, and the length) and the accuracy of the predicted distance will be determined in this research phase to identify factors that significantly affect the accuracy of predicting distance using the 3-D spatial modeling technique. Traditional statistical techniques include sample correlation coefficients and sample coefficients of determination will be used together with a grouping and comparison method to complete this task.

7) Task 7 Summary, Conclusions, and Recommendations

This research will be summarized. Conclusions will be drawn. Recommendations will be provided. The models, algorithms, and findings will all be well documented.

2.5 BENEFITS AND SIGNIFICANCE OF THE RESEARCH

By accomplishing the seven tasks to achieve the objectives of this research, it will have the following benefits and contributions:

- A new model will be developed to represent linear objects as three-dimensional linear objects and its accuracy will be evaluated. This model will be more accurate in representing road centerlines when compared with the traditional approaches that represent linear objects as either one-dimensional or two-dimensional linear objects because in reality, they are three-dimensional.
- This 3-D spatial model can be used to represent road centerlines as 3-D linear objects and predict their distances. This approach eliminates the need for fieldwork in measuring road centerline distance and provides consistent distance information, which could be used to update the existing transportation spatial databases. It is more efficient compared with existing technologies such as DMI that measures distance and requests tremendous field efforts. This approach can be also applied to other applications in which, distance prediction is involved (utilities, for example).
- Because this 3-D spatial model represents road centerlines as three-dimensional linear objects, information regarding to gradients, curves, and elevations is readily available to benefits applications other than

distance prediction. For example, this 3-D model can help sight distance determination in highway geometric design. In flood management and response, this 3-D model provides not only the accurate length of roads that are flooded, but also the locations of flooded road segments and the locations of dips. When combined with the surface data, it also helps identify the locations of natural dams and ponds in case of flooding. In case of bus routing, this 3-D road centerline model can help determining the optical path by minimizing the fuel costs via minimizing grade changes.

- In addition to the new tool to model linear objects and the new approach to determine their lengths, this research provides an in-depth view to spatial data quality issues including error source identification, error quantification, error propagation examination, and sensitivity analysis.
- The models and tools developed in this research can be used to develop new tools, or enhancements to existing tools, to improve spatial data representation, understanding, analysis, and processing.

2.6 EXPECTED RESEARCH RESULTS

- A three-dimensional spatial model to represent linear objects as three-dimensional linear objects.
- A new approach to determine the lengths for linear objects using the technique of 3-D spatial modeling, its accuracy evaluation, error identification and quantification, and quality control mechanisms for this approach.
- A new tool to predict road centerline distances that can be used to update transportation spatial databases.
- A new dataset that models road centerlines using the technique of 3-D spatial modeling.
- Evaluation of the effects from the quality of the input data (sensitivity analysis).
- Identification of significant factors.

3 SPATIAL MODELING

This chapter introduces the concepts of data models and data structures in the context of GIS. It also serves as a detailed literature review regarding 2-D modeling, 2.5-D modeling, 3-D modeling and its applications, and the spatial modeling in transportation.

3.1 DATA MODEL, DATA STRUCTURE, AND SPATIAL DATA MODEL

A model is an abstraction of the complex real world phenomena (Burrough and McDonnell 1998). A model provides a means to capture the real world phenomena in such a way that it is understandable to people, it can be stored in a computer, and it can be manipulated with the interaction of people and computer.

Data can be defined as known facts that can be recorded and that have implicit meaning (Elmasri and Navathe 2001). Geographical data are data that record the location and a value characterizing the geographical phenomenon (Burrough and McDonnell 1998). Data model is the abstraction and representation of real world phenomena according to a formalized, conceptual schema (Elmasri and Navathe 2001). In the context of GIS, a geographical data model (or spatial data model) is a formalized schema for representing data which has both location and characteristics, and is usually implemented using the geographical primitives of points, lines, and polygons, or discretized continuous fields while data structure refers to the organization of data in ways suitable for computer storage and manipulation (Burrough and McDonnell 1998).

Geographical phenomena require two descriptors to represent: what and where. More specifically, what is present and where it is. To determine where it is, the concepts of sphere or spheroid, projection, coordinate system, and datum were introduced into the field of geographic information long time ago, which will not be detailed herein. Whenever considering any space, two fundamental approaches are always used. The first approach perceives the space as being occupied by entities that are described by their attributes or properties, and whose position can be mapped using a geometric coordinate system. This is also more frequently referred to as the vector approach. The other approach imagines that the variation of an attribute of interest varies over

the space as some continuous mathematical function or field (Burrough and McDonnell 1998). This is also known as the field approach and the resulting data model is known as the raster data model.

3.2 2-D AND 2.5-D SPATIAL MODELING

Until recently, computers were of little assistance to three-dimensional spatial modeling and visualization due to expensive memories to handle the huge amount of data required by three-dimensional assessments, the slow computational speeds, low resolutions that were not enough for visualization (Turner 1989a, 1989b, 1989c, and 1989d). As a result, much experience with GISs was gained in two-dimensional and the two-dimensional GISs matured rapidly in the late 1980's and became widely accepted (Turner 1989a, 1989b, 1989c, and 1989d). Almost all of these are concerned solely with a two-dimensional abstraction of the world (Rhind 1989).

As stated earlier, in 2-D spatial modeling, the location is determined on a 2-D planimetric plane for both of the vector approach and the field approach, the corresponding spatial data models are vector data models and raster data models, respectively.

The vector data model represents the entities, or objects as points, lines, or polygons, which are all scale dependent. For example, rivers can be represented either as lines or polygons. A city or town can be represented as a point in a small scale map, but a polygon in a large scale map, which delineates its jurisdiction area from the surrounding areas. A point is zero-dimensional. It simply implies that the geographical extents of the object are limited to a location that can be specified by one set of X/Y-coordinates at the level of resolution of the abstraction (Burrough and McDonnell 1998). A line is one-dimensional and deploys sets of X/Y coordinate pairs that define a connected path through space but without true width. A polygon can be simply represented in terms of the set of X/Y coordinates of its boundary. It is usually referred to as two-dimensional. Other geometry information, topology, and attributes can be stored in an attribute table. This attribute table has a one to one relationship with the object it is referring to. In other words, one entity/object corresponds to one unique row in the attribute table while each row in the table corresponds to a unique entity/object.

With raster data model, the continuous space is discretized into sets of single basic units such as square, triangular, or hexagonal cells, or into irregular triangle or polygons (Burrough and McDonnell 1998). In case of using square as the basic unit, the cell location is determined by referring to the start point location (usually the lower-left corner of the space being represented), the cell size, and the row number and column number of a specific cell. A raster data model is usually used, but not limited to represent continuous surfaces (terrain modeling). Each cell has a single value for a specific attribute such as elevation in terrain modeling, pollution level in environmental applications, temperature, or moisture.

The triangular irregular network (TIN) is a data structure used to model continuous surfaces in terms of a simplified polygonal vector schema (Burrough and McDonnell 1998). It is based on the Thiessen/Dirichlet/Veronoi procedure. It is worth of mentioning here because it looks like a raster data model but it essentially deploys a vector schema. Rather than storing a single value for the basic unit according to a specific attribute, TIN stores the X/Y-coordinates and the attribute values for the three nodes of each triangle. The attribute values for a position other than the triangle nodes are interpolated from those values of the nodes of a triangle, in which this position is contained.

The 2.5-D techniques offer the opportunity to work with the elevations (the third spatial dimension) by modeling the elevation as a variable dependent on X/Y-coordinates. One popular example is the raster format surface models (USGS DEMs, for example), in which Z values defining a surface are projected onto a model and all three axes displayed in order to visualize 3-D objects with isometric models or diagrams (Raper 1989). Another example is the TIN described earlier (Li 1994). The disadvantages of using the 2.5-D techniques in representing 3-D objects and visualization include the limitations in representing multiple Z values and that they are limited to representing spatial objects with a planar continuity (Raper 1989). It is still very difficult to extract 3-D properties of spatial objects via this 2.5-D approach and thus, makes it very difficult to carry out three-dimensional spatial analyses.

3.3 3-D SPATIAL MODELING

The motivation for three-dimensional spatial modeling in GISs originated from the nature of the problems being solved by GISs. As stated earlier, all real world objects are three-dimensional objects in space (without referring to the temporal reference). Geoscientists are concerned with three-dimensional spatial observations, measurements, and explanations of a great variety of phenomena (Turner 1989a, 1989b, 1989c, and 1989d).

Prior to computers, geoscientists had used methods such as specialized maps, cross-sections, fence diagrams, and geometrical constructions such as stereonets to portray three-dimensional relationships on two-dimensional paper products (Turner 1989a, 1989b, 1989c, and 1989d). The advent of the modern computer workstation with enhanced memory and graphical capabilities have enabled the spatial modeling in a three-dimensional space. The development of general and readily available 3-D tools to handle spatial data has been described as a GIS research priority since 1988 (Rhind 1988).

This section describes the previous research efforts in 3-D spatial modeling in such a way that first, the key concepts in 3-D spatial modeling from the NATO Advanced Research Workshop in 1989 are presented, the applications of 3-D spatial modeling and visualization around 1990 are summarized, observations regarding researches and applications in 3-D spatial modeling around 1990 are obtained, recent research efforts (after 1990) are briefed, and finally observations regarding those recent research efforts are obtained.

3.3.1 NATO Advanced Research Workshop in 1989

The first formal discussion of 3-D geoscientific information systems appeared in 1989, which was entitled as “Three-Dimensional Modeling with Geoscientific Information Systems” and was supported by NATO Advanced Research Workshop (ARW). This meeting summarized the situation of 3-D spatial modeling at that time, which covered the problem definition of 3-D spatial modeling, the 3D GISs existing at that time, 3-D data structures, and applications of 3-D spatial modeling (Turner 1989a, 1989b, 1989c, and 1989d).

Rhind (Rhind 1989) differentiated designed 3-D objects from revealed 3-D objects. The designed 3-D objects were defined to include designed structures such as toxic waste disposal sites, quarries, and buildings, which would be the results of craftsmanship based upon an original, usually visually expressed, design. The revealed objects were defined as objects whose form is necessarily deduced from secondary evidence, often of a surrogate nature. Rhind (Rhing 1989) also defined the simplest 3-D GIS should be able to cope with 3-D data defined through scalar measurements on three orthogonal axes and restricted to single values of the Z for any X/Y position.

Rhind (Rhind 1989) and Kelk (Kelk 1989) both identified the 3-D GIS functionality, which differentiate 3-D visualization from 3-D analysis, modeling from visualization, and 3-D analysis from 3-D data structuring. In addition, Kelk (Kelk 1989) identified the problems inherent in geoscientific 3-D modeling including the complex spatial relationships, the sparse and random data, together with others.

This workshop also presented several papers focusing on three-dimensional data structures and display methods. These papers emphasized the volumetric solid modeling and octree structures and their extensions.

Gargantini (Gargantini 1989) represented the research works at the University of Western Ontario in terms of modeling natural objects. Wireframe, Generalized Cylinders, Boundary Representation (B-rep), Constructive Solid Geometry (CSG), Spatial Enumeration and Design-by-Feature are among the most well-known modeling paradigms (Mortenson 1985, Requicha 1980, Requicha and Voelcker 1982, Samet and Webber 1988). Gargantini (Gargantini 1989) represented his works in representing 3-D objects via the Spatial Enumeration or Volume Representation approach. In a Volume Representation, the entire object is explicitly stored via the location and size of its main components. Portions of the three-dimensional space can be stored as full or black if they belong to the object, empty or white otherwise (Gargantini 1989).

Gargantini (Gargantini 1989) also pointed out that octrees and related hierarchical structures are often used for encoding three-dimensional objects. The octree is a rooted, directed tree generated by recursive subdivision of an initial $2^n * 2^n * 2^n$ ($n > 0$) raster into nodes of octants. If all voxels forming a node are of the same attribute,

that node becomes a leaf with that attribute, otherwise the octant is subdivided into eight suboctants, the current node given eight sons, and the process repeated – at most n times (Gargantini 1989). In representing 3-D objects, usually two attributes are used, black for nodes which belong to the object, white for those which do not (Gargantini 1989). Figure 3.1 illustrates an object represented with $n = 2$ while Figure 3.2 shows its corresponding octree representation.

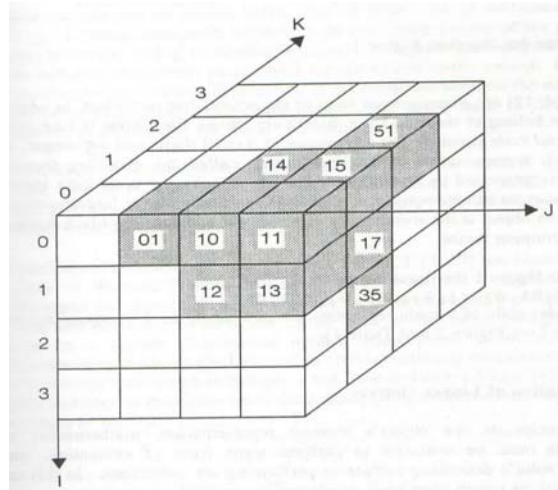


Figure 3.1 Representation of an Object with $n = 2$ (Modified after Gargantini 1989)

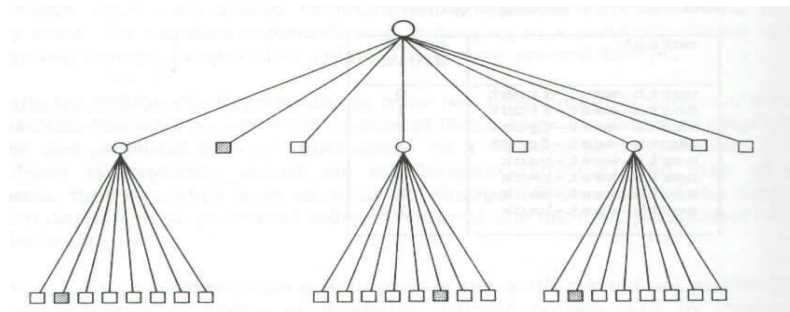


Figure 3.2 Octree for Object in Figure 3.1 (Modified after Gargantini 1989)

Gargantini (Gargantini 1989) also presented an experimental in-house software package named LINOCT that was developed at the University of Western Ontario for displaying octrees to realize 3D visualization.

Brunet (Brunet 1989) summarized several octree schemas to represent volume models in 3-D spatial modeling. Brunet classified the schemas used to represent rigid solids into ambiguous schemas and unambiguous schemas. The traditional system of specifying solids was the first method considered for storing objects in computer (Requicha and Voelcher 1982). However, it was announced to be inappropriate to handle 3-D geographic data due to its drawbacks including (1) the number of views and sections needed to generate an unambiguous model depends on the object to be designed, and (2) the modifications made to one view are difficult to apply to others without danger to the coherence of the system (Brunet 1989). Another ambiguous schema named Wireframe Modeling System came into use in the 1970's. The objects were represented by the set of vertices of the solid and by the set of edges which join pairs of vertices (Brunet 1989). It allows dihedral and isometric projections and perspectives of the solid. The most serious drawback of this system is that it is ambiguous for some objects and cannot guarantee the validity of the models and therefore, cannot give realistic visions.

Brunet (Brunet 1989) also summarized the main unambiguous modeling systems at that time. They included:

- 1) Primitive Instancing

This system describes solids formed by the grouping of objects from a finite number of families characterized by a finite number of parameters (Brunet 1989). This system is unambiguous, easy to validate, and simple to use, but it cannot support Boolean operation and therefore, it is inappropriate to support 3-D analysis (Brunet 1989).

- 2) Decomposition Models

Decomposition models represent solid objects by the union of a set of disjoint elementary cells without holes (Brunet 1989). The representation of the solid consists of the set of cells that are contained in it. Researchers have spent many efforts into these models and as a result, many modeling schemas have been produced for decomposition models.

The simplest decomposition for solids is the spatial enumeration, in which the cells (voxels) are elementary cubes of a fixed volume within a uniform mesh that divides space (Brunet 1989). Every cell can be simply represented by the coordinates of a certain singular point, such as the center of gravity (Requicha 1980).

Another schema is named bintrees that deploy recursive division schemas for the representation of solids. They divide an initial cube until homogeneous blocks inside or outside the solid are obtained (Brunet 1989). Bintrees can be used for the representation of two-dimensional pictures, 3-D solids, or even 4-D objects. However, geometric transformations are complex and adding difficulties to 3-D analyses (Brunet 1989). As a variation to bintrees, the binary space partition (BSP) methods arose as efficient rendering algorithms for static scenes in which the preprocessing of building the binary tree is worthwhile (Fuchs, Kedem, and Naylor 1983).

The third methods are named Octrees, which are also generated by recursive subdivision of a finite cubic universe (Brunet 1989). They resemble bintrees in that subdivision of heterogeneous nodes is performed in a standard way which does not depend on the object being modeled. The difference is that they are always divided into their eight octants – cubes of halved side dimensions and one eighth the previous volume – instead of halving the volume (Brunet 1989). In addition to the classical octrees, there are many variations such as vector octrees and face octrees. In this report, Brunet (Brunet 1989) detailed the main representations of these different octrees and associated algorithms.

3) Boundary Representations

In this model, the solids are determined by the set of points which belong to their surface; points inside the solid are separated from those outside (Brunet 1989). The boundary is represented by a disjoint set of faces bordered by one or more circular rings of edges which intersect in vertices. The information associated with the components of a surface (faces, edges, and vertices) has two parts. One is the geometry, which includes the dimension and mathematical equation (or location) in space of each component. The other is the topology that describes the connection between the elements (Eastman and Weiler 1979).

This system is unambiguous, non-unique, nor very concise (Brunet 1979). However, the calculation of geometric properties is very simple and visualization is simple, which makes it one of the most widely used representation schemes (Brunet 1989).

4) Constructive Solid Geometry

The constructive solid geometry (CSG) approach represents solids as a combination of basic solids through regularized Boolean operators (Requicha and Voelcker 1982). It is unambiguous, concise, and easy to validate but non unique. The visualization of the model with hidden lines and geometric calculations are rather complex (Brunet 1989).

Durst and Kunii (Durst and Kunii 1989) represented their researches in the quadtree, octree, and polytree models in representing 3-D objects. These models were examined in detail by these two authors. Improvements of these models, particularly for the storage requirements, were also proposed in this workshop.

3.3.2 Applications of 3-D Spatial Modeling

Around year 1990, the attention was focusing on the design and implementation of 3-D GISs in a range of geoscientific application areas. The developments of 3-D models and data structures were reported in the fields of oil exploration (Youngmann 1989), mining (Bak and Mill 1989), hydrogeology (Turner 1989a, 1989b, 1989c, and 1989d), geological modeling (Kelk 1989, Bristow and Raper 1991), environmental monitoring (Smith and Paradis 1989), civil engineering (Raper and Wainwright 1987), and landscape architecture (Batten 1989).

In the 1989 NATO Advance Research Workshop, applications of three-dimensional geoscientific modeling were presented. Sims (Sims 1989) described his experiences with two commercial systems, Interactive Volume Modeling (IVM) developed by Dynamic Graphics Inc. and Stratigraphic Geocellular Modeling (SGM) developed by Stratamodel Inc. in the field of oil industry. Harbaugh and Martinez (Harbaugh and Martinez 1989) discussed some problems in representing geological well data and seismic data in petroleum-bearing regions via 3-D GIS. Houlding (Houlding 1989) illustrated the application of 3-D modeling in the field of

mining. Turner (Turner 1989a, 1989b, 1989c, and 1989d) demonstrated the application of 3-D GIS to Hydrogeological studies.

In year 1989, Raper brought together the 3-D applications in GISs into a book titled “Three Dimensional Applications in Geographic Information Systems.” In this collection, Nicholas and Driel (Nicholas and Driel 1989) presented an overview of the available 3-D display tools and discussed some applications of these tools at the USGS. Gold (Gold 1989) discussed different approaches to contouring scattered datasets and argued for the use of adjacency rather than distance criteria in interpolation. Leenaers, Burrough, and Okx (Leenaers, Burrough, and Okx 1989) considered how co-kriging of elevation with lead contamination values can lead to better descriptions of contamination distribution than standard surface fitting algorithms. Dikau (Dikau 1989) showed the application of surface modeling to automatic landform classification from publicly available digital elevation models. McLaren and Kennie (McLaren and Kennie 1989) discussed the techniques for rendering images and illustrated some typical applications. Kraak (Kraak 1989) considered the problems of cartographic design in 3-D from the point of view perception. In addition, Turner (Turner 1989a, 1989b, 1989c, and 1989d) illustrated the application of 3D GISs to dynamic modeling in hydrogeology. Youngmann (Youngmann 1989) considered the data structure alternatives for the handling of geo-objects in 3-D in the petroleum exploration industry. Unger and others (Unger et al. 1989) showed the use of 3-D GISs in the geophysical modeling of the Earth’s crust in Central Maine.

Furthermore, some 3-D Geographic Information Systems were available by 1990, such as the previously mentioned IVM from Dynamic Graphics Inc. However, the majority were still in heavy customizations of existing systems by producing extensions and tools. The main objective in these customizations was inclined to 3-D visualization rather than 3-D analysis (Brunet 1989).

3.3.3 Observations Regarding Research and Applications of 3-D Spatial Modeling around 1990

The research efforts summarized in sections 3.3.1 and 3.3.2 are the pioneer researches in 3-D spatial modeling and applications in GISs. They provided the foundation for recent developments in this field. However, these

pioneer research projects and applications have limitations at many aspects due to the limitations from the computer hardware and software available at that time and knowledge in GISs and 3-D modeling during that time period. Important observations are obtained and listed as bellow after careful examination of these earlier research efforts in 3-D modeling around year 1990.

- The preliminary functionality requirements of 3-D spatial modeling were laid out.
- 3-D analysis was differentiated from 3-D rendering and visualization.
- Limitations from non-geoscientific approaches to 3-D modeling such as computer-aided design (CAD) techniques were recognized. New data structures and model were developed and used to overcome these limitations.
- These new data structures, models, and their applications focused on volumetric modeling of 3-D objects. Basic volumetric units such as cubes and spheres were deployed to model 3-D objects as 3-D polygons.
- The functionality of the earlier systems was limited to generating ‘pictures’ of the features under study which can not be analyzed or interrogated (Raper 1989). In other words, 3-D analyses were always left as application-specific.
- Research efforts that had been put to modeling linear objects in a 3-D space were almost nonexistent.
- Data quality issues in 3-D spatial modeling were recognized.

3.3.4 Recent Research Efforts in 3-D Spatial Modeling

Since the 1990’s, 3-D spatial modeling has attracted tremendous research interests. One prominent reason is that the effects from the third dimension are so significant for many applications, particularly those having their study areas located in mountainous areas, where the elevations must be taken into consideration and the problem need to be solved in a 3-D context (Li 1994, Xiang 1996, Barnes 2002, Berry 2003). As a result, many individual reports have appeared in academic and professional journals, magazines, and conference proceedings. New data structures and models were proposed based on improvements and modifications upon the earlier research results. The 3-D spatial modeling techniques were creatively applied to fields including geoscientific modeling, natural resource management, shoreline modeling and management, engineering, and others. 3-D GISs were being developed with some 3-D analysis functionality to solve complex 3-D problems.

In addition, with the rapidly developing computer technologies and remote sensing, the integration of key 3-D modeling concepts with them greatly improved the applications of 3-D modeling.

Li (1993, Li 1994, Li et al. 1994), together with other researchers published several articles focusing on 3-D modeling and 3-D GISs. It was pointed out that 3-D data structures in GISs were becoming of increasing interest in the GIS community for two reasons. First, there is a lack of real 3-D GIS data structures and systems. On the other hand, 3-D data structures are necessary in many applications such as most engineering applications, in which the project areas are relatively small and the vertical geometric information or the associated attribute information in this direction cannot be neglected (Li 1994). Li also identified the most significant difference between CAD packages and GIS was that in CAD, the geometric description of objects was the first priority while non-spatial information, especially GIS topologic and attribute information, was not considered (Li 1994). In other words, even though CAD has strong capability in 3-D modeling, it cannot substitute GISs to accomplish the tasks of 3-D representation and analysis.

Li (1994) defined 3-D geometric representations as geometric descriptions of objects for storage, geometric processing, display, and management of 3-D spatial data by computers in GISs. The 3-D geometric representations were classified into two groups, namely surface based representations and volume based representations, as illustrated in Figures 3.3, 3.4 and 3.5.

Li (Li 1994) also pointed out that the most significant difference between 2-D and 3-D GIS data acquisition was the consideration of elevation information of spatial objects. In case there are available elevation data, these data should be used to construct 3-D data in order to avoid the expensive cost of field survey (Li 1994).

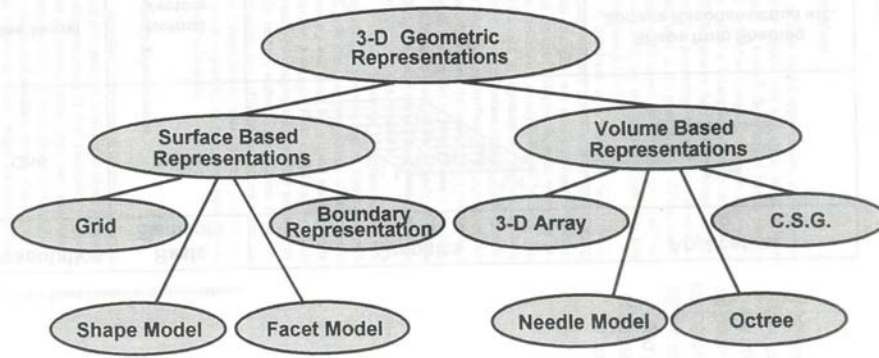


Figure 3.3 Classification of 3-D Geometric Representations (Li 1994)

Representation	Basic Elements	Graphics	Application								
Grid	Points		DEM Contour Lines 3-D Meshes etc.								
Shape Model	Normal Vectors		Shape from Shading Surface Reconstruction etc.								
Facet Model	Facets (Triangles)		DEM Contour Lines Shading Visualization Ray Tracing etc.								
Boundary Representation	Primitives	<table border="1"> <tr> <td>Volume</td> <td></td> </tr> <tr> <td>Face</td> <td></td> </tr> <tr> <td>Edge</td> <td></td> </tr> <tr> <td>Point</td> <td></td> </tr> </table>	Volume		Face		Edge		Point		Modeling Planning CAD Engineering etc.
Volume											
Face											
Edge											
Point											

Figure 3.4 Surface Based 3-D Geometric Representations (Li 1994)

Li (Li 1994) also recognized that for regular shaped objects (man-made objects such as buildings), mathematical functions could be deployed to represent them in a three-dimensional space. However, this might incur conversion and compatibility issues. Chen and Ikeda (Chen and Ikeda 1994) proposed a 3-D spatial modeling approach based on Delaunay tetrahedral tessellation (DTT) to model both regular and irregular

objects. DTT is an aggregate of connected but non-overlapping convex polyhedrons (Chen and Ikeda 1994). It has been concerned as one kind of simple data structure for spatial modeling due to its many advantages such as simplified complexes, the linear combination criterion, and easy to transform and visualize (Chen and Ikeda 1994). Researchers had tried to use this technology to model 3-D objects (Raper and Kelk 1991, Tsai 1993). Chen and Ikeda (Chen and Ikeda 1994) demonstrated a hybrid 3-D model of GISs based on a DTT that was generated by a raster-based algorithm. One advantage of this 3-D GIS is that it models both regular and irregular objects in a uniform data structure, which is also its disadvantage because it is difficult to differentiate regular objects such as buildings from the surrounding irregular objects such as the continuous Earth surface.

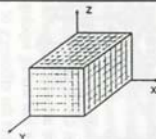
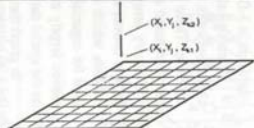
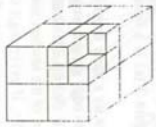
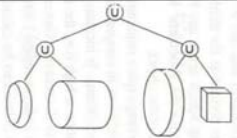
Representation	Basic Elements	Graphics	Application
3-D Array	Voxels		Intermediate Representation
Needle Model	Needles		Spatial Analysis Planning Engineering etc.
Octree	Octants		Dynamic Modeling Robotics Boolean Operations etc.
C.S.G.	Primitives		Planning Animation CAD Engineering etc.

Figure 3.5 Volume Based 3-D Geometric Representations (Li 1994)

One more interesting finding is that many research efforts went into the integration issues of elevations data with 2-D GISs and remote sensing technologies with 3-D GISs. The essential component of expanding 2-D GISs into 3-D GISs is to add the third dimension, elevation or height. On one hand, most 2-D GIS data are available. On the other hand, elevation data in the format of DEM or TIN are also available. Since data collection is very expensive, it is very natural that integrating the existing elevation data with 2-D GIS data into 3-D GISs has attracted many research efforts. For example, Fritsch and Schmidt developed 3-D query spaces

by integrating digital terrain model (DTM) in GISs (Fritsch and Schmidt 1993). A series of studies were reported, which concentrated on creating buffers that take into consideration of the effects from the slopes or elevation changes (Holder 1992, Stratton 1993, Xiang and Stratton 1994, Xiang 1996, Narumalani et al. 1997, Hiegl 2001, Barnes 2002). Tools aiming at creating 3-D buffers by taking into consideration of the elevation data in either raster or TIN format have also emerged (Larson 2003).

In addition, remote sensing technologies provide ways to view and extract 3-D objects from analogue or digital images. Researchers have recognized this capability and proposed many applications such as terrain modeling and 3-D object extraction (Gevers and Smeulders 1994, Dissard and Jamet 1995, Kim and Muller 1995, Roux et al. 1995, Haala and Anders 1997, Lammi 1997, Wang et al. 1997, Englert and Cremers 1997). The emerging photogrammetry technologies and the improvements on the traditional photogrammetry technologies such as aerial applications, terrestrial applications, close-range solutions, and laser scanning, make it easier to deliver 3-D geospatial solutions to various applications (Limp and Cothren 2003).

With rapid developments in computer technologies, geostatistical technologies, and the advancement of 3-D geological modeling systems, it was recognized as early as 1993 that 3-D spatial modeling was no longer just a research tool (Phillips 1993). However, regardless of the developments in 3-D spatial modeling, 3-D analytical software functions were practically nonexistent (Dobson 2003). The key to 3-D analysis lies in creating the same data structures – geometry, topology, and attributes – for the third dimension that is missing in 2-D GISs (Dobson 2003).

3.3.5 Observations Regarding Recent Research Efforts in 3-D Spatial Modeling

Reviewing the recent developments and research in 3-D spatial modeling, 3-D GISs, and their applications lead to following observations.

- Traditional 3-D data structures have been improved and enhanced while new 3-D data structures have been proposed.

- The research focus of 3-D data structure is still limited to model 3-D volumetric objects (3-D polygons or surfaces), which has been recognized as the most challenging and complicated scenario in 3-D modeling.
- The 3-D analytical functions are still almost nonexistent.
- The 3-D GIS applications still require brute-force solutions, which indicate tremendous customization and development efforts (Dobson 2003).
- Research efforts put to modeling linear objects in a 3-D space are still almost nonexistent.
- Integrating existing elevation data with 2-D GIS data has been approved to be a practical and economic way to conduct 3-D analyses.
- Data quality issues become so prominent that they are being dealt with as a separate research topic.

3.4 SPATIAL MODELING IN TRANSPORTATION

In the transportation context, three classes of GIS models are relevant: field models, discrete models, and network models (Goodchild 1992, 1998). Field models represent the continuous variation of a phenomenon over space. Terrain modeling is a typical example and is widely used in spatial modeling in transportation. Discrete models represent discrete entities (points, lines, or polygons) populate over space. These models can model spatial objects in transportation such as highway rest areas and toll barriers. Network models represent topologically connected linear entities such as roads that are fixed in the continuous reference surface (Thill 2001).

3.4.1 Modeling Surfaces

Surface modeling is a general term that is used to describe the process of representing a physical or artificially created surface by means of a mathematical expression. Terrain modeling is one particular category of surface modeling that deals with the specific problem of representing the surface of the Earth (Kennie and Petrie, 1991). A surface may be defined as a three-dimensional object with its third dimension varying continuously. A terrain is a surface object with elevation as the continuously varying third dimension. There are four different ways of modeling a surface: point models, contour line models, TIN, and DEM.

In modeling terrains, the continuously varying third dimension -- elevation 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 terrain models contain some inaccuracies, each of which depends on a number of factors (Shearer 1990), as the price to represent the real world objects with reduced but manageable information. Thus, the challenge in modeling surfaces is to provide high accuracy elevations with low consumption of resources.

The examples used in this section to illustrate the concepts in terrain modeling are extracted from the paper “An Exercise in Digital Elevation/Terrain Models: From Point to Mathematics” (Massasati 2000).

3.4.1.1 Breaklines and Mass Points

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. 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 and ridges. Breaklines and mass points provide input data to construct the final surface model.

3.4.1.2 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 3.6 provides an example of point model that

represents a pyramid. The dimensions of the pyramid are simplified and reduced entirely to values of 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.

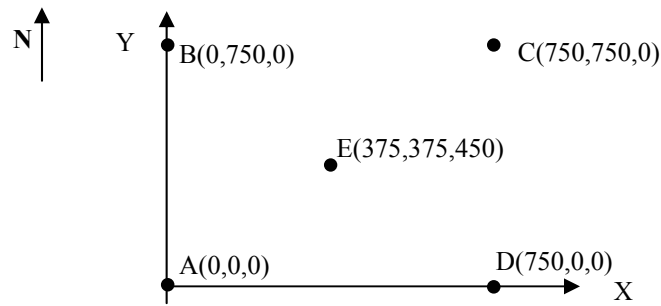


Figure 3.6 The Point Model for a Pyramid

3.4.1.3 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.

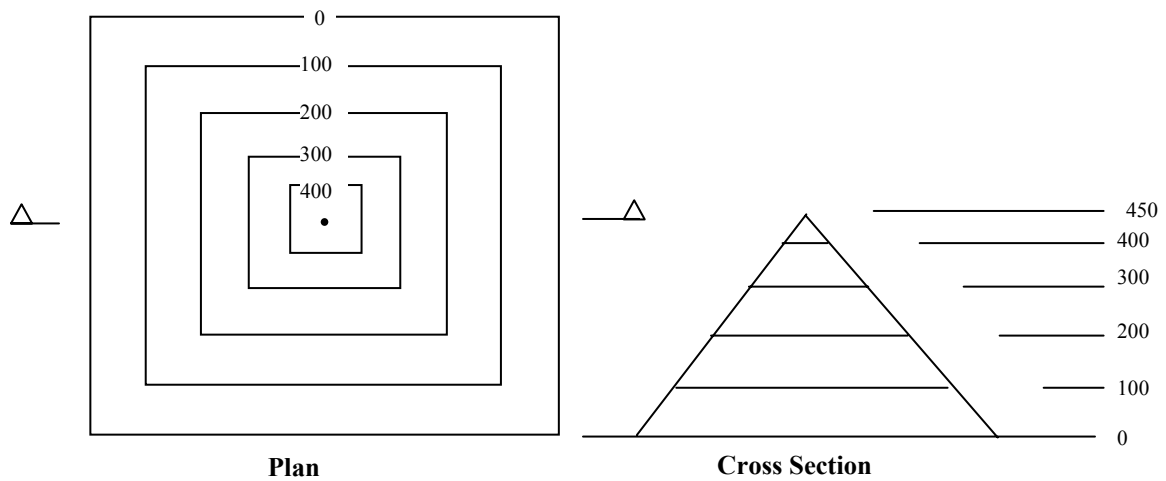


Figure 3.7 The Contour Line Model for a Pyramid

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 3.7 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.

3.4.1.4 DEM

DEM 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 in representing surfaces depends on several factors, one important factor being grid resolution (Heesom and Lamine 2001). DEM 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 DEM 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.

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 3.8 The DEM for the Pyramid

DEM 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 3.8 illustrates the DEM representing the surface of the pyramid using a 75×75 grid size. The maximum elevation within a cell is used as the elevation for that cell for illustration purposes.

3.4.1.5 TIN (Triangulated Irregular Network)

The Triangulated Irregular Network (or TIN) was designed by Peucker and co-workers (Peucker *et al.* 1978) for digital elevation modeling to accurately model surfaces while avoiding 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 3.9 provides the surface of the pyramid represented by a TIN.

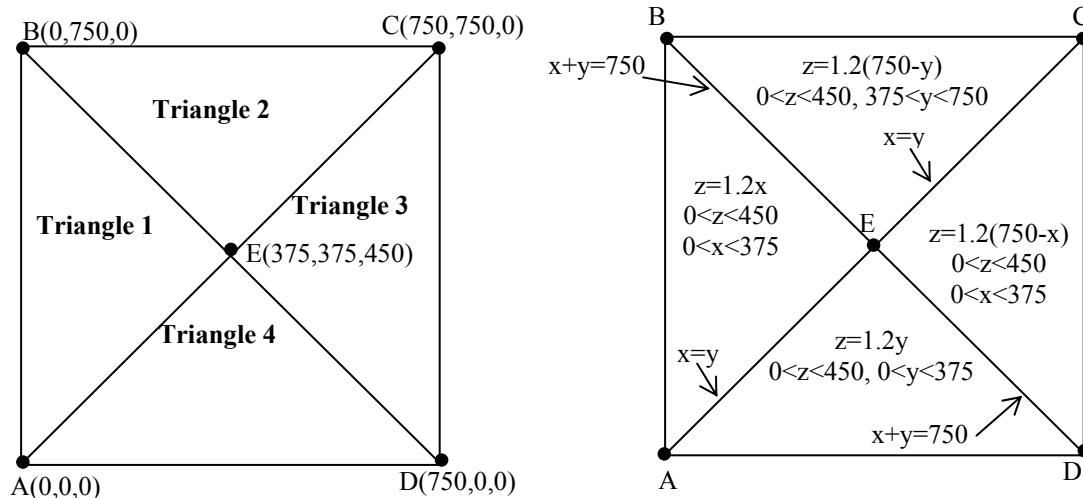


Figure 3.9 The TIN Model for the Pyramid

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. It is worth pointing out that TINs and DEMs 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/Y-coordinates analytically. For example, using the numerical values of Figures 3.9, each surface (of the pyramid) can be represented by the following mathematical model.

$$Z = a*X + b*Y + c \quad (3.1)$$

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 3.9 as shown in the following equations.

$$0 = a*0 + b*0 + c \quad (3.2)$$

$$0 = a*750 + b*0 + c \quad (3.3)$$

$$450 = a*375 + b*375 + c \quad (3.4)$$

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

$$Z = 1.2*Y, \quad \text{where: } 0 < Y < 375 \text{ and } 0 < Z < 450 \quad (3.5)$$

Likewise, equations for other triangular surfaces and lines bordering two triangular surfaces are acquired and represented in a like manner as shown in Figure 3.9.

3.4.1.6 Researches in Terrain Modeling

The subject of terrain modeling has been one of the continuously growth over the last 30 years. Growth in the applications of terrain modeling has been especially rapid in recent years as a result of the advances in computer graphics and imaging technologies and the parallel developments in the software used for display and visualization (Petrie and Kennie, 1991). As a result, digital terrain modeling (DTM) technologies are used routinely in many fields, especially in civil engineering such as landscape architecture and planning and cut and fill applications in construction.

However, there are only few publications on this subject area before year 1991. Davies and McCullagh's book, *Display and Analysis of Spatial Data*, published in 1975, dealt with the overall area of spatial data in its widest sense, but not with digital terrain modeling in particular (Davies and McCullagh, 1975). Since then, a few proceedings have appeared reporting the results of specialist workshops (ASPRS-ACSM, 1978, Kubik and Roy, 1986, Toomey, 1984). Another book authored by Lancaster and Salkauskas in year 1986 dealt with the purely mathematical aspects of curve and surface fitting, part of which has much relevance to terrain modeling (Lancaster and Salkauskas, 1986).

In year 1991, a book titled *Terrain Modeling in Surveying and Civil Engineering*, was published. This book was developed based on two three-day courses on Terrain Modeling in Surveying and Civil Engineering organized by Petrie and Kennie and given at the Universities of Surrey and Glasgow respectively in April and September 1987 (Petrie and Kennie, 1991). It provided an insight into the current state of the art in the areas of data acquisition and systems and software at that time. This book provided the basic theories in digital terrain modeling, data acquisition approaches and data displaying, and the systems and software existing by that time to aid in digital terrain modeling. Examples and case studies about the applications of terrain modeling in surveying and civil engineering were provided. Further more, this book revealed that the accuracy of terrain models was already an issue by that time.

The accuracy of digital terrain models (DTMs) is of concern to both DTM users and DTM producers (Li, 1994). The research direction in digital terrain modeling has shifted from the development of interpolation techniques to the assessment and control of DTM quality (accuracy) (Li, 1994). On one hand, there are numerous empirical studies in this area (Ackermann, 1978, 1979, Askermann and Stark, 1985, Ley, 1986, Day and Muller, 1988, Tuladhar and Makarovic, 1988, Li, 1990, 1992, Kumler, 1990). Among these studies, only Ackermann managed to propose an empirical model for a particular case (Ackermann, 1979). On the other hand, a few mathematical models have been developed through theoretical analysis (Makarovic, 1972, Kubik and Botman, 1976, Frederiksen, 1981). However, the evaluations by both Balce and Li indicate that no single one of these models is capable of producing reliable predictions (Balce, 1987, Li, 1990, 1993). As a supplement, Li

conducted a comparative study to determine the accuracy of digital terrain models that are derived from different data models (Li, 1994). It was argued by Li that it was still important to investigate accuracy issues in digital terrain modeling because many basic problems in this topic still remained unsolved and existing mathematical models were either incapable of producing reliable results or impractical to use (Li, 1994).

3.4.2 Modeling Discrete Features

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. Each object is described by its characteristics and properties, which are usually referred to as attributes. 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. DEM data are originally point data. Point data are the base to build TINs and to derive mathematical functions.

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). A linear object consists of a series of connected points (vertices). These vertices use coordinates to locate themselves spatially and in turn, to locate the linear object spatially. One very important feature in addition to locational coordinates for line features is length. This attribute allows users to measure the spatial extent of a line. Other than length that is the major concern of this research, shapes and orientations are also critical attributes for linear objects.

Polygon objects represent objects that occupy “two-dimensional” space. A polygon object can be delineated from the rest area via a series of linear objects that are connected and closed. 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.

Without taking into consideration of the particular characteristics such as connectivity and the dynamic nature of the attributes of the linear transportation objects (roads, for example), spatial objects in transportation can be modeled as points, lines, and polygons with associated attributes. While transportation features such as highway rest areas and toll barriers could be modeled as discrete points or polygons, the above-mentioned two properties of the linear objects in transportation indicate there must be another way to model these connected linear objects, which is described in the next subsection.

3.4.3 Modeling Transportation Networks

In modeling transportation networks, transportation engineers are more interested in the road network itself and the events on/along roads. The vector methods modeling transportation networks are too generic to support all transportation specific applications and support (Thill 2000).

The GIS network model represents topologically connected entities (roads, for example) that are fixed in the continuous surface. The network model is built around the concepts of arc and node and captures the topology in transportation network and thus, enables the movement analysis. In fact, many transportation applications only require a network model to represent data (Thill 2000). The following are some examples of transportation applications that only require network models to represent data (Thill 2000).

- Pavement and other facility management systems;
- Real-time and off-line routing procedures, including emergency vehicle dispatching and traffic assignment in the four-step urban transportation planning process;
- Web-based traffic information systems and trip planning engines;
- In-vehicle navigation systems; and

- Real-time congestion management and accident detection.

Figure 3.10 illustrates a simple network model with 5 bi-directional links connected by 4 nodes. The link table (Table 3.1) implicitly stores the connectivity (topology) by stating the two ending nodes for each link. In other words, two links with a common node are connected to each other. For illustration purpose, the travel impedance (time) for each link is included in the travel impedance table (Table 3.2). While these two tables stores the topology and impedance information, there are other tables that could be built to store this information, for example, a node table might be used to store connectivity information by stating all the links connected to a specific node. Similarly, instead of stating the travel impedance for each link, the travel impedance can be specified for pairs of nodes. The point here is that even for a very simple network as in Figure 3.10, there are still many design issues to be considered from the relational database management point of view.

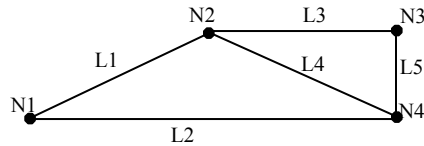


Figure 3.10 A Simple Network Model

Table 3.1 Link Table

Link	Node1	Node2
L1	N1	N2
L2	N1	N4
L3	N2	N3
L4	N2	N4
L5	N3	N4

Table 3.2 Impedance Table

Link	Impedance (minutes)
L1	12
L2	13
L3	24
L4	15
L5	19

In transportation research, there is a compelling need for attributing the infrastructure lines. As illustrated in the simple network model in Figure 3.10 and Tables 3.1 and 3.2, the travel impedance can be treated as an attribute belonging to each individual link. For the nodes, they could also have attributes such as traffic priorities and presence of traffic signals.

So far, the assumption about the links in the network models is that network links are homogeneous. In other words, each attribute for a specific link has a unique value, for example, the travel impedance attribute. However, this is not always the case. In transportation network models, this is always not the case. Number of lanes, pavement width, pavement conditions, and posted speeds are all but a few attributes that cannot be constrained to be constant between terminal nodes of a link (Thill 2000). Similarly, traffic parameters of speed, flow, and capacity cannot be expected to be constant between widely spaced junctions (Thill 2000). The dynamic nature of these attributes precludes the possibility to permanently edit the network to maintain the homogeneity of each link on each attribute. Instead, each attribute can be viewed as a spatial (linear) event occurring on the network. Thus, the problem becomes how to locate events on the different road network models.

3.4.3.1 Linear Referencing Systems (LRS)

The core for linear referencing is that the location of any unknown point along a linear feature can be determined by specifying the direction and distance from any known point to the unknown point (Vonderohe et al. 1995). This concept solves the problem of locating events along the network and is the basis for all linear referencing methods.

The development of measurement systems with road-based transportation networks in the United States dates from the earliest “post roads” in New England in the 1700’s, where distance reference marks were placed along the roadways to guide travelers (Geo Decisions 1997). In the 1800’s, the first national transportation project was the National Road, which linked Cumberland, Maryland with Wheeling, Virginia. Original milepost markers can still be found on that route (Geo Decisions 1997).

Linear referencing systems (LRSs) are used to manage and promote analysis of spatial data elements on transportation networks because inherent in these data collection and analysis efforts is a requirement to describe such data in terms of a location. The seminal work on LRSs was **Synthesis of Highway Practice 21: Highway Location Reference Methods**, published by the Highway Research Board in 1974 (Highway Location Reference Methods 1974). This report defined LRSs as well as the distinction between an LRS and a

linear referencing method. The LRSs encompass a larger set of office, field management, and integration procedures that include one or more linear referencing methods. The linear referencing system incorporates “procedures that relate all locations to each other...it includes techniques for storing, maintaining, and retrieving location information (O’Neill 1996)”. A location referencing method is “a way to identify a single location; i.e., to reference a specific position with respect to a known point (O’Neill 1996)”.

3.4.3.1.1 LRS Data Model

Transportation agencies from time to time have studied the location referencing methods they use and sought to adopt standards for those (Briggs and Chatfield 1987). During one study, the Michigan State Department of Transportation identified 38 location referencing methods in use by the agency (Vonderohe et al. 1995). Researchers have spent tremendous amount of efforts in resolving the multiplicity issue, which means the central need to integrate not only multiple scales of geography and cartographic (coordinate-based) data from multiple sources, but also multiple network models, each of them necessary for particular applications (Vonderohe et al. 1995). In the workshop held in Milwaukee, Wisconsin in 1994, a generic data model for LRSs was proposed, which was designed to be the basis upon which extensions could be made to meet specific needs of various applications. Figure 3.11 illustrates the cartographic presentation of the data model.

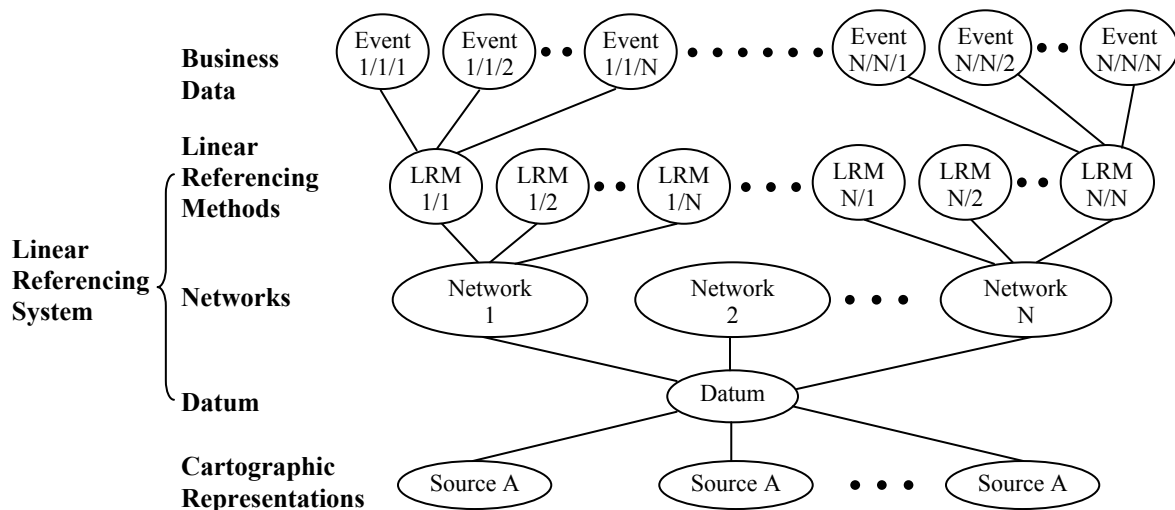


Figure 3.11 Conceptual Overview of the LRS Data Model (Modified after Vonderohe et al. 1995)

The central notion for this data model is a linear datum that supports multiple cartographic representations of any scales and multiple network models for various application areas (Vonderohe et al. 1995). The datum provides the basis for transforming among various linear referencing methods, network models, and cartographic representations. It also links the model to the real world via attributes that describe its location and spatial characteristics referencing to the real world (Vonderohe et al. 1995).

The primary spatial aspects of the model are topology in the datum and networks, distance measures in the datum, weights in the networks, and offsets to locate zero-dimensional objects (Vonderohe et al. 1995). Among these four primary spatial aspects of the LRS model, two of them are distance measurements (distance measures in the datum and offsets), which indicates the critical role of distance measurement in LRS.

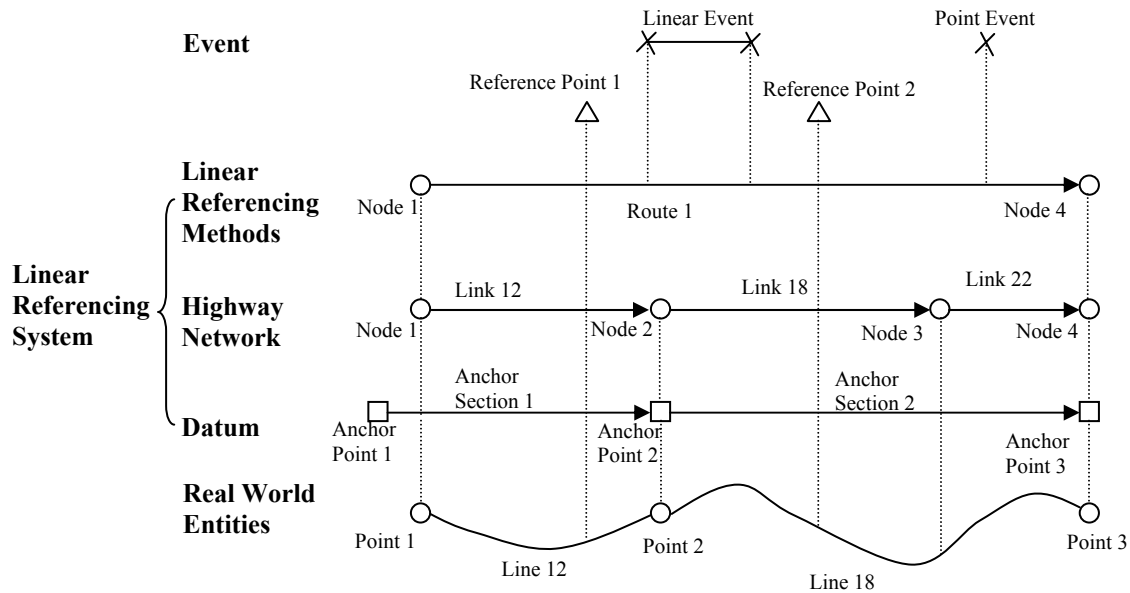


Figure 3.12 The Conceptual Model of a Linear Referencing System (Summarized after Vonderohe et al. 1995)

Figure 3.12 provides a conceptual model of a linear referencing system in which the datum is composed of anchor sections and anchor points and every component references to the linear datum. In this conceptual model, an anchor point is defined as a zero-dimensional location that can be uniquely identified in the real world; or its position can be determined and recovered in the field (Vonderohe et al. 1995). The anchor section is defined as the continuous, directed, non-branching linear feature, connecting two anchor points, whose real world length can be determined in the field (Vonderohe et al. 1995). Anchor points have location description

attribute to provide the necessary information to determine and recover the anchor points' positions in the field. Anchor sections are directed by specifying a "from" anchor point and a "to" anchor point. Anchor sections have a "distance" attribute, which is the length of the anchor section measured on the ground (Vonderohe et al. 1995).

3.4.3.1.2 Location Referencing Methods

As illustrated in the LRS data model, LRS incorporate various location referencing methods for specific transportation applications. This section provides a summary of major location referencing methods in use at state DOTs with general strengths and weakness, which is the result of a study conducted by Geo Decisions for NCDOT in 1997 (Geo Decisions 1997).

(1) Route-Milepost/Milepoint/Reference Post/Route-Segment-Offset (Figure 3.13)

These variations on similar approaches comprise the most commonly used referencing method in use at state DOTs (Geo Decision 1997). In general, a network is defined in terms of a route identifier with measured milepoint values. Field reference markers may be used as the primary method of implementation. For identification purposes, a route is specified as a collection of linear features variously defined as arcs, segments, links, or sections, and node features that provide additional linkage information. A beginning and ending measured distance is tied to the beginning and ending nodes. Events are located by offset distances that describe the geometric locations as measured distances along the route.

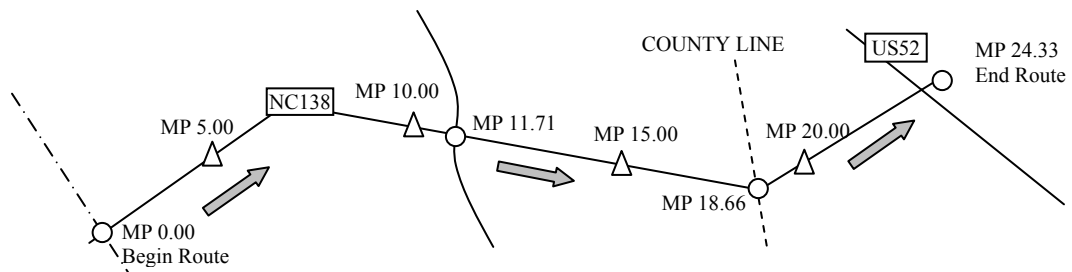


Figure 3.13 Milepost/Milepoint/Reference Post Location Reference Method (Modeled after Geo Decisions 1997)

(2) Link-Node (Figure 3.14)

A link-node system has much in common with the milepost system, but abstract relationships are defined among individual link and node features. For example, a point event may be described as occurring on link

963, halfway between nodes 3111 and 3112. While the functionality of locating data on a link-node system is sound, problems can occur in how the collection of links and nodes are defined, and how succeeding edits to the network are handled. Any data collected and referenced to links and nodes will have to be edited if the network definition changes.

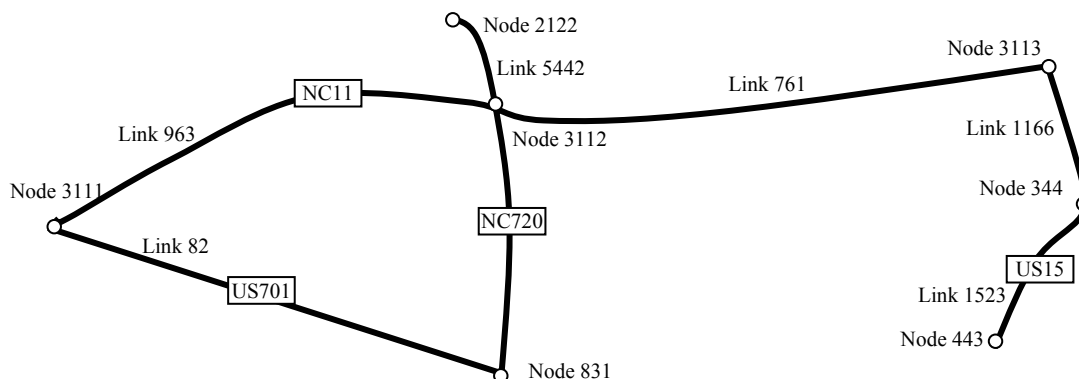


Figure 3.14 Link-Node Location Reference Method (Modeled after Geo Decisions 1997)

(3) Latitude-Longitude or Spatial (X-Y) Coordinates (Figure 3.15)

An increasingly common way to collect data in the field is through the use of global positioning systems (GPSs). A GPS collects data in latitude/longitude format. GPS data are usually transformed into another coordinate system as X/Y-coordinates for more practical use. As illustrated in Figure 3.15, this X/Y-coordinate information could be mapped to the road network directly. However, conversions are required to locate these X/Y coordinates on the network via distance measurements.

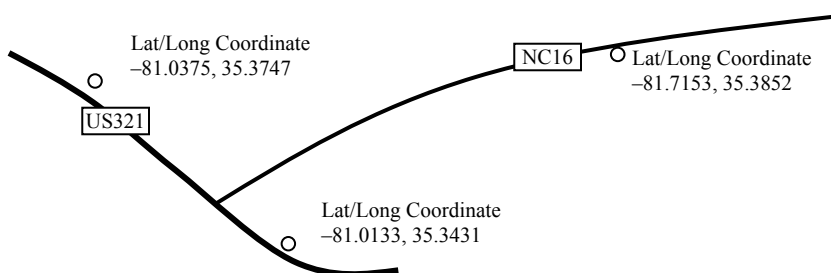


Figure 3.15 X-Y Point Location Reference Method (Modeled after Geo Decisions 1997)

(4) Address/Address Ranges/Intersections/Landmarks (Figure 3.16)

A common way of collecting data in the field, especially by public safety personnel, is through specification of address or address ranges, in conjunction with description of intersection locations. This system often

requires conversion to a format more meaningful for computerized analysis. Lookup tables often must be used to standardize address formats and location intersections. In addition, a method to include offset measurements must be incorporated. Figure 3.16 illustrates several examples for address locations.

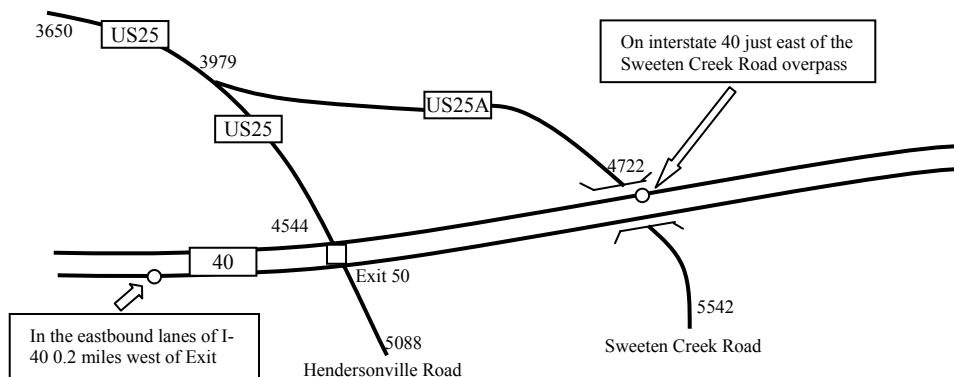


Figure 3.16 Address/Intersection Location Reference Method (Modeled after Geo Decisions 1997)

3.4.3.2 Current Research in LRS

The current research efforts in LRS could be categorized into the following three categories based on the LRS problems they are trying to solve.

(1) LRS Definition/Glossary

There has been considerable confusion in understanding, defining, and standardizing LRS-GIS related terms and concepts (Geo Decisions 1997). Various terms are used to describe the same LRS-GIS feature type, such as “arc”, “link”, and “segment”. To provide another example, “control point” and “control section” are often used interchangeably with “anchor point” and “anchor section”. FGDC and FHWA both have their own consistent definitions of LRS and LRS terms (FGDC 1994, FHWA 1998). They recognized the necessity and importance of having consistent definitions of LRS terms. However, detailed examinations regarding definitions from these two primary leading agencies in GIS-LRS research reveal that there are confusions about the terms and definitions (Geo Decisions 1997).

(2) Generic Linear Referencing Data Model

The NCHRP Report 359, “Adaptation of Geographic Information Systems for Transportation” (Vonderohe et al. 1993), is a landmark publication that spurred research in the area of linear referencing (FHWA 1998). It

recommends that transportation agencies develop conceptual organizing principles founded upon the notion of location as a data integrator.

There are many location referencing methods being employed in linear referencing systems and thus, the data integration and sharing has prompted the research efforts in searching for a generic reference data model as a tool to integrate these data and realize data sharing and transformation from one location referencing system to another (FHWA 1998).

A generic linear referencing data model was developed based primarily on the results of a workshop held in Milwaukee, Wisconsin in August 1994. This data model uses a single linear datum, based on anchor points and anchor sections, to associate transportation data with multiple cartographic representations and multiple network models as represented in Figures 3.11 and 3.12. The goal of this data model is to provide a generic data model that can be extended to meet specific needs of various applications (Vonderohe et al. 1995). The concept of linear datum enables the transformations between different location referencing methods, multiple networks, and cartographic representations at different scales.

Specifications and a prototype were developed for a linear referencing engine (LRS) as part of the GIS-T/ISTEA Management Systems Pooled Fund Study (PFS) (Fletcher 1995). The LRE was proposed as a robust data model framework in response to perceived needs among the GIS-T community for development of a standard model that incorporates different location referencing methods. The LRS was developed using Borland's Delphi software and demonstrated the successful intended transformations, but did not address many of the real world concerns associated with linear referencing (FHWA 1998).

In year 1997, a GIS-T enterprise data model was developed by Dueker and Butler (Dueker and Butler 1997). This is a more general model than the other two models in that it incorporates non-linear location referencing (for example, by GPS), area events, more detailed cartographic entities and non-transportation features (FHWA 1998). In Dueker and Butler's model, LRS components are related but generally fall into four

classes: geographic network, cartography, transportation network, and transportation topology (Kiel et al. 1999).

(3) Implementation Issues

This area of LRS research focuses on the physical implementation of LRS such as implementation strategies, transformation between two different location referencing methods, and location (distance and reference points) determinations.

Integration of disparate transportation data sources is a major ongoing concern at almost all departments of transportation today (Kiel et al. 1999). The researches focusing on data models mentioned above also provide implementation guides for each of the three generic data models described earlier.

On the other hand, researchers have recognized the importance of locating reference points and determining distances along linear network. In LRS, transportation data are usually referenced to highway networks by using a one-dimensional model (Baker and Blessing 1974). With this model, objects along a network are located using a set of known points on the network and distances and directions from the known points to the objects. An efficient way of distance (length) determination is a prerequisite in developing LRS due to the reliance of LRS on distance measurements.

Several studies were conducted to assess the accuracy issues in linear distance measurement for LRS. Quiroga (1999) studied the accuracy of linearly referenced data by using a GIS approach. In this study, the linear distances were computed using GIS rather than measured in field. Distances measured from four GIS packages are compared: ArcView v3.0, MapInfo v4.1, MGE v6.0, and TransCAD v3.1. This study concluded that GIS approach provided consistent distance measurements. It also revealed that the accuracy of computation depends on the type of models used to simulate the surface of the Earth.

Another closely related study was carried out by Rasdorf and Cai for NCDOT in 2001. This study evaluated the accuracy of length determination using a GIS approach. In this study, the Arc Macro Language (AML)

was used to develop customized program to carry out length calculation. This study also identified slope and distance as two factors that had significant effects on the accuracy of the linearly referenced data (Rasdorf and Cai 2001, Rasdorf et al. 2001).

With the rapid developments in GPS and its widely applications in transportation, a GPS provides another alternative to determine distance efficiently. The evaluation and assessment of GPS in collecting linear objects are being carried out widely. To appreciate the research efforts in this field, an example from Thailand is cited here. This study evaluates the accuracy of GPS in determining road centerline lengths with different real time GPS settings (Ganeshkumar 1999).

3.4.3.3 Federal Efforts

The Federal Highway Administration (FHWA) is coordinating works to improve LRS-GIS understanding (Federal Highway Administration 1998). FHWA requires all state transportation agencies to collect traffic data and other related information for highway links that comprise the National Highway Planning Network (NHPN). This information is gathered into FHWA's Highway Performance Monitoring System (HPMS) as the primary vehicle for the collection and reporting of data on highway conditions, performance, and usage (Geo Decisions 1997).

FHWA recognized the complexity and inconsistencies that exist in the ways transportation data are being collected and modeled spatially. FHWA identified the key to integrating and analyzing these data is the location and determined that linear referencing provides a set of methods and procedures for recording and retrieving locations along linear networks and thus, should be used to integrate transportation data spatially. The NHPN began with a manual LRS structure. In year 1996 computer software named LRSEDIT was released to assist in this task.

FHWA has also worked with the Oak Ridge National Laboratory and the USDOT's Bureau of Transportation Statistics to develop the HPMS structure and the LRSEDIT program. Both entities have also been involved with complementary research, especially in intelligent transportation system-related LRS issues. USDOT has

also sponsored creation of a National Transportation Atlas Database that will include LRS elements in its data structure.

In year 1998, FHWA published the “Linear Referencing Parishioners Handbook”, which provides background information about LRS, its applications in transportation via case studies conducted in four state DOTs. The relationship between LRSs and GISs and the implementation of LRSs in GISs are illustrated. Critical LRS implementation issues such as unique identifier, location and distance determinations, and quality control are included in this handbook (Federal Highway Administration 1998).

Federal Geographic Data Committee has identified transportation as one of the seven framework layers in the National Spatial Data Infrastructure (NSDI) (FGDC 1994). Its Ground Transportation Data Subcommittee has produced positions and recommendations documentation for LRS to be included in Spatial Data Transfer Standard (SDTS) efforts. The FGDC approach defines generic LRS terms and concepts, and expands the LRS scope to include multimodal transportation applications and ITS applications (FGDC 1994).

The Bureau of Transportation Statistics (BTS) from USDOT released in 1998 a ‘LRS in GIS’ CD-ROM to assist state and local transportation agencies and professionals with the implementation of linear referencing systems in GIS (FHWA 1998). The CD-ROM contains about 100 scanned documents relevant to LRS and GIS. It also includes a glossary of terms and a resource guide to deal with LRS implementation issues.

In addition, national transportation-related professional organizations including the American Association of State Highway and Transportation Organization (AASHTO) and the Transportation Research Board (TRB) have sponsored and coordinated research efforts on GIS-LRS. Along with other organizations, AASHTO and TRB are primary sponsors of the annual national GIS-T conference, which facilitates sharing of GIS-transportation knowledge, offers training, and serves as a forum for continuing research. TRB’s National Cooperative Highway Research Program Project 20-27 was initiated a few years ago to promote adoption of standards for GIS-T. This project has a wider scope than GIS-LRS only (Geo Decisions 1997).

3.4.4 3-D Spatial Modeling in GIS-T

Transportation applications possess a rich source of spatial data. This section introduces the recognition of the needs for 3-D spatial modeling in transportation and research efforts focusing on 3-D spatial modeling in GIS-T.

In GIS-T people are mostly concerned with the regularly shaped man-made objects, i.e. roads. With traditional 2-D GISs, roads are simplified into linear objects using the road centerline as the representative. In applications that have to take elevation into consideration, the integration of elevation data, which are usually in another data format such as contour lines, DEMs, or TINs, with the road data, is the most commonly taken approach.

An LRS data model still treats roads as linear objects, but in one dimension. Any event along a road segment is located by referencing the start point and the distance along the road. Researchers have recognized that in order to construct a high quality LRS database, the actual distance, or three-dimensional surface length should be used instead of the two-dimensional planimetric length, especially in mountainous areas with great changes in road slopes (Vonderohe et al. 1993, FHWA 1998, Rasdorf and Cai 2001, Rasdorf et al. 2003a, Rasdorf et al. 2003b). This indicates that a 3-D approach should be taken to model road centerlines.

In addition to more realistic representations and 3-D distance derivation, 3-D modeling benefits transportation applications by providing curvature, slope, and other road properties in a three-dimensional space. While applications such as emergency routing and flood response routing could greatly benefit from modeling road centerlines in a 3-D space, the research efforts focusing on modeling linear transportation objects in a 3-D space are almost nonexistent. Very little literature is available in this particular field.

However, this negligence is not total. Most 2-D GISs have built-in capability of deriving 3-D distance for linear objects using elevation datasets. Quiroga (Quiroga 1999) noticed this and evaluated the different accuracy levels that different commercial GISs can achieve in 3-D distance derivation. Rasdorf (Rasdorf and Cai 2001, Rasdorf et al. 2003a, Rasdorf et al. 2003b) recognized the benefits accompanying the 3-D approach to

determine road centerline distances while working on an LRS project for NCDOT. With the support from NCDOT, Rasdorf and Cai conducted a series of studies to address issues including how this 3-D approach could be implemented to determine road centerline 3-D distances, the accuracy level this approach could achieve, and how the errors in the source data affect the derived results. However, all these research projects put data quality as the first priority, leaving 3-D spatial modeling in GIS-T as the secondary issue. It leads to the belief that 3-D spatial modeling in GIS-T is becoming a new challenging research topic that requires tremendous research efforts.

4 DATA QUALITY

This chapter defines the concepts of data quality and spatial data accuracy. Previous researches in data quality parameter, accuracy evaluation, error sources and error modeling, and error propagations in the spatial analysis are summarized and discussed in this chapter. The federal efforts in developing data quality standards are included in this chapter. In addition, the specialty of data quality issues in GIS-T is presented and discussed.

4.1 DEFINITION OF DATA QUALITY AND SPATIAL DATA ACCURACY

This section defines the concepts of data quality, spatial data accuracy, and the parameters that could be employed to describe the data quality and the spatial data accuracy.

4.1.1 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-makings. It is clearly recognized that the validity of analysis is based on the validity of the data used to perform the analysis (Rasdorf and Cai 2001, Rasdorf et al. 2003a, Rasdorf et al. 2003b). 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 lead to the necessity for data quality control (Rasdorf and Cai 2001, Rasdorf et al. 2003a, Rasdorf et al. 2003b).

Data quality is a comprehensive measurement to assess the data suitability for use. 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 (SDTS 1997).

Positional accuracy is the positional closeness of an object as described in a spatial database to its real world position. It is usually derived by comparing the positional data to an independent and more accurate source (Rasdorf 2000). Attribute accuracy is defined as the accuracy of the attributes of a spatial object. 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 the target. Currently, it is most often the responsibility of the users to decide if a dataset is sufficient to meet their or their organization's quality requirements, and the standard may differ by application or use (Rasdorf 2000).

4.1.2 Errors, Accuracy, and Precision

Errors and uncertainties are intrinsic in spatial data and need to be addressed properly (Burrough and McDonnell 1998). The terms of error, accuracy, and precision are three related concepts and are sometimes used interchangeably. However, these three terms all have varying degrees of meanings, each with its own significance.

Error is the difference between the true and estimated values. Accuracy may be defined as the extent to which an estimated or observed value approaches the true value (Burrough and McDonnell 1998). Precision is the

dispersion of data from the actual data. It is estimated in terms of standard deviation of the observations over the mean. It also refers to the ability to display numbers to a certain number of decimal digits (Rasdorf 2000).

Data accuracy is a broad term that can be divided into positional accuracy (whether objects are in their correct positions), geometric accuracy (e.g., size or length of linear objects), interrelationship accuracy (topological) that describes the relationships between objects, attribute accuracy, and temporal accuracy

Positional accuracy can be expressed as two components – absolute positional accuracy and relative positional accuracy (Chong 1997). Absolute positional accuracy concentrates on how closely all positions on a map or data layer match to corresponding positions of features represented on the ground in a desired map projection system (Chong 1997). Relative positional accuracy considers how closely all the positions on a map or data layer represent their corresponding geometrical relationships on the ground (Chong 1997).

Geometric accuracy refers to the accuracy of geometric properties of spatial objects. Examples include length, direction, and orientation of linear objects, and area and perimeter of area objects. Interrelationship accuracy (topological) refers to the accuracy of topology, which describes the spatial relationship between spatial objects including connectivity, adjacency, and proximity. Attribute or thematic accuracy is defined as the accuracy of the attributes of its data (Burrough and McDonnell 1998). Temporal accuracy is defined as that part of the data's error that arises due to the temporary nature of the data. It is affected by the interaction between the duration of the recording interval and the rate of change in the event (Veregin 1999).

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 (Rasdorf and Cai 2001). 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 (Rasdorf et al. 2003a, Rasdorf et al. 2003b). 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.

4.1.3 Data Quality Parameters

The five parameters identified in SDTS are just a beginning point of evaluating data quality. Veregin proposed a different set of parameters other than the five components identified in SDTS to evaluate the data quality and thus, the fitness for use for a specific dataset in 1999 (Veregin 1999). These parameters are summarized in Figure 4.1, which consists of four components (accuracy, precision/resolution, consistency, and completeness) viewed from three aspects, spatial, temporal, and thematic.

This is a more comprehensive list of data quality parameters compared with the five parameters identified in SDTS. It is distinct in introducing the differences among spatial, temporal, and thematic quality parameters and providing systematic descriptions for each parameter. However, it suffers from two major drawbacks. First, it ignores the lineage, which usually provides important information about the origination and propagation of errors. Second, there are no well-developed methods or technologies for most parameters. Veregin recognized the difficulties in determining spatial accuracies for linear and polygon features though most data quality

researches concentrate on spatial accuracy issues. Furthermore, tests for some of the parameters such as thematic consistency are almost never carried out (Veregin 1999).

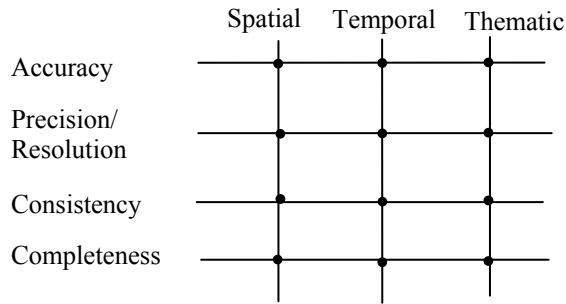


Figure 4.1 Data Quality Parameter Matrix

Data quality may be viewed as a dual-state phenomenon (Chrisman 1983). The first stage identifies quality characteristics by applying tests to data (Wu and Buttenfield 1994). In this stage data quality evaluation is concerned with extent and magnitude of errors, degree of accuracy, and other quality components that can be measured or quantified (Hu et al. 1990). The second stage, or the operation stage, seeks to understand data quality by reviewing the lineage associated with the data (Hu et al. 1990). Figure 4.2 summarizes the quality parameters proposed for two-stage accuracy evaluation by Wu and Buttenfield in 1994 (Wu and Buttenfield 1994).

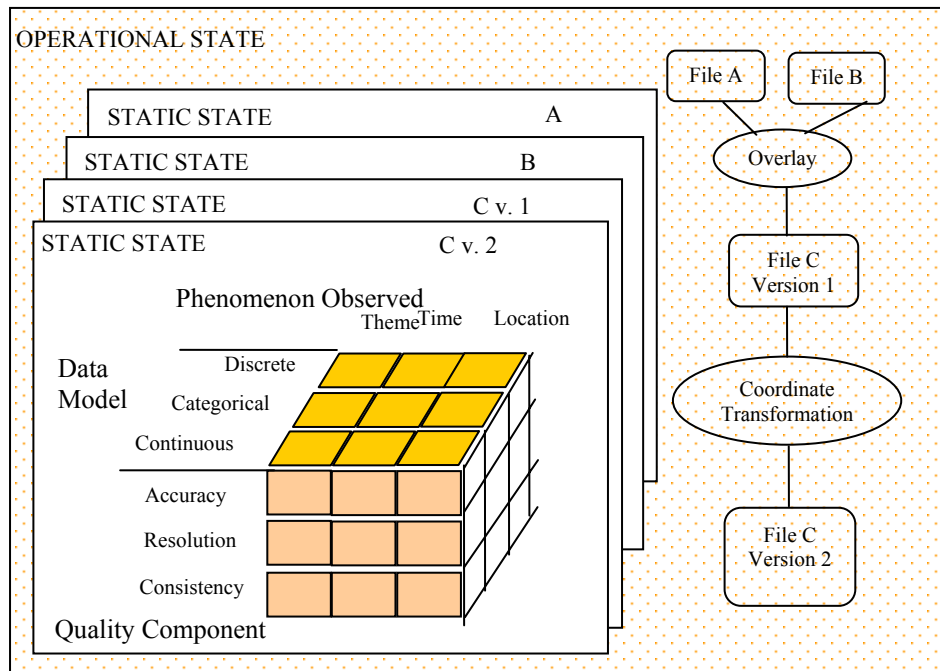


Figure 4.2 Static and Operational States of Data Quality Evaluation (Modeled after Wu and Buttenfield 1994)

This set of quality parameters describes data quality in two phases. In the static phase, a cubic matrix is presented to view data quality by taking into consideration quality components, data models, and phenomena observed. The lineage information is captured in the operational state, which is not testable as defined in SDTS (SDTS 1992). However, it shares one common drawback as the parameter list proposed by Veregin (1999), i.e., most parameters are not well-understood and are very difficult to be obtained or quantified.

To summarize, these data quality parameters all aim at providing a complete or at least a comprehensive list of quality parameters that might be useful in determining the “fitness for use”. However, the common problem is the acquisition and quantification for most parameters as pointed out earlier. In addition, “the fitness for use” also depends on the quality needs of different applications. It is unnecessary and impossible to evaluate all parameters for a specific spatial application. These comprehensive lists of data quality parameters should be used as references from which an appropriate subset could be extracted for quality evaluation for each individual application based on quality needs of that application.

4.1.4 Data Quality Standards

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). The SDTS is organized into the *base specifications* and *multiple profiles*. The *base specifications* currently cover logical specifications, spatial features, and using the general-purpose International Organization for Standardization (ISO) 8211 data exchange standard in an SDTS transfer (Miller and Shaw 2001). An *SDTS profile* is a well-defined subset of SDTS created for translating a specific type or model of spatial data. For example, the *Topological Vector Profile* (TVP) applies to vector geographic data in two-dimensional planimetric space, which is most commonly used in vector GIS software. Additional profiles are being drafted, reviewed, modified, and are expected to be approved in the near future. Among these, the *Transportation Network Profile* (TNP) is of the most interest to

the GIS-T user community. TNP is designed as vector data with network topology (Miller and Shaw 2001). In other words, the transportation spatial entities are modeled as point, line, and polygon features with lines connected to each other (topology).

Another example is the Geospatial Positioning Accuracy Standard that is purposed at providing consistency in reporting the accuracy of point geospatial data collected by different activities (FGDC-STD-007.3-1998). It provides sampling strategies and specifies positional accuracy in horizontal and vertical components. The accuracy statistic employed in this standard is root mean square error (RMSE), which in turn can be based upon to derive a 95% confidence level.

In order to maintain an agency's investment in its geospatial data and to encourage data sharing, U. S. President Clinton signed Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure (NSDI)," on April 11, 1994. This executive order instructs U. S. federal agencies to use a standard to document new geospatial data beginning in 1995, and to provide metadata to the public through the National Geospatial Data Clearinghouse (Miller and Shaw 2001). On June 8, 1994, FGDC approved the "Content Standards for Digital Geospatial Metadata," originally initiated in June 1992. Metadata describe the history, content, quality, condition, and other characteristics of data. The purpose is to provide a common set of terminology and definitions for metadata documentation. The following are the components in metadata defined in the *Metadata Content Standards* (MCS):

- Identification Information – describes the basic information about the dataset, such as the title, the geographic area covered, and rules for acquiring or using the data.
- Data Quality Information – provides an assessment of the quality of the dataset, such as positional and attribute accuracy, data completeness, data consistency, data sources, and methods used to produce the data. This data quality information follows the specifications in SDTS.
- Spatial Data Organization Information – describes the mechanism used to represent spatial information in the dataset, such as the vector or raster representation, street addresses or administration unit codes, and the number of spatial objects in the dataset.

- Spatial Reference Information – describes the reference frame and encoding systems, horizontal and vertical datums, and the coordinate system resolution.
- Entity and Attribute Information – provides information about the content of the dataset, such as the names and definitions of features, attributes, and attribute values.
- Distribution Information – provides information about data distribution, such as contact for obtaining the dataset, distribution media, and cost of acquiring the dataset.
- Metadata Reference Information – provides information on the currency of the metadata information and the responsibility party.

However, MCS does not specify the methods or approaches for obtaining the data quality information although a list of definitions is provided. It does not specify how the information should be organized in computer system and the means by which the information is transmitted to the user, either (Miller and Shaw 2001).

In addition to the above-mentioned three example standards, FGDC is working on new spatial data standards and reviewing and modifying standard drafts. Updated information for spatial data standards is available from FGDC website (www.fgdc.org).

4.2 MODELING ERRORS

Errors and uncertainties are inherent with spatial data. These errors and uncertainties contribute to the inaccuracy in spatial data and thus, must be understood and controlled in spatial analysis. Spatial data accuracy, error, and related issues have been previously studied. This section summarizes the previous research findings into three broad categories: error source identification, error analysis models, and error propagations.

4.2.1 Error Sources

Some researchers tried to identify error sources and error ranges. Amrhein and Schut discussed the range of errors that can accompany any datasets, 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). Rasdorf summarized all the error sources that could be encountered in data collection phase (Rasdorf 2000). It is not uncommon to find summarized error source information from GIS books, for example, Burrough and McDonnell identified 7 factors that affect the quality of spatial data (Burrough and McDonnell 1998). The findings from these research efforts are summarized as below.

Errors encountered in primary methods of data collection (data collected directly by aerial and terrestrial surveying and satellite imagery) include:

- Personal errors (biases introduced by the personnel who carry out the task of data collection)
- Instrumental errors (errors in instrument calibration, biased instruments, variations in instruments due to external factors)
- Environmental errors (solar illumination, clouds, wind, etc.)

Errors encountered in secondary methods for data collections (data are collected from images, charts, maps, graphs, etc.) include:

- All errors encountered in primary methods of data collection
- Compilation errors
- Errors in drawing
- Errors in map generalization and reproduction
- Errors due to material deformation
- Errors due to feature exaggeration
- Errors introduced due to the use of wrong scale
- Errors in digitizing or scanning
- Errors due to the uncertainties in the definition of a feature

Derived data are data generated by certain operations (functions or models) upon the input data. Errors in the derived data can be attributed to the errors in the input data, the quality of the operations, and the way the input data and the operations interact (Burrough and McDonnell 1998).

Discussions and descriptions of data quality control are also available in most GIS books and publications. For example, Burrough identified and described errors and uncertainties in GIS modeling and pointed out the stages where errors and uncertainties could creep in.

Some researchers studied the propagation of errors in spatial modeling (Heuvelink 1999). Error propagation illustrates that errors in GISs 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 GISs, a definition for formal spatial data standards has been proposed and discussed (Backe 1996).

4.2.2 Error Analysis and Accuracy Evaluation

Given a basic understanding of data quality issues in GISs, one must next evaluate, test, model and control errors and uncertainties. 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).

Since the emergence of GISs, researchers have put many efforts into understanding errors in GISs and the ways to avoid these errors to improve the accuracy and quality of spatial data. As a result of these efforts, some researchers have proposed solutions for modeling spatial data accuracy. Positional accuracy for points are well defined and widely applied. For points, error is usually defined as the discrepancy between the encoded location and the location as defined in the specification (Veregin 1999). The most common measures are horizontal errors (distance measured in x and y simultaneously) and vertical error (distance measured in z) as illustrated in Figure 4.3.

The measures in Figure 4.3 apply to an individual point only. For sets of points two common metrics are mean error, which tends to zero when 'bias' is absent, and root mean square error (RMSE), which is computed as the square root of the mean of the squared errors (Beard and Bittenfield 1999). RMSE is most commonly used in evaluating positional accuracies for point sets. Geospatial Positioning Accuracy Standard (FGDC-STD-007.3-

1998) adopted this approach as the standard approach to measure and report positional accuracy. It is also the standard way for USGS to evaluate and report the accuracy for its products of digital elevation models (DEMs). Recently, with the rapid development of global positioning systems (GPS), this RMSE approach for point positional accuracy is being widely adopted in evaluating and reporting the accuracy for GPS datasets.

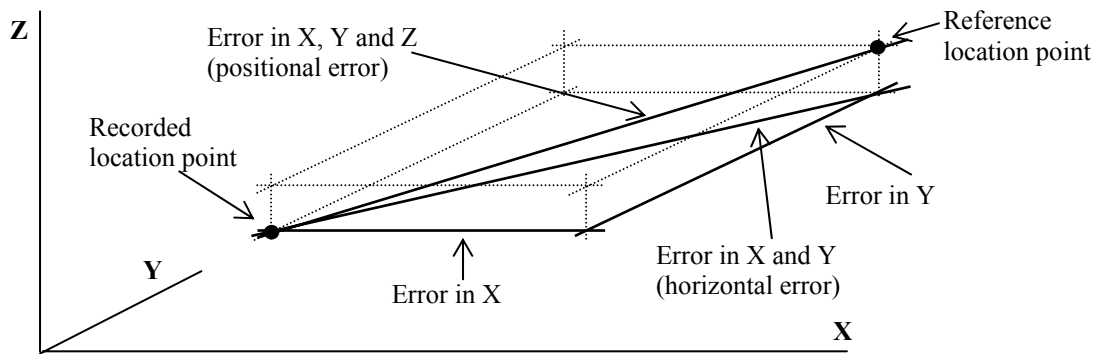


Figure 4.3 Measuring Components of Positional Accuracy for Point (Modeled after Veregin 1999)

This approach is limited to positional accuracies for points. Sampling strategies must be employed if this approach is used to evaluate positional accuracy for lines or polygons. Consequently, it can only provide limited accuracy information and has limited usage. In addition, the mean error matrix tends to be zero while positive and negative discrepancies cancel out the errors in random distributions and thus, is very limited in capturing the error magnificence. On the other hand, RMSE is a measure of the magnitude or error but it does not incorporate bias since the squaring eliminates the direction of the error (Veregin 1999).

Widely accepted approaches for positional accuracies of lines and areas still have to be developed (Veregin 1999). It was recognized that errors in lines arise from the errors in the points that define those lines (Veregin 1999). However, these points are not randomly selected and consequently, the errors present at points cannot be regarded as somehow typical of errors present in the line (Goodchild 1991b). It was also recognized that the approach of using point errors to approximate line errors suffers from two additional drawbacks. First, there is no simple statistical measure or error that can be adopted from statistics (Veregin 1999). Second, matching points encoded with reference points is problematic.

A natural extension to the approach mentioned above for point positional accuracy evaluation, which might be adopted for evaluating positional accuracy for lines is to define errors in position in terms of a bell-shaped probability surface (Goodchild 1991a). Figure 4.4 illustrates this extension.

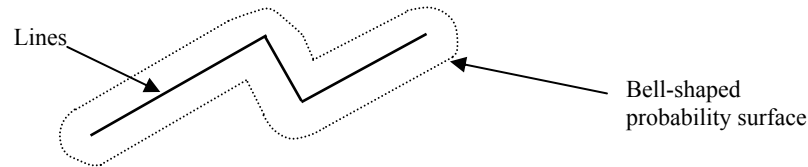


Figure 4.4 The Early Model of Positional Accuracy for Lines (Modeled after Goodchild 1991a)

The bell-shaped probability surface approach for modeling positional errors in lines takes a complete line into consideration rather than using the errors with a series of points and therefore, has surrogated the difficulty in modeling errors in lines. It defines the error using some variant of the epsilon band illustrated in Figure 4.4. Goodchild and Hunter had also developed an implementation method based upon the bell-shaped model around 1997 (Goodchild and Hunter 1997, Hunter 1999). This method estimates the percentage of the total length of the low accuracy representation of line that lies within a specified distance of the higher accuracy representation as illustrated in Figure 4.5.

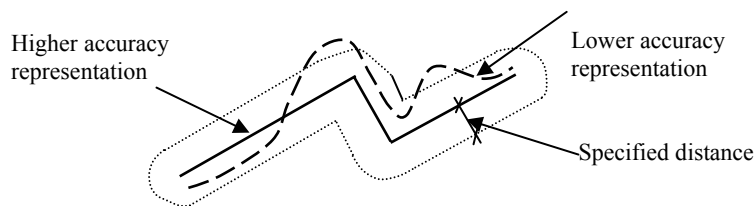


Figure 4.5 A Practical Error Evaluation Method for Lines

This approach can be easily implemented in most GIS software. It provides accuracy in the combination of two parameters, the specified distance and the corresponding percentage of the length that lies within this distance. However, there is no agreement as to the shape of the band and the distribution of errors within it (Veregin 1999). Early models assumed that the band had a uniform width and the errors within it had a normal distribution (Chrisman 1982). Recent studies revealed that both the distribution and the band itself might be non-uniform in shape as illustrated in Figure 4.6 (Caspary and Scheuring 1993).

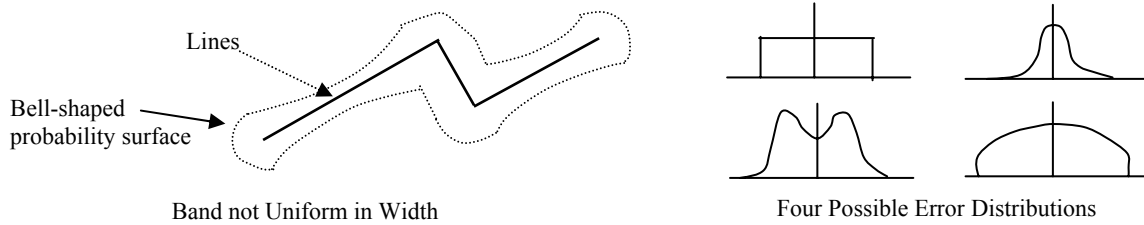


Figure 4.6 Variations in Band Shape and Error Distribution (Modeled after Veregin 1999)

The band concept for positional errors with lines can also be applied to determine the positional accuracy for polygons based on the concept that polygons are delineated from the rest via one self-closed line (the start and end of a line is the same point) or a set of connected lines (Goodchild et al. 1992). In other words, the positional accuracy of polygon edges is the indicator of the positional accuracy of polygon. Another natural extension is to apply buffers to the higher accuracy representation of a polygon and determine the percentage of the polygon area represented by a lower accuracy. The first approach suffers from the same limitations as those associated with line positional accuracy. The second approach also suffers from several limitations. For example, Figure 4.7 illustrates a situation that 100% of the polygon area as represent with a lower accuracy lies in the buffer of the polygon with a higher accuracy. However, it is obvious that the accuracy is no good at all. Consequently, there is no such model that evaluates positional accuracy for polygon based upon the band concept mentioned earlier. A logical surrogate would be to apply buffers and determine percentages of areas that are inside the buffers in two steps. First, a buffer is applied to the higher accuracy representation and determines the percentage area of the lower accuracy representation, which lies within that buffer. Second, the buffer is applied to the lower accuracy representation and determines the percentage area of the higher accuracy representation, which lies within this buffer. These two percentages together provide a fair evaluation of the positional accuracy for the polygon.

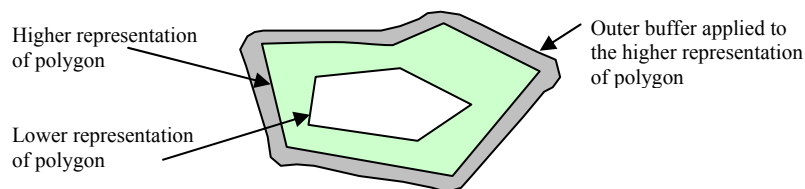


Figure 4.7 Illustration of the Dilemma in Determining Positional Accuracy for Polygons

A model was developed by Hunter and Goodchild around 1995 (Hunter and Goodchild 1995b, Hunter and Goodchild 1996, Hunter 1999), which corrects the positional errors for polygons based on the error statistics obtained from points in a grid-cell based approach. The model developed involves the creation of two independent, normally distributed, random error grids in the X and Y directions (Hunter 1999). These two grids are created based upon the error statistics obtained for points sampled from a vector data layer (it could be point, line, or polygon data layer). They are combined together to provide the two components of a set of simulated horizontal positional error vectors regularly distributed throughout the region of dataset (Hunter 1999). These two grids could be overlaid with the data to be corrected and the x and y shifts can be applied to the coordinates of each node and vertex in the dataset to create a new, but equally likely, version of it (Hunter 1999). This procedure could be repeated several times to assess the variability residing in the end product (Hunter and Goodchild 1996). This model provides a practical approach to correct vector data (it applies not only to polygons, but also to points and lines) and handle vector area uncertainties in GISs. However, this model still does not provide quantitative information about the positional accuracy for polygons.

The most widely applied accuracy assessment approach for thematic accuracy in land cover/use classification derived from remotely sensed data is error matrix/misclassification matrix. Misclassification matrix (Gendren van and Lock 1977, Mead and Szajgin 1982) provides empirical estimates of the probabilities of observing class *i* at a point where in reality class *j* is present for all *i* and *j*. This method involves two datasets, the classification dataset and the reference dataset (either obtained from field observation or derived from a higher accuracy dataset). The classification accuracy (thematic accuracy) is estimated via the comparisons of these two datasets for certain number of points. An error matrix can be constructed with the classification data as one dimension (row) and the reference data as the second dimension (column). Cell with row *i* and column *j* is filled with the number of points that are in class *j* in the reference data and class *i* in the classified data. Statistical measures such as user accuracy, producer accuracy, omission and commission errors, the overall accuracy, and Kappa statistic can be derived from the error matrix. Table 4.1 represents a sample error matrix and the calculation of user accuracy and producer accuracy.

Table 4.1 A Sample Error Matrix

		Reference Data				Total	User Accuracy
		Developed Area	Forest Area	Agricultural Area	Water		
Classified Data	Developed Area	34	4	1	1	40	34/40=85%
	Forest Area	2	32	2	1	37	86.5%
	Agricultural Area	4	4	25	1	34	73.5%
	Water	5	3	2	30	40	75%
Total		45	43	30	33	151	
Producer Accuracy		34/45=75.6%	74.4%	83.3%	90.9%		

The overall accuracy is calculated as $(34+32+25+30)/151=80.1\%$. The producer accuracy is defined as the probability that an area is actually class A in the field is classified into class A. For example, the producer accuracy for the Developed Area is 75.6%, which is the probability of a developed area being classified into the Developed Area class. The user accuracy is defined as the probability for an area that is classified into class A is actually in class A in the field. For example, the user accuracy for the Developed Area is 85%, which indicates there is an 85% probability that an area classified as Developed Area is actually Developed Area in the field. Similarly, omission error is the probability of failure to classify class A area into class A. Commission error is the probability of misclassifying an area that does not belong to class A into class A. In addition, the Kappa statistic is a discrete multivariate statistic that measures the agreement by excluding the chance agreement. It is a measure of the comparison between the classification and a completely random classification.

The error matrix is widely applied in evaluating and reporting thematic accuracy, especially in the land use/land cover classifications using remotely sensed data. It distinguishes producer accuracy from user accuracy in a meaningful way. However, it suffers from its implicit assumption that these probabilities are constant over the mapped area (Goodchild et al. 1992).

Goodchild, together with others, proposed an error model for categorical data around 1992 (Goodchild et al. 1992). This model can be applied to grid-based spatial database of categorical data. This model captures the probabilities for a cell to be in several classes. For example, given a cell that is with categorical data of developed area, it might include 80% of its area being developed area while the rest 20% being forest area. This model defines a vector of given probabilities in each cell, which determines the proportion of realizations in which the cell is assigned to each given class. It also involves a correlation parameter to determine the level

of dependence between outcomes in neighboring cells in any one realization. By capturing these variations this model is capable of evaluating the accuracy for categorical data (grid-based classification) via the area measurement (Goodchild et al. 1992). It is claimed that the probability vector and spatial autocorrelation parameter together characterize the model to provide an adequate description of generic problems of within polygon heterogeneity and transition across polygon boundaries. However, this model suffers from several drawbacks. This model is too complicated to be understood and widely applied. Second, the probability vector is difficult and time-consuming to acquire. Third, the autocorrelation parameter requires a lot of trials and errors.

Beard and Buttenfield summarized the graphical methods used to detect and evaluate errors in GISs around 1999. 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). Examples include the tools developed for graphical integration with GISs by Fisher (Fisher 1994a, 1994b), Goodchild et al. (Goodchild et al. 1994), Hunter and Goodchild (Hunter and Goodchild 1995a), MacEachren et al. (MacEachren et al. 1993), Mitasova et al. (Mitasova et al. 1993), and Paradis and Beard (Paradis and Beard 1994). MacEachren and others (MacEachren et al. 1993) developed a reliability visualization tool (RVIS) which supports several options for viewing data and metadata (reliability), which is estimated based on Kriging residuals and cross validation. This display option include side-by-side overlay and merged displays. Fisher developed a technique referred to as error animation to view the reliability of classified imagery (Fisher 1994a) and soil maps (Fisher 1994b). The uncertainty inherent in the assignment of a pixel to a class is conveyed by making the value or color of a pixel proportionate to the strength of it belonging to a particular class. The data quality filter (Paradise and Beard 1994) allows a user to specify a data quality parameter, a quality measure (RMSE, for example), and a threshold value. Only data meeting this threshold are displayed and consequently the users are informed of how many data do not meet the specific threshold according to the quality parameter and the quality measure specified. The visualization tools developed by Mitasova and others (Mitasova et al. 1995) incorporate multidimensional interpolation, visualization of the resulting model, and predictive accuracy of model results using cross-validation. The visualization tools allow cross-validation error to be view separately from the data and in combination with the data.

The graphical methods in error detection and modeling capture the quality information in a perceptive way (visual) and provide a fast communication channel (Beard and Buttenfield 1999). However, they suffer from several limitations. Due to the subjective human interpretations and judgments in applying these methods to examine data quality, some researchers even suggested that graphical approach was largely unscientific (Cox 1978). There are full of examples of techniques and their possible misinterpretations in the cartographic literature (Beard and Buttenfield 1999).

In addition to the above-mentioned error analysis models, comprehensive discussions of error modeling and management for GIS data were available in literature (Chapman et al. 1997, Fisher 1999). As one of the few critical authors in error modeling and management, Goodchild, together with others, published a series of publications concentrating on this field (Goodchild 1991a, Goodchild 1991b, Goodchild et al. 1992, Goodchild 1994, Goodchild 1995, Hunter and Goodchild 1995a, Hunter and Goodchild 1995b, Goodchild 1996a, Goodchild 1996b, Hunter and Goodchild 1996, Goodchild and Hunter 1997, Goodchild and Jeansoulin 1998).

The comprehensive literature review of data quality, error analysis, and accuracy assessment at the static stage leads to the following observations:

- The importance of data quality was recognized in the early years of GIS development. Data quality and its related issues have attracted tremendous research efforts.
- There is no agreement on the quality parameters that should be used to evaluate the fitness-for-use for a specific dataset, although SDTS identifies five components.
- Despite the tremendous research efforts on positional accuracy, there are no well-developed error models that are widely applied on lines and polygons.
- There are no well-developed approaches or methods to evaluate most data quality parameters other than positional accuracy and the thematic accuracy in classifications. Some data quality parameters are even announced to be impossible to be quantified.
- Geometric accuracy is treated in such a way that instead of evaluating the geometric accuracy directly, the accuracy for each individual geometric property (length and slope for lines, perimeter and area for

polygons) is evaluated and reported as geometric accuracy indicator. Each geometric property is evaluated in the same way as nonspatial attributes are evaluated.

- While the geometric accuracy is acquired by evaluating the accuracy for geometric attributes of spatial objects, the approaches and methods for evaluating attribute accuracy are almost never covered in GIS data quality researches. It is assumed that the attributes could always be evaluated easily and systematically using traditional statistical analyses. However, this is not always the case. There is a need to develop systematic yet simple enough approaches that could be employed to evaluate and report attribute accuracy.

4.2.3 Error Propagations

The previous section focuses on the static stage of data quality. This section focuses on the operational stage of data quality, or error propagation.

The derivation of new attributes/information from attributes/information already held in the GIS database is one of the most powerful capabilities of GISs (Heuvelink 1999). For example, elevation data in the form of DEMs can be used to derive maps of gradient and aspect. This kind of basic functions used for derivations are often provided as standard functions or operations in many GISs, under the name of ‘map algebra’ (Burrough and McDonnell 1998). In practice, many GIS operations are used in sequence to compute an attribute that is the result of a (computational) model (Heuvelink 1999). This concept is illustrated in Figure 4.8.

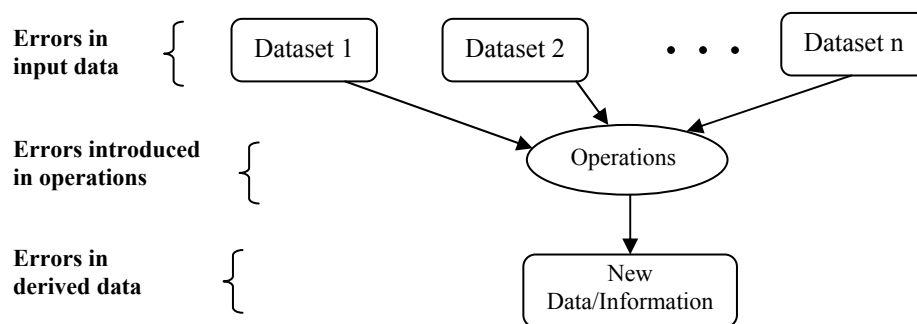


Figure 4.8 The Illustration of New Data Derivation

Burrough and McDonnell noticed that the quality of the results of quantitative models depends on three factors, the quality of the input data, the quality of the model/operation, and the way data and model interact (Burrough and McDonnell 1998). During GIS operations, errors in the input data accumulate with errors introduced in operations and affect the quality of the new derived data. This is defined as error propagation.

Error propagations have been researched for many years. As a summary, Burrough and McDonnell allocated one chapter in their book for error propagation issues, named as “Error Propagation in Numerical Modelling” (Burrough and McDonnell 1998). In this chapter, Burrough and McDonnell identified three components needed for error propagation analysis, i.e. sources of error estimates, error propagation theory, and error propagation tools. Sources of error estimates are obtained during the static stage of error and quality analysis. Two error propagation theories are presented in that chapter, Monte Carlo method and analytical approach. Monte Carlo method treats each attribute in the input data as having a Gaussian (normal) probability distribution function (PDF) with known mean and variance for each entity or cell. The arithmetical operation to derive new data is then carried out not with the single mean values but using a value that is drawn from the PDFs for each entity or cell. To take into consideration of the variations within the PDFs, the calculations are repeated many times in order to compute the mean result per entity or pixel and its standard deviation (Burrough and McDonnell 1998).

Monte Carlo methods of error analysis are straightforward and can be adapted to many kinds of numerical modeling (Burrough and McDonnell 1998). However, they are computing intensive and require considerable computing resources. Burrough and McDonnell also proposed analytical approaches that are based on the standard statistical theory of error propagation. They are computationally efficient and faster (Burrough and McDonnell 1998). Simple examples for applications of these analytical approaches are also provided to illustrate how these approaches could be used to model and evaluate error propagations in GIS spatial analysis.

Burrough and McDonnell also summarized the information about an error propagation tool named ADAM, which is a computer program written by Heuvelink and Wesseling (Heuvelink 1993). This tool can trace errors

through complex numerical models in 'point mode' that operate on the attributes of entities or on multiple raster overlays (Heuvelink 1993).

Heuvelink also summarized the researches and methods in error propagations (Heuvelink 1999). Heuvelink defined a stochastic error model for quantitative spatial attributes and compared two methods that are based on the stochastic error model, Monte Carlo method and Taylor series method. The main problem with the Taylor method is that the results are only approximate while it will not be easy to determine whether the approximations involved using this method is acceptable (Heuvelink 1999). While Monte Carlo method does not suffer from this problem, Heuvelink (1999) also pointed out that it is very time and resource consuming. It was suggested that the Taylor method may be used to obtain crude preliminary answers and the Monte Carlo method may be used for exact values or quantiles and/or percentiles (Heuvelink 1999).

The literature review regarding error propagations revealed that there are not many methods available to model error propagations due to the complexity and the number of variations in GIS operations and their combinations. Error propagations depend on the input data quality and the characteristics of the operations. There is no such thing that allows GIS users to input the input data quality information and specify the operations that will be carried out, and determines how errors would propagate into the end results. Error propagation is distinct at different operations or operation combinations and is heavily dependent on the error characteristics of the input data.

While the above mentioned methods in modeling error propagation are very theoretical, an empirical approach would be to reverse the procedure. Instead of modeling how the errors will propagate through GIS operations to affect the quality of end result, this approach evaluates the quality of the end product, together with the quality information of the input data, they will reveal important information about error propagations during this specific GIS spatial modeling and therefore, can be used to predict error propagations in similar GIS operations and to predict the quality of the end results.

4.3 DATA QUALITY ISSUES IN GIS-T

The above mentioned data quality issues all apply to GIS-T. In addition, there exist some data errors that are more specific to GIS-T applications (Miller and Shaw 2001). Miller and Shaw summarized the spatial modeling in transportation and provided an example of errors that are specific to GIS-T.

As mentioned earlier, transportation agencies deal with a wide range of attributes referenced along the linear transportation routes. These attributes are modeled as either point or linear events along the linear routes. Dynamic segmentation is a method designed for the handling of spatial data based on a linear referencing system (Miller and Shaw 2001). The basic design of a dynamic segmentation model is that each route is defined individually (a state highway, for example), and a collection of routes forming a route system (state highway system, for example). Locations of different attributes associated with a particular route are referenced by the linear measures along that route and are independent of other routes in the route system. This design offers a flexible way of defining independent routes in a route system, but it also causes potential data quality problem where two routes intersect with each other or share a common route segment (Miller and Shaw 2001).

To appreciate the specialty of the data quality issues in GIS-T, the example from Miller and Shaw (2001) is provided in Figure 4.9. Figure 4.9 shows a sample route system with three routes. Route 1 and Route 2 cross each other at Intersection B. Route 1 and Route 3 share a common segment between Intersection A and Intersection C. If Intersection B is referenced in both Route 1 and Route 2 in terms of its linear mileage along the two routes, it is very likely that Intersection B will show up as two separate, although often closely spaced, points on the display. This situation is due to the independent linear interpolations performed on the two routes and causes offsets from the exact intersection location in the digital database.

The error illustrated in Figure 4.9 could be fixed with a tolerance level built into the dynamic segmentation model, which will automatically resolves offsets smaller than a user-specific tolerance level (Miller and Shaw 2001). An alternative could be performed while the tolerance approach is not appropriate. This alternative works by calibrating each individual route through control points and with accurate positions and offsets.

Furthermore, if all lines have high positional accuracy, the error illustrated in Figure 4.9 could be dealt with highly accurate distance measurements.

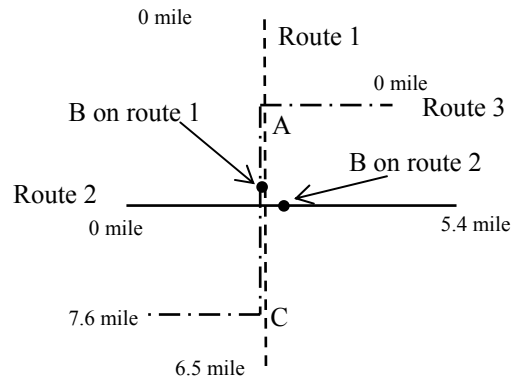


Figure 4.9 Example of Errors in LRS (Modeled after Miller and Shaw 2001)

Literature review regarding to the specialty of data quality issues in GIS-T revealed the extreme importance of positional accuracy for points and lines, which are the most common spatial objects in GIS-T, their geometric accuracy, and their topological accuracy. With the aid from GPSs, transportation engineers can obtain highly accurate and highly precise positional information for points. On the other hand, the topological (connectivity) error can be easily found or verified in field. For example, the overpass, ramp problem, no turn situation, and etc. can all be identified in field easily. The problem is the distance measurement (Rasdorf and Cai 2001, Rasdorf et al. 2003a, Rasdorf et al. 2003b). There exist many technologies or approaches to obtain distance (road surface length) measurement. However, there is no systematic description about what technologies or approaches could be employed in collecting distance measurements for transportation spatial database, how does each of these technologies or approaches work, how accurate is the measurement a specific technology or approach could achieve, what are the errors sources and how to control them in distance measurement, etc. The lack of answers to the above questions leads to the necessity of conducting researches focusing on the quality issues of distance measurement in transportation.

5 MEASUREMENT TECHNOLOGIES

This chapter describes different measurement technologies that could be deployed to measure the actual length of roads. For each technology, a brief discussion regarding its working principles, advantages and disadvantages, and data quality issues are provided.

5.1 INTRODUCTION

Roads are man-made regularly shaped objects. Road segments are 3-D objects that could be defined in a 3-D spatial referencing system. The length or distance of a road segment is an extremely important property for all applications in which road network is a component. Simply enough, a driver needs to know how long he/she has to drive from one place to another. In LRS, distance along road segments is an essential component in locating events along road network, which in turn is identified as the most critical issue in transportation planning (FHWA 1998). It leads to two questions. First, what is the representative line of a road segment? Second, how to determine the actual length or 3-D length of the representative line for given a road segment? Usually the road centerline is used as the representative of a road segment, which was defined earlier in Section 1.5. This answers the first question.

Regarding the second question, there are many technologies and approaches available to obtain the 3-D length for a road segment. The 3-D distance for a specific road segment can be determined by using its geometric design drawings, can be determined via ground surveying (using total station), can be determined by using global positioning systems (GPSs), and can be obtained via distance measurement instrument (DMI). Furthermore, it can be computed by 3-D analysis in a GIS using elevation data and the planimetric geometry of the road segment. The following sections describe the first four approaches/technologies in detail while leaving the last approach, the GIS approach, to be described in the next chapter.

5.2 LENGTHS DERIVED FROM LEGACY DRAWINGS OR MAPS

In highway design, the geometric properties such as cross section, profile, horizontal projection, and vertical alignment are all specified clearly in design drawings. These drawings might be produced manually and plotted

on sheets of paper. With the development of computer-aided design (CAD) and its software packages, presently, most of the design drawings are being produced with computers and stored as electronic files. This development improves the efficiency in drawing modification and reproducing. Even with the CAD technologies, paper drawings are still being produced by printing those electronic files. These indicate that the 3-D distance of road centerlines can be derived from the design drawings.

This approach suffers from several drawbacks described as bellow. These drawbacks make the approach of deriving road centerline from legacy drawings infeasible.

- 1) It is very time consuming. The whole process of retrieving the archived design drawings, reading drawings and obtaining necessary measures, and computing or calculating the distance is very tedious and requires lots of labor hours.
- 2) The design drawings are only kept for a limited time period, especially those paper drawings. Therefore, the availability is a major limitation of this approach.
- 3) The currency and correctness are always questionable. After the initial design and construction, many roads have undergone modifications such as expanding lanes and introducing new intersections with newly constructed roads, which lead to the inconsistency of the existing road geometry with the available design drawings.

A similar, but not the same approach to determine the length of road centerline is to measure and scale from paper maps. However, the major disadvantage is that the distance measured and scaled from paper maps is two-dimensional, or the planimetric length. While in areas with flat terrain this is not a serious issue, in mountainous areas with roads having steep slopes, the accuracy could be far away. In addition, the currency and correctness of paper maps are also questionable. There is always a time lag between the time a road segment is modified and the time this modification is reflected in paper maps. The position and shape of roads on maps might be adjusted or changed (smoothing, for example) in order to produce maps with a high visual quality.

5.3 GROUND SURVEYING

Ground surveying technologies are being widely used in capturing spatial data and helping numerous engineering applications. Among the surveying technologies, two of them are capable of distance measuring, taping and total station. Taping is one of the earliest surveying technologies, which is still being used in measuring relatively short distances in a small geographical scale. This section focuses on the technology of total station, which is more efficient to achieve a high level of accuracy in distance measurements.

5.3.1 Background

Surveying technologies and instruments are constantly changing and progressing. Traditionally, surveying had used analog methods for capturing data. Nowadays, electronic surveying equipment is being widely used. With the emergence of electronic theodolite and electronic distance measurement (EDM), it was common to use an electronic distance meter mounted above a theodolite as shown in Figure 5.1 in the mid 1970s. These early beginnings led to the invention of the magnificent total stations of today, which are used for running traverses, for making topographic maps, for laying out construction site, and for many other types of surveys. Figure 5.2 shows an example of the current total station.



Figure 5.1 Add-On Electronic Distance Measuring Instrument Mounted on a Theodolite (Photo Courtesy of Sokkia Corporation)



Figure 5.2 Current Total Station, Trimble 3600 Series (Photo courtesy of Trimble Navigation Limited)

5.3.2 Fundamentals of Total Station in Distance Measurement

Total stations are systems consisting of electronic distance measurement instrument (EDMI), electronic theodolite, electronic data collectors, and on-board microprocessor accompanied with computer programs. Prisms are the optional equipment that use along with total station in order to receive more accurate results. Only the electronic distance measurement instrument is described in detail while others are covered briefly.

5.3.2.1 Electronic Distance Measurement Instrument (EDMI)

Electronic distance measurement instrument (EDMI) is defined as an instrument that transmits a carrier signal of electromagnetic energy from its position to a receiver located at another position (Kavanagh 1997). The signal is returned from the receiver to the instrument such that two times the distance between the two positions can be measured. EDMIs, first introduced in the 1950s by the Geodimeter Inc. founders, has undergone continual refinements since those days. Current EDMIs use infrared light, laser light, or microwaves. The basic principle to measure the distance is the same by using the phase difference method as illustrated in Figure 5.3.

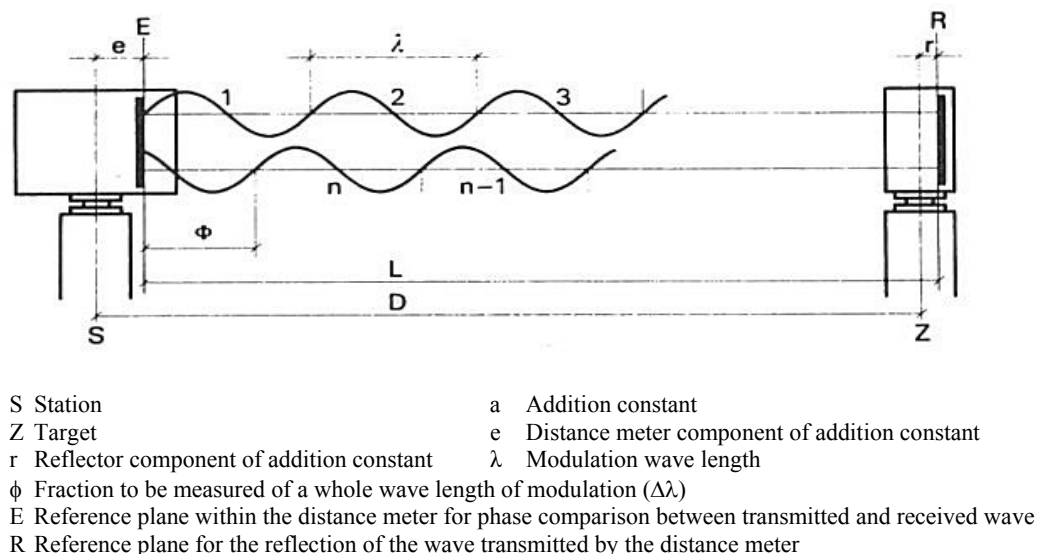


Figure 5.3 Principles of EDM Measurement (Courtesy of Kern Instruments-Leica) (Kavanagh 1997)

Figure 5.3 shows the modulated electromagnetic wave leaving the EDM and being reflected (light waves) or transmitted (microwaves) back to the EDM. It can be seen that the double distance ($2L$) is equal to a whole number of wavelengths ($n\lambda$), plus the partial wavelength (ϕ) occurring at the EDM according to Equation (5.1).

$$L = (n\lambda + \phi)/2 \text{ meters} \quad (5.1)$$

The partial wavelength (ϕ) is determined in the instrument by noting the phase delay required to precisely match up the transmitted and reflected or retransmitted waves. The phase difference method measured only the fraction of wavelength that is out of phase. The value of n is determined by the introduction of the multiple-frequency techniques (Rüeger 1996).

The smaller wavelength provides a more precise assessment of the fractional portion. The first number of each rough reading is added to the initial fine reading to give the total distance (Schofield 1993). For example, a frequency of 15 MHz is set up in the instrument, resulting in a half wavelength of 10 m. Let a full sweep of the phase meter represent this 10-m distance. The phase meter reading then gives the unit meter and decimal part of the meter in the measured distance from 0 to 9.999m. In a distance of 3485.276m, for example, this frequency would give the 5.276 part. Switching down to 1.5 Hz, the half wavelength is now 100m (100-m distance), which is resolved by the phase meter to give the tens of meter, which in this example is 80 (8 tens). The next frequency is then 0.15MHz, which in conjunction with the phase meter to give the hundreds of meters. In this example it is 400 (4 hundreds). Finally, a 15 kHz frequency will give the number of thousand meters in the distance, which in this example is 3000 (3 thousands). Therefore, the total distance is determined to be 3,485.276meters.

The 10-m, 100-m, 1,000-m, and 10,000-m distances in the previous example are called the unit length of an EDM instrument. The most important unit length of an instrument is always the smallest, which coincides with the highest frequency. This so-called “main” unit length is used for the fine measurement of distances (Rüeger 1996). The measurable distance is limited by the number of the unit length. From the previous example, as a result, it can measure distances less than 10,000-meter.

5.3.2.2 Electronic Angle Measurement/Electronic Theodolite

The electronic digital theodolite, first introduced in the late 1960s (Carl Zeiss Inc.), set the stage for modern field data collection and processing (Schofield 1993). This instrument is used in the measurement of angles. The horizontal and vertical circles of the instrument can be compared to circular protractors set in horizontal and vertical planes. Figure 5.4 shows the horizontal plane for angle measurement. According to the figure, observations to points A and C from B will give the horizontal angle $ABC = \theta$. The vertical angle of elevation to A is α and its zenith angle is Z_A . The horizontal angle between BA and BC (θ) is always measured even though points A, B, and C have different elevations.

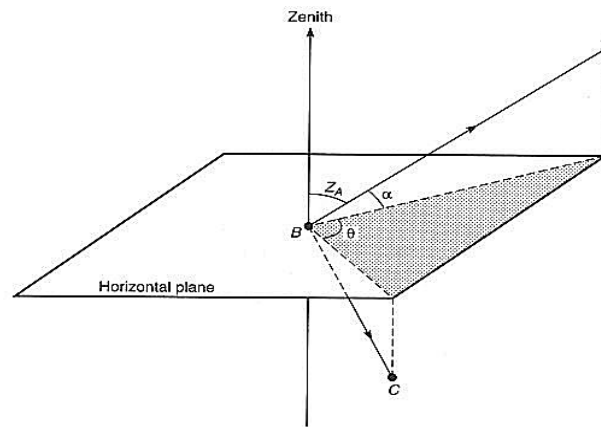


Figure 5.4 The Horizontal Plane for Angle Measurement (Schofield 1993)

The electronic theodolite contains circular encoders that sense the rotations of the spindles and the telescope, and then electronically convert these rotations into horizontal and vertical (or zenith) angles. After that, the values of the angles are displayed on liquid crystal displays (LCDs) or light emitting diode displays (LEDs). These readouts can be recorded in a conventional field book. They are also stored in a data collector for future printout or computations.

5.3.2.3 On-board Microprocessor

The functionality of the on-board microprocessor includes monitoring the instrument status (e.g., level and plumb orientation, batter status, return signal strength, and so on) and performing a variety of mathematical

operations, for example, averaging multiple angle measurements and averaging multiple distance measurements. In addition, the outputs stored in the data collector will be entered into a built-in microprocessor. The microprocessor converts the measured slope distance (S in Figure 5.5) to the horizontal distance (H in Figure 5.5) by using the measured vertical or zenith angle from the electronic theodolite (V or Z in Figure 5.5). It also computes the difference in elevation (DE of Figure 5.5) between the instrument center and the reflector target. If the elevation of the instrument center (the HI) and the height of the reflector target above ground are entered, the microprocessor computes the elevation of the target station.

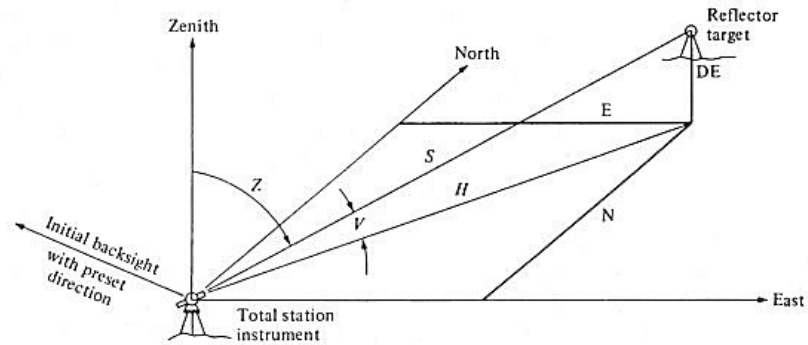


Figure 5.5 Illustration of Distance Calculation in Total Station (Moffitt and Bossler 1998)

5.3.2.4 Data Collector

The electronic data collector comes along with the advent of the electronic theodolite. This device is used to store the measured data. The information in the data collector can be interfaced directly to a computer. These can reduce a lot of tedious paper works. Also, errors that occur in transferring the field information from the field book to the computer are eliminated. Usually the data collector is a hand-held device connected to the on-board microprocessor by a cable. Some manufacturers also have the data collector built into the total station. Figure 5.6 shows an example of the handheld data collector.



Figure 5.6 Data Collector
(Courtesy of Trimble Navigation Limited)

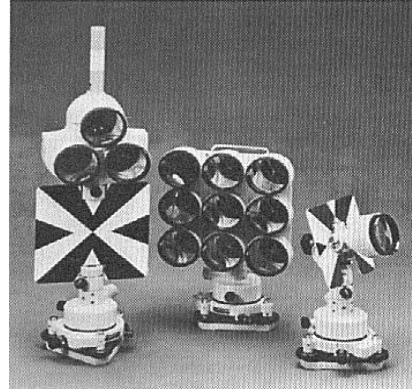


Figure 5.7 Prisms
(Courtesy of Topcon Instrument Corp., Paramus, NJ)

5.3.2.5 Prisms

Prisms are used with EDMs (light, laser, and infrared) to reflect the transmitted signals. A single reflector is a cube corner prism that has the characteristic of reflecting light rays back precisely in the same direction as they are received. This retro-direct capability means that the prism can be misaligned with respect to the EDM and still be effective. A cube corner prism is formed by cutting the corners off a solid glass cube. The quality of the prism is determined by the flatness of the surfaces and the perpendicularity of the 90° surfaces. Figure 5.7 shows various types of target and reflector systems. Prism can be tribrach-mounted on a tripod, centered by optical plummet, or attached to a prism pole held vertical on a point; however, prisms must be tribrach-mounted if a higher level of accuracy is required (Kavanagh 1997).

5.3.3 Data Quality Issues

This section describes the errors associated with total stations and the typical accuracy of total stations.

5.3.3.1 Errors

Errors associated with total stations can be categorized into two categories, instrumental errors and non-instrumental error.

5.3.3.1.1 Instrumental Errors

Instrumental errors are errors due to the inaccuracies introduced from any portion of the system. They are listed as below.

1) Collimation Error

Collimation error refers to the error that occurs in the observed angle due to the line of sight, or more correctly, the line of collimation, not being at 90° to the transit axis as shown in Figure 5.8.

2) Transit Axis Errors

Error will occur in the measurement of the horizontal angle if the transit axis is not 90° to the instrument axis as shown in Figure 5.9.

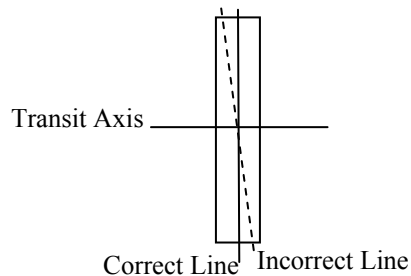


Figure 5.8 Plan View Showing Collimation Error (Worawat and Rasdorf 2003)

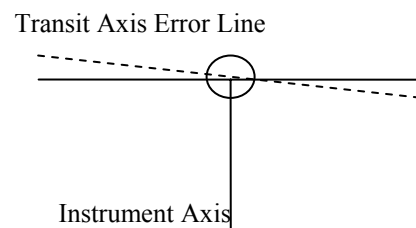


Figure 5.9 Front View Showing Transit Axis Error (Worawat and Rasdorf 2003)

3) Scale Errors

Scale errors can occur when the frequency of the emitted wave from EDM is unstable and the emitted wavelength is not in the normal value. The operational temperature is one of the factors that affect the frequency of the emitted wave. For this reason, the total station should be performed under the recommended operational temperature depending on model specifications. In addition, diode errors can cause scale errors since they are the causes that make the emitted wavelength different from its normal value.

5.3.3.1.2 Non-instrumental Errors

Non-instrumental errors are errors introduced by external factors of the system. They are listed as below.

1) Pointing Errors

This type of errors occurs because of both human inability to point from the instrument to the target and environmental conditions limiting the clear vision of the observed target. The best way to minimize pointing errors is to make the observation several times and use the average as the result.

2) Recording Errors

The two most common errors associated with fieldwork are reading an angle incorrectly and entering incorrect information into the field book (ASCE 2000). Although electronic data collection has already eliminated these errors, it is still possible for a surveyor to identify an object incorrectly such as making a shot to the wrong spot or inputting a wrong target height.

3) Atmospheric Errors

The temperature and air pressure can cause errors. Also, small errors can occur due to the direct sunlight because the sunlight will heat one side of the instrument. The heat is enough to cause an error. This problem could be solved by picking a shaded spot for the instrument.

5.3.3.2 Accuracy

The accuracy required in the measurement is determined by the purpose of the survey. The accuracy of total stations consists of two parts: angular measuring accuracy and linear (distance) measuring accuracy. Both parts of accuracy should be maintained in balance in order to obtain the precisely positioning measurement (Moffitt and Bossler 1998). The position of a point is determined by a horizontal angle from some line of reference and a distance from the vertex of the angle. According to Figure 5.10, a point P can be located by combining the angle NAP and the distance AP . The errors can occur in both the angle measurement and the distance measurement. If the angular error is $\pm e$ and it is supposed that there is no distance error, the position of P would be between P' and P'' . On the other hand, if it is assumed that there is no angular error, the effect of linear error would be to locate P somewhere between P_1 and P_2 . If the accuracy of the angular measurements is consistent with the accuracy of the distance measurements, the values of $P'P$, P_2P , $P''P$, and P_1P should all be approximately equal, and the located position of P will fall somewhere within an approximate square.

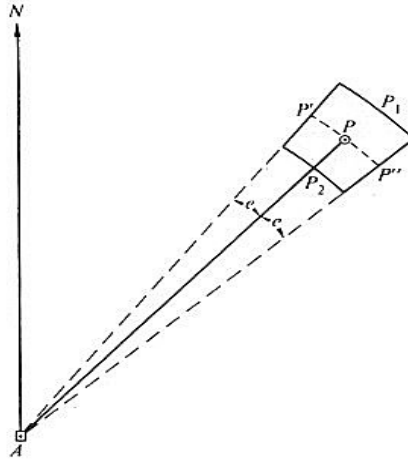


Figure 5.10 Relation between Angular and Linear Error (Moffitt and Bossler 1998)

According to a study based on specifications of market models, most models of total station can achieve the accuracy at the level of $2 \pm 2 \text{ppm} \cdot D$ when working with a reflector and at the level of $3 \pm 2 \text{ppm} \cdot D$ when without reflectors, where D = measuring distance (Worawat and Rasdorf 2003). For example, if a total station with the typical accuracy is used to measure a distance of 1000m, the accuracy would be ± 2.2 mm.

5.3.4 Advantages and Disadvantages

Total station in surveying is a well-established technology. It is capable of measuring distance, horizontal angle, and vertical angle between two locations. It can also achieve a high accuracy in both distance and angular measurements. However, the major limitation of this technology is that it can only measure the horizontal distance without taking into consideration of the slopes, or elevation changes between the station and the target. This makes it impractical to obtain the actual distance (3-D distance) between two locations. One possible approach to surrogate this limitation is to segment the route between two locations into smaller segments. For each of those smaller segments, the horizontal distance and the elevation difference between the start and end nodes can be obtained using the total station. By applying a mathematical function to compute the 3-D distance for each smaller segment and add them together, an approximate 3-D distance between two locations can be obtained, which is illustrated in Figure 5.11. Between locations A and B, three segments are

used to take into consideration of elevation changes. For each segment, its horizontal distance is obtained via Equation (5.2) and its 3-D distance is obtained via Equation (5.3).

$$\text{Horizontal } h_1 = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}, \text{ where } X_1/Y_1 \text{ are the } X/Y \text{ coordinates for one point and } X_2/Y_2 \text{ are the } X/Y \text{ coordinates for the other point} \quad (5.2)$$

$$\text{3-D distance} = \sqrt{h_1^2 + ed_1^2}, \text{ where } h_1 \text{ is the horizontal distance and } ed_1 \text{ is the change in elevations} \quad (5.3)$$

This surrogate is just an approximation. Numerous intermediate nodes and segmentation are required in order to achieve a high level of accuracy. It is very inefficient in obtaining a large amount of 3-D distance measurements.

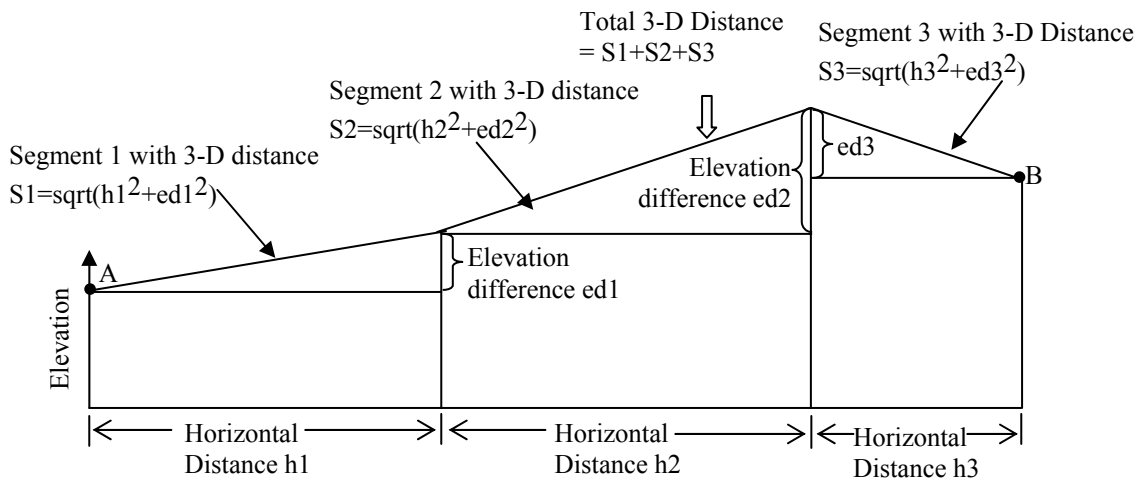


Figure 5.11 Approximation of 3-D Distance between Two Locations via Segmentation

5.4 GLOBAL POSITIONING SYSTEMS (GPSs)

Another technology that could be used to measure distances for linear objects is a global positioning system (GPS). This section describes the background information regarding GPSs, the fundamentals of how GPSs work, data quality issues, and using GPS to obtain distance measurements.

5.4.1 Background

During the 1980s the Department of Defense (DOD) began to launch satellites into space with the goal of being able to quickly and accurately locate positions on the Earth. In the mid-1980s, DOD began to implement a second-generation guidance system - *Navigation Satellite Timing and Ranging (NAVSTAR) Global Positioning System (GPS)*. In December 1993, the U. S. government officially declared that the system had reached its initial operational capability (IOC) with 26 satellites (23 Block II satellites) potentially available for tracking (Sickle 1996). Today there is a network of 27 of these orbiting satellites.

5.4.2 Fundamentals of GPS

GPS technologies have undergone rapid developments. However, the fundamentals are still the same. This section describes the basic principle of how GPSs work their components.

5.4.2.1 Basic Principle

The distance from a position on Earth to a satellite can be determined via satellite ranging (Sickle 1996), which measures time required for radio signals sent from the satellites to reach a specific position on the Earth surface and times that time with the traveling speed of the radio signals, which is 186,000 miles per second. This resulting value is referred to as *pseudorange* where the prefix *pseudo* means “false” because there are some inherent errors in the time measurement.

The basic working principle for GPSs is such that for a specific point position on the Earth surface, first, its distance to a satellite with known spatial position is determined. Second, by obtaining two more such distances to two other satellites with known spatial positions there will be three simultaneous equations that can be solved to obtain the point position in the X, Y, and Z directions. Since time measurements (and thus distance calculations) are not perfect, distance measurements are made to four or more satellites and this enables us to eliminate error in timing. In addition, computers in the receivers are set to keep subtle adjustments in order to make the measurements as optimal as possible.

5.4.2.2 GPS Elements

GPSs consist of three parts: the space segment, the user segment, and the control segment as illustrated in Figure 5.12 (Kennedy 1996).

5.4.2.2.1 Satellites (Space Segment)

GPS satellites are manufactured by Rockwell International, weigh about 1900 pounds, span 17 feet (with solar panels deployed), and orbit the Earth at 10,900 nautical miles (20,000 km) in a period of 12 hours (actually 11 hours, 58 minutes). The satellites' constellation consists of 24 satellites (including 3 spares) placed in six orbital planes (Figure 5.13). This configuration ensures that at least four satellites (the minimum number needed for precise measurements) are always potentially visible anywhere on Earth.

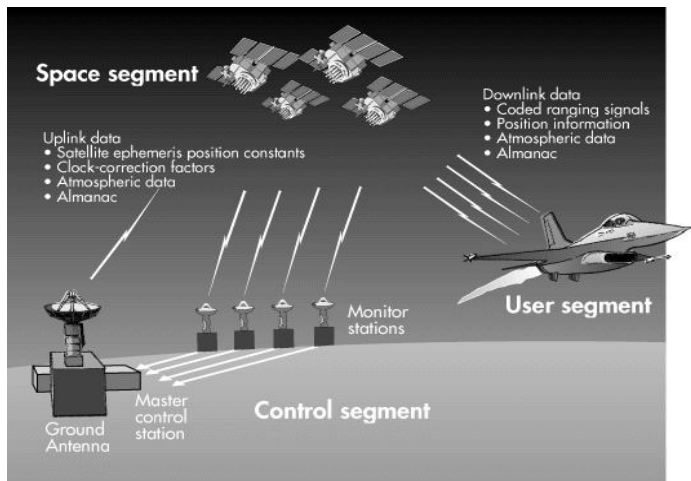


Figure 5.12 Three Major Segments of GPS
(Source: <http://www.aero.org/publications/GPSPRIMER>)



Figure 5.13 Satellite Constellation
(Source: <http://www.garmin.com>)

5.4.2.2.2 Receivers (User Segment)

GPS receivers are capable of generating and receiving signals to make positional measurements. GPS receivers range in ability (and cost) from survey-level receivers capable of use in surveys requiring high accuracy (sub centimeter accuracy) and costing more than \$20,000, to mapping and GIS receivers (sub meter accuracy)

costing about \$3,000 each, to marine navigation receivers (accuracy 100 to 200 m) costing about \$1,000, and finally to orienting (hiking) and low-precision mapping/GIS receivers costing only a few hundred dollars.

The major differences in the receivers are the number of channels available (the number of satellites that can be tracked at one time) and whether or not the receivers can observe both L1 and L2 frequencies ---- code phase and carrier phase. Generally speaking the higher-cost dual-frequency receivers requires much shorter observation times for positioning measurements than the less expensive single-frequency receivers.

5.4.2.2.3 Tracking Stations

In addition to satellites arrayed in space, a GPS includes tracking stations evenly spaced around the Earth. Stations are located at Colorado Springs, Colorado (the master control station), and on the islands of Ascension, Diego Garcia, Kwajalein, and Hawaii as shown in Figure 5.14. All satellites are observed at each station, with all the clock and ephemeris data being transmitted to the control station at Colorado Springs. The system is kept at peak efficiency, as corrective data are transmitted back to the satellites from Colorado Springs (and a few other ground stations) every few hours.



Figure 5.14 GPS Master Control and Monitor Station Network (Source: Peter H. Dana 5/27/1995)

5.4.3 Data Quality Issues

A GPS is not perfect in the performance. There are still some errors in the spatial data acquired, some are system errors, and others are random errors. Typically, these errors are grouped into 5 major groups described as below.

(1) Atmospheric Errors

The speed of light is a constant of 186,000 miles per hour when it passes through an ideal vacuum space. When light passes through denser mediums such as the heavily charged particles in the ionosphere (the outer part of the Earth's atmosphere) and the water vapor in the troposphere (the part of the Earth's atmosphere next to the Earth's surface), it slows down. Furthermore, the atmosphere is constantly changing.

(2) Multipath Errors

When transmitted signals arrive at the surface of the Earth they may be reflected from other objects before they reach the receiver, thus causing time values to be slightly large. Such errors are called multipath errors because the signals come to a receiver from more than one path as illustrated in Figure 5.15.

(3) Satellite Errors

Obviously timing is an extraordinarily important part of GPSs. Even though satellites are equipped with very accurate atomic clocks, they are not perfect. This can cause errors in the time measurement and consequently the distance measurement. In addition, a satellite might drift a little from its predicted orbit. The results are satellite ephemeris errors that affect the predicted satellite locations. An ephemeris (plural ephemerides) is an astronomical almanac that provides the positions of various bodies in space such as the sun, moon, planets, and other stars at certain time intervals.

(4) Receiver Errors

Receivers are equipped with less accurate clocks. This leads to errors in time measurement and consequently the distance obtained based on that time measurement. Furthermore, the existence of internal noise also introduces errors into the final results.

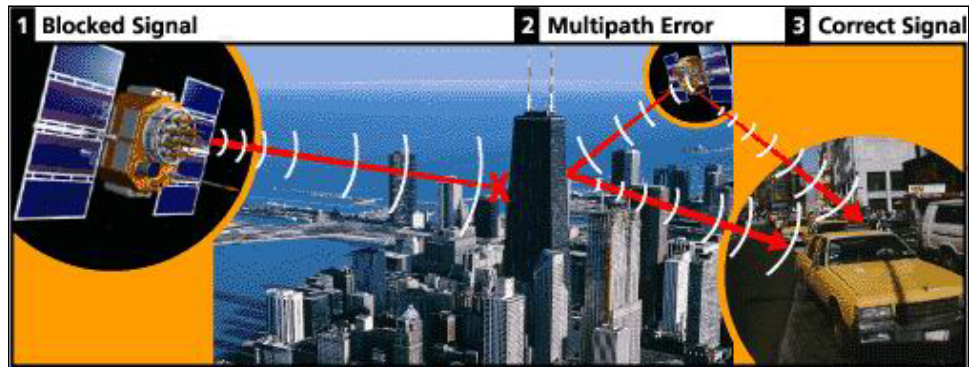


Figure 5.15 Illustration of Multipath Errors (Source: <http://www.garmin.com>)

(5) Selective Availability (SA)

Another type of error that is quite large is deliberately caused by the DOD and is referred to as selective availability (SA). The satellite system was conceived, developed, and paid for by the military establishment. In their opinions, enemies of the United States might use an extremely accurate system that is available to everyone for harmful purposes. DOD has the authority of altering satellite positions or radio broadcast messages from the satellites and consequently, resulting in errors in measurements with varying magnitudes. One thing need to be pointed out is that on May 1, 2000, President Bill Clinton announced that the intentional degradation of the civilian GPS signals would be stopped, allowing the accuracy of GPS to reach 10 meters compared to the previous accuracy of 100 meters due to SA. The long run goal is to exploit the advantages of GPSs to benefit the whole society.

Table 5.1 provides a summary of GPS error sources and the comparison of typical position accuracy between basic GPSs and differential GPSs. It is important to understand that the errors described cannot be determined once and used to correct the measurements made all that day. They are constantly changing and there must be several receivers working simultaneously at all times to do the job. This type of GPS application is usually referred to as a real-time kinematic GPS. Table 5.2 defines the different level of accuracies of GPSs with corresponding applications.

Table 5.1 Summary of GPS Error Sources and Error Budget (Worawat and Rasdorf 2003)

Typical Error in Meters (Per satellite)	Standard GPS (meters)	Differential GPS (meters)
Satellite Clocks	1.5	0
Orbit Errors	2.5	0
Ionosphere	5.0	0.4
Troposphere	0.5	0.2
Receiver Noise	0.3	0.3
Multipath	0.6	0.6
SA	30	0
Typical Positional Accuracy		
Horizontal	50	1.3
Vertical	78	2.0
3-D	93	2.8

Table 5.2 Accuracy and Application (Worawat and Rasdorf 2003)

Accuracy (Qualified)	Accuracy (Quantified)	Application
Very High	$\pm < 5$ mm	Survey Positioning
High	$\pm < 1$ meter	Mapping & GIS data
Moderate	$\pm < 20$ meters	Approximate location
Low	$\pm < 10$ to 200 m	Marine navigation

5.4.4 Using GPS to Measure Distance along Linear Objects

The main capability of GPSs is determining a point position of the Earth surface in the format of spherical coordinate system (position with latitude, longitude, and elevation). This absolute position can be converted into a more familiar format with X/Y/Z-coordinates by using projections, datums, and specific coordinate systems.

GPSs are also capable of measuring the 3-D distance along a linear object by taking an approach that is similar to what was illustrated in Figure 5.11. In other words, a GPS measures the 3-D distance of a linear object by providing X/Y/Z-coordinates for the start node, end node, and many intermediate nodes. A series of straight line segments are constructed by connecting the neighboring nodes to approximate the linear object being measured. The 3-D distance for each straight line segment can be computed using the mathematical function illustrated in Figure 5.11. These distances are added together to determine the 3-D length for the complete linear object.

5.4.5 Advantages and Disadvantages

GPSs are a relatively new technology that is still undergoing rapid developments. GPSs provide a very efficient way to obtain point positions. By taking a similar approach to the total station, GPSs are capable of determining the 3-D distance for a linear object via approximating the linear object with a series of connected shorter straight line segments.

While GPSs can achieve a very high accuracy level when obtaining point positions, its accuracy in determining 3-D length for linear objects decreases significantly because different modes are taken to obtain point positions and to determine lengths. When GPSs are used to obtain a point location, the receiver can stay at the point for a few minutes so that hundreds of readings will be taken and averaged to achieve a higher accuracy. This is not the case when determining lengths. Usually, only one reading will be taken for the start node, end node, and each of the intermediate nodes. Furthermore, it is very possible that when taking positions for some of those nodes, bad readings will step in, or there may not be sufficient number of satellites available to the receiver to obtain a reading. This is always the case when a GPS receiver is mounted on a vehicle and being moved fast to achieve efficiency.

5.5 DISTANCE MEASUREMENT INSTRUMENT (DMI)

The most commonly used technology in transportation to measure the 3-D distance of roads (linear objects) is DMI, which works similar to a vehicle odometer by driving along roads. This section describes the background of DMI, its working principles, and data quality issues.

5.5.1 Background

Nowadays there are many ways to measure distance. One of such methods is to use a DMI. The prototype preceding today's DMIs is the hand-wheel, which is illustrated in Figure 5.16. Today's DMIs work together with vehicles. Figure 5.17 shows a DMI attached to the car-dash board.

5.5.2 Fundamentals of DMI

A DMI works like a receiver to receive the pulses and calculate distance. For accurate electronic measurement of distance, a pulse generator, either a wheel sensor or a transmission sensor, and an electronic interface amplifier are required to work together with a DMI (Nu-Metrics 2002).



Figure 5.16 Hand Wheel
(Worawat and Rasdorf 2003)

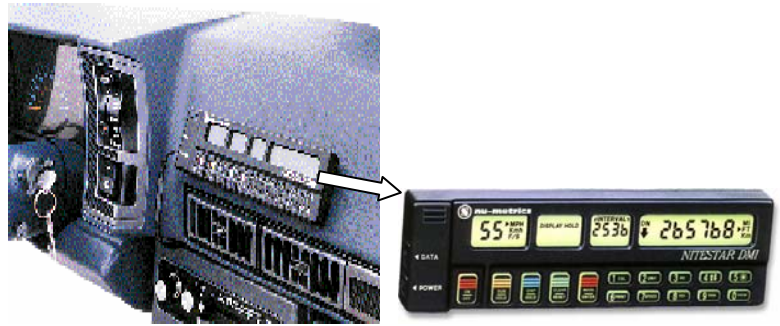


Figure 5.17 DMI Attached to Car Dash Board
(Courtesy of Nu-Metrics)

Figure 5.18 illustrates how a DMI works together with the pulse generator and the amplifier to obtain distances. An electrical impulse is generated by sensors when the vehicle is traveling. The generated pulses are then sent to the electronic interface amplifier. The electronic interface amplifier will divide and amplify the pulses into the suitable working rate. The pulses from the electronic interface amplifier are sent to the DMI. Finally, the DMI mathematically convert the pulses to represent the distance the vehicle has traveled.

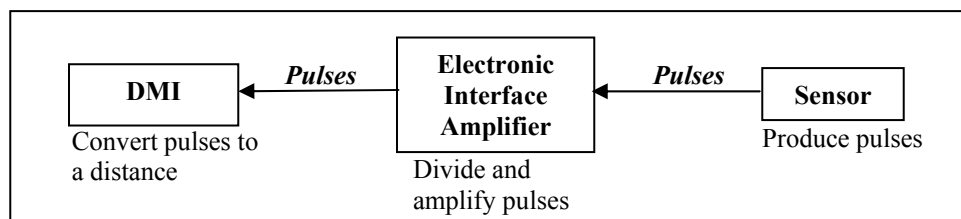


Figure 5.18 Using DMI to Calculate a Travel Distance (Worawat and Rasdorf 2003)

There are two types of sensors available to DMI, wheel sensor and transmission sensor. Most new vehicles (1992 and newer) have an electronic distance/ speed sensor inside their transmission, which sends pulses to the

vehicle's on-board computer for use in the speedometer, ABS braking system, engine/ transmission control, etc (Nu-Metrics 2002). For older vehicles (earlier than 1992 models), they need to be installed one of two types of sensors working with DMI, a wheel sensor or a transmission sensor.

5.5.2.1 Wheel Sensor

Figure 5.19 shows a wheel sensor attached to the wheel of a vehicle. For the wheel sensor, the vehicle tire is divided into pie sections by the equally spaced targets attached to the rim. The spacing distance between each target projects the proportional distance of the outer circumference of the tire. For example, if a 15-inch tire having an average rolling circumference of 7.16 feet is divided into 8 pie sections with 8 equally spaced targets, the distance between two neighboring targets is one eighth of the circumference, which is 0.895 feet.

An electrical impulse is generated when the metal target is in close proximity to the wheel sensor. As the vehicle moves, the sensor detects each wheel target as illustrated in Figure 5.19 and generates electrical pulse. When the measuring instrument receives the sensor pulses, the pulses are mathematically converted to represent the traveled distance.

5.5.2.2 Transmission Sensor

A transmission sensor sits in the vehicle's transmission as illustrated in Figure 5.20. Like the wheel sensor, an electrical impulse is generated when the metal target is in close proximity to the special magnetic disk rotates in the transmission sensor. As the vehicle moves, the sensor detects the disk rotation with the transmission sensor and generates electrical pulse. When the measuring instrument receives the sensor pulses, the pulses are mathematically converted to represent the traveled distance.

With the transmission sensor, six pulses are generated with each revolution of the internal disk inside the transmission shaft. Since most automobile speedometers are based on 1000 revolutions per mile, approximately 6000 pulses per mile are obtained. By dividing the pulses received (6000) into the course length (1 mile = 5,280 feet) each pulse represents 0.880 feet ($5,280 / 6,000 = 0.880$).

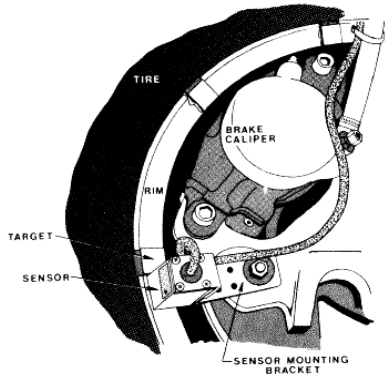


Figure 5.19 Wheel Sensor
(Courtesy of Nu-Metrics)

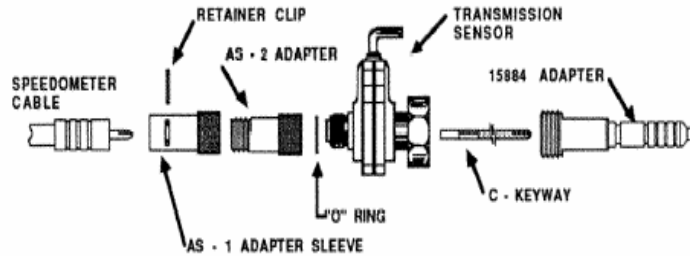


Figure 5.20 Transmission Sensor
(Courtesy of Nu-Metrics)

5.5.2.3 Electronic Interface Amplifier

The electronic interface amplifier takes pulses already generated by sensors and amplifies these signals. The pulse rate can vary from 4,000 to in excess of 100,000 pulses per mile, which is far more than the DMIs needs to accurately measure distance. The electronic interface amplifier will condition and amplify these signals for use by the DMI. Since the higher pulse rates are not required for accurate distance measurements, the electronic interface amplifier incorporates the divider circuit to reduce the pulses. This circuit changes the 100,000-pulse/mile-vehicle sensor pulse output to the 6,000 or so per mile needed by the DMI (Nu-Metrics 2002). Since the vehicle manufacturers are not concerned with accurate distance and speed-readings at very low speeds (e.g. a speedometer may not move below 5 mph), the signal strength at these low speeds is extremely minimal. An electronic transmission interface will amplify the low speed signals so that the accurate distance measurement can be obtained even at very low speeds.

5.5.2.4 Calibration

The DMI works by counting pulses produced either by an attached wheel sensor or a built-in transmission sensor. The number of pulses per mile produced is different when different vehicles are used. On one car it might be 5280 pulses per mile (i.e. 1 foot per pulse) while on another car it could be 10,560 (0.5 foot per pulse). The instrument counts pulses but the distance in feet (or miles) is preferred by converting the pulse count to an engineering measurement. If one travels over a known distance and the instrument counts the pulses over that

distance, it can then automatically figure out the conversion between pulses and feet. This value is called the *calibration number*. The *calibration number* for each vehicle will be the number of feet the vehicle travels between each pulse multiplied by 1000. This value varies between 400 and 1400 (.4 to 1.4 foot/pulse) (Jamar 2002). The determination of this value for the vehicle is called calibration.

Although DMIs are exceptionally well built, the effect of age, general use and wear may change the performance of DMIs. Therefore, it is essential to calibrate DMIs regularly and properly in order to accurately measure distance (Jamar 2002). In addition, it is very common to use a DMI on a plug-in basis among numbers of different vehicles that have been equipped to accept the instrument. This is a cost-efficient way but requires calibration before each use.

In order to calculate the calibration number for a particular vehicle, a calibration course must be established. The length of the course can be any known distance longer than 500 feet (1000 feet is ideal). The course should be straight with accurate distance.

5.5.3 Data Quality Issues

DMIs are probably the most accurate instrument available in the transportation field to measure 3-D distances for roads. However, errors can still occur and consequently inaccurate distance measurements appear. This section describes the error sources in DMI measurements and identifies the typical accuracy by examining the market models.

5.5.3.1 Errors

The error sources in DMI distance measurement are listed as below.

1) Signal Error

Due to various dynamic conditions of a vehicle in motion, there might be missing pulses. Occasional missing pulses can generate significant errors over long distances.

2) Speed of vehicle

The signal strength at the low speeds (e.g. less than 5 mph) is extremely minimal. That makes DMIs produce inaccurate distance information. A sensitive sensor is needed to solve this problem.

3) Calibration

The course of calibration can cause errors. According to the calibration procedure, the course should be straight and accurate. If it is too short or too long, it can also cause errors as well.

4) Equipment Errors

A DMI works with sensors, amplifiers, and the vehicle. If any of these are not in good shape, errors occur. For example, if the sensor is not in a good condition, it cannot generate pulses properly. In case a wheel sensor is used, if the tire does not have ample tread path or the tire pressure is not the maximum suggested by the tire manufacturer, the tire can make the wheel sensor work improperly.

5.5.3.2 Accuracy

Table 5.3 summarizes the accuracy that could be achieved by several market models. The typical accuracy of DMI in terms of repeatability is +/- 1.0 foot per mile. The repeatability is obtained by measuring the same distance with the same instrument several times. The maximum difference between two measured distances is the repeatability. For example, with the DMI that has the repeatability of +/- 1.0 foot per mile, it means that one can repeatedly use the DMI and get a difference of the returned distances within 1.0 foot of the actual one-mile distance.

Table 5.3 Accuracy of DMIs (Based on Market Models)

Manufacturer	Model	Accuracy*
Jamar	RAC100	1 foot/mile
	RAC200	1 foot/mile
Nu-Metrics	NS-50	1 foot/mile
	NS-60	1 foot/mile
Sun-Lap Technology	RoadTracker SL1000 DMI	1 foot/mile

*Note: Accuracy is measured in terms of repeatability.

5.5.4 Advantages and Disadvantages

The most prominent advantage of a DMI is that it measures directly the 3-D distance of linear objects. Unlike total station or GPS, in which a linear object is approximated using a series of connected straight line segments, a DMI is not limited to the shapes or curvatures of linear objects such as roads (Jamar 2002). This is why DMIs are being used widely in transportation to obtain road distance information.

A DMI measures distances in a more efficient way than total stations and GPSs. A DMI allows field personnel to measure distances while driving as fast or slow as necessary. However, even with this high efficiency, using DMIs to obtain large amount of distance information, for example, obtain the distance information for thousands of highway road segments in North Carolina, is still time-consuming and labor-intensive (Rasdorf and Cai 2001).

5.6 OBSERVATIONS

Examining the technologies that could be deployed to obtain 3-D distances for linear objects, especially transportation linear objects (roads), leads to following observations.

- Many technologies could be used to determine the 3-D distance for linear objects, even though some of them were not designed exclusively for this capability. The main objective of design drawings is to guide construction. The total station serves surveying purposes. The GPSs are aimed at obtaining point positions.
- Only DMIs aim at measuring 3-D linear distances. When using other technologies, approximations have to be made. The most popular way is to approximate a linear object with series of connected straight line segments.
- The approach of obtaining 3-D lengths from legacy drawings is infeasible due to its limitations mentioned earlier.
- The approach of using total stations to obtain distance information will interrupt traffic.
- All these technologies are time-consuming and labor-intensive when used on a large scale for data acquisition.

- There is a trade-off between efficiency and accuracy.
- When digital road data are available in two-dimension, how to locate the start and end nodes for a specific road segment in the field is very difficult, which introduces significant errors into the results.

These observations lead to our motivation of finding a more efficient approach to determine 3-D distances for road segments. GISs and 3-D spatial modeling attract our research efforts because most road data are being stored in GISs as two-dimensional linear objects. On the other hand, elevation datasets are also available to GISs in the formats of 3-D points, DEMs, TINs, or contour lines. 3-D spatial modeling enables representing roads as 3-D objects by combining their 2-D representation and the elevation datasets. By modeling roads in a three-dimensional space, the 3-D distance for a particular road segment can be easily obtained, together with many other geometric properties such as curvature and slope. This will greatly benefit the transportation field by providing better quality of data. Furthermore, it could be applied to linear objects other than roads, for example, utility pipes, and consequently, benefits numerous engineering applications of GISs.

6 GLOBAL 3-D MODELING CONCEPTS FOR LINEAR OBJECTS AND LENGTH PREDICTION

This chapter introduces the geometric properties of transportation linear objects (roads) and the global 3-D modeling concepts to represent road centerlines as 3-D lines and to predict their 3-D lengths.

6.1 ROAD GEOMETRY

Roads are man-made 3-D objects in space. They are usually regularly shaped. Following are some of the geometric properties that are necessary to describe the road segments as 3-D objects in a 3-D spatial referencing system. To help understand the concepts described here, please refer to Figures 1.4 and 1.5 in Section 1.5.

(1) Cross Section

Cross section refers to the road profile that is perpendicular to the centerline and extends to the limits of the right-of-way within which the road is constructed (Papacostas and Prevedouros 2001). The cross section of road includes the pavement, shoulder, clear zone, and additional R/W as required.

The super elevation is the lateral bank on the cross section of the curved paths (Papacostas and Prevedouros 2001). This concept has been introduced in almost all elementary kinetics and kinematics books and will not be discussed in detail here. The point is that the design of super elevation is always integrated with the design of horizontal alignment (Papacostas and Prevedouros 2001). Banking the cross section is needed on the curved portion of the facility but is not necessary along the tangent segments of the horizontal alignment (Papacostas and Prevedouros 2001). Consequently, the transition of the cross section from the normal crown on the tangent to a fully superelevated pavement on the curve is required. The length from the point where the banking starts to the point where a full banking is achieved is called superelevation, which is dependent on the rate at which the cross section is rotated (Papacostas and Prevedouros 2001).

(2) Horizontal Alignment

The horizontal alignment represents the projection of the road facility on a horizontal plane. Generally, it consists of straight-line segments (tangents) connected by circular curves directly (simple curves) or via intermediate transition curves (Papacostas and Prevedouros 2001).

(3) Vertical Alignment

The vertical alignment of highways consists of grade tangents connected with parabolic vertical curves. The desirable maximum design grades and gradient change depend on the facility type and vehicular characteristics (Papacostas and Prevedouros 2001).

6.2 ROAD CENTERLINE

While roads are 3-D spatial objects, they are usually simplified as linear objects using representative lines, which are named road centerlines, in 2-D GISs. The definition of road centerline was provided earlier in Section 1.5. Figure 1.6 in Section 1.5 explained how the road centerline was specified depending on the number of lanes a road segment have and whether or not the two directions of the same road segment are physically separated by a median and are modeled either as one or two linear objects.

In LRS, roads are actually modeled as one-dimensional linear objects. In other words, all events along a road segment are located by referencing to its start point with a distance measurement along the road segment. This is also the major difference between an LRS and a GIS. An LRS can be implemented without GISs, but GISs help LRS to realize visualization and support numerous spatial analyses (Geo Decisions 1997). However, in both 2-D GISs and LRSs, the representing lines used are the road centerlines.

By representing road segments as one-dimensional linear objects in LRSs without using GISs, the geometry properties listed in Section 6.1 are all lost. In case GISs are used to implement LRS, or in case of 2-D GISs, the geometry properties of roads listed in Section 6.1 are all lost except part of the horizontal alignment. While

these simplifications are satisfactory with some applications, it has been realized that the road centerlines are still three-dimensional objects that need to be modeled in a 3-D space.

6.3 3-D MODEL FOR ROAD CENTERLINE

Examining the previous researches in 3-D spatial modeling revealed that those research efforts focused on volumetric representation of 3-D objects, more specifically, 3-D polygons. There are almost no research efforts that have focused on modeling linear objects in the three-dimensional space. However, these previous researches and their proposed 3-D data structures and models are the basis for developing 3-D spatial models for linear objects. This section describes two models to represent road centerlines as 3-D linear objects in a 3-D space, variations of these two models when used with LRSs, and the relationship between these two models.

6.3.1 Point Model

As stated earlier, the most important point regarding 3-D spatial modeling is the treatment of the third dimension – elevation in a similar way as the other two planimetric dimensions. In 2-D modeling, a line object is constructed by connecting two or more points (vertices). The line segment connecting two neighboring vertices could be a straight line, a circular line, or a polygon line depending on the actual characteristics of the line object being constructed and the relative relationship among these vertices. Together, the point locations and the connecting line segments compose the geometry of the line object under construction. Figure 6.1 illustrates a 2-D line object constructed by 7 vertices being connected together. Vertices A and B are connected together with a straight line segment. Vertices B, C, and D are connected together with a circular line segment. Vertices D, E, F, and G are connected together via a polyline. These line segments can be easily described via mathematical functions in the format of $Y = f(X)$, where X and Y are the two planimetric dimensions in 2-D GISs. In other words, this point model consists of a set of points and a set of connecting lines that could be described with mathematical functions.

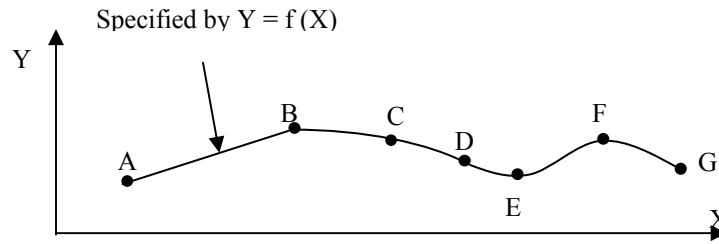


Figure 6.1 Line Construction in 2-D GIS

A natural extension to this principle is to represent a 3-D linear object by connecting two or more 3-D points. 3-D points are points that have X/Y/Z-coordinates specified in a 3-D space. Figure 6.2 illustrates the 3-D representation of the same line object as in Figure 6.1.

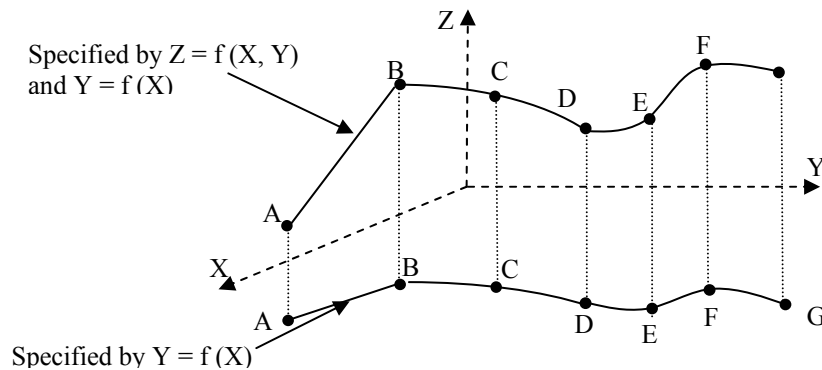


Figure 6.2 Line Construction in 3-D GIS

In Figure 6.2, line segments connecting these vertices can still be described via mathematical functions, but this time in the format of $Z = f(X, Y)$ and $Y = f(X)$, where X, Y, and Z are the three dimensions in 3-D GISs and the combination of two functions together specifies a 3-D line, not a 3-D surface in the three-dimensional space. Similarly, this point model consists of a set of 3-D points and a set of 3-D line segments that could be described with mathematical functions.

While in 2-D GISs, the location of a point is specified by X/Y-coordinates, in LRSs, the location of a point along a linear road segment is specified by a distance (the 3-D distance) along the road segment referring to the start point. In other words, only one measure (the distance) is needed to locate a point on a road segment,

leading to a one-dimensional model of three-dimensional linear object. If a similar approach is taken to model the planimetric projection of road centerline, all events along the planimetric projection of road centerline are located by a planimetric distance along that road centerline while referring to the start point. In this case, a 3-D point can be located in a three-dimensional space via the distance measure (planimetric distance) and associated elevation. This leads to a variation of the point model to represent road segments in a 3-D space as illustrated in Figure 6.3. Only one mathematical function is needed, which is in the format of $Z = f(D)$, where Z is the third dimension introduced (elevation) and D is the planimetric distance along the road segment referring to the start point. Under this circumstance, the 3-D line constructed with the Distance dimension and the Elevation dimension would appear different from that in Figure 6.2. By finding corresponding point on the Distance axis (which has a straight direction), it is similar of stretching the 2-D line with curvatures into a straight line. Similarly this point model variant also consists of a set of 3-D points and a set line segments.

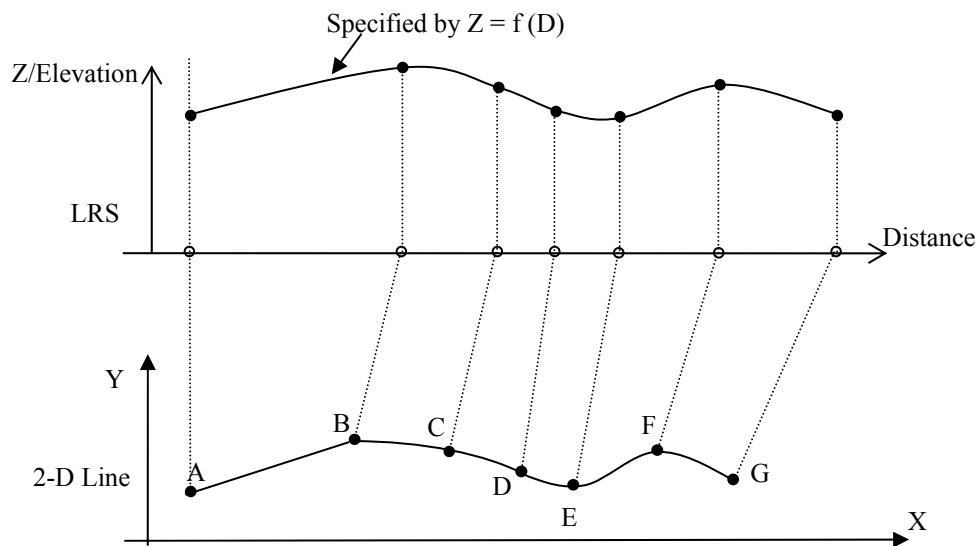


Figure 6.3 A Variation of 3-D Point Model with the Use of LRS

Another variation of the point model to represent 3-D linear objects is to simplify all line segments connecting neighboring vertices into straight lines. These vertices are in turn located via their planimetric distances along the road centerline to the start point of the road segment on which they are located. In this case, no mathematical function is necessary to describe the connecting line segments, even though it can be easily derived. This variation requires sufficient number of vertices in order to achieve a high accuracy. In other

words, this point model variation consists of 3-D points and straight line segments only. Generally, more 3-D points indicate a closer approximation to the original linear object. This variation may not be desirable in visualization, but it is very convenient for specific 3-D analyses such as distance prediction. This variation is very much interested because it is appropriate for modeling transportation linear objects for three major reasons. First, transportation linear objects are all man-made regularly shaped objects connected with straight lines and curves. Second, sufficient 3-D points could be obtained. And third, LRSs are being widely implemented at state DOTs in US. Many state DOTs now have road data in an LRS compatible format.

In case a point model or its variant is used to model linear objects in a 3-D space, all sample points could have their elevations directly obtained from spatial data collection activities or from available elevation datasets, while all other points could have their elevations interpolated from the neighboring sampling points and consequently, elevations for any points on the linear object can be obtained when needed.

6.3.2 Mathematical Model

While mathematical functions could be deployed to describe the connecting linear segments in the point model to represent 3-D linear objects in a three-dimensional space, the mathematical model is different from the point model in that the mathematical model treats a 3-D linear object as a single unit rather than a series of connected linear segments as in the case of point model. For a specific linear object, a function in the format of $Y = f(X)$ is specified to describe its planimetric geometry as in the 2-D GISs, and a function in the format of $Z = f(X, Y)$ is specified to introduce the third dimension in the model.

Similar to all kinds of interpolations in surface modeling that construct a continuous surface from sampling points, a best fitting three-dimensional line could be found to fit the 3-D points that belong to one linear object. This best fitting 3-D line is described via the mathematical functions mentioned earlier. The criteria for the best fitting line can be defined in many ways such as a smooth line that goes through all the points or a smooth line at a certain order that minimizes the total deviations from those points to the smooth line.

Based on the locating mechanism in LRSs, the mathematical model has a variation that is similar to the variation of the point model. In the variant, in order to find the best fitting 3-D line based on the 3-D points that belong to one linear object, a function in the format of $Z = f(D)$, where D is the planimetric distance along the planimetric projection of a linear object when referring to its start node, needs to be defined according to predefined criteria for best fitting (Figure 6.4). This looks almost the same as the variation in the point model. However, there is only one such function for one linear object in case of mathematical model. In case of point model, there will be several such functions for one linear object, one for each segment which, when connected together, forms one linear object. With such a mathematical model, any point on the line object can have its planimetric position located and its elevation calculated.

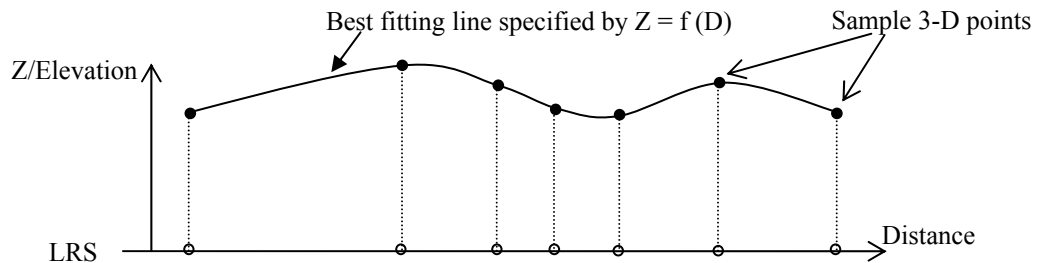


Figure 6.4 An Example of Mathematical Model in the Case of LRS Road Data

6.3.3 Relationship between Point Model and Mathematical Model

It is clear that the point model and its variations and the mathematical model and its variations are closely related to each other. The first thing needs to be pointed out is that in a 2-D GIS, a point is located via its X/Y-coordinates. If the LRS-based approach is used, a point along a linear object can be located via a planimetric distance along the linear object when referring to its start point. The conversion between these two can be easily implemented. In other words, for a point with X/Y-coordinates along a linear object, its planimetric distance along that linear object when referring to its start node can be easily obtained. On the other hand, if a point is located via its planimetric distance along a linear object referring to the start node of that object, its X/Y-coordinates can be easily obtained. This indicates that the conversion between a point model and its

variations is feasible. It also indicates that the conversion between a mathematical model and its variation is feasible.

Second, a mathematical model can be built from a point model by finding a best fitting line for those sample points and describing it with a mathematical function. Similarly, a point model can be built from a mathematical model by extracting 3-D points from the mathematical model. For example, if a mathematical model specifies a 3-D line with the function $Z = f(D)$, with an input value D_1 , the corresponding value Z_1 is obtained. Repeating the above step by using different D inputs D_2, D_3 , etc. leads to the extraction of many 3-D points, and therefore, building a point model for this 3-D line. Furthermore, values D_1, D_2, D_3 , etc. all correspond to pairs of X/Y -coordinates, which leads to building a variation of point model.

6.3.4 Model Selection

This research focuses on one variation of the point model, i.e. modeling a 3-D line using a set of 3-D points assuming they are connected with straight lines. In addition, the planimetric position of such a 3-D point is located via its planimetric distance along the line to the start point. By doing this, it is avoided to derive mathematical functions for best fitting line because there is no such thing as a universally applicable function that would fit all road segments.

For example, assuming there are only ten 3-D points (A_1 to A_{10}) on a road segment, it still has an infinite number of potential point models. Point model M_1 treats points A_1 to A_3 as one group and describe the road segment from A_1 to A_3 with one mathematical function. It treats points A_3 to A_{10} as another group and describes the road segment from A_3 to A_{10} with another mathematical function. Point model M_2 might treat points A_1 and A_2 as one group, points A_2, A_3 , and A_4 as one group, and points A_4 to A_{10} as one group. In other words, there will be an infinite number of point models. All of them have the same set of 3-D points but different sets of mathematical functions.

In addition, given a group of 3-D points, there are still an infinite number of mathematical functions, each of which represents a fitting line. Furthermore, there is no such agreement could be reached as to what criteria should be used to identify the “best” model for a given line segment. To summarize, using the variant of point model rather than the initial point model surrogates the difficulties of having an infinite number of potential models and filtering out the “best” model.

The mathematical model assumes only one mathematical function for a line segment with a set of 3-D points. This solves the problem of grouping neighboring points and deriving a mathematical function for each of these point groups. However, there are still an infinite number of potential mathematical models and there is no such agreement as what criteria should be used to identify the “best” fitting line.

To summarize, only the variant of point model that assumes two neighboring points are connect via a straight line avoids the involvement of complicated mathematical functions into the model. The initial point model, its first variant, and the mathematical model and its variations are all theoretically feasible but not practical. For these, assuming the difficulties in determining the “best” fitting lines could be solved, each road segment would have a distinct model. For an analysis that involves thousands of road segments, the appropriateness of applying these models is questionable. Therefore, the variant of point model, which assumes two neighboring points are connected via a straight line, is chosen to be the model to be implemented in the case study to model linear objects in a 3-D space, to predict their 3-D distances, and to derive other geometric properties such as curvature and slope.

6.4 MODEL CONSTRUCTION

Spatial data are needed in order to construct 3-D models for linear objects. While raw data can be collected from scratch, this is not a desirable way for several reasons. First, data collection is always expensive and time-consuming, even with the rapid developments of technologies in spatial data collection with increasing efficiency and decreasing cost. Second, most linear objects have already been captured and stored as two-dimensional digital data in computers. On the other hand, digital elevation data are also available in formats of

contour lines, DEMs, and TINs. It is obvious that these datasets could be utilized to construct 3-D models for linear objects in an efficient way. Therefore, this research focuses on developing and constructing 3-D models for linear objects using existing datasets. This section describes the datasets needed and how these datasets can be used to construct 3-D models for linear objects.

6.4.1 Source Data

Two datasets are required to construct 3-D models for transportation linear objects: planimetric road centerline data and elevation data. The planimetric road centerline data locate each particular road segment in a two-dimensional plane. The elevation data introduce the third dimension to the road segment and therefore, enable the construction of 3-D models.

As stated earlier, 2-D GISs model road segments as linear objects with vertices and lines connecting these vertices. A vertex is a point being used to model a linear object, which has X/Y-coordinates. The connecting line segments could be straight line, circular, or polyline. In the context of transportation road network, LRSs model road network in a different way. Any event along the road network is located by a distance measurement along the roads, referring to the start node of a specific road segment. Theoretically this distance is 3-D distance, or the surface distance. However, this is rarely the case. In most cases, 2-D planimetric distance is used. When the LRSs are implemented with GISs, these two different models can convert to each other. In other words, given the X/Y-coordinates of a point along a road segment, its planimetric distance to the start point of that road segment can be easily obtained. On the other hand, if the distance of a point on a road segment to the start node of that road segment is given, its X/Y-coordinates can also be obtained easily. To summarize, the two-dimensional road network data locate the road network and any point on that network on a two-dimensional planimetric plane.

Elevation data introduce the third dimension into 3-D modeling. The different elevation data models such as point models, DEMs, and TINs were introduced earlier. The ground surface is a continuous surface with elevations changing from one point to another point. There are an infinite number of points on a continuous

surface and therefore, it is impossible to measure the elevation for every single point. Generally, elevations at sampling points are obtained via stereo aerial photos, satellite images, scanning from satellites or aircrafts, measurements in the field, or combinations of these approaches (Burrough and McDonnell 1998). These point data (or point clouds) are the basis to produce DEMs and TINs by an approach named interpolation. Interpolation is the procedure of predicting the value of attributes at unsampled sites from measurements made at point locations within the same area or region (Burrough and McDonnell 1998). The rationale behind spatial interpolation is that , on average, elevations at points closer together in space are more likely to be similar than points further apart (Burrough and McDonnell).

There are many methods for interpolation. The first group is named global methods including classification using external information, trend surfaces on geometric coordinates, regression models on surrogate attributes, and methods of spectral analysis. The second group is named local deterministic methods including thiesen polygon and pycnophylactic methods, linear and inverse distance weighting, and thin plate splines. These interpolations were not developed specifically for terrain surface modeling, even though they can be applied successfully in terrain surface modeling so that elevation datasets such as DEMs and TINs can be developed from sampling points. With any elevation datasets, the elevation for a point within the region can be either directly obtained (which is rarely the case) or interpolated.

6.4.1.1 USGS DEM and National Elevation Dataset (NED)

The USGS has maintained a lead role in the collection and dissemination of digital elevation data for the nation since the mid 1970s (Osborn et al. 2001). This section describes the elevation datasets being developed by USGS, together with the efforts from many other federal agencies, and being distributed by USGS. More specifically, USGS DEMs and NED datasets are introduced in this section.

6.4.1.1.1 USGS DEMs

A digital elevation model (DEM) is a digital file consisting of terrain elevations for ground positions at regularly spaced horizontal intervals. The U.S. Geological Survey (USGS) is the federal agency that is in

charge of developing elevation datasets in the United States. The most popularly known elevation data developed by USGS are DEMs. According to USGS, a DEM is the digital cartographic representation of the elevation of the terrain at regularly spaced intervals in X and Y directions, using Z-values referenced to a common vertical datum (Maune et al. 2001). However, grid spacing, datum, coordinate systems, data formats, and other characteristics may vary widely, but normally following alternative specifications, with narrow grid spacing and State Plane coordinates for example. While UTM coordinates are commonly used for federal mapping programs that cross state boundaries, State Plane coordinates are normally preferred for state, county, and community mapping programs (Maune et al. 2001). State Plane scale factor errors, caused by forcing a spherical Earth to map as though the world is locally flat, are smaller than scale factor errors from UTM coordinates, and most local surveys are performed using the State Plane Coordinate System (SPCS) which is tailored to each state.

1) USGS DEM Products

The USGS produces five different digital elevation products. Although all are identical in the manner the data are structured, each varies in sampling interval, geographic reference system, areas of coverage, and accuracy; with the primary differing characteristic being the spacing, or sampling interval, of the data (USGS 2001b). A USGS digital elevation model is composed of integer values representing a gridded form of a topographic map hypsography overlay (USGS 2003). Following are the five DEM products from USGS:

- 7.5-Minute DEM 30- x 30-meter data spacing
- 1-Degree DEM 3- x 3-arc-second data spacing
- 2-Arc-Second DEM 2- x 2-arc-second data spacing
- 15-Minute Alaska DEM 2- x 3-arc-second data spacing
- 7.5-Minute Alaska DEM 1- x 2-arc-second data spacing

These digital cartographic/geographic data files are produced by the USGS as part of the National Mapping Program and are sold in 7.5-minute, 15-minute, 2-arc-second (also known as 30-minute), and 1-degree units. The 7.5- and 15-minute DEMs are included in the large scale category while 2-arc-second DEMs fall within the intermediate scale category and 1-degree DEMs fall within the small scale category.

Over the years, the USGS has collected digital elevation data using a number of production strategies including manual profiling from photogrammetric stereomodels; stereomodel digitizing of contours; digitizing topographic map contour plates; converting hypsographic and hydrographic tagged vector files; and performing autocorrelation via automated photogrammetric systems (USGS 2001b). Of these techniques, the derivation of DEMs from vector hypsographic and hydrographic data produces the most accurate model, and is the preferred method (USGS 2001b).

2) USGS DEM Data Organization

A DEM file is organized into three logical records (USGS 2003):

- Type A contains information defining the general characteristics of the DEM, including DEM name, boundaries, units of measurement, minimum and maximum elevations, projection parameters, and number of type B records. Each DEM file has one Type A record.
- Type B contains profiles of elevation data and associated header information. Each profile has a Type B record.
- Type C contains statistics on the accuracy of the data.

A DEM is in the format of a gridded surface interpolated from sampling points. A grid is a rectangular array of cells, each of which stores a value. In the case of USGS DEMs, each cell stores the elevation value for the center of the cell as illustrated in Figure 6.5. The elevation value for any point position other than those centers is assumed to be a value between those of adjacent cells and can be interpolated from these adjacent cell values (Maune et al. 2001).

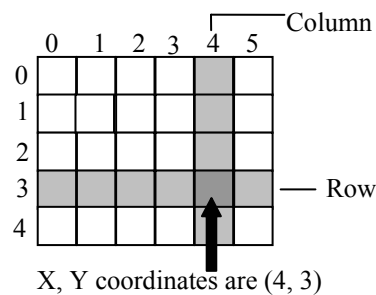


Figure 6.5 Illustration of the Basic Structure of a Grid (Modified after Maune et al. 2001)

DEM data are organized in three classification levels (USGS 2003):

- Level-1 DEMs are elevation datasets in a standardized format. The intent is to reserve this level for 7.5-minute DEMs which are created by scanning National High Altitude Photography (NHAP)/NAPP photography. A vertical RMSE of 7 meters is the desired accuracy standard. An RMSE of 15 meters is the maximum permitted.
- Level-2 DEMs are elevation datasets that have been processed or smoothed for consistency and edited to remove identifiable systematic errors. DEM data derived from hypsographic and hydrographic data digitizing, either photogrammetrically or from existing maps, are entered into the Level-2 category. An RMSE of one-half contour interval is the maximum permitted.
- Level-3 DEMs are derived from DLG data by incorporating selected elements from both hypsography (contours, spot elevations) and hydrography (lakes, shorelines, drainage). An RMSE of one-third of the contour interval is the maximum permitted.

DEM data are also organized into three scales (USGS 2003):

- Large scale: The majority of the 7.5-minute DEMs produced are categorized as Level-1 DEMs.
- Intermediate scale: All 2-arc-second DEMs derived from contours are Level 2. All 2-arc-second DEMs derived for 7.5-minute DEMs are Level 1.
- Small scale: All 1-degree DMA DTED-1 data have been classified as Level 3.

USGS is also providing DEMs in the Spatial Data Transfer Standard (SDTS) distribution format (USGS 2003). SDTS is designed as a mechanism for the transfer of spatial data between various computer systems. The SDTS format is designed to transfer data with complete content transfer (no loss of information) and to work between various platforms and various data models. However, there is limited software availability at this time which can accommodate the SDTS-formatted data (USGS 2003).

3) Data Sources of DEMs and Extraction of DEMs

DEM data sources include direct measurement in the field with theodolite and GPSs, stereo aerial photographs, scanner systems in aeroplanes and satellites, and the digitizing of contour lines on paper maps (Maune 2001). For systematic mapping, elevation data are derived by the methods of photogrammetry from overlapping stereoscopic aerial photographs and satellite imagery. For special purposes, other kinds of scanners can be used, such as airborne laser interferometry (LIDAR and IFSAR, for example) for high-accuracy surface measurements (Maune 2001). Sonar scanners mounted on boats, submarines, or hovercraft are also used for surveying the elevation patterns of sea and lake beds, or ground-penetrating radar and seismic technologies for mapping the elevation of sub-surface layers (Huff and Noll 2001).

The general procedure for DEM generation can be described as the following. First, spot elevations and breaklines are captured (Maune 2001). Alternatively, if photogrammetry is being used, then the elevations for all regularly spaced points can be observed. Second, a digital elevation matrix is collected and augmented with breaklines and spot elevations. And finally, interpolation, compilation, correction, and improvement are performed to finalize the DEM (Maune 2001).

It is obvious that the elevation measured might not distribute uniformly and coincidentally with the data points in a DEM. In other words, in a DEM that the elevation points are regularly spaced must estimate their elevations from the irregularly spaced data points by interpolation. Examples of interpolation include linear interpolation (inverse distance interpolation) and nonlinear interpolations (Maune et al. 2001). Nonlinear interpolations include splines and kriging. The assumption for inverse distance interpolation is that the value of an attribute Z at some unvisited point is a distance-weighted average of data points occurring within a neighborhood (Maune et al. 2001). Spline interpolations deploy a polynomial equation to fit all the data points in a specified region (Maune et al. 2001). This equation is then used to derive elevations for this region according to x and y coordinates and thus creates a smooth surface for this region (Maune et al. 2001). Splines capture surface trend. Kriging optimizes the interpolation procedure on the basis of the statistical nature of the surface. Kriging uses the idea of the regionalized variables, which varies from place to place

with some apparent continuity but cannot be modeled with a single smooth mathematical equation (Maune et al. 2001).

In order to produce DEMs, ground surveys can be conducted to acquire point elevations and locate breaklines (Maune et al. 2001). Ground survey is deemed as labor intensive and costly. Even with the developments and improvements of technologies such as global positioning systems (GPS), this approach is rarely used to produce DEM. In addition to money cost and time cost, accessibility is also a major reason contributing to its inappropriateness in DEM generation. It is listed here for completeness purpose. While developing DEMs, numerous ground survey points might be needed to control data processing and compilation.

Existing maps can be digitized to produce a digital copy of the paper map. Spot elevations and breaklines can be acquired from the digital map. These spot elevations and breaklines are then combined to produce the elevation matrix in the format of X/Y/Z-array. On the other hand, instead of digitizing the whole map, spot elevations and breaklines can be acquired by interpolation and visual examination on the paper map. In digitizing, only these points and breaklines are digitized to save digitizing time.

Photographs are the major source of DEMs. Photogrammetric methods for DEM extraction fall into a number of categories (Molander 2001):

- Stereocompilation methods including traditional photogrammetric capture with analytical workstations utilizing direct viewing of film imagery, and capture with digital photogrammetric workstation using either digitized or digital source imagery.
- Automatic collection of elevation data by digital correlation from digitized film or digital imagery. This is augmented by editing with analytical or digital compilation workstations.
- Hybrid approaches.

Photogrammetric approaches use the following imagery sources today (Molander 2001):

- Traditional aerial film cameras, used both by analytical and digital image workstations.

- High-resolution commercial remote sensing satellite systems generally limited to smaller scales, but affording direct digital imagery access and high quality imagery.
- Emerging digital airborne systems that are being positioned to replace film cameras.

Photogrammetrically compiled DEMs are comprised of mass points, three-dimensional line strings (break lines) and points of special significance (Molander 2001). These three types of three-dimensional features may also be supplemented with photogrammetric control points within the aerial triangulation solution and spot heights collected for cartographic purposes.

Conventional photogrammetric procedures for generating DEMs involve taking stereoscopic measurement of the elevation of terrain points with a precision stereo photogrammetric instrument (Molander 2001). The point distribution may be regular, quasi regular, or irregular, with a certain average density. The character and density of the DEM is dictated by the applications that it must support, and to an extent the application software. The most common practice is to collect breaklines, add mass points to provide an adequate density of points in areas not properly described by the breaklines, and finally to include points of special significance, typically hilltops and low points in depressions and saddles (Molander 2001).

Combining of SAR and interferometry techniques into a single system is called interferometric synthetic aperture radar (IFSAR or sometimes referred to as InSAR or ISAR), which operates at microwave frequencies (3-40,000 MHz) (Hensley et al. 2001). IFSAR measures the three dimensional locations of imaged points to a high degree of accuracy. In IFSAR, sophisticated image processing is required to form recognizable images from the raw data (Hensley et al. 2001). The relative three-dimensional location of the target point to the receiving antennas in the aircraft is in the raw data (Hensley et al. 2001). Meanwhile, the INU/GPS combination together with ground station provides accurate positional data for the receiving antennas, thus leading to positioning the target point.

Post processing of IFSAR is similar to standard photogrammetric or LIDAR post processing (Hensley et al. 2001). The first step is regridding. Smoothing of the interferometric phase measurements is done to reduce

phase noise and aid in the unwrapping process. Typically this involves spatially averaging the phase over a window by an amount that is set by the processor operator. This window is often larger than the post size of the DEM. Thus, the effective resolution of the DME may be less than post size depending on the spatial frequency spectrum of the underlying topography. Unwrapped phase measurements lie on a uniform grid in range and azimuth, however the position measurements determined from the phase measurements are distributed unevenly with respect to the processor ground projection. To obtain an elevation map on a uniform ground projection grid, the data is interpolated or resampled to the desired posts in the regriding process (Hensley et al. 2001).

The second step in post processing IFSAR data is map mosaic (Hensley et al. 2001). Mosaic is the process whereby multiple image and/or DEMs are merged into a single image and/or DEM with a common datum, map projection and data format (Hensley et al. 2001).

After these two steps, a DEM can be produced from the regrided data. However, there might be data gaps (Hensley et al. 2001). These gaps need to be filled by acquiring data over these gaps via using alternate data sources or using analytical methods to fill in the gaps. After gap filling, the final step in generating DEM is data editing, which corrects errors in the DEM detected during the quality control process or to manipulate height values so that they conform to a user prescribed mapping standard (Hensley et al. 2001). In addition, applications that require bare surface DEMs need to have IFSAR reflective surface elevation measurements corrected to bare surface elevations, which is called vegetation removal (Hensley et al. 2001). Vegetation removal involves identifying vegetated regions and then correcting the elevation measurements to the bare surface. Correction to bare surface elevations may employ algorithms similar to LIDAR and photogrammetric sensors where elevation measurements that penetrate to the bare surface are used in combination with surface fitting algorithm to make elevation adjustments.

SONAR is the acronym for Sound Navigation and Ranging. This technology employs the application of sound waves to detect the relative position between the sound wave origin and the target point (Huff and Noll 2001).

For these remote sensing technologies mentioned above and the one will be discussed in detail in the next (LIDAR), they have some common characteristics. All of them find the relative position between the observer point and the target point. Meanwhile, the three-dimensional position of observer point can be acquired using INU/GPS combination. Thus, using mathematical and geometric principles, the three-dimensional target position is derived. The processing of raw data collected in remote sensing leads to images and DEMs.

4) General Description of Errors

Errors step into DEMs through all stages of data collection, post processing, and DEM generation. Errors can be categorized into the errors existing in the data source and the errors that are introduced during DEM generation. These two categories of error can be further categorized into instrumental errors, human behavioral errors, and environment-related errors (weather, for example) (Daniel and Tennant 2001).

Errors existing in the data sources vary according to different data sources (Daniel and Tennant 2001). For errors in the data sources, the instrument errors include the errors related to theodolite in ground surveying, GPS in remote sensing, the internal navigation system errors in IFSAR, LIDAR, and photogrammetry, the sonar transmitter and receiver errors, the camera errors in photogrammetry, and etc. (Daniel and Tennant 2001). Human behavioral errors are those errors related to human operations, for example, the aircraft tilting and deviation is due to the pilot operation of the aircraft. Environment-related issues such as weather (sunlight, wind, snow, cloud, and etc.) also cause errors in photographs.

In DEM generation procedure, errors can step in at any time (Daniel and Tennant 2001). Errors can step in when collecting data points from the existing data sources. In addition to the errors existing in the data source, the extracted data contain errors according to the way in which the data are extracted (Daniel and Tennant 2001). For example, in manually extract data points viewing in stereoscope, ten operators might get ten different values for the same point. If the data source is paper map and this paper map is digitized into computer, then digitizing errors step into the DEM generation procedure. The digitizing errors have all three

categories of errors. The digitizer itself has instrumental errors. The operator of the digitizer contributes the human behavioral errors. The working environment (humidity and temperature cause expansion and shrinkage of paper maps, and etc.) introduces the errors related to working environment.

Another major category of errors encountered in DEM generation is the errors and approximations introduced by the interpolation methods (Daniel and Tennant 2001). As stated earlier, each interpolating methods estimates the elevations for unvisited point locations by interpolating the elevations for the visited points (points that have elevations). This procedure is an estimating procedure with errors varying according to the interpolating method being used.

It is worth mentioning that errors cannot be completely eliminated from geographic data. However, they can be controlled and improved. Some errors are totally avoidable while some can be improved via cautious calibrations. Before using any geographic data, a data quality evaluation must be performed to determine the appropriateness of a data set for a specific application (Rasdorf and Cai, 2001).

A photogrammetrically compiled DEM may also suffer from three different types of errors: random errors, blunders, and systematic errors, like other spatial datasets. Each of these errors influences the DEM in terms of planimetry and height. The following summarizes errors in the photogrammetrically compiled DEMs focusing primarily on the vertical error (Daniel and Tennant 2001).

- Random errors that result from accidental or unknown combinations of problems. Random errors remain in the data after blunders and systematic errors are removed (USGS 1997).
- Blunders, or vertical errors that exceed the maximum absolute error permitted (associated with the data collection process). Blunders can be easily identified and removed through visualization editing.
- Systematic errors that have fixed patterns and are usually related to collection procedures or photogrammetric systems in the DEM generation. These errors are not easily detectable and can introduce significant bias or artifacts in the final DEM product (Wechsler 1999). If the error can be identified, systematic error can be modeled, reduced or even eliminated.

Generally, as the DEM generation is conducted in the model/image space, any systematic errors in the stereo models or images will be reflected in the generated DEMs. These include (Daniel and Tennant 2001):

- Errors of camera calibration. When the camera calibration data is not accurate or up-to-date, the internal geometry of the bundle of rays will not be correctly reconstructed within the photogrammetric systems.
- Errors introduced by film and photomechanical process. The development of the film can result in deformations relative to the state of the film at the instant of exposure. Similarly, the production of diapositives for photogrammetric processing in analog systems is a source of additional error.
- Errors of the photogrammetric stereo plotters.
- Errors of scanners/scanning process.
- Errors of aero triangulation.
- Errors of exterior orientation parameters.
- Errors related to the earth curvature and atmospheric corrections.
- Influence of image matching algorithms.
- Errors of collection technique. Manually profiled DEMs typically display a “corn-row” effect when surface visualizations are generated. This effect is a result of more consistent elevation heights along the profile than across neighboring profiles and causes problems when deriving contours or drainage networks.
- Vegetation interference. The height and density of various types of vegetation in a project area can be the most significant contributor to systematic error in a DEM.
- Personal bias. When a photogrammetric technician collects DEMs on a stereo plotter, an operator-induced bias is introduced to the final DEMs. Each operator “reads” the height of the terrain slightly differently. The same operator may even read the terrain differently from day to day or even throughout the day.

The basic IFSAR errors are additive and have various sources (Daniel and Tennant 2001):

- Atmosphere
- Navigation measurement

- Baseline measurement
- Range measurement
- Phase – phase errors result from errors in image registration, motion compensation, image formation algorithms, target motion (temporal decorrelation), and target volumetric scatter.
- Differential phase errors including radome effects and multipath errors.

Most of the error sources listed above will contribute to the performance of IFSAR system. The largest contributions to systematic errors are (Daniel and Tennant 2001):

- Atmospheric errors (space borne systems)
- Navigation errors, including baseline estimation errors
- Range measurement errors
- Multipath or radome-included effects to the return phase.

5) DEM Accuracy

The method of determining DEM accuracy involves computation of the root-mean-square error (RMSE) for linearly interpolated elevations in the DEM and corresponding true elevations from the published maps. Test points are well distributed and representative of the terrain. Collection of test point data and comparison of the DEM to the quadrangle hypsography are conducted by USGS quality-control groups (USGS 2003). Following is the data accuracy information provided by USGS regarding its DEM products (USGS 2003).

a) Large scale DEMs

The vertical accuracy of 7.5-minute DEMs is equal to or better than 15 meters. A minimum of 28 test points per DEM is required (20 interior points and 8 edge points). The accuracy of the 7.5-minute DEM data, together with the data spacing, adequately support computer applications that analyze hypsographic features to a level of detail similar to manual interpretations of information as printed at map scales not larger than 1:24,000 scale.

Accuracy of the 15-minute DEMs is equal to or better than one-half of a contour interval of the 15-minute topographic quadrangle map. The accuracy of the 15-minute DEM data, together with the data spacing,

adequately support computer applications that analyze hypsographic features to a level of detail similar to manual interpretations of information as printed at map scales not larger than 1:63,360 scale. The plotting of contours from the 15-minute Alaska DEMs at scales larger than 1:63,360, or reliance on the elevation heights without incorporating the National Map Accuracy Standard (NMAS) horizontal error tolerance, will lead to less reliable results.

b) Intermediate scale

The 2-arc-second DEM accuracy is equal to or better than one-half of a contour interval of the 30- by 60-minute topographic quadrangle map. The accuracy of the 2-arc-second DEM data, together with the data spacing, adequately support computer applications that analyze hypsographic features to a level of detail similar to manual interpretations of information as printed at map scales not larger than 1:100,000 scale. The plotting of contours from 2-arc-second DEM data at scales larger than 1:100,000, or reliance on the elevation heights without incorporating the NMAS horizontal error tolerance, will lead to less reliable results.

c) Small scale

The 1-degree mosaic dataset spacing of elevation and profile data is the same as the 1-degree DEM contiguous U.S. data. The accuracy of the 1-degree DEM data (equal to or better than one-third of a contour interval), together with the data spacing, adequately support computer applications that analyze hypsographic features to a level of detail similar to manual interpretations of information as printed at map scales not larger than 1:250,000 scale. The plotting of contours from the 1-degree DEM at scales larger than 1:250,000, or reliance on the elevation heights without incorporating the NMAS horizontal error tolerance, will lead to less reliable results.

In addition, The USGS Data Users Guide 5 for DEM's (traditionally known as the "blue book"), last published in 1993, contains a general discussion of DEM accuracy and a brief description of the accuracy testing methodology (USGS 1993). It categorized DEMs into level 1, level 2 and level 3 DEMs, depending on data

characteristics, data purposes, and data sources of producing these DEMs. The level of a DEM is defined, in part, on the basis of its accuracy. For example, it describes level 1 DEMs as following (USGS 1993).

“Level 1 DEM's are elevation data sets in a standardized format. The intent is to reserve this level for 7.5-minute DEM's or equivalent that are derived from scanning National High-Altitude Photography Program, National Aerial Photography Program, or equivalent photography. A vertical RMSE of 7 m is the desired accuracy standard. A RMSE of 15 m is the maximum permitted. The intent for 7.5-minute DEM data at this level is that an absolute elevation error tolerance of 50 m (approximately three times the 15-meter RMSE) be set for blunders for any grid node when compared to the true elevation, or that an array of points not encompass more than 49 contiguous elevations and be in error by more than 21 m (three times the 7-m RMSE). Systematic errors within the stated accuracy standards are tolerated at this level.”

6.1.1.1.2 NED

USGS DEMs have several limitations. Elevations were initially only in whole meters to limit the number of significant figures for each elevation post and the resulting file size (Osborn et al. 2001). USGS has not used vertical units finer than decimeters in actual practice for standard DEM products (Osborn et al. 2001). The data format of standard USGS DEMs makes it very difficult for most people to interpret (Osborn et al. 2001). Furthermore, DEMs have been produced in such a format that the elevation data match the quadrangle format of the standard base map series for the United States, most users have the requirement to access and use multiple files of digital elevation data, usually in 7.5-minute tiles, to construct the elevation layer for their study areas. Due to the age of the data and the manner in which they were produced, there are challenges in assembling the DEM tiles (Osborn et al. 2001).

The rapid replacement of all 50,000 DEMs to a more useable format was not possible, but the alternative for USGS was to create a new dataset from the existing database that addressed many of the problems that have accumulated during the history of USGS DEM production (Osborn et al. 2001). This dataset is called the National Elevation Dataset (NED), which is considered by USGS to be the future direction for USGS DEM data (Osborn et al. 2001).

The USGS NED has been developed by merging the highest-resolution, best-quality elevation data available across the United States into a seamless raster format (USGS 2003a). NED is the result of the USGS effort to provide 1:24,000-scale Digital Elevation Model (DEM) data for the conterminous US and 1:63,360-scale DEM data for Alaska (USGS 2003a, USGS 2003b).

The National Elevation Dataset is a new raster product assembled by the USGS. It is designed to provide national elevation data in a seamless form with a consistent datum, elevation unit, and projection (USGS 2001c, USGS 2003b). In the NED assembly process, data corrections were made and the elevation values were converted to decimal meters as a consistent unit of measure. NAD83 was consistently used as the horizontal datum, and all the data were recast in a geographic projection (USGS 2001c, USGS 2003b). The vertical datum is North American Vertical Datum of 1988 (NAVD88). NED has a resolution of one arc-second (approximately 30 meters) for the conterminous United States, Hawaii, and Puerto Rico and a resolution of two arc-seconds for Alaska (USGS 2001c, USGS 2003b).

When a NED tile is assembled, the best available source data are selected according to the following criteria (ordered from first to last): 10-meter DEM, 30-meter level-2 DEM, 30-meter level-1 DEM, 2-arc-second DEM, 3-arc-second DEM (Osborn et al. 2001). For the continental US, the number of 10-meter source DEMs is increasing continually and therefore, the number of 30-meter DEMs used in NED production is decreasing. With the current NED, over 25% is derived from 10-meter source DEMs. In order to keep the data information of the 10-meter source DEMs, the NED production system has been modified to produce 1/3-arc-second data depending on the availability of 10-meter source DEMs to a specific area. As a result, a NED file might have multiple resolutions (Osborn et al. 2001).

NED is essentially a digital elevation model (DEM) in which terrain elevations for ground positions are sampled at either a regular or irregular horizontal interval (Anderson and Mikhail 1998, USGS 2001d). The terrain surface is represented by using an array with X, Y, and Z values. While DEMs are either in the USGS DEM format or SDTS format as stated earlier, NED offers three product format options as below (Osborn et al. 2001).

- ArcGrid, an Arc/Info proprietary format.
- Generic floating point raster: a non-proprietary format containing binary raster data in one file accompanied by separate ASCII descriptor files.
- Generic integer raster: a non-proprietary format containing binary raster data in one file accompanied by separate ASCII descriptor files.

Figure 6.6 illustrates a small piece of NED data, which is in the format of a generic integer raster, and its representation in a gridded cell format. The first two rows shown in Figure 6.6 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 down, 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).

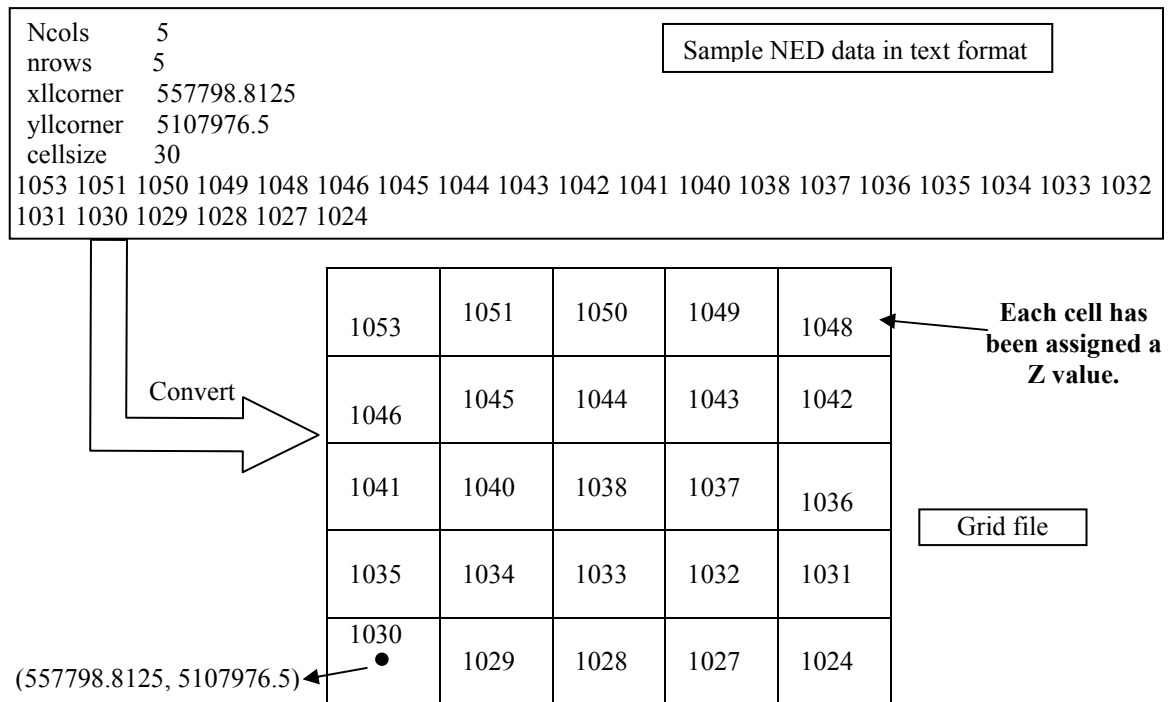


Figure 6.6 Conversion of NED Data into a Grid File (Modeled after Rasdorf et al. 2003a)

The accuracy of NED varies spatially due to the variable quality of the source DEMs and therefore, NED “inherits” the accuracy of the source DEMs (Osborn et al. 2001). Independent accuracy evaluation and assessment are an ongoing research topic of NED.

6.4.1.2 Light Detection and Ranging (LIDAR)

Over the last few years, airborne LIDAR technology has become the accurate, timely and economical way to capture elevation information for DEMs and DTMs (Hill et al. 2000). Elevation datasets produced from LIDAR could be mass points and DEMs with different resolutions.

6.4.1.2.1 Background

LIDAR systems emit rapid pulse of laser light to precisely measure distances from a sensor mounted in a port opening of an aircraft’s fuselage to targets on the ground. The technology of using lasers to measure distance, or laser range finding, has been around for a long time. The first optical laser was developed by Hughes Aircraft Company Laboratories in 1960 (Serr 2000). Early systems were designed to simply measure the distance between the aircraft and the ground from a single laser pulse, and were not effective for mapping. Modern LIDAR acquisition provides surface information for the Earth and features above the surface, such as vegetation canopy, building roof attribute, etc.

6.4.1.2.2 Fundamentals of LIDAR

LIDAR is also known as airborne laser mapping. A LIDAR is similar to the more familiar radar, and can be thought of as laser radar. A LIDAR transmits and receives electromagnetic radiation, but at a higher frequency than radar. A LIDAR operates in the ultraviolet, visible, and infrared region of the electromagnetic spectrum as shown in Figure 6.7.

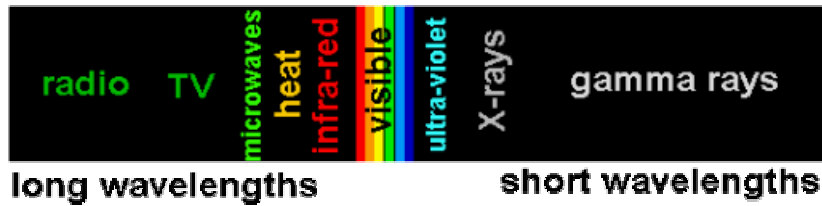


Figure 6.7 Illustration of Electromagnetic Spectrum (Worawat and Rasdorf 2003)

A modern LIDAR system consists of a GPS, an inertial navigation system/inertial measurement unit (INS/IMU), laser rangefinder, and supporting computer hardware and software, as illustrated in Figure 6.8. The GPS is used to determine a position of the sensor. Laser rangefinder is used to determine a distance between the sensor and the target. The orientation of the aircraft is controlled and determined by the INS. Combining all together, LIDAR provides the position of the target ground feature. Figure 6.9 is a depiction of the airborne LIDAR mapping in operation.

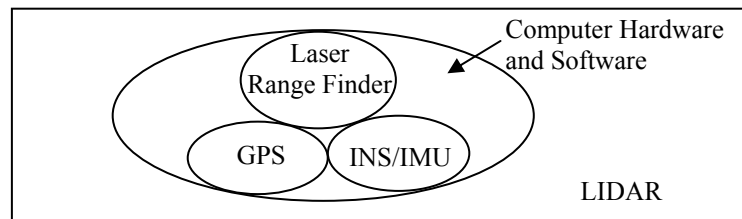


Figure 6.8 Components of a LIDAR System

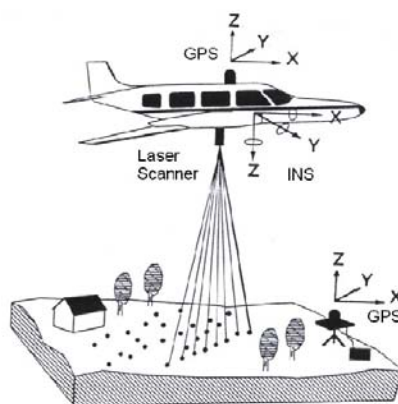


Figure 6.9 Operation of Airborne LIDAR in Mapping (Courtesy of NC Floodplain Mapping Program)

1) Laser Rangefinder

The principle of the laser rangefinder is that a laser pulse from a sensor is directed out of the aircraft by a multi-faceted rotating mirror. Ground features intercept the laser pulse and reflect it back to the aircraft. The receiver catches the reflected laser. By recording the time interval between the emission and reception of the laser pulse, the distance between the aircraft and the feature can be calculated as half of the time difference multiplied by the light speed.

The pulse rates of laser range from 2,000 to 33,000 pulses per second depending on each application and the manufacturers' designs. Even at these high pulse rates, the laser still has enough time to travel from the sensor to the ground and back before the next pulse is sent. A scanning mirror is used to direct the laser pulses back and forth across a wide swath underneath the path of the airplane. The swath width, the linear ground distance in the across-track direction, which is covered by a sensor on a single overpass, is dependent on the altitude of the aircraft and the scan angle. The reflected laser light from the ground follows the reverse optical path and is directed into a small telescope and then to the laser receiver. A series of overlapping and parallel swaths are conducted to cover the entire study area.

2) GPS

The laser distance measurements would be useless by themselves if there were no way to georeference the measurements into X/Y/Z data-points. The precise location of the laser sensor head and attitude of the aircraft must be accurately known in order to individually georeference each laser "hit". This is achieved by using a GPS.

The GPS is used to determine the position of the aircraft. The GPS antenna is usually mounted directly above the laser head. In addition to the onboard GPS receiver, there are one or more GPS ground reference stations. The ground stations identify and correct errors such as the satellite errors, atmospheric errors and so on, and send the correction back to the GPS receiver on board. These are operated by ground crews at known locations, such as USGS benchmarks or monuments, near or within the area of study. This process is called

differential correction. To provide a differential GPS solution for the aircraft position the GPS ground data is logged and post-processed together with the airborne GPS data (Flood 1997).

3) INS/IMU

The precise beam orientation at the time the laser pulse is sent is critical to correctly locating the reflection in a 3-D space. The angle of the laser scan mirror is easily measured at the time of each pulse. However, at the same time the aircraft is pitching, yawing, and rolling during the flight. Pitch, yaw, and roll represent a rotation of the aircraft about the Y, Z, and X axes, respectively, to the desired orientation. A very small change in attitude of the platform would result in a large change in the horizontal distance where the beam strikes the ground. The INS or IMU records the roll, pitch, and yaw of the aircraft at all times. While this would seem like a highly complex measurement, today's inertial systems are able to determine this to within 0.01 of a degree (Fowler 2000).

The INS/IMU units incorporate highly accurate accelerometers and horizontal and vertical gyroscopes, which are electro-magnetically held in suspension. Later during the post-processing phase, the information about the aircraft navigation is used to compute the exact angle and orientation of the beam relative to the coordinate system being used.

4) Supporting Computer Hardware and Software

The supporting computer hardware and software are used to post-process the raw LIDAR data. Generally there are three integrated datasets: rangefinding data, GPS data, and INS data. To achieve the end product, computers with accompanying software are used to store and analyze the raw data, and then store the results as end products.

6.4.1.2.3 Data Collection and Generation of End Product

Due to the dimension of the beam, the chances are good that the beam will hit more than one object on the way down and be reflected several times. In other words, there may be several "returns" for each pulse. For example, in forested areas, the first return may be from leaves at the top of the canopy, a second return may be

from a branch midway down and the last return may be the actual ground elevation. LIDAR systems have been developed to record as many as five returns from a single pulse (Turner 2000). The last return is used to develop DEMs representing bare earth. The time interval between the returns can also be interpreted to get other useful datasets such as height of building data and tree data.

Airborne LIDAR can quickly generate a large number of terrain data points. Commercial systems are available with typical data capture rates equivalent to 100km²/hr. The densities of the data points can be adjusted by changing the flying height, airspeed, scan angle or scan frequency. After raw data are downloaded into computer, post-processing of the raw data is done by using software supplied by the manufacturer of the LIDAR. Steps in post-processing include: auto-processing, calibration parameter checks, identification of systematic errors, and manual processing and data cleaning/anomaly classification. Post-processing costs more time than recording raw data. Typically the post-processing time is two to three hours for every hour of recorded flight data with additional processing time required for more sophisticated analysis such as target classification or vegetation removal (Flood 1997). The end product is normally a comma-delimited ASCII file in X/Y/Z format. Based on these data points, DEMs with different resolutions could be developed.

6.4.1.2.4 Advantages

LIDAR is a cost-effective alternative to conventional technologies for the creation of DEMs and DTMs at vertical accuracies from 15 centimeters to 100 centimeters (Hill et al. 2000). Technologies such as Photogrammetry at this level of accuracy are usually time consuming and costly to perform large-area project.

LIDAR data acquisition is less constrained by weather conditions. LIDAR data can be acquired both during the day and night because the selection of sun angle to avoid shadows is not an issue.

Raw LIDAR data are digital and can be acquired, processed and delivered in a matter of days or weeks instead of months. Delivery time is a function of the size of the study area and the complexity of associated products with added values.

Some of the LIDAR systems record multiple returns of laser, which record information for all significant Earth surface features such as resolution-dependent buildings, trees, levees, and other features in the raw LIDAR data in addition to the bare Earth data. This feature helps to create “bare Earth” DEMs below a tree canopy and also provides detailed information about tree density, height, canopy cover and biomass for characterizing the structure of vegetations and habitats. Mapping of this number and variety of Earth features is usually costly and time-prohibitive when using Photogrammetry (Hill et al. 2000).

The most important advantage of LIDAR is that LIDAR data can be much less expensive to acquire and process than conventional data for wide-area mapping and analysis projects. Traditional approaches for the development of precise elevation datasets have previously required the acquisition and analysis of many stereo pairs of aerial photography, considerable ground surveying, and extensive post-collection analysis (Hill et al. 2000). LIDAR helps to reduce this expensive, labor-intensive and time-consuming task to a cost-effective process.

6.4.1.2.5 Limitations

The laser in LIDAR systems that map Earth features generally operates in the near-infrared portion of the spectrum. Certain Earth features such as water (depending on the angle and turbidity), new asphalt, tar roofs, and some roof shingles, often absorb laser pulses in this wavelength. Moisture, including rain, clouds and fog, typically absorbs the laser signal. This can appear as holes in the dataset. However, this problem can be verified and edited using imagery and/or a triangular irregular network (TIN) model.

In addition, LIDAR datasets, especially when coupled with images, are quite large. Users must be prepared to augment computing resources to store and process the information. Frequently users may require enhancements to the basic LIDAR data. These requested enhancements might use a segmentation and/or compression process by using software to get the value-added products.

6.4.1.2.6 Errors and Accuracy

Errors contribute to the inaccuracy of LIDAR data can be categorized as along-track errors, which are defined as errors in the direction of the flight path of the aircraft carrying the LIDAR system, cross-track errors, which are defined as errors perpendicular to the flight path, and height errors (Daniel and Tennant 2001). Error sources are listed as below based on their effects (Daniel and Tennant 2001). Many error sources contribute to the inaccuracy in along-track direction, cross-track direction, and height direction and therefore, are repeated.

1) Sources of Along-Track Error

- GPS error – The error in determining the location of the scanning system with respect to the geodetic reference frame. In this case, the component is in the along-track direction.
- IMU pitch measurement error – The error in determining the pitch orientation of the scan pattern, affecting the measured location of the scan pattern in the along-track direction.
- Laser bore sight error – The error in determining the exact aim-point of the outbound laser pulse in the along-track direction.

2) Sources of Cross-Track Error

- GPS error – As above for the component in the cross-track direction.
- IMU roll measurement error – The error in determining the roll orientation of the scan pattern, affecting the measured location of points in the cross-track direction.
- Scan angle measurement error – The error in determining exact scan mirror position relative to the scanner housing, affecting the measured location of points in the cross-track direction.
- Rangefinder error – The error in determining the exact distance from the scan assembly to the ground.
- Laser bore sight error – As above for the component in the cross-track direction.
- Atmospheric refraction error – The error in which the true off-nadir angle from the scanner to the laser footprint's point of impact on the terrain is less than the scan angle indicated on the scanner hardware. The effect of refraction error is a cross-track error similar to that caused by IMU roll measurement error or scan angle measurement error.

- Refractive index error – An effect that induces a range measurement error due to variations in the speed of light as the laser pulse propagates through the atmosphere.

3) Sources of Height Error

- GPS error – As above for the height component.
- IMU roll measurement error – Causes the range vector to be displayed resulting in a height error as well as the cross-track direction error at off-nadir scan positions.
- Scan angle measurement error – As with the IMU roll measurement error above, results in the range vector being displaced resulting in a height error as well as a cross-track direction error.
- Rangefinder error – Results in a height error of equal magnitude at nadir. At off-nadir positions, the range error contribution to height error is reduced.
- Laser bore sight error – Similar to roll and scan angle error, results in the range vector being displaced resulting in a height error as well as a cross-track direction error.
- Atmospheric refraction error – As above, in this case being referred to the component in the height direction.
- Refractive index error – As above, in this case being referred to the component in the height direction.

Adding all these error effects together results in a typical 6-inch error budget in elevations and positions. However, the errors in the subsystems can and may cancel each other out. The best absolute vertical accuracy that can be guaranteed from current technology is ± 6 inches (± 15 centimeters) (Hill et al. 2000). With some LIDAR systems, particularly very low level flying configurations, this error may be reduced. But still, the absolute vertical accuracy will not be better than 4 inches. Examining the existing LIDAR systems available on market reveals that the vertical accuracy (RMSE) ranges from 10 to 30 cm (Worawat and Rasdorf 2003).

6.4.1.2.7 LIDAR in North Carolina

A project that collects Flood Insurance Rate Maps (FIRMs) in North Carolina is undergoing now, in which LIDAR point clouds and DEMs (20ft and 50ft resolutions) are being produced (NC Floodplain Mapping Program 2003).

In North Carolina, the majority of the state's Flood Insurance Rate Maps (FIRMs) were at least 10 years old, with many maps compiled from approximate studies in the 1970s without detailed hydrologic and hydraulic (H&H) analyses (Hill et al. 2000). This situation was revealed and called for attentions when the Hurricane Floyd hit North Carolina in 1999. It was clear that most of North Carolina needed to be re-mapped digitally, consistent with FEMA's Map Modernization Plan, using improved H&H modeling and analyses that define current flood risks with greater accuracy. The primary objective of this project is to obtain FIRMs. LIDAR is being used to obtain highly accurate elevation data in NC. The Resulting elevation point clouds and DEMs can be deployed in many applications. This project is paid attention to herein because our research will use the LIDAR elevation datasets it is producing.

6.4.2 Construction of 3-D Point Model

With planimetric data of transportation linear objects on one hand and elevation data (the continuous surface) on the other hand, the 3-D point model can be constructed. For any point on a specific linear object, first, its X/Y-coordinates can be obtained from planimetric data. Second, with its known X/Y-coordinates, its elevation or Z coordinate can be obtained from the elevation data. Repeating these two steps as many times as desired lead to a series of 3-D points along the linear object, each of which has X/Y/Z-coordinates.

In case of the variation of 3-D point model, i.e. an LRS based variation in which the planimetric position of a point is represented by a planimetric distance along the linear object referring to its start node, the planimetric distance measurement can be obtained from the planimetric data. Since this approach of locating a point on a linear object can be converted to the other approach with X/Y-coordinates, the elevation for this point can still be obtained from the elevation data. As a result, a series of 3-D points can also be generated. Under this

circumstance, the 3-D points still have Z coordinates but a planimetric distance measurement referring to the start node of the linear object replaces the X/Y-coordinates.

In addition to those intermediate 3-D points, the elevations for the start and end nodes of the linear object can be obtained from the elevation data and therefore, a complete set of 3-D points are generated to cover the complete linear object. Elevations for all other points can be interpolated linearly from the neighboring points as stated earlier. The problem becomes how dense should these 3-D points be in order to represent the linear object in a three-dimensional space with a reasonable accuracy. It depends on the resolution of the elevation data if they are DEMs and the density of the 3-D points if they are point clouds. As a general rule, in case DEMs are used as the elevation data to construct 3-D point model for linear objects, the planimetric interval between two neighboring 3-D points should be approximately the same as the resolution. Increasing the resolution (higher resolution with smaller size of grid cells) of a DEM will not increase its accuracy in representing the continuous surface (Burrough and McDonnell 1998). However, decreasing the resolution of a DEM will cause data lost and consequently decrease the accuracy of the elevation data to represent the continuous surface. This rule also applies to elevation point clouds and therefore, in case elevation point clouds are used as the elevation data, the planimetric interval between two neighboring 3-D points should be approximately the same as the average density of the point clouds.

6.4.3 Construction of 3-D Mathematical Model

As stated earlier, in order to construct a 3-D mathematical model it is still required to obtain a series of 3-D points either with X/Y/Z-coordinates or distance measure and Z-coordinate. Therefore, the guidelines and procedures stated in Section 6.4.2 still hold when constructing a 3-D mathematical model for liner objects. With the 3-D points available, mathematical functions in the format of $Z = f(X, Y)$ plus $Y = f(X)$ or $Z = f(\text{Distance})$ can be derived to best fit those 3-D points in a three-dimensional space. After the 3-D mathematical model is constructed, 3-D points can be generated from this model by applying the mathematical functions. These 3-D points can later be deployed to aid 3-D analyses such as predicting 3-D distances.

6.5 DISTANCE PREDICTION

After constructing the 3-D model for linear objects, either a point model or a mathematical model, this 3-D model can be deployed to predict the 3-D distance for the linear objects without field efforts. As stated earlier, the 3-D distance is the actual distance or surface distance of a linear object, not the planimetric or 2-D distance. While many technologies such as ground surveying, GPS, and DMI could be used to obtain 3-D distances, they are time-consuming and labor intensive, requiring tremendous field efforts. With our 3-D model approach, the 3-D distance for linear objects can be computed by computers and GIS programs without field efforts and therefore, it is an efficient approach that will save a lot of time and labor costs.

With our 3-D model, the approach of predicting distance is similar to the approximation concept described in Section 5.3.4 and Figure 5.11. In other words, first, 3-D points with X/Y/Z-coordinates are obtained from the 3-D point models or generated from the mathematical models. Second, the 3-D distance between two neighboring 3-D points is computed using Equation (5.3) herein.

3-D distance $d1 = \text{sqrt}((X1-X2)^2 + (Y1-Y2)^2 + (Z1-Z2)^2)$, where $X1/Y1/Z1$ are the X/Y/Z coordinates for one point and $X2/Y2/Z2$ are the X/Y/Z coordinates for the other point (5.3)

These 3-D distances of neighboring 3-D points are added together to obtain the 3-D distance for the complete linear object.

In case the variations of 3-D point model or 3-D mathematical models are used, i.e. LRS based approach, the 3-D points are located with planimetric distance and Z-coordinate, the 3-D distance between two neighboring 3-D points is obtained via Equation (5.4) as bellow.

3-D distance $d1 = \text{sqrt}((D1-D2)^2 + (Z1-Z2)^2)$, where $D1$ and $D2$ are the planimetric distances from the first point and the second point to the start node of the linear object, respectively, $Z1$ and $Z2$ are elevations for the first and second points, respectively, (5.4)

The 3-D distances are added together to obtain the 3-D distance for the complete linear object. In terms of accuracy in distance prediction, the variation models and Equation (5.4) are preferred because this approach takes into consideration the planimetric curvature of the linear objects. Because of this, theoretically, the 3-D distance computed using Equation (5.4) will be longer and more accurate than the 3-D distance computed using Equation (5.3) for the same linear object. Therefore, the computational model and implementation will focus on this variation model and Equation (5.4) as will be detailed in chapter 7.

7 COMPUTATIONAL IMPLEMENTATIONS OF 3-D MODELS

Based on the conceptual data models and the needed data sources represented in Chapter 6, this chapter provides computational implementations to represent linear objects in a three-dimensional space and realize three-dimensional analysis, more specifically, distance prediction. This chapter describes the development environment, how to derive 3-D points depending on different input elevation data sources, how to derive mathematical functions, and computing algorithms. In addition, the accuracy assessment of this 3-D spatial modeling approach is briefly discussed.

7.1 DEVELOPMENT ENVIRONMENT

As stated earlier, many commercial GIS packages are 2-D GISs. In order to carry out 3-D spatial modeling and 3-D analysis for specific applications, extensive customizations and development efforts are required. The computational model for representing transportation linear objects in a three-dimensional space for this research will be developed using ArcGIS and ArcObjects from Environmental Systems Research Institute, Inc. (ESRI), and Visual Basic as the programming language from Microsoft, which is built into ArcGIS. This decision is made based on the availability and capability of ArcGIS, ArcObjects, and Visual Basic for Applications (VBA), and the researchers' knowledge in these software packages.

ArcGIS is the desktop version of GIS software package developed by the vendor of ESRI. It is one of the most commonly used GIS software packages across the world. ArcGIS desktop consists of ArcMap, ArcCatalog, and ArcToolbox, and extensions (ArcGIS Desktop Help 2002). The first three components are the core components of ArcGIS (ArcGIS Desktop Help 2002). ArcCatalog is a windows-style program that enables the users to browse, preview, import and export GIS data, and connect to personal geodatabase and internet database server (ArcGIS Desktop Help 2002). ArcMap is the premier application for desktop GIS and mapping. It enables users to visualize, create, and analyze spatial data (ArcGIS Desktop Help 2002). ArcToolbox is the application that provides access to all ArcGIS analysis functions with powerful geoprocessing capabilities (ArcGIS Desktop Help 2002).

ArcObjects is the development platform for the ArcGIS family of applications such as ArcMap, ArcCatalog, and ArcToolbox. It is built using Microsoft's Component Object Model (COM) technology (Exploring ArcObjects 2001). While ArcGIS has limited spatial analysis functionality, the full geoprocessing and spatial analysis functionality can be realized via customizing and extending ArcGIS desktop using ArcObjects (Exploring ArcObjects 2001).

ArcObjects deploys the concepts of object-oriented programming and can be used in VBA that is built into ArcGIS. This research will develop 3-D models and carry out 3-D analysis of transportation linear objects by programming with VBA while deploying ArcObjects to develop customized programs. This is a popular way available to GIS developers to realize many spatial analysis functions that extend the functionality of ArcGIS desktop to carry out application-specific analysis tasks such as 3-D modeling and 3-D analysis in this research. In addition, the ArcGIS family software packages including ArcView, ArcEditor, and ArcGIS were developed via this approach (Exploring ArcObjects 2001) and therefore, programs developed in this research will be compatible with these software packages and can be readily utilized with other applications developed using ArcGIS family software packages.

ArcGIS desktop and ArcObjects development environments are available to students at North Carolina State University. Thanks to the Earth Observation Center at North Carolina State University, the researcher obtained the required license to install and run ArcGIS and ArcObjects with a laptop. The researcher is very interested in developing customized extensions to ArcGIS desktop via the ArcObjects development platform and has been using ArcGIS desktop package for more than two years.

To summarize, the computational models and algorithms for the 3-D spatial modeling and distance prediction of transportation linear objects of this research will be implemented via programming with VBA, utilizing ArcObjects and ArcGIS desktop.

7.2 OBTAINING 3-D POINTS ALONG LINEAR OBJECTS

Based on the conceptual 3-D model, the critical part of our 3-D modeling approach to represent linear objects in a three-dimensional space is to obtain a series of 3-D points along the linear objects. This section describes how to obtain the planimetric position and the elevation value for these points, or how to generate a series of 3-D points along the linear objects using the existing data sources.

7.2.1 Obtaining Planimetric Position

With ArcGIS, transportation linear objects are modeled and stored as linear features in shapefiles, coverages, or geodatabases. Whatever format that data might be, they could all be loaded into ArcMap and displayed. As illustrated in Figure 7.1, when the mouse is pointed to a point on a line, the status bar (bottom) shows the X/Y coordinates of that point.

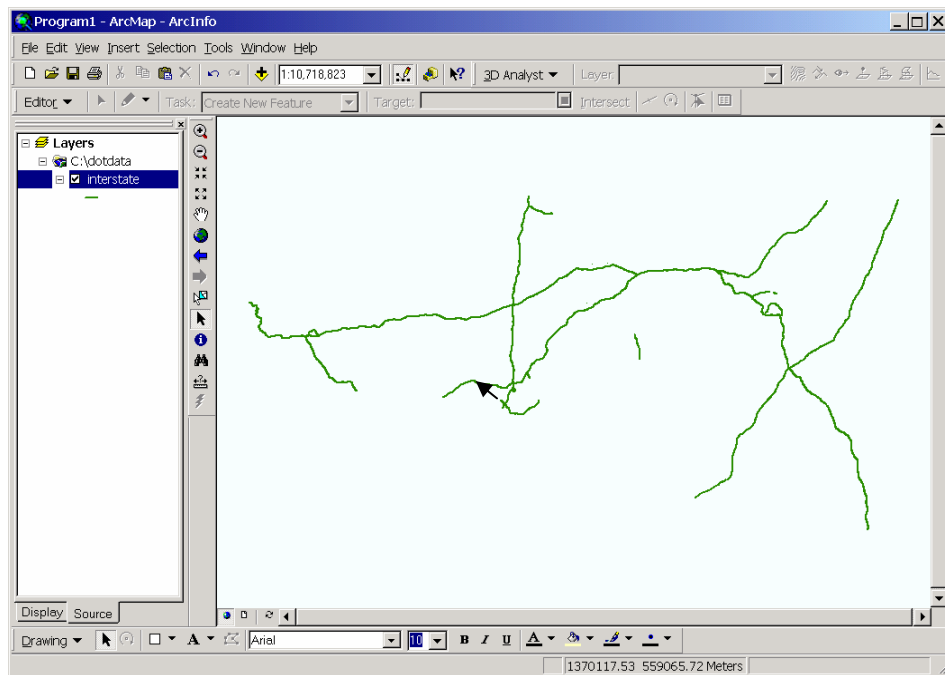


Figure 7.1 ArcMap Displaying X/Y Coordinates for a Point on a Line

In terms of programming with VBA and ArcObjects, three methods are related. The first one is the *QueryPoint* method, which provides access to the properties of a point including its X/Y-coordinates, along a linear object at a specified distance from its beginning. The second one is the *QueryPointAndDistance* method, which finds the closest point on the linear feature to the input point feature and determines properties of the found point such as its distance to the beginning of the linear feature along the linear feature and its X/Y-coordinates. The third is the *Touches* method, which determines if a given point is on a linear feature.

Figure 7.2 illustrates the spatial properties related to these methods. The linear object starts from point A and ends at point B. Point E touches or lies on the linear object. Point C does not touch the linear object. Point D is the point on the linear object which has the closest planimetric distance (d_1) to point C. The planimetric distance along the linear object is d_2 from point D to the start node A and is $d_2 + d_3$ from point E to the start node A.

For the given point E along the linear feature, there are two ways to spatially locate it. It can be located by its X/Y-coordinates or by its distance to the beginning of that linear feature along the linear feature. In case the point is located via its X/Y-coordinates, the *Touches* method will return a value of true and the *QueryPointAndDistance* method returns the distance from that point along the linear feature to the beginning of that linear feature. In case the point is located via its distance to the beginning of a linear feature along that linear feature, the *QueryPoint* method can be used to determine its X/Y-coordinates. This capability of ArcGIS proves that it is feasible and simple to convert between two different 2-D locating approaches, which were stated earlier.

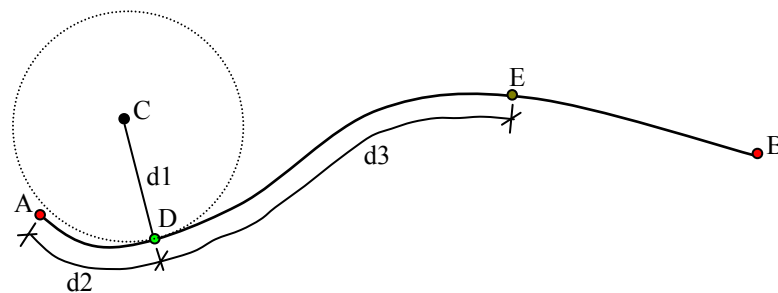


Figure 7.2 Illustration of the Spatial Properties and Spatial Relations for Points and a Linear Object

7.2.2 Obtaining Elevations

According to the conceptual 3-D models for transportation linear objects, the critical part is getting elevation values for a given point on a linear feature, whose planimetric position is given in either X/Y-coordinates or its distance along that linear object to its beginning. This section describes how to determine the elevation for a given point when referring to the underlying elevation datasets in two formats, point clouds and DEMs.

7.2.2.1 Obtaining Elevations from LIDAR Point Clouds

LIDAR point clouds refer to the elevation data that are in the format of numerous 3-D points with X/Y/Z-coordinates stored in text files. These 3-D points are generally neither uniformly distributed on the planimetric plane, nor in the three-dimensional space. These 3-D point clouds can be imported into ArcGIS as point features with Z-coordinate as the attribute of elevation.

Even though LIDAR point clouds have relatively high density (about 20ft between two neighboring points in the case of North Carolina Flood Mapping Program), it is rarely the case that many points would spatially lay on a linear feature. More specifically, it is rarely the case that LIDAR points would lay on the road centerline, even though many would lay on the road pavement. In order to obtain elevations for points (with unknown elevations) along a linear object from LIDAR point clouds (with available elevations), interpolations or approximations are needed.

Uniformly gridded DEMs could be produced based on the LIDAR point clouds via interpolations in post-processing. In other words, the DEMs are interpolated from the data points. Later when being used in 3-D analysis, the elevation for a specific point will have to be interpolated from the DEMs again. The question becomes why not use LIDAR points directly, or why should we use LIDAR DEMs that have gone through interpolations already? Since the elevation for a specific point can be interpolated from the LIDAR data points directly instead of going through one more interpolation as in case of DEMs, it is believed by the researchers that obtaining elevations from LIDAR data points directly is an innovative way to utilize LIDAR point clouds, is efficient, and is more accurate.

There are many different ways to obtain elevations for points along a linear feature, which vary in the interpolations and approximations. Two different approaches of obtaining elevations from LIDAR point clouds for points along a linear feature are described as bellow. The first approach adopts an interpolation method and is general enough to be applied to points along any linear objects. The second approach adopts an approximation method. It takes into consideration of the average density of LIDAR point clouds and the geometric properties of transportation linear objects. It is designed specifically to be utilized for transportation linear objects and will be used in the case study.

7.2.2.1.1 Interpolation Approach

This approach works in the way described as below.

First, for a linear object, a buffer with a certain size is applied. All LIDAR 3-D points within this buffer will be selected to obtain elevations for the points along this line segment. The purpose of doing this is to find LIDAR 3-D points that are relatively close to the linear object so that elevations from these 3-D points can be interpolated to obtain elevations for points along that linear object. The buffer size should be big enough to contain sufficient LIDAR 3-D points while small enough so that only those LIDAR 3-D points that are close to the linear object will be included.

Second, based on the spatial relationships among those LIDAR 3-D points and the line segment, the LIDAR 3-D points within the buffer are categorized into 3 groups, group A, B, and C as illustrated in the Figure 7.3. Group A includes all points on one side of the line, group B includes all points on the other side of the line, and group C includes all points lay exactly on the line, which is rarely the case.

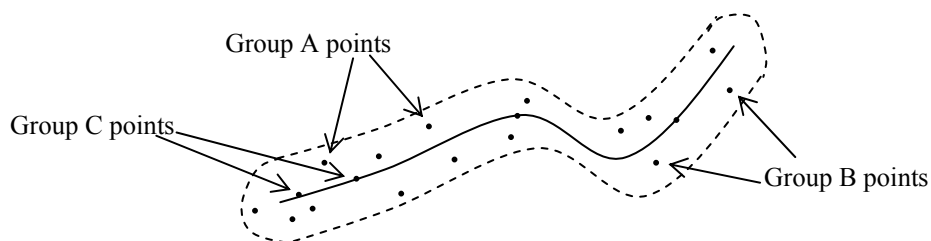


Figure 7.3 Three Groups of LIDAR Points Based on the Spatial Relationships

Third, for each point P in group A, its planimetric distances to all points in group B are determined. The point Q in group B, which has the shortest planimetric distance to point P will be identified. Repeat this step for all points in group A to identify their corresponding points in group B with the shortest planimetric distances. After this step, a few pairs of points are obtained. Each of these point pairs are connected with straight lines that intersect with the line segment as illustrated in Figure 7.4. The elevations for these intersections are linearly interpolated from the elevations of the corresponding points. The purpose of doing these is to find lines that connect two LIDAR 3-D points and intersect with the linear object so that the elevations for those intersections can be linearly interpolated from the corresponding pair of LIDAR 3-D points.

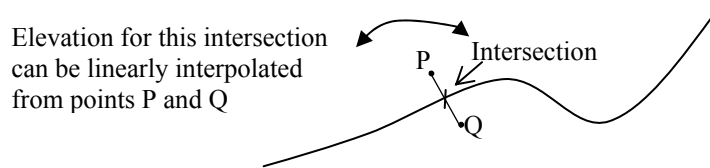


Figure 7.4 Illustration of Obtaining Intersections and Determining Their Elevations

Forth, all points in group C will be used together with those intersections identified in the previous step as 3-D points to represent the linear object.

Fifth, for the start node and the end node of the linear object, their elevations can be obtained by taking an (weighted) average, a maximum, or a minimum from the three or four closest LIDAR 3-D points based on the planimetric distances. Their elevations can also be obtained via interpolations from the surrounding LIDAR 3-D points.

Finally, the points used to represent this line segment in a 3-D space include all these intersections, start and end nodes, and all points in group C.

7.2.2.1.2 *Approximation Approach*

The interpolation approach described in the previous section can be applied to all linear objects when obtaining 3-D points along the linear objects from LIDAR 3-D points. This section describes another approach which adopts an approximation method while taking into consideration of the density of LIDAR 3-D points and the geometric properties of the transportation linear objects in North Carolina.

The steps taken in the approximation approach are described as bellow.

First, a buffer with a certain distance is applied to the road centerline. All LIDAR 3-D points contained in this buffer are obtained. The size of the buffer should be half of the size of the typical lane width so that this buffer would be an approximate representation of the road pavement.

For those LIDAR 3-D points in the buffer, they are snapped to the road centerline. More specifically, each LIDAR 3-D point identifies its shortest planimetric distance to the road centerline and the corresponding point along the road centerline. This point along the road centerline takes its X/Y-coordinates from the road centerline and its Z-coordinate from the LIDAR 3-D point.

After this approach, many 3-D points along the road centerline are obtained with X/Y/Z-coordinates. It is optional to obtain Z-coordinates for the start and end nodes because sufficient number of 3-D points will be obtained and the start and end nodes could be located in the three-dimensional space by examining the trend of the linear object represented by those 3-D points.

This approximation approach might be more accurate than the interpolation approach as mentioned earlier. One of such scenarios is illustrated in Figure 7.5, in which the vertical error due to snapping is smaller than the vertical error due to interpolation.

If the errors introduced into the 3D model due to this approximation are minor effects when compared to the vertical accuracy of LIDAR points (25 cm as specified in the North Carolina Flood Mapping Program), the snapping approach could be desirable.

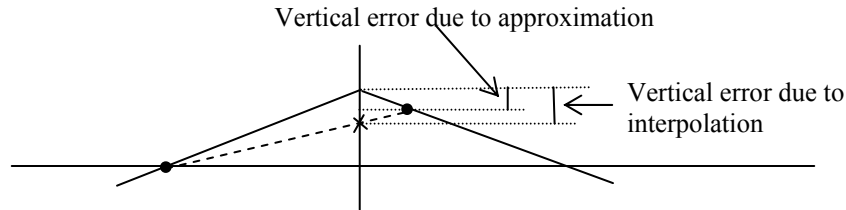


Figure 7.5 Cross-Sectional View of Comparing the Vertical Errors Due to Interpolation and Approximation

Transportation linear objects in this research, or roads, are three-dimensional objects in space with geometry properties such as horizontal and vertical alignments and the cross section. A road segment will have one or more lanes. The typical lane width for Interstate highways and US Routes in North Carolina is 12ft per lane and 10 ft for NC Routes. In case of two-lane Interstate highways, the width of pavement would be more than 24ft when taking into consideration of the paved shoulder. The maximum cross-sectional slope for Interstate Highways and US Routes in North Carolina is 8% and 4% for NC Routes. However, in a flat area like the eastern part of North Carolina, it would be reasonable to assume a 2% typical cross-sectional slope. This information is based on an informal conversation with Dr. Hummer from the Civil Engineering Department of North Carolina State University and Jay Bennett, the Design Engineer from NCDOT.

Since roads are man-made objects with smooth surface and small cross-sectional slope (typically 2%, maximum 8% for Interstate highways and US Routes), it is believed that taking the elevations of LIDAR 3-D point that are on the pavement surface and are close to the road centerline to approximate the elevations along the road centerline will provide a more accurate way to model the linear objects in 3-D after small adjustments. The maximum error due to the snapping approach with a typical 8% cross-sectional slope and a 12 ft lane width would be 0.96 ft (29.26 cm). In the case of a 2% typical cross-sectional slope and a 10 ft lane width, the maximum error due to the snapping approach would be 0.2 ft (6.10 cm). Regarding the geometry characteristics of road surface, the interpolation approach might be overdue. The approximation approach takes LIDAR points that are within a certain distance (planimetric) from the road centerline and snap these points to

road centerline. After snapping, these points obtain new X/Y-coordinates from the centerline while keeping its original Z-coordinate. Figure 7.5 provides a cross-sectional view, in which, the error of the interpolation approach is more significant than the error of the approximation approach, when the two input LIDAR points and the corresponding point on the road centerline are aligned in such a way that all three points are on a cross section that is vertical to the road centerline.

The situation illustrated in Figure 7.5 is rarely the case because usually the two input LIDAR points and the corresponding point on the road centerline form a surface that is vertical to the ground, but not a cross section that is vertical to the road centerline. Figure 7.6 demonstrates a more common situation.

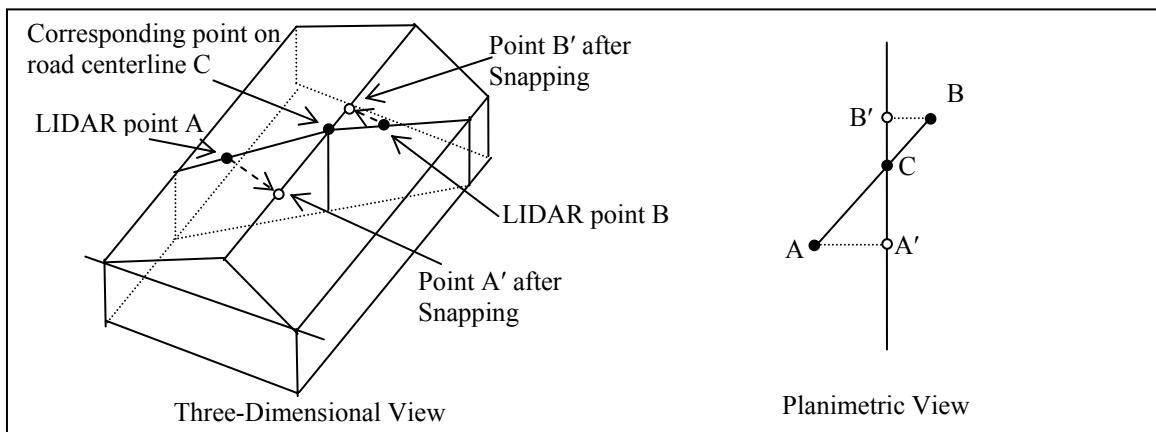


Figure 7.6 Comparison of the Interpolation Approach with the Approximation Approach

In Figure 7.6, assuming LIDAR points A and B have elevation values e_1 and e_2 , respectively, the interpolation approach will obtain the elevation value for point C as $[e_1 + (e_2 - e_1) * (\text{distance of AC}/\text{distance of AB})]$, where the distances are planimetric distances. The approximation approach will obtain two 3-D points along the road centerline, point B' (with elevation e_2) and point A' (with elevation e_1). However, if the elevation at point C were to be interpolated from points A' and B', it would be $[e_1 + (e_2 - e_1) * (\text{distance of AC}/\text{distance of AB})]$. It is obvious that $(\text{distance of AC}/\text{distance of AB})$ equals $(\text{distance of AC}/\text{distance of AB})$ and consequently, the elevation obtained for point C would be the same in both interpolation and approximation approaches. However, this is a simplified demonstration. It proves that the approximation approach would

provide at least the same accuracy, if not higher accuracy, as the interpolation approach. Furthermore, the approximation approach would generate as much as double number of 3-D points along the road centerline as that of the interpolation approach when the same set of LIDAR points are used. Generally, the more 3-D points along the road centerline, the more accurate will the 3-D model be when using a series of lines connected to each to model the 3-D line. Because of these reasons, the approximation approach will be adopted in this research to obtain 3-D points along the road centerline from the LIDAR points.

In addition, one of the prerequisites of using this snapping approach is that there will be sufficient LIDAR points on the road pavement. LIDAR 3-D points have a relatively high density, or there is a quite short planimetric distance between two neighboring LIDAR 3-D points. It is very possible that many of the LIDAR 3-D points will be on the road pavement. This is the case for our research study by examining the density of the LIDAR 3-D points (downloaded from www.ncfloodmaps.com) available for the Johnston County in North Carolina. Figure 7.7 illustrates the number of LIDAR points in the total of 263 tiles that completely cover the Johnston County. Each cell has the same dimension of 10,000ft * 10,000ft. Six data cells have no data available, which reduces the number of data cells from 263 to 257. The total LIDAR data points in these 257 data cells are 74394066. Taking the average values, there are 289471.07points/cell and 0.002895 points/ft*ft. Assuming these points are evenly distributed, the average distance between two neighboring LIDAR data points would be:

- (1) Cell with the lowest density: 31.89ft,
- (2) Cell with the second lowest density: 24.17ft,
- (3) Cell with the highest density: 16.73,
- (4) Cell with the second highest density: 16.77,
- (5) Average distance between two neighboring points is 18.59 ft = $\sqrt{\text{total area/total points}}$, or
- (6) Average distance between two neighboring points is 18.7 ft = average $\sqrt{\text{one cell area/points in this cell}}$.

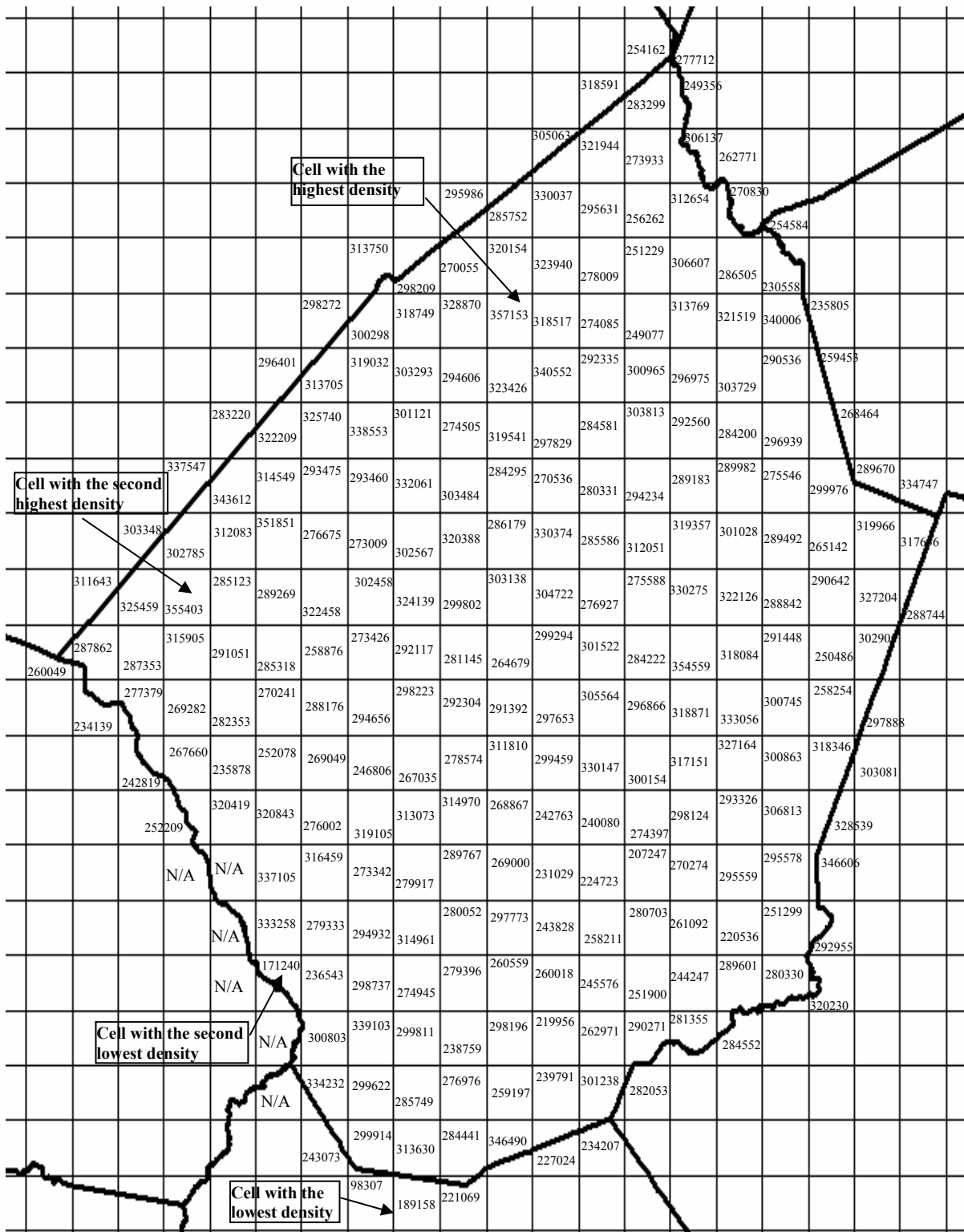


Figure 7.7 Illustration of the Density of LIDAR Points in Johnston County, North Carolina

7.2.2.2 Obtaining Elevations from LIDAR DEM

In addition to LIDAR point clouds, DEMs can be generated based on these point clouds by post-processing. When displayed and analyzed in GISs, these DEMs are in the same format as USGS DEMs and therefore, obtaining elevations from LIDAR DEMs follows the same procedure of obtaining elevations from USGS DEMs, which is described in the next subsection.

7.2.2.3 Obtaining Elevations from USGS DEMs and NED

The U.S. Geological Survey (USGS) is the federal agency that is in charge of developing elevation datasets in the United States. The most popularly known elevation data developed by USGS are DEMs. According to USGS, a DEM is the digital cartographic representation of the elevation of the terrain at regularly spaced intervals in X and Y directions, using Z-values referenced to a common vertical datum (Maune et al. 2001). However, grid spacing, datum, coordinate systems, data formats, and other characteristics may vary widely, but normally following alternative specifications, with narrow grid spacing and State Plane coordinates for example.

NED is a seamless raster data derived from source DEMs with correcting procedures and data quality control as described earlier. NED is essentially a DEM and therefore, the procedure of obtaining elevations is the same when working with DEMs and NED.

Both DEM data and NED data can be imported into ArcGIS in the grid format. The elevation for a specific point can be derived via a bilinear interpolation approach as illustrated in Figure 7.8, in which the elevation for point E is interpolated from the four surrounding points A, B, C and D.

In order to construct 3-D point model for transportation linear objects when using DEMs to provide elevation data, the 3-D points can be derived by creating a series of points along a linear object starting from its start node with a uniform planimetric interval along the linear object as illustrated in Figure 7.9 in which, four points are deployed including the start point A, end point B, and two intermediate points C and D. The planimetric

distance from C to A along the linear object is the same as the planimetric distance from D to C along the linear object. It is usually the case that the planimetric distance d' from B (the end point) to D (the last intermediate point) will be less than the uniform interval d being used to generate intermediate points because it is rarely the case that a linear object will have its planimetric distance as $n * \text{interval}$, where n is 1, 2, 3, For all these points, their X/Y-coordinates are determined by the linear object while their Z-coordinates can be interpolated from DEMs according to Figure 7.8.

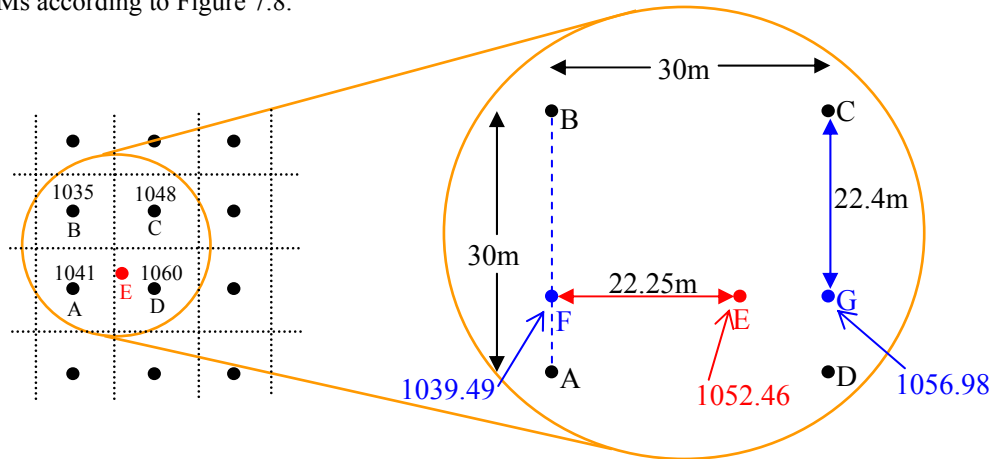


Figure 7.8 Bilinear Interpolation of Obtaining Elevations from DEMs (Rasdorf et al. 2003b)

The interval d being used to locate the intermediate points along the linear objects is dependent on the resolution of the DEM data being used. As a general rule, the optimum interval d would be the same as the resolution which is often represented by one dimension of the cell (for example, 30meters) of the DEM being used in order to achieve a reasonable accuracy and efficiency. As stated earlier that deriving a higher resolution DEM from a lower resolution DEM will not help improving the accuracy while deriving a lower resolution DEM from a higher resolution DEM loses information and therefore, affects the accuracy. This can be applied to determine the interval d that could be used to obtain the intermediate points along the linear objects. Obviously, the optimum interval d should be the same as the resolution of the DEM.

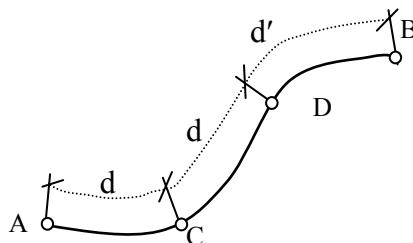


Figure 7.9 Obtaining Intermediate Points along a Linear Object

Finally, a number of 3-D points with X/Y/Z-coordinates along the linear objects are obtained as a result. In case NED dataset is used, the same procedure applies.

7.2.3 Deriving Mathematical Function and Adjusting 3-D Points

Whether LIDAR point clouds, DEMs, or NEDs are used to obtain elevation values for points along a linear object, the results are the same, i.e. a series of 3-D points with X/Y/Z-coordinates along the linear objects. In other words, the 3-D point model for linear objects is constructed. However, the difference between the 3-D point model constructed from LIDAR point clouds and that from DEMs and NEDs is that in the latter cases, the planimetric distances along the linear object between two neighboring points (except the end node) on the linear object are uniform, which is not the case when the 3-D point model is constructed using LIDAR point clouds.

As stated in the conceptual mathematical models, with a set of 3-D points with X/Y/Z-coordinates, mathematical functions in the format of $Z = f(X, Y)$ and $Y = f(X)$ can be obtained to fit these 3-D points. In case the 3-D points are associated with distance along the linear object and Z-coordinate, a mathematical function in the format of $Z = f(\text{distance})$ could be obtained. In our 3-D spatial modeling for linear objects, if DEMs are used to obtain elevations for points along the linear object with a uniform interval, no mathematical function will be derived in order to obtain the best-fitting line. In this case, the elevations are interpolated from DEMs and they are the best that could be derived from DEMs. However, in case LIDAR point clouds are used, an approximation approach is adopted, which indicates there is space for subtle adjustment. Figure 7.10 shows an example of deriving mathematical function from a set of 3-D points. The mathematical function is with an order of n : $Z = a_n * D^n + a_{n-1} * D^{n-1} + a_{n-2} * D^{n-2} + \dots + a_2 * D^2 + a_1 * D^1 + a_0$, where a_n , a_{n-1} , ...and a_0 are constants for a specific road segment, D is the distance from a point on the line to the start point along the line, and Z is the elevation of that point. The best-fitting line (defined with this function) will not pass all sample points. As stated earlier, there are many criteria in determining the best-fitting line. One example for defining the best-fitting line herein is such that (1) assuming there are n 3-D points on a specific linear object, (2) each point is located via distance and elevation, (3) the fitting line is described in the format of $Z = a_4 * D^4 + a_3 * D^3 +$

$a_2 * D^2 + a_1 * D^1 + a_0$, or order $n = 3$, (4) based on the distance value (D), new elevations for each of those 3-D points is derived from the line function, and (5) the total of the difference between the original elevation and the new elevation for all points are minimized.

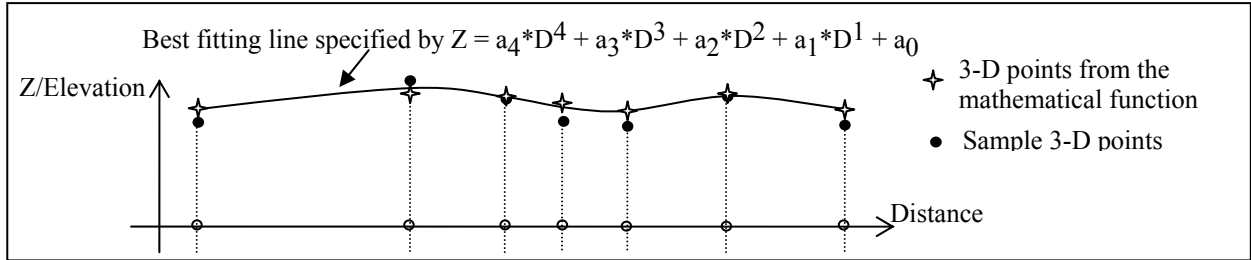


Figure 7.10 Defining Best-Fitting Line

The illustration here indicates the theoretical possibility of deriving a mathematical model from a set of 3-D points. However, since there are an infinite number of potential best-fitting lines and there is no such agreement as what criteria should be used in determining the optimum best-fitting line, this mathematical model will not be constructed and tested in this research. Theoretically, for a given set of 3-D points, a mathematical function at an infinite high order would enforce the defined best-fitting line to pass through all those points.

To summarize, in case DEMs or NEDs are used to obtain elevations for those points on a linear object, which are uniformly dispersed along the linear object (the planimetric distances along the linear object between two neighboring points are the same except for the distance between the end point and the last intermediate point), it is unnecessary to derive mathematical functions. These uniformly dispersed 3-D points will be used directly to represent the linear object in the three-dimensional space and to predict the 3-D distance.

In case LIDAR point clouds are used to obtain a series of 3-D points to approximate the road centerline in a three-dimensional space, a mathematical function could be derived. This mathematical function describes one of many potential best-fitting lines for these 3-D points. This best-fitting line may or may not pass every single input 3-D point. While this mathematical model represent a linear object in a three-dimensional space, it is inconvenient to use this function directly to carry out 3-D visualization and 3-D analysis tasks such distance prediction. A further step could be taken to obtain a series of 3-D points that are uniformly dispersed along the

linear object in a planimetric plane. For example, if the mathematical function is $Z = f(D)$, the interval is specified to be d , and assume the planimetric length of the linear object is $7.34d$, then a set of 9 3-D points can be generated, they are: $(0, f(0))$, $(d, f(d))$, $(2d, f(2d))$, $(3d, f(3d))$, $(4d, f(4d))$, $(5d, f(5d))$, $(6d, f(6d))$, $(7d, f(7d))$, and $(7.34d, f(7.34d))$.

The purpose of generating 3-D points from the mathematical functions when LIDAR point clouds are utilized to obtain elevations is to obtain a set of 3-D points that are uniformly dispersed along the linear object as in case of utilizing DEMs to obtain elevations and 3-D points along a linear object. However, in 3-D modeling and analysis, these newly generated 3-D points will not improve the accuracy because the mathematical function is derived from the initial 3-D points along a linear object and those newly generated 3-D points come from this mathematical function. Therefore, in this research, this mathematical function and adjustment are not taken. Rather, the initial set of 3-D points will be used to represent the linear object in a three-dimensional space and to predict 3-D distance.

Another adjustment could be applied when LIDAR point clouds are used to provide elevations. This adjustment takes into consideration of the typical cross-sectional slope of roads as illustrated in Figure 7.11. In the case road scenario 1 (centerline is the highest place) with typical cross-sectional slope of 2% and lane width of 12ft, the maximum adjustment would be $12 * 2\% = 0.24$ ft. In case of road scenario 2 with typical cross-sectional slope of 2% and lane width of 12ft, the maximum adjustment would be $\pm (12 * 2\%) = \pm 0.24$ ft. These maximum adjustments happen only when the LIDAR points are along the outer edge of the lane. Compared to the vertical accuracy of LIDAR points of the NC Flooding Mapping Program (25cm, approximately 0.82ft) (NCCTS 2003), these maximum adjustments are ignorable. Therefore, this adjustment will not be adopted, either. However, these subtle adjustments are worth further research efforts.

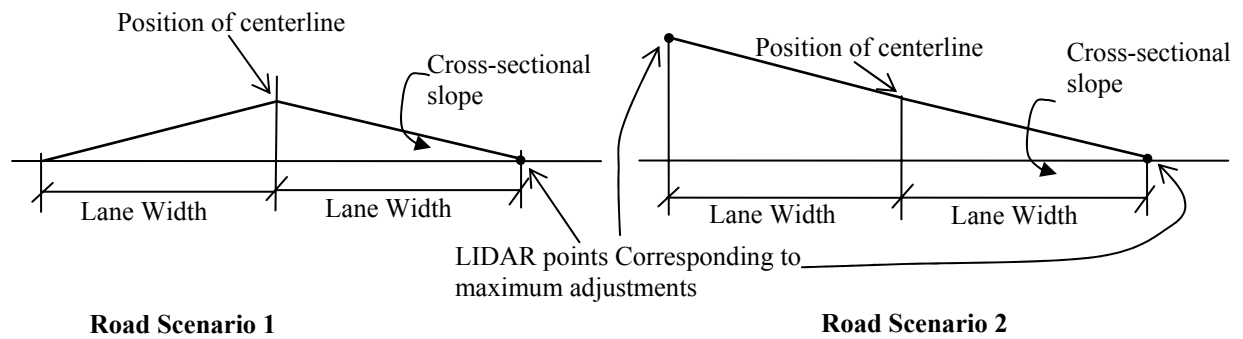


Figure 7.11 Maximum Vertical Adjustments

7.3 PREDICTING 3-D DISTANCE

Based on the conceptual model of 3-D distance prediction, the generated 3-D points will be used to calculate the three-dimensional distance for a linear object according to the concept illustrated in Figure 5.11 and Equations (5.3) and (5.4). As stated earlier, Equation (5.3) computes the three-dimensional distance by using a series of straight lines connected to each other at the 3-D points. Equation (5.4) takes a similar but more accurate approach by taking into consideration the planimetric geometric characteristics (curvature, for example) of a linear object. The difference between these two equations is which planimetric distance is used. Equation (5.3) uses the straight line distance between two neighboring points. Equation (5.4) uses the planimetric distance along the linear object following its curvature. In this research, Equation (5.4) is adopted. In other words, with those 3-D points with X/Y/Z-coordinates, the first step in distance prediction is to convert each 3-D point into distance/Z-coordinate format. For each 3-D point, its X/Y-coordinates are used to determine its planimetric distance along a linear object to the beginning of that linear object. After this step, Equation (5.4) will be used to predict its three-dimensional distance via the introduction of the third dimension – elevation.

7.4 ALGORITHMS

This section describes the complete procedure to carry out 3-D spatial modeling and 3-D distance prediction from road network data (polyline feature) and elevation data (LIDAR points, DEMs, and NEDs), from preprocessing to analysis.

7.4.1 Pre-Processing

The datasets being used in 3-D spatial modeling and analysis come from different data sources. The road network data come from state transportation agency (Department of Transportation). The LIDAR elevation datasets come from NC Flooding Mapping Program. The NED data come from USGS. These datasets were developed using their own projection method, datum, coordinate system, and units to meet their own purposes. It is a very common scenario in GIS that the datasets are developed with different spatial referencing systems. Therefore, it is critical for any GIS analyses that the datasets being used must be converted to the same spatial referencing system.

In this research, the 3-D spatial modeling and analysis are applied to the road network data by using different elevation datasets. Taking into consideration of the data format, data amount, and the complexity in converting data from one spatial referencing system to another, it is decided that the road network data will be reprojected and converted into the same spatial referencing system being used by the elevation datasets. In other words, when LIDAR points are serving as the elevation dataset, the road network data will be converted into the same spatial referencing system as being used by LIDAR points, and correspondingly, 3-D modeling and analysis will be carried out under that spatial referencing system. If NED is serving as the elevation data, the road network data will be converted into the spatial referencing system of the NED. Consequently, 3-D spatial modeling and analysis will be carried out under this spatial referencing system.

7.4.2 Processing

After preprocessing, sets of data with the same spatial referencing system are available (for example, road network data and LIDAR point data in the same spatial referencing system, road network data and DEM or NED data in the same spatial referencing system). This section provides the processing procedures with flow charts and descriptions.

7.4.2.1 Processing Procedure with LIDAR Point Data and Road Network Data

Based on the previous discussions, Figure 7.12 illustrates the processing procedure to model linear objects in a three-dimensional space and carry out 3-D distance prediction when working with LIDAR point data and road network data. The whole procedure is described as below.

- 1) From the road network data, an individual linear object is selected.
- 2) A buffer with the size of half of the typical road width is applied to this linear object.
- 3) The LIDAR points are overlaid with the buffer and all LIDAR points in the buffer are identified.
- 4) For each LIDAR point in the buffer, it is snapped to the linear object (i.e., find the closest point on the linear object) so that a set of points are obtained along the linear object. These points obtain their X/Y coordinates from the geometry of the linear object and Z coordinates from the corresponding LIDAR points. In addition, for each of these points, its planimetric distance (S) along the linear object from the start node of this linear object is obtained.
- 5) The start and end nodes of the linear object are identified. Start and end nodes obtain their X/Y coordinates from the geometry of the linear object. Their planimetric distances along the linear object to the start node are obtained. For the start node, that distance is 0 while for the end node, that distance is the planimetric length of that linear object. At this moment, their Z coordinates are assigned with the value of 9999 to indicate that their elevations will need to be interpolated later.
- 6) The elevations for the start and end nodes from step 5) are interpolated from neighboring points from step 4).
- 7) The points generated in step 4) are merged with the 3-D points from step 6) so that a set of 3-D points along the linear object, including its start and end nodes are obtained. The linear object can now be modeled in a three-dimensional space with these 3-D points.
- 8) For these 3-D points, their X/Y coordinates are converted into the planimetric distances along the linear object and therefore, each 3-D point is located via distance/Z-coordinate in a three-dimensional space.
- 9) Equation (5.4) is applied to these 3-D points to obtain 3-D distance for the linear object.
- 10) These steps are repeated for every individual linear object contained in the road network data. And finally all linear objects are modeled in a three-dimensional space and their 3-D distances predicted.

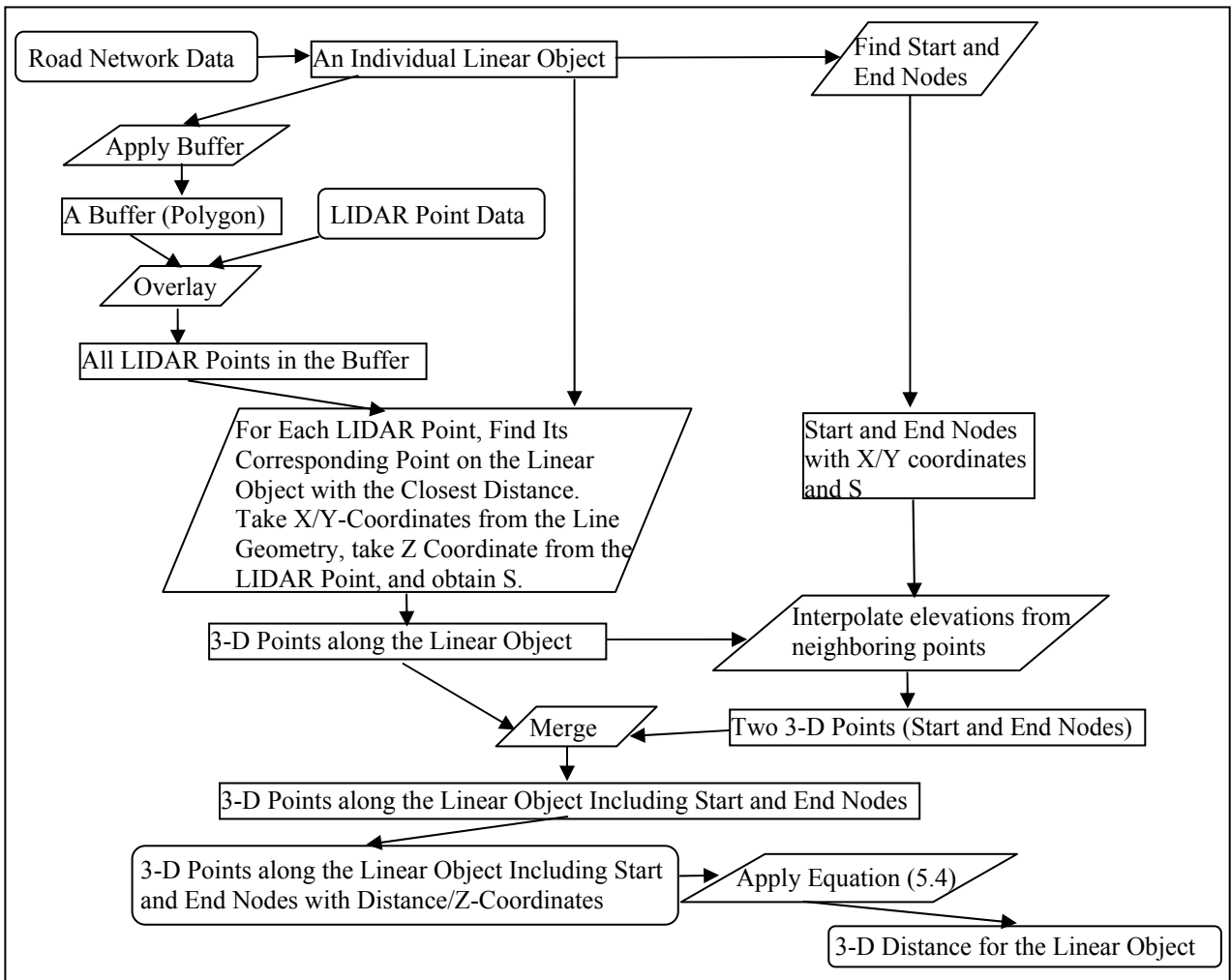


Figure 7.12 Processing Procedure for 3-D Spatial Modeling and 3-D Distance Prediction when Working with LIDAR Point Data and Road Network Data

7.4.2.2 Processing Procedure with DEM or NED Data and Road Network Data

Based on the previous discussions, Figure 7.13 illustrates the processing procedure to model linear objects in a three-dimensional space and carry out 3-D distance prediction when working with DEMs or NED elevation data and road network data. The whole procedure is described in the below.

- 1) From the road network data, an individual linear object is selected.
- 2) The start and end nodes of the linear object are identified.
- 3) Points that are uniformly distributed along the linear object with a specified planimetric interval along the linear objects are identified.

- 4) The points identified in steps 2) and 3) are merged together. These points obtain X/Y coordinates from the geometry of the linear object.
- 5) For each of these points, elevation value is interpolated from the DEMs or NEDs. After this step, a set of 3-D points along the linear object are obtained and could be used to represent the linear object in a three-dimensional space.
- 6) For these 3-D points, their X/Y coordinates are converted into the planimetric distances along the linear object and therefore, each 3-D point is located via distance/Z-coordinate in a three-dimensional space.
- 7) Equation (4) is applied to these 3-D points to obtain 3-D distance for the linear object.
- 8) These steps are repeated for every individual linear object contained in the road network data. And finally all linear objects are modeling in a three-dimensional space and their 3-D distances predicted.

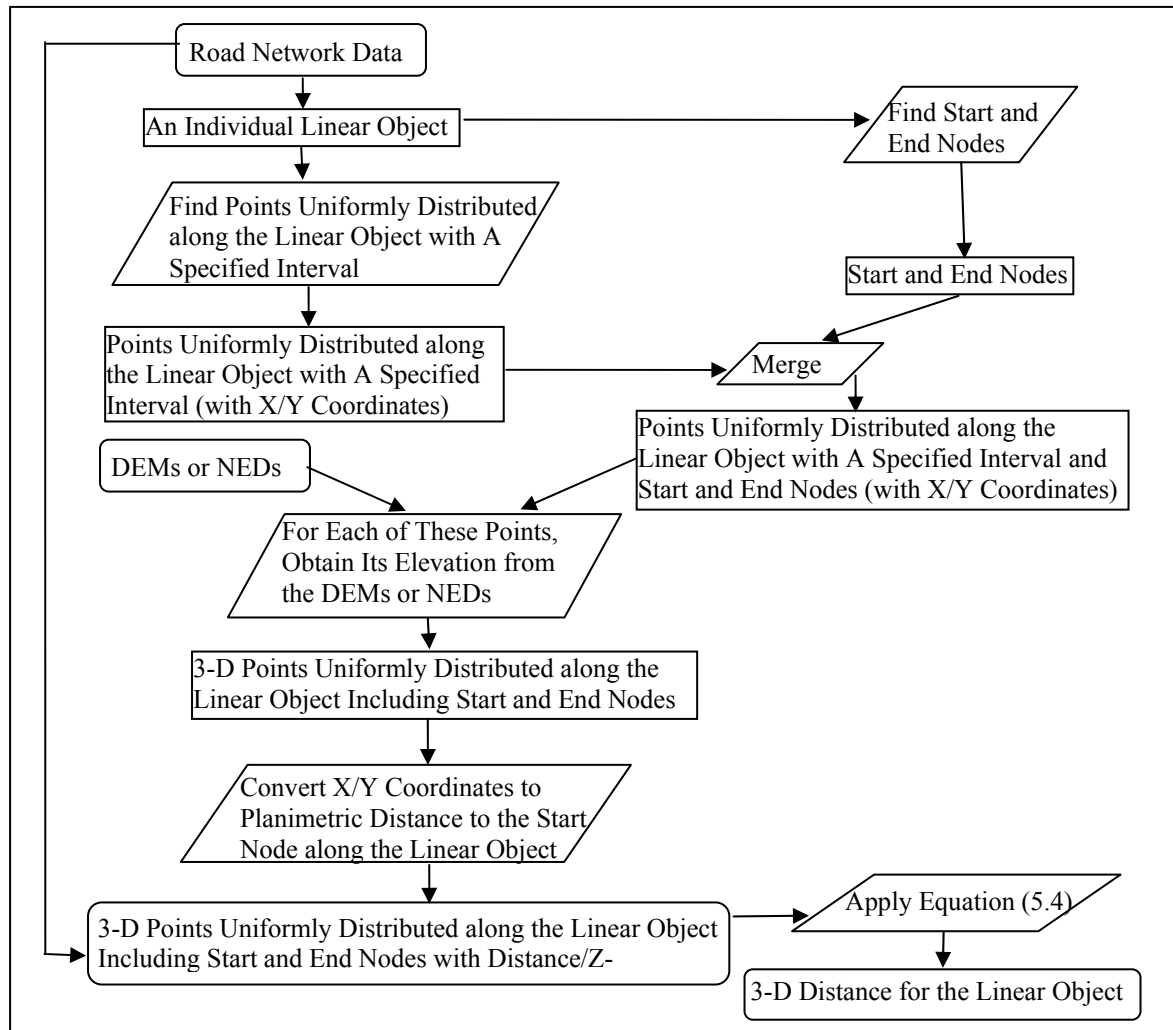


Figure 7.13 Processing Procedure for 3-D Spatial Modeling and 3-D Distance Prediction when Working with DEM Data

7.5 ACCURACY ASSESSMENT

With the newly developed 3-D spatial model to represent linear objects in a three-dimensional space and to predict their 3-D distance, the question becomes the accuracy of this 3-D model and the accuracy of the predicted 3-D distance.

The positional accuracy of the 3-D model is described as the spatial closeness between the 3-D model and its corresponding 3-D object being represented by this model. Since our 3-D model for a linear object is based on a set of 3-D points, the positional accuracy of the 3-D model can be evaluated by determining the positional accuracy of these 3-D points. Under ideal situation, the accuracy of each 3-D point can be assessed via the comparisons of its X/Y/Z-coordinates as in the 3-D model to its more accurate reference data with X/Y/Z-coordinates. For example, ground surveying or GPS can be deployed to obtain the X/Y/Z-coordinates for points. These X/Y/Z-coordinates from ground surveying or GPS become the reference data and serve as the more accurate data to evaluate the positional accuracy of the 3-D points in our 3-D model and consequently, the positional accuracy of our 3-D model in representing linear objects. The critical shortcoming of this approach is the identification of the corresponding point of a 3-D point of the 3-D model in the field. As illustrated in the previous discussions, the 3-D model developed in this research use a large amount of points along a linear object to represent it in a 3-D space. The start and end nodes could be unambiguously identified in the field with the help of road sections. However, for those intermediate points, identifying the corresponding points in the field is problematic (Goodchild and Hunter 1997, Hunter 1999).

An extension could be developed based on the concept that with errors a linear object is bell-shaped in space. Goodchild and Hunter (Goodchild and Hunter 1997, Hunter 1999) have proposed to estimate the percentage of the total length of the low accuracy representation of line that lies within a specified distance of the higher accuracy representation, which was illustrated in Figure 4.5. The extension works in terms of a 3-D space. It estimates the percentage of the total 3-D length of the low accuracy representation of line that lies within a specified distance (3-D buffer) of the higher accuracy representation. However, this approach works in a very

complicated way when being extended to a three-dimensional space. Furthermore, even in a 2-D space, there is no agreement as to the buffer size and the threshold value of the percentage. Its results are difficult to interpret and to utilize in engineering applications.

Examining our 3-D model and 3-D distance prediction reveals that the positional accuracies of the 3-D model and the accuracy of the 3-D distance prediction will be affected by the number of 3-D points, the planimetric intervals between two neighboring 3-D points, and the positional accuracies of these 3-D points. The positional accuracies of these 3-D points are determined by the positional accuracy of the planimetric road network data (linear features) and the vertical accuracy of the elevation data. It is obvious that the accuracy of the 3-D model and the accuracy of the 3-D distance prediction are closely related. Each is an indicator of the other. Generally, the higher the accuracy of the 3-D model, the higher the accuracy of the 3-D distance prediction, and vice versa.

As a surrogate, this research assesses the accuracy of the developed 3-D model by assessing the accuracy of its 3-D distance prediction via the comparisons to a more accurate dataset, the DMI measurements. As stated earlier, even though this is not a direct accuracy assessment of the 3-D model, the accuracy of its 3-D distance prediction is deemed as an indicator of the accuracy of the 3-D model and will be used to answer questions regarding the accuracy of the 3-D model.

8 CASE STUDY

This chapter describes a case study designed in the context of transportation to test and evaluate the 3-D spatial modeling and 3-D distance prediction models developed in this research when applied to transportation linear objects. Detailed information regarding the definition of the case study, its objective, the general approach, data sources and errors in the data source, and the steps taken to carry out 3-D spatial modeling and 3-D distance prediction, is provided in this chapter.

8.1 STUDY DEFINITION

The overall purpose of this case study is to test the 3-D spatial model and 3-D distance prediction models as proposed in the previous chapters, via the implementation of the proposed models in the context of transportation linear objects. Transportation possesses a rich source of spatial data, most of which are roads being modeled as two-dimensional linear objects in 2-D GISs. Modeling these linear objects in a three-dimensional space provides a more accurate and realistic representation of these 3-D objects with more accurate geometry measures (3-D distance and curvature, for example) and therefore, benefits numerous applications in which location and distance play critical roles such as emergency routing.

8.1.1 Study Problem Statement

Ground transportation, especially highway transportation is very popular in the United States. As an example of appreciation of the importance of transportation, transportation-related goods and service contributed \$1,050 billion to a \$9.87 trillion U. S. Gross Domestic Product (GDP) in year 2000, which is more than 10% (Office of Transportation Technologies, 2002).

There are a large amount of highways and local roads in the United States. To utilize the existing transportation systems efficiently and economically, a variety of transportation data must be in integrated and analyzed to support making decisions about where to make the best investment (FHWA 1998). This is a great challenge to transportation engineers with the increasing sophistication of transportation systems. The key to integrating these transportation data is where they are located. It cannot be overstated that location is the key to integrating

and analyzing transportation data, and thus to making decisions about transportation investments (FHWA 1998). Keeping records of these roads is a tough task facing state DOTs and local government, even with the rapid developments in transportation engineering and information technologies. Take NCDOT as an example it keeps records for about 79,000 miles of state-maintained roads in North Carolina.

Linear referencing, which remains the mainstay of most transportation data management systems, provides a set of methods and procedures for recording and retrieving locations along linear networks (Federal Highway Administration 1998). Linear referencing provides a foundation for the geographic information system for transportation (GIS-T) and therefore, is being employed widely in modeling transportation data, which is one of the seven framework layers identified in the National Spatial Data Infrastructure (FGDC 1994).

Two key issues in LRS are uniquely identifying a road segment and consistently and accurately referencing all point and linear events along the linear networks. The reference points in LRS are defined as those point features that could be easily located on the field (for example, road intersections) and their positions accurately acquired (Vonderohe et al 1995). Therefore, the key problem becomes the distance measurement, or road centerline 3-D length measurement (Rasdorf and Cai 2001). Other than locating events along linear objects, FGDC recommends including the length information, together with its accuracy information in the transportation framework data (FGDC 1999). In addition, the distance measurement of roads has political and economic meanings to this society. For example, NCDOT is getting one million dollars from the federal government for every mile of Interstate highways in North Carolina. Having accurate road distance measurements is a prerequisite of proper funding distribution nationwide.

A detailed examination of several approaches including extracting information from legacy design drawings, GPS, DMI, and ground surveying, which could be taken to measure road distances revealed that they are time-consuming and labor-intensive when dealing with large amount of road segments. The approach extracting information from legacy design drawings was identified as infeasible. An efficient way to obtain 3-D distances for road segments is needed to update and maintain transportation spatial databases. The 3-D spatial modeling approach attracts research interests because it provides an efficient way to obtain 3-D distances for road

centerlines by modeling them in a three-dimensional space. However, this approach still need to be tested and evaluated, especially the implementation details and the accuracy it could achieve.

8.1.2 Study Objectives

While the overall purpose of this case study is to test the 3-D spatial model and 3-D distance prediction and evaluate the distance accuracy for linear objects in the context of transportation spatial objects (roads), this case study aims at answering the following key questions to gain a comprehensive understanding of the 3-D approach and its accuracy in distance prediction in practice.

- How to implement the proposed 3-D spatial model and 3-D distance prediction with transportation linear objects in practice?
- How accurate is the 3-D spatial model in terms of distance prediction when compared with a higher accuracy distance dataset?
- What are the error sources in the input data? How do they affect the accuracy of the 3-D spatial model and the predicted 3-D distance?
- What are the quality control measures that could be applied to improve the accuracy of this 3-D spatial modeling approach in distance prediction?
- What is the sensitivity of this 3-D spatial modeling approach in distance prediction regarding to the varying accuracy in the input data?
- Are there any properties belonging to the linear objects or the way they are modeled, which significantly affect the accuracy of this 3-D spatial modeling approach in predicting distances?

The case study was designed in such a way that it would provide answers these questions by implementing the proposed 3-D spatial model and 3-D distance prediction with road segments and examining the results in detail.

8.1.3 Study Scope

The scope of this case study is determined based on the availability of required datasets of North Carolina: transportation road network, NED, and LIDAR data. The transportation road network and NED are available

statewide. This case study focuses on primary roads (Interstate highways, US routes, and NC routes) maintained by NCDOT. The restraint comes from LIDAR data, which come from the NC Flood Mapping Program. This program collects LIDAR data by river basins. Figure 8.1 presents the availability of LIDAR data in North Carolina by April 30, 2003, together with river basin and county information.

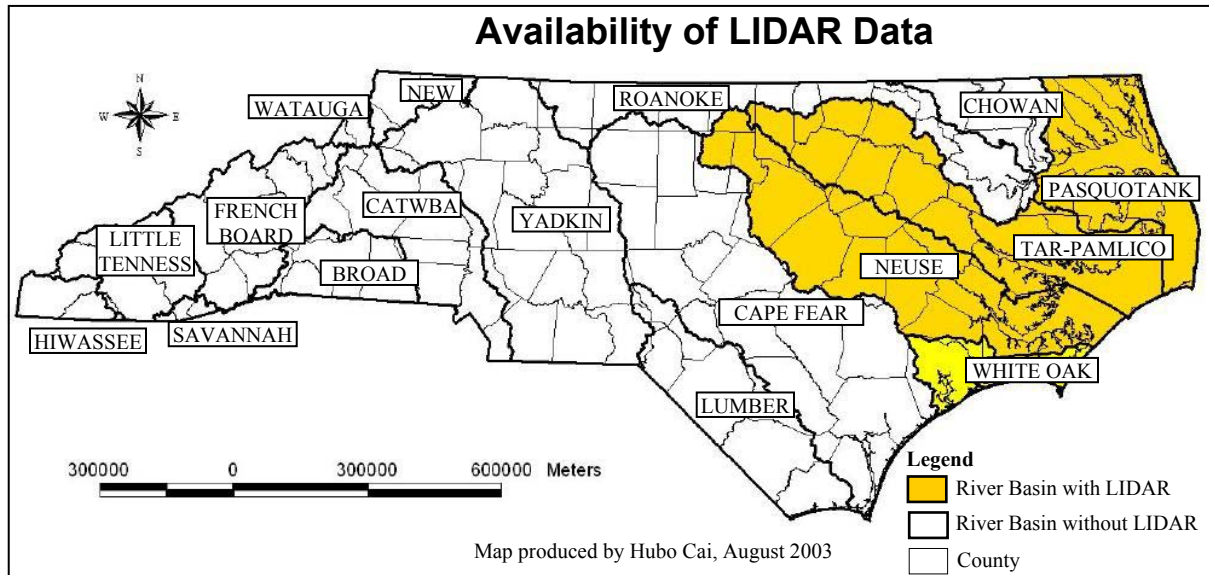


Figure 8.1 Availability of LIDAR Data in North Carolina

Figure 8.2 shows the study scope for this case study. Figure 8.3 provides an enlarged view. This study scope consists of portions of Interstate highways in 9 counties (Durham, Granville, Halifax, Johnston, Nash, Orange, Vance, Wake, and Wilson), and US routes and NC routes in Johnston County. This study scope is determined based on the following concerns.

- 1) Data availability, particularly, the availability of LIDAR data (all road segments in the study scope are surrounded with LIDAR data);
- 2) The variety of road types, i.e. Interstate highways, US routes, and NC routes;
- 3) Sample size, or the number of road segments of different road types.

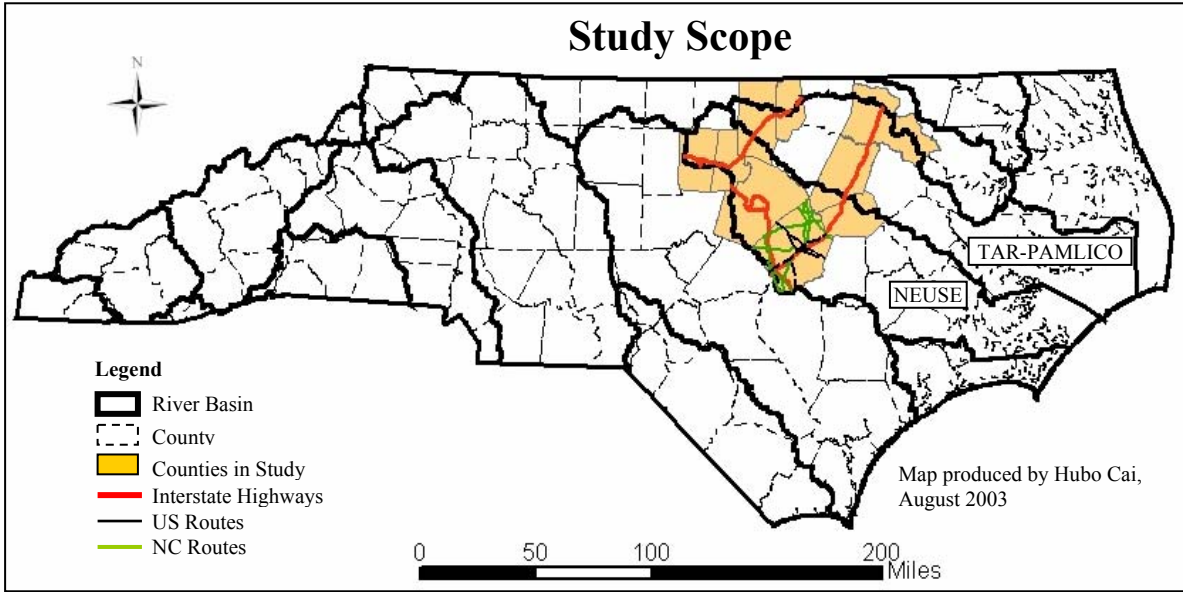


Figure 8.2 Study Scope for the Case Study

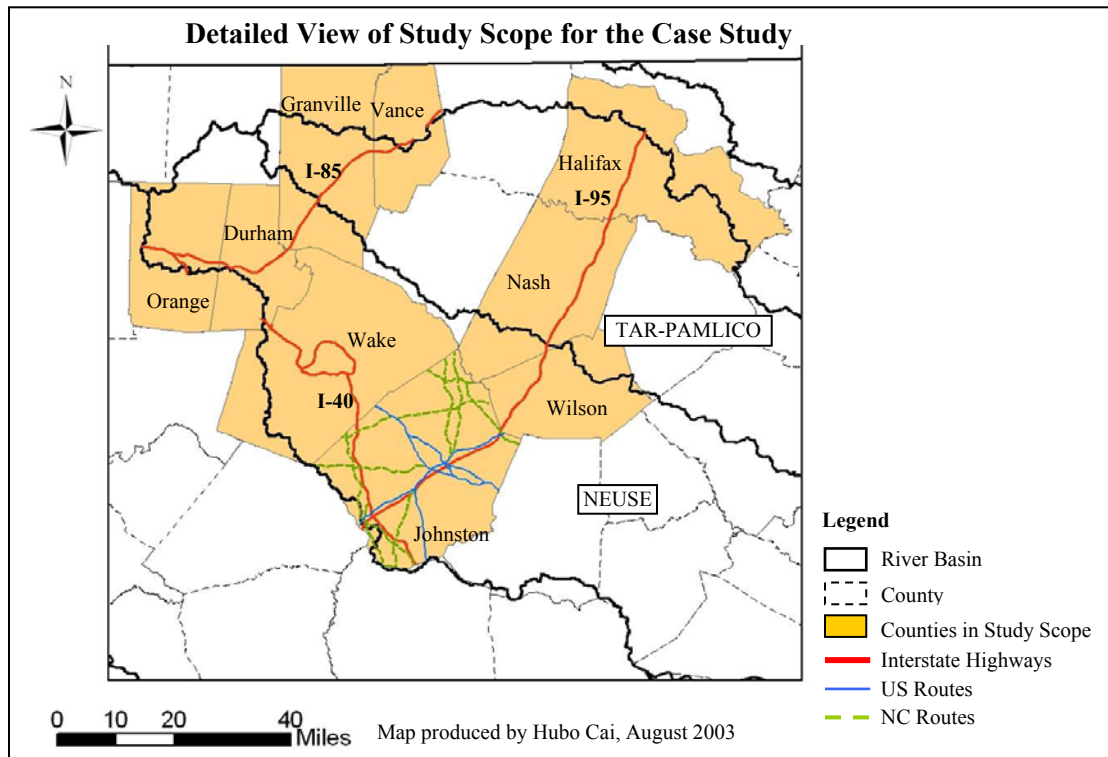


Figure 8.3 Detailed View of the Study Scope for the Case Study

Only Interstate highway segments within the Neuse River Basin and the Tar-Pamlico River Basin, which have LIDAR data available, are included in our study scope. Figures 8.2 and 8.3 illustrate that all US and NC routes in Johnston County are included in the study scope, even though there is no LIDAR data for small portions of Johnston County. Examining these US and NC routes in detail with the availability of LIDAR data reveals that they are surrounded by LIDAR data.

Table 8.1 provides the distribution of road segments included in the study scope. This distribution is based on the road types, i.e. Interstate highways, US routes, and NC routes.

Table 8.1 Distribution of State-Maintained Roads in the Study Scope

Road Type	Interstate Highway	US Route	NC Route	All
# of Segments	662	305	336	1303
Percentage	50.8%	23.4%	25.8%	100%

8.1.4 General Approach

The general approach that will be taken to carry out this case study includes the following steps.

- 1) Identifying information source (data source), including road network data, elevation datasets, and reference data.
- 2) Collecting source data and identifying and quantifying the accuracy of the collected data.
- 3) Preprocessing the collected source data so that the same spatial referencing system is used by all datasets that are involved in the modeling and analysis.
- 4) Developing algorithms under the specified development environment (ArcGIS, ArcObjects, and VBA) to implement 3-D spatial modeling and 3-D distance prediction when working with different elevation datasets.
- 5) Analyzing the results and evaluating the accuracy of the 3-D model and 3-D distance prediction by comparing the results to the reference data.
- 6) Assessing the sensitivity of the results regarding the varying accuracy levels of the elevation data.
- 7) Identifying significant factors (road geometry properties) that affect the accuracy of the 3-D modeling.

8.2 CASE STUDY INFORMATION SOURCES AND DATA COLLECTION

The input datasets of this case study include road network data and elevation data. Elevation datasets include NED, LIDAR point data, and LIDAR DEMs. The concepts of these data models and the technologies in producing these datasets were introduced in the earlier chapters. This section briefly describes each dataset being used in the case study from the aspect of its background information. The focus is on dataset specific characteristics such as density, resolution, spatial referencing systems, accuracy and error information, and data collection.

8.2.1 Road Network

This section describes the digitizing procedure being followed in the GIS unit of NCDOT to obtain digital road centerlines for all state-maintained roads in North Carolina, the digital orthophoto quarter-quadrangles (DOQQs) that are the source images being digitized, and information regarding obtaining the digital road centerline data.

8.2.1.1 Digitizing Road Centerlines from DOQQs

The GIS Unit of NCDOT is undertaking an LRS project, in which a critical part is cleaning the digital road network data for all state-maintained roads in North Carolina to ensure that all road centerlines are in correct positions. It has been realized that having road centerlines in correct positions is the prerequisite of implementing LRSs using GISs.

This cleaning task is implemented via a digitizing procedure. More specifically, road centerlines are digitized on-screen from DOQQs distributed by North Carolina Center for Geographic Information and Analysis (NCCGIA). Figure 8.4 shows an example of lines being digitized from DOQQs. The GIS Unit of NCDOT digitizes road centerlines county by county. There is no formal quality control procedure being followed. However, after the road centerlines are digitized, a project supervisor will check the results to see if they are satisfactory.



Figure 8.4 Digitizing Road Centerlines from DOQQs

The DOQQs being digitized in the GIS Unit of NCDOT to obtain digital road centerlines are 1993 grayscale DOQQs and 1998 color-infrared (CIR) DOQQs distributed by NCCGIA. These DOQQs have a ground resolution of 1 meter. They are georeferenced to State plane coordinates. The datum is NAD83. The measurement unit is either meters or feet. These DOQQs are compressed into JPEG format.

1993 grayscale DOQQs are available statewide in North Carolina while 1998 CIR DOQQs are only partially available in North Carolina. This information will be described in detail in the subsection with title of “DOQQs in North Carolina”. According to NCCGIA (<http://www.cgia.state.nc.us/cir98status/index.html>), flights were done for a statewide color infrared (CIR) DOQQ's coverage in January-March 1998 (with follow-up flights for the eastern Piedmont and other areas in early 1999). CIR data development status information is available from the CIR DOQQ status map. All 1998 CIR DOQQs covered under the North Carolina/USGS Joint Funding Agreement have been sent to CGIA and are available to NCDOT. The remaining DOQQs are being developed by the US Forest Service. No timetable has been sent to CGIA for delivery of these images.

8.2.1.2 DOQQs

DOQQs are digital images produced by the USGS. They contain orthorectified aerial photography at a resolution of 1 meter. DOQQs are black and white (B/W), or color-infrared (CIR) images (USGS 2003e). DOQQs combine the image characteristics of a photograph with the geometric properties of a map. DOQQs are generated by a rectification process to correct image displacements caused by camera lens distortion, camera tip and tilt, terrain relief, and scale (USGS 2003e). A digital orthophoto is a uniform scale photographic image and can be considered a photographic map. Standard DOQQs produced by the USGS have a spatial resolution of 1 meter and cover an area of 3.75-minute of latitude by 3.75-minute of longitude. The native format for DOQQs produced by USGS consists of an ASCII keyword header followed by a series of 8-bit binary image lines for Black/White and 24-bit band-interleaved-by-pixel (BIP) for color. They use the Universal Transverse Mercator (UTM) projection and are referenced to either the North American Datum (NAD) of 1927 (NAD27) or the NAD of 1983 (NAD83) (USGS 2003f).

With DOQQs, features are represented in their true ground position, making direct measurement of distance, areas, angles, and positions possible. Because DOQQs are photographic images, they also display features that may be omitted or generalized on other cartographic maps. This makes the digital orthophotos valuable as a layer in a geographic information system or as a tool for revision of other map materials such as digital line graphs and topographic maps. In other words, revision to digital vector file is one of many applications of DOQQs, which is also the procedure being used in the GIS Unit of NCDOT to obtain highly accurate digital road network data.

8.2.1.3 DOQQs in North Carolina

North Carolina has two generations of statewide photography used to produce digital orthophoto quarter-quadrangles (DOQQs). The first is black and white photography that was acquired in the spring of 1993. The second generation is color infrared acquired in the Spring of 1998 and 1999 (NCGS 2001).

Digital orthophotography is recognized throughout North Carolina as one of several critical geospatial data sets needed to effectively manage and use geographic information systems in statewide, regional and local databases and spatial applications (NCGS 2001). Orthophotography is needed as a base on which to collect and maintain other surface based data such as transportation, cadastral, and agricultural activities, and is highly effective when used in the presentation of natural phenomena, existing and historic conditions, and results of analysis.

The State Mapping Committee (SMAC) was directed by the NC Geographic Information Coordinating Council to pursue the development and acquisition of digital orthophotography for the state (NCGS 2001). The SMAC has been authorized to cooperate with the US Geological Survey on the development of Digital Orthophotography Quarter-Quadrangles (DOQQs) -- a project that satisfies SMAC requirements for digital imagery (NCGS 2001).

Joint Funding Agreements between the NC Geological Survey, the NC Department of Transportation, the NC Center for Geographic Information and Analysis, and the US Geological Survey were used beginning in 1995 to "jump start" the production of 3.75-minute DOQQs for the state (NCGS 2001). The DOQQs are produced from 1993 NAPP photography with a nominal pixel resolution of 1-3 meters. The source photography is black and white. The photographs were taken with a 6-inch lens with a nominal flight height was 20,000 feet. Most photography was taken in March-April, 1993 (leaf-off) (NCGS 2001).

There is a complete statewide coverage for the 1993 DOQQs. The 1993 B/W DOQQs can be ordered from the NCGS in CD-ROM format.

For the 1998 CIR DOQQs in North Carolina, the entire state was flown using color infrared film in the spring of 1998 (NCGS 2001). This is a jointly funded project with the US Geological Survey. Funds were committed to fly the entire state but additional funds were needed to produce DOQQs from the images. Photographs for the eastern and western thirds of the state were accepted from the 1998 flight. The central one-third of the state was re-flown in the spring of 1999. Also a few small areas were re-flown in 1999 (NCGS 2001). Due to funding issues, presently the 1998 CIR DOQQs only cover portions of North Carolina.

8.2.1.4 Description of the Road Centerline Data

The digital road centerline data were obtained from the GIS Unit of NCDOT. The data are in the format of ArcGIS line coverage with the Lambert Conformal Conic projection, the datum of NAD83, and the coordinate system of State plane coordinate system. The unit is meter. The data were obtained county by county. One coverage contains all state-maintained roads (Interstate highways, US routes, NC routes, and Secondary roads) in a particular county.

The road centerline data were initially structured in a link-node format. In case of the NCDOT digital road centerline data, whenever there is an intersection of two or more state-maintained roads, an intersection of state-maintained roads with a railroad, or an intersection of state-maintained roads with county lines (logical lines), from a planimetric view, this intersection becomes a node. In other words, it is a common situation that a node is not a physical connection, for example, an overpass or an underpass as illustrated in Figure 8.5. In Figure 8.5, road B has a bridge above road A. In reality, they are not physically intersected. However, from a planimetric point of view, they are intersected and therefore there is a node. A link is defined as a road segment on a road, which connects two neighboring nodes on that road.

In addition to road geometry, the digital road centerline has some other information in the attribute table. “FNODE” indicates the node from which a link starts. “TNODE” indicates the node at which a link ends. “INTERSECT1” and “INTERSECT2” specify the type of the “FNODE” and “TNODE” respectively. If the intersection is a physical connection between two roads, a unique number value is given. If the intersection is a bridge, the value becomes the bridge ID starting with “B”. If the intersection is a logical intersection of a state-maintained road with a county line, the value becomes the county line ID starting with “C”. For example, based on Figure 8.5, links 1 to 4 all have a node (either “FNODE” or “TNODE”) 1. However, the corresponding “INTERSECT1” or “INTERSECT2” for links 3 and 4 with have a value starting from “B” while for links 1 and 2, they are assigned with a number value. “RTTYPTXT1” represents the type of the road segment, i.e. “I” for Interstate Highways, “US” for US routes, “NC” for NC routes, and “SR” for secondary roads. “ROUTE1”

represents the primary route to which a link belongs. For example, all links on I-40 west have the same “ROUTE1” value of 10600040 while all links on I-40 east have the same “ROUTE1” value of 10000040.

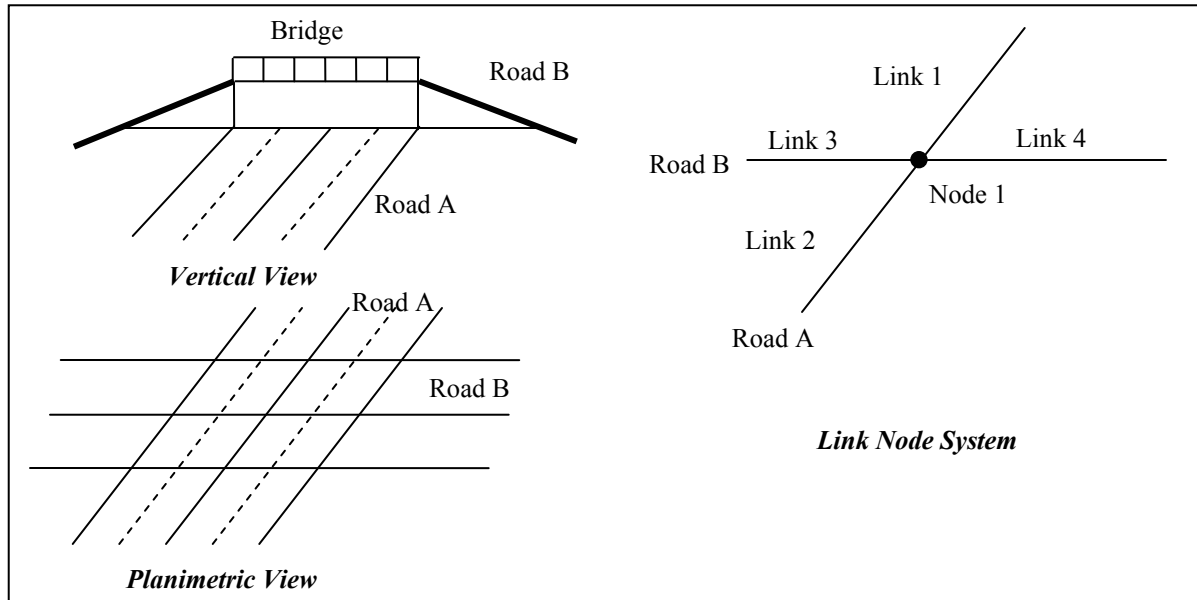


Figure 8.5 Illustration of Intersection, Link, and Node

8.2.1.5 Errors and Accuracy

The digital road centerline data from the GIS Unit of NCDOT were digitized on-screen from 1993 B/W DOQQs and 1998 CIR DOQQs. Consequently, there are two major error sources: errors in the DOQQs and errors introduced due to digitizing. According to document (USGS 2003g):

“Digital orthophoto quadrangles and quarter-quadrangles must meet horizontal National Map Accuracy Standards (NMAS) at 1:24,000 and 1:12,000 scale, respectively. The NMAS specify that 90 percent of the well-defined points tested must fall within 40 feet (1/50 inch) at 1:24,000 scale and 33.3 feet (1/30 inch) at the 1:12,000 scale. The vertical accuracy of the source DEM must be equivalent or better than a level 1 DEM, with a root-mean-square-error (RMSE) of no greater than 7.0 meters. The DOQ RMSE is the square root of the average of the squared discrepancies. These discrepancies are the differences in coordinate (X and Y) values derived by comparing the data being tested with values determined during aerotriangulation or by an

independent survey of higher accuracy. All remaining inputs and processes (i.e., aerotriangulation control and methodology, scanner and sensor calibrations) used in digital orthophoto production must be sufficiently accurate to ensure that the final product meets NMAS.”

For the digitizing part, the errors can be categorized into instrumental errors and human errors. Accuracies of digitizers available on market range from 0.127 to 0.508 mm depending upon each model (Worawat and Rasdorf 2003). Compared to the significance of human errors, these inaccuracies are negligible. In case on-screen digitizing is applied, the major contribution to the inaccuracy is still human errors. In digitizing vector line features from images, two most common types of errors are pseudo nodes and dangling nodes (Worawat and Rasdorf 2003). Pseudo nodes are nodes that are not in correct positions. A dangling node can be defined as a single node connected to a single line entity (Unesco 1999). Dangling nodes (usually called dangles) can result from three possible mistakes: failure to close a polygon, failure to connect the node to the object it was supposed to be connected to (called an undershoot), or going beyond the entity you were supposed to connect to (called an overshoot). Figure 8.6 illustrates a scenario with a pseudo node and two dangling nodes (overshoot and undershoot). It is worth pointing out that not all dangling nodes are digitizing errors. In case a node represents the start or the end of a line feature that is not topologically connected to another line feature, this node becomes a dangling node, but not an error.

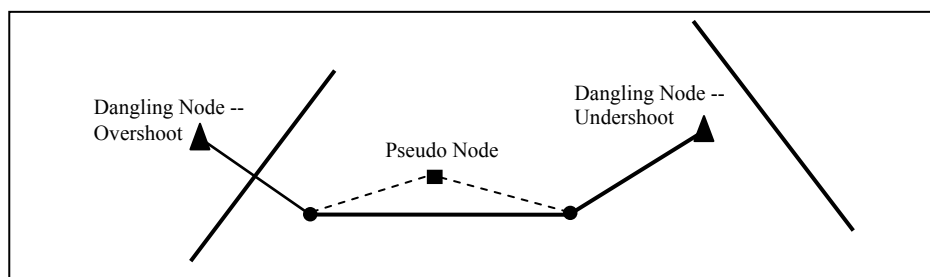


Figure 8.6 Illustration of Errors in Digitizing Linear Features

Currently there is no formal quality control procedure being followed in the GIS Unit of NCDOT when digitizing road centerlines from DOQQs. According to the digitizing procedure outlined earlier, the accuracy of the digitized road centerlines depends on individuals who are doing the digitizing job. Different individual

prefers different map display scales and different intervals between points to capture road curvature. All these are subjective judgments. Consequently, the accuracy varies by county and by road type since each individual was assigned to digitize specific types of roads in one or more counties, for example, all secondary roads in Johnston County.

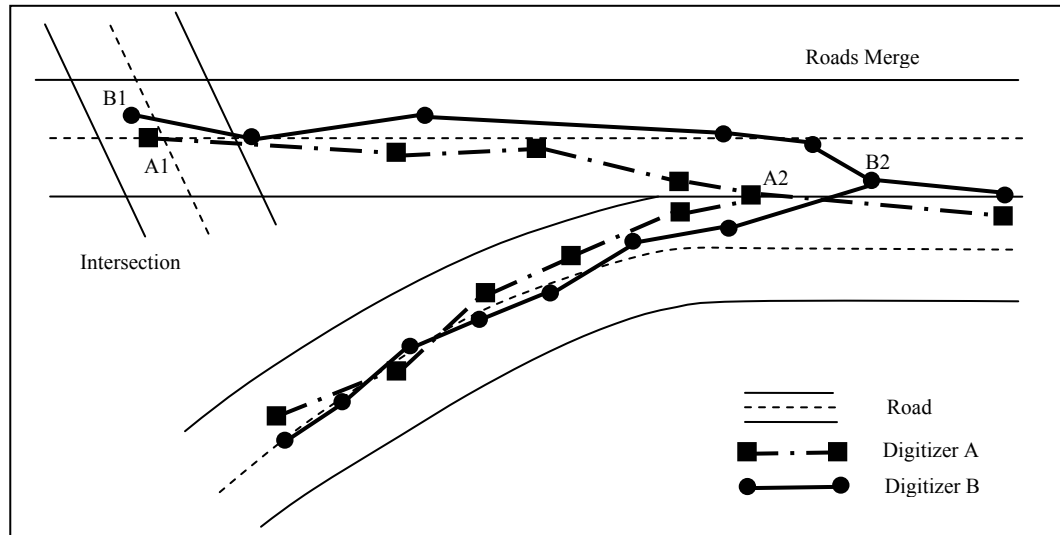


Figure 8.7 Illustration of Dilemmas in Digitizing Road Centerlines

Figure 8.7 illustrates a scenario in which two digitizers are digitizing the same line features but come up with two different sets of results. At the intersection, digitizer A determines that point A1 should be the point but digitizer B picks point B1. At the place where two roads merge, digitizer A picks point A2 while digitizer B chooses to use point B2. Please note the difference between the discrepancy between points A1 and B1 and the discrepancy between points A2 and B2. For the first discrepancy, theoretically there should be no disagreement regarding the position of the point where two road centerlines intersect. For the second discrepancy, the two centerlines for two merging roads continue in a parallel way and therefore, the position to be used to indicate the merging point is completely subjective. Other than those ending nodes, digitizers A and B also use two different sets of intermediate nodes and the resulting lines from digitizer A do not align perfectly with the resulting lines from digitizer B.

In addition, there is another type of positional errors related to county lines. As stated earlier, NCDOT maintains roads by county and wherever a road intersects with county lines, a point is created to indicate the

start point of one road segment in one county and the end point of another road segment in the other county. However, these county lines are logical lines. In other words, there is no such county line that could be physically located on field. On the other hand, there are signs along state-maintained roads indicating the delineation of two neighboring counties. The locations of these signs could be determined by using GPS. Interestingly, these locations do not match with the county lines as illustrated in Figure 8.8, in which the locations of signs delineating Orange County and Alamance County on I-40/85 do not match with the county line.



Figure 8.8 Illustration of Misalignment between Sign Locations and County Lines

This type of errors has more political meanings than spatial analyses. Ideally, either the locations of the signs could be changed or the county lines could be adjusted. However, in reality, this is a very sensitive political issue. No such changes are expected to happen in the near future. This issue will be addressed again in accuracy evaluation of 3-D road distances predicted by our 3-D models.

No accuracy evaluation regarding the positional accuracy of digitized road centerlines has been done in the GIS Unit of NCDOT. Based on conversations with mapping personnel in the GIS Unit, this digitizing procedure is

satisfactory in both the progress and the accuracy. The only concern is the currency of the DOQQs being used. There have been road modifications such as lane expanding and modifications in intersections since the date aerial photos were taken and DOQQs were produced. It is believed that the planimetric positional accuracy of road centerlines is no better than the positional accuracy of the source DOQQs.

8.2.2 Elevation Data

In order to carry out 3-D modeling, elevation datasets are required to add the third dimension – elevation. This section describes the elevation datasets being used in this case study, focusing on their spatial referencing systems, characteristics, the way they are obtained, and error and accuracy information.

8.2.2.1 NED

The USGS NED has been developed by merging the highest-resolution, best-quality elevation data available across the United States into a seamless raster format (USGS 2003a). NED is the result of the USGS effort to provide 1:24,000-scale Digital Elevation Model (DEM) data for the conterminous US and 1:63,360-scale DEM data for Alaska (USGS 2003a, USGS 2003b). The background information regarding NED was detailed earlier. This section describes how the needed NED data for this case study is obtained, description of the obtained data, and errors and accuracy information of the obtained NED data.

8.2.1.1.1 Data Collection and Description

USGS is distributing NED data via the National Map Seamless Data Distribution System (<http://seamless.usgs.gov/>). NED data for North Carolina can be obtained from the North Carolina State University Spatial Information Lab (<http://www.precisionag.ncsu.edu/>) and from the library data server of North Carolina State University (<ftp://rpp.lib.ncsu.edu/fedgov/usgs/ned/>).

The NED data being used in this case study are downloaded from North Carolina State University Spatial Information Lab (<http://www.precisionag.ncsu.edu/scripts/links/index.php?viewCat=51>) by counties. In other

words, for all nine counties in the study scope, NED data are downloaded for each of them. The downloaded files are in the format of interchange files (.e00 files). These interchange files can be easily converted into raster grid files. The horizontal datum is NAD83 (feet). The vertical datum is NAVD88. The projection is Lambert Conformal Conic projection. The coordinate system is State plane coordinate system with the horizontal units of feet and vertical units of meters. The grid spacing of the converted raster data is 1-arc-second (approximately 30 meters or 92.015319 feet).

8.2.1.1.2 Errors and Accuracy

Accuracy assessment and reporting for NED is a topic that is being actively pursued (Osborn et al. 2001). A plan for assessing the overall accuracy of the National Elevation Dataset (NED) based on independent high-accuracy geodetic control is currently under development (USGS 2002). The accuracy described will be in the terms used in the Federal Geographic Data Committee (FGDC) Content Standards for Framework Land Elevation Data (USGS 2002). Accuracy will be determined and reported using the methodology specified in the FGDC National Standard for Spatial Data Accuracy.

Until an independent overall assessment of the accuracy of NED is completed, it is best to refer to published information on the accuracy of the source digital elevation models (DEMs), from which NED was assembled. The primary source data were the USGS 7.5-minute (30-meter or 10-meter resolution) DEMs. Gaps in the 7.5-minute DEM coverage were filled with data from USGS 30-minute (2-arc-second resolution) or 1-degree (3-arc-second resolution) DEM's.

According to the metadata associated with the NED in North Carolina, the majority of the source DEMs used to produce NED in North Carolina are 30-meters DEMs (7.5-minute DEMs, Level 1 and Level 2). Level 1 DEMs are elevation data sets in a standardized format. This level includes 7.5-minute DEMs or an equivalent that is derived from stereo profiling or image correlation of National High Altitude Photography Program, National Aerial Photography Program, or equivalent photographs (National Mapping Program Standards 1998). For 7.5-minute DEMs derived from a photogrammetric source, 90 percent have a vertical accuracy of 7-meter RMSE or better and 10 percent are in the 8- to 15-meter range (USGS 2000). Level 2 DEMs are elevation data sets that

have been processed or smoothed for consistency and edited to remove identifiable systematic errors (National Mapping Program Standards 1998). DEM data derived from hypsographic and hydrographic data digitizing, either photogrammetrically or from existing maps, are entered into the level 2 category after review on a DEM editing system (National Mapping Program Standards 1998). An RMSE of one-half contour interval is the maximum permitted, with no errors greater than one contour interval (National Mapping Program Standards 1998).

These DEMs are photogrammetrically-collected DEMs. They suffer from three different types of errors: random errors, blunders, and systematic errors (Daniel and Tennant 2001). Random errors result from accidental or unknown combinations of problems and they remain in the data after blunders and systematic errors are removed (USGS 1997). Blunders are vertical errors that exceed the maximum absolute error permitted and are associated with the data collection process (Daniel and Tennant 2001). Blunders can be easily identified and removed through visualization and editing. Systematic errors are fixed patterns and are usually related to collection procedures or photogrammetric systems used in the DEM generation. Examples include errors of camera calibration, errors introduced by film and photomechanical process, errors of the photogrammetric stereo plotters, errors of scanners/scanning process, errors of aerotriangulation, errors of exterior orientation parameters, errors related to the Earth curvature and atmospheric corrections, errors due to image matching algorithms, errors of collection technique, errors due to vegetation interference, and personal bias (Daniel and Tennant 2001). If the error can be identified, systematic error can be modeled, reduced, or even eliminated (Daniel and Tennant 2001).

8.2.2.2 LIDAR

The LIDAR elevation data being used in this case study are obtainable from the North Carolina Floodplain Mapping Program (www.ncfloodmaps.com). This section describes the characteristics of the LIDAR elevation data from this program, the approach being used to obtain LIDAR data, and errors and accuracies of the obtained LIDAR data.

8.2.2.2.1 Data Collection and Description

The major purpose of NC floodplain mapping program is to conduct flood hazard analyses and produce digital flood insurance rate maps (DFIRMs). As part of this program, LIDAR is being used to generate digital elevation data (NCCTS 2003).

The initial set of LIDAR points are post-processed in order to generate bare Earth data, which are defined as digital elevation data of the terrain and free from vegetation, buildings, and other man-made structures (NCCTS 2003). To get these bare Earth elevation data, automated and manual post-processing are used to eliminate points that impinged on elevation features (NCCTS 2003). Automated post-processing includes computerized procedures that detect elevation changes that appear to be unnatural such as rooftops. Manual post-processing normally includes the overlay of data points on digital imagery so that the analyst can see where the laser points hit the ground.

As stated earlier, LIDAR cannot accurately delineate stream channels, shorelines, or ridge lines visible on photographic images and therefore, breaklines must be incorporated, which adds costs and delivery times. In case of the LIDAR bare Earth data produced by the North Carolina Floodplain Mapping Program, breaklines delineating water channels and cross-sectional data of water bodies are obtained via field surveys and examination of photographic images (NCCTS 2003). These breaklines are used together with the post-processed LIDAR points to produce other elevation datasets such as DEMs. Table 8.2 lists the available digital elevation data products from the North Carolina Floodplain Mapping Program (NCCTS 2003). These products are available for download at <http://www.ncfloodmaps.com>.

Table 8.2 Available Digital Elevation Data Products from the North Carolina Floodplain Mapping Program

Product Name	Tile Size	File Format(s)	Metadata
Bare Earth Mass Points	10,000 feet	ASCII	Yes
Bare Earth Breaklines	10,000 feet	ESRI Shapefile	Yes
50-foot Hydro-Corrected DEMs	10,000 feet	ESRI GRID, ASCII	Yes
20-foot DEMs	10,000 feet	ESRI GRID, ASCII	Yes

The LIDAR elevation data in this case study (bare Earth points, 50-ft DEMs, and 20-ft DEMs) are all downloaded from the address mentioned above. Since the data files are quite large regarding to their sizes, only tiles surrounding Interstate highways in the study scope and tiles covering Johnston County are downloaded, in order to reduce the amount of data to be downloaded.

The downloaded mass point data are in the format of ASCII files. Each point has X, Y, and Z-coordinates with points classified as non-Earth removed. The horizontal datum is NAD83. The projection is Lambert Conformal Conic projection. The coordinate system is State Plane (North Carolina). The vertical datum is NAVD88. Both vertical and horizontal units are feet. Average density of point in Johnston County is approximately 18.6 feet.

The downloaded 20-ft and 50-ft DEMs are in the format of ASCII files. They can be easily converted into grid files. These DEMs have the same spatial referencing system as the point data, i.e. the same projection, the same coordinate system, the same horizontal and vertical units, and the same datums. 20-ft and 50-ft represent the resolution of the DEMs (or cell size). According to the metadata, cell values in the DEM were derived from a Triangulated Irregular Network (TIN) produced from the bare Earth mass points and breaklines. Individual cell values are based on an average of 25 points inside each 20-ft DEM cell for 20-ft DEMs and an average of 25 points inside each 50-ft DEM cell for 50-ft DEMs.

A total of 186 tiles surrounding Interstate highways in the study scope and 257 tiles in Johnston County are downloaded.

8.2.2.2.2 Errors and Accuracy

While LIDAR technology is capable of providing highly accurate elevation data, it is not error free. The error sources contributing to the positional inaccuracy in the along-track direction, cross-track direction, and height of LIDAR points were described earlier in section 6.4.1.2.6. These LIDAR points are post-processed to generate bare Earth data as in the case of the North Carolina Floodplain Mapping Program. The purpose of conducting post-processing is to remove erroneous points or non-ground points. However, it is very possible that some

points are mistakenly removed while some others are ignored. In other words, some points that should not be removed might be removed while some points should not be removed are removed.

In the case of the North Carolina Floodplain Mapping, it is realized the importance of positional accuracy of LIDAR data. Typically, for hydraulic modeling, elevation data equivalent to 4 feet contours (RMSE = 37 cm) are appropriate for rolling to hilly terrain, and elevation data equivalent to 2 feet contours (RMSE = 18.5 cm) are appropriate for flat terrain. North Carolina specified a RMSE of 20 cm for coastal counties and a RMSE of 25 cm for inland counties (NCCTSMP 2001). A formal procedure is followed in checking the positional accuracy of the bare Earth points after post-processing. The procedure is summarized as below (NCCTSMP 2001).

- (1) All land is categorized into one of five major land cover categories representative of the floodplains: a) bare-Earth and low grass (e.g., sand, rock, plowed fields, lawns, golf courses); b) high grass, weeds, and crops (e.g., hay, corn, wheat, tobacco); c) brush lands and low trees; d) fully forested; and e) urban areas(vicinity of manmade structures, high density).
- (2) For each county, the NC Geodetic Survey (NCGS) will direct independent survey contractors to select 30 or more checkpoints, in each of the five land cover categories (10 of these will be held for subsequent quality control).
- (3) NCGS will contract for independent surveys for all checkpoints to 5-cm Local Network accuracy according to NOAA Technical Memorandum NOS NGS-58, "Guidelines for Establishing GPS-Derived Ellipsoid Heights (Standards: 2 cm and 5 cm)," Version 4.3, November 1997 (referred to as NGS-58).
- (4) Elevations for 20 of these checkpoints for each land cover category are interpolated from the post-processed bare Earth point data.
- (5) RMSE for all these points is computed and documented.
- (6) The worst 5% of these checkpoints are discarded and RMSE is computed for the remaining 95% checkpoints.

Accuracy assessment reports are accessible via <http://www.ncgs.state.nc.us/floodmap.html>. Table 8.3 provides the results for Durham County, Johnston County, and Wake County in the Neuse river basin. Figure 8.9 demonstrates that the RMSE varies among counties and among land types.

Table 8.3 Accuracy Assessment Results for Durham, Johnston, and Wake

Land Class	# of Points	Percentage	RMSE (cm)
Durham			
ALL	96	100	38.0
ALL	91	95	14.1
Grass	16	18	6.9
Weeds/Crop	16	18	10.3
Scrub	14	15	15.6
Forest	29	33	17.8
Built-up	10	11	13.6
Johnston			
ALL	224	100	20.5
ALL	213	95	14.6
Grass	43	19	10.9
Weeds/Crop	43	19	13.6
Scrub	34	15	18.9
Forest	47	21	16.2
Built-up	46	21	13.1
Wake			
ALL	132	100	16.3
ALL	125	95	13.2
Grass	23	17	13.2
Weeds/Crop	20	15	12.4
Scrub	15	11	15.8
Forest	43	33	13.0
Built-up	24	18	12.3

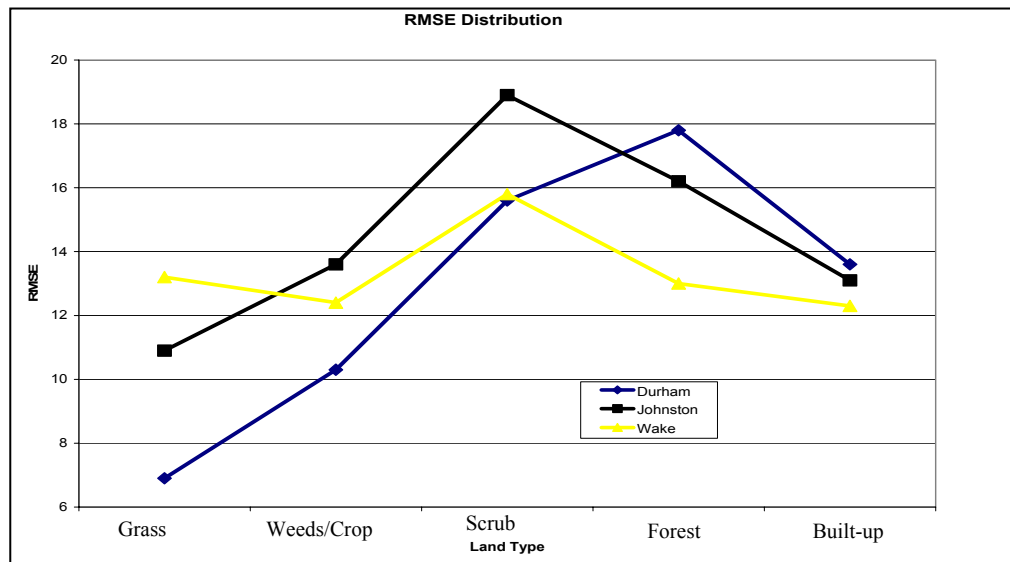


Figure 8.9 RMSE Trends

8.2.3 Reference Data (DMI Data)

The study scope for this case study includes Interstate highways in 9 counties and US and NC routes in Johnston County. In order to conduct accuracy assessment of 3-D distances predicted via 3-D modeling, DMI measurements that will serve as the reference data are needed for all these road segments. The GIS Unit of NCDOT has DMI length measurements for all Interstate highways in North Carolina. These DMI measurements were acquired in summer 2000 by NCDOT personnel with a precision of 0.01 mile. In other words, for each road segment defined as a link in the link-node format, it has a DMI measurement. Also, the GIS Unit of NCDOT has DMI measurements for US and NC routes in Johnston County, which were acquired in 1998. These DMI measurements have either 0.1 mile or 0.01 mile precision. All the DMI measurements are recorded on paper maps written by hand.

A preliminary examination of these DMI measurements reveals that they are not suitable to be used in this case study. First, they have varying precisions. Second, they are all written by hand on paper maps. They are not very well recorded. Some values are not clear enough to be identified. Some values are ambiguous in identifying the corresponding road segments. Third, there are missing DMI measurements for some road segments.

8.2.3.1 Data Collection

The researchers decided to collect DMI measurements in unit of foot for all road segments in the study scope via field work. The DMI used to collect measurements is NITESTAR[®] Model NS-60 from Nu-Metrics, Inc. This model has an accuracy of 1 foot/mile (repeatability) in measuring distances (Worawat and Rasdorf 2003, Nu-Metrics 1998). In order to obtain the accuracy specified by the manufacturer, calibrations are needed before measurements can be taken (Worawat and Rasdorf 2003, Nu-Metrics 1998). The calibration is done by NCDOT personnel working in the Road Inventory Section of the GIS Unit of NCDOT.

The data collection procedure is summarized as below. This procedure is determined based on all constraints in practice, for example, it is very dangerous to stop on Interstate highways without pulling over to shoulder.

- (1) A two-member data collection team is used, one for driving and the other for data collection.
- (2) Paper maps are prepared from a GIS. These paper maps are carefully labeled to help identification in field.
- (3) When doing field work, a laptop storing road network data is kept in the vehicle for reference purpose.
- (4) The measurements are done by road types, i.e., first, all Interstate highways in the study scope are measured; second, all US routes in the study scope are measured; and finally, all NC routes in the study scope are measured. Furthermore, when collecting DMI measurements, all road segments on one road or one direction of a road in case its two directions are modeled separately are measured continuously to minimize the errors of identifying start and end points.
- (5) When measuring the distance for a specific road segment, the vehicle stops on the shoulder at both the start and end points, whenever possible. Fine adjustments are made. The measured distance is recorded on the paper map. On the same day, the field measurements are input into computer.
- (6) In case there are more than two lanes, the vehicle is kept in the middle lane(s) and as close to the road center as possible. In case of two lane roads, the vehicle is kept in the outer lane in order to minimize the errors introduced due to pulling over to the shoulder.
- (7) No lane changing unless it's necessary such as pulling over to shoulder or trying to get on the road from the shoulder. In case of lane changing, the changing is finished in a distance as long as possible to minimize the errors introduced due to lane changing.
- (8) In order to avoid human errors, data recording and data inputting both undergoes a double-check procedure. Before recording a value on the paper map, both team members will check the readings. If there is a disagreement, this particular road segment will be measured again until an agreement is reached. After the measurements are input into the computer by one member, the other one will check the data in computer with those recorded on paper maps.

8.2.3.2 Errors and Accuracy

The manufacturer's specified accuracy is 1 foot/mile for NITESTAR[®] Model NS-60 from Nu-Metrics, Inc. However, the researchers believe this accuracy cannot be reached in practice even with careful calibrations. Errors introduced in the field measurements are summarized as below.

(1) Errors in identifying start and end points

Start and end points are those points where two or more road centerlines intersect from a planimetric point of view. In reality, the start or end points could be the physical intersection of two roads or just an underpass or an overpass. One special case is the point where two roads merge together (theoretically, the centerlines of these two roads are parallel to each other as illustrated in Figure 8.7. Under all these scenarios, the data collector must make a subjective judgment regarding the location of the start and end points while sitting in the vehicle. Furthermore, when collecting data on two roads that intersect with each other (assuming road R1 intersects with Road R2 at point A, point A is the start point of road segment S1 and the end point of road segment S2 on road R1, and point A is also the start point of road segment S3 and the end point of road segment S4 on road R2), the point A will be located once when working on road R1 and will be located again when working with R2. These two locations for the same point A1 will rarely match.

(2) Errors due to stop on shoulders at start and end points

Figure 8.10 illustrates a scenario in which errors are introduced due to stopping on shoulders at the start and end points. In Figure 8.10, the distance between points A and B is to be measured. The vehicle equipped with DMI stops on shoulder at point C, gets onto road via CD, moves on road along DE, pulls over to shoulder along EF, and stops on shoulder at point F. Obviously the measured distance for AB is actually CDEF, which is longer.

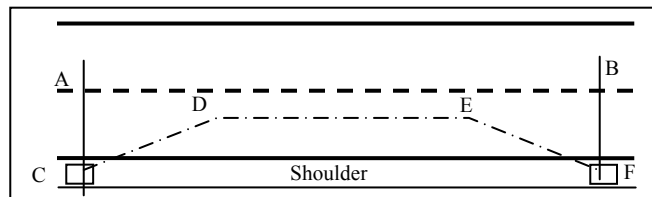


Figure 8.10 Illustration of Errors due to Stopping on Shoulders at Start and End Points

(3) Errors due to lane changing

This type of errors is similar to the errors introduced due to stopping on shoulders as illustrated in Figure 8.10 and therefore, are not repeated herein.

(4) Errors due to no stop at start or end points

It is not always possible to stop on shoulders. For example, there might be no shoulder at all or the shoulder is not wide enough for a vehicle to stop such as on some bridges of Interstate highways or a physical connection of two roads. In those cases, the measurement must be taken when the vehicle is moving. Assuming the vehicle is moving very slowly at 20 mile/hour, a one second time discrepancy in pushing the button to either start or end measurements will lead to a 30ft discrepancy.

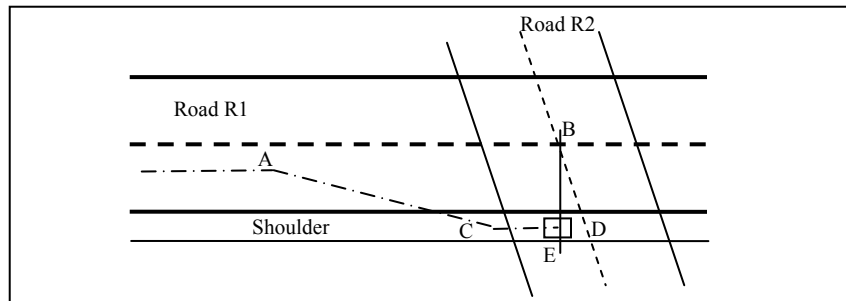


Figure 8.11 Quality Control in Identifying Start or End Points When Collecting DMI Measurements

Quality control measures could be taken to minimize some of the above errors. Some fine adjustments could also be taken. For errors in identifying start or end points, whenever there is a shoulder available for stop, the data collector will stop there and make find adjustment as illustrated in Figure 8.11. Assuming road R1 goes under road R2 (the intersection is actually not a physical intersection, but an underpass with a bridge on road R2) and the measurement is being taken for a road segment on road R1, which stops at point B. The data collector will pull over the vehicle along the relatively long line AC to stop the vehicle on the shoulder. By moving vehicle slowly along the shoulder under the bridge, point E rather than point D is located as the end point and the measurement is collected.

For errors introduced due to nonstop at start or end points, the vehicle is slowed down as much as possible when collecting data at those points. For errors introduced due to lane changing or pulling over to shoulder, this lane

changing or pulling over procedure takes a distance as long as possible to minimize errors illustrated in Figure 8.10. In addition, when collecting DMI measurements for short straight road segment, the measurements are taken when moving vehicle along the shoulder, if possible.

8.3 PRE-PROCESSING

This section describes the procedures followed to pre-process the obtained data, which enables them ready for 3-D modeling. After pre-processing, all data have the same projection (Lambert Conformal Conic Projection), the same horizontal datum (NAD83), the same horizontal unit (feet), and the same coordinate system (State Plane Coordinate System, North Carolina). LIDAR data and NED both have the same vertical datum (NAVD88). The vertical unit of LIDAR data is feet while the vertical unit of NED is meter. This discrepancy can be accommodated when model road centerlines in a 3-D space and predict their 3-D lengths.

8.3.1 Pre-Processing Road Centerline Data

The obtained road centerline data are in the format of ESRI coverage by county. In other words, a coverage contains all state-maintained roads in one county. The study scope includes primary roads (Interstate, US routes, NC routes) only. These roads will be grouped according to the road type and therefore, only primary roads were retrieved from the obtained road centerline data. In addition, they are separated according to road types. This pre-processing step leads to 9 shapefiles, each of which contains Interstate highways in the study scope for one county. Two other shapefiles contain all US routes in the study scope in Johnston County and all NC routes in Johnston County.

These shapefiles still deploy the link-node structure. Most of these links are short links less than 1 mile long. This link-node structure is the particular network model being used by NCDOT to manage its road network and road inventories. It has been recognized by FGDC that many transportation features are characterized by extensive linear network models (FGDC 1999). To promote data sharing, FGDC recommended defining road network by deploying the concepts of Framework Transportation Segment (FTSeg) and Framework Transportation Reference Point (FTRP) (FGDC 1999). An FTRP is defined as the specific location of one

endpoint of an FTSeg on a physical transportation system while an FTSeg is defined as a specified directed path between two FTRPs along a physical transportation system that identifies a unique segment of that physical system (FGDC 1999). In implementation, FGDC recommends that 1) An FTRP should be placed wherever a road crosses a jurisdictional boundary between two data authorities; 2) An FTRP should be placed wherever a road crosses a state border; 3) It is recommended to place an FTRP wherever a road crosses the boundary between two counties within a state; 4) An FTRP should be placed wherever two roads of similar functional class or importance cross one another at grade; and 5) An FTRP may be placed at grade-separated crossings such as overpass or underpass (FGDC 1999).

In order to promote data sharing and focus on future developments and applications, it was determined that this case study would follow FGDC recommendations while deploying the concepts of FTRP and FTSeg. In other words, FTRPs will be placed wherever a road crosses a county boundary, two primary roads intersect with each other, or at grade-separated crossings such as overpass or under pass. All connected links between two FTRPs are merged together into one FTSeg.

Figure 8.12 illustrates the algorithm developed to derive FTSegs from Interstate highway road links of the link-node system. The developed code to implement this algorithm for getting FTSegs for Interstate highways in Johnston County is attached as Appendix A. This algorithm works based on the characteristics of the link-node system and the concepts of FTSeg and FTRP. A link in the link-node system is directed from its FNODE to its TNODE. Each link has an attribute named ROUTE. All links having the same ROUTE value are connected links on a particular route with their BEGMP values indicating their beginning mileposts on that route.

The algorithm works in the way summarized as below, assuming there are two attribute tables, one for Interstate highway links and the other for links of US and NC routes. In addition, the merge value is a unique value indicating the identity of an FTSeg.

- 1) All interstate links having a particular ROUTE value are retrieved and sorted based on the BEGMP values, ascending.

- 2) Get the first link and assign it with the current merge value indicating the identity of the current FTSeg being derived. At the very beginning, this value is 0.
- 3) Query the Interstate highway link attribute table to find all Interstate highway links connected to the current link.
- 4) If there is no such link, then the search for links belong to the current FTSeg is completed. The merge value is increased by 1 to indicate this completion and a new FTSeg will be derived later with the new merge value. Go to step 7).
- 5) If there are two or more such links, then the search for links belonging to the current FTSeg is also completed. The merge value is increased by 1 to indicate this completion and a new FTSeg will be derived later with the new merge value. Go to step 7).
- 6) If there is only one such link, which is the next link in the sorted links, the USNC attribute table is queried to determine if there is a US or NC link connected to the current link. If there is such a link, the search for links belonging to the current FTSeg is also completed. The merge value is increased by 1 to indicate this completion and a new FTSeg will be derived later with the new merge value. If there is no such link, go to step 7).
- 7) Get the next link from the sorted links, assign the current merge value to it, and go to step 3) to repeat steps starting from step 3) until the last link of the sorted links is reached.

The whole procedure is conducted for all distinct ROUTE values. The result is that every link will be assigned with a merge value that indicates the identity of the FTSeg, to which this link belongs. Finally all links have the same merge value are merged together, resulting in 11 shapefiles for these newly derived FTSegs. Among these 11 shapefiles, 9 of them are Interstate highway FTSegs (one shapefile for each county), the rest 2 of them are US FTSegs and NC FTSegs. Since other datasets are using foot as the horizontal unit, all road shapefiles were reprojected into the State Plane Coordinate System (ft) using NAD83 as the datum.

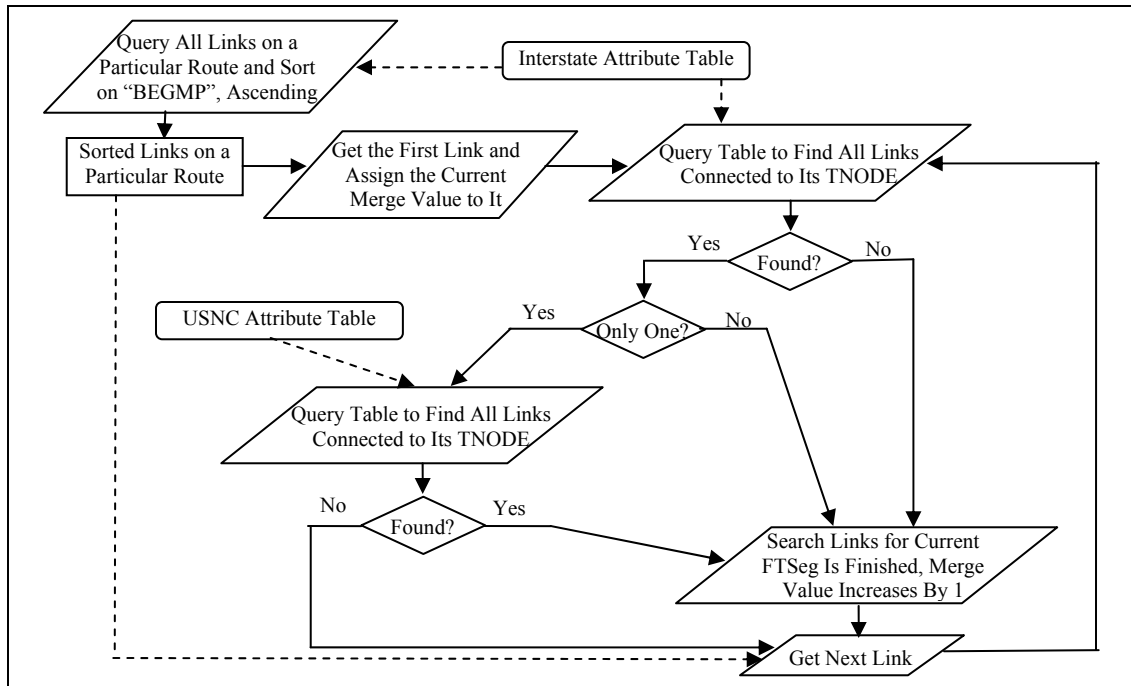


Figure 8.12 Algorithm for Deriving FTSEgs from Link-Node System

Table 8.4 provides the distribution of FTSEgs after the road centerline data are pre-processed (i.e. FTSEgs are derived from links). This distribution provides an adequate number of individual road segments (FTSEgs in this case) and a satisfactory variety of road types.

Table 8.4 Distribution of FTSEgs in the Study Scope

Road Type	Interstate Highway	US Route	NC Route	All
# of FTSEgs	184	78	48	310
Percentage	59.4%	25.2%	15.4%	100%

8.3.2 Pre-Processing LIDAR Data

The downloaded LIDAR data are compressed files. For each tile, there are three compressed files, one for bare-Earth mass point data, one for 20-ft LIDAR DEM, and one for 50-ft LIDAR DEM. All downloaded compressed files were uncompressed.

8.3.2.1 Pre-Processing LIDAR Mass Point Data

The uncompressed data files of LIDAR mass points are ASCII files containing X/Y/Z-coordinates for each LIDAR point. An ASCII file can be converted into point shapefile using ArcGIS. There is a tool in ArcToolbox to realize this conversion. However, due to the huge number of points in each file (typically more than 250,000 points), this tool was not working properly and therefore, another approach was taken. This approach first converted ASCII files into DBF files (database files). Each point in the ASCII file occupied a singular row (a record) in the corresponding DBF file. Second, all these DBF files were converted into point shapefile using ArcGIS.

Due to the file sizes and the number of points involved, before merging these points together, the number of points and consequently the file sizes were reduced by applying a 30-ft buffer to the road centerlines and only those points inside this buffer were kept. Finally, all these buffered points were merged together, resulting to a point shapefile containing all points that are close to the road centerlines in the study scope.

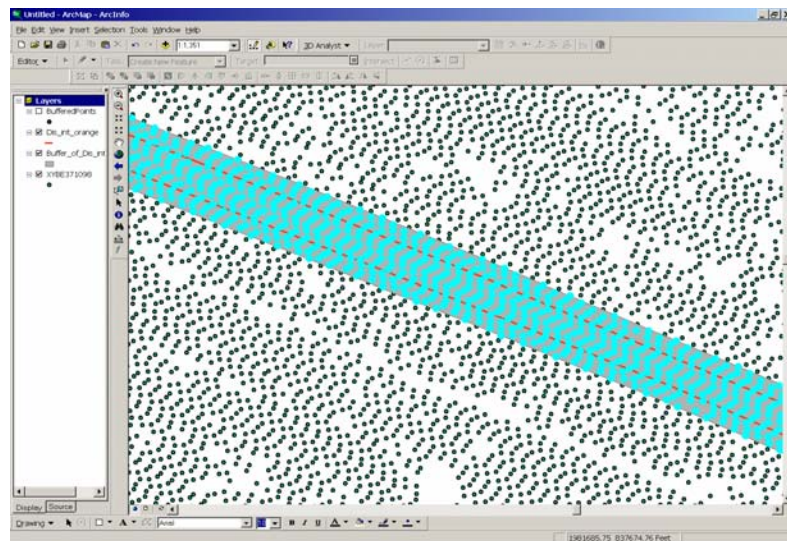


Figure 8.13 Illustration of Reducing the Number of LIDAR Points by Applying a Buffer – Part I

Figures 8.13 and 8.14 provide detailed views of the effect of reducing the number of points. Figure 8.13 shows the road centerlines, the 30-ft buffer applied to road centerlines, and all LIDAR points, with LIDAR points in

the buffer highlighted. Figure 8.14 shows the remaining LIDAR points after points outside the buffer are removed.

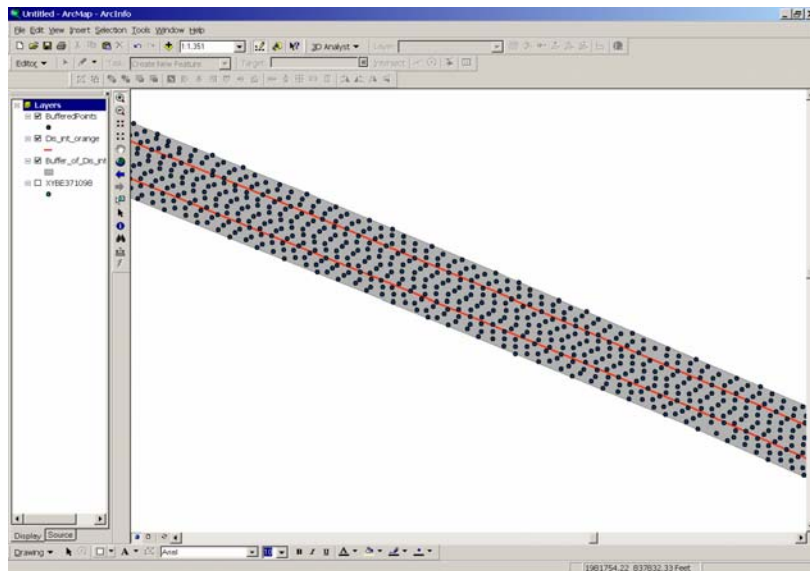


Figure 8.14 Illustration of Reducing the Number of LIDAR Points by Applying a Buffer – Part II

8.3.2.2 Pre-Processing LIDAR DEMs

The uncompressed DEMs were ASCII files. For each tile, there are two ASCII files, one for 20-ft DEM and the other for 50-ft DEM. All these ASCII files were converted into ESRI grids using ArcGIS. All elevation values are real numbers with four digits. Examining LIDAR DEMs reveals that the last two digits are all 0s.

Checking file sizes for these grids lead to the decision that all 20-ft DEM grids should be converted into integer grids, i.e. all cells have integer values. In order to keep the initial precision of the elevation values, this conversion was completed in such a way that a cell value in the converted grid is 100 times the original value of the corresponding cell in the grid before conversion. This conversion was not applied to 50-ft DEMs because the file size of 50-DEM was not a concern.

All 50-ft DEM grids were merged together to create a singular grid. All converted 20-ft DEM grids were merged to create another singular grid. The vertical unit is foot. According to the conversion applied to 20-ft DEM grids, when using these grids, the true elevation value should be the cell value divided by 100.

8.3.3 Pre-Processing NED Data

All downloaded NED data were interchange files (.e00 files). These interchange files were converted into ESRI grids. The vertical unit is meter. A conversion could have been done to convert the unit from meter to feet. But the decision was to keep the original unit and accommodate this discrepancy when modeling road centerlines in a 3-D space and predicting their 3-D lengths.

All NED grids were merged together to create a singular grid. This grid covered completely the 9 counties in the study scope of this case study.

8.4 MODELING ROAD CENTERLINES IN 3-D

After pre-processing, the road centerline data have the same spatial referencing system as that of the elevation datasets. They are suitable for modeling road centerlines in 3-D via spatial analyses. As stated earlier, a set of 3-D points on a particular road centerline will be used to model this road centerline in a 3-D space. This section describes how these 3-D points on road centerlines are obtained when working with different elevation datasets. Some sample results are included in this section. In addition, quality control measures taken when working with LIDAR points are also described.

8.4.1 Using LIDAR Points

The approximation approach described in section 7.2.2.1.2 was followed to obtain 3-D points along road centerlines when working with LIDAR point clouds. This section describes 4 steps that have been taken in the case study.

8.4.1.1 Snapping

The first step of implementing the approximation approach is to obtain 3-D points on road centerlines by a way that is similar to snapping LIDAR points to road centerlines. The concepts were illustrated in section 7.2.2.1.2. Figure 8.15 illustrates the algorithm for snapping. This algorithm requires a buffer size that is determined to be the typical lane width of the roads being modeled. More specifically, the buffer size is 12 ft when working with Interstate highway and US route FTSegs and 10 ft when working with NC route FTSegs.

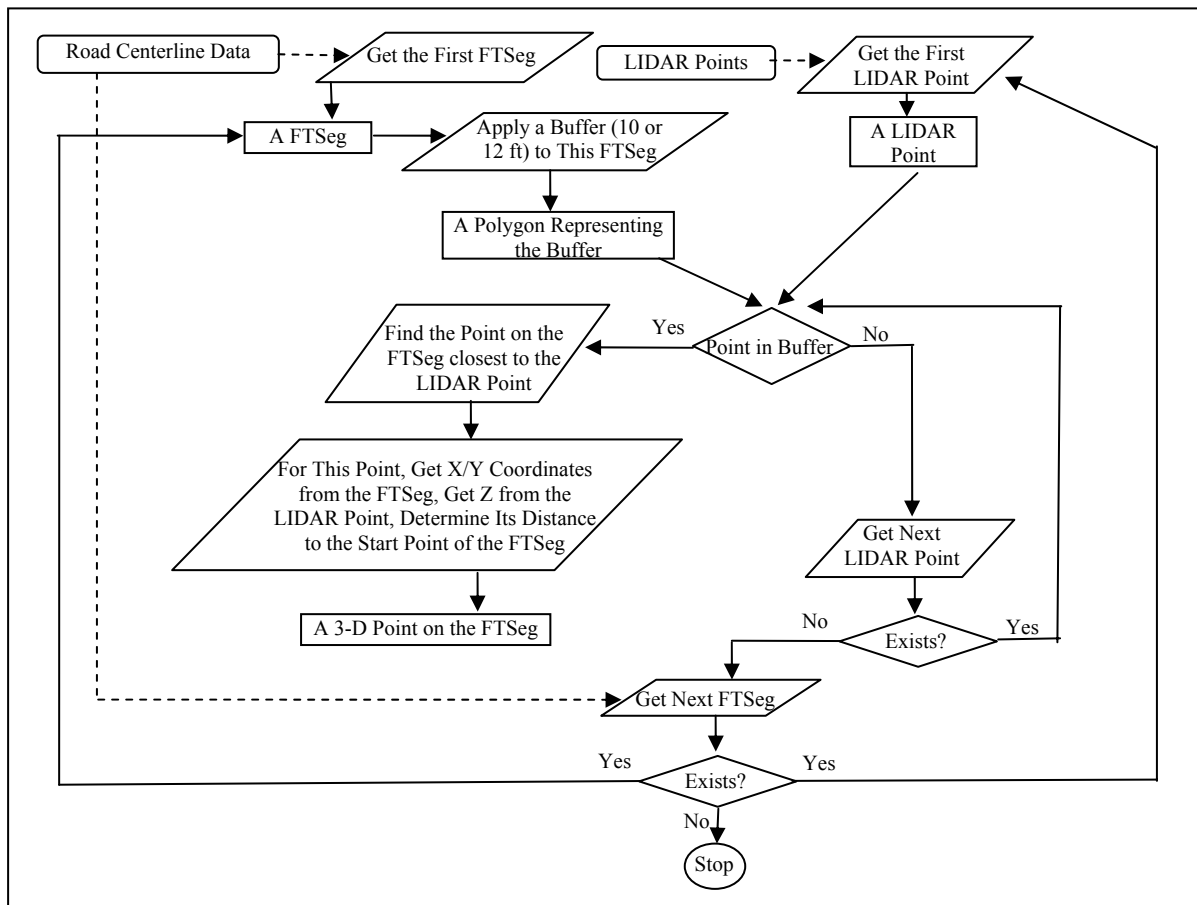


Figure 8.15 An Illustration of the Algorithm for Snapping

This algorithm includes two loops. The outer loop runs through FTSegs. The inner loop runs through all LIDAR points. To summarize, for each FTSeg, all LIDAR points are checked against it to determine if a snapping should be taken. The algorithm is described as below. The developed program code is attached as Appendix B.

- 1) The first FTSeg from the road centerline data is obtained.
- 2) A buffer with appropriate size is applied to this FTSeg.
- 3) The first LIDAR point from the LIDAR point data is obtained.
- 4) It is determined if this LIDAR point is inside the buffer or not.
- 5) If the LIDAR point is not inside the buffer and there are more LIDAR points left, the next LIDAR point is obtained and go to step 4).
- 6) If the LIDAR point is not inside the buffer and there is no more LIDAR point left, go to step 10)
- 7) If the point is inside the buffer, the point on the FTSeg, which is the closet point on the FTSeg to the LIDAR point is found. A new 3-D point is created by obtaining its X/Y coordinates from the point identified on the FTSeg and its 2-D distance along that FTSeg to its start point is derived. If this point coincides with the start or end point of that FTSeg, its elevation is assigned with a value of 999999. Otherwise, its elevation is taken from the corresponding LIDAR point.
- 8) If there are more LIDAR points left, get the next LIDAR point and go to step 4).
- 9) If there are no more LIDAR points left, go to step 10).
- 10) If there are more FTSegs, get the next FTSeg and go to step 2).
- 11) If there is no more FTSeg, stop.

The reason of assigning an elevation value of 999999 to the newly created 3-D point rather than assigning it with the elevation value of the corresponding LIDAR point in step 7) is to deal with scenarios such as the one illustrated in Figure 8.16. In Figure 8.16, LIDAR points A, B, and C are all inside the buffer applied to FTSeg F1. Point D is the end point of F1 and is also the closest point on F1 to points A, B, and C. In other words, the elevation (Z) value of point D could be assigned with the elevation of point A, point B, or point C, the average, the median, the minimum, or the maximum of these elevations. Furthermore, this scenario varies as the number of such LIDAR points and their spatial distribution vary. The number of such points might be 0, which makes any of the above options invalid. It is believed that the procedure described in the next subsection would be more appropriate in determining the elevation of the start and end points than taking an average elevation or choosing one out of many, based on the assumption of the 3-D model being used to model road centerlines in a

3-D space. The value of 999999 is chosen because there is no such elevation value so that 999999 could be used to signal the start and end points.

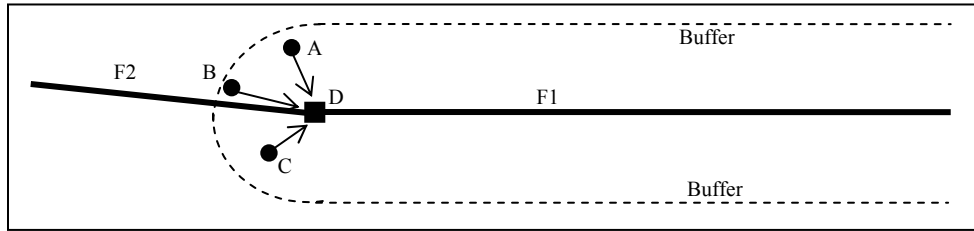


Figure 8.16 Illustration of the Dilemma in Determining Elevations for Start and End Points

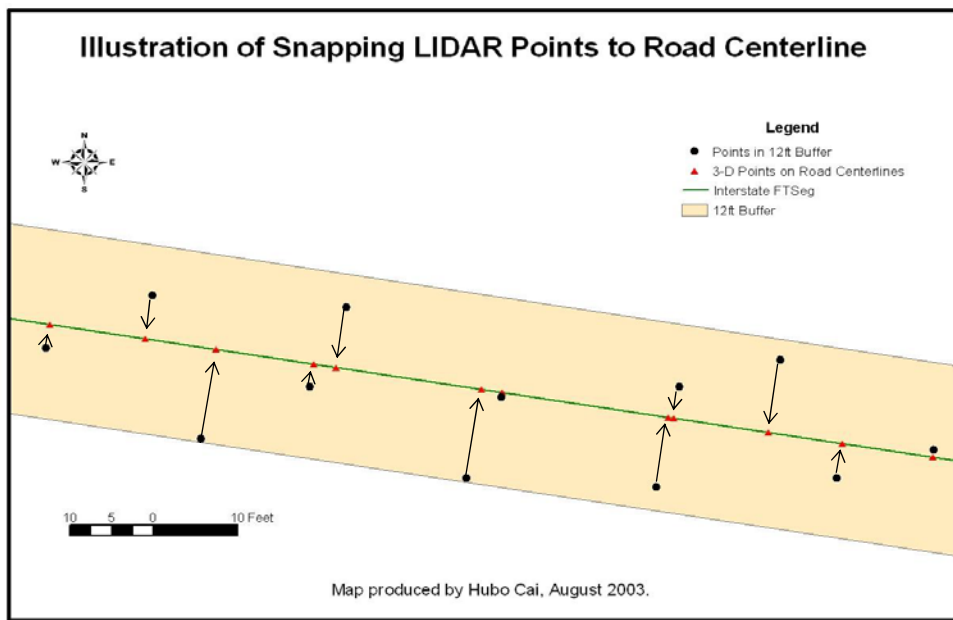


Figure 8.17 Illustration of Sample Results from Snapping

Figure 8.17 shows some sample results of this snapping for a short Interstate (I-40/85) road section in the Orange County to illustrate the LIDAR points inside the 12-ft buffer and 3-D points on the road centerline after snapping.

8.4.1.2 Obtaining Elevations at Start/End Points

One key assumption to the 3-D model chosen to represent road centerlines is that a road centerline can be approximated with a set of 3-D points connected with straight lines. In other words, the elevation for a point

between two neighboring 3-D points can be linearly interpolated. Similarly, the elevation of the start or end point of an FTSEg could be interpolated from neighboring points.

There are three approaches that were taken in this case study to obtain elevations for a start or end point of an FTSEg, depending on the number of neighboring FTSEgs connected at this point, as illustrated in Figure 8.18. In Figure 8.18, case A illustrates a scenario in which there are two neighboring FTSEgs (F1 and F2) connected at point C. Case B illustrates a scenario in which there is only one FTSEg connected to point C. Case C illustrates a scenario in which there are more than two neighboring FTSEgs connected at point C. An interpolation approach will be taken for case A. An extrapolation approach will be taken for case B. A weighted average approach will be taken for case C, in which, the elevation of a neighboring point is weighted according to its planimetric distance to the point whose elevation is being estimated, along the road centerline. In all cases, point C is the end point of FTSEg F1 and is the point whose elevation is being estimated.

In case A, Point A is the neighboring point to C on F1. Point B is the neighboring point to C on F2. Assuming point A has an elevation of E_A , point B has an elevation of E_B , the planimetric distance between points A and C along F1 is d_1 , and the planimetric distance between points B and C along F2 is d_2 , the elevation at point C is interpolated as $E_A + (E_A - E_B) * d_1 / (d_1 + d_2)$.

Case B shows that there is only one neighboring point to the end point C of F1. In other words, there is not another FTSEg F2 connected to F1 at point C and consequently, there is not another neighboring point on F2. However, there is another point D on F1, which is the second closest point to C on F1. In case B, assuming point A has an elevation of E_A , point D has an elevation of E_D , the planimetric distance between points A and D along F1 is d_1 , and the planimetric distance between points A and C along F1, the elevation for point C is extrapolated as $E_D + (E_A - E_D) * (d_1 + d_2) / d_1$.

In case C, there are two FTSEgs (F2 and F3) connected to F1 at point C. point E is the neighboring point to C on F3. Assuming point E has an elevation of E_E and the planimetric distance between points E and C along F3

is d_3 , the weighted average approach estimates the elevation for point C as $E_A * d_1 / (d_1 + d_2 + d_3) + E_B * d_1 / (d_1 + d_2 + d_3) + E_E * d_1 / (d_1 + d_2 + d_3)$.

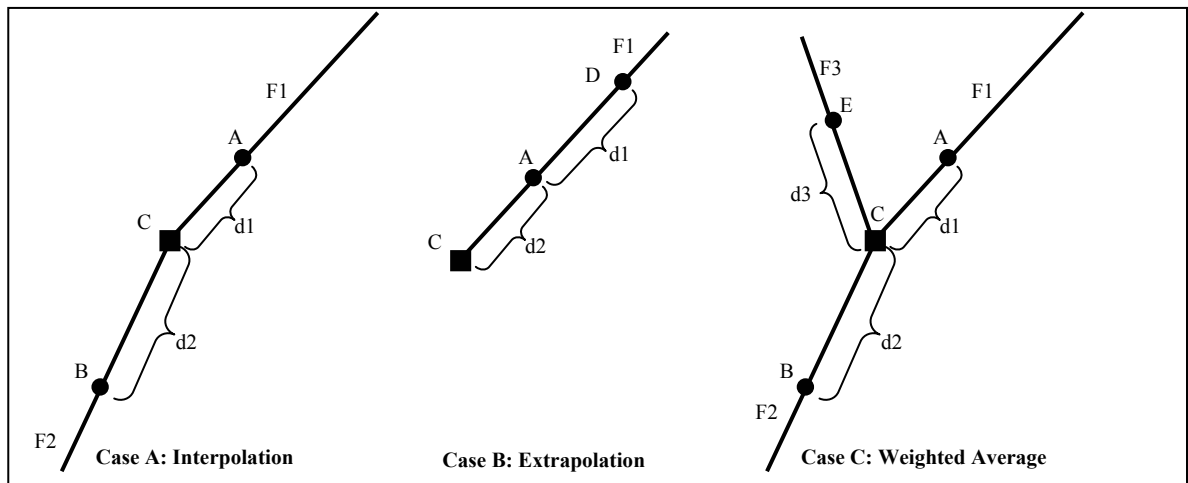


Figure 8.18 Illustration of Interpolation, Extrapolation, and Weighted Average Scenarios for End Points

The elevation for a start or end point is obtained via either an interpolation approach, and extrapolation approach, or a weighted average approach, depending on the number of neighboring FTSEgs connected at that particular start or end point. This step was completed manually in the case study. More specifically, the start and end points were identified via their attributes (a value of 999999 was assigned to all start and end points). Neighboring FTSEgs connected at a specific start or end point and corresponding neighboring points were visually identified in ArcMap. And finally, the elevation for this specific start of end point was estimated via either an interpolation approach, an extrapolation approach, or a weighted average approach, depending on the number of FTSEgs connected at this point.

This step was conducted by road types, i.e. when working with the start/end points of Interstate FTSEgs, only connected Interstate FTSEgs were counted for, when working with the start/end points of US FTSEgs, only connected US FTSEgs were counted for, and when working with NC FTSEgs, only connected NC FTSEgs were counted for. For example, in Figure 8.19, two Interstate FTSEgs (F1 and F2) and two US FTSEgs (F3 and F4) are planimetrically connected at point A. It is assumed that F1 ends at point A, F2 starts from point A, F3 ends

at point A, and F4 starts from point A. When working with Interstate FTSEgs, this belongs to the case A illustrated in Figure 8.18, i.e. only two FTSEgs are connected at point A.

When working with US FTSEgs, this also belongs to case A illustrated in Figure 8.18, or there are only two FTSEgs connected at point A. A total of 4 3-D points are obtained: one (P1) serves as the end point of F1; one (P2) serves as the start point of F; one (P3) serves as the end point of F3; the last one (P4) serves as the start point of F4. All 4 points have the same planimetric location (the same X/Y-coordinates). P1 and P2 would have the same elevation (Z-coordinate) interpolated from neighboring 3-D points (B and C) on Interstate FTSEgs. P3 and P4 would have the same elevation interpolated from neighboring 3-D points (E and D) on US FTSEgs. But these would be two different values, i.e. P1 and P3 would have different elevations.

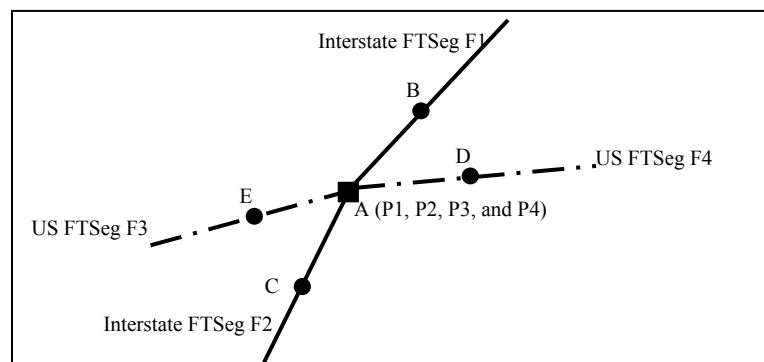


Figure 8.19 Illustration of Connected FTSEgs with Different Road Types

After the snapping step described previously and this step, each FTSEg in the study scope of this case study is completely covered by a set of 3-D points starting from its start point and ending at its end point. Each 3-D point is associated with the following information: X/Y/Z-coordinates, the identity of the FTSEg to which it belongs, the planimetric distance along the road centerline from this 3-D point to the start point of the FTSEg to which this 3-D point belongs, and the 2-D planimetric distance of that particular FTSEg.

8.4.1.3 Quality Control

As stated earlier, roads are man-made objects. Road centerlines are continuous smooth lines in a 3-D space. In the case of modeling a 3-D line with a series of 3-D points on that line, these 3-D points would follow the general trend of the line being modeled. In other words, plotting these 3-D points on a two-dimensional space such as X/Y plane, X/Z plane, and Y/Z plane, these points would illustrate the general trend of projecting the 3-D line onto the particular plane. Similarly, when plotting these 3-D points using a planar coordinate system with elevation (Z) as the vertical axis and the distance to the start point along the road centerline, these points would simulate a smooth line using an appropriate scale. An example is provided in Figure 8.20 (a NC route FTSeg with identity of 6 in Johnston County). The vertical axis is the elevation. The horizontal axis represents the planimetric distance from a 3-D point to the start point of the FTSeg along the road centerline of that FTSeg.

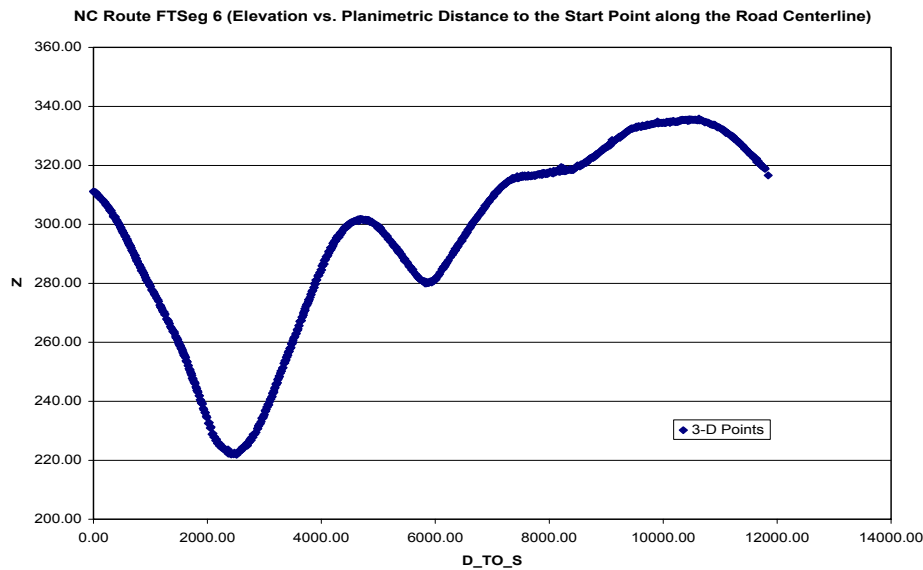


Figure 8.20 An Example of 3-D Points Simulating the Road Centerline

In this case study, such charts as the one shown in Figure 8.20 were obtained for all FTSegs. It revealed that it was not always the case that a series of 3-D points would simulate a smooth line. Some points were not following the general trend of a smooth line as shown in Figure 8.21. These points are defined as suspect points and require further investigations. It was found that most of the suspect points were either the first few points (including the start point) or the last few points (including the end point) of an FTSeg.

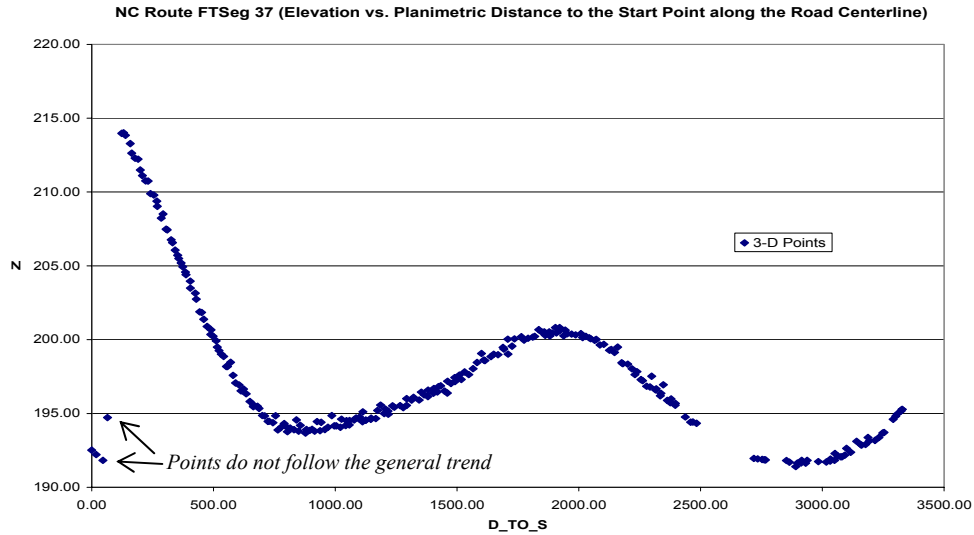


Figure 8.21 Illustration of Suspect Points

Examining the road centerline data and the 3-D points revealed that these suspect points were associated with a scenario illustrated in Figure 8.22. In Figure 8.22, there are 4 FTSegs connected planimetrically at point P1 and P2. P1 serves as the end point of F1 and the start point of F2. P2 serves as the end point of F3 and the start point of F4. P1 and P2 have the same planimetric location (same X/Y coordinates) but different elevations (assuming it's an overpass/underpass, F1 and F2 are actually connected via a bridge). L1 and L2 are two LIDAR points that are actually located on the pavement surface of F3 and are snapped to the F3 as points E1 and E2, respectively. However, since L1 and L2 are also inside the buffer applied to F1 and F2 and therefore, they are also mistakenly snapped to F1 as points D1 and D2. Similarly, points like L3 and L4 on the pavement surface of F4 might also be mistakenly snapped to F2 as points D3 and D4, respectively. The Elevation of point P1 is interpolated from D2 and D3 and consequently, resulting in suspect points D1, D2, and P1 in the chart of F1 and suspect points P1, D3, and D4 in the chart of F2. In case of the scenario as illustrated in Figure 8.21, the suspect points are identified from charts as shown in Figure 8.20. Elevations for these suspect points are linearly interpolated from points D5 and D6 with correct elevations.

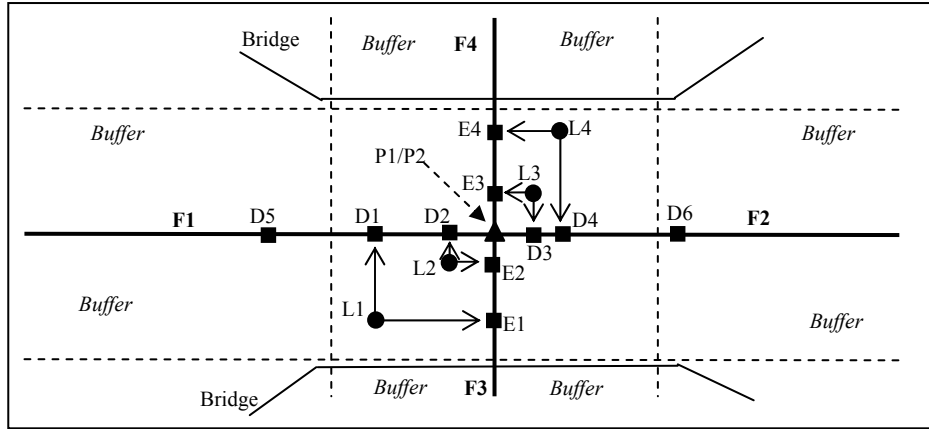


Figure 8.22 Illustration of a Typical Scenario Associated with Suspect Points

This quality control measure could be easily taken because the suspect points could be identified visually from charts such as the one shown in Figure 8.20 and elevations for these suspect points could be linearly interpolated from neighboring points with correct elevations.

8.4.1.4 Improvement

In addition to the quality control measure described in the previous subsection, there is another problem that provides an opportunity of further improvement. The scenario is illustrated in Figure 8.23. In Figure 8.23, the road segment has a curve. Points L1, L2, L3 and L4 are three LIDAR points. L1 and L2 are on the left side of the curve (inner side of the curve) while L3 and L4 are on the right side of the curve (outer side of the curve). Due to superelevation, L3 and L4 would have much higher elevations than L1 and L2. Snapping L1, L2, L3, and L4 to the road centerline results in 4 3-D points (D1, D2, D3, and D4) on the road centerline. D1 and D3 are planimetrically close to each other, but will have significantly different elevations (which is in contradiction with road centerline geometry in reality). Similarly, D2 and D4 are planimetrically close to each other but differ significantly in the elevations. It is believed that it would be more appropriate to use points A1 and A2 rather than using points D1, D2, D3, and D4 to model the road centerline in a 3-D space, in which, A1 is the middle point between D1 and D3 and A2 is the middle point between D2 and D4. The elevation of point A1 is determined as the average of the elevations of D1 and D3. The elevation of point A2 is the average of the elevations of D2 and D4.

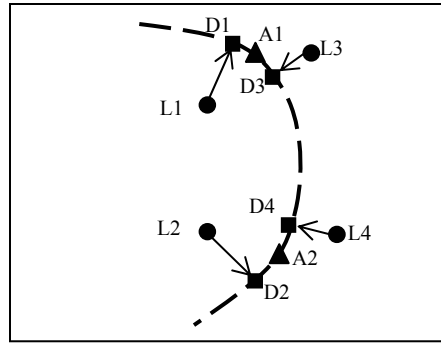


Figure 8.23 Illustration the Scenario for Further Improvement

Based on the above illustration, a further improvement measure was taken in this case study. This improvement works in such a way that wherever two points are close enough to each other on the road centerline of an FTSEg (except that one of these two points is the start point or the end point of that particular FTSEg), a new 3-D point would be generated from these two points based on an averaging approach. This new 3-D point would be used to model the road centerline in a 3-D space while the two original points would be discarded. For example, with points D1 and D2 on FTSEg F1, which are two points close to each other, a new 3-D point A would be generated. The elevation of point A would be the average of the elevations of points D1 and D2. The planimetric location of point A would be determined as the point between D1 and D2 on F1, which has its planimetric distance to the start point of F1 along F1 equivalent to the average of those two distances associated with D1 and D2.

Before this improvement measure could be applied, there is one more question need to be answered, i.e. how close should be deemed as close enough. The criteria were determined based on the average planimetric distances along the road centerline between two neighboring 3-D points. The average planimetric distance was calculated as the total of planimetric distances of road centerlines divided by the total number of 3-D points. According to Table 8.5, average planimetric distances vary by county and by road types. However, the average planimetric distances for 3-D points on Interstate highways in 9 counties are quite close with an overall average of 9.69 feet. The average planimetric distance for 3-D points on US routes is 9.77 feet, which is very close to that of 3-D points on Interstate highways. The average planimetric distance for 3-D points on NC routes is 11.74 feet. Based on this information, it was determined the threshold value is 3 feet for 3-D points on

Interstate highways or US routes and 4 feet for 3-D points on NC routes. In other words, if two neighboring 3-D points on Interstate highways has the planimetric distance between them along the road centerline is equal to or less than 3 feet, they are considered as close enough and the averaging approach would be taken.

Table 8.5 Average Density of 3-D Points before and after Further Improvement

Road Type	US	NC	Interstate									
			Durham	Gran	Halifax	John	Nash	Orange	Vance	Wake	Wilson	All
Before Further Improvement												
Points	71952	74490	17703	25192	19960	65333	29146	20078	12002	47540	18632	255586
Average Distance	9.77	11.74	10.09	9.89	8.91	9.58	9.51	9.96	10.35	9.87	9.30	9.69
After Further Improvement												
Points	60929	65626	15583	23489	18299	57859	26673	16816	10503	40269	17132	226623
Average Distance	11.54	13.33	11.47	10.61	9.72	10.82	10.39	11.90	11.83	11.65	10.12	10.92

This improvement measure was taken for all 3-D points on Interstate highways, US routes, and NC routes. It is obvious that the number of 3-D points would be reduced as shown in Table 8.5.

8.4.2 Using LIDAR DEMs

According to the conceptual and computational models described earlier, the major point of constructing 3-D models for road centerlines is to obtain 3-D points along road centerlines with a uniform interval, starting from the start point of an FTSeg. In other words, the planimetric distance between two neighboring points along the road centerline is constant except that distance between the end point and one right before it because it is rarely that a FTSeg would have its planimetric distance equaling to the interval multiplied by an integer. Similar to constructing 3-D models using LIDAR points, the planimetric location (X/Y-coordinates) are determined by the road centerlines. The difference is that these 3-D points are uniformly distributed along road centerlines.

The elevation of each of these 3-D points is interpolated from the DEMs. DEMs are in the format of grid files consisting of regularly shaped cells. The downloaded LIDAR DEMs being used in the case study are composed of rectangular cells with the same length and width (right rectangles). The cell size is defined as the resolution. The downloaded 50-ft DEMs have a resolution of 50 feet while the 20-ft DEMs have a resolution of 20 feet. Each cell has one elevation value averaging 25 LIDAR points in that cell.

The major concern is the interval and the interpolation method to be used in obtaining 3-D points along road centerlines. It was determined that the bilinear interpolation would be used in interpolating elevations from DEMs. In choosing an appropriate interval, the general rule is to use an interval as small as possible. However, when this interval reaches the size of cell resolution, there would be no more improvements if smaller intervals are used. In this case study, it was determined that when using 20-ft DEMs, the interval would be 20 feet and when using 50-ft DEMs, the interval would be 50 feet. In other words, the appropriate interval was determined based on the resolution of the DEMs being used. In addition, for comparison purposes, intervals at the half cell size would be also used to model road centerlines in a 3-D space. More specifically, for 20-ft DEMs, an interval of 10 feet would also be used while for 50-ft DEMs, an interval of 25 feet would also be used.

Based on the previous discussion, four point models would be constructed for a specific FTSeg when using LIDAR DEMs. The first one would be based on a 10-foot interval and the 20-ft DEMs. The second would be based on a 20-foot interval and the 20-ft DEMs. The third would be based on a 50-foot interval and the 50-ft DEMs. And the fourth model would be based on a 25-ft interval and the 50-ft DEMs.

Figure 8.24 illustrates the algorithm being used to construct 3-D point models based on a specified interval and a specific set of LIDAR DEMs. The algorithm is summarized as below. The developed program code to implement this algorithm is attached as Appendix C.

- 1) The first FTSeg is obtained from the road centerline.
- 2) The start point of the FTSeg is obtained with X/Y coordinates.
- 3) Obtain the elevation for this point from LIDAR DEM via bilinear interpolation. A 3-D point (P1) with X/Y/Z coordinates is obtained.
- 4) Find the next point on the FTSeg, which has its planimetric distance to point P1 obtained in the previous step along the road centerline equaling to the given interval.
- 5) If such a point is found, obtain its X/Y coordinates from the road centerline and go to step 3).
- 6) If such a point is not found, the end point of that particular FTSeg is obtained and its elevation interpolated from the LIDAR DEM.

- 7) Find the next FTSeg from the road centerline data. If found, get this FTSeg and go to step 2). If not found, stop.

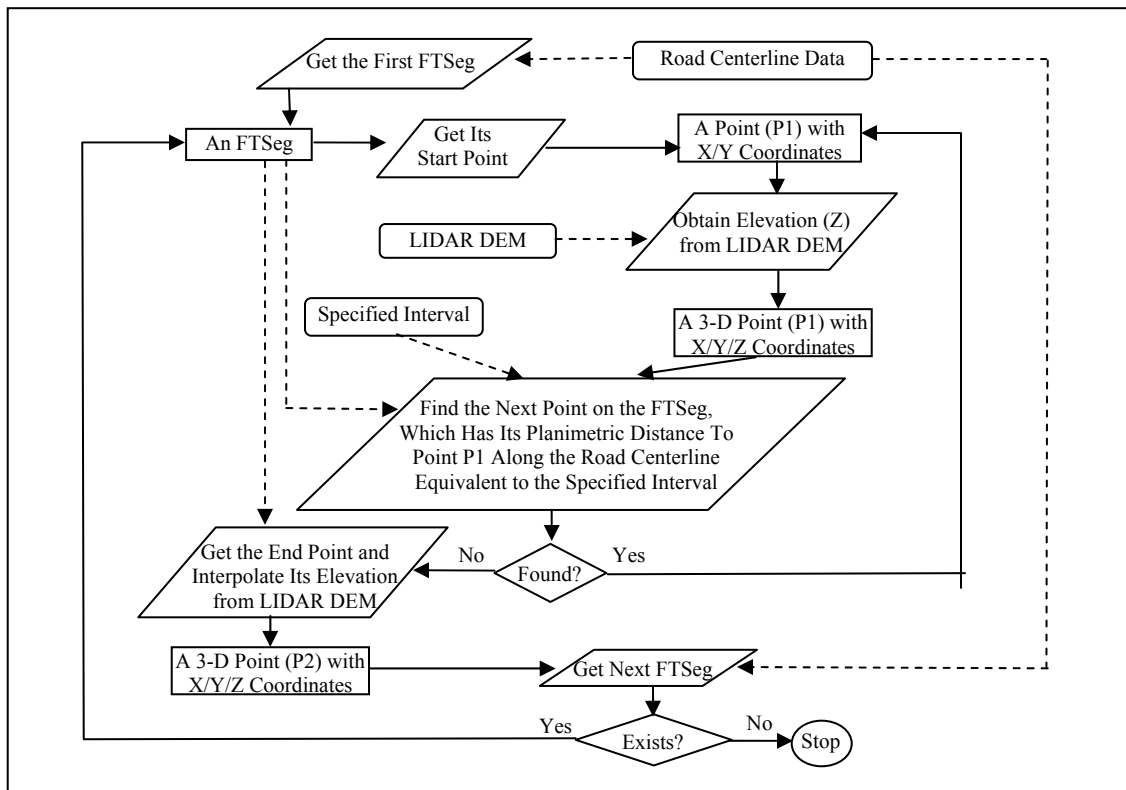


Figure 8.24 Algorithm of Working with LIDAR DEMs

This algorithm was carried out in this case study county by county and road type by road type. In other words, the algorithm was carried out for US routes in Johnston County, NC routes in Johnston County, Interstate highways in Johnston, Interstate highways in Durham County, and etc. It is worth pointing out here that the cell values in the 20-ft DEMs are not elevations, rather, they are 100 * elevations as described in pre-processing section. When obtaining elevations from 20-ft DEMs, this was taken into consideration. After interpolation, the obtained values were divided by 100 so that the 3-D points would have true elevations stored as their Z-coordinates.

By carrying out this algorithm, each FTSeg in the study scope of this case study is completely covered by a set of 3-D points starting from its start point and ending at its end point. These 3-D points are planimetrically uniformly distributed along road centerlines (except the end points). Similar to the 3-D points generated when

using LIDAR point data as the elevation dataset, each 3-D point is associated with the following information: X/Y/Z-coordinates, the identity of the FTSeg to which it belongs, the planimetric distance along the road centerline from this 3-D point to the start point of the FTSeg to which this 3-D point belongs, and the 2-D planimetric distance of that particular FTSeg.

8.4.3 Using NED

NED is essentially a DEM that consists of regularly shaped cells. The pre-processed NED in this case study is in the same format as the pre-processed LIDAR DEMs except the cell size, or the resolution. NED data in this case study has a resolution of 30 meters (approximately 92 feet). Consequently, the intervals used when working with NED were 15 meters (half of the cell size) and 30 meters (full cell size).

The algorithm being used to work with NED is the same as the one illustrated in the previous section and therefore, it is not repeated in this section. It is worth pointing out that the elevations in NED are in the unit of meters. For consistency purpose, this was taken into consideration when obtaining Z-coordinates for 3-D points. More specifically, the elevations of 3-D points were all converted from meter to foot so that the X/Y/Z-coordinates for those 3-D points being used to model road centerlines in a 3-D space all took the same measurement unit, foot, regardless of the different elevation datasets being used.

Similar to working with LIDAR point data and LIDAR DEMs, this algorithm was also carried out county by county and road type by road type when working with NED. After finishing this, each FTSeg in the study scope of this case study is completely covered by a set of 3-D points starting from its start point and ending at its end point. Each 3-D point is associated with the following information: X/Y/Z-coordinates, the identity of the FTSeg to which it belongs, the planimetric distance along the road centerline from this 3-D point to the start point of the FTSeg to which this 3-D point belongs, and the 2-D planimetric distance of that particular FTSeg.

To summarize, each FTSeg in the study scope of this scope would have seven point models to model it in a 3-D space. Among these seven point models, one was constructed using LIDAR point data, one was constructed

using LIDAR 20-ft DEMs with an interval of 10 feet (half of the cell size, or resolution); one was constructed using LIDAR 20-ft DEMs with an interval of 20 feet (full cell size, or resolution); one was constructed using LIDAR 50-ft DEMs with an interval of 25 feet (half of the cell size, or resolution); one was constructed using LIDAR 50-ft DEMs with an interval of 50 feet (full cell size, or resolution); one was constructed using NED with an interval of 15 meters (half of the cell size, or resolution); and one was constructed using NED with an interval of 30 meters (half of the cell size, or resolution),

8.5 3-D DISTANCE PREDICTION

After 3-D point models are constructed, each FTSeg in the case study scope is represented with seven point models, each of which has a different group of 3-D points. These 3-D points are planimetrically uniformly distributed along the road centerline if LIDAR DEMs or NED is used. When LIDAR point data are used, these 3-D points are not planimetrically uniformly distributed. Regardless of whether they are planimetrically distributed or not, these 3-D points will have the same set of attributes. Table 8.6 provides a sample attribute table, which contains all 3-D points belonging to a short FTSeg, obtained when working with LIDAR point data. Table 8.7 provides another sample data for the same FTSeg, obtained when working with LIDAR 20-ft DEM with a 20-ft interval. Only portions of all attributes are shown here for 3-D points. Attributes X, Y, and Z represent the location of the 3-D points. D_TO_S is the distance from a 3-D point to the start point of the FTSeg, to which it belongs, along the road centerline. All points belonging to the same FTSeg have the same values of DIST2D, which is the planimetric distance of that particular FTSeg, ROUTE, which identifies the corresponding LRS route. MERGE column holds the ID of the FTSeg.

Table 8.6 Sample 3-D Point Attributes (from LIDAR Point Data)

X	Y	Z	D_TO_S	DIST2D	ROUTE	MERGE
2131875.52	595999.45	263.49	0.00	58.34	30000027	1
2131883.71	595995.14	263.62	9.26	58.34	30000027	1
2131887.12	595993.34	263.07	13.11	58.34	30000027	1
2131903.74	595984.59	263.39	31.89	58.34	30000027	1
2131907.13	595982.80	262.59	35.73	58.34	30000027	1
2131923.57	595974.13	263.02	54.31	58.34	30000027	1
2131927.14	595972.26	262.24	58.34	58.34	30000027	1

Table 8.7 Sample 3-D Point Attributes (from LIDAR 20-ft DEM with a 20-ft Interval)

X	Y	Z	D TO S	DIST2D	ROUTE	MERGE
2131875.52	595999.45	263.39	0.00	58.34	30000027	1
2131893.22	595990.13	263.08	20.00	58.34	30000027	1
2131910.91	595980.80	262.84	40.00	58.34	30000027	1
2131927.14	595972.26	262.73	58.34	58.34	30000027	1

3-D distances for FTSegs after the 3-D point models are constructed are predicted based on the attributes of 3-D points as previously shown in Tables 13 and 14. The concept was described in detail earlier in Chapter 6. It approximates the 3-D road centerline with a group of 3-D points connected to each other with straight lines. When applying Equation (4), which takes into consideration of planimetric curvature and was described in Chapter 6, the 3-D distance for the FTSeg with identity of 1 (MERGE = 1) shown in Tables 13 and 14 can be computed as:

$$\begin{aligned} & \sqrt{((9.26 - 0.00) ^ 2 + (263.62 - 263.49) ^ 2)} + \sqrt{((13.11 - 9.26) ^ 2 + (263.07 - 263.62) ^ 2)} + \sqrt{((31.89 - 13.11) ^ 2 + (263.39 - 263.07) ^ 2)} \\ & + \sqrt{((35.73 - 31.89) ^ 2 + (262.59 - 263.39) ^ 2)} + \sqrt{((54.31 - 35.73) ^ 2 + (263.02 - 262.59) ^ 2)} \\ & + \sqrt{((58.34 - 54.31) ^ 2 + (262.24 - 263.02) ^ 2)} = 58.54, \end{aligned}$$

and

$$\begin{aligned} & \sqrt{((20.00 - 0.00) ^ 2 + (263.08 - 263.39) ^ 2)} + \sqrt{((40.00 - 20.00) ^ 2 + (262.84 - 263.08) ^ 2)} \\ & + \sqrt{((58.34 - 40.00) ^ 2 + (262.73 - 262.84) ^ 2)} = 58.34, \text{ respectively.} \end{aligned}$$

Figure 8.25 illustrates the algorithm for 3-D distance prediction based on the 3-D point data similar to those shown in Tables 13 and 14. The algorithm is summarized as below.

- 1) All 3-D points belonging to one FTSeg are retrieved from the 3-D point data.
- 2) These 3-D points are sorted based on their distances to the start point of that FTSeg along that FTSeg, ascending, resulting to a group of sorted 3-D points for that particular FTSeg.
- 3) Set the variables of SUM and DIST to 0. The variable SUM would be used to keep the total of the 3-D distances being calculated while the variable DIST would be used to keep the 3-D distance for the current road section defined as the road centerline between two neighboring 3-D points.
- 4) The first 3-D point coinciding to the start point of that FTSeg is retrieved and assigned to Point 1.
- 5) Find the next 3-D point.

- 6) If it is not found, i.e., there is no more 3-D point left, which belongs to the current FTSeg, the 3-D distance prediction for the current FTSeg is complete and assigned to the current FTSeg. Go to step 9).
- 7) If it is found, it is assigned to Point 2. The 3-D distance between Point 1 and Point 2 is calculated and assigned to DIST. The variable SUM is updated by adding this DIST value to it.
- 8) Assign Point 2 to Point 1 and then assign Point 2 to nothing. Go to step 6).
- 9) If there are still more FTSegs, retrieve all 3-D points belonging to the next FTSeg and go to step 2).
- 10) If there are no more FTSegs, 3-D distance prediction for all FTSegs is completed. The procedure stops.

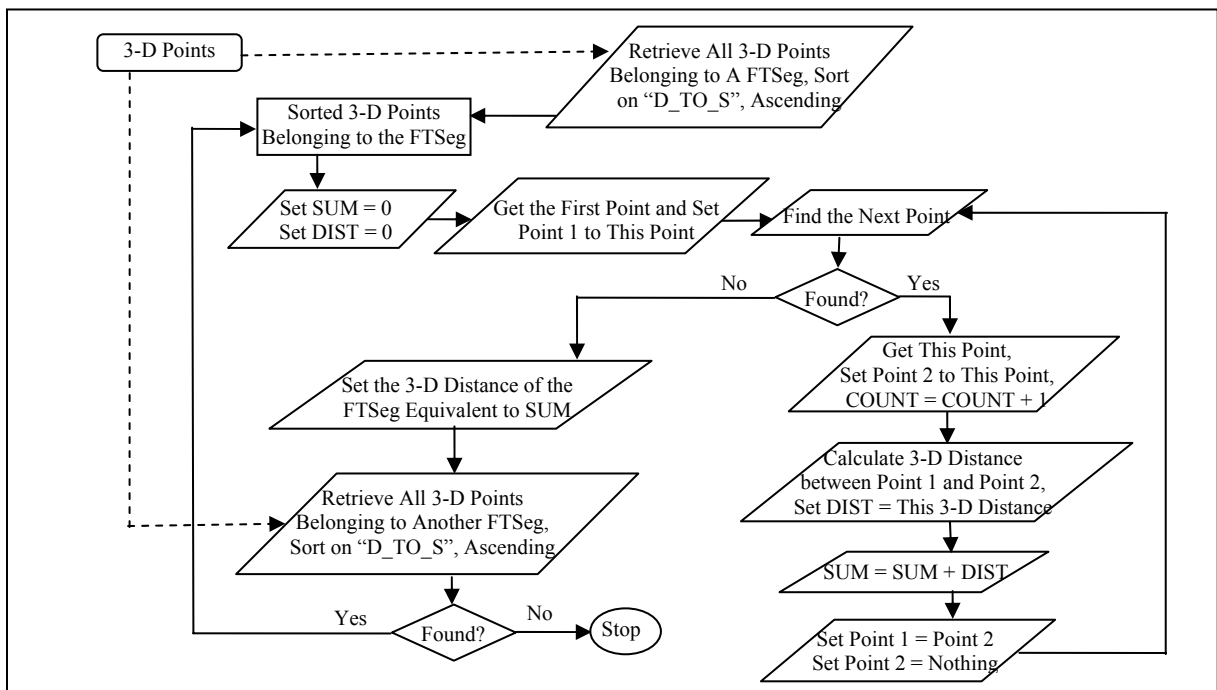


Figure 8.25 Algorithm for 3-D Distance Prediction

This algorithm applies to one 3-D point file, which consists of groups of 3-D points, of which, each group of 3-D points belongs to one particular FTSeg. In the case study, this algorithm was carried out for each FTSeg with each set of its 3-D points, i.e. that FTSeg would have seven predicted 3-D distances since it was modeled with seven point models and each point model was associated with a set of 3-D points.

9 ACCURACY EVALUATION AND ERROR ANALYSIS

After the pre-processing and processing procedures described in chapter 8 are completed, each FTSeg in the study scope of this case study has seven point models depending on the source elevation datasets and the intervals being used when working with DEMs. Each point model for an FTSeg contains a group of 3-D points. Each point model is also associated with a predicted 3-D distance for that FTSeg. In addition, each FTSeg has a reference 3-D distance that was obtained by DMI.

This chapter describes the accuracy evaluation of these predicted 3-D distances by comparing them with corresponding DMI measurements. Brief discussions regarding the statistical methods being used are provided. Accuracy evaluation results are presented. Errors in the source data and errors introduced in modeling are also described. Furthermore, sensitivity analysis is also performed to analyze the results in detail to evaluate the effects on the accuracy of the predicted 3-D distances from varying accuracies and resolutions in the source elevation datasets and varying intervals being used in modeling road centerlines in a 3-D space when working with DEMs.

9.1 COMBINATION OF FTSEGS

As stated earlier, there is a problem with FTSegs touching county boundaries. The location of signs along roads, which delineate two neighboring counties, does not match with the intersections of county line data with road centerline data as illustrated in Figure 8.8. According to data collection procedures described in chapter 8, the road centerlines were digitized county by county based on DOQQs and the digital county line data. However, DMI measures for these FTSegs were obtained via field work, based on the location of county boundary signs along roads. It is inappropriate to evaluate the accuracy of those predicted 3-D distances by comparing them with corresponding DMI measurements, as illustrated in Figure 9.1. Based on the county line, FTSeg F1 starts from point A and ends at point B. FTSeg F2 starts from point B and ends at point C. Consequently, the predicted distance for F1 is the 3-D distance for road section AB and the predicted distance for F2 is the 3-D distance for road section BC. However, when obtaining DMI measurements, the sign delineating two neighboring counties is located at point D and consequently, the distance measured in field for

F1 is the distance of AD and for F2 it is DC. Obviously, there would be significant errors when comparing the predicted distances with the DMI measured distances because distances of two different road sections are being compared.

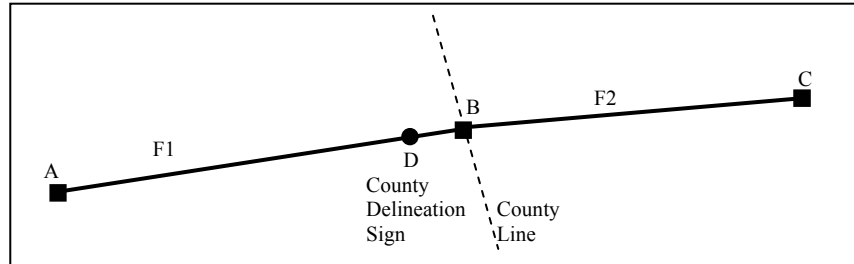


Figure 9.1 The Problem Associated with Accuracy Evaluation for FTSEgs Touching County Boundaries

However, it is clear that if these two FTSEgs are treated as one FTSeg by combining these two FTSEgs into one, this problem is avoided. This combined FTSeg will have its predicted 3-D distance as the sum of the predicted 3-D distances of F1 and F2. The corresponding DMI measurement would also be the sum of the DMI measurements of F1 and F2. The errors introduced due to the misalignment of county delineation signs with digital county lines are thus eliminated. This combined FTSeg would be deemed as a singular FTSeg in the accuracy assessment.

In this case study, all FTSEgs touching county boundaries were identified. For each of these FTSEgs, if there was a neighboring FTSeg in the neighboring county, which was also in the study scope, they were combined into one FTSeg; if there was no neighboring FTSeg in the neighboring county in the study scope, this FTSeg was discarded from accuracy evaluation.

Examining the road centerline data in this case study revealed that there were 38 Interstate FTSEgs touching county boundaries. Among these 38 FTSEgs, 32 of them composed 16 pairs of neighboring FTSEgs connecting at county boundaries and were combined into 16 FTSEgs. The rest 6 did not have neighboring FTSEgs included in the study scope and therefore, were discarded. It was also identified there were 12 NC FTSEgs and 7 US FTSEgs that touched county boundaries. For US and NC routes, the study scope was limited to one county only and therefore, there were no neighboring FTSEgs could be combined with these 12 NC FTSEgs and 7 US

FTSegs. Consequently, these 17 FTSegs were discarded from accuracy assessment. In addition, there were 2 pairs of Interstate FTSegs in Durham County, which had confusions in identifying their connecting points due to the recent modifications of the intersections of US 70 and I 85 in Durham County. Each of these two pairs of FTSegs was also combined into one FTSeg. Similarly 2 pairs of Interstate FTSegs in Johnston County were combined into two FTSegs. Table 9.1 shows the FTSeg distributions before and after this combination procedure.

Table 9.1 FTSeg Distribution before and after Combination

	Road Type	Interstate Highway	US Route	NC Route	All
Before Combination	# of FTSegs	184	78	48	310
	Percentage	59.4%	25.2%	15.4%	100%
After Combination	# of FTSegs	158	71	36	265
	Percentage	59.6%	26.8%	13.6%	100%

9.2 STATISTICAL METHODS APPLIED TO THE CASE STUDY

This section describes the statistical methods being used to evaluate the accuracy of the predicted 3-D distances when compared with the reference data. The methods being used in this case study include descriptive statistics, statistical inferences, frequency analysis, 100% and 95% RMSEs, and Analysis of Variance (ANOVA).

9.2.1 Describing Samples

From the statistical point of view, a population is defined as a collection of measurements about which an inference is desired while a sample consists of a finite number of measurements chosen from a population (Rao 1998). A number calculated using all the measurements in a population is called a parameter of that population while a number calculated using all the measurements in a sample is called a statistic (Rao 1998). The difference between a parameter and a statistic is that a parameter is calculated using all the measurements in a population while the statistic uses only the sample values.

In our case study, any group of errors being evaluated is a sample rather than a population. A sample can be described with numerical values (statistics) or graphically (histogram and box-plot, for example). Descriptive statistics are measures used to describe a sample (Rao 1998). The descriptive statistics being used in the case study in evaluating accuracies and errors include mean, median, and sample percentiles for describing the central tendency of a sample, standard deviation and variance for describing the dispersion of a sample, and skewness for describing the symmetry of a sample distribution.

The sample mean is the average of all measurements in a sample. The sample median is such a value that no more than half of the measurements in the sample are greater than or equal to it and no more than half of the measurements in the sample are less than or equal to it. A percentile is defined in a similar way. For example, the 25 percentile is defined as such a value that no more than 25% of the measurements in the sample are less than or equal to it and no more than $(100\% - 25\%) = 75\%$ of the measurements in the sample are greater than or equal to it. The 25, 50, and 75 are the most commonly used percentiles and are called the first, second, and the third quartiles (Rao 1998). The second quartile, or the 50 percentile, corresponds to the median of the sample. In this case study, the percentiles of 0 (minimum), 25, 50, 75, and 100 (maximum) were used.

The standard deviation of a quantitative sample is defined as the positive square root of a statistic called the sample variance, which in turn is the average of the squared differences between the measurements in the sample and the sample mean.

According to Microsoft Excel Help, skewness characterizes the degree of asymmetry of a distribution around its mean. The skew is zero when the error distribution is perfectly symmetrical. Positive skewness indicates a distribution with an asymmetric tail extending toward more positive values. Negative skewness indicates a distribution with an asymmetric tail extending toward more negative values. Skewness is the condition of being asymmetrical, and departing from the expected normal distribution shown by the “bell curve.” When the absolute value of the skew exceeds a value such as 0.5 (FEMA’s empirically derived criterion), it means the error distribution is sufficiently asymmetrical to cause concern that the dataset may not represent a normal distribution (Daniel and Tennant 2001).

In addition to the above descriptive statistics, in this case study, samples are also described graphically by histograms. A histogram is a useful tool to describe graphically the frequency distribution of a sample (Rao 1998). It is a bar-graph in which the lengths of the bars represent the proportions of the measurements from a sample in the various categories (Rao 1998).

9.2.2 Statistical Inferences

Statistical inferential procedures utilize samples to draw conclusions about populations (Rao 1998). After getting descriptive statistics for a sample as described in the previous section, conclusions regarding the population parameters can be drawn for the corresponding population. The statistical inferential problems can be classified into one of the following three types (Rao 1998):

- Estimating parameters of target population;
- Testing hypotheses about target populations; and
- Predicting responses that will be observed in the future.

For example, assuming a sample is obtained from a population and the sample mean is obtained, in order to provide information regarding the population mean, the objective of estimation is to determine a set of numbers that can be considered as plausible values for the population mean. A hypothesis test is to verify a claim – a hypothesis – about the true value of the population mean. The objective of prediction is to predict the future response. In this case study, interval estimation (or confidence interval) and hypothesis test were used.

An interval estimation (or confidence interval) of a parameter of a population is an interval with two endpoints as statistics calculated from random samples. The endpoints are selected in such a way that the interval will contain the true value of the population parameter with a predetermined probability that is called the confidence efficient and is usually expressed as $1 - \alpha$, where α is a number between 0 and 1.

Statistical hypothesis tests deal with two hypotheses: a research hypothesis, denoted by H_1 ; and a null hypothesis, denoted by H_0 (Rao 1998). Usually the research hypothesis is a hypothesis that needs confirmation while a null hypothesis is a hypothesis that is true if and only if a research hypothesis is false (Rao 1998). A test of a null hypothesis H_0 is a statistical inferential procedure that can be used to see if the observed data provide sufficient evidence to reject H_0 in favor of H_1 . A hypothesis test has its significance level (or simply level) that is a number α such that the probability that the test will falsely support the research hypothesis does not exceed α . In symbols, α is a number such that

$$\Pr \{\text{test will reject } H_0 \text{ when } H_0 \text{ true}\} \leq \alpha.$$

A hypothesis test works in such a way that first, a null hypothesis and a research hypothesis are designed; second, an appropriate statistic is calculated using the sample data; third, a critical value is obtained based on the sampling distribution of the statistic being calculated, the null hypothesis, and a confidence coefficient; forth, the calculated statistic is compared to the critical value; and finally, a conclusion is reached.

Both interval estimation and hypothesis test deal with obtaining information regarding population parameters while using the corresponding statistics from samples. Consequently, the appropriate estimating and testing procedures are dependent on assumptions regarding the populations and corresponding samples.

9.2.3 ANOVA

The ANOVA is usually performed to evaluate the equality of the expected responses for different treatments in an experimental design (Rao 1998). An ANOVA can be viewed as an extension of pairwise testing. In a pairwise test, the null hypothesis is in a format of $\mu_1 = \mu_2$, where μ_1 is the mean of population 1 and μ_2 is the mean of population 2. With ANOVA, assuming there are seven samples (each represents a treatment), the null hypothesis becomes $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$, where $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6,$ and μ_7 are the means of the populations corresponding to these seven samples. It is obvious that when there are two populations, the result of ANOVA is the same as the pairwise t-test.

The result of an ANOVA is usually represented as the one shown in Table 9.2, in which t is the number of treatments, N is the number of total observations/measurements, $SS[T]$ is the treatment sum of squares, $SS[E]$ is the error sum of squares, $SS[TOT]$ is the total sum of squares, $MS[T]$ is the mean square for treatments, which is calculated as $SS[T] / (t - 1)$, $MS[E]$ is the mean square for errors, which is calculated as $SS[E] / (N - t)$, F is the statistic calculated from the samples for testing the null hypothesis, and F_c is the critical value to which F is compared to reach a conclusion (Rao 1998). Calculations of these values can be found in numerous statistical textbooks and therefore, are not included herein.

Table 9.2 ANOVA Result

Source of variation	Degrees of freedom (df)	Sums of squares (SS)	Mean squares (MS)	F	F _c
Treatment	$t - 1$	$SS[T]$	$MS[T]$	$MS[T]/MS[E]$	$F(t - 1, N - t, \alpha)$
Error	$N - t$	$SS[E]$	$MS[E]$		
Total	$N - 1$	$SS[TOT]$			

There are three assumptions of ANOVA for completely randomized designs (CRD) with fixed treatments. First, the populations have normal distributions. Second, the population variances are equal. And third, the observed responses are independent random samples from populations (Rao 1998). These assumptions are not always the case in reality. As a result, the appropriateness of using ANOVA is dependent on each application. Generally speaking, the ANOVA F-test is not very sensitive to violations of the assumption of a normal distribution (Rao 1998). It is also insensitive to moderate violations of the equal-variance assumption, provided that the sample sizes per group are equal and not too small (Rao 1998).

In this case study, each FTSeg is associated with 8 distances, of which 7 are predicted 3-D distances and 1 is the reference distance (DMI measured distance). From a statistical point of view, each of these 8 distances could be viewed as a treatment, resulting in a total of 8 treatments. ANOVA could be performed to evaluate the equality of the expected responses from these 8 treatments. However, the sample distribution of distances is not a normal distribution (it should not be). Also, the focus here is the accuracy of the predicted 3-D distance, not directly the distance itself. Rather than performing ANOVA on the distances directly, in this case study, ANOVA was performed on the resulting error sets by comparing the predicted distances to the DMI measured

distance. There are 7 error sets, each of which can be deemed as a treatment. Consequently, ANOVA could be performed to test the equality of the expected responses (errors) from these seven treatments.

9.2.4 100% RMSE and 95% RMSE

The statistical methods described in the previous sections are fundamental from a statistical point of view. These methods were not initially designed for use in spatial accuracy evaluation and error analysis. This section describes a method that is being widely used in GISs for accuracy assessment, especially in evaluating the accuracy of point locations by comparing them with a higher accuracy dataset (Daniel and Tennant 2001). For example, in evaluating the vertical accuracy of DEMs, assuming there are n checkpoints and each of these checkpoints has two elevation values, one from DEM and one from geodetic survey. The difference between these two elevations is error and totally, there are n vertical errors from e_1 to e_n . RMSE is calculated using the equation below.

$$\text{RMSE} = [(e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2)/n]^{1/2}; \quad \text{where } e_1, e_2, e_3, \dots, e_n \text{ are errors.} \quad (9.1)$$

In evaluating the positional accuracy (horizontal or vertical) for points, a 100% RMSE is the RMSE calculated from all checkpoints while a 95% RMSE is the RMSE calculated from the “best” 95% of the checkpoints (the other 5% “worst” points are discarded). While RMSEs with other percentages such as 90% can also be used, the 100% and 95% RMSEs are the most commonly used ones in evaluating and reporting point positional accuracy. The percent RMSEs are different from the percentiles in that the percent RMSEs are calculated values based on the differences between two datasets while percentiles are values delineating the values in one dataset.

In this case study, the accuracy being evaluated is the accuracy of predicted 3-D distances, when compared to DMI measurements (the reference data). In other words, it is the difference between two datasets that is being evaluated. From a statistical point of view, RMSE is useful in describing differences between two datasets. This is the main reason that in the spatial world, this statistic is widely used to describe positional accuracy, including horizontal and vertical accuracy and would be used in this case study to evaluate the accuracy of the

predicted 3-D distances. In this case study, 100% and 95% RMSEs would be calculated for those predicted 3-D distances when compared with DMI measured distances.

9.2.5 Frequency Analysis

One more simple, but useful tool in accuracy evaluation is frequency analysis. This method works in such a way that given a set of errors, first, they are ordered from “best” to “worst”; second, ranges are designed, and finally errors in a particular range are counted and the percentage is calculated.

In this case study, frequency analyses would be conducted for each set of the 7 error sets mentioned earlier. For example, in this case study, when evaluating the accuracy of the predicted 3-D distances by using LIDAR point data, these predicted distances are compared to corresponding DMI measurements (differences are errors). Follow the procedure described earlier, the occurrences of errors in designed ranges could be counted and corresponding percentages calculated to finish the frequency analysis.

9.3 ACCURACY EVALUATION RESULTS

This section describes the results of applying some of the previously described statistical methods. More specifically, accuracy evaluation focuses on the accuracy of predicted 3-D distances by comparing them to the corresponding DMI measurements (differences are errors).

9.3.1 General Information

In this case study, each FTSeg has eight 3-D distances:

- DMI measured distance;
- 3-D distance predicted using LIDAR point data;
- 3-D distance predicted using LIDAR 20-ft DEM with a half cell size interval (10-ft);
- 3-D distance predicted using LIDAR 20-ft DEM with a full cell size interval (20-ft);
- 3-D distance predicted using LIDAR 50-ft DEM with a half cell size interval (25-ft);

- 3-D distance predicted using LIDAR 50-ft DEM with a full cell size interval (50-ft);
- 3-D distance predicted using 30-m NED with a half cell size interval (15-m); and
- 3-D distance predicted using 30-m NED with a full cell size interval (30-m).

In this case study, the objective is to evaluate the accuracy of predicted 3-D distances when compared with DMI measurements. DMI measured distances serve as the baseline distances (reference data). Accuracies of all the predicted 3-D distances are evaluated by comparisons with the DMI data. Accuracy evaluation of a particular set of predicted 3-D distances in this case study is done based on the differences between that particular set of predicted 3-D distances and the DMI measurements and the corresponding proportional differences (differences divided by the corresponding DMI measurements). Consequently there were seven sets of differences and seven sets of proportional differences as below. It is these errors, whether expressed as the difference or the proportional difference, that would be evaluated.

- Differences and proportional differences between DMI measured distances and 3-D distances predicted using LIDAR point data;
- Differences and proportional differences between DMI measured distances and 3-D distances predicted using LIDAR 20-ft DEM with a half cell size interval (10-ft);
- Differences and proportional differences between DMI measured distances and 3-D distances predicted using LIDAR 20-ft DEM with a full cell size interval (20-ft);
- Differences and proportional differences between DMI measured distances and 3-D distances predicted using LIDAR 50-ft DEM with a half cell size interval (25-ft);
- Differences and proportional differences between DMI measured distances and 3-D distances predicted using LIDAR 50-ft DEM with a full cell size interval (50-ft);
- Differences and proportional differences between DMI measured distances and 3-D distances predicted using 30-m NED with a half cell size interval (15-m); and
- Differences and proportional differences between DMI measured distances and 3-D distances predicted using 30-m NED with a full cell size interval (30-m).

In addition, FTSegs are categorized based on road types, resulting into four groups: All FTSegs, Interstate highway FTSegs, US routes FTSegs, and NC route FTSegs. The group of All FTSegs includes Interstate FTSegs, US FTSegs, and NC FTSegs. As a result, there are a total of 28 samples of errors (differences) and a total of 28 samples of proportional errors (proportional differences) in this case study as illustrated in Figure 9.2.

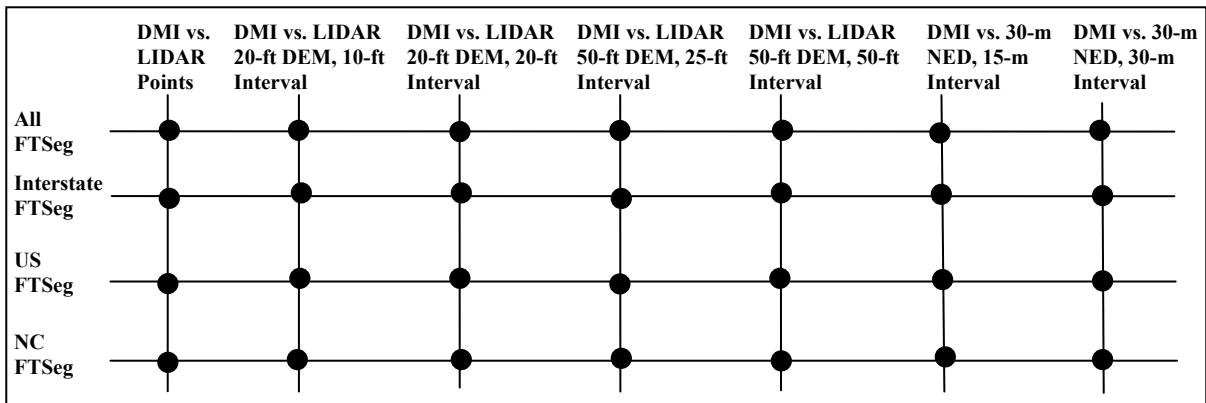


Figure 9.2 Grouping Errors into Samples

In this case study, descriptive statistics and histograms for each sample in Figure 9.2 will be derived, from both aspects of the difference and the proportional difference. Based on the sample data, estimates of the corresponding population means will be obtained. 100% and 95% RMSEs based on the difference and the proportional difference will be calculated for each sample. And finally, frequency analyses for each sample will be performed for both the difference and the proportional difference.

9.3.2 Accuracy Assessment

This section presents the results of conducting accuracy assessment for each set of predicted 3-D distances by applying some of the statistical methods described earlier. Whenever a confidence coefficient α is needed, α is assumed to be 0.05, which is most commonly used from a statistics point of view. This section simply represents the results without analyzing them in detail. Comprehensive and detailed analyses of these results will be represented in the next section.

9.3.2.1 Descriptive Statistics

This subsection presents the results of descriptive statistics in the case study for the differences and proportional differences between a particular set of predicted 3-D distances and the DMI measured distances. The purpose of doing this is to quantitatively describe all the samples in the case study. The information presented in this subsection would be used for further analyses as described in the next sections. The unit for differences is foot. The unit for proportional differences is foot per thousand feet (foot/1,000 feet).

(1) 3-D Distances predicted Using LIDAR Point Data versus DMI Distances

The descriptive statistics for both differences and proportional differences between the 3-D distances predicted using LDIAR point data and the DMI measured distances are shown in Table 9.3.

Table 9.3 Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (LIDAR Point Data) and DMI Measurements

Error Format	Road Type	Central Tendency		Sample Percentiles				Dispersion		Skew
		Mean	Median	0	25	75	100	Standard Deviation	Variance	
Differences	All	-8.43	-4.94	-71.40	-26.22	7.55	53.48	24.28	589.33	-0.30
	*Inter	-9.93	-5.53	-71.40	-26.21	5.85	48.29	22.76	517.96	-0.39
	**US	-10.81	-8.68	-71.15	-28.16	6.63	40.61	26.29	691.38	-0.19
	***NC	2.86	5.29	-54.38	-11.55	21.67	53.48	24.21	586.26	-0.41
Proportional Differences	All	-6.48	-0.72	-352.48	-305.48	-78.41	292.61	50.12	2511.56	-1.71
	*Inter	-1.17	-0.69	-164.46	-2.94	0.72	114.12	32.14	1033.13	-1.20
	**US	-15.02	-1.63	-352.48	-264.84	-201.84	85.54	62.36	3888.81	-3.02
	***NC	-12.92	0.36	-235.63	-12.81	1.18	292.61	78.83	6214.30	0.50

*Inter indicates Interstate FTSEgs **US indicates US FTSEgs ***NC indicates NC FTSEgs

A perfect symmetrical distribution would have 0 as the value for the skew. A high skew value indicates there is a large variation between the mean and the median. Based on Table 9.3, it is observed that the samples representing proportional differences could all be viewed as sufficiently asymmetrical, using ± 0.5 as the criterion (FEMA's empirically derived criterion), causing the concern that the proportional differences may not represent normal distributions.

It is observed that the majority of the differences and proportional differences are negative (as indicated by medians) values indicating underestimates, i.e. the 3-D distances predicted using LIDAR point data are shorter than the DMI measured distances.

It is also observed that the variance of proportional differences were larger than the corresponding variance of differences. In other words, the proportional differences were more dispersed than the differences. This could also be observed when examining the minimum (0 percentile) and maximum (100 percentile) values. The proportional differences have a wider range than the differences.

(2) 3-D Distance Predicted Using LIDAR 20-ft DEM with an Interval of 10-ft versus DMI Distances

The descriptive statistics for both differences and proportional differences between the 3-D distances predicted using LDIAR 20-ft DEM with an interval of 10-ft and the DMI measured distances are shown in Table 9.4.

Table 9.4 Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 10-ft Interval) and DMI Measurements

Error Format	Road Type	Central Tendency		Sample Percentiles				Dispersion		Skew
		Mean	Median	0	25	75	100	Standard Deviation	Variance	
Differences	All	-16.28	-8.99	-105.38	-35.54	3.80	42.46	29.11	847.21	-0.71
	Inter	-18.86	-12.16	-105.38	-36.81	2.50	37.91	27.64	763.71	-0.78
	US	-16.79	-9.82	-100.02	-40.74	5.81	42.46	33.65	1132.17	-0.58
	NC	-3.95	-4.55	-70.18	-13.74	11.45	39.21	22.66	513.41	-0.58
Proportional Differences	All	-6.43	-1.33	-352.63	-305.61	-79.09	292.53	50.97	2597.89	-1.56
	Inter	-1.01	-1.13	-164.55	-3.63	0.25	128.09	33.25	1105.53	-0.83
	US	-15.08	-3.20	-352.63	-265.10	-201.11	105.07	63.66	4053.20	-2.77
	NC	-13.16	-0.26	-235.92	-9.17	0.82	292.53	78.73	6198.73	0.51

Based on Table 9.4, it is observed that the samples representing differences and proportional differences could all be viewed as sufficiently asymmetrical, using ± 0.5 as the criterion causing the concern that the differences and proportional differences may not represent normal distributions.

It is observed that the majority of the differences and proportional differences are negative values indicating underestimates, i.e. the 3-D distances predicted using LIDAR 20-ft DEM data and a 10-ft interval are shorter than the DMI measured distances.

It is also observed that the variance of proportional differences is larger than the corresponding variance of differences. In other words, the proportional differences are more dispersed than the differences. This could

also be observed when examining the minimum (0 percentile) and maximum (100 percentile) values. The proportional differences have a wider range than the proportional differences.

(3) 3-D Distance Predicted Using LIDAR 20-ft DEM with an Interval of 20-ft versus DMI Distances

The descriptive statistics for both differences and proportional differences between the 3-D distances predicted using LDIAR 20-ft DEM with an interval of 20-ft and the DMI measured distances are shown in Table 9.5.

Table 9.5 Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 20-ft Interval) and DMI Measurements

Error Format	Road Type	Central Tendency		Sample Percentiles				Dispersion		Skew
		Mean	Median	0	25	75	100	Standard Deviation	Variance	
Differences	All	-16.82	-9.45	-107.11	-35.98	3.44	42.44	29.29	857.99	-0.72
	Inter	-19.52	-12.43	-107.11	-38.62	2.14	37.55	27.91	779.11	-0.78
	US	-17.03	-9.82	-101.18	-41.32	5.81	42.44	33.78	1141.42	-0.60
	NC	-4.51	-4.98	-71.53	-13.74	9.61	37.70	22.48	505.14	-0.64
Proportional Differences	All	-6.51	-1.34	-352.63	-6.51	0.78	292.53	50.94	2595.12	-1.56
	Inter	-1.11	-1.17	-164.56	-3.71	0.23	127.86	33.21	1102.99	-0.84
	US	-15.15	-3.21	-352.63	-17.52	5.53	105.01	63.64	4049.73	-2.77
	NC	-13.21	-0.28	-235.92	-9.17	0.82	292.53	78.73	6197.96	0.51

Similar observations to those obtained for differences and proportional differences between predicted 3-D distances using LIDAR 20-ft DEM data and a 10-ft interval and the DMI measured distances could be obtained herein for differences and proportional differences between predicted 3-D distances using LIDAR 20-ft DEM and a 20-ft interval and the DMI measured distances. More specifically, differences and proportional differences are sufficiently asymmetrical; the majority of the differences and proportional differences are negative values indicating underestimates; and the proportional differences are more dispersed than the differences.

(4) 3-D distance predicted using LIDAR 50-ft DEM with an Interval of 25-ft versus DMI Distances

The descriptive statistics for both differences and proportional differences between the 3-D distances predicted using LDIAR 50-ft DEM with an interval of 25-ft and the DMI measured distances are shown in Table 9.6.

Table 9.6 Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 25-ft Interval) and DMI Measurements

Error Format	Road Type	Central Tendency		Sample Percentiles				Dispersion		Skew
		Mean	Median	0	25	75	100	Standard Deviation	Variance	
Differences	All	-0.87	-0.74	-87.00	-17.80	13.36	160.13	33.95	1152.59	0.75
	Inter	5.44	2.37	-77.33	-9.37	16.78	160.13	35.54	1262.92	1.12
	US	-15.19	-9.72	-87.00	-38.12	5.42	41.64	30.44	926.40	-0.36
	NC	0.21	-4.05	-62.27	-13.72	13.26	47.44	24.33	591.74	-0.23
Proportional Differences	All	-5.34	-0.06	-352.64	-4.53	3.03	292.45	50.60	2560.72	-1.69
	Inter	0.62	0.31	-164.61	-1.86	3.13	118.49	32.42	1051.23	-1.18
	US	-14.76	-1.97	-352.64	-15.68	7.48	89.00	63.24	3999.50	-2.88
	NC	-12.93	-0.12	-235.99	-9.17	1.01	292.45	78.74	6200.68	0.50

Based on Table 9.6, it is observed that the majority of the differences and proportional differences are negative values indicating underestimates and the proportional differences are more dispersed than the differences. The proportional differences are sufficiently asymmetrical. When all FTSEgs are viewed as one group without discriminating road types, the differences are also sufficiently asymmetrical. However, when road types are under consideration, only the differences for the group of Interstate highway FTSEgs are sufficiently asymmetrical.

(5) 3-D Distance Predicted Using LIDAR 50-ft DEM with an Interval of 50-ft versus DMI Distances

The descriptive statistics for both differences and proportional differences between the 3-D distances predicted using LDIAR 50-ft DEM with an interval of 50-ft and the DMI measured distances are shown in Table 9.7.

Table 9.7 Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 50-ft Interval) and DMI Measurements

Error Format	Road Type	Central Tendency		Sample Percentiles				Dispersion		Skew
		Mean	Median	0	25	75	100	Standard Deviation	Variance	
Differences	All	-3.57	-2.72	-93.24	-23.86	11.13	168.69	34.03	1158.29	1.02
	Inter	1.59	-0.85	-80.78	-15.28	13.79	168.69	35.91	1289.40	1.54
	US	-16.09	-9.72	-93.24	-38.84	5.40	41.55	31.36	983.19	-0.43
	NC	-1.54	-4.40	-65.36	-13.78	12.58	45.46	23.33	544.23	-0.42
Proportional Differences	All	-5.83	-0.39	-352.64	-5.29	2.23	292.45	50.50	2551.44	-1.69
	Inter	-0.10	-0.11	-164.61	-2.26	2.02	118.43	32.34	1045.93	-1.15
	US	-14.91	-2.11	-352.64	-15.72	7.19	88.44	63.15	3987.55	-2.89
	NC	-13.08	-0.17	-236.04	-9.22	0.94	292.45	78.79	6208.51	0.50

The observations based on Table 9.7 are the same as those obtained based on Table 9.6 and therefore, are not repeated herein.

(6) 3-D Distance Predicted Using 30-m NED with an Interval of 15-m versus DMI Distances

The descriptive statistics for both differences and proportional differences between the 3-D distances predicted using 30-m NED with an interval of 15-m and the DMI measured distances are shown in Table 9.8.

Table 9.8 Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (30-m NED with 15-m Interval) and DMI Measurements

Error Format	Road Type	Central Tendency		Sample Percentiles				Dispersion		Skew
		Mean	Median	0	25	75	100	Standard Deviation	Variance	
Differences	All	-18.63	-10.89	-109.64	-38.80	3.14	40.06	30.31	918.97	-0.72
	Inter	-21.56	-13.34	-103.75	-40.44	0.74	39.05	29.46	868.06	-0.69
	US	-18.97	-9.83	-109.64	-45.06	4.93	40.06	33.96	1152.98	-0.64
	NC	-5.14	-5.14	-74.38	-13.79	9.96	33.16	22.49	505.64	-0.82
Proportional Differences	All	-7.48	-1.56	-352.55	-6.71	0.71	292.19	50.22	2522.00	-1.66
	Inter	-2.11	-1.26	-164.63	-3.84	0.17	115.12	32.32	1044.85	-1.08
	US	-16.42	-3.35	-352.55	-17.79	4.73	85.54	62.45	3899.85	-2.96
	NC	-13.36	-0.24	-235.99	-13.31	0.77	292.19	78.80	6208.73	0.51

Based on Table 9.8, it is observed that differences and proportional differences are sufficiently asymmetrical; the majority of the differences and proportional differences are negative values indicating underestimates; and the proportional differences are more dispersed than the differences.

(7) 3-D Distance Predicted Using 30-m NED with an Interval of 30-m versus DMI Distances

The descriptive statistics for both differences and proportional differences between the 3-D distances predicted using 30-m NED with an interval of 30-m and the DMI measured distances are shown in Table 9.9.

Table 9.9 Descriptive Statistics for Differences and Proportional Differences between 3-D Distances (30-m NED with 30-m Interval) and DMI Measurements

Error Format	Road Type	Central Tendency		Sample Percentiles				Dispersion		Skew
		Mean	Median	0	25	75	100	Standard Deviation	Variance	
Differences	All	-18.92	-11.21	-109.07	-38.84	3.14	40.13	30.38	922.80	-0.72
	Inter	-22.01	-13.89	-104.09	-41.29	0.53	36.07	29.62	877.35	-0.69
	US	-18.87	-9.82	-109.07	-45.02	4.96	40.13	33.84	1144.99	-0.63
	NC	-5.41	-5.55	-75.17	-13.79	8.22	32.88	22.46	504.25	-0.84
Proportional Differences	All	-7.49	-1.58	-352.55	-6.73	0.70	292.19	50.22	2521.75	-1.66
	Inter	-2.14	-1.29	-164.63	-3.84	0.15	115.08	32.32	1044.67	-1.07
	US	-16.41	-3.35	-352.55	-17.79	4.74	85.54	62.45	3900.23	-2.96
	NC	-13.38	-0.26	-235.99	-13.33	0.73	292.19	78.79	6208.30	0.51

Observations based on Table 9.9 are the same as those obtained based on Table 9.8 and therefore, are not repeated herein.

Based on the observations, it is concluded that the majority of the predicted 3-D distances are shorter than the DMI measured distances and the proportional differences are more dispersed than the differences, regardless of the elevation datasets and the intervals being used. It is also concluded that the proportional differences are all sufficiently asymmetrical based on the ± 0.5 skew criterion. However, a few samples of differences are not sufficiently asymmetrical based on this skew criterion, if the discrimination of road types is considered.

It is worth pointing out that based on the descriptive statistics, the observations for tables using the same elevation dataset but different intervals are very close. For example, observations based on Tables 9.6 and 9.7 are the same (both using LIDAR 50-ft DEMs) and observations based on Tables 9.8 and 9.9 are the same (both using NED data).

9.3.2.2 Histograms

This subsection presents all the histograms obtained for the samples in the case study. The purpose of doing this was to graphically describe the sample distribution. The histogram for a normal distribution would simulate a bell-shaped frequency distribution. For each sample, differences are grouped into 11 intervals with the same interval width based on the range except the first and the last intervals. The first interval is in the format of $(-\infty, a \text{ value}]$. The last interval is in the format of $(a \text{ value}, +\infty)$. For each interval, the number of differences grouped into it is counted (frequency) and relative frequency is calculated. For consistency purposes, only the histograms based on relative frequencies are included in this subsection. Since the only purpose of distribution histograms in this case study is to graphically describe the sample distributions, the X-axis is not labeled, i.e. the values of the interval ends are not provided.

For convenience purposes, a naming schema is designed as shown in Table 9.10. Each abbreviation contains three or four letter groups separated by underscores. The first letter group indicates the error format (i.e. difference or proportional difference). The second letter group indicates the road type (A for all roads, I for

Interstate highways, US for US routes, NC for NC routes). The third letter group indicates the elevation dataset being used (i.e. LP for LIDAR point data, L20 for LIDAR 20-ft DEM, L50 for LIDAR 50-ft DEM, and N for NED). The fourth letter group indicates the intervals taken, if there are intervals (10 for 10-ft interval, 20 for 20-ft interval, 25 for 25-ft interval, 50 for 50-ft interval, 15 for 15-m interval, and 30 for 30-m interval).

Table 9.10 Naming Schema

Error Format	Road Type	LIDAR 20-ft DEM		LIDAR 50-ft DEM	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval
Difference	ALL	D A L20 10	D A L20 20	D A L50 25	D A L50 50
	Interstate	D I L20 10	D I L20 20	D I L50 25	D I L50 50
	US	D US L20 10	D US L20 20	D US L50 25	D US L50 50
	NC	D NC L20 10	D NC L20 20	D NC L50 25	D NC L50 50
Proportional Difference	ALL	PD A L20 10	PD A L20 20	PD A L50 25	PD A L50 50
	Interstate	PD I L20 10	PD I L20 20	PD I L50 25	PD I L50 50
	US	PD US L20 10	PD US L20 20	PD US L50 25	PD US L50 50
	NC	PD NC L20 10	PD NC L20 20	PD NC L50 25	PD NC L50 50

Table 9.10 (Continued)

Error Format	Road Type	LIDAR Point Data	NED	
			15-m Interval	30-m Interval
Difference	All	D A LP	D A N 15	D A N 30
	Interstate	D I LP	D I N 15	D I N 30
	US	D US LP	D US N 15	D US N 30
	NC	D NC LP	D NC N 15	D NC N 30
Proportional Difference	All	PD A LP	PD A N 15	PD A N 30
	Interstate	PD I LP	PD I N 15	PD I N 30
	US	PD US LP	PD US N 15	PD US N 30
	NC	PD NC LP	PD NC N 15	PD NC N 30

(1) 3-D Distances predicted Using LIDAR Point Data versus DMI Distances

Figure 9.3 illustrates the distribution histograms for differences between 3-D distances predicted using LIDAR point data and the DMI measured distances.

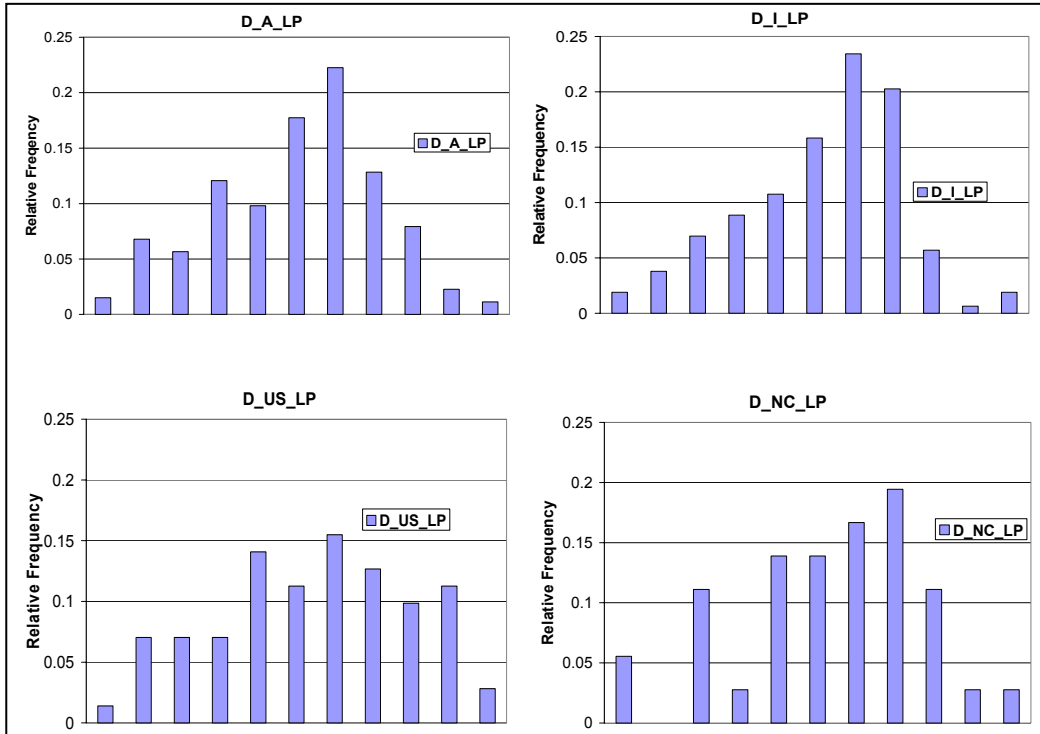


Figure 9.3 Distribution Histograms (LIDAR Point Data, Differences)

Figure 9.4 shows the distribution histograms for the proportional differences between predicted 3-D distances using LIDAR point data and the DMI measured distances.

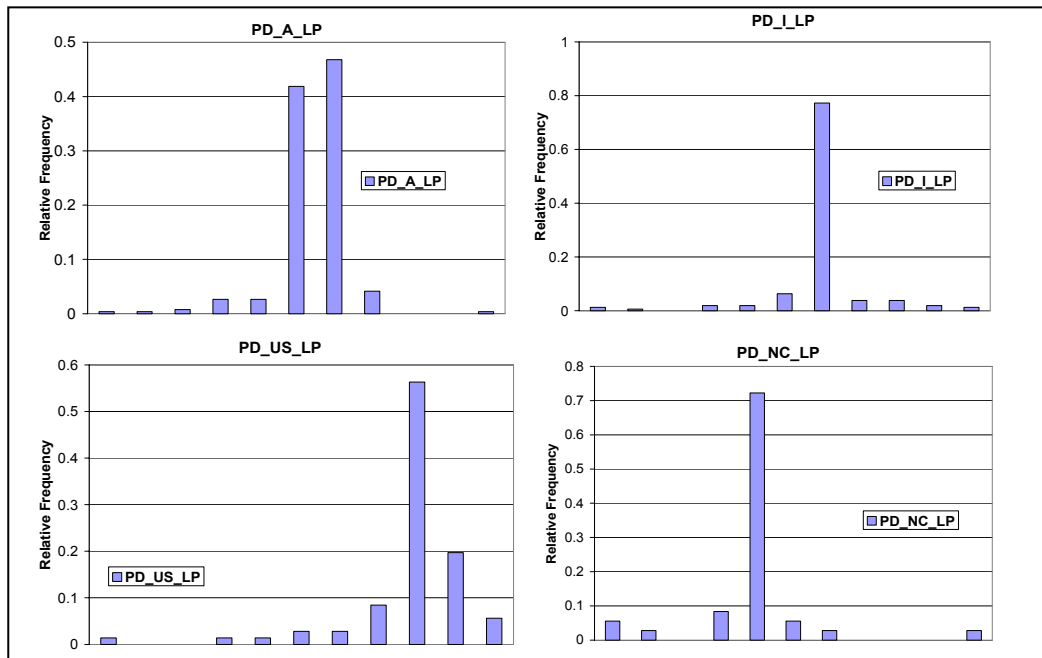


Figure 9.4 Distribution Histograms (LIDAR Point Data, Proportional Differences)

(2) 3-D Distance Predicted Using LIDAR 20-ft DEM with an Interval of 10-ft versus DMI Distances

Figures 9.5 and 9.6 illustrate distribution histograms for differences and proportional differences between predicted 3-D distances using LIDAR 20-ft DEM with a 10-ft interval and DMI measured distances, respectively.

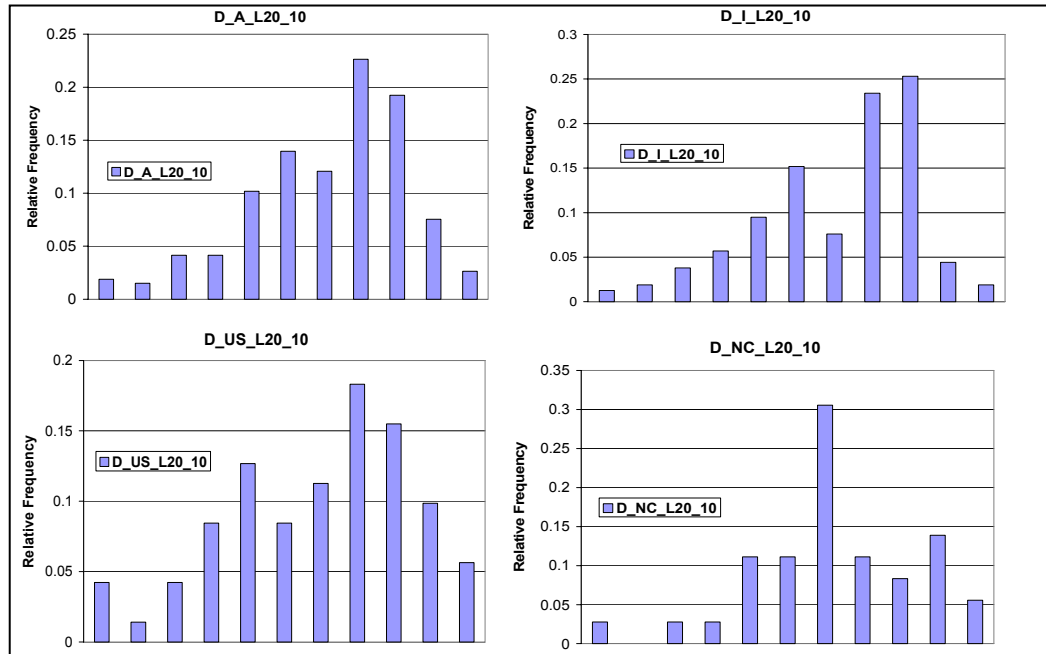


Figure 9.5 Distribution Histograms (LIDAR 20-ft DEM, 10-ft Interval, Differences)

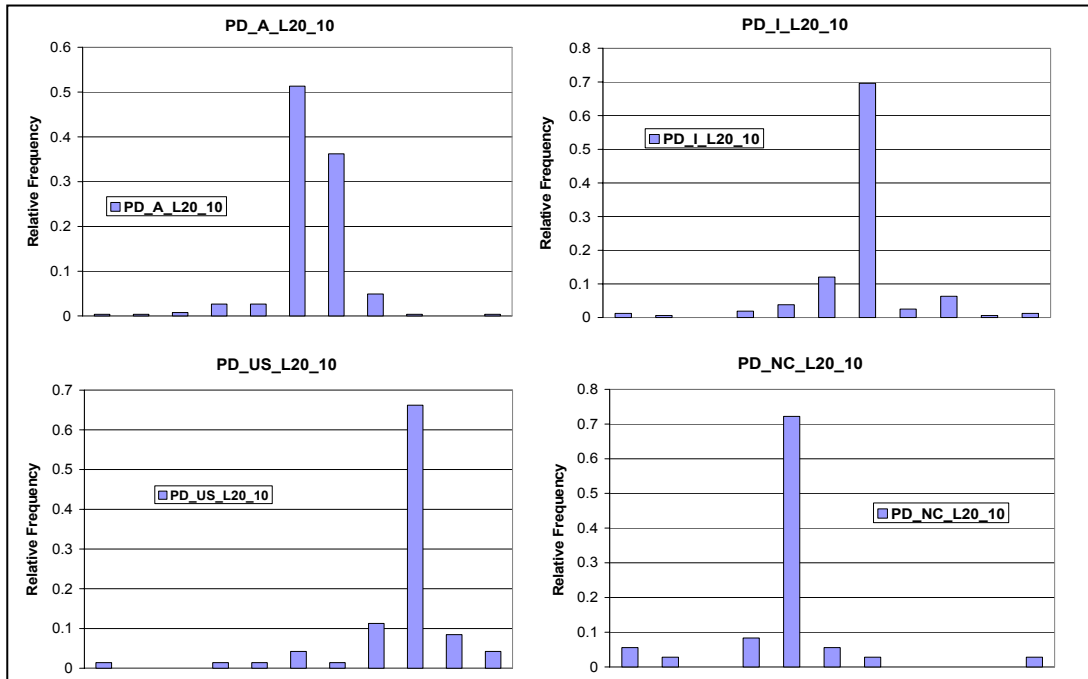


Figure 9.6 Distribution Histograms (LIDAR 20-ft DEM, 10-ft Interval, Proportional Differences)

(3) 3-D Distance Predicted Using LIDAR 20-ft DEM with an Interval of 20-ft versus DMI Distances

Figures 9.7 and 9.8 illustrate the distribution histograms for differences and proportional differences between predicted 3-D distances using LIDAR 20-ft DEM with an interval of 20-ft and the DMI measured distances, respectively.

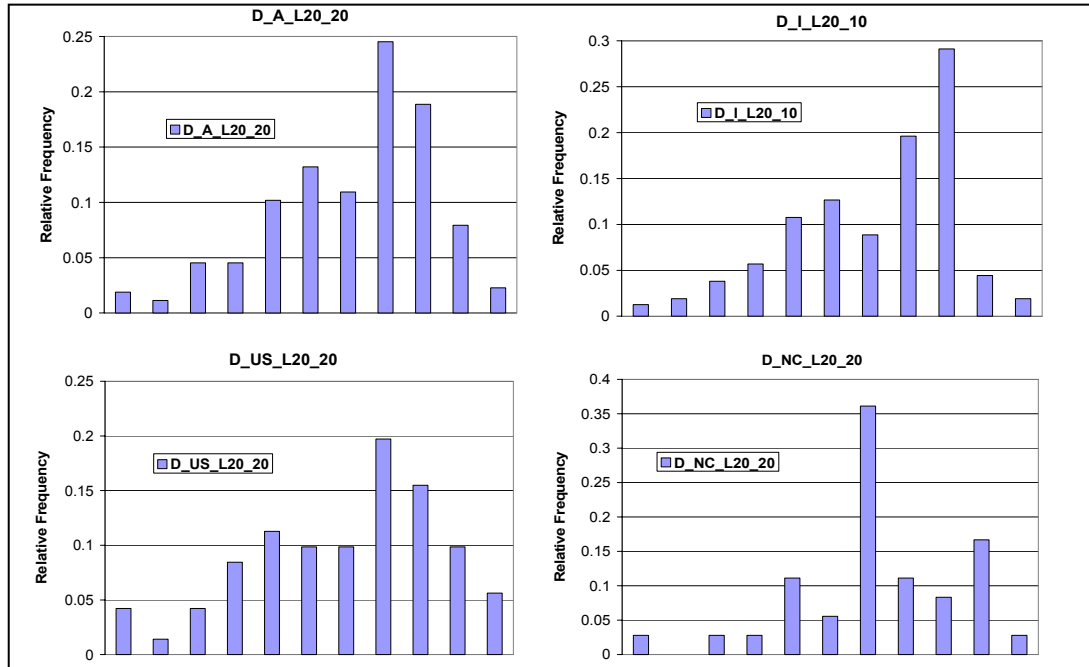


Figure 9.7 Distribution Histograms (LIDAR 20-ft DEM, 20-ft Interval, Differences)

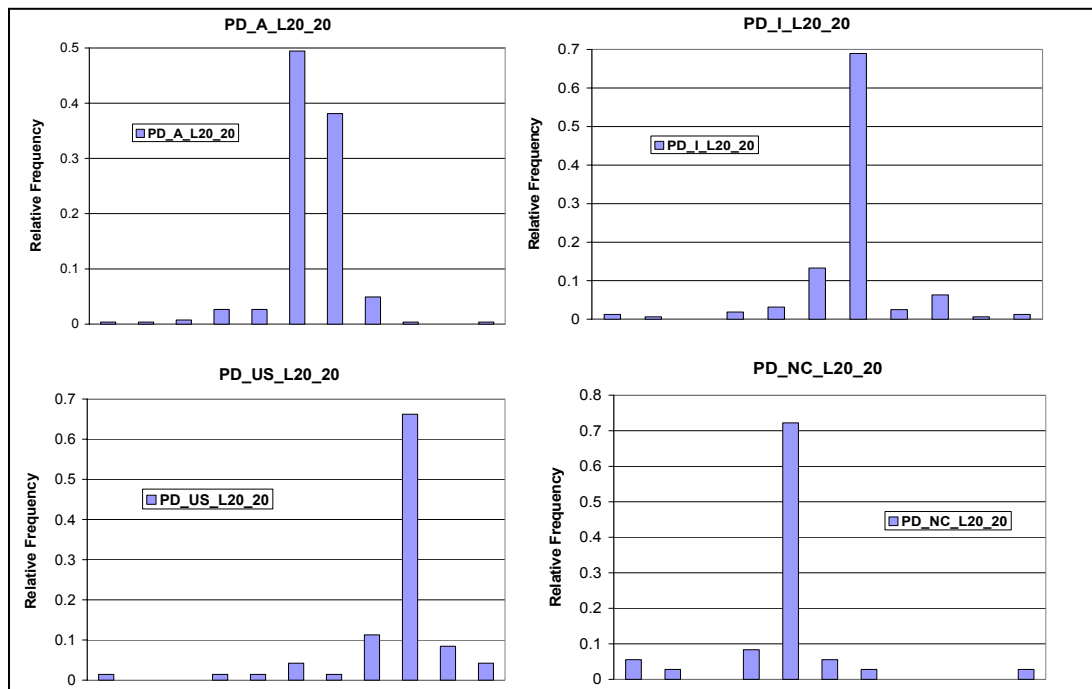


Figure 9.8 Distribution Histograms (LIDAR 20-ft DEM, 20-ft Interval, Proportional Differences)

(4) 3-D distance predicted using LIDAR 50-ft DEM with an Interval of 25-ft versus DMI Distances

Figures 9.9 and 9.10 illustrate the distribution histograms for differences and proportional differences between predicted 3-D distances using LIDAR 50-ft DEM and a 25-ft interval and the DMI measured distances, respectively.

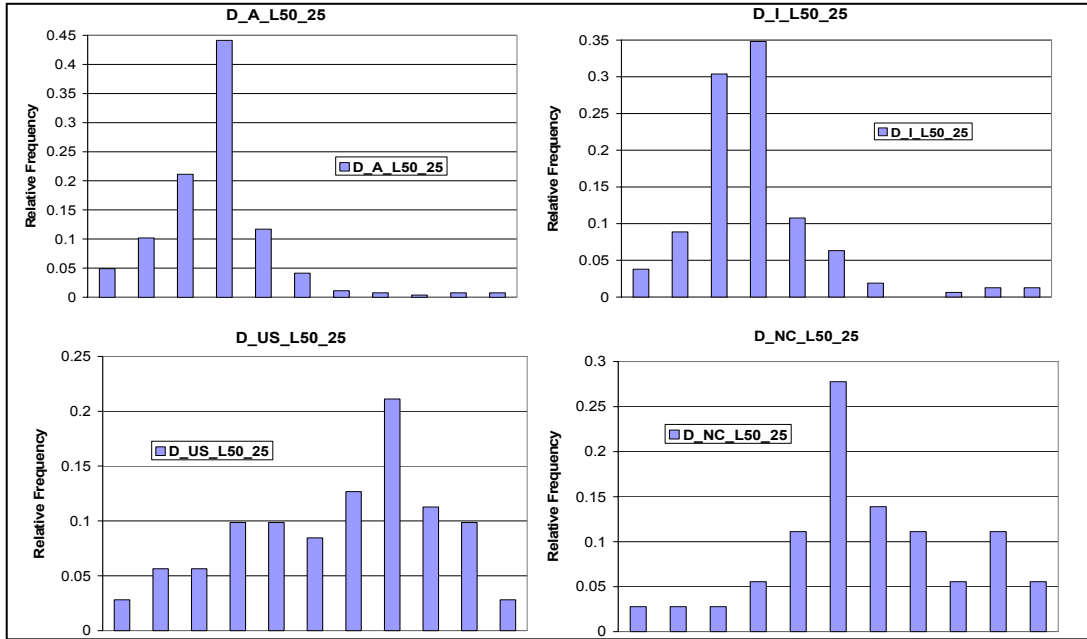


Figure 9.9 Distribution Histograms (LIDAR 50-ft DEM, 25-ft Interval, Differences)

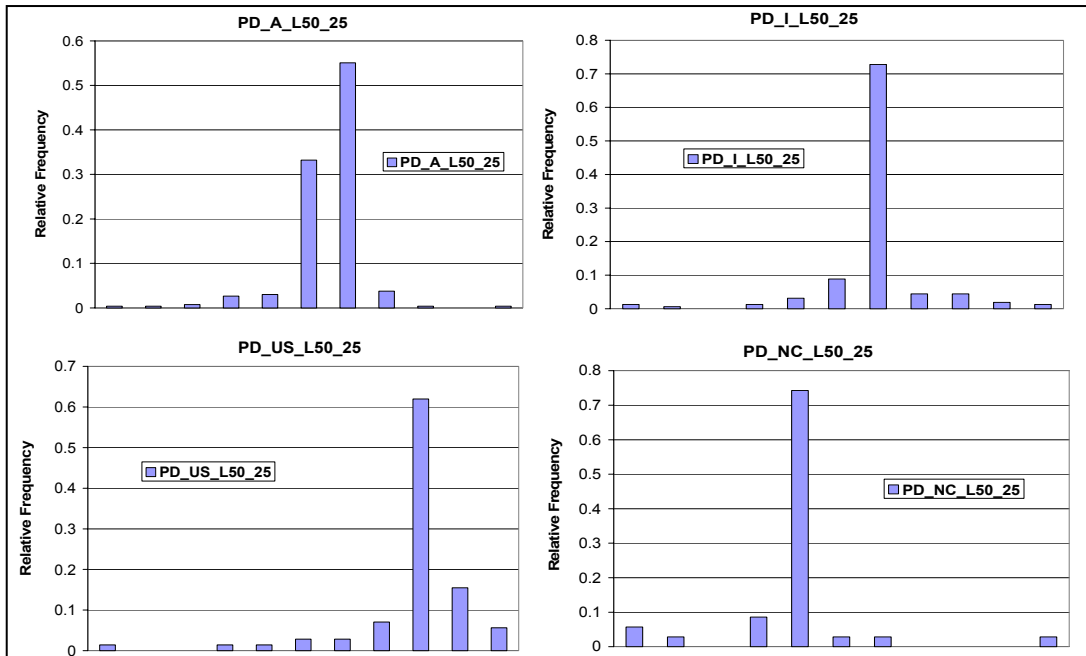


Figure 9.10 Distribution Histograms (LIDAR 50-ft DEM, 25-ft Interval, Proportional Differences)

(5) 3-D Distance Predicted Using LIDAR 50-ft DEM with an Interval of 50-ft versus DMI Distances

Figures 9.11 and 9.12 illustrate the distribution histograms for differences and proportional differences between predicted 3-D distances using LIDAR 50-ft DEM and a 50-ft interval and the DMI measured distances, respectively.

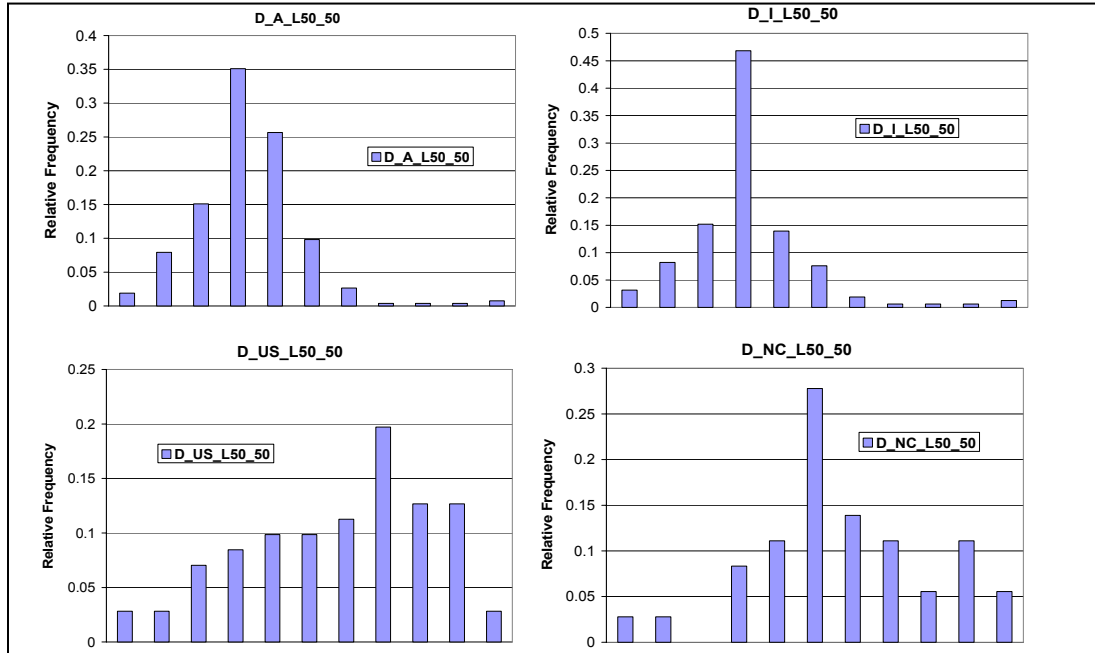


Figure 9.11 Distribution Histograms (LIDAR 50-ft DEM, 50-ft Interval, Differences)

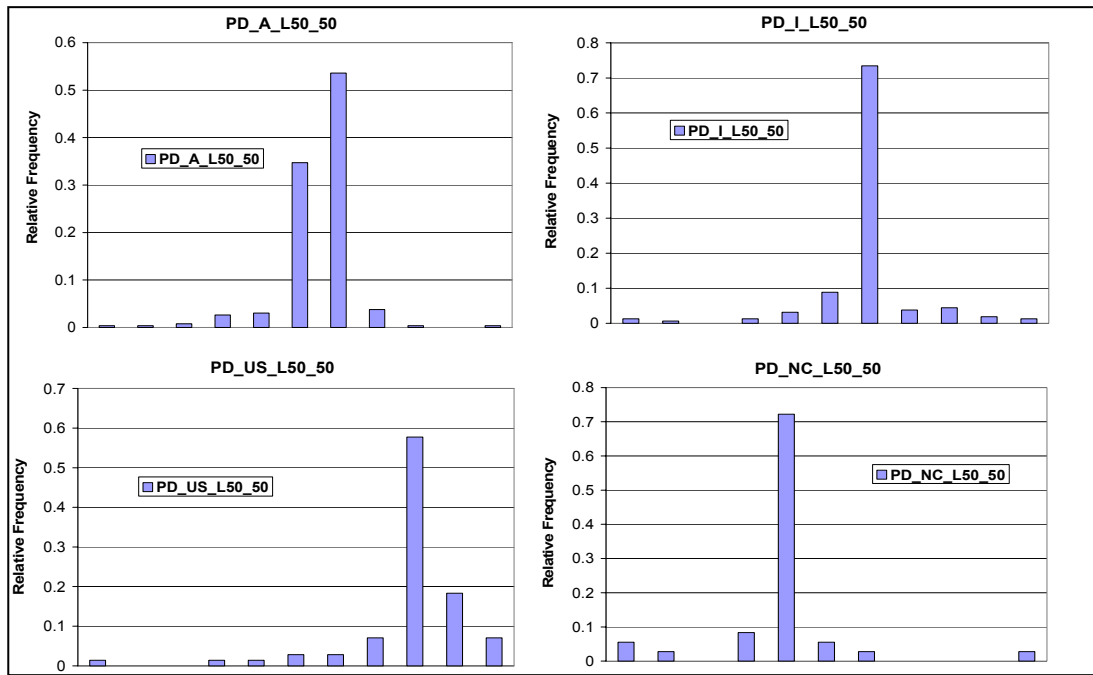


Figure 9.12 Distribution Histograms (LIDAR 50-ft DEM, 50-ft Interval, Proportional Differences)

(6) 3-D Distance Predicted Using 30-m NED with an Interval of 15-m versus DMI Distances

Figures 9.13 and 9.14 illustrate the distribution histograms for differences and proportional differences between predicted 3-D distances using NED and a 15-m interval and the DMI measured distances, respectively.

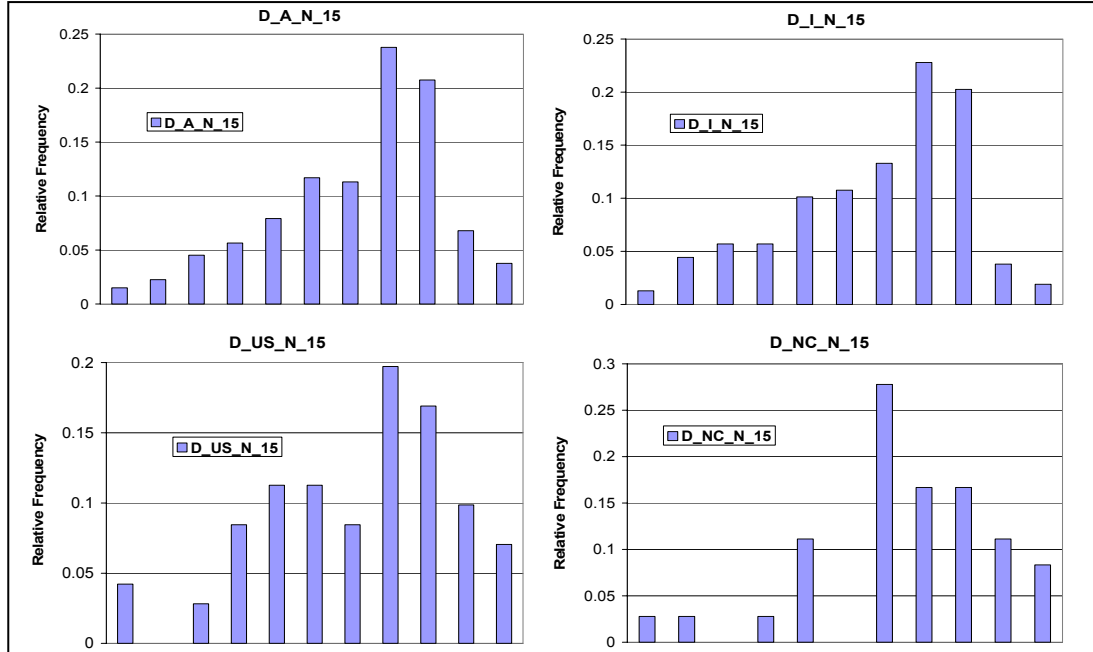


Figure 9.13 Distribution Histograms (NED, 15-m Interval, Differences)

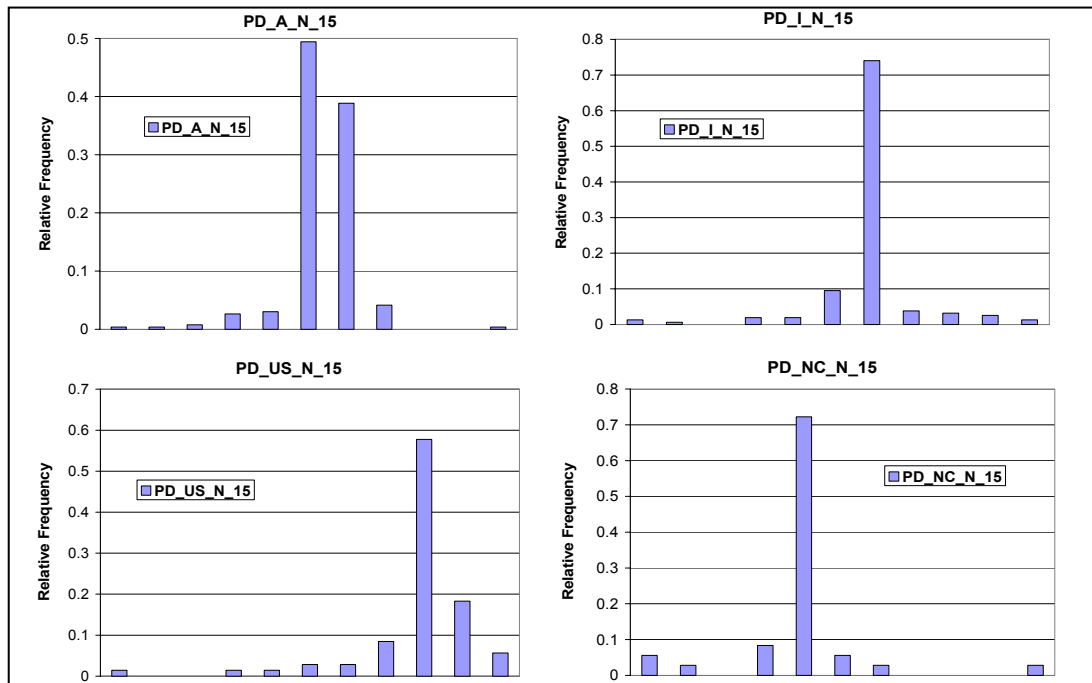


Figure 9.14 Distribution Histograms (NED, 15-m Interval, Proportional Differences)

(7) 3-D Distance Predicted Using 30-m NED with an Interval of 30-m versus DMI Distances

Figures 9.15 and 9.16 illustrate the distribution histograms for differences and proportional differences between predicted 3-D distances using NED and a 30-m interval and the DMI measured distances, respectively.

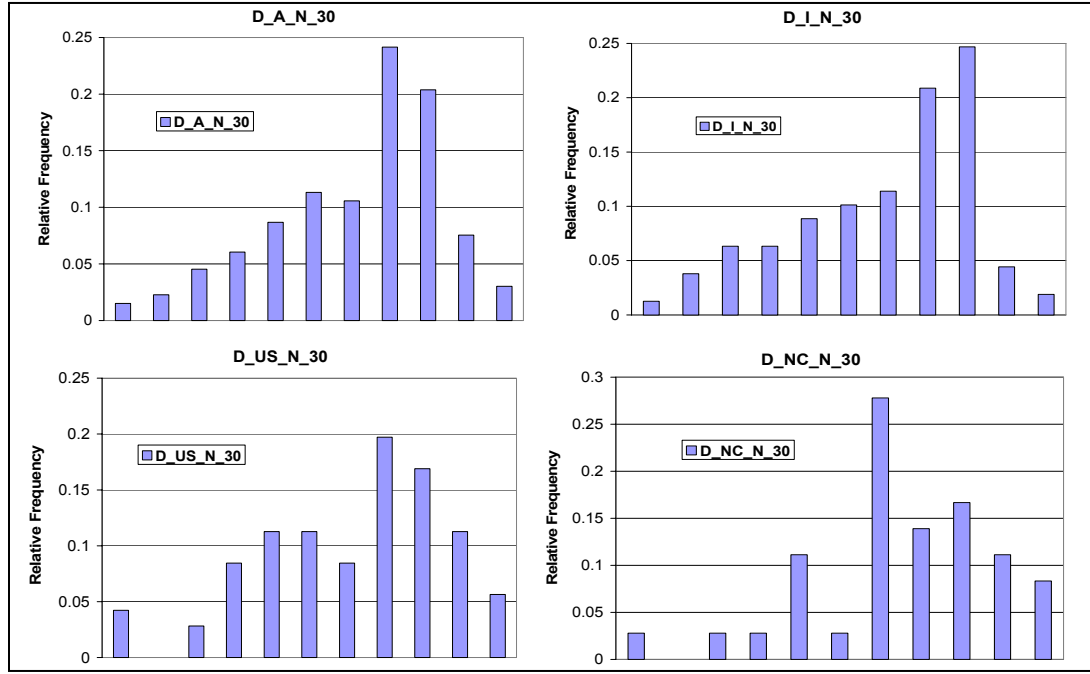


Figure 9.15 Distribution Histograms (NED, 30-m Interval, Differences)

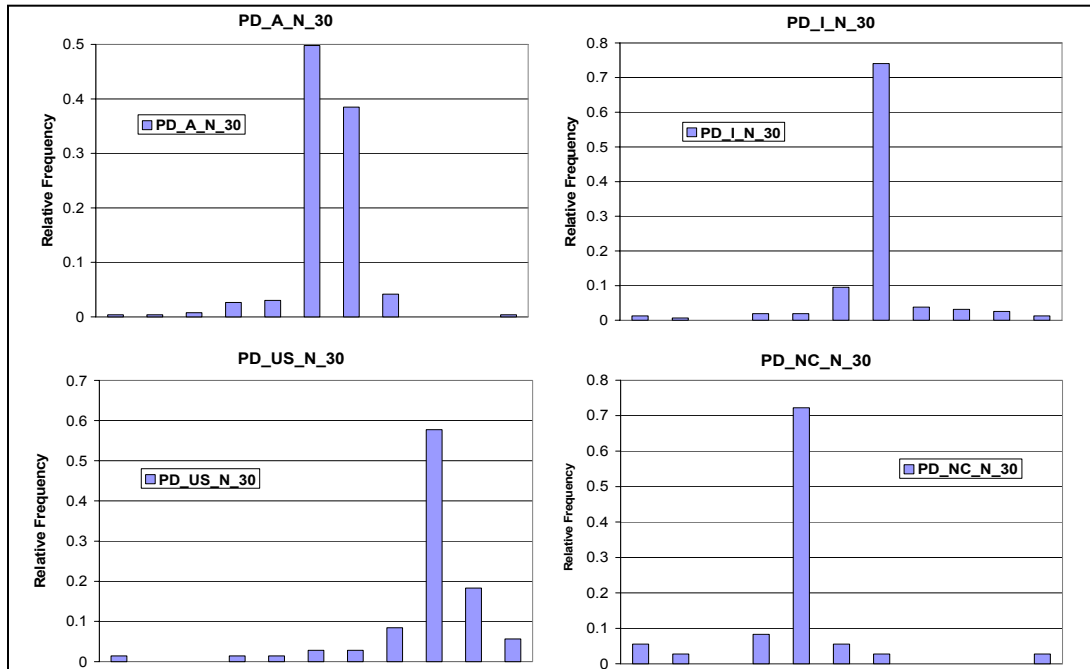


Figure 9.16 Distribution Histograms (NED, 30-m Interval, Proportional Differences)

Based on these histograms, it is observed that given a specific road type (e.g. Interstate FTSEgs, US FTSEgs, NC FTSEgs, or All FTSEgs), the distributions of differences between predicted 3-D distances and the DMI measured distances are similar to each other regardless of the elevation datasets and the model parameters (different intervals, if apply). This is also true in the case of the proportional differences.

It is also observed that none of the samples could be assumed with a normal distribution. The distribution histogram would illustrate a “bell-shape” for a normal distribution. None of the histograms shows such a shape. The asymmetry and skewness are obvious with data tailed either to the left side or the right side.

The initial impression is that the proportional differences are more clustered together than the differences. In other words, the differences are more dispersed than the proportional differences, which is in contradiction to the observations obtained based on descriptive statistics. Examining the data in detail revealed that the proportional errors have a much wider data range and by using the same number of intervals, the interval width for the proportional differences is much bigger than that for the differences and as a result, distribution details for those data categorized into the middle ones (highest bars) are lost. For example, in Figure 9.4, the width of the interval corresponding to the highest bar for PD_A_LP is 59 while the width of the interval for D_A_LP is 8. In other words, approximately 7 intervals (out of a total of 11 intervals) from D_A_LP corresponded to 1 interval from PD_A_LP, and consequently, the distribution details for the data being categorized into the corresponding interval of the highest bar are lost, leading to the false impression that differences are more dispersed than proportional differences.

9.3.2.3 Hypothesis Tests and Confidence Intervals

The population parameter being tested herein is population mean, based on the sample data. In this case study, the difference and proportional difference between a predicted 3-D distance and the DMI measured distance can be viewed as the differences between paired observations from a statistics point of view. The test procedure is summarized as below.

Suppose Y and S are the mean and standard deviation of a random sample of n differences between paired observations and the population of differences has the mean of μ and the standard deviation of σ , the statistic for testing hypotheses about the population mean of the difference population is $t = Y/(S/\sqrt{n})$ (Rao 1998). Table 9.11 shows the different null hypotheses (H_0), corresponding research hypotheses (H_1), and criteria for accepting or rejecting H_0 at level α . The confidence interval for the population mean at the level of α is from $(Y - t(n-1, \alpha/2) * (S/\sqrt{n}))$ to $(Y + t(n-1, \alpha/2) * (S/\sqrt{n}))$ (Rao 1998).

Table 9.11 The Criteria for α -level Hypothesis Tests (Rao 1998)

Hypothesis		Statistic	Critical Value(s)	Reject H_0 at level α if
H_0	H_1			
$\mu \leq 0$	$\mu > 0$	t	$t(n-1, \alpha)$	$t > t(n-1, \alpha)$
$\mu \geq 0$	$\mu < 0$	t	$-t(n-1, \alpha)$	$t < -t(n-1, \alpha)$
$\mu = 0$	$\mu \neq 0$	t	$\pm t(n-1, \alpha/2)$	$t > t(n-1, \alpha/2)$ or $t < -t(n-1, \alpha/2)$

The hypothesis tests in this subsection focuses on the relationship between population means and 0 at the confidence level of 95% ($\alpha = 5\%$). More specifically, the focus herein is to judge if the population mean is less than 0, less than or equal to 0, equal to 0, unequal to 0, greater than 0, or greater than or equal to 0. The results of these tests are represented below for all samples of differences and proportional differences between predicted 3-D distances using different elevation datasets and modeling procedures and the DMI measured 3-D distances. Together, confidence intervals for population means at the same confidence level are shown in the format of $Y \pm t(n-1, \alpha/2)$.

(1) 3-D Distances predicted Using LIDAR Point Data versus DMI Distances

Table 9.12 shows the testing results for differences and proportional differences between 3-D distances predicted using LIDAR point data and the DMI measured distances.

Based on Table 9.12, it was observed that at the level of $\alpha = 0.05$ (95% confidence)

- Population mean of D_A_LP is less than 0;
- Population mean of D_I_LP is less than 0;
- Population mean of D_US_LP is less than 0;

- Population mean of D_NC_LP is equal to 0;
- Population mean of PD_A_LP is less than 0;
- Population mean of PD_I_LP is equal to 0;
- Population mean of PD_US_LP is less than 0; and
- Population mean of PD_NC_LP is equal to 0.

Table 9.12 Hypothesis Test Results (LIDAR Point Data)

Sample	Hypothesis		Statistic	Critical Value(s)	Reject or Accept at $\alpha = 5\%$	Confidence Interval
	H ₀	H ₁				
D_A_LP	$\mu \leq 0$	$\mu > 0$	-5.6509	1.6476	A*	-8.43 ± 2.94
	$\mu \geq 0$	$\mu < 0$	-5.6509	-1.6476	R**	
	$\mu = 0$	$\mu \neq 0$	-5.6509	±1.9642	R	
D_I_LP	$\mu \leq 0$	$\mu > 0$	-5.4826	1.6558	A	-9.93 ± 3.58
	$\mu \geq 0$	$\mu < 0$	-5.4826	-1.6558	R	
	$\mu = 0$	$\mu \neq 0$	-5.4826	±1.9771	R	
D_US_LP	$\mu \leq 0$	$\mu > 0$	-3.4651	1.6691	A	-10.81 ± 6.22
	$\mu \geq 0$	$\mu < 0$	-3.4651	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-3.4651	±1.9979	R	
D_NC_LP	$\mu \leq 0$	$\mu > 0$	0.7088	1.6906	A	2.86 ± 8.19
	$\mu \geq 0$	$\mu < 0$	0.7088	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	0.7088	±2.0317	A	
PD_A_LP	$\mu \leq 0$	$\mu > 0$	-2.1033	1.6476	A	-6.48 ± 6.06
	$\mu \geq 0$	$\mu < 0$	-2.1033	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-2.1033	±1.9642	R	
PD_I_LP	$\mu \leq 0$	$\mu > 0$	-0.4562	1.6558	A	-1.17 ± 5.05
	$\mu \geq 0$	$\mu < 0$	-0.4562	-1.6558	A	
	$\mu = 0$	$\mu \neq 0$	-0.4562	±1.9771	A	
PD_US_LP	$\mu \leq 0$	$\mu > 0$	-2.0295	1.6691	A	-15.02 ± 14.76
	$\mu \geq 0$	$\mu < 0$	-2.0295	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-2.0295	±1.9979	R	
PD_NC_LP	$\mu \leq 0$	$\mu > 0$	-0.9836	1.6906	A	-12.92 ± 26.67
	$\mu \geq 0$	$\mu < 0$	-0.9836	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-0.9836	±2.0317	A	

*Accept null hypothesis H₀.

**Reject null hypothesis H₀.

(2) 3-D Distance Predicted Using LIDAR 20-ft DEM with an Interval of 10-ft versus DMI Distances

Table 9.13 shows the testing results for differences and proportional differences between 3-D distances predicted using LIDAR 20-ft DEM with a 10-ft interval and the DMI measured distances.

Based on Table 9.13, it is observed that at the level of $\alpha = 0.05$ (95% confidence)

- Population mean of D_A_L20_10 is less than 0;
- Population mean of D_I_L20_10 is less than 0;

- Population mean of D_US_L20_10 is less than 0;
- Population mean of D_NC_L20_10 is equal to 0;
- Population mean of PD_A_L20_10 is less than 0;
- Population mean of PD_I_L20_10 is equal to 0;
- Population mean of PD_US_L20_10 is less than 0; and
- Population mean of PD_NC_L20_20 is equal to 0.

Table 9.13 Hypothesis Test Results (LIDAR 20-ft DEM, 10-ft Interval)

Sample	Hypothesis		Statistic	Critical Value(s)	Reject or Accept	Confidence Interval
	H ₀	H ₁				
D_A_L20_10	$\mu \leq 0$	$\mu > 0$	-9.1040	1.6476	A	-16.28 ± 3.52
	$\mu \geq 0$	$\mu < 0$	-9.1040	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-9.1040	±1.9642	R	
D_I_L20_10	$\mu \leq 0$	$\mu > 0$	-8.5763	1.6558	A	-18.86 ± 4.34
	$\mu \geq 0$	$\mu < 0$	-8.5763	-1.6558	R	
	$\mu = 0$	$\mu \neq 0$	-8.5763	±1.9771	R	
D_US_L20_10	$\mu \leq 0$	$\mu > 0$	-4.2055	1.6691	A	-16.79 ± 7.96
	$\mu \geq 0$	$\mu < 0$	-4.2055	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-4.2055	±1.9979	R	
D_NC_L20_10	$\mu \leq 0$	$\mu > 0$	-1.0460	1.6906	A	-3.95 ± 7.67
	$\mu \geq 0$	$\mu < 0$	-1.0460	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-1.0460	±2.0317	A	
PD_A_L20_10	$\mu \leq 0$	$\mu > 0$	-2.0537	1.6476	A	-6.43 ± 6.16
	$\mu \geq 0$	$\mu < 0$	-2.0537	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-2.0537	±1.9642	R	
PD_I_L20_10	$\mu \leq 0$	$\mu > 0$	-0.3809	1.6558	A	-1.01 ± 5.23
	$\mu \geq 0$	$\mu < 0$	-0.3809	-1.6558	A	
	$\mu = 0$	$\mu \neq 0$	-0.3809	±1.9771	A	
PD_US_L20_10	$\mu \leq 0$	$\mu > 0$	-1.9963	1.6691	A	-15.08 ± 15.07
	$\mu \geq 0$	$\mu < 0$	-1.9963	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-1.9963	±1.9979	A	
PD_NC_L20_10	$\mu \leq 0$	$\mu > 0$	-1.0032	1.6906	A	-13.16 ± 26.64
	$\mu \geq 0$	$\mu < 0$	-1.0032	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-1.0032	±2.0317	A	

(3) 3-D Distance Predicted Using LIDAR 20-ft DEM with an Interval of 20-ft versus DMI Distances

Table 9.14 shows the testing results for differences and proportional differences between 3-D distances predicted using LIDAR 20-ft DEM with a 20-ft interval and the DMI measured distances. Based on Table 9.14, it is observed that at the level of $\alpha = 0.05$ (95% confidence) the test results for populations means of the differences and the proportional differences between the predicted 3-D distances using LIDAR 20-ft DEMs with an interval of 20-ft interval are the same as those with the use of LIDAR 20-ft DEMs with an interval of 10-ft and therefore, are not repeated herein.

Table 9.14 Hypothesis Test Results (LIDAR 20-ft DEM, 20-ft Interval)

Sample	Hypothesis		Statistic	Critical Value(s)	Reject or Accept	Confidence Interval
	H ₀	H ₁				
D_A_L20_20	$\mu \leq 0$	$\mu > 0$	-9.3452	1.6476	A	-16.82 ± 3.54
	$\mu \geq 0$	$\mu < 0$	-9.3452	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-9.3452	±1.9642	R	
D_I_L20_20	$\mu \leq 0$	$\mu > 0$	-8.7912	1.6558	A	-19.52 ± 4.39
	$\mu \geq 0$	$\mu < 0$	-8.7912	-1.6558	R	
	$\mu = 0$	$\mu \neq 0$	-8.7912	±1.9771	R	
D_US_L20_20	$\mu \leq 0$	$\mu > 0$	-4.2473	1.6691	A	-17.03 ± 8.00
	$\mu \geq 0$	$\mu < 0$	-4.2473	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-4.2473	±1.9979	R	
D_NC_L20_20	$\mu \leq 0$	$\mu > 0$	-1.2050	1.6906	A	-4.51 ± 7.60
	$\mu \geq 0$	$\mu < 0$	-1.2050	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-1.2050	±2.0317	A	
PD_A_L20_20	$\mu \leq 0$	$\mu > 0$	-2.0816	1.6476	A	-6.51 ± 6.16
	$\mu \geq 0$	$\mu < 0$	-2.0816	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-2.0816	±1.9642	R	
PD_I_L20_20	$\mu \leq 0$	$\mu > 0$	-0.4187	1.6558	A	-1.11 ± 5.22
	$\mu \geq 0$	$\mu < 0$	-0.4187	-1.6558	A	
	$\mu = 0$	$\mu \neq 0$	-0.4187	±1.9771	A	
PD_US_L20_20	$\mu \leq 0$	$\mu > 0$	-2.0064	1.6691	A	-15.15 ± 15.06
	$\mu \geq 0$	$\mu < 0$	-2.0064	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-2.0064	±1.9979	R	
PD_NC_L20_20	$\mu \leq 0$	$\mu > 0$	-1.0069	1.6906	A	-13.21 ± 26.64
	$\mu \geq 0$	$\mu < 0$	-1.0069	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-1.0069	±2.0317	A	

(4) 3-D distance predicted using LIDAR 50-ft DEM with an Interval of 25-ft versus DMI Distances

Table 9.15 shows the testing results for differences and proportional differences between 3-D distances predicted using LIDAR 50-ft DEM with a 25-ft interval and the DMI measured distances.

Based on Table 9.15, it is observed that at the level of $\alpha = 0.05$ (95% confidence)

- Population mean of D_A_L50_25 is equal to 0;
- Population mean of D_I_L50_25 is greater than 0;
- Population mean of D_US_L50_25 is less than 0;
- Population mean of D_NC_L50_25 is equal to 0;
- Population mean of PD_A_L50_25 is less than 0;
- Population mean of PD_I_L50_25 is equal to 0;
- Population mean of PD_US_L50_25 is less than 0; and
- Population mean of PD_NC_L50_50 is equal to 0.

Table 9.15 Hypothesis Test Results (LIDAR 50-ft DEM, 25-ft Interval)

Sample	Hypothesis		Statistic	Critical Value(s)	Reject or Accept	Confidence Interval
	H ₀	H ₁				
D_A_L50_25	$\mu \leq 0$	$\mu > 0$	0.5505	1.6476	A	1.34 ± 4.79
	$\mu \geq 0$	$\mu < 0$	0.5505	-1.6476	A	
	$\mu = 0$	$\mu \neq 0$	0.5505	±1.9642	A	
D_I_L50_25	$\mu \leq 0$	$\mu > 0$	2.5946	1.6558	R	9.03 ± 6.87
	$\mu \geq 0$	$\mu < 0$	2.5946	-1.6558	A	
	$\mu = 0$	$\mu \neq 0$	2.5946	±1.9771	R	
D_US_L50_25	$\mu \leq 0$	$\mu > 0$	-4.2062	1.6691	A	-15.19 ± 7.20
	$\mu \geq 0$	$\mu < 0$	-4.2062	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-4.2062	±1.9979	R	
D_NC_L50_25	$\mu \leq 0$	$\mu > 0$	0.0510	1.6906	A	0.21 ± 8.23
	$\mu \geq 0$	$\mu < 0$	0.0510	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	0.0510	±2.0317	A	
PD_A_L50_25	$\mu \leq 0$	$\mu > 0$	-1.7177	1.6476	A	-5.34 ± 6.12
	$\mu \geq 0$	$\mu < 0$	-1.7177	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-1.7177	±1.9642	A	
PD_I_L50_25	$\mu \leq 0$	$\mu > 0$	0.2421	1.6558	A	0.63 ± 5.10
	$\mu \geq 0$	$\mu < 0$	0.2421	-1.6558	A	
	$\mu = 0$	$\mu \neq 0$	0.2421	±1.9771	A	
PD_US_L50_25	$\mu \leq 0$	$\mu > 0$	-1.9671	1.6691	A	-14.76 ± 14.97
	$\mu \geq 0$	$\mu < 0$	-1.9671	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-1.9671	±1.9979	A	
PD_NC_L50_25	$\mu \leq 0$	$\mu > 0$	-0.9850	1.6906	A	-12.93 ± 26.64
	$\mu \geq 0$	$\mu < 0$	-0.9850	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-0.9850	±2.0317	A	

(5) 3-D Distance Predicted Using LIDAR 50-ft DEM with an Interval of 50-ft versus DMI Distances

Table 9.16 shows the testing results for differences and proportional differences between 3-D distances predicted using LIDAR 50-ft DEM with a 50-ft interval and the DMI measured distances.

Based on Table 9.16, it is observed that at the level of $\alpha = 0.05$ (95% confidence)

- Population mean of D_A_L50_50 is less than 0;
- Population mean of D_I_L50_50 is equal to 0;
- Population mean of D_US_L50_50 is less than 0;
- Population mean of D_NC_L50_50 is equal to 0;
- Population mean of PD_A_L50_50 is less than 0;
- Population mean of PD_I_L50_50 is equal to 0;
- Population mean of PD_US_L50_50 is less than 0; and
- Population mean of PD_NC_L50_50 is equal to 0.

Table 9.16 Hypothesis Test Results (LIDAR 50-ft DEM, 50-ft Interval)

Sample	Hypothesis		Statistic	Critical Value(s)	Reject or Accept	Confidence Interval
	H ₀	H ₁				
D_A_L50_50	$\mu \leq 0$	$\mu > 0$	-1.7079	1.6476	A	-3.57 ± 2.09
	$\mu \geq 0$	$\mu < 0$	-1.7079	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-1.7079	±1.9642	A	
D_I_L50_50	$\mu \leq 0$	$\mu > 0$	0.5569	1.6558	A	1.59 ± 5.64
	$\mu \geq 0$	$\mu < 0$	0.5569	-1.6558	A	
	$\mu = 0$	$\mu \neq 0$	0.5569	±1.9771	A	
D_US_L50_50	$\mu \leq 0$	$\mu > 0$	-4.3231	1.6691	A	-16.09 ± 7.42
	$\mu \geq 0$	$\mu < 0$	-4.3231	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-4.3231	±1.9979	R	
D_NC_L50_50	$\mu \leq 0$	$\mu > 0$	-0.3954	1.6906	A	-1.54 ± 3.89
	$\mu \geq 0$	$\mu < 0$	-0.3954	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-0.3954	±2.0317	A	
PD_A_L50_50	$\mu \leq 0$	$\mu > 0$	-1.8788	1.6476	A	-5.83 ± 6.11
	$\mu \geq 0$	$\mu < 0$	-1.8788	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-1.8788	±1.9642	A	
PD_I_L50_50	$\mu \leq 0$	$\mu > 0$	-0.0376	1.6558	A	-0.10 ± 5.08
	$\mu \geq 0$	$\mu < 0$	-0.0376	-1.6558	A	
	$\mu = 0$	$\mu \neq 0$	-0.0376	±1.9771	A	
PD_US_L50_50	$\mu \leq 0$	$\mu > 0$	-1.9900	1.6691	A	-14.91 ± 14.95
	$\mu \geq 0$	$\mu < 0$	-1.9900	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-1.9900	±1.9979	A	
PD_NC_L50_50	$\mu \leq 0$	$\mu > 0$	-0.9957	1.6906	A	-13.08 ± 26.66
	$\mu \geq 0$	$\mu < 0$	-0.9957	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-0.9957	±2.0317	A	

(6) 3-D Distance Predicted Using 30-m NED with an Interval of 15-m versus DMI Distances

Table 9.17 shows the testing results for differences and proportional differences between 3-D distances predicted using NED 30-m DEM with a 15-m interval and the DMI measured distances.

Based on Table 9.17, it is observed that at the level of $\alpha = 0.05$ (95% confidence)

- Population mean of D_A_N_15 is less than 0;
- Population mean of D_I_N_15 is equal to 0;
- Population mean of D_US_N_15 is less than 0;
- Population mean of D_NC_N_15 is equal to 0;
- Population mean of PD_A_N_15 is less than 0;
- Population mean of PD_I_N_15 is equal to 0;
- Population mean of PD_US_N_15 is less than 0; and
- Population mean of PD_NC_N_15 is equal to 0.

Table 9.17 Hypothesis Test Results (NED 30-m DEM, 15-m Interval)

Sample	Hypothesis		Statistic	Critical Value(s)	Reject or Accept	Confidence Interval
	H ₀	H ₁				
D_A_N_15	$\mu \leq 0$	$\mu > 0$	-10.0064	1.6476	A	-18.63 ± 3.67
	$\mu \geq 0$	$\mu < 0$	-10.0064	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-10.0064	±1.9642	R	
D_I_N_15	$\mu \leq 0$	$\mu > 0$	-9.1975	1.6558	A	-21.56 ± 4.63
	$\mu \geq 0$	$\mu < 0$	-9.1975	-1.6558	R	
	$\mu = 0$	$\mu \neq 0$	-9.1975	±1.9771	R	
D_US_N_15	$\mu \leq 0$	$\mu > 0$	-9.1975	1.6691	A	-18.97 ± 8.04
	$\mu \geq 0$	$\mu < 0$	-9.1975	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-9.1975	±1.9979	R	
D_NC_N_15	$\mu \leq 0$	$\mu > 0$	-1.3710	1.6906	A	-5.14 ± 7.61
	$\mu \geq 0$	$\mu < 0$	-1.3710	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-1.3710	±2.0317	A	
PD_A_N_15	$\mu \leq 0$	$\mu > 0$	-2.4231	1.6476	A	-7.48 ± 6.07
	$\mu \geq 0$	$\mu < 0$	-2.4231	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-2.4231	±1.9642	R	
PD_I_N_15	$\mu \leq 0$	$\mu > 0$	-0.8220	1.6558	A	-2.11 ± 5.08
	$\mu \geq 0$	$\mu < 0$	-0.8220	-1.6558	A	
	$\mu = 0$	$\mu \neq 0$	-0.8220	±1.9771	A	
PD_US_N_15	$\mu \leq 0$	$\mu > 0$	-2.2156	1.6691	A	-16.42 ± 14.78
	$\mu \geq 0$	$\mu < 0$	-2.2156	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-2.2156	±1.9979	R	
PD_NC_N_15	$\mu \leq 0$	$\mu > 0$	-1.0176	1.6906	A	-13.36 ± 26.66
	$\mu \geq 0$	$\mu < 0$	-1.0176	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-1.0176	±2.0317	A	

(7) 3-D Distance Predicted Using 30-m NED with an Interval of 30-m versus DMI Distances

Table 9.18 shows the testing results for differences and proportional differences between 3-D distances predicted using NED 30-m DEM with a 30-m interval and the DMI measured distances.

Based on Table 9.18, it is observed that at the level of $\alpha = 0.05$ (95% confidence)

- Population mean of D_A_N_30 is less than 0;
- Population mean of D_I_N_30 is equal to 0;
- Population mean of D_US_N_30 is less than 0;
- Population mean of D_NC_N_30 is equal to 0;
- Population mean of PD_A_N_30 is less than 0;
- Population mean of PD_I_N_30 is equal to 0;
- Population mean of PD_US_N_30 is less than 0; and
- Population mean of PD_NC_N_30 is less than 0.

Table 9.18 Hypothesis Test Results (NED 30-m DEM, 30-m Interval)

Sample	Hypothesis		Statistic	Critical Value(s)	Reject or Accept	Confidence Interval
	H ₀	H ₁				
D_A_N_30	$\mu \leq 0$	$\mu > 0$	-10.1363	1.6476	A	-18.92 ± 3.67
	$\mu \geq 0$	$\mu < 0$	-10.1363	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-10.1363	±1.9642	R	
D_I_N_30	$\mu \leq 0$	$\mu > 0$	-9.3415	1.6558	A	-22.01 ± 4.65
	$\mu \geq 0$	$\mu < 0$	-9.3415	-1.6558	R	
	$\mu = 0$	$\mu \neq 0$	-9.3415	±1.9771	R	
D_US_N_30	$\mu \leq 0$	$\mu > 0$	-4.6983	1.6691	A	-18.87 ± 8.01
	$\mu \geq 0$	$\mu < 0$	-4.6983	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-4.6983	±1.9979	R	
D_NC_N_30	$\mu \leq 0$	$\mu > 0$	-1.4467	1.6906	A	-5.41 ± 7.60
	$\mu \geq 0$	$\mu < 0$	-1.4467	-1.6906	A	
	$\mu = 0$	$\mu \neq 0$	-1.4467	±2.0317	A	
PD_A_N_30	$\mu \leq 0$	$\mu > 0$	-2.4278	1.6476	A	-7.49 ± 6.07
	$\mu \geq 0$	$\mu < 0$	-2.4278	-1.6476	R	
	$\mu = 0$	$\mu \neq 0$	-2.4278	±1.9642	R	
PD_I_N_30	$\mu \leq 0$	$\mu > 0$	-0.8318	1.6558	A	-2.14 ± 5.08
	$\mu \geq 0$	$\mu < 0$	-0.8318	-1.6558	A	
	$\mu = 0$	$\mu \neq 0$	-0.8318	±1.9771	A	
PD_US_N_30	$\mu \leq 0$	$\mu > 0$	-2.2139	1.6691	A	-16.41 ± 14.78
	$\mu \geq 0$	$\mu < 0$	-2.2139	-1.6691	R	
	$\mu = 0$	$\mu \neq 0$	-2.2139	±1.9979	R	
PD_NC_N_30	$\mu \leq 0$	$\mu > 0$	-2.2139	1.6906	A	-13.38 ± 26.66
	$\mu \geq 0$	$\mu < 0$	-2.2139	-1.6906	R	
	$\mu = 0$	$\mu \neq 0$	-2.2139	±2.0317	R	

These observations regarding hypothesis tests performed for the population means reveals that all population means (except D_I_L50_25) are less than or equal to 0, which confirms the underestimate observation that was obtained by examining the descriptive statistics earlier. The confidence interval widths for the population means of the proportional differences are wider than the corresponding interval widths for the population means of the differences, which also confirms the dispersion observation obtained earlier (proportional differences are more dispersed than differences). Both the test results and the confidence intervals showed consistencies. When using the same elevation dataset but two different modeling procedures (two different intervals) for FTSEgs of a particular road type (Interstate highway FTSEgs, US FTSEgs, NC FTSEgs, or All FTSEgs), the resulting two confidence intervals are almost the same as each other.

9.3.2.4 100% and 95% RMSEs

The RMSE is widely applied in evaluating point positional accuracies in GISs, for example, the horizontal and vertical accuracies of DEMs could be evaluated using this approach, together with appropriate sampling strategies (choosing checkpoints) (Daniel and Tennant 2001). A 100% RMSE is calculated using 100% of the

checkpoints while a 95% RMSE is calculated using the “best” 95% of the checkpoints (the “worst” 5% of the checkpoints are discarded). RMSEs at other percentages could also be derived from the checkpoints. The major reason of discarding some worst checkpoints is that statistics such as sample mean, sample standard deviation, and RMSE are all sensitive to outliers, and consequently, the existence of outliers may result in misleading statistics and misleading observations based on these misleading statistics.

It is always a debate whether outliers should be discarded or not from RMSE calculations and accuracy report (Daniel and Tennant 2001). To address this issue, a method called DEM “5-Step” Accuracy Assessment and Report, has been developed as a simple, practical, and standardized approach to DEM accuracy assessment that negates the debate as to whether or not outliers should be discarded (Daniel and Tennant 2001). This method works in such a way that: (1) the 100% RMSE and corresponding standard deviation are calculated; (2) a “3-sigma” value is established by multiplying the standard deviation by 3; (3) errors are checked with this “3-sigma” value. If there are no errors bigger than this “3-sigma” value, the accuracy is computed as $1.96 * RMSE$ and reported. If there are errors bigger than this “3-sigma” value, it is believed that the dataset includes statistical outliers that depart from a normal distribution. As a result, 95% RMSE would be calculated and the check procedure is applied again. If at the 95% level, the check fails again, a 90% RMSE would be calculated and the check procedure is applied again. If it still fails at the 90% level, further options such as improvement measures should be considered (Daniel and Tennant 2001).

The utilization of the “3-sigma” value in checking data distributions is based on the empirical rule in statistics. The empirical rule states that for many (but not all) populations encountered in practice, approximately 68% and 95% of the measurements are, respectively, within a distance of one and two standard deviations from the population mean; and nearly all of the measurements are within a distance of three standard deviations of the population mean (Rao 1998). This rule works particularly well for populations that are mound-shaped (a normal population is mound-shaped).

An RMSE approach modified based on the DEM “5-Step” Accuracy Assessment and Report method has been developed in this case study to evaluate the accuracy of predicted 3-D distances when compared to DMI

measured distances, based on the histograms presented earlier and the empirical rule in statistics. It is observed that the samples in this case study do not follow a normal distribution. However, they are all mound-shaped. In other words, it is valid to rely on the empirical rule in accuracy assessment. The modified procedure calculated both 100% RMSE and 95% RMSE and corresponding reported accuracies ($1.96 * RMSE$) for all samples. The corresponding numbers of outliers (by applying the empirical rule) are provided to obtain a comprehensive evaluation of the accuracy of the predicted 3-D distances. In addition, the mean and absolute mean (mean of the absolute values) are included. The absolute mean focused on the magnitude of differences and proportional differences. The unit for RMSEs from the aspect of the difference is foot. The unit for RMSEs from the aspect of the proportional difference is foot/1,000 feet.

(1) 3-D Distances predicted Using LIDAR Point Data versus DMI Distances

Tables 9.19 and 9.20 provide RMSEs, reported accuracies, number and percentage of outliers, mean, and absolute means for differences and proportional differences between predicted 3-D distances using LIDAR point data and the DMI measurements, using 100% and the “best” 95% of the data, respectively.

Table 9.19 100% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR Point Data) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	25.65	50.27	0	0%	-8.43	19.92
	Inter	24.76	48.53	0	0%	-9.93	18.64
	US	28.26	55.39	0	0%	-10.81	22.71
	NC	24.04	47.12	0	0%	2.86	20.06
Proportional Differences	All	50.44	98.86	7	2.64%	-6.48	21.56
	Inter	32.06	62.84	5	3.16%	-1.17	14.09
	US	63.72	124.89	2	2.82%	-15.02	31.22
	NC	78.80	154.45	1	2.78%	-12.92	35.25

Table 9.20 95% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR Point Data) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	22.48	44.06	0	0%	-5.74	17.85
	Inter	21.39	41.92	0	0%	-7.32	16.49
	US	25.28	49.55	0	0%	-8.14	20.67
	NC	21.32	41.79	0	0%	3.34	18.27
Proportional Differences	All	24.90	48.80	6	2.38%	-0.50	12.91
	Inter	18.19	35.65	6	4.00%	0.62	8.79
	US	35.47	69.52	1	1.48%	-4.03	21.08
	NC	51.93	101.78	1	2.92%	-16.62	23.01

It is observed that for differences, there are no outliers in both tables. However, for proportional differences, there are outliers with small percentages in both tables.

(2) 3-D Distance Predicted Using LIDAR 20-ft DEM with an Interval of 10-ft versus DMI Distances

Tables 9.21 and 9.22 provide RMSEs, reported accuracies, number and percentage of outliers, mean, and absolute means for differences and proportional differences between predicted 3-D distances using LIDAR 20-ft DEM with a 10-ft interval and the DMI measurements, using 100% and the “best” 95% of the data, respectively.

Table 9.21 100% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 10-ft Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	33.30	65.27	1	0.38%	-16.28	24.26
	Inter	33.38	65.42	1	0.63%	-18.86	24.00
	US	37.39	73.28	0	0%	-16.79	28.40
	NC	22.69	44.47	0	0%	-3.95	17.28
Proportional Differences	All	51.28	100.51	7	2.64%	-6.43	22.52
	Inter	33.16	64.99	5	3.16%	-1.01	15.10
	US	64.99	127.38	2	2.82%	-15.08	32.77
	NC	78.74	154.33	1	2.78%	-13.16	34.90

Table 9.22 95% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 10-ft Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	27.65	54.19	0	0%	-12.57	20.98
	Inter	27.95	54.78	0	0%	-15.34	20.76
	US	31.31	61.37	0	0%	-12.60	24.82
	NC	18.56	36.38	0	0%	-0.99	15.02
Proportional Differences	All	26.36	51.67	6	2.38%	-0.50	13.87
	Inter	19.76	38.73	6	4.00%	-1.13	9.80
	US	37.73	73.95	1	1.48%	-4.07	22.69
	NC	51.82	101.57	1	2.92%	-16.86	22.63

It is observed that for differences, there are outliers (1 in All FTSEgs and 1 in Interstate Highway FTSEgs) in the 100% table. These outliers are discarded in the 95% table. For proportional differences, there are outliers with small percentages in both tables.

(3) 3-D Distance Predicted Using LIDAR 20-ft DEM with an Interval of 20-ft versus DMI Distances

Tables 9.23 and 9.24 provide RMSEs, reported accuracies, number and percentage of outliers, mean, and absolute means for differences and proportional differences between predicted 3-D distances using LIDAR 20-ft DEM with a 20-ft interval and the DMI measurements, using 100% and the “best” 95% of the data, respectively.

Table 9.23 100% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 20-ft Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	33.73	66.11	1	0.38%	-16.82	24.55
	Inter	33.99	66.62	1	0.63%	-19.52	24.45
	US	37.62	73.74	0	0%	-17.03	28.50
	NC	22.62	44.34	0	0%	-4.51	17.19
Proportional Differences	All	51.26	100.47	7	2.64%	-6.51	22.52
	Inter	33.12	64.92	5	3.16%	-1.11	15.10
	US	64.98	127.36	2	2.82%	-15.15	32.75
	NC	78.74	154.33	1	2.78%	-13.21	34.90

Table 9.24 95% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 20-ft DEM with 20-ft Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	28.01	54.90	0	0%	-13.08	21.22
	Inter	28.53	55.92	0	0%	-15.99	21.17
	US	31.42	61.58	0	0%	-12.80	24.87
	NC	18.31	35.89	0	0%	-1.54	14.89
Proportional Differences	All	26.33	51.61	5	1.99%	-0.59	13.87
	Inter	19.75	38.71	6	4.00%	-1.22	9.81
	US	37.72	73.93	1	1.48%	-4.15	22.67
	NC	51.82	101.57	1	2.92%	-16.91	22.64

It is observed that for differences, there are outliers (1 in All FTSEgs and 1 in Interstate Highway FTSEgs) in the 100% table. These outliers are discarded in the 95% table. For proportional differences, there are outliers with small percentages in both tables.

(4) 3-D distance predicted using LIDAR 50-ft DEM with an Interval of 25-ft versus DMI Distances

Tables 9.25 and 9.26 provide RMSEs, reported accuracies, number and percentage of outliers, mean, and absolute means for differences and proportional differences between predicted 3-D distances using LIDAR 50-

ft DEM with a 25-ft interval and the DMI measurements, using 100% and the “best” 95% of the data, respectively.

Table 9.25 100% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 25-ft Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	33.90	66.44	5	1.89%	-0.87	23.64
	Inter	35.84	70.25	5	3.16%	5.44	24.72
	US	33.83	66.31	0	0%	-15.19	26.33
	NC	23.99	47.02	0	0%	0.21	18.53
Proportional Differences	All	50.79	99.55	7	2.64%	-5.34	22.13
	Inter	32.33	63.37	5	3.16%	0.62	14.71
	US	64.51	126.44	2	2.82%	-14.76	32.13
	NC	78.71	154.27	1	2.78%	-12.93	35.01

Table 9.26 95% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 25-ft Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	27.71	54.31	0	0%	-2.82	20.62
	Inter	28.75	56.35	0	0%	1.71	20.46
	US	29.83	58.47	0	0%	-11.92	23.66
	NC	20.91	40.98	0	0%	3.17	16.55
Proportional Differences	All	25.58	50.14	6	2.38%	0.68	13.51
	Inter	18.80	36.85	5	3.33%	1.48	9.48
	US	36.87	72.27	1	1.48%	-3.74	22.02
	NC	51.78	101.49	1	2.92%	-16.26	22.26

It is observed that for differences, there are outliers (5 in All FTSEgs and 5 in Interstate Highway FTSEgs) in the 100% table. These outliers are discarded in the 95% table. For proportional differences, there are outliers with small percentages in both tables.

(5) 3-D Distance Predicted Using LIDAR 50-ft DEM with an Interval of 50-ft versus DMI Distances

Tables 9.27 and 9.28 provide RMSEs, reported accuracies, number and percentage of outliers, mean, and absolute means for differences and proportional differences between predicted 3-D distances using LIDAR 50-ft DEM with a 50-ft interval and the DMI measurements, using 100% and the “best” 95% of the data, respectively.

Table 9.27 100% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 50-ft Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	34.16	66.95	4	1.51%	-3.57	23.58
	Inter	35.83	70.23	3	1.90%	1.59	23.37
	US	35.04	68.68	0	0%	-16.09	27.01
	NC	23.05	45.18	0	0%	-1.54	17.72
Proportional Differences	All	50.75	99.47	7	2.64%	-5.83	21.94
	Inter	32.24	63.19	5	3.16%	-0.10	14.41
	US	64.45	126.32	2	2.82%	-14.91	32.09
	NC	78.78	154.41	1	2.78%	-13.08	35.01

Table 9.28 95% RMSE for Differences and Proportional Differences between 3-D Distances (LIDAR 50-ft DEM with 50-ft Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	26.01	50.98	0	0%	-4.43	19.73
	Inter	25.38	49.74	0	0%	-2.13	18.82
	US	30.51	59.80	0	0%	-12.59	24.09
	NC	19.47	38.16	0	0%	1.43	15.61
Proportional Differences	All	25.45	49.88	6	2.38%	0.17	13.30
	Inter	18.64	36.53	5	3.33%	0.73	9.16
	US	36.76	72.05	1	1.48%	-3.90	21.98
	NC	51.89	101.70	1	2.92%	-16.76	22.75

It is observed that for differences, there are outliers (4 in All FTSEgs and 3 in Interstate Highway FTSEgs) in the 100% table. These outliers are discarded in the 95% table. For proportional differences, there are outliers with small percentages in both tables.

(6) 3-D Distance Predicted Using 30-m NED with an Interval of 15-m versus DMI Distances

Tables 9.29 and 9.30 provide RMSEs, reported accuracies, number and percentage of outliers, mean, and absolute means for differences and proportional differences between predicted 3-D distances using LIDAR NED with a 15-m interval and the DMI measurements, using 100% and the “best” 95% of the data, respectively.

Table 9.29 100% RMSE for Differences and Proportional Differences between 3-D Distances (30-m NED with 15-m Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	35.53	69.64	1	0.38%	-18.63	25.82
	Inter	36.43	71.40	0	0%	-21.56	26.37
	US	38.69	75.83	0	0%	-18.97	28.94
	NC	22.76	44.61	1	2.78%	-5.14	17.22
Proportional Differences	All	50.68	99.33	7	2.64%	-7.48	22.11
	Inter	32.29	63.29	5	3.16%	-2.11	14.68
	US	64.14	125.71	2	2.82%	-16.42	32.02
	NC	78.83	154.51	1	2.78%	-13.36	35.13

Table 9.30 95% RMSE for Differences and Proportional Differences between 3-D Distances (30-m NED with 15-m Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	29.95	58.70	0	0%	-14.88	22.44
	Inter	31.38	61.50	0	0%	-18.05	23.12
	US	32.29	63.29	0	0%	-14.70	25.20
	NC	17.98	35.24	0	0%	-2.05	14.77
Proportional Differences	All	25.35	49.69	5	1.99%	-1.55	13.48
	Inter	18.52	36.30	4	2.66%	-0.40	9.39
	US	36.19	70.93	1	1.48%	-5.48	21.91
	NC	52.02	101.96	1	2.92%	-17.06	22.89

It is observed that for differences, there are outliers (1 in All FTSEgs and 1 in NC FTSEgs) in the 100% table. These outliers are discarded in the 95% table. For proportional differences, there are outliers with small percentages in both tables.

(7) 3-D Distance Predicted Using 30-m NED with an Interval of 30-m versus DMI Distances

Tables 9.31 and 9.32 provide RMSEs, reported accuracies, number and percentage of outliers, mean, and absolute means for differences and proportional differences between predicted 3-D distances using LIDAR NED with a 30-m interval and the DMI measurements, using 100% and the “best” 95% of the data, respectively.

Table 9.31 100% RMSE for Differences and Proportional Differences between 3-D Distances (30-m NED with 30-m Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	35.74	70.05	0	0%	-18.92	26.00
	Inter	36.83	72.19	0	0%	-22.01	26.71
	US	38.53	75.52	0	0%	-18.87	28.87
	NC	22.79	44.67	1	2.78%	-5.41	17.22
Proportional Differences	All	50.68	99.33	7	2.64%	-7.49	22.11
	Inter	32.29	63.29	5	3.16%	-2.14	14.69
	US	64.14	125.71	2	2.82%	-16.41	32.02
	NC	78.83	154.51	1	2.78%	-13.38	35.13

Table 9.32 95% RMSE for Differences and Proportional Differences between 3-D Distances (30-m NED with 30-m Interval) and DMI Measurements

Error Format	Road Type	RMSE	Reported Accuracy	# of Outliers	Percentage of Outliers	Mean	Absolute Mean
Differences	All	30.16	59.11	0	0%	-15.16	22.61
	Inter	31.77	62.27	0	0%	-18.49	23.44
	US	32.21	63.13	0	0%	-14.63	25.16
	NC	17.92	35.12	0	0%	-2.31	14.74
Proportional Differences	All	25.35	49.69	5	1.99%	-1.56	13.49
	Inter	18.52	36.30	4	2.66%	-0.43	9.40
	US	36.19	70.93	1	1.48%	-5.47	21.91
	NC	52.02	101.96	1	2.92%	-17.08	22.89

It is observed that for differences, there is 1 outlier (in NC FTSegs) in the 100% table. It is discarded in the 95% table. For proportional differences, there are outliers with small percentages in both tables.

It is quite obvious that the distributions of outliers are quite close to each other regardless of the elevation datasets and the model parameters (different intervals, if apply) being used. There are fewer outliers for differences in the 100% tables. These outliers are discarded when using the “best” 95% data as indicated by no outliers for differences in the 95% tables. It is also observed that by using the “best” 95% data, there is still almost the same number of outliers for proportional differences as that of using 100% of the data. This confirms the observation obtained earlier, which stated that proportional differences were more dispersed than differences.

9.3.2.5 Frequency Analysis

In this case study, frequency analyses are conducted to quantitatively describe the distribution of differences and proportional differences. It differs from RMSEs and descriptive statistics such as means in that it focused on the distribution of the magnitude of differences and proportional differences rather than getting a singular value for a complete sample. In other words, given a sample, there would be one mean and one RMSE. However, the frequency analysis would provide information regarding to the percentages of the measurements in that sample falling into different groups. In a sense, it is similar to the distribution histogram, which was used earlier. However, it differed from distribution histograms in that a distribution histogram would assume a uniform interval width but with a frequency analysis, non-uniform intervals could be used. In addition, with a distribution histogram, a measurement from the sample fell into exactly one interval and was counted once. However, with a frequency analysis, accumulative values could be used and a measurement could be counted more than once.

In this case study, accumulative approach is taken for frequency analyses. The groups for differences are defined as $[-5, 5]$, $[-10, 10]$, $[-20, 20]$, $[-30, 30]$, $[-50, 50]$, and $(-\infty, -50)$ and $(50, +\infty)$, with unit of foot. The groups for proportional differences are defined as $[-1, 1]$, $[-5, 5]$, $[-10, 10]$, $[-20, 20]$, $[-30, 30]$, $[-50, 50]$, $[-100, 100]$, and $(-\infty, -100)$ and $(100, +\infty)$, with unit of foot/1,000 feet. The group of $[-5, 5]$ indicates a group that has all measurements from a sample, which are greater than or equal to -5 and less than or equal to 5. The group of $(-\infty, -50)$ and $(50, +\infty)$ indicates a group that has all measurements from a sample, which are either less than -50 or greater than 50.

(1) 3-D Distances predicted Using LIDAR Point Data versus DMI Distances

Table 9.33 shows the results of frequency analysis for the differences and proportional differences between predicted 3-D distances using LIDAR point data and the DMI measured distances.

Based on Table 9.33, it is observed that

- The majority (over 50%) of the differences are in the $[-20, 20]$ error range.
- Over 90% of the differences are in the error range of $[-50, 50]$.

- The majority of proportional differences are in the [-5, 5] error range when all FTSEgs are considered together.
- When the road type of FTSEgs are considered, the majority of Interstate FTSEgs and NC FTSEgs are still in the [-5, 5] error range. However, for US FTSEgs, the corresponding error range becomes [-10, 10].
- For the proportional differences, about 86% are in the [-50, 50] error range and over 90% are in the [-100, 100] error range.

Table 9.33 Results of Frequency Analysis (LIDAR Point Data)

Error Format	Groups	All FTSEgs		Interstate FTSEgs		US FTSEgs		NC FTSEgs	
		#	%	#	%	#	%	#	%
Differences	[-5, 5]	52	19.62%	33	20.89%	13	18.31%	6	16.67%
	[-10, 10]	97	36.60%	68	43.04%	20	28.17%	9	25.00%
	[-20, 20]	151	56.98%	98	62.03%	35	49.30%	18	50.00%
	[-30, 30]	205	77.36%	122	77.22%	52	73.24%	31	86.11%
	[-50, 50]	249	93.96%	150	94.94%	65	91.55%	34	94.44%
	$(-\infty, -50)$ and $(50, +\infty)$	16	6.04%	8	5.06%	6	8.45%	2	5.56%
Proportional Differences	[-1, 1]	64	24.15%	49	31.01%	4	5.63%	11	30.56%
	[-5, 5]	153	57.74%	107	67.72%	25	35.21%	21	58.33%
	[-10, 10]	177	66.79%	118	74.68%	35	49.30%	24	66.67%
	[-20, 20]	201	75.85%	129	81.65%	46	64.79%	26	72.22%
	[-30, 30]	211	79.62%	133	84.18%	51	71.83%	27	75.00%
	[-50, 50]	228	86.04%	140	88.61%	60	84.51%	28	77.78%
	[-100, 100]	250	94.34%	153	96.84%	65	91.55%	32	88.89%
$(-\infty, -100)$ and $(100, +\infty)$	15	5.66%	15	9.49%	6	8.45%	4	11.11%	

indicates the number of FTSEgs. % indicates the percentage.

(2) 3-D Distance Predicted Using LIDAR 20-ft DEM with an Interval of 10-ft versus DMI Distances

Table 9.34 shows the results of frequency analysis for the differences and proportional differences between predicted 3-D distances using LIDAR 20-ft DEM with a 10-ft interval and the DMI measured distances.

Based on Table 9.34, it is observed that

- The majority (over 50%) of the differences are in the [-20, 20] error range.
- When road types are considered, the majority of the differences for Interstate FTSEgs and NC FTSEgs are still in the [-20, 20] error range. However, for US FTSEgs, the corresponding percentage (46.48%) is less than 50%, but quite close.
- Over 85% of the differences are in the error range of [-50, 50].
- The majority of proportional differences are in the [-5, 5] error range when all FTSEgs are considered together.

- When the road type of FTSEgs are considered, the majority of Interstate FTSEgs and NC FTSEgs are still in the [-5, 5] error range. However, for US FTSEgs, the corresponding error range becomes [-20, 20]. Furthermore, the corresponding percentage of proportional differences from US FTSEgs to the [-10, 10] error range is close to 50% (47.89%).
- For the proportional differences, about 84% are in the [-50, 50] error range and over 90% are in the [-100, 100] error range.

Table 9.34 Results of Frequency Analysis (LIDAR 20-ft DEM, 10-ft Interval)

Error Format	Groups	All FTSEgs		Interstate FTSEgs		US FTSEgs		NC FTSEgs	
		#	%	#	%	#	%	#	%
Differences	[-5, 5]	61	23.02%	39	24.68%	12	16.90%	10	27.78%
	[-10, 10]	97	36.60%	63	39.87%	19	26.76%	15	41.67%
	[-20, 20]	146	55.09%	91	57.59%	33	46.48%	22	61.11%
	[-30, 30]	184	69.43%	109	68.99%	44	61.97%	31	86.11%
	[-50, 50]	227	85.66%	134	84.81%	58	81.69%	35	97.22%
	$(-\infty, -50)$ and $(50, +\infty)$	38	14.34%	24	15.19%	13	18.31%	1	2.78%
Proportional Differences	[-1, 1]	53	20.00%	38	24.05%	3	4.23%	12	33.33%
	[-5, 5]	153	57.74%	106	67.09%	26	36.62%	21	58.33%
	[-10, 10]	174	65.66%	115	72.78%	34	47.89%	25	69.44%
	[-20, 20]	199	75.09%	126	79.75%	47	66.20%	26	72.22%
	[-30, 30]	209	78.87%	131	82.91%	51	71.83%	27	75.00%
	[-50, 50]	225	84.91%	138	87.34%	59	83.10%	28	77.78%
	[-100, 100]	249	93.96%	153	96.84%	64	90.14%	32	88.89%
$(-\infty, -100)$ and $(100, +\infty)$	16	6.04%	5	3.16%	7	9.86%	4	11.11%	

(3) 3-D Distance Predicted Using LIDAR 20-ft DEM with an Interval of 20-ft versus DMI Distances

Table 9.35 shows the results of frequency analysis for the differences and proportional differences between predicted 3-D distances using LIDAR 20-ft DEM with a 20-ft interval and the DMI measured distances.

Based on Table 9.35, it is observed that

- The majority (over 50%) of the differences are in the [-20, 20] error range.
- When road types were considered, the majority of the differences for Interstate FTSEgs and NC FTSEgs are still in the [-20, 20] error range. However, for US FTSEgs, the corresponding percentage (46.48%) is less than 50%, but quite close.
- About 85% of the differences are in the error range of [-50, 50].
- The majority of proportional differences are in the [-5, 5] error range when all FTSEgs are considered together.

- When the road type of FTSegs are considered, the majority of Interstate FTSegs and NC FTSegs are still in the [-5, 5] error range. However, for US FTSegs, the corresponding error range becomes [-20, 20]. Furthermore, the corresponding percentage of proportional differences from US FTSegs to the [-10, 10] error range is close to 50% (47.89%).
- For the proportional differences, about 84% are in the [-50, 50] error range and over 90% are in the [-100, 100] error range.

Table 9.35 Results of Frequency Analysis (LIDAR 20-ft DEM, 20-ft Interval)

Error Format	Groups	All FTSegs		Interstate FTSegs		US FTSegs		NC FTSegs	
		#	%	#	%	#	%	#	%
Differences	[-5, 5]	60	22.64%	38	24.05%	12	16.90%	10	27.78%
	[-10, 10]	97	36.60%	64	40.51%	19	26.76%	14	38.89%
	[-20, 20]	147	55.47%	92	58.23%	33	46.48%	22	61.11%
	[-30, 30]	183	69.06%	107	67.72%	44	61.97%	32	88.89%
	[-50, 50]	225	84.91%	132	83.54%	58	81.69%	35	97.22%
	$(-\infty, -50)$ and $(50, +\infty)$	40	15.09%	26	16.46%	13	18.31%	1	2.78%
Proportional Differences	[-1, 1]	54	20.38%	38	24.05%	3	4.23%	13	36.11%
	[-5, 5]	153	57.74%	106	67.09%	26	36.62%	21	58.33%
	[-10, 10]	174	65.66%	115	72.78%	34	47.89%	25	69.44%
	[-20, 20]	199	75.09%	126	79.75%	47	66.20%	26	72.22%
	[-30, 30]	209	78.87%	131	82.91%	51	71.83%	27	75.00%
	[-50, 50]	226	85.28%	139	87.97%	59	83.10%	28	77.78%
	[-100, 100]	249	93.96%	153	96.84%	64	90.14%	32	88.89%
	$(-\infty, -100)$ and $(100, +\infty)$	16	6.04%	5	3.16%	7	9.86%	4	11.11%

(4) 3-D distance predicted using LIDAR 50-ft DEM with an Interval of 25-ft versus DMI Distances

Table 9.36 shows the results of frequency analysis for the differences and proportional differences between predicted 3-D distances using LIDAR 50-ft DEM with a 25-ft interval and the DMI measured distances.

Based on Table 9.36, it is observed that

- The majority (over 50%) of the differences are in the [-20, 20] error range.
- When road types were considered, the majority of the differences for Interstate FTSegs and NC FTSegs are still in the [-20, 20] error range. However, for US FTSegs, the corresponding percentage (46.48%) is less than 50%, but quite close.
- About 85% of the differences are in the error range of [-50, 50].
- The majority of proportional differences are in the [-5, 5] error range when all FTSegs are considered together.

- When the road type of FTSEgs are considered, the majority of Interstate FTSEgs and NC FTSEgs are still in the [-5, 5] error range. However, for US FTSEgs, the corresponding error range becomes [-20, 20]. Furthermore, the corresponding percentage of proportional differences from US FTSEgs to the [-10, 10] error range is close to 50% (47.89%).
- For the proportional differences, about 86% are in the [-50, 50] error range and over 90% are in the [-100, 100] error range.

Table 9.36 Results of Frequency Analysis (LIDAR 50-ft DEM, 25-ft Interval)

Error Format	Groups	All FTSEgs		Interstate FTSEgs		US FTSEgs		NC FTSEgs	
		#	%	#	%	#	%	#	%
Differences	[-5, 5]	60	22.64%	38	24.05%	12	16.90%	10	27.78%
	[-10, 10]	105	39.62%	73	46.20%	19	26.76%	13	36.11%
	[-20, 20]	155	58.49%	101	63.92%	33	46.48%	21	58.33%
	[-30, 30]	183	71.70%	106	67.09%	46	64.79%	31	86.11%
	[-50, 50]	227	85.66%	133	84.18%	59	83.10%	35	97.22%
	$(-\infty, -50)$ and $(50, +\infty)$	38	14.34%	25	15.82%	12	16.90%	1	2.78%
Proportional Differences	[-1, 1]	60	22.64%	45	28.48%	4	5.63%	11	30.56%
	[-5, 5]	148	55.85%	102	64.56%	25	35.21%	21	58.33%
	[-10, 10]	174	65.66%	115	72.78%	34	47.89%	25	69.44%
	[-20, 20]	200	75.47%	126	79.75%	48	67.61%	26	72.22%
	[-30, 30]	211	79.62%	132	83.54%	52	73.24%	27	75.00%
	[-50, 50]	227	85.66%	140	88.61%	59	83.10%	28	77.78%
	[-100, 100]	250	94.34%	153	96.84%	65	91.55%	32	88.89%
$(-\infty, -100)$ and $(100, +\infty)$	15	5.66%	5	3.16%	6	8.45%	4	11.11%	

(5) 3-D Distance Predicted Using LIDAR 50-ft DEM with an Interval of 50-ft versus DMI Distances

Table 9.37 shows the results of frequency analysis for the differences and proportional differences between predicted 3-D distances using LIDAR 50-ft DEM with a 50-ft interval and the DMI measured distances.

Based on Table 9.37, it is observed that

- The majority (over 50%) of the differences are in the [-20, 20] error range.
- When road types are considered, the majority of the differences for Interstate FTSEgs and NC FTSEgs are still in the [-20, 20] error range. However, for US FTSEgs, the corresponding percentage (46.48%) is less than 50%, but quite close.
- About 87% of the differences are in the error range of [-50, 50].
- The majority of proportional differences are in the [-5, 5] error range when all FTSEgs are considered together.

- When the road type of FTSegs are considered, the majority of Interstate FTSegs and NC FTSegs are still in the [-5, 5] error range. However, for US FTSegs, the corresponding error range becomes [-20, 20]. Furthermore, the corresponding percentage of proportional differences from US FTSegs to the [-10, 10] error range is close to 50% (46.48%).
- For the proportional differences, about 86% are in the [-50, 50] error range and over 90% are in the [-100, 100] error range.

Table 9.37 Results of Frequency Analysis (LIDAR 50-ft DEM, 50-ft Interval)

Error Format	Groups	All FTSegs		Interstate FTSegs		US FTSegs		NC FTSegs	
		#	%	#	%	#	%	#	%
Differences	[-5, 5]	60	22.64%	39	24.68%	12	16.90%	9	25.00%
	[-10, 10]	99	37.36%	65	41.14%	20	28.17%	14	38.89%
	[-20, 20]	149	56.23%	95	60.13%	33	46.48%	21	58.33%
	[-30, 30]	190	71.70%	114	72.15%	45	63.38%	31	86.11%
	[-50, 50]	230	86.79%	137	86.71%	58	81.69%	35	97.22%
	$(-\infty, -50)$ and $(50, +\infty)$	35	13.21%	21	13.29%	13	18.31%	1	2.78%
Proportional Differences	[-1, 1]	64	24.15%	48	30.38%	4	5.63%	12	33.33%
	[-5, 5]	152	57.36%	106	67.09%	25	35.21%	21	58.33%
	[-10, 10]	174	65.66%	116	73.42%	33	46.48%	25	69.44%
	[-20, 20]	201	75.85%	127	80.38%	48	67.61%	26	72.22%
	[-30, 30]	211	79.62%	132	83.54%	52	73.24%	27	75.00%
	[-50, 50]	227	85.66%	140	88.61%	59	83.10%	28	77.78%
	[-100, 100]	250	94.34%	153	96.84%	65	91.55%	32	88.89%
	$(-\infty, -100)$ and $(100, +\infty)$	15	5.66%	5	3.16%	6	8.45%	4	11.11%

(6) 3-D Distance Predicted Using 30-m NED with an Interval of 15-m versus DMI Distances

Table 9.38 shows the results of frequency analysis for the differences and proportional differences between predicted 3-D distances using NED with a 15-m interval and the DMI measured distances.

Based on Table 9.38, it is observed that

- The majority (over 50%) of the differences are in the [-20, 20] error range.
- When road types are considered, the majority of the differences for Interstate FTSegs and NC FTSegs are still in the [-20, 20] error range. However, for US FTSegs, the corresponding percentage (43.66%) is less than 50%, but quite close.
- About 83% of the differences are in the error range of [-50, 50].
- The majority of proportional differences are in the [-5, 5] error range when all FTSegs are considered together.

- When the road type of FTSEgs are considered, the majority of Interstate FTSEgs and NC FTSEgs are still in the [-5, 5] error range. However, for US FTSEgs, the corresponding error range becomes [-20, 20]. Furthermore, the corresponding percentage of proportional differences from US FTSEgs to the [-10, 10] error range is close to 50% (46.48%).
- For the proportional differences, about 86% are in the [-50, 50] error range and over 90% are in the [-100, 100] error range.

Table 9.38 Results of Frequency Analysis (NED, 15-m Interval)

Error Format	Groups	All FTSEgs		Interstate FTSEgs		US FTSEgs		NC FTSEgs	
		#	%	#	%	#	%	#	%
Differences	[-5, 5]	50	18.87%	32	20.25%	12	16.90%	6	16.67%
	[-10, 10]	88	33.21%	58	36.71%	17	23.94%	13	36.11%
	[-20, 20]	135	50.94%	84	53.16%	31	43.66%	20	55.56%
	[-30, 30]	176	66.42%	102	64.56%	44	61.97%	30	83.33%
	[-50, 50]	220	83.02%	130	82.28%	56	78.87%	34	94.44%
	$(-\infty, -50)$ and $(50, +\infty)$	45	16.98%	28	17.72%	15	21.13%	2	5.56%
Proportional Differences	[-1, 1]	53	20.00%	38	24.05%	2	2.82%	13	36.11%
	[-5, 5]	151	56.98%	105	66.46%	25	35.21%	21	58.33%
	[-10, 10]	172	64.91%	115	72.78%	33	46.48%	24	66.67%
	[-20, 20]	199	75.09%	126	79.75%	47	66.20%	26	72.22%
	[-30, 30]	211	79.62%	132	83.54%	52	73.24%	27	75.00%
	[-50, 50]	227	85.66%	140	88.61%	59	83.10%	28	77.78%
	[-100, 100]	250	94.34%	153	96.84%	65	91.55%	32	88.89%
	$(-\infty, -100)$ and $(100, +\infty)$	15	5.66%	5	3.16%	6	8.45%	4	11.11%

(7) 3-D Distance Predicted Using 30-m NED with an Interval of 30-m versus DMI Distances

Table 9.39 shows the results of frequency analysis for the differences and proportional differences between predicted 3-D distances using NED with a 30-m interval and the DMI measured distances.

Based on Table 9.39, it is observed that

- The majority (over 50%) of the differences are in the [-20, 20] error range.
- When road types are considered, the majority of the differences for Interstate FTSEgs and NC FTSEgs are still in the [-20, 20] error range. However, for US FTSEgs, the corresponding percentage (43.66%) is less than 50%, but quite close.
- About 82% of the differences are in the error range of [-50, 50].
- The majority of proportional differences are in the [-5, 5] error range when all FTSEgs are considered together.

- When the road type of FTSegs are considered, the majority of Interstate FTSegs and NC FTSegs are still in the [-5, 5] error range. However, for US FTSegs, the corresponding error range becomes [-20, 20]. Furthermore, the corresponding percentage of proportional differences from US FTSegs to the [-10, 10] error range is close to 50% (46.48%).
- For the proportional differences, about 86% are in the [-50, 50] error range and over 90% are in the [-100, 100] error range.

Table 9.39 Results of Frequency Analysis (NED, 30-m Interval)

Error Format	Groups	All FTSegs		Interstate FTSegs		US FTSegs		NC FTSegs	
		#	%	#	%	#	%	#	%
Differences	[-5, 5]	49	18.49%	31	19.62%	12	16.90%	6	16.67%
	[-10, 10]	86	32.45%	58	36.71%	16	22.54%	12	33.33%
	[-20, 20]	134	50.57%	83	52.53%	31	43.66%	20	55.56%
	[-30, 30]	176	66.42%	102	64.56%	44	61.97%	30	83.33%
	[-50, 50]	218	82.26%	128	81.01%	56	78.87%	34	94.44%
	$(-\infty, -50)$ and $(50, +\infty)$	47	17.74%	30	18.99%	15	21.13%	2	5.56%
Proportional Differences	[-1, 1]	53	20.00%	38	24.05%	2	2.82%	13	36.11%
	[-5, 5]	151	56.98%	105	66.46%	25	35.21%	21	58.33%
	[-10, 10]	172	64.91%	115	72.78%	33	46.48%	24	66.67%
	[-20, 20]	199	75.09%	126	79.75%	47	66.20%	26	72.22%
	[-30, 30]	211	79.62%	132	83.54%	52	73.24%	27	75.00%
	[-50, 50]	227	85.66%	140	88.61%	59	83.10%	28	77.78%
	[-100, 100]	250	94.34%	153	96.84%	65	91.55%	32	88.89%
	$(-\infty, -100)$ and $(100, +\infty)$	15	5.66%	5	3.16%	6	8.45%	4	11.11%

It is found based on the observations for each table presented above that most of the observations are the same, especially those observations regarding the majority issues. In other words, the differences in the percentages might be quite subtle. In addition, it is also observed that for two tables showing the frequency analysis results, which used the same elevation dataset but different model parameters (different intervals), the results are almost the same. Consequently, it lead to the thought that the effects of using different model parameters but the same elevation dataset on the predicted 3-D distances might be minor compared to the effects of using different elevation datasets on the predicted 3-D distances.

9.4 RESULT ANALYSIS AND SENSITIVITY ANALYSIS

The previous section represented the results of accuracy assessments by deploying analysis methods described earlier. The observations provided were based on each individual sample, i.e. comparing predicted 3-D

distances with DMI measured distances. This section analyzes the accuracy assessment results by summarizing these results together and comparing these results so that the effects of using different elevation datasets and model parameters on the accuracy of the predicted 3-D distances when compared to DMI measured distances could be evaluated. In other words, the previous section focused on evaluating the accuracy of the predicted 3-D distances using a particular elevation dataset and a particular model parameter by examining the differences between the predicted 3-D distances and the DMI measured distances. This section focuses on examining the differences of those differences.

This section takes such an approach that assuming A1 is the set of predicted 3-D distances using elevation dataset E1 and model parameter P1, A2 is the set of predicted 3-D distances using elevation dataset E2 and model parameter A2, and A is the set of 3-D distances measured by DMI, the difference between A1 and A are obtained and named as D1 and the difference between A2 and A are obtained and named as D2. The effects on the accuracy of predicted 3-D distances from the use of different elevation datasets (E1 and E2) and model parameters (P1 and P2) are evaluated by comparing D1 with D2. More specifically, D1 was examined in the previous section to evaluate the accuracy of the predicted 3-D distances using elevation dataset D1 and model parameter P1. D2 was examined in the previous section to evaluate the accuracy of the predicted 3-D distances using elevation dataset E2 and model parameter P2. It is the differences between D1 and D2 that would be examined in this section.

It is worth pointing out that an alternative approach would be to compare A1 and A2 directly without going through the procedure mentioned above. However, there would be no reference data involved and therefore, the results would only tell how significant the differences would be but there would be no way to tell which one would be better. As a result, in this case study, this alternative was not adopted.

The comparisons of this section included comparisons of means (100% and 95%), medians, absolute means (100% and 95%), RMSEs (100% and 95%), and frequencies.

9.4.1 Comparison of Means

The major purpose of comparing means is to determine if the population mean of the differences between the predicted 3-D distances using a particular elevation dataset and a particular model parameter and the DMI measured distances is significantly different from another population mean of the differences between the predicted 3-D distances using a different elevation dataset, a different model parameter, or both and the DMI measured distances. By doing so, the effects on the accuracy of using different elevation datasets and different model parameters could be examined from the aspect of error means.

As represented earlier, each sample had a 100% mean and a 95% mean. This section includes three subsections. The first subsection focuses on ANOVA on 100% means without differentiation in road types. The second subsection focuses on tabular and graphical representations of the 100% means. The third subsection focuses on tabular and graphical representations of the 95% means.

9.4.1.1 ANOVA

ANOVA can be performed to evaluate the effects of different treatments as stated earlier. In this case study, the use of a particular elevation dataset with a particular model parameter (interval) could be viewed as one treatment and consequently, there are 7 treatments in this case study (LIDAR point data, LIDAR 20-ft DEM with a 10-ft interval, LIDAR 20-ft DEM with a 20-ft interval, LIDAR 50-ft DEM with a 25-ft interval, LIDAR 50-ft DEM with a 50-ft interval, NED with a 15-m interval, and NED with a 30-m interval).

ANOVA can be conducted with any number of treatments. If the number of treatments is two, ANOVA is the same as pairwise t-testing. One assumption of ANOVA is the normal distribution of the populations under analysis. As represented earlier, the descriptive statistics and the distribution histograms revealed that this assumption was not valid in this case study. Another assumption of ANOVA is the equal-variance, which is not the case in this case study either. However, ANOVA is not very sensitive to violations of the assumption of a normal distribution and the assumption of equal-variance, provided the sample sizes per group are equal and not too small (Rao 1998). This is the reason that why in this case study ANOVA is performed on samples without

differentiation in road types (to assure a big enough sample size). Each sample has the same sample size of 265.

Table 9.40 shows the result of ANOVA analysis for differences with the null hypothesis of $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$, at the level of $\alpha = 0.05$, when all seven treatments in this case study are taken into consideration.

Table 9.40 Result of ANOVA for 100% Means (Differences, All Samples Considered)

<i>Source of variation</i>	<i>Degrees of freedom (df)</i>	<i>Sums of squares (SS)</i>	<i>Mean squares (MS)</i>	<i>F</i>	<i>F_c</i>
Treatment	104427.66	6.00	17404.61	17.7454	2.1035
Error	1812506.16	1848.00	980.80		
Total	1916933.83	1854.00			

Since F-statistic equals to 17.7454, which is much larger than the critical F value (2.1035), the null hypothesis is rejected at the confidence level of 95%. It could be concluded that from the aspect of the population mean of differences, the accuracies of the predicted 3-D distances using different elevation datasets and model parameters are not simultaneously the same. More specifically, from the aspect of the population mean of differences, the accuracies of the predicted 3-D distances are dependent on the elevation dataset and the model parameter being used.

Table 9.41 shows the result of ANOVA analysis for proportional differences with the null hypothesis of $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$, at the level of $\alpha = 0.05$, when all seven treatments in this case study are taken into consideration.

Table 9.41 Result of ANOVA for 100% Means (Proportional Differences, All Samples Considered)

<i>Source of variation</i>	<i>Degrees of freedom (df)</i>	<i>Sums of squares (SS)</i>	<i>Mean squares (MS)</i>	<i>F</i>	<i>F_c</i>
Treatment	988.86	6.00	164.81	0.0646	2.1035
Error	4715166.55	1848.00	2551.50		
Total	4716155.40	1854.00			

Since F-statistic equals to 0.0646, which is smaller than the critical F value (2.1035), the null hypothesis is accepted at the confidence level of 95%. It could be concluded that from the aspect of the population mean of proportional differences, the accuracies of the predicted 3-D distances using different elevation datasets and

model parameters are simultaneously the same. More specifically, from the aspect of the population mean of differences, the accuracies of the predicted 3-D distances are independent on the elevation dataset and the model parameter being used. Based on observations from Table 9.41, it is determined that pairwise comparisons of 100% means for proportional differences are unnecessary.

Table 9.42 provides the results of ANOVA for all pairwise comparisons of 100% means for differences with the null hypothesis of $\mu_1 = \mu_2$.

Table 9.42 Results of ANOVA for Pairwise Comparisons of 100% Means for Differences

Sample 1	Sample 2	<i>F</i>	<i>F_c</i>	Accept or Reject
D_A_LP	D A L20 10	11.3710	3.8591	Reject
	D A L20 20	12.8832	3.8591	Reject
	D A L50 25	11.7030	3.8591	Reject
	D A L50 50	3.5764	3.8591	Accept
	D A N 15	18.3042	3.8591	Reject
	D A N 30	19.2777	3.8591	Reject
D_A_L20_10	D A L20 20	0.0448	3.8591	Accept
	D A L50 25	34.0175	3.8591	Reject
	D A L50 50	21.3379	3.8591	Reject
	D A N 15	0.8327	3.8591	Accept
	D A N 30	1.0411	3.8591	Accept
D_A_L20_20	D A L50 25	35.9627	3.8591	Reject
	D A L50 50	23.0558	3.8591	Reject
	D A N 15	0.4933	3.8591	Accept
	D A N 30	0.6562	3.8591	Accept
D_A_L50_25	D A L50 50	2.3418	3.8591	Accept
	D A N 15	42.4625	3.8591	Reject
	D A N 30	43.5995	3.8591	Reject
D_A_L50_50	D A N 15	28.9469	3.8591	Reject
	D A N 30	29.9824	3.8591	Reject
D A N 15	D A N 30	0.0114	3.8591	Accept

Based on Table 9.42, it is observed that

- If two samples use the same elevation dataset but different model parameters (different intervals), i.e. D_A_L20_10 versus D_A_L20_20, D_A_L50_25 versus D_A_L50_50, and D_A_N_15 versus D_A_N_30, the null hypothesis of $\mu_1 = \mu_2$ is accepted with a confidence level of 95% ($\alpha = 0.05$).
- If one of the two samples being compared uses the LIDAR point data, the null hypothesis is always rejected with a confidence level of 95%.
- If one of the two samples being compared used the LIDAR 50-ft DEM, the null hypothesis is always rejected with a confidence level of 95%.

- For the two samples being compared, if one uses LIDAR 20-ft DEM and the other uses NED, the null hypothesis is always accepted with a confidence level of 95%.

With these observations, it is concluded that based on the 100% sample means, from the aspect of means of differences between predicted 3-D distances and DMI measured distances, the accuracies are not dependent on the interval being used, if the same elevation dataset is being used. However, the accuracies are dependent on the elevation dataset being used except that there is no statistically significant difference in the means of differences between samples using LIDAR 20-ft DEM and NED.

9.4.1.2 Comparison of 100% Means

The ANOVA presented earlier focused on determining if two or more means were significantly different, without differentiation of road types. This section compares the 100% means for samples when road types are taken into consideration. Table 9.43 provides the summary of all sample means of differences with rows specifying road types and columns specifying the elevation datasets and model parameters used. Figure 9.17 presents these sample means of differences graphically. In Figure 9.17, Y-axis represented the sample means. X-axis represented the elevation datasets and the intervals being used in predicting 3-D distances. LP indicated LIDAR point data. L20/10 indicated LIDAR 10-ft DEM with a 10-ft Interval. L20/20 indicated LIDAR 20-ft DEM with a 20-ft interval. L50/25 and L50/50 indicated LIDAR 50-ft DEM with a 25-ft interval and LIDAR 50-ft DEM with a 50-ft interval, respectively. Similarly, N15 indicated NED with a 15-m interval and N30 indicated NED with a 30-m interval.

Table 9.43 Summary of 100% Sample Means of Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	-8.43	-16.28	-16.82	-0.87	-3.57	-18.63	-18.92
Interstate FTSegs	-9.93	-18.86	-19.52	5.44	1.59	-21.56	-22.01
US FTSegs	-10.81	-16.79	-17.03	-15.19	-16.09	-18.97	-18.87
NC FTSegs	2.86	-3.95	-4.51	0.21	-1.54	-5.14	-5.41

It is observed from Figure 9.17 that the trend lines connecting means for a specific road type all followed the same trend, i.e. from left to right (from LIDAR point data to LIDAR 20-ft data, to LIDAR 50-ft data, and finally to NED), the lines go down first, then up, and then down again. In the cases of NC FTSEgs and US FTSEgs, the largest means correspond to the samples using LIDAR point data. In the cases of All FTSEgs and Interstate FTSEgs, the largest means correspond to the samples using LIDAR 50-ft DEM with an interval of 25-ft. In all cases the smallest means correspond to the sample of using NED with an interval of 30-m.

Figure 9.17 illustrated that sample means vary with different elevation datasets and intervals. The sample means also vary with road types. The trend lines for US FTSEgs and NC FTSEgs are almost parallel to each other. The trend lines for All FTSEgs and Interstate FTSEgs are close to each other because the majority (60%) of the FTSEgs is Interstate FTSEgs.

If only magnitude is concerned, it is observed that the samples using NED occupied the biggest values, regardless of road types. On the other hand, samples using LIDAR point data and LIDAR 50-ft DEMs occupied the smallest values, regardless of road types.

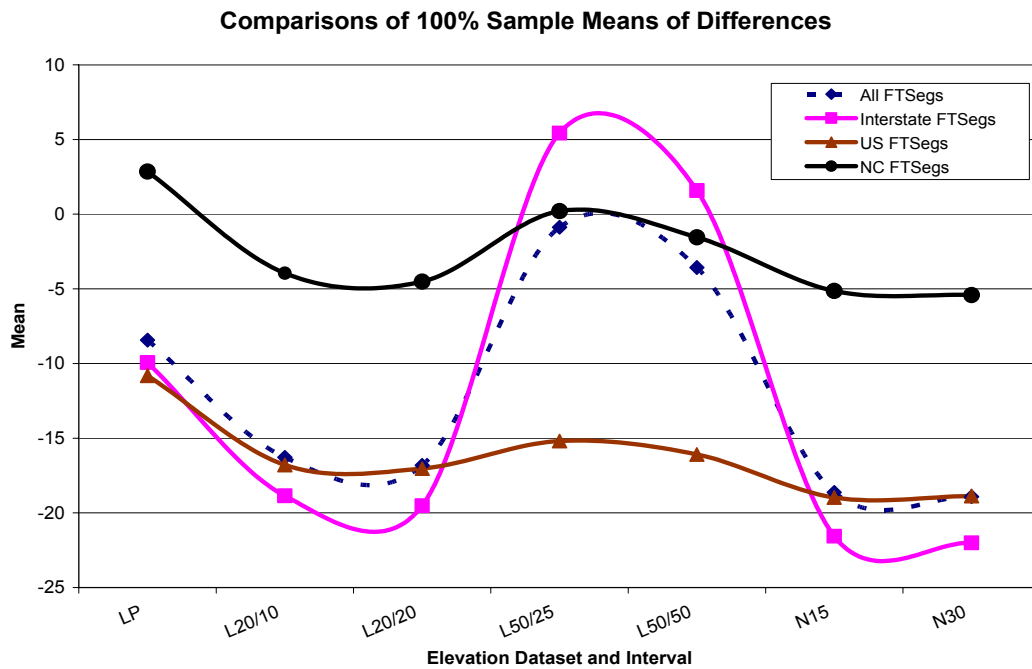


Figure 9.17 Comparisons of 100% Sample Means of Differences

Table 9.44 provides the summary of all sample means of proportional differences with rows specifying the road types and the columns specifying the elevation datasets and model parameters used. Figure 9.18 presents these sample means of proportional differences graphically.

Table 9.44 Summary of 100% Sample Means of Proportional Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	-6.48	-6.43	-6.51	-5.34	-5.83	-7.48	-7.49
Interstate FTSEgs	-1.17	-1.01	-1.11	0.62	-0.10	-2.11	-2.14
US FTSEgs	-15.02	-15.08	-15.15	-14.76	-14.91	-16.42	-16.41
NC FTSEgs	-12.92	-13.16	-13.21	-12.93	-13.08	-13.36	-13.38

Figure 9.18 illustrates that the trend lines connecting sample means of proportional differences have a different pattern than that of the differences. The variations of the means of proportional differences due to the use of different elevation datasets and intervals are very small as indicated by the trend lines that are almost flat lines shown in Figure 9.18. The four trend lines are almost parallel to each other. However, it is always the case that given an elevation dataset and a particular interval, the sample mean of the proportional differences for Interstate FTSEgs is always the highest one, followed by the mean of proportional differences of NC FTSEgs while the mean of proportional differences of US FTSEgs is always the lowest one. Furthermore, all sample means of proportional differences (except the one corresponding to the sample mean of proportional differences of Interstate FTSEgs when using LIDAR 20-ft DEM with an interval of 25-ft) are less than 0. In other words, a higher sample mean in this case is associated with a smaller magnitude of the sample mean. It could be concluded that from the aspect of sample means of proportional differences, the Interstate FTSEgs are the most accurate group with the smallest magnitude of sample means of proportional differences while the US FTSEgs are the least accurate group with the biggest magnitude of sample means of proportional differences.

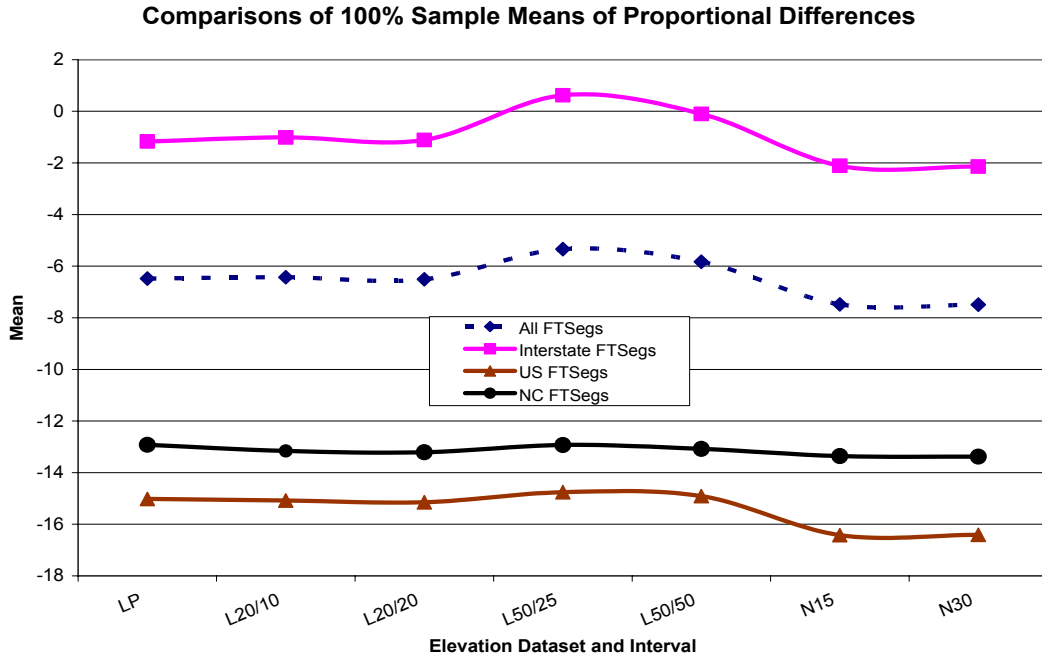


Figure 9.18 Comparisons of 100% Sample Means of Proportional Differences

9.4.1.3 Comparison of 95% Means

It is of concern that means are sensitive to outliers in the samples. As a surrogate, for each sample, the “best” 95% of the data are identified and kept while the “worst” 5% of the data are discarded. Means of the remaining 95% of the data are also calculated in this case study. Table 9.45 summarizes these 95% means of differences for all samples in this case study. Figure 9.19 graphically illustrates the variations of the 95% means of differences due to the use of different elevation datasets and intervals and due to the different road types.

Table 9.45 Summary of 95% Sample Means of Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	-5.74	-12.57	-13.08	-2.82	-4.43	-14.88	-15.16
Interstate FTSEgs	-7.32	-15.34	-15.99	1.71	-2.13	-18.05	-18.49
US FTSEgs	-8.14	-12.60	-12.80	-11.92	-12.59	-14.70	-14.63
NC FTSEgs	3.34	-0.99	-1.54	3.17	1.43	-2.05	-2.31

Comparing Figure 9.19 with Figure 9.18 (100% means) reveals that the patterns of the trend lines are almost the same. However, Figure 9.19 demonstrates less variations of the sample means (95% means) of differences for

All FTSegs and Interstate FTSegs than those variations of the 100% sample means of differences, which was indicated by the shapes of the trend lines. For US FTSegs and NC FTSegs, the trend lines are still parallel to each other. When compared to those two trend lines in Figure 9.18, it is obvious that the 95% means of the US FTSegs and NC FTSegs are higher than the corresponding 100% means.

If only the magnitude of the means is concerned, the observations are similar to those obtained based on 100% means and therefore, are not repeated herein.

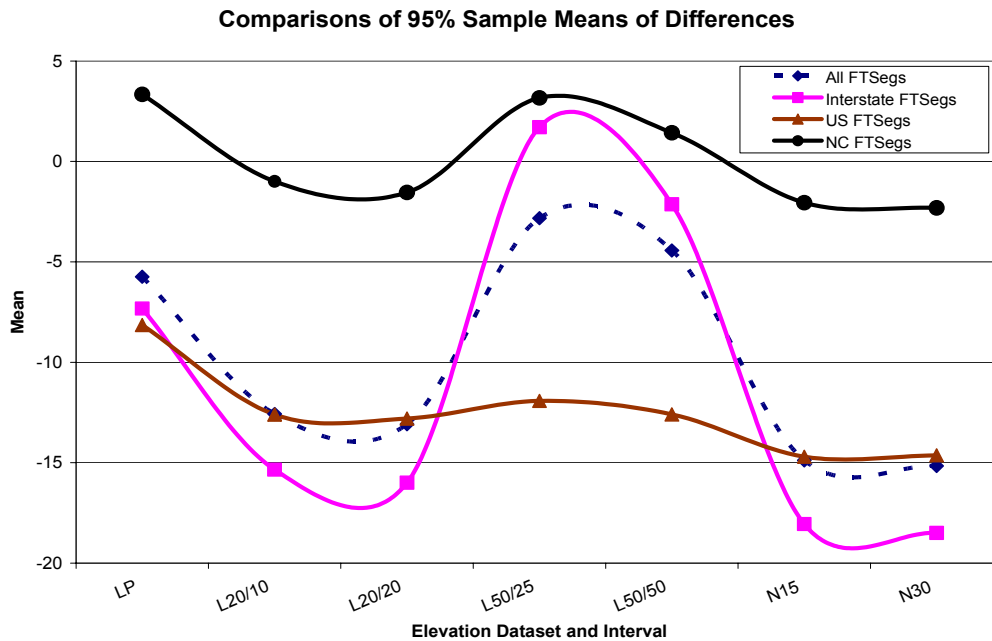


Figure 9.19 Comparisons of 95% Sample Means of Differences

Table 9.46 summarizes these 95% means of proportional differences for all samples in this case study. Figure 9.20 graphically illustrates the variations of the 95% means of the proportional differences due to the use of different elevation datasets and intervals and due to the different road types.

Similar to Figure 9.18, in Figure 9.20, all four trend lines are still kind of parallel to each other. However, different from Figure 9.18, Figure 9.20 illustrates that the trend lines of Interstate FTSegs and All FTSegs are closer to each other. The trend lines for All FTSegs and US FTSegs are closer to the straight line corresponding to 0. The trend line for NC FTSegs is further away from 0. If only the magnitude of the means is concerned,

the group of Interstate FTSEgs is still the group with the best accuracy. The group of NC FTSEgs becomes the group with the worst accuracy.

Table 9.46 Summary of 95% Sample Means of Proportional Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	-0.50	-0.50	-0.59	0.68	0.17	-1.55	-1.56
Interstate FTSEgs	0.62	-1.13	-1.22	1.48	0.73	-0.40	-0.43
US FTSEgs	-4.03	-4.07	-4.15	-3.74	-3.90	-5.48	-5.47
NC FTSEgs	-16.62	-16.86	-16.91	-16.26	-16.76	-17.06	-17.08

Comparisons of 95% Sample Means of Proportional Differences

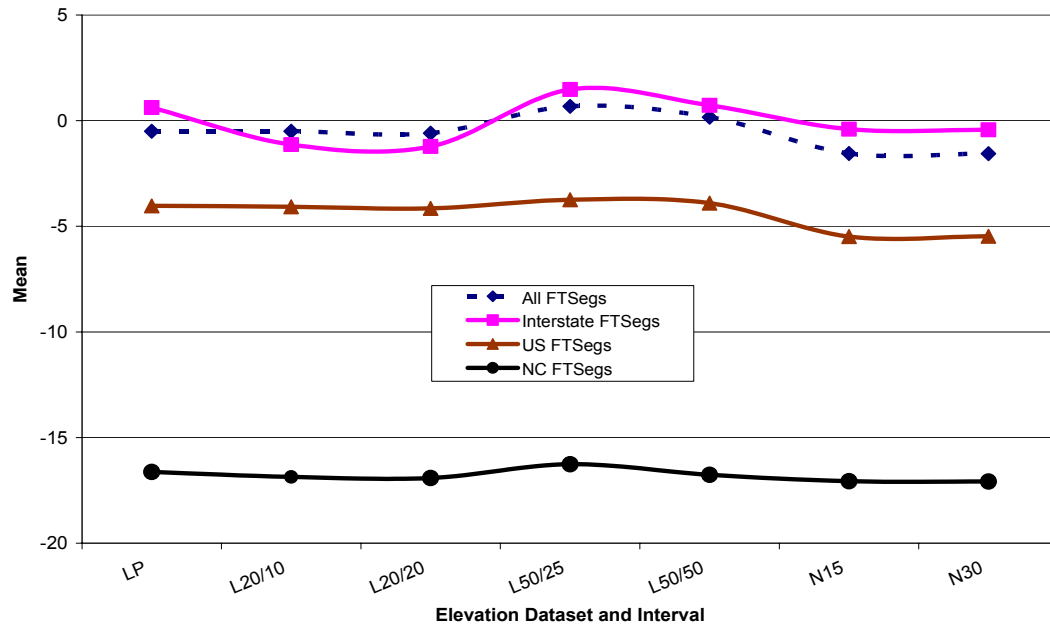


Figure 9.20 Comparisons of 95% Sample Means of Proportional Differences

9.4.2 Comparison of Medians

As stated earlier, a mean was sensitive to outliers. This is also the reason that when means were compared, not only 100% of the data were used to calculate the mean, but also the “best” 95% of the data were used to compute the 95% mean so that the outliers were discarded. On the other hand, median is not sensitive to outliers and therefore, medians based on 100% of the data are compared to evaluate the effects on the accuracy

of the predicted 3-D distances when compared to DMI measured distances, from the use of different elevation datasets and intervals and from the road types.

Table 9.47 summarizes the medians for all samples. Figure 9.21 presents the comparison of these medians graphically.

Table 9.47 Summary of Medians of Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	-4.94	-8.99	-9.45	-0.74	-2.72	-10.89	-11.21
Interstate FTSEgs	-5.53	-12.16	-12.43	2.37	-0.85	-13.34	-13.89
US FTSEgs	-8.68	-9.82	-9.82	-9.72	-9.72	-9.83	-9.82
NC FTSEgs	5.29	-4.55	-4.98	-4.05	-4.40	-5.14	-5.55

Comparisons of Medians of Differences

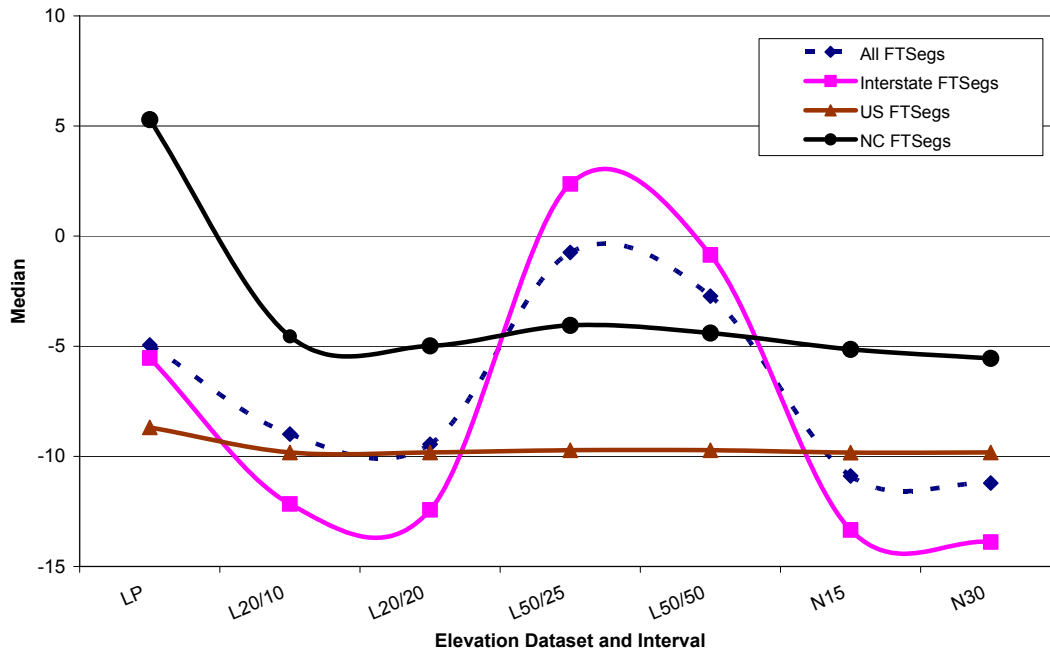


Figure 9.21 Comparison of Medians of Differences

Figure 9.21 illustrates that for All FTSEgs, the variation of medians due to the use of different elevation datasets and intervals is quite significant. However, the major contributor to this variation is the Interstate FTSEgs. For US FTSEgs, the medians are quite close when different elevation datasets and intervals are used, except the one using LIDAR point data. This is also true with the medians of NC FTSEgs. The pattern of the trend lines is

similar to the pattern of the trend lines shown in Figures 9.17 and 9.19, in which means were compared. In addition, the medians vary with different road types when using the same elevation dataset and the same interval. It is also observed that the medians of a specific type of FTSegs are close to each other if the same elevation dataset, but different intervals were used.

Table 9.48 summarizes the medians for all samples. Figure 9.22 presents the comparison of these medians graphically.

Table 9.48 Summary of Medians of Proportional Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	-0.72	-1.33	-1.34	-0.06	-0.39	-1.56	-1.58
Interstate FTSegs	-0.69	-1.13	-1.17	0.31	-0.11	-1.26	-1.29
US FTSegs	-1.63	-3.20	-3.21	-1.97	-2.11	-3.35	-3.35
NC FTSegs	0.36	-0.26	-0.28	-0.12	-0.17	-0.24	-0.26

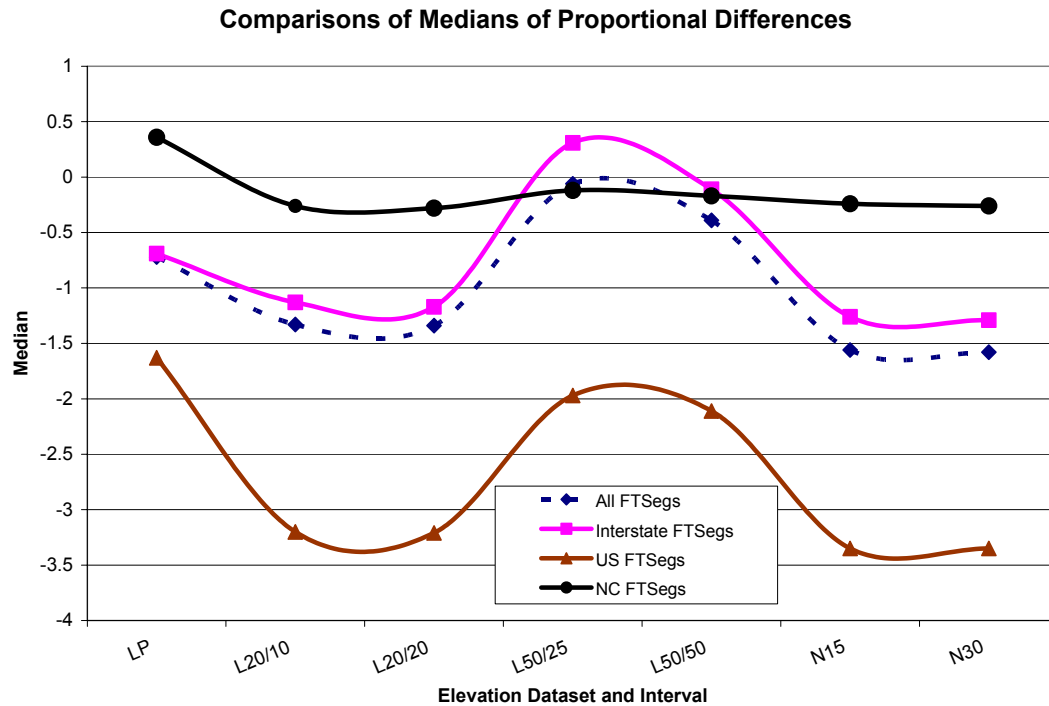


Figure 9.22 Comparison of Medians of Proportional Differences

Figure 9.22 illustrates that the medians of proportional differences vary with different elevation datasets and road types. The trend lines for US FTSegs and Interstate FTSegs are almost parallel to each other while the

trend line for NC FTSEgs shows a different pattern. For NC FTSEgs, the medians are still quite close to each other even different elevation datasets and intervals are used except using LIDAR point data. Different from the medians of differences showed in Figure 9.21, Figure 9.22 demonstrates that for US FTSEgs and NC FTSEgs, the variations of the proportional differences are quite significant when different elevation datasets were used. However, it is still true that for any type of FTSEgs, if the same elevation dataset but different intervals are used, the corresponding two medians are still close to each other.

9.4.3 Comparison of Absolute Means

In many cases, the concern is the magnitude of errors. In calculating means, the errors with positive values and errors with negative values would cancel out each other and as a result, the calculated mean would tend to be close to 0. As a surrogate, absolute means are means that are calculated using the absolute values of the errors. In other words, only the magnitude of errors would be used in calculating the absolute means so that errors with positive values and errors with negative values would not cancel out each other. It is obvious that the absolute mean is also sensitive to outliers and consequently, 100% absolute means and 95% absolute means for both differences and proportional differences are calculated and compared in this case study.

Table 9.49 summarizes the 100% absolute means for differences. Figure 9.23 graphically illustrates the comparisons of these 100% absolute means for differences.

Table 9.49 Summary of 100% Absolute Means for Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	19.92	24.26	24.55	23.64	23.58	25.82	26.00
Interstate FTSEgs	18.64	24.00	24.45	24.72	23.37	26.37	26.71
US FTSEgs	22.71	28.40	28.50	26.33	27.01	28.94	28.87
NC FTSEgs	20.06	17.28	17.19	18.53	17.72	17.22	17.22

Based on Figure 9.23 and Table 9.49, it is observed that the sample using LIDAR point data has the lowest absolute mean when All FTSEgs are considered. The samples using NED have the highest absolute means. All other samples using LIDAR DEMs have absolute means close to each other. This is also the trend when examining the Interstate FTSEgs. For US FTSEgs, the sample using the LIDAR point data still occupies the

lowest absolute mean while the other samples have absolute means close to each other. The exception is the NC FTSEgs, in which, the sample using LIDAR point data has a higher absolute mean than the samples using LIDAR DEMs and NED. In addition, when using LIDAR point data, the absolute means of Interstate FTSEgs, US FTSEgs, and NC FTSEgs are close to each other while using other elevation data, the absolute mean of US FTSEgs are the highest followed by the absolute means of Interstate FTSEgs and the absolute means of NC FTSEgs are the lowest.

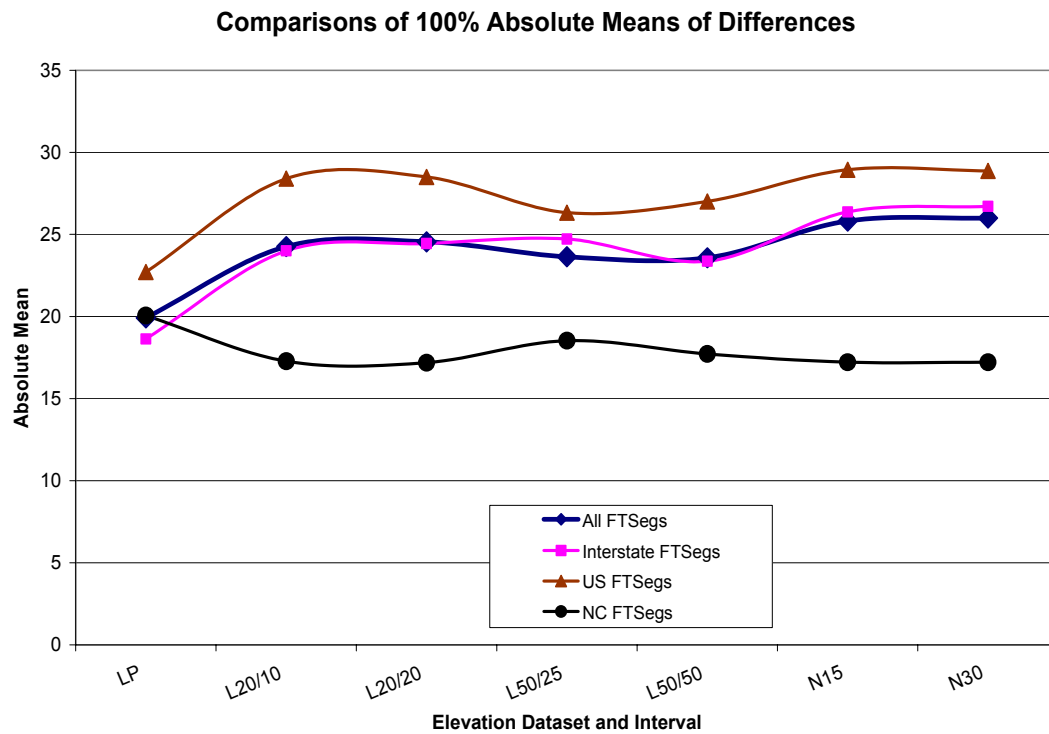


Figure 9.23 Comparison of 100% Absolute Means of Differences

Table 9.50 summarizes the 95% absolute means for differences. Figure 9.24 graphically illustrates the comparisons of these 95% absolute means for differences.

Table 9.50 Summary of 95% Absolute Means for Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	17.85	20.98	21.22	20.62	19.73	22.44	22.61
Interstate FTSEgs	16.49	20.76	21.17	20.46	18.82	23.12	23.49
US FTSEgs	20.67	24.82	24.87	23.66	24.09	25.20	25.16
NC FTSEgs	18.27	15.02	14.89	16.55	15.61	14.77	14.74

Comparing Figure 9.24 with Figure 9.23 reveals that the pattern of the trend lines is almost the same except that in Figure 9.24 these lines are closer to each other because the “worst” 5% of the data are discarded. Same observations could be derived as those from Figure 9.23 and therefore, they are not repeated herein.

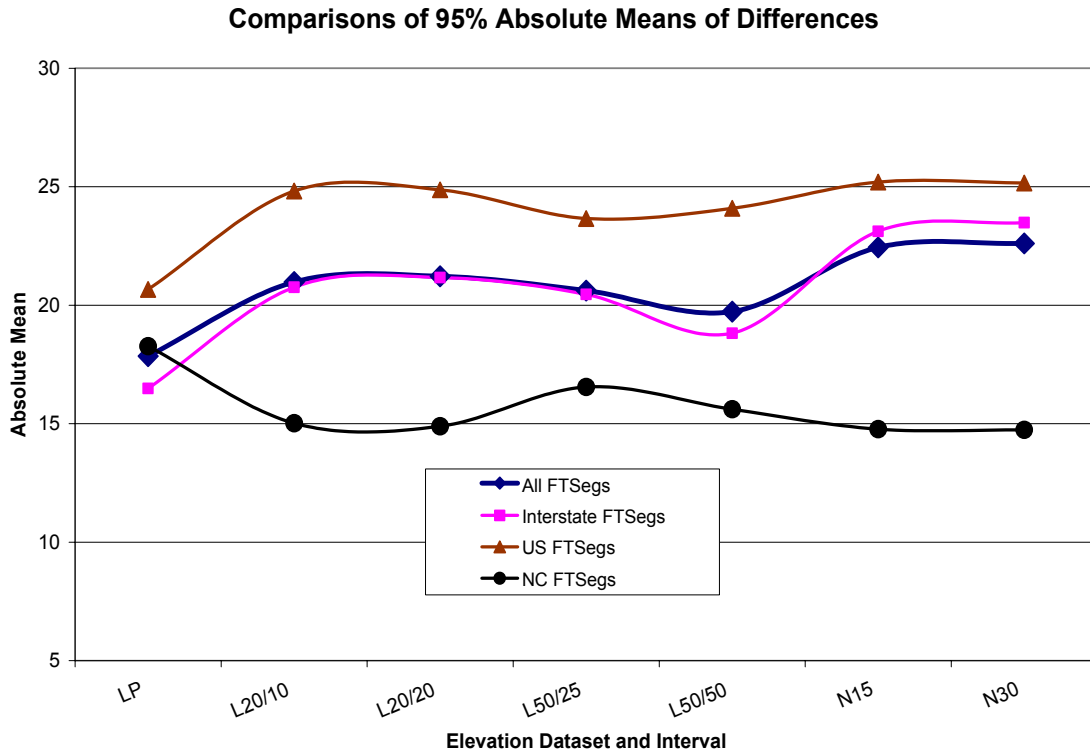


Figure 9.24 Comparison of 95% Absolute Means of Differences

Table 9.51 summarizes the 100% absolute means for proportional differences. Figure 9.25 graphically illustrates the comparisons of these 100% absolute means for proportional differences.

Table 9.51 Summary of 100% Absolute Means for Proportional Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	21.56	22.52	22.52	22.13	21.94	22.11	22.11
Interstate FTSEgs	14.09	15.10	15.10	14.71	14.41	14.68	14.69
US FTSEgs	31.22	32.77	32.75	32.13	32.09	32.02	32.02
NC FTSEgs	35.25	34.90	34.90	35.01	35.01	35.13	35.13

Based on Figure 9.25, it is observed that the trend lines are almost all parallel to each other with the trend line for NC FTSEgs on the top, followed by the trend line for US FTSEgs and the trend line for All FTSEgs, and the trend line for Interstate FTSEgs at the bottom. These trend lines are quite flat indicating that the differences in the absolute means for samples using different elevation datasets and different intervals are very small. However, the absolute mean of the sample using the LIDAR point data is still lower than the others. The only exception is the NC FTSEgs, in which the sample using LIDAR point data has the highest absolute mean. Still, the absolute means of NC FTSEgs using different elevation dataset and intervals are very close to each other.

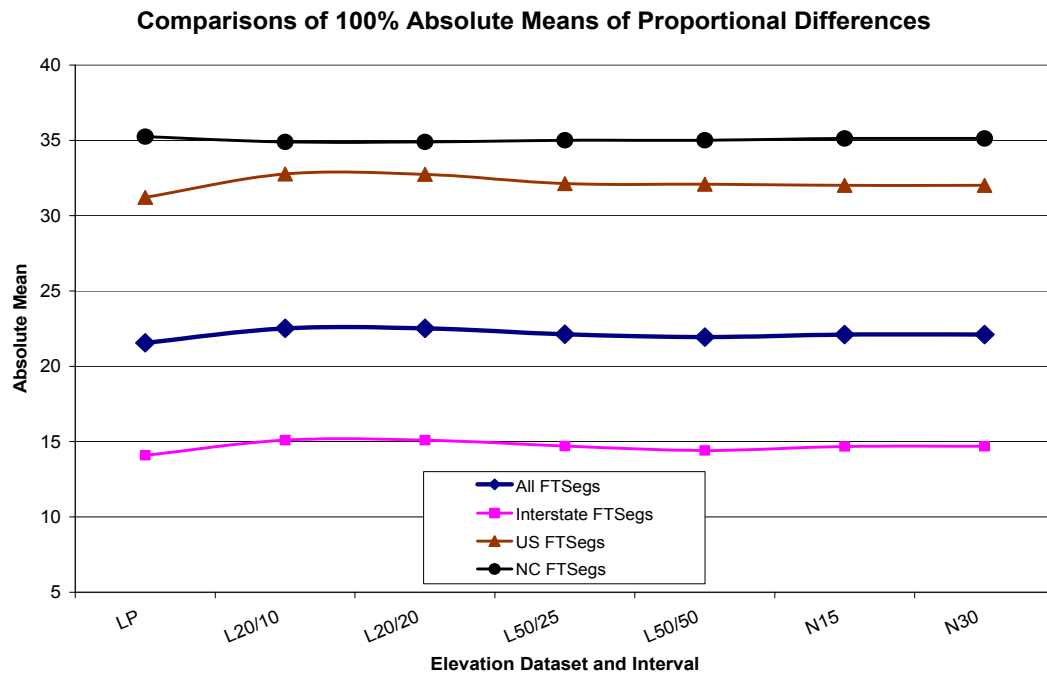


Figure 9.25 Comparison of 100% Absolute Means of Proportional Differences

Table 9.52 summarizes the 95% absolute means for proportional differences. Figure 9.26 graphically illustrates the comparisons of these 95% absolute means for proportional differences.

Table 9.52 Summary of 95% Absolute Means for Proportional Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	12.91	13.87	13.87	13.51	13.30	13.48	13.49
Interstate FTSEgs	8.79	9.80	9.81	9.48	9.16	9.39	9.40
US FTSEgs	21.08	22.69	22.67	22.02	21.98	21.91	21.91
NC FTSEgs	23.01	22.63	22.64	22.26	22.75	22.89	22.89

Figure 9.26 illustrates a pattern of the trend lines similar to that illustrated in Figure 9.25 except the trend line of US FTSEgs and the trend line of NC FTSEgs were closer to each other than in Figure 9.25. Similar observations to those obtained from Figure 9.25 could be derived from Figure 9.26 and therefore, are not repeated herein.

The observations based on the comparisons of absolute means lead to the conclusion that when comparing the absolute means for samples using different elevation datasets and intervals, the patterns shown with 100% absolute means and 95% absolute means are very close to each other. Consequently, observations based on these comparisons would be the same.

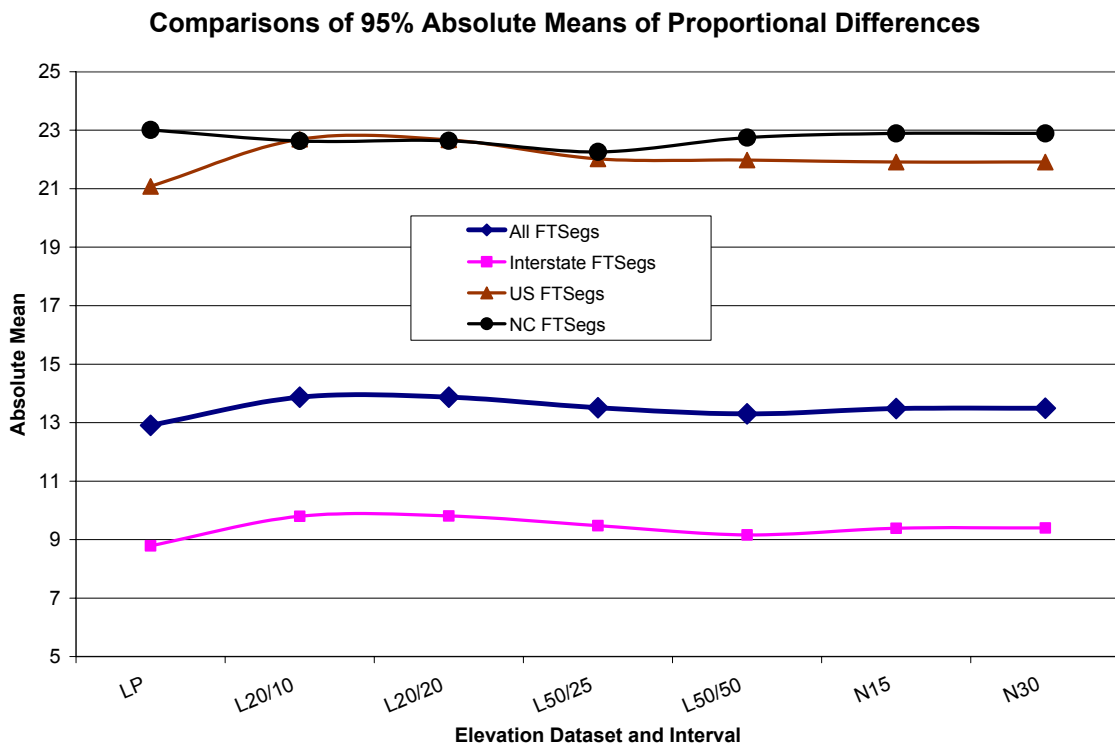


Figure 9.26 Comparison of 95% Absolute Means of Proportional Differences

9.4.4 Comparison of RMSEs

The comparisons of means, medians, and absolute means lead to some useful observations of the accuracy of the predicted 3-D distances using different elevation datasets and model parameters when compared to DMI

measured distances. However, none of these were a universally agreed upon parameter that could describe accuracies very well. In reality, it is rarely the case that a mean, a median, or an absolute mean would be used singularly to describe the accuracy of a dataset. The major contribution of comparing these above-mentioned parameters in this case study was to reveal the distributions of the errors of the predicted 3-D distances using different elevation dataset and intervals when compared to DMI measured distances so that it could be determined if the effects on the accuracy due to the use of different elevation datasets and intervals were the same. In other words, if the sample mean of the differences between dataset A and the reference dataset R is bigger than the sample mean of the differences between dataset B and the reference dataset R, it could be only concluded that the accuracy of dataset A is different from the accuracy of the dataset B when compared to the reference dataset R, but there is no way to say that dataset A has a better accuracy than dataset B, and vice versa.

On the other hand, RMSE is a widely used measure in evaluating accuracies and errors. Different from the comparison of means, medians, and absolute means, a bigger RMSE is associated with a worse accuracy while a smaller RMSE is associated with a better accuracy. However, RMSE is also sensitive to outliers and consequently, both 100% RMSEs and 95% RMSEs are compared in this case study.

9.4.4.1 Comparison of 100% RMSEs

Table 9.53 summarizes the 100% RMSEs for differences. Figure 9.27 graphically illustrates the comparisons of these 100% RMSEs for differences.

Table 9.53 Summary of 100% RMSEs for Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	25.65	33.30	33.73	33.90	34.16	35.53	35.74
Interstate FTSegs	24.76	33.38	33.99	35.84	35.83	36.43	36.83
US FTSegs	28.26	37.39	37.62	33.83	35.04	38.69	38.53
NC FTSegs	24.04	22.69	22.62	23.99	23.05	22.76	22.79

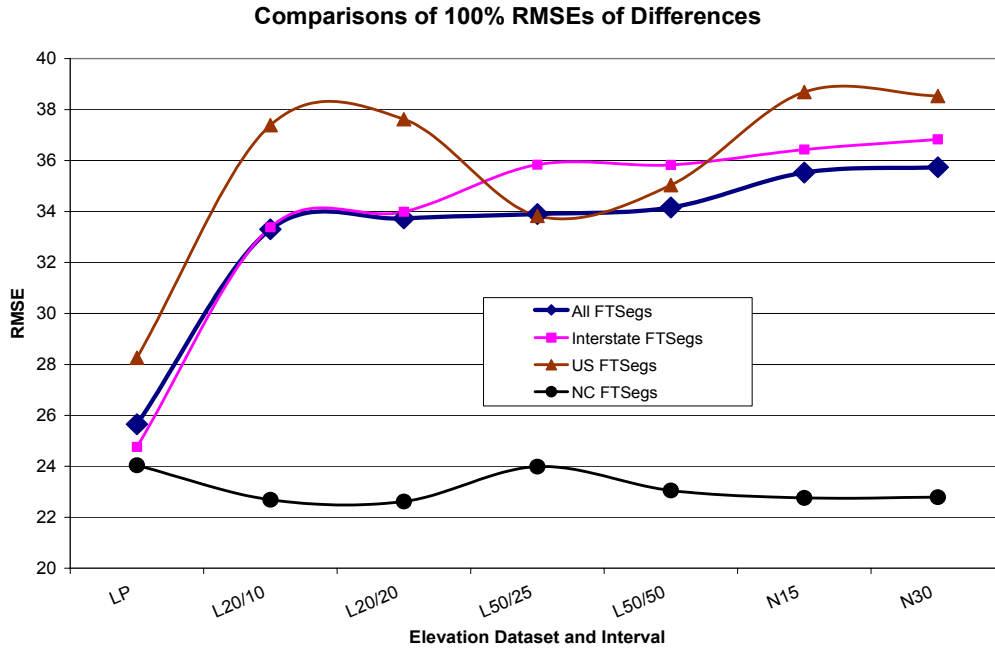


Figure 9.27 Comparison of 100% RMSEs for Differences

Based on Table 9.53 and Figure 9.27, it is observed that for All FTSEgs, the sample using the LIDAR point data has the lowest RMSE (the highest accuracy). The RMSEs of samples using LIDAR DEMs are higher than the RMSE of the sample using LIDAR point data, but lower than the RMSEs of the samples using NED. While the RMSEs for samples using LIDAR DEMs are quite close to each other, the trend line for All FTSEgs illustrates that RMSEs increase with the increase in the resolution of DEM and NED and the increase in the interval size.

For Interstate FTSEgs, the trend line is quite close to the trend line for All FTSEgs except that the RMSEs of samples using LIDAR 50-ft DEMs are closer to the RMSEs of samples using NED than the RMSEs of samples using LIDAR 20-ft DEMS.

For US FTSEgs, the RMSE of the sample using LIDAR point data is still the lowest while the RMSEs of the samples using NED are higher than the others. The exception is the samples using LIDAR 50-ft DEM. The RMSEs of these two samples are higher than the RMSE of the sample using LIDAR point data, but lower than the others.

For NC FTSegs, the RMSE of the sample using LIDAR point data is higher than the others. The RMSEs of the samples using LIDAR 20-ft DEMs are the lowest. However, the values of these RMSEs are very close to each other.

Table 9.54 summarizes the 100% RMSEs for proportional differences. Figure 9.28 graphically illustrates the comparisons of these 100% RMSEs for differences.

Table 9.54 Summary of 100% RMSEs for Proportional Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	50.44	51.28	51.26	50.79	50.75	50.68	50.68
Interstate FTSegs	32.06	33.16	33.12	32.33	32.24	32.29	32.29
US FTSegs	63.72	64.99	64.98	64.51	64.45	64.14	64.14
NC FTSegs	78.80	78.84	78.84	78.71	78.78	78.83	78.83

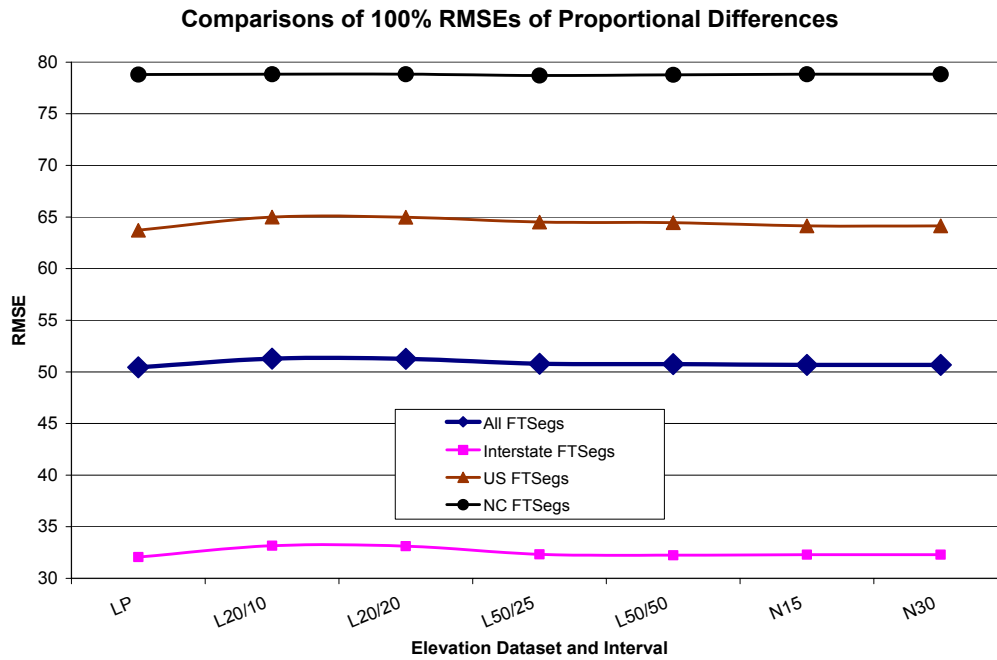


Figure 9.28 Comparison of 100% RMSEs for Proportional Differences

Based on Figure 9.28 and Table 9.54, it is observed that the trend lines for All FTSegs, Interstate FTSegs, US FTSegs, and NC FTSegs are almost parallel to each other. The RMSEs of the samples of a particular road type but using different elevation datasets and intervals are close to each other. When all roads are considered, the

sample using LIDAR point data had the lowest RMSE. The lowest RMSEs correspond to the samples using LDIAR point data for Interstate FTSegs and US FTSegs and the samples using LIDAR 20-ft DEMs for NC FTSegs.

9.4.4.2 Comparison of 95% RMSEs

95% RMSEs were obtained by removing the “worst” 5% of the data in the samples. Table 9.55 illustrated the comparisons based on 95% RMSEs for differences. Figure 9.29 graphically represented these comparisons.

Table 9.55 Summary of 95% RMSEs for Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	22.48	27.65	28.01	27.71	26.01	29.95	30.16
Interstate FTSegs	21.39	27.95	28.53	28.75	25.38	31.38	31.77
US FTSegs	25.28	31.31	31.42	29.83	30.51	32.29	32.21
NC FTSegs	21.32	18.56	18.31	20.91	19.47	17.98	17.92

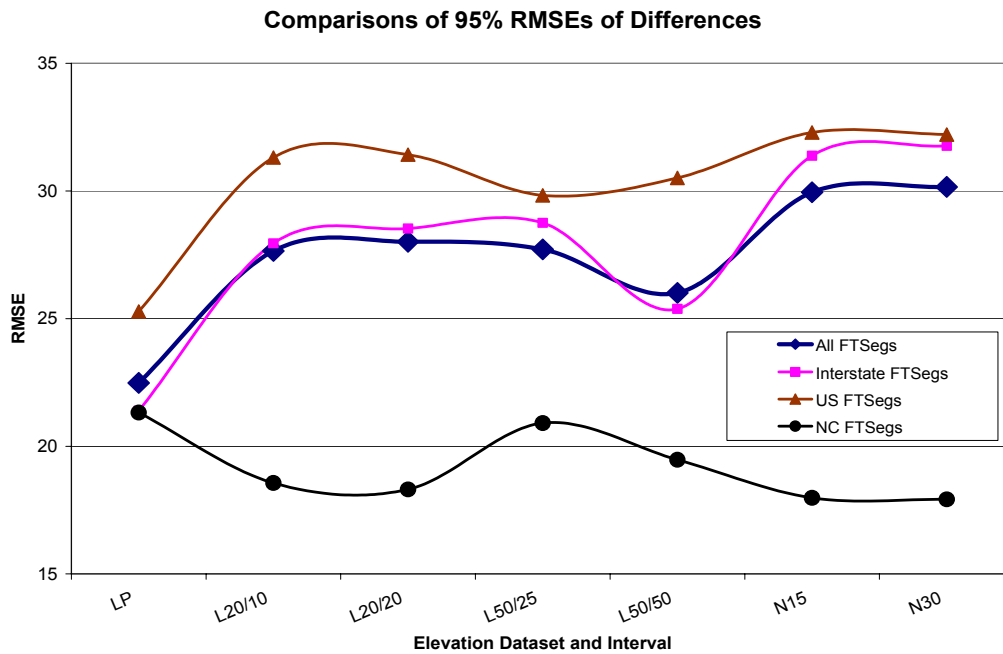


Figure 9.29 Comparison of 95% RMSEs for Differences

Based on Table 9.55 and Figure 9.29, it is observed that by discarding the “worst” 5% of the data, the values of RMSEs are smaller than the corresponding RMSEs when 100% of the data are used. But still, the pattern of the trend lines is similar to the pattern of the trend lines for 100% RMSEs. The RMSEs of the samples using LIDAR point data are the lowest (the highest accuracy) while the RMSEs of the samples using NED are the highest (the lowest accuracy). The exception is the NC FTSegs, in which the values of RMSEs of the samples using LIDAR point data and LIDAR 50-ft DEMs are close to each other while the RMSEs of the samples using LIDAR 20-ft DEMs and NED are close to each other but lower than the other RMSEs.

Table 9.56 summarizes the comparisons based on 95% RMSEs for proportional differences. Figure 9.30 graphically represents these comparisons.

Table 9.56 Summary of 95% RMSEs for Proportional Differences

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	24.90	26.36	26.33	25.58	25.45	25.35	25.25
Interstate FTSegs	18.19	19.76	19.75	18.80	18.64	18.52	18.52
US FTSegs	35.47	37.73	37.72	36.87	36.76	36.19	36.19
NC FTSegs	51.93	51.82	51.82	51.78	51.89	52.02	52.02

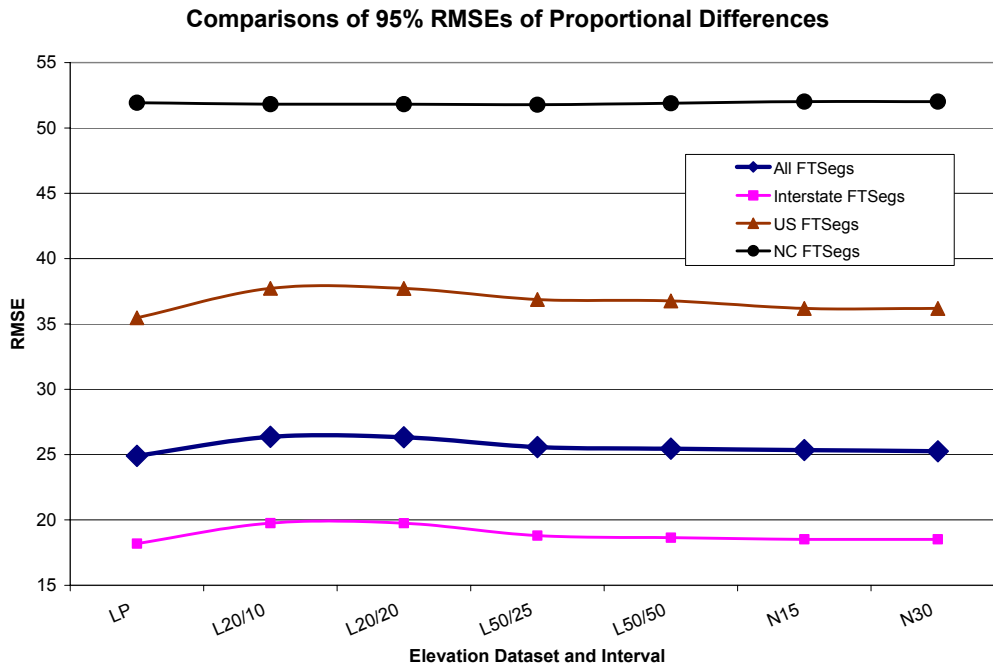


Figure 9.30 Comparison of 95% RMSEs for Proportional Differences

Based on Table 9.56 and Figure 9.30, it is observed that the trend lines for All FTSEgs, Interstate FTSEgs, US FTSEgs, and NC FTSEgs are almost parallel to each other. The RMSEs of the samples of a particular road type but using different elevation datasets and intervals are close to each other. When all roads are considered, the sample using LIDAR point data has the lowest RMSE. The lowest RMSEs corresponded to the samples using LIDAR point data for Interstate FTSEgs and US FTSEgs and the samples using LIDAR 50-ft DEMs for NC FTSEgs. However, the RMSE values for NC FTSEgs are very close to each other. In all cases, the RMSEs of the samples using NED are higher than the corresponding RMSEs of the samples using LIDAR point data. It is also worth pointing out herein that by removing the “worst” 5% of the data, the improvements on the RMSEs for proportional differences are significant and more significant than those for differences.

9.4.5 Comparison of Frequencies

As stated earlier, statistics such as means, medians, and RMSEs all focused on getting a singular measure for a sample while the frequency analysis focused on counting the number of occurrences from a sample given a range and calculating the percentage. Consequently, frequency analysis could provide useful information regarding the error distribution in this case study. Furthermore, accuracy requirements are dependent on applications. With frequency analysis, a particular application might specify a particular error range as the required accuracy and consequently, the corresponding percentage of occurrences in this given error range would be a direct measure of the accuracy.

The results of the frequency analyses would be compared in this section to evaluate the effects of using different elevation datasets and intervals and the effects of road types on the accuracy of the predicted 3-D distances. In this section, for each of the given error ranges (which was used in conducting frequency analyses), the percentages of the FTSEgs falling into a particular error range, when using different elevation datasets and intervals and from different road types, are compared.

9.4.5.1 Comparison of Frequencies for Differences

Table 9.57 summarizes the percentages for differences within the error range of $[-5, 5]$ for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals.

Figure 9.31 illustrates comparisons of these percentages.

Table 9.57 Summary of Percentages for Differences ($[-5, 5]$)

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	19.62%	23.02%	22.64%	22.64%	22.64%	18.87%	18.49%
Interstate FTSegs	20.89%	24.68%	24.05%	24.05%	24.68%	20.25%	19.62%
US FTSegs	18.31%	16.90%	16.90%	16.90%	16.90%	16.90%	16.90%
NC FTSegs	16.67%	27.78%	27.78%	27.78%	25.00%	16.67%	16.67%

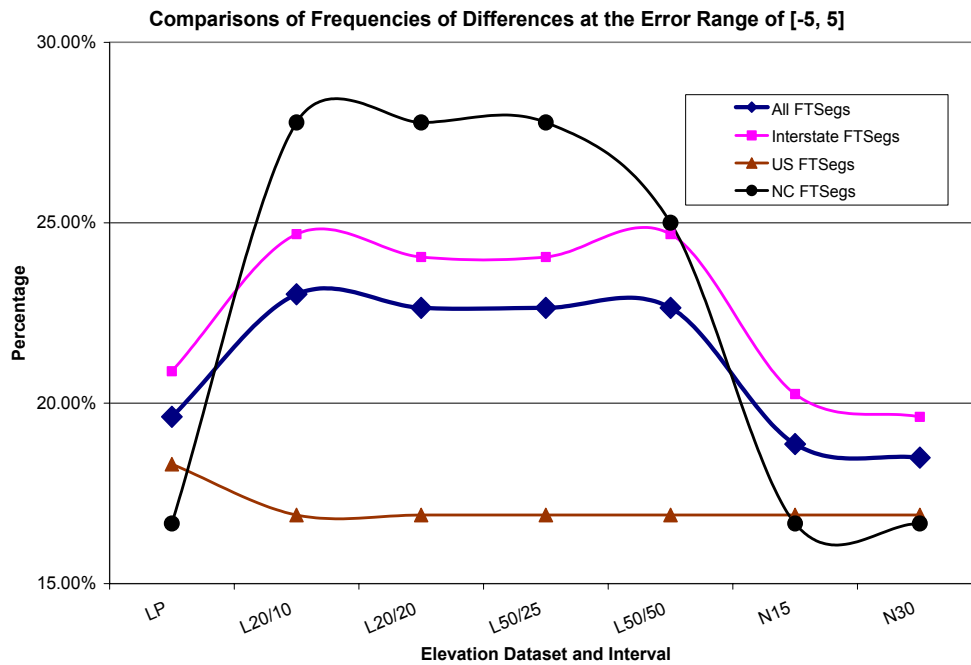


Figure 9.31 Illustration of Percentages at the Error Range of $[-5, 5]$ for Differences

Based on Table 9.57 and Figure 9.31, it is observed that the samples using LIDAR DEMs are the groups with the highest percentages (the best accuracy) at the given error range of $[-5, 5]$. The samples using NED are the groups with the lowest percentages (the worst accuracy) at the given error range of $[-5, 5]$. The samples using LIDAR point data have their percentages higher (better accuracy) than those of the samples using NED for a

particular road type. When only US FTSegs are concerned, the sample using LIDAR point data has the highest percentage when compared to the percentages of samples using other elevation datasets and intervals.

Table 9.58 summarizes the percentages for differences within the error range of [-10, 10] for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.32 illustrates comparisons of these percentages.

Table 9.58 Summary of Percentages for Differences ([-10, 10])

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	36.60%	36.60%	36.60%	39.62%	37.36%	33.21%	32.45%
Interstate FTSegs	43.04%	39.87%	40.51%	46.20%	41.14%	36.71%	36.71%
US FTSegs	28.17%	26.76%	26.76%	26.76%	28.17%	23.94%	22.54%
NC FTSegs	25.00%	41.67%	38.89%	36.11%	38.89%	36.11%	33.33%

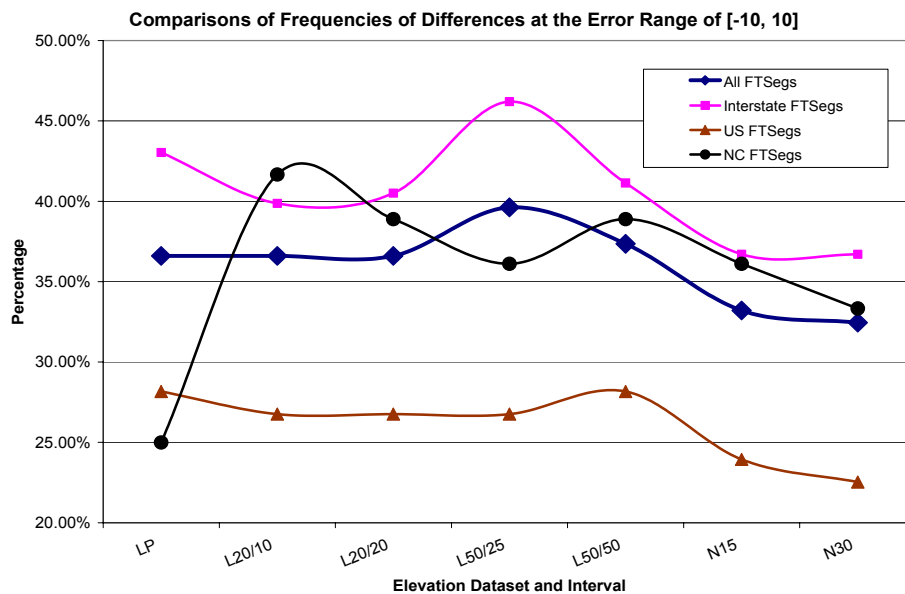


Figure 9.32 Illustration of Percentages at the Error Range of [-10, 10] for Differences

Based on Table 9.58 and Figure 9.32, it is observed that if All FTSegs are concerned, the percentages of the samples using LIDAR data are higher than the percentages of the samples using NED data. The percentage of the sample using LIDAR 50-ft DEM with an interval of 25-ft is the highest. If Interstate FTSegs are concerned, the percentages of the samples using LIDAR data are higher than the percentages of the samples using NED

data. The percentage of the sample using LIDAR 50-ft DEM with an interval of 25-ft is the highest. If US FTSegs are concerned, the percentages of the samples using LIDAR data are still higher than the percentages of the samples using NED data. The percentage of the sample using LIDAR 50-ft DEM with an interval of 50-ft is the highest. If NC FTSegs are concerned, the percentage of the sample using LIDAR point data is the lowest. In addition, the trend line of Interstate FTSegs is the highest (with higher percentages) and the trend line of US FTSegs is the lowest (with lower percentages).

Table 9.59 summarizes the percentages for differences with the error range of [-20, 20] for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.33 illustrated comparisons of these percentages.

Table 9.59 Summary of Percentages for Differences ([-20, 20])

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	56.98%	55.09%	55.47%	58.49%	56.23%	50.94%	50.57%
Interstate FTSegs	62.03%	57.59%	58.23%	63.92%	60.13%	53.16%	52.53%
US FTSegs	49.30%	46.48%	46.48%	46.48%	46.48%	43.66%	43.66%
NC FTSegs	50.00%	61.11%	61.11%	58.33%	58.33%	55.56%	55.56%

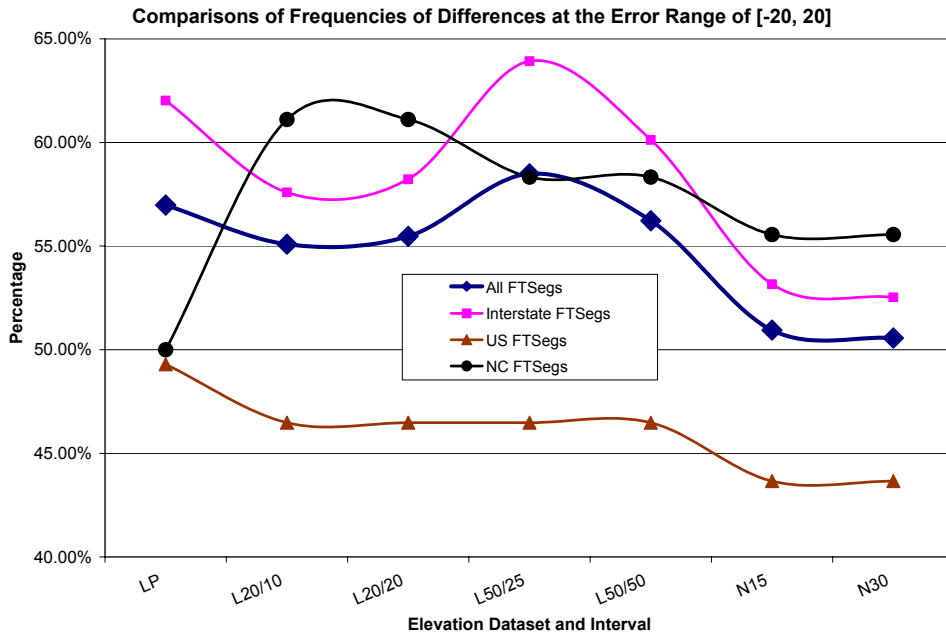


Figure 9.33 Illustration of Percentages at the Error Range of [-20, 20] for Differences

Based on Table 9.59 and Figure 9.33, it is observed that when All FTSEgs are concerned, the percentages of the samples using LIDAR data are significantly higher than the percentages of the samples using NED data. The percentages of the sample using LIDAR point data and LIDAR 50-ft DEM with an interval of 25-ft are the highest. When Interstate FTSEgs are concerned, same observations could be derived. When US FTSEgs are concerned, the percentages of the samples using LIDAR data are still higher than the percentages of the samples using NED data. The percentage from the sample using LIDAR point data is the highest. When NC FTSEgs are concerned, the percentages from the samples using LIDAR 20-ft DEMs are the highest while the percentage from the sample using LIDAR point data is the lowest.

Table 9.60 summarizes the percentages for differences with the error range of $[-30, 30]$ for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.34 illustrates comparisons of these percentages.

Table 9.60 Summary of Percentages for Differences ($[-30, 30]$)

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	77.36%	69.43%	69.06%	69.06%	71.70%	66.42%	66.42%
Interstate FTSEgs	77.22%	68.99%	67.72%	67.09%	72.15%	64.56%	64.56%
US FTSEgs	73.24%	61.97%	61.97%	64.79%	63.38%	61.97%	61.97%
NC FTSEgs	86.11%	86.11%	88.89%	86.11%	86.11%	83.33%	83.33%

Based on Table 9.60 and Figure 9.34, it is observed that for All FTSEgs, the percentages from the samples using LIDAR data are higher than the percentages from samples using NED data. The percentage of the sample using LIDAR point data is the highest. For Interstate FTSEgs, same observations could be derived. For US FTSEgs, the percentage from the sample using LIDAR point data is still the highest while the percentages from other samples are close to each other. For NC FTSEgs, the percentages from samples using LIDAR data are higher than the percentages from samples using NED data. However, the highest percentage corresponds to the sample using LIDAR 20-ft DEM with a 20-ft interval. In addition, the trend line for US FTSEgs is the lowest line with the lowest percentages and therefore, the lowest accuracy while the trend line for NC FTSEgs is the highest line with the highest percentage and therefore, the highest accuracy at this error range.

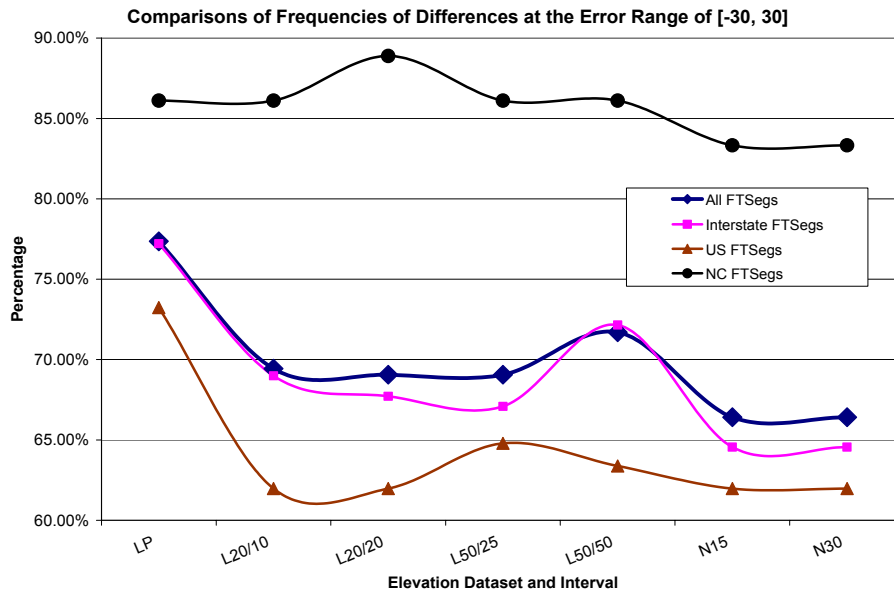


Figure 9.34 Illustration of Percentages at the Error Range of [-30, 30] for Differences

Table 9.61 summarizes the percentages for differences within the error range of [-50, 50] for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.35 illustrates comparisons of these percentages.

Table 9.61 Summary of Percentages for Differences ([-50, 50])

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	93.96%	85.66%	84.91%	85.66%	86.79%	83.02%	82.26%
Interstate FTSegs	94.94%	84.81%	83.54%	84.18%	86.71%	82.28%	81.01%
US FTSegs	91.55%	81.69%	81.69%	83.10%	81.69%	78.87%	78.87%
NC FTSegs	94.44%	97.22%	97.22%	97.22%	97.22%	94.44%	94.44%

From Figure 9.35, it is obvious that the percentages from samples using LIDAR data are higher than the percentages from samples using NED data. The percentages from the sample using LIDAR point data are the highest except in the case of NC FTSegs, in which, the percentages from samples using LIDAR DEMs are close to each other but higher than the percentage from the sample using LIDAR point data. In addition, the trend line for US FTSegs is still the lowest line with the lowest percentages and therefore, the lowest accuracy while the trend line for NC FTSegs is still the highest line with the highest percentage and therefore, the highest accuracy.

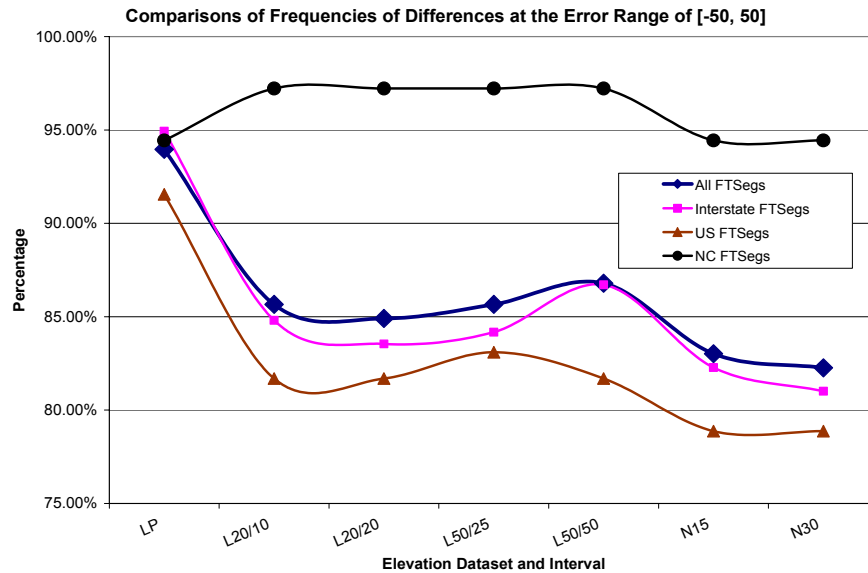


Figure 9.35 Illustration of Percentages at the Error Range of [-50, 50] for Differences

Table 9.62 summarizes the percentages for differences with the error range of $(-\infty, -50)$ and $(50, +\infty)$ for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.36 illustrates comparisons of these percentages.

Table 9.62 Summary of Percentages for Differences $(-\infty, -50)$ and $(50, +\infty)$

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	6.04%	14.34%	15.09%	14.34%	13.21%	16.98%	17.74%
Interstate FTSEgs	5.06%	15.19%	16.46%	15.82%	13.29%	17.72%	18.99%
US FTSEgs	8.45%	18.31%	18.31%	16.90%	18.31%	21.13%	21.13%
NC FTSEgs	5.56%	2.78%	2.78%	2.78%	2.78%	5.56%	5.56%

The percentages in Table 9.62 are obtained by subtracting 100% with those percentages from Table 9.61. A higher percentage in Table 9.62 indicates a higher percentage of errors falling into the error range of $(-\infty, -50)$ and $(50, +\infty)$, or big errors. More specifically, a higher percentage in Table 9.62 means a lower accuracy.

Based on Table 9.62 and Figure 9.36, it is obvious that samples using LIDAR data have better accuracies than samples using NED data. The samples using LIDAR point data have the best accuracies in the cases of All

FTSegs, Interstate FTSegs, and US FTSegs. For NC FTSegs, the samples using LIDAR DEMs have the best accuracy.

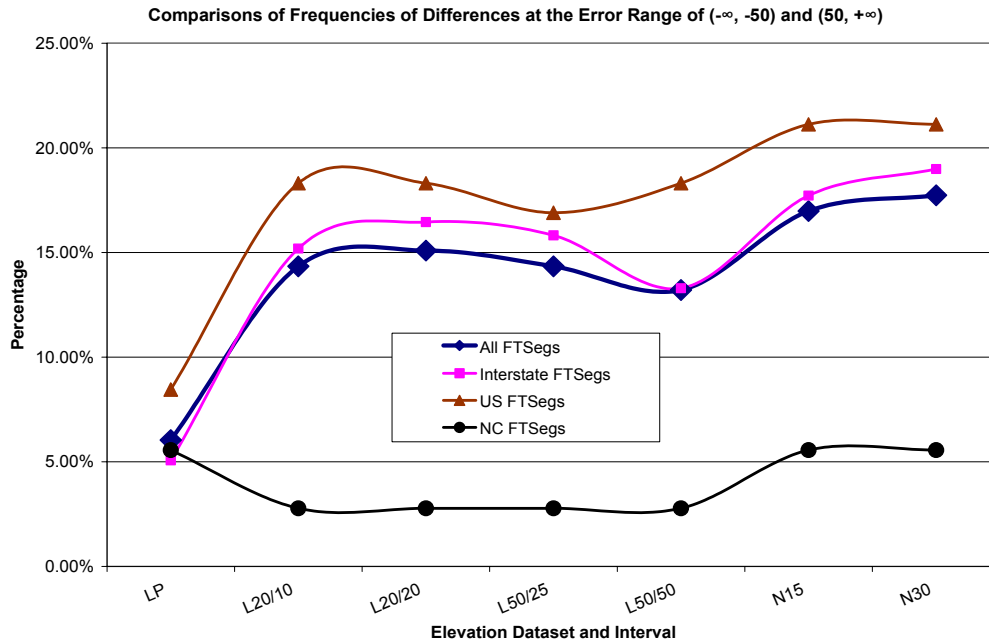


Figure 9.36 Illustration of Percentages at the Error Range of $(-\infty, -50)$ and $(50, +\infty)$ for Differences

9.4.5.2 Comparison of Frequencies for Proportional Differences

This section describes the comparisons based on percentages for proportional differences. The error ranges are defined using values with the unit of foot/1,000 feet.

Table 9.63 summarizes the percentages for proportional differences with the error range of $[-1, 1]$ for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.37 illustrates comparisons of these percentages.

Table 9.63 Summary of Percentages for Proportional Differences $[-1, 1]$

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	24.15%	20.00%	20.38%	22.64%	24.15%	20.00%	20.00%
Interstate FTSegs	31.01%	24.05%	24.05%	28.48%	30.38%	24.05%	24.05%
US FTSegs	5.63%	4.23%	4.23%	5.63%	5.63%	2.82%	2.82%
NC FTSegs	30.56%	33.33%	36.11%	30.56%	33.33%	36.11%	36.11%

Based on Table 9.63 and Figure 9.37, it is observed that for All FTSEgs, the percentage from the sample using LIDAR point data is the highest. The percentages from samples using LIDAR data are higher than those percentages from samples using NED data. These observations are also true in the cases of Interstate FTSEgs and US FTSEgs. With NC FTSEgs, the percentages from the samples using NED are the highest. In addition, the trend line for US FTSEgs is the lowest line with the lowest percentages while the trend line for NC FTSEgs is the highest line with the highest percentages.

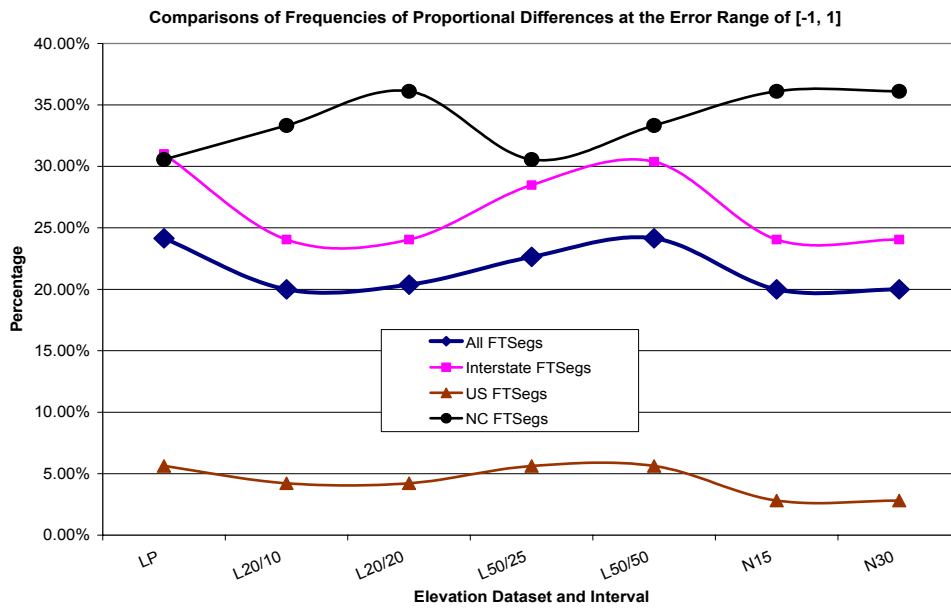


Figure 9.37 Illustration of Percentages at the Error Range of [-1, 1] for Proportional Differences

Table 9.64 summarizes the percentages for proportional differences within the error range of [-5, 5] for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.38 illustrates comparisons of these percentages.

Table 9.64 Summary of Percentages for Proportional Differences ([-5, 5])

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	57.74%	57.74%	57.74%	55.85%	57.36%	56.98%	56.98%
Interstate FTSEgs	67.72%	67.09%	67.09%	64.56%	67.09%	66.46%	66.46%
US FTSEgs	35.21%	36.62%	36.62%	35.21%	35.21%	35.21%	35.21%
NC FTSEgs	58.33%	58.33%	58.33%	58.33%	58.33%	58.33%	58.33%

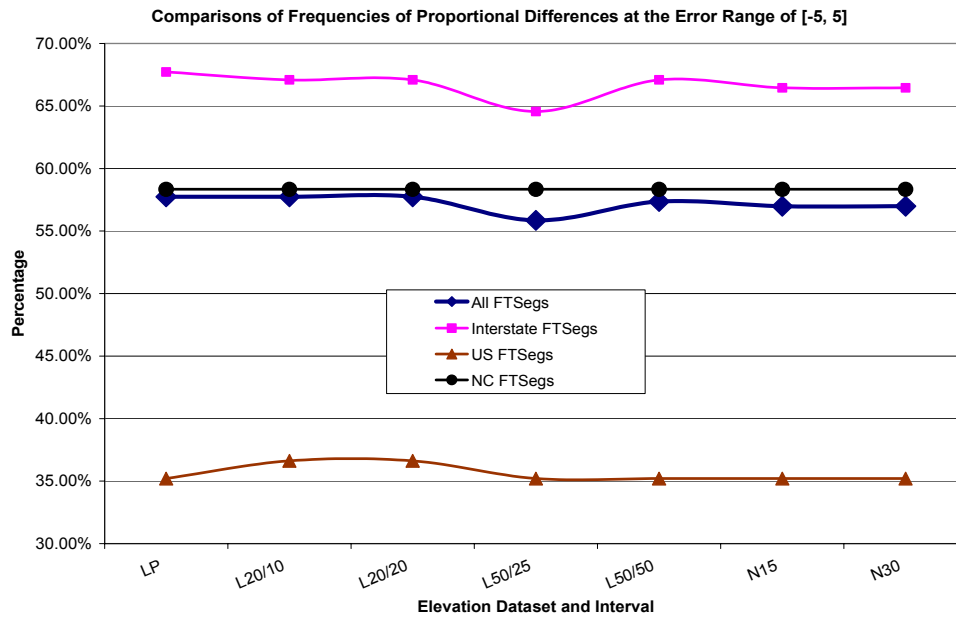


Figure 9.38 Illustration of Percentages at the Error Range of [-5, 5] for Proportional Differences

Figure 9.38 illustrates that the trend lines for Interstate FTSEgs, US FTSEgs, and NC FTSEgs are almost parallel to each other. The trend line for US FTSEgs is the lowest line with the lowest percentage. The trend line for NC FTSEgs is the highest line with the highest percentages. The variation of percentages at this error range for a particular road type is quite small.

Table 9.65 summarizes the percentages for proportional differences within the error range of [-10, 10] for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.39 illustrates comparisons of these percentages.

Table 9.65 Summary of Percentages for Proportional Differences ([-10, 10])

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	66.79%	65.66%	65.66%	65.66%	65.66%	64.91%	64.91%
Interstate FTSEgs	74.68%	72.78%	72.78%	72.78%	73.42%	72.78%	72.78%
US FTSEgs	49.30%	47.89%	47.89%	47.89%	46.48%	46.48%	46.48%
NC FTSEgs	66.67%	69.44%	69.44%	69.44%	69.44%	66.67%	66.67%

Based on Table 9.65 and Figure 9.39, it is observed at the error range of [-10, 10] for proportional differences, the percentages from samples using LIDAR data are higher than those percentages from samples using NED

data. The percentages from samples using LIDAR point data are highest in the cases of All FTSegs, Interstate FTSegs, and US FTSegs. In the case of NC FTSegs, the percentages from samples using LIDAR DEMs are close to each other but higher than the others. Still, the trend line for US FTSegs is the lowest line with the lowest percentages while the trend line for NC FTSegs is the highest line with the highest percentages.

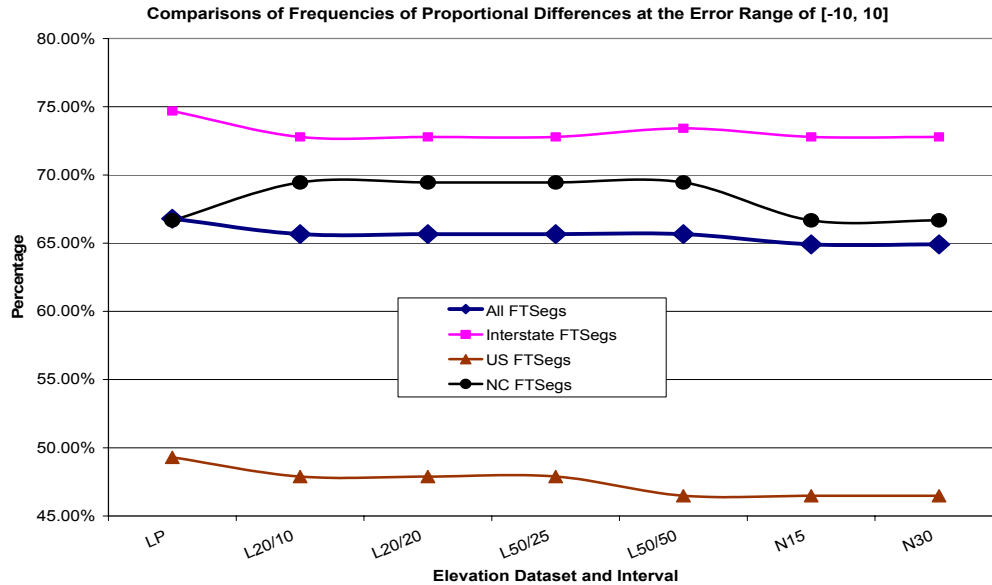


Figure 9.39 Illustration of Percentages at the Error Range of [-10, 10] for Proportional Differences

Table 9.66 summarized the percentages for proportional differences with the error range of [-20, 20] for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.40 illustrated comparisons of these percentages.

Table 9.66 Summary of Percentages for Proportional Differences ([-20, 20])

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	75.85%	75.09%	75.09%	75.47%	75.85%	75.09%	75.09%
Interstate FTSegs	81.65%	79.75%	79.75%	79.75%	80.38%	79.75%	79.75%
US FTSegs	64.79%	66.20%	66.20%	67.61%	67.61%	66.20%	66.20%
NC FTSegs	72.22%	72.22%	72.22%	72.22%	72.22%	72.22%	72.22%

Based on Table 9.66 and Figure 9.40, it is observed that in the cases of All FTSegs and Interstate FTSegs, the percentages from samples using different elevation datasets and intervals are quite close to each other with the percentages of the samples using LIDAR point data as the highest. In addition, the percentages from samples

using LIDAR data are higher than the percentages from samples using NED data. In the case of US FTSEgs, the highest percentage corresponds to the samples using LIDAR 50-ft DEM. In the case of NC FTSEgs, all the percentages are the same. Furthermore, it was observed that the trend line for US FTSEgs is the lowest line with the lowest percentages while the trend line for NC FTSEgs is the highest line with the highest percentages.

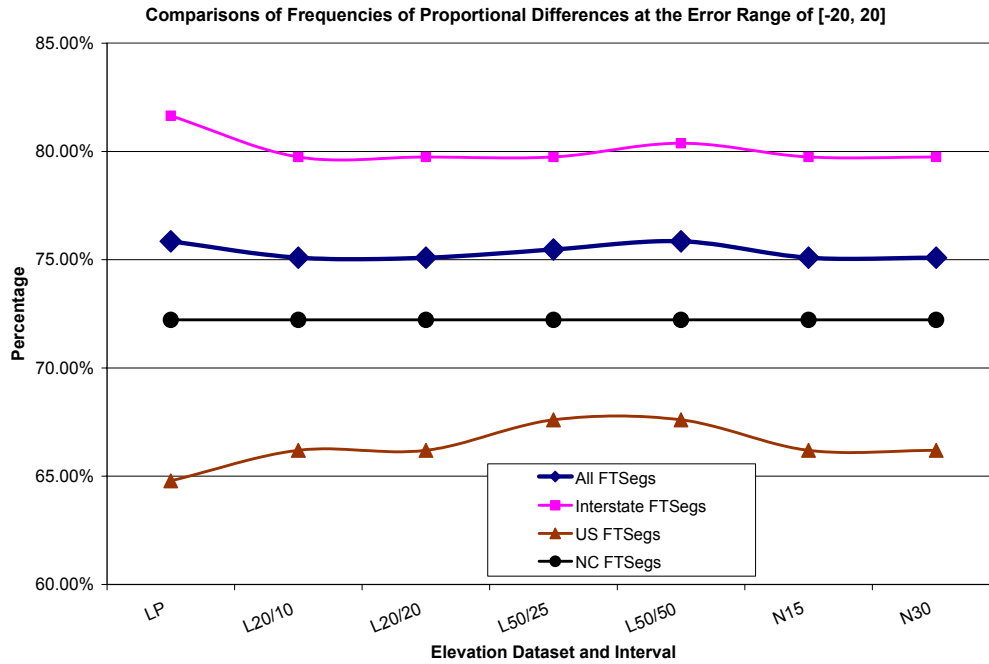


Figure 9.40 Illustration of Percentages at the Error Range of [-20, 20] for Proportional Differences

Table 9.67 summarizes the percentages for proportional differences with the error range of [-30, 30] for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.41 illustrates comparisons of these percentages.

Table 9.67 Summary of Percentages for Proportional Differences ([-30, 30])

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	79.62%	78.87%	78.87%	79.62%	79.62%	79.62%	79.62%
Interstate FTSEgs	84.18%	82.91%	82.91%	83.54%	83.54%	83.54%	83.54%
US FTSEgs	71.83%	71.83%	71.83%	73.24%	73.24%	73.24%	73.24%
NC FTSEgs	75.00%	75.00%	75.00%	75.00%	75.00%	75.00%	75.00%

Based on Table 9.67 and Figure 9.41, it is observed that at this error range of proportional differences, the percentages from samples of a particular road type are close to each other. In the case of NC FTSEgs, the percentages are the same. In the cases of Interstate FTSEgs and All FTSEgs, the percentages from samples using LIDAR point data are the highest while the percentages from samples using LIDAR 20-ft DEMs are the lowest. Still, the trend line for US FTSEgs is the lowest line with the lowest percentages while the trend line for NC FTSEgs is the highest line with the highest percentages.

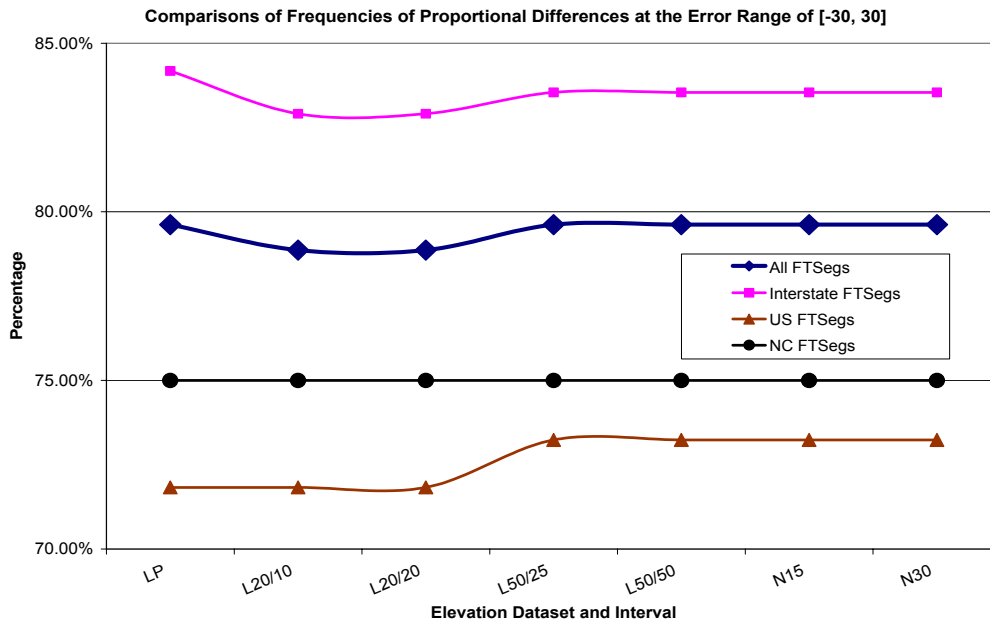


Figure 9.41 Illustration of Percentages at the Error Range of [-30, 30] for Proportional Differences

Table 9.68 summarizes the percentages for proportional differences with the error range of [-50, 50] for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.42 illustrates comparisons of these percentages.

Table 9.68 Summary of Percentages for Proportional Differences ([-50, 50])

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	86.04%	84.91%	85.28%	85.66%	85.66%	85.66%	85.66%
Interstate FTSEgs	88.61%	87.34%	87.97%	88.61%	88.61%	88.61%	88.61%
US FTSEgs	84.51%	83.10%	83.10%	83.10%	83.10%	83.10%	83.10%
NC FTSEgs	77.78%	77.78%	77.78%	77.78%	77.78%	77.78%	77.78%

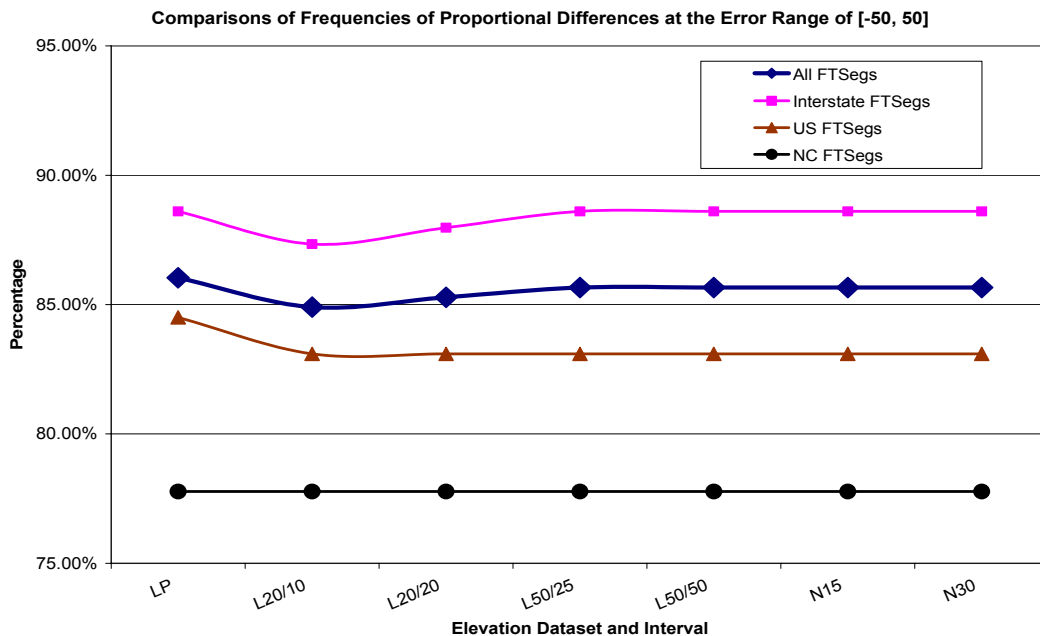


Figure 9.42 Illustration of Percentages at the Error Range of [-50, 50] for Proportional Differences

Based on Table 9.68 and Figure 9.42, it is observed that at this error range, the variations of percentages from samples of a particular road type with varying elevation datasets and intervals are very little. In other words, it would be reasonable to say that at this error range, the effects on the accuracies of predicted 3-D distances due to the use of different elevation datasets and intervals are the same. Examining the subtle variations, it is found that for NC FTSegs, all the percentages are the same. For All FTSegs, Interstate FTSegs, and US FTSegs, the percentages of the samples using LIDAR point data are among the highest. The trend line for US FTSegs becomes the lowest line with the lowest percentages while the trend line for Interstate FTSegs becomes the highest line with the highest percentages.

Table 9.69 summarizes the percentages for proportional differences with the error range of [-100, 100] for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. Figure 9.43 illustrates comparisons of these percentages.

Table 9.69 Summary of Percentages for Proportional Differences ([-100, 100])

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSEgs	94.34%	93.96%	93.96%	94.34%	94.34%	94.34%	94.34%
Interstate FTSEgs	96.84%	96.84%	96.84%	96.84%	96.84%	96.84%	96.84%
US FTSEgs	91.55%	90.14%	90.14%	91.55%	91.55%	91.55%	91.55%
NC FTSEgs	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%

Based on Table 9.69 and Figure 9.43, it is obvious that at this error range, the variations of percentages from samples of a particular road type are very little (the percentages were almost the same). One exception is in the case of US FTSEgs, in which the percentages corresponding to samples using LIDAR 20-ft DEMs are lower than the others. Furthermore, the trend line for NC FTSEgs is still the lowest line with the lowest percentages while the trend line for Interstate FTSEgs is the highest line with the highest percentages.

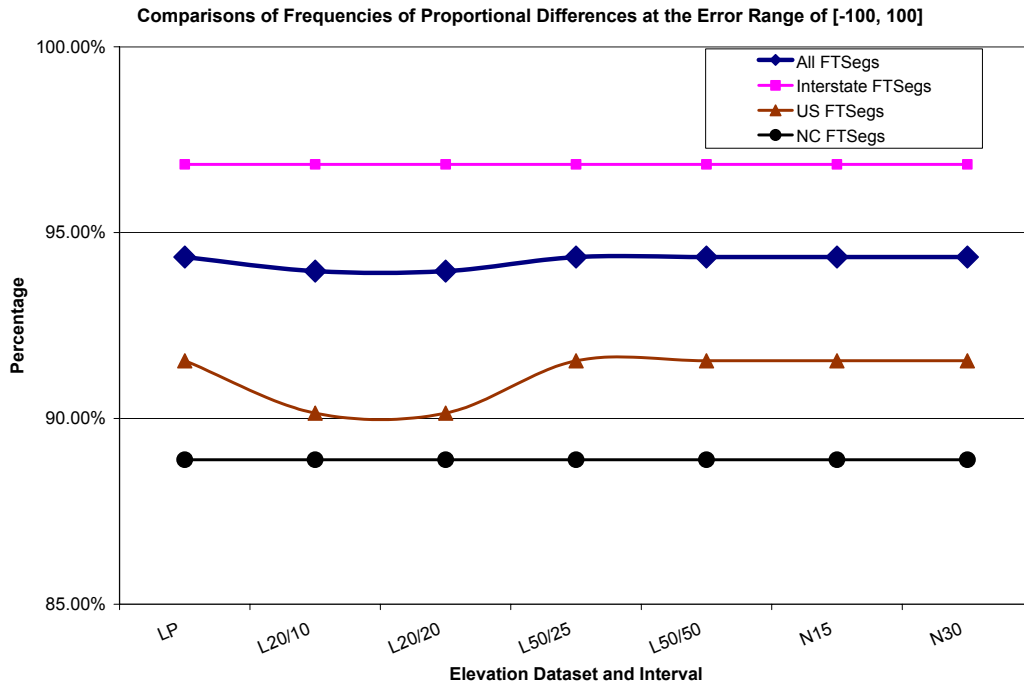


Figure 9.43 Illustration of Percentages at the Error Range of [-100, 100] for Proportional Differences

Table 9.70 summarizes the percentages for proportional differences with the error range of $(-\infty, -100)$ and $(100, +\infty)$ for all the predicted 3-D distances compared with DMI measured distances, grouped by the use of elevation datasets and intervals. The relationship between the percentages in Table 9.70 and the percentages in Table 9.69 was the same as the relationship between the percentages in Table 9.62 to the percentages in Table 9.61.

Consequently, no figure is obtained for comparisons of these percentages and no further observations are provided herein.

Table 9.70 Summary of Percentages for Differences((-∞, -100) and (100, +∞))

Road Type	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
All FTSegs	5.66%	6.04%	6.04%	5.66%	5.66%	5.66%	5.66%
Interstate FTSegs	3.16%	3.16%	3.16%	3.16%	3.16%	3.16%	3.16%
US FTSegs	8.45%	9.86%	9.86%	8.45%	8.45%	8.45%	8.45%
NC FTSegs	11.11%	11.11%	11.11%	11.11%	11.11%	11.11%	11.11%

9.5 ERROR ANALYSIS AND ERROR PROPAGATION

There are two major error sources in the predicted 3-D distances for linear objects using the 3-D modeling approach developed in this research: errors from the input elevation data and planimetric road centerline data (static errors) and errors introduced by the 3-D model (operational errors).

As stated earlier, in this research, a particular linear object was modeled in a three-dimensional space with a set of 3-D points (with X, Y, and Z-coordinates) connected by straight lines. These 3-D points obtain X and Y-coordinates (planimetric position) from the planimetric line data and consequently, the horizontal positional accuracy of these 3-D points is dependent on the positional accuracy of the planimetric line data. These 3-D points obtain their Z coordinate from the input elevation datasets and therefore, the vertical positional accuracy of these 3-D points is dependent on the vertical accuracy of the input elevation dataset and the way the elevations are obtained (in the case of using LIDAR DEMs or NED, the bilinear interpolation method; in the case of LIDAR point data, the buffering and snapping method). To summarize, the errors from the input elevation data and the planimetric line data are represented as the positional errors of the 3-D points being used to model the linear objects in a three-dimensional space. These positional errors of the 3-D points affect the accuracy of the predicted 3-D distances using the 3-D models in this research.

The positional error for 3-D points could be modeled as cylinders as shown in Figure 9.44. The cylinder is the 3-dimensional range in which a point with specified horizontal and vertical accuracies could appear. Assuming the black circles are the true positions of 3-D points, cylinders could be created using these true positions as the

center. Half of the cylinder height corresponds to the vertical error and the radius of the cylinder corresponds to the horizontal error. For a given 3-D point, its position could be anywhere inside the cylinder. Figure 9.44 also illustrates that based on these assumptions, two lines could be constructed by connecting two sets of 3-D points using straight line segments, of which one is the longest and one is the shortest. The length of the third line that connects the set of 3-D points with true positions would fall in this range.

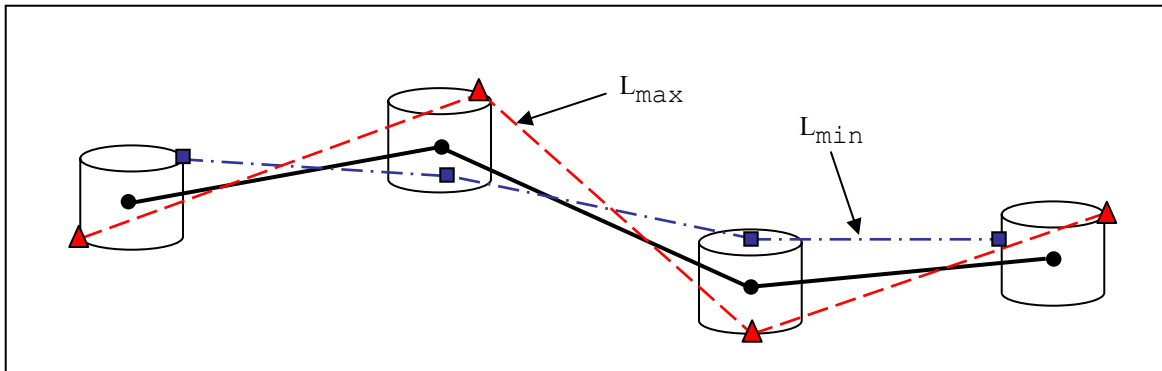


Figure 9.44 Illustration of the Effects from the Point Positional Errors on the Predicted 3-D Distances

On the other hand, by modeling a 3-D line with a set of 3-D points connected by straight lines, errors are also introduced into the predicted 3-D distance using this 3-D model approach as illustrated in Figure 9.45. Figure 9.45 shows the vertical profile of a linear object being modeled using two sets of 3-D points with different intervals. The true line is longer than any of these two series of straight lines. Generally, more points (and consequently more straight line segments) indicate a closer approximation of the 3-D line constructed using these 3-D points to approximate the true 3-D line.

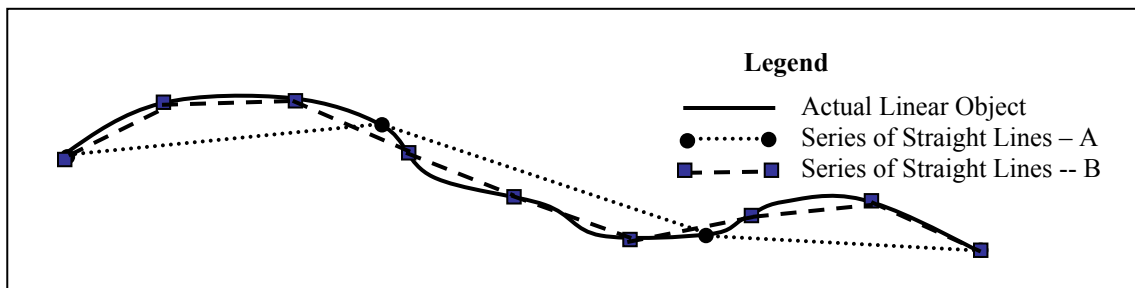


Figure 9.45 Illustration of the Effects of the 3-D model on the Predicted 3-D Distances

The 3-D model in this research, which uses a set of connected straight line segments to approximate the true line, would have a negative effect on the predicted 3-D distance using this 3-D model as illustrated in Figure 9.45. In other words, by using this 3-D model, the predicted 3-D distance for a linear object would be shorter than its true length. On the other hand, the effects on the predicted 3-D distances from the point positional errors could be positive or negative. The positive effects would compensate for the negative effects from the 3-D model while the negative effects would make the negative effects more significant.

Figure 9.46 illustrates the effects from the point positional errors and the effects from the 3-D model by using a series of connected straight line segments to approximate the true 3-D line. In Figure 9.46, there are three lines. One is the true line. Another line is the line consisting of straight line segments constructed using the 3-D model approach, assuming there is no effect from the point positional errors. The third line is the line constructed using the 3-D model, assuming there are point positional errors. It is obvious that the predicted 3-D distance (the length of the line constructed using the 3-D model while assuming there were point positional errors) would be different from the true 3-D distance.

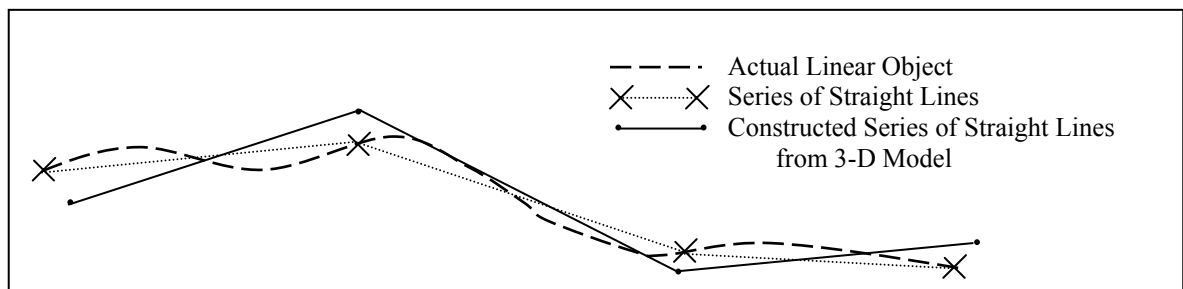


Figure 9.46 Illustration of the Difference between the True 3-D Line and the Constructed 3-D Line

The effects on the accuracy of the predicted 3-D distances from the vertical point positional errors have been evaluated in accuracy assessment and sensitivity analysis, which focused on evaluating the effects from the use of different elevation datasets. As stated earlier, the vertical accuracy of the 3-D points being used in modeling linear objects in a three-dimensional space was dependent on the vertical accuracy of the elevation dataset being used and different elevation datasets had different vertical accuracies, evaluating the effects on the accuracy of the predicted 3-D distances from the use of different elevation datasets is in a sense the same as evaluating the

effects on the accuracy of the predicted 3-D distances from the vertical point positional accuracy. However, the effects on the predicted 3-D distances from the horizontal point positional errors are not evaluated in this research because the horizontal accuracy of the 3-D points is dependent on the horizontal accuracy of the planimetric line data and in this case study, the same line data are used in modeling linear objects in a three-dimensional space and predicting their 3-D distances using the 3-D approach.

The effects on the accuracy of the predicted 3-D distances from the 3-D model in this research, i.e. the use of a set of connected straight line segments to approximate the true 3-D line would be evaluated in the next chapter, Identification of Significant Factors. As stated earlier, with the 3-D model in this research, if more 3-D points are used to construct a 3-D line to approximate the true 3-D line, the constructed 3-D line would be closer to the true 3-D line and consequently, the distance of the constructed 3-D line would be closer to the true 3-D distance. In the next chapter, the effects on the accuracy of the predicted 3-D distances from the number of 3-D points being used would be determined.

9.6 CONCLUSIONS OF ACCURACY ASSESSMENT AND SENSITIVITY ANALYSIS

In this chapter, the accuracy of the predicted 3-D distances using different elevation datasets and intervals was determined by comparing the predicted 3-D distances to the reference data, DMI measured distances. The errors of the predicted 3-D distances are represented as differences between the predicted 3-D distances and the DMI measured distances and proportional differences that are obtained by dividing the differences by the corresponding DMI measured distances. Statistics such as means, medians, percentiles, standard deviations, skewness, absolute means, and RMSEs are used to describe these errors. In addition, the errors due to the use of different elevation datasets and intervals are compared in numerous ways to evaluate the effects (the sensitivity) on the accuracy of the predicted 3-D distances from the use of different elevation datasets with varying accuracies and resolutions. Furthermore, it is also described that how the errors of the input data propagate into the final product (the predicted 3-D distances) through operations on the input data, and how the assumptions of the 3-D model in this research affect the accuracy of the predicted 3-D distances.

This section provides conclusions about errors and resulting accuracies in the predicted 3-D distances when compared with the DMI measured distances, based on the comprehensive analyses presented in this chapter and on the observations obtained from these analyses.

It is concluded that

- (1) Errors of the predicted 3-D distances, when compared to DMI measured distances, either in the format of differences or proportional differences, are not normally distributed. This conclusion is based on the descriptive statistics, especially the skew, and the distribution histograms.

- (2) The accuracy of the predicted 3-D distances, when compared with DMI measured distances, varies with the use of different elevation datasets. In this case study, four elevation dataset are used, the LIDAR point data, the LIDAR 20-ft DEMs, the LIDAR 50-ft DEMs, and the NED data. The accuracy of the predicted 3-D distances is dependent on the elevation dataset being used.

- (3) The higher the accuracy of the elevation dataset being used, the higher the accuracy of the predicted 3-D distances, when compared with DMI measured distances. In this case study, LIDAR data have higher vertical accuracy than the NED data. The accuracy assessment and sensitivity analysis presented earlier in this chapter demonstrates that the accuracy of the 3-D distances predicted using LDIAR data is higher than the accuracy of the predicted 3-D distances using NED data.

From the aspect of differences using the 100% RMSE as the measure of the accuracy, the use of LIDAR point data in this case study improves the accuracy by about 28% compared to the use of NED data. The use of LIDAR DEMs improves the accuracy by about 6% compared to the use of NED data.

From the aspect of differences using the 95% RMSE as the measure of the accuracy, the use of LIDAR point data in this case study improves the accuracy by about 25% compared to the use of NED data. The use of LIDAR DEMs improves the accuracy by about 9% compared to the use of NED data.

From the aspect of proportional differences, the improvements due to the use of higher accurate elevation datasets are not as significant as those shown for differences above. This is the case because the majority (53%) of the FTSEgs in this case study are longer than 5,000 ft, about 73% are longer than 1,000 ft, and 43% are longer than 10,000 ft.

(4) It is concluded that the 3-D distances predicted using LIDAR point data with the snapping approach in this research have the best accuracy when compared to DMI measured distances.

(5) While the accuracy of the predicted 3-D distances is dependent on the elevation dataset being used, if the same elevation dataset is used, but two different intervals are taken, one equivalent to half of the cell size and the other equivalent to the full cell size of the elevation data, the accuracies of the predicted 3-D distances are almost the same. In other words, the accuracy of the predicted 3-D distances is not dependent on the interval sizes if they are less than or equal to the cell size of the elevation data.

(6) The accuracy of the predicted 3-D distances is also dependent on the road types if road types are taken into consideration. Given the use of a particular elevation dataset, from the aspect of differences, NC FTSEgs have the best accuracy, followed by Interstate FTSEgs. US FTSEgs have the worst accuracy. From the aspect of proportional differences, Interstate FTSEgs have the best accuracy while NC FTSEgs have the worst accuracy. The accuracy of US FTSEgs is close to but better than the accuracy of the NC FTSEgs.

(7) Examining the predicted 3-D distances in detail, together with the statistics used in the accuracy assessment leads to an impression that overall, the predicted 3-D distances using LIDAR data (LIDAR point data and LIDAR DEMs) are close to each other, but significantly different from using NED data. A further investigation has been conducted by examining the vertical differences of the 3-D points. This investigation randomly picks 2 Interstate FTSEgs from Johnston County. These 2 FTSEgs have 2 sets of 3-D points obtained using LIDAR point data and the snapping approach mentioned earlier. For each of these 3-D points, three more elevation values are obtained, one from LIDAR 20-ft DEM, one from LIDAR 50- ft DEM, and the other from NED data. It is found that the vertical differences of these 3-D points are 0.35 ft (RMSE) between

the use of LIDAR point data and LIDAR 20-ft DEMs, 3.5 ft (RMSE) between the use of LIDAR point data and the use of LIDAR 50-ft DEM, but 11.7 ft between the use of LIDAR point data and the use of NED data. This confirms that the accuracies of the predicted 3-D distances using LIDAR data are close to each other but much better than those accuracies of the predicted 3-D distances using NED data.

(8) Another further investigation was conducted to determine the significance of the errors in the road length information if 2-D lengths were to be used as road distances. This investigation calculated the difference between a 2-D length and the DMI measured distance and the proportional difference based on the DMI measured distance. 100% RMSEs were obtained and shown in Table 9.71. For differences, the unit is foot. For proportional differences, the unit is foot/1,000 feet.

Examining these RMSEs leads to the appreciation of the inappropriateness of using 2-D length to approximate the actual road distances. Comparing these RMSEs with those corresponding 100% RMSEs from those predicted 3-D distances using different elevation datasets and intervals reveals that from the aspect of difference, the improvements on the accuracy of distances due to the use of predicted 3-D distances range from 13.13% (with the use of NED data) to 37.65% (with the use of LIDAR point data); from the aspect of proportional difference, the improvements are not significant because most FTSEgs are relatively long FTSEgs.

Table 9.71 RMSEs from Comparing 2-D Lengths with DMI Measured Distances

Road Type	RMSEs	
	Difference (foot)	Proportional Difference (foot/1,000 feet)
All FTSEgs	41.14	50.66
Interstate FTSEgs	44.03	32.26
US FTSEgs	41.05	64.15
NC FTSEgs	25.05	78.80

10 IDENTIFICATION OF SIGNIFICANT FACTORS

This chapter describes the factors that might be closely related to the accuracy of the predicted 3-D distances. In this chapter, the factors under consideration are proposed and defined and their relationships with the accuracy of the predicted 3-D distances are determined.

10.1 INTRODUCTION

The accuracy of the predicted 3-D distances was evaluated in chapter 9 by comparing with the DMI measured distances. It was concluded that the accuracy varied with different elevation datasets being used. It was also concluded that different road types had different accuracies of the 3-D distances predicted using the 3-D approach.

It is also concerned that the accuracy of the predicted 3-D distances might also be related to certain properties of the linear objects such as the slope, slope changes, and distances. In addition, it is also concerned that the number of 3-D points being used to represent a 3-D line might also be related to the accuracy of the predicted 3-D distances.

The major purpose of identifying significant factors is to identify factors that have certain relationships with the accuracy of the predicted 3-D distances. More specifically, this chapter focuses on determining if there is a linear association between a factor and the accuracy of the predicted 3-D distances. By doing so, cautions could be applied to those linear objects that occupy certain properties and corresponding quality control measures could be taken to further improve the accuracy of the predicted 3-D distances using the 3-D approach.

10.2 FACTORS UNDER CONSIDERATION

In this research, factors under consideration are determined based on the characteristics of the 3-D model being used to represent linear objects in a 3-D space and to predict their 3-D distances. As stated earlier, in this research, a 3-D line was modeled as a group of 3-D points connected with straight line segments. Generally, a longer 3-D line would have more 3-D points and straight line segments connecting these 3-D points and

consequently, the more significant of the approximation. Also, if a linear object is more complex from the aspect of the vertical profile, using connected straight lines would have a more significant approximation.

Factors under consideration in this research include the distance, the slope and slope change, the number of points being used, and the average planimetric distance along a linear object of the 3-D points being used to model the 3-D line. In this research, the distance is defined as the DMI measured distance. The number of 3-D points is defined as the number of 3-D points being used to model a 3-D line using LIDAR point data and consequently, this factor would be evaluated for the accuracy of the predicted 3-D distances only in the case of using LIDAR point data. Generally, the more 3-D points would be obtained for a longer linear object using the snapping approach. To surrogate this situation, the average planimetric distance along a linear object of the 3-D points being used to model the 3-D line is also evaluated. This average distance is a measure of the density of the 3-D points along a linear object. It is worth to be pointed out here that in the cases of using LDIAR DEMs and NED, evaluating the factor of the average planimetric distance of the 3-D points would be meaningless because in these cases, 3-D points are planimetrically uniformly distributed along the linear objects and consequently, there would be almost no difference of the average planimetric distance of the 3-D points when a particular DEM or NED dataset with a particular interval is used. Also, evaluating the factor of the number of 3-D points would be the same as evaluating the factor of the distance.

The vertical complexity of a linear object could be described using the slope and the slope change. In this research, most linear objects are relatively long and therefore, it is unreasonable to simply take the start point and the end point of a linear object and calculate the slope using these two points. Rather, in this research, two slopes are obtained for each linear object based on the 3-D points obtained using LIDAR point data and the snapping approach, the average slope and the weighted average slope. The average slope is calculated as the average of slopes of all the straight lines connecting the 3-D points for a linear object. The weighted average slope is calculated by using the 3-D distance of those straight lines as the weight and taking a weighted average as illustrated in Figure 10.1.

In Figure 10.1, there are three 3-D points including the start and the end point. The average slope would be calculated as $(S1 + S2)/2$ while the weighted average would be calculated as $(S1 * D1 + S2 * D2)/(D1 + D2)$, where $S1 = E1/PD1$ and $S2 = E2/PD2$. The researchers also considered using the planimetric distance as the weight. In this case, the weighted average slope would be calculated as $(S1 * PD1 + S2 * PD2)/(PD1 + PD2)$, where $S1 = E1/PD1$ and $S2 = E2/PD2$, which would be equivalent to $(E1 + E2)/(PD1 + PD2)$. In other words, the result would be the same as using the start and end points only to calculate the slope. Consequently, this research uses the 3-D distance of the straight line as the weight, not the planimetric distance of the straight line.

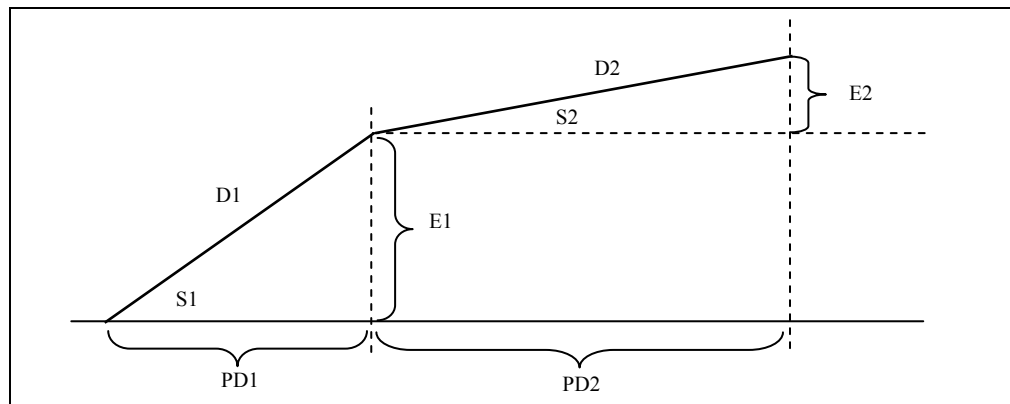


Figure 10.1 Illustration of Slope and Slope Change Calculation

Another measure of the vertical complexity of a 3-D line is the slope change, which is defined as the change of the slope of the current straight line from the slope of the preceding straight line. For example, the slope change for the second straight line in Figure 10.1 would be $(S2 - S1)$. Since there is no preceding straight line for the very first straight line with the start point of a 3-D line as the start point, it is assumed that the slope change for this particular straight line would be $(S1 - 0)$, i.e. from slope of 0 to slope $S1$. The absolute values of these slope changes are averaged to obtain the average slope change in this research. Similarly, these slope changes are also weighted by the corresponding 3-D distances of the straight lines and a weighted average slope is obtained.

Figure 10.2 illustrates the algorithm used to calculate average slope, weighted average slope, average slope change, and weighted average slope change in this research. The developed program code is attached as Appendix D.

The algorithm is summarized as below.

- 1) All 3-D points belonging to one FTSeg are retrieved from the 3-D point data.
- 2) These 3-D points are sorted based on their distances to the start point of that FTSeg along that FTSeg, ascending, resulting to a group of sorted 3-D points for that particular FTSeg.
- 3) Set the variables of SUM and DIST to 0. The variable SUM would be used to keep the total of the 3-D distances being calculated while the variable DIST would be used to keep the 3-D distance for the current road section defined as the road centerline between two neighboring 3-D points.
- 4) Set the variables of SUM_Slope, Slope1, Slope2, SUM_Slope_W, and Slope_W to 0. The variable of SUM_Slope would be used to keep the sum of the slopes. The variable Slope1 would be used to keep the slope of the preceding straight line segment. The variable Slope2 would be used to keep the slope of the current straight line segment. The variable of Slope_W would be used to keep the weighted slope of the current straight line segment. The variable of SUM_Slope_W would be used to keep the sum of the weighted slopes.
- 5) Set the variables of SUM_Slope_CH, Slope_CH, SUM_Slope_CH_W, and Slope_CH_W to 0. The variable of SUM_Slope_CH would be used to keep the sum of the slope changes. The variable of Slope_CH would be used to keep the slope change from the current straight line segment from the preceding straight line segment. The variable of SUM_Slope_CH_W would be used to keep the sum of the weighted slope changes. The variable of Slope_CH_W would be used to keep the weighted slope change of the current straight line segment.
- 6) Set the variable of COUNT to 0. This variable would be used to keep the total number of straight line segments.
- 7) The first 3-D point coinciding to the start point of that FTSeg is retrieved and assigned to Point 1.
- 8) Find the next 3-D point.
- 9) If it is not found, i.e., there is no more 3-D point left, which belongs to the current FTSeg, the calculations for the sums of the slopes, slope changes, weighted slopes, and weighted slope changes for the current FTSeg is complete. The average slope, the weighted average slope, the average slope change,

and the weighted average slope for the current FTSeg were calculated and assigned to the current FTSeg.

Go to step 12).

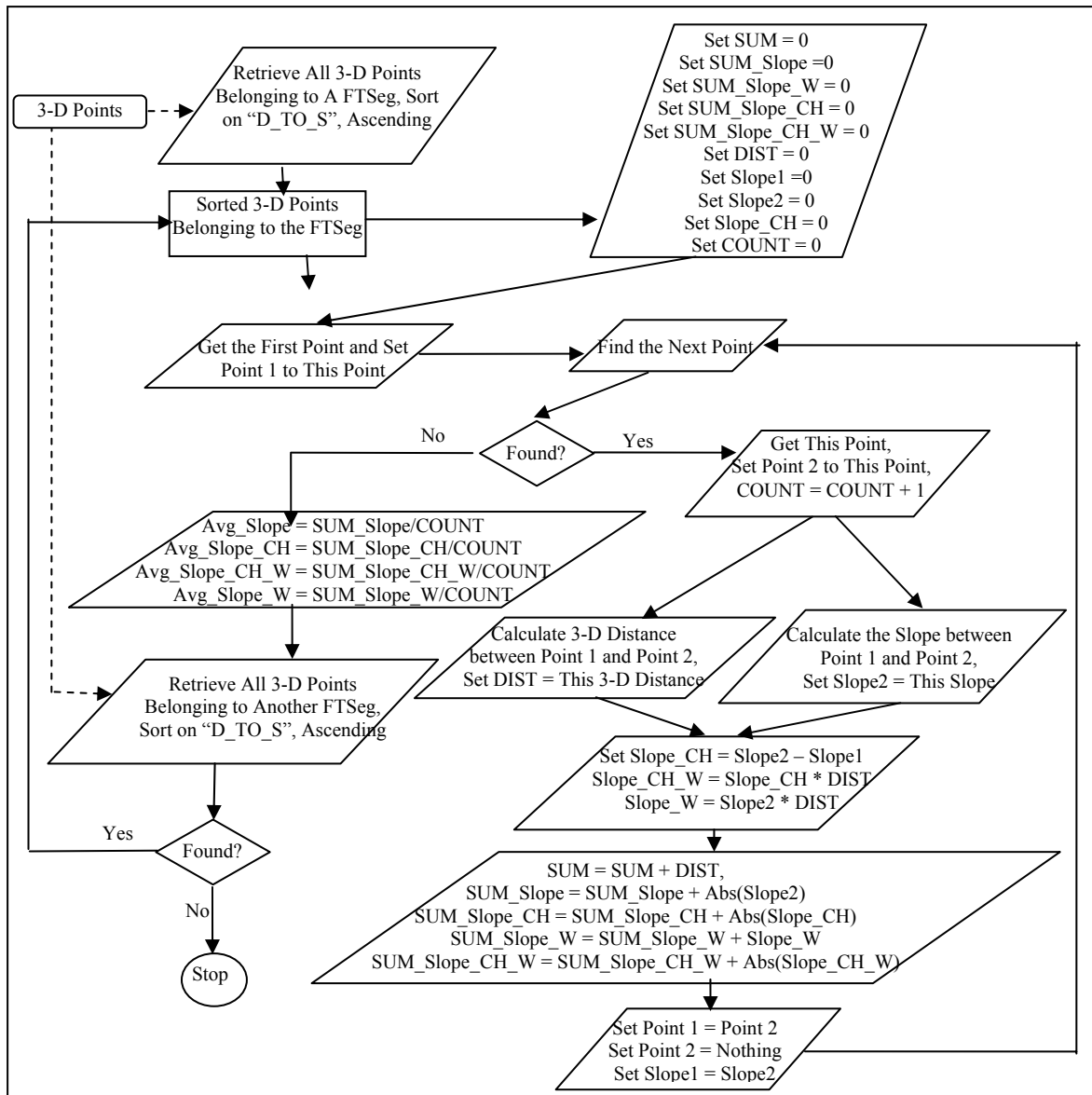


Figure 10.2 The Algorithm of Calculating Average Slopes, Weighted Average Slopes, Slope Changes, and Weighted Slope Changes

- 10) If it is found, it is assigned to Point 2. The variable COUNT is increased by 1. The 3-D distance of the straight line segment connecting Point 1 and Point 2 is calculated and assigned to DIST. The variable SUM is updated by adding this DIST value to it. The slope and the weighted slope for the current

straight line are calculated and the corresponding sums are updated. The slope change and the weighted slope change for the current straight line are calculated and the corresponding sums are updated.

- 11) Assign Point 2 to Point 1 and then assign Point 2 to nothing. Assign the value of Slope2 to Slope 1 and Go to step 8).
- 12) If there are still more FTSegs, retrieve all 3-D points belonging to the next FTSeg and go to step 2).
- 13) If there are no more FTSegs, calculations of average slopes, weighted average slopes, average slope changes, weighted average slope changes for all FTSegs are completed. The procedure stops.

10.3 ANALYSIS METHODS

In this case study, two methods are used to determine if a factor is a significant factor, i.e. if there is a linear association between a factor under consideration and the accuracy of the predicted 3-D distances when compared with DMI measured distances. The first one is the calculations of sample correlation coefficient and coefficient of determination of two variables, one is the factor under consideration and the other is one format of the errors in the predicted 3-D distances such as the difference and the proportional difference between the predicted 3-D distances and the DMI measured distances.

Assuming $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ denote a random sample of n pairs of observed values of a pair of continuous variables (X, Y) , and S_{xy} is defined as the sample covariance of these two variables ($\text{cov}(X, Y)$) and is calculated as $(1/n) * \sum (x_i - \mu_x) * (y_j - \mu_y)$, where μ_x and μ_y are the means of variables X and Y , respectively, the sample correlation coefficient between X and Y (r_{xy}) would be calculated as $\text{cov}(X, Y) / (\sigma_x * \sigma_y)$ (Rao 1998). In this research, variable X is the accuracy of the predicted 3-D distances (the difference or the proportional difference between predicted 3-D distances and the DMI measured distances). Variable Y is the factor under consideration such as the distance or the average slope. The sample determination is calculated as the square of the sample correlation coefficient.

According to Microsoft Excel Help, the correlation coefficient could be used to determine whether two ranges of data move together — that is, whether large values of one set are associated with large values of the other

(positive correlation), whether small values of one set are associated with large values of the other (negative correlation), or whether values in both sets are unrelated (correlation near zero). The value of the sample correlation coefficient ranges from -1 to 1. Consequently, the value of the sample coefficient of determination ranges from 0 to 1. It could be understood in such a way that the closer to 1 the value of the sample coefficient of determination is, the stronger the linear association is. If the value of the sample coefficient of determination is 1, variables X and Y are said to be perfectly correlated (Rao 1998). If the value of the sample coefficient of determination is 0, variables X and Y are said to have no linear association. The sign of the sample correlation coefficient determines if there is a positive or a negative linear association between two variables.

Furthermore, the sample coefficient of determination is the measure of the percentage of the variation in one variable that could be explained by the other variable (Rao 1998). In this research, it is determined that if the sample correlation coefficient of two variables, one is the factor under consideration and the other is the error of the predicted 3-D distance when compared to DMI measured distance, either in the format of difference or proportional difference, is greater than or equal to 0.30 or less than or equal to -0.30, this factor would be considered as a significant factor. The corresponding sample coefficient of determination would be 0.09. In other words, greater than or equal to 9% of the variation in the error could be explained by this factor.

The second method is the grouping and comparison method. With this method, rather than grouping roads into different road types, roads are grouped based on the value of the factor under consideration. RMSEs for these groups are obtained and compared. Given a particular set of predicted 3-D distances determined by the use of elevation dataset and the model parameter, the variation of the RMSEs among the groups could be evaluated and the effects of the factor under consideration on the accuracy of the predicted 3-D distances could be determined. For example, if the distance is the factor under consideration, all FTSEgs could be grouped based on their DMI measured distances and the comparison would be based on the RMSEs of these groups.

10.4 SIGNIFICANCE DETERMINATION

This section describes the results of determining significant factors using the methods described above. Based on the results, observations are obtained and presented. For each of the factors under consideration, the analyses using the differences between the predicted 3-D distances and the DMI measured distances as the error format and the analyses using the proportional differences as the error format are conducted separately.

10.4.1 Sample Correlation Coefficient and Sample Coefficient of Determination

This subsection evaluates the significance of the factors under consideration by using the first method, or the calculation of sample correlation coefficients and sample coefficients of determination.

10.4.1.1 Distance

This subsection represents the results of analyzing the sample data to determine if the distance factor is a significant factor.

Table 10.1 summarizes the results of sample correlation coefficients and sample coefficients of determination, when the distance and the difference are two variables under consideration.

Table 10.1 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Difference with the Factor of Distance

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	-0.07	0.00	-0.34	0.12	-0.36	0.13	0.16	0.03	0.06	0.00	-0.35	0.13	-0.37	0.13
Inter	-0.11	0.01	-0.37	0.14	-0.39	0.15	0.27	0.07	0.16	0.03	-0.38	0.14	-0.39	0.16
US	-0.44	0.19	-0.65	0.42	-0.65	0.42	-0.54	0.29	-0.58	0.34	-0.67	0.44	-0.66	0.44
NC	0.42	0.17	0.08	0.01	0.06	0.00	0.28	0.08	0.22	0.05	0.06	0.00	0.05	0.00

It is observed based on Table 10.1 that from the aspect of the difference, there are negative linear associations between the distance factor and the error (in the format of the difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance when using LIDAR 20-ft DEMs and NED. The magnitudes of the sample correlation coefficients exceed 0.3. There is a negative linear association between the

distance factor and the error of the predicted 3-D distance when using LIDAR point data, but the magnitude of the sample correlation coefficient is less than 0.3. There is a positive linear association between the distance factor and the error of the predicted 3-D distance when using LIDAR 50-ft DEM, but the magnitude is less than 0.3. It is also observed that the sample correlation coefficients vary with the road types.

Other than the difference and the proportional difference, their magnitudes are also concerned because in reality, a smaller error means its magnitude is smaller. It is obvious that if two errors are both positive, then the error with the smaller value is the error whose magnitude is smaller and is considered to be the smaller error. If two errors are both negative, then the error with the larger value is the error whose magnitude was smaller and is considered to be the smaller error. If two errors consist of one positive error and one negative error, their magnitudes must be compared to determine which one is the smaller error.

Table 10.2 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the absolute values (magnitudes) of differences, where the factor under consideration is the distance factor.

Table 10.2 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Difference with the Factor of Distance

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	0.31	0.10	0.36	0.13	0.37	0.14	0.35	0.12	0.34	0.11	0.40	0.16	0.41	0.16
Inter	0.29	0.09	0.38	0.14	0.39	0.15	0.37	0.14	0.36	0.13	0.42	0.18	0.43	0.19
US	0.49	0.24	0.71	0.50	0.71	0.51	0.59	0.35	0.63	0.40	0.73	0.53	0.72	0.52
NC	0.37	0.14	0.05	0.00	0.04	0.00	0.15	0.02	0.08	0.01	0.05	0.00	0.06	0.00

It is observed based on Table 10.2 that from the aspect of the magnitude of the difference, there are positive linear associations between the distance factor and the error (in the format of the absolute difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance, regardless of the different combinations of elevation dataset and model parameters. All the magnitudes of the sample correlation coefficients exceed 0.3.

Table 10.3 summarizes the results of sample correlation coefficients and sample coefficients of determination based on proportional differences, where the factor under consideration is the distance factor.

Table 10.3 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Proportional Difference with the Factor of Distance

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	0.10	0.01	0.09	0.01	0.09	0.01	0.09	0.01	0.09	0.01	0.10	0.01	0.10	0.10
Inter	0.02	0.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.02	0.02
US	0.13	0.02	0.12	0.02	0.12	0.02	0.12	0.02	0.12	0.02	0.14	0.02	0.14	0.13
NC	0.15	0.02	0.15	0.02	0.15	0.02	0.15	0.02	0.15	0.02	0.15	0.02	0.15	0.15

Based on Table 10.3, it is observed that from the aspect of the proportional difference, there are positive linear associations between the distance factor and the error of the predicted 3-D distance. However, all the magnitudes of the sample correlation coefficients are less than 0.3.

Table 10.4 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the absolute values (magnitudes) of proportional differences, where the factor under consideration is the distance factor.

Table 10.4 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Proportional Difference with the Factor of Distance

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	-0.38	0.14	-0.39	0.15	-0.38	0.15	-0.38	0.14	-0.38	0.14	-0.38	0.14	-0.38	0.14
Inter	-0.44	0.19	-0.45	0.20	-0.45	0.20	-0.44	0.20	-0.44	0.19	-0.44	0.20	-0.44	0.20
US	-0.31	0.10	-0.31	0.10	-0.31	0.10	-0.31	0.10	-0.31	0.10	-0.31	0.10	-0.31	0.10
NC	-0.44	0.20	-0.44	0.19	-0.44	0.19	-0.44	0.19	-0.44	0.19	-0.44	0.20	-0.44	0.20

Based on Table 10.4, it is observed that from the aspect of the magnitudes of the proportional errors (the absolute proportional differences), there are negative linear association between the distance factor and the error of the predicted 3-D distances, regardless of the different combinations of elevation datasets and model parameters. All the magnitudes of the sample correlation coefficients exceed 0.3.

It is concluded that the distance factor is significant to the accuracy of the predicted 3-D distance. Considering the magnitudes of errors, there is a positive linear association between the error of the predicted 3-D distance and the distance factor from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the longer the FTSeg is, the bigger the error it has and the smaller the proportional effort it has.

10.4.1.2 Average Slope

This subsection represents the results of analyzing the sample data to determine if the average slope is a significant factor.

Table 10.5 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the difference, where the factor under consideration is the average slope.

Table 10.5 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Difference with the Factor of the Average Slope

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	-0.22	0.05	-0.33	0.11	-0.33	0.11	-0.15	0.02	-0.20	0.04	-0.34	0.12	-0.34	0.12
Inter	-0.28	0.08	-0.45	0.20	-0.45	0.20	-0.14	0.02	-0.21	0.05	-0.45	0.20	-0.45	0.20
US	-0.16	0.03	-0.18	0.03	-0.18	0.03	-0.22	0.05	-0.21	0.05	-0.20	0.04	-0.20	0.04
NC	0.05	0.00	0.02	0.00	0.02	0.00	0.02	0.00	0.02	0.00	0.02	0.00	0.02	0.00

It is observed based on Table 10.5 that the linear associations between the error (in the format of the difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the average slope are negative. The magnitudes of the sample correlation coefficients exceed 0.3 for samples using LIDAR 20-ft DEMs and NED, but less than 0.3 for samples using LIDAR point data and LIDAR 50-ft DEMs. The associations vary with different road types as indicated by the varying values of the sample correlation coefficients.

Table 10.6 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the absolute difference, where the factor under consideration is the average slope.

Table 10.6 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Difference with the Factor of the Average Slope

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	0.24	0.06	0.36	0.13	0.36	0.13	0.19	0.04	0.25	0.06	0.36	0.13	0.36	0.13
Inter	0.31	0.09	0.47	0.22	0.47	0.22	0.20	0.04	0.28	0.08	0.47	0.22	0.47	0.22
US	0.18	0.03	0.19	0.04	0.19	0.04	0.23	0.05	0.22	0.05	0.20	0.04	0.20	0.04
NC	-0.07	0.00	-0.08	0.01	-0.08	0.01	-0.05	0.00	-0.06	0.00	-0.07	0.00	-0.07	0.00

It is observed based on Table 10.6 that the linear associations between the error (in the format of the absolute difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the average slope are positive. The magnitudes of the sample correlation coefficients exceed 0.3 for samples using LIDAR 20-ft DEMs and NED, but less than 0.3 (but close to 0.3) for samples using LIDAR point data and LIDAR 50-ft DEMs. The associations vary with different road types as indicated by the varying values of the sample correlation coefficients.

Table 10.7 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the proportional difference, where the factor under consideration is the average slope.

Table 10.7 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Proportional Difference with the Factor of the Average Slope

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	0.00	0.00	-0.05	0.00	-0.05	0.00	-0.04	0.00	-0.04	0.00	-0.04	0.00	-0.04	0.00
Inter	-0.01	0.00	-0.09	0.01	-0.09	0.01	-0.06	0.00	-0.06	0.00	-0.06	0.00	-0.07	0.00
US	-0.02	0.00	-0.06	0.00	-0.06	0.00	-0.08	0.01	-0.08	0.01	-0.06	0.00	-0.06	0.00
NC	0.07	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.05	0.00	0.05	0.00

Based on Table 10.7, it is observed that from the aspect of the proportional difference, there is no linear association between the error (in the format of the proportional difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the average slope, as indicated by the small magnitudes of the sample correlation coefficients and the sample coefficients of determination.

Table 10.8 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the absolute proportional difference, where the factor under consideration is the average slope.

Table 10.8 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Proportional Difference with the Factor of the Average Slope

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{XY}	r^2_{XY}	r_{XY}	r^2_{XY}	r_{XY}	r^2_{XY}	r_{XY}	r^2_{XY}	r_{XY}	r^2_{XY}	r_{XY}	r^2_{XY}	r_{XY}	r^2_{XY}
All	-0.25	0.06	-0.23	0.05	-0.23	0.05	-0.23	0.05	-0.22	0.05	-0.22	0.05	-0.22	0.05
Inter	-0.38	0.15	-0.35	0.12	-0.35	0.12	-0.36	0.13	-0.35	0.12	-0.34	0.11	-0.34	0.11
US	-0.21	0.04	-0.19	0.04	-0.19	0.03	-0.18	0.03	-0.17	0.03	-0.17	0.03	-0.17	0.03
NC	-0.15	0.02	-0.16	0.03	-0.16	0.03	-0.16	0.03	-0.16	0.03	-0.16	0.03	-0.16	0.03

It is observed based on Table 10.8 that the linear associations between the error (in the format of the absolute proportional difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the average slope are negative. All magnitudes of the sample correlation coefficients are less than 0.3, but close to 0.3.

It is concluded that the average slope could be considered as a significant factor to the accuracy of the predicted 3-D distances when compared to DMI measured distances. Considering the magnitudes of errors, there is a positive linear association between the error of the predicted 3-D distance and the average slope factor from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the larger the average slope of the FTSeg is, the bigger the error it has and the smaller the proportional effort it has.

10.4.1.3 Weighted Average Slope

This subsection represents the results of analyzing the sample data to determine if the weighted average slope is a significant factor.

Table 10.9 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the difference, where the factor under consideration is the weighted average slope.

Table 10.9 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Difference with the Factor of the Weighted Average Slope

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	-0.19	0.04	-0.31	0.10	-0.31	0.10	-0.14	0.02	-0.19	0.04	-0.32	0.11	-0.32	0.10
Inter	-0.26	0.07	-0.46	0.21	-0.46	0.21	-0.14	0.02	-0.22	0.05	-0.46	0.21	-0.46	0.21
US	-0.15	0.02	-0.16	0.03	-0.16	0.03	-0.20	0.04	-0.20	0.04	-0.18	0.03	-0.18	0.03
NC	0.07	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.07	0.00	0.07	0.00

It is observed based on Table 10.9 that the linear associations between the error (in the format of the difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the weighted average slope are negative. The magnitudes of the sample correlation coefficients exceed 0.3 for samples using LIDAR 20-ft DEMs and NED, but less than 0.3 for samples using LIDAR point data and LIDAR 50-ft DEMs. The associations vary with different road types as indicated by the varying values of the sample correlation coefficients.

Table 10.10 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the absolute difference, where the factor under consideration is the weighted average slope.

Table 10.10 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Difference with the Factor of the Weighted Average Slope

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	0.20	0.04	0.33	0.11	0.33	0.11	0.17	0.03	0.23	0.05	0.34	0.12	0.34	0.12
Inter	0.28	0.08	0.48	0.23	0.48	0.23	0.19	0.04	0.28	0.08	0.48	0.23	0.47	0.22
US	0.13	0.02	0.15	0.02	0.15	0.02	0.19	0.04	0.18	0.03	0.17	0.03	0.17	0.03
NC	-0.13	0.02	-0.12	0.01	-0.12	0.01	-0.09	0.01	-0.10	0.01	-0.11	0.01	-0.11	0.01

It is observed based on Table 10.10 that the linear associations between the error (in the format of the absolute difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the weighted average slope are positive. The magnitudes of the sample correlation coefficients exceed 0.3 for samples using LIDAR 20-ft DEMs and NED, but less than 0.3 (but close to 0.3) for samples using LIDAR point data and LIDAR 50-ft DEMs. The associations vary with different road types as indicated by the varying values of the sample correlation coefficients.

Table 10.11 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the proportional difference, where the factor under consideration is the weighted average slope.

Table 10.11 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Proportional Difference with the Factor of the Weighted Average Slope

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	-0.03	0.00	-0.07	0.01	-0.07	0.01	-0.06	0.00	-0.07	0.00	-0.07	0.00	-0.07	0.00
Inter	-0.05	0.00	-0.13	0.02	-0.13	0.02	-0.10	0.01	-0.10	0.01	-0.11	0.01	-0.11	0.01
US	-0.07	0.00	-0.11	0.01	-0.11	0.01	-0.12	0.01	-0.12	0.01	-0.10	0.01	-0.10	0.01
NC	0.07	0.01	0.06	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.06	0.00

Based on Table 10.11, it is observed that from the aspect of the proportional difference, there is no linear association between the error (in the format of the proportional difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the weighted average slope, as indicated by the small magnitudes of the sample correlation coefficients and the sample coefficients of determination.

Table 10.12 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the absolute proportional difference, where the factor under consideration is the weighted average slope.

Table 10.12 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Proportional Difference with the Factor of the Weighted Average Slope

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	-0.21	0.05	-0.20	0.04	-0.20	0.04	-0.20	0.04	-0.19	0.04	-0.19	0.04	-0.19	0.04
Inter	-0.39	0.15	-0.36	0.13	-0.36	0.13	-0.37	0.14	-0.36	0.13	-0.35	0.12	-0.35	0.12
US	-0.15	0.02	-0.13	0.02	-0.13	0.02	-0.12	0.02	-0.12	0.01	-0.12	0.01	-0.12	0.01
NC	-0.06	0.00	-0.13	0.02	-0.07	0.01	-0.07	0.01	-0.07	0.01	-0.07	0.01	-0.07	0.01

It is observed based on Table 10.12 that the linear associations between the error (in the format of the absolute proportional difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the weighted average slope are negative. All magnitudes of the sample correlation coefficients are less than 0.3, but close to 0.3.

The conclusion regarding the significance of the weighted average slope is the same as the average slope. More specifically, the weighted average slope could be considered as a significant factor to the accuracy of the predicted 3-D distances when compared to DMI measured distances. Considering the magnitudes of errors, there is a positive linear association between the error of the predicted 3-D distance and the weighted average slope factor from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the larger the weighted average slope of the FTSeg is, the bigger the error it has and the smaller the proportional effort it has.

10.4.1.4 Average Slope Change

This subsection represents the results of analyzing the sample data to determine if the average slope change is a significant factor.

Table 10.13 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the difference, where the factor under consideration is the average slope change.

Table 10.13 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Difference with the Factor of the Average Slope Change

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	-0.24	0.06	-0.36	0.13	-0.36	0.13	-0.15	0.02	-0.21	0.04	-0.37	0.14	-0.37	0.14
Inter	-0.26	0.07	-0.46	0.21	-0.46	0.21	-0.14	0.02	-0.21	0.04	-0.45	0.20	-0.45	0.20
US	-0.19	0.04	-0.21	0.04	-0.21	0.05	-0.25	0.06	-0.24	0.06	-0.23	0.05	-0.23	0.05
NC	0.09	0.01	-0.03	0.00	-0.04	0.00	0.01	0.00	0.00	0.00	-0.04	0.00	-0.05	0.00

It is observed based on Table 10.13 that the linear associations between the error (in the format of the difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the average slope change are negative. The magnitudes of the sample correlation coefficients exceed 0.3 for samples using LIDAR 20-ft DEMs and NED, but less than 0.3 for samples using LIDAR point data and LIDAR 50-ft DEMs. The associations vary with different road types as indicated by the varying values of the sample correlation coefficients.

Table 10.14 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the absolute difference, where the factor under consideration is the average slope change.

Table 10.14 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Difference with the Factor of the Average Slope Change

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	0.26	0.07	0.39	0.16	0.39	0.16	0.22	0.05	0.29	0.08	0.40	0.16	0.40	0.16
Inter	0.31	0.10	0.49	0.24	0.49	0.24	0.21	0.04	0.30	0.09	0.48	0.23	0.48	0.23
US	0.21	0.05	0.22	0.05	0.22	0.05	0.27	0.07	0.26	0.07	0.23	0.05	0.23	0.06
NC	0.16	0.03	0.08	0.01	0.09	0.01	0.10	0.01	0.09	0.01	0.09	0.01	0.09	0.01

It is observed based on Table 10.14 that the linear associations between the error (in the format of the absolute difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the average slope change are positive. The magnitudes of the sample correlation coefficients exceed 0.3 for samples using LIDAR 20-ft DEMs and NED, but less than 0.3 (but close to 0.3) for samples using LIDAR point data and LIDAR 50-ft DEMs. The associations vary with different road types as indicated by the varying values of the sample correlation coefficients.

Table 10.15 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the proportional difference, where the factor under consideration is the average slope change.

Table 10.15 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Proportional Difference with the Factor of the Average Slope Change

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	0.02	0.00	-0.02	0.00	-0.02	0.00	-0.02	0.00	-0.02	0.00	-0.01	0.00	-0.01	0.00
Inter	0.00	0.00	-0.08	0.01	-0.08	0.01	-0.05	0.00	-0.05	0.00	-0.05	0.00	-0.05	0.00
US	0.02	0.00	-0.02	0.00	-0.02	0.00	-0.04	0.00	-0.04	0.00	-0.02	0.00	-0.02	0.00
NC	0.10	0.01	0.09	0.01	0.09	0.01	0.09	0.01	0.09	0.01	0.09	0.01	0.09	0.01

Based on Table 10.15, it is observed that from the aspect of the proportional difference, there is no linear association between the error (in the format of the proportional difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the average slope change, as

indicated by the small magnitudes of the sample correlation coefficients and the sample coefficients of determination.

Table 10.16 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the absolute proportional difference, where the factor under consideration is the average slope change.

Table 10.16 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Proportional Difference with the Factor of the Average Slope Change

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	-0.30	0.09	-0.29	0.08	-0.29	0.08	-0.28	0.08	-0.28	0.08	-0.27	0.08	-0.27	0.08
Inter	-0.40	0.16	-0.37	0.14	-0.37	0.14	-0.38	0.14	-0.37	0.13	-0.36	0.13	-0.36	0.13
US	-0.26	0.07	-0.24	0.06	-0.24	0.06	-0.23	0.05	-0.23	0.05	-0.22	0.05	-0.22	0.05
NC	-0.42	0.18	-0.43	0.18	-0.43	0.18	-0.43	0.18	-0.43	0.18	-0.43	0.19	-0.43	0.19

It is observed based on Table 10.16 that the linear associations between the error (in the format of the absolute proportional difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the average slope change are negative. All magnitudes of the sample correlation coefficients are less than 0.3 (but very close to 0.3) or equal to 0.3

The conclusion regarding the significance of the average slope change is the same as the average slope and the weighted average slope. More specifically, the average slope change could be considered as a significant factor to the accuracy of the predicted 3-D distances when compared to DMI measured distances. Considering the magnitudes of errors, there is a positive linear association between the error of the predicted 3-D distance and the average slope change factor from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the larger the average slope change of the FTSeg is, the bigger the error it has and the smaller the proportional effort it has.

10.4.1.5 Weighted Average Slope Change

This subsection represents the results of analyzing the sample data to determine if the weighted average slope change is a significant factor.

Table 10.17 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the difference, where the factor under consideration is the weighted average slope change.

Table 10.17 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Difference with the Factor of the Weighted Average Slope Change

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	-0.25	0.06	-0.39	0.15	-0.39	0.15	-0.17	0.03	-0.23	0.05	-0.39	0.16	-0.39	0.16
Inter	-0.27	0.07	-0.49	0.24	-0.49	0.24	-0.14	0.02	-0.22	0.05	-0.49	0.24	-0.49	0.24
US	-0.24	0.06	-0.25	0.06	-0.25	0.06	-0.29	0.09	-0.29	0.08	-0.26	0.07	-0.27	0.07
NC	0.10	0.01	0.03	0.00	0.02	0.00	0.04	0.00	0.04	0.00	0.01	0.00	0.01	0.00

It is observed based on Table 10.17 that the linear associations between the error (in the format of the difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the weighted average slope change are negative. The magnitudes of the sample correlation coefficients exceed 0.3 for samples using LIDAR 20-ft DEMs and NED, but less than 0.3 for samples using LIDAR point data and LIDAR 50-ft DEMs. The associations vary with different road types as indicated by the varying values of the sample correlation coefficients.

Table 10.18 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the absolute difference, where the factor under consideration is the weighted average slope change.

Table 10.18 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Difference with the Factor of the Weighted Average Slope Change

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	0.27	0.07	0.41	0.17	0.41	0.17	0.23	0.05	0.29	0.09	0.42	0.17	0.42	0.17
Inter	0.32	0.10	0.52	0.27	0.52	0.27	0.22	0.05	0.31	0.09	0.51	0.26	0.51	0.26
US	0.24	0.06	0.24	0.06	0.24	0.06	0.29	0.08	0.28	0.08	0.26	0.07	0.26	0.07
NC	0.02	0.00	-0.04	0.00	-0.04	0.00	-0.02	0.00	-0.03	0.00	-0.02	0.00	-0.02	0.00

It is observed based on Table 10.18 that the linear associations between the error (in the format of the absolute difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the weighted average slope change are positive. The magnitudes of the sample correlation

coefficients exceed 0.3 for samples using LIDAR 20-ft DEMs and NED, but less than 0.3 (but close to 0.3) for samples using LIDAR point data and LIDAR 50-ft DEMs. The associations vary with different road types as indicated by the varying values of the sample correlation coefficients.

Table 10.19 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the proportional difference, where the factor under consideration is the weighted average slope change.

Table 10.19 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Proportional Difference with the Factor of the Weighted Average Slope Change

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	0.00	0.00	-0.05	0.00	-0.05	0.00	-0.04	0.00	-0.04	0.00	-0.04	0.00	-0.04	0.00
Inter	-0.03	0.00	-0.11	0.01	-0.11	0.01	-0.08	0.01	-0.08	0.01	-0.09	0.01	-0.09	0.01
US	0.00	0.00	-0.05	0.00	-0.05	0.00	-0.06	0.00	-0.06	0.00	-0.04	0.00	-0.04	0.00
NC	0.00	0.00	0.08	0.01	0.08	0.01	0.08	0.01	0.08	0.01	0.07	0.01	0.07	0.01

Based on Table 10.19, it is observed that from the aspect of the proportional difference, there is no linear association between the error (in the format of the proportional difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the weighted average slope change, as indicated by the small magnitudes of the sample correlation coefficients and the sample coefficients of determination.

Table 10.20 summarizes the results of sample correlation coefficients and sample coefficients of determination based on the absolute proportional difference, where the factor under consideration is the weighted average slope change.

Table 10.20 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination Based on the Absolute Proportional Difference with the Factor of the Weighted Average Slope Change

FTSeg Type	LIDAR Point Data		LIDAR 20-ft DEM				LIDAR 50-ft DEM				NED			
			10-ft Interval		20-ft Interval		25-ft Interval		50-ft Interval		15-m Interval		30-m Interval	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All	-0.30	0.09	-0.28	0.08	-0.28	0.08	-0.28	0.08	-0.27	0.07	-0.27	0.07	-0.27	0.07
Inter	-0.42	0.18	-0.39	0.15	-0.39	0.15	-0.39	0.16	-0.39	0.15	-0.37	0.14	-0.37	0.14
US	-0.24	0.06	-0.22	0.05	-0.22	0.05	-0.21	0.04	-0.21	0.04	-0.20	0.04	-0.20	0.04
NC	-0.35	0.12	-0.36	0.13	-0.36	0.13	-0.36	0.13	-0.36	0.13	-0.36	0.13	-0.36	0.13

It is observed based on Table 10.20 that the linear associations between the error (in the format of the absolute proportional difference between the predicted 3-D distance and the DMI measured distance) of the predicted 3-D distance and the factor of the weighted average slope change are negative. All magnitudes of the sample correlation coefficients are less than 0.3 (but very close to 0.3) or equal to 0.3

The conclusion regarding the significance of the weighted average slope change is the same as the average slope, the weighted average slope, and the average slope change. More specifically, the weighted average slope change could be considered as a significant factor to the accuracy of the predicted 3-D distances when compared to DMI measured distances. Considering the magnitudes of errors, there is a positive linear association between the error of the predicted 3-D distance and the weighted average slope change factor from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the larger the weighted average slope change of the FTSeg is, the bigger the error it has and the smaller the proportional effort it has.

10.4.1.6 Number of 3-D Points and Average Density of 3-D Points

This subsection represents the results of analyzing the sample data to determine if the number of 3-D points and the average density of the 3-D points along linear objects are significant factors. As stated earlier, the evaluation only applied to the accuracy of the predicted 3-D distances using LIDAR point data.

Table 10.21 summarizes the results of the sample correlation coefficients and sample coefficients of determination based on the difference, the absolute difference, the proportional difference, and the absolute proportional difference, with the number of 3-D points on the linear object as the factor under consideration..

Based on Table 10.21, it is observed that when the difference and the proportional difference are considered, there is no linear association between the accuracy of the predicted 3-D distances using LIDAR point data when compared to the DMI measured distances and the factor of the number of 3-D points, as indicated by the very small values of the magnitudes of the sample correlation coefficients and the sample coefficients of

determination. However, when considering the error magnitudes (the absolute difference and the absolute proportional difference), there is a positive linear association between the factor of the number of 3-D points and the error (absolute difference) and a negative linear association between the factor of the number of 3-D points and the proportional error (absolute proportional difference). Of the corresponding two sample correlation coefficients, the positive one has its magnitude less than but very close to 0.3 while the negative one has its magnitude exceeding 0.3.

Table 10.21 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination with the Factor of the Number of 3-D Points for Errors of Using LIDAR Point Data

Road Type	Difference		Absolute Difference		Proportional Difference		Absolute Proportional Difference	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All FTSegs	-0.07	0.00	0.29	0.08	0.10	0.01	-0.37	0.14
Interstate FTSegs	-0.06	0.00	0.25	0.06	0.02	0.00	-0.43	0.18
US FTSegs	-0.44	0.20	0.49	0.24	0.13	0.02	-0.30	0.09
NC FTSegs	0.43	0.19	0.38	0.15	0.15	0.02	-0.44	0.19

It is concluded that the factor of the number of 3-D points is a significant factor to the accuracy of the predicted 3-D distance using LDIAR point data when compared to the DMI measured distance. Considering the magnitudes of errors, there is a positive linear association between the error of the predicted 3-D distance and the factor of the number of 3-D points from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the more the 3-D points being used to represent an FTSeg in a 3-D space, the bigger the error it has and the smaller the proportional effort it has.

Generally speaking, the longer the linear object, the more 3-D points would be on it when using LIDAR point data and the snapping approach to model the linear object in a 3-D space. It is believed that treating the density of the 3-D points on a linear object as a factor under consideration would be a measure of the density of 3-D points, which excludes the effects from the length of the linear object on the number of 3-D points being used to model that linear object. In this case study, the average density of 3-D points on a linear object is calculated as the average planimetric distance along the linear object between two neighboring 3-D points. Table 10.22 summarizes the results of the sample correlation coefficients and sample coefficients of determination based on

the difference, the absolute difference, the proportional difference, and the absolute proportional difference, with the average density of 3-D points on the linear object as the factor under consideration.

Table 10.22 Summary of Sample Correlation Coefficients and Sample Coefficients of Determination with the Factor of the Average Density of 3-D Points for Errors of Using LIDAR Point Data

Road Type	Difference		Absolute Difference		Proportional Difference		Absolute Proportional Difference	
	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}	r_{xy}	r^2_{xy}
All FTSegs	0.13	0.02	-0.35	0.12	0.09	0.01	0.30	0.09
Interstate FTSegs	0.16	0.03	-0.33	0.11	0.08	0.01	0.64	0.41
US FTSegs	0.19	0.04	-0.36	0.13	0.27	0.07	0.13	0.02
NC FTSegs	-0.12	0.01	-0.47	0.22	-0.08	0.01	0.10	0.01

Based on Table 10.22, it is observed that when the difference and the proportional difference are considered, there is no linear association between the accuracy of the predicted 3-D distances using LIDAR point data when compared to the DMI measured distances and the factor of the average density of 3-D points, as indicated by the very small values of the magnitudes of the sample correlation coefficients and the sample coefficients of determination. However, when considering the error magnitudes (the absolute difference and the absolute proportional difference), there is a negative linear association between the factor of the average density of 3-D points and the error (absolute difference) and a positive linear association between the factor of the average density of 3-D points and the proportional error (absolute proportional difference). Of the corresponding two sample correlation coefficients, both have their magnitudes exceeding 0.3.

It is concluded that the factor of the average density of 3-D points is a significant factor to the accuracy of the predicted 3-D distance using LDIAR point data when compared to the DMI measured distance. Considering the magnitudes of errors, there is a positive linear association between the error of the predicted 3-D distance and the factor of the average density of 3-D points from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In this case study, the average density is represented with the average planimetric distance between two neighboring 3-D points along a linear object. In other words, the higher the density, the smaller the corresponding average distance. More specifically, the higher the average density of the 3-D points being used to represent a FTSeg in a 3-D space, the bigger the error it has and the smaller the proportional effort it has.

10.4.2 Grouping and Comparison

This section describes the significance evaluation of the factors under consideration by using the grouping and comparison method. Each factor is evaluated for the errors of the predicted 3-D distances in the formats of the difference and the proportional difference when compared to the DMI measured distances. The difference is in the unit of foot while the proportional difference is in the unit of foot per 1,000 feet.

10.4.2.1 Distance

This subsection evaluates if the distance is a significant factor to the accuracy of the predicted 3-D distances when compared to DMI measured distances.

Table 10.23 illustrates the designed groups based on the distance. The first group (Group 1) is used to include all very short FTSEgs. The last group (Group 7) is used to include all very long FTSEgs. The symbol “(” means “greater than”. The symbol “)” means “less than”. The symbol “[” means “less than or equal to”. Table 10.24 summarizes the RMSEs based on the difference of these groups. Figure 10.3 graphically illustrates the comparisons of these RMSEs.

Table 10.23 Illustration of Groups Based on the Distance

Group Name	Distance Range (ft)	Number of FTSEgs	Percentage
Group 1	(0, 100]	46	17.36%
Group 2	(100, 1,000]	26	9.81%
Group 3	(1,000, 5,000]	52	19.62%
Group 4	(5,000, 10,000]	28	10.57%
Group 5	(10,000, 20,000]	38	14.34%
Group 6	(20,000, 30,000]	32	12.08%
Group 7	(30,000, +∞)	43	16.23%
Total	--	265	100%

Table 10.24 Summary of RMSEs from the Difference for Groups Based on the Distance

Group	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
Group 1	6.75	6.84	6.84	6.78	6.78	6.77	6.77
Group 2	14.10	16.11	16.12	16.06	16.07	16.03	16.04
Group 3	26.73	32.80	32.93	31.56	31.63	33.39	33.42
Group 4	28.15	32.56	32.71	29.45	29.70	34.48	34.53
Group 5	27.76	32.19	32.62	37.92	31.69	37.95	38.10
Group 6	32.52	44.51	45.30	65.97	52.60	47.93	48.29
Group 7	32.23	47.40	48.24	56.33	47.48	49.72	50.20

It is observed from Figure 10.3 that the general trend of the RMSEs indicates that the longer the distance, the bigger the RMSE, given a particular elevation dataset and a model parameter. The two groups having the shortest distances are the two groups having the smallest RMSEs. The two groups having the longest distances are the two groups having the largest RMSEs. The other three groups having intermediate distances are the groups having intermediate RMSEs.

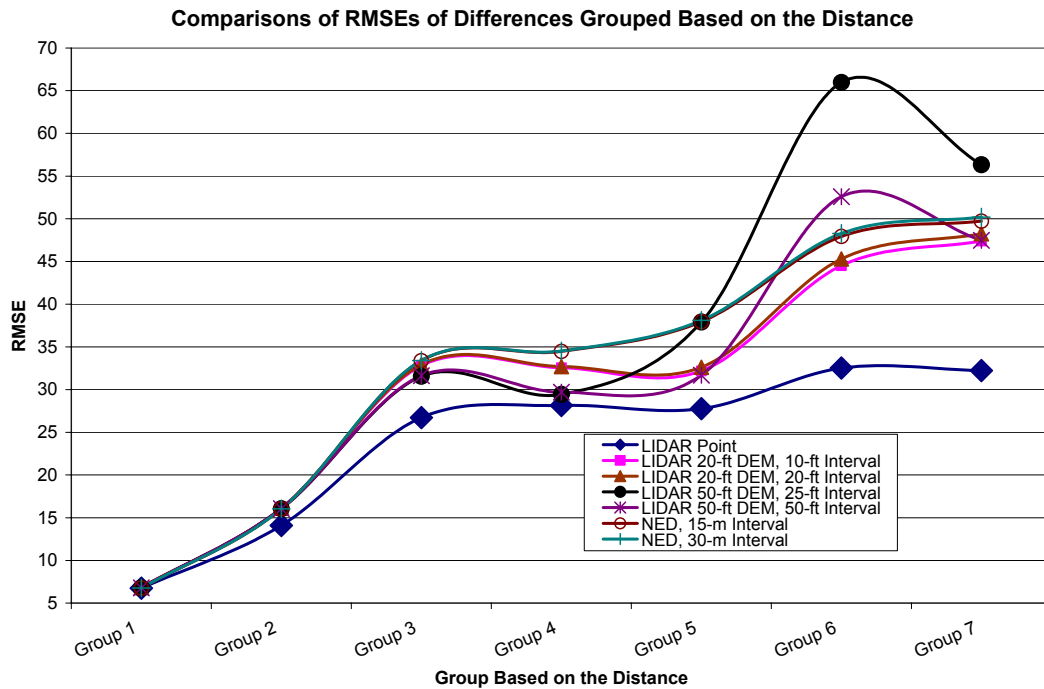


Figure 10.3 Comparison of RMSEs of the Difference for Groups Based on the Distance

Table 10.25 summarized the RMSEs based on the proportional difference of these groups. The unit was still foot per 1,000 feet. Figure 10.4 graphically illustrated the comparisons of these RMSEs.

Table 10.25 Summary of RMSEs from the Proportional Difference for Groups Based on the Distance

Group	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
Group 1	116.40	117.94	117.90	116.96	116.92	116.61	116.61
Group 2	41.30	43.25	43.19	42.00	41.83	42.27	42.27
Group 3	10.54	11.83	11.84	11.87	11.86	11.95	11.96
Group 4	4.23	5.06	5.09	4.46	4.61	5.40	5.41
Group 5	1.95	2.27	2.29	2.60	2.22	2.62	2.63
Group 6	1.32	1.76	1.79	2.51	1.98	1.92	1.93
Group 7	0.80	1.22	1.24	1.35	1.17	1.23	1.24

From Figure 10.4, it was obvious that the seven trend lines corresponding to the use of seven different combinations of elevation datasets and model parameters were clustered together. The general trend for these seven trend lines indicated that groups having longer distances were associated with smaller RMSEs of the proportional difference while groups having shorter distances were associated with larger RMSEs of the proportional difference.

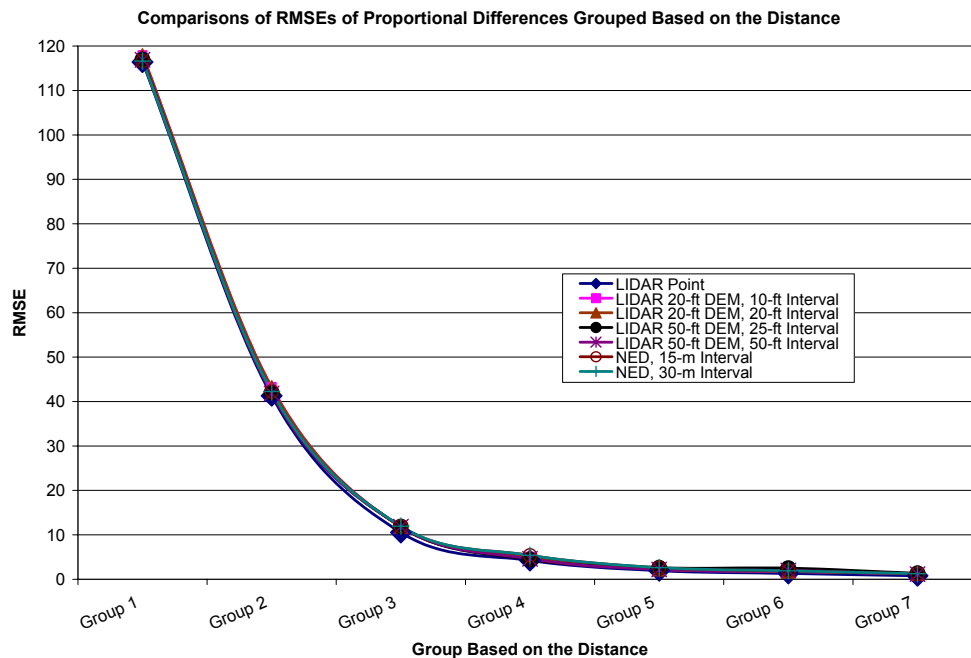


Figure 10.4 Comparison of RMSEs of the Proportional Difference for Groups Based on the Distance

It was concluded that the distance was a significant factor to the accuracy of the predicted 3-D distances when compared to DMI measured distances, regardless of the different elevation datasets and intervals being used. From the aspect of the difference, the distance had a positive association with the error (difference), i.e. the longer the distance, the larger the error (difference). From the aspect of the proportional difference, the distance had a negative association with the error (proportional difference), i.e. the longer the distance, the smaller the error (proportional difference). It was reasonable to obtain these conclusions because a proportional difference was obtained by dividing the difference by the distance.

10.4.2.2 Average Slope

This subsection evaluates if the average slope is a significant factor to the accuracy of the predicted 3-D distances when compared with DMI measured distances.

Table 10.26 illustrates the designed groups based on the average slope. The first group (Group 1) is used to include very flat FTSEgs. The last group (Group 5) is used to include steeper FTSEgs. It is noted that most of the FTSEgs in this case study are flat (81.51% of the FTSEgs having the average slope less than or equal to 5%).

Table 10.26 Illustration of Groups Based on the Average Slope

Group Name	Average Slope(%)	Number of FTSEgs	Percentage
Group 1	(0, 2]	38	14.34%
Group 2	(2, 3]	42	15.85%
Group 3	(3, 4]	77	29.06%
Group 4	(4, 5]	59	22.26%
Group 5	(5, +∞)	49	18.49%
Total	--	265	100%

Table 10.27 summarizes the RMSEs based on the difference of these groups. Figure 10.5 graphically illustrates the comparisons of these RMSEs.

Table 10.27 Summary of RMSEs from the Difference for Groups Based on the Average Slope

Group	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
Group 1	6.35	6.63	6.62	6.45	6.43	6.35	6.35
Group 2	21.40	19.47	19.45	26.59	23.68	20.25	20.29
Group 3	29.21	34.81	35.26	48.69	38.26	36.23	36.61
Group 4	28.82	39.36	40.03	49.59	43.90	43.45	43.58
Group 5	28.20	43.23	43.71	33.90	34.29	45.93	46.15

Based on Table 10.27 and Figure 10.5, it is observed that the effects on the accuracy of the predicted 3-D distances from the average slope depend on the elevation dataset being used, even though the general trend showed that groups with steep slopes are associated with big RMSEs (low accuracies) while groups with flat slopes are associated with small RMSEs (high accuracy).

In the cases of using LIDAR 20-ft DEMs and NED, there is a positive association between the average slope and the errors of the predicted 3-D distances from the aspect of the difference between the predicted 3-D distance and the DMI measured distance. In other words, groups with steep slopes are the groups having high RMSEs (low accuracies).

In the cases of using LIDAR 50-ft DEMs, the average slope of 5% is the threshold. For groups having slopes less than or equal to 5%, the groups with steep slopes are associated with high RMSEs while the groups having flat slopes are associated with low RMSEs. Interestingly, the group with the highest slopes has their RMSEs smaller than those of the groups having slopes between 3-5%, but bigger than the groups having slopes less than or equal to 3%.

In the case of using LIDAR point data, the group with the lowest slopes is the group associated with the smallest RMSE. For other groups, the RMSEs are quite close to each other.

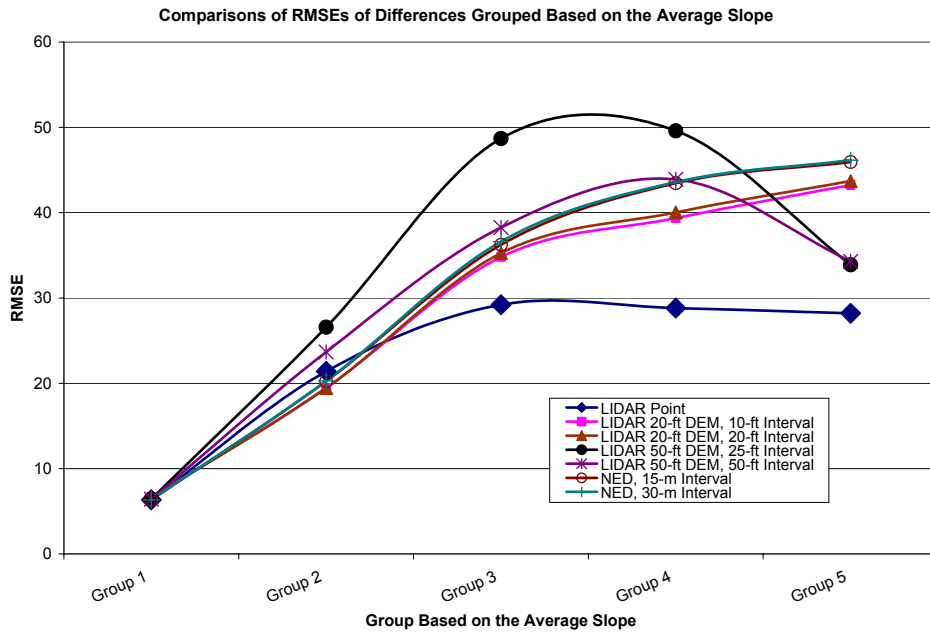


Figure 10.5 Comparison of RMSEs of the Difference for Groups Based on the Average Slope

Table 10.28 summarizes the RMSEs based on the proportional difference of these groups. Figure 10.6 graphically illustrates the comparisons of these RMSEs.

Table 10.28 Summary of RMSEs from the Proportional Difference for Groups Based on the Average Slope

Group	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
Group 1	80.88	83.40	83.35	81.60	81.49	81.04	81.04
Group 2	82.11	82.15	82.15	82.04	81.98	82.14	82.14
Group 3	34.08	34.49	34.47	34.33	34.33	34.17	34.17
Group 4	25.89	26.03	26.04	26.06	26.04	26.14	26.14
Group 5	16.55	18.88	18.88	18.69	18.72	18.90	18.90

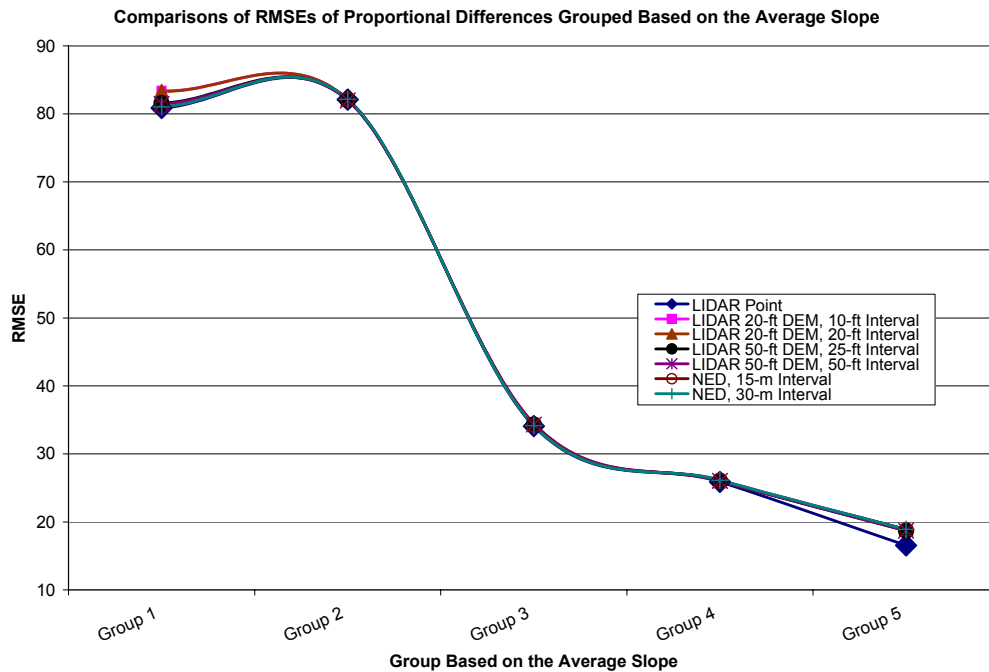


Figure 10.6 Comparison of RMSEs of the Proportional Difference for Groups Based on the Average Slope

From Figure 10.6, it is obvious that the seven trend lines corresponding to the use of seven different combinations of elevation datasets and model parameters are clustered together. The general trend for these seven trend lines indicates that groups having steep slopes are associated with smaller RMSEs of the proportional difference while groups having flat slopes are associated with larger RMSEs of the proportional difference. Group 1 and Group 2 are the two most flat groups with small slopes. The associated RMSEs of the

proportional difference to these two groups are close to each other, but higher than the RMSEs of the other three groups, of which the three RMSEs could be considered to have close values.

It is concluded that the average slope is a significant factor to the accuracy of the predicted 3-D distances when compared to DMI measured distances. From the aspect of the proportional difference, the average slope has a negative association with the error (proportional difference), i.e. the steeper the average slope, the smaller the error (proportional difference).

From the aspect of the difference, it is observed that the average slope is still a significant factor. However, the effects from the average slope on the accuracy of the predicted 3-D distances depend on the elevation dataset being used. But the overall trend shows that there is a positive association between the average slope and the accuracy of the predicted 3-D distances. In other words, groups with steep slopes are associated with big RMSEs (low accuracies) while groups with flat slopes are associated with small RMSEs (high accuracy).

One exception is in the cases of using LIDAR 50-ft DEMs, where the average slope of 5% is the threshold. For groups having slopes less than or equal to 5%, the groups with steep slopes are associated with high RMSEs while the groups having flat slopes are associated with low RMSEs. Interestingly, the group with the highest slopes has its RMSEs smaller than those of the groups having slopes between 3-5%, but bigger than the groups having slopes less than or equal to 3%.

In the case of using LIDAR point data, the group with the lowest slopes is the group associated with the smallest RMSE. For other groups, the RMSEs are quite close to each other.

10.4.2.3 Weighted Average Slope

This subsection evaluates if the weighted average slope is a significant factor to the accuracy of the predicted 3-D distances when compared with DMI measured distances.

Table 10.29 illustrates the designed groups based on the average slope. The first group (Group 1) is used to include very flat FTSEgs based on the weighted average slope. The last group (Group 6) is used to include steeper FTSEgs based on the weighted average slope. It is noted that most of the FTSEgs in this case study are flat (88.30% of the FTSEgs having the weighted average slope less than or equal to 4%).

Table 10.29 Illustration of Groups Based on the Weighted Average Slope

Group Name	Weighted Average Slope (%)	Number of FTSEgs	Percentage
Group 1	(0, 2]	46	17.36%
Group 2	(2, 2.5]	35	13.21%
Group 3	(2.5, 3]	75	28.30%
Group 4	(3, 3.5]	34	12.83%
Group 5	(3.5, 4]	31	11.70%
Group 6	(4, +∞)	44	16.60%
Total	--	265	100%

Table 10.30 summarizes the RMSEs based on the difference of these groups. Figure 10.7 graphically illustrates the comparisons of these RMSEs.

Table 10.30 Summary of RMSEs from the Difference for Groups Based on the Weighted Average Slope

Group	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
Group 1	11.68	11.67	11.66	13.35	12.81	11.63	11.64
Group 2	26.73	24.71	24.71	30.96	28.02	25.24	25.23
Group 3	27.55	34.75	35.14	49.19	38.82	34.52	34.79
Group 4	30.46	42.40	43.11	55.89	49.59	47.31	47.54
Group 5	27.38	32.73	33.52	36.59	30.60	40.82	41.13
Group 6	26.77	43.47	43.96	33.26	33.92	45.39	45.62

Based on Figure 10.7, it is observed that for groups from Group 1 to Group 4, there is a clear trend that the steeper the weighted average slope of the group, the bigger the RMSE based on the difference, given a particular combination of the elevation dataset and the model parameter. Group 4 is always the group with the biggest RMSEs of the difference in all cases. The Group 5 and Group 6 have RMSEs smaller than Group 4. The trend of Group 4, Group 5, and Group 6 vary with the different combinations of elevation datasets and

model parameters. When using LIDAR 20-ft DEMs, LIDAR 50-ft DEM with a 50-ft interval, and NED, Group 5 has its RMSE lower than the corresponding RMSE of Group 6. When using LIDAR point data and LIDAR 50-ft DEM with a 50-ft interval, the trend of Group 4, Group 5, and Group 6 is such that the steeper the weighted average slope of the group, the smaller the RMSEs of the difference.

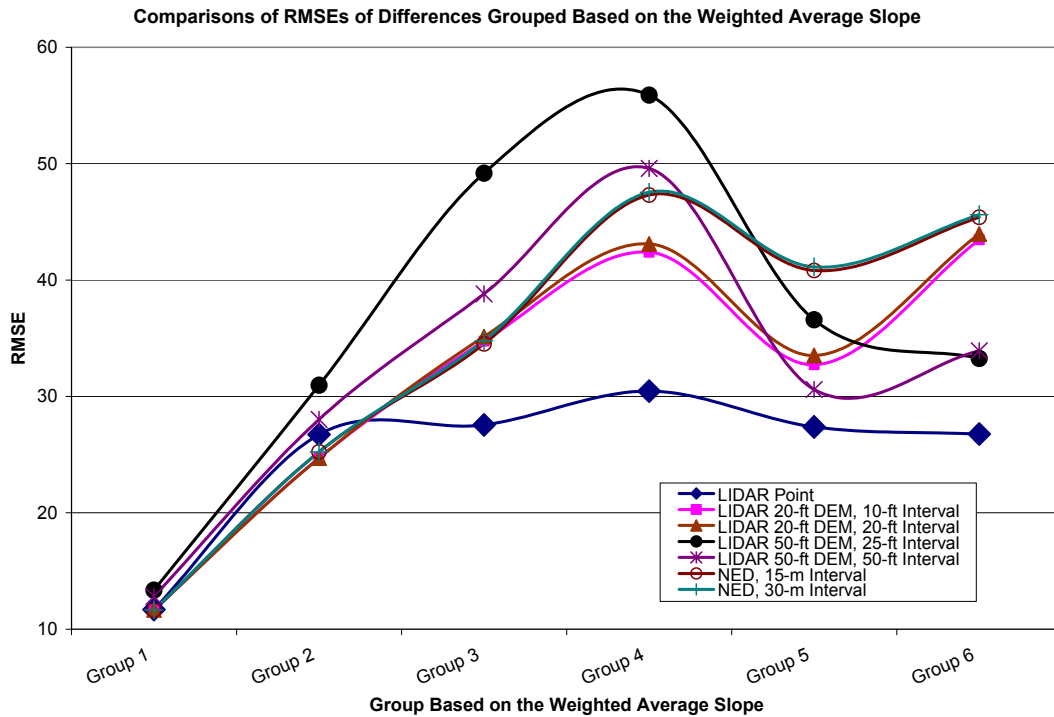


Figure 10.7 Comparison of RMSEs of the Difference for Groups Based on the Weighted Average Slope

Table 10.31 summarizes the RMSEs based on the proportional difference of these groups. Figure 10.8 graphically illustrates the comparisons of these RMSEs.

Table 10.31 Summary of RMSEs from the Proportional Difference for Groups Based on the Weighted Average Slope

Group	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
Group 1	90.07	91.96	91.92	90.61	90.54	90.19	90.19
Group 2	49.50	49.52	49.52	49.36	49.21	49.47	49.47
Group 3	45.60	45.88	45.85	45.76	45.78	45.67	45.67
Group 4	3.93	3.98	4.00	4.14	4.11	4.39	4.38
Group 5	22.05	22.62	22.58	22.10	22.11	22.14	22.14
Group 6	31.11	32.58	32.60	32.62	32.63	32.83	32.83

From Figure 10.8, it is obvious that the seven trend lines for the seven combinations of elevation datasets and model parameter are clustered together indicating the similarity among them when comparing RMSEs of the proportional difference for groups based on the weighted average slope. For groups of Group 1 to Group 4, there is a clear overall trend, i.e. the steeper the weighted average slope of the group, the smaller the RMSE of the proportional difference. For groups of Group 4 to Group 6, there is also a clear overall trend. But this time, the steeper the weighted average slope of the group, the bigger the RMSE of the proportional difference.

It is concluded that the weighted average slope could also be considered as a significant factor that has a close relationship with the accuracy of the predicted 3-D distances when compared with DMI measured distances. However, this relationship could not be simply described as positive linear or negative linear. In other words, this relationship is not linear.

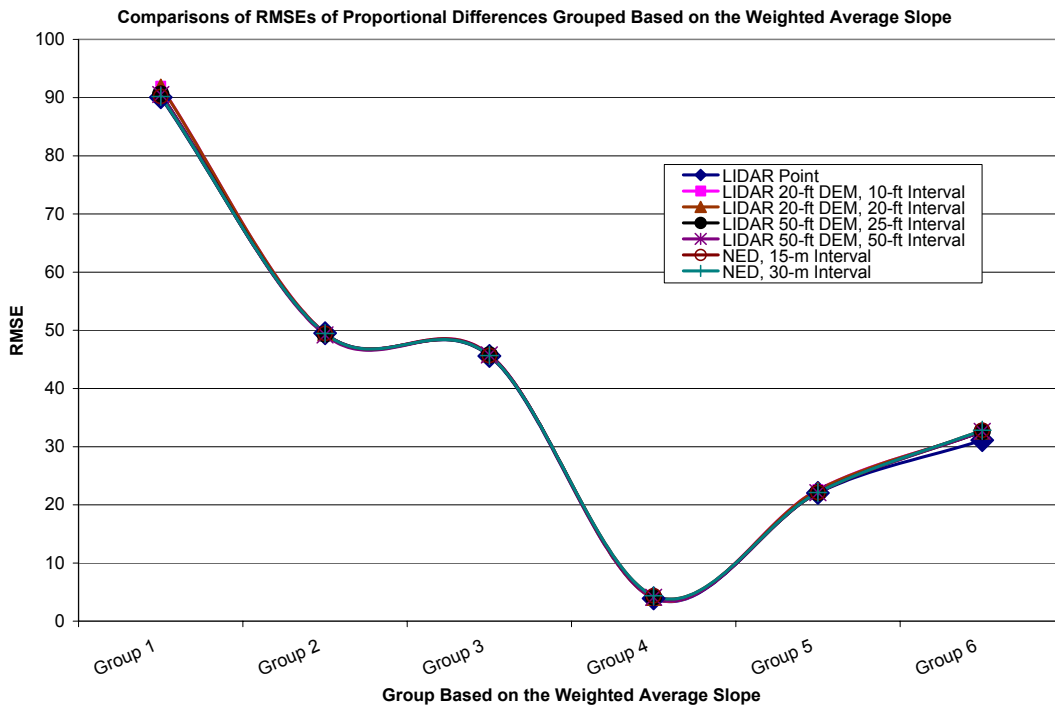


Figure 10.8 Comparison of RMSEs of the Proportional Difference for Groups Based on the Weighted Average Slope

10.4.2.4 Average Slope Change

This subsection evaluates if the average slope change is a significant factor to the accuracy of the predicted 3-D distances when compared with DMI measured distances.

Table 10.32 illustrates the designed groups based on the average slope change. The first group (Group 1) is used to include FTSEgs that had small average slope changes. The last group (Group 5) is used to include FTSEgs that had large average slope changes.

Table 10.32 Illustration of Groups Based on the Average Slope Change

Group Name	Average Slope Change (%)	Number of FTSEgs	Percentage
Group 1	(0, 3]	41	15.47%
Group 2	(3, 5]	42	15.85%
Group 3	(5, 6]	48	18.11%
Group 4	(6, 8]	68	25.66%
Group 5	(8, +∞)	66	24.91%
Total	--	265	100%

Table 10.33 summarizes the RMSEs based on the difference of these groups. Figure 10.9 graphically illustrates the comparisons of these RMSEs.

Table 10.33 Summary of RMSEs from the Difference for Groups Based on the Average Slope Change

Group	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
Group 1	6.89	7.13	7.12	6.97	6.96	6.90	6.90
Group 2	21.51	18.89	18.71	24.32	22.07	18.90	18.86
Group 3	27.16	28.14	28.35	47.71	37.29	29.95	30.14
Group 4	30.93	43.11	43.85	44.45	38.03	45.56	45.92
Group 5	28.22	41.27	41.80	46.83	42.93	44.97	45.16

Based on Figure 10.9, it is observed that the overall trend is such that groups that have higher values of the average slope changes have higher values of RMSEs (lower accuracies). However, there is small variation in the RMSEs of Group 4 and Group 5 when using LIDAR 20-ft DEMs and NED. There is a small variation in the RMSEs of Group 3, Group 4, and Group 5 when using LIDAR point data and LIDAR 50-ft DEMS. In other words, when the average slope change reaches 6% (for the uses of LIDAR 20-ft DEMs and NED) or 5% (for

the uses of LIDAR point data and LIDAR 50-ft DEMs), the effects on the accuracy of the predicted 3-D distances from the increase in the average slope change are almost nonexistent and could be ignored.

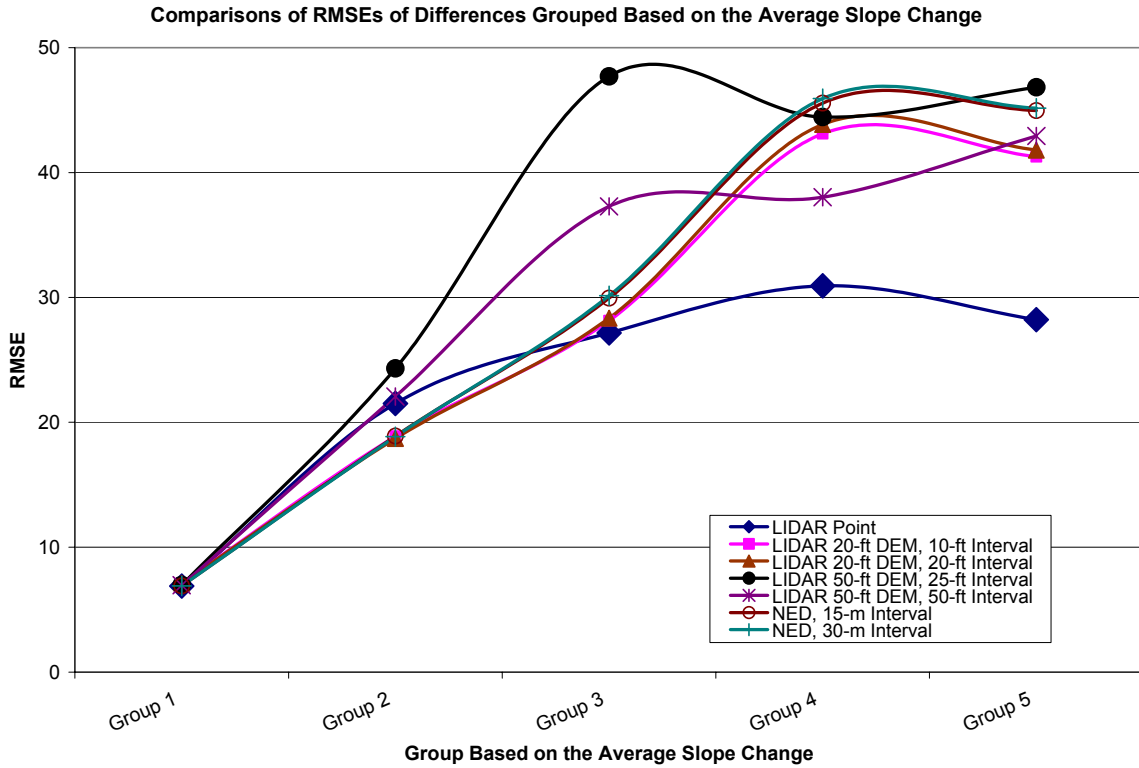


Figure 10.9 Comparison of RMSEs of the Difference for Groups Based on the Average Slope Change

Table 10.34 summarizes the RMSEs based on the proportional difference of these groups. Figure 10.10 graphically illustrates the comparisons of these RMSEs.

Table 10.34 Summary of RMSEs from the Proportional Difference for Groups Based on the Average Slope Change

Group	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
Group 1	98.90	100.83	100.79	99.42	99.37	99.03	99.03
Group 2	74.56	74.68	74.68	74.53	74.46	74.64	74.65
Group 3	15.47	15.69	15.69	15.73	15.69	15.61	15.62
Group 4	13.92	14.70	14.66	14.44	14.42	14.13	14.12
Group 5	15.10	17.05	17.05	16.94	16.94	17.07	17.07

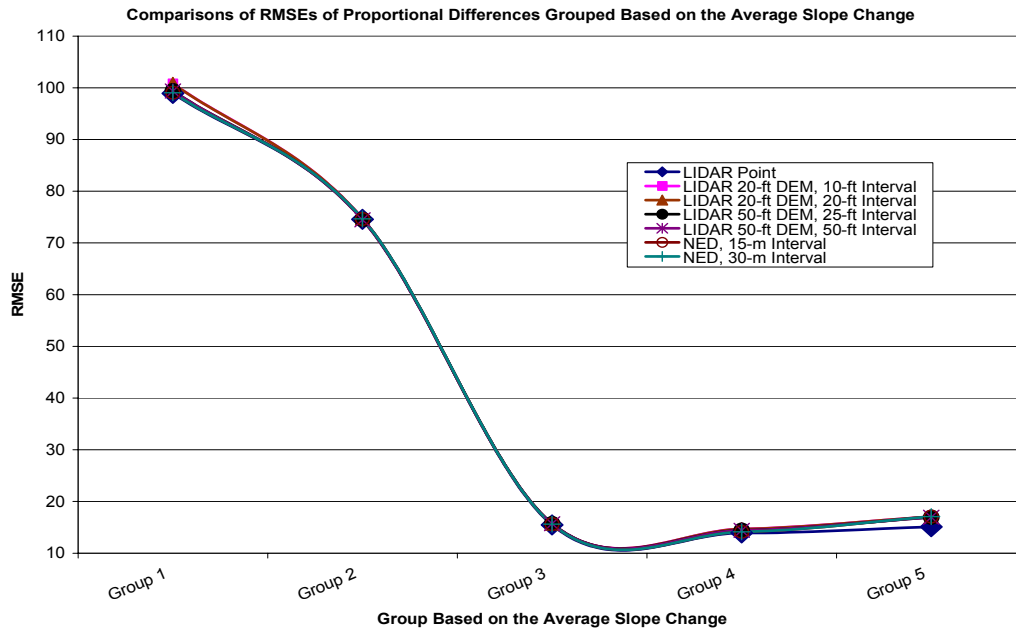


Figure 10.10 Comparison of RMSEs of the Proportional Difference for Groups Based on the Average Slope Change

Based on Figure 10.34, it was observed that the seven trend lines for the seven combinations of different elevation datasets and model parameters were clustered together. The overall trend was such that groups that had higher average slope changes had lower RMSEs (higher accuracies). Similar to the observation from the aspect of the difference, when the average slope change reached 5% (Group 3, Group 4, and Group 5), there was small variation in the RMSEs. In other words, when the average slope change reached 5%, the effects on the accuracy of the predicted 3-D distances from the increase in the average slope changed were almost nonexistent and could be ignored.

It was concluded that the average slope change could be considered as a significant factor. From the aspect of the difference, there was a positive association between the error and the average slope change. From the aspect of the proportional difference, there was a negative association between the error and the average slope change. Furthermore, when the average slope change reached 5%, the effects on the accuracy of the predicted 3-D distances from the increase in the average slope change were almost nonexistent and could be ignored.

10.4.2.5 Weighted Average Slope Change

This subsection evaluates if the weighted average slope change is a significant factor to the accuracy of the predicted 3-D distances when compared with DMI measured distances.

Table 10.35 illustrates the designed groups based on the weighted average slope change. The first group (Group 1) is used to include FTSEgs that have small weighted average slope changes. The last group (Group 5) is used to include FTSEgs that have large weighted average slope changes.

Table 10.35 Illustration of Groups Based on the Weighted Average Slope Change

Group Name	Weighted Average Slope Change (%)	Number of FTSEgs	Percentage
Group 1	(0, 3]	47	17.74%
Group 2	(3, 5]	65	24.53%
Group 3	(5, 6]	54	20.38%
Group 4	(6, 8]	59	22.26%
Group 5	(8, +∞)	40	15.09%
Total	--	265	100%

Table 10.36 summarizes the RMSEs based on the difference of these groups. Figure 10.11 graphically illustrates the comparisons of these RMSEs.

Table 10.36 Summary of RMSEs from the Difference for Groups Based on the Weighted Average Slope Change

Group	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
Group 1	10.11	10.11	10.11	10.04	10.03	10.16	10.16
Group 2	24.45	23.17	23.20	30.29	26.23	23.91	23.99
Group 3	31.09	38.89	39.40	53.71	42.86	40.11	40.41
Group 4	27.37	40.48	41.21	49.55	42.45	44.47	44.78
Group 5	29.30	43.55	44.05	35.45	36.80	47.19	47.35

Based on Figure 10.11 and Table 10.36, it is observed that the overall trend is such that groups that have larger values in the weighted average slope changes have higher RMSEs (lower accuracies). When the weighted average slope reaches 5% (Group 3, Group 4, and Group 5), there is very small variations in the RMSEs with the change of the weighted average slope changes. With the use LIDAR point data and LIDAR 50-ft DEMs, the RMSEs of Group 4 and Group 5 (larger values in the weighted average slope changes) are even a littler smaller than Group 3 (smaller values in the weighted average slope change). In other words, when the weighted

average slope change reaches 5%, the effects on the accuracy of the predicted 3-D distances from the increase in the weighted average slope changed are almost nonexistent and could be ignored.

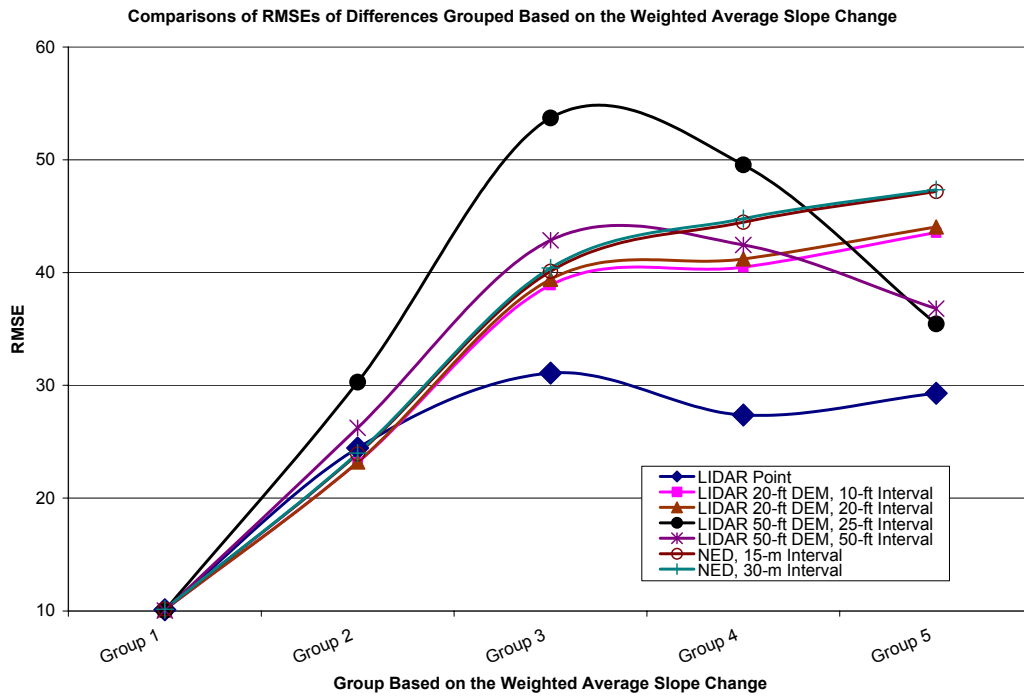


Figure 10.11 Comparison of RMSEs of the Difference for Groups Based on the Weighted Average Slope Change

Table 10.37 summarizes the RMSEs based on the proportional difference of these groups. Figure 10.12 graphically illustrates the comparisons of these RMSEs.

Table 10.37 Summary of RMSEs from the Proportional Difference for Groups Based on the Weighted Average Slope Change

Group	LIDAR Point Data	LIDAR 20-ft DEM		LIDAR 50-ft DEM		NED	
		10-ft Interval	20-ft Interval	25-ft Interval	50-ft Interval	15-m Interval	30-m Interval
Group 1	104.36	105.97	105.93	104.74	104.64	104.46	104.46
Group 2	44.33	44.71	44.69	44.56	44.57	44.45	44.45
Group 3	14.12	14.15	14.15	14.15	14.11	14.13	14.13
Group 4	14.54	14.78	14.79	14.86	14.81	14.82	14.82
Group 5	16.83	19.60	19.60	19.39	19.44	19.73	19.73

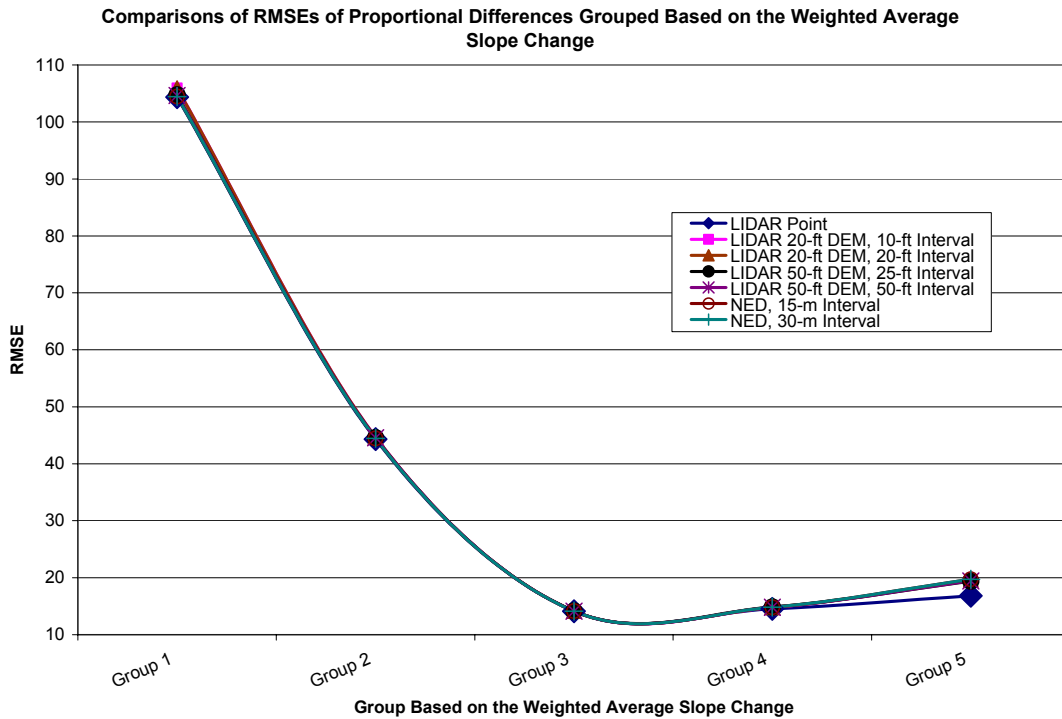


Figure 10.12 Comparison of RMSEs of the Proportional Difference for Groups Based on the Weighted Average Slope Change

Based on Figure 10.12 and Table 10.37, it was observed that the seven trend lines of the seven different combinations of elevation datasets and model parameters were clustered together. The overall trend was such that groups that had larger values in the weighted average slope changes had lower RMSEs (higher accuracies). When the weighted average slope changes were lower than 6% (Group 1, Group 2, and Group 3), there were significant drops in the RMSEs with the increases in the weighted average slope changes. When the weighted average slope changes reached 5% (Group 3, Group 4, and Group 5), there was a very small variation in the RMSEs with the changes in the weighted average slope changes. However, the trend became such that the RMSEs increased a little with the increases in the weighted average slope changes.

It was concluded that the weighted average slope change could be considered as a significant factor. From the aspect of the difference, there was a positive association between the error and the weighted average slope change. From the aspect of the proportional difference, there was a negative association between the error and

the weighted average slope change. Furthermore, when the weighted average slope change reached 5%, the effects on the accuracy of the predicted 3-D distances from the increase in the weighted average slope change were almost nonexistent and could be ignored.

10.4.2.6 Number of 3-D Points and Average Density of 3-D Points

This subsection evaluates if the factors of the number of 3-D points and the average density of 3-D points to model a linear object in a 3-D space are significant factors to the accuracy of the predicted 3-D distances using LIDAR point data when compared to the DMI measured distances. The effects of these two factors on the accuracy of the predicted 3-D distances using LIDAR point data are examined in detail.

Tables 10.38 and 10.39 show the grouping of FTSegs based on the number of 3-D points being used and the average density of the 3-D points, respectively.

Table 10.38 Illustration of Grouping Based on the Number of 3-D Points

Group Name	Number of Points	Number of FTSegs	Percentage
Group 1	(0, 10]	54	20.38%
Group 2	(10, 100]	23	8.68%
Group 3	(100, 500]	54	20.38%
Group 4	(500, 1,000]	26	9.81%
Group 5	(1,000, 2,000]	42	15.85%
Group 6	(2,000, 3,000]	33	12.45%
Group 7	(3,000, +∞)	33	12.45%
Total	--	265	100%

Table 10.39 Illustration of Grouping Based on the Average Density of 3-D Points

Group Name	Average Density of 3-D Points	Number of FTSegs	Percentage
Group 1	[7.5, 10]	21	7.92%
Group 2	(10, 11]	65	24.53%
Group 3	(11, 12]	53	20.00%
Group 4	(12, 13]	43	16.23%
Group 5	(13, 20]	41	15.47%
Group 6	(20, +∞)	42	15.85%
Total	--	265	100%

Table 10.40 summarizes the RMSEs of both the difference and the proportional difference between the predicted 3-D distances using LIDAR point data and the DMI measured distances for the groups based the number of 3-D points. Figure 10.13 graphically illustrates the comparisons.

Table 10.40 Summary of the RMSEs for Groups Based on the Number of 3-D Points

Group	RMSEs of the Difference	RMSEs of the Proportional Difference
Group 1	6.90	109.88
Group 2	16.12	26.75
Group 3	27.79	9.74
Group 4	27.36	3.67
Group 5	30.35	1.85
Group 6	29.52	1.06
Group 7	33.60	0.80

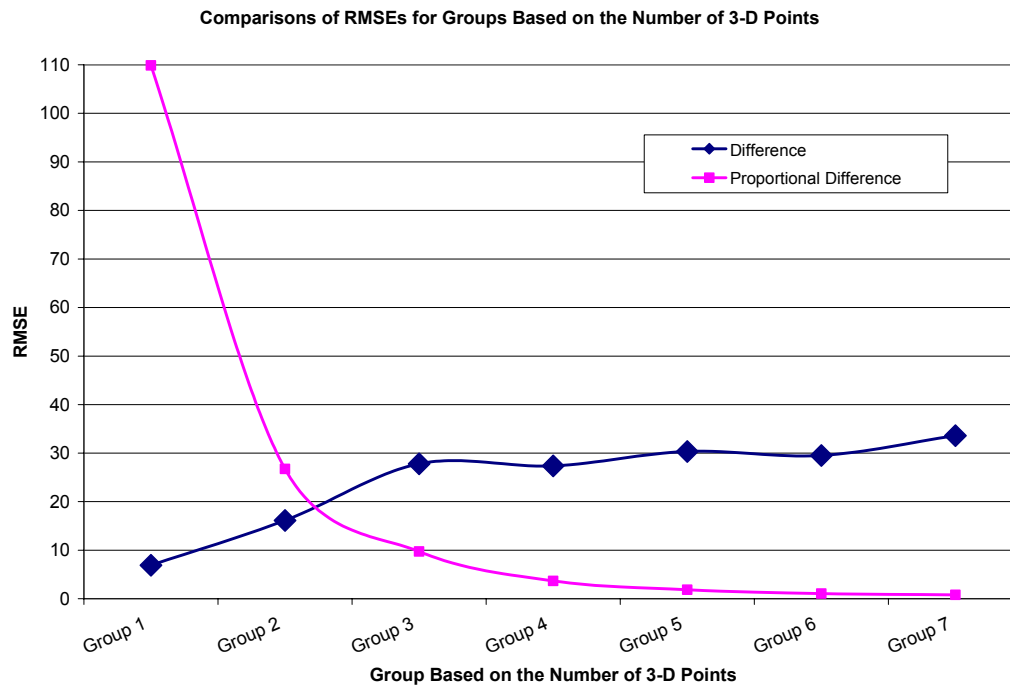


Figure 10.13 Comparison of RMSEs of the Difference and Proportional Difference for Groups Based on the Number of 3-D Points

It is observed based on Figure 10.13 that from the aspect of the difference, the overall trend is such that the RMSEs increase with the increase in the number of 3-D points. Groups that include FTSEgs with larger numbers of 3-D points have higher RMSEs (lower accuracies). There is small variation among the RMSEs of Group 3, Group 4, Group 5, and Group 6. In other words, when the number of 3-D points being used to model

a linear object in a 3-D space reaches 500, from the aspect of the difference, the effects on the accuracy of the predicted 3-D distances using LIDAR point data when compared to DMI measured distances are almost nonexistent and could be ignored.

From the aspect of the proportional difference, the overall trend is the reverse to the overall trend observed from the aspect of the difference. The RMSEs decrease with the increase in the number of 3-D points. Groups that include FTSEgs with larger numbers of 3-D points have lower RMSEs (higher accuracies). There is a small variation among the RMSEs of Group 4, Group 5, and Group 6. In other words, when the number of 3-D points being used to model a linear object in a 3-D space reaches 1,000, from the aspect of the proportional difference, the effects on the accuracy of the predicted 3-D distances using LIDAR point data when compared to DMI measured distances are almost nonexistent and could be ignored.

Table 10.41 summarizes the RMSEs of both the difference and the proportional difference between the predicted 3-D distances using LIDAR point data and the DMI measured distances for the groups based the average density of 3-D points. Figure 10.14 graphically illustrates the comparisons.

Table 10.41 Summary of the RMSEs for Groups Based on the Average Density of 3-D Points

Group	RMSEs of the Difference	RMSEs of the Proportional Difference
Group 1	11.92	109.56
Group 2	26.56	8.63
Group 3	32.23	45.36
Group 4	30.15	38.48
Group 5	26.67	34.57
Group 6	7.34	68.23

Based on Figure 10.14, it is observed that from the aspect of the difference, the RMSEs of Group 2, Group 3, Group 4, and Group 5 are close to each other, but higher than the RMSEs of Group 1 and Group 6. In other words, when the average density of the 3-D points is very high or very low (two extremes), the accuracy of the predicted 3-D distances are relatively better. From the aspect of the proportional difference, there is no clear pattern that could be observed. Examining the raw data in detail revealed that there are a few of FTSEgs that are short (less than 100 feet) with an about average density of 3-D points (around 10 feet), a small value of the

difference (around 20 feet), but a huge value of the proportional difference (around 200 ft/1,000 ft). RMSEs are sensitive to these outliers and consequently, the pattern of the RMSEs versus the average density is obscured.

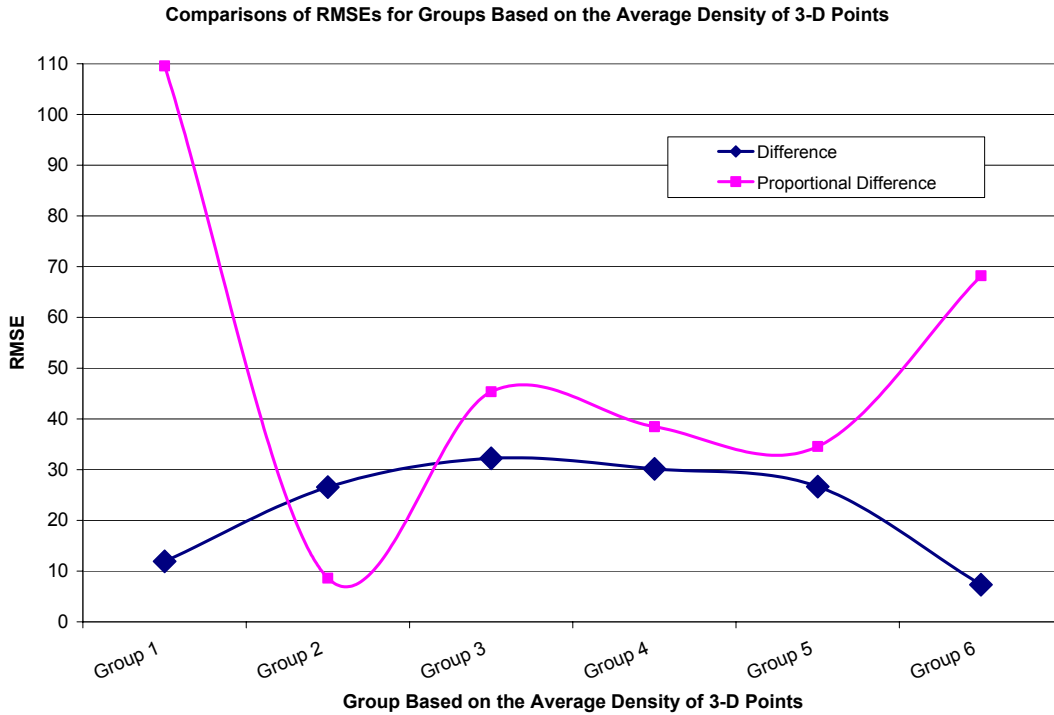


Figure 10.14 Comparison of RMSEs of the Difference and Proportional Difference for Groups Based on the Average Density of 3-D Points

10.5 CONCLUSIONS AND LIMITATIONS

In this chapter the significance of a factor to the accuracy of the predicted 3-D distance when compared to the DMI measured distance is evaluated using two methods. The first method calculates the sample correlation coefficient and the sample coefficient of determination between the factor and the error of the predicted 3-D distance. The second method takes a grouping and comparison approach. All FTSegs are grouped based on the values of the factor under consideration. The errors of these groups are quantified as RMSEs. These RMSEs are compared and the effects on the accuracy of the predicted 3-D distance from the factor under evaluation are evaluated. The factors under consideration in this case study include the distance, the average slope, the

weighted average slope, the average slope change, the weighted average slope change, and the number of 3-D points and the average density of the 3-D points (only applied to the case of using LIDAR point data).

Based on the results of the first method, following conclusions are obtained when considering the magnitudes of errors only.

(1) The factor of distance is significant to the accuracy of the predicted 3-D distance when compared to the DMI measured distance. There is a positive linear association between the error of the predicted 3-D distance and the factor of the distance from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the longer the distance of the FTSeg, the bigger the error it has and the smaller the proportional effort it has.

(2) The factor of the average slope is significant to the accuracy of the predicted 3-D distance when compared to the DMI measured distance. There is a positive linear association between the error of the predicted 3-D distance and the factor of the average slope from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the larger the average slope of the FTSeg, the bigger the error it has and the smaller the proportional effort it has.

(3) The factor of the weighted average slope is significant to the accuracy of the predicted 3-D distance when compared to the DMI measured distance. There is a positive linear association between the error of the predicted 3-D distance and the factor of the weighted average slope from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the larger the weighted average slope of the FTSeg, the bigger the error it has and the smaller the proportional effort it has.

(4) The factor of average slope change is significant to the accuracy of the predicted 3-D distance when compared to the DMI measured distance. There is a positive linear association between the error of the predicted 3-D distance and the factor of the average slope change from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the

larger the average slope change of the FTSeg, the bigger the error it has and the smaller the proportional effort it has.

(5) The factor of the weighted average slope change is significant to the accuracy of the predicted 3-D distance when compared to the DMI measured distance. There is a positive linear association between the error of the predicted 3-D distance and the factor of the weighted average slope change from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the larger the weighted average slope change of the FTSeg, the bigger the error it has and the smaller the proportional effort it has.

(6) The factor of the number of 3-D points is significant to the accuracy of the predicted 3-D distance using LDIAR point data and the snapping approach when compared to the DMI measured distance. There is a positive linear association between the error of the predicted 3-D distance and the factor of the number of 3-D points from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the more the number of 3-D points being used to represent the FTSeg in a 3-D space, the bigger the error it has and the smaller the proportional effort it has.

(7) The factor of the average density of 3-D points is significant to the accuracy of the predicted 3-D distance using LDIAR point data and the snapping approach when compared to the DMI measured distance. There is a positive linear association between the error of the predicted 3-D distance and the factor of the average density of 3-D points from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the higher the density (the smaller the average planimetric distance between two neighboring 3-D points along the linear object) of the 3-D points being used to represent the FTSeg in a 3-D space, the bigger the error it has and the smaller the proportional effort it has.

The conclusions of all the slope factors including the average slope, the weighted average slope, the average slope change, and the weighted average slope change are the same. All these factors are used to capture the

vertical geometric complexity of 3-D linear objects. It leads to the belief that there is a consistency among these factors in describing the vertical geometric complexity of the 3-D lines. The conclusions of the factors of the number of 3-D points and the distance reveal that even though LIDAR points are not uniformly distributed, a longer FTSeg would have more 3-D points on it than a shorter FTSeg.

The observations from the second method regarding these factors confirm the above conclusions. Furthermore, it is concluded that

(1) With the factor of the distance, when the distance reaches 5,000 ft, from the aspect of the proportional difference, the effects on the error of the predicted 3-D distance when compared to the DMI measured distance from the increase of the distance are almost nonexistent and therefore, could be ignored.

(2) With the factors of the average slope change and the weighted average slope change, when the value of the average slope change or the weighted average slope change reaches 5%, the effects on the error of the predicted 3-D distance when compared to the DMI measured distance from the increase of the average slope change or the weighted average slope change are almost nonexistent and therefore, could be ignored.

(3) With the factor of the number of 3-D points being used to model a linear object in a 3-D space, when 500 points are reached from the aspect of the difference and when 1,000 points are reached from the aspect of the proportional difference, the effects on the error of the predicted 3-D distance when compared to the DMI measured distance from the increase of the number of 3-D points being used are almost nonexistent and therefore, could be ignored.

It is also realized that there are limitations in using those two methods to determine the significance of the factors to and evaluate their effects on the accuracy of the predicted 3-D distance when compared to the DMI measured distance.

The correlation coefficient is a measure of the strength of linear association between two variables (Rao 1998). The absence of correlation ($r_{XY} = 0$) does not mean that the variables are not associated, but just suggests that

the variables are not linearly associated with each other. This limitation is considered in this case study. As a supplementary, the second method (the grouping and comparison method) is used to evaluate the effects on the accuracy of the predicted 3-D distance from the factors under consideration. However, it is also recognized that the grouping is subjective and different results or even contradictory results are be obtained if the same sample data are grouped in different ways. In addition, this grouping and comparison method might lead to no clear observation regarding the association between two variables, even though there is a certain association, as illustrated in the case of using this method to evaluate the effects on the accuracy from the factor of the average density of 3-D points.

Another limitation of using of using the sample correlation coefficients and the sample coefficients of determination to determine the significance of a factor is the lack of a reference value (the baseline value), based on which it could be said that a factor is significant or not. The sample correlation coefficient tells if two variables are positively linearly associated (positive values) or negatively linearly associated (negative values). The sample coefficient of determination tells the strength of the linear association. It ranges from 0 to 1. The closer to 1 the sample coefficient of determination is, the stronger the linear association between two variables. However, there is no such value that has been agreed upon that if the sample coefficient of determination is greater than or equal to it, the two variables being evaluated could be considered as the significant factor to each other. In this case study, it is determined that if the magnitude of the sample correlation coefficient exceeds 0.3 (the corresponding value of the sample coefficient of determination would be 0.09), the factor under evaluation would be considered as a significant factor to the accuracy of the predicted 3-D distances. It is obvious that if another value were to be chosen, the observations and conclusions would be different.

11 USE OF THE MODEL TO ASSESS HIGHWAY FLOODING

This chapter describes the application of the 3-D road centerline in flood mapping, more specifically, the use of the 3-D road centerline in determinations of road segments under flood. Two models are developed and implemented in this application. The first model is the model that predicts the flood extent given a flood level. The second model is the model that determines the road segments under flood using the results of the first model. The developed models are tested, the results are presented and analyzed, and conclusions are obtained.

11.1 INTRODUCTION

Flooding and flash flooding poses serious hazards to human populations in many parts of the world. According to FEMA, floods are the second most common and widespread of all natural disasters (Noman et al. 2001). Take North Carolina as an example, North Carolina faces extreme hazards and consequences from hurricanes and flooding. Since 1989, there have been 14 federally declared disasters in North Carolina. Damage from Hurricane Floyd alone has reached \$3.5 billion. Hurricane Floyd destroyed 4,117 uninsured and under-insured homes (NCCTSFMP 2000).

Under a flooding scenario, it is critical to identify road segments that are flooded and road segments that are not flooded so that rescue and response routes could be determined and rescue personnel and supplies could be distributed promptly and in a timely manner. However, currently there is no such information system that would predict road segments under flood given a flood level and provide the road flood information to rescue activities. It is not unusual for a rescue team to go along a road for a while before they realize that part of that road is under flood and becomes a barrier. This will cause delays in rescue. These delays are very expensive. They cost not only property damages, but also lives. Therefore, it is extremely important to establish such a prediction system to help the rescue activities under a flooding scenario.

Most states in the US collect and maintain flood maps. It is quite natural to turn to these flood maps for help in determining rescue and response routes under a flooding scenario. This idea has been denied for a few reasons as below. First, flood maps categorize areas into 100-year flood zones, 50-year flood zones, 20-year flood

zones, etc. They predict the flood risk for areas rather than anything else. Therefore, it is unsuitable to use flood maps to identify road segments under flood. Second, when a flood occurs, road segments in the flooded area might not be flooded at all or just partially flooded and could be used as the rescue routes.

It is obvious that in order to determine flooded road segments, two information sources are required, the flood depth and extent information and the 3-D road data. As stated earlier, this research focused on modeling linear objects in a 3-D space and in the case study, road centerlines were the linear objects that were modeled in a 3-D space using the 3-D model developed in this research. The 3-D road centerline data from the case study could be readily utilized in determining flood road segments with the flood depth and extent information. The major goal of the study presented in this chapter is to develop and test models and algorithms to deploy the 3-D road centerline data to determine flooded road segments. By doing so, the flooded road segments could be determined in a timely manner to help determining rescue routes.

11.2 FLOOD EXTENT AND DEPTH DETERMINATION AND FLOODED ROAD SEGMENT IDENTIFICATION

This section describes the problem of determining flooded road segments in detail. As stated earlier, there were two tasks in order to determine flooded road segment. First, the flood extent and depth need to be predicted. Second, based on the flood extent and depth information, the 3-D road centerline data could be used to determine flooded road segments. In this section, the scenarios of water flooding and its impacts on roads are described in detail.

A water body becomes a water body only because part of the area is lower than the surrounding areas as in the cases of lakes, rivers, and streams. It is obvious that all these water bodies occupy certain areas. In the sense of spatial modeling, these water bodies could be modeled as polygons. However, similar to the way that is being used to model roads, rivers and streams could be modeled as linear objects using polylines. In addition, a polygon can be modeled as polylines connected together as illustrated in Figure 11.1. In case a river is quite

wide, which makes it unreasonable to be modeled as a singular polyline, two polylines running roughly parallel to each other could be used as illustrated in Figure 11.1.

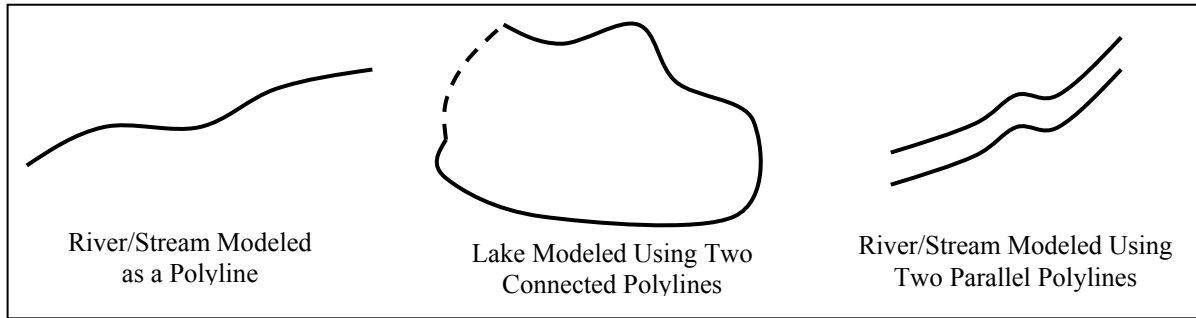


Figure 11.1 Modeling Water Bodies Using Polylines

In order to predict road segments under flood, it is necessary to examine the situation in detail. In reality, this problem consists of two parts: first, the water body floods (water level is higher than normal) and water flows to the surrounding areas; second, the flowing water hits roads and floods road segments. The extent to which water flows to the surrounding areas depends on the surface or elevation changes of the surrounding areas. Determining which road segments are under flooding depends on the elevations of the road and the flood level. Figure 11.2 provides a cross-sectional view of how water body floods and water runs to the surrounding areas. In Figure 11.2, the water level increases from the normal water level to the flooded water level when flooding. It is obvious that there are surrounding areas that were not under water before flooding but are under water after flooding (areas having elevations higher than the blue line but lower than the red line). The dash line represents another flood level, in which area A is not flooded even though its elevation is lower than the yellow line, because area B becomes a natural barrier for the water to flood into area A.

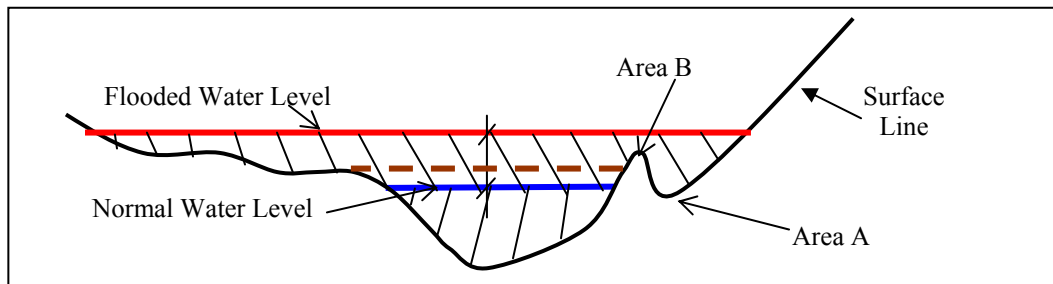


Figure 11.2 Cross-sectional View of Flooding

Figure 11.3 provides a profile of a road segment with several portions under flood. As illustrated in Figure 11.3, part of the road segment may not be flooded even though it is in the flooded area because a road is a three-dimensional object with elevations changing along it.

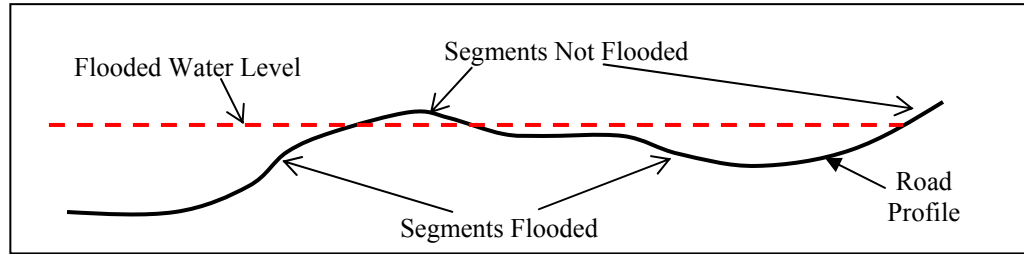


Figure 11.3 Illustration of Flooded and Not Flooded Road Segments in the Flooded Area

According to Figures 11.2 and 11.3, in order to predict flooded road segments, at least two models need to be developed. The first one will predict the extent to which water flows when flooding based on the elevations of the surrounding areas and the flooded water level. The second will identify flooded road segments for roads running through the flooded areas.

11.3 MODELS AND ALGORITHMS

This section describes the conceptual models, algorithms, and computational models developed in this study to predict flood extent with flood depth information and to determine flooded road segments.

11.3.1 Conceptual Models

This subsection describes the conceptual models for flood extent prediction and flooded road segments identification.

11.3.1.1 The Flood Extent Prediction Model

Flood extent prediction, or floodplain delineation is the process of determining inundation extent and depth by comparing river water levels with ground surface elevations (Noman et al. 2001).

The traditional method of floodplain delineation is to work on a topographic map (Noman et al. 2001). This method consists of marking the water levels from the observations of river stages the topographic map, extending the water level until impeded by the higher elevation contour, tracing the contour lines to delineate the floodplain, and manually producing a flood extent map. In an automated system, the procedure is still the same but the topographic map is replaced with a digital terrain model (DTM) and the hydraulic model results either replace or supplement the observations of river stages to obtain water levels (Noman et al. 2001). Based on the water level points, a water level surface is created and compared to the DTM to produce an automated flood depth and extent map. A fairly recent development in the field of floodplain delineation has been the coupling of hydraulic modeling and GISs (Tate et al. 2001). With GISs, the stream cross-section parameters are extracted from a digital terrain model and imported into a hydraulic model to be processed to produce results regarding the floodplains (Tate et al. 2001).

In this study, an automated procedure is taken to obtain floodplains and depth information. But this automated procedure is different from the one mentioned above in the way the water levels are obtained. The model developed in this study does not create a water surface level in the raster format and compare the water surface level with the DTM to delineate floodplains. Rather, the floodplains are directly delineated from the high-resolution DTM (in this study, LIDAR 20-ft DEMs) using a simplified, but practical approach, which is developed based on the actual mechanism of flooding. The resulting floodplains are represented by numerous small polygons. Each polygon is associated with the flood depth information.

It is realized that water level (the elevation of the water surface) varies throughout a water body as in the case of river. It is infeasible to assume a uniform water level in a relatively large area. Rather than obtaining water levels at different locations via direct measurements or using the results from a hydraulic model, a surrogate is taken to establish the flood extent prediction model by assuming that at a given point on the a polyline being used to model a water body, the water level equals the elevation of that point. This elevation comes from elevation data such as LIDAR DEMs (this assumption is confirmed by the communications with the personnel from the North Carolina Floodplain Mapping Program in the case of using LIDAR data in North Carolina).

Figure 11.4 illustrates how the flooded area is predicted with a given flood level (the increase of water level when flooding from the normal water level). In Figure 11.4, the river is represented with a polyline. A few points that are uniformly distributed along the polyline (with the same interval) starting from the start point of that polyline are obtained. In addition, the start and end points of that polyline are obtained. At each of these points, a line normal to the river polyline is constructed. For example, at point A, a normal line L_1 is constructed. Along this line L_1 , two points (A_1 and A_2) are identified. These two points are on two different sides of the river polyline and have elevations equaling the elevation at point A plus the given flood. For example, if point A has an elevation of 105 ft and the given flood level is 2 ft, points A_1 and A_2 will have the same elevation of 107 ft. After points A_1 and A_2 are identified, a straight line between these two points is constructed. A set of straight lines will be produced by performing this process with every point and consequently a set of neighboring polygons are constructed by closing neighboring lines. For example, two polygons (polygon $A_1A_2B_2B_1$ and polygon $B_1B_2C_2C_1$) can be constructed by closing lines A_1A_2 with B_1B_2 and by closing lines B_1B_2 with C_1C_2 . Each of these polygons will have a water level, which takes the average of elevations at points A and B plus the given flood level.

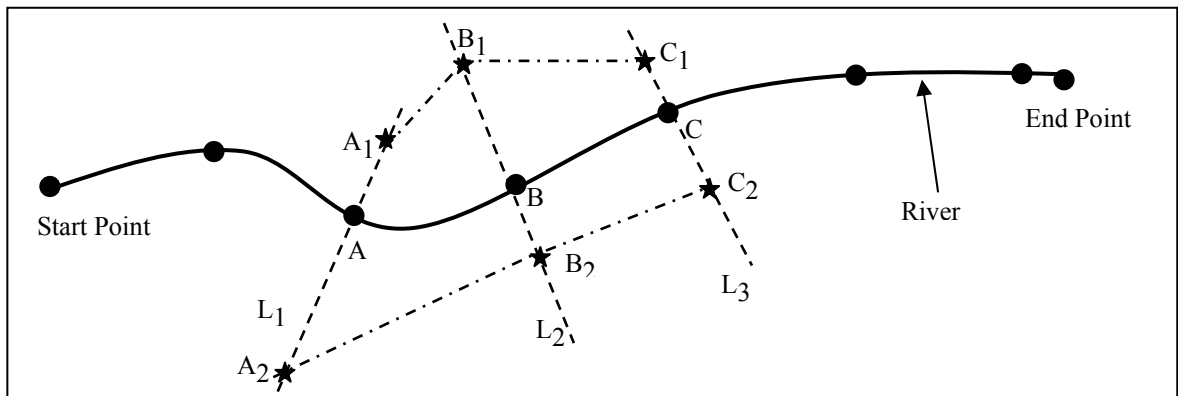


Figure 11.4 Conceptual Model for Predicting Flood Extent

According to this flood extent prediction model, the issue of continuously varying water level along a river water body is addressed. In case the water body is a lake that is being described using connected polylines or the water body is still a river but being described using approximately parallel polylines, the polygons surrounding the polylines describing the water body will still be constructed. However, only the parts that are

located at the outer direction will be taken. For example, only the polygons A_1ABB_1 and B_1BCC_1 will be used, or only the polygons AA_2B_2B and BB_2C_2C will be used, depending on the spatial relationship between the polylines being used to describe a water body and the actual location of the water body.

In addition, there are two constraints from the reality. The first constraint is based on the cause of a water body, i.e. the elevation of the place where the water body is located is lower than the elevations of the surrounding areas. The second constraint comes from the characteristic of a water body, i.e. water always runs from higher levels to lower levels, or along the water running direction, the water surface levels are always decreasing.

11.3.1.2 The Flooded Road Segment Identification Model

As stated earlier, if a road segment is within a flooded area, it does not mean that the whole road segment is under flood. In this study, the flooded area is predicted by using the flood extent prediction model described in the previous section. This flood extent prediction model uses digital elevation data (LIDAR 20-ft DEMs, more specifically) to determine the flood extent. LIDAR DEMs are in the format of grid files consisting of numerous uniform rectangular cells. Each cell possesses a single value (representative elevation). In the flood extent prediction model, it is this representative elevation that is used to determine if a cell is under flood or not. However, it is obvious that the elevations of points within a single cell still vary and therefore, for a road segment in the flooded area, it is very possible that only parts of that road segment is flooded as illustrated in Figure 11.5. It is the main goal of our flooded road segment identification model to identify those parts and differentiate them from the rest of the road segment.

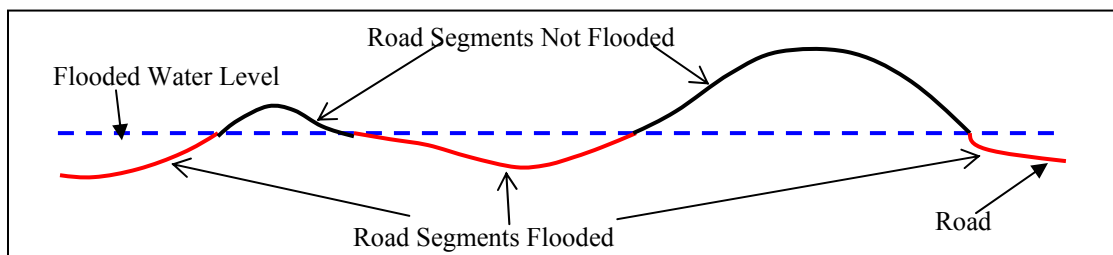


Figure 11.5 Conceptual Model for Identifying Flooded Road Segments

11.3.2 Computational Model

After the conceptual models are established to address the problems we are facing, with several assumptions to facilitate problem solving, computational models and algorithms will need to be developed. This section describes the developing environment, data sources, and algorithms. Developed codes are attached to this dissertation as appendices.

11.3.2.1 Development Environment

The development environment chosen in this research project is ArcGIS from Environmental Systems and Research Institute (ESRI) and Visual Basic from Microsoft, which is embedded in ArcGIS. This decision is made based on the source data format, the software availability, and the researchers' knowledge and skills in programming.

The version being used is ArcGIS 8.3. It consists of ArcMap, ArcCatalog, ArcTools, etc. ArcMap, ArcCatalog, and ArcTools are used a lot for data management, visual inspection, and analysis. The customized programs are developed using Visual Basic and ArcObjects.

11.3.2.2 Data Sources

Three major data sets are used in this project. They are water body data, LIDAR DEMs, and 3-D Road data.

The water body data are in the format of a line shapefile. They are provided by NCDOT. The LIDAR DEMs are downloaded from www.ncfloodmaps.com, which is undertaking the flooding mapping program in North Carolina to obtain elevation data with a high accuracy and consequently, to produce flood maps statewide. The downloaded LIDAR DEMs have a resolution of 20 ft. The road data are 3-D road centerline data obtained in this research. Each road segment is associated with a set of 3-D points (with X/Y/Z-coordinates) along the road segment. For these 3-D points, their X/Y-coordinates come from NCDOT LRS road data and their Z-

coordinates come from 3-D LIDAR points via a snapping approach. The average density of these 3-D points along road segments is approximately 15 ft.

To summarize, the water body data are two-dimensional line data. The road data are three-dimensional data, which are represented as straight line segments connecting 3-D points along road segments. The LIDAR DEMs are the elevation data being used to describe the surface.

11.3.2.3 Pre-Processing

Before these data sets could be used to perform analysis, some fundamental preprocessing measures are taken. The LIDAR DEMs are organized as individual grid files. Each of these grid files covers a rectangular area with the size of 10,000ft * 10,000ft. After downloading these individual files, they are merged together to generate a single but huge file.

The coordinate system being used is the State Plane Coordinate System in feet. The horizontal datum is NAD83. The vertical datum is North American Vertical Datum of 1988 (NAVD88) in feet. All datasets are reprojected into the same coordinate system. After preprocessing, these three data sets are ready for the flood extent analysis and flood road segment analysis.

11.3.2.4 Algorithms for Flood Extent and Depth Prediction

This section describes the algorithms developed to predict the flood extent and depth using an elevation dataset and a given flood level. Figures 11.6 and 11.7 describe the algorithm being used to predict the flood extent with given flood level, water body polyline, and elevation from LIDAR DEMs.

Figure 11.6 illustrates the algorithm used to obtain points along a polyline. In Figure 11.6, from the water body data (a line shapefile), a feature is obtained. This feature is a polyline. By identifying the start and end nodes of this polyline and identifying points along the polyline, which are uniformly distributed at the specified interval from the start point, a set of points are obtained for a polyline. Repeating this procedure for all features in the

water body data results in sets of points. Each set is associated with a polyline feature describing the water body. These points are distributed along the polyline features uniformly at the specified interval.

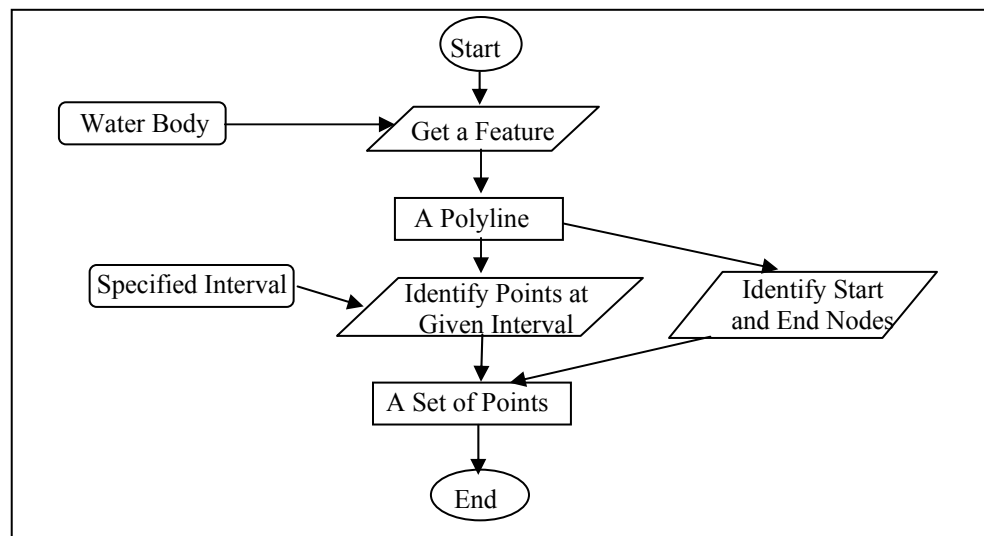


Figure 11.6 Obtaining Points along A Polyline

Figure 11.7 describes the procedure of constructing normal lines to the water body polyline at those points identified in Figure 11.6. First, a set of points is taken and ordered in ascending according to their distances to the start node of the corresponding polyline. For each of these points, a line normal to the polyline is constructed based on the curvature of the polyline. The elevation at such a point is determined based on the LIDAR DEMs. Adding together this elevation with the given flood level leads to the target elevation. Two points having the target elevation on the normal line are identified.

For example, assuming the original point is point A with elevation E_1 , two new points A1 and A2 will be identified on the normal line. Both of these two points have the target elevation, which equals the sum of the elevation at point A and the given food level. Furthermore, there will be no points between A and A1 and between A and A2 along the normal line, which have elevations higher than the target elevation. In other words, A1 should be the first point that reaches the target elevation starting from A in one direction. Similarly, A2 should be the first point that reaches the target elevation starting from A in another direction. After A1 and A2 are identified, a straight line connecting A1 and A2 can be constructed and later on, this line will be used

together with another such line being constructed at the neighboring points of point A to construct polygons that represent the flood extent.

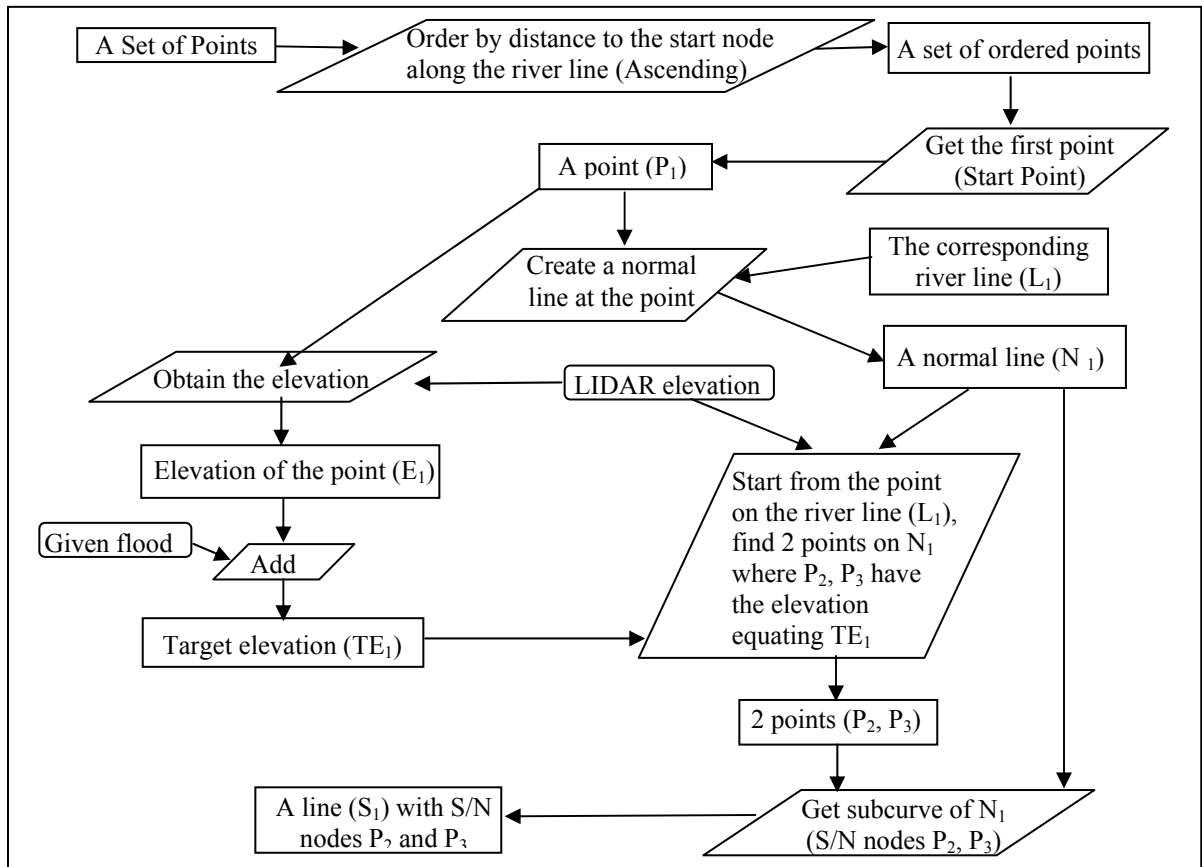


Figure 11.7 The Procedure of Predicting Flood Extent

The critical part of this procedure is how the two points A1 and A2 are identified based on the elevation data. Figure 11.8 illustrates this identification procedure in detail. In Figure 11.8, line L1 is the normal line constructed at point A. Point A1 is the point that has the target elevation and is the point need to be identified. The procedure is described as below:

- 1) The elevation of the cell in which point A is located is obtained.
- 2) Elevation at point A is added with given flood level to obtain the target elevation.
- 3) Starting from point A along the normal line, point T1 is identified. The distance between A and T1 equals the resolution of the elevation grid file (the height or width of a cell).

- 4) The elevation of the cell containing point T1 is obtained and compared to the target elevation. If this elevation is lower than the target elevation, another point T2 along the normal line is obtained. The distance between T2 and T1 equals the resolution of the elevation grid file. The elevation of T2 is compared to the target elevation. This step is repeated until the first point having elevation higher than the target elevation is obtained.
- 5) Assuming the first point having elevation higher than the target elevation is T3, it indicates that point T2 has an elevation lower than the target elevation. Taking a linear interpolation approach to identify a point A1 between T3 and T2 on the normal line, which has its elevation equaling to the target elevation. This point A1 is the point need to be identified.
- 6) Repeat steps 3) to 5) to identify point A2 on the other side.

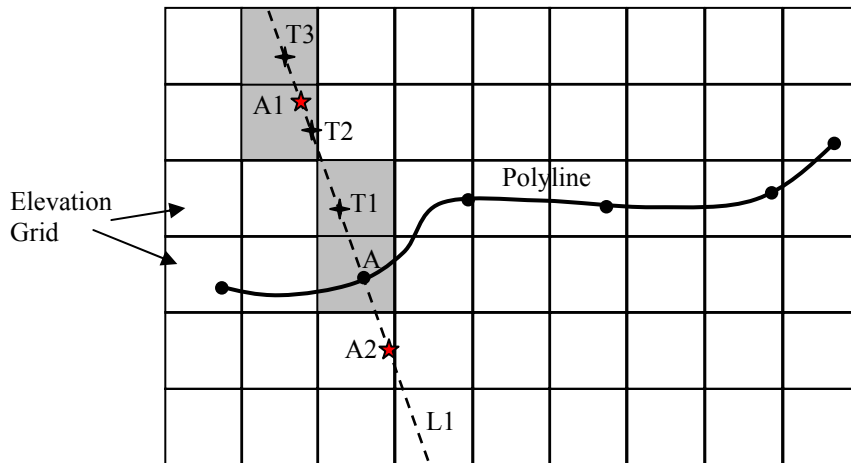


Figure 11.8 Identifying Points with Target Elevation

11.3.2.5 Enforcement of Constraints

The algorithm illustrated in Figure 11.7 utilizes the results from the execution of the algorithm illustrated in Figure 11.6, i.e. sets of points along the polyline. The two constraints stated earlier are enforced by examining the elevations of the resulting points from the algorithm illustrated in Figure 11.6. These resulting points must be examined before the execution of the algorithm illustrated in Figure 11.7 to assure the constraints are enforced.

The first constraint states that the elevation of a point on the polyline representing the water body must be lower than the surrounding areas. This constraint is enforced in this study by comparing the elevation value of the cell (from the grid representing the surface elevations) containing that particular point with the elevations of the eight neighboring cells. If there is a neighboring cell whose elevation is lower, this particular point is discarded.

The second constraint states that water is running in the direction of high to low. Figure 11.9 illustrates this situation. Theoretically, if water runs from point A1 to point A9, then the water levels of points A1 to A9 should appear as scenario 1. If water runs from point A9 to point A1, then the water levels of points A1 to A9 should appear as scenario 2.

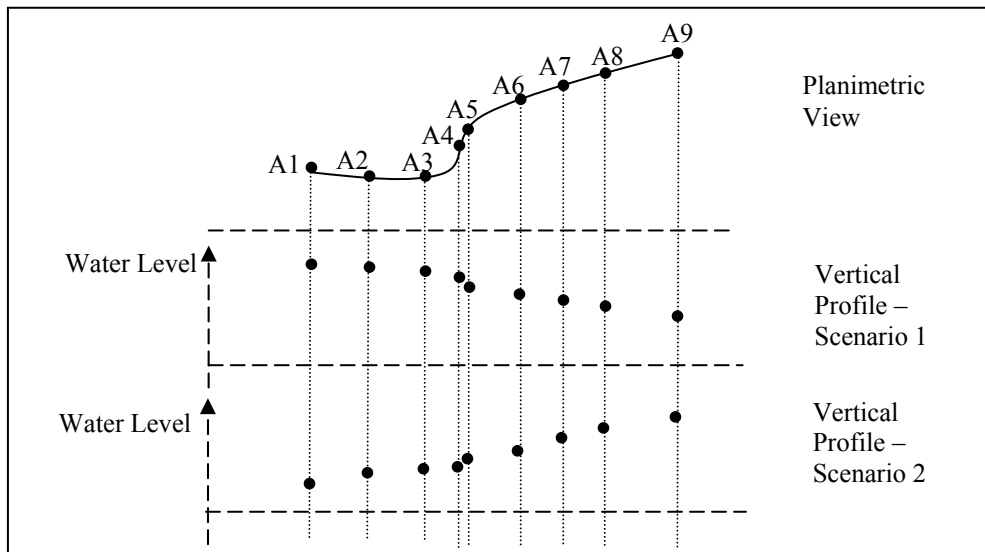


Figure 11.9 Illustration of Water Flow

Examining the points along the polyline after enforcing the first constraint reveals that it is not always the case that these points would follow the water flow rule. In order to ensure that water is always running downward, the elevations of the points along the polyline are adjusted. This adjustment could be implemented with two different algorithms. Both of them need two constants, an error tolerance (in the units of ft/ft) and a maximum drop (in the units of ft/ft). Figure 11.10 illustrates the concepts of error tolerance and maximum drop.

In Figure 11.10, a straight line is used to represent the vertical profile of water. The horizontal axis represents the planimetric distance to the start point of the polyline, to which the points belong. The vertical axis represents the water level.

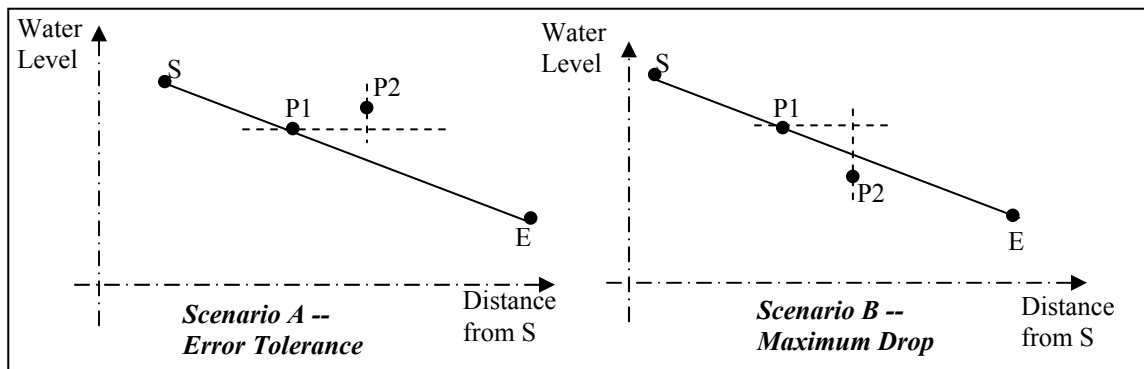


Figure 11.10 Error Tolerance and Maximum Drop

It assumes that water runs from the start point S to the end point E. Points P1 and P2 are two neighboring points that are used to illustrate two scenarios of errors. Ideally, all these points are on the straight line representing the vertical water profile. Scenario A illustrates that point P2 is higher than point P1. Scenario B illustrates that point P2 is lower than point P1. In case of scenario A, the error tolerance is applied. Assuming the planimetric distance from point P1 to S is $D1$, the planimetric distance from point P2 to S is $D2$, point P1 has a water level of $W1$, point P2 has a water level of $W2$, $W2$ is higher than $W1$, and an error tolerance of $E1$, if $W2 \leq (W1 + E1 * (D2 - D1))$, it is within the given error tolerance and is deemed to be acceptable. On the other hand, if $W2 > (W1 + E1 * (D2 - D1))$, it exceeds the error tolerance and countermeasures will be taken as described later. In case of scenario B, the maximum drop from point P1 to point P2 is checked. If $(W1 - W2)/(D2 - D1) \leq \text{Maximum Drop (MD)}$, it is acceptable. Otherwise, it is deemed that the maximum drop is exceeded and countermeasures will have to be taken as described later.

Two different algorithms are taken to deal with neighboring points either exceeding the given error tolerance or the maximum drop.

- 1) Algorithm 1

In case the error tolerance is exceeded, algorithm 1 assigns the water level of P1 to point P2. In case the maximum drop is exceeded, algorithm 1 assigns point P2 with a water level that is equivalent to (water level of point P1 – Maximum Drop * (D2 – D1)). The code for fulfilling algorithm 1 is attached as Appendix E.

2) Algorithm 2

In case the error tolerance is exceeded, algorithm 2 assigns the water level of P1 to point P2, which is the same as algorithm 1. In case the maximum drop is exceeded, algorithm 1 assigns point P1 with a water level that is equivalent to (water level of point P2 + Maximum Drop * (D2 – D1)), which is different from algorithm 1. In other words, this is the only part that is different from algorithm 1. Consequently, the code for fulfilling algorithm 2 is almost the same as that for algorithm 1 and is not attached as an independent appendix to this thesis. Rather, it is encoded into the code attached as Appendix E.

Using these algorithms, the flood extent is predicted. The predicted flood extent consists of numerous polygons. Each of these polygons has a flood level. This set of data will be used later to identify road segments under flood. The program developed for flood extent prediction using Visual Basic and ArcObjects is attached as Appendix F.

11.3.2.6 Algorithms for Identifying Flooded Road Segments

After the flood extent and depth information is obtained by implementing the flood extent prediction model as described above, a set of small polygons are created. These polygons represent the areas under flood with the given flood level. Each polygon is associated with a value representing the flooded water level. This information is used together with the 3-D road centerline data to identify road segments that are flooded.

Figures 11.11 and 11.12 illustrate, respectively, part I and part II of the algorithm for identifying portions of a road segment that are under flood when that road segment runs through the flooded area. The algorithm part I is simply an overlay operation, i.e. the road centerline data are overlaid with the polygon data representing the flood extent. By doing this, road segments in the flood extent are identified. These identified road segments are associated with the information including the IDs of the corresponding 3-D road features and flood levels.

More specifically, each of the road segments identified as in the flood extent has the information in the following format: road segment A corresponds to road feature with ID F, starts from S1, and ends at S2; the corresponding water flood level is E. Points S1 and S2 are located via their distances to the start point of the road feature on which they are located.

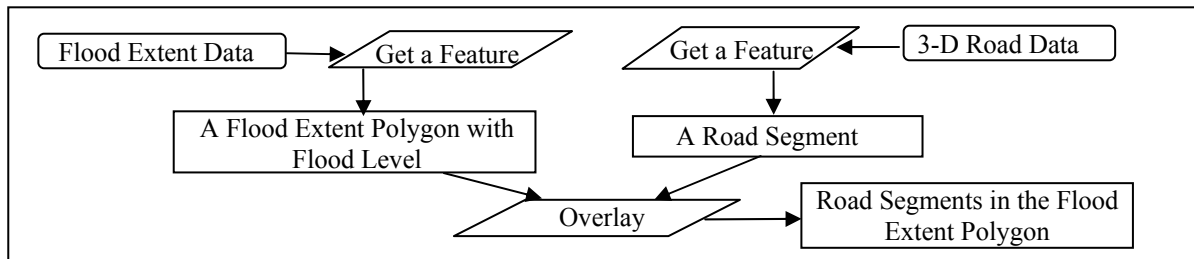


Figure 11.11 Flood Road Segment Identification Algorithm, Part I

For each of the road segments identified in Part I of the algorithm, Part II is performed as illustrated in Figure 11.12. First, a set of 3-D points between the start and end points of the segment are obtained from the 3-D road data. In addition, the start and end points of the identified road segment are obtained and their elevations linearly interpolated from the 3-D road data. With a set of 3-D points covering the complete road segment, all portions of this road segment that are under flood are identified (refer to Figure 11.5 for concepts). This is the critical part of the algorithm and it is illustrated in Figure 11.13.

In Figure 11.13, the black dots represent all 3-D points on the road segment. Except the start and end points, all these 3-D points come directly from the 3-D road data. For the start and end points, their elevations are obtained by linear interpolation. For example, for the start point, its elevation is linearly interpolated from its two neighboring 3-D points from the 3-D road data. One 3-D point is right before the start point and the other is right after the start point based on their distances to the start point of the corresponding 3-D road feature.

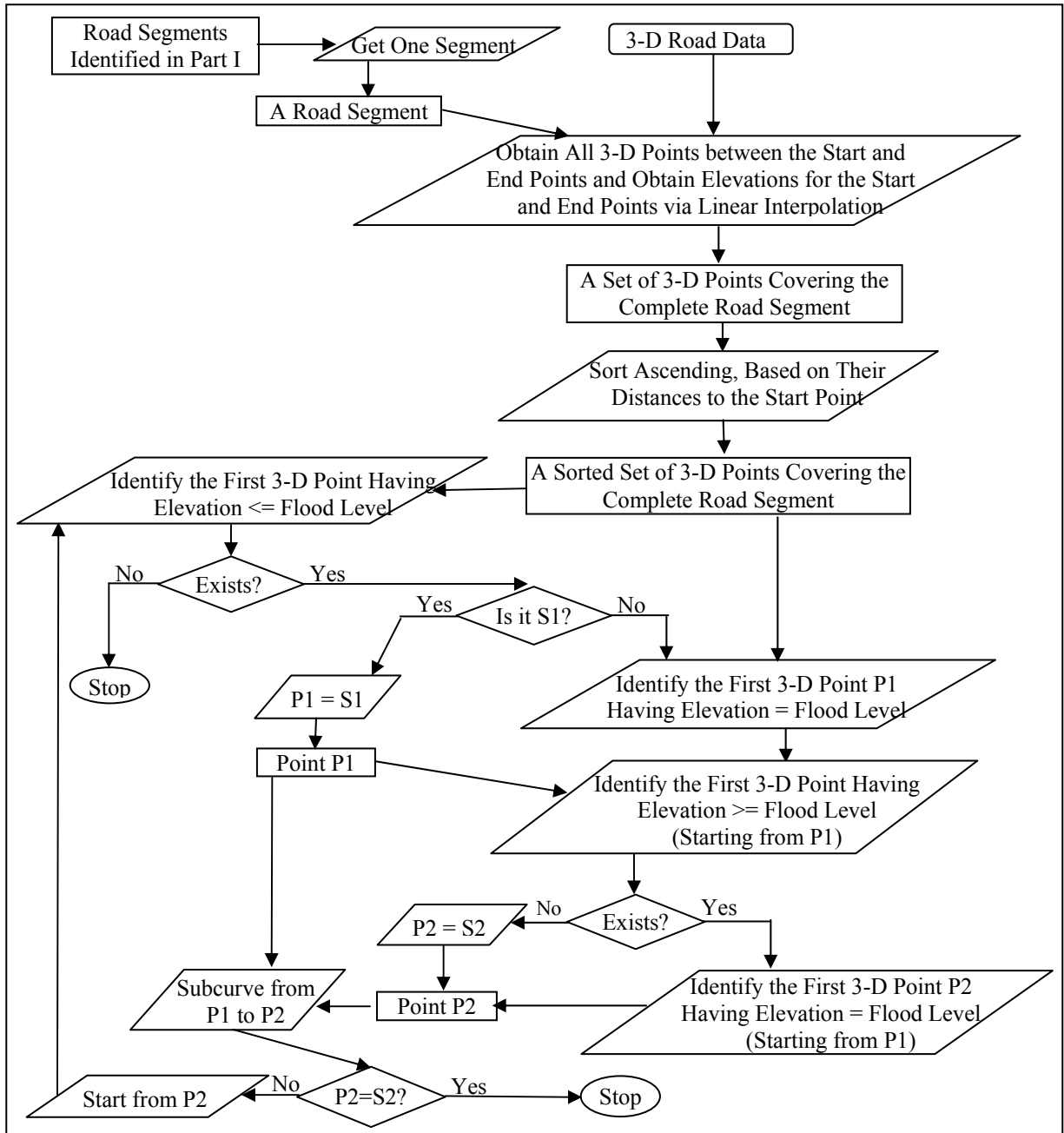


Figure 11.12 Flood Road Segment Identification Algorithm, Part II

The essence of identifying flooded road segments is to identify those points with elevations equaling the flooded water level, which are represented with light blue stars. According to the algorithm described earlier, the first point with elevation less than or equal to the flooded water level is S1 and therefore, S1 is assigned to point P1. Starting from P1, the first point with elevation greater than or equal to the flooded water level is point

T1. A linear interpolation will be used on point T1 and point T2, which is the 3-D point right before T1. By doing this, point P2 is located. Its elevation equals to the flooded water level. After identifying points P1 and P2 a sub-curve of the road segment starting from P1 and ending at P2 is generated. This sub-curve is under flood.

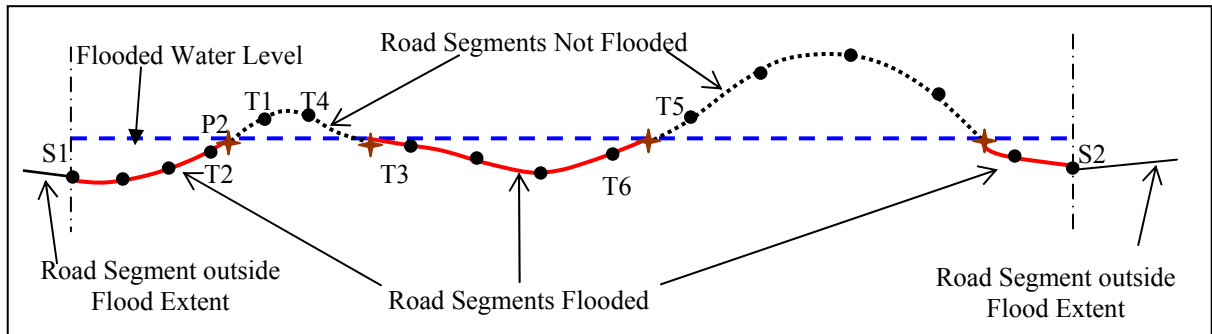


Figure 11.13 An Illustration of Identifying Flooded Road Segment Algorithm

According to the algorithm illustrated in Figure 11.12, since P2 is not S2, this procedure will continue starting from point P2. Consequently, points T3 and T5 will be found, new P1 and P2 are obtained, and a new sub-curve will be obtained. Repeating this procedure will identify all flooded portions of that road segment.

Again, a program developed in Visual Basic using ArcObjects is attached as Appendix G.

11.4 TESTING

This section describes the results of testing the flood extent prediction model, the model for identifying road segments under flood, and their corresponding algorithms.

11.4.1 Study Area

The study area for testing the models and algorithms is chosen to be part of the Wilson County in North Carolina. There are a few reasons of choosing the study area as illustrated in Figure 11.14. First, according to the conversation with NCDOT personnel, there are high-quality water body data in this area. Second, LIDAR data are available in this area. Third, 3-D road data are available in this area. Fourth, there are roads that are

close to the water body and therefore, will produce meaningful results. While only portions of roads and water bodies in Wilson County are included in the testing scope, the LIDAR elevation data are countywide, or they cover the complete county of Wilson.

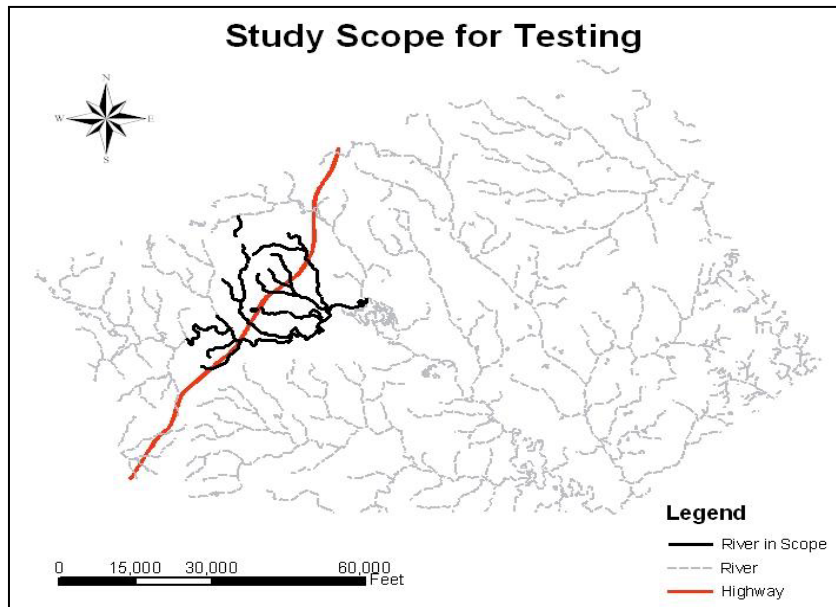


Figure 11.14 Study Scope for Testing in Wilson County

11.4.2 Results

This subsection describes the results of testing our models and algorithms in the study area. The flood extent prediction model and its associated algorithm are tested with the water body data in the study scope. The initial interval of the points along the river is 20 ft, which is the same as the cell size of the elevation data (LIDAR 20-ft DEMs). The flood level is 2 ft. Figures 11.15 and 11.16 show the results of flood extent prediction, using algorithm 1 and algorithm 2 to deal with the error tolerance and the maximum drop issues. Both algorithms use an error tolerance of 1ft/10ft and a maximum drop of 1ft/50ft. Figures 11.17 and 11.18 provide detailed views from the results shown in Figures 11.15 and 11.16, respectively.

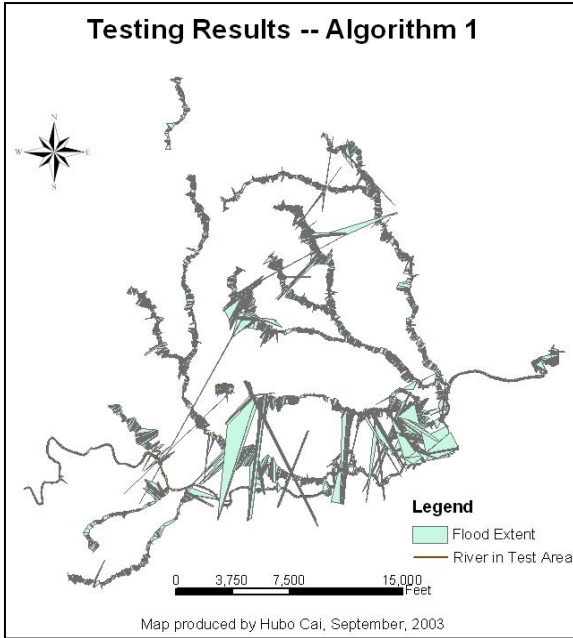


Figure 11.15 Testing Results – Algorithm 1

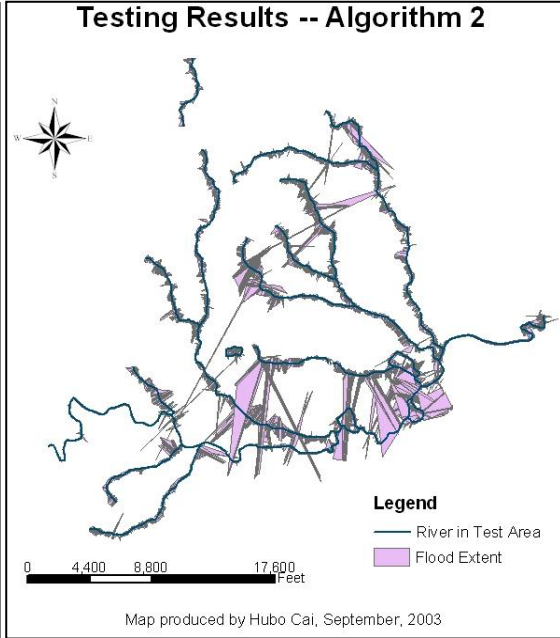


Figure 11.16 Testing Results – Algorithm 2

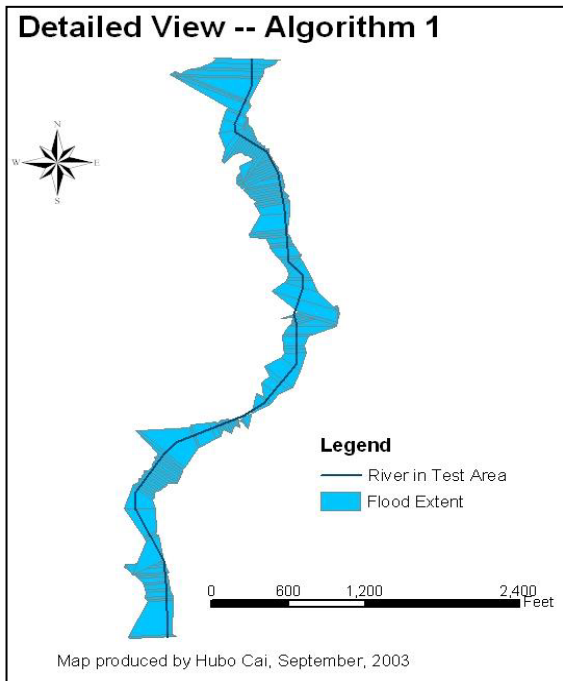


Figure 11.17 Detailed View – Algorithm 1

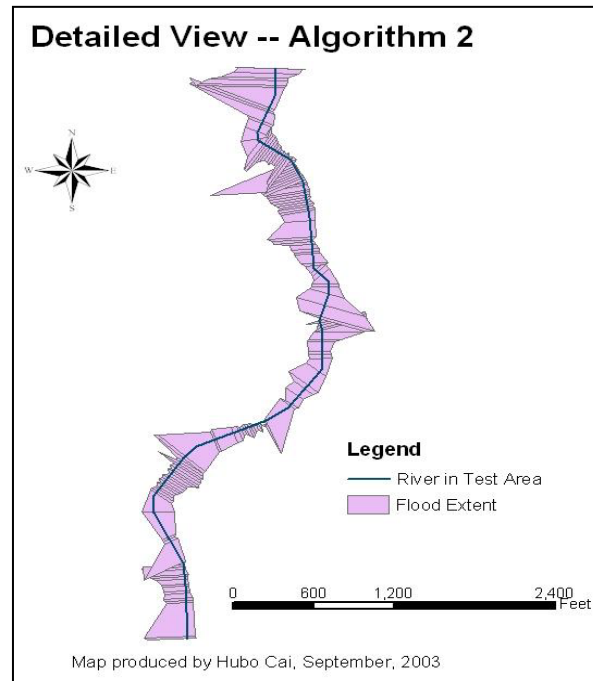


Figure 11.18 Detailed View – Algorithm 2

After the flood extent is determined with given flood level and interval, the flooded road segment identification model and its associated algorithm can be applied to obtain determine flooded road segments. In applying this model and its algorithm, the 3-D road data mentioned earlier play a critical role.

Figure 11.19 shows the results of applying this model when using the results of the flood extent model with the flood level of 2 ft, using algorithm 1. Figure 11.20 shows the results of applying this flooded road segment identification model using the results of the flood extent model with the flood level of 2 ft, using algorithm 2. Figure 11.21 provides a detailed view of the identified flooded road segments. It is obvious that for a road segment within the flood extent it does not mean that this road segment is completely flooded.

Table 11.1 provides part of the attribute table of the resulting identified flooded road segment, which contains helpful information to identify the flooded road segments easily. The column with the heading of “MERGE” identifies the road feature in the 3-D road data, of which the flooded road segments come. The columns with headings of “Start” and “End” specifies the planimetric distances from the start and end points of the flooded road segment to the start of that particular road feature. Consequently, with this information, the flooded road segments can be easily identified in addition to the visual result as shown in Figure 11.21 and therefore, can be combined with other data for further analyses.

Table 11.1 Part of the Attribute Table of the Identified Flooded Road Segments

FID	Shape	MERGE	Start	End
0	Polyline	1	16268.6676	16727.8315
1	Polyline	1	25156.4506	25380.1103
2	Polyline	1	28723.0097	29402.4092
3	Polyline	1	27629.4739	28723.0097
4	Polyline	2	16993.5179	17597.3257
5	Polyline	2	17597.3257	18524.5620
6	Polyline	2	11142.0085	11908.6555
7	Polyline	2	21084.6812	21183.0234
8	Polyline	2	10422.6277	10742.9855
9	Polyline	2	9247.3923	9790.3087
10	Polyline	2	7324.4328	7902.3860

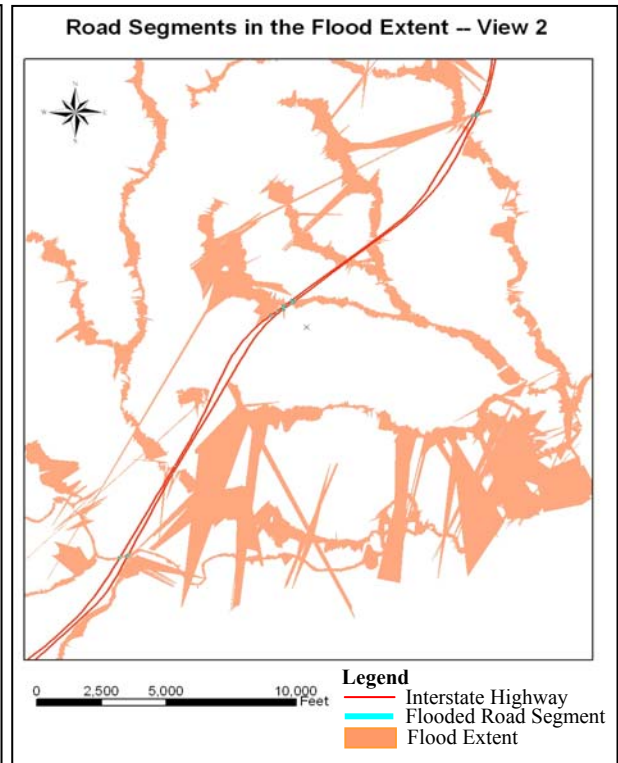
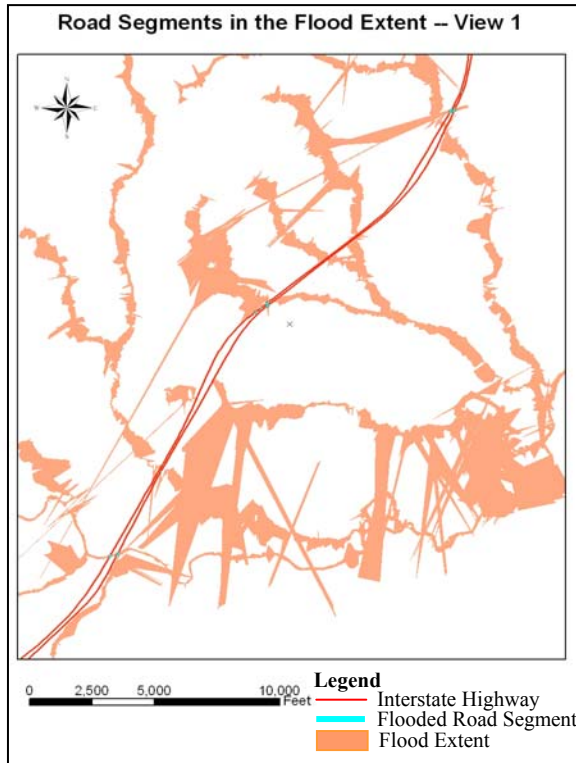


Figure 11.19 Flooded Road Segments–Algorithm 1 Figure 11.20 Flooded Road Segments–Algorithm 2

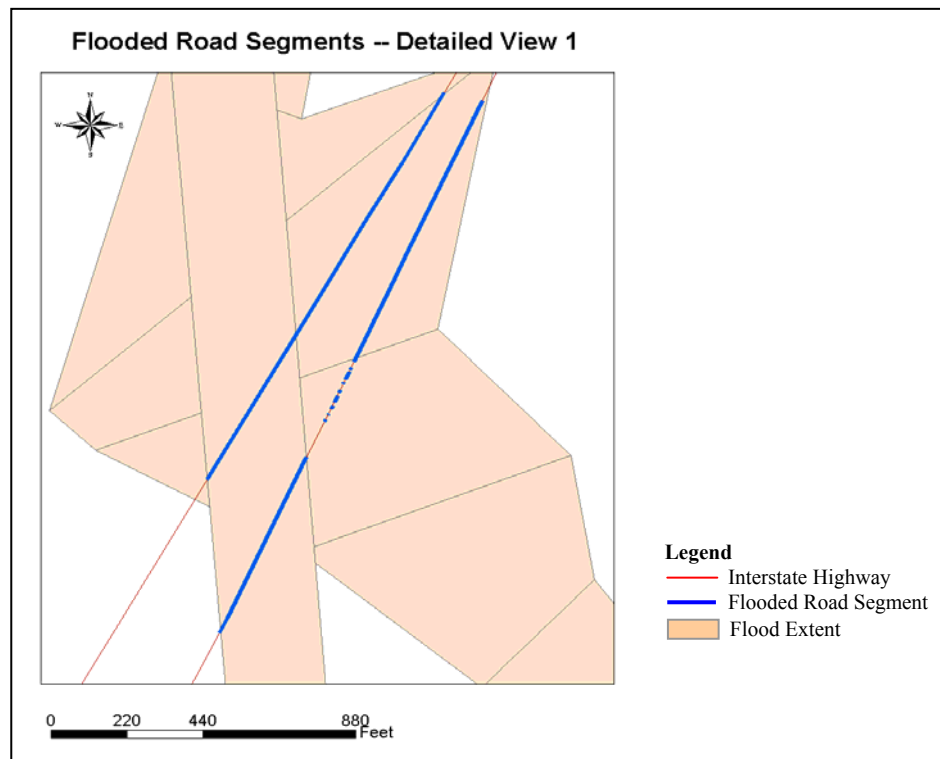


Figure 11.21 Detailed View of the Flooded Road Segments

11.5 LIMITATIONS AND CONCLUSIONS

It is recognized that the two models developed in this study have limitations. The flood extent prediction model assumes a uniform flood level (a uniform water surface level increase) for all water bodies in the study scope. In reality, this is not the case. Rather, the water surface level should be obtained from field observations of river stages or via the deployment of hydraulic models. Furthermore, in enforcing those two constraints, an error tolerance and a maximum drop are used. In testing, the values of the error tolerance and the maximum drop are determined based on the conversations with NCDOT personnel. These values will vary if areas other than the testing area are interested. Detailed examination of the testing results also reveals that there are some other potential problems with the flood extent prediction model in this study. To summarize, the flood extent prediction model developed in this study could be improved further or could be replaced by other models.

However, the major purpose of this study is to deploy the usefulness of the 3-D road centerline data in identifying flooded road segments under a flooding scenario. Theoretically, it could be assumed that the flood extent and depth information should be available outside this study. The only purpose of developing the flood extent prediction model in this study is to provide flood extent and depth information, which is needed in order to determine the flooded road segments. The flood extent and depth information obtained in the testing might not be perfect, but it does not mitigate the usefulness of the 3-D road centerline data in determining flooded road segments under a given flooding situation, as illustrated in this study. The problems with the flood extent prediction model are not in the scope of this research and therefore, are not further addressed.

It is concluded based on the testing results that modeling transportation linear objects (road centerlines) in a 3-D space and its resulting 3-D road centerline data are useful in determining flooded road segments and consequently, useful information could be obtained to determine the rescue route for the delivery of rescue personnel and supplies. Compared to the use of planimetric road centerline data, the use of 3-D road centerline provides more detailed information regarding the flooded road segments. It is believed that with accurate flood extent and depth information, the use of 3-D road centerline data would significantly benefit emergency response and management activities such as those in the scenario of hurricanes and floods.

12 SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This chapter summarizes this research, presents the conclusions obtained in this research, provides recommendations, and points out future research directions. Key findings from this research are emphasized.

12.1 SUMMARY

Real world objects are all three-dimensional (3-D) objects. Numerous applications in geographic information systems (GIS) require modeling and representing spatial objects in 3-D space with elevation (or height) and geometric information such as gradients, aspects, and actual lengths and areas together with their planimetric positional information. However, many current geographic information systems (GIS) can only represent two-dimensional (2-D) information by projecting the 3-D objects onto a planar surface and referencing them to a known coordinate system (Burrough and McDonnell 1998). The GIS community has been struggling with solving complex problems dealing with 3-D objects in a 2-D approach (terrain modeling, for example).

With rapid developments in computer technologies, the 3-D spatial modeling has been a fairly recent development in GIS, as illustrated by the increasingly available 3-D GISs. A comprehensive literature review of the 3-D spatial modeling revealed that it has been widely used in modeling three-dimensional polygon (or area) objects so that important information about gradients, aspects, and volumes is captured. This 3-D spatial modeling technique can also be applied in modeling three-dimensional linear objects and capture geometric information about these linear objects such as gradients, curves, and distances, which are very important to applications such as flood response and management, highway geometric designs, routing procedures, road inventory, and etc. Despite the needs for modeling linear objects in 3-D space and the benefits it would bring, it has been largely neglected in the past few decades.

The major goal of this research is to answer questions regarding how to model and represent linear objects as 3-D linear objects in a 3-D space and how accurate this model would be in predicting their lengths. In this research, 3-D models (the point model and its variant and the mathematical model and its variant) were proposed based on previous researches and evaluated. The relationship between a point model or its variant

with a mathematical model or its variant was revealed. This research determined to implement a variant of the point model to model the linear objects in a 3-D space. More specifically, a 3-D linear object was modeled as a series of straight lines connecting a group of 3-D points that had X/Y/Z-coordinates. Based on this model, a calculation mechanism was proposed to predict the 3-D distance for linear objects using a 3-D approach.

To implement this variant of point model to model linear objects in a 3-D space, a computational model was developed with implementation details including the required input data (elevation datasets and planimetric line data), developing environments (ArcGIS, ArcObjects, and Visual Basic). The currently available elevation datasets (USGS DEMs, NED, and LIDAR) were described in detail, focusing on the error sources, data formats and structures. The technologies used to produce these elevation datasets were also introduced. With elevation datasets in different formats, two different approaches were proposed: one for working with LIDAR point data and one for working with DEMs. These two approaches differed in the way that the third dimension (elevation) was introduced. More specifically, in this research, a buffering and snapping approach was proposed to model linear objects in a 3-D space when using LIDAR point data. When working with DEMs, the elevations were introduced using a bilinear interpolation method to points planimetrically uniformly distributed along the linear object. The buffering and snapping approach proposed to work with LIDAR point data in modeling road centerlines in a 3-D space was a creative approach that took into consideration of the characteristics of roads and the density of LIDAR points. Furthermore, mathematical equations were proposed to predict 3-D distances for linear objects using a 3-D approach.

A case study was designed to test the developed 3-D model. The case study was in the transportation field because transportation occupies a rich source of linear objects (roads). A fairly recent development in GIS-T has been the deployment of LRS to store and manage road inventory data. With an LRS, all events along roads are located by distances along the roads to a reference point. Modeling transportation linear objects in a 3-D space and predicting their 3-D distances using the proposed 3-D approach had an important benefit, i.e. road distances could be predicted without field work so that significant savings in cost and manpower could be achieved. As a deviation, existing technologies that could be used to obtain road distances were described and

evaluated. Their advantages, disadvantages, and limitations were discussed. This also served as an illustration of the significance of this research.

The case study scope was determined based on the availability of LIDAR data. The North Carolina Floodplain Mapping Program is collecting LIDAR data in North Carolina. This is still an ongoing project and therefore, LIDAR data are not available statewide in North Carolina. As a result, the study scope of the case study in this research included most Interstate highway road segments in nine counties in North Carolina and most US and NC routes in Johnston County of North Carolina. The sample size and the sample variety were also considered when determining this study scope.

After the study scope was determined, the required elevation datasets and planimetric road centerline data were obtained. In this case study, two different elevation datasets were used: LIDAR elevation datasets and NED. LIDAR elevation datasets were further categorized into LIDAR point data, LIDAR 20-ft DEMs, and LIDAR 50-ft DEMs. The planimetric road centerline data were obtained from the GIS unit of NCDOT, which were digitized from DOQQs. The road centerline data were originally in the format of a link-node system, consisting of numerous short road segments. For open data (data sharing) purposes, FTSegs were extracted from the original road centerline data. The obtained elevation datasets and the planimetric road centerline data were preprocessed so that all datasets were using the same spatial referencing system.

After preprocessing, algorithms and program codes were developed and executed to construct 3-D models for linear objects based on the developed conceptual model and the computational model. In this case study, when using DEMs to construct 3-D models, two intervals were taken to locate points planimetrically along that linear object, one at the full cell size of the DEM and the other at the half cell size of the DEM. When using LIDAR point data, the buffer sizes were dependent on the road types. The typical lane width for Interstate highways and US routes in North Carolina is 12 ft. For NC routes, it is 10 ft. Assuming two lane roads, the buffer size used when constructing 3-D lines for Interstate FTSegs and US FTSegs was 12 ft. It was 10 ft for NC FTSegs. As a result, each linear object was represented by a series of straight lines and a series of 3-D points. These 3-D points were along the linear object and connected by the corresponding series of straight lines. The 3-D

distance for a linear object was calculated based on these 3-D points and straight lines. Consequently, each linear object in this case study was associated with 7 calculated 3-D distances, one from the use of LIDAR point data, one from the use of LIDAR 20-ft DEM and an interval of 10 ft, one from the use of LIDAR D20-ft DEM and an interval of 20 ft, one from the use of LIDAR 50-ft DEM with an interval of 25 ft, one from the use of LIDAR 50-ft DEM and an interval of 50 ft, one from the use of NED and an interval of 15 meters, and one from the use of NED and an interval of 30 meters.

A major concern in GIS applications is the accuracy. In this case study, the accuracy of the predicted 3-D distances was evaluated by comparing to the reference data – DMI measured distances. The DMI measured distances were obtained by fieldwork in the summer of 2003. The researchers encountered some problems when doing field work. The biggest problem was the identification of the start or end point for a FTSeg in the field. To avoid these problems, a data collection strategy was designed and followed.

The accuracy of the predicted 3-D distances was assessed by comparisons to the DMI measured distances. In this case study, errors of predicted 3-D distances were represented in two formats: the difference and the proportional difference based on the DMI measured distance between the predicted 3-D distance and the DMI measured distance. In addition, road types were taken into consideration, resulting in 4 different groups of FTSegs: All FTSegs, Interstate FTSegs, US FTSegs, and NC FTSegs.

The assessment methods used included calculations of descriptive statistics, error distribution histograms, hypothesis tests, frequency analyses, and calculations of 100% RMSEs and 95% RMSEs. The descriptive statistics and error distribution histograms were used to determine if the errors in the predicted 3-D distances were randomly distributed. The hypothesis tests against sample means were carried out to determine if the means of errors were equivalent to 0. The frequency analyses were conducted to obtain a detailed view of the error distribution based on certain error ranges. The RMSEs were calculated to provide quantitative values for the error magnitudes.

In order to evaluate the effects from the use of different elevation datasets and intervals, the results of the accuracy assessment were analyzed in detail and a sensitivity analysis was conducted. Hypothesis tests and ANOVA against sample means were carried out to determine if errors of the predicted 3-D distances using different elevation datasets and intervals were same from the aspect of statistical distribution. The sample means, absolute sample means, frequencies for all error ranges used, and RMSEs from the samples using different elevation datasets and intervals were compared and comparisons charts were derived to illustrate the patterns and trends.

In addition, error sources in the input data (elevation data and road centerline data) were identified and described. The effects on the accuracy of the predicted 3-D distances from these errors (error propagations) were discussed. Errors introduced from the 3-D model developed in this research and their effects on the accuracy of the predicted 3-D distances were also discussed. Limitations of the methods being used herein were also discussed.

This case study also included significant factor analyses. The purpose of significant factor analyses was to determine if there was an association between a factor under consideration and the error of the predicted 3-D distance. The factors under consideration in this case study were all related to the geometry of the road centerlines. They included the distance, the average slope and the weighted average slope, the average and weighted slope changes, the number of 3-D points and the average density of the 3-D points being used to model a linear object in a 3-D space. All the slope-related factors were measures of the vertical complexity of the 3-D line. The number and average density of the 3-D points being used to model a linear object are measures of the effects from the integration of the input LIDAR point data and the creative way of utilizing LIDAR point data (the buffering and snapping approach) developed in this research.

The methods being used included calculation and comparisons of sample correlation coefficients and sample coefficients of determination, and the grouping and comparison method. Relationships between the accuracy of the predicted 3-D distance and a geometric characteristic a linear object occupied were revealed. Consequently, precautions could be applied to such linear objects when using the methods developed in this research to predict

their 3-D distances and the accuracy of the predicted 3-D distances could be further improved, if needed. However, the limitations of the methods being used herein were also discussed.

As an appreciation of the usefulness and significance of this research, the resulting 3-D road centerlines were used in a floodplain mapping application. More specifically, the resulting 3-D road centerlines from this research were deployed to determine road segments under flood given a flooding scenario. Due to the lack of flood extent and depth information, a flood extent prediction model was developed together with the flooded road segment identification model. Corresponding algorithms and program codes were developed. The developed models and algorithms were tested in part of the Wilson County in North Carolina. By doing so, the usefulness of modeling road centerlines in a 3-D space in identifying flooded road segment and supporting rescue route determination was illustrated.

12.2 CONCLUSIONS

This research developed 3-D models to model linear objects in a 3-D space and to predict their 3-D distances. The developed models and algorithms were tested in a case study designed in the transportation context. Analyzing the results comprehensively lead to many interesting and meaningful observations. The conclusions reached in this research are listed as below.

- 1) With the current computer technologies and available datasets, it is technically feasible to model linear objects in a 3-D space, using a point model as developed in this research. A linear object can be modeled in a 3-D space using a series of 3-D points with X/Y/Z coordinates and connected to via a series of straight lines.
- 2) In order to model linear objects in a 3-D space, two datasets are required. One is the planimetric line data. The other is the elevation data. The 3-D points being used to model a linear object in a 3-D space obtain their X/Y coordinates from the line data and their Z coordinates from the elevation dataset.
- 3) Geometric properties of 3-D lines could be calculated using the 3-D approach, such as 3-D distance, slope and slope changes, and 3-D curvatures.

- 4) Other than using elevation datasets in the traditional format (DEM), 3-D line models could be also constructed using LIDAR point data with the buffering and snapping approach as developed in this research. This approach is developed based on the characteristics of the linear objects being modeled (roads occupy certain widths) and the density of LIDAR points.
- 5) The prerequisite for the use of the 3-D model and the 3-D approach to predict 3-D distances for linear objects is that the line data must be planimetrically correct. As illustrated in this research, the 3-D points being used to model a linear object in a 3-D space obtain their X/Y coordinates from the planimetric line data. Their Z coordinates are obtained from the elevation dataset based on their X/Y coordinates. Consequently, without planimetrically correct line data, the resulting 3-D lines and the predicted 3-D distances would be erroneous. In the case of using the buffering and snapping approach developed in this research, when working with LIDAR point data, if the lines are not in correct locations, LIDAR points that should not be snapped would be snapped while LIDAR points should be snapped might not be used. Also, the LIDAR points are snapped to the planimetric lines and without planimetrically correct line data, the snapping approach would be erroneous.
- 6) There is no universally agreed upon set of accuracy assessment method or approach, which could be applied in all cases of accuracy assessments for GIS applications. In other words, the use of a particular set of accuracy assessment methods or approaches is dependent on the applications.
- 7) Errors of the predicted 3-D distances, when compared to DMI measured distances, either in the format of differences or proportional differences, were not normally distributed. This conclusion was based on the descriptive statistics, especially the skew, and the distribution histograms.
- 8) The accuracy of the predicted 3-D distances, when compared with DMI measured distances, varied with the use of different elevation datasets. In this case study, four elevation dataset were used, the LIDAR point

data, the LIDAR 20-ft DEMs, the LIDAR 50-ft DEMs, and the NED data. The accuracy of the predicted 3-D distances was dependent on the elevation dataset being used.

9) Using 100% RMSEs as the accuracy measure, when compared to DMI measured distances, the RMSE of the predicted 3-D distances using LIDAR point data was 25.65 feet for the difference and 50.44 feet/1,000 feet for the proportional difference. With the use of LIDAR 20-ft DEMs with a 10-ft interval, the RMSE was 33.30 feet for the difference and 51.28 feet/1,000 feet for the proportional difference. With the use of LIDAR 20-ft DEMs with a 20-ft interval, the RMSE was 33.73 feet for the difference and 51.26 feet/1,000 feet for the proportional difference. With the use of LIDAR 50-ft DEMs with a 25-ft interval, the RMSE was 33.90 feet for the difference and 50.79 feet/1,000 feet for the proportional difference. With the use of LIDAR 50-ft DEMs with a 50-ft interval, the RMSE was 34.16 feet for the difference and 50.75 feet/1,000 feet for the proportional difference. With the use of NED with a 15-m interval, the RMSE was 35.53 feet for the difference and 50.68 feet/1,000 feet for the proportional difference. With the use of NED with a 30-m interval, the RMSE was 35.74 feet for difference and 50.68 feet/1,000 feet for the proportional difference.

10) The higher the accuracy of the elevation dataset being used, the higher the accuracy of the predicted 3-D distances, when compared with DMI measured distances. In this case study, LIDAR data had higher vertical accuracy than the NED data. The accuracy assessment and sensitivity analysis presented earlier in this chapter demonstrated that the accuracy of the 3-D distances predicted using LIDAR data was higher than the accuracy of the predicted 3-D distances using NED data.

From the aspect of differences using the 100% RMSE as the measure of the accuracy, the use of LIDAR point data in this case study improved the accuracy by about 28% compared to the use of NED data. The use of LIDAR DEMs improved the accuracy by about 6% compared to the use of NED data.

From the aspect of differences using the 95% RMSE as the measure of the accuracy, the use of LIDAR point data in this case study improved the accuracy by about 25% compared to the use of NED data. The use of LIDAR DEMs improved the accuracy by about 9% compared to the use of NED data.

From the aspect of proportional differences, the improvements due to the use of higher accurate elevation datasets were not as significant as those shown for differences above. This was the case because the majority (53%) of the FTSEgs in this case study were longer than 5,000 ft and about 73% were longer than 1,000 ft and 43% were longer than 10,000 ft.

11) It was concluded that the 3-D distances predicted using LIDAR point data with the snapping approach in this research had the best accuracy when compared to DMI measured distances.

12) While the accuracy of the predicted 3-D distances was dependent on the elevation dataset being used, if the same elevation dataset was used, but two different intervals were taken, one equivalent to half of the cell size and the other equivalent to the full cell size of the elevation data, the accuracies of the predicted 3-D distances were almost the same. In other words, the accuracy of the predicted 3-D distances was not dependent on the interval sizes if they were less than or equal to the cell size of the elevation data.

13) The accuracy of the predicted 3-D distances was also dependent on the road types. From the aspect of difference, NC FTSEgs had the best accuracy, followed by Interstate FTSEgs. US FTSEgs had the worst accuracy. From the aspect of proportional difference, Interstate FTSEgs had the best accuracy while NC FTSEgs had the worst accuracy. The accuracy of US FTSEgs was close to but better than the accuracy of the NC FTSEgs.

14) Examining the predicted 3-D distances in detail, together with the statistics used in the accuracy assessment lead to an impression that overall, the predicted 3-D distances using LIDAR data (LIDAR point data and LIDAR DEMs) were close to each other, but significantly different from using NED data.

15) The distance factor was significant to the accuracy of the predicted 3-D distance when compared to the DMI measured distance. There was a positive linear association between the error of the predicted 3-D distance and the factor of the distance from the aspect of the absolute difference but a negative linear

association from the aspect of the absolute proportional difference. In other words, the longer the distance of the FTSeg, the bigger the error it had and the smaller the proportional error it had.

16) The factor of the average slope was significant to the accuracy of the predicted 3-D distance when compared to the DMI measured distance. There was a positive linear association between the error of the predicted 3-D distance and the factor of the average slope from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the larger the average slope of the FTSeg, the bigger the error it had and the smaller the proportional effort it had.

17) The factor of the weighted average slope was significant to the accuracy of the predicted 3-D distance when compared to the DMI measured distance. There was a positive linear association between the error of the predicted 3-D distance and the factor of the weighted average slope from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the larger the weighted average slope of the FTSeg, the bigger the error it had and the smaller the proportional effort it had.

18) The factor of average slope change was significant to the accuracy of the predicted 3-D distance when compared to the DMI measured distance. There was a positive linear association between the error of the predicted 3-D distance and the factor of the average slope change from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the larger the average slope change of the FTSeg, the bigger the error it had and the smaller the proportional effort it had.

19) The factor of the weighted average slope change was significant to the accuracy of the predicted 3-D distance when compared to the DMI measured distance. There was a positive linear association between the error of the predicted 3-D distance and the factor of the weighted average slope change from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference.

In other words, the larger the weighted average slope change of the FTSeg, the bigger the error it had and the smaller the proportional effort it had.

20) The factor of the number of 3-D points was significant to the accuracy of the predicted 3-D distance using LDIAR point data and the snapping approach when compared to the DMI measured distance. There was a positive linear association between the error of the predicted 3-D distance and the factor of the number of 3-D points from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the more the number of 3-D points being used to represent the FTSeg in a 3-D space, the bigger the error it had and the smaller the proportional effort it had.

21) The factor of the average density of 3-D points was significant to the accuracy of the predicted 3-D distance using LDIAR point data and the snapping approach when compared to the DMI measured distance. There was a positive linear association between the error of the predicted 3-D distance and the factor of the average density of 3-D points from the aspect of the absolute difference but a negative linear association from the aspect of the absolute proportional difference. In other words, the higher the density (the smaller the average planimetric distance between two neighboring 3-D points along the linear object) of the 3-D points being used to represent the FTSeg in a 3-D space, the bigger the error it had and the smaller the proportional effort it had.

21) The conclusions of all the slope factors including the average slope, the weighted average slope, the average slope change, and the weighted average slope change were the same. All these factors were used to capture the vertical geometric complexity of 3-D linear objects. It leads to the belief that there was a consistency among these factors in describing the vertical geometric complexity of the 3-D lines. The conclusions of the factors of the number of 3-D points and the distance revealed that even though LIDAR points were not uniformly distributed, a longer FTSeg would have more 3-D points on it than a shorter FTSeg.

22) With the factor of the distance, when the distance reached 5,000 ft, from the aspect of the proportional difference, the effects on the error of the predicted 3-D distance when compared to the DMI measured distance from the increase of the distance were almost nonexistent and therefore, could be ignored.

23) With the factors of the average slope change and the weighted average slope change, when the value of the average slope change or the weighted average slope change reached 5%, the effects on the error of the predicted 3-D distance when compared to the DMI measured distance from the increase of the average slope change or the weighted average slope change were almost nonexistent and therefore, could be ignored.

24) With the factor of the number of 3-D points being used to model a linear object in a 3-D space, when 500 points were reached from the aspect of the difference and when 1,000 points were reached from the aspect of the proportional difference, the effects on the error of the predicted 3-D distance when compared to the DMI measured distance from the increase of the number of 3-D points being used were almost nonexistent and therefore, could be ignored.

25) The resulting 3-D lines and their associated 3-D point data could be used in other applications such as in identifying flooded road segment given flood extent and depth information so that rescue routes could be determined.

12.3 KEY FINDINGS

Key findings from this research are emphasized in this section as below.

- 1) The 3-D modeling approach developed in this research provides an efficient way to predict 3-D distances for road centerlines. Using this approach, all distances are calculated by computer. There is no more field measurement needed. It indicates significant savings in time and efforts.

- 2) Modeling road centerlines in a 3-D space by using the approach developed in this research results in 3-D road centerline data, which could be used in many real applications such as determining flooded roads and the exact depth of the flooding on roads.
- 3) The accuracy of the predicted 3-D distance depends on the elevation dataset being used. It is found in this research that the use of LIDAR point data improves the accuracy by about 28% compared to the use of NED data. The use of LIDAR DEMs improves the accuracy by about 6% compared to the use of NED data.
- 4) Compared to the use of 2-D length to approximate the 3-D distance, the use of LIDAR point data improves the accuracy by 38%. The use of LIDAR DEMs improves the accuracy by 19%. The use of NED improves the accuracy by 13%.
- 5) The preference order of the elevation datasets to be used in modeling road centerlines as 3-D lines and predicting their distances is LIDAR point data, LIDAR DEMs, and NED (from higher preference to lower preference). LIDAR point data are the most preferred elevation dataset to be used.

12.4 RECOMMENDATIONS

Based on the results of this research, it is recommended that:

- 1) Agents maintaining road inventories should consider modeling road centerlines in a 3-D space using the models developed in this research and predicting their 3-D distances using the 3-D approach as proposed in this research, unless they have the required resources (personnel, DMI, GPS, etc.) to measure road distances via field efforts and time and costs are not issues. Modeling road centerlines in a 3-D space using the existing data and predicting their 3-D distances would provide reasonable accuracies, but significantly save time and money and minimize required field efforts.

2) Practitioners interested in modeling road centerlines in a 3-D space should consider adopting the program codes developed in this research with minor modifications, rather than developing their own program codes from scratch. It was estimated that it took approximately 200 man hours in developing these program codes, given the developers are proficient in ArcGIS, ArcObjects, and programming with Visual Basic. Even though the program codes developed in this research were based on the data structure (link-node structure) of the road centerline data maintained in NCDOT, they could be adopted in deriving 3-D road centerlines and predicting their 3-D distances, given these road centerlines are modeled as linear objects.

3) In addition to the 3-D distance, modeling road centerlines in a 3-D space also enables the derivation of other information such as gradients, slopes, three-dimensional curvatures, and etc. to facilitate applications such as highway geometric design, determination of sight distance, but routing, etc. 3-D road centerlines should also be used in determining flooded road segments in a flooding scenario to help determine rescue routes.

4) The models and algorithms proposed in this research should be also used to model linear objects other than road centerlines in a 3-D space, such as stream lines and power lines. Generally speaking, any objects modeled as lines in a 2-D GIS could be modeled in a 3-D space using the methods proposed in this research.

5) In order to use the methods developed in this research, the planimetric line data should be spatially corrected. This is the prerequisite of modeling linear objects in a 3-D space using the models and algorithms developed in this research.

6) Which elevation dataset to use should be determined based on the data availability and the accuracy requirements. The use of LIDAR elevation datasets indicated the higher accuracy in the predicted 3-D distances when compared to the use of NED. However, NED data are available nationwide but LIDAR data are not. Furthermore, the use of LIDAR point data resulted in the best accuracy of the predicted 3-D distances. However, it requires more efforts to use LIDAR point data to model linear objects in a 3-D space than the use of LIDAR DEMs or NED. Therefore, the appropriateness of using a particular elevation dataset

must be evaluated. Also, elevation datasets are usually bare-Earth data and therefore, they should not be used to model linear objects above ground or underground such as power lines.

7) When taking the buffering and snapping approach developed in this research to model road centerline in a 3-D space using LIDAR point data, an appropriate buffer should be chosen based on the characteristics of the roads being modeled. In addition, if linear objects other than road centerlines are to be modeled, the appropriateness of this buffering and snapping approach should be evaluated.

8) When using the methods developed in this research to predict 3-D distances, precautions must be paid to certain linear objects based on the significant factor information so that the accuracy of the predicted 3-D distances could be further improved.

12.5 FUTURE WORK

It is recognized that this research has certain limitations. The directions in which future study would offer considerable rewards are listed as below.

1) This research used a point model to model linear objects in a 3-D space. Models in other formats could also be developed. For example, based on the point model, trend lines could be derived and be described with mathematical functions.

2) Accuracy assessment methods other than those used in this research could be developed and used to evaluate the accuracy of the predicted 3-D distances. Also, this research did not directly measure the accuracy of the developed model. Rather, the accuracy of the predicted 3-D distance was treated as an indicator of the accuracy of the developed model. For example, a three-dimensional buffer could be applied to the true 3-D line and the percentage of the modeled 3-D line in the buffer could be calculated as a measure of the accuracy of the 3-D model.

- 3) This research evaluated the effects on the accuracy of the predicted 3-D distances from the use of different elevation datasets with different vertical accuracies and resolutions. But it did not specifically evaluate the effects on the accuracy of the predicted 3-D distances from the accuracy of the line data (in this research, only one line dataset was used).
- 4) The significance of factors other than those considered in this research could be evaluated, for example, the planimetric curvature or the three-dimensional curvature or the number of bridges on a particular linear object.
- 5) In evaluating the significance of factors, methods other than those used in this research could be deployed. For example, instead of evaluating the significance for each individual factor, multivariate models could be adopted to evaluate the significance for several factors under consideration. The resulting significance measures could be compared to reveal information such as which factor is the most significant one.
- 6) This research focused on evaluating the linear association between a factor under consideration and the accuracy of the predicted 3-D distance. However, a factor might have a nonlinear association with the accuracy of the predicted 3-D distance and therefore, other statistical methods could be used reveal those nonlinear associations.
- 7) Due to the limitations in the availability of LDIAR data, the case study in this research was constrained to the eastern part of North Carolina, which is a relatively flat area. Testing the developed models and evaluating the accuracy of the predicted 3-D distances in a mountainous area, for example, the western part of North Carolina, would supplement the findings from this research.

REFERENCES

- Ackermann, F., "Experimental Investigation into the Accuracy of Contouring from DTM," *Photogrammetry Engineering Remote Sensing*, 44(12), pages 1537-1548, 1978.
- Ackermann, F., "The Accuracy of Digital Terrain Models," *Proceedings, 37th Photogrammetric Week*, University of Stuttgart, pages 113-143, 1979.
- Ackermann, F. and Stark, E., "Digital Terrain Model from Spacelab Metric Camera Photographs," *Proceedings, Metric Camera Workshop*, ESA, pages 9-12, 1985.
- Airborne 1 Corporation, "Lidar Accuracy, an Airborne 1 Perspective," Los Angeles, CA, 2000.
- ArcGIS Desktop Help 2002*, Environmental Systems Research Institute, Inc. (ESRI), 380 New York Street, Redlands, California, 92373, USA.
- Anderson, J. M. and Mikhail E. M., *Surveying Theory and Practice*, 7th Edition, WCB/McGraw-Hill, Chapter 13, pages 742-779, 1998.
- ASCE, *Topographic Surveying*, Technical Engineering and Design Guides as Adapted from the US Army Corps of Engineers, No. 29, American Society of Civil Engineers, Reston, Virginia, 2000.
- ASPRS-ACSM, *Proceedings of the Digital Terrain Models (DTM) Symposium*, St. Louis, Missouri, pages 567, 1978.
- Backe, K., "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, 1996.
- Bak, P. and Mill, A., "Three Dimensional Representation in a Geoscientific Resource Management System for the Minerals Industry," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 155-182, 1989.
- Baker, W. T. and Blessing, W. E., *NCHRP Synthesis of Highway Practice 21: Highway Location Reference Methods*, Transportation Research Board, National Research Council, Washington, D. C., page 29, 1994.
- Balce, A., "Determination of Optimum Sampling Interval in Grid Digital Elevation Models (DEM) data acquisition," *Photogrammetry Engineering Remote Sensing*, 53(3), pages 323-330, 1987.
- Barnes, S., "River Buffer Rules Face Opposition from Property Rights Advocates," *Watauga Mountain Times*, Issue 14, April 18, 2002.
- Batten, L. G., "National Capital Urban Planning Project: Development of a 3D GIS," *Proceedings of GIS/LIS 89*, Orlando, FA, pages 781-786, November 1989.
- Beard, M. K. and Battenfield, B. P., *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, 1999.
- Berry, J. K., "Beware Slope's Slippery Slope," *GeoWorld*, pages 22-29, January 2003.
- Briggs, D. W. and Chatfield, B. V., "Integrated Highway Information Systems," *Report 133, National Cooperative Research Program*, National Academy Press, Washington, D. C., 1987.
- Bristow, C. S. and Raper, J. F., "Modeling 3D Reservoir Geometry: a New Approach Using IVM," *Geological Society of London Special Publication "Advances in Petroleum Geology"*, 1991.
- Burrough, P. A. and McDonnell, R. A., *Principles of Geographical Information Systems*, Oxford University Press, New York, 1998.

Burrough, P. A., *Principles of Geographical Information Systems for Land Resources Assessment*, Oxford University Press, Oxford, 1986.

Caspary, W. and Scheuring, R., "Positional Accuracy in Spatial Databases," *Computers, Environment and Urban Systems*, Volume 17, pages 103-110, 1993.

Chen, X., Ikeda, K., Yamakita, K., and Nasu, M., "Three-Dimensional Modeling of GIS Based on Delaunay Tetrahedral Tessellations," *ISPRS Commission III Symposium, Spatial Information from Digital Photogrammetry and Computer Vision*, Volume 2357, Part 3/1, pages 132-139, 1994.

Chong, A. K., "A Field Check Sampling Procedure to Evaluate the Positional Accuracy of Digital Database," *Proceedings of ASPRS/ACSM, Volume 2, Annual Convention and Exposition of Technical Papers*, Seattle, Washington, page 1, 1997.

Chrisman, N. R., "A Theory of Cartographic Error and Its Measurement in Digital Databases," *Proceedings, AUTO-CARTO 5*, pages 159-168, 1982.

Chrisman, R. N., "The Role of Quality Information in the Long-Term Functioning of A Geographic Information System," *Proceedings: AUTO-CARTO 6*, Bethesda, MD: ACSM, pages 301-312, (meeting held in Ottawa), 1983.

Chrisman, R. N., *Exploring Geographic Information Systems*, 2nd Edition, Wiley, New York, 2002.

Cox, D. R., "Some Remarks on the Role of Statistics of Graphical Methods," *Applied Statistics*, Volume 27, page 9, 1978.

Dana, P. H., "Global Positioning System Overview," *The Geographer's Craft Project*, Department of Geography, The University of Colorado at Boulder.

Daniel, C. and K. Tennant, *DEM Quality Assessment, Digital Elevation Model Technologies and Applications: the DEM Users Manual*, American society for Photogrammetry and Remote Sensing, Bethesda, MA, 2001

Day, T. and Muller, J., "Quality Assessment of Digital Elevation Models Produced by Automatic Stereomatchers from SPOT Image Pairs," *Photogrammetry Records*, 12(72), pages 797-808, 1988.

Davis, J. C. and McCullagh, M. J., *Display and Analysis of Spatial Data*, John Wiley and Sons, London, pages 338, 1975.

Dikau, R., "The Application of A Digital Relief Model to Lnadform Analysis in Geomorphology," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 51-78, 1989.

Dilke, O. A. W., *The Roman Land Surveyors. An Introduction to the Agrimensores*, David and Charles, Newton Abbot, 1971.

Dissard, O. and Jamet, O., "3D Reconstruction of Buildings from Stereo Images Using Both Monocular Analysis and Stereo Matching: An Assessment within the Context of Cartographic Production," *Proceedings of SPIE --- The International Society for Optical Engineering, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, Volume 2486, pages 251-262, Orlando, Florida, April 1995.

Dueker, K. J., "Geographic Information Systems and Computer-aided Mapping," *Journal of the American Planning Association* 53, pages 383-390

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

Dueker, K. J. and Kjerne, D., "Multipurpose Cadastre: Terms and Definitions ," *Proceedings Annual Convention ACSM-ASPRS*, Volume 5, pages 94-103, 1998.

- Durst, J. M. and Kunii, L. T., "Methods for the Efficient Storage and Manipulation of Spatial Geological Data," Chapter 14, *Three-Dimensional Modeling with Geoscientific Information Systems*, Editor Turner, A. K., NATO ASI Series, Kluwer Academic Publishers, pages 189-214, 1989.
- Eastman, CH. And Weiler, K., "Geometric Modeling using the Euler Operators," *Proceedings, First Annual Conference of Computer Graphics in CAD/CAM System*, Cambridge, Massachusetts, pages 248-259, 1979.
- Elmasri, R. and Navathe, B. S., *Fundamentals of Database Systems*, Pearson Education (Singapore) Pte. Ltd., Indian Branch, 482 F.I.E. Patparganj, Delhi 110 092, India.
- Englert, R. and Bremers, A. B., "Improving Reconstruction of Man-Made Objects from Sensor Images by Machine Learning," *Proceedings of SPIE --- The International Society for Optical Engineering, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, Volume 3072, pages 263-275, Orlando, Florida, April 1997.
- Exploring ArcObjects*, Editor Michael Zeiler, Environmental Systems Research Institute, Inc. (ESRI), 380 New York Street, Redlands, California, 92373, USA, 2001.
- Federal Geographic Data Committee (FGDC), *National Spatial Data Infrastructure Framework Transportation Identification Standard – Working Draft*, FGDC Ground Transportation Subcommittee, May 1999.
- Federal Geographic Data Committee (FGDC), *Subcommittee Reports: FGDC Ground Transportation Subcommittee Positions and Recommendations on Linear Referencing Systems*, FGDC Ground Transportation Subcommittee, 1994.
- Federal Highway Administration, *Linear Referencing Practitioners Handbook*, prepared for FHWA contract #DTFH61-95-C-00169, Task Order 5 by GIS/Trans, Ltd. 1998.
- FGDC-STD-002, "Spatial Data Transfer Standards," Federal Geographic Data Committee (a modified version was adopted as ANSI NCITS 320:1998), 1998.
- FGDC-STD-007.3-1998, "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, 1998.
- Fisher, P., "Visualization of the Reliability in Classified Remotely Sensed Images," *Photogrammetric Engineering and Remote Sensing*, Volume 60, pages 905-910, 1994a.
- Fisher, P., "Visualizing the Uncertainty of Soil Maps by Animation," *Cartographic*, Volume 30, pages 20-27, 1994b.
- Fletcher, D. R., "Modeling GIS Transportation Networks," *Proceedings Volume II, Urban and Regional Information Systems Association, 25th Annual Conference*, Fort Lauderdale, FL, 1987.
- Fletcher, D. R., *GIS-T Pooled Fund Study Phase B Summary Report*, Alliance for Transportation Research, Albuquerque NM, 1995.
- Fletcher, D. R., "Geographic Information Systems for Transportation: A Look Forward," *Transportation in the New Millennium*, Transportation Research Board, Washington, DC, 2000.
- Flood, M., and B. Gutelius, "Commercial Implication of Topographic Terrain Mapping Using Scanning Airborne Laser Radar," *Photogrammetric Engineering and Remote Sensing, ASPRS, Journal of The American Society for Photogrammetry and Remote Sensing*, Vol. 63, No. 4, April 1997
- Fowler, R.A., "LIDAR for Flood Mapping," *Earth Observation Magazine*, July 2000.
- Frederiksen, P., "Terrain Analysis and Accuracy Prediction by Means of Fourier Transformation," *Photogrammetria*, 36, pages 145-157, 1981.

- Fritsch, D. and Schmidt, D., "Digital Terrain Model (DTM) Integration and Three-Dimensional Query Spaces in Geographic Information Systems," *ISPRS Commission III Symposium, Spatial Information from Digital Photogrammetry and Computer Vision*, Volume 2357, Part 3/1, pages 235-242, 1994.
- Fuchs, H., Kedem, Z., and Naylor, B. F., "Near Real-Time Shaded Display of Rigid Objects," *ACM Computer Graphics*, Volume 17, Number 3, pages 65-72, 1983.
- Genderen van, J. L. and Lock, B. F., "Testing Land-use Map Accuracy," *Photogrammetric Engineering and Remote Sensing*, Volume 43, pages 1135-1137, 1977.
- Geo Decisions, "Linear Referencing Systems, Review and Recommendation for GIS and Database Management Systems," unpublished material, Geo Decisions, a division of Gannett Fleming, Inc. 1997.
- Gevers, T. and Smeulders, A. W. M., "Efficient Recognition of 3D Rigid Solid Objects from 2D Projective Images Based on Projective Invariant Descriptions," *ISPRS Commission III Symposium, Spatial Information from Digital Photogrammetry and Computer Vision*, Volume 2357, Part 3/1, pages 296-302, 1994.
- Gold, M. C., "Surface Interpolation, Spatial Adjacency and GIS," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 21-36, 1989.
- Goodchild, M. F., "Issues of Quality and Uncertainty," In J. C. Muller (Ed.), *Advances in Cartography*, Elsevier, New York, pages 113-139, 1991a.
- Goodchild, M. F., "Issues of Quality and Uncertainty," *Proceedings, State of Indiana Geographic Information System Conference*, Indianapolis, Indiana University GIS Alliance, pages 17-53, 1991b.
- Goodchild, M. F., "Geographical Data Modeling," *Computers and Geosciences*, volume 18, pages 401-408, 1992.
- Goodchild, M. F., "Spatial Accuracy," *Proceedings, 22nd Annual Conference, Australasian Urban and Regional Information Systems Association*, Sydney, addendum, 1994.
- Goodchild, M. F., "Attribute Accuracy," In *Elements of Spatial Data Quality*, S. C. Guptill and J. L. Morrison (Eds.), Elsevier, New York, pages 59-80, 1995.
- Goodchild, M. F., "Generalization, Uncertainty, and Error Modeling," *Proceedings, GIS/LIS 96*, Denver, pages 765-774, 1996a.
- Goodchild, M. F., "Communicating the Results of Accuracy Assessment: Metadata, Digital Libraries, and Assessing Fitness for Use," *Proceedings, Second International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Fort Collins, CO, 1996b.
- Goodchild, M. F., "Geographic Information Systems and Disaggregate Transportation Planning," *Geographical Systems*, volume 5, pages 19-44, 1998.
- Goodchild, M. F., Battenfield, B., and Wood, J., "Introduction to Visualizing Data Validity," In Hearnshaw H. Unwin D (eds) *Visualization in Geographic Information Systems*, Chichester, John Wiley & Sons, pages 141-149, 1994.
- Goodchild, M. F., Guoqing, S., and Shiren, Y., "Development and Test of an Error Model for Categorical Data," *International Journal of Geographic Information Systems*, Volume 6, Number 2, Pages 87-104, 1992.
- Goodchild, M. F. and Hunter, G. J., "A Simple Positional Accuracy Measure for Linear Features," *International Journal of Geographical Information Systems*, Volume 11, No. 3, pages 299-306, 1997.
- Goodchild, M. F. and Jeansoulin, R. (Eds.), *Data Quality in Geographic Information*, Editions Hermes, Paris, 1998.

Harbaugh, J. W. and Martinez, P. A., "Two Major Problems in Representing Geological Well Data and Seismic Data in Petroleum-Bearing Regions via 3-D Geographic Information Systems," Chapter 19, *Three-Dimensional Modeling with Geoscientific Information Systems*, Editor Turner, A. K., NATO ASI Series, Kluwer Academic Publishers, pages 291-302, 1989.

Heesom D. and Lamine M., "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, 2001.

Hensley, S., Munjy, R., and Rosen, P., Maune, "Interferometric Synthetic Aperture Radar (IFSAR)," in *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, Editor Maune, F. D., American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland 20814, pages 142-200, 2001.

Heuvelink, G. B. M., *Error Propagation in Quantitative Spatial Modeling*, KNAG, University of Utrecht Pub. 163, 1993.

Heuvelink, G. B. M., *Geographic Information Systems – Principles and Technical Issues – Volume 1 – Detecting and Evaluating Errors by Geographical Methods*, John Wiley & Sons, New York, Chapter 14, Pages 207-217, 1999.

Hiegl, E., "A GIS Journey: From the Allegheny National Forest to the New River," *Master's Internship Report*, Appalachian State University, 2001.

Highway Location Reference Methods, *NCHRP Synthesis of Highway Practice No. 21*, Highway Research Board, 1974.

Hill, J.M., L.A.Graham, R.H. Henry, D.M. Cotter, A. Ding, and D. Young, "Wide-Area Topographic Mapping and Applications Using Airborne Light Detection and Ranging (LIDAR) Technology," *Photogrammetric Engineering & Remote Sensing, ASPRS, Journal of The American Society for Photogrammetry and Remote Sensing*, Vol. 66, No. 8, August 2000.

Hodgkiss, A. G., *Understanding Maps*, Dawson, Folkstone, 1981.

Holder, W. J., "Assessment of Riparian Buffers Effectiveness Using GIS Technology: Laurel Springs, Ashe County North Carolina," *Master's Thesis*, Appalachian State University, 1992.

Houlding, S., "The Application of 3-D Computer Modeling Techniques to Mining," Chapter 20, *Three-Dimensional Modeling with Geoscientific Information Systems*, Editor Turner, A. K., NATO ASI Series, Kluwer Academic Publishers, pages 303-325, 1989.

Huff, C. L. and Noll, T. G., "Sonar" in *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, Editor Maune, F. D., American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland 20814, pages 307-332, 2001.

Huh, Y. U., Keller, F. R., Redman, T. C., and Watkins, A. R., "Data Quality," *Information and Software Technology*, Volume 32, No. 8, pages 559-565, 1990.

Hunter, G. J., "New Tools for Handling Spatial Data Quality: Moving from Academic Concepts to Practical Reality," *URISA Journal*, Volume 11, No. 2, pages 25-34, 1999.

Hunter, G. J. and Goodchild, M. F., "Dealing with Error in Spatial Databases – A Simple Case Study," *Photogrammetric Engineering and Remote Sensing*, Volume 61, No. 5, pages 529-537, 1995a.

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

Hunter, G. J. and Goodchild, M. F., "A New Model for Handling Vector Data Uncertainty in Geographic Information Systems," *Journal of the Urban and Regional Information Systems Association*, Volume 8, No. 1, pages 51-57, 1996.

Jaala, N. and Anders, K.-H, "Acquisition of 3D Urban Models by Analysis of Aerial Images, Digital Surface Models, and Existing 2D Building Information," *Proceedings of SPIE --- The International Society for Optical Engineering, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, Volume 3072, pages 212-222, Orlando, Florida, April 1997.

Jamar Technology, Inc., <http://www.jamartech.com/dmimanual.pdf>, Modified February 13, 2002, Accessed March 6, 2002.

Karaak, J. M., "Computer-Assisted Cartographical 3D Imaging Techniques," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 99-114, 1989.

Karimi, H. A., Hummer, J. E., and Khattak, A. J., "Collection and Presentation of Roadway Inventory Data," NCHRP Report 437, National Research Council, Transportation Research Board, Washington, D.C. page 11, 2000.

Karimi H. A., A. J. Khattak, and J. E. Hummer, "Evaluation of Mobile Mapping Systems for Roadway Data Collection," *ASCE, Journal of Computing in Civil Engineering*, Vol. 14, No. 3, July 2000, pp. 168-173.

Kavanagh, B.F., *Surveying with Construction Applications*, 3rd Edition, Prentice-Hall, Inc., New Jersey, Columbus, Ohio, 1997.

Kelk, B., "3-D Modeling with Geoscientific Information Systems: the Problem," Chapter 4, *Three-Dimensional Modeling with Geoscientific Information Systems*, Editor Turner, A. K., NATO ASI Series, Kluwer Academic Publishers, pages 29-37, 1989.

Kennedy, M., *The Global Positioning System and GIS*, Ann Arbor Press, Inc., Chelsea, Michigan, 1996.

Kennie, T. J. M. and Petrie, G., "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, 1990.

Kiel, D., Rasdorf, W., Shuller, E., and Poole, R., "Linear Referencing System for North Carolina Department of Transportation," *Transportation Research Record 1660*, Transportation Research Board, National Research Council, Washington, D. C., pages 108-113, 1999.

Khisty C. J., *Transportation Engineering: An Introduction*, Prentice Hall, Englewood Cliffs, New Jersey 07632, 1990.

Kim, T. and Muller, J.-P. A., "Automated Building Height Estimation and Object Extraction from Multiresolution Imagery," *Proceedings of SPIE --- The International Society for Optical Engineering, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, Volume 2486, pages 267-276, Orlando, Florida, April 1995.

Kubik, K. and Botman, A., "Interpolation Accuracy for Topographic and Geological Surfaces," *ITC J.*, 1976(2), pages 236-274, 1976.

Kubik, K. and Roy, B. C., Digital Terrain Model Workshop Proceedings, Ohio State University, Columbus, Ohio, page 227, 1986.

Kumler, M., "A Quantitative Comparison of Regular and Irregular Digital Terrain Models," *GIS/LIS'90*, Volume 1, pages 255-264, 1990.

Lammi, J., "Automatic Building Extraction Using A Combination of Spatial Data and Digital Photogrammetry," *Proceedings of SPIE --- The International Society for Optical Engineering, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, Volume 3072, pages 223-230, Orlando, Florida, April 1997.

- Lancaster, P. and Salkauskas, K., *Curve and Surface Fitting: An Introduction*, Academic Press, London, page 280, 1986.
- Larson, C. M., "Creating Buffers on Surfaces," *Presentation at North Carolina 2003 GIS Conference*, Winston-Salem, North Carolina, February 2003.
- Leenaers, H., Burrough, A. P., and Okx, P. J., "Efficient Mapping of Heavy Metal Pollution on Floodplains by Co-kriging from Elevation Data," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 37-50, 1989.
- Ley, R., "Accuracy Assessment of Digital Terrain Models," *Auto Carto London*, 1, pages 455-464, 1986.
- Li, R., "Generation of Geometric Representations of 3-D Objects in CAD/CAM by Digital Photogrammetry," *ISPRS Journal of Photogrammetry and Remote Sensing*, 48(5), pages 2-11, 1993.
- Li, R., "Data Structures and Application Issues in 3-D Geographic Information Systems," *Geomatica*, Volume 48, Number 3, pages 209-224, Summer 1994.
- Li, R., Qian, L., Chen, Y., and Hughes, D., "Approaching 3-D Data Structures in GIS," *GIS 94 and ISPRS Com. II Symposium*, Ottawa, pages 1669-1680, June 1994.
- Li, Zhilin, *Sampling Strategy and Accuracy Assessment for Digital Terrain Modeling*, Ph. D. thesis, University of Glasgow, Glasgow, page 299, 1990.
- Li, Zhilin, "Variation of the Accuracy of Digital Terrain Models with Sampling Interval," *Photogrammetry Records*, 14(79), pages 113-127, 1992.
- Li, Zhilin, "Theoretical Models of the Accuracy of Digital Terrain Models: An Evaluation and Some Observations," *Photogrammetry Records*, October-issue, 1993.
- Limp, W. F. and Cothren, J., "3-D Emerging Photogrammetry Technologies Make It Easier Than Ever to Deliver Geospatial Solutions," *GeoWorld*, pages 30-33, January 2003.
- MacEachren, A. M., Howard, D., Wyss, M. von, Asko, D., and Taormino, T., "Visualizing the Health of Chesapeake Bay: An Uncertain Endeavor," *Proceedings GIS/LIS 93 Minneapolis*, pages 449-458, 1993.
- Makarovic, B., "Information Transfer in Construction of Data from Sampled Points," *Photogrammetria*, 28(4), pages 111-130, 1972.
- Massasati, Ahmad S., "An Exercise in Digital Elevation/Terrain Models: From Point to Mathematics," *NCGLA Core Curriculum in GIScience*, <http://www.ncgis.ucsb.edu/giscc/extra/e001/eoo1.html>, 2000, accessed June 26, 2001.
- Maune, D., *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, Editor Maune, F. D., American Society for Photogrammetry and Remote Sensing, 5410 Grosvenor Lane, Suite 210, Bethesda, MD 20814, 2001.
- Maune, D. F., Kopp, S. M., Crawford, C. A., and Zervas, C. E., "Introduction," in *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, Editor Maune, F. D., American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland 20814, pages 1-31, 2001.
- McCormac, J., *Surveying*, 4th Edition, John Wiley & Sons, Inc. New York, NY, 1999.
- McLaren, A. R. and Kennie, J. M. T., "Visualization of Digital Terrain Models: Techniques and Applications," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 79-98, 1989.
- Mead, R. A. and Szajgin, J., "Landsat Classification Accuracy Assessment Procedures," *Photogrammetric Engineering and Remote Sensing*, Volume 48, pages 139-141, 1982.

- Miller, C. and LaFlamme, R. A., "The Digital Terrain Model ---- Theory and Applications," *Photogrammetric Engineering* 24(3), pages 433-442, 1958.
- Miller, H. J. and Shaw, S. L., *Geographic Information Systems for Transportation: Principles and Applications*, Oxford University Press, New York, 2001.
- Mitasova, H., Mitas, L., Brown, W., Gerdes, D. P., Kosinovsky I., Baker, T., "Modeling Spatially and Temporally Distributed Phenomena: New Methods and Tools for GRASS GIS," *International Journal of Geographical Information Systems*, Volume 9, pages 433-446, 1995.
- Mortenson, M. E., *Geometric Modeling*, John Wiley, 1985.
- Moffitt, F.H., J.D.Bossler, *Surveying*, 10th Edition, Addison Wesley Longman, Inc., 1998.
- Molander, W. C., "Photogrammetry," in *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, Editor Maune, F. D., American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland 20814, pages 121-141, 2001.
- Narumalani, S., Zhou, Y., and Jensen, J. R., "Application of Remote Sensing and Geographic Information Systems to the Delineation and Analysis of Riparian Buffer Zones," *Aquatic Botany*, Volume 58, pages 393-409, 1997.
- National Mapping Program Standards, *National Mapping Program Technical Instructions, Part 2, Specifications for Digital Elevation Models*, <http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/dem/2DEM0198.PDF>, January 1998, Accessed May 2003.
- NC Floodplain Mapping Program, <http://www.ncfloodmaps.com>, Accessed May 2003.
- NCCTS, "LIDAR and Digital Elevation Data Fact Sheet," North Carolina Cooperating Technical State, http://www.ncfloodmaps.com/pubdocs/lidar_final_jan03.pdf, January 2003, Accessed May 2003.
- NCCTSMP, "Issue 5: Quality Control of LIDAR Elevation Data in North Carolina," http://www.ncgs.state.nc.us/flood/ip05_nc_lidar_qc.dbf, October 2001.
- NCGS, "Digital Orthophoto Quarter-Quadrangles (DOQQ): 1:12,000-Scale Images," NC Geological Survey, <http://www.geology.enr.state.nc.us/doqq.html>, Last Modified February 27, 2001, Accessed August 25, 2003, 2001.
- Nicholas, J. and Driel, V., "Three Dimensional Display of Geologic Data," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 1-10, 1989.
- Noman, S. N., Nelson, J. E., and Zundel, K. A., "Review of Automated Floodplain Delineation from Digital Terrain Models," *Journal of Water Resources Planning and Management*, American Society of Civil Engineers, Volume 127, No. 6, pages 394-402, 2001.
- NCCTSMP, "Overview of the North Carolina Flood Mapping Program," North Carolina Cooperating Technical State Flood Mapping Program, <http://www.ncfloodmaps.com/pubdocs/NCFPMPHndOut.pdf>, October 2, 2000.
- Nu-Metrics, Inc., http://www.nu-metrics.com/pdf/nite_man_2002.pdf, Copyright Nu-Metrics 2002, Accessed March 6, 2002.
- Nu-Metrics, Inc., *NITESTAR[®], Accurate Distance Measuring from Any Vehicle, Model No. NS-50/NS-60*, Nu-Metrics, Inc. Uniontown, PA 15401, 1998.
- Nyerges, T. L. and Dueker, K. J., *Geographic Information Systems in Transportation*, Office of Planning HPN-22, Federal Highway Administration, U. S. Department of Transportation, Washington, DC, 1988.

Office of Transportation Technologies, *Fact of the Week: Transportation and the GDP*, <http://www.ott.doe.gov/facts/archives/fotw223.shtml>, last updated 07/08/2002, visited 10/06/2002.

O'Neill, Wende, "UDOT Enterprise Location Referencing," presented at *Enterprise Location Referencing Systems: Policies, Procedures, and Standards for Implementation*, workshop notes, 1996.

Osborn, K., List, J., Gesch, D., Crowe, J., Merrill, G., Constance, E., Mauck, J., Lund, C., Caruso, V., and Kosovich, J., "National Digital Elevation Program (NDEP)," in *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, Editor Maune, D. F., American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland 20814, pages 83-120, 2001.

Papacostas, C. S. and Prevedouros, P. D., *Transportation Engineering & Planning*, 3rd edition, Prentice-Hall, Inc. Upper Saddle River, New Jersey 07458, 2001.

Paradise, J., Beard, M. K., "Visualization of Data Quality for the Decision-maker: A Data Quality Filter," *Journal of the Urban and Regional Information Systems Association*, Volume 6, pages 25-34, 1994.

Peucker, T. K. et. al., "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, 1978.

Quiroga, C. A., "Accuracy of Linearly Referenced Data by Using Geographic Information Systems," *Transportation Research Record 1660*, Transportation Research Board, National Research Council, Washington, D. C., pages 100-107, 1999.

Rao, P. V., *Statistical Research Methods in the Life Sciences*, Brooks/Cole Publishing Company, 511 Forest Lodge Road, Pacific Grove, CA 93950, 1998.

Raper, J. F., "Key 3D Modeling Concepts for Geoscientific Analysis," Chapter 15, *Three-Dimensional Modeling with Geoscientific Information Systems*, Editor Turner, A. K., NATO ASI Series, Kluwer Academic Publishers, pages 215-232, 1989.

Raper, J. and Kelk, B., "Three-Dimensional GIS," In: *Geographical Information Systems: Principles and Applications*, Longman, London, Volume 1, pages 299-317, 1991.

Raper, J. F. and Wainwright, D. E., "The Use of the Geotechnical Database GEOSHARE in Site Investigation Data Management," *Quarterly Journal of Engineering Geology* 20, pages 221-230, 1987.

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

Rasdorf, W. J. and Cai, H., "A GIS Data Quality Case Study: Accuracy Comparison of Length Measurements Using GIS/NED and DMI," Technical Report, Department of Civil Engineering, North Carolina State University, 2001.

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

Rasdorf, W., Cai, H., Tilley, C., Brun, S., Karimi, H., and Robson, F., "Accuracy Assessment of Road Length Measurement Using DEM," *Journal of Surveying Engineering*, American Society of Civil Engineers, submitted, 2003b.

Requicha, A. A. G., "Representation of Rigid Solids: Theory, Methods, and Systems," *Computing Surveys* 12:4, pages 437-464, 1980.

Requicha, A. A. G. and Voelcker, H. B., "Solid Modeling: A Historical Summary and Contemporary Assessment," *IEEE Computing Graphics Applications* 2, 2, pages 9-24, March 1982.

- Rhind, D. W., "Geographical Information Systems in Britain," *Quantitative Geography*, Edited by Wrigley, N. and Bennett, R. J., 17-39, Routledge and Kegan Paul, London, 1981.
- Rhind, D. W., "A GIS Research Agenda," *International Journal of GIS*, 2, pages 23-28, 1988.
- Rhind, D. W., "Spatial Data Handling in the Geoscience," Chapter 3, *Three-Dimensional Modeling with Geoscientific Information Systems*, Editor Turner, A. K., NATO ASI Series, Kluwer Academic Publishers, pages 13-27, 1989.
- Ries, T., "Design Requirements for Location as A Foundation for Transportation Information Systems," *Proceedings, AASHTO Symposium on Geographic Information Systems in Transportation*, Portland, OR, 1993.
- Roux, M., Hsieh, Y. C., and McKeown, D. M. Jr., "Performance Analysis of Object-Space Matching for Building Extraction Using Several Images," *Proceedings of SPIE --- The International Society for Optical Engineering, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, Volume 2486, pages 277-297, Orlando, Florida, April 1995.
- Rüeger, J.M., *Electronic Distance Measurement*, 4th Edition, Springer-Verlag Berlin Heidelberg New York, Germany, 1996.
- Samet, H., "Connected Component Labeling using Quadrees," *Journal ACM* 28, 3, pages 487-501, July 1981.
- Samet, H. and Webber, R. E., "Hierarchical Data Structures and Algorithms for Computer Graphics, Part I: Fundamentals," *IEEE Computing Graphics Applications* 8, 3, pages 46-68, May 1988.
- Scarponcini, P., "Generalized Model for Linear Referencing," *Proceeding of the Seventh International Symposium, ACM GIS'99*, pages 53-59, 1999.
- Schofield, W., *Engineering Surveying: Theory and Examination Problems for students*, 4th Edition, Butterworth-Heinemann Ltd., Oxford, Great Britain, 1993.
- SDTS, "American National Standard for Information Systems – Spatial Data Transfer Standard (SDTS) – Part 1. Logical Specifications," Draft for Review, Published by American National Standard Institute, New York, page 5, 1997.
- Serr, David, "Use of LIDAR in creating Accurate Terrain Elevation Models for Floodplain Mapping," *ES771 Term Project*, December 7, 2000.
- Shearer, J. W., "The Accuracy of Digital Terrain Model," *Terrain Modeling in Surveying and Civil Engineering*, T. J. M. Kenzie and G. Petrie, editors, Whittles Publication, Pages 315-334, 1990.
- Sims, D. L., "Applications of 3-D Geoscientific Modeling for Hydrocarbon Exploration," Chapter 18, *Three-Dimensional Modeling with Geoscientific Information Systems*, Editor Turner, A. K., NATO ASI Series, Kluwer Academic Publishers, pages 285-289, 1989.
- Smith, R. D. and Paradis, R. A., "Three-dimensional GIS for the Earth Science," *Three Dimensional Applications in Geographic Information Systems*, Edited by Raper, J., Taylor and Francis Ltd., Basingstoke, Hampshire, Chapter 11, pages 149-154, 1989.
- Smith, D. R. and Paradis, A. R., "Three-Dimensional GIS for the Earth Sciences," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 149-154, 1989.
- Spatial Data Transfer Standard (SDTS), *Spatial Data Transfer Standard* (Federal Information Processing Standards FIPS-173), U. S. Geological Survey, Reston, VA, 1992.
- Sickle, J. V., *GPS for Land Surveyors*, Ann Arbor Press, Inc, Chelsea, Michigan, 1996.

- Stough, R. R., *Intelligent Transportation Systems: Cases and Policies*, Edward Elgar Publishing Limited, Glensanda House, Montpellier Parade, Cheltenham, Glos GL501UA, UK, 2001.
- Stratton, W. L., "Generating Variable Width Stream Buffers Using A Geographic Information System," *Master's Thesis*, University of North Carolina at Charlotte, Charlotte, North Carolina, 1993.
- Sun-Lab Technology, Inc., <http://www.sun-lab.com/pdfs/brochure.pdf>, Copyright January 2001, Accessed March 6, 2002.
- Tate, C. E., Maidment, R. D., Olivera, F., and Anderson, J. D., "Creating a Terrain Model for Floodplain Mapping," *Journal of Hydrologic Engineering*, American Society of Civil Engineers, Volume 7, No. 2, pages 100-108, 2001.
- The Pavement Management Section, Maine Department of Transportation, Feb 2002, <http://www.state.me.us/mdot/planning/pavement/pmspage.htm>.
- Thill, J. -C., "Geographic Information Systems for Transportation in Perspective," *Geographic Information Systems in Transportation Research*, Elsevier Science Ltd., 2000.
- Torlegard, K., Ostman, A. and Lindgren, R., "A Comparative Test of Photogrammetrically Sampled Digital Elevation Models," *Photogrammetria*, 41(1), pages 1-16, 1986.
- Toomey, M., *Digital Elevation Model Workshop Proceedings*, Edmonton, Alberta, page 231, 1984.
- Tsai, V., "Delaunay Triangulations in TIN Creation: An Overview and A Linear-Time Algorithm," *International Journal of Geographic Information Systems*, 7(6), pages 365-388, 1989.
- Turner, A. K., "LIDAR Provides Better DEM Data," *GeoWorld*, page 30, November 2000.
- Turner A. K., "The Role of Three-dimensional Geographic Information Systems in Subsurface Characterization for Hydrogeological Applications," *Three Dimensional Applications in Geographic Information Systems*, Edited by Raper, J., Taylor and Francis Ltd., Basingstoke, Hampshire, Chapter 8, pages 115-127, 1989a.
- Turner, A. K., "Preface," Chapter 1, *Three-Dimensional Modeling with Geoscientific Information Systems*, Editor Turner, A. K., NATO ASI Series, Kluwer Academic Publishers, pages 1-5, 1989b.
- Turner, A. K., "The Role of Three-Dimensional Geographic Information Systems in Subsurface Characterization for Hydrogeological Applications," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 115-128, 1989c.
- Turner, A. K., "Applications of Three-Dimensional Geoscientific Mapping and Modeling Systems to Hydrogeological Studies," Chapter 21, *Three-Dimensional Modeling with Geoscientific Information Systems*, Editor Turner, A. K., NATO ASI Series, Kluwer Academic Publishers, pages 327-364, 1989d.
- Unesco, "Topology and Spatial Relationships," *Unesco Training Module Part G*, http://ioc.unesco.org/oceanteacher/resourcekit/Module2/GIS/Module/Module_g/index.html, Last Modified 1999, Accessed November 19, 2002, 1999.
- Unger, D. J., Liberty M. L., Phillips, D. J. and Wright, E. B., "Creating a 3-Dimensional Transect of the Earth's Crust from Craton to Ocean Basin across the N. Appalachian Orogen," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 137-148, 1989.
- USGS, *Standards for Digital Elevation Models*, Department of the Interior, U.S. Geologic Survey, 1997.
- USGS, "Topographic Maps," <http://mcmweb.er.usgs.gov/topomaps/>, accessed August 16, 2001, 2001a.
- USGS, "Digital Elevation Models," http://rockyweb.cr.usgs.gov/elevation/dpi_dem.html, Last modified November 2001, Accessed March 2003, 2001b.

USGS, "USGS Digital Elevation Data Model," http://edcwww.cr.usgs.gov/glis/hyper/guide/usgs_dem, Accessed March 2003a.

USGS, "National Elevation Dataset," November 2002, <http://gisdata.usgs.net/ned/>, Accessed March 26, 2003b.

USGS, "National Elevation Dataset," USGS Mapping Applications Center, December 22, 2000, <http://mac.usgs.gov/mac/isb/pubs/factsheets/fs14899.html>, Accessed October 15, 2001c.

USGS, "Digital Elevation Model," Western Geographic Science Center, March 28, 2001, <http://craterlake.wr.usgs.gov/dem.shtml>, Accessed June 2, 2001d.

USGS, "Digital Orthophoto Quadrangle (DOQs)," <http://edc.usgs.gov/products/aerial/doq.html>, Last Modified March 14, 2003, Accessed August 25, 2003e.

USGS, "Digital Orthophoto Quadrangle," <http://www-wmc.wr.usgs.gov/doq/>, Last Modified June 6, 2003, Accessed August 25, 2003f.

USGS, "National Mapping Program Standards -- Digital Orthophoto Standards, Part 2: Specifications, Standards for Digital Orthophotos," <http://rmmcweb.cr.usgs.gov/public/nmpstds/acrodocs/doq/2DOQ1296.PDF>, Accessed August 25, 2003g.

USGS, "Accuracy of the National Elevation Dataset," <http://edcnts12.cr.usgs.gov/ned/Accuracy.asp>, Last Modified on November 6, 2002, Accessed on August 23, 2003. 2002.

USGS, "US GeoData Digital Elevation Models, Fact Sheet 040-00 (April 2000)," <http://mac.usgs.gov/mac/isb/pubs/factsheets/fs04000.html#accuracy>, Last modified: December 31, 2002, Accessed May 2003.

USGS, *Data Users Guide 5*, US Geological Survey, 1993, <ftp://mapping.usgs.gov/pub/ti/DEM/demguide/>, Accessed August, 2003, 1993.

USGS, *Standards for Digital Elevation Models*, Department of Interior, U.S. Geological Survey, Reston, VA, 1997.

Veregin, H., *Geographical Information Systems – Principles and Technical Issues*, Volume 1, "Data Quality Parameters," John Wiley & Sons, Chapter 12, pages 177-189, 1999.

Vonderohe, A. P., Travis, L., Smith, R. L., and Tsai, V., "Adaptation of Geographic Information Systems for Transportation," *NCHRP Report 359*, Transportation Research Board, National Research Council, Washington, DC, 1993.

Vonderohe, A. P., Chou, C.-L., Sun, F., and Adams, T., "Results of A Workshop on A General Data Model for Linear Referencing Systems," *AASHTO Symposium on Geographic Information Systems in Transportation*, 1995.

Wang, X., Hanson, A. R., Collins, R. T., and Dehart, J., "Surface Microstructure Extraction from Multiple Aerial Images," *Proceedings of SPIE --- The International Society for Optical Engineering, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, Volume 3072, pages 251-262, Orlando, Florida, April 1997.

Waters, N. M., "Transportation GIS: GIS-T," *Geographic Information Systems, Volume 2: Management Issues and Applications*, Wiley, New York, pages 827-844, 1999.

Wechsler, S. P., "Digital Elevation Model (DEM) Uncertainty: Evaluation and Effect on Topographic Parameters," *Proceedings of ESRI User Conference*, 1999.

Worawat, T. and Rasdorf, W., "Evaluation of Spatial Measurement Technologies," Master Project Report, Department of Civil Engineering, North Carolina State University, January 2003.

Wu, C. V. and Buttenfield, B. P., "Spatial Data Quality and Its Evaluation," *Computers, Environment and Urban Systems*, Volume 18, No. 3, pages 153-164, 1994.

Xiang, W., "GIS-Based Riparian Buffer Analysis: Injecting Geographic Information into Landscape Planning," *Landscape and Urban Planning*, Volume 34, pages 1-10, 1996.

Xiang, W. and Stratton, W. L., "The B-Function and Variable Stream Buffer Mapping: A Note on 'A GIS Method for Riparian Water Quality Buffer Generation'," *International Journal of Geographic Information Systems*, Volume 10, Number 4, pages 499-510, 1994.

Youngmann, C., "Spatial Data Structures for Modeling Subsurface Features," *Three Dimensional Applications in Geographical Information Systems*, Editor Raper, J. F., Taylor and Francis, London, pages 129-136, 1989.

APPENDICES

APPENDIX A PROGRAM CODE FOR EXTRACTION OF FTSEGS FROM THE LINK-NODE SYSTEM

'Program Name: FTSegExtraction

'This program extracts FTSEgs from a link-node system by identifying links that belong to one FTSeg and 'assigning them with the same MERGE value. Later on, these links can be merged together to create a FTSeg, 'using the tool of dissolving features based on attribute from ArcGIS.

'This program requires two tables. One is the attribute table of the links (all NC Route links, for example) to be 'merged. The other is the table containing all other links (all Interstate highway links and US Route links, for 'example) that help to determine if the current link belongs to the current FTSeg, or a new FTSeg should be 'created.

'This program was developed by Hubo Cai during the summer of 2003.

Option Explicit

'Main sub

Sub FTSegExtraction()

'Create workspace and open tables

Dim pworkspacefactory As IWorkspaceFactory

Set pworkspacefactory = New ShapefileWorkspaceFactory

Dim pfeatureworkspace As IFeatureWorkspace

Set pfeatureworkspace = pworkspacefactory.OpenFromFile ("S:\PALRS\users\Hubo\road_meter\JOHNSTON",
0)

'Open the table for NC route links in Johnston County

Dim myintertable As ITable

Set myintertable = pfeatureworkspace.OpenTable ("NC_JOHNSTON")

'Open the table for Interstate and US route links in Johnston County

Dim myusnctable As ITable

Set myusnctable = pfeatureworkspace.OpenTable ("INTERUS_JOHNSTON")

'Define tablesort

Dim pTableSort As ITableSort

Set pTableSort = New esriCore.TableSort

Dim pQueryFilter As IQueryFilter

Set pQueryFilter = New QueryFilter

Dim DIMENSION As Integer

DIMENSION = 11

'Define a dynamic array to hold all route values

Dim ROUTE () As Long

ReDim ROUTE(DIMENSION)

ROUTE (0) = 30000027

ROUTE (1) = 30000039

ROUTE (2) = 30000042

ROUTE (3) = 30000050

ROUTE (4) = 30000055

ROUTE (5) = 30000096

ROUTE (6) = 30000210

ROUTE (7) = 30000222

ROUTE (8) = 30000231

ROUTE (9) = 30000242

ROUTE (10) = 30600042

Dim routecount As Integer

routecount = UBound(ROUTE)


```

Dim i AS Integer

'Start with MERGE 0
Dim mergevalue As Integer
mergevalue = 0

For i = 0 To routecount
'Query the table to find all links belonging to the same Route
pQueryFilter.whereClause = "ROUTE1 = " & ROUTE(i)

'Sort the table based on the BEGMP1
With pTableSort
.Fields = "BEGMP1"
.Ascending("BEGMP1") = True
.Set .QueryFilter = pQueryFilter
.Set .Table = myintertable
End With
pTableSort.Sort Nothing

Dim pCursor As ICursor
Set pCursor = pTableSort.Rows
Dim rowcount As Integer
'Define pindex to refer to the column MERGE
Dim pindex As Integer
pindex = myintertable.FindField("MERGE")

Dim pRow As IRow
Set pRow = pCursor.NextRow
'Variable CFNode indicates the From Node of the current link
Dim CFNode As Integer

Do While Not pRow Is Nothing
    CFNode = pRow.value(3)
    'Assign the current mergevalue to the current link
    pRow.value(pindex) = mergevalue
    pRow.Store

    'Based on the From Node of the current link, query the current table to see if
    'some links either start or end at this node
    Dim iwhereclause As String
    iwhereclause = "FNODE_ = " & CFNode & "OR TNODE_ = " & CFNode
    Dim intercursor As ICursor
    Set intercursor = AttributeQuery(myintertable, iwhereclause)
    Dim interrow As IRow
    Set interrow = intercursor.NextRow

    rowcount = 0
    'Check if there are exactly one NC Route link that start from the current FNode
    'If there are 0 or more than 1 links, next link would have a different merge value

    Do While Not interrow Is Nothing
        rowcount = rowcount + 1
        Set interrow = intercursor.NextRow
    Loop
    'MsgBox "how many rows are selected: " & rowcount

```

```

If rowcount <> 2 Then
    mergevalue = mergevalue + 1
Else
    'Check if there is an Interstate or US route link that start from the current FNode
    'If there are, next NC Route link would have a different merge value
    Dim UNwhereclause As String
    UNwhereclause = "FNODE_ = " & CFNode & " OR TNODE_ = " & CFNode
    Dim usncursor As ICursor
    'Call the AttributeQuery function to query table
    Set usncursor = AttributeQuery(myusnctable, UNwhereclause)
    Dim usncrow As IRow
    Set usncrow = usncursor.NextRow
    Dim usncrowcount As Integer
    usncrowcount = 0
    Do While Not usncrow Is Nothing
        usncrowcount = usncrowcount + 1
        Set usncrow = usncursor.NextRow
    Loop

    If usncrowcount > 0 Then
        mergevalue = mergevalue + 1
    Else
        End If
    End If
End If

Set pRow = pCursor.NextRow
Loop
Next
End Sub

'Public function to query table, given a whereclause
Public Function AttributeQuery(pTable As esriCore.ITable, Optional whereClause As String = "")
As esriCore.ICursor

    Dim pQueryFilter As esriCore.IQueryFilter
    Dim pCursor As esriCore.ICursor
    ' create a query filter
    Set pQueryFilter = New esriCore.QueryFilter

    ' create the where statement
    pQueryFilter.whereClause = whereClause

    ' query the table passed into the function and use a cursor to hold the results
    Set pCursor = pTable.Search(pQueryFilter, False)
    Set AttributeQuery = pCursor
End Function

```

APPENDIX B PROGRAM CODE FOR SNAPPING LIDAR POINTS TO LINEAR FEATURES

Program Name: Snapping

This program snaps LIDAR 3-D points to line features based on a given buffer size. It checks each LIDAR point with each line feature to see if a LIDAR point is in the buffer. If it is, the point is snapped to the line. The resulting point is the point on the line feature, but closest to the LIDAR point. The X/Y coordinates of this new point are determined by the line feature. Its Z coordinate (elevation) is the same as the LIDAR point being snapped.

This program was developed by Hubo Cai during the summer of 2003.

Option Explicit

Sub Snapping ()

'set work space and open shapefiles and table

Dim pworkspacefactory As IWorkspaceFactory

Set pworkspacefactory = New ShapefileWorkspaceFactory

Dim pfeatureworkspace As IFeatureWorkspace

Set pfeatureworkspace = pworkspacefactory.OpenFromFile ("s:\palrs\users\hubo\Nroad_english\VANCE", 0)

'Open line feature shapefile

Dim linefeatureclass As IFeatureClass

Set linefeatureclass = pfeatureworkspace.OpenFeatureClass ("Dis_int_vance")

'Open LIDAR point shapefile

Dim pointfeatureclass As IFeatureClass

Set pointfeatureclass = pfeatureworkspace.OpenFeatureClass ("LIDAR_30ft_vance")

'Open the target table to store the sanpped points

Dim mytable As ITable

Set mytable = pfeatureworkspace.OpenTable ("snap_vance")

'Set datasets and assign the feature classes opened earlier to them

Dim linedataset As IDataset

Set linedataset = linefeatureclass

Dim pointdataset As IDataset

Set pointdataset = pointfeatureclass

'Define all the intermediate variables that will be used in the program

Dim linefeature As IFeature

Dim myline As IPolycurve

Dim mypoint As IPoint

Dim pointfeature As IFeature

Dim linetotal As Long

Dim pointtotal As Long

Dim linecount As Long

Dim pointcount As Long

Dim ibuffer As ITopologicalOperator

Dim istartbuffer As ITopologicalOperator

Dim iendbuffer As ITopologicalOperator

Dim ppolygon As IPolygon

Dim pstartpolygon As IPolygon

Dim pendpolygon As IPolygon

Dim iib As Boolean

Dim instartbuffer As Boolean

Dim inendbuffer As Boolean

Dim pointinbuffer As IRelationalOperator

Dim pointinstartbuffer As IRelationalOperator

Dim pointinendbuffer As IRelationalOperator

```

Dim myrow As IRow
Dim outpoint As IPoint
Dim distalongcurve As Double
Dim distfromcurve As Double
Dim rightside As Boolean
Dim mystartpoint As IPoint
Dim myendpoint As IPoint
Dim startelevationsum As Double
Dim startelevationcount As Integer
Dim endelevationsum As Double
Dim endelevationcount As Integer
Dim overlapstart As Boolean
Dim overlapend As Boolean
Dim tostartorenddist As Double
Dim havingpointinstartbuffer As Boolean
Dim havingpointinendbuffer As Boolean

```

```

'Count the total number of lines and points for the line and point feature classes
linetotal = linefeatureclass.FeatureCount(Nothing)
pointtotal = pointfeatureclass.FeatureCount(Nothing)

```

```

'Start the FOR loop to snap LIDAR points to corresponding lines
For linecount = 0 To linetotal - 1

```

```

    'Variables overlapstart and overlapend are used to indicate if the resulting
    'point on the line feature is the start or end points of that line feature
    overlapstart = False
    overlapend = False
    havingpointinstartbuffer = False
    havingpointinendbuffer = False

```

```

    'Get a line feature from the line feature class
    Set linefeature = linefeatureclass.GetFeature(linecount)
    Set myline = linefeature.Shape
    Set ibuffer = linefeature.Shape
    'Create a 12 ft buffer to the current line feature
    Set ppolygon = ibuffer.Buffer(12)
    Set pointinbuffer = ppolygon
    'Get the start and the end point of the current line feature
    Set mystartpoint = myline.FromPoint
    Set myendpoint = myline.ToPoint

```

```

    'Create a row in the target table for the start point
    'Assign it with elevation value of 999999
    Set myrow = mytable.CreateRow
    myrow.Value (1) = mystartpoint.X
    myrow.Value (2) = mystartpoint.Y
    myrow.Value (3) = 999999
    myrow.Value (5) = 0
    myrow.Value (4) = myline.Length
    myrow.Value (6) = linefeature.Value (9)
    myrow.Value (7) = linefeature.Value (2)
    myrow.Value (8) = linefeature.Value (4)
    myrow.Value (9) = linefeature.Value (5)
    myrow.Value (10) = linefeature.Value (7)
    myrow.Value (11) = linefeature.Value (8)
    myrow.Store

```

```

'Create a row for the end point
Set myrow = mytable.CreateRow
  myrow.Value (1) = myendpoint.X
  myrow.Value (2) = myendpoint.Y
  myrow.Value (3) = 999999
  myrow.Value (5) = myline.Length
  myrow.Value (4) = myline.Length
  myrow.Value (6) = linefeature.Value (9)
  myrow.Value (7) = linefeature.Value (2)
  myrow.Value (8) = linefeature.Value (4)
  myrow.Value (9) = linefeature.Value (5)
  myrow.Value (10) = linefeature.Value (7)
  myrow.Value (11) = linefeature.Value (8)
myrow.Store

'Check with all LIDAR points
For pointcount = 0 To pointtotal - 1
  Set pointfeature = pointfeatureclass.GetFeature (pointcount)
  'Determine if the point is in the buffer or not
  iib = pointinbuffer.Contains(pointfeature.Shape)
  overlapstart = False
  overlapend = False

  If iib Then
    'Point is in the buffer, continue snapping
    Set mypoint = pointfeature.Shape
    Set outpoint = New Point
    myline.QueryPointAndDistance esriNoExtension, mypoint, False, outpoint, distalongcurve,
      distfromcurve, rightside

    'Determine if the resulting point is the start point of the line feature
    If outpoint.X = mystartpoint.X And outpoint.Y = mystartpoint.Y Then
      overlapstart = True
    Else
      overlapstart = False
    End If

    'Determine if the resulting point is the end point of the line feature
    If outpoint.X = myendpoint.X And outpoint.Y = myendpoint.Y Then
      overlapend = True
    Else
      overlapend = False
    End If

    If overlapstart Or overlapend Then
      'If the resulting point is the start or the end point, do nothing and go to next LIDAR point
      Else
        'The resulting point is not the start point, nor the end point. Create a new row in the
        'target table
        Set myrow = mytable.CreateRow
        myrow.Value (1) = outpoint.X
        myrow.Value (2) = outpoint.Y
        myrow.Value (3) = pointfeature.Value (4)
        myrow.Value (5) = distalongcurve
        myrow.Value (4) = myline.Length

```

```
myrow.Value (6) = linefeature.Value (9)
myrow.Value (7) = linefeature.Value (2)
myrow.Value (8) = linefeature.Value (4)
myrow.Value (9) = linefeature.Value (5)
myrow.Value (10) = linefeature.Value (7)
myrow.Value (11) = linefeature.Value (8)
myrow.Store
End If
Else
'Point is not in the buffer, do nothing and go to the next LIDAR point
End If
Next
Next
End Sub
```

APPENDIX C PROGRAM CODE FOR WORKING WITH DEMS TO OBTAIN 3-D POINTS ALONG LINES

'Program Name: PointsFromRaster

'This program obtains elevations for points planimetrically uniformly distributed along the lines from raster data 'via a bilinear interpolation.

'This program was developed by Hubo Cai during the summer of 2003.

Option Explicit

'The main function creates a workspace and opens all files

Sub PointsFromRaster ()

Dim pworkspacefactory As IWorkspaceFactory

Set pworkspacefactory = New ShapefileWorkspaceFactory

Dim pfeatureworkspace As IFeatureWorkspace

Set pfeatureworkspace = pworkspacefactory.OpenFromFile ("s:\palrs\users\hubo\Nroad_english\VANCE", 0)

Dim linefeatureclass As IFeatureClass

Set linefeatureclass = pfeatureworkspace.OpenFeatureClass ("Dis_int_vance")

Dim pointTable As ITable

Set pointTable = pfeatureworkspace.OpenTable ("raster_25ft")

'Call the OpenRasterDataset function to open the raster file

Dim praster As IRasterDataset

Dim ppraster As IRaster

Set ppraster = OpenRasterDataset ("s:\palrs\users\Hubo\LIDAR_50ft\grid", "final").CreateDefaultRaster

ppraster.ResampleMethod = RSP_BilinearInterpolation

'Call the ExtractValueTOPoint function to obtain elevations for points along the linear features

Call ExtractValueTOPoint (ppraster, linefeatureclass, pointTable, "elev_1")

End Sub

'This public function opens the raster file

Public Function OpenRasterDataset (sPath As String, sFileName As String) As IRasterDataset

'sPath: directory where dataset resides

'sFileName: name of the raster dataset

On Error GoTo ErrorHandler

'Create RasterWorkSpaceFactory

Dim pWSF As IWorkspaceFactory

Set pWSF = New RasterWorkspaceFactory

'Get RasterWorkspace

Dim pRasWS As IRasterWorkspace

If pWSF.IsWorkspace (sPath) Then

Set pRasWS = pWSF.OpenFromFile (sPath, 0)

Set OpenRasterDataset = pRasWS.OpenRasterDataset (sFileName)

End If

'Release memory

Set pRasWS = Nothing

Set pWSF = Nothing

Exit Function

ErrorHandler:

Set OpenRasterDataset = Nothing

End Function

'This public function extracts elevations from raster file to points using bilinear interpolation

```

Sub ExtractValueTOPoint (pInRaster As IRaster, pInFeatureClass As IFeatureClass, pIntable As ITable,
                        sFieldName As String)
    'pInRaster: input raster
    'pInFeatureClass: input line feature class
    'sFieldName: name of the field that stores the values

    'Define field name
    Dim pFld As IFieldEdit
    Set pFld = New Field
    pFld.Name = sFieldName

    'Define field type based on the value type of the raster file
    Dim pProp As IRasterProps
    Set pProp = pInRaster
    If pProp.PixelType = PT_FLOAT Or pProp.PixelType = PT_DOUBLE Or pProp.PixelType Then
        pFld.Type = esriFieldTypeDouble
        pFld.Length = 24
        pFld.Required = 8
    ElseIf pProp.PixelType = PT_CHAR Or pProp.PixelType = PT_UCHAR Then
        MsgBox "Data type not compatible for calculation, exit procedure"
        Exit Sub
    Else 'for integer case
        pFld.Type = esriFieldTypeInteger
        pFld.Length = 24
        pFld.Required = 0
    End If

    'Add field
    pIntable.AddField pFld

    'Get field index
    Dim FieldIndex As Integer
    FieldIndex = pIntable.FindField (sFieldName)
    If FieldIndex < 0 Then Exit Sub

    'Create a raster layer and QI for IIdentify interface
    Dim pRLayer As IRasterLayer
    Set pRLayer = New RasterLayer
    pInRaster.ResampleMethod = RSP_BilinearInterpolation
    pRLayer.CreateFromRaster pInRaster
    Dim pIdentify As IIdentify
    Set pIdentify = pRLayer

    'Based on the property of the raster file, determine the cell size and the top-left point
    Dim pProperty As IRasterProps
    Set pProperty = pInRaster
    Dim cellheight As Double
    cellheight = pProperty.Height
    cellheight = (pProperty.Extent.YMAX - pProperty.Extent.YMin) / cellheight
    Dim cellwidth As Double
    cellwidth = pProperty.Width
    cellwidth = (pProperty.Extent.XMax - pProperty.Extent.XMin) / cellwidth
    Dim XMIN As Double
    XMIN = pProperty.Extent.XMIN
    Dim YMAX As Double
    YMAX = pProperty.Extent.YMAX

```


'I1 and I2 will be used to locate the surrounding cells for the target point
'and determine the centerpoints for these surrounding cells

```
Dim I1 As Integer
Dim I2 As Integer
Dim D1 As Double
Dim D2 As Double
```

'pIDArray and pRIDObj are used for the target point while the others are used for the surrounding cell centers.

```
Dim pIDArray As IArray
Dim pRIDObj As IRasterIdentifyObj
Dim pIDArray1 As IArray
Dim pRIDObj1 As IRasterIdentifyObj
Dim pIDArray2 As IArray
Dim pRIDObj2 As IRasterIdentifyObj
Dim pIDArray3 As IArray
Dim pRIDObj3 As IRasterIdentifyObj
Dim pIDArray4 As IArray
Dim pRIDObj4 As IRasterIdentifyObj
```

```
Dim i As Long
Dim j As Long
Dim pPoint As IPoint
Dim pFeature As IFeature
```

'pNewPoint is the target point, pNewPoint1, 2, 3, 4 are the center points of the surrounding cells.

```
Dim pNewPoint As IPoint
Set pNewPoint = New Point
Dim pNewPoint1 As IPoint
Set pNewPoint1 = New Point
Dim pNewPoint2 As IPoint
Set pNewPoint2 = New Point
Dim pNewPoint3 As IPoint
Set pNewPoint3 = New Point
Dim pNewPoint4 As IPoint
Set pNewPoint4 = New Point
```

'Loop through each line in the feature class and obtain points
'including start point and end point and points with cellsize interval

```
Dim NumOfRow As Integer
NumOfRow = pInFeatureClass.FeatureCount (Nothing)
Dim pLine As IPolyline
Set pLine = New Polyline
Dim length2d As Double
```

```
For i = 0 To NumOfRow - 1
```

```
    'Get line
    Set pFeature = pInFeatureClass.GetFeature (i)
    Set pLine = pFeature.Shape
    length2d = pLine.Length
    'jj and jjj will be used to determine how many points will be obtained
    Dim jj As Double
    Dim jjj As Integer
    Dim interval As Integer
    interval = cellwidth / 2
```

```

jj = length2d / interval
jjj = length2d / interval

If (jj - jjj) > 0 Then
    jjj = jjj + 1
Else
End If

For j = 0 To jjj
    Dim myrow As IRow
    Set myrow = pIntable.CreateRow
    myrow.value (5) = pFeature.value (2)
    myrow.value (1) = pLine.Length

    If j = 0 Then
        'Get the start point
        Set pPoint = pLine.FromPoint
        myrow.value (2) = 0
        myrow.value (3) = pLine.FromPoint.X
        myrow.value (4) = pLine.FromPoint.Y
    ElseIf j >= jjj Then
        'Get the end point
        Set pPoint = pLine.ToPoint
        myrow.value (2) = pLine.Length
        myrow.value (3) = pLine.ToPoint.X
        myrow.value (4) = pLine.ToPoint.Y
    Else
        'Get intermediate points based on the given interval
        pLine.QueryPoint esriNoExtension, j * interval, False, pPoint
        myrow.value (2) = j * interval
        myrow.value (3) = pPoint.X
        myrow.value (4) = pPoint.Y
    End If

    pNewPoint.X = pPoint.X
    pNewPoint.Y = pPoint.Y

    'Get RasterIdentifyObject on that point
    Set pIDArray = pIdentify.Identify (pNewPoint)
    I1 = (pNewPoint.X - XMIN) / cellwidth
    I2 = (YMAX - pNewPoint.Y) / cellheight

    'Identify the center points of the surrounding cells
    pNewPoint1.X = XMIN + (I1 - 0.5) * cellwidth
    pNewPoint1.Y = YMAX - (I2 - 0.5) * cellheight
    pNewPoint2.X = XMIN + (I1 + 0.5) * cellwidth
    pNewPoint2.Y = YMAX - (I2 - 0.5) * cellheight
    pNewPoint3.X = XMIN + (I1 - 0.5) * cellwidth
    pNewPoint3.Y = YMAX - (I2 + 0.5) * cellheight
    pNewPoint4.X = XMIN + (I1 + 0.5) * cellwidth
    pNewPoint4.Y = YMAX - (I2 + 0.5) * cellheight

    'Get RasterIdentifyObject on these four points
    Set pIDArray1 = pIdentify.Identify (pNewPoint1)
    Set pIDArray2 = pIdentify.Identify (pNewPoint2)
    Set pIDArray3 = pIdentify.Identify (pNewPoint3)

```

```
Set pIDArray4 = pIdentify.Identify (pNewPoint4)
```

```
'Eletotal and Elecount will be used to deal with no surrounding cell or surrounding cell with no data cases
```

```
Dim Eletotal As Double
```

```
Eletotal = 0
```

```
Dim Elecount As Integer
```

```
Elecount = 0
```

```
'In case one or more surrounding cells not exist, an average of the existing ones is taken
```

```
If (pIDArray1 Is Nothing) Or (pIDArray2 Is Nothing) Or (pIDArray3 Is Nothing) Or (pIDArray4 Is  
Nothing) Then
```

```
    If Not pIDArray1 Is Nothing Then
```

```
        Set pRIDObj1 = pIDArray1.Element (0)
```

```
        If pRIDObj1.Name <> "NoData" Then
```

```
            Elecount = Elecount + 1
```

```
            Eletotal = Eletotal + CDb1 (pRIDObj1.Name)
```

```
        End If
```

```
    End If
```

```
    If Not pIDArray2 Is Nothing Then
```

```
        Set pRIDObj2 = pIDArray2.Element (0)
```

```
        If pRIDObj2.Name <> "NoData" Then
```

```
            Elecount = Elecount + 1
```

```
            Eletotal = Eletotal + CDb1 (pRIDObj2.Name)
```

```
        End If
```

```
    End If
```

```
    If Not pIDArray3 Is Nothing Then
```

```
        Set pRIDObj3 = pIDArray3.Element (0)
```

```
        If pRIDObj3.Name <> "NoData" Then
```

```
            Elecount = Elecount + 1
```

```
            Eletotal = Eletotal + CDb1 (pRIDObj3.Name)
```

```
        End If
```

```
    End If
```

```
    If Not pIDArray4 Is Nothing Then
```

```
        Set pRIDObj4 = pIDArray4.Element (0)
```

```
        If pRIDObj4.Name <> "NoData" Then
```

```
            Elecount = Elecount + 1
```

```
            Eletotal = Eletotal + CDb1(pRIDObj4.Name)
```

```
        End If
```

```
    End If
```

```
    If Elecount > 0 Then
```

```
        myrow.value (FieldIndex) = Eletotal / Elecount
```

```
        myrow.value (6) = Eletotal / Elecount
```

```
        myrow.Store
```

```
    End If
```

```
Else
```

```
    Set pRIDObj1 = pIDArray1.Element (0)
```

```
    Set pRIDObj2 = pIDArray2.Element (0)
```

```
    Set pRIDObj3 = pIDArray3.Element (0)
```

```
    Set pRIDObj4 = pIDArray4.Element (0)
```

```
'In case one or more surrounding cells have no data, an average is taken
```

```

If pRIDObj1.Name = "NoData" Or pRIDObj2.Name = "NoData" Or pRIDObj3.Name = "NoData" Or
    pRIDObj4.Name = "NoData" Then
    If pRIDObj1.Name <> "NoData" Then
        Elecount = Elecount + 1
        Eletotal = Eletotal + CDb1 (pRIDObj1.Name)
    End If

    If pRIDObj2.Name <> "NoData" Then
        Elecount = Elecount + 1
        Eletotal = Eletotal + CDb1 (pRIDObj2.Name)
    End If

    If pRIDObj3.Name <> "NoData" Then
        Elecount = Elecount + 1
        Eletotal = Eletotal + CDb1 (pRIDObj3.Name)
    End If
    If pRIDObj4.Name <> "NoData" Then
        Elecount = Elecount + 1
        Eletotal = Eletotal + CDb1 (pRIDObj4.Name)
    End If
    If Elecount > 0 Then
        myrow.value (FieldIndex) = Eletotal / Elecount
        myrow.value (6) = Eletotal / Elecount
        myrow.Store
    End If
Else
    'All 4 surrounding cells exist and have values, elevation for the target point is obtained via bilinear
    'interpolation
    'E1 to E4 will be used to store the elevations for the surrounding cells
    Dim E1 As Double
    Dim E2 As Double
    Dim E3 As Double
    Dim E4 As Double
    'EE1 and EE2 will be used to store the elevations for the intermediate points
    Dim EE1 As Double
    Dim EE2 As Double
    'EE will be used to store the elevation for the target point after bilinear interpolation
    Dim EE As Double
    E1 = CDb1 (pRIDObj1.Name)
    E2 = CDb1 (pRIDObj2.Name)
    E3 = CDb1 (pRIDObj3.Name)
    E4 = CDb1 (pRIDObj4.Name)
    EE1 = E1 + (pNewPoint.X - pNewPoint1.X) * (E2 - E1) / cellwidth
    EE2 = E3 + (pNewPoint.X - pNewPoint3.X) * (E4 - E3) / cellwidth
    EE = EE2 + (EE1 - EE2) * (pNewPoint.Y - pNewPoint4.Y) / cellheight
    myrow.value (FieldIndex) = EE
    myrow.value (6) = EE
    myrow.Store
End If
End If
myrow.Store
Next j
Next i
End Sub

```

APPENDIX D PROGRAM CODE FOR CALCULATIONS OF AVERAGE, WEIGHTED AVERAGE, AVERAGE SLOPE CHANGE, AND WEIGHTED AVERAGE SLOPE CHANGE

'Program Name: Slope

'Computes the average slope, weighted average slope, average slope change, and weighted average slope change from the 3-D points representing FTSEgs in a 3-D space.

'This program was developed by Hubo Cai during th summer of 2003.

Option Explicit

Sub slope ()

'Create a workspace to open the attribute table of the 3-D road centerline data and the target table to store the calculations

Dim pworkspacefactory As IWorkspaceFactory

Set pworkspacefactory = New ShapefileWorkspaceFactory

Dim pfeatureworkspace As IFeatureWorkspace

Set pfeatureworkspace = pworkspacefactory.OpenFromFile ("c:\dotdata\studyscope", 0)

'Open the 3-D road centerline data (3-D point data)

Dim mytable As ITable

Set mytable = pfeatureworkspace.OpenTable ("snap_johnston_nc_4ft")

MsgBox "how many rows: " & mytable.RowCount (Nothing)

'Open the target table to store the calculations

Dim mytargettable As ITable

Set mytargettable = pfeatureworkspace.OpenTable ("slope_johnston_nc_average")

'Define all variables

Dim sumslopedifference As Double

Dim sumslope As Double

Dim absoluteslope As Double

Dim currentslopedifference As Double

Dim slope As Double

Dim currentslope As Double

Dim current2d As Double

Dim currentelevation As Double

Dim MERGE As Integer

Dim CONDITION As String

Dim pointcount As Long

Dim segcount As Long

Dim sdifference As Double

Dim DIST As Double

Dim SUM As Double

Dim slope1 As Double

Dim slope2 As Double

Dim SUM_Slope_W As Double

Dim SUM_Slope_CH_W As Double

'Loop through the table based on the MERGE value (the ID of the FTSEgs)

For MERGE = 0 To 100

 'Initialize the variables

 sumslopedifference = 0

 sumslope = 0

 currentslope = 0

 current2d = 0

 currentelevation = 0

```

pointcount = 0
DIST = 0
SUM = 0
slope1 = 0
slope2 = 0
SUM_Slope_W = 0
SUM_Slope_CH_W = 0

'Define the tablesort where clause
CONDITION = "MERGE = " & MERGE
'Create a tablesort to retrieve all 3-D points belonging to the current FTSeg
Dim pTableSort As ITableSort
Set pTableSort = New esriCore.TableSort
Dim pQueryFilter As IQueryFilter
Set pQueryFilter = New QueryFilter
pQueryFilter.WhereClause = CONDITION

With pTableSort
    .Fields = "D_TO_S"
    .Ascending ("D_TO_S") = True
    Set .QueryFilter = pQueryFilter
    Set .Table = mytable
End With

pTableSort.Sort Nothing

Dim pCursor As ICursor
Set pCursor = pTableSort.Rows
Dim pRow As IRow
Dim tRow As IRow
Set pRow = pCursor.NextRow

'Start from the start point, go through neighboring points and do calculations
If Not pRow Is Nothing Then
    currentelevation = pRow.Value (3)
    current2d = pRow.Value (5)
    pointcount = pointcount + 1
    Set tRow = mytargettable.CreateRow
    tRow.Value (1) = pRow.Value (7)
    Set pRow = pCursor.NextRow

    Do While Not pRow Is Nothing
        'Variable sdifference is the current 2D distance between two neighboring 3D points
        sdifference = pRow.Value (5) - current2d
        If sdifference > 0 Then
            'DIST is the 3-D distance between two neighboring points
            DIST = Sqr((pRow.Value (3) - currentelevation) * (pRow.Value (3) - currentelevation) + (pRow.Value
                (5) - current2d) * (pRow.Value (5) - current2d))
            'SUM is the sum of the 3D distances
            SUM = SUM + DIST
            'Variable slope is the current slope between two neighboring points
            slope = (pRow.Value (3) - currentelevation) / (pRow.Value (5) - current2d)
            'Variable currentslopedifference is the slope change from the current slope to the slope of the previous
            one
            currentslopedifference = Abs (slope - currentslope)
            absoluteslope = Abs (slope)
        End If
    End Do
End If

```

```

'Variable sumslopedifference holds the sum of the slope changes
sumslopedifference = sumslopedifference + currentslopedifference
'Variable sumslope holds the sum of the absolute slopes
sumslope = sumslope + absoluteslope
'SUM_Slope_W and SUM_Slope_CH_W hold the sum of the weighted slopes and the sum of the
  weighted slope changes, respectively
SUM_Slope_W = SUM_Slope_W + absoluteslope * DIST
SUM_Slope_CH_W = SUM_Slope_CH_W + currentslopedifference * DIST
'MsgBox "2D Difference is: " & (pRow.Value (2) - current2d) & "3D Distance is: " & sum
  pointcount = pointcount + 1
  currentslope = slope
Else
  'Two points are too close and do nothing
  MsgBox "Points too close!"
End If

current2d = pRow.Value (5)
currentelevation = pRow.Value (3)
Set pRow = pCursor.NextRow
Loop

'The total number of segments is calculated as the number of 3-D points subtracted by 1
segcount = pointcount - 1
'Calculate all the average values and store them to the target table
tRow.Value (2) = sumslopedifference
tRow.Value (3) = segcount
tRow.Value (4) = sumslopedifference / segcount
tRow.Value (5) = sumslope
tRow.Value (6) = sumslope / segcount
tRow.Value (7) = SUM_Slope_W / SUM
tRow.Value (8) = SUM_Slope_CH_W / SUM
tRow.Value (9) = SUM
tRow.Value (10) = SUM_Slope_W
tRow.Value (11) = SUM_Slope_CH_W
tRow.Store
Else
  'No FTSeg with the current MERGE value exists. Do nothing and go to the next MERGE value
End If
Next
End Sub

```

APPENDIX E PROGRAM CODE FOR CHECKING THE ERROR TOLERANCE AND THE MAXIMUM DROP AND MAKING ADJUSTMENTS

'Program Name: Clean_methods

'For all the points along river polylines, check the error tolerance and the maximum drop and make 'corresponding adjustments.

'This program was developed by Hubo Cai during the summer of 2003.

Option Explicit

Sub Clean_methods()

'Open a workspace and open the point data

Dim pworkspacefactory As IWorkspaceFactory

Set pworkspacefactory = New ShapefileWorkspaceFactory

Dim pfeatureworkspace As IFeatureWorkspace

Set pfeatureworkspace = pworkspacefactory.OpenFromFile ("S:\PALRS\users\Hubo\Hydro", 0)

Dim mytable As ITable

Set mytable = pfeatureworkspace.OpenTable ("PointAlongRiverClean_method1")

'Define variables. Variable totalchanged is used to keep track of the number of points being changed

Dim RiverID As Integer

Dim elev1 As Double

Dim elev2 As Double

Dim CONDITION As String

Dim totalchanged As Long

totalchanged = 0

'Loop through each river

For RiverID = 0 To 67

 CONDITION = "RiverID = " & RiverID

 'elev1 and elev2 are used to store the average elevations of the first 3 points and the last 3 points, respectively

 elev1 = 0

 elev2 = 0

 'Define a tablesort

 Dim pTableSort As ITableSort

 Set pTableSort = New esriCore.TableSort

 Dim pQueryFilter As IQueryFilter

 Set pQueryFilter = New QueryFilter

 pQueryFilter.whereClause = CONDITION

 'Sort 1 to get the first three points

 With pTableSort

 .Fields = "D_TO_S"

 .Ascending ("D_TO_S") = True

 Set .QueryFilter = pQueryFilter

 Set .Table = mytable

 End With

 pTableSort.Sort Nothing

 Dim pCursor As ICursor

 Set pCursor = pTableSort.Rows

 Dim pRow As IRow

 Dim trow As IRow

 Set pRow = pCursor.NextRow


```

Dim count As Integer
count = 0
Dim i As Integer

'get the average elevation of the first three points
If Not pRow Is Nothing Then
    Do While (Not pRow Is Nothing) And (count < 3)
        elev1 = elev1 + pRow.value (1)
        count = count + 1
        Set pRow = pCursor.NextRow
    Loop
    elev1 = elev1 / count
Else
    elev1 = 9999
End If

'Sort 2 to get the last three points
With pTableSort
    .Fields = "D_TO_S"
    .Ascending ("D_TO_S") = False
    Set .QueryFilter = pQueryFilter
    Set .Table = mytable
End With

pTableSort.Sort Nothing
Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow
count = 0

'Get the average elevation of the last three points
If Not pRow Is Nothing Then
    Do While (Not pRow Is Nothing) And (count < 3)
        elev2 = elev2 + pRow.value (1)
        count = count + 1
        Set pRow = pCursor.NextRow
    Loop
    elev2 = elev2 / count
Else
    elev2 = 9999
End If

Dim down As Boolean
Dim upper As Boolean
Dim flat As Boolean

'Determine if the water is running from start to the end or from the end to the start
If elev1 = 9999 Or elev2 = 999 Then
    down = False
    upper = False
    MsgBox "some points are missing"
    GoTo eex
Else
    If Abs(elev1 - elev2) <= 10 Then 'difference is less than 1ft
        down = False
        upper = False
        flat = True
    End If
End If

```

```

ElseIf elev1 >= elev2 Then
    down = True
    upper = False
Else
    down = False
    upper = True
End If
End If

'Start cleaning
'Based on the flowing direction, sort the points based on their distance to the start point so that the order is
always from higher elevations to lower elevations
If upper Then
    With pTableSort
        .Fields = "D_TO_S"
        .Ascending ("D_TO_S") = False
        Set .QueryFilter = pQueryFilter
        Set .Table = mytable
    End With
Else
    With pTableSort
        .Fields = "D_TO_S"
        .Ascending ("D_TO_S") = True
        Set .QueryFilter = pQueryFilter
        Set .Table = mytable
    End With
End If

pTableSort.Sort Nothing
Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow
Dim targetrow As IRow
Dim maxdrop As Double
Dim tolerance As Double
'Define the maximum drop and the error tolerance
maxdrop = 0.1 '10%
tolerance = 0.02 '1ft/50ft

If Not pRow Is Nothing Then
    Set targetrow = pRow
    Set pRow = pCursor.NextRow
    Do While Not pRow Is Nothing
        'check with error tolerance
        If (pRow.value (1) - targetrow.value (1)) >= 0 Then
            If Abs((pRow.value (1) - targetrow.value (1)) / ((pRow.value (2) - targetrow.value (2)) * 10)) <=
                tolerance Then
                'Within error tolerance, ok, do nothing
            Else
                'Error tolerance is exceeded. Assign the elevation of the succeeding point to the current point
                pRow.value (1) = targetrow.value (1)
                pRow.Store
                totalchanged = totalchanged + 1
            End If
            Set targetrow = pRow
            Set pRow = pCursor.NextRow
        Else

```

```

'check with maximum drop
If Abs((pRow.value (1) - targetrow.value (1)) / ((pRow.value (2) - targetrow.value (2)) * 10)) <=
    maxdrop Then
'Within the maximum drop range, ok, do nothing
Else
    'Algorithm 1, adjust the elevation of the current point
    pRow.value (1) = targetrow.value (1) - maxdrop * Abs(pRow.value (2) - targetrow.value (2)) * 10
    pRow.Store
    totalchanged = totalchanged + 1

    'Algorithm 2, adjust the elevation of the succeeding point
    '(targetrow.value (1) = pRow.value (1) + maxdrop * Abs(pRow.value (2) - targetrow.value (2)) *
        10}
    '{targetrow.Store}
    '{totalchanged = totalchanged + 1}
End If

    Set targetrow = pRow
    Set pRow = pCursor.NextRow
End If
Loop
Else
End If

eex: MsgBox "Some points are missing, pay attention to river with ID: " & RiverID
Next
End Sub

```

APPENDIX F PROGRAM CODE FOR FLOOD EXTENT PREDICTION

'Program Name: FloodExtentPrediction

'This program is developed to predict flood extent with given LIDAR 20ft DEM and point data along 'waterbodies '(polylines)

'This program was developed by Hubo Cai during the summer of 2003

Option Explicit

Sub FloodExtentPrediction ()

'Define raster data set and open raster data by calling the OpenRasterDataset Sub

Dim praster As IRasterDataset

Dim ppraster As IRaster

Set ppraster = OpenRasterDataset ("s:\palrs\users\Hubo\Hydro", "elev_wilson").CreateDefaultRaster

ppraster.ResampleMethod = RSP_BilinearInterpolation

'Define workspace and open the waterbody data and the resulting polygon data

Dim pworkspacefactory As IWorkspaceFactory

Set pworkspacefactory = New ShapefileWorkspaceFactory

Dim pfeatureworkspace As IFeatureWorkspace

Set pfeatureworkspace = pworkspacefactory.OpenFromFile ("S:\PALRS\users\Hubo\Hydro", 0)

Dim polygonfeatureclass As IFeatureClass

Set polygonfeatureclass = pfeatureworkspace.OpenFeatureClass ("Testing_results_clean_method2")

Dim newfeature As IFeature

Dim mypointcollection As IPointCollection

Dim mypolygon As IPolygon

'a Linefeature represents a waterbody polyline

'linetotal and linecount are used to control the loop

Dim linefeatureclass As IFeatureClass

Set linefeatureclass = pfeatureworkspace.OpenFeatureClass ("Testing_riv_3")

Dim linetotal As Integer

linetotal = linefeatureclass.FeatureCount (Nothing)

Dim linecount As Integer

Dim myfeature As IFeature

Dim mypointtable As ITable

Set mypointtable = pfeatureworkspace.OpenTable ("PointAlongRiverClean_method2")

Dim flood1 As Double

Dim flood2 As Double

Dim reachingboundary As Long

reachingboundary = 0

For linecount = 0 To linetotal - 1

Set myfeature = linefeatureclass.GetFeature (linecount)

Dim myline As IPolyline

Set myline = myfeature.Shape

Dim elevadd As Double

'Assume the given flood is 2 ft

'Our elevation data were multiplied by 10 and therefore, 2 ft equals 20 now

elevadd = 20

Dim count As Integer

'Point1, point2, point3, and point4 will be used to represent the vertices

'of the polygon to be constructed.

'midpoint represents a point along the waterbody polyline, through which
'the normal line will be built.

```
Dim point1 As IPoint  
Dim point2 As IPoint  
Dim point3 As IPoint  
Dim point4 As IPoint  
Set point1 = New Point  
Set point2 = New Point  
Set point3 = New Point  
Set point4 = New Point  
Dim midpoint As IPoint
```

```
Dim CONDITION As String  
CONDITION = "RiverID = " & myfeature.value(0)  
Dim pTableSort As ITableSort  
Set pTableSort = New esriCore.TableSort  
Dim pQueryFilter As IQueryFilter  
Set pQueryFilter = New QueryFilter  
pQueryFilter.whereClause = CONDITION
```

```
With pTableSort  
    .Fields = "D_TO_S"  
    .Ascending ("D_TO_S") = True  
    Set .QueryFilter = pQueryFilter  
    Set .Table = mypointtable  
End With
```

```
pTableSort.Sort Nothing  
Dim pCursor As ICursor  
Set pCursor = pTableSort.Rows  
Dim pRow As IRow  
Set pRow = pCursor.NextRow  
Dim delete1 As Boolean  
Dim delete2 As Boolean
```

```
If Not pRow Is Nothing Then  
    'Built the normal line through the start point of a waterbody polyline feature  
    Dim pnormal1 As ILine  
    Set pnormal1 = New Line  
    myline.QueryNormal esriNoExtension, pRow.value (2), False, 100, pnormal1  
    Set point1 = pnormal1.ToPoint  
    Dim pnormal2 As ILine  
    Set pnormal2 = New Line  
    Set midpoint = New Point  
    Set midpoint = myline.FromPoint
```

```
'The normal line starts from the midpoint to one direction.  
'Since we need the normal line in both directions, point2 is  
'temporarily used to extent the line to the other direction.  
point2.X = midpoint.X * 2 - point1.X  
point2.Y = midpoint.Y * 2 - point1.Y  
pnormal2.FromPoint = midpoint  
pnormal2.ToPoint = point2
```

```
'Call the mypoint sub to determine point1 and point2 on the normal lines  
'The new point1 and point2 having their elevations equaling the target elevation.
```

```
Call mypoint(delete1, pRow.value (1), point1, ppraster, pnormal1, midpoint, elevadd, flood1)
Call mypoint(delete2, pRow.value (1), point2, ppraster, pnormal2, midpoint, elevadd, flood1)
```

```
If delete1 Then
    reachingboundary = reachingboundary + 1
Else
End If
```

```
If delete2 Then
    reachingboundary = reachingboundary + 1
Else
End If
```

```
Dim pnormal3 As ILine
Dim pnormal4 As ILine
Set pnormal3 = New Line
Set pnormal4 = New Line
```

```
'Continue to construct normal line and to determine point3 and point4
'at the next point along the water body polyline
```

```
Set pRow = pCursor.NextRow
```

```
Do While Not pRow Is Nothing
```

```
    myline.QueryNormal esriNoExtension, pRow.value (2), False, 100, pnormal3
```

```
    myline.QueryPoint esriNoExtension, pRow.value (2), False, midpoint
```

```
    Set point4 = pnormal3.ToPoint
```

```
    Set pnormal4 = New Line
```

```
    Set point3 = New Point
```

```
    point3.X = midpoint.X * 2 - point4.X
```

```
    point3.Y = midpoint.Y * 2 - point4.Y
```

```
    pnormal4.FromPoint = midpoint
```

```
    pnormal4.ToPoint = point3
```

```
'Call the mypoint sub to determine point3 and point4 on the normal lines
```

```
'The new point1 and point2 having their elevations equaling the target elevation.
```

```
Call mypoint(delete1, pRow.value (1), point3, ppraster, pnormal4, midpoint, elevadd, flood2)
```

```
Call mypoint(delete2, pRow.value (1), point4, ppraster, pnormal3, midpoint, elevadd, flood2)
```

```
If delete1 Then
    reachingboundary = reachingboundary + 1
Else
End If
```

```
If delete2 Then
    reachingboundary = reachingboundary + 1
Else
End If
```

```
'After 4 vertices are determined, a new polygon is constructed
```

```
Set mypolygon = New Polygon
```

```
Set mypointcollection = mypolygon
```

```
If mypointcollection.pointcount > 0 Then
```

```
    mypointcollection.RemovePoints 0, mypointcollection.pointcount
```

```
Else
End If
```

```

If delete1 Or delete2 Then
  If delete1 Then
    mypointcollection.AddPoints 1, point1
    mypointcollection.AddPoints 1, point2
    mypointcollection.AddPoints 1, point4
    mypolygon.Close

    Set newfeature = polygonfeatureclass.CreateFeature
    Set newfeature.Shape = mypolygon
    newfeature.Store
    newfeature.value (3) = myfeature.value (0)
    newfeature.value (4) = (flood1 * 2 + flood2) / 3
    newfeature.Store

    Set point1 = New Point
    'Set point2 = New point
    Set point1 = point4
    'Set point2 = point3
    flood1 = flood2
    Set pRow = pCursor.NextRow
  Else
    mypointcollection.AddPoints 1, point1
    mypointcollection.AddPoints 1, point2
    mypointcollection.AddPoints 1, point3
    mypolygon.Close
    Set newfeature = polygonfeatureclass.CreateFeature
    Set newfeature.Shape = mypolygon
    newfeature.Store
    newfeature.value (3) = myfeature.value(0)
    newfeature.value (4) = (flood1 * 2 + flood2) / 3
    newfeature.Store
    'Set point1 = New point
    Set point2 = New Point
    'Set point1 = point4
    Set point2 = point3
    flood1 = flood2
    Set pRow = pCursor.NextRow
  End If
Else
  mypointcollection.AddPoints 1, point1
  mypointcollection.AddPoints 1, point2
  mypointcollection.AddPoints 1, point3
  mypointcollection.AddPoints 1, point4
  mypolygon.Close

  Set newfeature = polygonfeatureclass.CreateFeature
  Set newfeature.Shape = mypolygon
  newfeature.Store
  newfeature.value (3) = myfeature.value (0)
  newfeature.value (4) = (flood1 + flood2) / 2
  newfeature.Store
  'After the new polygon is constructed and stored, point3 and point4
  'are assigned to point2 and point1, respectively.
  'The procedure continues to identify new point3 and point4 at the next
  'point along the water body polyline and consequently, new polygon will
  'be constructed.

```

```

        'Doing the inner loop will construct polygons surrounding the water body.
        Set point1 = New Point
        Set point2 = New Point
        Set point1 = point4
        Set point2 = point3
        flood1 = flood2
        Set pRow = pCursor.NextRow
    End If
Loop
Else
End If
Next
MsgBox "total reaching boundary: " & reachingboundary
End Sub

Public Function OpenRasterDataset (sPath As String, sFileName As String) As IRasterDataset
    ' sPath: directory where dataset resides
    ' sFileName: name of the raster dataset
    On Error GoTo ErrorHandler
    ' Create RasterWorkspaceFactory
    Dim pWSF As IWorkspaceFactory
    Set pWSF = New RasterWorkspaceFactory
    ' Get RasterWorkspace
    Dim pRasWS As IRasterWorkspace
    If pWSF.IsWorkspace (sPath) Then
        Set pRasWS = pWSF.OpenFromFile (sPath, 0)
        Set OpenRasterDataset = pRasWS.OpenRasterDataset (sFileName)
    End If
    ' Release memory
    Set pRasWS = Nothing
    Set pWSF = Nothing
    Exit Function
ErrorHandler:
    Set OpenRasterDataset = Nothing
End Function

'This sub aims at determining two points on the normal line, which have their elevations
'equalling to the target elevation.
Public Sub mypoint(delete As Boolean, base_elev As Double, outpoint As IPoint, pInRaster As IRaster,
    pInLine As ILine, ppInPoint As IPoint, elevadd As Double, flood As Double)
    ' pInRaster: input raster
    ' pInFeatureClass: input line feature class
    ' sFieldName: name of the field that stores the values

    delete = False
    ' Create a raster layer and QI for IIdentify interface
    Dim pRLayer As IRasterLayer
    Set pRLayer = New RasterLayer
    pInRaster.ResampleMethod = RSP_BilinearInterpolation
    pRLayer.CreateFromRaster pInRaster
    Dim pIdentify As IIdentify
    Set pIdentify = pRLayer
    Dim pProperty As IRasterProps
    Set pProperty = pInRaster
    Dim cellheight As Double
    cellheight = pProperty.Height

```



```

cellheight = (pProperty.Extent.YMAX - pProperty.Extent.YMin) / cellheight
'Determine cell width
Dim cellwidth As Double
cellwidth = pProperty.Width
cellwidth = (pProperty.Extent.XMax - pProperty.Extent.XMIN) / cellwidth
'Determine the upper-left corner of the raster data
Dim XMIN As Double
XMIN = pProperty.Extent.XMIN
Dim YMAX As Double
YMAX = pProperty.Extent.YMAX

```

```

'I1 and I2 will be used to locate the surrounding cells for the target point
'and determine the centerpoints for these surrounding cells
Dim I1 As Integer
Dim I2 As Integer
Dim D1 As Double
Dim D2 As Double

```

'pIDArray and pRIDObj are used for the target point while the others are used for the surrounding cell centers.

```

Dim pIDArray As IArray
Dim pRIDObj As IRasterIdentifyObj
Dim pIDArray1 As IArray
Dim pRIDObj1 As IRasterIdentifyObj
Dim pIDArray2 As IArray
Dim pRIDObj2 As IRasterIdentifyObj

```

```

Dim pNewPoint1 As IPoint
Set pNewPoint1 = New Point
Dim pNewPoint2 As IPoint
Set pNewPoint2 = New Point

```

```

'Get RasterIdentifyObject on the point on the river line
Dim pInPoint As IPoint
Set pInPoint = New Point
Set pInPoint = ppInPoint
'MsgBox "Do we have pInPoint?" & pInPoint.X & ", " & pInPoint.Y
Set pIDArray = pIdentify.Identify(pInPoint)
If pIDArray Is Nothing Then
    MsgBox "No elevation is found, exit sub!"
    Exit Sub
Else
    'MsgBox "GOOD POINT!"
End If

```

```

'Determine the elevation of the midpoint and consequently
'the target elevation
'Set pRIDObj = pIDArray.Element (0)
'Dim base_elev As Double
'base_elev = CDb1 (pRIDObj.Name)
Dim target_elev As Double
target_elev = base_elev + elevadd
Dim currentelev As Double
Dim elev1 As Double
Dim elev2 As Double

```

```

'Identify two points on line with one elevation lower than the target elevation
'and the other with elevation higher than the target elevation
Dim ii As Integer
ii = 0
'The Do loop searches for the target point along the normal line, starting from
'the mid point. Each search step has the distance increase equalling to cell
'width. The target point is located via linear interpolation.
Do
  ii = ii + 1
  pInLine.QueryPoint esriExtendTangentAtTo, ii * cellwidth, False, pNewPoint2
  Set pIDArray2 = pIdentify.Identify (pNewPoint2)
  If pIDArray2 Is Nothing Then
    MsgBox "ii is: " & ii
    Exit Do
  Else
    Set pRIDObj2 = pIDArray2.Element (0)
    If pRIDObj2.Name = "NoData" Then
      MsgBox "new ii is: " & ii
      Exit Do
    Else
      currentelev = CDBl (pRIDObj2.Name)
      elev2 = currentelev
    End If
  End If
Loop While (currentelev <= target_elev) And (Not pIDArray2 Is Nothing) And (pRIDObj2.Name <>
  "NoData")

pInLine.QueryPoint esriExtendTangentAtTo, (ii - 1) * cellwidth, False, pNewPoint1
Set pIDArray1 = pIdentify.Identify (pNewPoint1)
Set pRIDObj1 = pIDArray1.Element (0)
elev1 = CDBl (pRIDObj1.Name)

Set outputpoint = New Point
If pIDArray2 Is Nothing Or pRIDObj2.Name = "NoData" Then
  'The boundary is reached
  delete = True
  outputpoint.X = pNewPoint1.X
  outputpoint.Y = pNewPoint1.Y
Else
  delete = False
  If Abs(elev2 - elev1) < 0.01 Then
    'Two points having elevation very close to each other
    'the target point is located in the middle of the
    'two points
    outputpoint.X = (pNewPoint1.X + pNewPoint2.X) / 2
    outputpoint.Y = (pNewPoint1.Y + pNewPoint2.Y) / 2
  Else
    'Linear interpolation is applied
    outputpoint.X = pNewPoint1.X + (pNewPoint2.X - pNewPoint1.X) * (target_elev - elev1) / (elev2 - elev1)
    outputpoint.Y = pNewPoint1.Y + (pNewPoint2.Y - pNewPoint1.Y) * (target_elev - elev1) / (elev2 - elev1)
  End If
End If
'This flood is the flooded water level of the new polygon being constructed.
flood = target_elev
MsgBox "Base elevatio is: " & base_elev & "; flood elevation is: " & target_elev
End Sub

```

APPENDIX G PROGRAM CODE FOR FLOODED ROAD SEGMENT IDENTIFICATION

'Program Name: FloodedRoadSegmentIdentification

'This program is developed to identify flooded road segments given the resulting polygons representing flood extent and the 3-D road data.

'This program was developed by Hubo Cai during the summer of 2003

Option Explicit

'The main sub determines the road segments in the flood extents. For each of the road segments, the public sub 'myflood is called to determine the flooded road segments within the road segments in the flood extent.

Sub FloodedRoadSegmentIdentification ()

'Open workspace and all the files

Dim pworkspacefactory As IWorkspaceFactory

Set pworkspacefactory = New ShapefileWorkspaceFactory

Dim pfeatureworkspace As IFeatureWorkspace

Set pfeatureworkspace = pworkspacefactory.OpenFromFile ("s:\palrs\users\hubo\Hydro", 0)

'Open 3-D road data file

Dim linefeatureclass As IFeatureClass

Set linefeatureclass = pfeatureworkspace.OpenFeatureClass ("Inter_Wilson")

'Open the flood extent file

Dim polygonfeatureclass As IFeatureClass

Set polygonfeatureclass = pfeatureworkspace.OpenFeatureClass ("Testing_results_500_4")

'Open the new road segment file to store identified flooded road segments

Dim newlinefeatureclass As IFeatureClass

Set newlinefeatureclass = pfeatureworkspace.OpenFeatureClass ("road1")

Dim mylinefeature As IFeature

Dim myline As IPolyline

'Variable linecount is used to identify the number of polylines in the 3-D road data

Dim linecount As Long

linecount = linefeatureclass.FeatureCount (Nothing)

Dim flood As Double

Dim lineID As Long

'The outer loop works with each feature from the 3-D road data

For lineID = 0 To linecount - 1

Set mylinefeature = linefeatureclass.GetFeature (lineID)

Set myline = New Polyline

Set myline = mylinefeature.Shape

Dim myoperator As ITopologicalOperator

Set myoperator = myline

Dim mypolygonfeature As IFeature

'The variable polygoncount is used to count the number of polygons in the flood extent file

Dim polygoncount As Long

polygoncount = polygonfeatureclass.FeatureCount(Nothing)

Dim polygonID As Long

'The inner loop works with every polygon from the flood extent file

For polygonID = 0 To polygoncount - 1

Set mypolygonfeature = polygonfeatureclass.GetFeature (polygonID)

'Obtain the flood level stored with each polygon

flood = mypolygonfeature.value (4) / 10

Dim mypolygon As IPolygon

```

Set mypolygon = New Polygon
Set mypolygon = mypolygonfeature.Shape
Dim newline As IPolyline
Set newline = New Polyline
Dim newfeature As IFeature
'Object newmultipoints will be used to store points for constructing flooded road segments
Dim newmultipoint As IMultipoint
Set newmultipoints = New Multipoint
Dim newpointcollection As IPointCollection
Dim mygeometry As IGeometry
Set mygeometry = New Multipoint
'Each road feature is intersected with each polygon
Set mygeometry = myoperator.intersect (mypolygon, 1)
Set newpointcollection = mygeometry
If newpointcollection.pointcount = 0 Then
Else
    'Work with the intersecting points to figure out the points
    'needed to construct flooded road segments
    Dim ii As Integer
    ii = newpointcollection.pointcount Mod 2
    If ii = 0 Then
    Else
        'identify start and end nodes
        'Judge if start or end nodes are in the polygon
        'Determine if the start or end node should be used to construct flooded road segments
        Dim startnode As IPoint
        Dim endnode As IPoint
        Set startnode = New Point
        Set endnode = New Point
        Set startnode = myline.FromPoint
        Set endnode = myline.ToPoint
        Dim test As IRelationalOperator
        Set test = mypolygon
        Dim startin As Boolean
        Dim endin As Boolean
        startin = test.Contains (startnode)
        endin = test.Contains (endnode)
        If startin Then
            newpointcollection.AddPoint startnode
        ElseIf endin Then
            newpointcollection.AddPoint endnode
        Else
            MsgBox "something wrong!"
            MsgBox "line ID is:" & mylinefeature.value (0)
            MsgBox "polygon ID is:" & mypolygonfeature.value (0)
        Exit Sub
        End If
    End If
End If

Dim pointcount As Integer
pointcount = newpointcollection.pointcount

'The two arrays defined hereby will be used to store those points
'in the order of their distances to the start point of the road segment
'from which they are identified.
'By doing this, all intersected road segments are identified.

```

```

Dim orderedpointcollection (100) As IPoint
Dim dist (100) As Double
Dim i As Integer
Dim out As IPoint
Dim dis1 As Double
Dim dis2 As Double
Dim bright As Boolean

For i = 0 To pointcount - 1
    Set out = New Point
    myline.QueryPointAndDistance esriNoExtension, newpointcollection.Point (i), False, out, dis1, dis2,
                                bright
    dist (i) = dis1
Next

Dim outerloop As Integer
Dim interloop As Integer
Dim min As Double

For outerloop = 0 To pointcount - 2
    min = dist (outerloop)
    For interloop = outerloop + 1 To pointcount - 1
        If dist (interloop) < min Then
            min = dist (interloop)
            dist (interloop) = dist (outerloop)
            dist (outerloop) = min
        Else
            End If
    Next
Next

Dim newlinecount As Integer
newlinecount = pointcount / 2
Dim linecreation As Integer

For linecreation = 0 To newlinecount - 1
    Set newline = New Polyline
    'Call the myflood function to determine the flooded road segments
    'for all road segments that are in the flood extent
    Call myflood(mylinefeature.value (2), dist(linecreation * 2), dist(linecreation * 2 + 1), flood, myline,
                newlinefeatureclass)
Next
End If
Next
Next
End Sub

```

'This public works with the 3-D road data to determine flooded road segments within the road segments in the 'flood extent

```

Public Sub myflood(linemerger As Long, dist1 As Double, dist2 As Double, flood As Double, myline As
                  IPolyline, newlinefeatureclass As IFeatureClass)
    'Open workspace and needed files
    Dim fpworkspacefactory As IWorkspaceFactory
    Set fpworkspacefactory = New ShapefileWorkspaceFactory
    Dim fpfeatureworkspace As IFeatureWorkspace

```

```

Set fpfeatureworkspace = fpworkspacefactory.OpenFromFile
    ("S:\PALRS\users\Hubo\Nroad_english\WILSON", 0)
'This table contains 3-D road data
Dim fmytable As ITable
Set fmytable = fpfeatureworkspace.OpenTable("snap_wilson")
Dim newfeature As IFeature
'Variable CONDITION is that condition that will be used to retrieve all 3-D points enclosed
'by the start and end points of a road segment in the flood extent
Dim CONDITION As String
    CONDITION = "MERGE = " & linemerge & " AND D_TO_S >= " & dist1 & " AND D_TO_S <= " &
        dist2
'Variables CONDITION1 and CONDITION2 will store the necessary conditions for retrieving 3-D
'points from the table, which are used to determine the elevations for the start and end points
'of the road segment in the flood extent, via linear interpolation from two neighboring points.
Dim CONDITION1 As String
Dim CONDITION2 As String
    CONDITION1 = "MERGE = " & linemerge & " AND D_TO_S <= " & dist1
    CONDITION2 = "MERGE = " & linemerge & " AND D_TO_S >= " & dist2

Dim pTableSort As ITableSort
Set pTableSort = New esriCore.TableSort
Dim pQueryFilter As IQueryFilter
Set pQueryFilter = New QueryFilter
pQueryFilter.whereClause = CONDITION1

Dim newline As IPolyline
Set newline = New Polyline

Dim e1 As Double
Dim e2 As Double
Dim ss1 As Double
Dim ss2 As Double

'Obtain the elevation and distance the 3-D point right before dist1 (the start
'point of the road segment in the flood extent
With pTableSort
    .Fields = "D_TO_S"
    .Ascending ("D_TO_S") = True
    Set .QueryFilter = pQueryFilter
    Set .Table = fmytable
End With
pTableSort.Sort Nothing
Dim pCursor As ICursor
Set pCursor = pTableSort.Rows
Dim pRow As IRow
Set pRow = pCursor.NextRow

Do While Not pRow Is Nothing
    e1 = pRow.value (3)
    ss1 = pRow.value (5)
    Set pRow = pCursor.NextRow
Loop

'Obtain the elevation and distance to the start point for the point right after dist2
pQueryFilter.whereClause = CONDITION2
With pTableSort

```

```

.Fields = "D_TO_S"
.Ascending ("D_TO_S") = True
Set .QueryFilter = pQueryFilter
Set .Table = fmytable
End With
pTableSort.Sort Nothing
Set pCursor = Nothing
Set pRow = Nothing
Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow

If (Not pRow Is Nothing) Then
    e2 = pRow.value (3)
    ss2 = pRow.value (5)
Else
    MsgBox "Something wrong!"
    Exit Sub
End If

'Finding 3-D points between dist1 and dist2
pQueryFilter.whereClause = CONDITION
With pTableSort
    .Fields = "D_TO_S"
    .Ascending ("D_TO_S") = True
    Set .QueryFilter = pQueryFilter
    Set .Table = fmytable
End With
pTableSort.Sort Nothing
Set pCursor = Nothing
Set pRow = Nothing
Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow

Dim EE1 As Double
Dim nopoint As Boolean

If (Not pRow Is Nothing) Then
    'Determining the elevation for the start point via linear interpolation
    Dim difference1 As Double
    difference1 = pRow.value (5) - ss1
    If Abs (difference1) < 0.001 Then
        EE1 = pRow.value (3)
    Else
        EE1 = e1 + (pRow.value (3) - e1) * (dist1 - ss1) / (pRow.value (5) - ss1)
    End If
Else
    'In case there is no point between dist1 and dist2, the whole road segment between
    'dist1 and dist2 is assumed to be flooded
    nopoint = True
    Set newline = Nothing
    If dist1 <= dist2 Then
        myline.GetSubcurve dist1, dist2, False, newline
        Set newfeature = newlinefeatureclass.CreateFeature
        Set newfeature.Shape = newline
        newfeature.value (2) = linemerge
        newfeature.value (3) = dist1
    End If
End If

```

```

        newfeature.value (4) = dist2
        newfeature.Store
        'MsgBox "create line at position 1"
        Exit Sub
    Else
    End If
End If

With pTableSort
    .Fields = "D_TO_S"
    .Ascending("D_TO_S") = False
    Set .QueryFilter = pQueryFilter
    Set .Table = fmytable
End With
pTableSort.Sort Nothing
Set pCursor = Nothing
Set pRow = Nothing
Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow

Dim EE2 As Double
'Determining the elevation of the end point
If (Not pRow Is Nothing) Then
    Dim difference2 As Double
    difference2 = ss2 - pRow.value (5)
    If Abs (difference1) < 0.001 Then
        EE2 = pRow.value (3)
    Else
        EE2 = e2 - (e2 - pRow.value (3)) * (dist2 - ss2) / (ss2 - pRow.value (5))
    End If
Else
    'There is no point between dist1 and dist2 and shoule already be dealt with
    Exit Sub
End If

'Start to identify flooded road segments for the road segment in the flood extent
'substart and subend are used to label the start and end points for the flooded segments
Dim substart As Double
Dim subend As Double

Dim previousrow As IRow
Dim currentrow As IRow
Dim afterrow As IRow

If EE1 < flood Then
    'ss1 is the start point for the first flooded segment
    substart = ss1
Else
    'If ss1 is not the start point for the first flooded segment,
    'the first point having elevation lower than the flood water level is identified
    'the first point having elevation equaling to the flood water level is identified vai
    'linear interpolation
    With pTableSort
        .Fields = "D_TO_S"
        .Ascending ("D_TO_S") = True
        Set .QueryFilter = pQueryFilter
    End With

```



```

    Set .Table = fmytable
End With
pTableSort.Sort Nothing
Set pCursor = Nothing
Set pRow = Nothing
Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow
If Not pRow Is Nothing Then
    Set currentrow = pRow
Else
    Exit Sub
End If

Do While (Not pRow Is Nothing)
    If pRow.value (3) > flood Then
        Set currentrow = pRow
        Set pRow = pCursor.NextRow
    Else
        Exit Do
    End If
Loop

If (Not pRow Is Nothing) Then
    'afterrow indicates the first point having its elevation lower than the flood water level
    Set afterrow = pRow
    If (pRow Is currentrow) Then
        'This indicates the start point is the point right before afterrow
        substart = ss1 + (afterrow.value (5) - ss1) * (EE1 - flood) / (EE1 - afterrow.value (3))
    Else
        'This indicates the current row represents the point right before afterrow
        Set previousrow = currentrow
        If Abs(previousrow.value (3) - afterrow.value (3)) < 0.0001 Then
            substart = previousrow.value (5)
        Else
            substart = previousrow.value(5) + (afterrow.value(5) - previousrow.value(5)) * (previousrow.value(3)
            - flood) / (previousrow.value(3) - afterrow.value(3))
        End If
    End If
Else
    'This indicates the failure to find a point having its elevation lower than the flood water level
    'and therefore, the sub is ended. There is no flooded road segment
    Exit Sub
End If
End If

'After identifying the sub start point, it continues to search for the sub end point
'tempRow is a variable used to store the current position of pRow
Dim tempRow As IRow
Set tempRow = pRow

Set pRow = pCursor.NextRow
If (Not pRow Is Nothing) Then
    Set currentrow = pRow
Else
    'Reaches the end
    'use the ss2 point to get the flooded road segment

```

```

If EE2 <= flood Then
    subend = ss2
Else
    subend = tempRow.value(5) + (flood - tempRow.value(3)) * (ss2 - tempRow.value(5)) / (EE2 -
        tempRow.value(3))
End If

'Creating a linear feature for the identified flooded road segment
If substart <= subend Then
    myline.GetSubcurve substart, subend, False, newline
    Set newfeature = newlinefeatureclass.CreateFeature
    Set newfeature.Shape = newline
    newfeature.value (2) = linemerge
    newfeature.value (3) = substart
    newfeature.value (4) = subend
    newfeature.Store
    Set newline = Nothing
Else
    Exit Sub
End If
Exit Sub
End If

'It continues to find the first point with elevation higher than the flood water level
Do While (Not pRow Is Nothing)
    If pRow.value (3) < flood Then
        Set currentrow = pRow
        Set pRow = pCursor.NextRow
        MsgBox "currentrow is not nothing" & currentrow.Value (5)
    Else
        Exit Do
    End If
Loop

'Determining the sub end point, which has its elevation equaling to the flood water level
If (Not pRow Is Nothing) Then
    Set afterrow = pRow
    If (currentrow Is pRow) Then
        Set previousrow = tempRow
    Else
        Set previousrow = currentrow
    End If
    'Set afterrow = pRow
    If Abs(previousrow.value(3) - afterrow.value(3)) < 0.0001 Then
        subend = afterrow.value(5)
    Else
        subend = previousrow.value(5) + (afterrow.value(5) - previousrow.value(5)) * (previousrow.value(3) -
            flood) / (previousrow.value(3) - afterrow.value(3))
    End If

'Creat a new line feature for the identified flooded road segment
If substart <= subend Then
    myline.GetSubcurve substart, subend, False, newline
    Set newfeature = newlinefeatureclass.CreateFeature
    Set newfeature.Shape = newline
    newfeature.value (2) = linemerge

```

```

    newfeature.value (3) = substart
    newfeature.value (4) = subend
    newfeature.Store
    'MsgBox "create line at position 2"
    Set newline = Nothing
Else
End If
Else
'It indicates that it reaches the end, but doesn't find a point with elevation
'higher than the flood water level
'This means that ss2 and currentrow should be used to determine the flooded road segment
If EE2 <= flood Then
    subend = ss2
Else
    subend = currentrow.value(5) + (flood - currentrow.value(3)) * (ss2 - currentrow.value(5)) / (EE2 -
        currentrow.value(3))
End If

If substart <= subend Then
    myline.GetSubcurve substart, subend, False, newline
    Set newfeature = newlinefeatureclass.CreateFeature
    Set newfeature.Shape = newline
    newfeature.value (2) = linemerge
    newfeature.value (3) = substart
    newfeature.value (4) = subend
    newfeature.Store
    Set newline = Nothing
Else
End If
Exit Sub
End If

'Continue the loops to find the next flooded road segment
Do While (Not pRow Is Nothing)
    Set currentrow = pRow
    Set pRow = pCursor.NextRow

    Do While (Not pRow Is Nothing)
        If pRow.value(3) > flood Then
            Set currentrow = pRow
            Set pRow = pCursor.NextRow
        Else
            Exit Do
        End If
    Loop

    If (Not pRow Is Nothing) Then
        Set previousrow = currentrow
        Set afterrow = pRow
        If Abs(previousrow.value(3) - afterrow.value(3)) < 0.0001 Then
            substart = previousrow.value(5)
        Else
            substart = previousrow.value(5) + (afterrow.value(5) - previousrow.value(5)) * (previousrow.value(3) -
                flood) / (previousrow.value(3) - afterrow.value(3))
        End If
        Set currentrow = pRow
    
```

```

Else
  If EE2 < flood Then
    subend = ss2
    substart = currentrow.value(5) + (flood - currentrow.value(3)) * (ss2 - currentrow.value(5)) / (EE2 -
      currentrow.value(3))
    If substart <= subend Then
      myline.GetSubcurve substart, subend, False, newline
      Set newfeature = newlinefeatureclass.CreateFeature
      Set newfeature.Shape = newline
      newfeature.value (2) = linemerge
      newfeature.value (3) = substart
      newfeature.value (4) = subend
      newfeature.Store
      Set newline = Nothing
    Else
      End If
  Else
    End If
  Exit Sub
End If

Set pRow = pCursor.NextRow
If pRow Is Nothing Then
  If ss2 <= flood Then
    subend = ss2
  Else
    subend = currentrow.value(5) + (ss2 - currentrow.value(5)) * (flood - currentrow.value(3)) / (EE2 -
      currentrow.value(3))
  End If

  Set newline = Nothing
  If substart <= subend Then
    myline.GetSubcurve substart, subend, False, newline
    Set newfeature = newlinefeatureclass.CreateFeature
    Set newfeature.Shape = newline
    newfeature.value (2) = linemerge
    newfeature.value (3) = substart
    newfeature.value (4) = subend
    newfeature.Store
    Set newline = Nothing
  Exit Sub
  Else
    End If
Else
  Do While (Not pRow Is Nothing)
    If pRow.value (3) < flood Then
      Set currentrow = pRow
      Set pRow = pCursor.NextRow
    Else
      Exit Do
    End If
  Loop
End If

If Not pRow Is Nothing Then
  Set previousrow = currentrow

```

```

Set afterrow = pRow
If Abs(previousrow.value (3) - afterrow.value (3)) < 0.0001 Then
    subend = afterrow.value(5)
Else
    subend = previousrow.value(5) + (afterrow.value(5) - previousrow.value(5)) * (previousrow.value(3) -
        flood) / (previousrow.value(3) - afterrow.value(3))
End If
Else
    If EE2 > flood Then
        subend = currentrow.value(5) + (ss2 - currentrow.value(5)) * (flood - currentrow.value(3)) / (EE2 -
            currentrow.value(3))
    Else
        Exit Sub
    End If
End If

Set newline = Nothing
If substart <= subend Then
    myline.GetSubcurve substart, subend, False, newline
    Set newfeature = newlinefeatureclass.CreateFeature
    Set newfeature.Shape = newline
    newfeature.value (2) = linemerge
    newfeature.value (3) = substart
    newfeature.value (4) = subend
    newfeature.Store
    Set newline = Nothing
Else
    End If
Loop
End Sub

```