

DATA COLLECTION ISSUES AND THEIR IMPACTS ON THE PAVEMENT
MANAGEMENT DECISIONS

By

Christopher Michael Dean

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Civil Engineering

2012

ABSTRACT

DATA COLLECTION ISSUES AND THEIR IMPACTS ON THE PAVEMENT MANAGEMENT DECISIONS

By

Christopher Michael Dean

The management of the highway network presents major and complex problems that challenge highway administrators. The problems are compounded by the long-term economic climate, which forced highway administrators to reduce costs. Some have adopted data sampling, others decreased the data collection frequency, and still others reduced the entire pavement management system operation. Unfortunately, these actions were made without detailed evaluation of their impacts on the pavement management decisions. Consequently, time series pavement condition and distress data were obtained from four State Highway Agencies and the Minnesota Road Research project (MnROAD). The data were used to evaluate the impacts of data sampling, pavement analysis length, and data imputation on the variability of the data and the accuracy of pavement management decisions.

It is shown that data sampling causes various degrees of errors in pavement management decisions, including the identification of project boundaries, selection of treatment type, and allocation of treatment funds. Furthermore, data sampling yields longer pavement analysis lengths. It is shown that longer analysis lengths do not necessarily improve the analyses. On the contrary, longer analyses lengths have significant adverse impacts on various pavement management decisions. Finally, some of the measured data in the databases were randomly amputated to simulate missing data and imputed using different methods to bridge the gap. It is shown that all methods yield differences between the measured and the imputed data. However, in general, such differences do not significantly impact the pavement management decisions.

TO EVERYONE THAT LOVES AND BELIEVES IN ME

ACKNOWLEDGMENTS

I would like to extend my deepest appreciation to my advisor, Dr. Gilbert Y. Baladi, for his support, patience, and encouragement throughout my graduate studies. His commitment to excellence and ever-lasting excitement for research are unparalleled. Without his guidance, this thesis would not have been possible. I would also like to thank him for the innumerable lessons that he has taught me throughout my studies. These lessons are timeless and will continue to serve me throughout my professional career. I am forever indebted to him and grateful for all the opportunities that he bestowed upon me.

I would also like to thank the other members of my advisory committee, Dr. Neeraj Buch and Dr. Karim Chatti. Their technical knowledge and support has been crucial to my success. Many thanks also to Dr. Syed Haider for his research guidance and support.

It is also important to acknowledge the Federal Highway Administration (FHWA) for sponsoring and funding this study. Thanks to the Colorado Department of Transportation (CDOT), the Louisiana Department of Transportation and Development (LADOTD), the Michigan Department of Transportation (MDOT), the Washington State Department of Transportation (WSDOT), and the Minnesota Road Research project (MnROAD) for their cooperation during this study. Finally, special thanks to my friend and co-researcher Tyler Dawson. Many thanks also to Adam Beach, Matthew McCloskey, Ryan Muscott, Corbin St. Aubin, and Nick Tecca for their efforts in this study.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	x
CHAPTER 1	
INTRODUCTION & RESEARCH PLAN	1
1.1 Introduction	1
1.2 Problem Statement and Research Plan	3
1.3 Objectives	5
1.4 Thesis Layout	5
CHAPTER 2	
LITERATURE REVIEW	6
2.1 Introduction to Pavement Management	6
2.1.1 The Need to Manage Pavements	7
2.1.2 Pavement Management versus Managing Pavement	7
2.1.3 Acts of Managing Pavements	8
2.1.4 Network-Level Pavement Management	10
2.1.5 Project-Level Pavement Management	10
2.1.6 Commonalities and Differences Between Network and Project-Level Pavement Management	11
2.2 Data Elements and Data Collection Procedures	13
2.3 Data Collection Frequency	22
2.4 Pavement Project Boundaries (Data Delineation)	24
2.5 Pavement Performance Data Sampling	27
2.6 Pavement Survey and Analysis Length	29
2.7 Data Imputation	31
2.8 Location Referencing Systems	35
2.9 Pavement Performance Data Modeling and Remaining Service Life (RSL)	42
2.9.1 The Need for Data Modeling	42
2.9.2 International Roughness Index (IRI) Modeling	43
2.9.3 Rut Depth Modeling	43
2.9.4 Crack Modeling	44
2.9.5 Uses of Data Modeling	46
2.10 Life Cycle Cost Analysis (LCCA)	49
CHAPTER 3	
DATA MINING	55
3.1 Foreword	55
3.2 Introduction	55
3.3 Data Format, Restructuring, and Unification	56
3.3.1 Pavement Condition and Distress Data from Four State Highway Agencies	62

3.3.2	Minnesota Road Research Project (MnROAD)	70
3.4	Treated Pavement Section Identification	71
3.5	Time Series Data Restructuring	73
3.6	Cost Data.....	74
CHAPTER 4		
	DATA ANALYSES & DISCUSSION	77
4.1	Introduction and Research Objectives	77
4.2	Pavement Performance Data Sampling	78
4.2.1	Fixed Sampling.....	80
4.2.2	Random Sampling	102
4.2.3	Sample Size	106
4.3	Pavement Performance Data Sampling: Condition State Concept	109
4.4	Pavement Project Boundary Analysis.....	124
4.4.1	Pavement Project Boundary Analysis: Condition and Distress Data Approach	128
4.4.2	Pavement Project Boundary Analysis: RSL Concept	136
4.5	The Hidden Costs of Pavement Performance Data Sampling	140
4.6	Pavement Analysis Length	144
4.6.1	The Impacts of Pavement Analysis Length on the Variability and Modeling of Pavement Performance Data.....	145
4.6.2	The Impacts of Pavement Analysis Length on PMS Decisions	159
4.7	Pavement Performance Data Imputation	161
4.7.1	Linear Interpolation Imputation	163
4.7.2	Regression Imputation.....	166
4.7.3	Moving Regression Imputation	170
4.7.4	Multiple Regressions Imputation	173
4.7.5	Discussion	178
4.7.6	The No-Action (No Imputation) Approach.....	182
4.7.7	Comparison of the Four Imputation Methods and the No-Action Approach.....	183
4.7.8	Advantages and Disadvantages of the Four Imputation Methods.....	188
CHAPTER 5		
	SUMMARY, CONCLUSIONS, & RECOMMENDATIONS	192
5.1	Summary	192
5.2	Conclusions.....	194
5.3	Recommendations.....	197
APPENDIX A		
	PMS DATA	200
APPENDIX B		
	PAVEMENT PERFORMANCE DATA SAMPLING FIGURES	208
APPENDIX C		
	PAVEMENT ANALYSIS LENGTH FIGURES	264

APPENDIX D	
PAVEMENT PERFORMANCE DATA IMPUTATION FIGURES.....	269
REFERENCES	286

LIST OF TABLES

Table 2.1 Pavement inventory and monitoring data	15
Table 2.2 Main advantage(s) and disadvantage(s) of each location reference system	40
Table 2.3 LADOTD linear referencing system error (after Dawson 2012).....	41
Table 2.4 Mathematical functions used to model the data of the indicated distress type.....	45
Table 2.5 Example distress thresholds.....	48
Table 3.1 Pavement condition and distress data stored by various SHAs	57
Table 3.2 The pavement condition and distress data received from the four SHAs.....	63
Table 3.3 MnROAD database items	71
Table 3.4 The number of pavement projects and total length identified in each state	73
Table 3.5 Typical preventive maintenance treatment costs in 2009 for the state of Michigan (Baladi and Dean 2011)	75
Table 3.6 Typical preventive maintenance treatment costs in 2010 for the state of Michigan (Baladi and Dean 2011)	76
Table 3.7 Typical material and total treatment costs for the state of Michigan (Baladi & Dean 2011)	76
Table 4.1 Mathematical functions used to model the pavement condition and distress data	112
Table 4.2 Pavement condition state, RSL range, and probable pavement treatment actions	113
Table 4.3 Percent of each road changing certain number of condition states based on the IRI data	115
Table 4.4 Solution sequence of the AASHTO unit delineation method (after AASHTO 1993)	125
Table 4.5 Example of the solution sequence of the AASHTO unit delineation method using IRI data for 2 miles of HWY 24 in Colorado.....	126
Table 4.6 Total pavement expenditure, data digitization, and total data collection costs for three states (Fillastre 2011, Hartgen et al. 2009)	142

Table 4.7 Total expenditures, potential savings, and potential misallocation of funds as a result of the use of ten percent sampling for various pavement networks	143
Table 4.8 The number of pavement projects and their total length identified for analysis in each state	146
Table 4.9 Number of analysis sections for various analysis lengths along a 2.2 mile long project	149
Table 4.10 Coefficients of variation for ten pavement analysis lengths based on three years of IRI and longitudinal cracking data along US-2 in Washington, BMP 89.3 to 99.0.....	157
Table 4.11 Linear interpolation imputation example.....	164
Table 4.12 Regression imputation example.....	167
Table 4.13 Moving regression imputation example	171
Table 4.14 Multiple regressions imputation example.....	175
Table 4.15 Differences and percent differences between the imputed and the actual IRI data for four imputation methods (three rates of deterioration and no variability).....	181
Table 4.16 Differences and percent differences between the imputed and the actual IRI data for four imputation methods (three rates of deterioration and variable data).....	181
Table 4.17 Advantages and disadvantages of four data imputation methods.....	189
Table A.1 Measured pavement condition data along I-70 in Colorado.....	202
Table A.2 Formatted pavement condition data along I-70 in Colorado	203
Table A.3 Formatted pavement condition data along US-80 in Louisiana.....	204
Table A.4 Formatted pavement condition data along I-69BL in Michigan.....	205
Table A.5 Formatted pavement condition data along I-5 in Washington.....	206
Table A.6 Formatted pavement condition data along cell 1 at MnROAD	207

LIST OF FIGURES

Figure 2.1 Network and project-levels pavement management (AASHTO 1993, Hudson et al. 1979)	12
Figure 2.2 Percent of SHAs using the stated method of pavement condition data collection	21
Figure 2.3 Area scan image for fully-automated data collection (McGhee 2004)	21
Figure 2.4 Percent of SHAs using the indicated PMS data collection frequency.....	23
Figure 2.5 A typical pavement response variable versus distance for data delineation (after AASHTO 1993)	24
Figure 2.6 The Idealized Approach for data delineation (AASHTO 1993)	26
Figure 2.7 Mean absolute error for four imputation methods (after Farhan and Fwa 2012).....	36
Figure 2.8 Percent of SHAs utilizing stated location referencing systems (after Flintsch et al. 2004)	37
Figure 2.9 Illustrative example of yearly error of linear referencing system (Dawson 2012).....	41
Figure 2.10 Illustration of the mathematical functions used to model IRI, rutting, and cracking and to predict future conditions	45
Figure 2.11 Illustration of the Remaining Service Life using three mathematical models	50
Figure 2.12 Components of user costs (after Morgado and Neves 2009)	52
Figure 3.1 Time series transverse cracking data for each severity level and the sum of all levels, Colorado, HWY 24, direction 2, BMP 329.9	60
Figure 3.2 Cumulative time series transverse cracking data showing individual transverse crack severity level and the sum of all severity levels, Colorado, HWY 24, direction 2, BMP 329.9 .	60
Figure 3.3 The BMP of the first ten 0.1 mile long pavement segments along M-39, control section 82192, direction 1, Michigan.....	68
Figure 3.4 Longitudinal cracking data after adjusting the BMP for the first ten 0.1 mile long pavement segments along M-39, control section 82192, direction 1, Michigan	68
Figure 4.1 Continuous and sampled IRI data along four roads in four different states	83

Figure 4.2 Continuous and sampled transverse crack length data along four roads in four different states	84
Figure 4.3 Sampled (squares) and the range of the continuous IRI data for each mile	87
Figure 4.4 Sampled (squares) and the range of the continuous transverse cracking data for each mile	88
Figure 4.5 Continuous time series IRI data as percent of the sampled data along a portion of the flexible HWY 24 in Colorado	91
Figure 4.6 Continuous time series IRI data as percent of the sampled data along a portion of the flexible LA-34 (control section 0.12304) in Louisiana	92
Figure 4.7 Continuous time series transverse crack data as percent of the sampled data along a portion of the rigid I-69 (control section 77024) in Michigan	93
Figure 4.8 Continuous time series transverse crack data as percent of the sampled data along the composite SRID 005 in Washington	94
Figure 4.9 Maximum, minimum, and average percent differences between the continuous and the sampled transverse cracking, relative to the sampled data, Colorado	96
Figure 4.10 Maximum, minimum, and average percent differences between the continuous and the sampled transverse cracking, relative to the sampled data, Louisiana	97
Figure 4.11 Maximum, minimum, and average differences between the continuous and the sampled transverse cracking, relative to the sampled data, Michigan	98
Figure 4.12 Maximum, minimum, and average percent differences between the continuous and the sampled transverse cracking, relative to the sampled data, Washington	99
Figure 4.13 Continuous and randomly sampled IRI data per mile along four roads in Colorado	104
Figure 4.14 Continuous and randomly sampled transverse cracking data along four roads in Colorado	105
Figure 4.15 Maximum, minimum, and average continuous/sampled transverse cracking ratio versus sample size along two roads in Michigan	108
Figure 4.16 Percent of two roads in Colorado and two in Louisiana versus the numbers of condition state change, IRI data	116
Figure 4.17 Percent of two roads in Michigan and two in Washington versus the numbers of condition state change, IRI data	117

Figure 4.18 Distribution of the numbers of condition state change based on IRI data for the approximately 97 mile network	120
Figure 4.19 Percent of two roads in Washington versus the numbers of condition state change, rut depth and longitudinal cracking data.....	123
Figure 4.20 Cumulative differences versus beginning mile point along the flexible HWY 24 in Colorado showing three pavement projects	127
Figure 4.21 Cumulative differences versus beginning mile point along the flexible HWY 24 in Colorado showing two pavement projects	127
Figure 4.22 Project boundaries based on the AASHTO unit delineation method and the continuous IRI data along 13 miles of the flexible HWY 24 in Colorado	130
Figure 4.23 Project boundaries based on the AASHTO unit delineation method and the ten percent sampled IRI data along 13 miles of the flexible HWY 24 in Colorado	130
Figure 4.24 Project boundaries based on the continuous (100 percent sampled) and the sampled IRI data along a section of the flexible HWY 24 in Colorado.....	131
Figure 4.25 Project boundaries based on the continuous (100 percent sampled) and the sampled IRI data along a section of the composite I-94 in Michigan.....	132
Figure 4.26 Project boundaries based on the continuous (100 percent sampled) and sampled IRI data along a section of the composite SRID 161 in Washington.....	132
Figure 4.27 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (IRI) data along a section of the flexible HWY 24 in Colorado	138
Figure 4.28 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (IRI) data along a section of the composite I-94 in Michigan	138
Figure 4.29 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (IRI) data along a section of the composite SRID 161 in Washington.....	139
Figure 4.30 Percent acceptance versus pavement analysis length, thin HMA overlay, Colorado	151
Figure 4.31 Percent acceptance versus pavement analysis length, thin HMA overlay, Louisiana	151
Figure 4.32 Percent acceptance versus pavement analysis length, thin HMA overlay, Washington	152
Figure 4.33 Percent acceptance versus pavement analysis length, single chip seal, Colorado ..	152

Figure 4.34 Percent acceptance versus pavement analysis length, single chip seal, Louisiana .	153
Figure 4.35 Percent acceptance versus pavement analysis length, single chip seal, Washington	153
Figure 4.36 Ratio of the coefficients of variation for the longitudinal cracking and IRI data versus pavement analysis length along US-2 in Washington, BMP 89.3 to 99.0.....	157
Figure 4.37 The difference between the imputed and measured data in percent of the measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the linear interpolation imputation method.....	164
Figure 4.38 The difference between the RSL values based on imputed and measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the linear interpolation imputation method	165
Figure 4.39 Example regression parameter calculation for regression imputation, missing data at year 7.....	167
Figure 4.40 The difference between the imputed and measured data in percent of the measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the regression imputation method	168
Figure 4.41 The difference between the RSL values based on imputed and measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the regression imputation method.....	169
Figure 4.42 The difference between the imputed and measured data in percent of the measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the moving regression imputation method.....	172
Figure 4.43 The difference between the RSL values based on imputed and measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the moving regression imputation method	173
Figure 4.44 The difference between the imputed and measured data in percent of the measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the multiple regressions imputation method	176
Figure 4.45 The difference between the RSL values based on imputed and measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the multiple regressions imputation method	177
Figure 4.46 IRI data having three deterioration rates and no variability	179
Figure 4.47 Variable IRI data having three deterioration rates	180

Figure 4.48 Maximum, minimum, and average percent difference between imputed and measured data at BMPs 321, 322, 323, 324, and 325 along HWY 24 in Colorado, using four imputation methods.....	184
Figure 4.49 Differences between the RSL based on the stated imputation method (second available data point imputed) and that based on the complete measured data at BMPs 321, 322, 323, 324, and 325 along HWY 24 in Colorado	185
Figure 4.50 Differences between the RSL based on the stated imputation method (third available data point imputed) and that based on the complete measured data at BMPs 321, 322, 323, 324, and 325 along HWY 24 in Colorado	185
Figure 4.51 Differences between the RSL based on the stated imputation method (fourth available data point imputed) and that based on the complete measured data at BMPs 321, 322, 323, 324, and 325 along HWY 24 in Colorado	186
Figure 4.52 Maximum, minimum, and average difference between RSL values based on imputed and measured data at BMPs 321, 322, 323, 324, and 325 along HWY 24 in Colorado, using four imputation methods.....	187
Figure B.1 Continuous and sampled IRI data along four roads in Colorado.....	209
Figure B.2 Continuous and sampled IRI data along four roads in Louisiana.....	210
Figure B.3 Continuous and sampled IRI data per mile along four roads in Michigan	211
Figure B.4 Continuous and sampled IRI data per mile along four roads in Washington	212
Figure B.5 Continuous and sampled transverse crack data along four roads in Colorado	213
Figure B.6 Continuous and sampled transverse crack data per mile along four roads in Louisiana	214
Figure B.7 Continuous and sampled transverse crack data per mile along four roads in Michigan	215
Figure B.8 Continuous and sampled transverse crack data per mile along four roads in Washington	216
Figure B.9 Sampled (squares) and continuous IRI data per mile along four roads in Colorado	217
Figure B.10 Sampled (squares) and continuous IRI data per mile along four roads in Louisiana	218
Figure B.11 Sampled (squares) and continuous IRI data per mile along four roads in Michigan	219

Figure B.12 Sampled (squares) and continuous IRI data per mile along four roads in Washington	220
Figure B.13 Sampled (squares) and continuous transverse crack data per mile along four roads in Colorado.....	221
Figure B.14 Sampled (squares) and continuous transverse crack data per mile four roads in Louisiana.....	222
Figure B.15 Sampled (squares) and continuous transverse crack data per mile along four roads in Michigan	223
Figure B.16 Sampled (squares) and continuous transverse crack data per mile along four roads in Washington	224
Figure B.17 Continuous time series IRI data as percent of the sampled data along HWY 24 (flexible) in Colorado.....	225
Figure B.18 Continuous time series IRI data as percent of the sampled data along I-70 (rigid) in Colorado.....	226
Figure B.19 Continuous time series IRI data as percent of the sampled data along HWY 36 (composite) in Colorado	227
Figure B.20 Continuous time series IRI data as percent of the sampled data along HWY 71 (flexible) in Colorado.....	228
Figure B.21 Continuous time series IRI data as percent of the sampled data along LA-34 (flexible) in Louisiana.....	229
Figure B.22 Continuous time series IRI data as percent of the sampled data along LA-83 (rigid) in Louisiana.....	230
Figure B.23 Continuous time series IRI data as percent of the sampled data along LA-1 (composite) in Louisiana.....	231
Figure B.24 Continuous time series IRI data as percent of the sampled data along LA-526 (rigid) in Louisiana.....	232
Figure B.25 Continuous time series IRI data as percent of the sampled data along I-94 (composite) in Michigan	233
Figure B.26 Continuous time series IRI data as percent of the sampled data along I-69 (CS 12033, rigid) in Michigan	234

Figure B.27 Continuous time series IRI data as percent of the sampled data along US-31 (flexible) in Michigan	235
Figure B.28 Continuous time series IRI data as percent of the sampled data along I-69 (CS 77024, rigid) in Michigan	236
Figure B.29 Continuous time series IRI data as a percent of the sampled data along SRID 005 (composite) in Washington	237
Figure B.30 Continuous time series IRI data as a percent of the sampled data along SRID 161 (composite) in Washington	238
Figure B.31 Continuous time series IRI data as a percent of the sampled data along SRID 005 (flexible) in Washington	239
Figure B.32 Continuous time series IRI data as a percent of the sampled data along SRID 082 (rigid) in Washington	240
Figure B.33 Continuous time series transverse crack data as percent of the sampled data along HWY 24 (flexible) in Colorado	241
Figure B.34 Continuous time series transverse crack data as percent of the sampled data along I-70 (rigid) in Colorado	242
Figure B.35 Continuous time series transverse crack data as percent of the sampled data along HWY 36 (composite) in Colorado	243
Figure B.36 Continuous time series transverse crack data as percent of the sampled data along HWY 71 (flexible) in Colorado	244
Figure B.37 Continuous time series transverse crack data as percent of the sampled data along LA-34 (flexible) in Louisiana	245
Figure B.38 Continuous time series transverse crack data as percent of the sampled data along LA-83 (rigid) in Louisiana	246
Figure B.39 Continuous time series transverse crack data as percent of the sampled data along LA-1 (composite) in Louisiana	247
Figure B.40 Continuous time series transverse crack data as percent of the sampled data along LA-526 (rigid) in Louisiana	248
Figure B.41 Continuous time series transverse crack data as percent of the sampled data along I-94 (composite) in Michigan	249

Figure B.42 Continuous time series transverse crack data as percent of the sampled data long I-69 (CS 12033, rigid) in Michigan	250
Figure B.43 Continuous time series transverse crack data as percent of the sampled data along US-31 (flexible) in Michigan.....	251
Figure B.44 Continuous time series transverse crack data as percent of the sampled data along I-69 (CS 77024, rigid) in Michigan	252
Figure B.45 Continuous time series transverse crack data as percent of the sampled data along SRID 005 (composite) in Washington	253
Figure B.46 Continuous time series transverse crack data as percent of the sampled data along SRID 161 (composite) in Washington.....	254
Figure B.47 Continuous time series transverse crack data as percent of the sampled data along SRID 005 (flexible) in Washington	255
Figure B.48 Continuous time series transverse crack data as percent of the sampled data along SRID 082 (rigid) in Washington.....	256
Figure B.49 Continuous and randomly sampled IRI data per mile along four roads in Louisiana	257
Figure B.50 Continuous and randomly sampled IRI data per mile along four roads in Michigan	258
Figure B.51 Continuous and randomly sampled transverse crack data along four roads in Louisiana.....	259
Figure B.52 Continuous and randomly sampled transverse crack data along four roads in Michigan	260
Figure B.53 Project boundaries based on the continuous (100 percent sampled) and the sampled rut depth data along a section of the flexible HWY 24 in Colorado.....	261
Figure B.54 Project boundaries based on the continuous (100 percent sampled) and the sampled transverse cracking data along a section of the flexible HWY 24 in Colorado.....	261
Figure B.55 Project boundaries based on the continuous (100 percent sampled) and the sampled rut depth data along a section of the composite SRID 161 in Washington	262
Figure B.56 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (rut depth) data along a section of the flexible HWY 24 in Colorado.....	262

Figure B.57 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (transverse cracking) data along a section of the flexible HWY 24 in Colorado	263
Figure B.58 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (rut depth) data along a section of the composite SRID 161 in Washington.....	263
Figure C.1 Percent acceptance versus pavement analysis length, thin mill and fill, Colorado ..	265
Figure C.2 Percent acceptance versus pavement analysis length, thick HMA overlay, Louisiana	265
Figure C.3 Percent acceptance versus pavement analysis length, double chip seal, Louisiana .	266
Figure C.4 Percent acceptance versus pavement analysis length, thin mill and fill, Louisiana .	266
Figure C.5 Percent acceptance versus pavement analysis length, thick mill and fill, Louisiana	267
Figure C.6 Percent acceptance versus pavement analysis length, thick HMA overlay, Washington	267
Figure C.7 Percent acceptance versus pavement analysis length, thin mill and fill, Washington	268
Figure D.1 The difference between the imputed and measured data in percent of the measured data versus the BMP along LA 21 (control section 0.02902) in Louisiana, three data points imputed using the linear interpolation method	270
Figure D.2 The difference between the imputed and measured data in percent of the measured data versus cells 1, 2, 3, 5, and 14 at MnROAD, three data points imputed using the linear interpolation method	270
Figure D.3 The difference between the imputed and measured data in percent of the measured data versus the BMP along SRID 099 in Washington, three data points imputed using the linear interpolation method	271
Figure D.4 The difference between the RSL values based on imputed and measured data versus the BMP along LA 21 (CS 0.02902) in Louisiana, three data points imputed using the linear interpolation method	271
Figure D.5 The difference between the RSL values based on imputed and measured data versus the cell number at MnROAD, three data points imputed using the linear interpolation method	272
Figure D.6 The difference between the RSL values based on imputed and measured data versus the BMP along SRID 099 in Washington, three data points imputed using the linear interpolation method.....	272

Figure D.7 The difference between the imputed and measured data in percent of the measured data versus the BMP along LA 21 (control section 0.02902) in Louisiana, three data points imputed using the regression method	273
Figure D.8 The difference between the imputed and measured data in percent of the measured data versus cells 1, 2, 3, 5, and 14 at MnROAD, three data points imputed using the regression method.....	273
Figure D.9 The difference between the imputed and measured data in percent of the measured data versus the BMP along SRID 099 in Washington, three data points imputed using the regression method	274
Figure D.10 The difference between the RSL values based on imputed and measured data versus the BMP along LA 21 (CS 0.02902) in Louisiana, three data points imputed using the regression method.....	274
Figure D.11 The difference between the RSL values based on imputed and measured data versus the cell number at MnROAD, three data points imputed using the regression method	275
Figure D.12 The difference between the RSL values based on imputed and measured data versus the BMP along SRID 099 in Washington, three data points imputed using the regression method	275
Figure D.13 The difference between the imputed and measured data in percent of the measured data versus the BMP along LA 21 (control section 0.02902) in Louisiana, three data points imputed using the moving regression method	276
Figure D.14 The difference between the imputed and measured data in percent of the measured data versus cells 1, 2, 3, 5, and 14 at MnROAD, three data points imputed using the moving regression method	276
Figure D.15 The difference between the imputed and measured data in percent of the measured data versus the BMP along SRID 099 in Washington, three data points imputed using the moving regression method	277
Figure D.16 The difference between the RSL values based on imputed and measured data versus the BMP along LA 21 (CS 0.02902) in Louisiana, three data points imputed using the moving regression method	277
Figure D.17 The difference between the RSL values based on imputed and measured data versus the cell number at MnROAD, three data points imputed using the moving regression method	278
Figure D.18 The difference between the RSL values based on imputed and measured data versus the BMP along SRID 099 in Washington, three data points imputed using the moving regression method.....	278

Figure D.19 The difference between the imputed and measured data in percent of the measured data versus the BMP along LA 21 (control section 0.02902) in Louisiana, three data points imputed using the multiple regressions method.....	279
Figure D.20 The difference between the imputed and measured data in percent of the measured data versus cells 1, 2, 3, 5, and 14 at MnROAD, three data points imputed using the multiple regressions method.....	279
Figure D.21 The difference between the imputed and measured data in percent of the measured data versus the BMP along SRID 099 in Washington, three data points imputed using the multiple regressions method	280
Figure D.22 The difference between the RSL values based on imputed and measured data versus the BMP along LA 21 (CS 0.02902) in Louisiana, three data points imputed using the multiple regressions method.....	280
Figure D.23 The difference between the RSL values based on imputed and measured data versus the cell number at MnROAD, three data points imputed using the multiple regressions method	281
Figure D.24 The difference between the RSL values based on imputed and measured data versus the BMP along SRID 099 in Washington, three data points imputed using the multiple regressions method.....	281
Figure D.25 Maximum, minimum, and average percent difference between imputed and measured data at BMPs 3, 4, 5, 6, and 7 along LA 21 (CS 0.02902) in Louisiana, using four imputation methods.....	282
Figure D.26 Maximum, minimum, and average percent difference between imputed and measured data at cells 1, 2, 3, 5, and 14 at MnROAD, using four imputation methods.....	282
Figure D.27 Maximum, minimum, and average percent difference between imputed and measured data at BMPs 10.5, 10.6, 11.2, 11.6, and 11.7 along SRID 099 in Washington, using four imputation methods	283
Figure D.28 Maximum, minimum, and average difference between RSL values based on imputed and measured data at BMPs 3, 4, 5, 6, and 7 along LA 21 (CS 0.02902) in Louisiana, using four imputation methods	283
Figure D.29 Maximum, minimum, and average difference between RSL values based on imputed and measured data at cells 1, 2, 3, 5, and 14 at MnROAD, using four imputation methods	284
Figure D.30 Maximum, minimum, and average difference between RSL values based on imputed and measured data at BMPs 10.5, 10.6, 11.2, 11.6, and 11.7 along SRID 099 in Washington, using four imputation methods	284

CHAPTER 1

INTRODUCTION & RESEARCH PLAN

1.1 Introduction

The highway systems in the United States and most industrialized nations are considered to be the largest investments ever made in the history of these nations (Baladi et al. 1992). The investments are on-going as the highway systems deteriorate and cannot be made to last forever. A constant flow of money is required on an annual basis to properly and continually maintain, rehabilitate, and reconstruct the existing highway systems and expand them in areas of growth. Given the complexity of the highway systems, the size of the required annual investment, and the state of the economy, the management of the highway systems presents major and complex problems that challenge highway administrators. The problems could be solved and the challenge could be met through the design and implementation of an effective, accurate, comprehensive, and integrated pavement management system (PMS).

The most critical parts of a PMS are considered to be the collection, storage, and analysis of the pavement performance data (Smith et al. 1998). Each of these parts is briefly described below:

1. Data collection - Nearly all State Highway Agencies (SHAs) collect pavement performance data at 1-, 2-, or 3-year frequency. Some SHAs such as the Washington State Department of Transportation (WSDOT) survey the pavement network each year. Others survey half the pavement network each year and still others survey one third of the pavement network each year. Regardless of the data collection frequency, most agencies use semi- or fully-automated means of data collection along the outer traffic (driving) lane (McGhee 2004). Further, some SHAs use sampling techniques where the

pavement performance data are collected along a short pavement segment (such as 100, 200, 300 feet, 0.1 mile, etc.) of each one mile and the data are assumed to represent the pavement conditions along the entire mile (Zimmerman 1995). Given the current financial situation of most SHAs and in order to reduce data collection costs, many administrators are under tremendous pressure to sample the pavement performance data. Unfortunately, the effects of the pavement performance data sampling on the accuracy of the PMS decisions are not well known. Such effects should be studied in detail before a decision to sample is made.

2. Data storage – The data storage practices vary greatly from one SHA to another. Some SHAs survey and store the pavement performance data for each 0.1 mile pavement segment along the network while others use greater lengths. For those SHAs that use sampling, the decision to store the data along different lengths is dictated by the sampling technique. For instance, if the conditions along a short pavement segment (i.e. 0.1 mile) are assumed to represent the conditions along 1 mile, then the storage length is also 1 mile. Alternatively, if the conditions along a short pavement segment (i.e. 250 feet) are assumed to represent the conditions along 0.5 mile, then the storage length is also 0.5 mile. It was thought that, for performance data stored along short pavement segments, the variability of the pavement condition data from one year to the next makes the analysis of pavement performance data a difficult task. To reduce data variability and improve data analysis, some advocate data storage and analysis over greater lengths such as half or one mile. Longer analysis lengths may not improve the data analysis and may disadvantage or jeopardize the accuracy of the PMS decisions. Hence, in this research, the impacts, if any,

of data analysis lengths on the accuracy of the data analysis results and on the PMS decisions are assessed and discussed.

3. Data analysis - For certain data collection cycles, some data elements of some pavement segments along the network could be missing for various reasons. These include skipping one data collection cycle for certain pavement sections, equipment malfunctioning, failure to properly record the data, loss of data due to human error, delays in issuing the data collection/storage contract, construction zones and so forth. This poses a serious issue because, typically, at least three data points are required in order to properly model pavement performance data. Because some data could be missing and less than three data points may be available, the question may arise of how to analyze the incomplete data sets. This question could be answered by evaluating how to impute the missing data so that analysis of the time-series performance data could be improved leading to better and more accurate PMS decisions (Bennett 2004, Flintsch and McGhee 2009, Zhang and Smadi 2009). Thus, a study is necessary to evaluate the effects of various data imputation techniques on the accuracy of the pavement performance data.

Given the SHAs practice regarding data collection, storage, and analysis the goal of this research study is to evaluate the impacts of pavement condition data collection practices and data analysis techniques on the accuracy of the PMS decisions.

1.2 Problem Statement and Research Plan

The long-term economic climate has caused some highway administrators to consider pavement performance data sampling as an alternative to reduce data collection costs. However the effects of data sampling on the accuracy of the PMS decisions are not well documented. It is not clear that data sampling ultimately reduces the overall PMS costs. Furthermore, for those

SHAs that sample the pavement performance data, the pavement lengths for which the performance data are stored and analyzed are dictated by the sampling techniques used as stated in item 2 in the previous section. Longer analysis lengths may not improve the data analysis results and may artificially decrease the data variability. Thus, the impact of the pavement analysis length on the variability of the pavement performance data and, ultimately, on the accuracy of the PMS decisions should be evaluated before a data analysis length is selected. Finally, for various reasons, data from some data collection cycles along certain pavement segments may not be documented or are missing, which makes it difficult or impossible to model and analyze the data. In order to solve this problem, the effects of different data imputation techniques on the accuracy of the pavement performance data should be evaluated. Accordingly, a comprehensive research plan was designed to address the three problems through the analyses of the impacts of data sampling, data analysis length, and data imputation on the accuracy of the pavement performance data and the PMS decisions.

The research plan consists of the three tasks stated below.

Task 1: Use continuously collected pavement performance data to simulate sampling and study the effects of sampling on the accuracy of PMS decisions.

Task 2: Use the measured and stored PMS data along each 0.1 mile pavement segment to calculate the average or cumulative data along longer segments and hence to analyze the impact of data analysis lengths on the accuracy of PMS decisions.

Task 3: Scrutinize the various data imputation methods to determine whether or not these methods can be used to impute pavement performance data with little to no impacts on the accuracy of the pavement performance data and PMS decisions

1.3 Objectives

The objectives of this research study are to:

1. Assess the impacts of pavement performance data sampling on the accuracy of PMS decisions and quantify the hidden costs of data sampling.
2. Analyze the effects of data analysis length on the variability of the pavement condition data and on the accuracy of the PMS decisions.
3. Study the impacts of various data imputation methodologies on the variability and accuracy of the pavement condition data and the corresponding PMS decisions.

1.4 Thesis Layout

This thesis consists of five chapters, four appendices, and references as listed below.

Chapter 1 – Introduction & Research Plan

Chapter 2 – Literature Review

Chapter 3 – Data Mining

Chapter 4 – Data Analyses & Discussion

Chapter 5 – Summary, Conclusions, & Recommendations

Appendix A – PMS Data

Appendix B – Pavement Performance Data Sampling Figures

Appendix C – Pavement Analysis Length Figures

Appendix D – Pavement Performance Data Imputation Figures

References

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction to Pavement Management

The cost-effective construction, maintenance, rehabilitation, and preservation of pavements in road systems involve complex decisions to keep the systems performing at reasonable levels. In the past, these decisions have been left to road supervisors who relied on their extensive knowledge and experience (FHWA 2011). Since the early 1980s however, highway administrators and engineers have recognized the need to use a pavement management system (PMS) to aid in the decision-making process (Zimmerman and Peshkin 2004).

A PMS can be described as a “coordinated set of activities directed toward achieving the best value possible for the available public funds in providing and operating smooth, safe, and economical pavements” (Hudson et al. 1979). Although each PMS has its own level of complexity depending on the geographical, financial, and political conditions in which it operates, most systems include five key components (Peterson 1987):

1. Pavement surveys related to condition and serviceability.
2. Database containing all pavement related information.
3. Analysis scheme.
4. Decision criteria.
5. Implementation procedures.

The FHWA databank indicates that each of the 50 states, the District of Columbia, and Puerto Rico have some kind of PMS. Forty-two states report that their systems include a method of prioritization and 20 states include an optimization provision for purposes of budget planning and project programming (Finn 1998). Thus, it is clear that most states recognize the importance

of a PMS; however the degree to which the different components of a PMS are utilized by the various highway agencies varies significantly.

2.1.1 The Need to Manage Pavements

The FHWA estimates that the roadway systems in the United States consist of approximately 8,500,000 lane-miles (WSDOT 2010). Furthermore, the systems are considered to be the single largest investment ever made in the history of industrialized nations (Baladi et al. 1992). This investment is on-going as the highway system deteriorates and cannot be made to last forever. A constant flow of money is required to properly and continually maintain, rehabilitate, reconstruct, and preserve the existing pavements while designing and constructing new pavements in areas of growth. Given the lack of appropriate funding and the size of the required investment, management of the pavements present a major and complex puzzle that challenges highway administrators and engineers. Only a high quality and comprehensive PMS can satisfy the need to manage pavements and direct the investment where needed to make proper decisions at the proper time (Baladi et al. 1992).

2.1.2 Pavement Management versus Managing Pavement

Managing pavement consists of the various acts of planning, designing, constructing, maintaining, evaluating, rehabilitating, and preserving the pavements. On the other hand, pavement management is the coordination of all the activities involved in the planning, financing, designing, constructing, maintaining, evaluating, and rehabilitating the pavement portion of a public works program. In other words, pavement management is essentially the comprehensive directing of the various actions involved in managing pavements. Further, the establishment of documented procedures regarding the comprehensive coordination of all the pavement management activities is referred to as a PMS (Dawson 2012).

2.1.3 Acts of Managing Pavements

In 1903, the creation of the Office of Road Inquiry of the Department of Agriculture marked the beginning of the Federal Government's involvement in highway activities on a continuing basis. In 1918, the office became the Bureau of Public Roads (BPR) of the Department of Agriculture. The Great Depression and President Roosevelt's "New Deal" pushed for the creation of massive public works projects; one focus was the creation of highways across the entire nation. Under the reorganization of the Federal Government of 1939, the BPR became the Public Roads Administration (PRA), which was then transferred to the Federal Works Agency. Another reorganization of the government in 1949 renamed the PRA to become the BPR again and transferred the bureau to the Department of Commerce. In 1967, the BPR was again transferred to the newly formed Department of Transportation. Finally, in 1970 the BPR was again renamed to its present designation, the Federal Highway Administration (FHWA) (Baladi et al. 1992, Dawson 2012, Nostrand 1992).

The Federal-Aid Highway Act of 1916 marked the beginning of the modern era of Federal-aid for highway activities by authorizing the expenditure of \$75 million for rural highway improvements. The enactment of the gasoline tax by various states in 1919 helped states secure matching funds for highway activities. In 1934, Congress authorized the expenditure by the states of up to 1.5 percent of the Federal-aid funds for highway planning and surveying. After World War II, the involvement of the government in the construction and improvement of the highway system increased substantially. The 1944 act provided approximately \$500 million per year for 3 years, dedicated to the improvement of the highway system. In 1952, an act directed the Secretary of Commerce to study highway financing and assess the condition of the highway system. As a result, in 1955, the Clay Committee (Advisory

Committee to President Eisenhower on a National Highway Program) reported the need for modernization of the highway system (Baladi et al. 1992).

The Clay Committee's report spurred the development of one of the most important acts in U.S. transportation history; the Federal-Aid Highway Act of 1956. In this act, proceeds from highway-user taxes were earmarked and placed into a newly created special trust fund, referred to as the "Highway Trust Fund." The 1956 act authorized the design and construction of the approximately 41,000-mile National Interstate and Defense Highway System to be completed within a period of 15 years and made funds available to the states on a 90-10 matching basis (Baladi et al. 1992, Dawson 2012).

One downfall of the Federal-Aid Highway Act of 1956 is that it did not include matching funds for the maintenance and rehabilitation of the highway systems. Therefore, the Federal-Aid Highway Act of 1976 introduced the 3R program; resurfacing, restoration, and rehabilitation. The program was financed using federal highway funds. Eventually, in 1981, congress added the fourth 'R' to the program; reconstruction. Finally, on May 21, 1980, the FHWA issued a Federal Highway Program Manual addition titled 'pavement management' that encouraged all states to strengthen their system of selecting pavement projects by developing a PMS (Nostrand 1992).

By the end of the 1980s more than half of the states were developing or implementing a PMS. In 1989, the FHWA issued a policy requiring all states to have a PMS that would cover principal arterials under the states' jurisdiction. The scope of federal and state involvement in PMS expanded when Congress passed the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991. This act required that all states have a PMS that covers all Federal-Aid highways. In December 1993, the FHWA issued a regulation regarding all management

systems. Section 500.207 contains the required components of a PMS for all Federal-Aid highways: data collection, data analyses, and update (Botelho 1994).

2.1.4 Network-Level Pavement Management

The pavement management process has two basic working levels, network and project levels. Each of these levels results in different sets of decisions through different levels of analysis of the collected data using a shared database. At the network-level, the primary goal of a PMS is to assess the overall condition of the pavement network and produce a prioritized work plan by examining various time and budget constraint scenarios. The products of network-level evaluation and analysis include:

1. Documenting the distribution of the pavement conditions along the network.
2. Assessing the impact of budget levels on the health of the pavement network.
3. Generating a pavement fix (reconstruction, rehabilitation and preservation) strategy to maximize the longevity of the pavement network and minimize costs. Such strategy identifies candidate pavement projects or sections based on pavement condition levels or brackets and it does not include the specific locations of the projects.
4. The main disadvantage of network-level pavement management is that the design models are simple and do not adequately consider the detailed factors that are required for project-level analysis (ASTM 2000, Fwa 2006, Reigle 2000).

2.1.5 Project-Level Pavement Management

At the project level, detailed consideration is given to alternative design, construction, maintenance, or rehabilitation activities for particular roadway sections or projects. In essence, a project-level approach utilizes data for individual pavement sections to determine the optimum maintenance and rehabilitation strategies for priority projects. Project-level pavement

management models are typically complex, dealing with technical concerns and requiring detailed information such as pavement structural thicknesses, time-series distress and roughness data, as well as deflection data among many other things. Thus, project-level pavement management requires specific detailed information so that cost-effective decisions can be made regarding pavement treatment alternatives (AASHTO 1993, Pavement Interactive 2008, Reigle 2000).

Nevertheless, the products of project-level analysis and evaluation include:

1. The selection of candidate pavement projects and their boundaries.
2. The selection of treatment types and time.
3. Alternative treatment designs and their associated costs.
4. The estimation of the needed future treatments and their associated costs (results of life cycle cost analysis)

2.1.6 Commonalities and Differences Between Network and Project-Level Pavement Management

Network and project-level pavement management differ in regards to the required amount of details. For instance, network-level management requires generalized models to assess the overall condition of the pavement network while project-level management requires specific distress and condition models to assess the exact condition of one pavement project. The management systems for the network and project-levels are mutually dependant, though. One of the many functions of the network-level PMS is to identify pavement sections within a network of pavements that require immediate maintenance or rehabilitation actions. In turn, project-level management may be responsible to determine the optimum pavement treatment time and types for each specific pavement section identified at the network-level. Thus, detailed data are used at

the project-level to develop specific pavement treatment designs for pavement sections that were identified or ‘flagged’ at the network-level. Hence, it can be stated that network-level pavement management provides an overall ‘picture’ of the pavement network while project-level management helps paint a detailed ‘picture’ of specific pavement sections. The two, though, share the same data system and are mutually dependant (Reigle 2000). Figure 2.1 displays the general functions of the two different management levels while linking them both to the same data system.

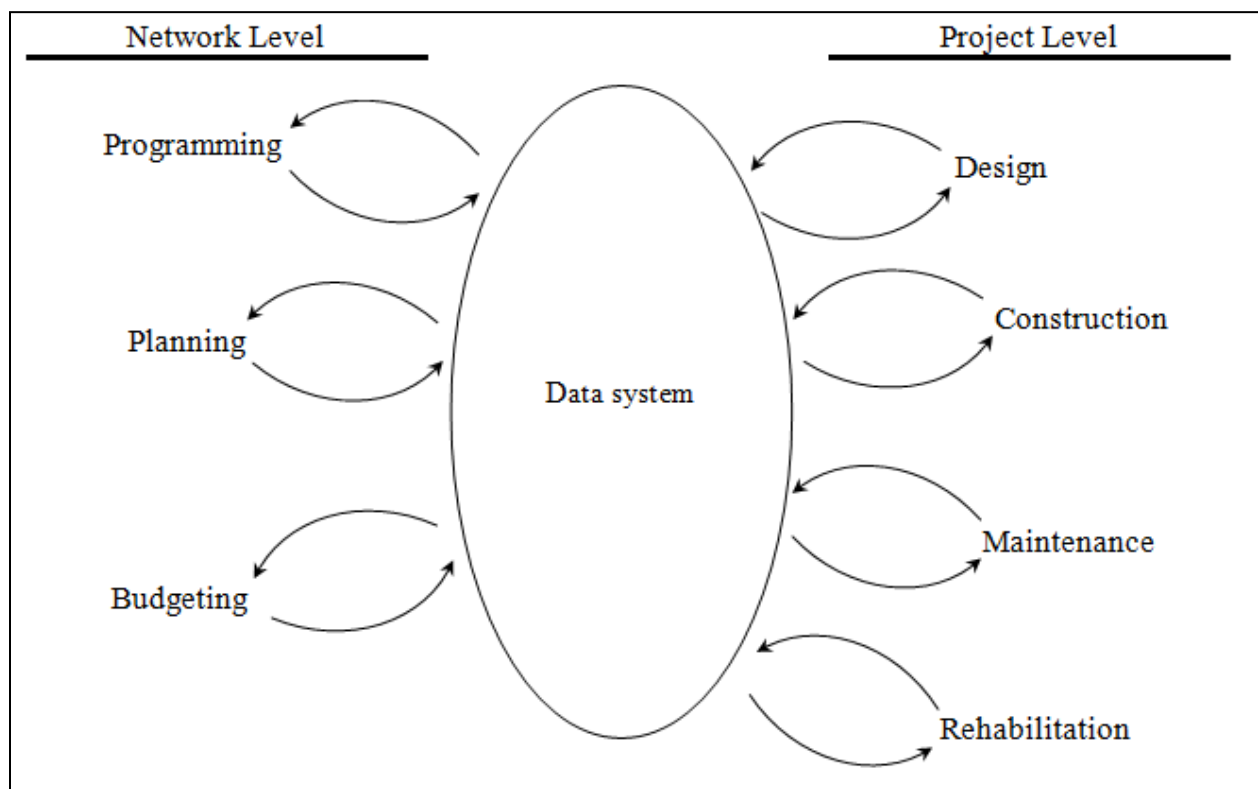


Figure 2.1 Network and project-levels pavement management (AASHTO 1993, Hudson et al. 1979)

It is the opinion of this author that the terms network- and project-level pavement management are confusing and do not express the true nature of PMS. The two terms should read data analysis at the network and project levels. To illustrate, assume that a State Highway Agency (SHA) collects and stores pavement data along each 0.1 mile pavement segment. The

data could be analyzed at the project-level to determine the pavement rate of deterioration, the pavement remaining service life (RSL) and so forth. Results of the analysis at the project-level could be summarized to study the distribution of the pavement conditions along the network. Such distribution could be used for planning, programming and budgeting analysis. This scenario would generate compatible and consistent network and project level analysis based on the same basic pavement data.

2.2 Data Elements and Data Collection Procedures

The collection and storage of pavement management data are the most costly parts of the implementation of a pavement management system (PMS) and are typically the most expensive parts of keeping a PMS operating. The various data elements of a successful PMS can be collected using costly or inexpensive methods. The resolution, accuracy, and detail of the collected data are highly dependent on the selected collection method (Smith et al. 1998). In recent years, the adoption of methods and technologies associated with improved pavement management data collection and analysis have improved the function of most PMSs (McGhee 2004).

Pavement condition surveys and related data collection are considered to be key components for any successful PMS (WSDOT 2012, Peterson 1987). The pavement condition surveys and related data collection include inventory data and condition/distress data. In general, inventory data describe the physical characteristics of the pavement and usually do not change with the time between rehabilitation and significant maintenance activities. The inventory data can be divided into four categories (Baladi et al. 1992):

1. As-built data including pavement type (flexible, rigid, composite), width of highway and shoulder, number of lanes, thickness of each pavement layer, type of asphalt or concrete

mix design, joint spacing, type of dowel and tie bars, temperature steel if any, drainage type (cross drain, edge drain, drainable layers, etc.), quality control and quality assurance data, incentive payments, if any, and so forth.

2. Historical data including years of last major activity (rehabilitation, construction, widening, etc.), materials used in last major activity, type of last improvements, type of rehabilitation/maintenance projects and their associated costs, etc..
3. Environmental data including detailed past climatologically-collected data, which could be obtained from the local office of the National Oceanic and Atmospheric Administration (NOAA).
4. Other data including specialized data requiring substantial details including material test type, material test method, material test procedure, and equipment.

Detailed inventory data are listed in Table 2.1.

In addition to inventory data, a comprehensive PMS database should also include monitoring data. It is desirable for the monitoring data to include (Baladi et al. 1992):

1. Pavement distress and roughness.
2. Surface friction.
3. Drainage conditions.
4. Deflection measurements.
5. Traffic and load information.

Detailed monitoring data are also listed in Table 2.1. Note that data elements presented in Table 2.1 represent the visions of the author as well as the visions of many staff of various SHAs.

Therefore, data in the table must be reviewed on a continuous basis and modified when the need arises.

Table 2.1 Pavement inventory and monitoring data

Data items	Example uses of data
<p><u>Inventory data</u></p> <ul style="list-style-type: none"> • Reference location • Pavement type • As constructed asphalt concrete layer <ul style="list-style-type: none"> ○ Layer thickness ○ Material properties ○ Asphalt mix data ○ Number and thicknesses of AC courses ○ QC and QA data ○ Cost per ton ○ Deviation from design • As constructed PCC layer <ul style="list-style-type: none"> ○ Layer thickness ○ Concrete properties ○ Concrete mix data ○ Joint spacing ○ Dowel and tie bars ○ QC and QA data ○ Cost per ton ○ Deviation from design • Base layer <ul style="list-style-type: none"> ○ Layer thickness ○ Material properties ○ QC and QA data ○ Cost per ton ○ Deviation from design • Subbase layer <ul style="list-style-type: none"> ○ Layer thickness ○ Material properties ○ QC and QA data ○ Cost per ton ○ Deviation from design • Roadbed soil <ul style="list-style-type: none"> ○ Soil classification ○ Roadbed modulus ○ Hydraulic conductivity ○ Cost of rolling per lane-mile 	<ul style="list-style-type: none"> • Locate pavement segments and link various data files • Deterioration models • Backcalculation of layer moduli and deterioration models • Deterioration models and quality control • Deterioration models and quality control • Deterioration models and quality control • Deterioration models • Life cycle cost • Link pavement design to management • Backcalculation of layer moduli and deterioration models • Deterioration models and quality control • Deterioration models and quality control • Deterioration models • Deterioration models • Life cycle cost • Link pavement design to management • Backcalculation of layer moduli and deterioration models • Deterioration models and quality control • Deterioration models and quality control • Life cycle cost • Link pavement design to management • Backcalculation of layer moduli and deterioration models • Deterioration models and quality control • Deterioration models and quality control • Life cycle cost • Link pavement design to management

Table 2.1 (cont'd)

Data items	Example uses of data
<p><u>Inventory data</u> (continued)</p> <ul style="list-style-type: none"> • Traffic <ul style="list-style-type: none"> ○ ADT ○ Mixed traffic volume and weight ○ ESAL • Environment <ul style="list-style-type: none"> ○ Daily low and high temperatures ○ Rain fall ○ Snow fall ○ Depth of frost penetration • Drainage type <ul style="list-style-type: none"> ○ Location ○ Drainage type ○ Outlets ○ Discharge 	<ul style="list-style-type: none"> • Capacity, safety and planning • Traffic spectrum for design purposes using the M-EPDG • 1993 AASHTO Design Guide and deterioration models • Causes of pavement distresses • Causes of pavement distresses • Causes of pavement distresses • Causes of pavement distresses and deterioration models • Rehabilitation and maintenance • Performance of drainage • Maintenance • Drainage quality
<p><u>Cracking data</u></p> <ul style="list-style-type: none"> • Alligator (bottom up) • Alligator (top-down) • Block • Blow-up • Corner break • Durability “D” • Longitudinal • Mat • Reactive aggregate • Reflective • Slippage • Transverse 	<ul style="list-style-type: none"> • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action

Table 2.1 (cont'd)

Data items	Example uses of data
<u>Joint data</u> <ul style="list-style-type: none"> • Construction joint in AC pavements • Transverse joints in PCC pavements • Longitudinal joints in PCC pavements • Joint spalling 	<ul style="list-style-type: none"> • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action • Performance models, RSL, and selection of rehabilitation action
<u>Other condition data</u> <ul style="list-style-type: none"> • Asphalt hardening/oxidation • Bleeding • Corrugation • Depression • Lane/shoulder drop off or heave • Lane/shoulder separation • Patching • Polished aggregate • Potholes • Pumping • Raveling • Scaling • Spalling • Segregation • Swell • Weathering/stripping 	<ul style="list-style-type: none"> • Performance models and selection of rehabilitation action • Safety problem • Performance models and selection of rehabilitation action • Performance models and selection of rehabilitation action • Safety problem • Safety problem • Performance models and selection of rehabilitation action • Safety problem • Safety and ride problem, effects of freeze-thaw cycles • Voids under slab, select the proper action • Safety and ride quality issues • Concrete mix and construction problems • Deterioration rates • Causes of cracks and potholes • Lower layer problems, selection of proper action • Presence of water under the asphalt mat
<u>Rehabilitation actions</u> <ul style="list-style-type: none"> • Type • Materials • Year of inception • Year of construction • Expected/design life • Location reference • Cost 	<ul style="list-style-type: none"> • Determine its service life and performance • Deterioration models • Determine the pavement conditions at inception • Determine the pavement prior and after construction • Identify the proper pavement project • Life cycle cost

Table 2.1 (cont'd)

Data items	Example uses of data
<u>Pavement preservation actions</u> <ul style="list-style-type: none"> • Type • Materials • Year of inception • Year of construction • Expected/design life • Location reference • Cost 	<ul style="list-style-type: none"> • Determine its service life and performance • Deterioration models • Determine the pavement conditions at inception • Determine the pavement prior and after construction • Link performance to design • Identify the proper pavement project/segment • Life cycle cost
<u>Pavement maintenance actions</u> <ul style="list-style-type: none"> • Type • Materials • Year of application • Expected/design life, when applicable • Location reference • Cost 	<ul style="list-style-type: none"> • Determine its service life and performance • Deterioration models • Determine the effects on the pavement conditions • Link performance to design • Identify the proper pavement project/segment • Life cycle cost
<u>Other data</u> <ul style="list-style-type: none"> • Date of data collection • Air and/or pavement temperature • Reference location* • Accident type, date and location • Material specifications • Special pavement project provisions 	<ul style="list-style-type: none"> • Safety versus pavement conditions • Impact of specifications on performance • Track the propagation of distress • Effects of the provisions on pavement performance • Seasonal impacts on the data • Effects of temperature on crack opening
<p>* It is desirable to include reference location for the beginning and ending of each condition along the pavement structure. Such data are essential to study the propagation of cracks and their transformation from low to medium and to high severity levels. Most semi-automated equipment that are equipped to videotape the pavement surface or to measure the longitudinal and transverse profiles are equipped with a GPS unit and are capable of recording positions every 6-inches along the pavement.</p>	

The methods used by SHAs to collect pavement condition data vary from manual (wind-shield survey), to semi-automated (videotaping and digitizing the data by reviewing the videotapes), and to automated (computer identification/recognition of the images of the pavement distress). Typically, the manual distress survey is conducted in different ways as stated below (Baladi et al. 2009, FHWA 1995).

1. Walking survey of 100 percent of the pavement surface in which all distress types, severities, and quantities are measured and recorded. Some map the distress whereas others do not.
2. Walking survey of samples of the pavement surface in which all distress types, severities, and quantities within the sampled areas are measured or estimated and recorded.
3. Driving survey in which distress types, severities, and quantities are estimated while driving on the shoulder at creep speed. Periodic shoulder stops are made at selected areas to measure the distress and check the estimates.
4. Driving survey in which some distress types, severities, and quantities are estimated while driving at the posted speed limit. Periodic shoulder stops are made to verify the estimates.
5. Driving survey during which the pavement is assigned a general categorical description (excellent, very good, good, fair, poor, or very poor) or sufficiency rating number without identifying individual distress types and their severities and extents. The survey is done at the posted speed limit.

In general, the results of the manual distress surveys can be questionable due to the impact of human subjectivity. Many aspects of the data collection process (such as time of day, training, experience, fatigue, weather, and supervision of the raters) affect the results. For this

reason, the state-of-the-practice in data collection has shifted towards semi- and fully-automated techniques (FHWA 1995, Freeman and Ragsdale 2003, Smith and Change-Albitres 2007). The semi-automated data collection technique consists of obtaining analog distress images on film or high resolution video. Sometime later, trained observers view the images and identify and digitize the type, severity, and extent or quantity of each individual distress (FHWA 1995). In the fully-automated technique, images are collected on film or high resolution video and image analyzers identify and digitize the type, severity, and extent or quantity of each individual distress in real-time or at a later time in the office (Freeman and Ragsdale 2003).

In the past few years, the National Center for Pavement Perseveration (NCPPI) conducted a survey of 43 SHAs in order to assess their PMS practices, including data collection practices. The results of the survey indicated that, as expected, data collection methods vary from manual, to semi-automated, and to automated. This can be seen in Figure 2.2. Approximately 35 percent of the SHAs surveyed utilize automated data collection means while 42 percent use both automated and wind shield (manual) collection techniques (Dawson 2012, Galehouse 2010).

In an effort to improve the reliability of the collected pavement condition data, there has been a recent move for all SHAs to utilize fully automated data collection techniques. In this manner, as it was mentioned previously, an imaging system for cracking survey would be fully automated and the data digitized in real-time. In addition, laser sensors are used to measure to measure the pavement longitudinal and transverse profiles and to calculate the pavement rut depths and roughness such as the International Roughness Index (IRI) (Wang and Gong 2002). Data image processing techniques are used to identify and analyze the distress type, severity, and quantity. An example of this is shown below in Figure 2.3.

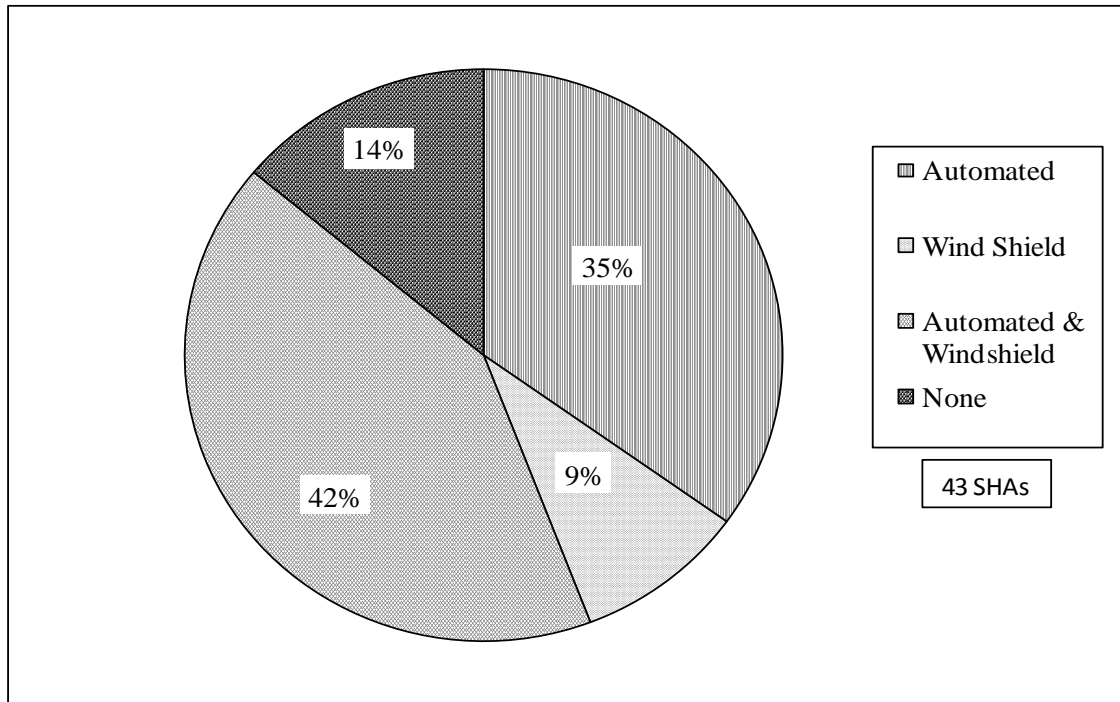


Figure 2.2 Percent of SHAs using the stated method of pavement condition data collection

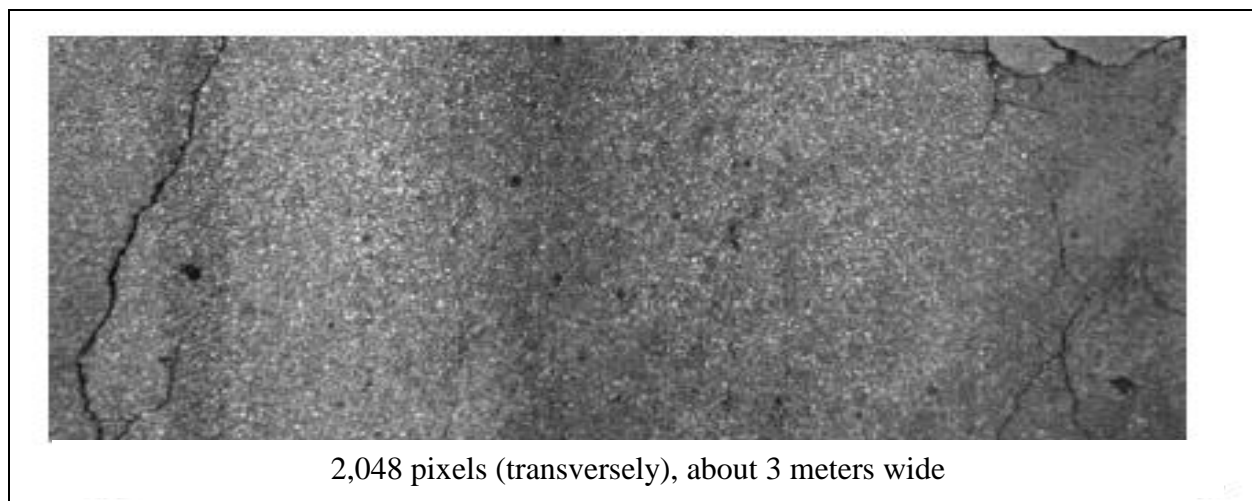


Figure 2.3 Area scan image for fully-automated data collection (McGhee 2004)

A significant advantage of the fully-automated technique is that it removes the subjective distress identification and rating from the various pavement management personnel. Thus, one would expect the data to be more reliable and consistent than manual and semi-automated condition data. However, although a fully-automated technique is optimal for PMS data collection, significant improvement needs to occur such that crack recognition becomes more reliable. Currently, ‘tight’ cracks may be missed due to minimal crack width. Further, especially in rigid pavements, it is possible that more distress will be recorded than is actually exists due to the presence of the sawn joints (Wang and Smadi 2011). Thus, fully-automated data collection and distress identification is the optimal procedure for SHAs but significant gains in technology need to be made before this can be fully-implemented with confidence.

2.3 Data Collection Frequency

The goals and objectives of a particular SHA dictate the type and details of the pavement distress data to be collected, as well as the method and frequency of data collection. In general, the frequency of pavement distress data collection can be divided into three categories (Haider et al. 2007):

1. Short frequency (1 to 2 years); the data are typically used to develop or calibrate pavement performance models.
2. Long frequency (3 to 4 years); the data are typically used for assessing the general condition of the pavement network.
3. Localized monitoring; the data are typically used to address unexpected pavement performance.

In general, pavement conditions are monitored by most SHAs at 1-, 2-, or 3-year frequencies (McGhee 2004). An NCPP survey of 43 SHAs found that approximately 63 percent

of the surveyed agencies use a PMS data collection frequency of 1 year whereas 28 percent of the agencies collect data at a 2 year frequency as shown in Figure 2.4 (Galehouse 2010). It should be noted that some SHAs collect sensor-based data (roughness and rut depth) more frequently than image-based (cracking) data. However, it has been shown that, to make accurate pavement decisions; image based condition data (cracking) should be collected every year and sensor data could be collected less frequently (Haider et al. 2010).

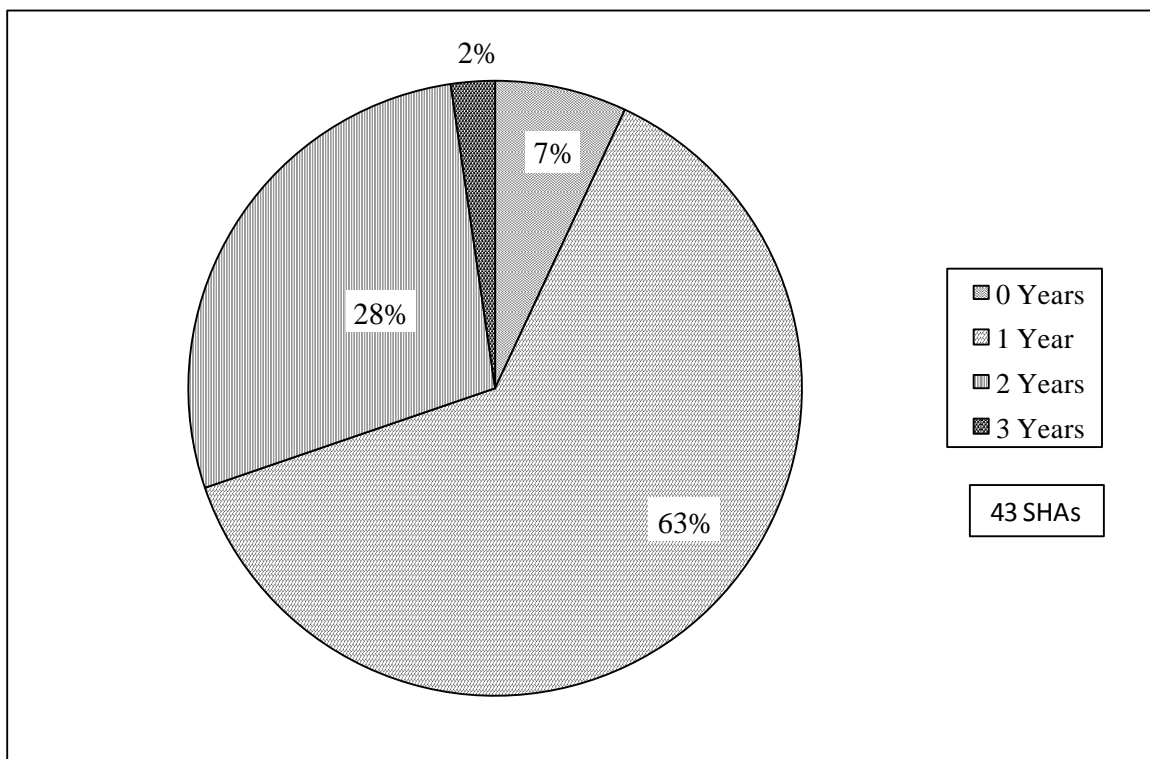


Figure 2.4 Percent of SHAs using the indicated PMS data collection frequency

2.4 Pavement Project Boundaries (Data Delineation)

When selecting pavement projects for treatments, especially those that are of great significance, pavement monitoring activities are undertaken to obtain measurements which assess various pavement response variables. These variables include pavement deflection, serviceability, rut depth, and friction number among many others. In general, the measured response variable indicates changes from one location to another with some points experiencing changes of major magnitude. At the locations of significant changes in the response variable, the overall performance of the pavement segments on either side of the locations will be noticeably different. Hence, it is likely that the locations of these significant changes define the pavement project boundaries. Figure 2.5 illustrates a typical plot of a response variable as a function of distance along a highway segment. Four moderate to significant changes in the response variable can be seen at certain locations. Hence, four separate pavement treatments may be warranted (i.e., four separate overlay design thicknesses) (AASHTO 1993).

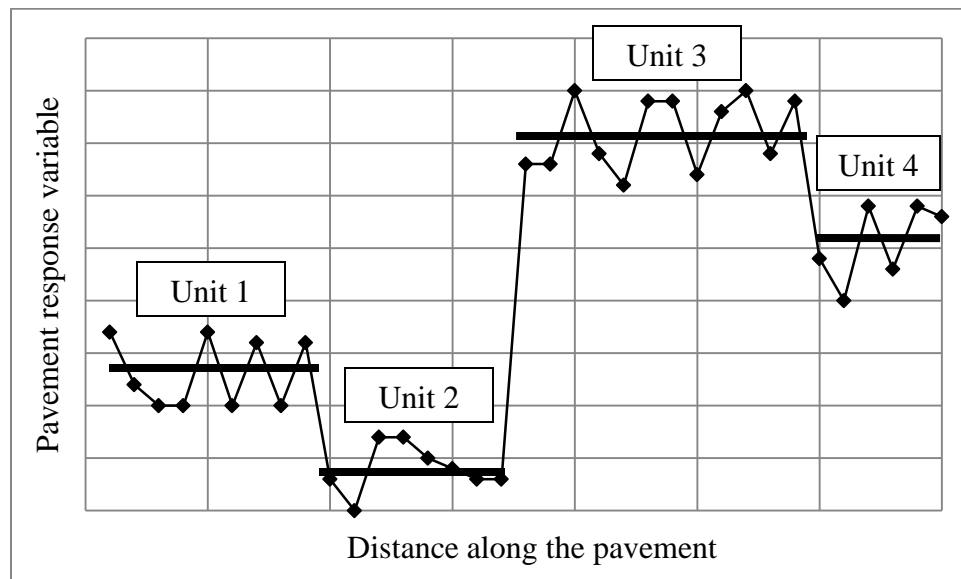


Figure 2.5 A typical pavement response variable versus distance for data delineation (after AASHTO 1993)

Two data delineation methods exist to help SHA employees identify the locations of significant changes in the response vehicle and hence, locate the pavement project boundaries. These methods (sometimes referred to as methods of unit delineation) are called the Idealized Approach and the Measured Pavement Response Approach.

For the Idealized Approach, the engineer should isolate each unique factor influencing potential pavement performance. These factors include:

1. Pavement type.
2. Construction history (including rehabilitation and major maintenance).
3. Pavement cross section (layer material type/thickness).
4. Subgrade (foundation).
5. Traffic.
6. Pavement Condition.

Under ideal circumstances, the historic pavement database is used to evaluate these factors. The analysis units, or the project boundaries, are determined by a unique combination of each of the pavement performance factors. An example of such method is shown in Figure 2.6. If accurate records have been kept, this idealized approach based on historical data has more merit in delineating the data than a procedure which relies on performance indicators. This is because changes in one or more design factors (which indicate points of delineation) are not always evident through observation (AASHTO 1993).

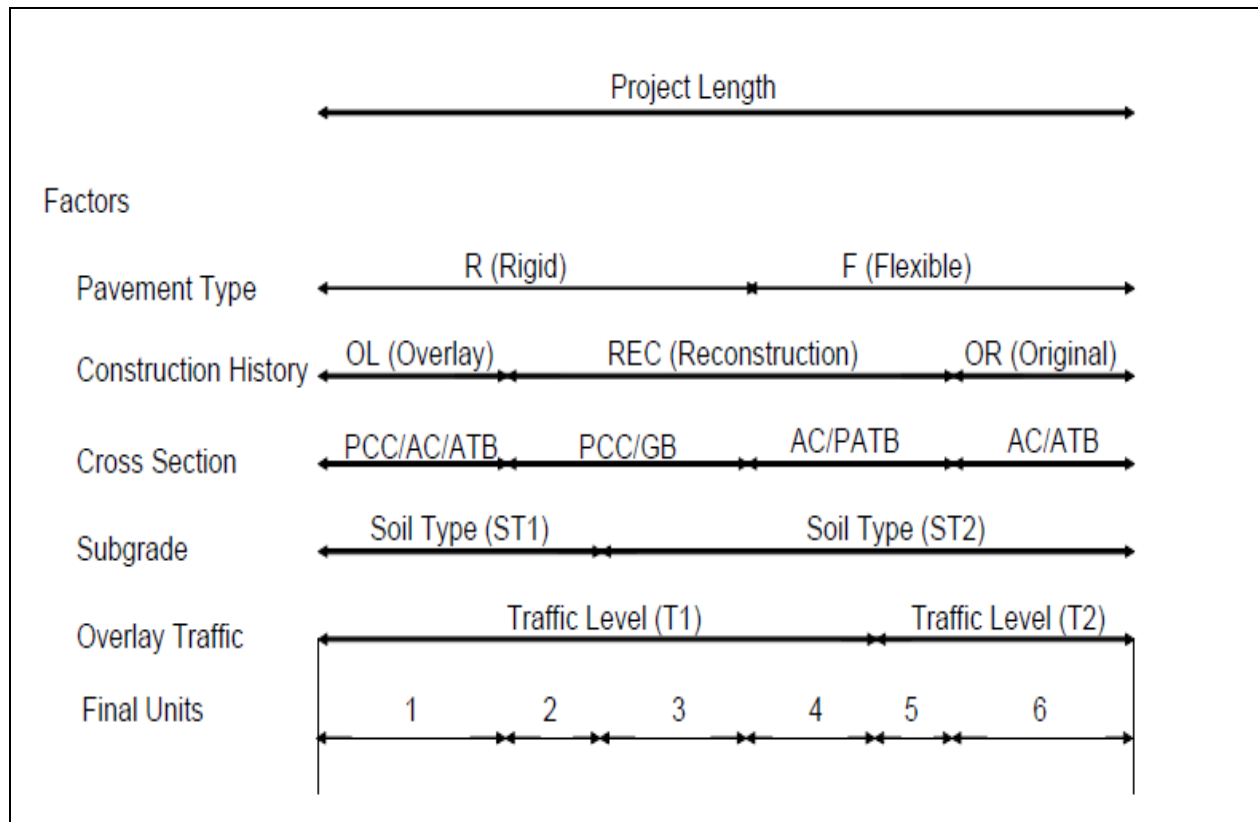


Figure 2.6 The Idealized Approach for data delineation (AASHTO 1993)

The Measured Pavement Response Approach is a favored alternative to the Idealized Approach because one may not be able to accurately determine the practical extent of the factors influencing pavement performance described previously and must rely upon the analysis of a measured pavement response variable (i.e. deflection) for unit delineation. The SHA engineer should first develop a plot of the measured pavement response variable as a function of the distance along the project, as shown in Figure 2.5. Once the plot of a pavement response variable has been completed, it could be used to delineate the data using several techniques. The simplest of these techniques is visual examination to determine where relatively unique similarity between data points occurs.

In addition, several analytical techniques are available for data delineation; the AASHTO 1993 recommended technique is the cumulative difference, which is readily programmable in computers. This method was used in this study and it is detailed in Chapter 4 of this thesis. The cumulative difference method removes the subjectivity of data delineation through mathematical analysis. Nonetheless, other data delineation methods that exist include grouping pavement segments with similar conditions and/or rates of deterioration, and selecting boundaries based on a visual inspection without quantifying the conditions of the pavement section.

2.5 Pavement Performance Data Sampling

In recent years, administrators of various SHAs have been asked to reduce the costs of pavement distress data collection and decrease the time period between data collection and data digitization. It is thought by many that the reduction in costs and time could be easily accomplished by sampling the pavement distress data. For some SHAs, the term “sampling” implies continuously videotaping the pavement surface conditions along the entire pavement network and digitizing the data using a sampling technique. That is, digitize the data along a short pavement segment (such as 100, 200, 300 feet, or 0.1 mile) of each one mile and assume the digitized data represent the pavement conditions along the entire mile. For other SHAs, “sampling” implies surveying short pavement segments (e.g., wind shield survey or walking survey) and assigning the results to a longer pavement section. Regardless of the sampling technique or method used, it is evident that sampling reduces the direct costs and time for data digitization. However, the short- and long-term costs and benefits of sampling cannot be determined unless the impacts of sampling on the accuracy of the pavement decision processes are analyzed.

The collection and evaluation of the time series pavement condition data is an essential part of the selection of cost-effective pavement preservation strategy and, in general, substantially influence the entire pavement decision process (Cafisco et al. 2002). The pavement condition data collection practice is typically limited by the available resources, practicality, and other constraints within SHAs. The National Cooperative Highway Research Program (NCHRP) synthesis of highway practice 222 states that some of the practical constraints can be attributed to the size and variability of the conditions of the pavement network. The accuracy of the data is limited by the frequency of the data collection and by the representation of the selected samples (Zimmerman 1995). Sampling of pavement condition data requires less time and money than continuous data collection. However, the question becomes could sampling deliver accurate pavement condition prediction without excessive amounts of data (Robertson et al. 2004)? In the NCHRP 2004 synthesis of highway practice 334, 42 states, the District of Columbia, 2 FHWA offices, 10 Canadian provinces and territories, and Transport Canada (airfields) were surveyed regarding pavement condition data collection (McGhee 2004). It was found that:

1. Most agencies use an automated means for data collection along the entire outer traffic lane every other year.
2. For pavement cracking:
 - a) Nine agencies survey 100% of the lane to be evaluated.
 - b) Three agencies collect cracking data along samples of varying length.
 - c) Five agencies sample 10% to 30% of the roadway using a random sampling technique.

- d) Others reported that they videotape 100% of the survey lane but, for each one mile, they digitize the data along 50 to 1,000 feet segments. The data along these segments are assumed to represent the pavement condition along the entire mile.
3. For pavement roughness:
- a) Many agencies collect the data along the entire surveying lane and they report the data for each 0.1-mile interval.
 - b) The Canadian provinces report the roughness data at 50-meter to 100-meter intervals.
 - c) In the District of Columbia, the reporting intervals are measured by one city block.
 - d) The State of Arizona uses a reporting interval of 1 mile.

In another report, the Ministère des Transports du Québec reports the use of sampling of a 30-meter pavement section for every 100-meter pavement segment (30 percent sampling) (Tremblay et al. 2004). The above information indicates that the sampling and reporting intervals vary greatly from one road agency to another, with the intent to properly assess the current pavement condition and reduce the time and cost of the data collection.

2.6 Pavement Survey and Analysis Length

As it was stated in the previous section, the collection and analyses of pavement condition data are essential steps to support decisions regarding the selection of cost-effective pavement treatments (Cafisco et al. 2002). Some SHAs collect, store, and analyze pavement condition data along 0.1 mile pavement survey or segment lengths while others use different survey lengths. In reality, the pavement data collection practice is limited by resources, practicality, and economical constraints within the SHAs. The NCHRP synthesis of highway

practice 334 indicates that data collection practices varies greatly from one highway agency to another. For example, some collect, store, and analyze the data for each 0.1 mile pavement segment while others use different lengths (McGhee 2004). As a result, the data storage needs vary significantly. For example, a SHA that collects, stores, and analyzes pavement condition data for each 0.1 mile pavement segment requires ten times the data storage capacity compared to another agency that uses 1 mile intervals. The consequences of reporting the data using longer pavement segments include the loss of detail of the pavement surface conditions along the road. In more general terms, data that are collected, reported, and analyzed along longer pavement segments may lack the necessary details to accurately understand the variability and distribution of the pavement conditions along a given pavement section. In addition, the pavement survey or analysis length restricts the determination of project boundaries and data delineation. To illustrate, data stored along each 0.1 mile pavement segment would lead to project boundaries for project length within 0.1 mile distance (e.g., 6.3, 7.1, or 1.5 miles). On the other hand, the project boundaries based on data stored and analyzed along each 1 mile pavement segment are restricted to whole mile numbers such as 6, 8, or 3 miles.

Furthermore, the variability of pavement condition data along short survey and analysis lengths complicates the modeling of pavement conditions over time and the assessment of the overall pavement performance. Recently, some engineers and highway administrators have suggested that increasing the pavement survey and analysis lengths over which the conditions are analyzed may decrease the data variability and simplify the modeling of pavement conditions over time (Fillastre 2011). The thought behind this is that longer analysis lengths decrease data variability and simplify the modeling of pavement condition leading to improved PMS decisions

and proper management of the pavement network. It is shown in Chapter 4 that the pavement analysis length does not have any impact on the accuracy of the PMS decision or operation.

2.7 Data Imputation

Some SHAs collect the time-series pavement condition data on a bi-yearly basis and store the data for each 0.1 mile pavement segment. For some 0.1 mile pavement segments along a network, a data collection cycle for certain years may be skipped or some of the distress or condition data are missing or not reported. Such scenarios are due to various reasons including equipment malfunction, failure to properly record the data, loss of data due to human error, delays in issuing the data collection contract, construction zones, outliers, and so forth. Hence, for certain years, some time series data could be missing from the database. Some of the missing data elements could be crucial to commence data analysis. As a result, some of the SHAs staff or the PMS manager or engineers may impute the missing data element so that analysis of the time-series condition data could be improved leading to better and more accurate PMS decisions (Bennett 2004, Farhan and Fwa 2012, Flintsch and McGhee 2009, Zhang and Smadi 2009).

Data imputation is the practice of ‘filling in’ missing data with plausible values. Data imputation is an attractive approach to analyzing incomplete data sets. When completed properly, imputation solves the missing data problem. However, a naïve or unprincipled imputation methods could create more problems than they solve by distorting estimates and increasing the standard errors (Little and Rubin 1987). Currently, though, the principles of statistical quality assurance in terms of imputation of missing data are well developed but their application and performance to the imputation of pavement management data is unclear (Farhan and Fwa 2012).

Before one can accurately impute the missing data, it is important to understand why the data are missing. It was stated that missing data elements can informally be thought of as being caused by some combination of random processes, processes which are measured, and processes which are not measured (Graham et al. 2003, Wayman 2003). More specifically, the missing data mechanisms can be placed in one of the following three categories (Little and Rubin 1987, Wayman 2003):

1. Missing completely at random (MCAR) – The MCAR data cases are no different than non-missing cases, in terms of the analysis being performed. These cases can be thought of as randomly missing from the data and the only real penalty in failing to account for missing data is loss of statistical significance or power.
2. Missing at random (MAR) – The MAR data depends on known values and is described fully by variables observed in the data set. Thus, accounting for the values which ‘cause’ the missing data will produce unbiased results in the analysis.
3. Missing not at random (MNAR) – The MNAR data can be missing in an unmeasured fashion that is often termed ‘non-ignorable.’ The missing data depends on events or items for which the person analyzing the data cannot quantify. Thus, this is the most damaging mechanism for missing data.

Depending on the missing data mechanism, different data imputation techniques could be used to estimate the values of the missing data. In general, all data imputation techniques can be considered as either ‘single’ or ‘multiple’ imputation. For single imputation, a definite value is substituted for a missing value based on an established procedure. After such substitution, standard statistical methods for complete data analysis can be used on the entire data set. The single imputation method is widely used because of its simplicity and efficiency. This method

does not require the user to have extensive statistical knowledge (Zhang et al. 2008). However, it has been shown that single imputation can reduce the variance of the variable in question, creating a bias in the data. Nonetheless, the most common single imputation techniques are summarized below (Farhan and Fwa 2012, Little and Rubin 1987, and Zhang et al. 2008).

1. Mean substitution – In this approach, the missing physical values are imputed using the mean value of a data set over time. It has been shown that this approach adds no new information since the overall mean will remain constant and the variance of the data set will be decreased proportionally to the number of missing data. Furthermore, given that pavement distress values increase over time, imputation by mean substitution would result in considerable errors. Regardless, Equation 2.1 can be used for mean substitution.

$$\bar{x} = \frac{1}{n} * \sum_{i=1}^n x_i \quad \text{Equation 2.1}$$

2. Linear interpolation/extrapolation – The missing data are computed by interpolating between, or extrapolating from the adjacent, available data points. Graphically, this amounts to substituting missing data by connecting a straight line to the two adjacent data points near the missing data. This method assumes a linear correlation in the data.

Equation 2.2 can be used for linear interpolation and extrapolation.

$$y_2 = \frac{(x_2 - x_1) * (y_3 - y_1)}{(x_3 - x_1)} + y_1 \quad \text{Equation 2.2}$$

3. Regression substitution – The missing data are imputed using the proper regression line based on available information. For example, missing IRI data could be imputed by modeling the available data as a function of time using an exponential function, missing

rut depth data could be imputed by modeling the available data as a function of time using a power function, and cracking based on a logistic or s-shaped curve.

Although single imputation is often preferred because it does not require significant statistical knowledge and is relatively simple, the decrease in variance of the complete data set as a result of the imputed values poses problems for some researchers and SHA employees. To solve this problem, the multiple imputation method is recommended. Multiple imputations can be accomplished using the following steps:

1. Missing values for any variable are predicted using existing values from other variables.
2. The predicted values are substituted for the missing values to complete the data set.
3. This process is performed multiple times, producing multiple imputed data sets.
4. Standard statistical analysis is carried out on each imputed data set, producing multiple analysis results.
5. Results of the statistical analyses are then combined to produce one overall analysis.

Essentially, the multiple imputation technique is used to produce plausible versions of the substituted values, and thus, different plausible versions of how the data may appear in the population. Averaging over these versions could result in more accurate conclusions (Wayman 2003). However multiple imputation techniques require significant statistical knowledge and can be hard to implement.

In a research study published recently, Farhan and Fwa (2012) utilized a highway pavement condition survey database consisting of 29 years of data to assess the performance of various imputation approaches in estimating missing distress data. The distress data that were analyzed include alligator, longitudinal, and transverse cracking as well as IRI and rut depth. To assess each imputation method, data points were artificially removed from the pavement

performance database. Each triplet was intended to represent three consecutive time periods, or data collection cycles. Hence, for each distress type, 12 of the available 29 data points were removed from the database. Using four different imputation techniques (mean substitution, linear interpolation, regression substitution, and multiple imputations), the imputed data were compared to the measured data using mean percent absolute error. Few results of the study are summarized in Figure 2.7. As it can be seen in the figure, the mean substitution method resulted in the highest error, followed by the regression substitution method and the interpolation method. The multiple imputation method proposed in the study yielded the smallest errors for all distress types (Farhan and Fwa 2012). However, besides the rut depth data, the differences in the error between multiple imputation and regression substitution or interpolation are not significant. Given that multiple imputation requires substantial statistical background, it is the opinion of the author of this thesis that regression substitution or interpolation would suffice.

2.8 Location Referencing Systems

One of the basic requirements of a successful PMS is to have an efficient pavement condition data collection program to help facilitate accurate pavement management decisions (Farhan and Fwa 2012). In fact, any recommendations or decisions regarding pavement treatments must be based on the analysis of quantifiable data including the pavement condition data (Holt and Gramling 1991). The procedure for which SHAs store the pavement condition data has evolved over time, however one issue persists: the accuracy of the location reference systems (Flintsch et al. 2004). The use of an accurate location reference system is essential for pavement management engineers to evaluate the proper time-series pavement condition data and to formulate correct and cost-effective pavement management decisions.

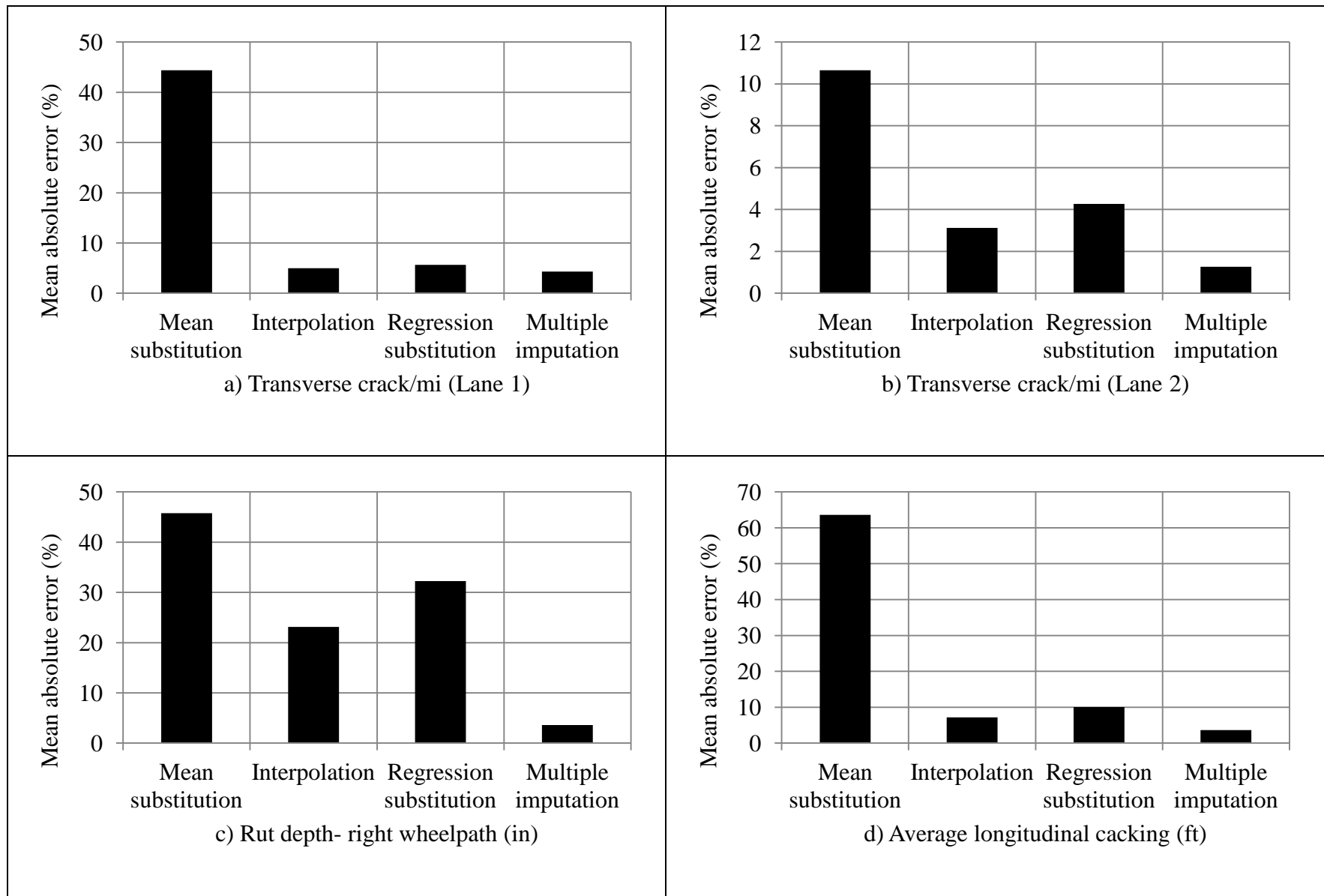


Figure 2.7 Mean absolute error for four imputation methods (after Farhan and Fwa 2012)

A location referencing system is formally defined as “an automated system used to manage the collection and storage of spatial information at exact locations” (Iowa DOT 1998). The proper use of a location referencing system greatly impacts the utility of the pavement condition data collected and stored in the PMS database (Deighton and Blake 1993). In general, three types of location referencing systems exist and are used in the PMS databases:

1. Linear location referencing system
2. Global positioning system (GPS)
3. Geographical information system (GIS)

It should be noted that a GIS is considered to be a relational database that includes the location reference system while the linear location referencing and GPS are strictly location referencing systems. Figure 2.8 displays the results of a survey of 46 SHAs regarding the types of location referencing systems used to support pavement data collection and storage activities (Flintsch et al. 2004). The data in the figure imply that some SHAs use dual or triple location referencing systems.

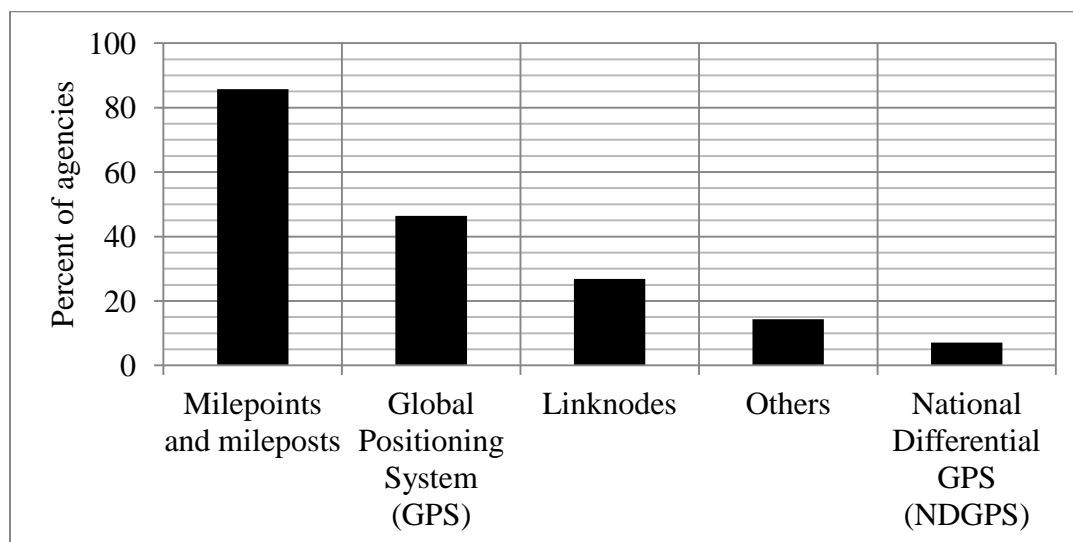


Figure 2.8 Percent of SHAs utilizing stated location referencing systems (after Flintsch et al. 2004)

Indeed, traditionally, some SHAs reference the pavement condition data using various linear location referencing techniques. These include link/node locations, reference points or offsets, and most commonly route names and mileposts or milepoints. For the route name and milepost referencing method, each roadway is identified using a unique name and the distance along the route is specified from a particular origin. The distance units are either marked with signs along the route to mark data collection points in the field (Flintsch et al. 2004) or not marked and the distance from the origin is estimated using the car odometer. The main advantage of a linear referencing system is that it is well-developed and understood by most of the agency staff. One significant shortcoming is that mileposts lack accuracy and hence, the same point on the pavement structure may move substantially from one data collection cycle to the next due to various reasons including realignment, rehabilitation, or other pavement activities (Flintsch et al. 2004). This effectively changes the size and location of individual pavement sections and negatively impacts the analysis of time-series condition data.

The second most common location referencing system used by SHAs is the GPS. The GPS is ‘a satellite-based navigation system made up of a network of 24 satellites placed into orbit by the U.S. Department of Defense. GPS satellites circle the earth twice a day in a very precise orbit and transmit signal information to earth. The GPS receivers take this information and use triangulation to calculate the user’s exact location’ (Garmin 2012). Thus, GPS holds a significant advantage over linear referencing systems in that it is considered to be accurate to within approximately 50 feet (Garmin 2012). Furthermore, the use of GPS as a location reference system allows for more automation and less user input in the field. The differential cost to outfit pavement management users with GPS units is affordable and data collection with GPS will result in time and costs savings as well as more accurate data (McNerney and Rioux 2000).

The fact that GPS technology is a relatively new component of a PMS, though, poses a significant problem. The past historical pavement condition data of a particular SHA were collected and stored using a certain location referencing system (likely a linear system). Hence, the location reference system must be converted properly to integrate the data with the GPS database. This conversion is almost always complicated and can create significant discrepancies in the time-series data. The shortcomings of the GPS include potential problems with the signal due to heavy foliage and/or buildings blocking satellite signals as well as general equipment malfunction (Flintsch et al. 2004). Further, the agency staff must be trained to use the GPS properly.

The least-commonly used reference location system is the GIS. A GIS is ‘a computerized database management system for accumulating, storage, retrieval, analysis, and display of spatial data’ (Elhadi 2009). In a GIS, textual databases can be combined with digitized maps to enable the display of various data on a map. A GIS has minimal location referencing error due to its ability to collate and cross-reference many types of data by location (Foote and Huebner 1995). Besides unparalleled accuracy, the advantages of a GIS include the ability to visually display the results of pavement management analyses on a map of the highway network, view network conditions through dynamic color-coding, and access particular pavement section data through the map interface (Elhadi 2009). Hence it is clear that GIS is a powerful tool that can be used significantly in pavement management. However, the current structure of most PMSs adversely influences the use of GIS. The staff of the agency must be subjected to training before a GIS can be fully implemented. Finally, the integration of historically collected pavement condition data in the GIS database may be a difficult task and in some scenarios may not be possible.

For the benefits of the readers, the main advantages and disadvantages of each of the previously described location referencing systems are summarized in Table 2.2.

Table 2.2 Main advantage(s) and disadvantage(s) of each location reference system

Location reference system	Main advantage(s)	Main disadvantage(s)
Linear	Ease-of-use and typically physically visible roadside markers	Lack of accuracy, especially from year-to-year
GPS	Improved accuracy and automated input	Integration of data with historical linearly-referenced data and need for technological knowledge
GIS	Minimal error and improved integration of PMS elements in 'visual' database	Cost, integration of data with historically linearly referenced data, and technological knowledge

Case Study - Louisiana Department of Transportation and Development (LADOTD)

As previously mentioned, the most common location referencing system in use by SHAs is a linear system consisting of route names and mile posts or milepoints. This location referencing method lacks accuracy and adversely impacts the collection and organization of time-series pavement condition data. Essentially, the size and location of pavement segments can change from one year to the next as shown in Figure 2.9 (Dawson 2012).

In an effort to quantify the error of a linear location reference system, Dawson utilized pavement condition data from the Louisiana Department of Transportation and Development (LADOTD). The quantification of error was accomplished by comparing the Beginning Mile Points (BMPs) and recorded GPS data of 37,843 0.1 mile long pavement segments using the Pythagorean Theorem. It was found that only 36.8 percent of 0.1 mile pavement segments have error ranging from 0 to ± 50 feet. Furthermore, the error in the linear referencing system is as much as approximately ± 530 feet. This means that the pavement condition data recorded for a

0.1 mile long pavement segment one year may be recorded for an entirely different 0.1 mile long pavement segment the next year. Detailed results of the error calculations are listed in Table 2.3.



Figure 2.9 Illustrative example of yearly error of linear referencing system (Dawson 2012)

Table 2.3 LADOTD linear referencing system error (after Dawson 2012)

Range of distance between the assigned BMPs (feet)	The number of pavement segments from the LADOTD included in the analyses and the percentage within each range
	Total population of 0.1 mile long pavement segments is 37,843
$\pm \leq 10$	15.0 %
± 11 to 50	21.8 %
± 51 to 100	26.7 %
± 101 to 250	23.6 %
± 251 to 528	8.4 %
$\pm > 529$	4.5 %

The analyses results listed in Table 2.3 highlight the main disadvantage of the linear referencing system in that it may provide inconsistent location references from one year to the next. However, despite the possibility for significant error, the linear system is the most commonly used system amongst SHAs. This is no surprise because, as noted earlier, the system has been established and used for a long period of time. In recent years though, some SHAs have begun to shift towards the use of GPS to improve accuracy in location referencing. However the problem still exists to integrate the historical linearly referenced data and the new GPS data. In order to solve this problem, several agencies have started using dual location referencing systems (linear and GPS) as a first step towards total conversion to GPS.

2.9 Pavement Performance Data Modeling and Remaining Service Life (RSL)

The importance of pavement performance data modeling and the concept of remaining service life are discussed in the following sections.

2.9.1 The Need for Data Modeling

As it was mentioned previously, cost-effective pavement management recommendations and/or decisions must be based on the analysis of quantifiable data, including the pavement condition data (Holt and Gramling 1991). The condition data of any pavement segment include two sets of information, the condition at a given time (which can be determined using the collected pavement condition data from one data collection cycle) and the pavement rate of deterioration over time (which can be determined from a minimum of three cycles of pavement condition data). Thus, information about past and current pavement condition are often needed to assess the rate of deterioration of a given pavement segment and to predict its future conditions. Different pavement segments may have the same condition in one year but their rates of deterioration could be significantly different. Therefore, in order to assess pavement

performance over time it is necessary to model the data using the appropriate mathematical functions. In other terms, curves are fitted through the various measures of pavement condition to show past and predicted pavement performance (AASHTO 2001).

2.9.2 International Roughness Index (IRI) Modeling

One of the pavement surface conditions is pavement roughness or smoothness measures. The Highway Performance Monitoring System (HPMS) of the Federal Highway Administration (FHWA 2005) requires the states to provide consistent, comparable, realistic, universal and practical pavement surface roughness measurements. These measurements of road roughness are referred to as the International Roughness Index (IRI). Lower IRI values indicate a smoother pavement surface and better ride quality. Typically, the IRI is expressed in terms of inch/mile, meter/kilometer, or millimeter/meter (Lenz 2011). In general, the IRI increases exponentially over time (as the pavement ages, it develops cracks, and becomes uneven) (Dawson 2012). Therefore, pavement roughness (IRI) data are typically modeled as a function of time using an exponential function (Smith et al. 2002).

2.9.3 Rut Depth Modeling

Rut depth, or rutting, is a load related distress and is the result of the accumulation of plastic (permanent) deformations in the pavement layers along the wheel paths of flexible and composite pavements. The rutting in composite pavements are typically confined to the HMA overlay, whereas rutting in flexible pavements could be the accumulation of rut in all layers and the roadbed soil. Rigid pavements are not susceptible to rutting caused by plastic deformation or lateral movement of materials due to traffic loading. Rigid pavement rutting is typically caused by mechanical dislodging due to studded tire wear (Dean et al. 2011, Naiel 2010). Pavement rutting is a functional and structural distress that could lead to hazardous conditions when the rut

channel is filled with water causing hydroplaning. Rutting typically occurs early in the service life of asphalt and/or composite pavements and its progression decreases over time as the pavement materials become denser due to loading. Hence, a power function is typically used to model rut depth data as a function of time (Odermatt et al. 1999).

2.9.4 Crack Modeling

The most common pavement distress, regardless of pavement type, is cracking. Various types of cracking exist including fatigue (alligator), longitudinal, and transverse cracking (FHWA 2003). Pavement cracking could be caused by insufficient pavement thickness, poor drainage, low material strength, unexpected heavy traffic loads, and/or aging among many other things. In any case, a cracked pavement should be investigated to determine the root cause of cracking. Regardless of the cracking type (i.e. alligator, longitudinal, and transverse cracking), the propagation and deterioration of cracks typically follows three stages. In the first stage, few cracks appear in the early pavement service life. In the second stage, new cracks appear and old cracks propagate and deteriorate rapidly. In the third stage, the pavement reaches crack saturation status and the number of new cracks decreases significantly. Thus a logistic, S-shaped curve is typically used to model cracking data as a function of time (Dawson 2012, Yang 2004). In most cases, though, not enough time-series data points are available to model the s-shaped function. In such scenario, early cracking data can be modeled using the exponential function whereas for older pavement sections, the data could be modeled using a power function.

The appropriate mathematical models for roughness, rut depth, and cracking are summarized in Table 2.4 and shown in Figure 2.10.

Table 2.4 Mathematical functions used to model the data of the indicated distress type

Pavement distress type and units*	Model form	Generic equation
IRI (inch/mile)	Exponential	$IRI = \alpha \exp^{t\beta}$
Rut depth (inch)	Power	$Rut = \gamma t^{\omega}$
Cracking (feet, feet ² , or number of cracks)	Logistic (S-shaped)	$Crack = \frac{Max}{1 + \exp^{(\theta + \mu t)}}$
Where, α , β , γ , ω , θ , and μ are regression parameters, t = elapsed time (year), and Max = the maximum value of cracking (the threshold)		
*English units are displayed for each distress type, however metric units are used as well		

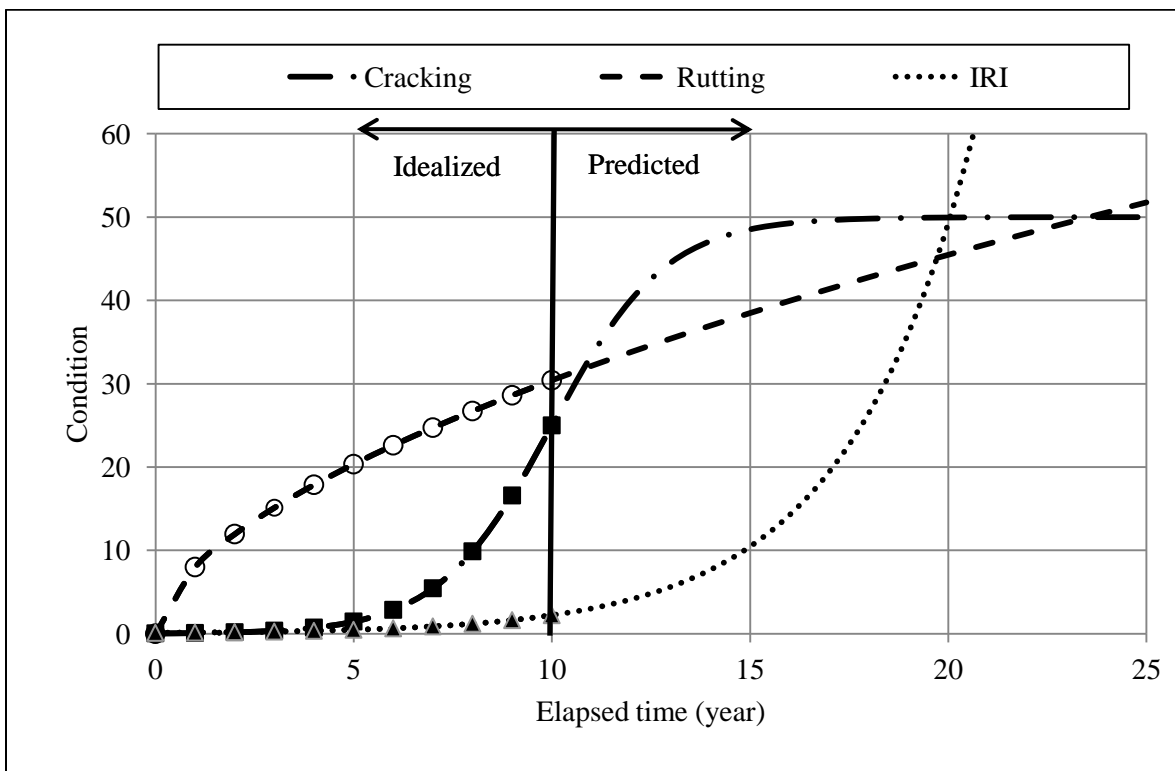


Figure 2.10 Illustration of the mathematical functions used to model IRI, rutting, and cracking and to predict future conditions

2.9.5 Uses of Data Modeling

The modeling of pavement performance data is essential to pavement management at both the network and project levels. In its broadest sense, the major purposes of monitoring and modeling pavement performance data are to (Lytton 1987):

1. Objectively determine the current pavement conditions of each pavement section along the network.
2. Estimate the pavement rate of deterioration and future conditions.
3. Estimate the pavement remaining service life (RSL) and the time for pavement treatment.

The resulting information is used to formulate a management plan of actions. The specific use of pavement performance data modeling at the network level includes:

1. Prioritizing pavement treatment projects.
2. Determining budgetary needs or constraints.
3. Assessing the overall health of the pavement network.
4. Communicating with lawmakers and the users.

The specific uses of pavement performance data modeling at the project level include (Lytton 1987):

1. Analysis of specific pavement segments.
2. Design of detailed pavement treatments for pavement projects.
3. Analysis of life cycle costs for various pavement treatment design alternatives.
4. Estimation of pavement treatment benefits.

One common project and network level pavement performance parameter that is calculated by modeling the pavement condition data is referred to as the Remaining Service Life (RSL). For each pavement distress type, the RSL of a pavement section is the time in years

between now or the year when the latest distress data point was collected and the year when the pavement distress reaches a certain distress threshold value (Baladi et al. 1991). Hence, one RSL value can be calculated for each distress type and the lowest value should be assigned to the pavement section in question. When the RSL of a pavement section reaches zero value, it implies that the pavement has started to provide substandard service quality in terms of ride quality, safety, functionality, structural capacity or any combination thereof.

A significant advantage of the RSL concept is that it combines the severity and extent of the distress as well as the rate of the pavement deterioration. Thus, the RSL is preferred over other modeling techniques such as pavement distress indices. Various pavement sections may have similar distress indices but significantly different rates of deterioration. Thus, the RSL solves the main problem of the individual and combined distress indices, allowing SHAs to make proper pavement management decisions regarding the past and future service life of the pavement structures.

The steps enumerated below can be used to calculate the RSL based on the time-series distress data.

1. Data Mining – Construction and maintenance records should be examined to identify the boundaries of pavement projects. The available time series distress data for each 0.1 mile pavement segment along these projects should be downloaded from the database.
2. Threshold – For each distress type, establish a distress threshold value such that when the pavement conditions reach the threshold value the pavement is in substandard condition and in need of immediate action. Examples of distress threshold values are listed in Table 2.5 (Titus-Glover et al. 2010).

Table 2.5 Example distress thresholds

Distress	Threshold
IRI	200 inch/mile
Rut Depth	0.5 inch
Alligator Cracking	10% (105.6 feet or 316.8 feet ²)
Longitudinal Cracking	700 feet
Transverse Cracking	700 feet or 58 full-width cracks

3. Acceptance Criteria – For each pavement segment within a given pavement section or project, subject the available time series pavement distress and condition data (from step 1) to the two acceptance criteria stated below. The objective is to determine whether or not portions of the project should be excluded from the analyses (Dawson 2012, Dean 2012). The two acceptance criteria are:
 1. Three data points – For treated pavement sections, a minimum of three data points before treatment (BT) and three data points after treatment (AT) are required to model the data using a non-linear function; any model can fit one or two data points. For pavement sections that were not subjected to treatment or no distress data are available before treatment, a minimum of three data points are required after construction.
 2. Positive regression parameters – Positive regression parameters of the pavement deterioration model are required for time-series distress data that are consistent for pavement sections that did not receive any treatment. Negative regression parameters imply that the pavement is self-healing with time (decreasing distress) without any treatments, which is not reasonable and cannot be used to study the pavement deterioration.

4. Data Modeling – For the pavement segments that satisfy the two acceptance criteria, model the time series pavement condition data using the proper mathematical function listed in Table 2.4 and shown in Figure 2.10.
5. RSL – For each pavement segment and for each equation (distress type) obtained in the previous step, calculate the time (t_{th}) at which the distress is expected to reach its threshold value. The RSL is the difference in years between the calculated time and the pavement surface age as stated in Equation 2.3.

$$RSL = (t_{th} - SA) \leq [DSL - SA] \quad \text{Equation 2.3}$$

Where t_{th} = time at which the threshold value is reached in years;

SA = surface age in years; and

DSL = design service life in years.

It should be noted that the constraint on the RSL value to be equal to or less than the difference between the DSL and SA is mainly for new pavement sections where the pavement shows little to no deterioration. Such constrain should not be used when the pavement section in question starts to show measurable amount of distress. Further, the inclusion of the DSL in the equation enhances the feedback from the pavement management to the pavement design process. Figure 2.11 shows an idealized illustration of the RSL using the three mathematical models.

2.10 Life Cycle Cost Analysis (LCCA)

In general, highway pavements are designed and constructed to provide services for a limited time period, often referred to as the pavement service life. Over time, the combined effects of traffic loads and environmental factors accelerate the pavement deterioration rates and reduce its level of serviceability. Maintenance, preservation, and rehabilitation treatments are designed and applied to pavement sections to slow their rates of deterioration and to extend their

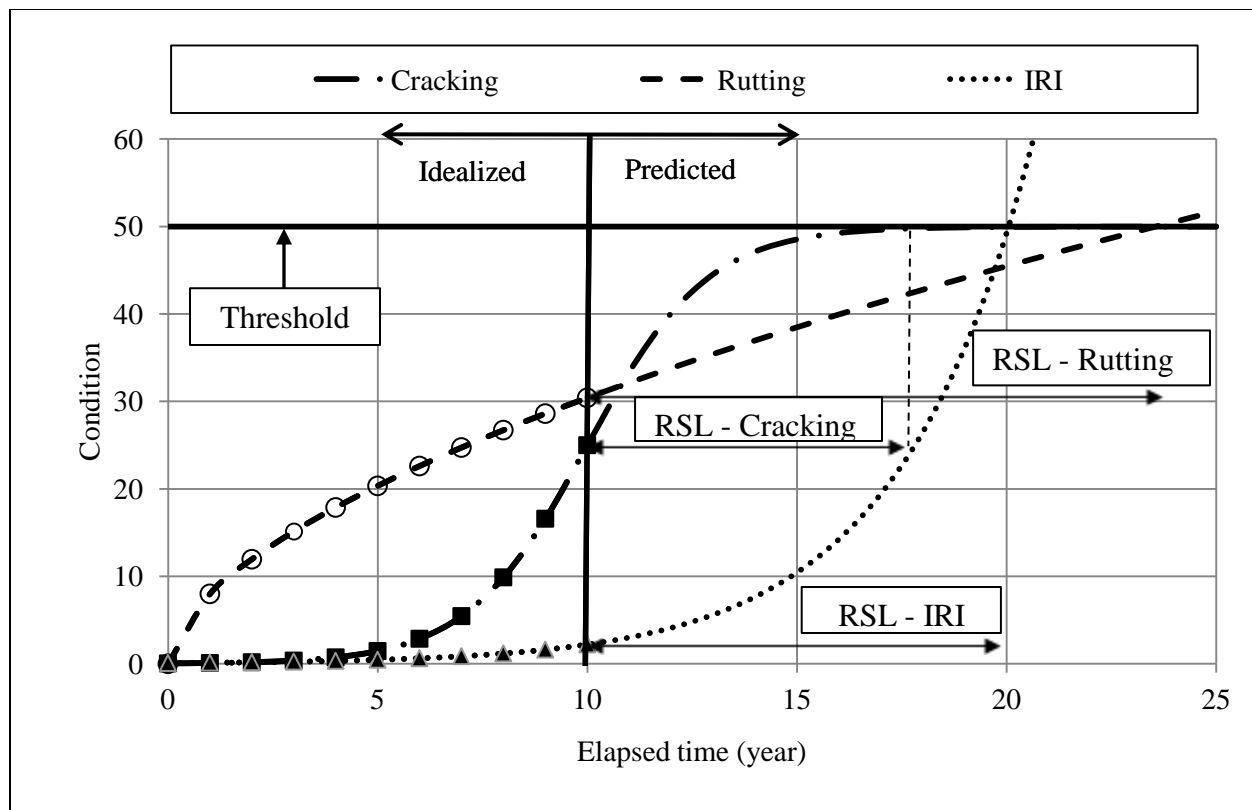


Figure 2.11 Illustration of the Remaining Service Life using three mathematical models

service lives. The application of any pavement treatment requires traffic control (lane closures and/or detours), which significantly impacts traffic flow, increases travel time, and increases vehicle operating costs (VOC). The costs and benefits of pavement treatments are comprised of many elements including:

1. The agency costs of the pavement treatment, which consist of many attributes including:
 - Material and contractual costs.
 - The cost of traffic control in the work zone, which is defined as an area along the highway system where maintenance and construction operations adversely affect the number of lanes opened to traffic or affect the operational characteristics of traffic flow through the work zone (Chien et al. 2002).

- Quality assurance and quality control (QA/QC) costs.
 - The costs of future treatments.
2. The agency benefits, which could be measured by the life of the treatment or the life extension of the treated pavement sections. It should be noted that, for any given treatment, the net benefits to the agency should be calculated by the impact of the given treatment on the weighted average longevity (or the remaining service life) of the pavement network.
 3. The user costs, which are also comprised of many attributes including (Daniels et al. 2000, Lewis 1999):
 - Time delay user costs or work zone user costs, which are defined as the associated costs of time delays due to lane closures because of roadway construction, rehabilitation, and maintenance activities.
 - Costs incurred by those highway users who cannot use the facility because of either agency or self imposed detour requirements.
 - Vehicle operating costs in terms of fuel, wear and tear, and depreciation over the delay periods.
 - Accident costs.
 - Environmental costs due to air pollution caused by excessive uses of gasoline or diesel fuel due to lower speed and time delay, including noise pollution.
 4. The user benefits which are comprised of improved serviceability and ride quality that would lower the VOC and improve traffic flow.

The user costs due to planned changes in highway capacity or improvement in pavement condition by means of road maintenance, rehabilitation and/or reconstruction can also be divided

into two categories; work zone costs and non-work zone costs (Daniels et al. 2000, Lewis 1999). These two general categories and their various components are shown in Figure 2.12.

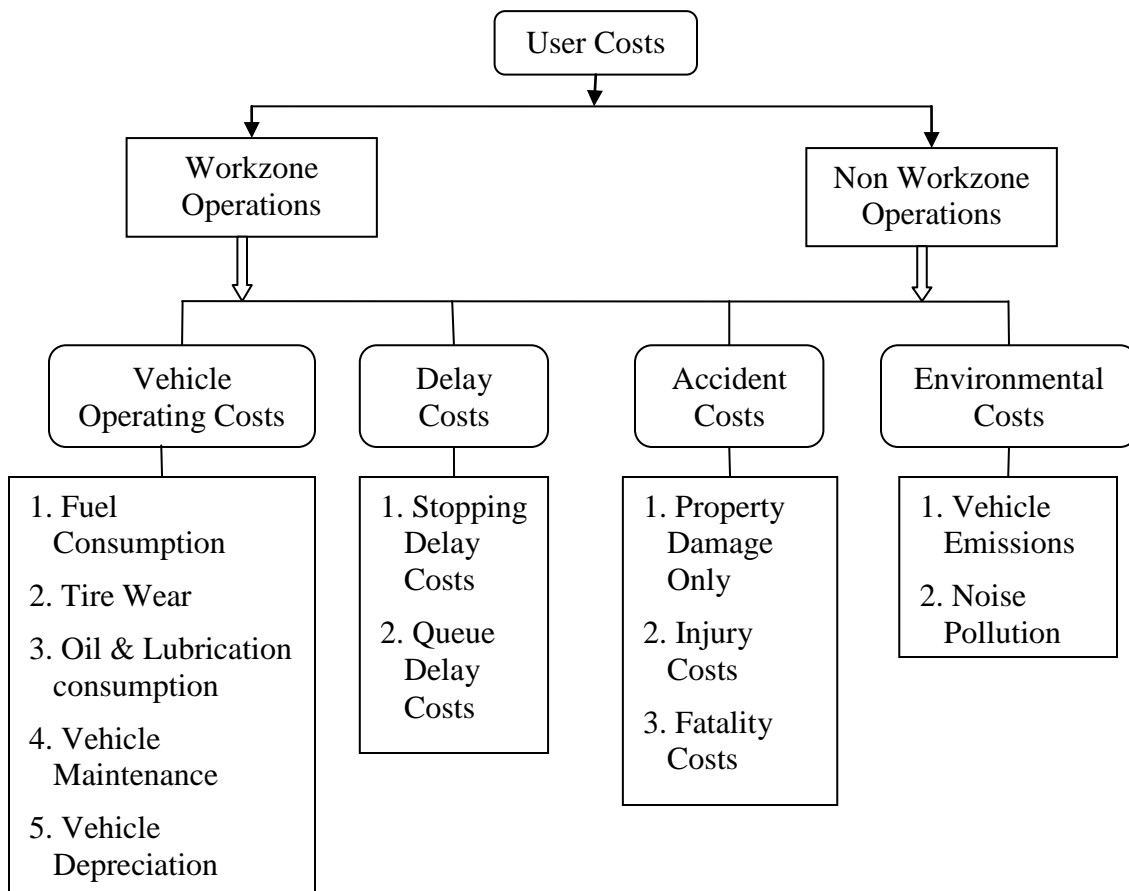


Figure 2.12 Components of user costs (after Morgado and Neves 2009)

Estimation of the work zone user costs requires significant and detailed information on the work zone length, speed restrictions, lane closures, VOC, and traffic capacity in the work zone. For instance, work zone length (the length of the pavement that is subjected to maintenance) has a significant impact on the user costs when the demand is greater than the capacity of the traffic control practices. Hence, it is important to estimate an optimal work zone length which minimizes both user and construction costs. Various models have been developed

to estimate optimal work zone lengths for both two and four lane freeways by minimizing both construction and additional user costs (Hall et al. 2003).

Traffic control procedures also have a significant impact on the agency and user costs in work zones. Some traffic control procedures may accelerate construction, therefore decreasing the time the work zone is in place. In this way, the agency cost increases while the user cost decreases. On the other hand, some traffic control procedures may cause the duration of the work zone to increase, which may cause the agency cost to decrease and the user cost to increase. Hence, there is a direct relationship between traffic control procedures, agency cost, and user cost. Analyses indicate that the cost effectiveness of traffic control procedures depends on several site factors and should be evaluated on a site-by-site basis (Dudek et al. 1986).

In addition, as shown in Figure 2.12, non-work zone user costs are comprised of the costs associated with standard travel such as VOC, accident, and environmental costs. The delay time is assumed to be zero. The VOC are also a function of several other variables including vehicle type, vehicle speed, speed changes, gradient, curvature, and pavement surface (Morgado and Neves 2009).

Given that LCCA requires extensive data from both the agency and user-perspectives, it is usually only performed to consider pavement treatment alternatives for pavement projects greater than one million dollars in value. When performed properly, LCCA has applications for many areas of interest to State and local transportation agencies. Common applications of LCCA include the following (FHWA 2011):

1. Designing, selecting, and documenting the most affordable means of accomplishing a specified project or objective. For instance, if a bridge must be replaced, LCCA can be

used to select the replacement option that would cost the least over the expected life of the bridge.

2. Evaluating pavement preservation strategies. The costs of each strategy can be evaluated relative to the expected effects it will have on delaying the costs of expensive rehabilitations or reconstructions.
3. Performing value engineering (VE). Value engineering must be applied to all Federal-aid highway projects on the National Highway System with an estimated cost of \$25 million or more. Among other requirements, the VE team must consider the lowest life cycle cost means of accomplishing a project.
4. Project planning and implementation, especially the use and timing of work zones. LCCA allows the analyst to balance higher agency and/or contractor costs associated with off-peak work hours against reduced traveler delay costs associated with fewer work zones during peak hours.

CHAPTER 3

DATA MINING

3.1 Foreword

The data mining phase of this research study was completed through the combined effort of several individuals to organize and store massive amounts of pavement management system (PMS) data obtained from various sources. Hence, the material in this chapter is the combined work of various individuals including the author of this thesis, Dr. Tyler Dawson, and Dr. Gilbert Baladi et al (Dawson 2012, Baladi et al. 2012). Nonetheless, the complete data mining materials are included in this thesis for the convenience of the reader and to ensure clarity and promote understanding.

3.2 Introduction

The pavement condition and distress, treatment types, and cost data used in this research study were obtained from the PMS units of the following State Highway Agencies (SHAs) and pavement study.

1. The Colorado Department of Transportation (CDOT).
2. The Louisiana Department of Transportation and Development (LADOTD).
3. The Michigan Department of Transportation (MDOT).
4. The Washington State Department of Transportation (WSDOT).
5. The Minnesota Road Research Project (MnROAD).

The geographical locations of the four SHAs and MnROAD are distributed throughout the USA with varying network size, traffic, climate, etc. Hence, the pavement data are somewhat representative of the pavement network in the USA.

At the onset of this research, pavement condition and distress data for only nine pavement sections were requested from each of the four SHAs. These data were received and analyzed; later the entire PMS databases were requested to broaden the extent of the analyses and to incorporate the entire sets of available data from each SHA. The databases were received in various formats (Microsoft Excel spreadsheets, Microsoft Access files, text documents, and so forth). Further, the pavement condition and distress data were expressed using various measurement units (SI and English) and various formats such as area, length, count, average, maximum, etc. The data were transferred to Microsoft Excel files and the units of measurements were converted and unified for analyses.

3.3 Data Format, Restructuring, and Unification

The first step in support of the pavement condition and distress data analyses, discussed in Chapter 4, was to convert each of the four PMS databases obtained from the four SHAs, to one uniform format for analyses. The database conversions were conducted as listed below:

1. The pavement condition and distress data from the States of Colorado, Louisiana, and Washington were received in Microsoft Access format. The data analyses (see Chapter 4) were conducted in-part using the Matlab computer program, which does not read Microsoft Access files. Hence, the data were converted to Microsoft Excel files. In addition, the data from Washington and Louisiana were stored in separate Excel files, one file for each data collection year. This was accomplished by sorting the data by the data collection year and copying each year individually to a new Excel file. The data from Colorado were already stored by year.
2. The pavement condition and distress data received from MDOT were stored in several formats including Microsoft Excel and Text files such as EI2, WI2, ER3, WR3, etc.

Numerous attempts were made to convert the data without any success. Consequently, the data were left in their original formats.

As stated earlier, the pavement condition and distress data were reported using various units of measurements and measurement types as listed in Table 3.1. Since the data from each SHA were used in this study, and in order to conduct uniform analyses, some of the data elements were re-arranged and/or re-configured to produce uniform data structure using uniform units of measurement. The restructuring and re-configuration did not affect the accuracy and/or the variability of the data. Such restructuring and re-configuration were dependent on the pavement condition or distress type as detailed below.

Table 3.1 Pavement condition and distress data stored by various SHAs

Condition or distress	Units and types of measurements reported by SHAs				Units used in the analysis
	SI	English	Types of measurements		
Roughness	m/km	in/mi	Each wheel path	Average wheel paths	Average IRI (inch/mile)
Rut depth	mm	inch	Maximum and average rut depth	Average rut depth	Average rut depth (inch)
Alligator cracks	m and m ²	ft and ft ²	Percent of length	Percent of area	Length along the pavement segment (ft)
Longitudinal cracks	m	ft	Cumulative length along the pavement segment		Cumulative length along the pavement segment (ft)
Transverse cracks	m and count	ft and count	Count	Cumulative length	Cumulative length along the pavement segment (ft)

- Roughness – Each of the four SHAs collects pavement roughness data in terms of the International Roughness Index (IRI). The data are recorded in either SI units (m/km) or English units (in/mi). In this study the data reported in SI units were converted to English units. In addition, the roughness data for each 0.1 mile long pavement segment were reported, depending on the SHA, for each wheel path and/or the average IRI of the two

wheel paths along the pavement segment in question. The average IRI measurements of the two wheel paths per pavement segment were used in the analysis.

- **Rut Depth** – The rut depth data are collected and stored in either SI units (mm) or English units (in). In this study the data reported in SI units were converted to English units. In addition, for each 0.1 mile long pavement segment, the rut depth data were reported, depending on the SHA, as the average of each wheel path, the average of the pavement segment, or the average and the maximum rut depths of the pavement segment in question. The average rut depth measurements per pavement segment were used in this study.
- **Alligator Cracking** - The alligator cracking data are collected and stored in terms of length (m or ft), area (m^2 or ft^2), percent of the pavement length, or percent of area. In addition, the alligator cracking data for each 0.1 mile long pavement segment were reported, depending on the SHA, over the entire pavement width (12 feet) or in the wheel paths (3 foot width per wheel path) of the pavement segment in question. In this study the alligator cracking data were converted to linear feet by dividing the area by the width. For example, a pavement segment with 600 ft^2 of alligator cracking, and full lane width measurement criteria, was divided by 12 feet, yielding 50 linear feet of alligator cracking.
- **Longitudinal Cracking** - The cumulative length of longitudinal cracks are collected and stored in either SI units (m) or English units (ft). In this study the cumulative longitudinal crack length data were converted to English units and used in the analysis.
- **Transverse Cracking** - The transverse cracking data are collected and stored in either SI units (m), English units (ft), or by count (number). The latter was reported as the number of transverse cracks in certain length categories (such as 0 to 1 foot, 2 to 3 feet, etc.) or

simply the number of transverse cracks without crack lengths. In this study, the cumulative length of transverse cracks in feet was used in the analysis. The cumulative length of transverse cracks was calculated by multiplying the crack count by the average reported length. For example, one crack in the 0 to 1 foot long category was counted as 0.5 foot long, while a crack in the 6 to 12 feet category was counted as a 9 feet long crack. Transverse crack data reported by count only (no crack lengths were included), were assumed to extend across the 12 feet wide lane and therefore their cumulative length was calculated by multiplying the crack count by 12 feet.

- Severity levels – Most cracking data are stored under three separate severity levels: low, medium, and high. The problem of such data is that the crack severity level rating is a function of the judgment of the surveyor who is reviewing and digitizing the electronic images. Such judgment is a function of the degree of training and experience of the individual surveyor. In addition, the same pavement segment may not be reviewed by the same surveyor each year or each data collection cycle. Thus, a crack may be labeled high severity in one year and medium severity next year and vice versa. Figures 3.1 and 3.2 depict the time series data for each transverse crack severity level along a portion of highway 24 in Colorado. The data in the figures show that:

1. The transverse crack length changes from one data collection cycle to the next without a pavement treatment. For example, the high severity transverse crack length is about 150 feet in year 2000, then only about 10 feet in 2001, and is finally absent in year 2002. Likewise, the medium severity transverse crack length is about 130 feet in year 2000, 150 feet in year 2001, and only about 10 feet in year 2002.

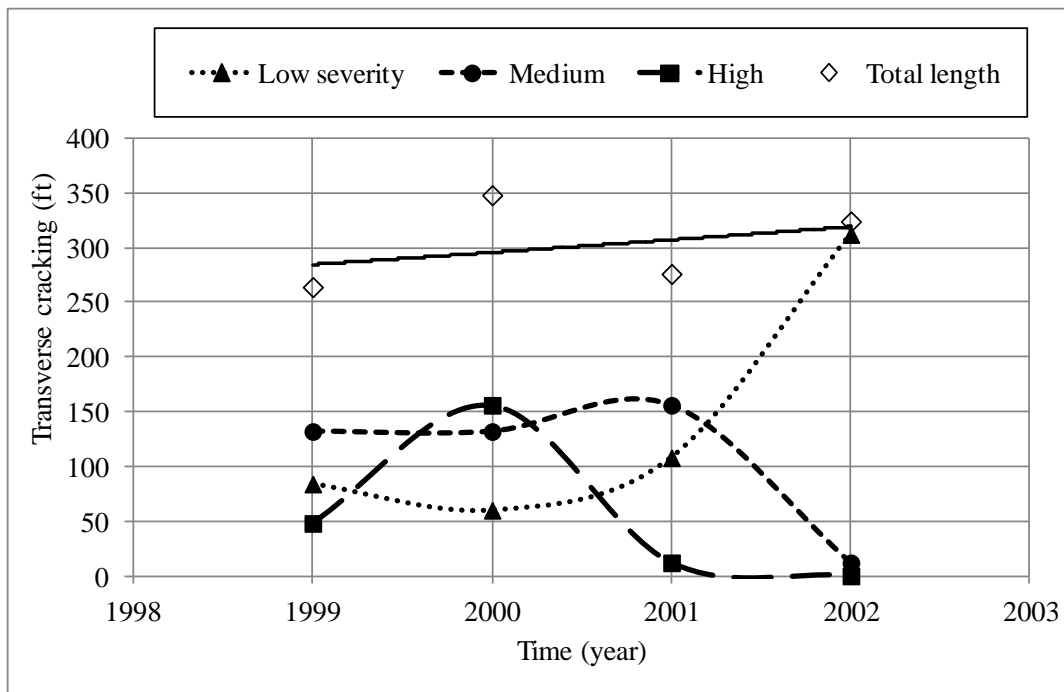


Figure 3.1 Time series transverse cracking data for each severity level and the sum of all levels, Colorado, HWY 24, direction 2, BMP 329.9

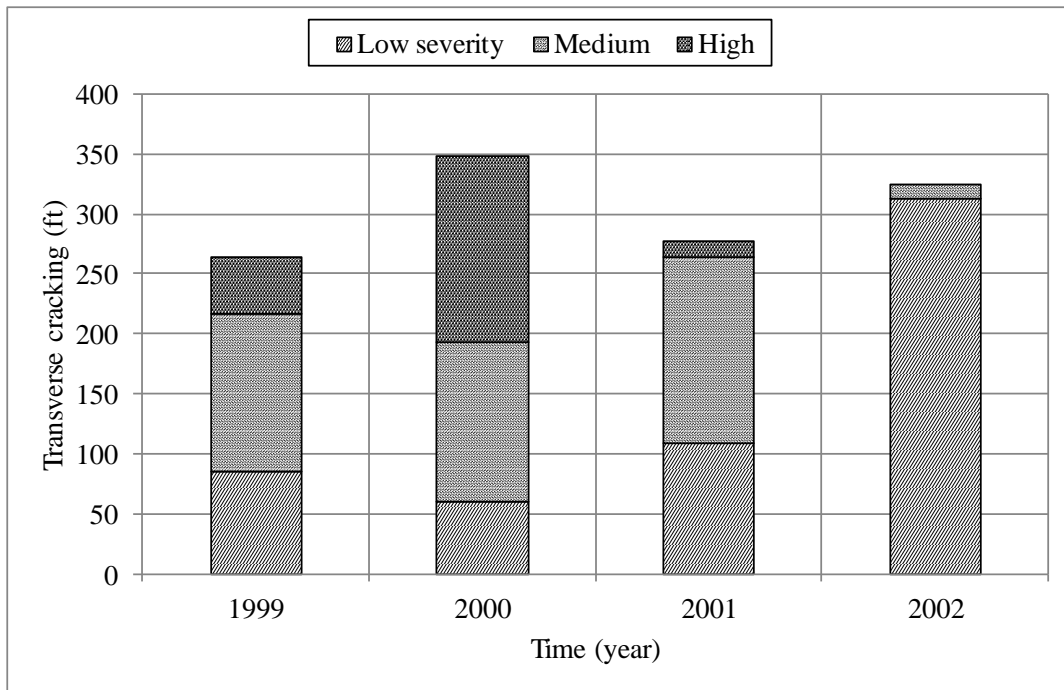


Figure 3.2 Cumulative time series transverse cracking data showing individual transverse crack severity level and the sum of all severity levels, Colorado, HWY 24, direction 2, BMP 329.9

On the other hand, the low severity transverse crack length increases from about 60 feet, to 110 feet, and to 310 feet over the same time period. The reduction in the lengths of medium and high severity transverse cracks could be attributed to the rater assigning low severity to the same cracks which were previously assigned medium or high severity. It could also be related to the pavement temperature at the time of data collection. Higher temperature causes decreases in crack opening and hence, in severity level.

2. The individual severity levels show high variability when modeled in time series, as indicated by exponential models determined for each transverse crack length severity level. In fact, the models indicate that the medium and high severity transverse crack length is decreasing with time without a pavement treatment.

To overcome the two problems of data inconsistency and high variability, the sum of all severity levels for each crack type was obtained and used in the analysis. This yielded more consistent data that can be modeled for analysis as shown in Figures 3.1 and 3.2.

The crack severity level data could be used, though, to roughly estimate the amount of required work. For example, so many feet of cracks in the medium and high severity levels need to be sealed or patched. Low severity cracks are typically not sealed or patched. Likewise, for concrete pavements, low severity transverse cracks may be subjected to dowel bar retrofit, while medium and high severity cracks are typically not.

A summary of the data characteristics and information of each of the four SHAs and how they were re-arranged and re-structured to yield uniform sets of data for analysis are detailed in the next few subsections.

3.3.1 Pavement Condition and Distress Data from Four State Highway Agencies

Table 3.2 lists the pavement condition and distress types, severity levels, and measurement units reported by each of the four SHAs. For each agency, some statistical information regarding the pavement network and a summary of the data characteristics and how they were reformatted and restructured to the uniform set of data are detailed in the next few subsections.

a) Colorado Department of Transportation (CDOT)

CDOT has 9,134 miles (22,912 lane-miles) of pavement and 3,406 bridges under its jurisdiction. The annual traffic is in excess of 48 billion vehicle-miles. Forty percent of that traffic is carried on the Interstate system (I-25, I-70, I-76, I-225, and I-270) which accounts for about ten percent of the network length. In 2009, 7.2 million miles of pavement were cleared of snow and 248,000 tons of asphalt materials and 178,800 gallons of liquid asphalt were used to repair damaged pavement surfaces (Hartgen et. al. 2009, CDOT 2010).

Pavement condition and distress data for three flexible, three rigid, and three composite pavement sections were originally requested and received from CDOT. Later, upon request CDOT provided their entire pavement condition and distress database covering the period from 1998 through 2009. The condition and distress data were collected every year, every other year, or every third year depending on the road and direction of travel. The data were collected on a continuous basis and recorded in the database for each 0.1 mile long pavement segment. The applicable data types, severity levels, and measurement units contained in the database are listed in Table 3.2.

Table 3.2 The pavement condition and distress data received from the four SHAs

SHA	Pavement condition information		
	Condition or distress type	Severity level	Measurement unit
CDOT	Alligator cracking		Feet ² / segment
	IRI		Inch/mile
	Longitudinal cracking		Feet/ segment
	Rut depth		Inch
	Transverse cracking		Count/ segment
LADOTD	Alligator cracking	Low, medium, and high	Feet ² (wheel paths)/ segment
	IRI		Inch/ mile
	Longitudinal cracking	Low, medium, and high	Feet/ segment
	Rut depth in right or left wheel path		Inch
	Transverse cracking	Low, medium, and high	Feet/ segment
MDOT	Alligator cracking consists of two categories (right and/or left wheel path)	25 levels depending on the length and associated distress	Percent of length
	IRI		Inch/ mile
	Longitudinal cracking consists of 3 categories for rigid and 5 for flexible & composite pavements depending on location in the lane	49 levels for flexible, rigid and composite pavements depending on crack length and its associated distress	Percent of length
	Rut depth		Inch
	Transverse cracking consists of two categories for rigid, composite and flexible pavements depending on crack opening for rigid and composite, and crack irregularity for flexible pavements	Twelve levels for rigid and composite and 16 levels for flexible depending on length and width of the associated distresses	Count per segment
WSDOT	Alligator cracking	Low, medium, and high	Percent of 2 wheel paths of a segment
	IRI		Meter/ kilometer
	Longitudinal cracking	Low, medium, and high	% of segment
	Rut depth		Millimeter
	Transverse cracking	Low, medium, and high	Count/ segment

The data were adjusted where necessary to unify the measurement units and severity levels as discussed above. The alligator cracking data were provided as lane area (ft^2) and were divided by the lane width (12 ft) to obtain linear feet. The reported number of transverse cracks was multiplied by the lane width (12 ft) to obtain cumulative linear feet. The IRI, longitudinal cracking, and rut depth data were not adjusted. Note that no distress severity levels were reported with the full database. Examples of the original and converted data are listed in Tables A.1 and A.2 of Appendix A and the entire database is available upon request from the Department of Civil & Environmental Engineering at Michigan State University (MSU) (CDOT 2010).

b) Louisiana Department of Transportation and Development (LADOTD)

LADOTD has 15,987 miles (38,458 lane-miles) of pavement and 8,060 bridges under its jurisdiction. The large number of bridges is due to the extensive areas of land near the Mississippi River and Delta, which are near or below sea level. The Interstate system (I-10, I-12, I-20, I-49, I-55, I-59, I-110, I-210, I-220, I-310, I-510, I-610, and I-910) makes up about six percent of the pavement network length. In 2006, approximately six percent of the network was maintained through sealing or resurfacing (LADOTD 2007, Hartgen et. al. 2009).

Pavement condition and distress data for three flexible, three rigid, and three composite pavement sections were originally requested and received from LADOTD. Later, upon request LADOTD provided their entire pavement condition and distress database covering the period from 1995 through 2009. The condition and distress data were recorded in 1995, 1997, 2000, 2003, 2005, 2007, and 2009. The data were collected on a continuous basis and stored in the database for each 0.1 mile long pavement segment. The applicable data types, severity levels, and measurement units contained in the database are listed in Table 3.2. The alligator cracking data were provided as wheel path area (ft^2) and were divided by the assumed wheel path width

of 3 ft to obtain linear feet. The rut depth data were provided as the average of each wheel path. The average of the averages of the two wheel paths was taken as the average rut depth. The IRI, longitudinal cracking, and transverse cracking data were not adjusted. Note that the low, medium, and high severity distresses were summed. Examples of the formatted data are listed in Table A.3 of Appendix A and the entire database is available from the Department of Civil & Environmental Engineering at MSU (LADOTD 2010).

It should be noted that rut depth data were collected by ultra-sounding in 1993 and 1995, then by 3-point laser system from 1997 to 2005, then by continuous laser scanning after 2005. There is no correlation between equipment. An error in the rut depth collection software was discovered by the staff of LADOTD and the rut depth data collected after 2005 were corrected and replaced the original data. It should also be noted that sealed cracks are not counted as cracks.

c) Michigan Department of Transportation (MDOT)

MDOT has about 9,700 miles (27,503 lane-miles) of pavement and 5,400 bridges under its jurisdiction. The Interstate system consists of I-69, I-75, I-94, I-96, I-194, I-275, I-475, I-496, I-675, and I-696. In 2006, about 400 miles of pavement and 300 bridges were improved (MDOT 2010, Hartgen et. al. 2009).

Pavement condition and distress data for three flexible, three rigid, and three composite pavement sections were originally requested and received from MDOT. Later, upon request MDOT provided their entire pavement condition and distress database covering the period from 1992 through 2009. The condition and distress data were typically recorded every other year, alternating annually for each direction of travel. The data were collected on a continuous basis and recorded for each 0.1 mile long pavement segment. The applicable data types, severity

levels, and measurement units contained in the database are listed in Table 3.2. The alligator and longitudinal cracking data were provided as decimal percent of the length of each segment (528 feet) and were multiplied by the segment length to obtain linear feet. The transverse cracking data were provided as the count of cracks with certain range of length and were multiplied by the average length in each range to obtain linear feet. The IRI and rut depth data were provided in inch/mile, and inch, respectively and hence, were not adjusted. Note that the cracking data are collected under several classifications and the length was summed. Examples of the formatted are listed in Table A.4 of Appendix A and the entire database is available from the Department of Civil & Environmental Engineering at MSU (MDOT 2010).

The MDOT pavement condition and distress data are stored in several formats scattered among thousands of individual files. Each file contains portions of the pavement condition and distress database. For example, the IRI, rut depth, and cracking data are each stored separately in different data files having different formats for each 0.1 mile long pavement segment. In addition, the majority of each data type is stored in different files for each control section. For a given control section and a given pavement data collection cycle, the linear location reference system in the IRI and rut depth files often do not match that in the cracking files. For example, the beginning mile points (BMPs) for the sensor collected data (IRI and rut depth) and the videotaped data (cracking) are not consistent along portions of M-39, control section 82192, direction 1. In 2001, a portion of the videotaped data were collected for the pavement segments with BMPs 0.0, 0.131, 0.256 and so on; while the sensor collected data were collected for pavement segments with BMPs 0.0, 0.1, 0.2 and so on. To further complicate the situation, similar problems exist in time series where the BMPs of the pavement segments float back and forth from one data collection cycle to the next. Figure 3.3 shows the BMPs of the first 10

pavement segments along M-39, control section 82192, direction 1 from 5 data collection cycles (1999 to 2007). The data in the figure indicate that the BMPs of the pavement segments shift significantly back and forth from one year to the next. Further, few of the pavement segments are more or less than 0.1 mile long and some of the segments were not surveyed in few years. Countless attempts were made to adjust the BMPs using especially written Matlab based computer programs to match the data for each pavement segment within each data collection cycle (sensor and videotaped data) and between various data collection cycles. Some of the data were matched by shifting each BMP along the pavement to coincide with a reference BMP (the oldest data collection cycle). In the example, the data collected from 2001 to 2007 were shifted forward or backward to match the BMPs of the data collected in 1999. The difficulties arise because the required shift in the BMPs for each 0.1 mile long pavement segment within a data collection cycle is not constant. Further, for the same 0.1 mile long pavement segment, the required shift in the BMP from one data collection cycle to the next is not constant. Such shifts were often greater than 0.1 mile. Nevertheless, Figure 3.4 shows the resulting time series longitudinal crack data after adjusting the BMP of each 0.1 mile long pavement segment by hand. The data in the figure indicate that the shifted BMPs result in highly variable time series cracking data. Such variability did not allow:

1. Modeling the data as a function of time.
2. Calculating consistent rate of deterioration.
3. Estimating the benefits of treatments.

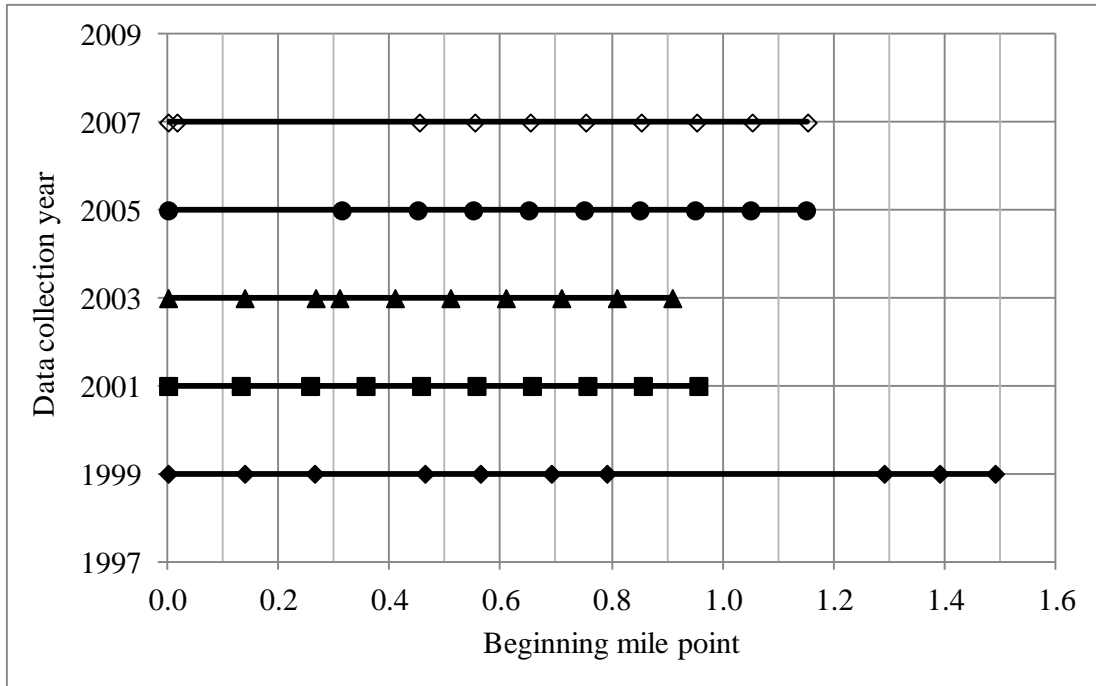


Figure 3.3 The BMP of the first ten 0.1 mile long pavement segments along M-39, control section 82192, direction 1, Michigan

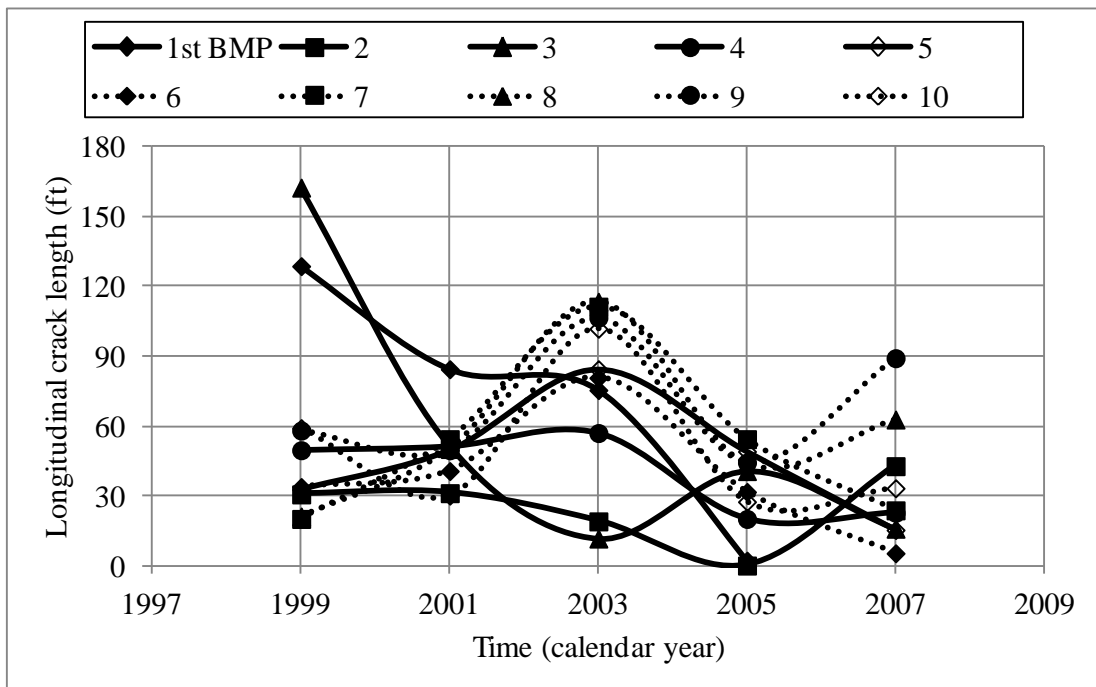


Figure 3.4 Longitudinal cracking data after adjusting the BMP for the first ten 0.1 mile long pavement segments along M-39, control section 82192, direction 1, Michigan

Finally, for about a four month period, countless attempts were made to unify the BMP of each 0.1 mile long pavement segment between the various data collection cycles. Unfortunately, none of these attempts were even partially successful. Therefore, it was decided to exclude the MDOT pavement condition and distress data from most analyses in this study.

d) Washington State Department of Transportation (WSDOT)

WSDOT has about 7,000 miles (18,392 lane-miles) of pavement and 3,400 bridges under its jurisdiction. The Interstate system consists of I-5, I-10, I-82, I-90, I-405, and I-705 (Hartgen et. al. 2009, WSDOT 2010).

Similar to the other three states, pavement condition and distress data for three flexible, three rigid, and three composite pavement sections were originally requested and received from WSDOT. Later, upon request WSDOT provided their entire pavement condition and distress database covering the period from 1969 through 2007. The condition and distress data were collected and recorded primarily every other year from 1969 to 1988 and every year starting in 1989. The data were collected on a continuous basis and recorded for each 0.1 mile long pavement segment. The applicable data types, severity levels, and measurement units contained in the database are listed in Table 3.2. The alligator cracking data were provided as percent of the length of the two wheel paths in the segment (1028 feet). In this study, the data were divided by 100 and multiplied by the segment wheel path length to obtain linear feet. The Longitudinal cracking data were provided as percent of the length of each segment (528 feet). Once again, the data were divided by 100 and multiplied by the segment length (528 feet) to obtain linear feet. The transverse cracking data were provided as the count of the number of cracks and were multiplied by the lane width (12 feet) to obtain linear feet. Note that, in the analysis of the cracking data, the sum of the low, medium, and high severity levels was used. The IRI and rut

depth data, which are recorded in SI units (m/km and mm) were converted to English units (in/mile and inch), respectively. Examples of the formatted data are listed in Table A.5 of Appendix A and the entire database is available from the Department of Civil & Environmental Engineering at MSU (WSDOT 2010).

3.3.2 Minnesota Road Research Project (MnROAD)

The MnROAD test facility is located near Albertville, Minnesota. The pavement testing facility was constructed between 1992 and 1994 through a joint venture between the Minnesota Department of Transportation (MnDOT) and the Minnesota Local Road Research Board (LRRB). The test facility consists of two road sections divided into fifty five 500-foot long segments or “cells”. One road section is a 3.5 mile mainline section along I-94 carrying about 28,500 vehicles per day. The other is a 2.5 mile long low volume road utilizing a controlled five-axle semi tractor-trailer (MnROAD 2008, MnDOT 2011). The pavement condition and distress types, severity levels, and measurement units contained in the database are listed in Table 3.3. The longitudinal, transverse, and alligator cracking data were not restructured or reconfigured. Note that, in the data analysis of each cell, the sum of the low, medium, and high severity cracking data was used. The IRI and rut depth data were converted from SI to English units. Examples of the formatted data are listed in Table A.6 of Appendix A and the entire database is available from the Department of Civil & Environmental Engineering at MSU.

Table 3.3 MnROAD database items

Condition or distress type	Severity level	Measurement unit
Alligator cracking	Low, medium, and high	Feet/cell
IRI		Meter/ kilometer
Longitudinal cracking (wheel path or non wheel path and sealed or not sealed)	Low, medium, and high	Feet/cell
Rut depth		Millimeter
Transverse cracking (sealed or not sealed)	Low, medium, and high	Feet/cell

3.4 Treated Pavement Section Identification

Following the conversion and restructuring of the pavement condition and distress databases, exhaustive searches were conducted of the CDOT, LADOTD, and WSDOT databases and data files to identify pavement sections that were subjected in the past to one of the following six most common treatment types:

1. Thin (< 2.5 inch) hot mix asphalt (HMA) overlay of asphalt surfaced pavement.
2. Thick (≥ 2.5 inch) HMA overlay of asphalt surfaced pavement.
3. Single chip seal.
4. Double chip seal.
5. Thin (< 2.5 inch) mill and fill of asphalt surfaced pavement.
6. Thick (≥ 2.5 inch) mill and fill of asphalt surfaced pavement.

It should be noted that, the pavement treatment files provided by each of the three SHAs were used in the identification of treated pavement sections. This identification consists of determining the acceptable time window of the treatment calendar year such that the database likely contains three or more data collection cycles before treatment (BT) and three or more data collection cycles after treatment (AT). For example, if the database contains data collection cycles for the calendar years 1994, 1996, 1999, 2000, 2002, 2004, 2006, and 2008, then the

acceptable treatment time window is between the years 2000 and 2003. This increases the probabilities that every 0.1 mile long pavement segment has three or more available data points BT and AT. In general, the section identification implies that for those states that collect data every other year, an absolute minimum of 12 years of time series data (6 years BT and 6 years AT) are required. Whereas for the State of Washington, which collects the data every year, an absolute minimum of 6 years of data are required (3 years BT and 3 years AT).

It is important to note that, in this study as well as in others, such data search is difficult, laborious, and time consuming. The difficulties that were, or could be, encountered include:

- For some pavement sections, the PMS database of some SHAs, especially those who collect pavement condition and distress data on a bi-annual basis, did not contain the minimum of three time series data points BT and/or AT.
- Many pavement sections were subjected to treatment more frequently than the required time span listed above.
- For a significant number of pavement sections, accurate and consistent records regarding the applications of routine, reactive, and preventive maintenance treatments do not exist in the database.
- For those SHAs who collect pavement condition and distress data using a sampling technique, the availability of three time series data points could be significantly reduced.

Nevertheless, for each SHA, the number of pavement sections or projects that were identified for analysis, the cumulative length of these sections for each treatment type, and the total length of all projects of all SHAs identified for analyses are listed in Table 3.4.

Table 3.4 The number of pavement projects and total length identified in each state

Treatment type	Number and total length of pavement projects in three SHAs						Total, all SHAs	
	Colorado		Louisiana		Washington			
	Projects	Length	Projects	Length	Projects	Length	Projects	Length
1	70	104.6	18	45.5	48	130.9	136	281
2			59	252	4	22.7	63	274.7
3	69	495.8	46	242.1	7	20.4	122	758.3
4			12	40.5			12	40.5
5	2	9.1	11	36.5	21	93	34	138.6
6			32	139			32	139
Total	141	609.5	178	755.6	80	267	399	1632.1
Treatment type: 1 = Thin (< 2.5 inch) HMA overlay of asphalt surfaced pavements; 2 = Thick (≥ 2.5 inch) HMA overlay of asphalt surfaced pavements; 3 = Single chip seal; 4 = Double chip seal; 5 = Thin (< 2.5 inch) mill and fill of asphalt surfaced pavements; 6 = Thick (≥ 2.5 inch) mill and fill of asphalt surfaced pavements.								

3.5 Time Series Data Restructuring

For each treated pavement section identified above, the pavement condition and distress data for each 0.1 mile long pavement segment along the section were restructured in a time series format. A Matlab based computer program was written to search the database and retrieve the corresponding condition and distress data based on the user inputs of:

- The beginning and the ending mile point (BMP and EMP) of the pavement project.
- The route and control section number and the direction of traffic.

For each pavement section/project, the output of the Matlab program consists of one Excel file containing six worksheets; one sheet per condition and distress type and one listing the treatment type, treatment time, and project boundaries (BMP and EMP).

3.6 Cost Data

Limited cost data were received from the three SHAs. When cost data were available they were extremely limited in detail. For example, the treatment cost consists of the total contract cost. The cost data did not include the breakdown of cost relative to materials, pre-overlay activities, safety improvements (e.g., guardrails), signing, mobilization charges, incentive/disincentives, and so forth. Detailed cost data are required to determine the cost-effectiveness of each treatment. Nevertheless, when available, only the material and labor costs data were used to determine the cost-effectiveness of the treatment. Extraneous cost items in the contract (guardrails, sign replacement, etc.) were not included in the analyses. Similarly, the equipment mobilization cost is a significant portion of the cost of short pavement projects and small portions of longer projects. In most cases, this information was not available, but under ideal circumstances should be included in the database and in the analyses. Further, since the pavement conditions and distress levels along a given pavement project vary substantially, the amount of pre-treatment work and cost would also vary. Hence, a complete set of cost data should include the cost of the treatment along each 0.1 mile of the pavement section. This would allow analysis of the optimum time at which the pavement treatment is the most cost effective.

The lack of sufficient and detailed cost data did not allow comprehensive analysis of the treatment costs in this study. Regardless, analyses of treatment benefits (effectiveness) were undertaken. The analysis procedures and the results are detailed and discussed in Chapter 4. Note that, Baladi & Dean 2011 obtained, from MDOT, general cost data of various preventive maintenance treatment actions. The data are summarized in Tables 3.5 and 3.6 for the years 2009 and 2010, respectively. The data in the tables detail the entire range of project costs, which include project locations and possible provisions of the contracts. Table 3.7 summarizes the

general material costs and the total costs including materials, engineering fees, culverts and slope work, etc. Finally, equipment mobilization is typically about \$50,000. The cost data (listed in Tables 3.5 and 3.6) were used to conduct life cycle cost analysis (LCCA) of light (frequent) and heavy (less frequent) pavement treatments in urban and rural areas. A summary of the analyses procedures and results is presented in Chapter 4. Detailed LCCA results are discussed elsewhere (Baladi et al. 2012, Dean 2012).

Table 3.5 Typical preventive maintenance treatment costs in 2009 for the state of Michigan (Baladi and Dean 2011)

Treatment type	Cost per unit in 2009		
	Minimum	Average	Maximum
Cold milling bituminous/HMA surface	\$0.26/yd ² \$0.01/ton	\$0.69/yd ² \$6.34/ton	\$7.43/yd ² \$11.50/ton
Non structural bituminous/HMA overlay	\$39.54/ton	\$52.73/ton	\$100.85/ton
Over band crack fill	\$2,145/RBMI	\$4,191/RBMI	\$13,150/RBMI
Bituminous crack treatment	\$3,092/RBMI	\$4,210/RBMI	\$17,000/RBMI
Chip seal	\$1.24/yd ²	\$1.38/yd ²	\$1.66/ yd ²
Double chip seal	\$2.57/ yd ²	\$2.78/ yd ²	\$2.98/ yd ²
Micro-surface with Warranty	\$2.55/ yd ²	\$3.09/ yd ²	\$3.91/ yd ²
HMA ultra-thin with low volume warranty	\$2.08/ yd ²	\$2.35/ yd ²	\$2.70/ yd ²
HMA ultra-thin with medium volume warranty	\$2.72/ yd ²	\$2.86/ yd ²	\$3.07/ yd ²
Paver placed surface seal	\$5.19/ yd ²	\$5.74/ yd ²	\$6.00/ yd ²
Diamond grinding concrete pavement	\$2.90/ yd ²	\$3.14/ yd ²	\$3.18/ yd ²
Resealing transverse pavement joints with hot-poured rubber	\$0.69/ft	\$0.97/ft	\$2.00/ft
Resealing longitudinal joints with hot-poured rubber	\$0.59/ft	\$0.92/ft	\$1.50/ft
Crack sealing concrete pavement	\$1.00/ft	\$1.12/ft	\$2.00/ft
Yd ² = square yard with 1 inch depth, RBMI = roadbed mile, ft = foot			

Table 3.6 Typical preventive maintenance treatment costs in 2010 for the state of Michigan (Baladi and Dean 2011)

Treatment type	Cost per unit in 2010		
	Minimum	Average	Maximum
Cold milling bituminous/HMA surface	\$0.01/ yd ² \$0.01/ton	\$0.72/ yd ² \$0.45/ton	\$7.00/ yd ² \$1.00/ton
Non structural bituminous/HMA overlay	\$45.00/ton	\$57.54/ton	\$110.00/ton
Over band crack fill	\$2,935/rbmi	\$4,418/rbmi	\$13,600/rbmi
Bituminous crack treatment	\$1,757/rbmi	\$4,447/rbmi	\$12,541/rbmi
Chip seal	\$1.28/ yd ²	\$1.51/ yd ²	\$1.65/ yd ²
Double chip seal	\$2.69/ yd ²	\$2.75/ yd ²	\$2.81/ yd ²
Micro-surface with Warranty	\$2.55/ yd ²	\$3.48/ yd ²	\$5.00/ yd ²
HMA ultra-thin with low volume warranty	\$2.50/ yd ²	\$2.58/ yd ²	\$3.00/ yd ²
HMA ultra-thin with medium volume warranty	\$2.55/ yd ²	\$2.58/ yd ²	\$2.60/ yd ²
Paver placed surface seal	\$5.19/ yd ²	\$5.74/ yd ²	\$6.00/ yd ²
Diamond grinding concrete pavement	\$2.90/ yd ²	\$3.14/ yd ²	\$3.18/ yd ²
Resealing transverse pavement joints with hot-poured rubber	\$0.95/ft	\$1.88/ft	\$4.00/ft
Resealing longitudinal joints with hot-poured rubber	\$0.96/ft	\$1.03/ft	\$2.40/ft
Crack sealing concrete pavement	\$0.10/ft	\$0.58/ft	\$2.47/ft
Yd ² = square yard with 1 inch depth, rbmi = roadbed mile, ft = foot			

Table 3.7 Typical material and total treatment costs for the state of Michigan (Baladi & Dean 2011)

Treatment type	Road type	
	Freeway	Non-Freeway
	Material cost only	
HMA Reconstruction	\$476,000/lane-mile	\$272,000/lane-mile
PCC Reconstruction	\$359,000/lane-mile	\$325,000/lane-mile
Rubblization & HMA overlay	\$238,000/lane-mile	\$263,000/lane-mile
Unbonded PCC Overlay	\$263,000/lane-mile	\$262,000/lane-mile
	Total cost including materials, engineering fees, culverts and slope work, etc.	
HMA Reconstruction	\$1,000,000/mile	\$1,200,000/mile
PCC Reconstruction	\$1,300,000/mile	-

CHAPTER 4

DATA ANALYSES & DISCUSSION

4.1 Introduction and Research Objectives

In this chapter, the analysis methods used in this research and the analysis results are presented and discussed. The analyses methods, results, and discussion are based on the problem statement, research plan, and objectives included in Chapter 1. For convenience, and to facilitate complete understanding, the problem statement and the research objectives are repeated below.

The long-term economic climate has caused some highway administrators to consider pavement performance data sampling as an alternative to reduce data collection costs. However the effects of data sampling on the accuracy of the PMS decisions are not well documented. It is not clear that data sampling ultimately reduces the overall PMS costs. Furthermore, for those SHAs that sample the pavement performance data, the pavement lengths for which the performance data are stored and analyzed are dictated by the sampling techniques used as stated in Chapter 1. Longer analysis lengths may not improve the data analysis results and may artificially decrease the data variability. Thus, the impact of the pavement analysis length on the variability of the pavement performance data and, ultimately, on the accuracy of the PMS decisions should be evaluated before a data analysis length is selected. Finally, for various reasons, data from some data collection cycles along certain pavement segments may not be documented or are missing, which makes it difficult or impossible to model and analyze the data. In order to solve this problem, the effects of different data imputation techniques on the accuracy of the pavement performance data should be evaluated. Accordingly, a comprehensive research plan was designed to address the three problems through the analyses of the impacts of data

sampling, data analysis length, and data imputation on the accuracy of the pavement performance data and the PMS decisions.

The specific objectives of the research are stated below.

1. Assess the impacts of pavement performance data sampling on the accuracy of PMS decisions and quantify the hidden costs of data sampling.
2. Analyze the effects of data analysis length on the variability of the pavement condition data and on the accuracy of the PMS decisions.
3. Study the impacts of various data imputation methodologies on the variability and accuracy of the pavement condition data and the corresponding PMS decisions.

4.2 Pavement Performance Data Sampling

As stated in Chapter 3, some SHAs collect pavement condition and distress data on a continuous basis and store the data over short pavement segments (0.1 mile, 0.5 mile, or 1 mile), whereas others use sampling techniques to reduce the time and costs of data collection. It is typically assumed that the pavement conditions of one mile, two miles, or longer pavement sections can be represented by the pavement conditions along shorter pavement segments such as 100, 200, or 500 ft. Some SHAs obtain electronic images of the pavement surface on a continuous basis along the entire network but only digitize a small “representative” sample of the data. The assumption herein is that digitized sample of the pavement performance data represents the conditions of a longer pavement section. Regardless, pavement performance data sampling is typically accomplished using either randomly selected short pavement segments or physically fixed short pavement segments.

In this research, time series pavement condition and distress data that were continuously collected along four road networks were obtained. The four sources of the data (see Chapter 3)

are the Colorado Department of Transportation (CDOT), the Louisiana Department of Transportation and Development (LADOTD), the Michigan Department of Transportation (MDOT), and the Washington State Department of Transportation (WSDOT). The pavement condition data of the four SHAs are reported for each 0.1 mile segment along their pavement networks. The International Roughness Index (IRI) and the transverse and longitudinal cracking data received from the four SHAs were used to analyze the effects of sampling and sample size on the accuracy of the pavement management decisions. This was accomplished by sampling the continuous data along several miles of various pavement sections using two sampling techniques; fixed and random sampling. The effects of sample size on the PMS decisions were studied using the fixed sampling technique and the sample size varied from 10 to 60 percent using 10 percent increments. The six sample sizes were obtained as follows:

1. Ten percent sampling: the pavement condition and distress data of the first 0.1 mile long pavement segment of each one mile.
2. Twenty percent sampling: the pavement condition and distress data of the first two 0.1 mile long pavement segment of each mile.
3. Thirty percent sampling: the pavement condition and distress data of the first three 0.1 mile long pavement segments of each mile.
4. Forty percent sampling: the pavement condition and distress data of the first four 0.1 mile long pavement segments of each mile.
5. Fifty percent sampling: the pavement condition and distress data of the first five 0.1 mile long pavement segments of each mile.
6. Sixty percent sampling: the pavement condition and distress data of the first six 0.1 mile long pavement segments of each mile.

For each sample size, it was assumed that the pavement condition and distress of the sampled pavement segments represent the pavement condition and distress of the entire mile. In the random sampling technique, for each mile of pavement, the pavement conditions and distress of a randomly selected 0.1 mile long pavement segment along each mile were assumed to represent the pavement conditions and distress of the entire mile.

4.2.1 Fixed Sampling

Recall that CDOT, LADOTD, MDOT, and WSDOT report and store the measured IRI and pavement distress data for each 0.1 mile long pavement segment along the road network. Such data are labeled herein as continuous data. Using the continuous pavement condition and distress data from the four states, the effects of 10 percent sampling on the accuracy of the pavement distress data were analyzed. In this manner, for each data collection cycle, the pavement data of the first 0.1 mile long pavement segment of each 1 mile road section were assumed to represent the pavement conditions and distress along the entire mile. For each analyzed road in the four SHAs, the same analysis steps were used and they are outlined below.

Step 1 – Differences Between the Sampled and the Continuous Data – In this step, the differences between the ten percent sampled and the continuous pavement condition and distress data were studied. Recall that the pavement condition and distress data along the first 0.1 mile long pavement segment of each one mile pavement section were assumed to represent the conditions and distress along the entire mile. Hence, the sampled data consist of one data point for each one mile of road whereas the continuous data provide ten data points per mile. The solid lines in Figures 4.1 and 4.2 depict, respectively, the sampled IRI and transverse cracking data while the individual data points in the figures represent the continuous IRI and transverse cracking data for each 0.1 mile long pavement segment of the road. It is important to note that, in

most SHAs, the longitudinal and transverse pavement profiles along the road network are measured by the same vehicle. The two profiles are automatically analyzed and the IRI and rut depth data are recorded for each road survey segment within the road network. Hence, most SHAs do not sample IRI or rut depth data. The IRI data are used herein to show the effects of sampling on sensor-collected data whereas the transverse cracking data are used to show the impact of sampling on image based data.

The data in Figure 4.1 indicate that the higher is the variability of the measured IRI data, the higher are the differences between the sampled and the continuous IRI data. To illustrate, the degree of uniformity of the continuous IRI data in Figure 4.1 varies substantially amongst the four roads as does the degree at which the sampled data represent the continuous data as stated below.

1. The degree of uniformity of the continuous IRI data in Figure 4.1d for the twenty miles long composite pavement section along SRID 005 in Washington is the highest amongst the four roads in Figure 4.1. Likewise, the sampled data represent the continuous data throughout the twenty miles section of the road.
2. Except for about eleven 0.1 mile long pavement segments (localized segments), the degree of uniformity of the continuous IRI data in Figure 4.1a for the thirteen miles long flexible pavement section along HWY 24 in Colorado is very good. Likewise, the sampled data represent the continuous data throughout the thirteen miles section except at the localized segments of the road. At some of these localized areas, the continuous data are more than four times higher than the sampled data.
3. The uniformity of the continuous IRI data for the eighteen miles long composite pavement section along I-94 in Michigan is similar to that of HWY 24 in Colorado as

shown in Figure 4.1c. The exception is that more localized data variability can be found along I-94 than along HWY 24. Likewise, the sampled data represent the uniform continuous data throughout the eighteen miles section except at the localized areas where the data variability is high. At some of these areas, the continuous data are more than twice and less than half of the sampled data.

4. Figure 4.1b shows high variability of the continuous IRI data of the 11 miles long flexible pavement section along LA-34 in Louisiana. The figure also shows that the sampled data do not represent the continuous data along the majority of the eleven mile road section.

Similar results were obtained for other roads and are shown in Figures B.1 through B.4 of Appendix B.

Figure 4.2 depicts the sampled and the continuous transverse cracking data along the same four road sections shown in Figure 4.1. The data in Figure 4.2 indicate that the variability of the continuous transverse cracking data is much higher than the IRI data of Figure 4.1. Consequently, the sampled data in Figure 4.2 often misrepresent the continuous data for almost all roads that were analyzed. For example, the sampled transverse cracking data for HWY 24 in Colorado (Figure 4.2a) either significantly overestimate or underestimate the continuous data for almost each mile point. This scenario is substantially magnified along LA-34 in Louisiana where the continuous transverse cracking data are highly variable as shown in Figure 4.2b. At some locations, the continuous transverse cracking data indicate 0 feet of cracking while the sampled data indicate 1000 feet of cracking. In a few cases, though, the differences between the sampled and the continuous transverse cracking data are not significant. This can be seen from mile points 8 to 18 along I-94 in Michigan (Figure 4.2c) as well as from mile points 62 through 70 and 72

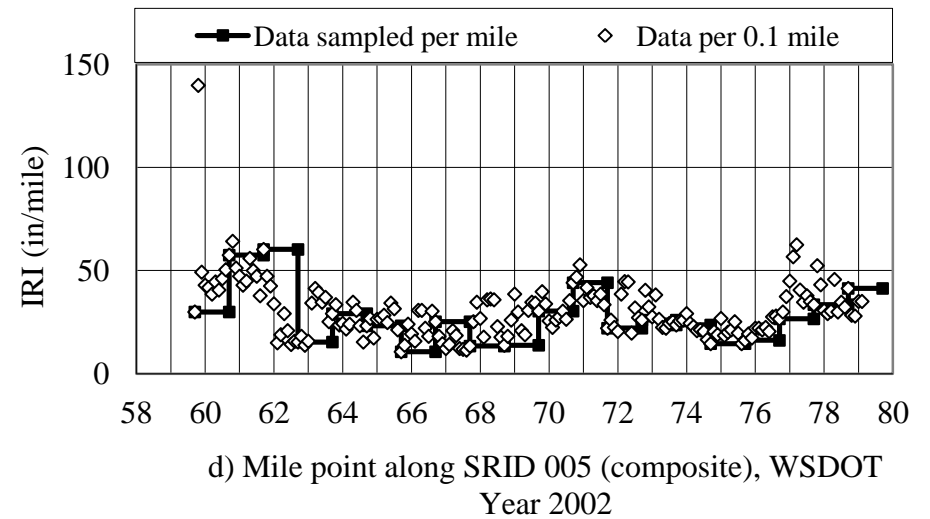
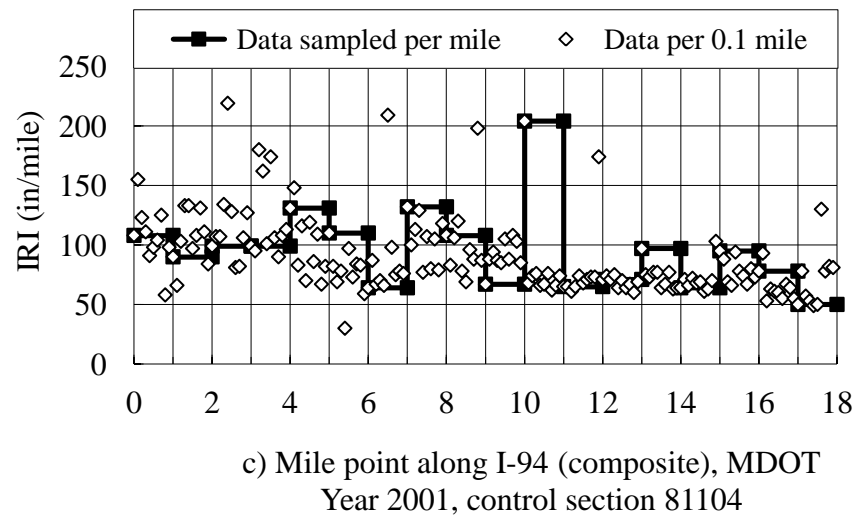
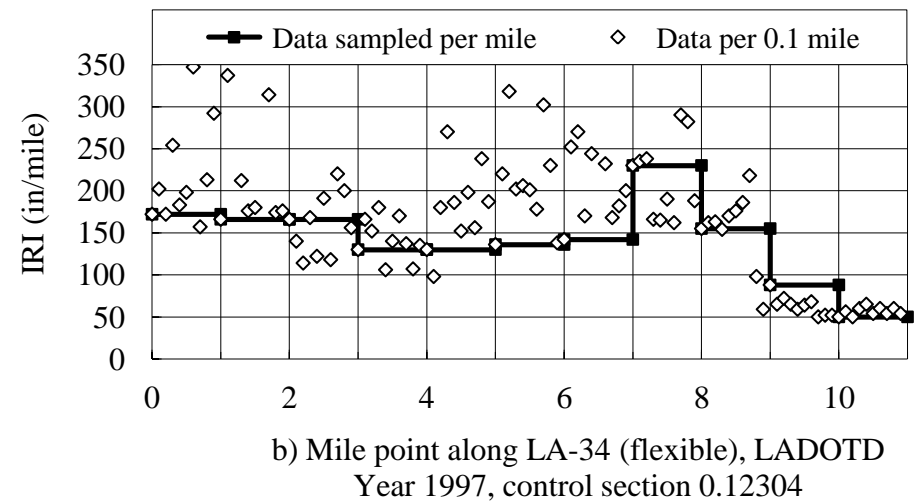
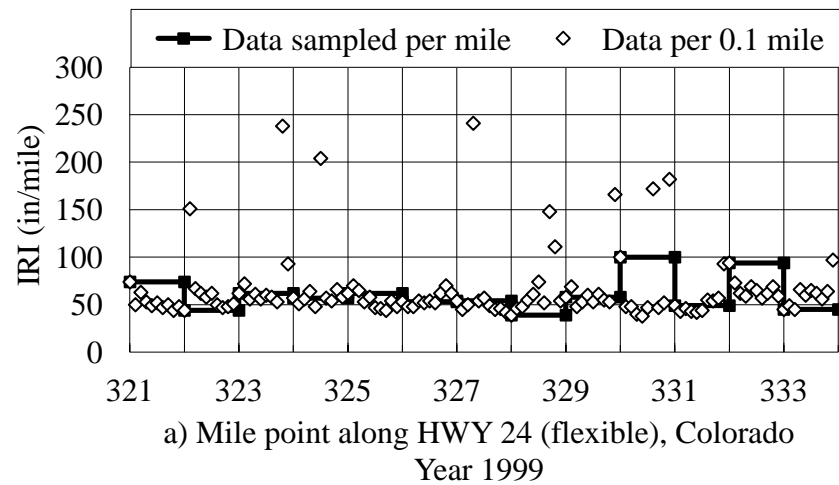


Figure 4.1 Continuous and sampled IRI data along four roads in four different states

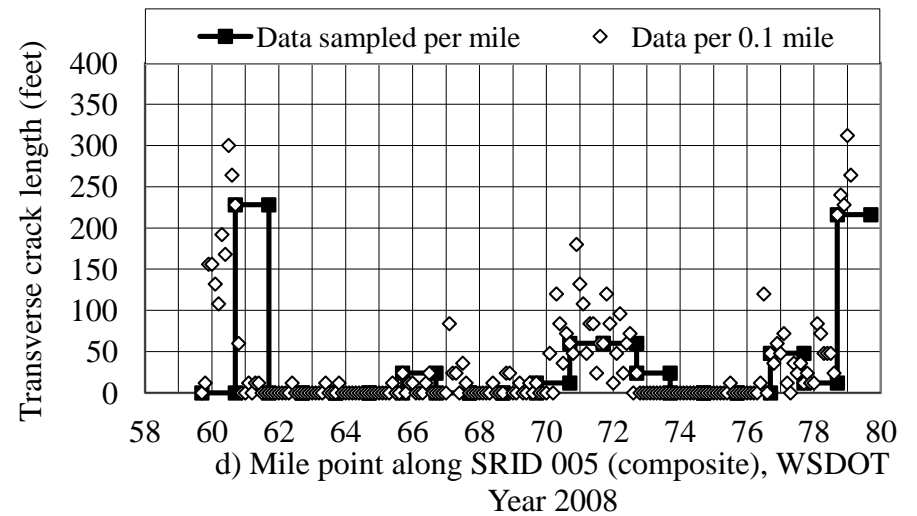
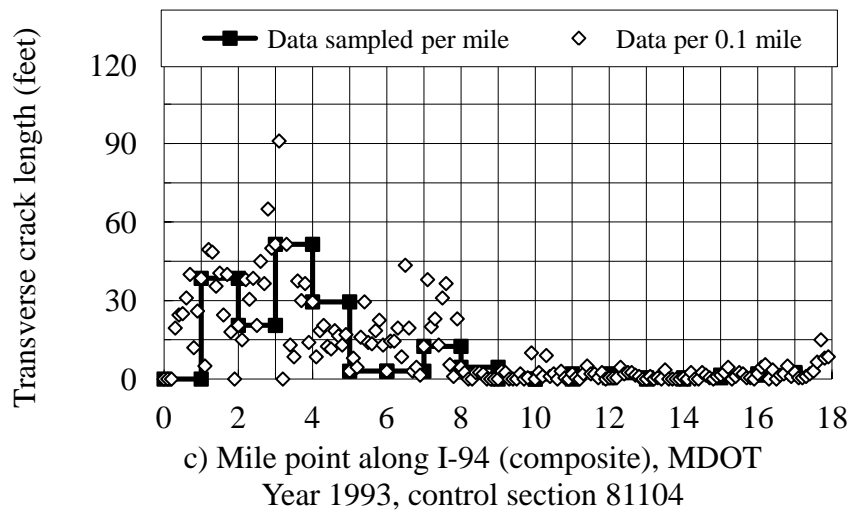
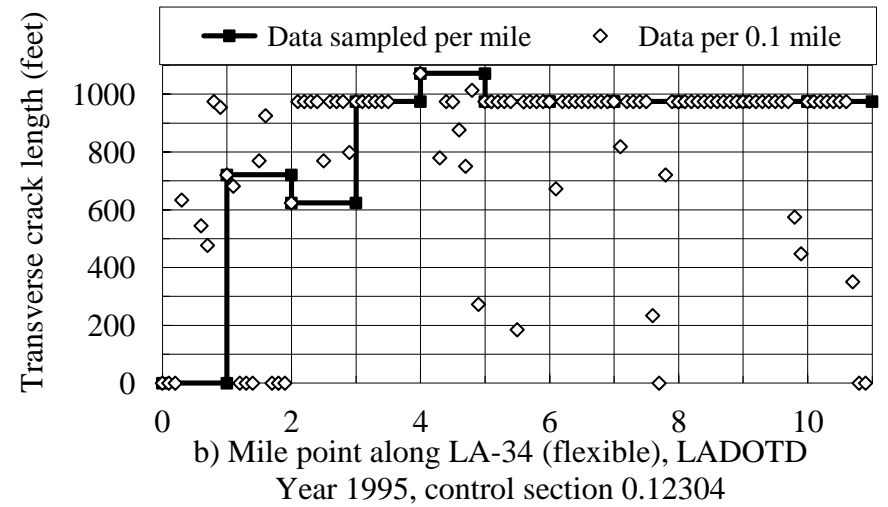
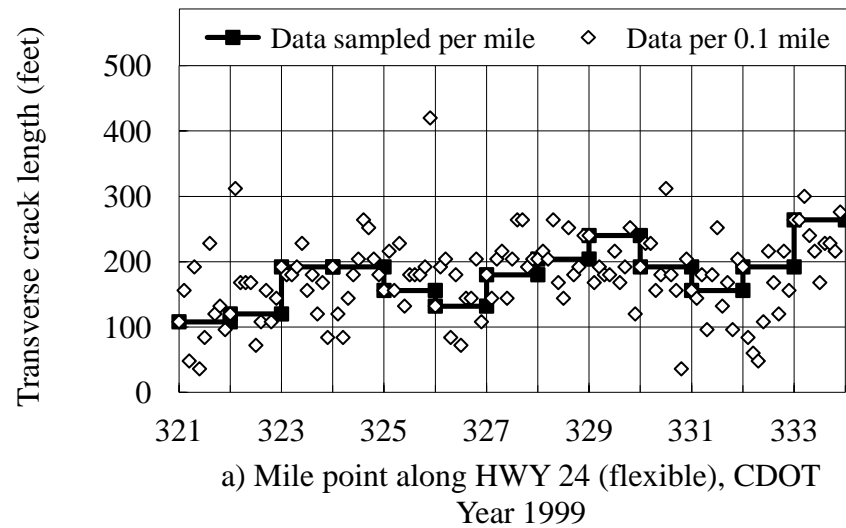


Figure 4.2 Continuous and sampled transverse crack length data along four roads in four different states

through 76 along SRID 005 in Washington (Figure 4.2d). Nevertheless, similar results were obtained for other roads and are shown in Figures B.5 through B.8 of Appendix B.

The above discussion regarding the sampling of the IRI and transverse cracking data indicates that:

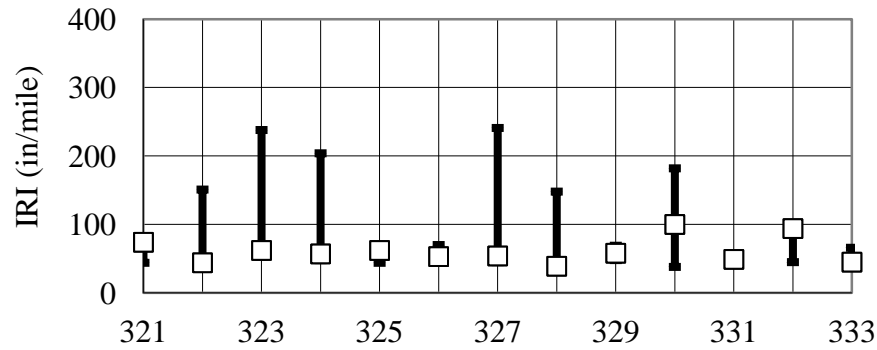
1. For uniform pavement condition and distress (such as in a newly constructed or rehabilitated pavement sections), the sampled data accurately represent the continuous data. Otherwise, when the pavement condition and distress are highly variable, the sampled data significantly misrepresent the pavement condition and distress.
2. In general, ten percent sampling appears to represent the sensor-based continuous data more accurately than the image-based data (cracking). This is no surprise given the methods that are used to collect the data. The IRI data are collected by laser sensors using fully-automated techniques. On the other hand, the cracking data are collected using semi-automated techniques that require the input of trained SHA or contractor employees to review and digitize the video images of the pavement surface. Thus, human subjectivity impacts the cracking data as one employee may view crack lengths and severity levels different than another employee. As a result, the variability of the cracking data is relatively high and consequently, the ability of ten percent sampling to represent the continuous pavement distress is questionable.

Step 2 – One Year Data Variability/Range – In this step, for a given data collection year, and for each one mile section along the road, the sampled and the range of the continuous IRI and transverse cracking data were compared and studied. Figures 4.3 and 4.4 depict the sampled and the range of the continuous IRI and transverse cracking data, respectively. In the two figures, the sampled data for each one mile of the four roads are represented by a square

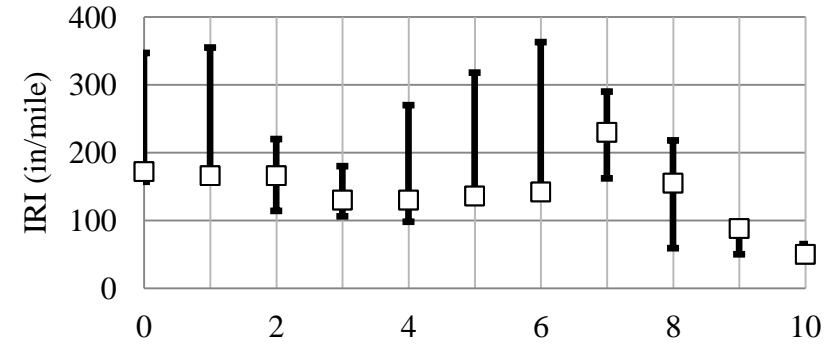
whereas the range of the continuous data along each mile is represented by the solid straight lines. The intent of these figures is to convey the fact that the ten percent sampled data may be at the low or high end or at the average of the continuous data. Stated differently, the sampled data could underestimate or overestimate the continuous data or the true pavement conditions and distress. For example, Figure 4.3a shows that for most mile points along HWY 24 in Colorado, ten percent sampling results in underestimation of the continuous IRI data. That is, the sampled data fall on the low end of the continuous data range. A similar scenario can be seen in Figure 4.3b for the IRI data along LA-34 in Louisiana. On the other hand, the sampled data along I-94 in Michigan appear to underestimate the IRI at some mile points and overestimate the data at other mile points.

Perhaps the results are even more evident for the cracking data. Considering the data shown in Figure 4.4b, it can be seen that the ten percent sampled transverse cracking data could be as little as 0 feet of cracking for one mile of road while the continuous transverse cracking data show a range from zero to 1500 feet for the same one mile pavement section. The reverse scenario can also be seen in the same figure for the same road. The sampled data indicate that approximately 1000 feet of transverse cracking is present along one mile of LA-34 starting at mile point 7. However, the continuous data show that transverse cracking varies from zero to 1000 feet along the same one mile of road. The above discussion indicates that ten percent sampling may significantly underestimate or overestimate the true pavement conditions and distress.

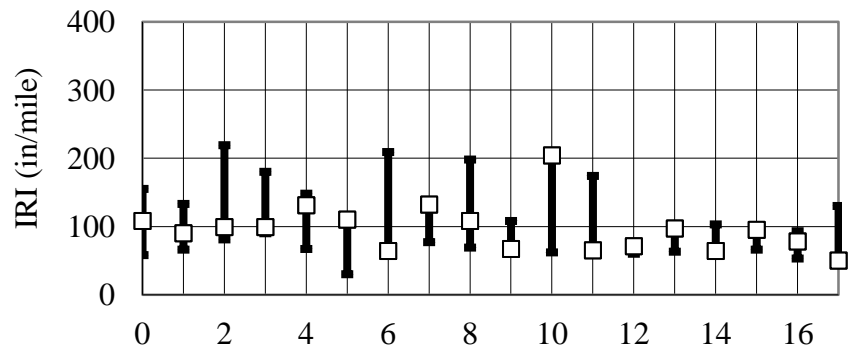
The data in Figures 4.3 and 4.4 also indicate that data variability impacts the degree of accuracy at which ten percent sampled data represent the continuous data. In the figures, the length of each solid line represent the range or the variability of the ten data points along each



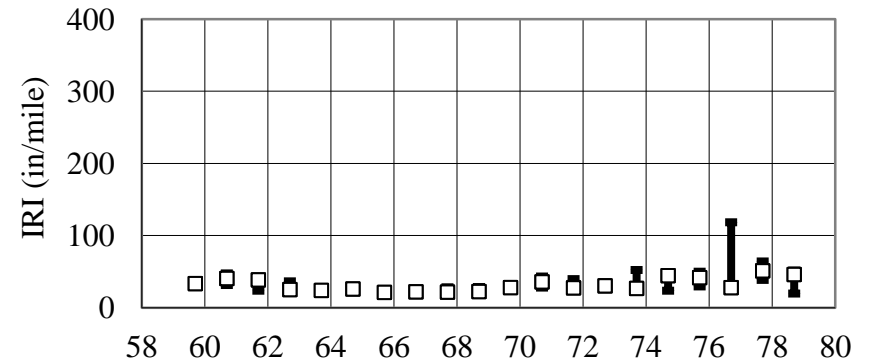
a) Mile point along HWY 24 (flexible), CDOT
Year 1999



b) Mile point along LA-34 (flexible), LADOTD
Year 1997, control section 0.12304



c) Mile point along I-94 (composite), MDOT
Year 2001, control section 81104



d) Mile point along SRID 005 (composite),
WSDOT Year 1999

Figure 4.3 Sampled (squares) and the range of the continuous IRI data for each mile

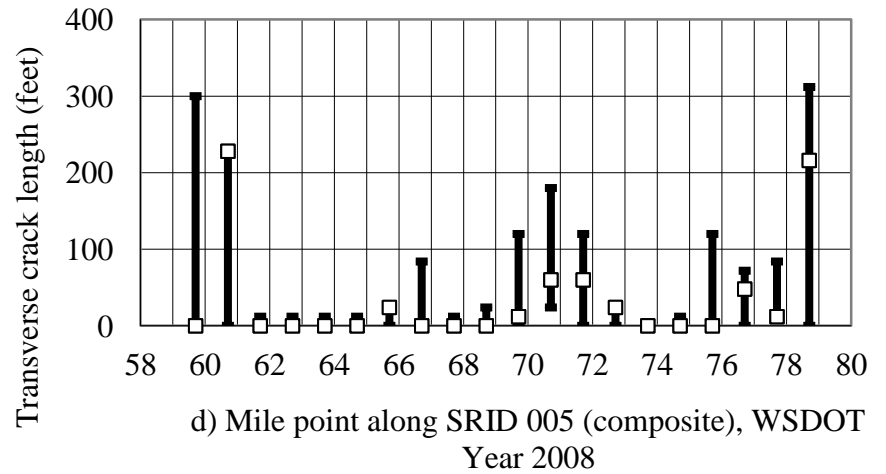
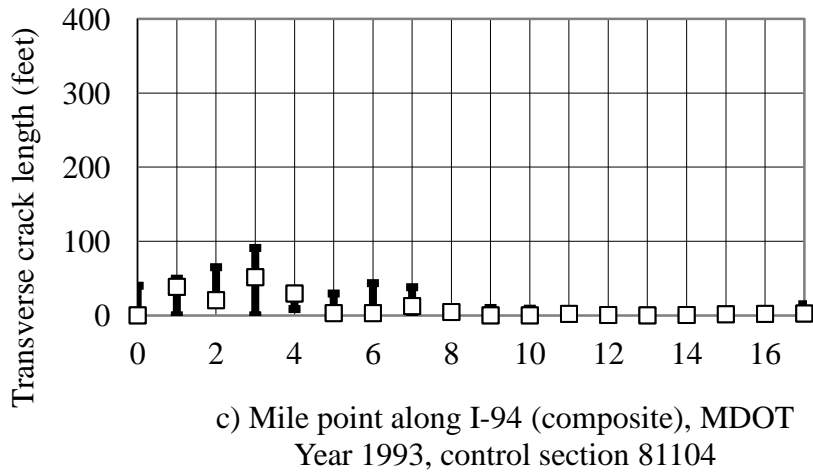
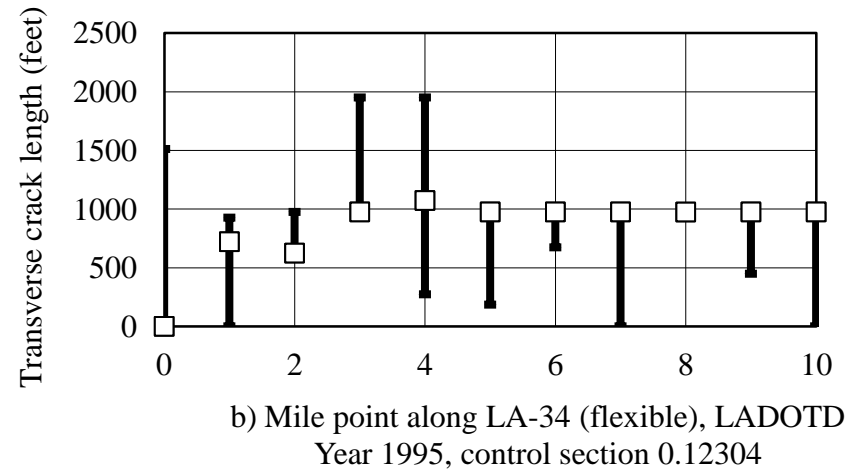
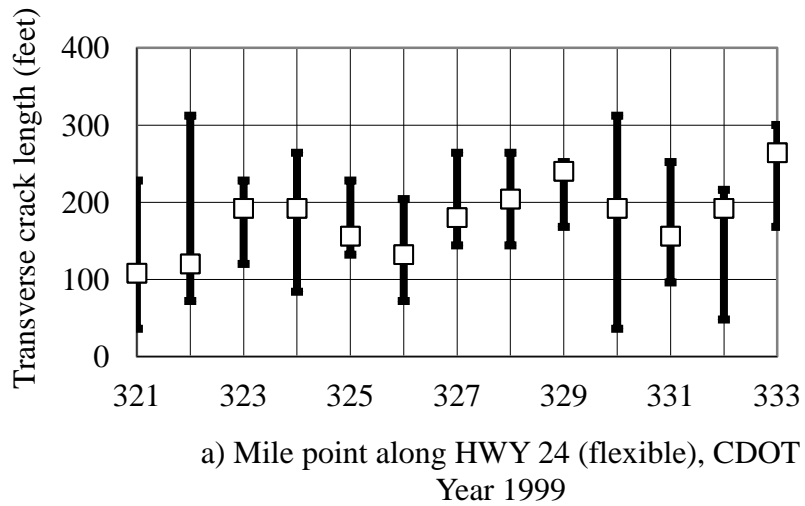


Figure 4.4 Sampled (squares) and the range of the continuous transverse cracking data for each mile

one mile of road. Thus, shorter solid lines indicate smaller range or lower variability and longer lines indicate greater range or higher variability of the continuous data. The squares in the figures, on the other hand, represent the sampled data. It can be seen that the sampled data (the squares) represent the continuous data when the solid lines are short or the variability of the continuous data is low. The above observations can be clearly seen in Figure 4.4d for the transverse cracking data along SRID 005 in Washington. Smaller ranges of the transverse cracking data such as at mile points 62 through 66 yield sampled data that are representative of the continuous data. Wider ranges of the transverse cracking data such as at mile points 60, 61, and 79 yielded sampled data that are not representative of the continuous data or the true distribution of the pavement conditions and distress. Similar results were found for other roads and are shown in Figures B.9 through B.16 of Appendix B.

Step 3 – Time Series Data Variability - The previous two steps addressed the differences between the samples and the continuous IRI and transverse cracking data for a given year. In this step, the effects of sampling on the differences between the continuous and the sampled data were analyzed using the time series (multi-year) data for various roads from the four SHAs. To accomplish this, the IRI and the transverse cracking data were sampled for each data collection cycle or distress survey year. Similar to the previous two steps, for each year of data and for each one-mile pavement section, the measured data along the first 0.1 mile long segment was assumed to represent the pavement conditions along the entire one mile long section. For each one mile pavement section and for each data collection year, the sampled data were divided by itself producing the constant 100 percent line along the entire road. The continuous data for each 0.1 mile long pavement segment within each one mile section was also divided by the sampled data of that mile. This yielded the data points shown by the various

symbols in the figures. Each data symbol represents one data collection year as indicated in the legend of each figure. Four examples of the results are shown in Figures 4.5 through 4.8.

The data in Figures 4.5 and 4.6 indicate that, for some miles along the roads, the continuous IRI data may be more than 200 percent higher than the sampled data. For instance, the continuous IRI data along mile point 321 of I-94 in Michigan (Figure 4.5) range from approximately 120 percent to nearly 240 percent the sampled data. The data along mile point 330 of the same road can be as much as 250 percent the sampled data. Similar results can be seen for various mile points along the roads in the other states. Regardless, in general, it can be stated that most continuously measured IRI data lie within ± 50 percent of the ten percent sampled data.

The data in Figures 4.7 and 4.8 show similar results for the transverse cracking data. However, the differences between the continuously measured transverse cracking data relative to the sampled data increase significantly. In some cases along each of the roads, the continuous data may be more than 1000 percent greater than the sampled data. In fact, as it can be seen for I-69 in Michigan shown in Figure 4.7, the continuous transverse cracking data may be as much as approximately 9500 percent greater than the sampled data. Regardless, despite the significant variation for some locations along each of the roads, it can be stated that, in general, most continuously measured transverse cracking data are within ± 100 percent of the ten percent sampled data. It should be noted that the above discussion addresses two roads from two states for IRI and two roads for the other two states for transverse cracking. The main reason is the size of the figures and to avoid unnecessary repetition. Nevertheless, similar results for other roads were obtained and are shown in Figures B.17 through B.48 in Appendix B.

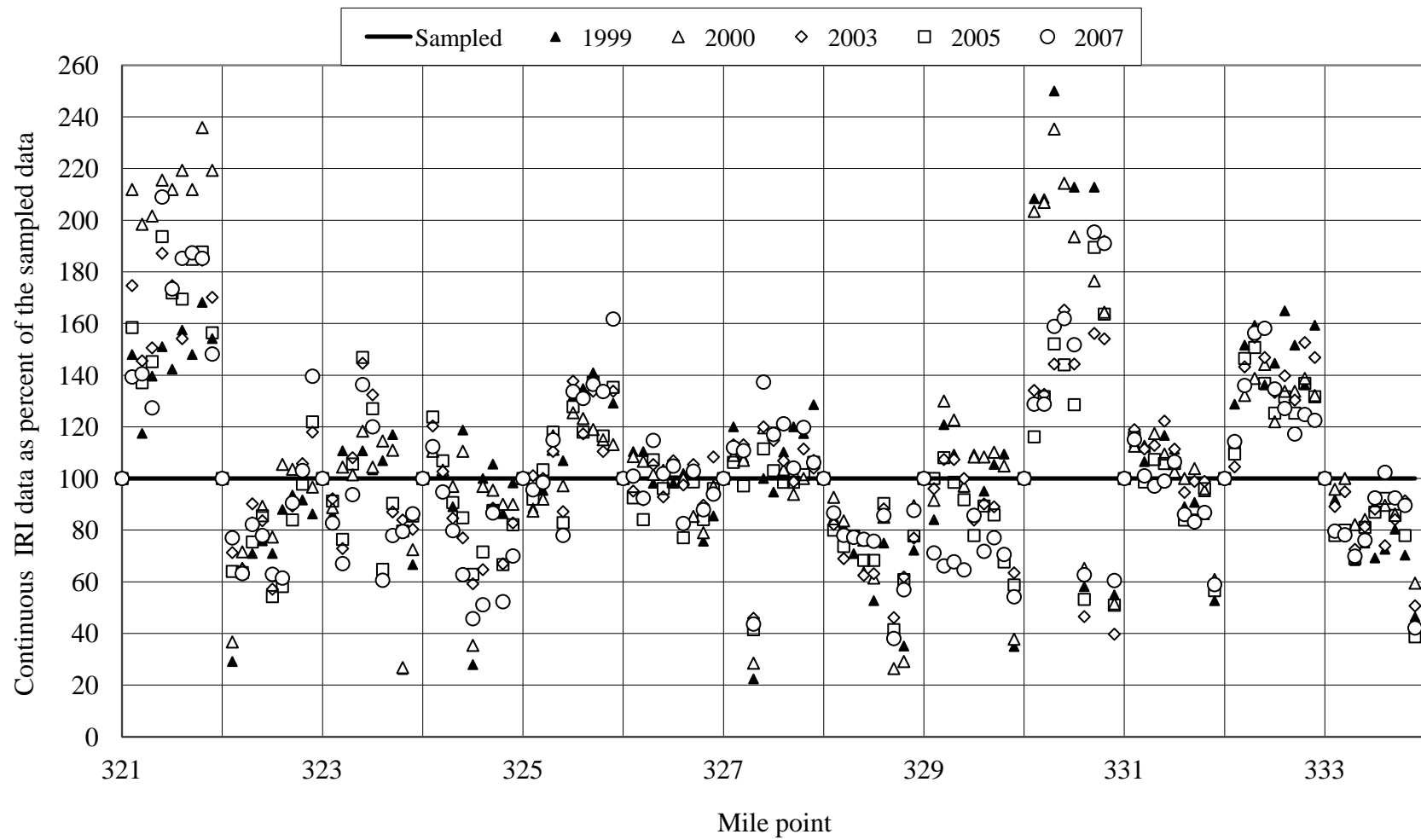


Figure 4.5 Continuous time series IRI data as percent of the sampled data along a portion of the flexible HWY 24 in Colorado

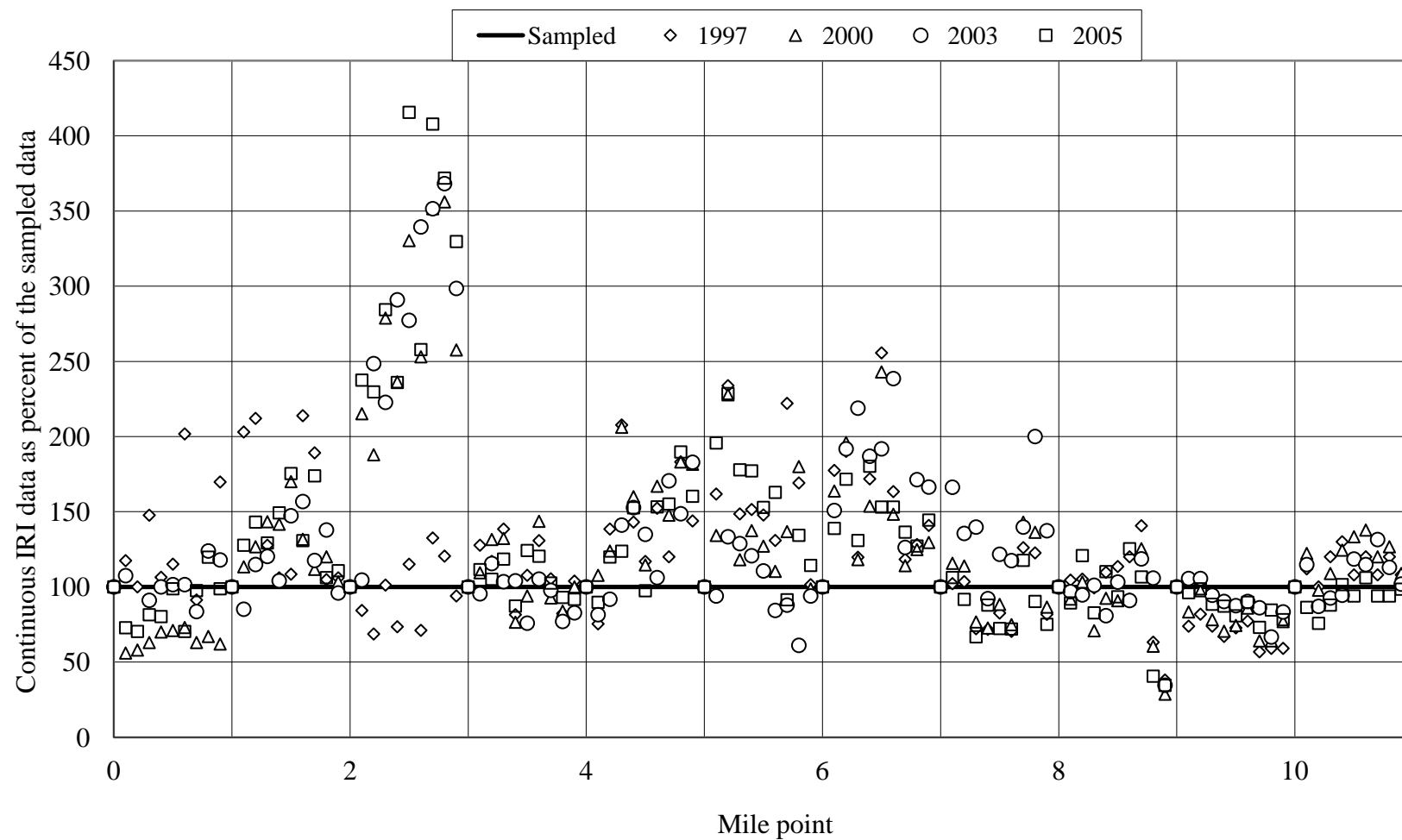


Figure 4.6 Continuous time series IRI data as percent of the sampled data along a portion of the flexible LA-34 (control section 0.12304) in Louisiana

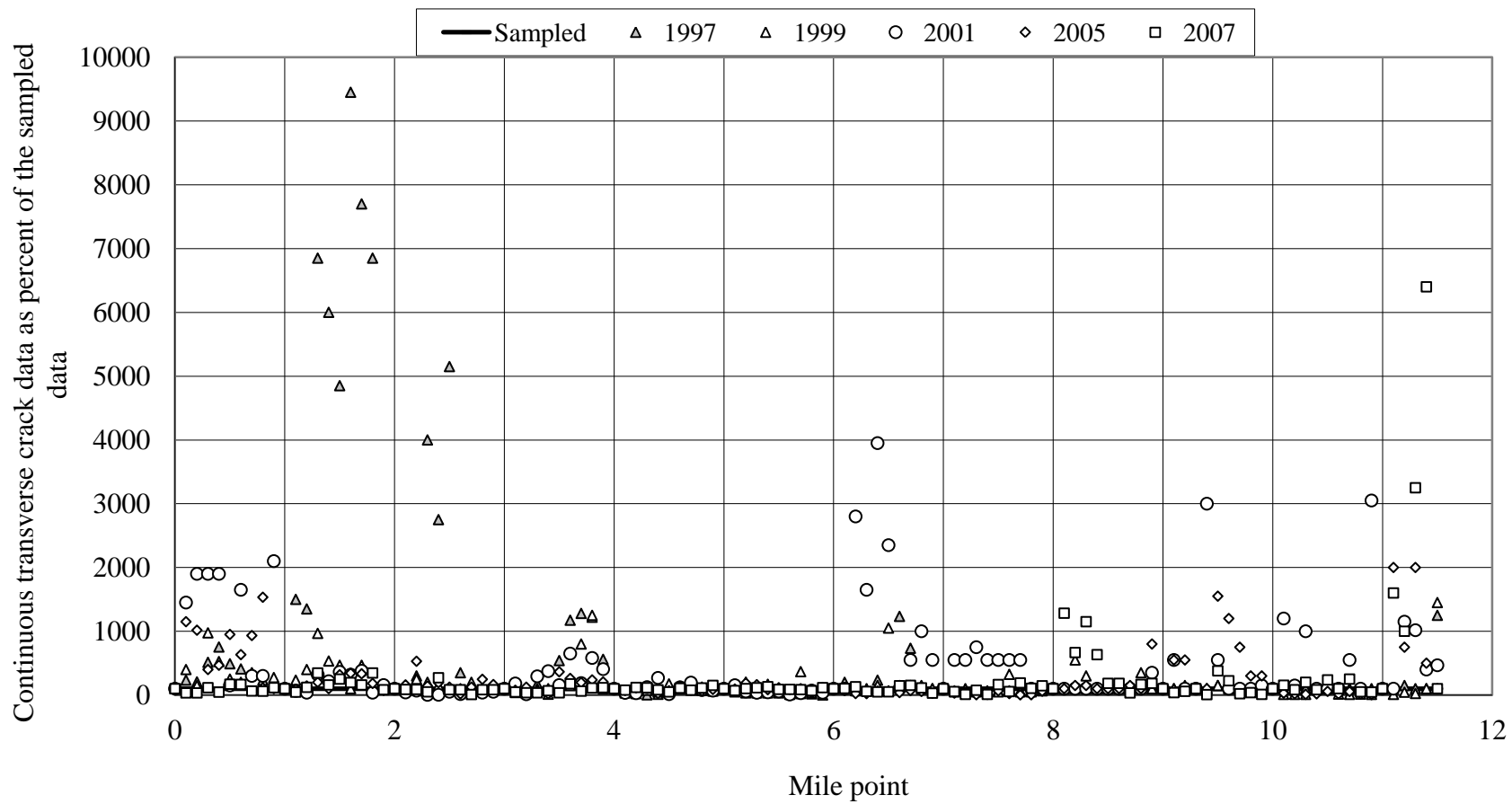


Figure 4.7 Continuous time series transverse crack data as percent of the sampled data along a portion of the rigid I-69 (control section 77024) in Michigan

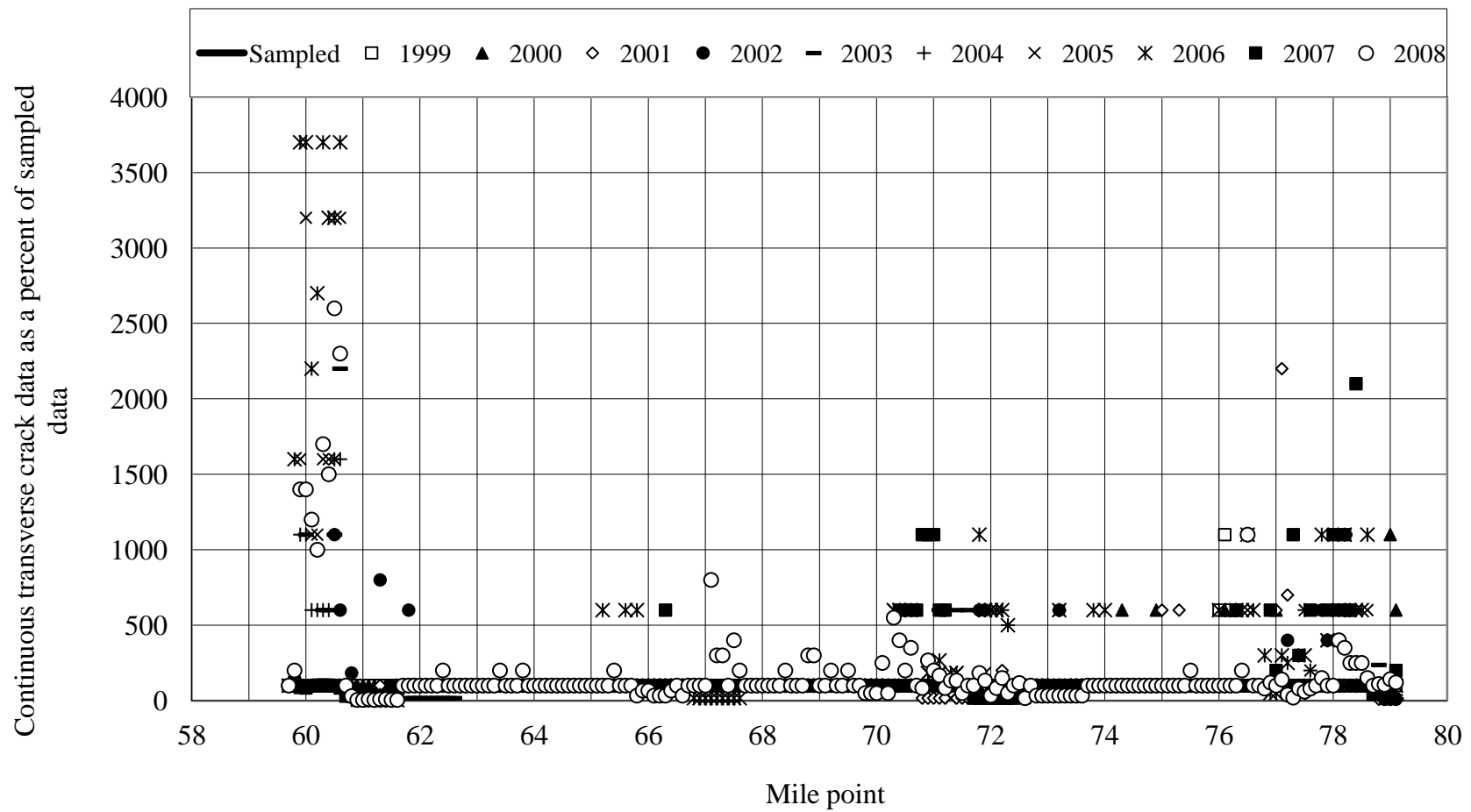


Figure 4.8 Continuous time series transverse crack data as percent of the sampled data along the composite SRID 005 in Washington

Furthermore, one would expect the degree at which the ten percent sampled data represent the continuous data to decrease as the pavement deteriorates over time without any applied treatments. That is, the pavement condition and distress data increase and show higher variability along the road as the pavement ages. Thus, the ten percent sampled data would likely become less representative of the continuous data as shown in the previous sections. In order to prove this point, the impacts of ten percent sampling on the historical time-series transverse cracking data were also analyzed using the percent differences between the continuous and sampled data relative to the sampled data of each pavement section. The maximum, minimum, and average percentages were calculated and plotted for each data collection year and for various roads in each state as shown in Figures 4.9 through 4.12.

The data in Figure 4.9a, for HWY 71 in Colorado, show that the range of the percent difference between the continuous and the sampled transverse crack data increases from year 1999 to 2000 but decreases from year 2000 to 2001. From year 2001 to 2002, the range of the percent difference increases again before decreasing significantly in year 2003. This can be explained using the CDOT maintenance records. The records indicate that crack sealing treatments were applied in 2002. Hence, the range of the percent difference between the continuous and the sample data decreases in year 2003 (CDOT does not count sealed transverse cracks). Stated differently, the crack sealing treatment in 2002 decreased the variability of the transverse cracking data and hence, the differences between the sampled and continuous data also decreased. It should be noted that the reduction in the range of the percent differences between the sampled and the continuous data from 2000 to 2001 could be due to either data variability or a treatment that was applied but not documented.

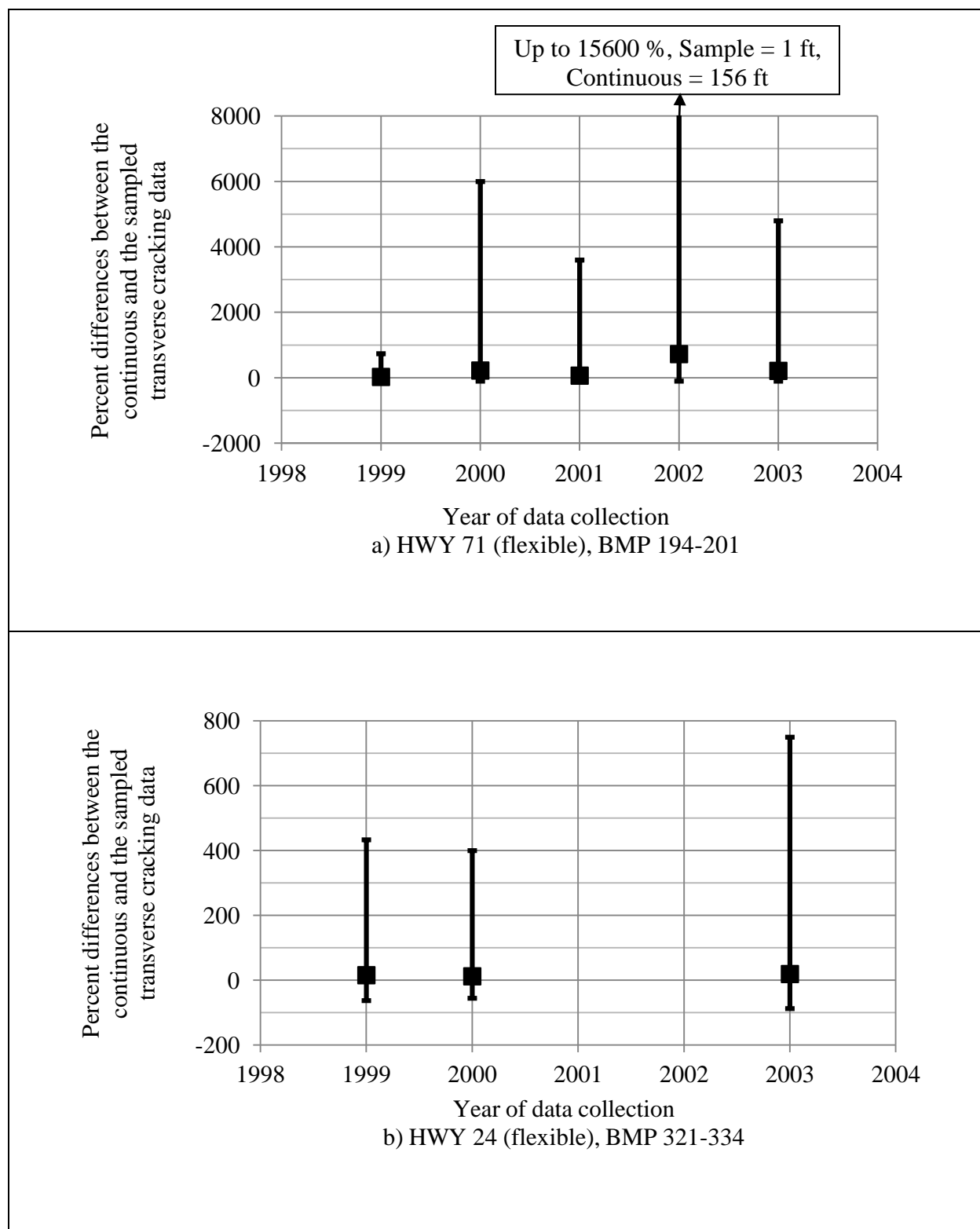


Figure 4.9 Maximum, minimum, and average percent differences between the continuous and the sampled transverse cracking, relative to the sampled data, Colorado

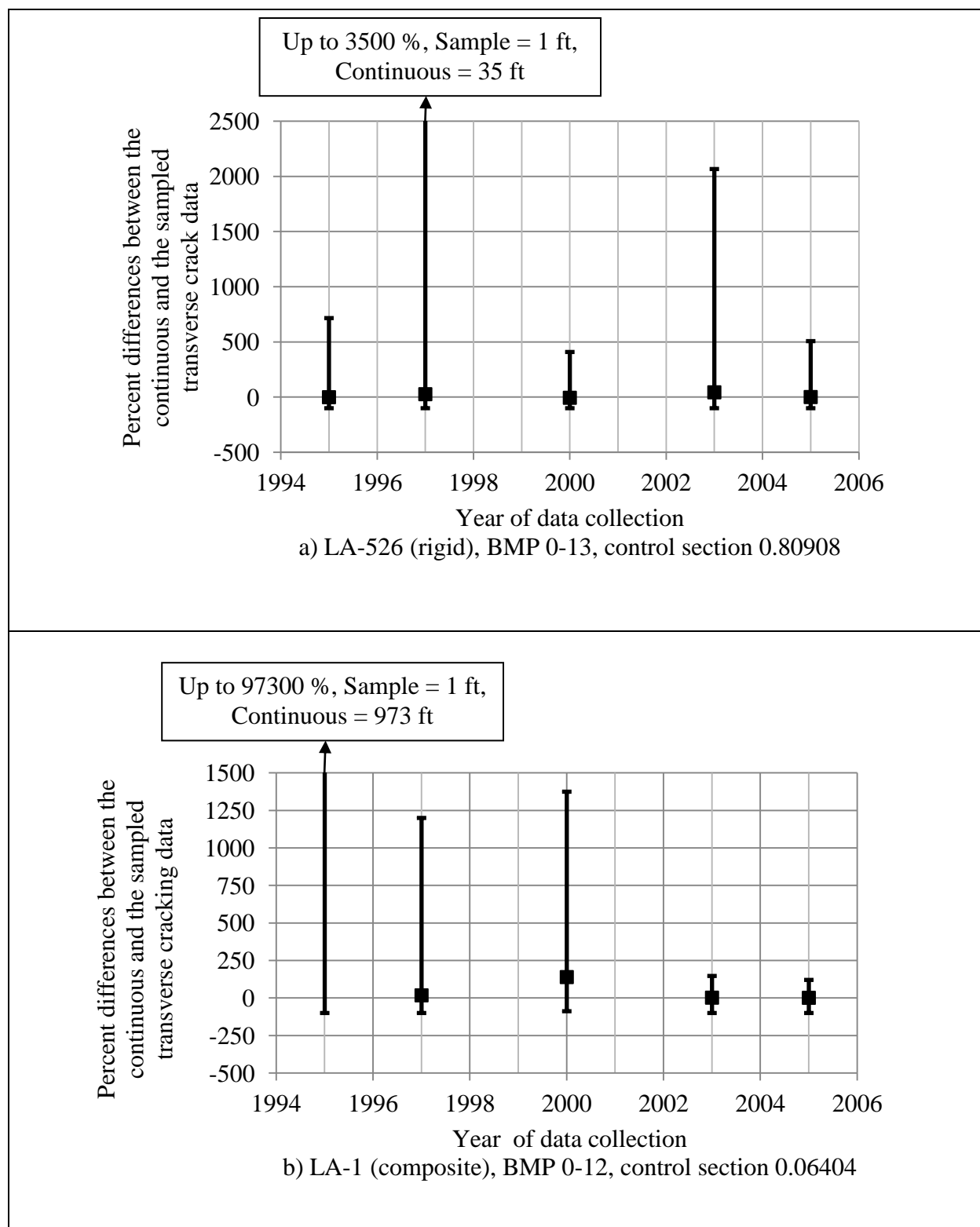


Figure 4.10 Maximum, minimum, and average percent differences between the continuous and the sampled transverse cracking, relative to the sampled data, Louisiana

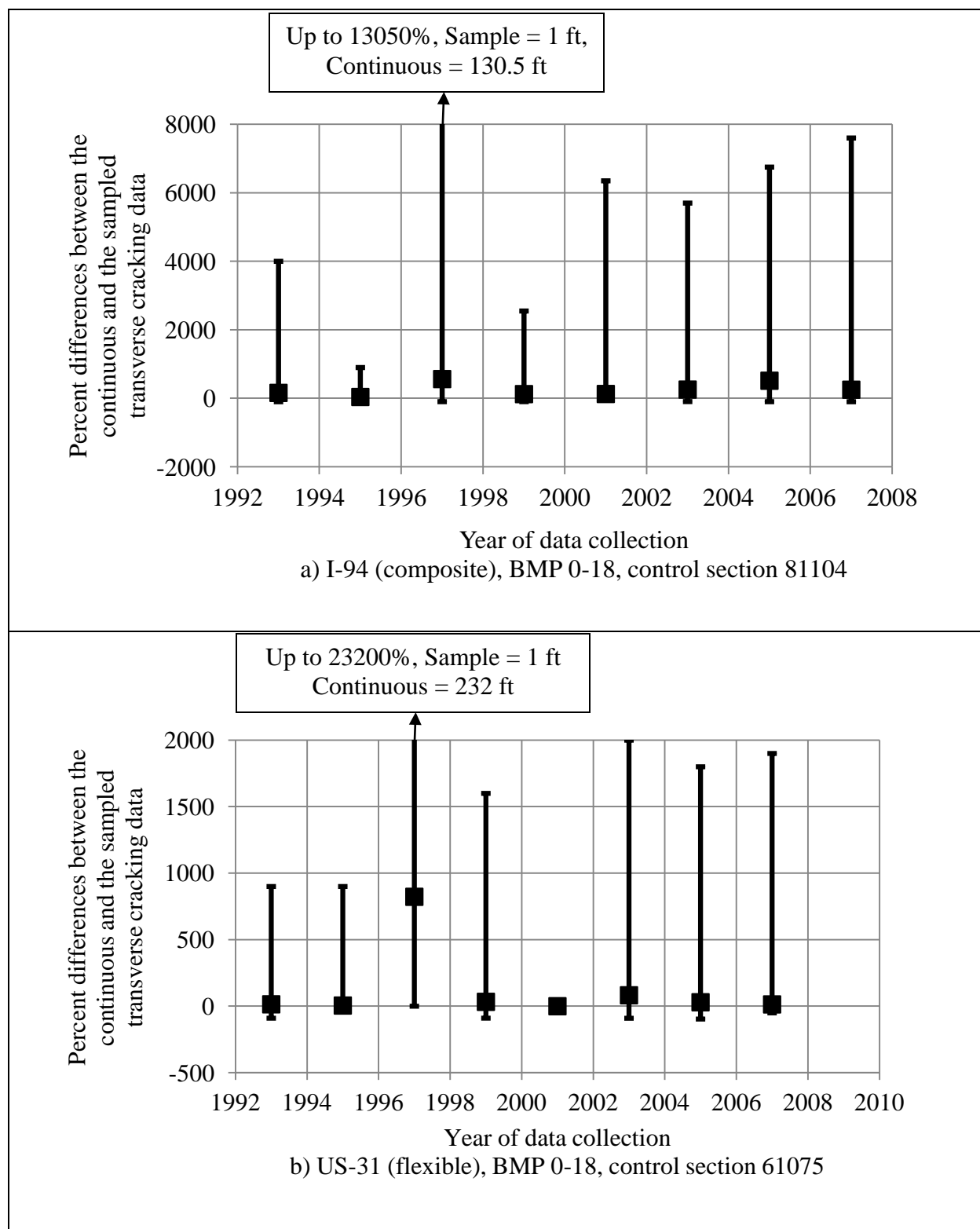


Figure 4.11 Maximum, minimum, and average differences between the continuous and the sampled transverse cracking, relative to the sampled data, Michigan

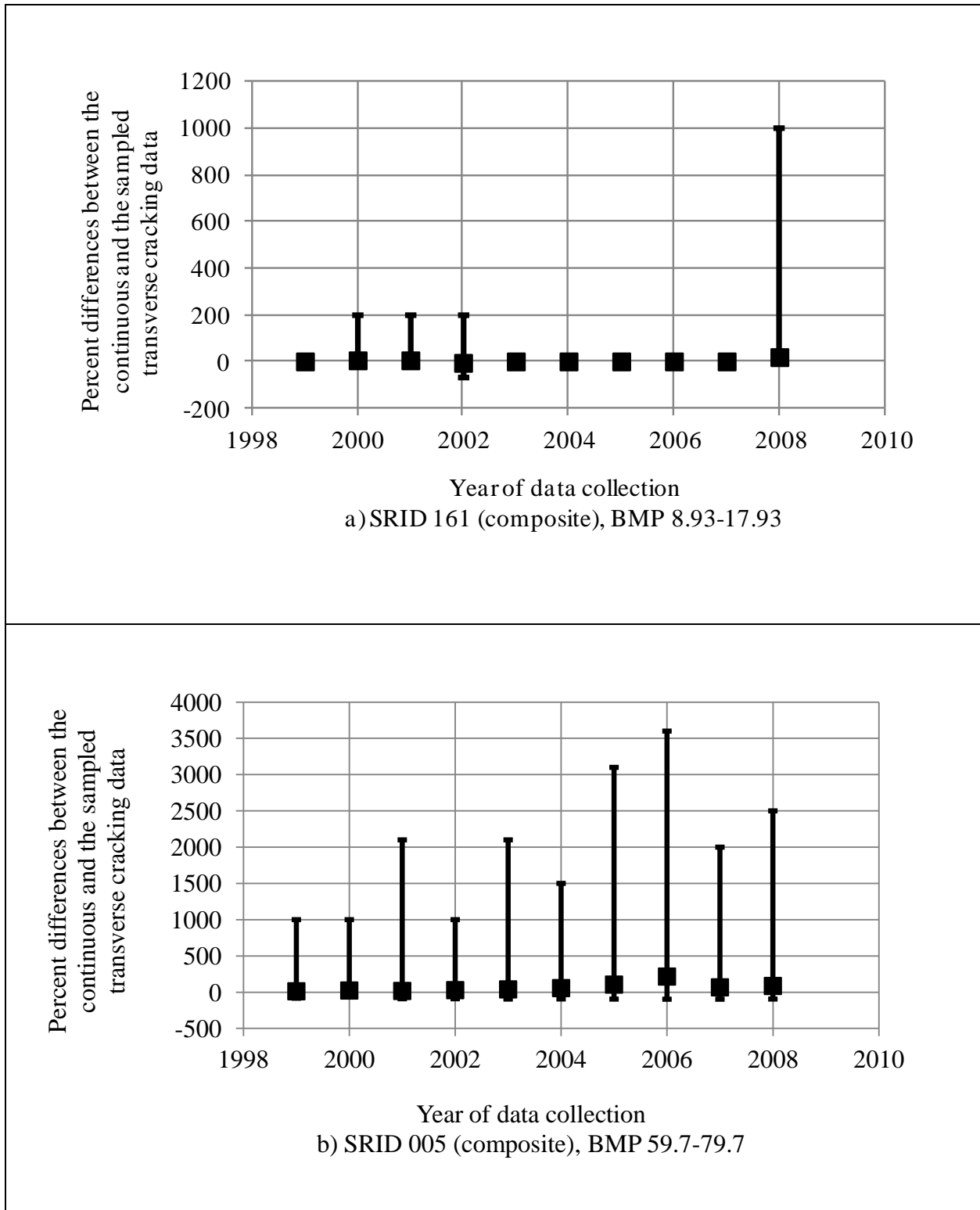


Figure 4.12 Maximum, minimum, and average percent differences between the continuous and the sampled transverse cracking, relative to the sampled data, Washington

No maintenance records were available for HWY 24 in Colorado (Figure 4.9b); however, it is likely that no treatments were applied between 1999 and 2003 (no distress data were collected in the years 2001 and 2002). This is evidenced by the fact that the range of the percent differences between the continuous and the sampled data increases from 1999 to 2003. Furthermore, the average transverse crack length increases from approximately 177 feet in 1999 to nearly 231 feet in 2003. Thus, the transverse cracks are becoming more variable over time and the range of the percent differences increases.

Similarly, the data in Figure 4.10a for LA-526 in Louisiana can also be explained using the maintenance records of the LADOTD. The records indicate that concrete pavement patching treatments were applied in 1998, 1999, 2000, and 2005. Hence, the decreases in the range of percent differences between the continuous and the sampled data from 1997 to 2000 as well as from 2003 to 2006 coincide with the reduction in pavement distress variability due to the patching treatments. Stated differently, as the transverse cracking becomes more uniform (due to pavement treatments), the sampled data more accurately represent the continuous data.

No maintenance records were available for LA-1 (Figure 4.10b) but it is likely that some type of pavement treatment was applied between years 2000 and 2003. This is because the range of the percent differences between the continuous and the sampled data increase from 1997 to 2000 but shows a significant decrease in 2003. Maintenance records were also not available from MDOT for I-94. However, the data in Figure 4.11a appear to correspond to pavement treatments between 1993 and 1995 as well as between 1997 and 1999. Further, the ten percent sampled data do not accurately represent the continuous data along the newly reconstructed section of US-31 in Michigan shown in Figure 4.11b. The premature transverse cracking data along the road are highly variable with an anomaly in 1997.

Finally, the data in Figure 4.12a show little to no percent differences between the continuous and the sampled data from years 2003 to 2007 along SRID 161 in Washington. The maintenance records from WSDOT show that in 2004 and 2005 the existing roadway was resurfaced and planed. It is possible that this work could have begun in 2003 and was not documented until 2004 and 2005. Thus, the variability of the transverse cracking data was reduced causing the percent differences between the continuous and the sampled data to decrease.

WSDOT maintenance records do not show any treatments for SRID 005 (Figure 4.12b) between the years 1998 and 2008. The slight decrease in the range of percent differences between the continuous and the sampled data from one year to the next (i.e. 2001 to 2002, 2003 to 2004, and 2006 to 2007) is likely due to the transverse cracking data variability.

It is important to note that, in most cases of Figures 4.9 through 4.12, the averages of the percent differences between the continuous and the sampled transverse cracking data are almost zero. Yet, the ranges of the percent differences are significant. Hence, it is misleading to compare the average of the sampled data to the average of the continuous data.

Regardless, in general, the data in Figures 4.9 through 4.12 indicate that as the pavement ages and the pavement conditions become more variable, sampled data do not accurately represent the continuous data. Hence, the errors due to sampling increase as the pavement ages and the pavement surface conditions and distress show higher degrees of variability. Inversely, after the application of pavement treatment actions, the pavement conditions become more uniform and the sampled data more accurately represent the continuous data. Therefore, as it has been stated in previous sections, the accuracy of ten percent sampling is a function of the data variability along the road.

4.2.2 Random Sampling

In the previous section of this thesis, fixed sampling technique was used where the pavement condition and distress of the first 0.1 mile long segment of each one mile pavement section were assumed to represent the pavement conditions and distress along the entire one mile. In this section, random sampling technique was used. Since each one mile of road consists of ten 0.1 mile pavement segments, for each survey year and for each one mile section along various roads from the four states, a random number between one and ten was generated. The random number was then used to select the 0.1 mile pavement segment along each mile that would be used as the sample. For example, if the random number was four then the pavement condition and distress data of the fourth 0.1 mile long segment of a given one mile road section was assumed to represent the pavement condition and distress along that entire mile. Hence, the randomly selected 0.1 mile long pavement segment may or may not be the same segment from one data collection cycle to the next.

For each analyzed road, the differences between the ten percent sampled and the continuous pavement condition and distress data were studied. For each one mile of road, the randomly sampled data consists of one data point whereas the continuous data provides ten points. The solid lines shown in Figure 4.13 depict the sampled IRI data while the individual data points in the figure represent the continuous IRI data for each 0.1 mile segment of the road. Likewise, the sampled and continuous transverse cracking data along four roads are shown in Figure 4.14.

The data in Figures 4.13 and 4.14 show similar results to those that have been discussed in the previous section for the fixed sampling analysis. To avoid redundancy, only four roads from the CDOT are shown in the figures. Similar results were obtained for the other roads and

are shown in Figures B.49 through B.52 of Appendix B. Examination of the data in Figures 4.13 and 4.14 indicate that, for both IRI and transverse cracking, the higher is the variability of the continuous data, the higher are the differences between the sampled and the continuous data. The specific observations are stated below.

1. Except for approximately thirteen 0.1 mile long pavement segments (localized segments), the continuous IRI data in Figure 4.13a for the thirteen miles long flexible pavement section along HWY 24 in Colorado are uniform. Consequently, the sampled data represent the continuous data throughout the thirteen miles section except at localized segments of the road. At some of these localized areas, the continuous data are more than four times greater than the sampled data.
2. The measured IRI data for the twelve miles long rigid pavement section along I-70 shown in Figure 4.13b are almost perfectly uniform. Thus, except for very few cases, the sampled data represent the continuous data throughout the length of the road.
3. The IRI data along the 6 mile long section of HWY 36 in Colorado shown in Figure 4.13c are also uniform. Consequently, except for few localized areas, the sampled data represent the continuous data.
4. Figure 4.13d shows high variability of the continuous IRI data of the seven miles long flexible pavement section along HWY 71 in Colorado. The figure also shows that the sampled data do not represent the continuous data along a majority of the road section; some sampled data are twice or half the continuous data.
5. The transverse cracking data along the twelve miles long rigid pavement section of I-70 in Colorado displayed in Figure 4.14b show no variation. Hence, the sampled data perfectly represent the measured data.

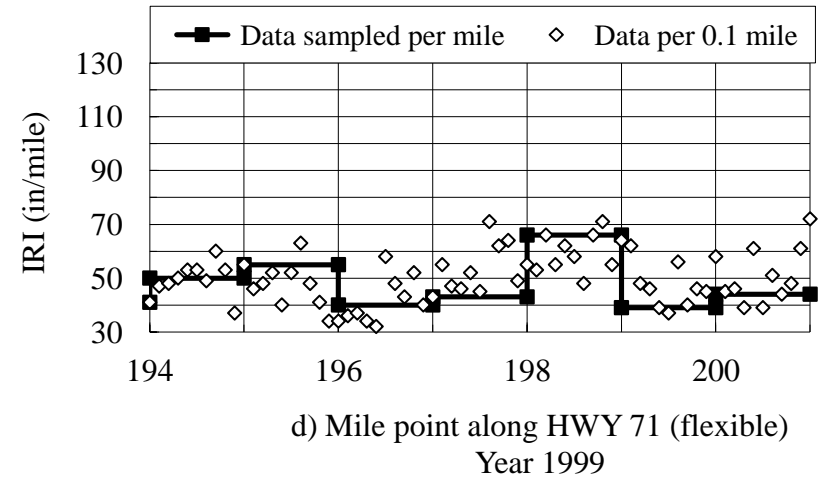
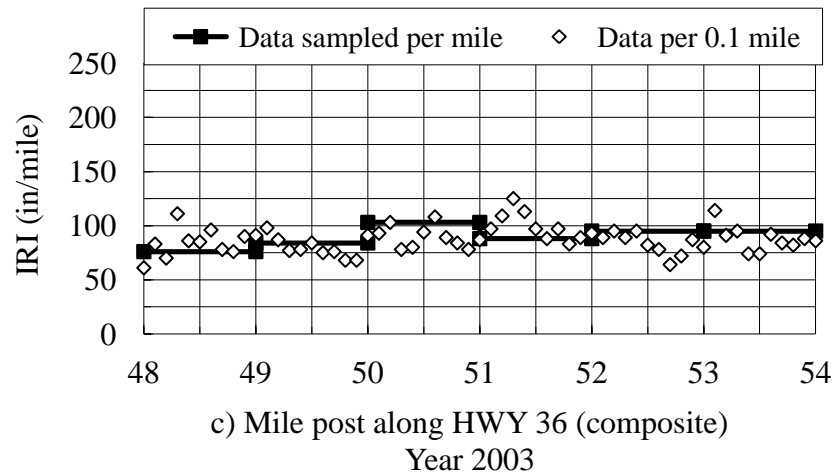
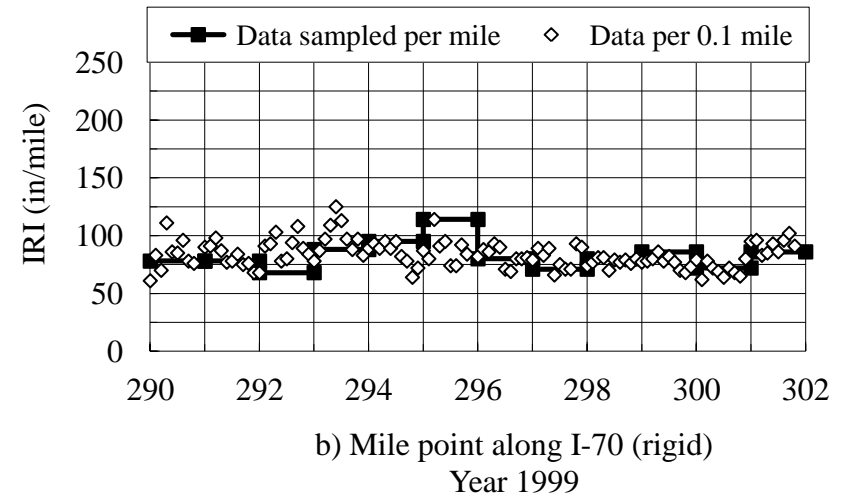
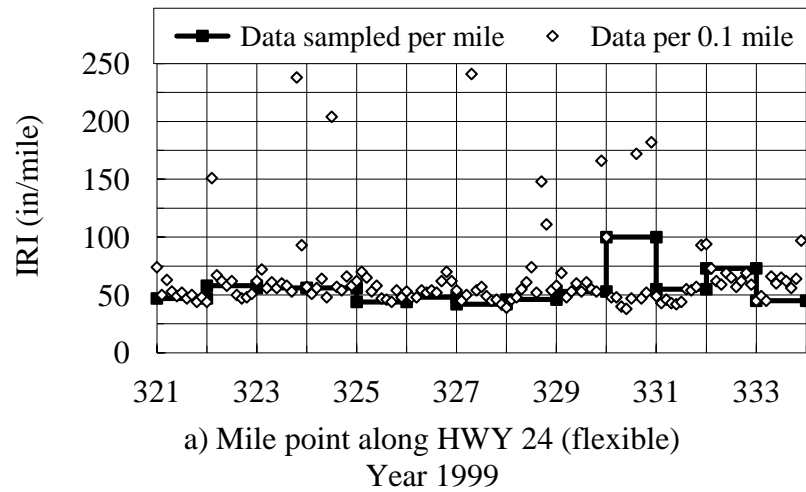


Figure 4.13 Continuous and randomly sampled IRI data per mile along four roads in Colorado

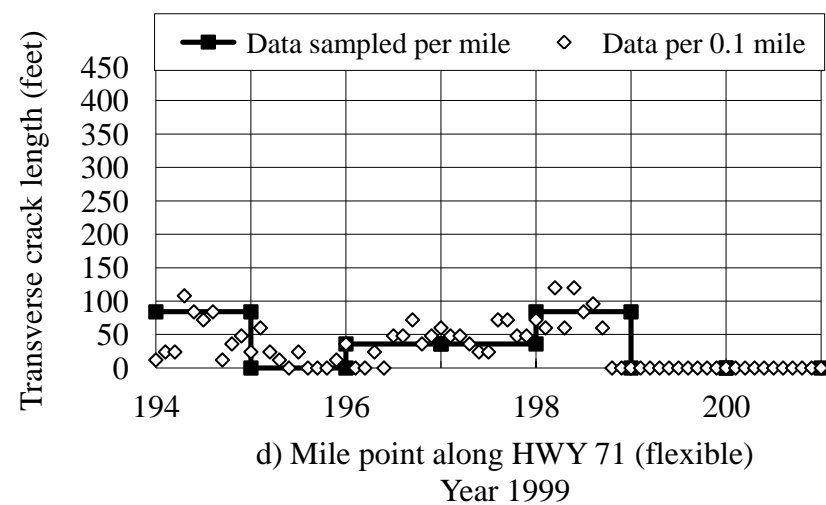
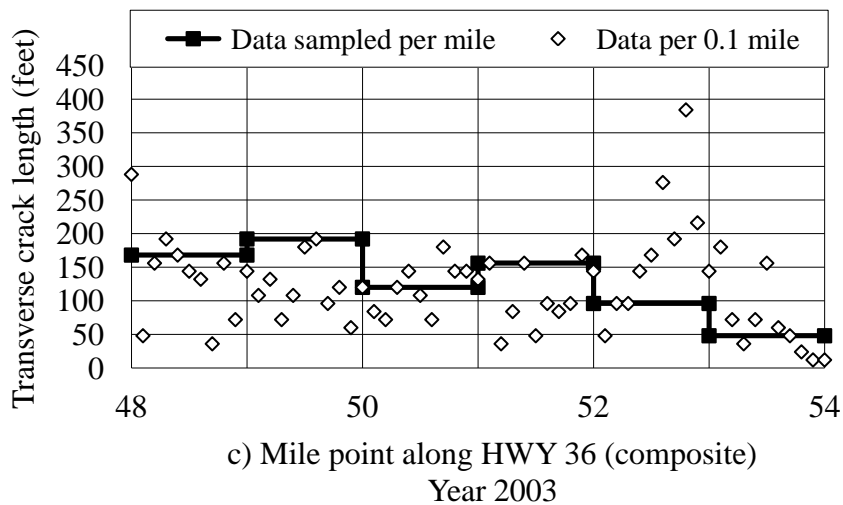
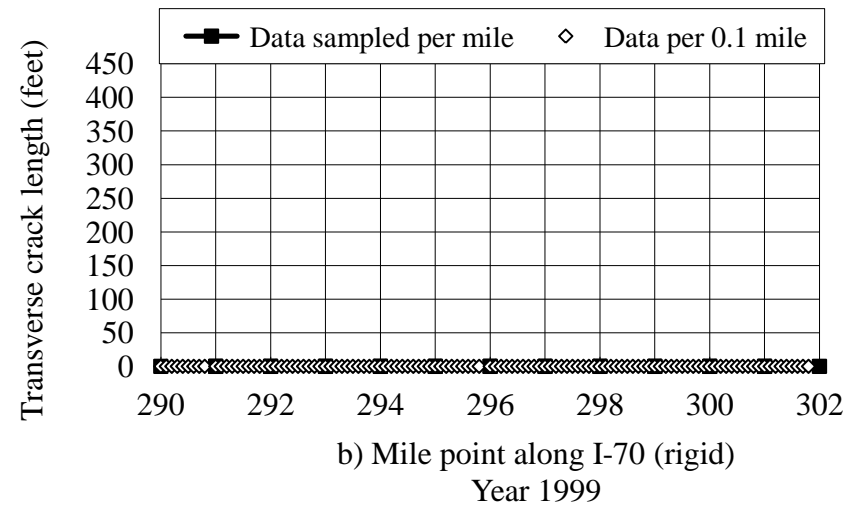
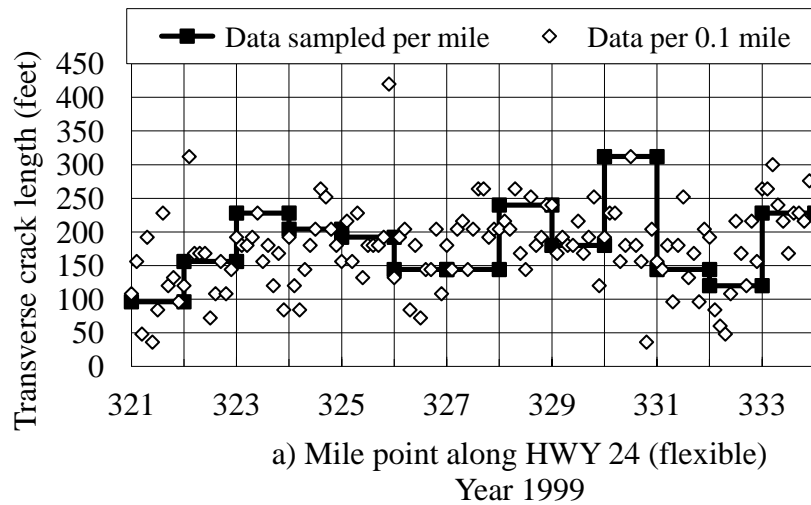


Figure 4.14 Continuous and randomly sampled transverse cracking data along four roads in Colorado

6. The variability of the continuous transverse cracking data for the other three roads shown in Figures 4.14a, c, and d is much higher than the IRI data of Figure 4.13 or the transverse cracking data of Figure 4.14b. As a result, the sampled data for these roads often misrepresent the continuous data.

The six observations regarding the random sampling of the IRI and transverse cracking data indicates that, as is the case for fixed sampling:

1. For uniform pavement condition and distress (such as in a newly constructed or rehabilitated pavement sections), the sampled data accurately represent the continuous data. Otherwise, when the pavement condition and distress are highly variable, the sampled data significantly misrepresent the pavement condition and distress.
2. In general, ten percent sampling appears to more accurately represent sensor-based data (IRI) as opposed to image-based data (cracking). This is mainly due to the data collection methodologies as discussed in section 4.2.1 above.

4.2.3 Sample Size

Given that the results presented in the previous two sections of this thesis indicate that in most cases, ten percent sampling does not accurately represent variable pavement conditions and distress, additional analysis was performed to determine the impacts of sample size on the differences between the sampled and the continuous data. Six sample sizes were used, ten percent (the results are presented in the previous sections), twenty, thirty, forty, fifty, and sixty percent. Results of the analysis for two roads in Michigan are shown in Figure 4.15. In the analysis, it was assumed that for each one mile of pavement, the average distress along the sampled lengths of 0.2, 0.3, 0.4, 0.5, and 0.6 mile of each mile independently represent the distress along the entire mile. After calculating the average transverse cracking length along each

sample length, the continuous data for each 0.1 mile was divided by the sampled data. The results for two pavement sections; one along I-94 (Figure 4.15a) and the other along US-31 (Figure 4.15b) in Michigan show similar trends. As it was expected, the differences between the sampled and the continuous data decreases as the sample size increases. For example, consider Figure 4.15a for a section of the composite pavement along I-94. The data in the figure indicate that at 10 percent sample size, the continuous data could be as high as nearly 70 times the sampled data. In this scenario, the sampled data indicate a transverse crack length of 1 foot while the sampled data indicate 68.5 feet of transverse cracking. On the other hand, for the 60 percent sample size, the continuous data could be as high as about ten times the sampled data. The results for the flexible pavement section along US-31 indicate that the continuous data could be as much as 17 times higher than the ten percent sampled data and only ten times higher for the sixty percent sampling.

The sampling of pavement condition and distress data certainly reduces the cost of data collection. However, the above observations indicate that the sampled data could produce both accurate and misleading representations of the continuous data. Thus, there is a certain risk associated with data sampling. The problem herein is the balance between the acceptable level of risk and the level of cost savings. It is clear, from the data presented above from four SHAs, that the risk level of data sampling is high when the pavement conditions and distress along the road are highly variable. On the other hand, the risk level is relatively low for uniform pavement conditions and distress. Consequently, the problem of data sampling cannot be addressed accurately and comprehensively unless the impacts of data sampling on the accuracy of the pavement management decisions can be determined. Knowing such impacts, the hidden cost of sampling and the cost savings could be calculated and compared. Therefore, the impacts of

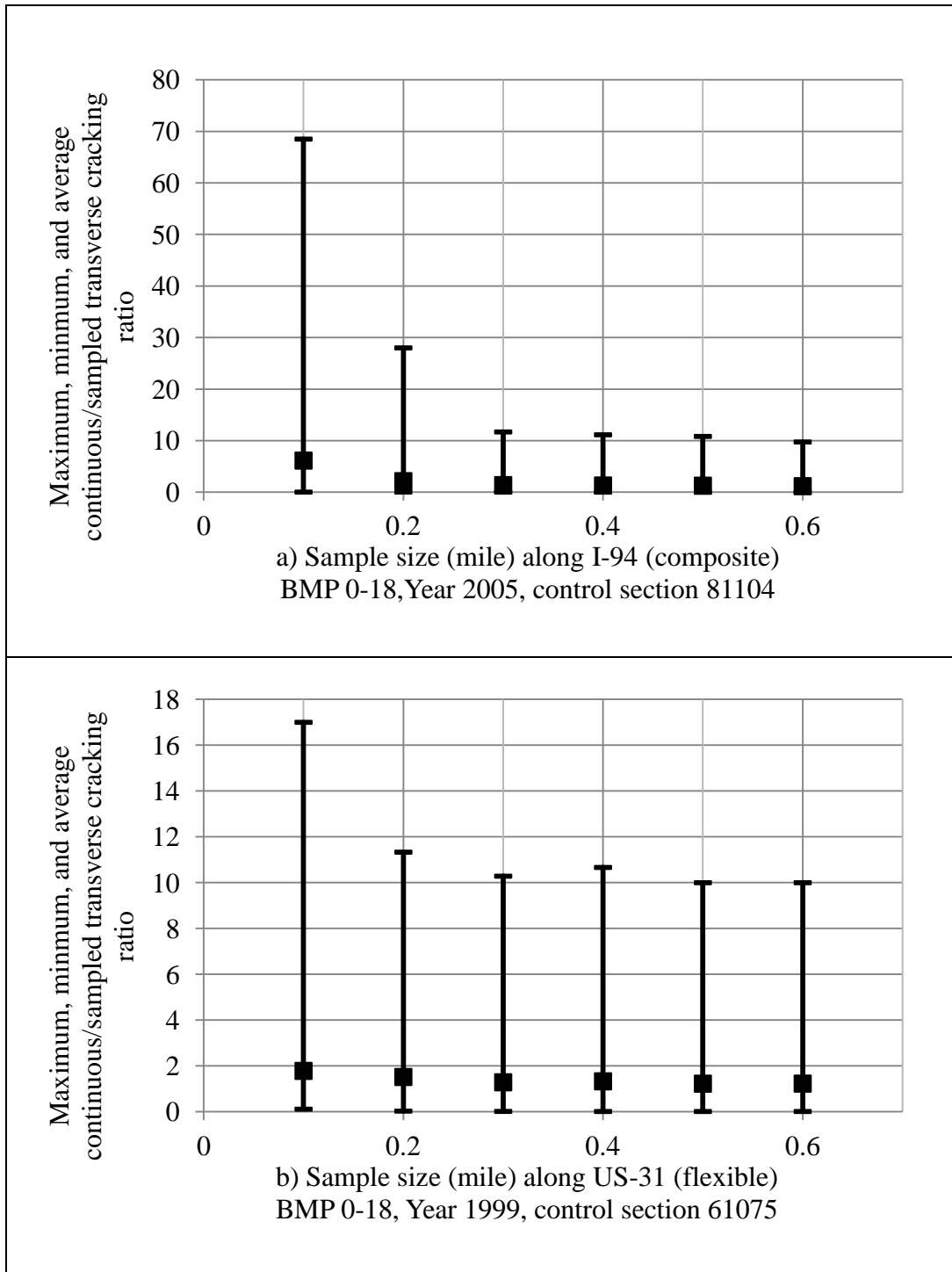


Figure 4.15 Maximum, minimum, and average continuous/sampled transverse cracking ratio versus sample size along two roads in Michigan

sampling on the accuracy of pavement management decisions are addressed in the next section. Whereas the hidden cost of sampling is addressed in section 4.5 of this thesis.

4.3 Pavement Performance Data Sampling: Condition State Concept

In the previous sections, the differences between the continuous and the sampled pavement condition and distress data were addressed. In this section, the effects of sampling the pavement condition and distress data on the accuracy of pavement management decisions are analyzed and discussed. The analysis is based on the magnitudes of the pavement condition or distress and the pavement rates of deterioration. The condition or distress and the rate of deterioration for each pavement segment in question are combined using the remaining service life (RSL) concept. Various RSL brackets (ranges) are then used to form the pavement condition states. The condition states are used to group pavement segments with similar RSL values as explained in the next section. Please note that detailed discussion of the RSL concept can be found in section 9.5 of Chapter 2.

For each road included in the analysis from each of the four SHAs, the analysis was conducted using the same steps presented below.

Step 1 – Pavement Sections Identification – Construction and maintenance records obtained from the CDOT, the LADOTD, the MDOT, and the WSDOT were examined to identify a minimum of 6 miles long pavement sections where no maintenance or preservation actions have been taken during at least three or more consecutive data collection cycles.

Step 2 – Acceptance Criteria – For each pavement section identified in step one, each of the time series pavement condition and distress data were subjected to the two acceptance criteria presented below. Note that, for each pavement condition and distress type, any 0.1 mile long pavement segment that did not pass the two acceptance criteria was excluded from any further

analysis. Further, for both acceptance criterion, a 0.1 mile long pavement segment could be accepted based on the IRI and/or the transverse cracking data and rejected based on other condition or distress type(s).

1. **First Acceptance Criterion - Three Data Points** – For each 0.1 mile long pavement segment and for each pavement condition and distress type, the time-series data were independently examined to determine whether or not the database contains a minimum of three data points. The three data points are required to model the pavement condition and distress using non-linear functions (any type of curve can be fit to two or one data points). This acceptance criterion is not redundant to step 1 above because, during any data collection cycle, the data along certain 0.1 mile long pavement segments may not have been collected due to equipment malfunctioning, construction zones, traffic flow and so forth.
2. **Second Acceptance Criterion - Positive Regression Parameters** – All pavement condition and distress data of each 0.1 mile long pavement segment that passed the first acceptance criteria were subjected to this second acceptance criterion. The criterion specifies that the time series data of each pavement condition and distress should independently have positive slope (increasing condition or distress). Positive slope implies positive regression parameters of the non-linear mathematical functions used to model the data. Negative regression parameters imply that the pavement is healing itself over time (improving condition or decreasing distress) without any treatments, which is not reasonable. Negative slopes could be caused by temperature and/or moisture variations from one data collection cycle to the next (which affect crack openings in all pavement types and curling in rigid pavements), equipment calibration, and data

variability due to the image digitization processes from one data collection cycle to the next.

Step 3 – Calculating RSL – For each 0.1 mile long pavement segment of each identified pavement section in step 1, and for the pavement condition and distress data that passed the two acceptance criterion, the data were modeled using the proper mathematical function (see Table 4.1). Note that, for cracking, the logistic function listed in Table 4.1 was divided into two parts, exponential for newly constructed or treated pavement sections and power for older sections. The main reason is that, in most cases, the four databases used in this study did not contain the necessary number of time-series data points between two consecutive treatment actions to support the data modeling using logistic function. Such function requires a minimum of five data points to estimate the values of the regression parameters with an acceptable confidence level. Nevertheless, in most cases, the 0.1 mile long pavement segments were newly treated; hence, the cracking data were modeled using the exponential function. The regression parameters of all models were calculated using the Microsoft Excel program. Based on the regression parameters, the RSL of each 0.1 mile pavement section was calculated as the difference in time between the last available data point and the time when the pavement conditions or distress are expected to reach the pre-specified threshold value. The pre-specified threshold values used in the calculation of RSL were 200 in/mile for IRI, 0.25 inch for rutting, and 250 feet of cracks for all cracking data.

Step 4 – Simulating Sampling Using the Calculated RSL Values – The RSL values of each 0.1 mile long pavement segment (using the continuous data) along each road included in the analysis that were calculated in step 3 were used to simulate ten percent sampling. It was assumed that for each one mile section of the road in question, the RSL of the first 0.1 mile

Table 4.1 Mathematical functions used to model the pavement condition and distress data

Pavement condition or distress type and units	Model form	Generic equation
IRI (inch/mile)	Exponential	$IRI = \alpha \exp^{t\beta}$
Rut depth (inch)	Power	$Rut = \gamma t^{\omega}$
Cracking (feet)	Logistic (s-shaped)	$Crack = \frac{Max}{\delta + \exp^{(\theta + \mu t)}}$
Where, α , β , γ , ω , θ , δ , and μ are regression parameters, t = elapsed time (year), and Max = the maximum value of cracking (crack saturation threshold)		

segment of each one mile represents the RSL of the entire mile. For example, if the RSL of beginning mile point (BMP) 0 was equal to 5 years then the RSL value of mile BMP 0 through BMP 0.9 was also assumed to be 5 years.

Step 5 – Pavement Condition State Based on RSL Brackets - The calculated RSL values were grouped into six brackets as listed in Table 4.2. Each of the six brackets is referred to as a pavement condition state. As indicated in the table, the RSL range of the condition states increases as the RSL value increases. This is because the accuracy of the calculated RSL values decreases as the RSL value increases (prediction over long time periods). In addition, Table 4.2 provides a list of possible pavement treatment actions based on the pavement condition state and RSL range. Explanation of the pavement condition states and the possible pavement treatment actions listed in the table is provided below.

Table 4.2 Pavement condition state, RSL range, and probable pavement treatment actions

Pavement condition state	RSL range (year)	Pavement treatment actions
1	0 - 2	Reconstruction or major rehabilitation
2	3-5	Rehabilitation
3	6-10	Preservation
4	11-15	Preservation
5	16-25	Do nothing or light maintenance
6	>25	Do nothing

- Pavement condition state one (RSL range 0 to 2 years) - In most cases, the pavement segments in this condition state have deteriorated to the point that light rehabilitation or preservation actions are not cost-effective. Hence, reconstruction or major rehabilitation is the most probable action.
- Pavement condition state two (RSL range 3 to 5 years) – Rehabilitation – Pavement segments in this condition state have experienced significant deterioration and require some rehabilitation actions.
- Pavement condition state three (RSL range 6 to 10 years) – Pavement segments in this condition state have experienced moderate deterioration and distress. Therefore, preservation actions would likely extend the life of the pavement in a cost-effective manner.
- Pavement condition state four (RSL range 11 to 15 years) – Pavement segments in this condition state have experienced minor distress and deterioration and their condition state could be substantially improved by the application of certain preservation actions. These actions would likely extend the life of the pavement and decrease its rate of deterioration.

- Pavement condition states five and six (RSL range 16 to 25 years and greater than 25 years) - Pavement segments within these condition states are in good to excellent condition and may have minimal distress such as minor cracking. Light maintenance activities, such as crack sealing, are the most likely cost-effective options.

Nevertheless, the continuous and sampled RSL values were grouped into condition states listed in Table 4.2. The purpose of such grouping is to determine the effects of sampling on the accuracy of the pavement management decisions.

Step 6 – Number of 0.1 Mile Long Pavement Segments in Each Condition State –

For each road and each RSL bracket, the number of pavement segments within each condition state were separately counted for the continuous and sampled data. The same numbers were then divided by the total number of 0.1 mile long pavement segments along the road section in question to obtain the percent of each road in the different condition states.

Step 7 – Changes in Condition State Due to Sampling – A change in the condition state of a pavement segment, with no applied treatment, implies that the pavement segment has deteriorated from one condition state to the next. For example, if the condition state of one 0.1 mile long pavement segment was 3 in 2005 and 1 in 2007, then the pavement condition state has changed by 2 states in two years. In this study, for each data collection cycle, the condition state based on the continuous data was compared to that based on the sampled data. Changes in the condition states due to sampling imply that sampling affects the accuracy of the pavement management decisions since project selection is based on the pavement condition state and its distribution. Nevertheless, for each 0.1 mile pavement segment of each road included in the analysis, the condition state of the sampled data was subtracted from the condition state of the continuous data as shown by Equation 4.1. The results (the number of condition states changed)

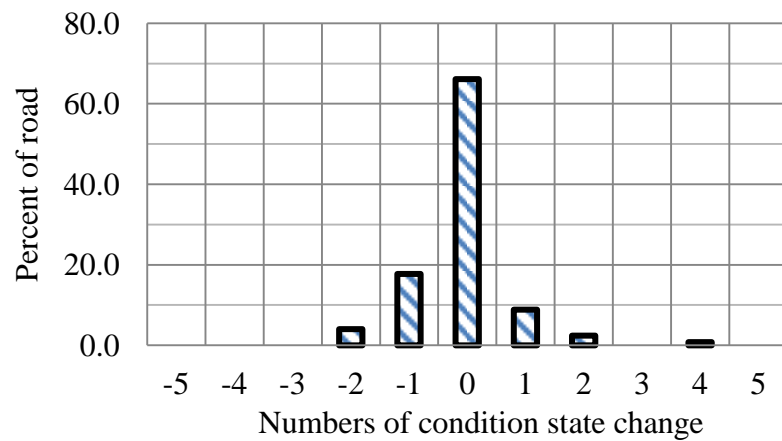
could be negative or positive. For instance, if the condition state based on the continuous data is 1 and the sample indicated a condition state of 5, the number of condition states changed is -4. On the other hand, if the continuous condition state is 5 and the sample indicates a condition state of 1, there is a condition state change of + 4.

$$\text{Changes in condition state} = \text{continuous condition state} - \text{sampled condition state} \quad \text{Eq. 4.1}$$

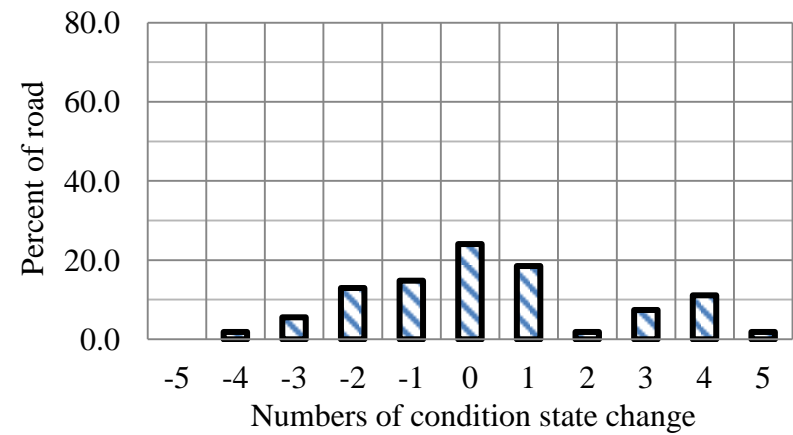
Results of the analyses for eight roads, two roads from each of the four SHAs, shown as the percent of the road that changed a certain number of condition states based on the IRI data are listed in Table 4.3 and depicted in Figures 4.16 and 4.17.

Table 4.3 Percent of each road changing certain number of condition states based on the IRI data

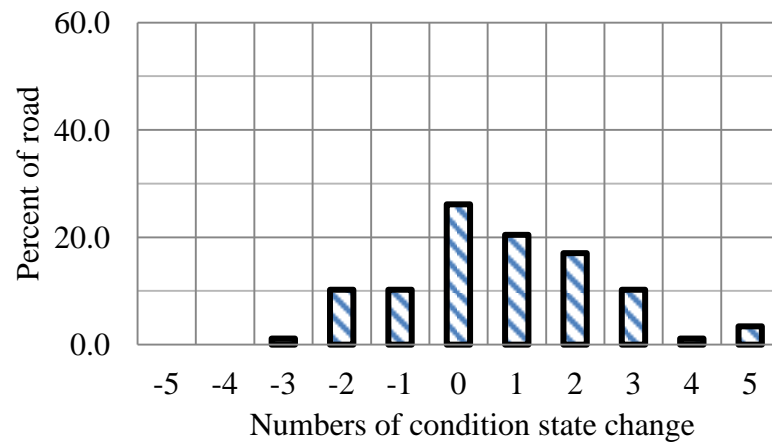
State	Road	Length (mile)	Percent of each road that changed the stated number of condition states										
			-5	-4	-3	-2	-1	0	1	2	3	4	5
CO	HWY 24	13	0.0	0.0	0.0	3.9	17.1	65.9	8.5	2.3	0.8	1.6	0.0
	I-70	6	0.0	1.7	5.1	11.9	16.9	22.0	18.6	3.4	8.5	10.2	1.7
LA	US-90	11	2.0	0.0	2.0	10.9	8.9	30.7	17.8	14.9	8.9	1.0	3.0
	US-80	14	0.0	0.0	1.4	5.1	13.0	34.8	23.9	13.0	5.8	1.4	1.4
MI	I-94	18	0.0	0.0	0.0	5.0	12.8	52.8	18.3	6.1	5.0	0.0	0.0
	US-131	16	0.0	0.0	0.0	3.8	16.3	53.8	15.6	5.6	5.0	0.0	0.0
WA	SRID 161	9	0.0	0.0	3.8	8.8	21.3	47.5	17.5	1.3	0.0	0.0	0.0
	SRID 020	8	1.3	2.5	5.0	2.5	13.8	58.8	10.0	3.8	2.5	0.0	0.0
Total network		95	0.1	0.4	1.5	6.0	15.0	47.8	16.7	6.5	4.3	1.2	0.6



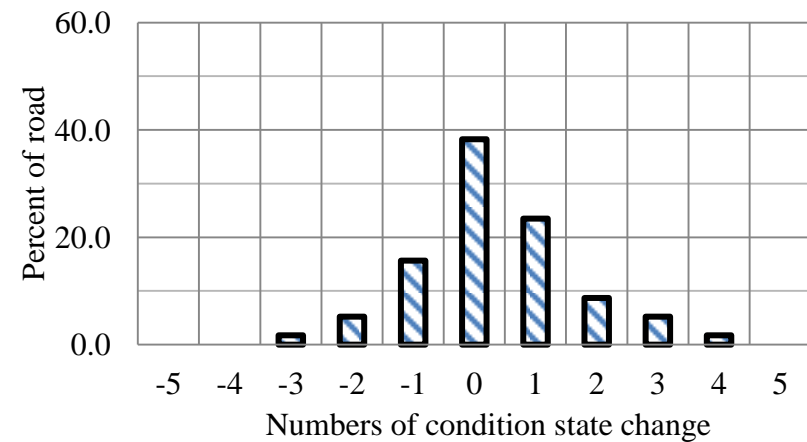
a) HWY 24 (flexible), BMP 321-334, CDOT



b) I-70 (flexible), BMP 261-267, CDOT

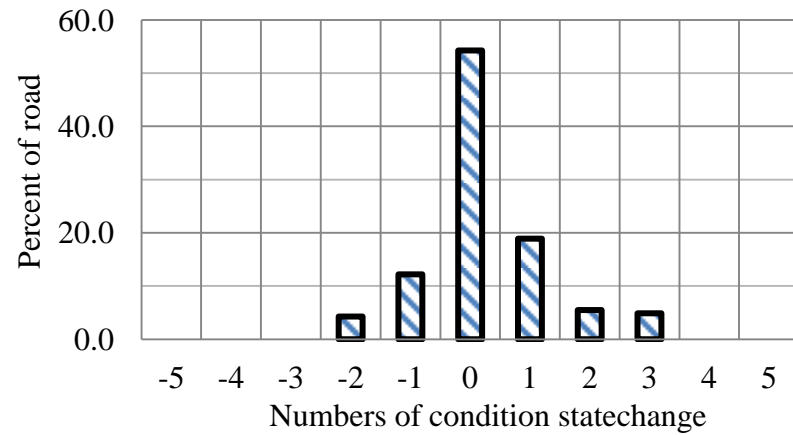


c) US-90 (flexible), BMP 0-11, control section 0.0031, LADOTD

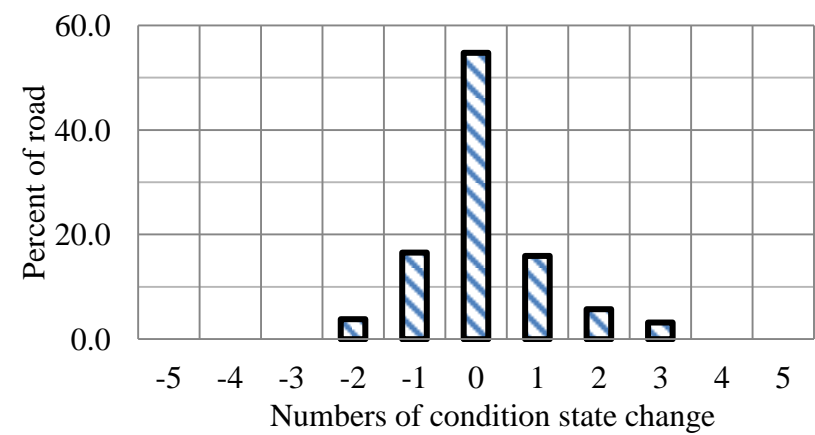


d) US-80 (flexible), BMP 0-14, control section 0.00107, LADOTD

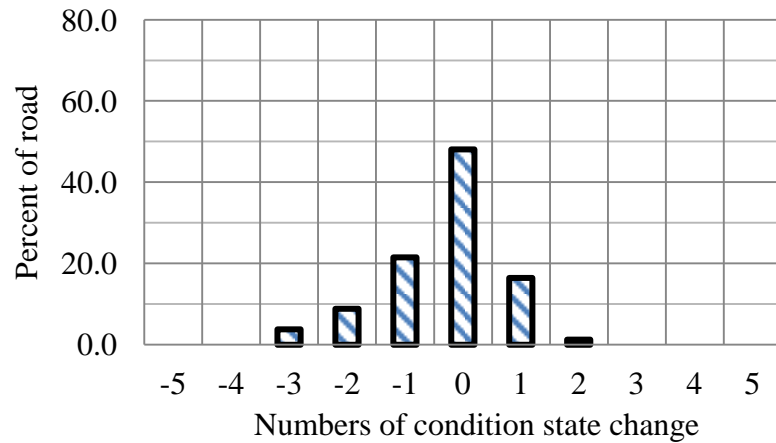
Figure 4.16 Percent of two roads in Colorado and two in Louisiana versus the numbers of condition state change, IRI data



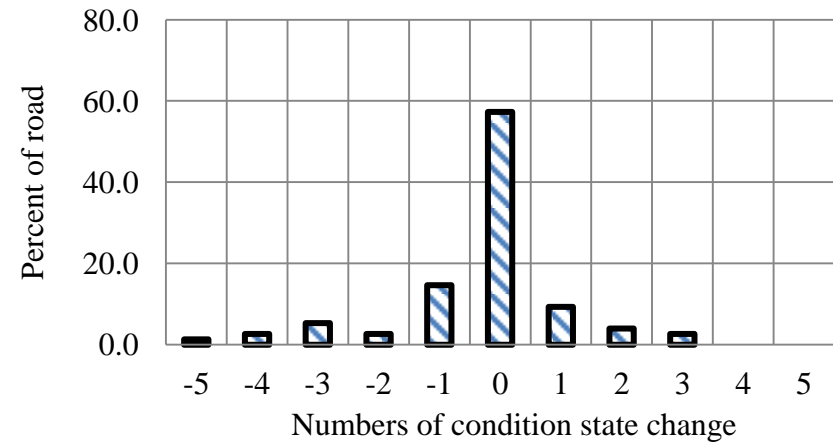
a) I-94 (composite), BMP 0-18, control section 81104, MDOT



b) US-31(flexible), BMP 0-18, control section 61075, MDOT



c) SRID 161(composite), BMP 8.93-17.93, WSDOT



d) SRID 020 (flexible), BMP 32.61 to 40.61, WSDOT

Figure 4.17 Percent of two roads in Michigan and two in Washington versus the numbers of condition state change, IRI data

Examination of the data listed in Table 4.3 and depicted in Figures 4.16 and 4.17 indicate that the percent of each road where the sampled and the continuous data produce the same condition states varies from 24.1 percent for I-70 in Colorado to about 66.1 percent for HWY 24 in Colorado. The percentage for each road is detailed below.

- For HWY 24 in Colorado shown in Figure 4.16a, the sampled and continuous data show no condition state differences for about 66 percent of the pavement segments. That is, ten percent sampling produces no condition state changes for about 66 percent of the 13 mile section of HWY 24 whereas 34 percent of the pavement segments change condition states.
- For the 6 miles long section along I-70 in Colorado shown in Figure 4.16b, only 22 percent of the section does not change condition states due to sampling while 78 percent do change condition states as a result of sampling.
- For the 11 miles long section along US-90 in Louisiana shown in Figure 4.16c about 26 percent of the pavement segments do not change pavement condition states due to sampling. However, 74 percent of the pavement section changes condition states.
- Figure 4.16d, for US-80 in Louisiana, shows that about 38 percent of the road does not change pavement condition states due to sampling, whereas 62 percent changes condition states.
- For I-94 and US-31 in Michigan shown in Figures 4.17a and 4.17b, respectively, and for the eight miles long section along SRID 020 in Washington shown in Figure 4.17d, approximately 55 percent of the pavement sections do not change condition states from the sampled to the continuous data. Stated differently, ten percent sampling precipitates

differences in the sampled and continuous condition states for about 45 percent of each of the three roads.

- For the nine miles long section along SRID 161 in Washington shown in Figure 4.17c, approximately 48 percent of the pavement section does not change condition states from the sampled to the continuous data.

Finally, the eight roads listed in Table 4.3 were combined as one network and the percentages of the network that changed the indicated number of condition states are listed in the bottom row of the table and are shown in Figure 4.18. The data listed at the bottom row of Table 4.3 indicate that the effects of ten percent sampling on the condition states are:

- About 48 percent of the eight roads (the network) did not change condition states. That is the condition states based on the sampled data were the same as those based on the continuous data. Please note that the 48 percent is an average value, the actual percentage along the network varies from 66.1 to 24.1 percent.
- The condition state of about 32 percent of the road network changed one state.
- The condition state of about 13 percent of the eight roads changed up or down by 2 states which is equivalent to approximately 8 years.
- The condition states of the remaining 7 percent of the eight roads changes up or down by 3 or more states. This is equivalent to 13 or more years.

In addition, as it was expected, the percent of the network that changed the various condition state numbers is almost normally distributed amongst all condition states as shown in Figure 4.18. The normal distribution of the differences between the condition states based on the continuous data and those based on the sampled data may lead some to conclude that on average, the sampling error is zero. Such conclusion will be unfounded and misleading.

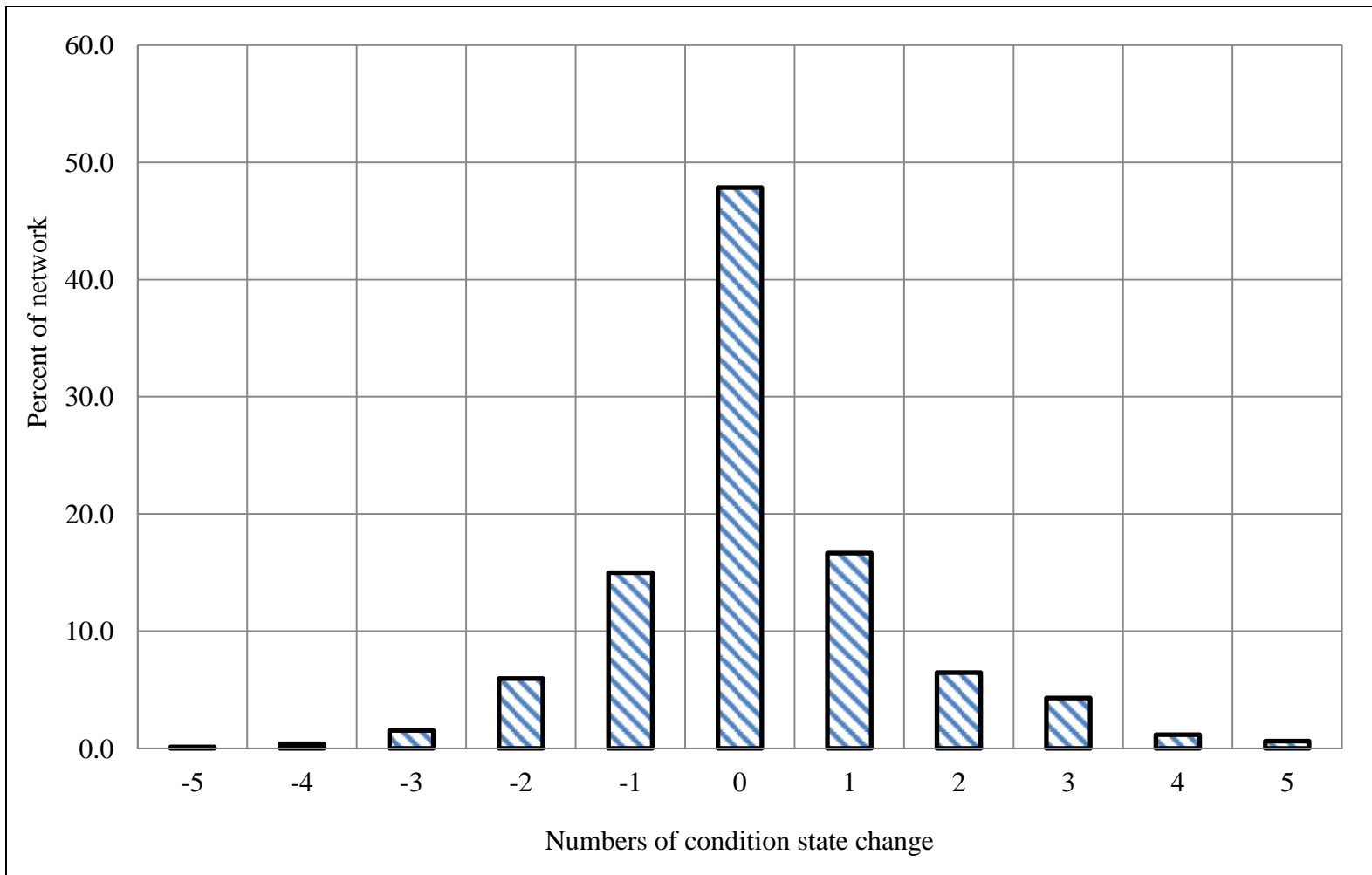


Figure 4.18 Distribution of the numbers of condition state change based on IRI data for the approximately 97 mile network

Indeed, ten percent sampling of the pavement condition data may cause errors in one or all of the following decisions:

- The selection of optimum time and space (project boundaries) for the pavement preservation and rehabilitation programs since ten percent sampling causes over and/or underestimation of the condition state (RSL) of certain percentages of the road.
- The selection of the pavement treatment type since the sampled pavement condition and distress data may not represent the actual pavement conditions and distress and their distribution along the pavement network.
- The selection of the pavement treatment strategy since ten percent sampling yields inaccurate distribution of the pavement condition states before treatment.

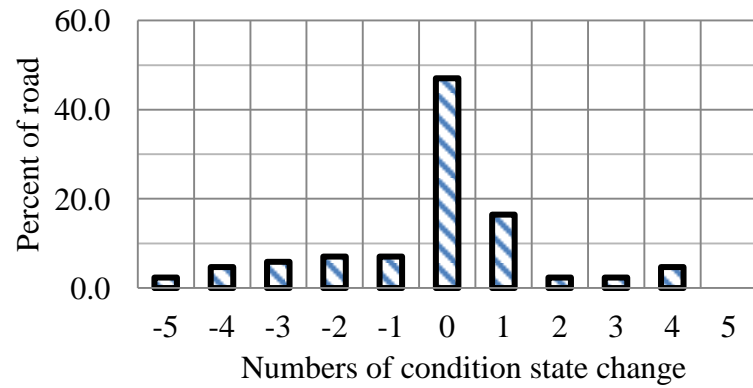
Similar condition state analyses were performed on the rut depth and longitudinal cracking data of the four SHAs. To avoid repetition and facilitate clarity, results of the analyses along two roads (SRID 020 and SRID 005 from the WSDOT) are presented and discussed. Figure 4.19 depicts the percent of road versus the numbers of condition states changed based on the rut depth and longitudinal cracking data. Figures 4.19a and 4.19b show approximately 50 percent of the road along SRID 020 and about 25 percent of SRID 005 did not change condition states due to sampling. These results were expected and are similar to the IRI results because both data (rut depth and IRI) are automatically collected by sensors as described previously in this chapter. The condition states of the remaining 50 percent of SRID 020 and the remaining 75 percent of SRID 005 changed from ± 1 to ± 4 condition states. Similar to the IRI scenarios presented above, the condition states changes would result in different pavement management decisions. Finally, given that the variability of sensor collected data is much lower than the

image collected data, it should be expected that the effects of sampling on the image based data is much higher than the sensor collected data.

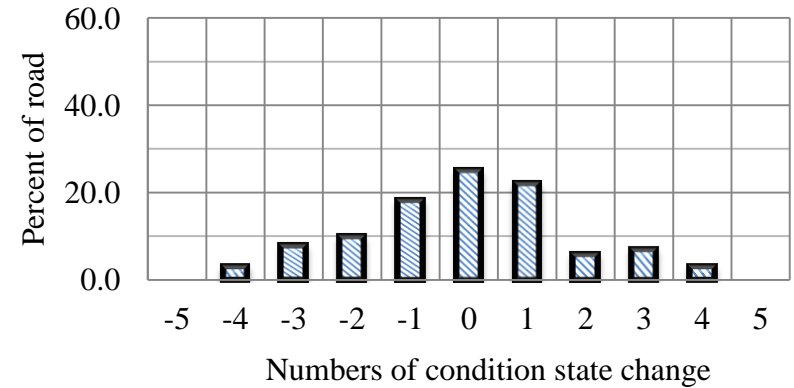
Figures 4.19c and 4.19d depict the results of the analyses of the longitudinal cracking data along sections of SRID 020 and SRID 005. It can be seen that ten percent sampling of the longitudinal cracking data caused the condition states based on the continuous data to:

- Change for about 50 percent of SRID 020 while the condition states of the other 50 percent remained the same.
- Change for about 70 percent of SRID 005 while the condition states of the other 30 percent remained the same.
- Change by 5 condition states or ± 20 years for about 20 percent of SRID 020 and more than 10 percent of SRID 005.

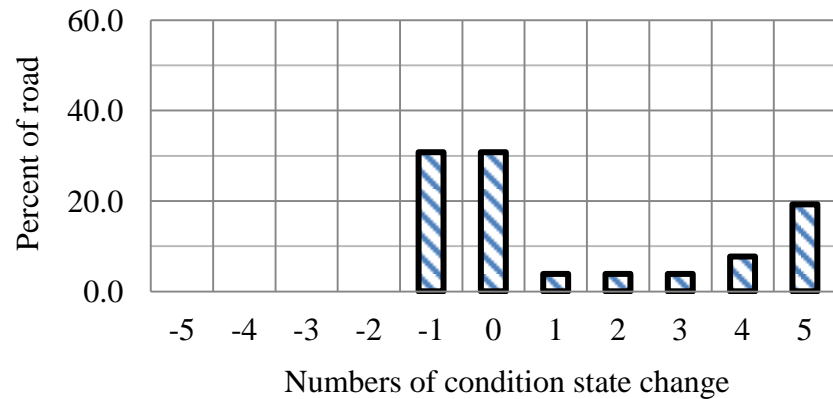
Relative to the IRI and rut depth data, the higher differences between the RSL brackets based on the sampled and those based on the continuous data were expected and confirm the statement made earlier that image based data have higher variability and therefore the sampled data have higher errors. The high percentages of the two roads changing condition states would pose significant problems for the SHAs and their decisions regarding the selection of cost-effective fix time, project boundaries (discussed in the next section), and fix type. If one is to examine the bigger picture, such errors force the SHA staff in the various regions to lose confidence in the PMS data and hence, ignore it. In fact, this could be one of the most damaging scenarios for any highway agency. Pavement management decisions may be made without any reference to the data; hence, making the data unnecessary and ineffective.



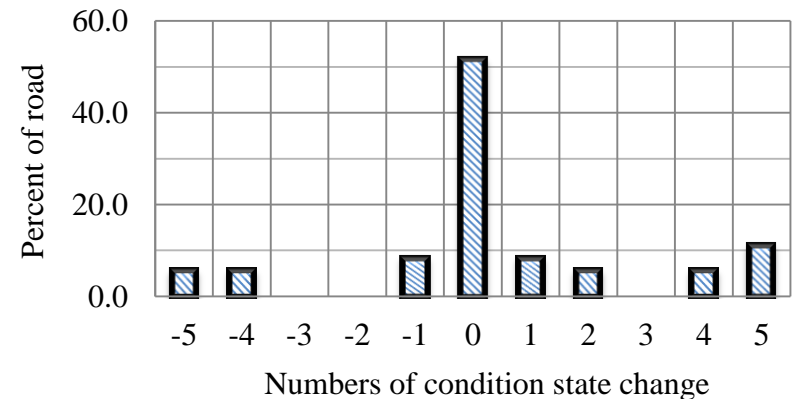
a) SRID 020 (flexible), BMP 32.61 to 40.61, rut depth



b) SRID 005 (composite), BMP 59.7 to 79.7, rut depth



c) SRID 020 (flexible), BMP 32.61 to 40.61, longitudinal cracking



d) SRID 005 (composite), BMP 59.7 to 79.7, longitudinal cracking

Figure 4.19 Percent of two roads in Washington versus the numbers of condition state change, rut depth and longitudinal cracking data

4.4 Pavement Project Boundary Analysis

In this section, the effects of 10, 20, 30, 40, 50, and 60 percent sample sizes on the selection of project boundaries are analyzed and discussed. The analyses were conducted using:

1. The continuous and sampled pavement condition and distress data for six pavement sections; two sections from each of the CDOT, MDOT, and WSDOT.
2. The RSL values based on the continuous and the sampled data for the six road sections.
3. The AASHTO unit delineation by cumulative differences method, referred to herein as the AASHTO unit delineation method, to determine the project boundaries based on uniform pavement sections. The method consists on the calculation of the various variables or parameters listed in Table 4.4. The project boundaries are defined by the locations where the slope of the variable Z_x as a function of the distance X along the project changes sign from positive to negative and vice versa (ARA, Inc. 2004).

Table 4.4 illustrates the equations and solution sequence required to execute the AASHTO unit delineation method. The table and entries are self-explanatory. In order to promote clarity and understanding, Table 4.5 illustrates the solution sequence using IRI data from a two mile long section of the flexible HWY 24 in Colorado. The results (the variable Z_x as a function of X or BMP) are shown in Figure 4.20 where the boundaries of three projects are defined by the three solid lines in the figure. One project extends from BMP 321 to BMP 322, one from 322 to 322.7, and the third from 322.7 to 323. In some scenarios and for cost-effective purposes, some SHAs specify a minimum pavement project length. In this study, a minimum project length of one mile was used. Hence, the three projects in Figure 4.20 were consolidated into the two projects shown in Figure 4.21.

Table 4.4 Solution sequence of the AASHTO unit delineation method (after AASHTO 1993)

	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9
i	BMP	Pavement condition or distress (r_i)	Interval number (n)	Interval length (Δx_i)	Cumulative interval length (CD)	Average interval condition or distress (\bar{r}_i)	(Pavement condition or distress)* (Interval length) (a_i)	Cumulative area (CA)	Cumulative difference variable $Z_x = \text{Column 8} - F * \text{Column 5}$
1	BMP ₁	r_1	N_1	Δx_1	$CD_1 = \sum_{i=1}^n \Delta X_i$	$\bar{r}_1 = \frac{1}{2} \sum_{i=1}^i r_i$	$a_1 = \bar{r}_1 \Delta x_1$	$CA_1 = \sum_{i=1}^i a_i$	$Z_{x1} = CA_1 - F * CD_1$
2	BMP ₂	r_2	N_2	Δx_2	$CD_2 = \sum_{i=1}^n \Delta X_i$	$\bar{r}_2 = \frac{1}{2} \sum_{i=1}^i r_i$	$a_2 = \bar{r}_2 \Delta x_2$	$CA_2 = \sum_{i=1}^i a_i$	$Z_{x2} = CA_2 - F * CD_2$
3	BMP ₃	r_3	N_3	Δx_3	$CD_3 = \sum_{i=1}^n \Delta X_i$	$\bar{r}_3 = \frac{1}{2} \sum_{i=1}^i r_i$	$a_3 = \bar{r}_3 \Delta x_3$	$CA_3 = \sum_{i=1}^i a_i$	$Z_{x3} = CA_3 - F * CD_3$
...
n	BMP _n	r_n	N_n	Δx_n	$CD_n = \sum_{i=1}^n \Delta X_i$	$\bar{r}_n = \frac{1}{2} \sum_{i=1}^i r_i$	$a_n = \bar{r}_n \Delta x_n$	$CA_n = \sum_{i=1}^i a_i$	$Z_{xn} = CA_n - F * CD_n$
BMP = Beginning mile point Interval number = the data point number along the project								$F = \frac{CA_n}{CD_n}$	

Table 4.5 Example of the solution sequence of the AASHTO unit delineation method using IRI data for 2 miles of HWY 24 in Colorado

i	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9
	BMP	IRI (in/mile)	Interval number	Interval length (mile)	Cumulative interval length (mile)	Average interval IRI (in/mile)	Area = IRI*(interval length)	Cumulative area	Z _x value
1	321	163	1	0.1	0.1	163	16.3	16.3	2.49
2	321.1	117	2	0.1	0.2	140	14.0	30.3	2.67
3	321.2	116	3	0.1	0.3	116.5	11.7	42.0	0.51
4	321.3	128	4	0.1	0.4	122	12.2	54.2	-1.10
5	321.4	78	5	0.1	0.5	103	10.3	64.5	-4.61
6	321.5	94	6	0.1	0.6	86	8.6	73.1	-9.83
7	321.6	88	7	0.1	0.7	91	9.1	82.2	-14.54
8	321.7	87	8	0.1	0.8	87.5	8.7	90.9	-19.60
9	321.8	88	9	0.1	0.9	87.5	8.8	99.7	-24.66
10	321.9	110	10	0.1	1	99	9.9	109.6	-28.58
11	322	134	11	0.1	1.1	122	12.2	121.8	-30.19
12	322.1	174	12	0.1	1.2	154	15.4	137.2	-28.60
13	322.2	212	13	0.1	1.3	193	19.3	156.5	-23.11
14	322.3	163	14	0.1	1.4	187.5	18.8	175.2	-18.17
15	322.4	172	15	0.1	1.5	167.5	16.8	192.0	-15.24
16	322.5	213	16	0.1	1.6	192.5	19.3	211.2	-9.80
17	322.6	218	17	0.1	1.7	215.5	21.6	232.8	-2.06
18	322.7	148	18	0.1	1.8	183	18.3	251.1	2.42
19	322.8	130	19	0.1	1.9	139	13.9	265.0	2.51
20	322.9	96	20	0.1	2.0	113	11.3	276.3	0.00
BMP = Beginning mile point Interval number = the data point number along the project							CA _n = 276.3		
							F = 138.125		

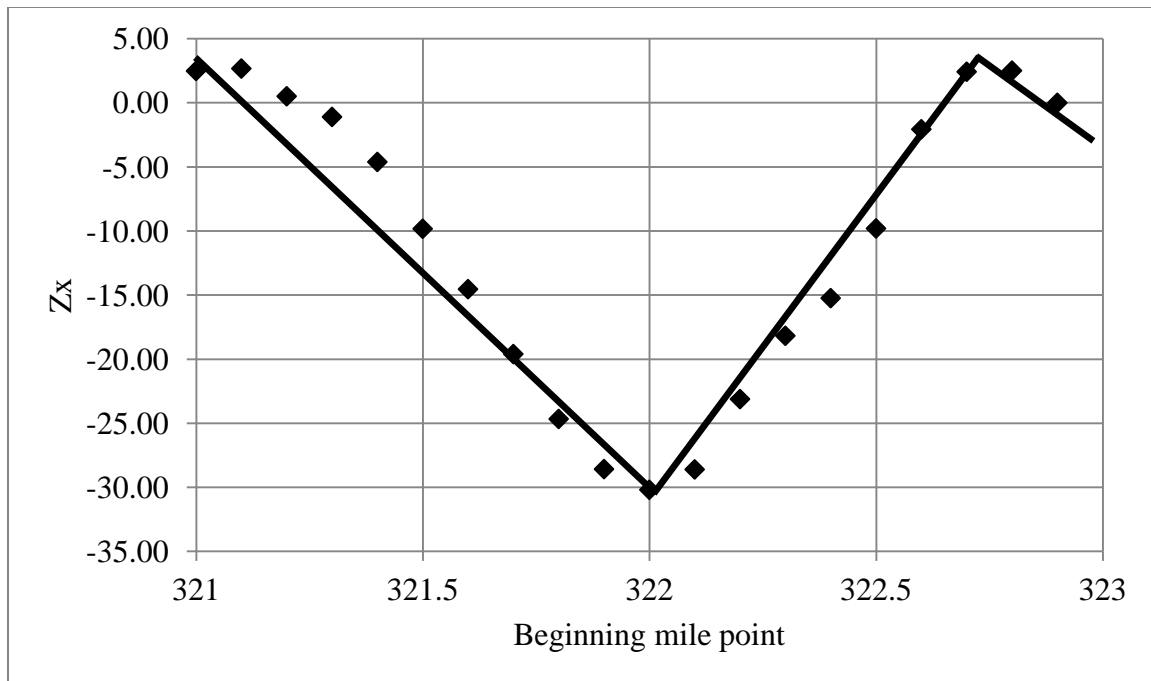


Figure 4.20 Cumulative differences versus beginning mile point along the flexible HWY 24 in Colorado showing three pavement projects

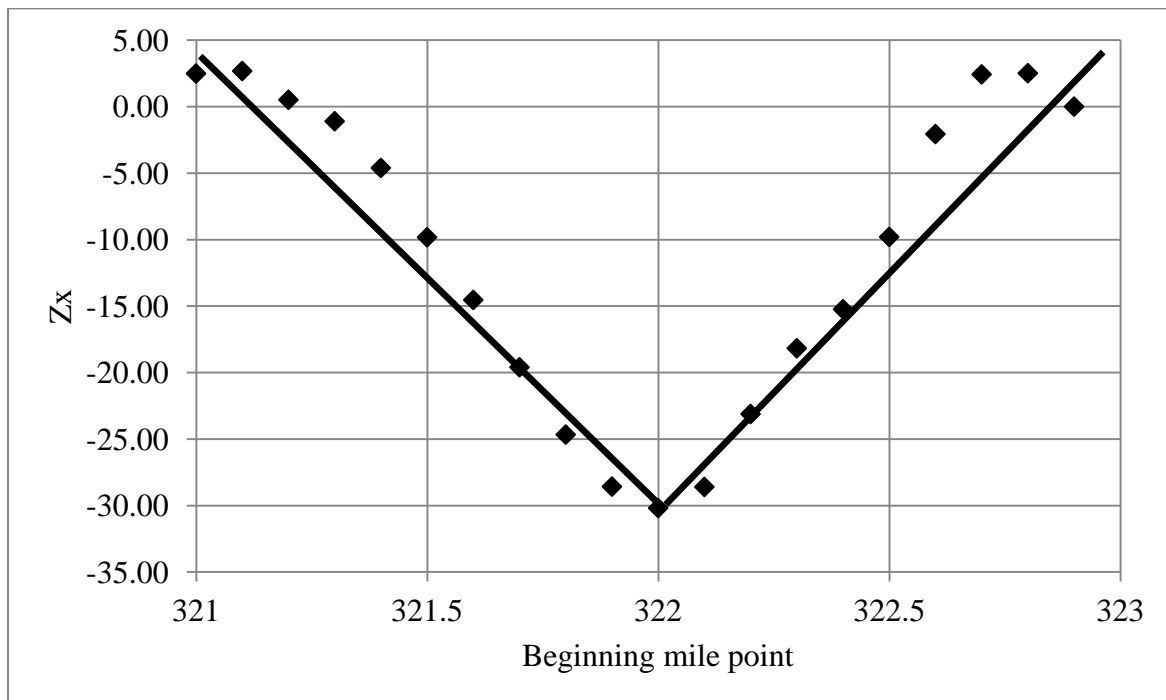


Figure 4.21 Cumulative differences versus beginning mile point along the flexible HWY 24 in Colorado showing two pavement projects

In reality, various scenarios exist for short projects. These scenarios include:

1. The pavement conditions and distress along the short project do not require any action. In this case, the short project is a “do-nothing” project.
2. The pavement conditions and distress along the short project require treatment actions while the adjacent long project is a “do-nothing” project. In this case, the short project could be handled by the SHA or by contractor depending on the pavement needs.
3. The pavement conditions and distress along the short project require treatment actions similar to those of the adjacent long project. In this case, the two projects could be combined.
4. The pavement conditions and distress along the short project require treatment actions that are substantially different than those along adjacent long projects. In this case, the two projects can be combined into one contract that specifies the various treatments and their locations.

As stated earlier, pavement project boundary analyses were completed for six pavement sections, two from each of the CDOT, MDOT, and WSDOT. The analyses were conducted using the continuous and sampled pavement condition and distress data and RSL data, coupled with the AASHTO unit delineation method described above. Results of the analyses are presented and discussed in the following subsections.

4.4.1 Pavement Project Boundary Analysis: Condition and Distress Data Approach

For each of the six pavement sections, the AASHTO unit delineation method was used to delineate the continuous and the sampled pavement condition and distress data. The sampled data for each sample size were obtained by assuming that for each one mile of pavement, the average condition and distress along the sampled lengths of 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6 mile of

each mile independently represent the condition and distress along the entire mile. The effects of 10, 20, 30, 40, 50, and 60 percent sample sizes on the identification of pavement project boundaries were then analyzed by comparing the project boundaries identified using the continuous condition and distress data to those using the sampled data. The analyses steps detailed below were used for each pavement section.

Step 1 – Solution Sequence – For the continuous and sampled pavement condition and distress data along each of the six pavement sections from the three SHAs, the solution sequence of the AASHTO unit delineation method displayed in Table 4.4 and illustrated in Table 4.5 was used to calculate the cumulative difference value at each BMP.

Step 2 – Project Boundary Identification – For the continuous and sampled pavement condition and distress data along each pavement section, the cumulative difference values were plotted as a function of the distance along the pavement section (the BMP). Example plots for a portion of HWY 24 in Colorado are shown in Figures 4.22 and 4.23 where the cumulative differences are based on the continuous and ten percent sampled IRI data, respectively. Plots for the other sample sizes are similar. Using visual inspection, the project boundaries were defined at the locations where there are changes in the slope (positive to negative or vice versa) of the cumulative differences along the length of the pavement section. The black lines in Figures 4.22 and 4.23 indicate the various pavement projects and their boundaries. Note that where the data indicate a pavement project length of less than one mile, the project was included as a part of a longer adjacent project. It is important to notice that, for the continuous pavement condition and distress data, ten cumulative difference values exist for each one mile along the pavement section. However, for the sampled data, only one cumulative difference value exists for each one

mile. This is because the sampled data represent the pavement conditions or distress along the entire mile.

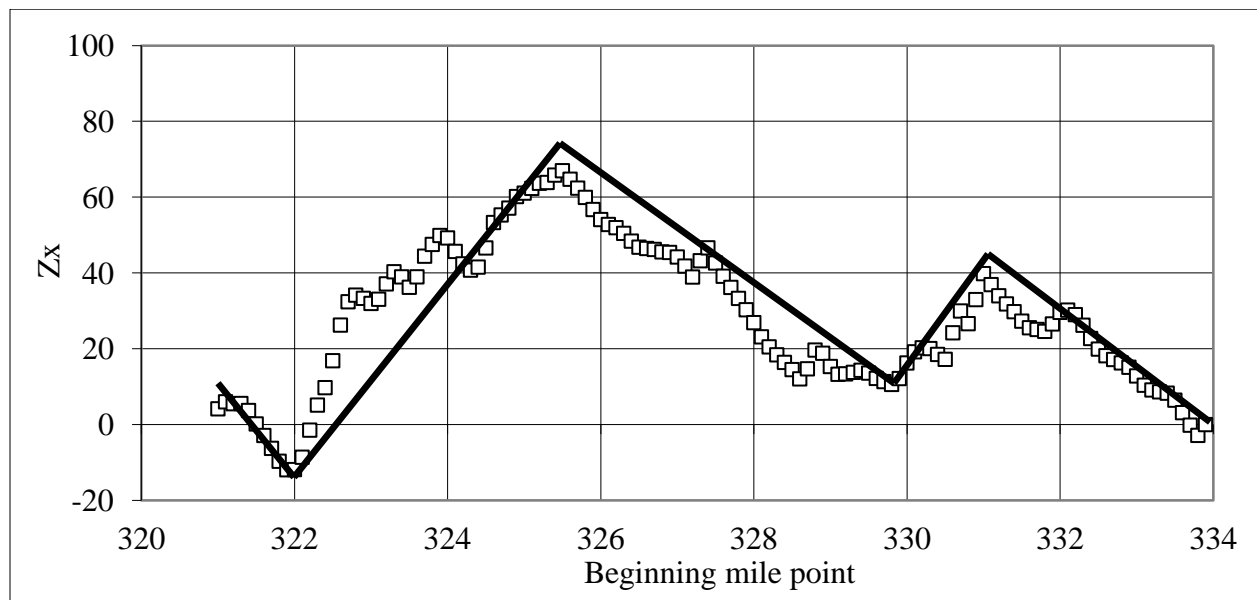


Figure 4.22 Project boundaries based on the AASHTO unit delineation method and the continuous IRI data along 13 miles of the flexible HWY 24 in Colorado

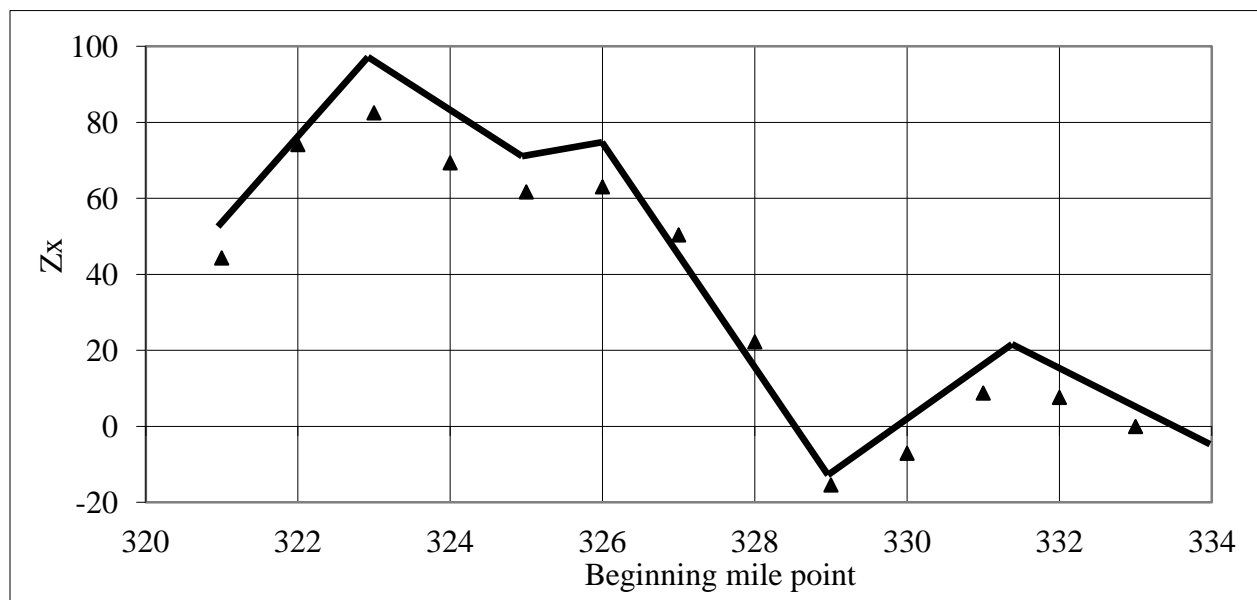


Figure 4.23 Project boundaries based on the AASHTO unit delineation method and the ten percent sampled IRI data along 13 miles of the flexible HWY 24 in Colorado

Step 3 – Comparison of the Results Based on the Pavement Condition and Distress –

To facilitate clarity and understanding, the project boundaries based on the continuous pavement condition and distress data (100 percent sampled) and on the 10, 30, and 60 percent sampled data were combined into one figure. Results of the 20, 40, and 50 percent sample sizes were similar to the results of the other sample sizes and thus, they are not included in the figure to avoid congestion. Figures 4.24, 4.25, and 4.26 show the pavement project boundaries based on the continuous and sampled IRI data along a 13 miles long section of HWY 24 in Colorado, 18 miles section of I-94 in Michigan, and 8 miles long section of SRID 161 in Washington, respectively. In each figure, the data points represent the identified pavement project boundaries. Results of the analysis for the other types of pavement condition and distress data are similar to those shown for the IRI data and are shown in Figures B.53 through B.55 of Appendix B.

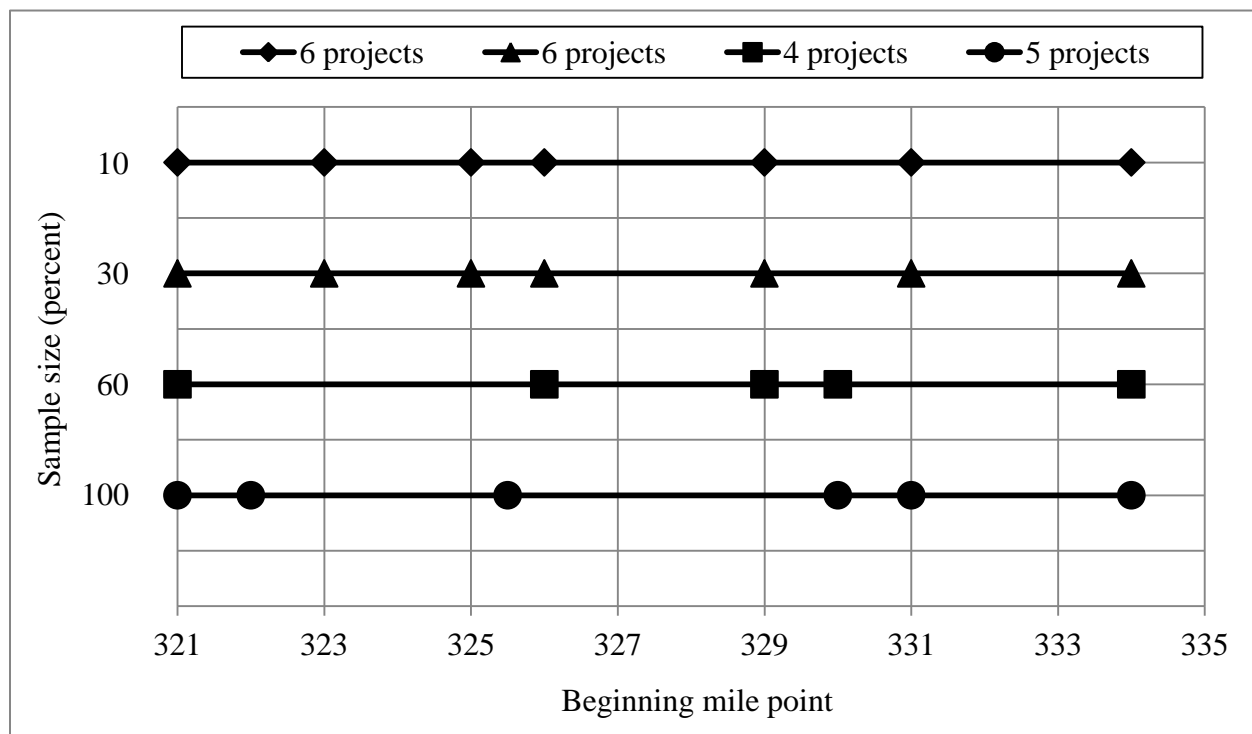


Figure 4.24 Project boundaries based on the continuous (100 percent sampled) and the sampled IRI data along a section of the flexible HWY 24 in Colorado

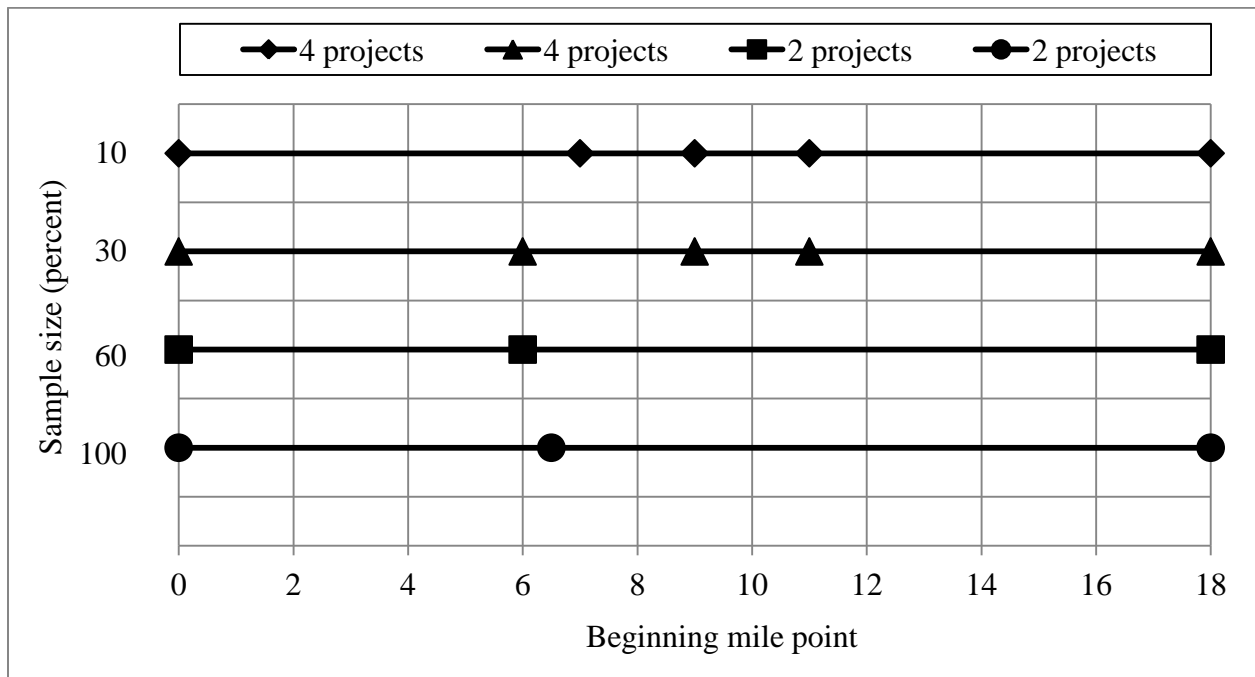


Figure 4.25 Project boundaries based on the continuous (100 percent sampled) and the sampled IRI data along a section of the composite I-94 in Michigan

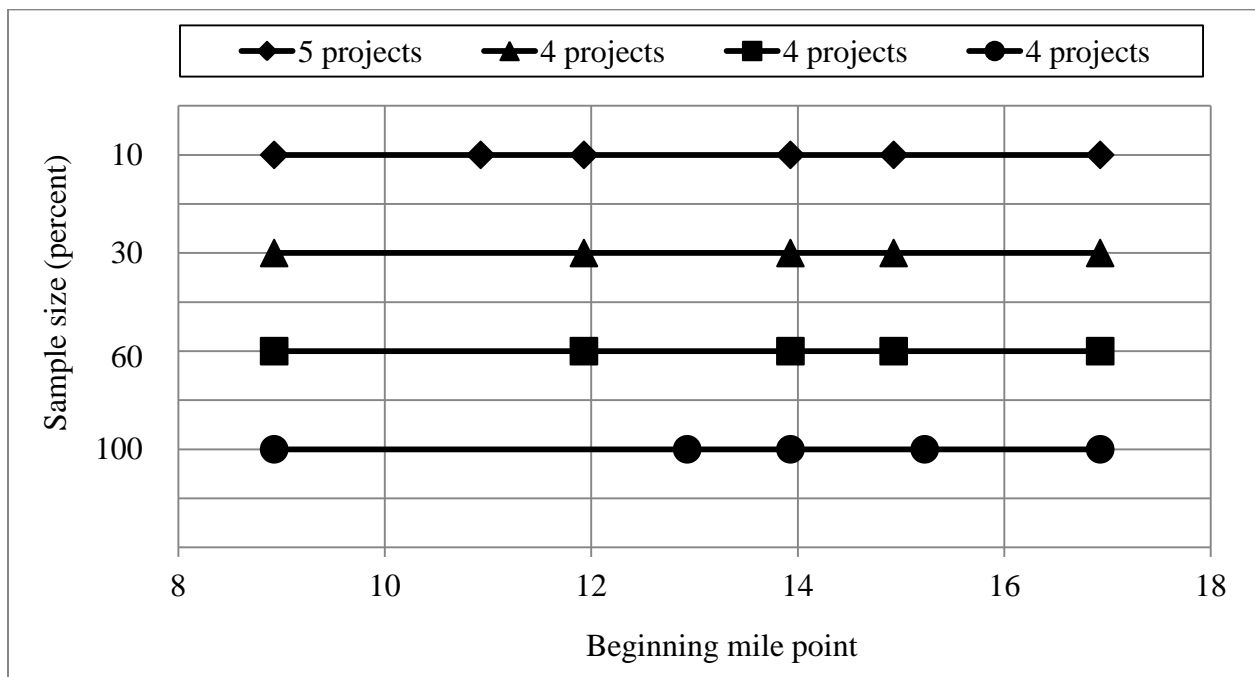


Figure 4.26 Project boundaries based on the continuous (100 percent sampled) and sampled IRI data along a section of the composite SRID 161 in Washington

Examination of the data in Figures 4.24 through 4.26 indicates that:

- For HWY 24 in Colorado, shown in Figure 4.24, five pavement projects are identified based on the continuous IRI data, six on the 10 and 30 percent sampled data and only four on the 60 percent sampled data. Hence each of the 10, 30, and 60 percent sampled data produce an absolute 20 percent difference in the number of projects relative to the continuous data. Further, the first pavement project shown in Figure 4.24 extends from BMP 321 to BMP 322 based on the continuous IRI data. Whereas the 10 and 30 percent sampled data show that the same project extends from BMP 321 to BMP 323; a 100 percent difference relative to the continuous data. The 60 percent sampled data show that the project extends from BMP 321 to BMP 326; a 500 percent difference relative to the continuous data.
- For I-94 in Michigan, shown in Figure 4.25, two pavement projects are identified based on the continuous and the 60 percent sampled IRI data and four projects based on the 10 and 30 percent sampled data. Thus, the 10 and 30 percent sampled data exhibit 100 percent difference relative to the number of projects identified using the continuous data. Further, the length of one of the pavement projects identified along I-94 varies from 11.5 miles long based on the continuous data to twelve miles long based on the 60 percent sampled data and to seven miles long based on the 10 and 30 percent sample sizes. Thus, depending on the sample size, the sampled data produce an absolute difference between about 4 and 39 percent relative to the continuous data.
- Similarly, the results for SRID 161 in Washington, depicted in Figure 4.26, indicate that the variation in the number of pavement projects identified based on the continuous and sampled data varies from 4 based on the continuous and the 30 and 60 percent sampled

data to 5 based on the 10 percent sampled data (a 25 percent difference). The identified pavement project starting at BMP 8.93 varies in length from 4 miles based on the continuous data to 3 miles based on the 30 and 60 percent sample sizes (a 25 percent difference) and to 2 miles based on the 10 percent sample size (50 percent difference).

The above discussion indicates that the pavement project boundaries and the project lengths based on the continuous pavement condition and distress data varies significantly from those based on the sampled data. These variations are a function of the data variability along the road. Higher variability produces higher variations in the project boundaries and project length. These findings have significant implications for SHAs. For instance, if the sampled data indicate that a pavement treatment project is only one mile long, but in reality, the project is three miles long, the SHA may only budget one third of the necessary funds to properly treat the pavement. The opposite scenario is also true; if the sampled data indicate that the pavement project is three miles long, but in reality, the project is only one mile long, the SHA may budget three times the necessary funds to properly treat the pavement. However, in most cases, the above scenarios do not happen. The reason is that most SHAs rely heavily on field inspection before the allocation of pavement expenditures. The main consequence of the differences between the sampled data and the reality in the field is that most of the highway agency staff would lose faith in the PMS data and disregard it. For an ideal PMS scenario, the pavement condition and distress data in the PMS database should reflect the real scenario in the field. Analysis of the data should produce the same or similar results that can be obtained through field inspection. Nevertheless, the use of the sampled data to identify pavement project boundaries may produce significantly different results from the field inspection and could result in significant misallocation of pavement expenditures. This topic is discussed in detail in section 4.5.

Once again, the author realizes that the current state-of-the-practice for pavement project boundary identification in most SHAs does not rely solely on the pavement condition and distress data. Rather, pavement projects are typically identified by employees of the SHAs who travel the pavement network and identify the boundaries by visual inspection of the pavement conditions. The short coming of this approach is that the visual inspection does not tell the full story regarding the pavement condition and distress. The missing element is the pavement rate of deterioration, which can be obtained from the time series pavement condition and distress data. However, if the PMS data show significant differences in the pavement condition and distress than those observed in the field, the employees will lose faith and disregard the PMS data. Unfortunately, the hidden cost of such loss in faith is very high (to the order of several magnitudes higher than the savings due to sampling) and cannot be accurately estimated. In addition, given the current financial situation of many SHAs, some could opt to reduce costs by choosing to make pavement project boundary decisions based solely on the pavement condition and distress data. If the SHA chose to further reduce costs by sampling the pavement condition and distress data, the accuracy and cost-effectiveness of the pavement project boundary identification could be lost and significant misallocation of pavement expenditures could occur as explained in section 4.5.

In this section, results of the analyses of the differences between the continuous and the sampled data in identifying pavement project boundaries based on the pavement condition and distress data were presented and discussed. In the next section, the RSL concept is utilized in order to further examine the impacts of data sampling on the identification of pavement project boundaries. The analysis steps, results, and discussion based on the RSL concept are detailed in the next subsection.

4.4.2 Pavement Project Boundary Analysis: RSL Concept

The AASHTO unit delineation method described in the previous sections was also used to delineate the continuous and sampled RSL data of six pavement sections, two from each of the CDOT, MDOT, and WSDOT. The effects of 10, 20, 30, 40, 50, and 60 percent sample sizes on the identification of pavement project boundaries were analyzed and compared to the project boundaries identified using the continuous RSL data. For each sample size, it was assumed that for each one mile of pavement, the average RSL along the sampled lengths of 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6 mile of each mile independently represent the RSL along the entire mile. Note that the RSL was calculated according to the methods detailed in steps 2 and 3 of section 4.3.1. Regardless, for each pavement section from the three SHAs, the same analysis steps were used and they are detailed below.

Step 1 – Solution Sequence – For the continuous and sampled RSL data along each of the six pavement sections from the three SHAs, the solution sequence of the AASHTO unit delineation method displayed in Table 4.4 and illustrated in Table 4.5 was used to calculate the cumulative differences at each BMP.

Step 2 – Project Boundary Identification – For the continuous and sampled RSL data along each pavement section, the cumulative difference values were plotted as a function of the distance along the pavement section (the BMP), similar to Figures 4.22 and 4.23. Using visual inspection, the project boundaries were defined at the locations where there are positive to negative changes in the slope, or vice versa, of the cumulative differences along the length of the pavement section. Note that, as was the case for the condition and distress data, where the data indicate a pavement project length of less than one mile, the project was included as a part of a longer adjacent project.

Step 3 – Comparison of the Results Based on the RSL – To facilitate clarity and understanding, the project boundaries based on the continuous RSL data (100 percent sampled) and on the 10, 30, and 60 percent sampled data were combined into one figure. The 20, 40, and 50 percent sample sizes showed similar results that were not included in the figure to avoid congestion. Figures 4.27, 4.28, and 4.29 show the pavement project boundaries for the continuous and sampled RSL data based on the IRI along sections of HWY 24 in Colorado, I-94 in Michigan, and SRID 161 in Washington, respectively. In each figure, the data points represent the identified pavement project boundaries. Results of the analysis for the RSL based on the other types of pavement condition and distress data are similar to those shown in Figures 4.27 through 4.29 and are shown in Figures B.56 through B.58 of Appendix B.

Examination of the data in Figures 4.27 through 4.29 indicates that:

- For HWY 24 in Colorado, shown in Figure 4.27, four pavement projects are identified based on the continuous and the 10 percent sampled RSL data. Whereas six projects are identified based on the 30 and 60 percent sampled data. Thus, the 30 and 60 percent sample sizes have 50 percent difference relative to the number of projects based on the continuous data. Further, one of the pavement projects shown in Figure 4.27 is six miles long and extends from BMP 321 to BMP 327 based on the continuous and the 10 percent sampled RSL data. The same project is only two miles long and it extends from BMP 321 to BMP 323 based on the 30 and 60 percent sampled data. Thus, the absolute difference between the 30 and 60 percent sample sizes and the continuous data is nearly 67 percent.

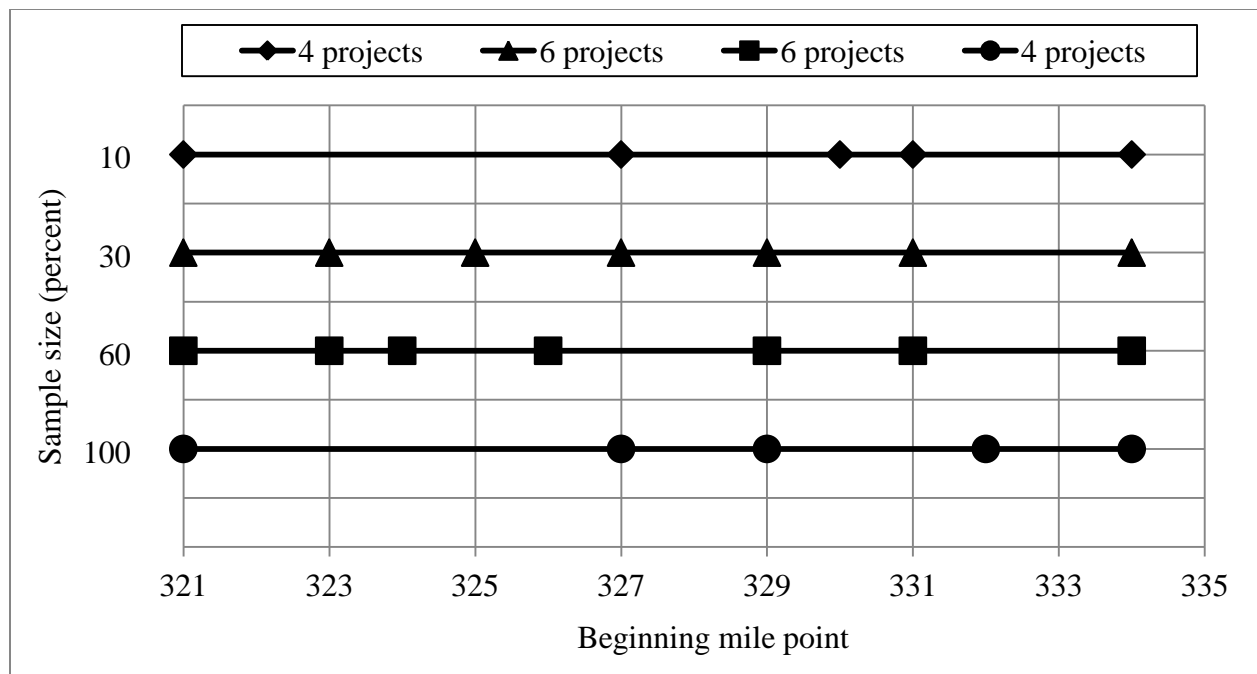


Figure 4.27 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (IRI) data along a section of the flexible HWY 24 in Colorado

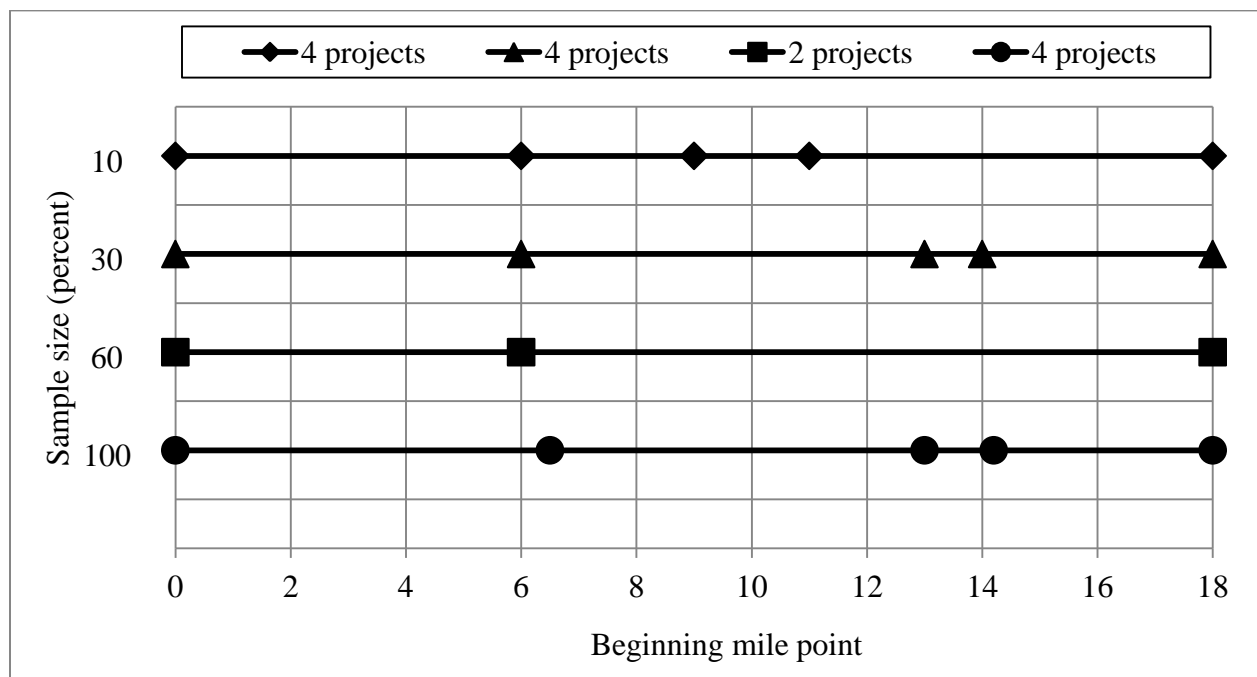


Figure 4.28 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (IRI) data along a section of the composite I-94 in Michigan

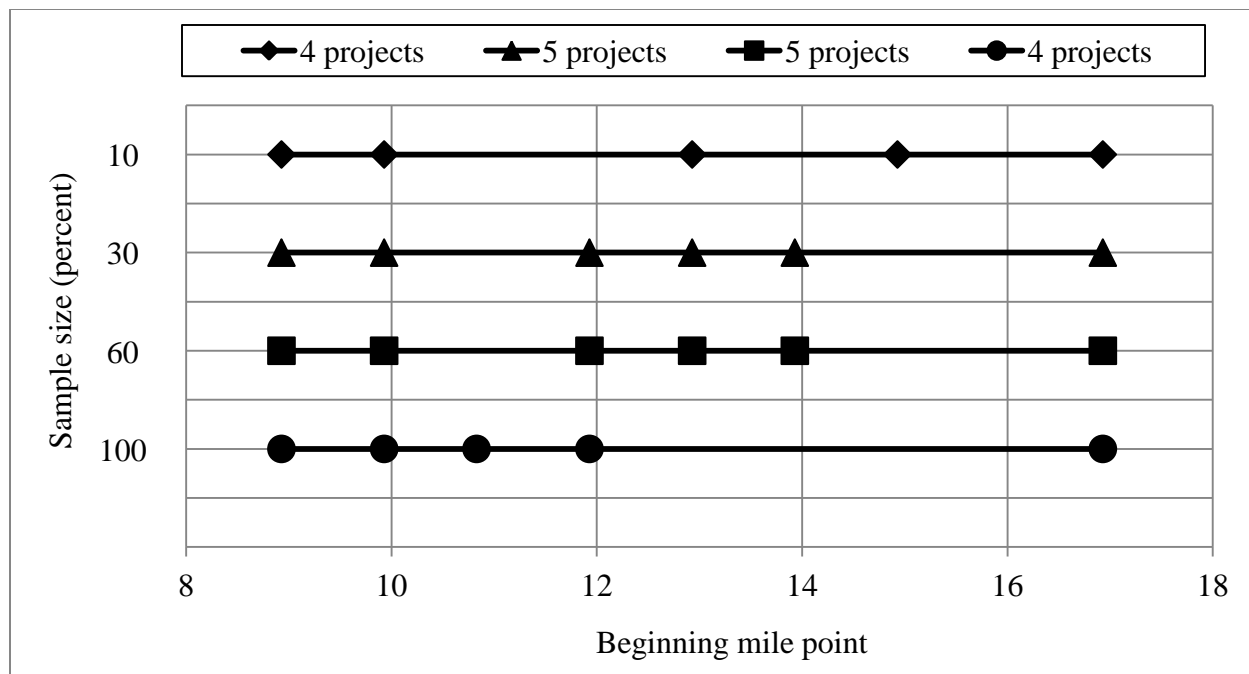


Figure 4.29 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (IRI) data along a section of the composite SRID 161 in Washington

- For I-94 in Michigan, shown in Figure 4.28, four pavement projects are identified based on the continuous, 10 percent, and 30 percent sampled RSL data. However, only two projects are identified based on the 60 percent sampled data (50 percent absolute difference). In addition, the four miles long pavement project, which extends from BMP 14 to BMP 18 based on the continuous data is seven miles long based on 10 percent sampled RSL, 4 miles long based on the 30 percent sampled RSL, and 12 miles long based on the 60 percent sampled RSL. Thus, depending on the sample size, the length of a project based on the sampled data could be three times the length of the same project based on the continuous data.
- For the 8 miles long pavement section along SRID 161 in Washington, shown in Figure 4.29, four pavement projects are identified based on the continuous and 10 percent sampled RSL data while five projects are identified based on the 30 and 60 percent

sample size. Stated differently, the 30 and 60 percent sample sizes exhibit 25 percent difference from the continuous data. Further, the continuous and each of the sample size data produce one mile long project starting at BMP 9. On the other hand, the continuous data also produced a 5 mile long project starting at BMP 12. This project is broken into three different projects based on the 30 and 60 percent sample size and to two and part of a third project based on the 10 percent sample size. In all cases, there are significant differences in the various project lengths depending on the sample size and the variability of the data along the project.

The implications of the above findings are very much similar to those based on the pavement condition and distress data. The next section details the hidden costs of data sampling and possible implications for SHAs.

4.5 The Hidden Costs of Pavement Performance Data Sampling

The analyses results presented and discussed in the previous sections showed that significant differences between the continuous and sampled data sets exist regardless of the data analysis methods or sample sizes. These differences were shown using the pavement condition and distress data, condition state concept, and RSL concept, as well as the AASHTO unit delineation method. Given that the continuous data express the distribution of the pavement condition and distress along the road more accurately than the sampled data; using sampled data only may lead to inaccurate pavement management decisions relative to the true pavement needs. Although data sampling decreases the cost of data collection, its hidden cost may be several folds higher than the savings.

To illustrate this point, consider the total pavement expenditures and the costs of data collection in three SHAs listed in Table 4.6. The total pavement expenditures, calculated as the

total disbursement less the administration and bridge disbursements for each state, ranges from 22,542 \$/lane-mile in Colorado to 13,048 \$/lane-mile in Louisiana and to 17,871 \$/lane-mile for Washington (Hartgen et al. 2009). Whereas the data collection cost ranges from 39.75 \$/lane-mile to 77.67 \$/lane-mile and the data digitization cost ranges from 11.25 \$/lane-mile to 40.00 \$/lane-mile depending on the SHA (Fillastre 2011, Hartgen et al. 2009).

Suppose that the three SHAs listed in Table 4.6 sample the pavement condition and distress data using a ten percent sample size. This entails obtaining pavement images along the entire pavement network and digitizing only ten percent of the pavement network. Using the cost data of each of the three SHAs and the network total lane-miles listed in Table 4.6, the following calculations were made and the results are listed in Table 4.7:

- The total pavement expenditure.
- The cost of obtaining pavement images.
- The cost of data digitization.
- The total data collection costs.

As it was mentioned previously, the total pavement expenditure was estimated as the total disbursement less the administration and bridge disbursements for each state (Hartgen et al. 2009). The potential total savings due to ten percent sampling were also calculated and are listed in Table 4.7. For each state, 90 percent of the data digitization costs could be saved as a result of ten percent sampling. On average, this savings is calculated at about \$709,000. However, if ten percent sampling were to cause a minimum error of only 10 percent, the average potential misallocation, or the hidden cost of sampling based on the total pavement expenditure would be greater than \$44 million. Stated differently, on average, the potential hidden cost (misallocation of funds) due to sampling is more than 63 times the savings. Hence, the potential misallocation

of funds significantly outweighs the savings in data digitization costs. This scenario becomes even more dramatic considering that ten percent sampling produces significantly higher errors than the assumed 10 percent minimum error. It should be noted, though, that the misallocated expenditures may not be totally lost. Rather, the expenditures could be partially lost due to less cost-effective treatment timing and types. Still, even if only a portion of the misallocated expenditures are totally lost, the lost funds significantly outweigh the savings due to ten percent sampling. The most hidden cost of sampling, though, may be the loss of faith in the PMS due to inaccurate decisions based on the sampled data.

Table 4.6 Total pavement expenditure, data digitization, and total data collection costs for three states (Fillastre 2011, Hartgen et al. 2009)

State	Network (lane-miles)	Total pavement expenditure (\$/lane-mile)	Data digitization costs (\$/lane-mile)	Total data collection cost (\$/lane- mile)
Colorado	22,912	22,542	11.25	39.75
Louisiana	38,458	13,048	35.59	77.67
Washington	18,392	17,871	40.00	59.00
Average	26,587	17,820	28.95	58.81

Table 4.7 Total expenditures, potential savings, and potential misallocation of funds as a result of the use of ten percent sampling for various pavement networks

State	Network (lane- miles)	Total pavement expenditure (\$)	Data imaging cost (\$)	Data digitization cost (\$)	Total data collection cost (\$)	Potential total savings due to ten percent sampling (\$)	Potential misallocation of expenditures based on 10 percent error (\$)	Ratio of potential misallocation and total savings for ten percent sampling
Colorado	22,912	516,482,304	652,992	257,760	910,752	231,984	51,648,230	222.6
Louisiana	38,458	501,799,984	1,618,313	1,368,720	2,987,033	1,231,848	50,179,998	40.7
Washington	18,392	328,683,432	349,448	735,680	1,085,128	662,112	32,868,343	49.6
Average	26,587	448,988,573	873,584	787,387	1,660,971	708,648	44,898,857	63.4

4.6 Pavement Analysis Length

Traditionally, the collected pavement condition and distress data are subjected to various types of analysis. The most popular analysis type includes evaluation of the pavement performance over time. In the current state-of-the-practice, the length of the pavement section to be subjected to such performance analysis is constrained by the data collection procedures and by the analysis details. Specifically, these constraints include:

1. The pavement survey length along which the pavement condition and distress data are measured and stored in the database. The length over which the condition and distress data are analyzed, termed the pavement analysis length, cannot be shorter than the pavement survey length. However, it can be longer by averaging the data over two or more adjacent survey lengths. For example, if the pavement survey length is 0.1 mile, the pavement analysis length is 0.1 mile or longer.
2. The employed sampling technique and the length of the pavement section where the collected distress data along the sample size are assumed to represent the pavement condition and distress along that section. The pavement analysis length cannot be shorter than that section; it could be longer by averaging the data over two or more adjacent sections. For example, a SHA could survey one 0.1 mile long pavement segment of each one mile of road and assume the condition and distress along the one mile are the same as those along the 0.1 mile. Hence, the data are stored for each one mile of road. The analysis length cannot be shorter than 1 mile but it can be much longer if the data from two or more adjacent sections are averaged.
3. The opinions of the technical staff of the agency. Some believe that data averaging over longer pavement sections could decrease the data variability and improve data modeling.

The effects of data sampling on the variability of the pavement condition data and on the accuracy of the PMS decisions are discussed in early parts of this chapter. In this section, the effects of the data analysis length on the variability and modeling of the pavement condition and distress data along projects are presented and discussed. The impacts of the pavement analysis length on the accuracy of the PMS decisions are also scrutinized.

4.6.1 The Impacts of Pavement Analysis Length on the Variability and Modeling of Pavement Performance Data

Recall that the CDOT, LADOTD, and WSDOT collect and store the pavement condition and distress data for each 0.1 mile long pavement segment along the entire pavement network. The data from more than 1,632 miles of pavement sections (16,320 0.1 mile long pavement segments). Thus, for each data collection cycle, 16,320 data points were used to analyze the data over pavement analysis lengths of 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 mile. The data analysis steps presented below were used to scrutinize the pavement condition data modeling and variability for each of the analysis lengths.

Step 1 – Data Mining – Construction and maintenance records obtained from the CDOT, LADOTD, and WSDOT were examined to identify the boundaries (BMPs and EMPs) of pavement projects where treatments were conducted in the past. The common treatments that were considered include:

1. Thin (< 2.5 inch) HMA overlay of asphalt surfaced pavements.
2. Thick (≥ 2.5 inch) HMA overlay of asphalt surfaced pavements.
3. Single chip seal.
4. Double chip seal.
5. Thin (< 2.5 inch) mill and fill of asphalt surfaced pavements.

6. Thick (≥ 2.5 inch) mill and fill of asphalt surfaced pavements.

For each SHA and for each treatment type, the number of pavement sections or projects that were identified for analysis, their cumulative length, and the total length of all projects in the three states are listed in Table 4.8.

Table 4.8 The number of pavement projects and their total length identified for analysis in each state

Treatment type	Number and total length (miles) of pavement projects in each state						The number of projects and their length (miles) in all states	
	Colorado		Louisiana		Washington			
	Projects	Length	Projects	Length	Projects	Length	Projects	Length
1	70	104.6	18	45.5	48	130.9	136	281
2			59	252	4	22.7	63	274.7
3	69	495.8	46	242.1	7	20.4	122	758.3
4			12	40.5			12	40.5
5	2	9.1	11	36.5	21	93	34	138.6
6			32	139			32	139
Total	141	609.5	178	755.6	80	267	399	1,632.1
Treatment type: 1 = Thin (< 2.5 inch) HMA overlay of asphalt surfaced pavements; 2 = Thick (≥ 2.5 inch) HMA overlay of asphalt surfaced pavements; 3 = Single chip seal; 4 = Double chip seal; 5 = Thin (< 2.5 inch) mill and fill of asphalt surfaced pavements; 6 = Thick (≥ 2.5 inch) mill and fill of asphalt surfaced pavements.								

Step 2 – Analysis Lengths – The time series condition and distress data for each 0.1 mile long pavement segment of each of the 399 pavement projects identified in step one were downloaded from the PMS databases of each state. Prior to the data analysis, a thorough examination of the pavement condition and distress data indicated that:

1. The database has less than three data points for some of the 0.1 mile long pavement segments and hence, these segments cannot be included in the analysis (see acceptance criteria 1 in step 4 below).

2. The rate of deterioration of a significant number of 0.1 mile long pavement segments is negative (self healing) and hence, the data cannot be modeled.
3. The first data collection cycle after treatment was accomplished a few months to two years after treatment. Consequently, the first measured pavement condition and distress data do not correspond to the time immediately after treatment. Although, the initial pavement condition and distress immediately after treatment are not known, for some treatments it can be assumed based on common sense and the state-of-the-practice. For example, for thin and thick HMA overlay treatments and for thin and thick mill and fill treatments, it can be reasonably assumed that the rut depth immediately after treatment is 0.0 inch. In this study, a rut depth of 0.01 inch rut depth at 0.01 year after treatment was assumed (0.0 inch rut depth could not be assigned because of power function modeling requirements). Such assumption increased the number of data points to be included in the analysis, and increased the number of 0.1 mile long pavement segments that passed the second acceptance criterion (see step 4 below). The above scenario implies that, in this study, the rut depth data was modified by adding 0.01 inch rut depth data point immediately after the treatment was completed.

Depending on the pavement condition and distress types, two methods were used to calculate the pavement condition and distress data for pavement analysis lengths of 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 mile. The two methods are enumerated below.

1. IRI and rut depth – The IRI and rut depth data for each analysis length were calculated as the respective average IRI or rut depth of the corresponding number of 0.1 mile long segments making up the analysis length. For example, the IRI and rut depth data for each one mile long analysis length is the average of the ten data points corresponding to the ten 0.1 mile long pavement segments along the one mile.

2. Cracking – The cracking data for each analysis length were calculated as the sum of the data of the corresponding number of 0.1 mile long pavement segments. For example, the cracking data for each one mile analysis length is the sum of the ten data points corresponding to the ten 0.1 mile pavement segments along the mile.

Step 3 – Number of Analysis Sections - Although the analysis length methodology is simple in nature, it is necessary to address the analysis technique employed at the end of pavement projects. To illustrate, consider a 2.2 mile long pavement treatment project. When the data is analyzed over a length of 0.5 miles, four 0.5 mile analysis lengths were obtained accounting for the first two miles of the project. The 0.2 mile long left over pavement segment was analyzed using the data from two 0.1 mile pavement segments and is counted as 0.4 or 40 percent of the 0.5 mile length. Therefore, in this example, the 2.2 mile long project length was analyzed using 4.4 analysis lengths of 0.5 mile. In more general terms, the remaining length at the end of a pavement treatment project is analyzed and counted as a fraction of an analyzed section. For illustrative purposes and for each analysis length, the number of analysis sections for the 2.2 mile long project is listed in Table 4.9. The same methodology was used for each pavement treatment project regardless of the pavement or the treatment types.

Note that if the data of a certain 0.1 mile long pavement segment was missing, they were treated as a blank cell during the analysis. For instance, if an analysis length included one or more of 0.1 mile pavement segments without data, the value of the pavement condition and distress for the analysis length was calculated as the average or the sum (depending on the condition or distress type) of the remaining 0.1 mile long pavement segments with data. If all of the 0.1 mile long segments did not have data, the analysis length was assigned no data (blank cell). It is estimated that this method was used to account for blank cells in less than one percent of the analyzed data.

Table 4.9 Number of analysis sections for various analysis lengths along a 2.2 mile long project

Analysis length (mile)	Number of analysis sections for a 2.2 mile long pavement project
0.1	22.00
0.2	11.00
0.3	7.30
0.4	5.50
0.5	4.40
0.6	3.67
0.7	3.14
0.8	2.75
0.9	2.44
1.0	2.20

Step 4 – Data Acceptance Criteria – For each pavement treatment and analysis length identified in steps one and two, it is possible that some data elements were not collected along portions of the project in some years. It is also possible that the collected pavement condition and distress data show decreasing distress over time without the application of any pavement treatment. For these reasons, one may not be able to model the pavement condition data. Therefore, for each pavement project and analysis length, the pavement condition and distress were subjected to two acceptance criteria tests to determine whether or not the data could be modeled using non-linear mathematical functions. The two acceptance criteria are:

1. Acceptance criteria one - Three data points – A minimum of three data points before treatment (BT) and a minimum of three data points after treatment (AT) are required to fit any non-linear functions, since any model can be fit to two or one data points. Any pavement segment that did not meet this criterion was excluded from any further analysis.
2. Acceptance criteria two - Positive regression parameters – Positive BT and AT regression parameters of the exponential and power functions are required for time series distress data that are consistent for pavement sections that did not receive any treatment. Negative regression

parameters imply that the pavement is healing itself with time (decreasing distress) without any treatments, which is not reasonable and would bias the analysis. One of the reasons for negative regression parameters is the data variability over time.

Step 5 – Percent Acceptance – For each pavement treatment project (treatment type) and analysis length in each state, and for each condition and distress type, the percent of the pavement segments that passed the two acceptance criterion was calculated. The results for two common treatment types (thin HMA overlay and single chip seal) from each state are shown in Figures 4.30 through 4.35. Figures 4.30 through 4.32 display the results of the thin HMA overlay treatment in Colorado, Louisiana, and Washington, respectively. Figures 4.33 through 4.35 display the results for the single chip seal treatment in each state. Results of the analysis for the other treatment types in each state are similar and are shown in Figures C.1 through C.7 of Appendix C.

It should be noted that an insignificant number of 0.1 mile pavement segments were rejected based on acceptance criteria one. Most pavement segments had at least three data points BT and three data points AT. Therefore, most pavement segments were rejected because they did not satisfy acceptance criteria 2 (positive regression parameters). It is estimated that, of those pavement segments that were rejected, 90 percent did not exhibit positive time series regression parameters.

Examination of the data shown in Figures 4.30 through 4.32 for the thin HMA overlay treatment in Colorado, Louisiana, and Washington indicates that the percent acceptance for most pavement conditions and distress types fluctuate up and down with increasing analysis length without following a specific trend. Hence, the results indicate that increasing the pavement analysis length does not necessarily increase the number of pavement sections that can be modeled. For instance, the IRI data shown in Figure 4.30 indicates that:

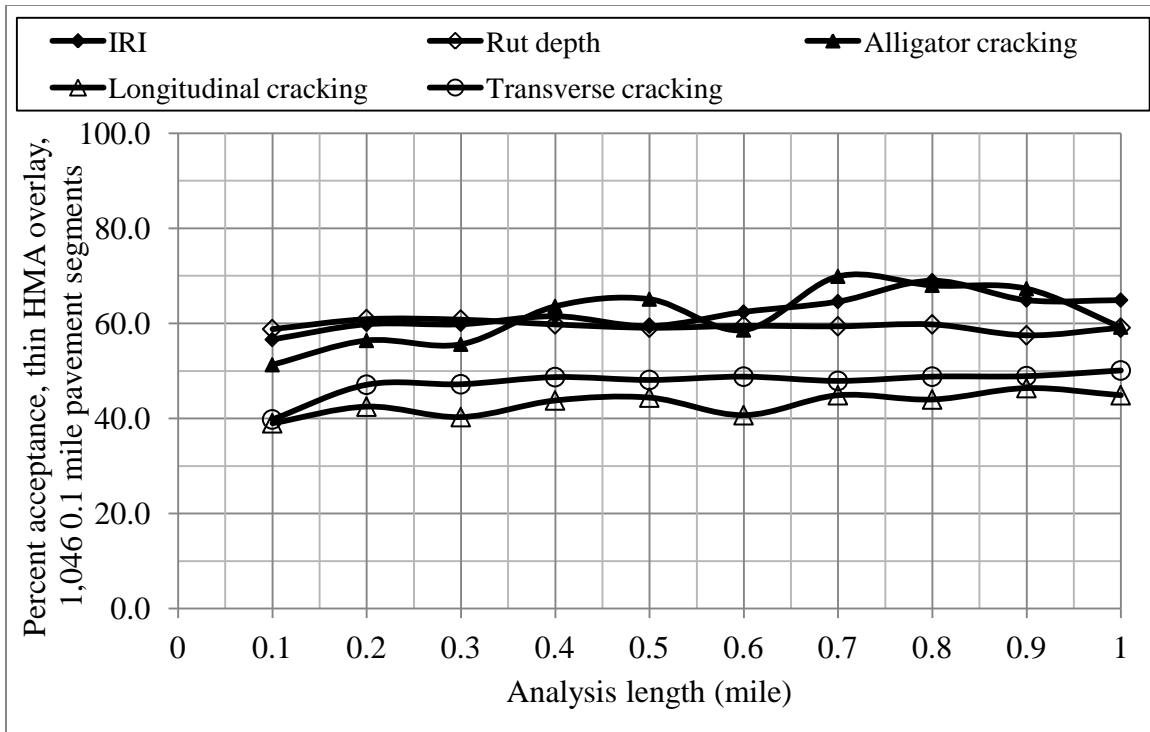


Figure 4.30 Percent acceptance versus pavement analysis length, thin HMA overlay, Colorado

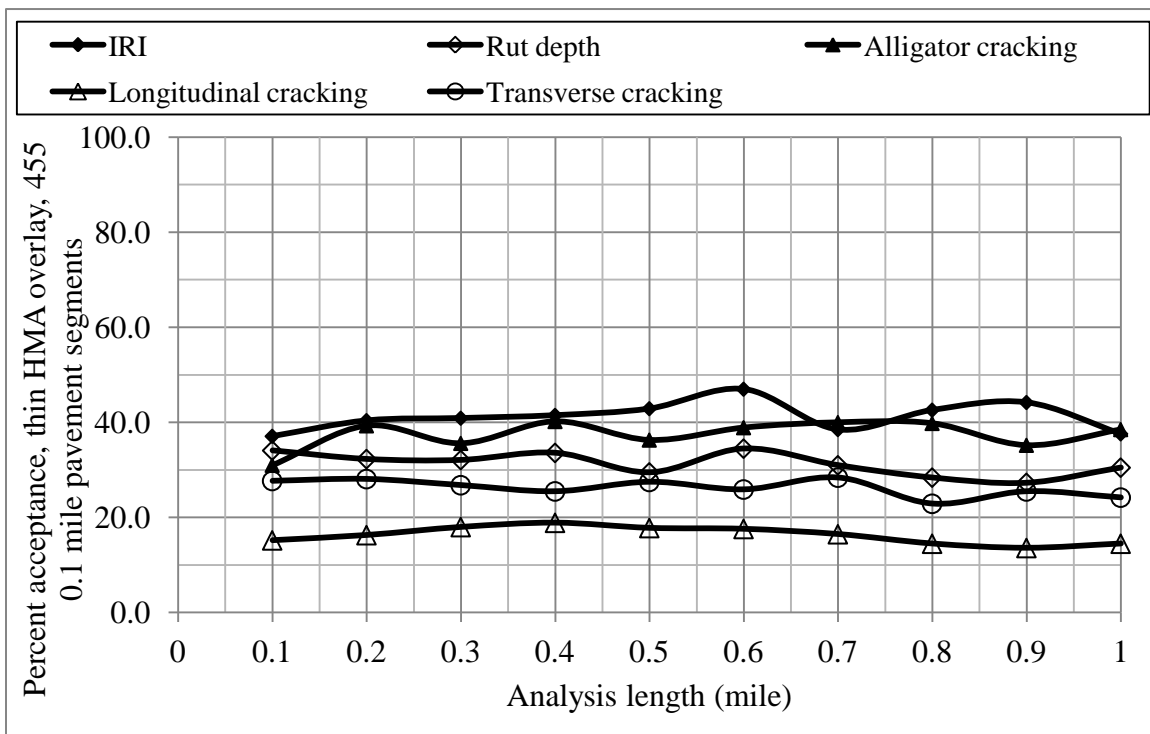


Figure 4.31 Percent acceptance versus pavement analysis length, thin HMA overlay, Louisiana

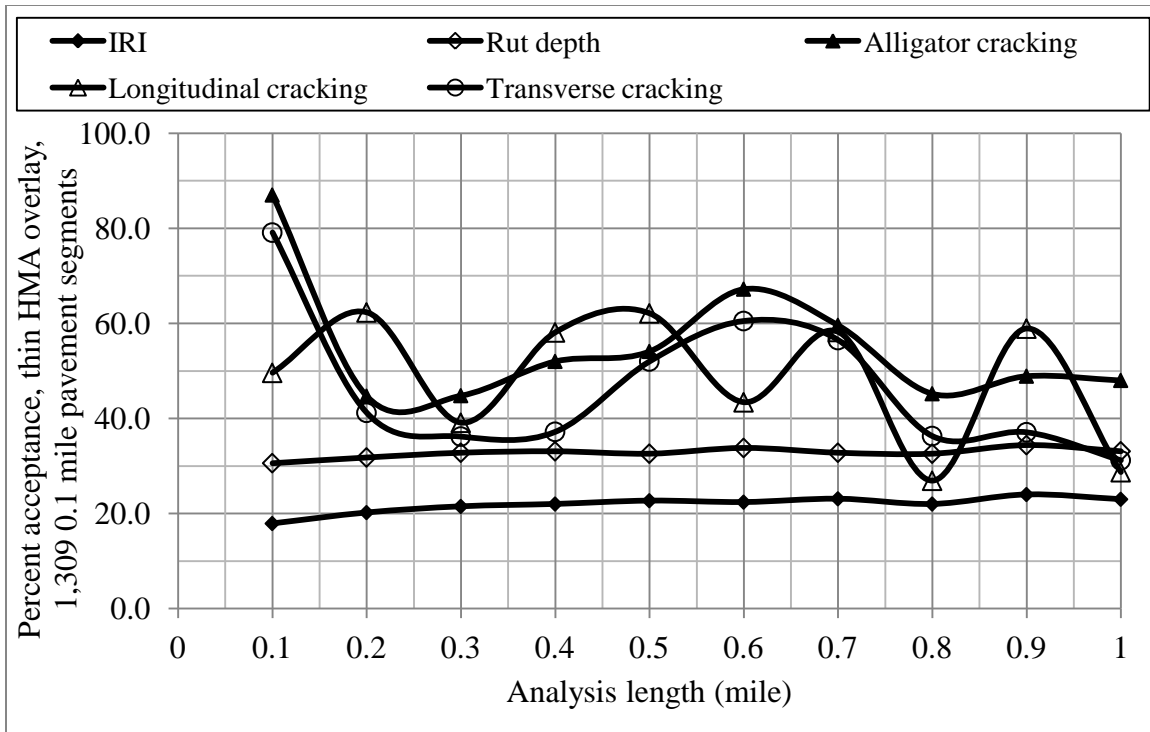


Figure 4.32 Percent acceptance versus pavement analysis length, thin HMA overlay, Washington

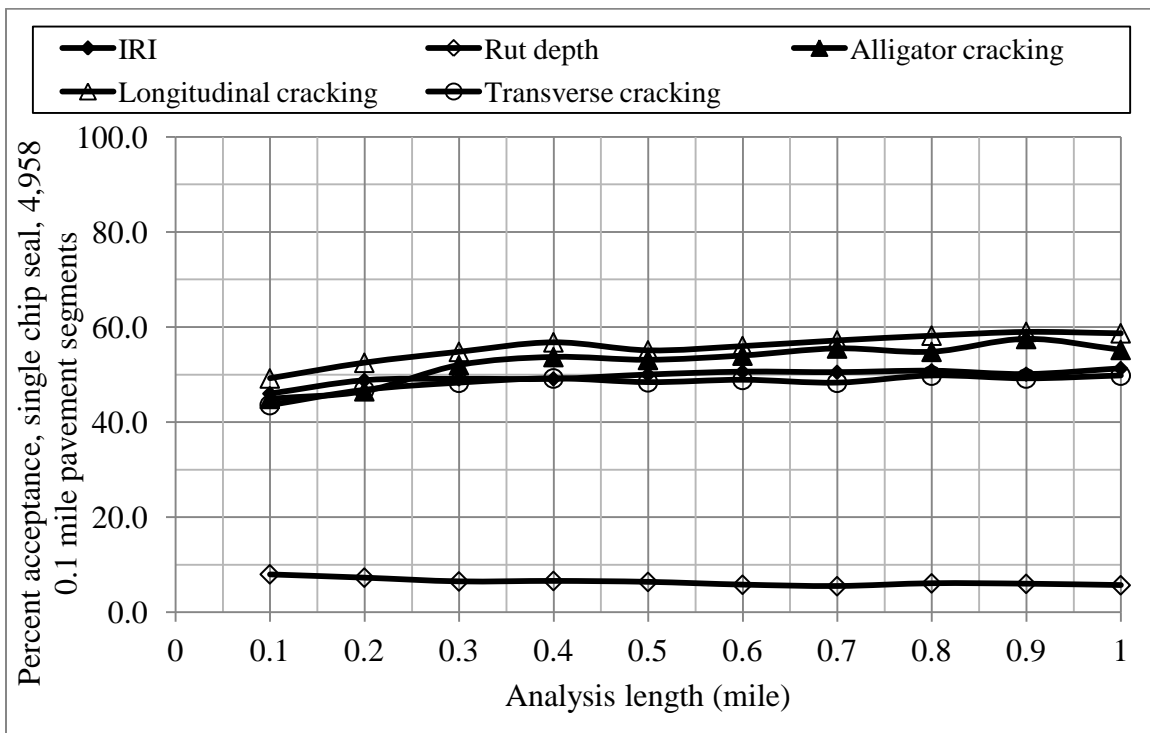


Figure 4.33 Percent acceptance versus pavement analysis length, single chip seal, Colorado

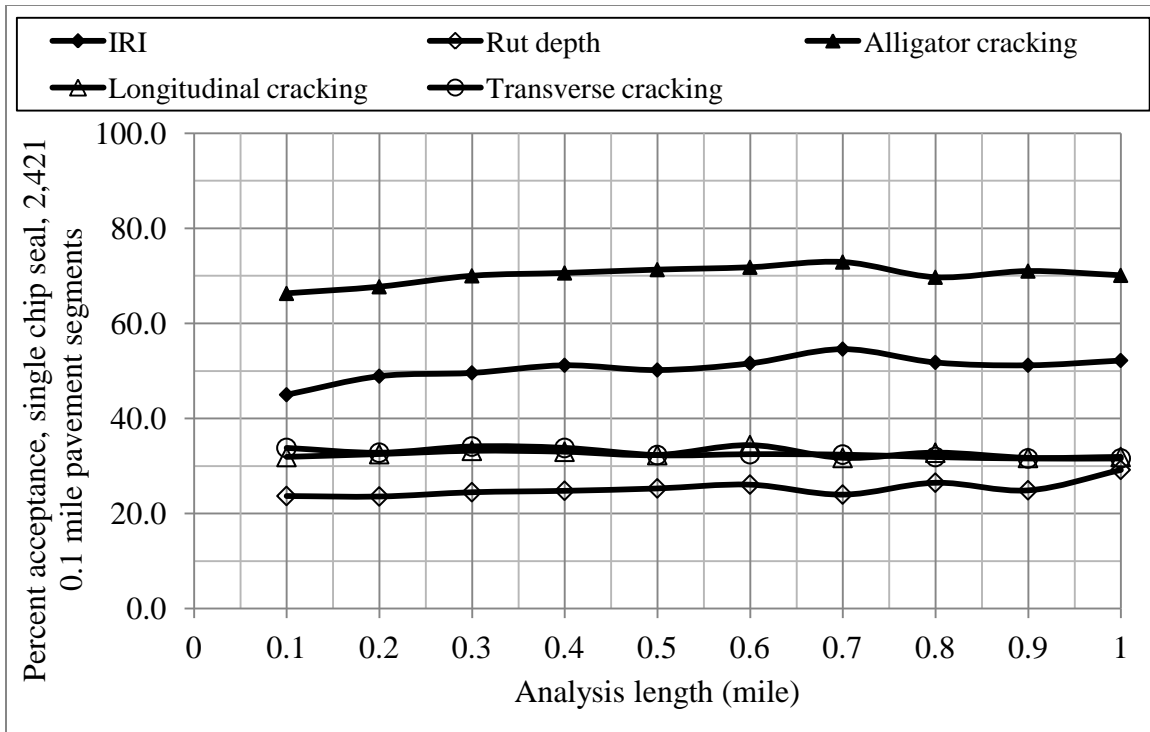


Figure 4.34 Percent acceptance versus pavement analysis length, single chip seal, Louisiana

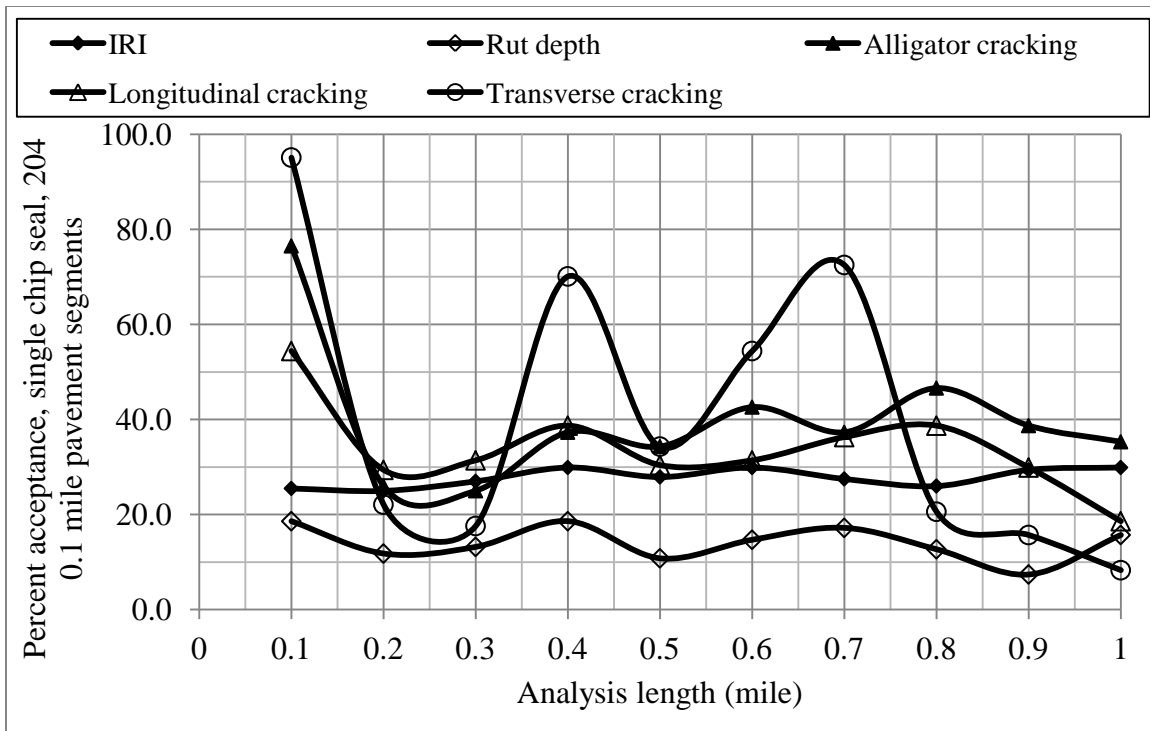


Figure 4.35 Percent acceptance versus pavement analysis length, single chip seal, Washington

1. The percent acceptance remains around 60 percent as the analysis lengths increases from 0.1 mile to 0.5 mile.
2. The percent acceptance increases from about 60 to about 70 percent as the analysis length increases from 0.5 mile to 0.8 mile.
3. The percent acceptance decreases from about 70 to about 65 percent as the analysis length increases from 0.8 to 1.0 mile.

On the other hand, the percent acceptance data, for the rut depth in Figure 4.30 are almost constant and they do not change as the analysis length increases from 0.1 mile to one mile. Similar trends can be seen for the cracking data in the figure. Figure 4.31 depicts the analysis results of the percent acceptance for the thin HMA overlay treatment in Louisiana. It can be seen that the results are similar to those from the same treatment type in Colorado. Increases in the analysis length from 0.1 to one mile do not precipitate significant increases in the percent acceptance.

Similar scenarios can be seen in Figure 4.32 for the IRI and rut depth data of the thin HMA overlay treatment in Washington. The percent acceptance for the IRI data increases from slightly less than 20 percent at 0.1 mile analysis length to slightly greater than 20 percent at one mile analysis length. Furthermore, the percent acceptance for the rut depth data increases from about 30 percent at 0.1 mile analysis length to slightly greater than 30 percent at one mile analysis length. However, the percent acceptance for the cracking data is highly variable and shows no consistent trend with the analysis length.

As mentioned previously, the low percentages of acceptance are mainly due to the second acceptance criteria of positive regression parameters. This is especially true for the time series cracking data. That is, a significant percentage of the time series cracking data do not display

trends that are consistent with pavements that have received no treatment. This can be attributed to several different factors, including data variability. Such variability is a function of many things including the data collection procedures, ambient conditions (temperature and moisture) during the data collection which vary from one year to the next, improved equipment and so forth. The data collection procedure has more impact on the cracking data than on sensor collected data such as IRI and rut depth. The reason is the semi-automated crack data collection technique used to measure and record cracking data versus the fully automated technique for the sensor based data. In the semi-automated technique, a data collection vehicle travels along the pavement and obtains electronic images of the pavement surface. The images are reviewed by trained personnel who digitize the cracking data and record them in the database. Depending on the personnel, cracks could be rated differently due to the inherent human subjectivity of the crack rating. Other variables such as the temperature at the time of the data collection, moisture, lighting, and camera type could also add variability to the subjective crack rating. Thus the trends in the cracking data are somewhat subjective from one year to the next and the percent acceptance is highly variable from one pavement analysis length to another. Unlike the image based data, the IRI and rut depth data are collected and recorded using fully automated procedures. Lasers mounted on the front of a data collection vehicle measure and recorded the IRI and rut depth along the pavement. Thus, these pavement conditions are not subjectively measured and recorded. Hence, compared to the cracking data, the IRI and rut depth exhibit more consistent trends, and the percent acceptance is not as variable from one pavement analysis length to another.

Note that the cracking data for the thin HMA overlay treatments in Colorado and Louisiana, shown in Figures 4.30 and 4.31, respectively, exhibit less variability from one pavement analysis length to another compared to the cracking data from Washington. This could be caused by various reasons including the pavement surface age and the methods of data

collection (sealed cracks are not counted in Louisiana and Colorado). In summary, the percent acceptance based on the cracking data varies significantly from one data analysis length to the next. On the other hand, the percent acceptance based on the IRI and rut depth is almost constant and it shows no specific relationship to the pavement analysis length.

At the onset of this analysis, it was thought that averaging or summing the pavement condition and distress data over long pavement segments will decrease the data variability along the project and over time and would increase the percent of the data that satisfy the two acceptance criteria. Results of the analysis indicate that:

1. As it was expected, the data variability along the project decreases as the pavement analysis length increases. This is evidenced by the fact that, in most cases as the pavement analysis length increases the coefficient of variation of the data decreases. An example of such trend is shown in Table 4.10 for the IRI and longitudinal cracking data along a 9.6 mile portion of US 2 in Washington. For each data collection year shown in the table, the coefficient of variation generally decreases as the pavement analysis length increases. Furthermore, the data in this table also support the discussion regarding the variability of sensor-based data versus image-based data. That is, the data in Table 4.10 indicate that the coefficients of variation of the cracking data are much higher than those of the IRI data. To illustrate, the ratios of the coefficients of variation of the cracking data to those of the IRI data were calculated and are shown in Figure 4.36 for three years of data collection. It can be seen from the figure that the ratio of the coefficient of variation varies from just above 4 to just above 9.
2. The percent acceptance is not impacted by the pavement analysis length. That is, increasing the pavement analysis length does not change the overall trend of the

Table 4.10 Coefficients of variation for ten pavement analysis lengths based on three years of IRI and longitudinal cracking data along US-2 in Washington, BMP 89.3 to 99.0

Pavement analysis length (mile)	Coefficient of variation for each distress type and data collection year					
	IRI			Longitudinal cracking		
	2003	2004	2005	2003	2004	2005
0.1	0.28	0.32	0.27	2.58	1.80	1.90
0.2	0.25	0.28	0.23	1.77	1.58	1.44
0.3	0.22	0.25	0.21	1.40	1.39	1.24
0.4	0.19	0.25	0.20	1.33	1.23	0.94
0.5	0.16	0.22	0.19	1.34	1.05	1.03
0.6	0.16	0.21	0.17	1.06	1.02	0.86
0.7	0.14	0.18	0.14	1.05	1.13	0.68
0.8	0.14	0.17	0.15	0.90	0.87	0.64
0.9	0.13	0.18	0.14	0.99	0.99	0.72
1.0	0.15	0.18	0.15	0.92	0.98	0.73

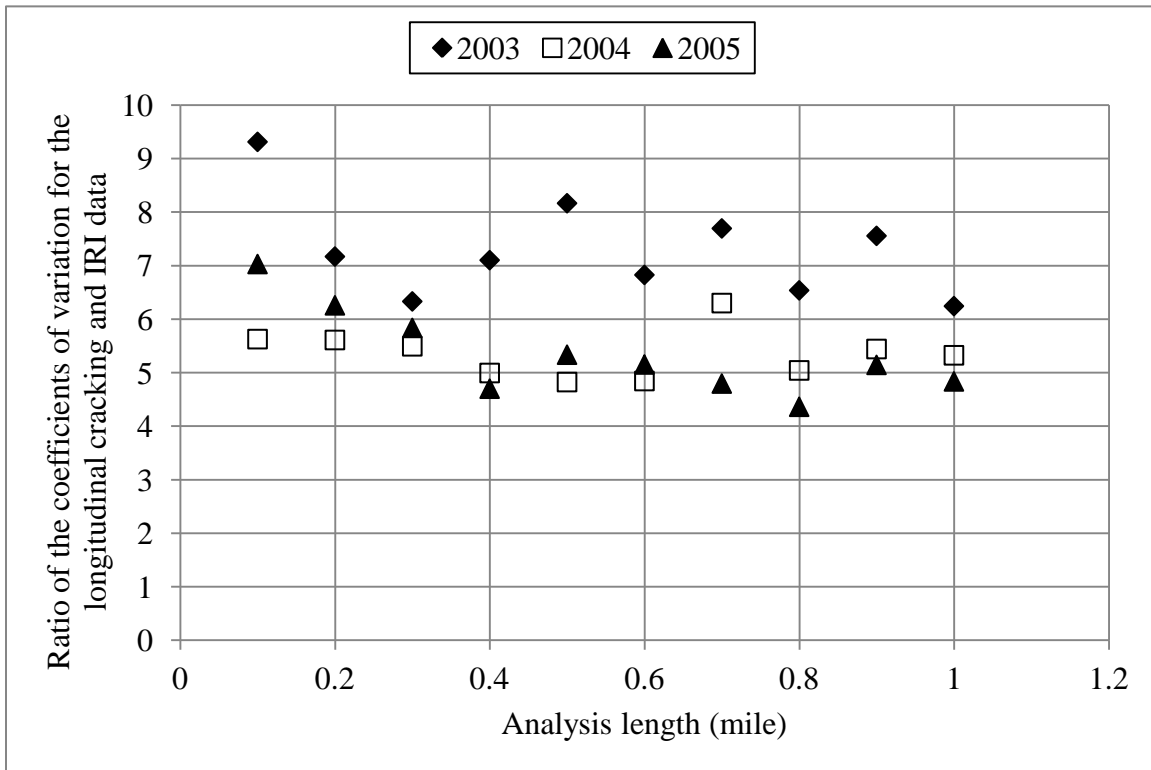


Figure 4.36 Ratio of the coefficients of variation for the longitudinal cracking and IRI data versus pavement analysis length along US-2 in Washington, BMP 89.3 to 99.0

pavement condition and distress data. Stated differently, if the time series data of the 0.1 mile long pavement segments show decreasing trend from one year to the next, averaging or summing the pavement condition or distress over longer analysis lengths does not, in most cases, change the overall trend of the data. Hence, the percent acceptance is unchanged as shown in Figures 4.30 through 4.32.

The same trends described above for the thin HMA overlay treatment can also be seen for the single chip seal treatment in each of the three states. In every case shown in Figures 4.33 and 4.34, the pavement analysis length does not significantly impact the percent acceptance. The same can be said for the percent acceptance of the IRI and rut depth data for the single chip seal treatment in Washington as shown in Figure 4.35. The figure also shows that the percent acceptance for the cracking data is extremely variable and not consistent from one analysis length to another. It is important to note, though, that the data in Figure 4.35 consists of only 204 0.1 mile long pavement segments. Thus, some of the variability in the figure may be attributable to the small number of analyzed pavement segments.

Regardless, the data in Figures 4.30 through 4.35 as well as in Figures C.1 through C.7 in Appendix C indicate that:

1. In general, increasing the pavement analysis length does not necessarily increase the percent acceptance. In other terms, the ability of the data to be analyzed or modeled does not improve as the analysis length increases.
2. For sensor collected data (IRI and rut depth), the percent acceptance is almost constant and independent of the pavement analysis length.
3. For image based data (cracking), the percent acceptance is variable and not related to the pavement analysis length.

4.6.2 The Impacts of Pavement Analysis Length on PMS Decisions

Although the pavement analysis length does not appear to significantly impact the percent of pavement segments that can be modeled or analyzed, the analysis length does have significant impacts on the PMS decisions. Specifically, the pavement analysis length impacts the PMS in three different scenarios.

1. The selection of project boundaries which are directly related to the pavement analysis length. The length over which the pavement condition and distress data are analyzed dictates the selection of the project boundaries. To illustrate, suppose that a SHA chooses to analyze the pavement condition and distress data along each one mile of the road network. This causes the pavement project length to be whole miles such as 1, 2, 3 or more miles. The selection of a project length of 4.5 mile will be arbitrary and not based on the results of the data analysis. On the other hand, if the SHA chose to analyze the data along each 0.1 mile of the road network, the minimum pavement project length would be 0.1 mile. The above scenarios indicate that the pavement analysis length impacts the decisions regarding project boundaries. Therefore, the indirect consequence is that the analysis length also impacts the decisions regarding pavement treatment funds. Longer analysis lengths result in longer pavement treatment projects that, in turn, require the allocation of more pavement treatment funds.
2. The identification of hot spots or areas of localized poor condition or high distress. The averaging or summing of the pavement condition and distress data along longer analysis lengths would artificially decrease the variability of the data along the analysis length. The averaging and summing processes tend to hide areas of localized high distress or poor condition. Again suppose that a SHA chooses to store the pavement condition and

distress data along each one mile of the road network. Such data cannot be used to identify localized hotspots where the pavement conditions and/or distress along a short pavement segment (such as 0.2 mile) are very high. Thus, the PMS database will not have the detailed pavement performance data to identify troubled or highly distressed areas.

3. The statistical significance of the number of analyzed sections or segments. That is the longer is the pavement analysis length, the less is the number of pavement sections that can be subjected to analysis. To illustrate, consider a 7 mile long project. If the analysis length is 1 mile, there are 7 one mile long pavement sections to be analyzed. This is statistically insignificant number. On the other hand if the data is stored and analyzed for each 0.1 mile long pavement segments, there are 70 segments to be analyzed.

Statistically, PMS decisions that are based on 70 data points would have a greater significance than decisions based on 7 data points. Thus, in general, as the pavement analysis length increases, the statistical significance or the confidence level in the PMS decisions decreases as less data points are used to formulate the decisions.

Given the discussion in the previous paragraphs, it is fair to say that the selection of longer pavement analysis lengths could create multiple problems. In the worst case, longer pavement analysis lengths could result in longer pavement treatment projects. Within these longer pavement treatment projects, localized areas of increased distress or poor condition could be masked. Furthermore, the PMS decisions regarding the distress and condition along these segments would be based on less data points and the statistical significance or confidence in the decisions would be lacking. To avoid these problems, and others, it is most beneficial for SHAs to store and analyze the pavement condition and distress data over short pavement lengths such as 0.1 mile.

4.7 Pavement Performance Data Imputation

For various reasons, the pavement performance data from some data collection cycles along certain pavement segments or sections may not be documented in the database. These include equipment malfunctioning, failure to properly record the data, loss of data to human errors, construction zones and so forth. Missing data may pose serious issues regarding the modeling and analysis of the time-series pavement performance data especially when less than three data points are available. Hence, the question herein is what can be done to relatively accurately substitute the missing data elements?

One possible answer to this question is data imputation. The statistics and various methodologies supporting the principles of data imputation are well developed and documented in the literature. However, their applications to the imputation of accurate time dependent pavement performance data are not clear. Therefore, in this study, various analysis techniques were used to examine the effects of four different data imputation methodologies on the accuracy of the imputed pavement performance data. In the analysis, the time dependent IRI data of flexible, composite, and rigid pavement sections were used. The IRI data were obtained from the CDOT, the LADOTD, and the WSDOT as well as from cells of the MnROAD experiment. The four data imputation methods that were analyzed are linear interpolation, regression, moving regression, and multiple regressions. For each analyzed pavement segment, the same analysis steps listed below were used.

Step 1 – Data Limitation and Amputation – In general, for each of the 0.1 mile long pavement segments along the roads in each of the three SHAs that were used in the analysis, only five time series IRI data points were available from five data collection cycles. In reality, more data points were available in the database; however, the additional data points were

collected either before or after treatments were applied. Pavement treatment actions cause changes in the pavement condition and distress and in the pavement rate of deterioration and thus, data imputation cannot be performed using the entire set of data. Regardless, for each analyzed pavement segment, the second, third, and fourth available IRI data points were independently removed (amputated) from the time series data. Hence, three amputated data sets were created; one missing the second data point, one missing the third, and one missing the fourth data point.

Step 2 – Data Imputation – For each amputated data set created in step 1, the linear interpolation, regression, moving regression, and multiple regressions imputation methods were used to independently impute the second, third, and fourth time series IRI data points.

Step 3 – Differences in Imputed and Measured Data – For the second, third, and fourth data points, the imputed and the measured data were compared using two analysis techniques:

1. The percent difference of the imputed data point relative to the measured one.
2. The differences between the RSL value that was estimated using the measured data set and a threshold value of 200 inch/mile and the three RSL values estimated using the same threshold value and each of the imputed data sets. These differences were then compared to determine the effects of the various data imputation methods on the PMS decisions.

Note that the RSL concept is described in detail in section 2.9 of this thesis.

Detailed descriptions, examples, results, and discussion of each of the analyzed data imputation method are presented in the next four sections. It should be noted that the name of each imputation method used in this thesis was selected to promote clarity. The names do not necessarily reference those stated in the literature.

4.7.1 Linear Interpolation Imputation

The linear interpolation imputation method is simple, easy to use, and is based on the following two assumptions;

1. The missing data point is located between two available data points.
2. The two available data points and the missing one can be modeled as a linear function of time.

Hence, the missing data point is imputed by linearly interpolating between two available data points. Graphically, this amounts to imputing the missing data point by connecting a straight line between the two data points on either side of the missing one as stated in Equation 4.1.

$$IRI(t) = \frac{[IRI(t+1) - IRI(t-1)][(t) - (t-1)]}{[(t+1) - (t-1)]} \quad \text{Equation 4.1}$$

Where $IRI(t)$ = the missing IRI data at time t

$IRI(t+1)$ = the available IRI data at time $(t+1)$

$IRI(t-1)$ = the available IRI data at time $(t-1)$

An example of the linear interpolation imputation method is shown in Table 4.11. In this example, the fourth available time-series IRI data was amputated and imputed. The same analysis method was used to amputate and consequently impute the second and third data points.

For the three imputed data points (the second, third and fourth), the percent differences between the measured and the imputed IRI data for five 0.1 mile long pavement segments along HWY 24 in Colorado are shown in Figure 4.37. It can be seen in the figure that, for these five BMPs, the percent differences between the imputed and the measured data range from about -30 to +30 percent and that most imputed data are within -15 to +20 percent of the measured data.

Table 4.11 Linear interpolation imputation example

Time series data number	Data collection year	Elapsed time (year)	Available IRI data set (in/mile)	Amputated IRI data set (in/mile)	Imputed IRI data set (in/mile)	Equation
1	1999	1	74	74	74	
2	2000	2	125	125	125	
3	2003	5	131	131	131	
4	2005	7	122		147	$147 = 131 + \frac{(163 - 131) * (7 - 5)}{(9 - 5)}$
5	2007	9	163	163	163	

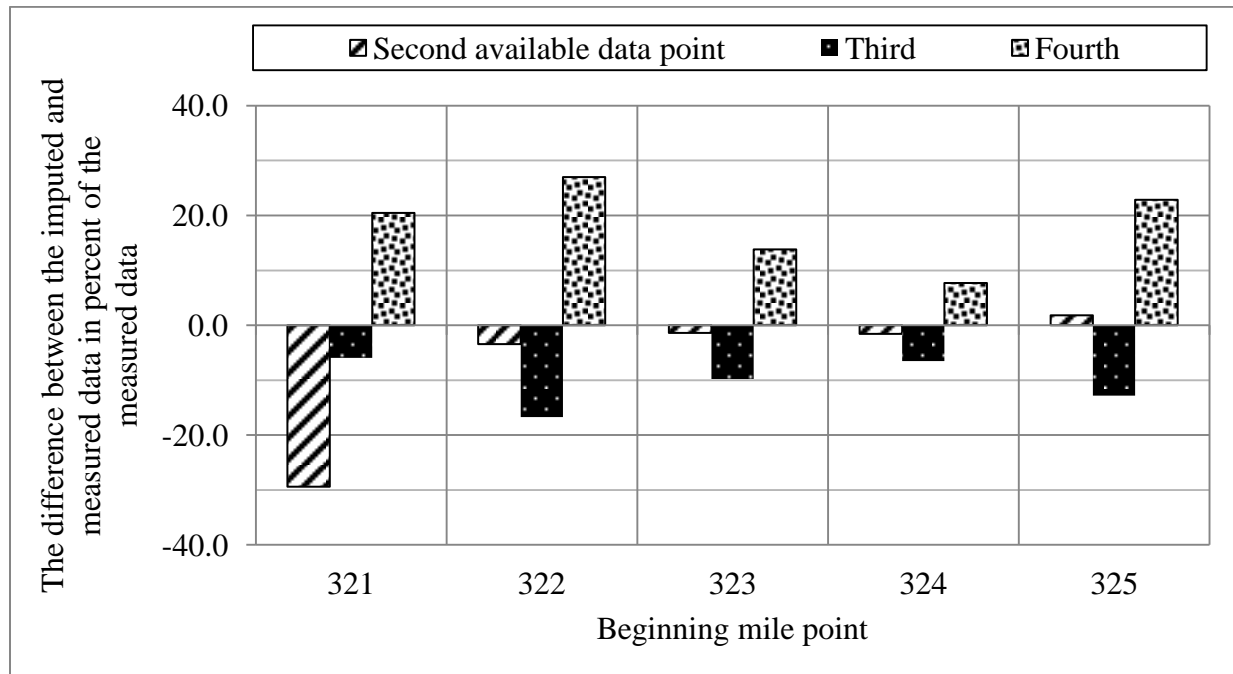


Figure 4.37 The difference between the imputed and measured data in percent of the measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the linear interpolation imputation method

As stated earlier, in the second analysis technique, the measured data points were used to calculate the RSL of each pavement segment. Further, the remaining measured data points and each of the imputed data points were also used to calculate the RSL of each pavement segment. The differences in the RSL values based on the measured and the measured and imputed IRI data were then calculated and are shown in Figure 4.38. The data in the figure indicate that the maximum difference between the RSL values based on the measured and the imputed and the measured IRI data for the five 0.1 long pavement segments along HWY 24 in Colorado is only 1.5 years. Further the majority of the differences in the RSL values are less than 0.5 year. For most cases, such differences in the RSL values would have insignificant impacts on the condition state (see section 4.3) of the pavement segments and hence, would not affect the PMS decisions.

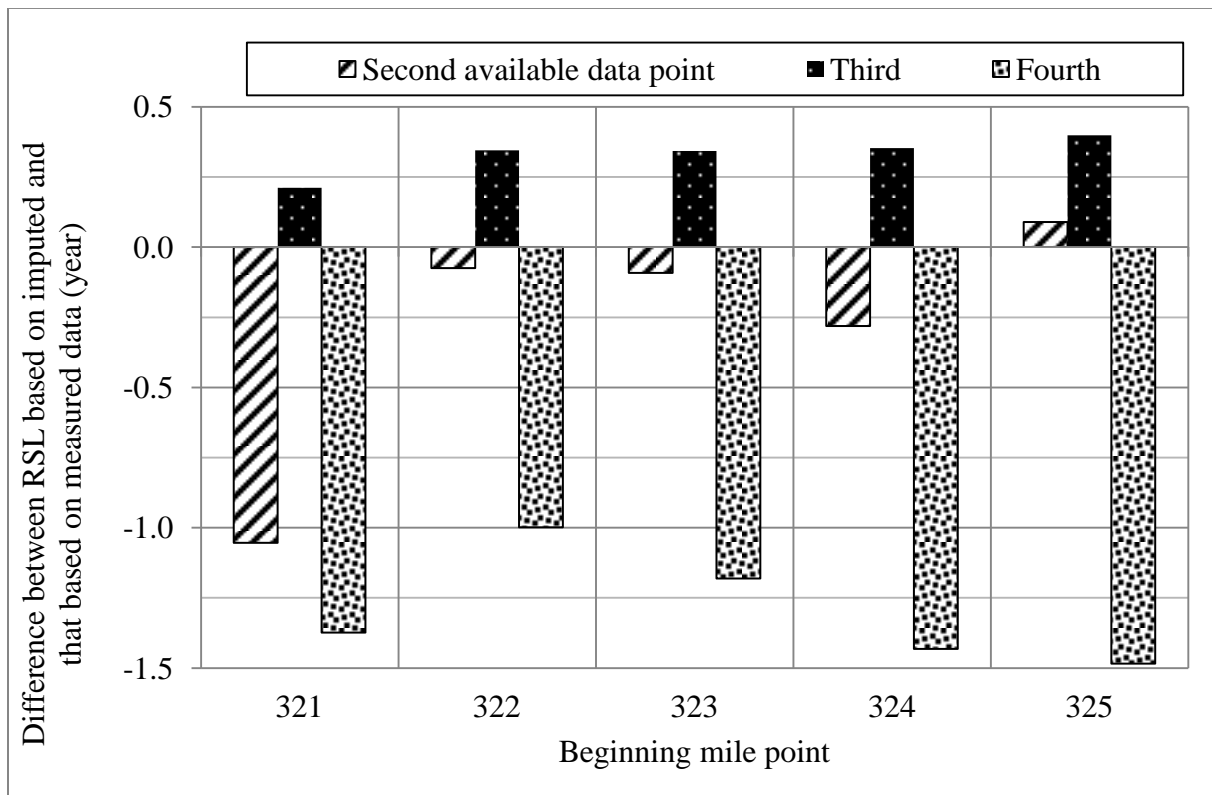


Figure 4.38 The difference between the RSL values based on imputed and measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the linear interpolation imputation method

The above example indicates that although the linear interpolation imputation method produces ± 30 percent difference between the measured and the imputed IRI data, the effect of the method on RSL (and PMS decisions) is negligible. Nevertheless, Figures D.1 through D.6 of Appendix D show similar results that were obtained for 15 pavement segments along other roads located in the other states and the MnROAD experiment.

4.7.2 Regression Imputation

The regression imputation method is used to impute missing data using the proper regression function, which is obtained based on the available information at the time of imputation. For example, consider the data shown in Figure 4.39 and listed in Table 4.12, after amputating the fourth data point (collected in 2005), the remaining three data points (measured in years 1999, 2000, and 2003) were used to calculate the regression parameters of the exponential function shown in the figure. The amputated (missing) IRI data point was then calculated (imputed) using the exponential function and its regression parameters shown in Figure 4.39 and it is listed in Table 4.12. The same analysis method was used to independently amputate and consequently impute the second and the third data points. It is important to note that three data points are required to perform regression imputation using any non-linear function. This implies that, depending on the year of the missing data point, a SHA has to wait anywhere from zero to four years to impute the missing data. To illustrate, if the second data point in a time series data is missing, and if the state collects data every other year, the missing data point can be imputed using the regression method after four years when two additional data collection cycles are completed. On the other hand, if the fourth data point is missing (three data points are available), the missing data point can be imputed immediately using the regression method presented above.

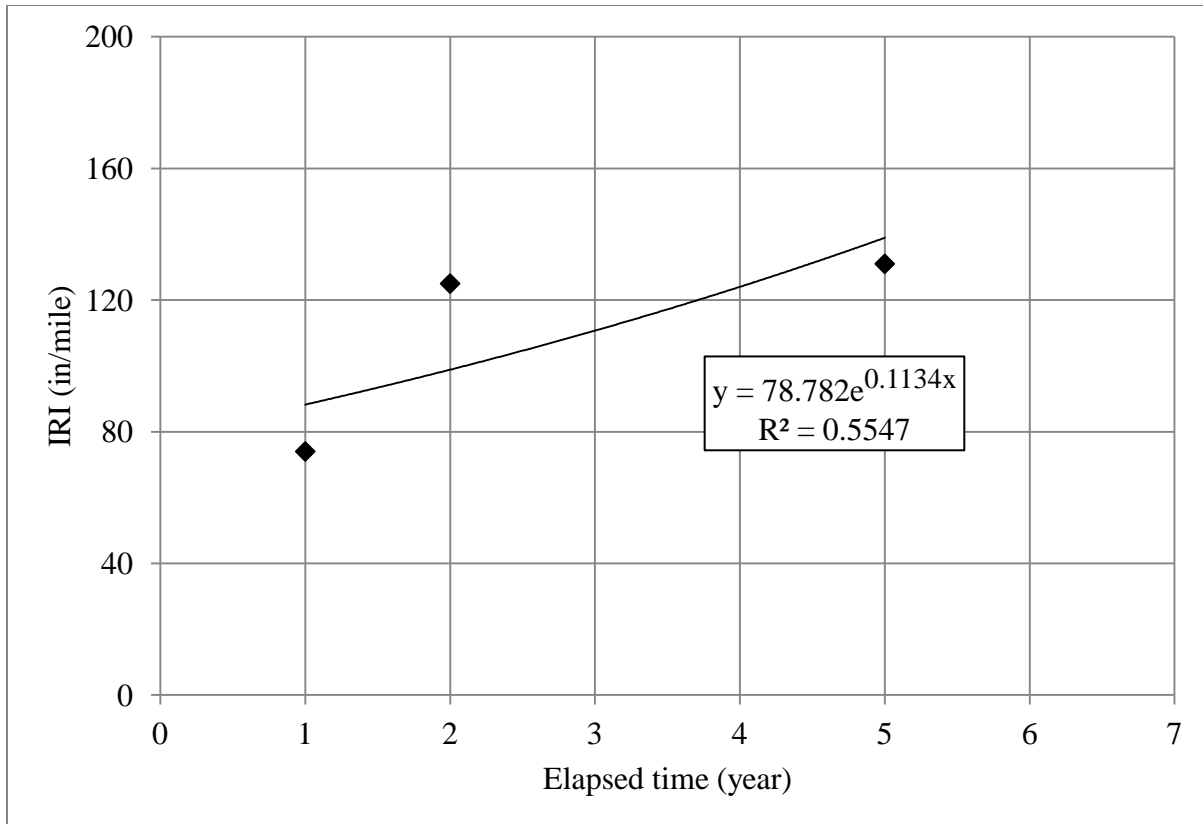


Figure 4.39 Example regression parameter calculation for regression imputation, missing data at year 7

Table 4.12 Regression imputation example

Time series data number	Data collection year	Elapsed time (year)	Available IRI data set (in/mile)	Amputated IRI data set (in/mile)	Imputed IRI data set (in/mile)	Equation
1	1999	1	74	74	74	
2	2000	2	125	125	125	
3	2003	5	131	131	131	
4	2005	7	122		174.2	$174.2 = 78.782e^{(.1134 \times 7)}$

Like the linear interpolation method, for the three imputed data points, the percent differences between the measured and the imputed IRI data for five 0.1 mile long pavement segments along HWY 24 in Colorado are shown in Figure 4.40. Note that these are the same five BMPs that were presented for the linear interpolation imputation method. It can be seen in the figure that the percent differences between the imputed and measured data range from less than -30 to approximately +50 percent. Most of the imputed data, though, lie within ± 20 percent of the measured data.

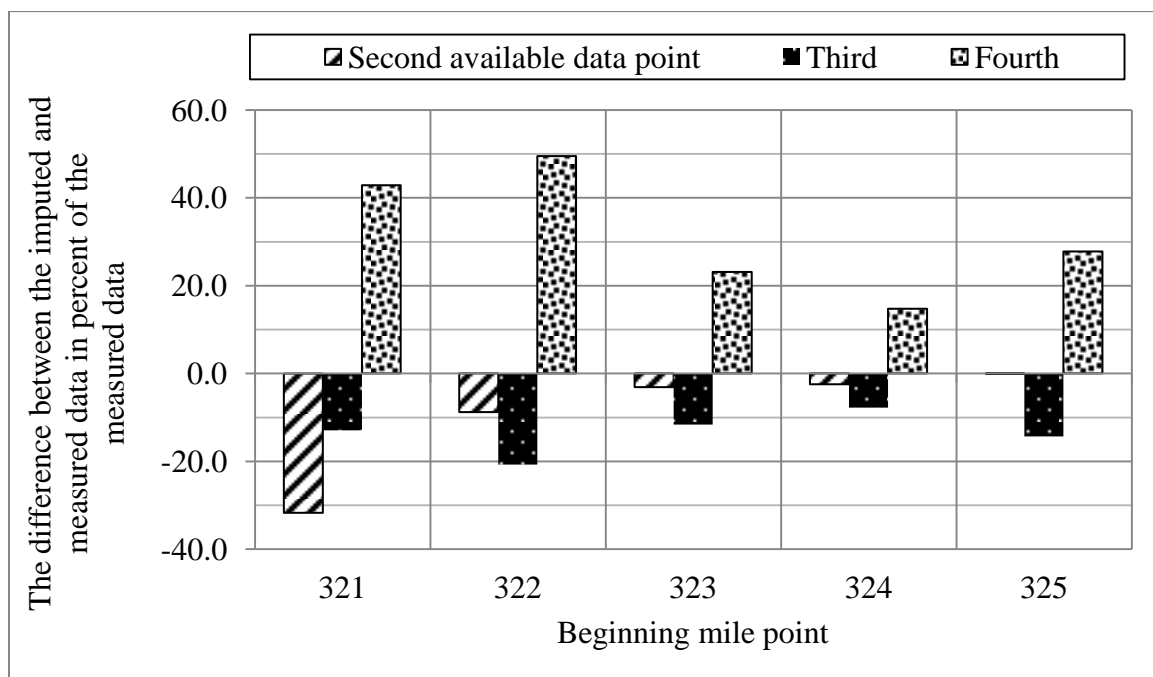


Figure 4.40 The difference between the imputed and measured data in percent of the measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the regression imputation method

Further, the RSL of each pavement segment was calculated based on a threshold value of 200 inch/mile and the measured and the measured and imputed data points. The differences in the RSL values based on the measured and the measured and imputed IRI data are shown in Figure 4.41. The data in the figure indicate that the maximum difference between the RSL values based on the measured and the imputed and measured IRI data for the five 0.1 mile long

segments along HWY 24 in Colorado is nearly 2.5 years. However, the majority of the differences in the RSL values are less than 0.5 year. Once again, in most scenarios, such differences in RSL values would have insignificant impacts on the condition state (see section 4.3) of the pavement segments and would not affect PMS decisions.

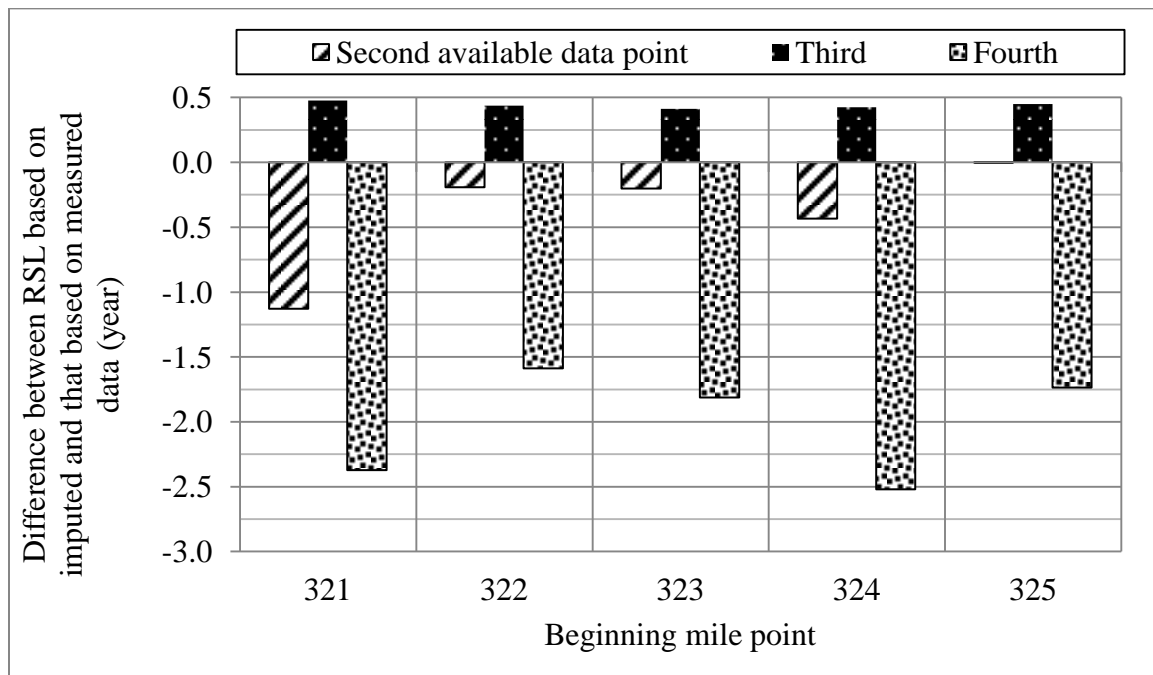


Figure 4.41 The difference between the RSL values based on imputed and measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the regression imputation method

The above example indicates that although the regression imputation method produces, in general, ± 20 percent differences between the measured and imputed IRI data, the effect of the method on the RSL (and PMS decisions) is negligible. Figures D.7 through D.12 of Appendix D show similar results that were obtained for 15 pavement segments along other roads located in the other states and the MnROAD experiment.

4.7.3 Moving Regression Imputation

The moving regression method used in this thesis is a modification of the regression method stated earlier. It uses the proper regression function based on the available information. In this modified method, the imputed data points are updated every time a new data point is measured and becomes available. That is, the imputed data are ‘updated’ based on the recalculated values of the regression parameters of the function after a new data point becomes available. An example of the moving regression imputation method is shown in Table 4.13. In this case, the data at year 7 (fourth data point) was amputated and labeled as missing. The missing data point was then imputed using the regression parameters of the exponential model based on the first, second, and third data points. This is the same procedure as for the regression imputation method presented in the previous section. However, when the fifth data point becomes available, the fourth data point is re-imputed using the regression parameters of the exponential model based on the first, second, third, and fifth data points. The new imputed value (moving regression 2) replaces the first imputed value and the cycle will continue when the sixth data point becomes available. Note that the results of moving regression 1 (the first imputed value) in Table 4.13 and the imputed value using the regression imputation method (Table 4.12) are the same.

For the three imputed data points, the percent differences between the measured and moving regression imputed IRI data for five 0.1 mile long pavement segments along HWY 24 in Colorado are shown in Figure 4.42. It can be seen in this figure that the percent difference between the imputed and measured IRI data ranges from slightly less than -30 percent to about +35 percent with the majority of the imputed data lying within ± 20 percent of the measured data.

Table 4.13 Moving regression imputation example

	Time series data number	Data collection year	Elapsed time (year)	Available IRI data set (in/mile)	Amputated IRI data set (in/mile)	Imputed IRI data set (in/mile)	Equation
Moving regression number 1	1	1999	1	74	74	74	
	2	2000	2	125	125	125	
	3	2003	5	131	131	131	
	4	2005	7	122		174.2	$174.2 = 78.782e^{(.1134*7)}$
Moving regression number 2	1	1999	1	74	74	74	
	2	2000	2	125	125	125	
	3	2003	5	131	131	131	
	4	2005	7	122		146.7	$146.7 = 85.313e^{(.0774*7)}$
	5	2007	9	163	163	163	

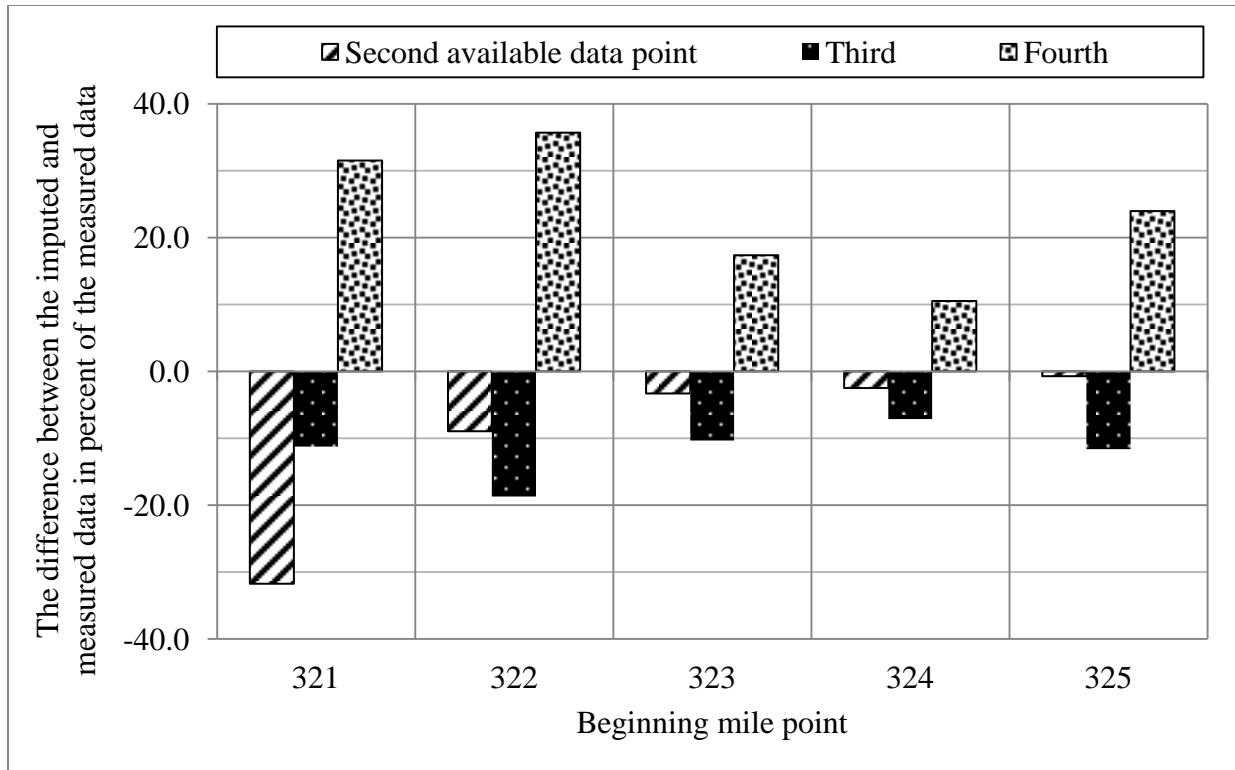


Figure 4.42 The difference between the imputed and measured data in percent of the measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the moving regression imputation method

Similar to the two imputation methods presented previously, the measured and measured and imputed data points were used to calculate the RSL of each pavement segment based on an IRI threshold of 200 inch/mile. The differences in the RSL values based on the measured and measured and imputed IRI data are shown in Figure 4.43. The data in the figure show that the maximum difference between the RSL values based on the measured and the measured and imputed IRI is less than 1.5 years. In fact, most of the differences in the RSL values are less than 1 year. Differences of this magnitude would not cause any significant changes in the condition state (see section 4.3) of the pavement segment. Thus, changes in PMS decisions are not expected.

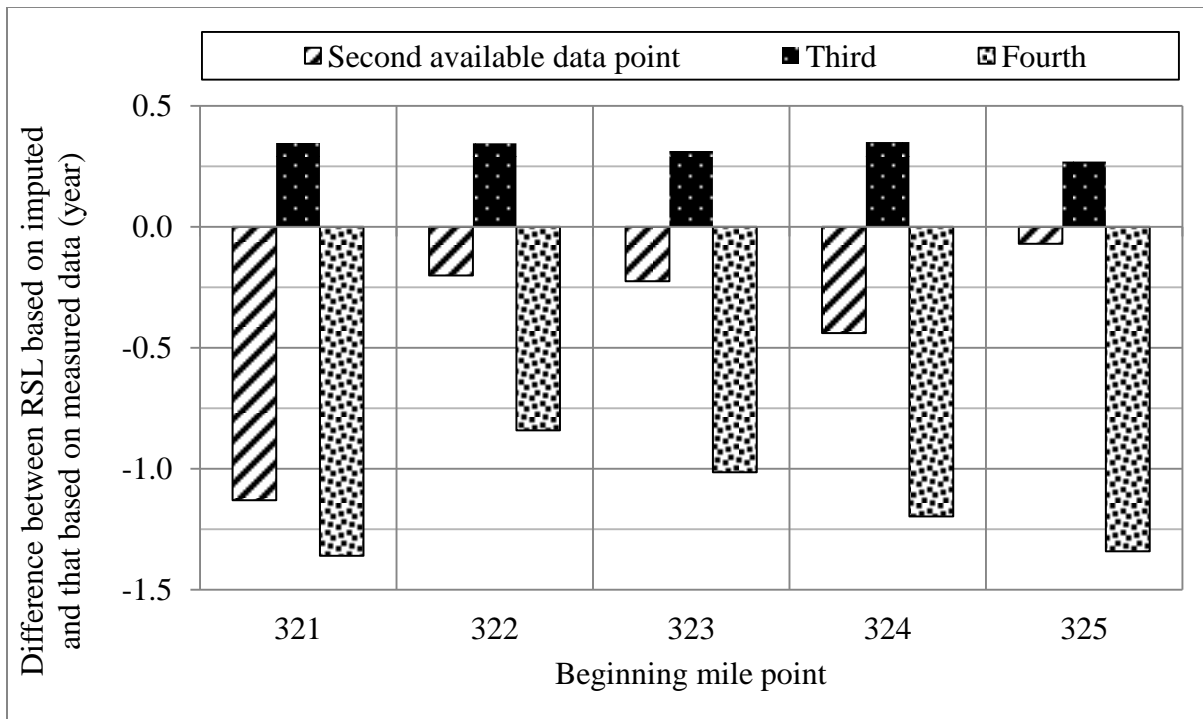


Figure 4.43 The difference between the RSL values based on imputed and measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the moving regression imputation method

Although the above example indicates that the moving regression imputation method produces ± 20 percent difference between the measured and the imputed IRI data, the effect of the method on the RSL, and PMS decision, is negligible. Figures D.13 through D.18 in Appendix D show similar results that were obtained for 15 pavement segments along other roads located in the other states and the MnROAD experiment.

4.7.4 Multiple Regressions Imputation

The multiple regressions imputation method used in this study is another slight modification of the regression method. That is the method is slightly different than the regression and the moving regression methods. It is used to impute missing data using the proper regression function based on the available information. In the multiple regressions method, the first value of the missing data point is calculated using at least three available data points. The second value is

calculated as the average of the first value and the updated value as one additional data point becomes available. The third value is calculated as the average of the first, second and the updated value as another data point becomes available and so forth. Thus, the calculation steps used in the multiple regressions method are the same as those used in the moving regression and the added step is for the calculation of the average of multiple values based on the regression models assumed to represent the missing data. Perhaps this is most easily understood through the example displayed in Table 4.14. In this case, the data at year 7 (fourth data point) is amputated and assumed missing. The fourth data point is first imputed using the regression parameters of the exponential model based on the first, second, and third data points (multiple regression number 1 in Table 4.14). This is the same procedure as for the regression and the moving regression imputation methods. However, when the fifth data point becomes available, the missing fourth data point is calculated using the regression parameters of the exponential model based on the first, second, third, and fifth data points (multiple regression number 2 in Table 4.14). The difference from the moving regression method is that the final imputed value is the average of the two calculated values based on the two regression models. In the example shown in Table 4.14, the multiple regression number 1 produced an IRI value of 174.2 in/mile at year 7. When the fifth data point is included, the model yields an IRI value of 146.7 inch/mile. The average of these two values (160.5 in/mile) is the imputed data at year 7 for multiple regression 2 as shown in the table. When the sixth data point becomes available, the parameter of the exponential function are recalculated using the first, second, third, fifth and sixth data points and the value of the missing data point (the fourth one in this example) is the average of the three imputed values. It should be noted that the results of the first regression step (multiple regression 1, moving regression 1 in Table 4.13, and regression imputation in Table 4.12 are the same.

Table 4.14 Multiple regressions imputation example

	Time series data number	Data collection year	Elapsed time (year)	Available IRI data set (in/mile)	Amputated IRI data set (in/mile)	Imputed IRI data set (in/mile)	Equation
Multiple regression number 1	1	1999	1	74	74	74	
	2	2000	2	125	125	125	
	3	2003	5	131	131	131	
	4	2005	7	122		174.2	$174.2 = 78.782e^{(.1134*7)}$
Multiple regression number 2	1	1999	1	74	74	74	
	2	2000	2	125	125	125	
	3	2003	5	131	131	131	
	4	2005	7	122		160.5	$146.7 = 85.313e^{(.0774*7)}$ 160.5 = Average (174.2, 146.7)
	5	2007	9	163	163	163	

For the three imputed data points, the percent differences between the measured and the multiple regressions imputed IRI data for five 0.1 mile long pavement segments along HWY 24 in Colorado are shown in Figure 4.44. It can be seen in this figure that the percent difference between the imputed and the measured IRI data ranges from less than -35 percent to about +35 percent with the majority of the imputed data lying within approximately ± 20 percent of the measured data.

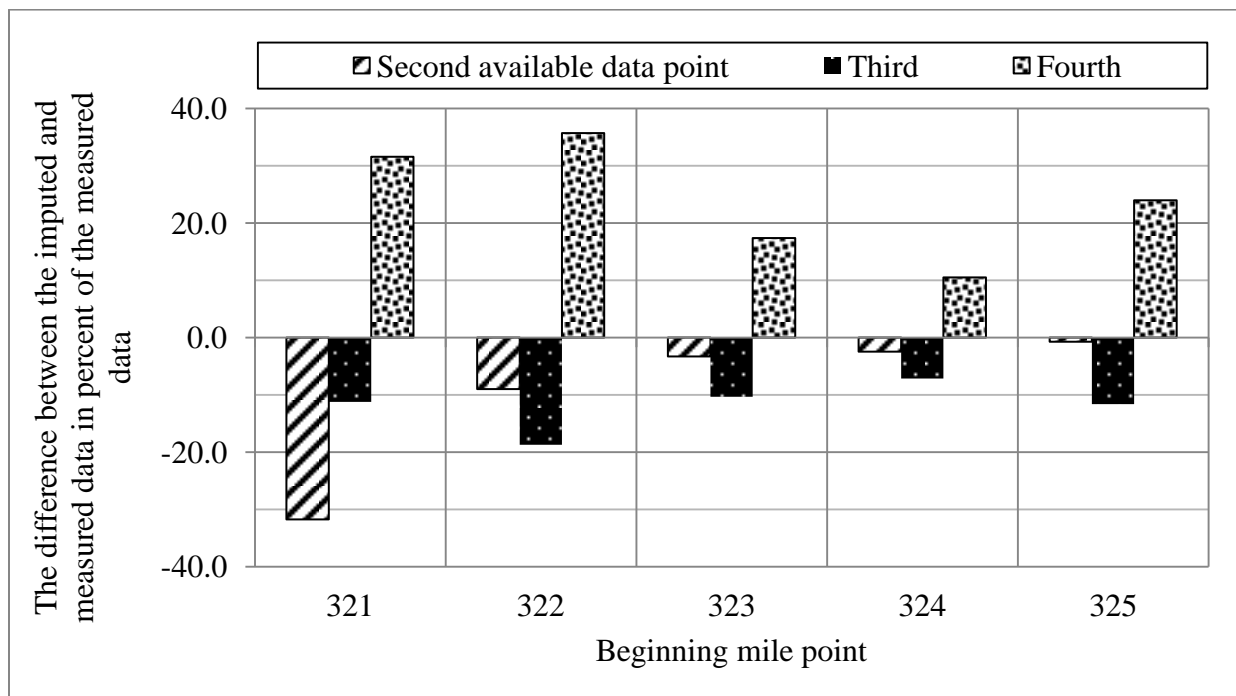


Figure 4.44 The difference between the imputed and measured data in percent of the measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the multiple regressions imputation method

The measured and the imputed and measured data points were used to calculate the RSL of each pavement segment based on an IRI threshold value of 200 inch/mile. The differences in the RSL values based on the measured and the imputed and measured IRI data are shown in Figure 4.45. The data in the figure show that the maximum difference between the RSL values based on the measured and the imputed and measured IRI data is less than 2 years. Most of the

differences in the RSL values are less than 0.5 year. Differences of this magnitude would not cause any significant changes in the condition state (see section 4.3) of the pavement segments and thus, changes in PMS decisions would not be expected.

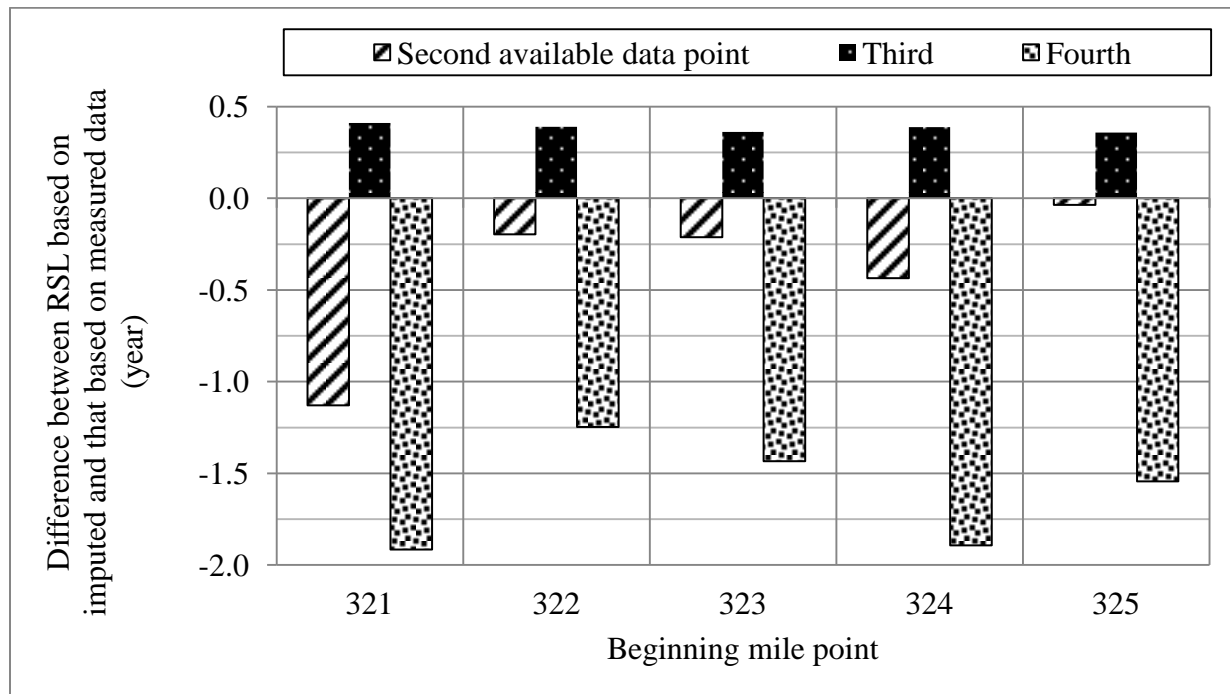


Figure 4.45 The difference between the RSL values based on imputed and measured data versus the BMP along HWY 24 in Colorado, three data points imputed using the multiple regressions imputation method

Similar to the other imputation methods described previously, the multiple regressions imputation may produce ± 20 percent differences between the measured and the imputed IRI data, its effect on the RSL is negligible. Hence, the effect of the multiple imputation method on PMS decisions is also negligible. Figures D.19 through D.24 of Appendix D indicate that similar results were obtained for 15 pavement segments along other roads in the other states and along the MnROAD experiment.

4.7.5 Discussion

In the previous subsections, four data imputation methods were presented along with examples using five 0.1 mile long pavement segments along HWY 24 in Colorado. Additional pavement segments were also analyzed where few data points were amputated and then imputed using each of the four methods. Although the five 0.1 mile pavement segments do not present a statistically significant number of segments to support conclusions, the combined analysis results of the five segments and those reported in Appendix D (15 segments), increase the confidence level in the results and in the observations and discussion presented in this section. As stated previously, after imputing the data, the imputed values were compared to the measured ones using two analysis techniques:

1. The differences in the values of the imputed and the measured data points.
2. The differences between the RSL values estimated using the entire measured data set and the measured and imputed data set.

Results of the two analyses were presented in the previous four sections. Based on thorough examination of the results of the analysis of the four imputation methods, the following observations and their associated discussion were made.

1. The accuracy of the imputed data points is a function of the variability of the measured data. Low variability data causes the accuracy of the imputed data to be very high. To illustrate, consider Figures 4.46 and 4.47. In the first figure, three sets of six theoretical IRI data points are shown having three deterioration rates but no variability over time. While Figure 4.47 depicts three sets of six time series IRI data points having three deterioration rates and variability over time. For each set of the six data points in each figure, the fourth data point was amputated and then imputed using the four methods. The

results are listed in Tables 4.15 and 4.16. It can be seen from Table 4.15 that except for the linear interpolation, the other imputation methods yielded perfect results (no differences between the theoretical and the imputed data points). The linear interpolation imputation method yielded slight differences (up to 12.8 percent) between the theoretical and the imputed data. The results listed in Table 4.16 tell a different story. All imputation methods produced differences between the imputed and the real data.

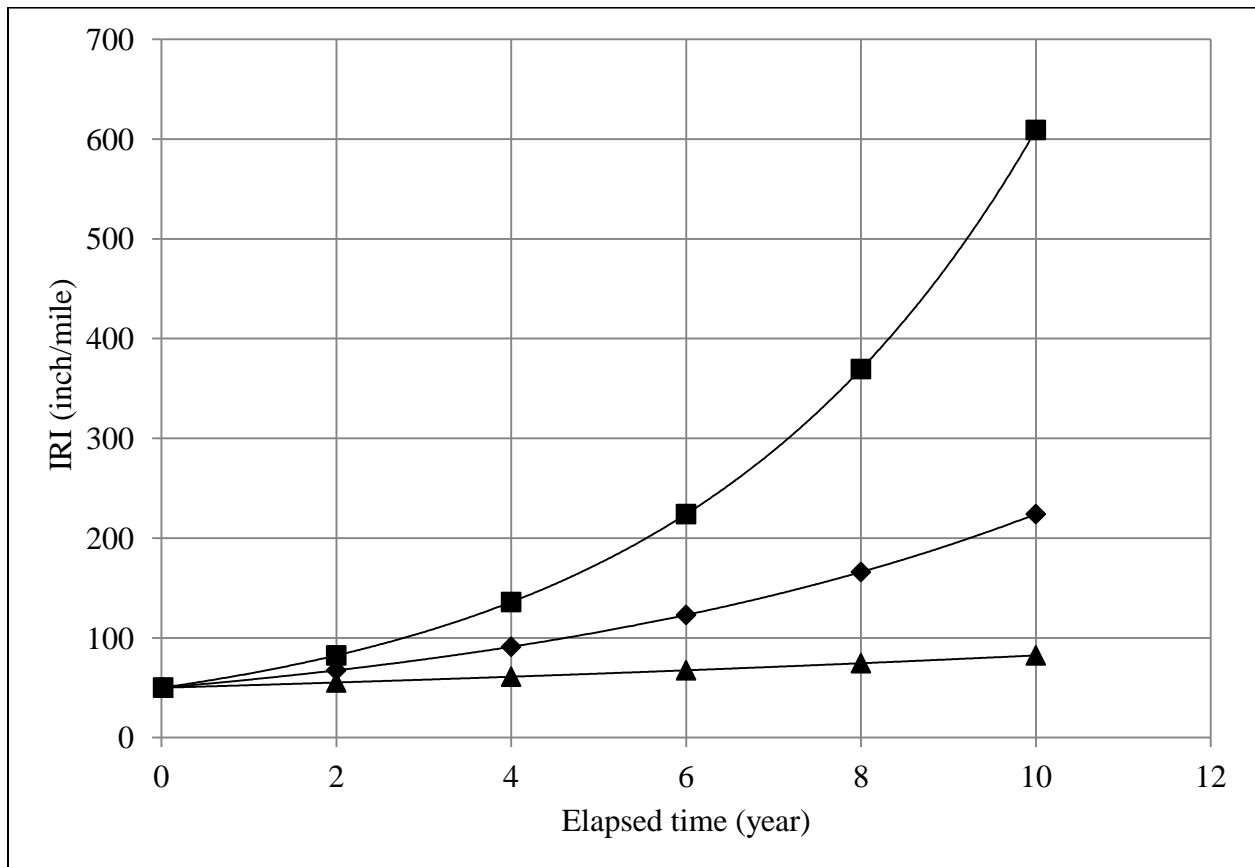


Figure 4.46 IRI data having three deterioration rates and no variability

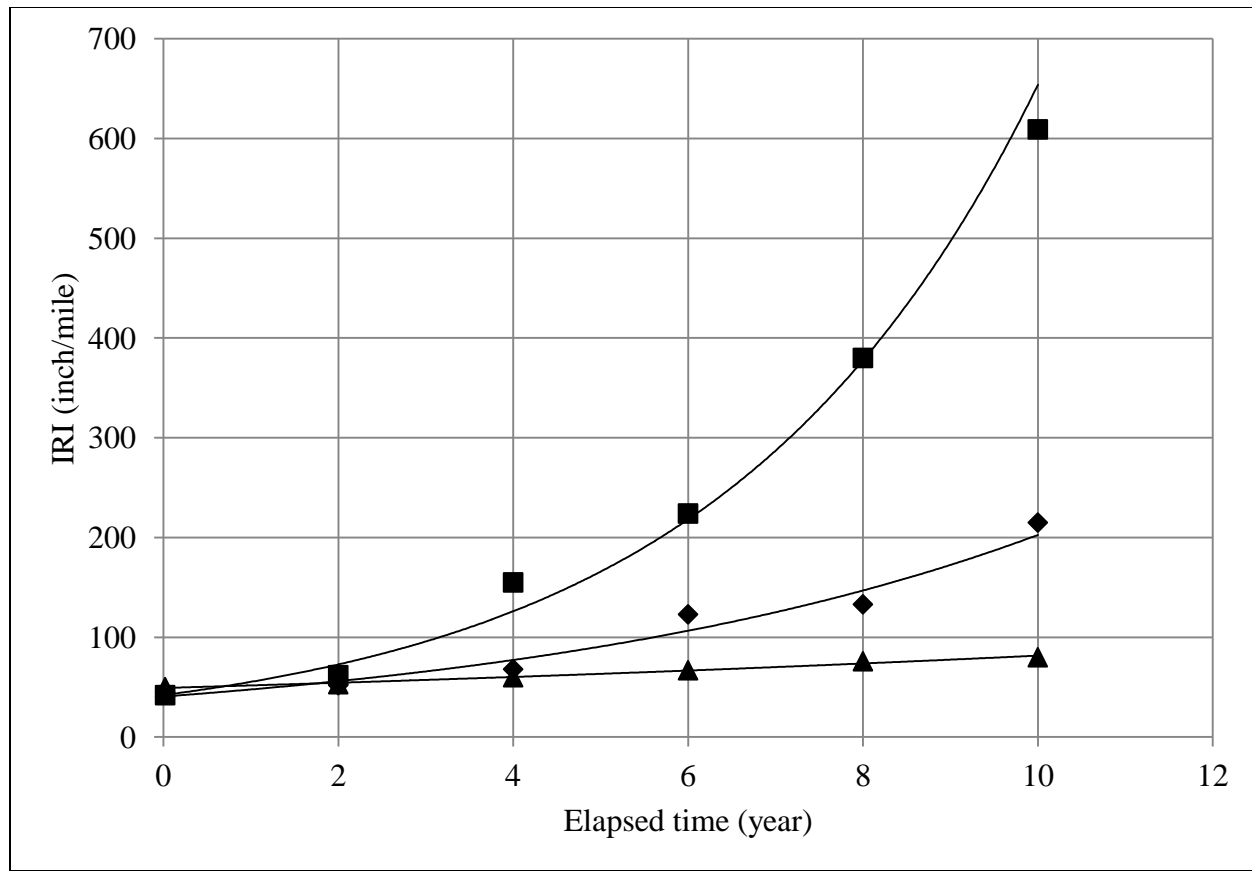


Figure 4.47 Variable IRI data having three deterioration rates

2. The accuracy of the imputed data points is a function of the pavement rate of deterioration. This functionality varies from one imputation method to the next. The results listed in Table 4.16 indicate that:
 - a) For the linear imputation method, the percent difference between the imputed and the actual data increases as the rate of deterioration increases. This was expected because the data follow an exponential function while the imputation method is based on a linear relationship.
 - b) For the other three imputation methods, the percent differences between the imputed and the actual data fluctuate up and down as the rate of deterioration

Table 4.15 Differences and percent differences between the imputed and the actual IRI data for four imputation methods (three rates of deterioration and no variability)

Elapsed time (year)	Deterioration rate			Deterioration rate	Differences and percent differences between the imputed and the actual IRI data for four imputation methods							
	Low	Medium	High		Linear interpolation		Regression		Moving regression		Multiple regressions	
					Difference	(%)	Difference	(%)	Difference	(%)	Difference	(%)
0.02	50	50	50	Low	0.3	0.5	0	0	0	0	0	0
2	55	67	82	Medium	6	4.5	0	0	0	0	0	0
4	61	91	136	High	9	12.8	0	0	0	0	0	0
6	67	123	224									
8	75	166	369									
10	82	224	609									

Table 4.16 Differences and percent differences between the imputed and the actual IRI data for four imputation methods (three rates of deterioration and variable data)

Elapsed time (year)	Deterioration rate			Deterioration rate	Differences and percent differences between the imputed and the actual IRI data for four imputation methods							
	Low	Medium	High		Linear interpolation		Regression		Moving regression		Multiple regressions	
					Difference	(%)	Difference	(%)	Difference	(%)	Difference	(%)
0.02	50	45	42	Low	1	1	-2	-3	-0.4	-1	-1	-1
2	53	52	62	Medium	-22	-18	-41	-33	-20	-16	-29	-24
4	60	68	155	High	43	19	50	22	7	-13	15	7
6	67	123	224									
8	76	133	380									
10	80	215	609									

increases. It appears that the accuracy of the three imputation methods is a function of the interaction between data variability and the rate of deterioration.

- c) The ranges of the differences and the percent differences appear to be the smallest for the moving imputation method amongst the four methods. This is similar to the findings presented in the previous sections based on the five 0.1 mile long pavement segments along HWY 24 in Colorado.
- d) The accuracy of the imputation method is a function of the number of available data points. Higher number of data points produces more accurate imputation. This was expected because higher number of data points defines the trend of the IRI better than fewer data points. The implication of this is that the accuracy of the imputed data is a function of the data collection frequency.

After a thorough examination of the results of the various imputation methods, one question arises; what if no imputation action is taken and the missing data remain missing? This issue is addressed in the next section.

4.7.6 The No-Action (No Imputation) Approach

For various reasons, including time and technological restraints, some SHAs may choose not to impute missing pavement performance data. This is referred to as the no-action approach. That is, the missing data remains missing and the available time series pavement performance data are used for pavement performance modeling. This implies that the number of available time series data points may be limited if, for some reason, the data are not collected along a pavement segment during a data collection cycle. Results of the no-action approach and comparison of the results with the four imputation methods are presented in the next section.

4.7.7 Comparison of the Four Imputation Methods and the No-Action Approach

In the previous sections, the impacts of each of four imputation methods on the differences between the measured and the imputed IRI data and the impacts of each method on the calculated RSL values were presented and discussed for five BMPs along the same road in Colorado. The impacts of data variability, the pavement rate of deterioration and the data collection frequency on the accuracy of the imputed data using the four imputation methods were also addressed. In this section, the four imputation methods and the no-action approach are compared.

For each imputation method, the differences between the measured and the imputed IRI data for the five 0.1 mile pavement segments along HWY 24 in Colorado shown in Figures 4.37, 4.40, 4.42, and 4.44 were combined to determine the maximum, minimum, and the average percent differences between the imputed and the measured data. The results of this combination are shown in Figure 4.48. It can be seen in the figure that the moving regression imputation method produces the smallest percent differences range (slightly less than -30 percent to about +20 percent). The next widest percent differences belongs to the linear interpolation method (-30 to about +30 percent). The range in the differences between the measured and the imputed data for the multiple regressions imputation method varies from about -30 to +35 percent, whereas for the regression imputation method from about -30 to about +50 percent. For all four methods, the average percent difference between the imputed and the measured data is near zero. The misleading average was expected and is mainly due to the positive and negative percent differences between the measured and the imputed data. Keep in mind that the distributions of the percent differences for the five BMPs along HWY 24 in Colorado are shown in Figures 4.37, 4.40, 4.42, and 4.44. It can be seen that most of the percent differences for the five BMPs fall

within a much smaller range than is depicted by the maximum and minimum percent errors shown in Figure 4.48. Note that the no-action approach is not shown in Figure 4.48. The reason is that it is not possible to calculate the percent differences between the imputed and the measured data given that no data are imputed.

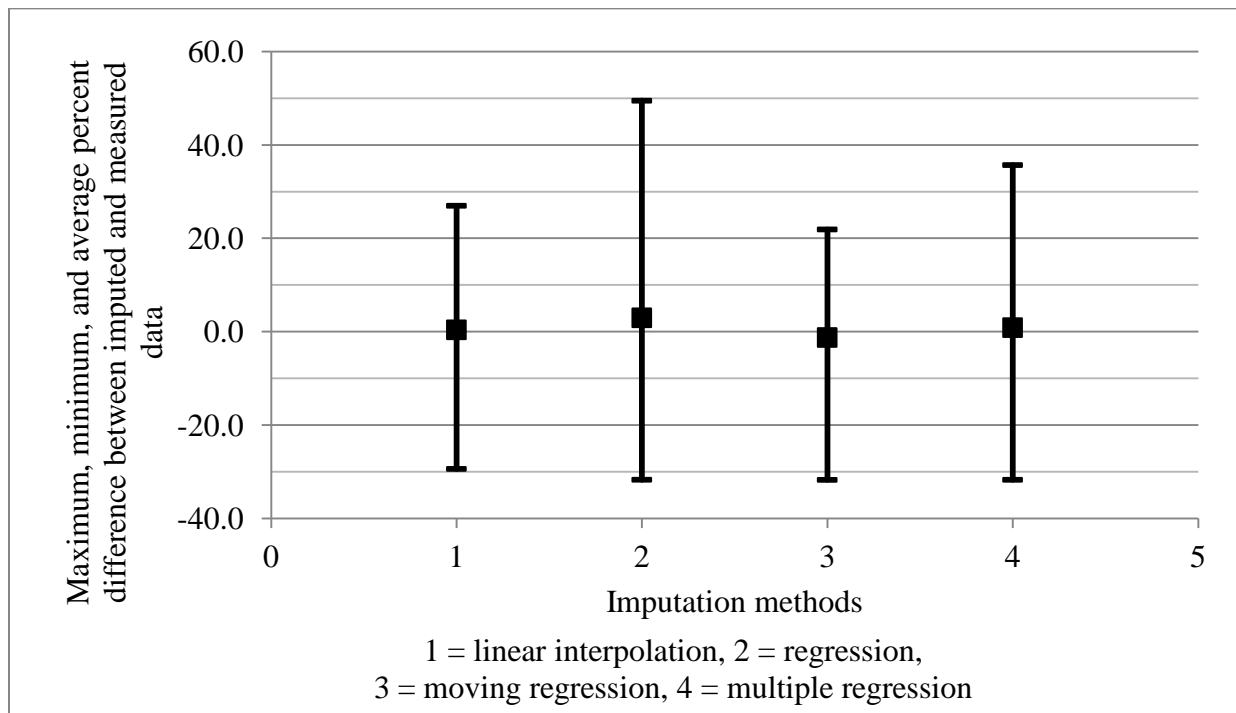


Figure 4.48 Maximum, minimum, and average percent difference between imputed and measured data at BMPs 321, 322, 323, 324, and 325 along HWY 24 in Colorado, using four imputation methods

On the other hand the differences between the RSL values based on the four imputation methods, the no-action approach, and on the complete measured data sets are shown in Figures 4.49 through 4.51. Figures 4.49 through 4.51 display the results from amputating and imputing the second, third, and fourth available data point, respectively. It can be seen that, for each of the imputed data points, the differences in the RSL based on each imputation method and the no-action are not significant. Furthermore, these differences would not have measurable impacts on the PMS decisions.

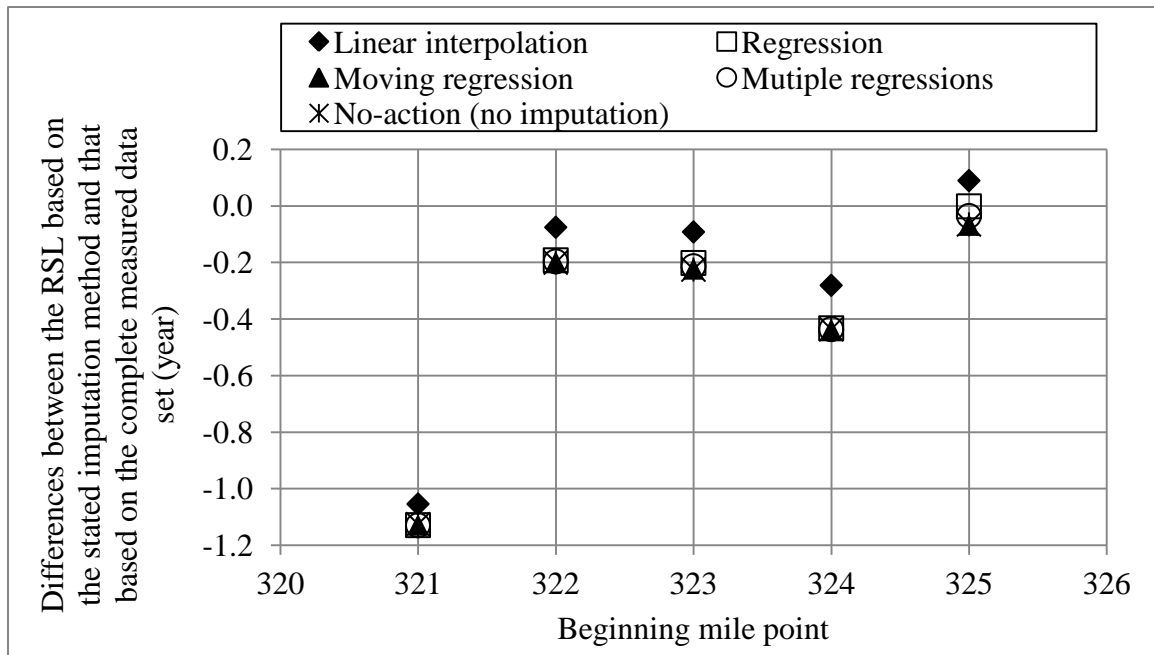


Figure 4.49 Differences between the RSL based on the stated imputation method (second available data point imputed) and that based on the complete measured data at BMPs 321, 322, 323, 324, and 325 along HWY 24 in Colorado

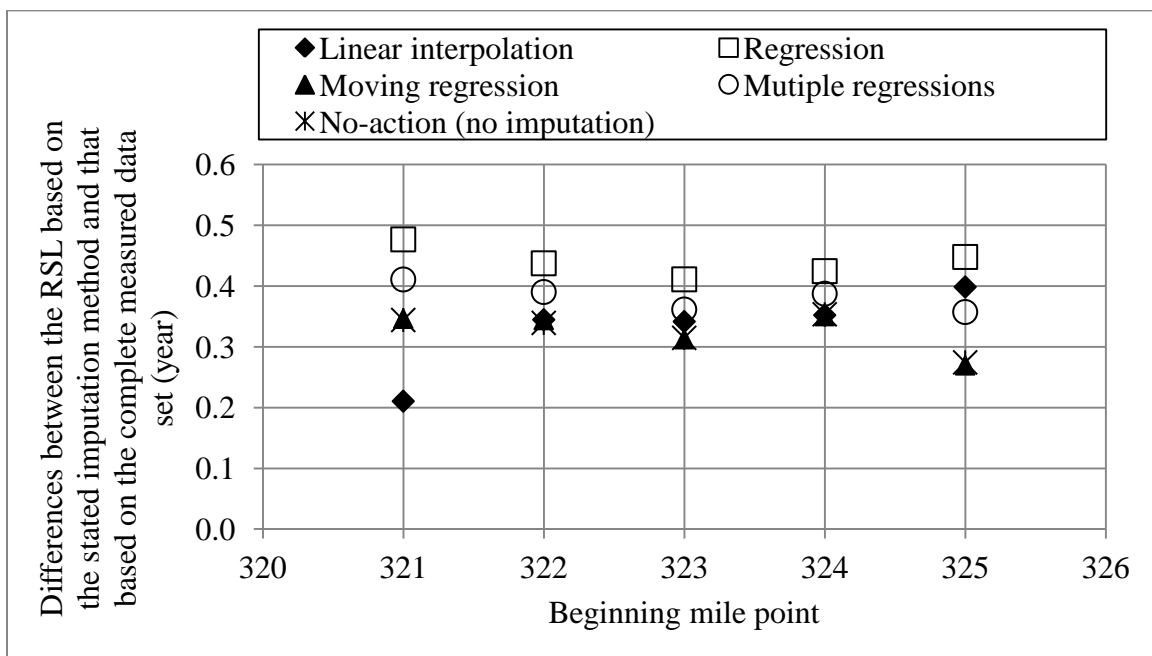


Figure 4.50 Differences between the RSL based on the stated imputation method (third available data point imputed) and that based on the complete measured data at BMPs 321, 322, 323, 324, and 325 along HWY 24 in Colorado

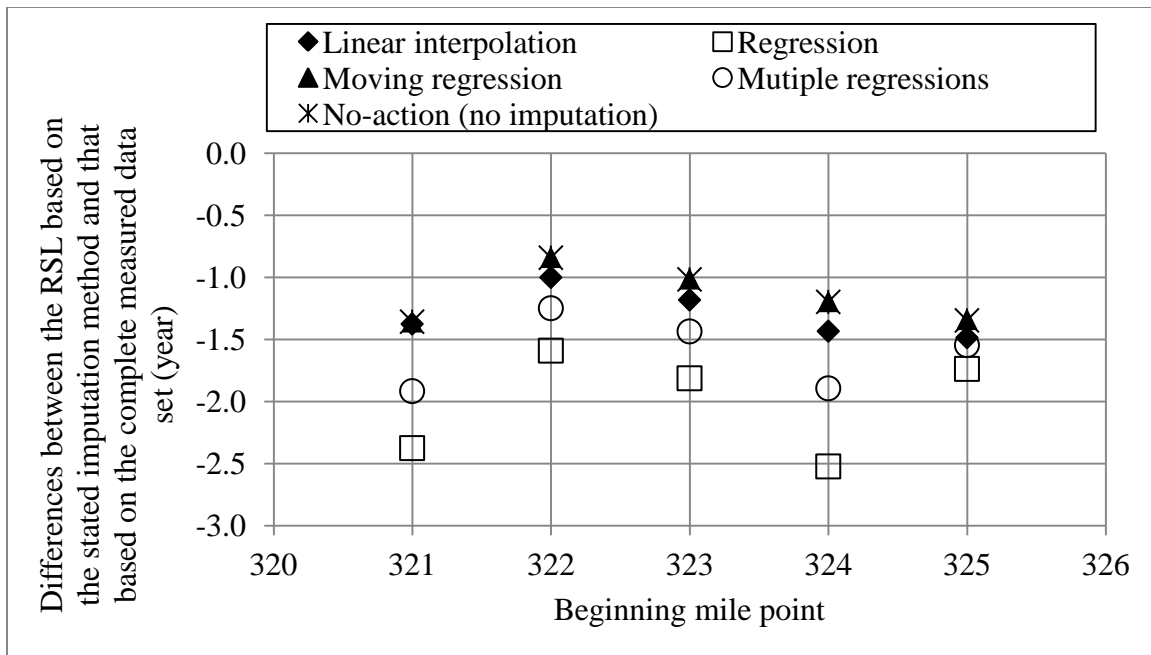


Figure 4.51 Differences between the RSL based on the stated imputation method (fourth available data point imputed) and that based on the complete measured data at BMPs 321, 322, 323, 324, and 325 along HWY 24 in Colorado

The differences between the RSL values based on the measured and those based on the imputed and measured IRI data shown in Figures 4.38, 4.41, 4.43, and 4.45 were combined to determine the maximum, minimum, and average RSL differences shown in Figure 4.52. The range in the differences of the no-action approach relative to the full set of the measured data is also shown in Figure 4.52. It can be seen from the figure that the no-action approach and the moving regression imputation method produces the smallest range of RSL differences (slightly less than 1.5 years) followed by the linear interpolation method (about 1.5 years), the multiple regressions imputation method (slightly less than 2 years), and by the regression imputation method (about 2.5 years). In any case, these differences would not have significant effects on the condition state of the pavement segments (see section 4.3). Similar results were obtained for the pavement segments along other roads in the other states and the MnROAD experiment and are shown in Figures D.25 through D.30 in Appendix D.

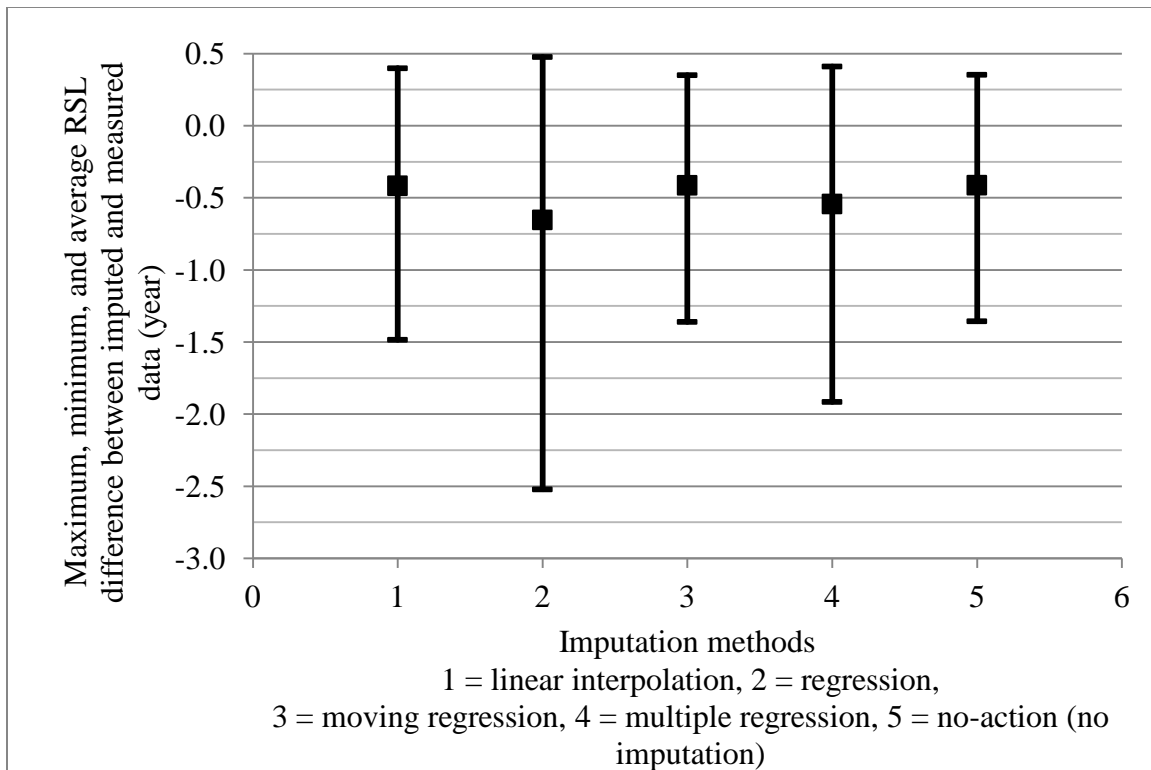


Figure 4.52 Maximum, minimum, and average difference between RSL values based on imputed and measured data at BMPs 321, 322, 323, 324, and 325 along HWY 24 in Colorado, using four imputation methods

It should be noted that, although the no-action approach produces small RSL differences relative to the complete measured data, in some cases, data imputation is desirable. The decision to impute or not to impute, as well as the imputation method to be used, depends on several factors including:

1. The agency's desire to impute missing data.
2. The number of missing and available data points. For example, if the missing data point causes the number of available data points to decrease below three, the only available imputation option is linear interpolation. However, the imputed data point will add no more information and the linear model based on the two available data points will be exactly the same as that based on the two available data points and the imputed data.

If there are three or more data points available in the database, the missing data point or points can be neglected (no-action), or imputed using either the multiple or the moving regression method. These imputation methods generally generate the smallest errors.

One important point that should be noted is that some SHAs employees may lose confidence in the data analyses results if the analyses are based on time series data sets having missing data elements. Imputing the data bridges the gaps reasonably, which may increase the confidence in the analyses results increased. Thus, for various reasons, it may be beneficial to impute the missing data points.

4.7.8 Advantages and Disadvantages of the Four Imputation Methods

Given that the analysis results show that the impacts of each of the four data imputation methods on the calculated RSL values are insignificant, it is fair to say that a SHA could use any of the four methods without expecting any significant changes in the PMS decisions. However, each of the four imputation methods has advantages and disadvantages that should be evaluated before a data imputation method is selected to impute missing pavement condition and distress data points. The advantages and disadvantages of each imputation method include those listed in Table 4.17 below.

The most simple of the four imputation methods that was analyzed is the linear interpolation method. In general, the application of this method is simple, easily understood, and only requires two data points. However, the missing data point that requires imputation must be between two existing data points. Stated differently, linear interpolation imputation cannot be used to impute any data points at the end of a time series data set by extrapolation. Furthermore, because this method assumes a linear rate of pavement deterioration, the results of the imputation could be significantly impacted by localized data variability. If the two data points used in the

Table 4.17 Advantages and disadvantages of four data imputation methods

Imputation method	Advantages	Disadvantages
Linear interpolation	<ul style="list-style-type: none"> • Easy to calculate and understand • Expresses pavement deterioration using simple linear model • Requires only two data points 	<ul style="list-style-type: none"> • Missing data must be between two measured data points and hence it cannot be used to impute any end data (extrapolation) • Assumes linear pavement rate of deterioration • Could be significantly impacted by localized data variability
Regression	<ul style="list-style-type: none"> • Based on the available time-series data (distress versus time) and the non-linear pavement rate of deterioration. 	<ul style="list-style-type: none"> • Requires modeling of the time series data • Cannot be used if less than three data points are available • Could be significantly impacted by localized data variability
Moving regression	<ul style="list-style-type: none"> • Based on the available time-series data (distress versus time) and the non-linear pavement rate of deterioration. • The imputed values are ‘updated’ as more data are collected 	<ul style="list-style-type: none"> • Requires modeling of the time series data • Cannot be used if less than three data points are available
Multiple regressions	<ul style="list-style-type: none"> • Based on the available time-series data (distress versus time) and the non-linear pavement rate of deterioration. • The imputed values are ‘updated’ as more data are collected • The averaging of the multiple imputed values mitigates the effects of significant data variability 	<ul style="list-style-type: none"> • Requires modeling of the time series data • Cannot be used if less than three data points are available

imputation method have significant localized variability, the imputed data between these two points will also be affected by the localized variability. This can be seen clearly by comparing the data shown in Tables 4.15 and 4.16. The advantages of simplicity and understandability are countered by the impacts of localized data variability and the ability to only impute data between two existing data points. Extrapolation may produce significant error in the imputed data.

In comparison, the mathematical modeling requirements of the regression, moving regression, and multiple regressions methods cause these methods to be more time-consuming and laborious than the linear interpolation method. Without modeling software, the application of these methods to the pavement condition and distress data becomes problematic. This, however, could be easily solved by using the proper available software, which can be automated to analyze each pavement project along the entire database. In addition, similar to linear interpolation, these methods are impacted by localized data variability. The data variability decreases the accuracy of the non-linear regression models and impacts the differences between the imputed and the measured data. However, the moving regression and multiple regressions methods mitigate the effects of localized data variability because the data are re-imputed as more data become available. Examples of such mitigation are shown in Table 4.16. For the same deterioration rate, the moving and multiple regressions imputation methods produce lower percent differences than the regression imputation method. Hence, the imputed data are not as susceptible to the effect of localized data variability from one year to the next.

Regardless, it is the opinion of the author that, since the impacts of the linear interpolation and regression methods (regression, moving, and multiple regressions imputation) on the RSL (and hence, PMS decisions) are similar, the linear interpolation imputation method could prove more useful for implementation in SHAs. For extremely large pavement networks, the time and modeling requirements for the regression imputation methods could become overwhelming. The linear interpolation imputation method could be implemented in a much more timely manner while producing results that are reasonable and do not change the condition state (or PMS decision) for the pavement segments. However, if the SHA wishes to computerize the data imputation process, the moving regression or multiple regressions imputation methods could be used to better account for the pavement deterioration rates and mitigate the effects of

localized data variability from one year to the next. This would also allow for imputation of missing data that are not between two measured data points. Therefore, depending on the available technological resources, time constraints, and imputation preferences of different SHAs, the moving and multiple regressions imputation methods could also be selected to impute missing data points that are reasonable and do not affect the PMS decisions.

CHAPTER 5

SUMMARY, CONCLUSIONS, & RECOMMENDATIONS

5.1 Summary

Some SHAs collect pavement performance data along the entire road network whereas others use sampling techniques. The long-term economic climate has caused some highway administrators to consider pavement performance data sampling as an alternative to reduce data collection costs. The problem is that the effects of data sampling on the accuracy of the PMS decisions are not well documented. It is not clear that data sampling ultimately reduces the overall PMS costs. The problem is compounded in that data sampling affects the pavement length for which the performance data are stored and analyzed. Long analysis lengths may not improve the data analysis results and may artificially decrease the data variability. In addition, some pavement performance data from one or more data collection cycles along certain pavement segments may not be documented or are missing, which makes it difficult or impossible to model and analyze the data.

To address these issues, pavement management system (PMS) databases were obtained from four State Highway Agencies; the Colorado Department of Transportation (CDOT), the Louisiana Department of Transportation and Development (LADOTD), the Michigan Department of Transportation (MDOT), and the Washington State Department of Transportation (WSDOT). In addition, the pavement performance data from the Minnesota Road Research Project (MnROAD) were also obtained. The four SHAs collect and store the pavement condition and distress data for each 0.1 mile long pavement segment along their entire road network (no sampling) on an annual or bi-annual basis. The pavement condition and distress data

from the four databases were used to simulate various sampling techniques and to study the impacts of sampling on:

1. The differences between the continuously measured and the sampled pavement condition and distress data.
2. The variability of the pavement condition and distress data along various pavement projects.
3. The remaining service life (RSL) and the condition state of each 0.1 mile long pavement segment along various pavement projects.
4. The accuracy of the pavement management decisions regarding the selection of project boundaries and time of treatment.

The results from the various sampling analyses were combined to quantify the potential savings and the hidden cost of sampling pavement performance data and its potential impacts on the SHAs.

Additionally, in the current state-of-the-practice, the length of pavement sections subjected to performance analysis is constrained by the data collection procedures, such as the sampling technique, and by the analysis details. To address the issue of pavement analysis length, pavement condition and distress data from each of the four SHAs were analyzed (the IRI data were averaged and the cracking data summed) over 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 mile pavement lengths. The data for each analysis length were scrutinized to determine whether or not the data could be modeled using the proper non-linear mathematical function. Based on the results, the impacts of the pavement analysis length on the PMS decisions were estimated and discussed in detail.

Finally, for various reasons, the pavement performance data from some data collection cycles along certain pavement segments or sections may not be documented in the database. Missing data may pose serious issues regarding the modeling and analysis of the time series pavement performance data especially when the missing data decrease the number of available data points to less than three. To help solve these issues, some of the measured pavement performance data from the four SHAs and the MnROAD databases were amputated to simulate missing data points. These data were consequently imputed using four data imputation techniques. The differences between the imputed and the measured data were then analyzed and their impacts on the pavement management decisions were documented and discussed.

5.2 Conclusions

Based on the results of the analyses conducted in this study, conclusions regarding general issues, data sampling, pavement analysis length, and data imputation were drawn and are presented below.

General Issues

- Certain data elements such as pavement treatment costs, location, and pavement surface age are currently not included in the PMS database.
- The use of linear location referencing systems, such as beginning and ending mile points, cause significant errors in the location of a pavement segment from one data collection cycle to the next.
- Robust quality assurance and quality control procedures for the pavement performance data should be developed and implemented to increase the overall reliability of the data. This is especially important for the cracking data.

Data Sampling

- For uniform pavement condition and distress along given pavement projects, such as in newly constructed or rehabilitated pavement sections, the sampled pavement performance data accurately represent the continuous data.
- As the pavement ages and/or the pavement conditions and distress become more variable, sampled data do not accurately represent the continuous data. Hence, the errors caused by data sampling increase as the pavement ages and/or the pavement conditions and distresses become more variable.
- Pavement performance data sampling causes various degrees of errors in one or all of the following decisions:
 - The assessment of the true variability of the pavement performance data along a project and along the road network.
 - The selection of optimum time and space (project boundaries).
 - The selection of pavement treatment type.
 - The selection of pavement treatment strategy.
 - The allocation of pavement treatment funds.

Pavement Analysis Length

- Longer pavement analysis lengths artificially decrease the variability of the pavement performance data along given projects and along the network.
- In general, increasing the length over which the pavement performance data are collected, stored, and analyzed does not necessarily improve or increase the percent of pavement segments that can be modeled and analyzed. Specifically, the pavement analysis length

has different effects on the pavement performance data depending on the data collection procedures as detailed below.

- For sensor collected data, the percent of pavement segments that can be modeled and analyzed is almost constant and independent of the pavement analysis length.
- For image based data, the percent of pavement segments that can be modeled and analyzed is variable and not related to the pavement analysis length.
- The pavement analysis length has significant adverse impacts on various PMS decisions including:
 - The selection of project boundaries.
 - The identification of hot spots or areas of localized poor condition or increased distress.
 - The statistical significance of the number of pavement segments or sections that can be subjected to analysis. That is, longer pavement analysis lengths decrease the sample size.

Data Imputation

- All available imputation techniques produce differences between the imputed and the measured data. These differences are a function of several factors including:
 - The variability of the measured data. Low data variability causes the accuracy of the imputed data to be very high.
 - The pavement rate of deterioration. Higher pavement deterioration rates decrease the accuracy of the imputed data.
- The impacts of each of the four data imputation methods analyzed in this study (linear interpolation, regression, moving regression, and multiple regressions) on the calculated

RSL values, and PMS decisions, are insignificant and similar to that if not action is taken. Differences between the imputed and the measured data do not necessarily produce differences in PMS decisions.

5.3 Recommendations

Based on the results of the analyses and conclusions, it is recommended that State Highway Agencies:

- Expand the PMS databases to an integrated database that includes detailed data regarding pavement treatment, maintenance, preservation, rehabilitation and reconstruction types and costs, location, pavement surface age, material properties, traffic, incentive and disincentive payments, and all other pavement related data.
- Use a global positioning system (GPS) as a location reference system in order to provide more consistent location of pavement segments and treatment locations from one data collection cycle to the next.
- Collect pavement condition and distress data on continuous basis along the entire road network.
- Store the pavement performance data in the database for each 0.1 mile long pavement segment.
- Analyze the pavement condition and distress data for each 0.1 mile long pavement segment along projects or along the road network. The 0.1 mile analysis length provides more flexibility and accuracy in the selection of pavement project boundaries, the identification of hot spots or areas of localized poor condition or increased distress, and improves the statistical significance of decisions based on the data.

- Select, if needed, the data imputation methods (i.e. linear interpolation, regression, moving regression, or multiple regressions) based on the technological resources, time constraints, and imputation preferences of the SHA. Each method has advantages and disadvantages depending on the situation, although all methods cause insignificant differences in the PMS decisions.

APPENDICES

APPENDIX A

PMS DATA

The PMS databases collected for this study are extremely large and would take many thousands of pages to include on paper. Hence, the databases are available, by request, from the Department of Civil & Environmental Engineering at Michigan State University. To request the data, please contact:

Department of Civil & Environmental Engineering

C/O Dr. Gilbert Baladi

Michigan State University

Engineering Building

428 S. Shaw Lane, Room 3546

East Lansing, MI 48824

Tel: (517) 355-5107 Fax: (517) 432-1827 E-mail: cee@egr.msu.edu

For convenience, examples of the pavement condition and distress data are included below. Tables A.1 and A.2 contain the measured and formatted condition and distress data for a section of I-70 from the Colorado Department of Transportation (CDOT). Similar examples of the formatted condition and distress data are listed in Tables A.3 through A.6 for the Louisiana Department of Transportation and Development (LADOTD), the Michigan Department of Transportation (MDOT), the Washington State Department of Transportation (WSDOT), and the Minnesota Road Research Project (MnROAD).

Table A.1 Measured pavement condition data along I-70 in Colorado

Route	Control section	Mile points		Direction to survey	Year	IRI (in/mi)	Rut depth (in)	Alligator cracking (ft ²)	Longitudinal cracking (ft)	Transverse cracking (count)
		Beginning	Ending							
I-70	070A	4.8	4.9	1	1998	98	0.4	0	3	1
I-70	070A	4.9	5	1	1998	114	0.4	0	1	1
I-70	070A	5	5.1	1	1998	94	0.4	0	8	2
I-70	070A	5.1	5.2	1	1998	99	0.4	0	1	0
I-70	070A	5.2	5.3	1	1998	98	0.3	0	0	0
I-70	070A	5.3	5.4	1	1998	78	0.3	0	7	0
I-70	070A	5.4	5.5	1	1998	80	0.3	0	0	1
I-70	070A	5.5	5.6	1	1998	89	0.3	0	0	0
I-70	070A	5.6	5.7	1	1998	100	0.3	0	0	0
I-70	070A	5.7	5.8	1	1998	104	0.3	0	0	0
I-70	070A	5.8	5.9	1	1998	95	0.3	31	0	0
I-70	070A	5.9	6	1	1998	86	0.3	0	0	0
I-70	070A	6	6.1	1	1998	112	0.3	0	0	0
I-70	070A	6.1	6.2	1	1998	81	0.3	0	3	0
I-70	070A	6.2	6.3	1	1998	85	0.3	0	0	0
I-70	070A	6.3	6.4	1	1998	85	0.3	0	0	1
I-70	070A	6.4	6.5	1	1998	103	0.3	0	2	0
I-70	070A	6.5	6.6	1	1998	78	0.3	0	2	0
I-70	070A	6.6	6.7	1	1998	93	0.4	100	14	0
I-70	070A	6.7	6.8	1	1998	86	0.4	0	0	1
I-70	070A	6.8	6.9	1	1998	71	0.4	0	0	0
I-70	070A	6.9	7	1	1998	92	0.4	0	0	1
I-70	070A	7	7.1	1	1998	95	0.4	2	0	0
I-70	070A	7.1	7.2	1	1998	90	0.4	0	0	0
I-70	070A	7.2	7.3	1	1998	76	0.4	0	0	0
I-70	070A	7.3	7.4	1	1998	93	0.3	0	0	0

Table A.2 Formatted pavement condition data along I-70 in Colorado

Route	Control section	Mile points		Direction to survey	Year	IRI (in/mi)	Rut depth (in)	Alligator cracking (ft)	Longitudinal cracking (ft)	Transverse cracking (ft)
		Beginning	Ending							
I-70	070A	4.8	4.9	1	1998	98	0.4	0	3	12
I-70	070A	4.9	5	1	1998	114	0.4	0	1	12
I-70	070A	5	5.1	1	1998	94	0.4	0	8	24
I-70	070A	5.1	5.2	1	1998	99	0.4	0	1	0
I-70	070A	5.2	5.3	1	1998	98	0.3	0	0	0
I-70	070A	5.3	5.4	1	1998	78	0.3	0	7	0
I-70	070A	5.4	5.5	1	1998	80	0.3	0	0	12
I-70	070A	5.5	5.6	1	1998	89	0.3	0	0	0
I-70	070A	5.6	5.7	1	1998	100	0.3	0	0	0
I-70	070A	5.7	5.8	1	1998	104	0.3	0	0	0
I-70	070A	5.8	5.9	1	1998	95	0.3	2.6	0	0
I-70	070A	5.9	6	1	1998	86	0.3	0	0	0
I-70	070A	6	6.1	1	1998	112	0.3	0	0	0
I-70	070A	6.1	6.2	1	1998	81	0.3	0	3	0
I-70	070A	6.2	6.3	1	1998	85	0.3	0	0	0
I-70	070A	6.3	6.4	1	1998	85	0.3	0	0	12
I-70	070A	6.4	6.5	1	1998	103	0.3	0	2	0
I-70	070A	6.5	6.6	1	1998	78	0.3	0	2	0
I-70	070A	6.6	6.7	1	1998	93	0.4	8.3	14	0
I-70	070A	6.7	6.8	1	1998	86	0.4	0	0	12
I-70	070A	6.8	6.9	1	1998	71	0.4	0	0	0
I-70	070A	6.9	7	1	1998	92	0.4	0	0	12
I-70	070A	7	7.1	1	1998	95	0.4	0.3	0	0
I-70	070A	7.1	7.2	1	1998	90	0.4	0	0	0
I-70	070A	7.2	7.3	1	1998	76	0.4	0	0	0
I-70	070A	7.3	7.4	1	1998	93	0.3	0	0	0

Table A.3 Formatted pavement condition data along US-80 in Louisiana

Route	Control section	Mile points		Direction to survey	Year	IRI (in/mi)	Rut depth (in)	Alligator cracking (ft)	Longitudinal cracking (ft)	Transverse cracking (ft)
		Beginning	Ending							
US-80	001-05	0	0.1	1	2005	172	0.11	69	100	404
US-80	001-05	0.1	0.2	1	2005	171	0.1	120	140	470
US-80	001-05	0.2	0.3	1	2005	138	0.08	110	87	345
US-80	001-05	0.3	0.4	1	2005	136	0.06	51	148	278
US-80	001-05	0.4	0.	1	2005	241	0.15	43	194	400
US-80	001-05	0.5	0.6	1	2005	105	0.07	349	56	304
US-80	001-05	0.6	0.7	1	2005	127	0.11	381	0	330
US-80	001-05	0.7	0.8	1	2005	135	0.21	35	71	120
US-80	001-05	0.8	0.9	1	2005	136	0.14	86	38	56
US-80	001-05	0.9	1	1	2005	74	0.21	7	9	59
US-80	001-05	1	1.1	1	2005	105	0.25	55	0	31
US-80	001-05	1.1	1.2	1	2005	109	0.14	29	0	33
US-80	001-05	1.2	1.3	1	2005	119	0.15	54	38	81
US-80	001-05	1.3	1.4	1	2005	120	0.15	4	13	77
US-80	001-05	1.4	1.5	1	2005	107	0.12	0	16	79
US-80	001-05	1.5	1.6	1	2005	102	0.08	13	0	28
US-80	001-05	1.6	1.7	1	2005	118	0.12	134	35	93
US-80	001-05	1.7	1.8	1	2005	114	0.22	102	161	204
US-80	001-05	1.8	1.9	1	2005	-	0.32	12	86	64
US-80	001-05	1.9	2	1	2005	140	0.23	216	125	176
US-80	001-05	2	2.1	1	2005	192	0.21	49	0	24
US-80	001-05	2.1	2.2	1	2005	163	0.3	62	0	0
US-80	001-05	2.2	2.3	1	2005	217	0.41	48	0	8
US-80	001-05	2.3	2.4	1	2005	194	0.35	72	0	33
US-80	001-05	2.4	2.5	1	2005	191	0.31	0	0	0
US-80	001-05	2.5	2.6	1	2005	244	0.24	0	0	0

Table A.4 Formatted pavement condition data along I-69BL in Michigan

Route	Control section	Mile points		Direction to survey	Year	IRI (in/mi)	Rut depth (in)	Alligator cracking (ft)	Longitudinal cracking (ft)	Transverse cracking (ft)
		Beginning	Ending							
I69BL	33043	0.000	0.100	1	2000	202	0.94	-	25	22
I69BL	33043	0.100	0.200	1	2000	143	0.59	-	41	18
I69BL	33043	0.200	0.300	1	2000	175	0.59	-	68	37
I69BL	33043	0.300	0.400	1	2000	258	0.58	-	80	36
I69BL	33043	0.400	0.500	1	2000	182	0.67	-	76	29
I69BL	33043	0.500	0.600	1	2000	188	0.65	-	81	33
I69BL	33043	0.600	0.700	1	2000	170	0.69	-	74	47
I69BL	33043	0.700	0.800	1	2000	112	0.36	-	76	38
I69BL	33043	0.800	0.900	1	2000	144	0.42	-	103	30
I69BL	33043	0.900	1.000	1	2000	189	0.61	-	106	56
I69BL	33043	1.000	1.100	1	2000	207	0.86	-	106	35
I69BL	33043	1.100	1.200	1	2000	166	0.67	-	34	35
I69BL	33043	1.200	1.300	1	2000	232	0.30	-	99	45
I69BL	33043	1.300	1.400	1	2000	173	0.52	-	59	9
I69BL	33043	1.400	1.500	1	2000	174	0.61	-	64	1
I69BL	33043	1.500	1.600	1	2000	89	0.33	-	57	-
I69BL	33043	1.600	1.700	1	2000	307	1.30	-	74	1
I69BL	33043	1.700	1.800	1	2000	243	1.23	-	73	34
I69BL	33043	1.800	1.900	1	2000	200	0.76	-	54	27
I69BL	33043	1.900	2.000	1	2000	100	0.18	-	35	4
I69BL	33043	2.000	2.100	1	2000	74	0.15	-	32	16
I69BL	33043	2.100	2.200	1	2000	116	0.16	-	61	-
I69BL	33043	2.200	2.300	1	2000	67	0.20	-	54	3
I69BL	33043	2.300	2.458	1	2000	61	0.43	-	107	14
I69BL	33043	2.458	2.558	1	2000	202	0.94	-	2	22
I69BL	33043	2.558	2.658	1	2000	143	0.59	-	25	18

Table A.5 Formatted pavement condition data along I-5 in Washington

Route	Control section	Mile points		Direction to survey	Year	IRI (in/mi)	Rut depth (in)	Alligator cracking (ft)	Longitudinal cracking (ft)	Transverse cracking (ft)
		Beginning	Ending							
I-5	-	5.3	5.4	1	2003	94	0.04	0	0	0
I-5	-	5.4	5.5	1	2003	123	0.04	0	132	0
I-5	-	5.5	5.6	1	2003	67	0.16	0	359.04	0
I-5	-	5.6	5.7	1	2003	68	0.31	0	73.92	0
I-5	-	5.7	5.8	1	2003	95	0.35	0	0	0
I-5	-	5.8	5.9	1	2003	77	0.31	0	0	0
I-5	-	5.9	6	1	2003	75	0.20	0	21.12	0
I-5	-	6	6.1	1	2003	81	0.31	137.28	15.84	0
I-5	-	6.1	6.2	1	2003	85	0.31	42.24	0	0
I-5	-	6.2	6.3	1	2003	91	0.46	0	42.24	0
I-5	-	6.3	6.4	1	2003	131	0.41	0	21.12	48
I-5	-	6.4	6.5	1	2003	89	0.24	0	163.68	12
I-5	-	6.5	6.6	1	2003	68	0.28	0	68.64	0
I-5	-	6.6	6.7	1	2003	67	0.24	0	63.36	0
I-5	-	6.7	6.8	1	2003	78	0.35	0	52.8	0
I-5	-	6.8	6.9	1	2003	73	0.28	0	0	0
I-5	-	6.9	7	1	2003	74	0.20	0	52.8	0
I-5	-	7	7.1	1	2003	95	0.12	0	132	0
I-5	-	7.1	7.2	1	2003	67	0.12	0	0	12
I-5	-	7.2	7.3	1	2003	85	0.12	0	0	0
I-5	-	7.3	7.4	1	2003	81	0.20	0	0	0
I-5	-	7.4	7.5	1	2003	106	0.31	0	0	12
I-5	-	7.5	7.6	1	2003	84	0.35	0	0	24
I-5	-	7.6	7.7	1	2003	85	0.24	10.56	42.24	48
I-5	-	7.7	7.8	1	2003	80	0.16	0	5.28	12
I-5	-	7.8	7.9	1	2003	68	0.04	0	0	0

Table A.6 Formatted pavement condition data along cell 1 at MnROAD

Cell	Date	IRI (in/mi)	Rut depth (in)	Alligator cracking (ft)	Longitudinal cracking (ft)	Transverse cracking (ft)
1	11-FEB-94	94	0.04	0	0	0
1	21-OCT-94	123	0.04	0	0	0
1	20-APR-95	67	0.16	0	0	0
1	04-NOV-95	68	0.31	0	0	0
1	13-FEB-96	95	0.35	0	0	170
1	01-MAR-96	77	0.31	0	0	184
1	13-MAR-96	75	0.20	0	0	184
1	18-APR-96	81	0.31	0	0	190
1	13-NOV-96	85	0.31	0	0	190
1	30-APR-97	91	0.46	0	0	190
1	15-NOV-97	131	0.41	0	0	190
1	29-APR-98	89	0.24	0	0	191
1	02-OCT-98	68	0.28	0	0	2
1	20-APR-99	67	0.24	0	0	2
1	22-OCT-99	78	0.35	0	0	2
1	16-MAY-00	73	0.28	0	0	2
1	04-JAN-01	74	0.20	0	0	2
1	30-APR-01	95	0.12	0	0	2
1	23-OCT-01	67	0.12	0	0	2
1	20-APR-02	85	0.12	0	0	14
1	08-OCT-02	81	0.20	0	0	14
1	18-APR-03	106	0.31	0	40	38
1	15-OCT-03	84	0.35	0	20	38
1	21-APR-04	85	0.24	0	20	160
1	14-APR-05	80	0.16	0	5	204

APPENDIX B

PAVEMENT PERFORMANCE DATA SAMPLING FIGURES

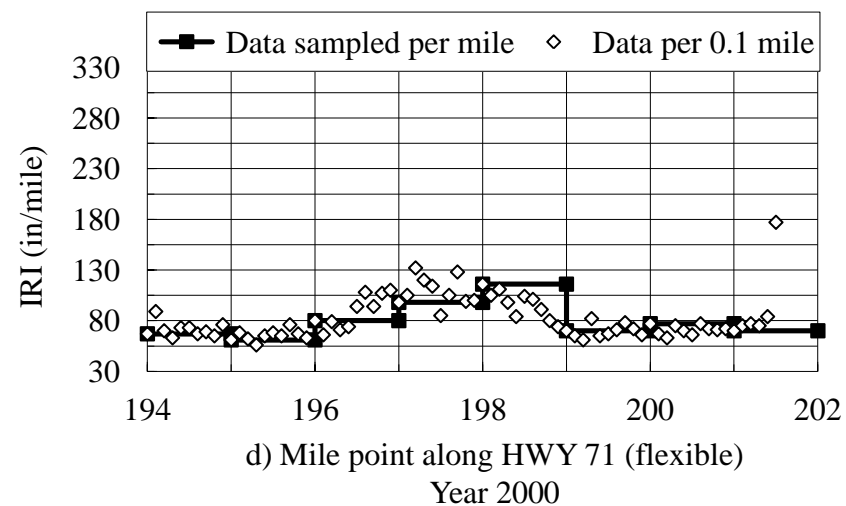
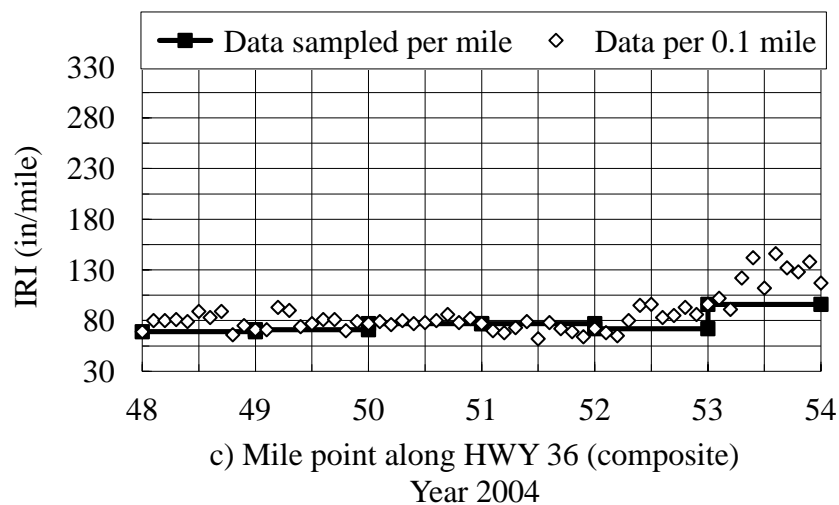
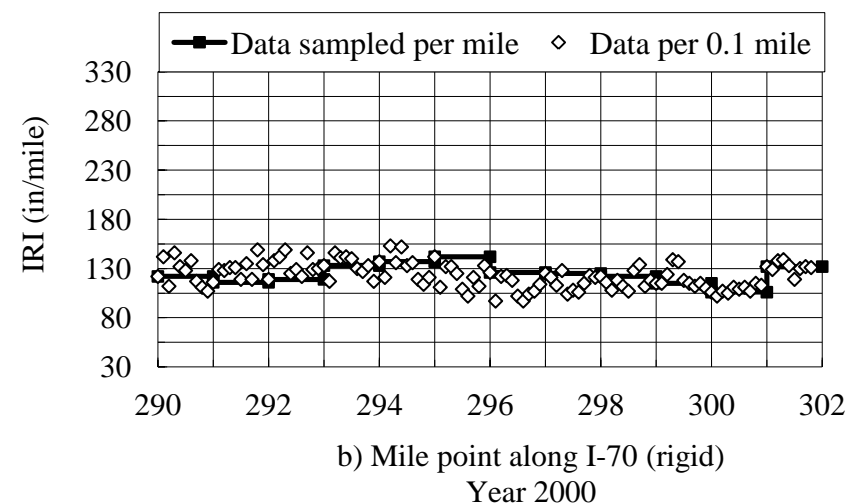
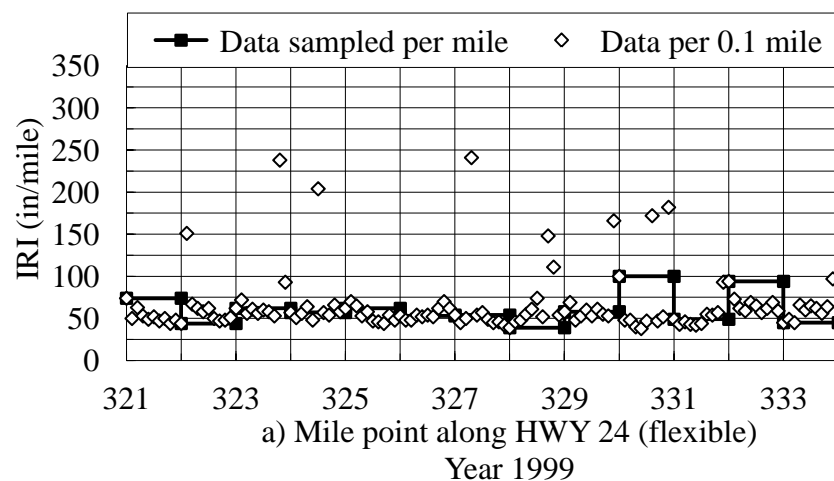
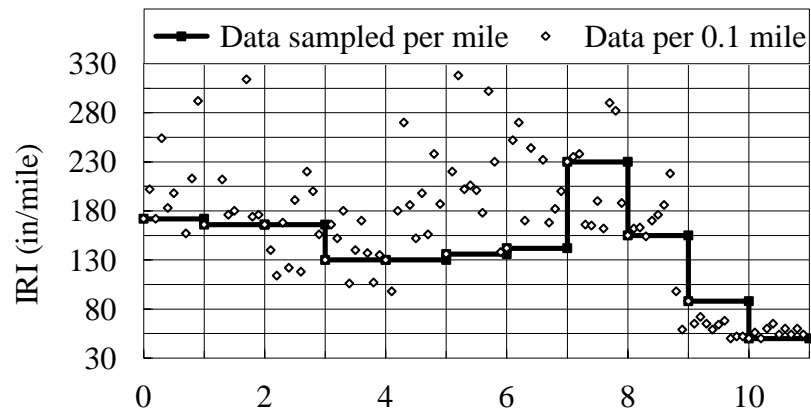
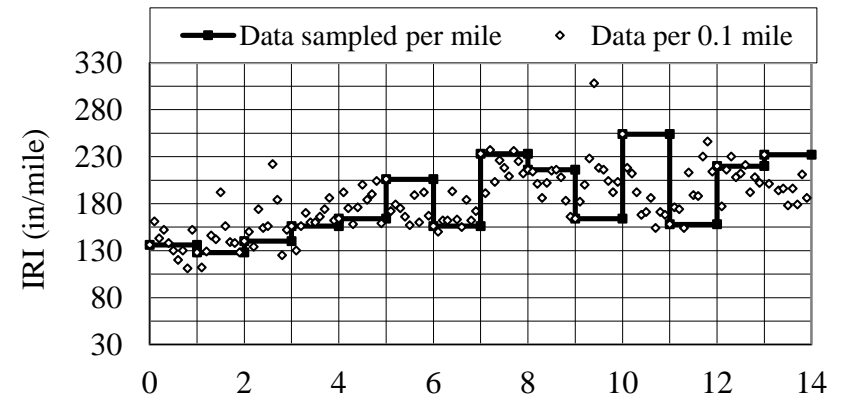


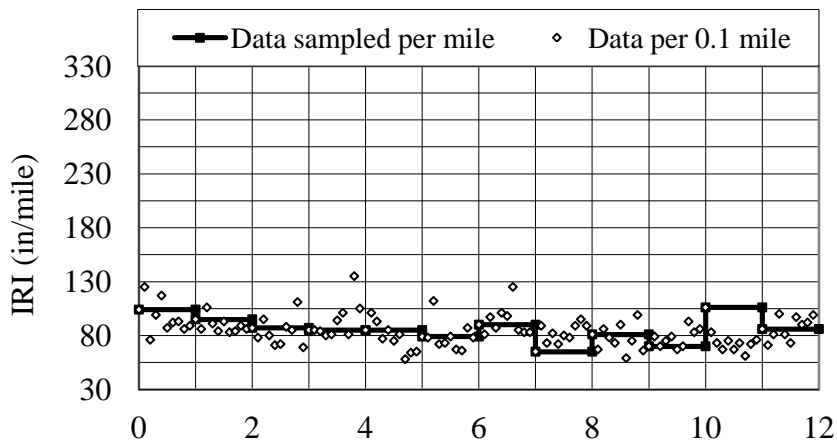
Figure B.1 Continuous and sampled IRI data along four roads in Colorado



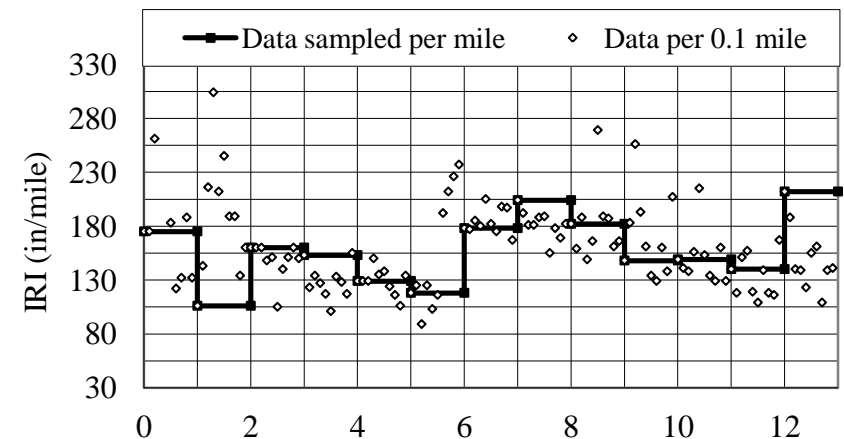
a) Mile point along LA-34 (flexible)
Year 1999



b) Mile point along LA-83 (rigid)
Year 1997

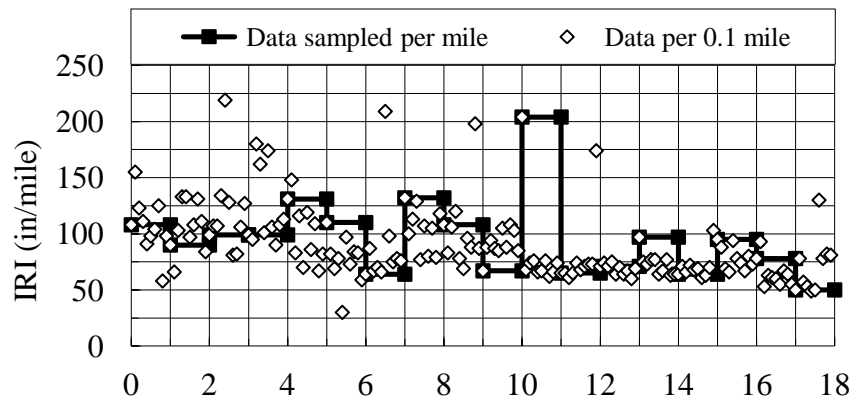


c) Mile point along LA-1 (composite)
Year 2000

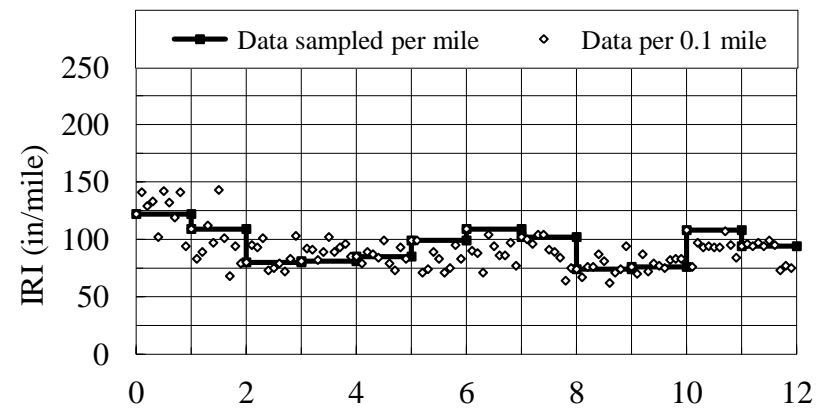


d) Mile point along LA-526 (rigid)
Year 1995

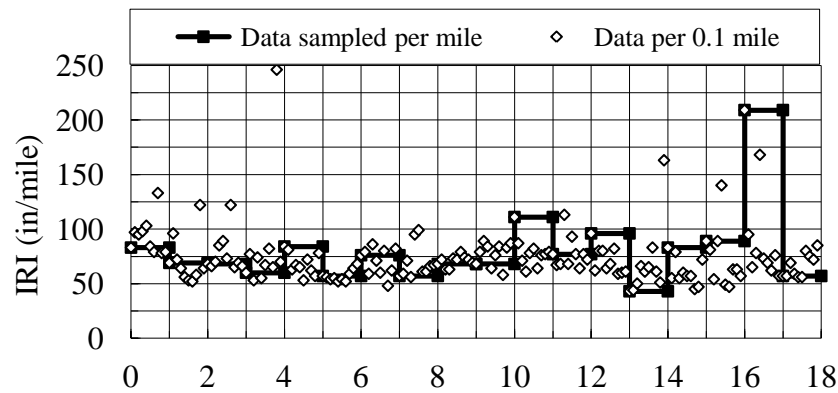
Figure B.2 Continuous and sampled IRI data along four roads in Louisiana



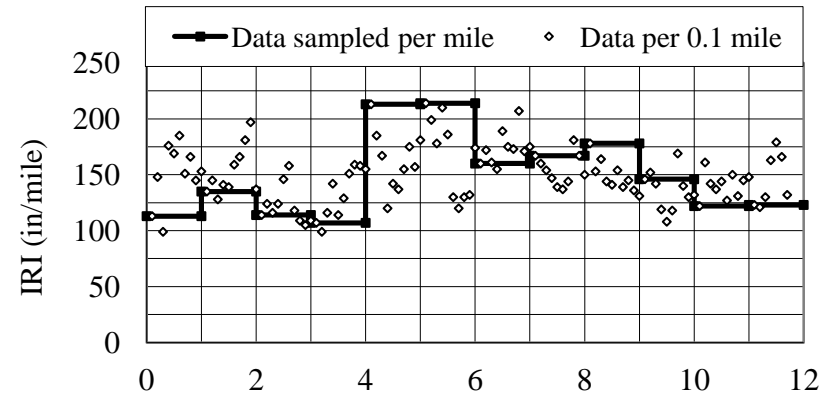
a) Mile point along I-94 (composite)
Year 2001



b) Mile point along I-69, control section 12033
(rigid) Year 2001



c) Mile point along US-31 (flexible)
Year 2001



d) Mile Point along I-69, control section 77024
(rigid) Year 2001

Figure B.3 Continuous and sampled IRI data per mile along four roads in Michigan

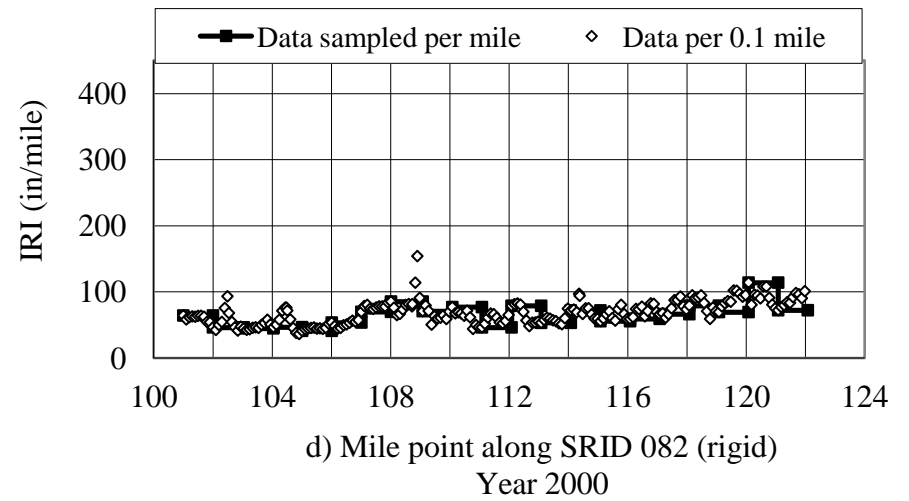
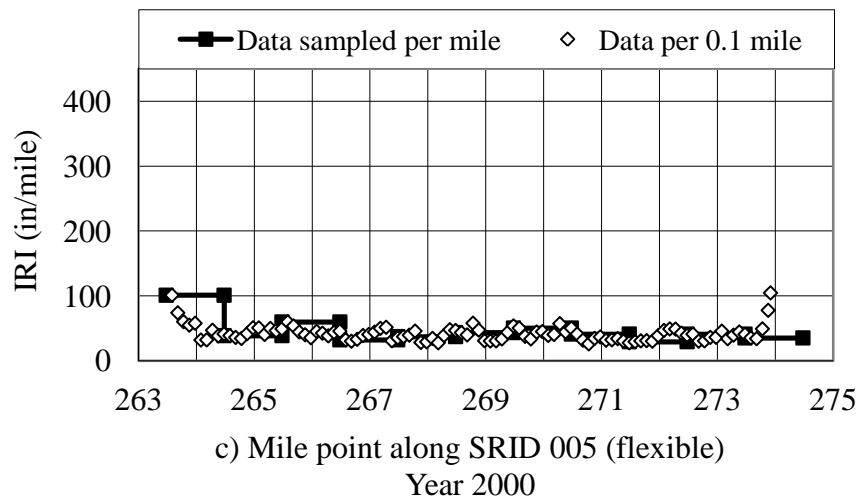
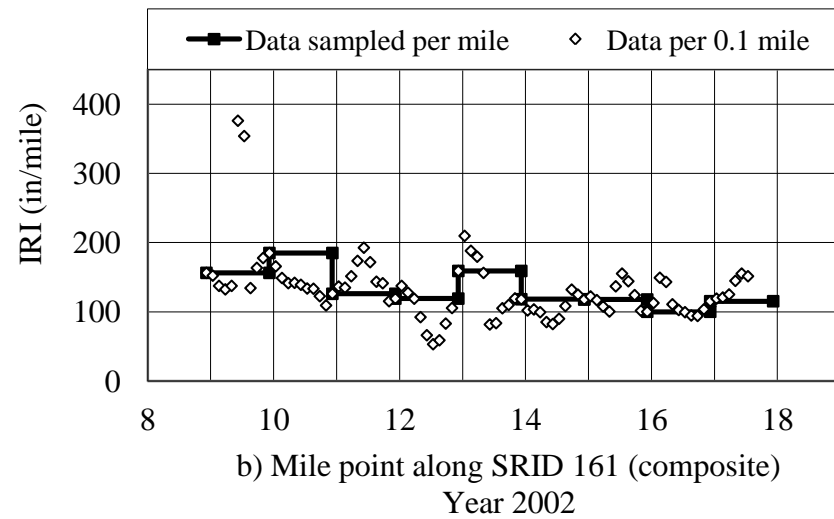
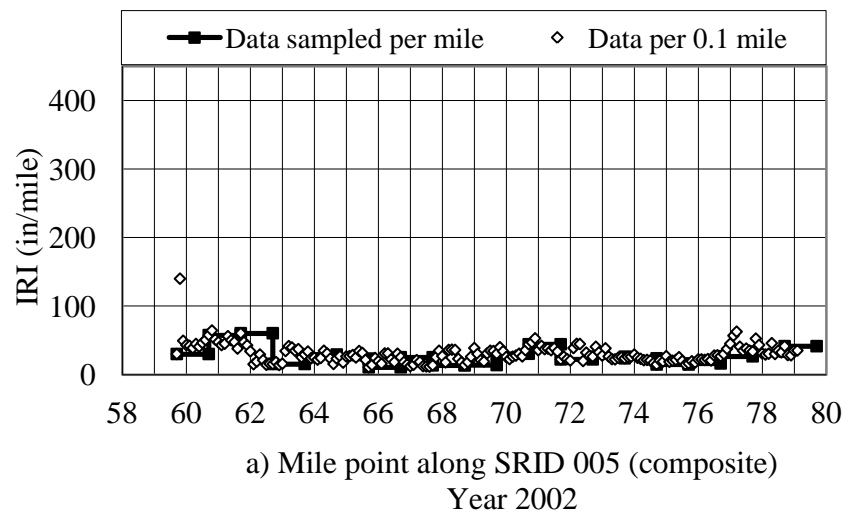


Figure B.4 Continuous and sampled IRI data per mile along four roads in Washington

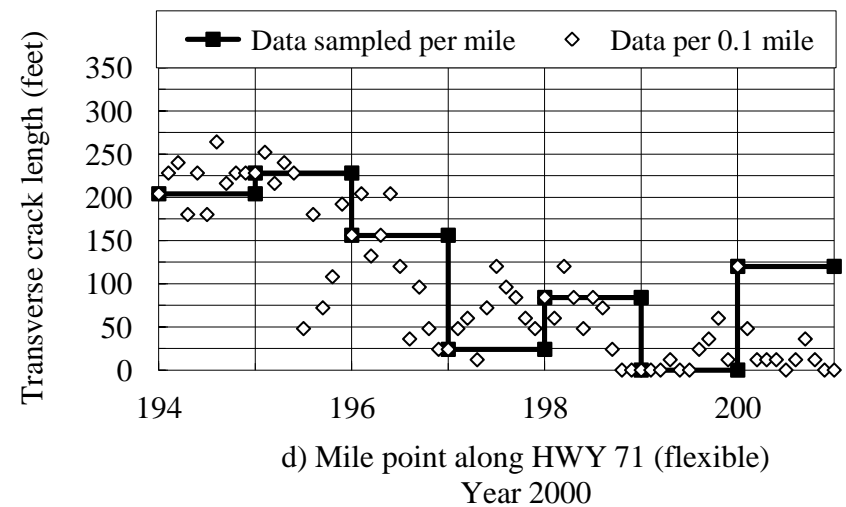
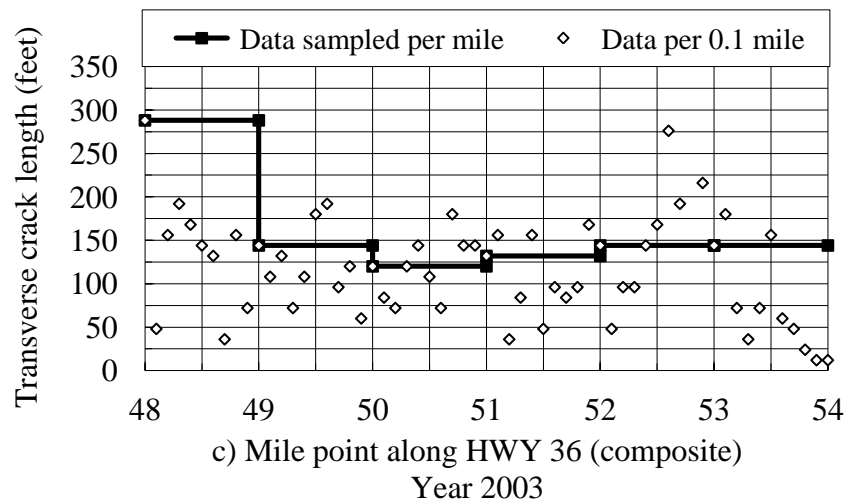
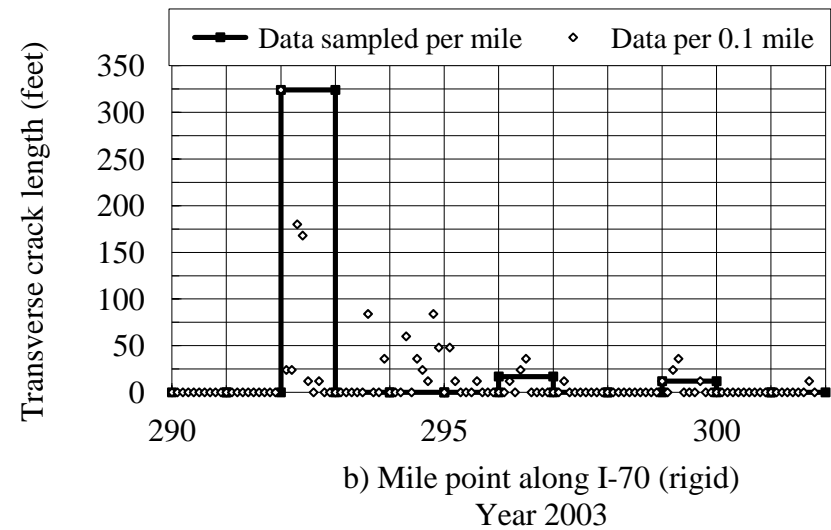
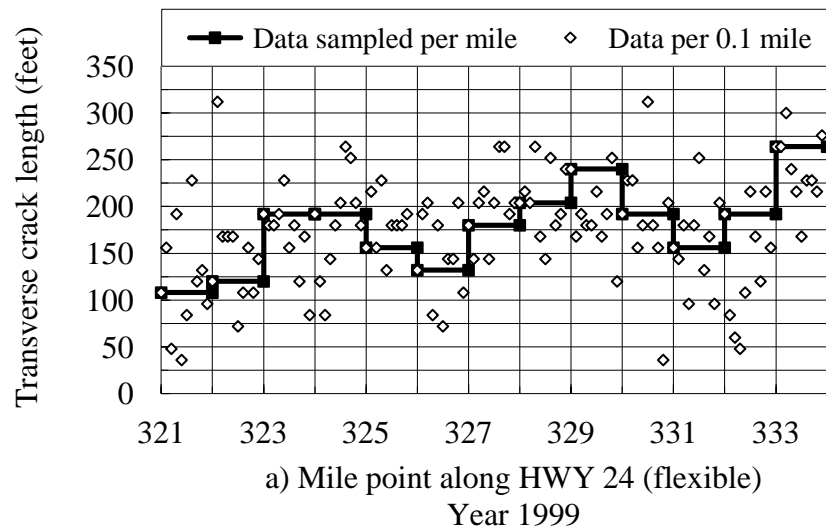


Figure B.5 Continuous and sampled transverse crack data along four roads in Colorado

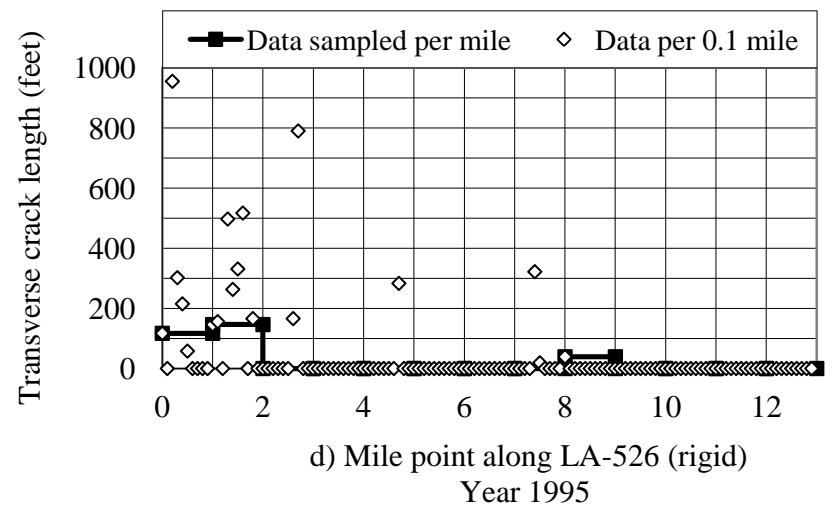
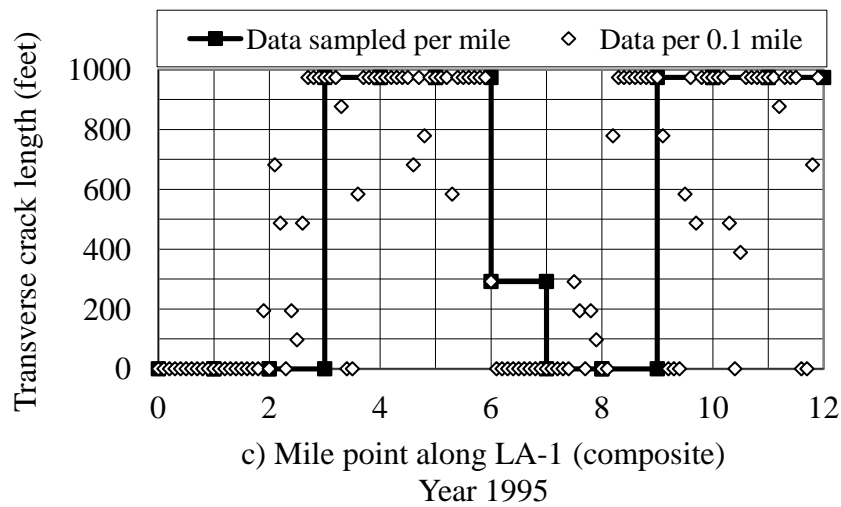
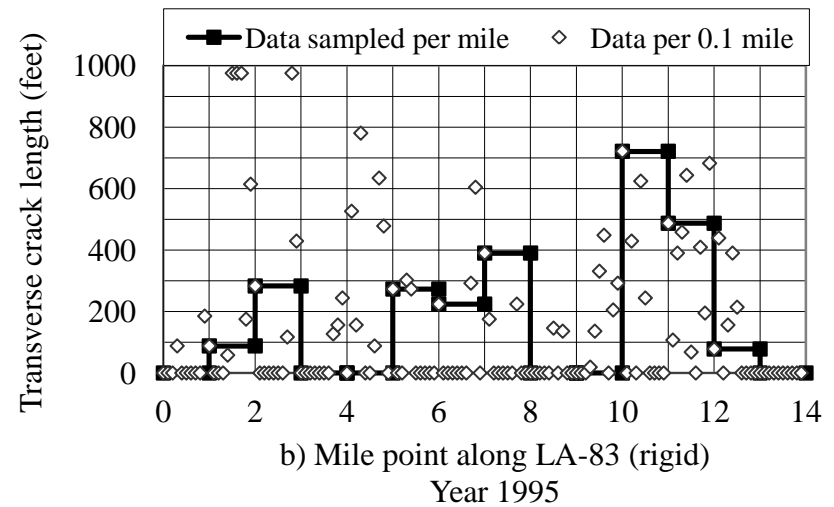
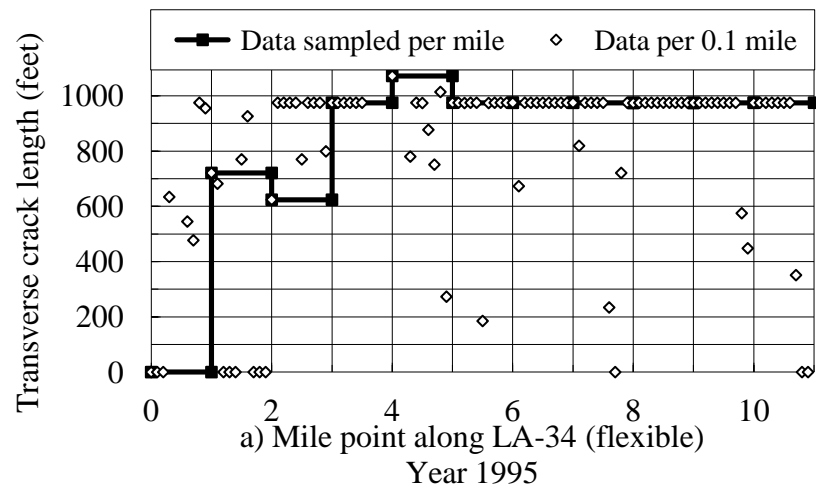


Figure B.6 Continuous and sampled transverse crack data per mile along four roads in Louisiana

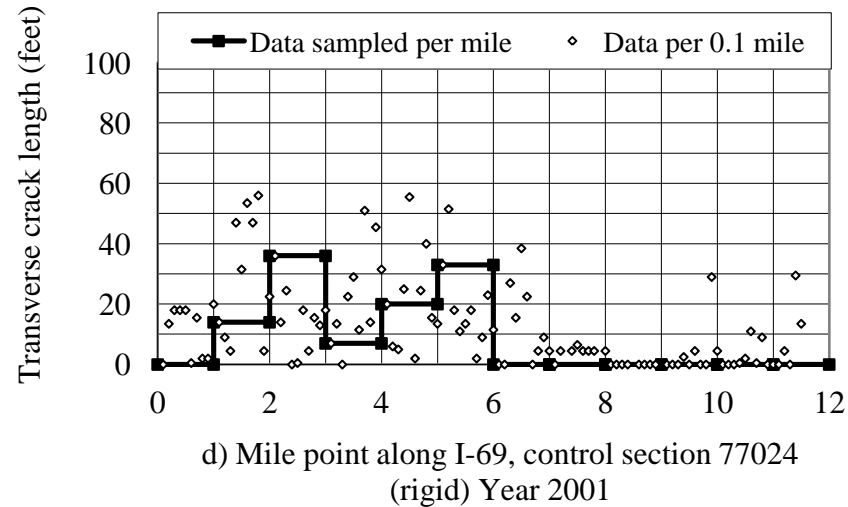
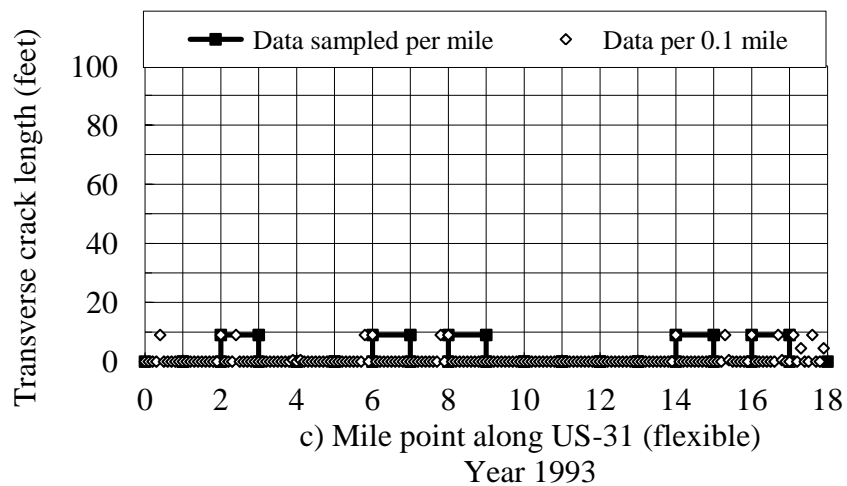
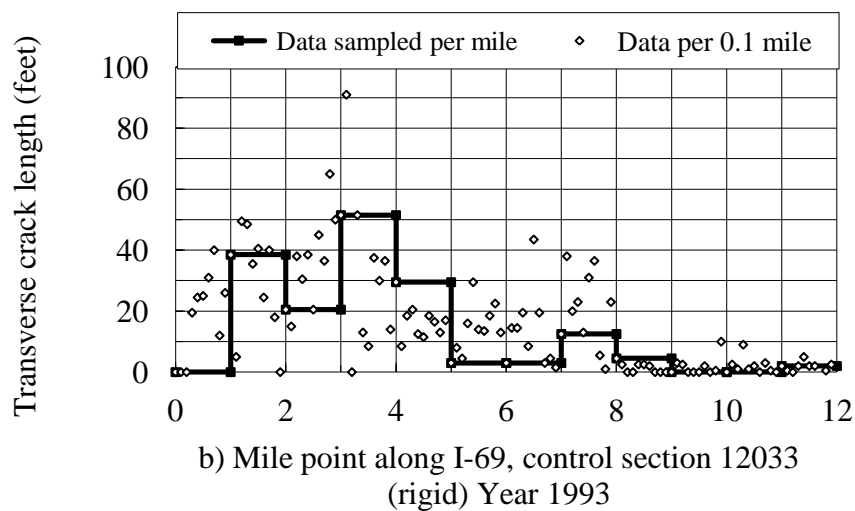
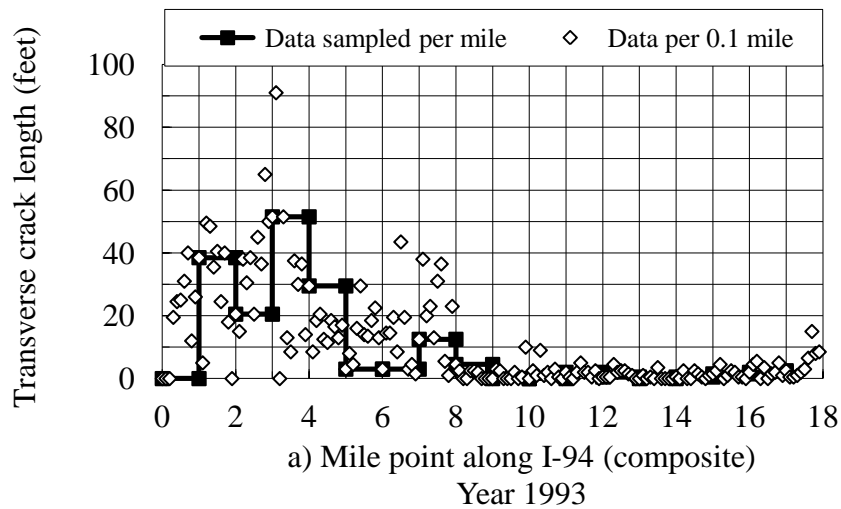


Figure B.7 Continuous and sampled transverse crack data per mile along four roads in Michigan

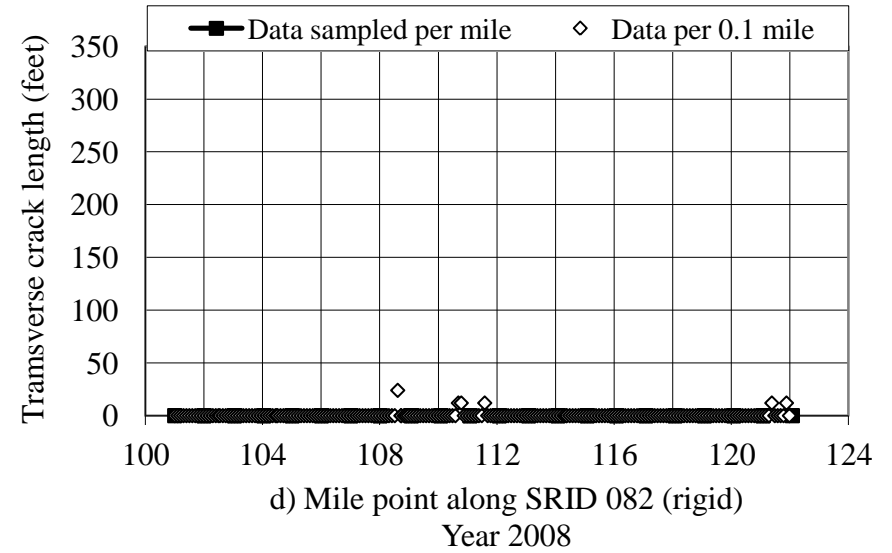
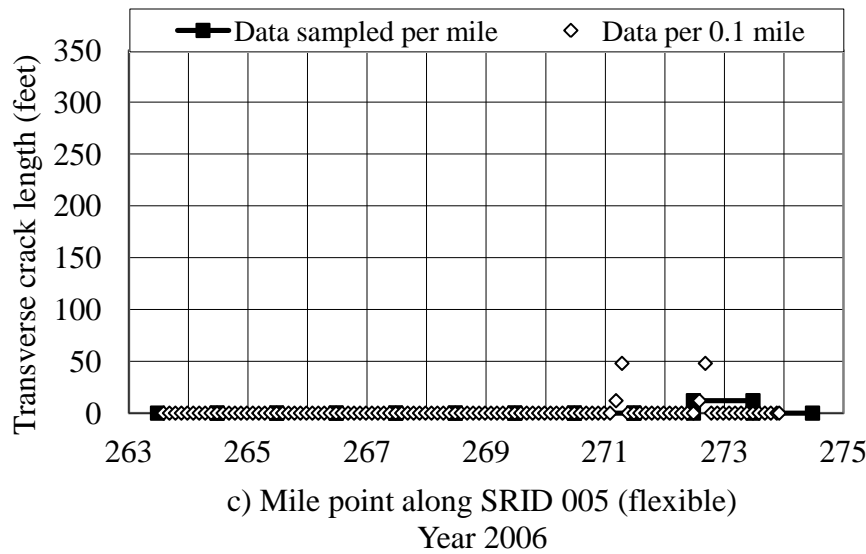
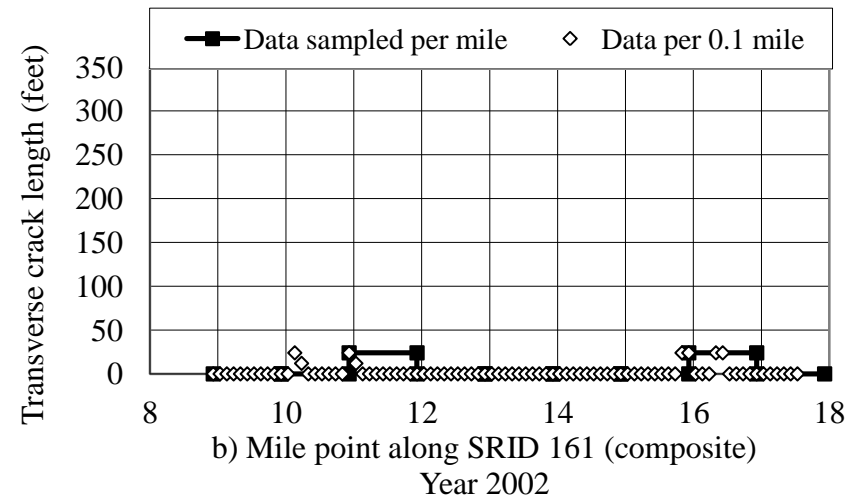
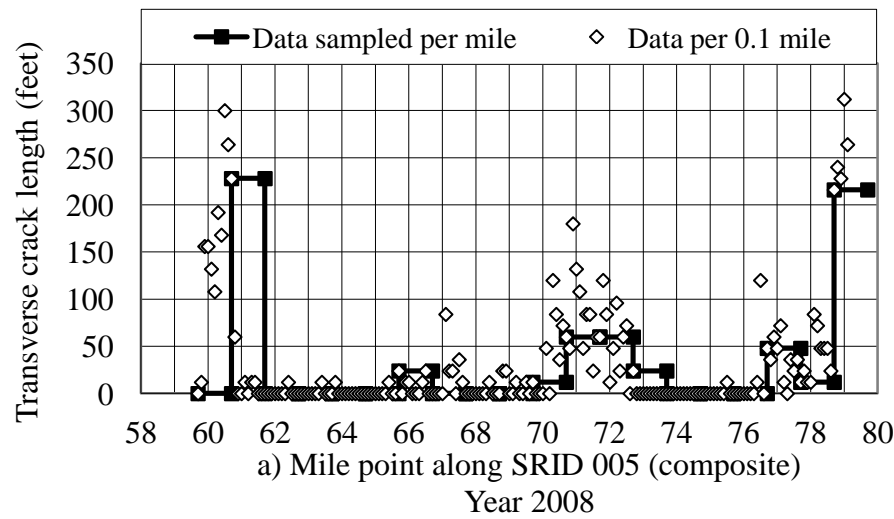


Figure B.8 Continuous and sampled transverse crack data per mile along four roads in Washington

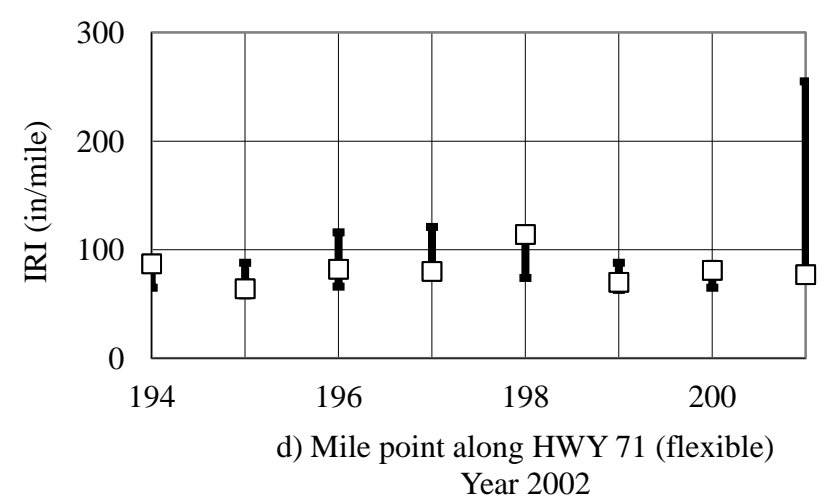
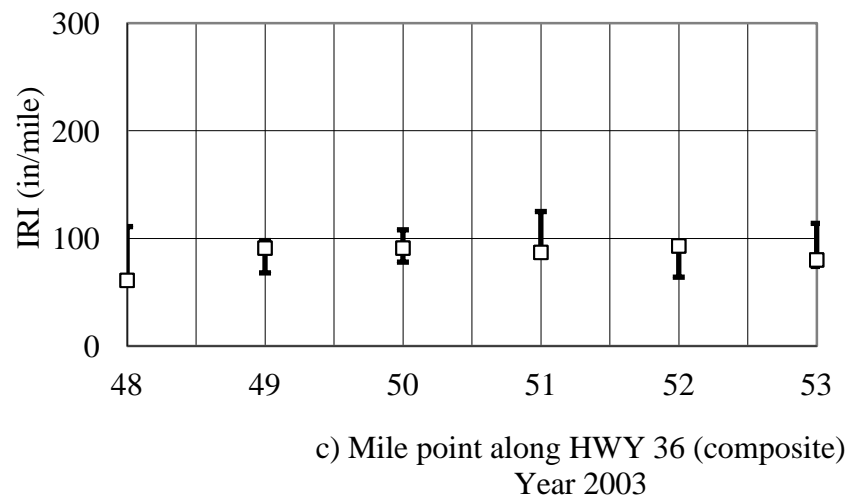
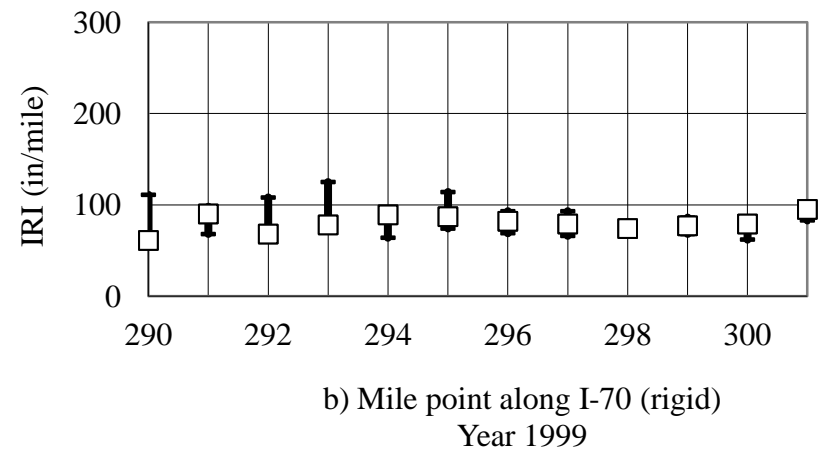
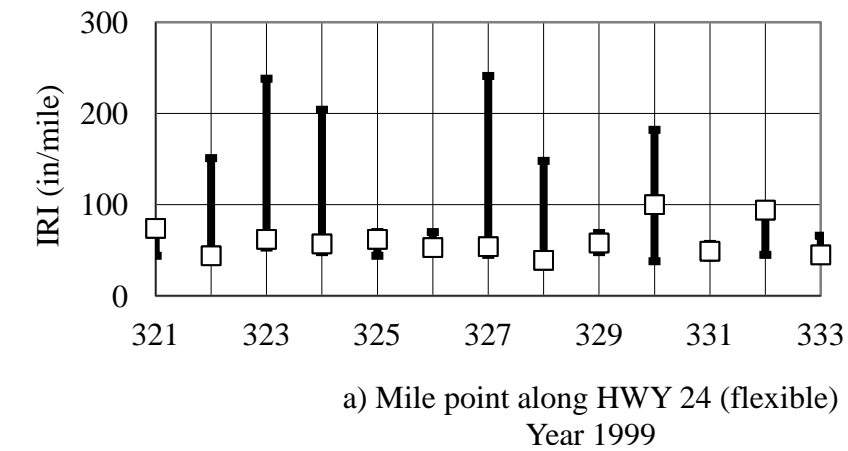
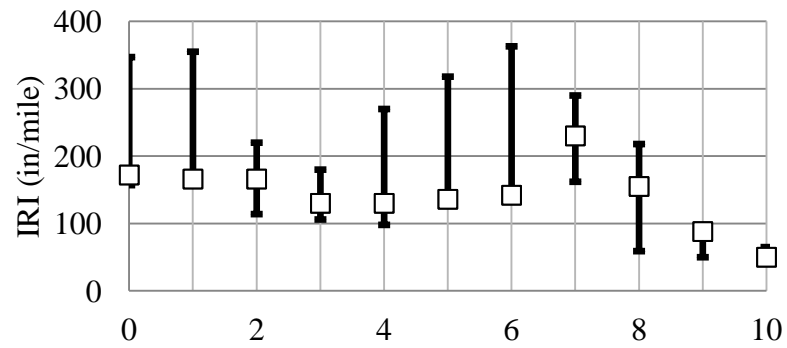
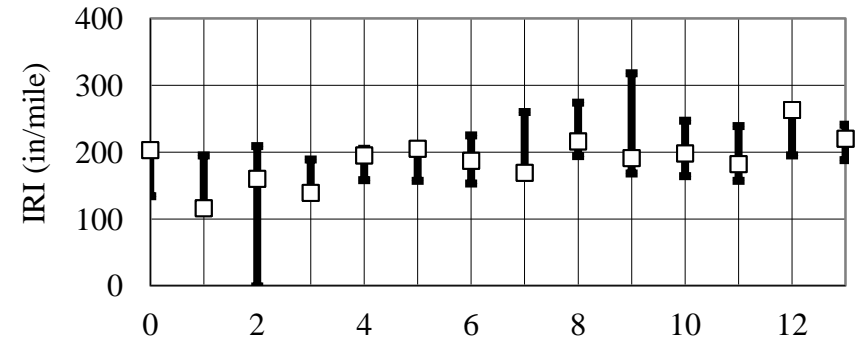


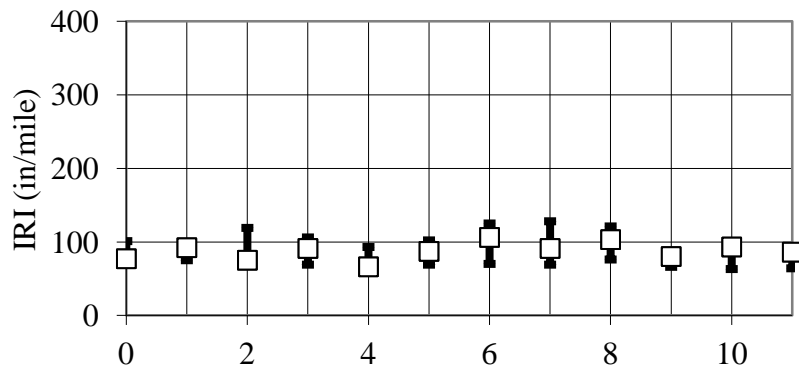
Figure B.9 Sampled (squares) and continuous IRI data per mile along four roads in Colorado



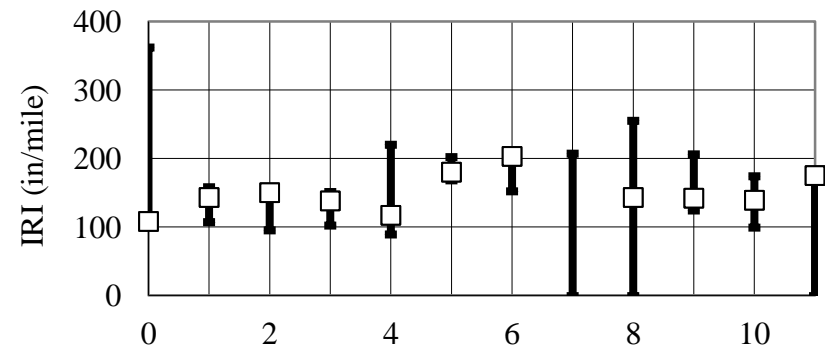
a) Mile point along LA-34 (flexible)
Year 1997



b) Mile point along LA-83 (rigid)
Year 2005

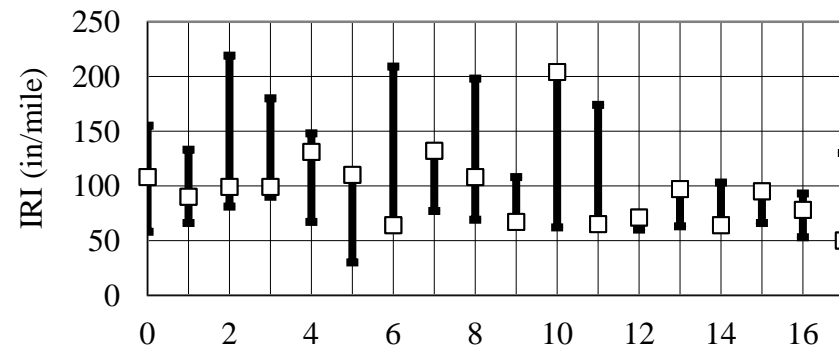


c) Mile point along LA-1 (composite)
Year 1995

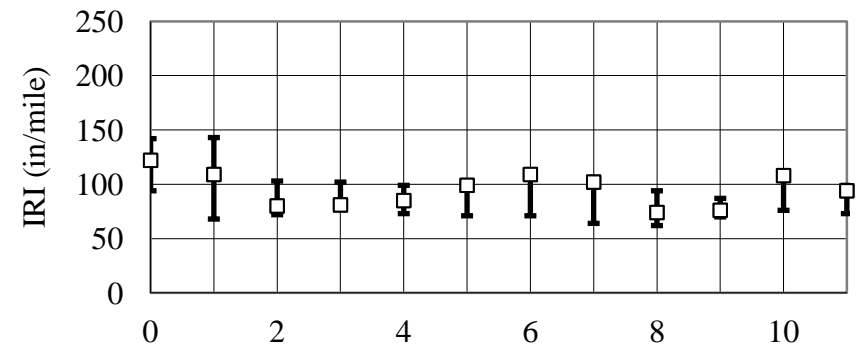


d) Mile point along LA-526 (rigid)
Year 2003

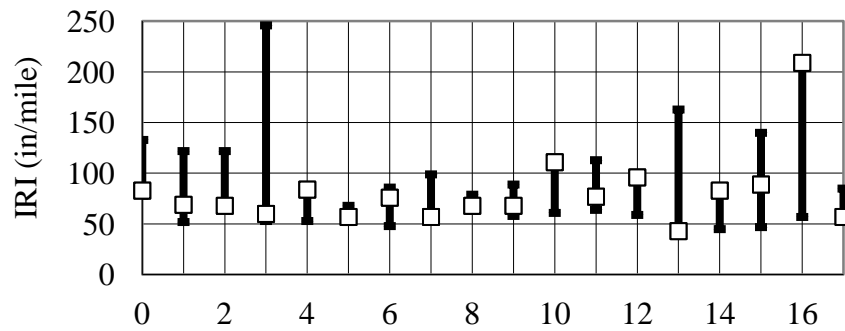
Figure B.10 Sampled (squares) and continuous IRI data per mile along four roads in Louisiana



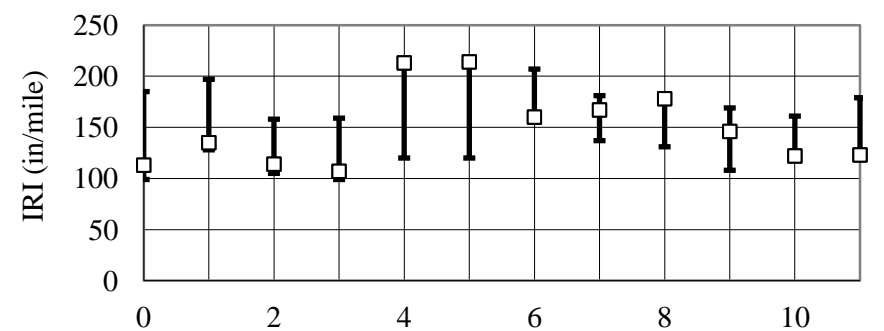
a) Mile point along I-94 (composite)
Year 2001



b) Mile point along I-69, control section 12033 (rigid)
Year 2001



c) Mile point along US-31 (flexible)
Year 2001



d) Mile point along I-69, control section 77024 (rigid)
Year 2001

Figure B.11 Sampled (squares) and continuous IRI data per mile along four roads in Michigan

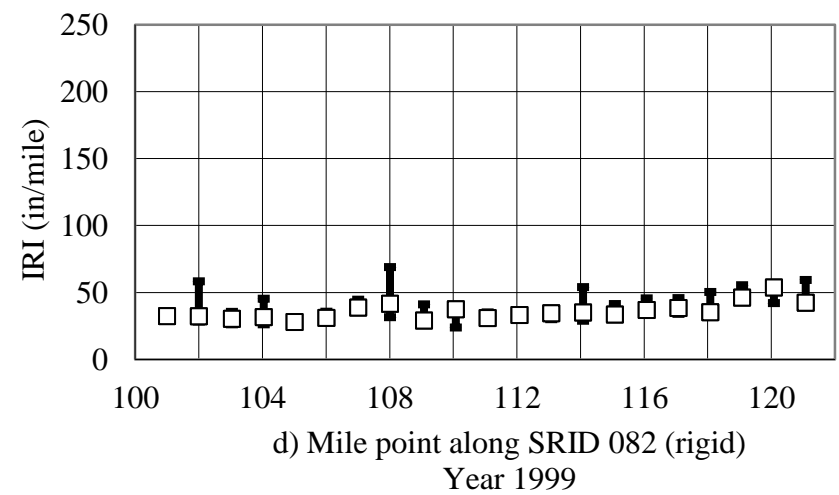
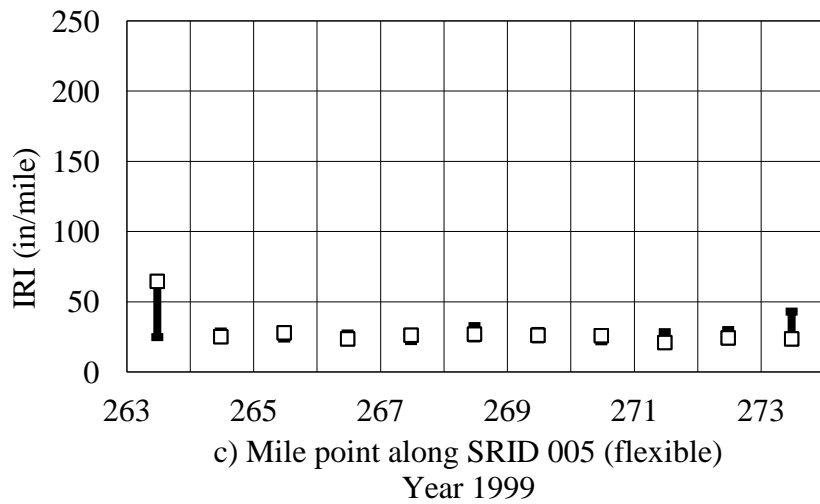
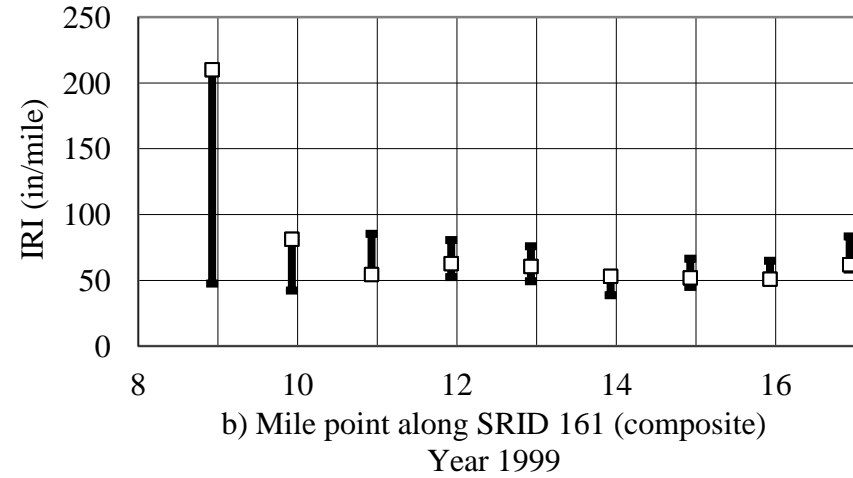
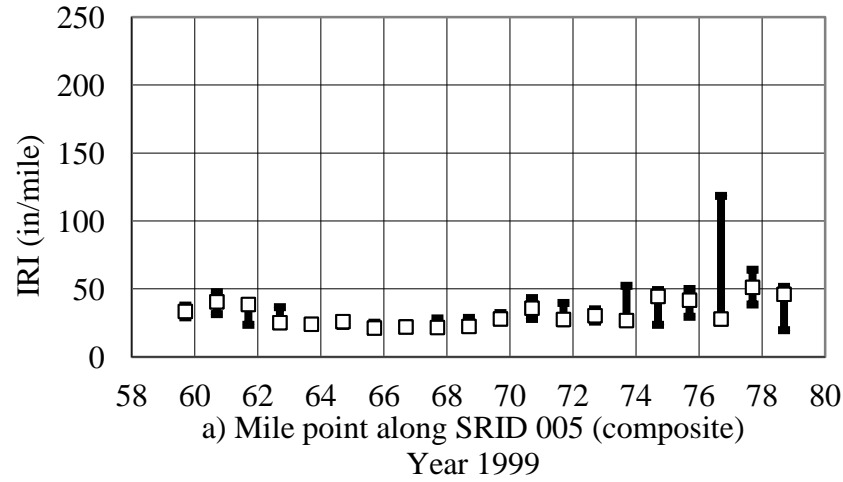
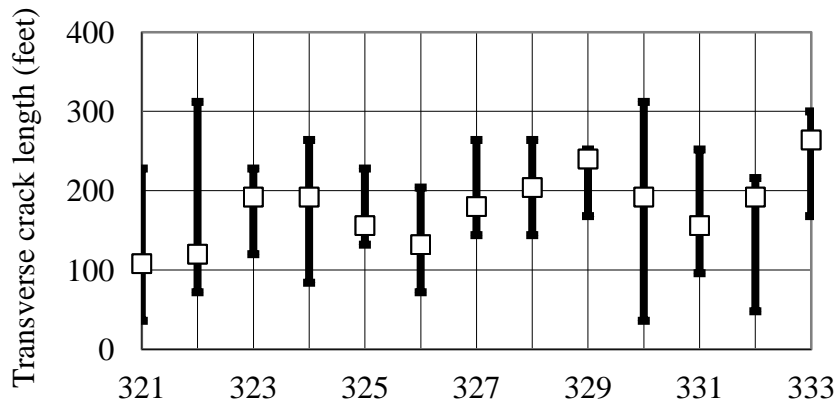
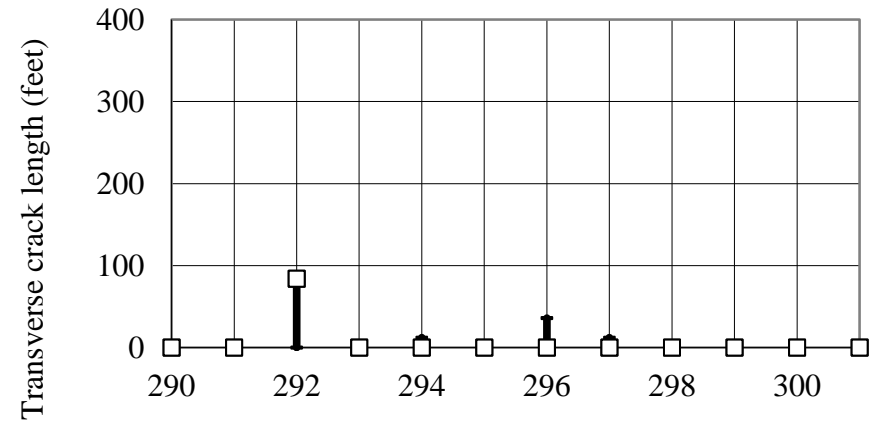


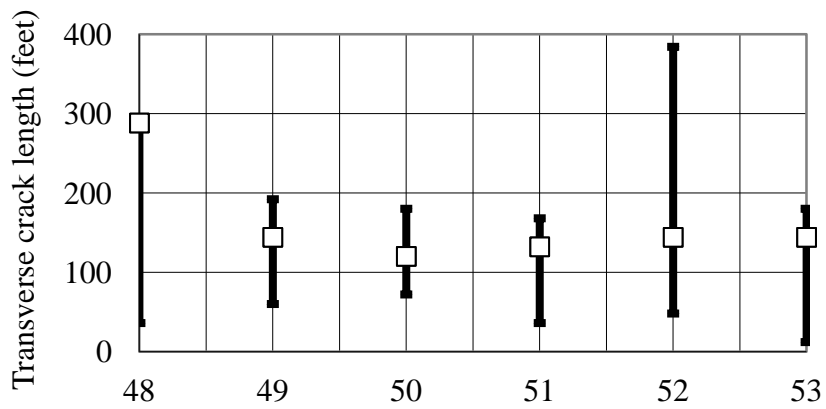
Figure B.12 Sampled (squares) and continuous IRI data per mile along four roads in Washington



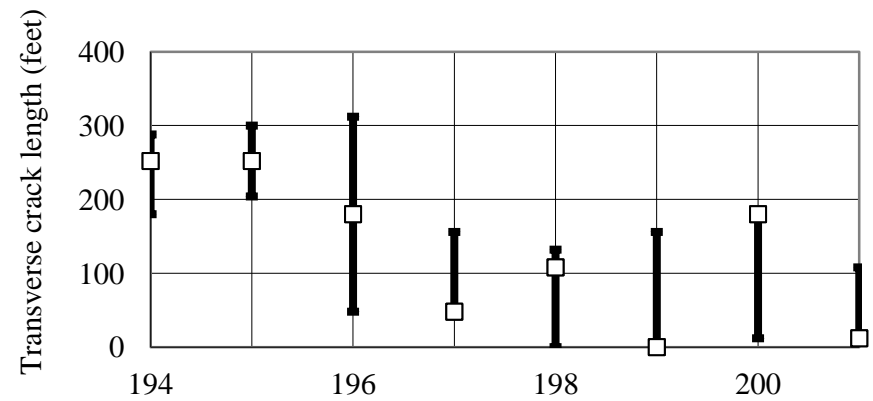
a) Mile point along HWY 24 (flexible)
Year 1999



b) Mile point along I-70 (rigid)
Year 2000

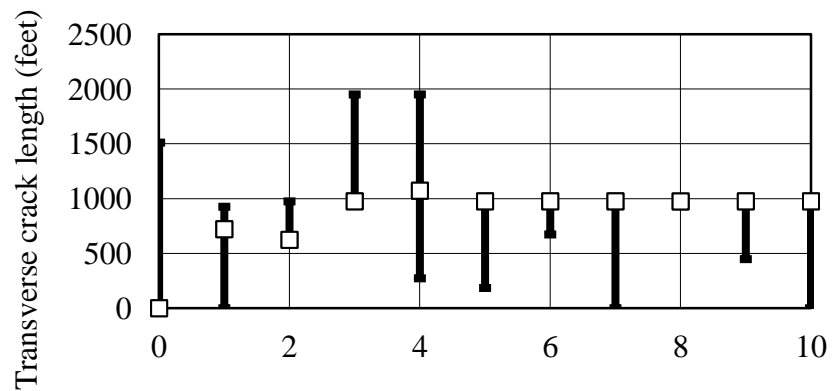


c) Mile point along HWY 36 (composite)
Year 2003

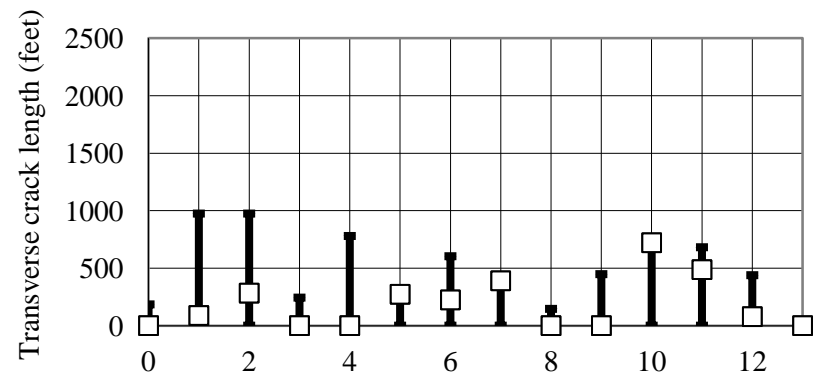


d) Mile point along HWY 71 (flexible)
Year 2002

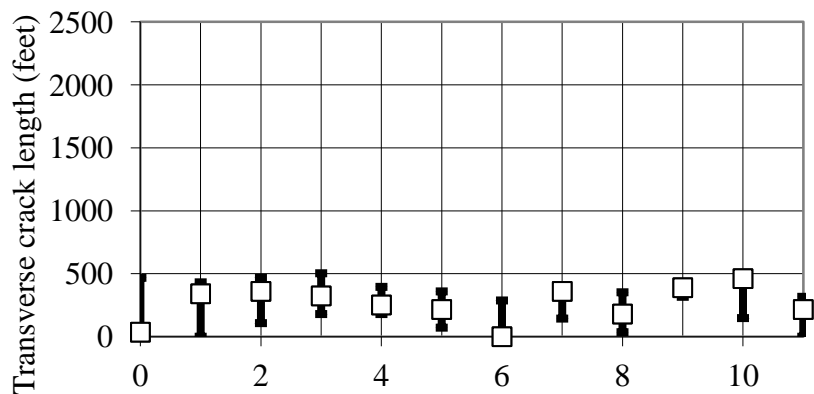
Figure B.13 Sampled (squares) and continuous transverse crack data per mile along four roads in Colorado



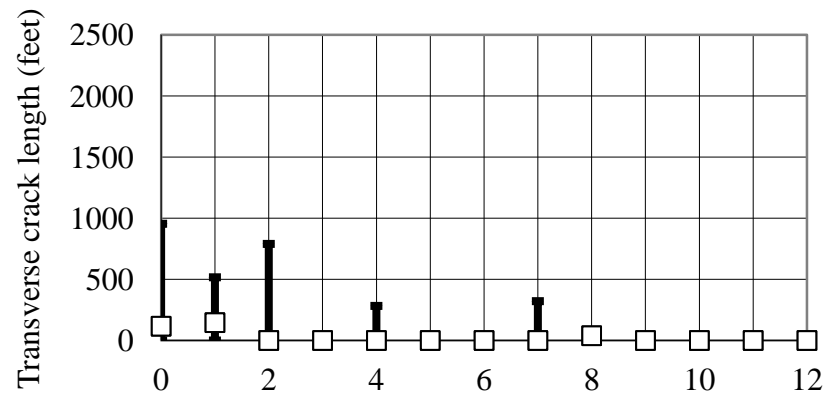
a) Mile point along LA-34 (flexible)
Year 1995



b) Mile point along LA-83 (rigid)
Year 1995



c) Mile point along LA-1 (composite)
Year 1997



d) Mile point along LA-526 (rigid)
Year 1995

Figure B.14 Sampled (squares) and continuous transverse crack data per mile four roads in Louisiana

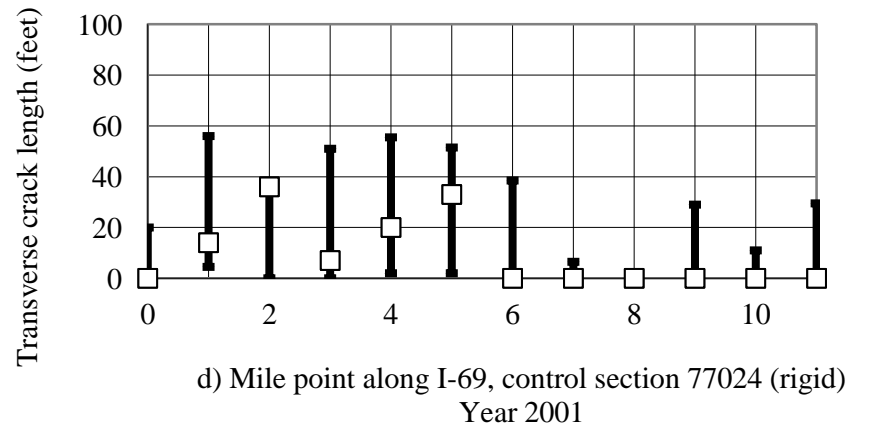
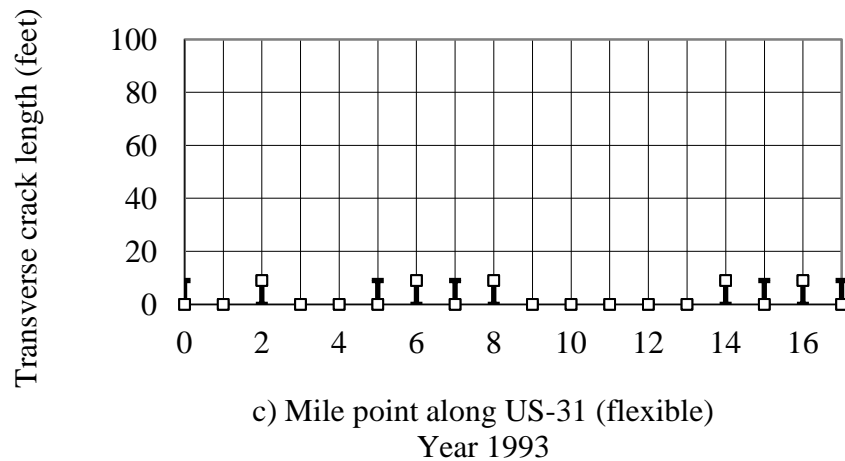
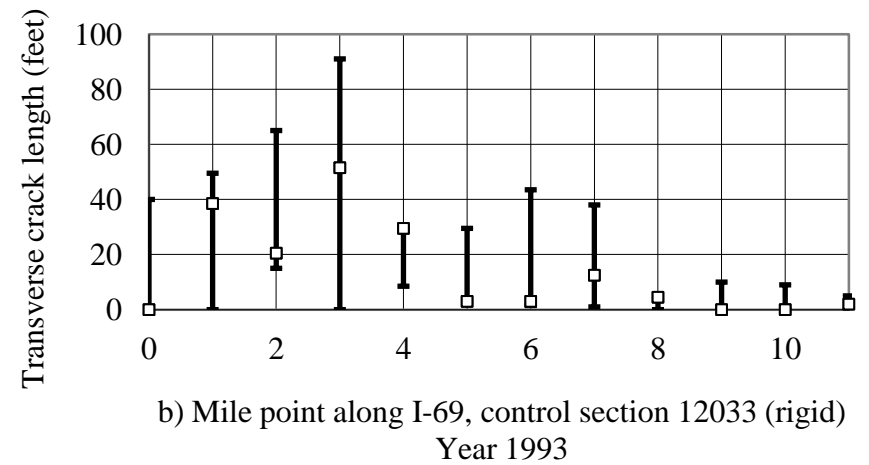
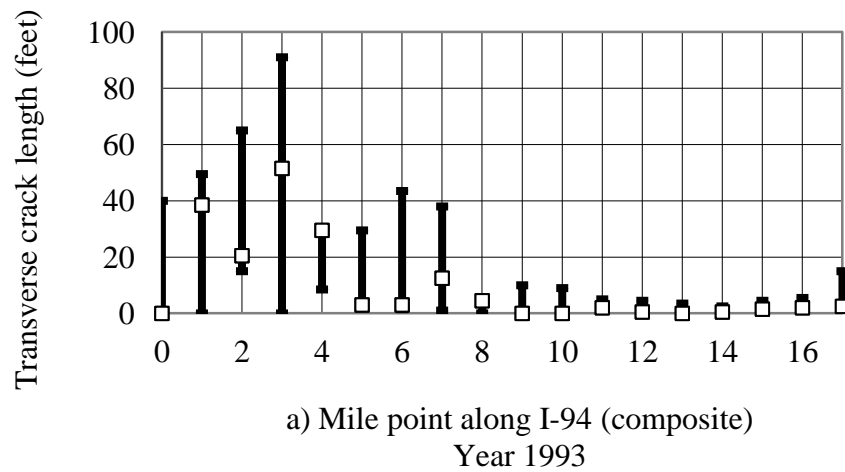


Figure B.15 Sampled (squares) and continuous transverse crack data per mile along four roads in Michigan

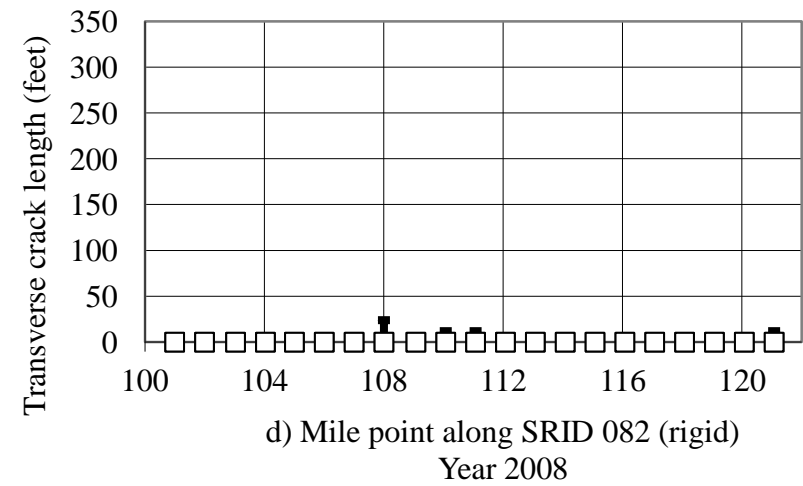
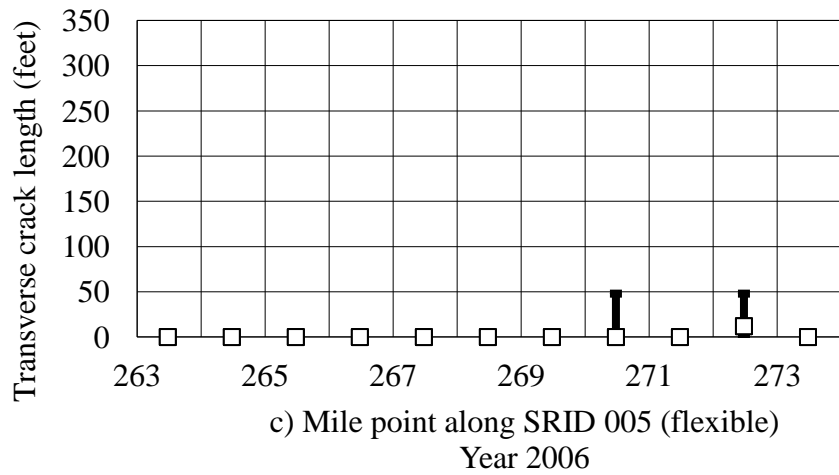
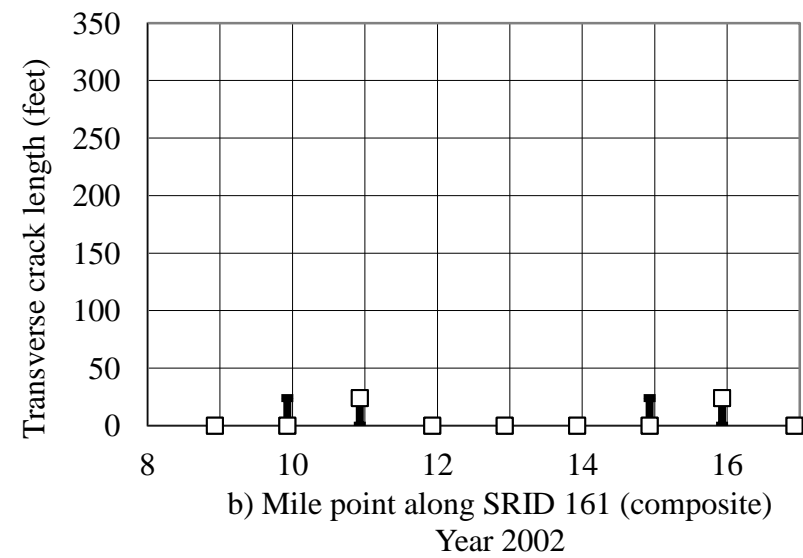
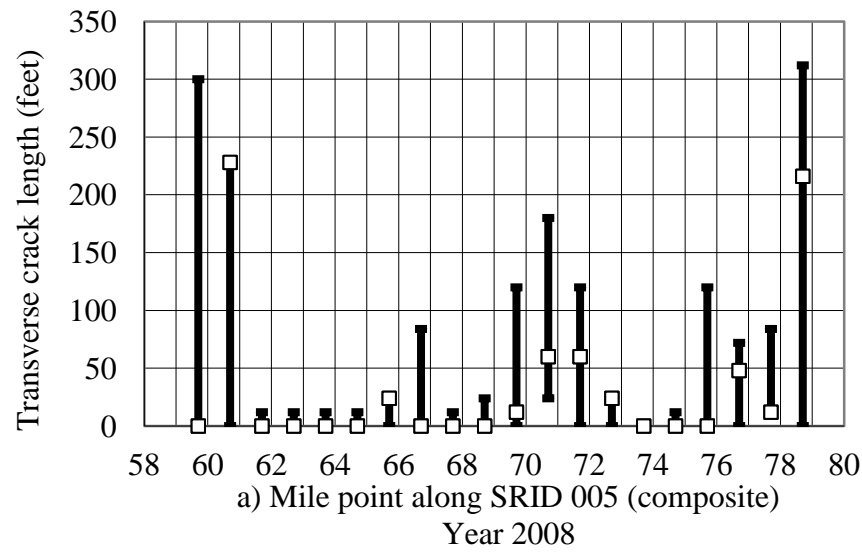


Figure B.16 Sampled (squares) and continuous transverse crack data per mile along four roads in Washington

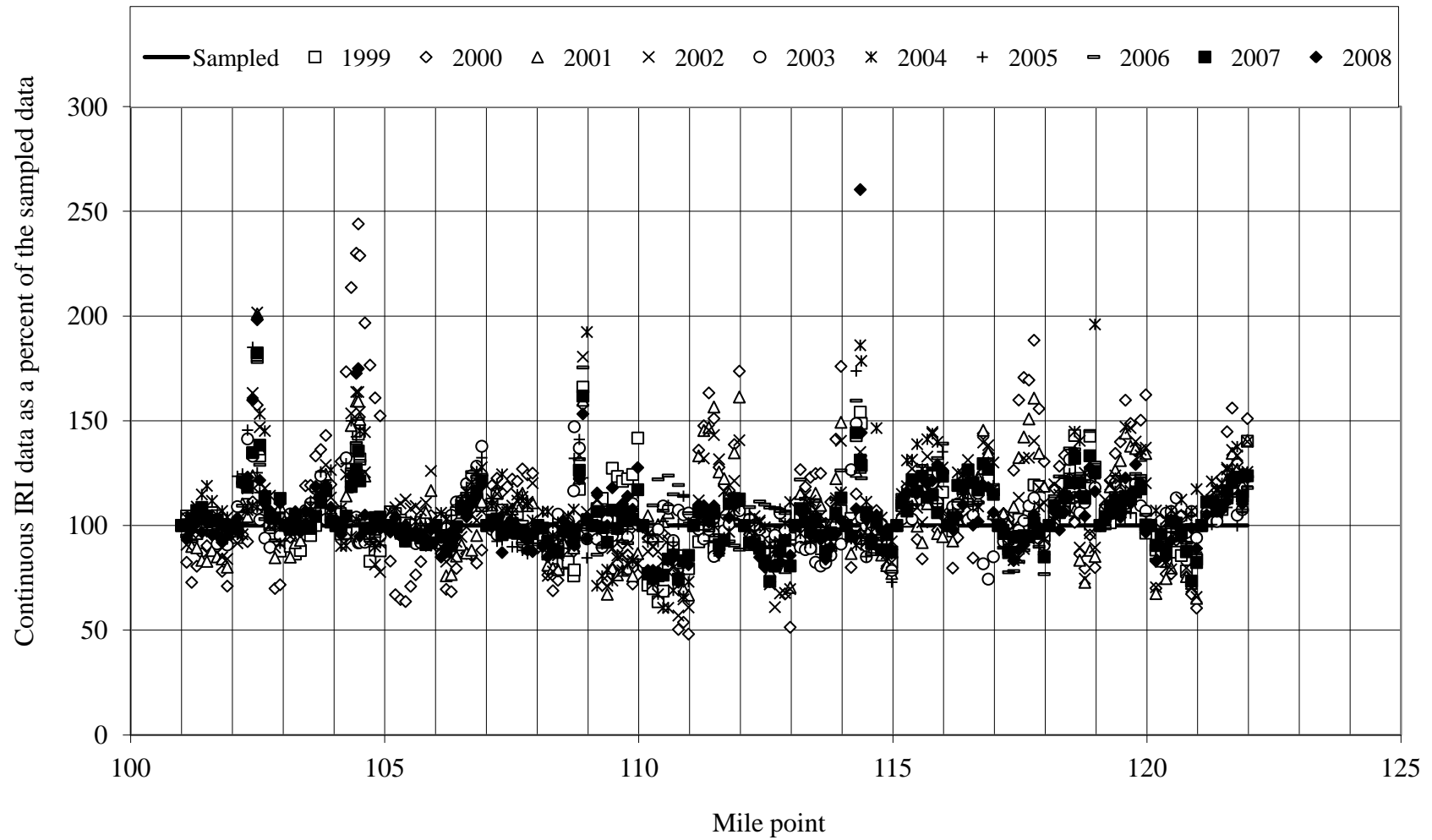


Figure B.17 Continuous time series IRI data as percent of the sampled data along HWY 24 (flexible) in Colorado

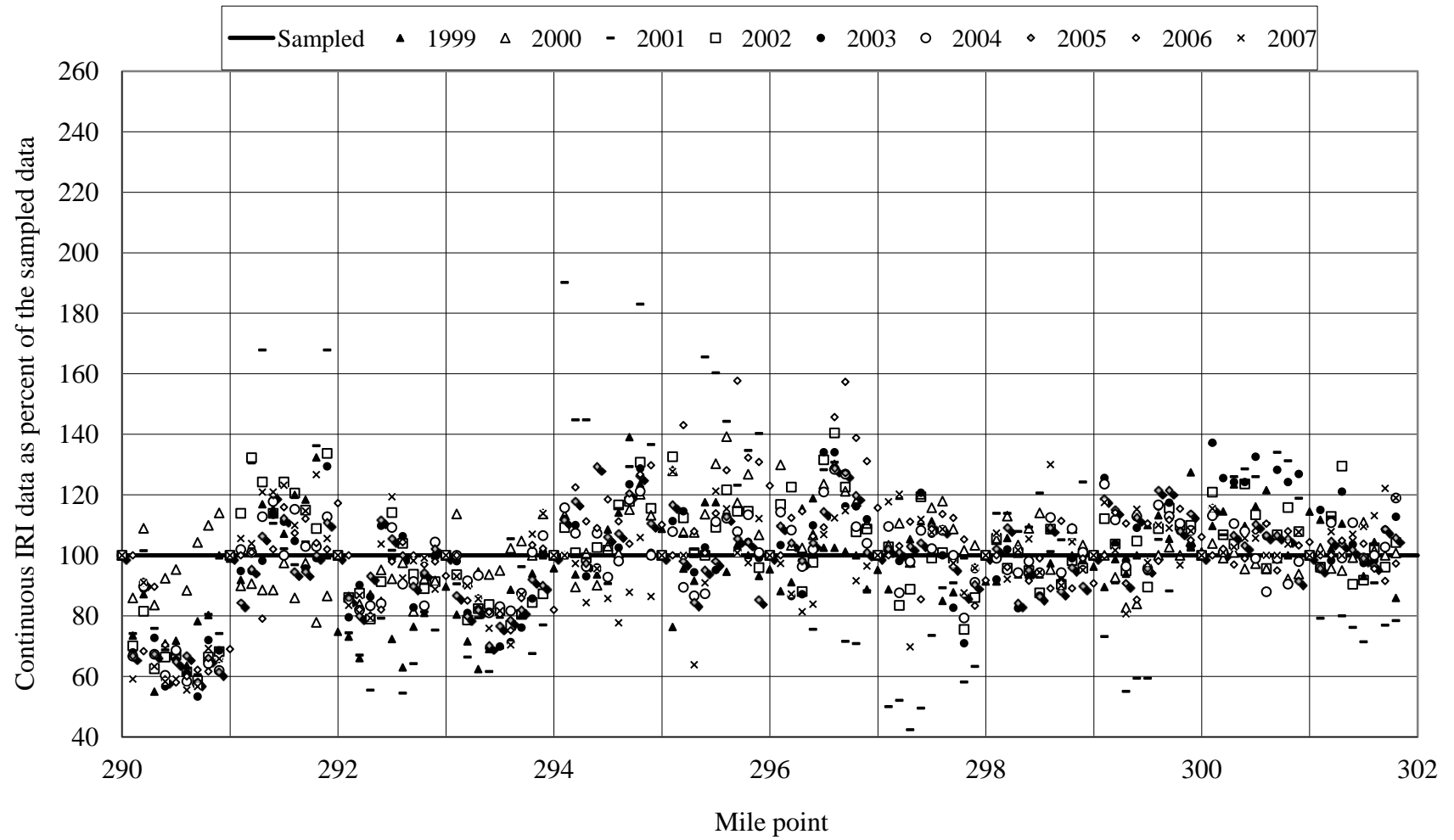


Figure B.18 Continuous time series IRI data as percent of the sampled data along I-70 (rigid) in Colorado

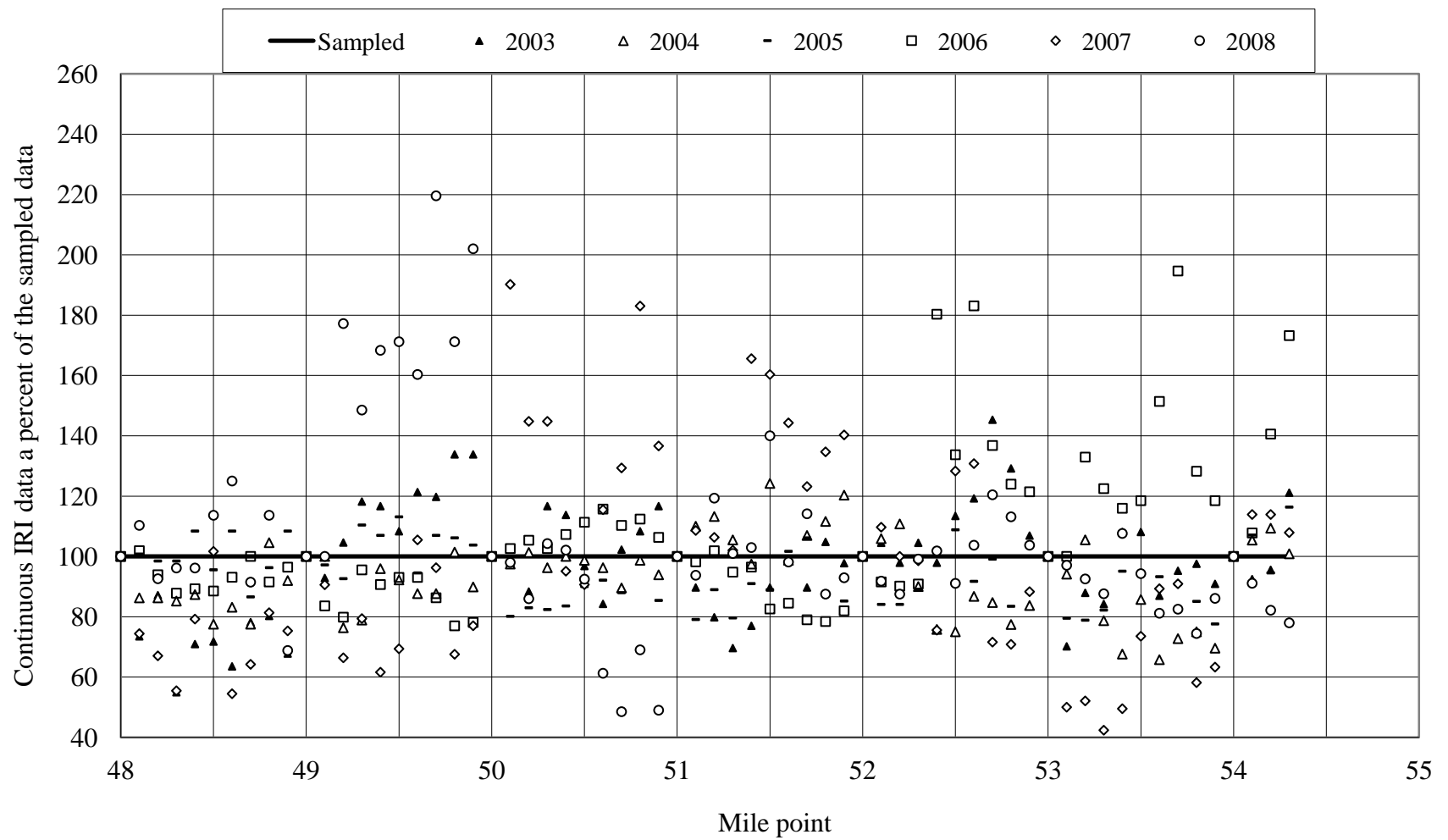


Figure B.19 Continuous time series IRI data as percent of the sampled data along HWY 36 (composite) in Colorado

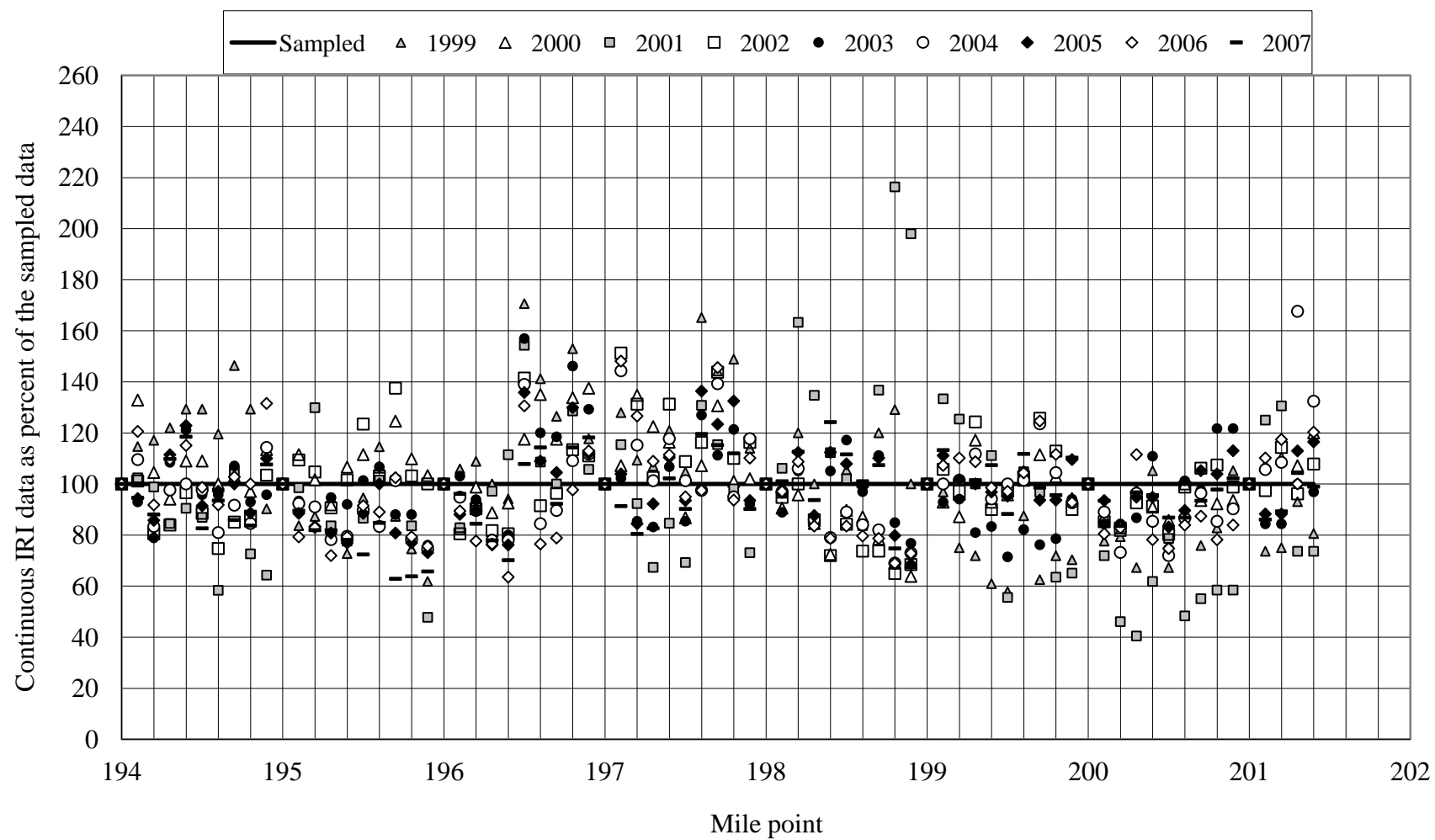


Figure B.20 Continuous time series IRI data as percent of the sampled data along HWY 71 (flexible) in Colorado

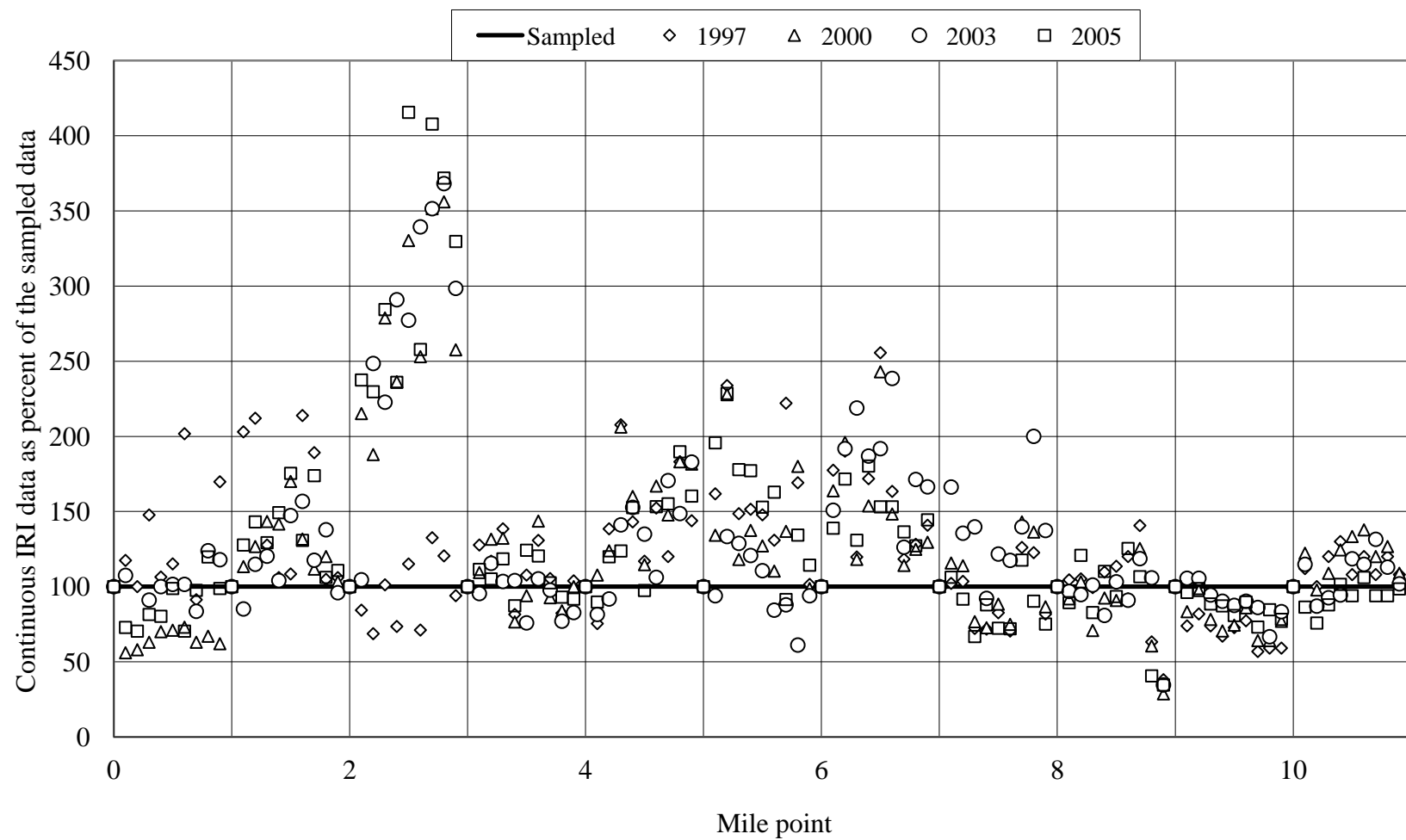


Figure B.21 Continuous time series IRI data as percent of the sampled data along LA-34 (flexible) in Louisiana

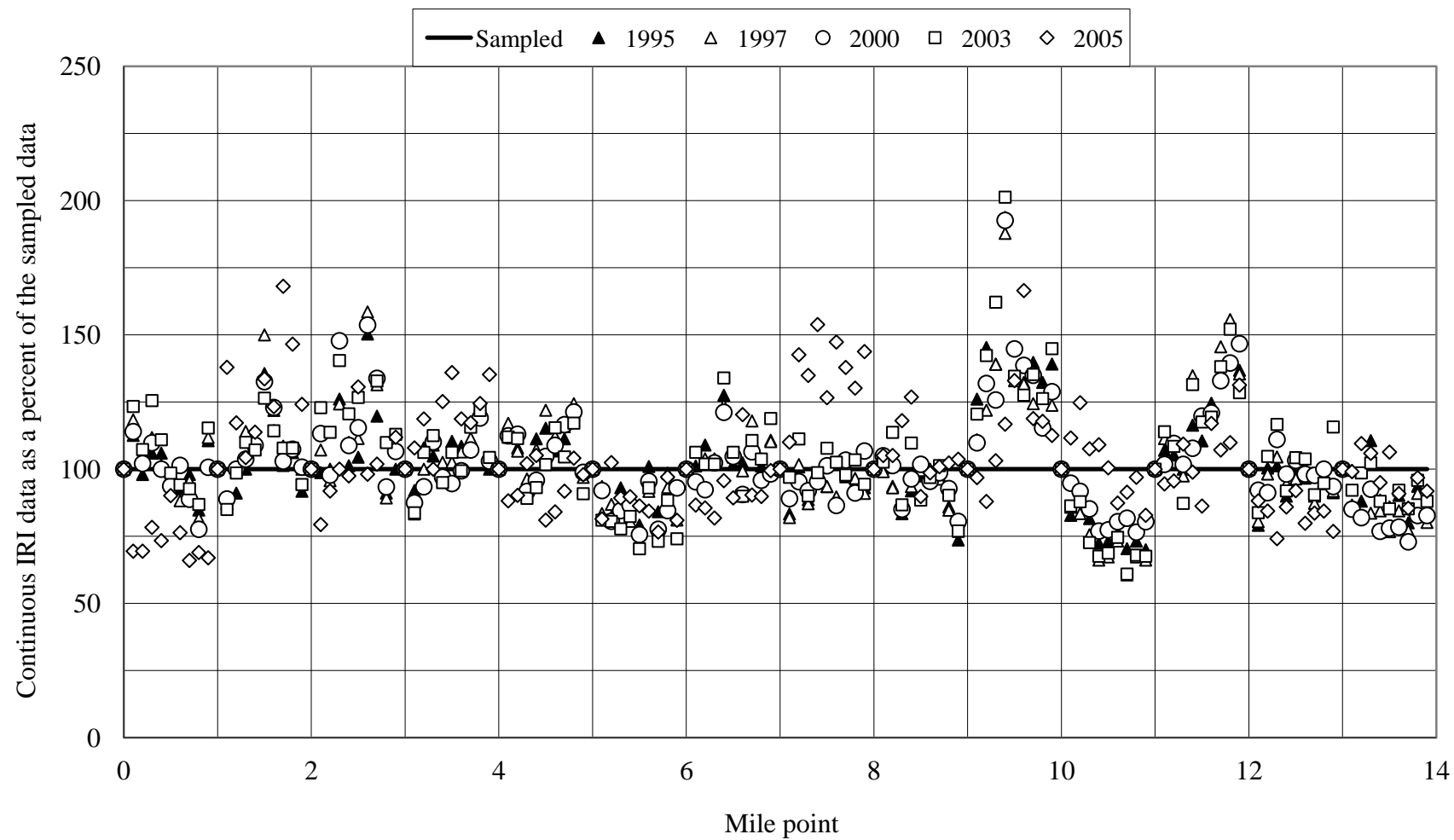


Figure B.22 Continuous time series IRI data as percent of the sampled data along LA-83 (rigid) in Louisiana

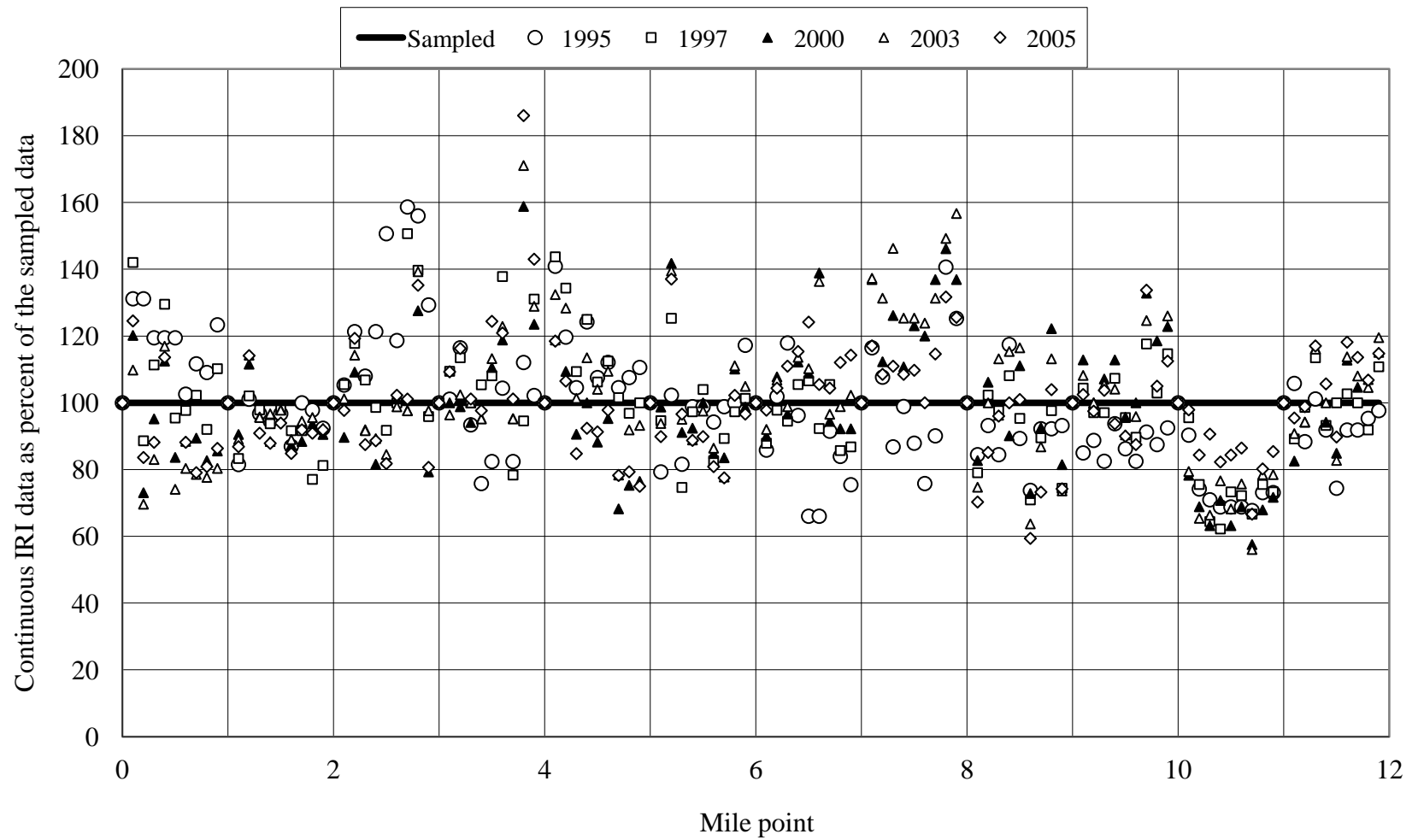


Figure B.23 Continuous time series IRI data as percent of the sampled data along LA-1 (composite) in Louisiana

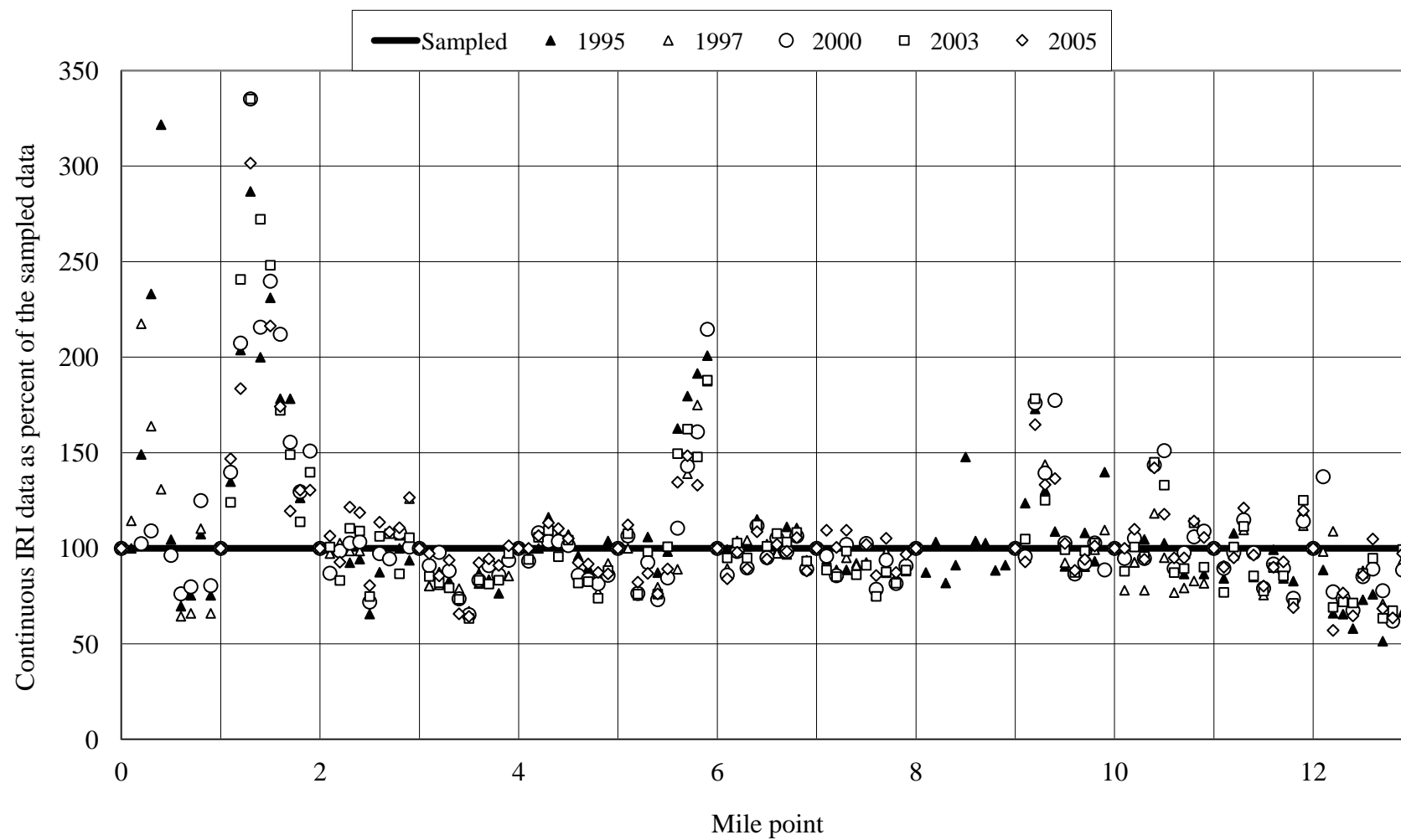


Figure B.24 Continuous time series IRI data as percent of the sampled data along LA-526 (rigid) in Louisiana

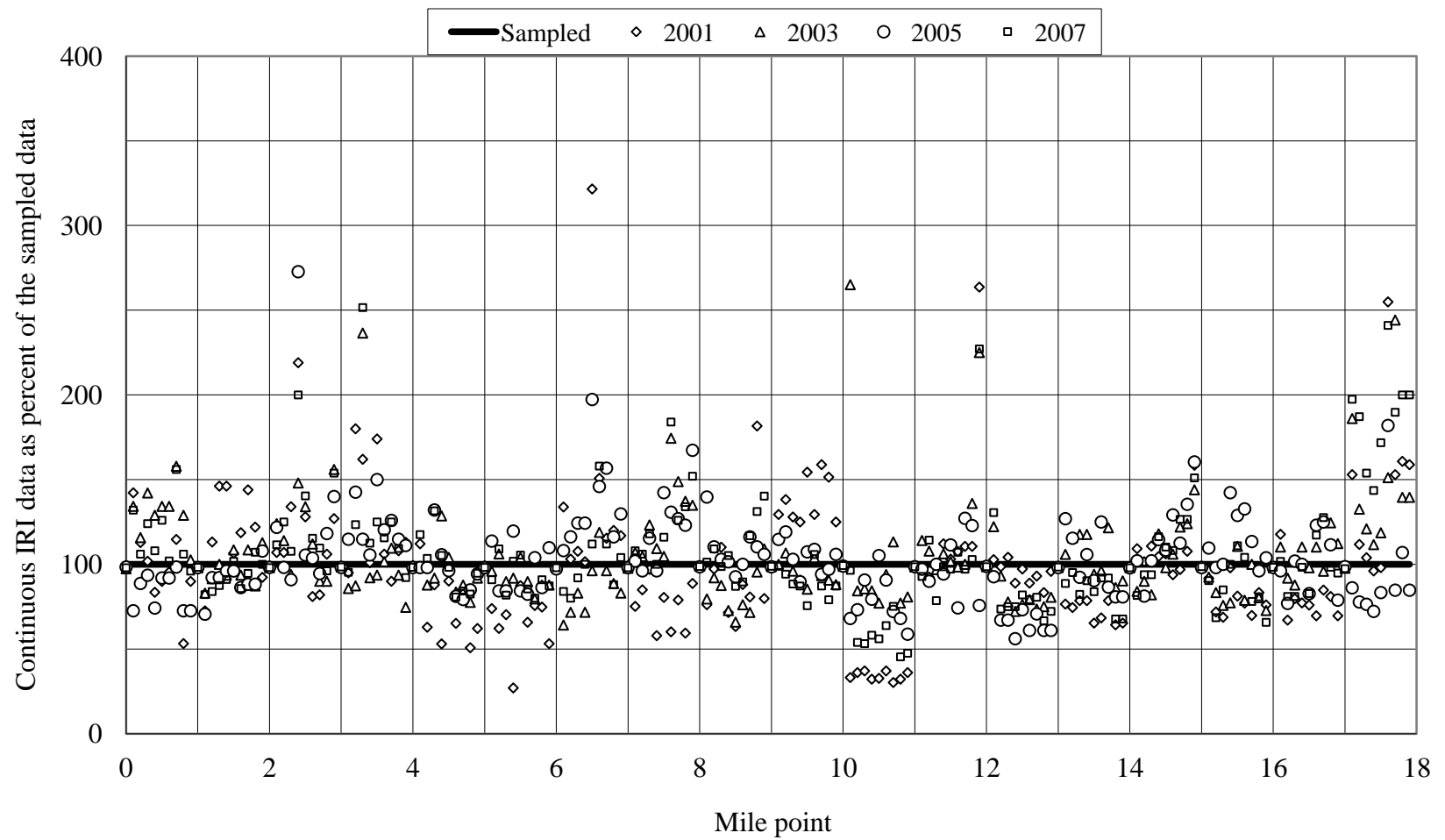


Figure B.25 Continuous time series IRI data as percent of the sampled data along I-94 (composite) in Michigan

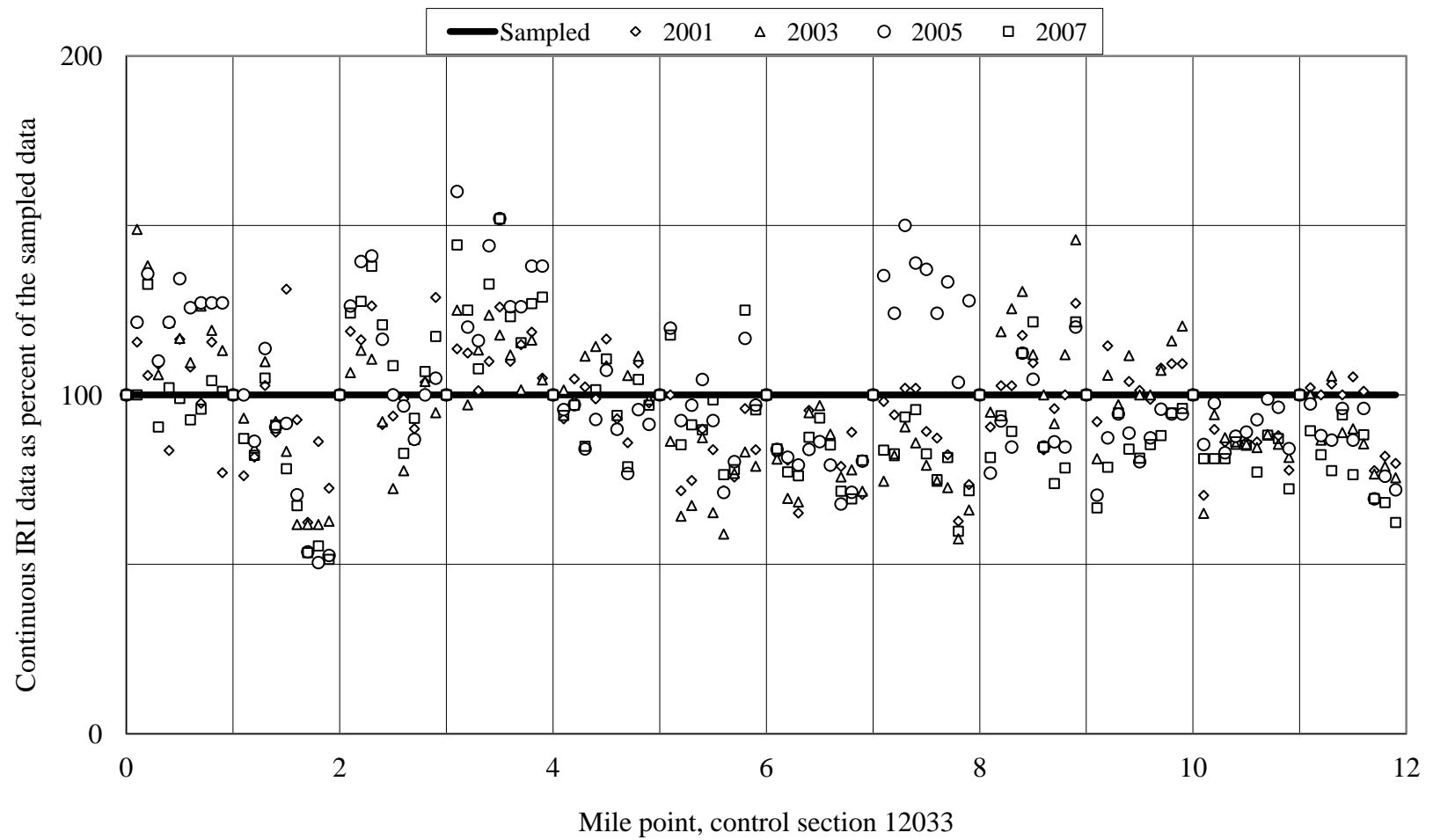


Figure B.26 Continuous time series IRI data as percent of the sampled data along I-69 (CS 12033, rigid) in Michigan

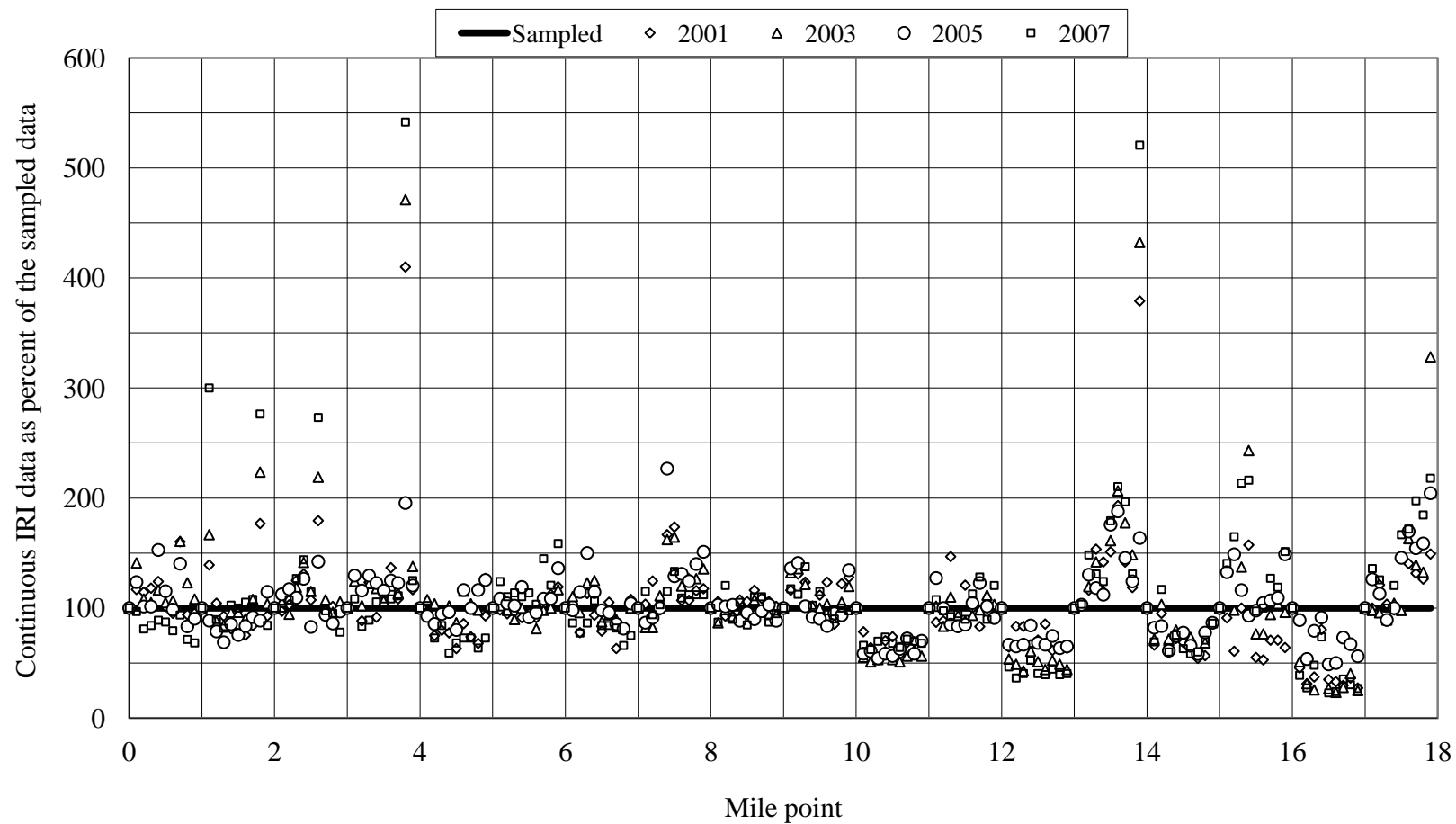


Figure B.27 Continuous time series IRI data as percent of the sampled data along US-31 (flexible) in Michigan

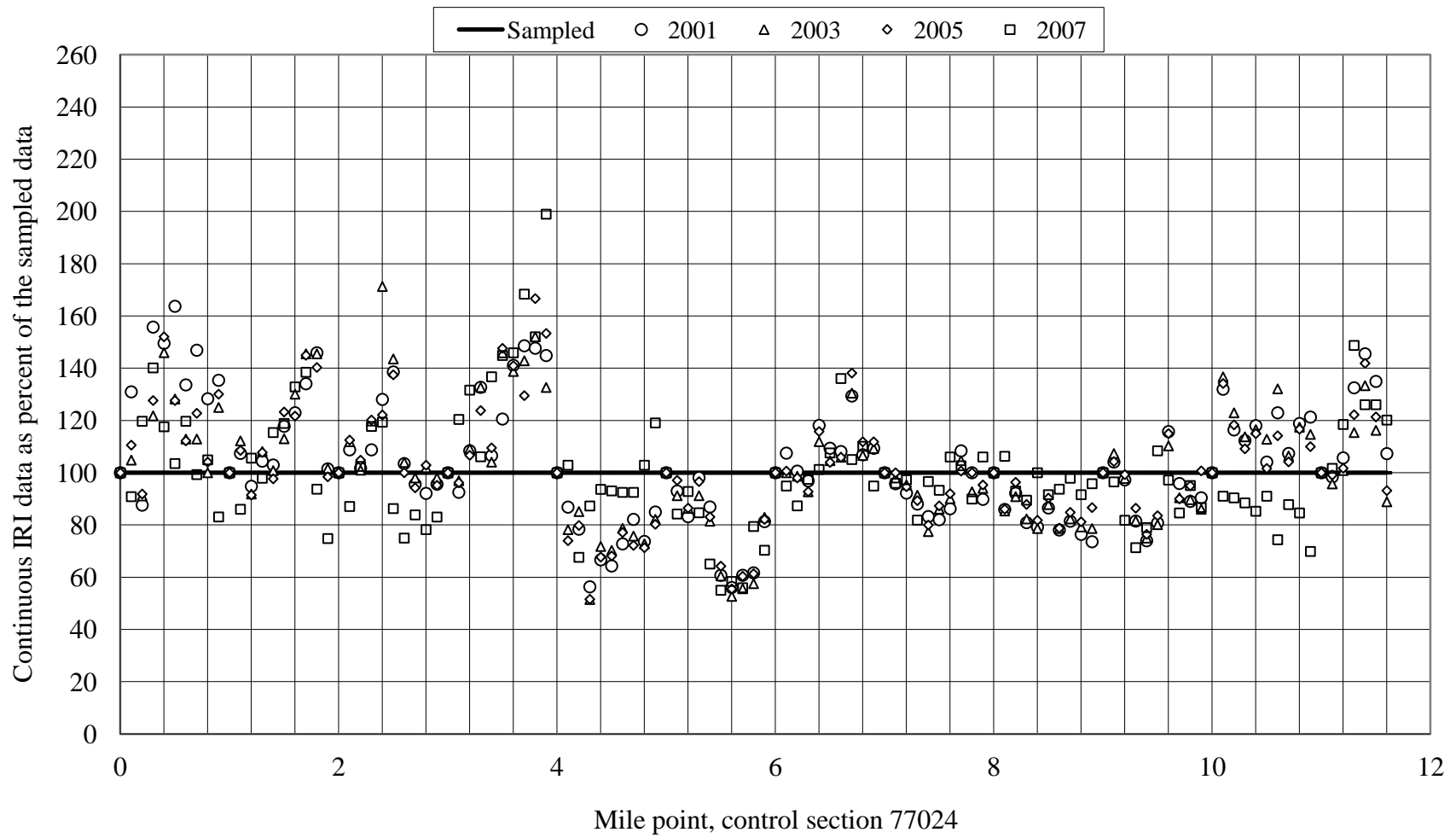


Figure B.28 Continuous time series IRI data as percent of the sampled data along I-69 (CS 77024, rigid) in Michigan

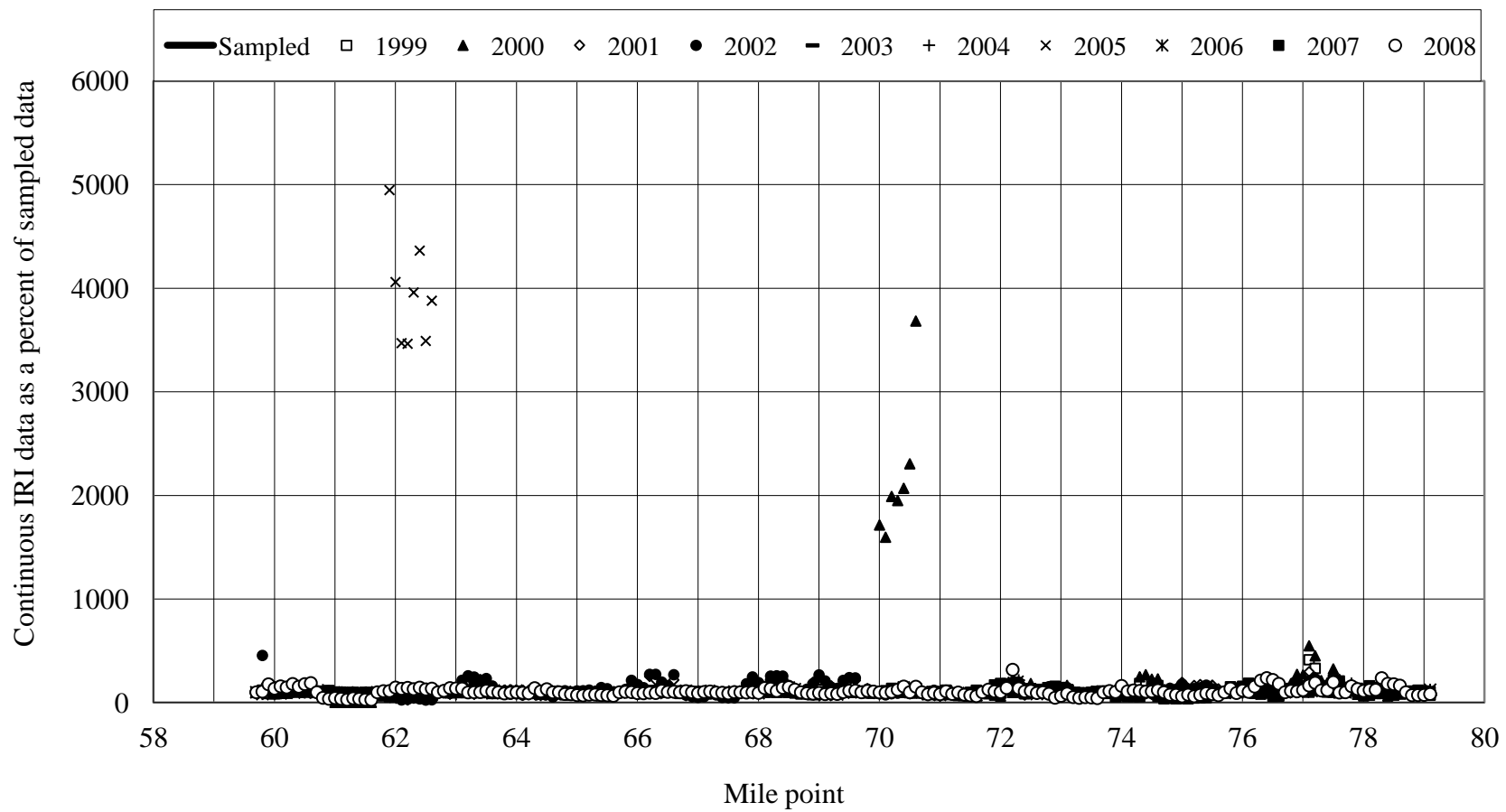


Figure B.29 Continuous time series IRI data as a percent of the sampled data along SRID 005 (composite) in Washington

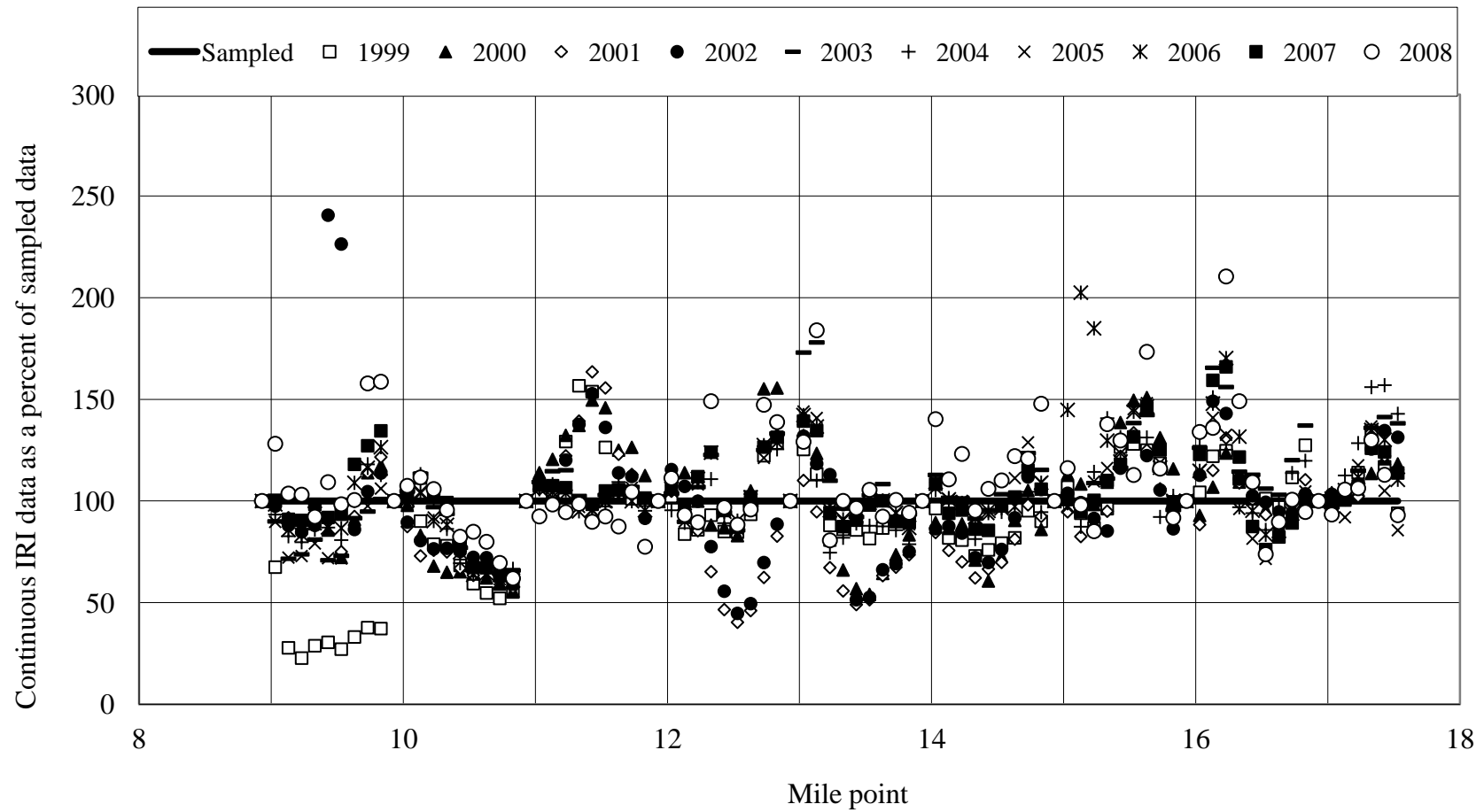


Figure B.30 Continuous time series IRI data as a percent of the sampled data along SRID 161 (composite) in Washington

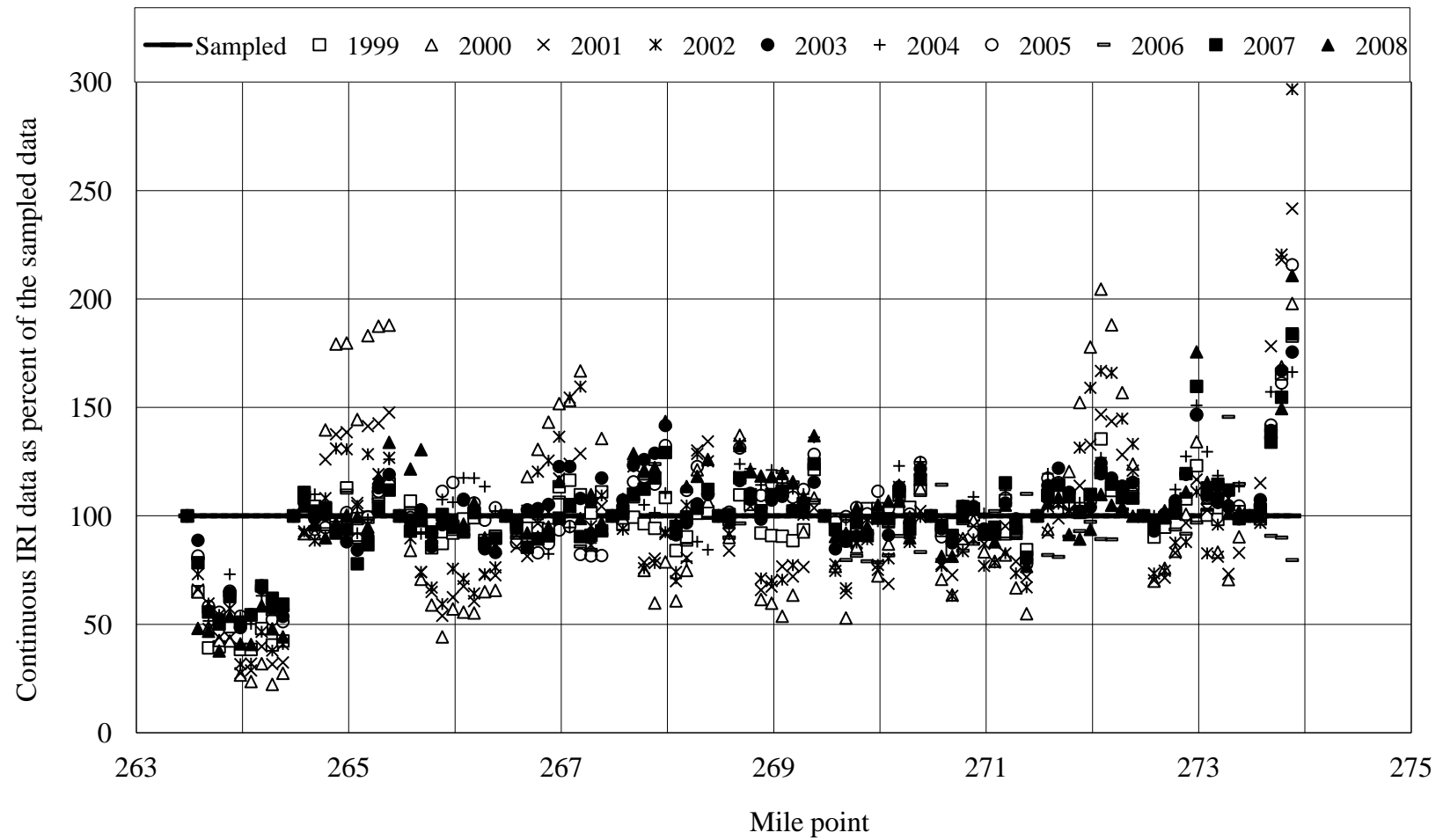


Figure B.31 Continuous time series IRI data as a percent of the sampled data along SRID 005 (flexible) in Washington

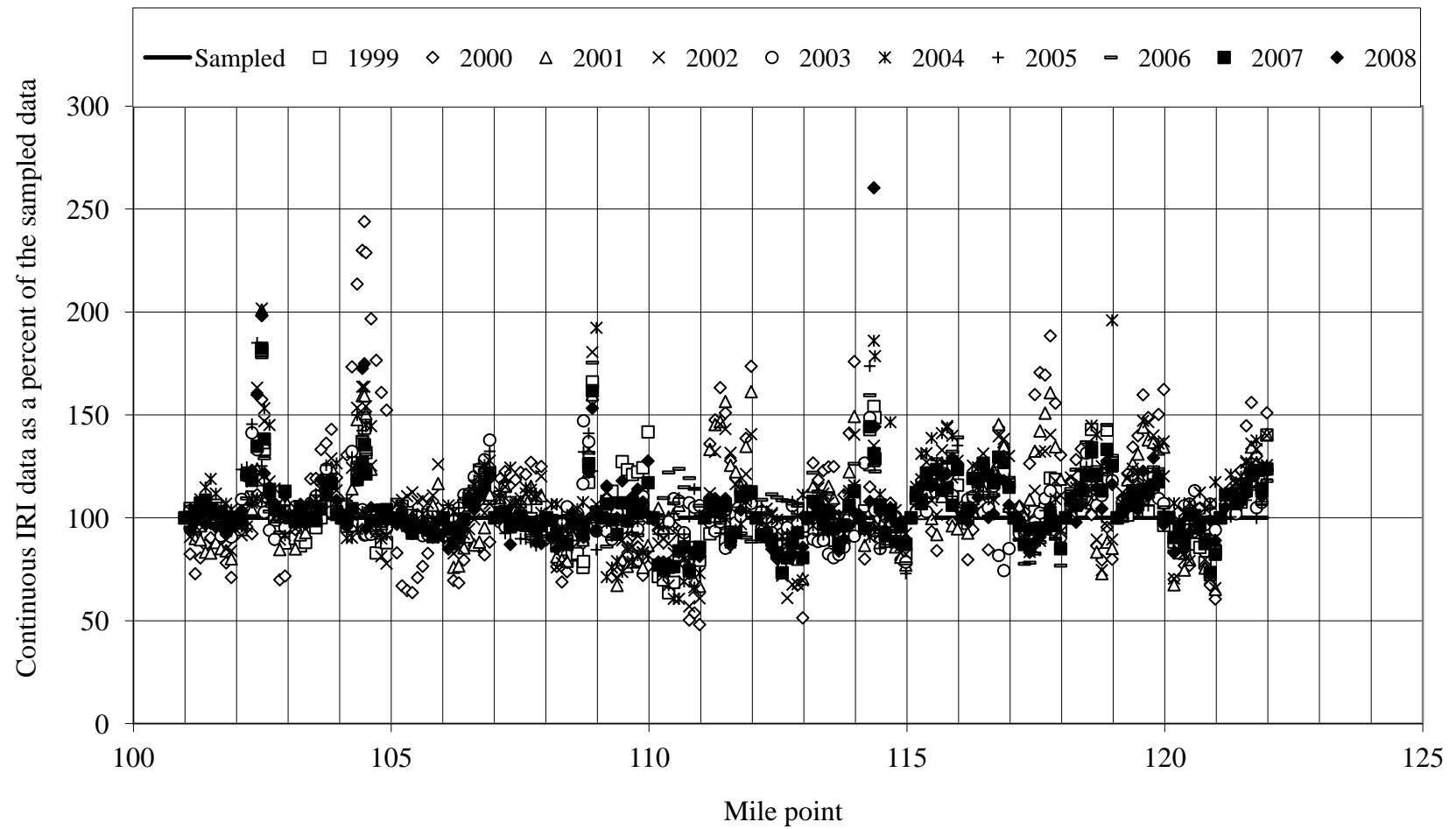


Figure B.32 Continuous time series IRI data as a percent of the sampled data along SRID 082 (rigid) in Washington

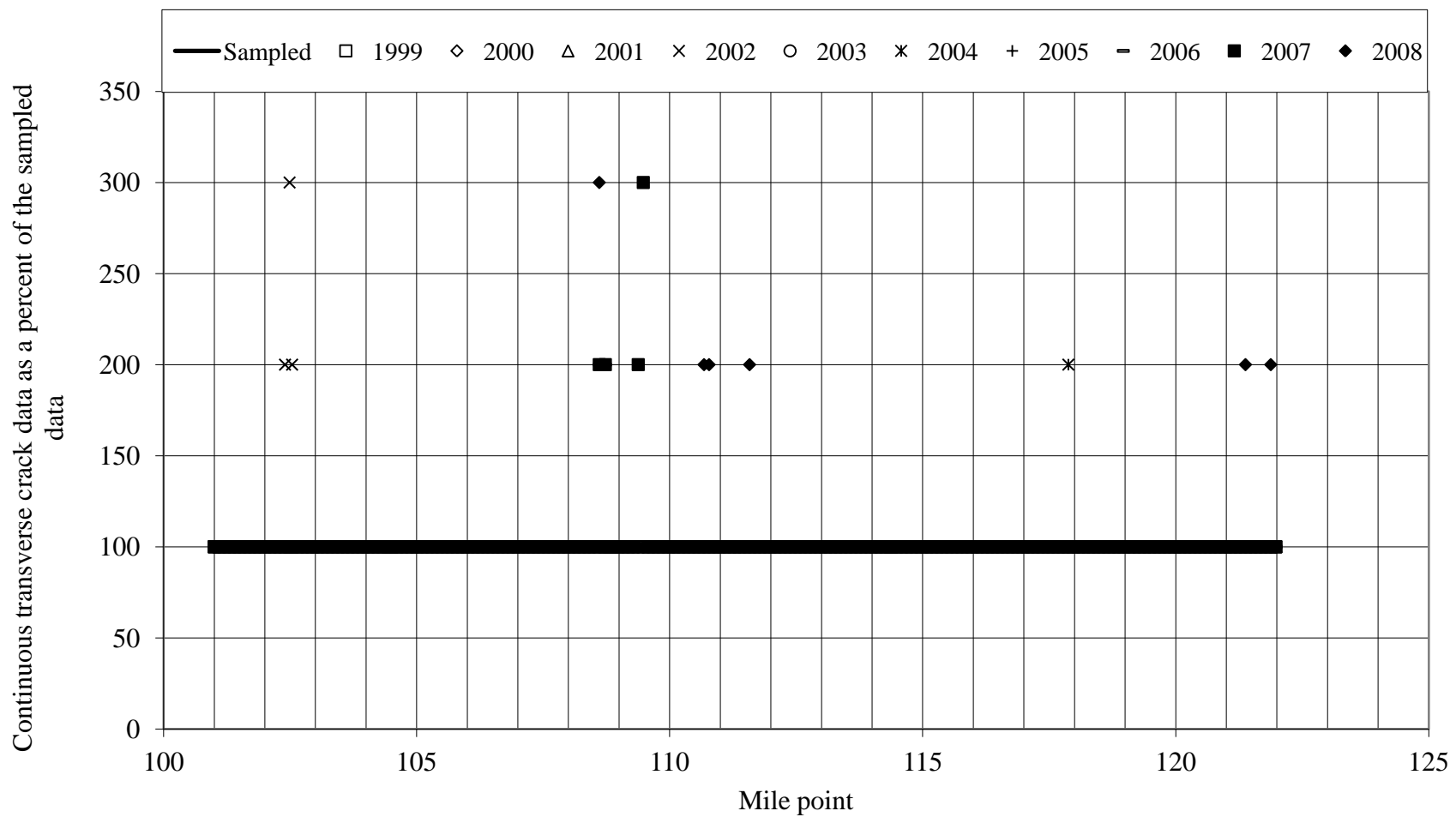


Figure B.33 Continuous time series transverse crack data as percent of the sampled data along HWY 24 (flexible) in Colorado

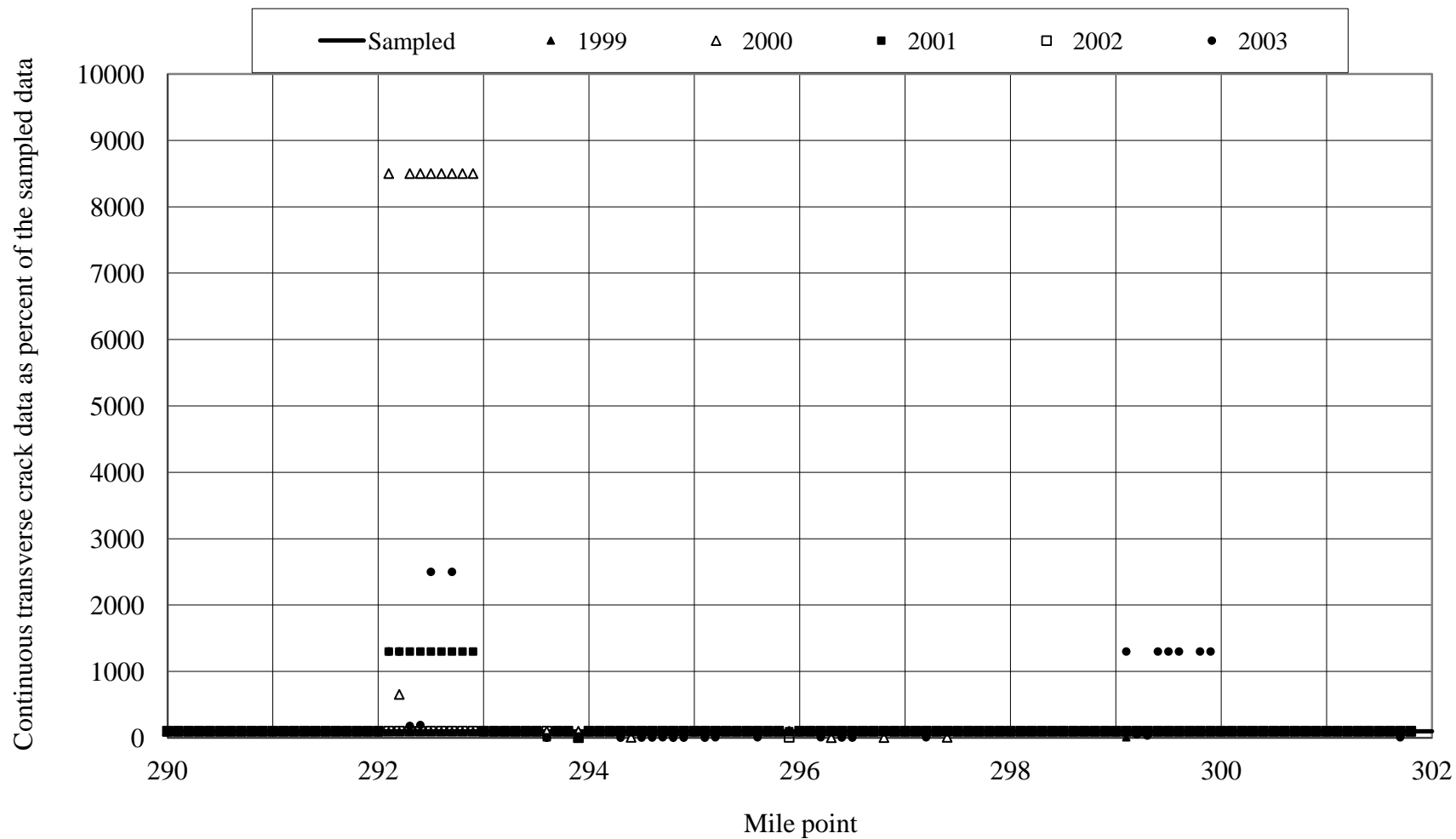


Figure B.34 Continuous time series transverse crack data as percent of the sampled data along I-70 (rigid) in Colorado

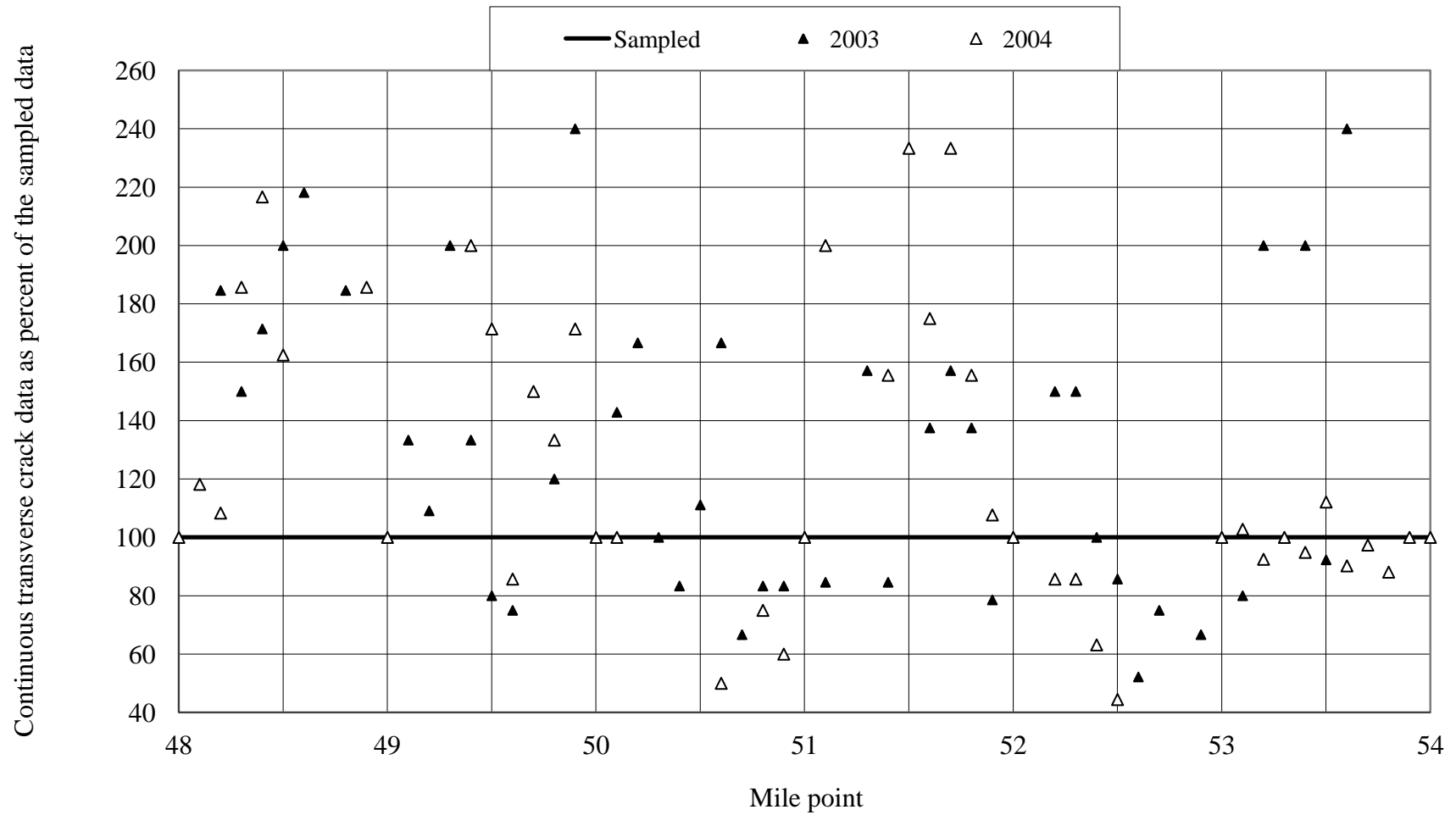


Figure B.35 Continuous time series transverse crack data as percent of the sampled data along HWY 36 (composite) in Colorado

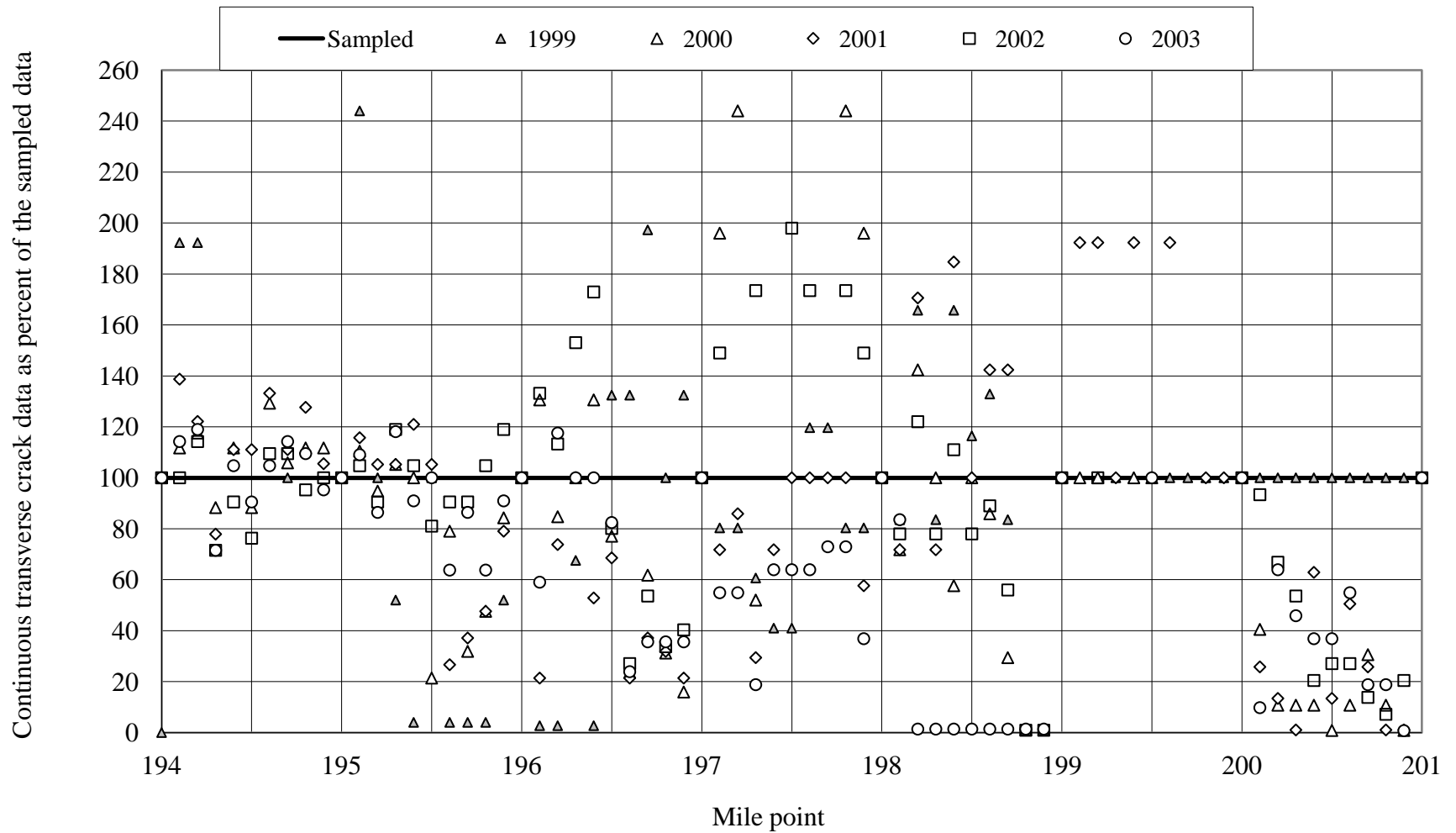


Figure B.36 Continuous time series transverse crack data as percent of the sampled data along HWY 71 (flexible) in Colorado

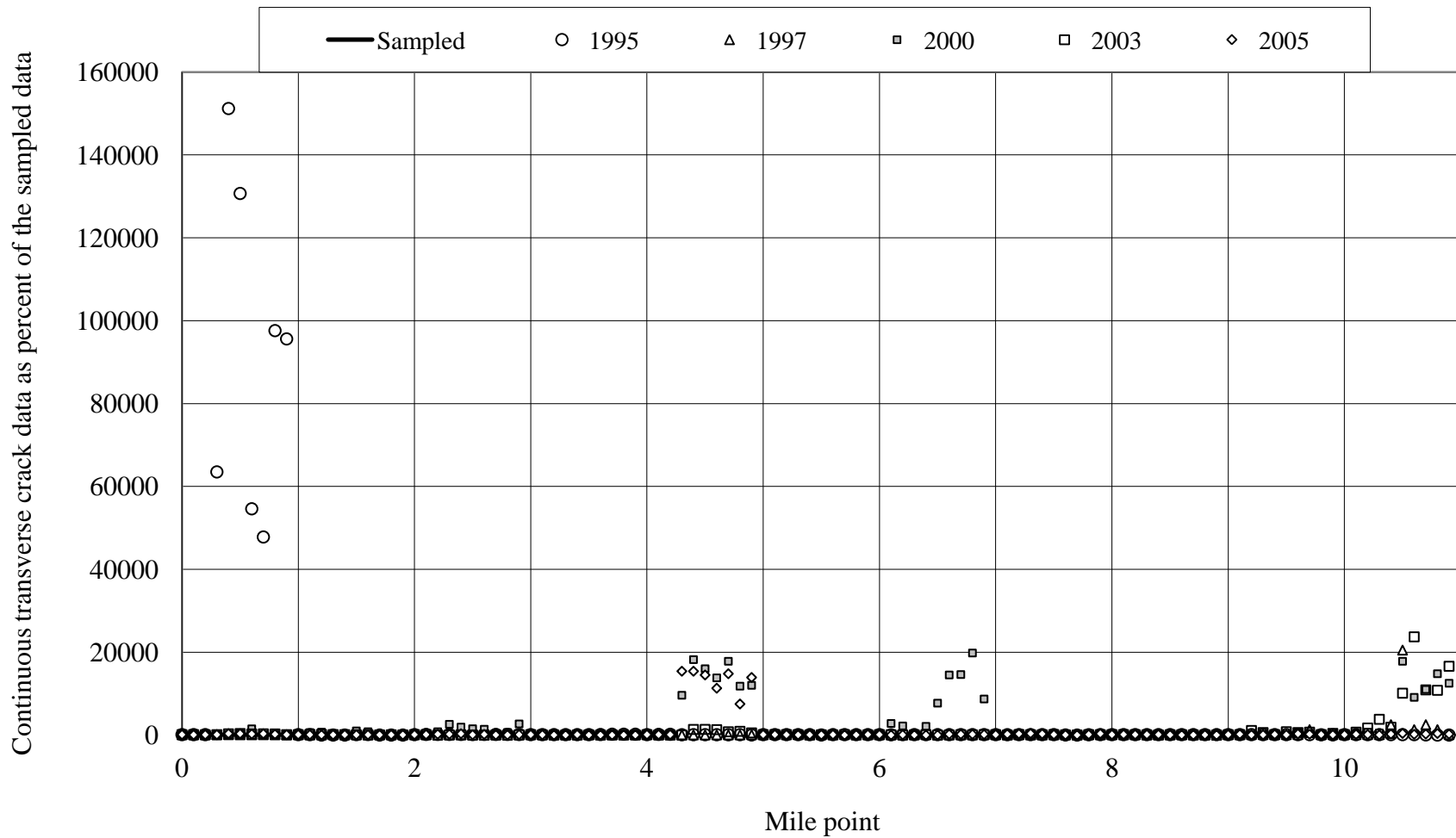


Figure B.37 Continuous time series transverse crack data as percent of the sampled data along LA-34 (flexible) in Louisiana

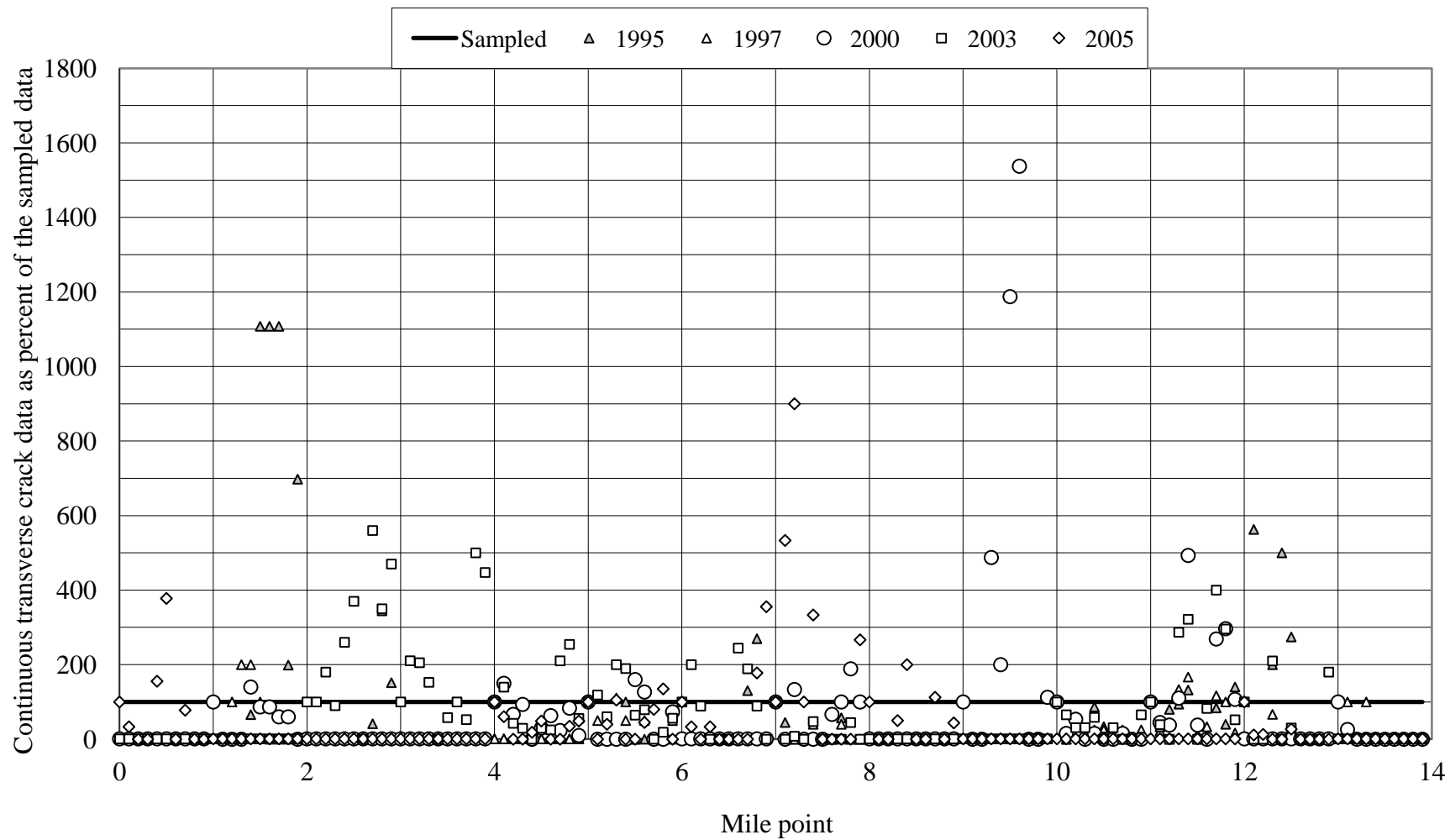


Figure B.38 Continuous time series transverse crack data as percent of the sampled data along LA-83 (rigid) in Louisiana

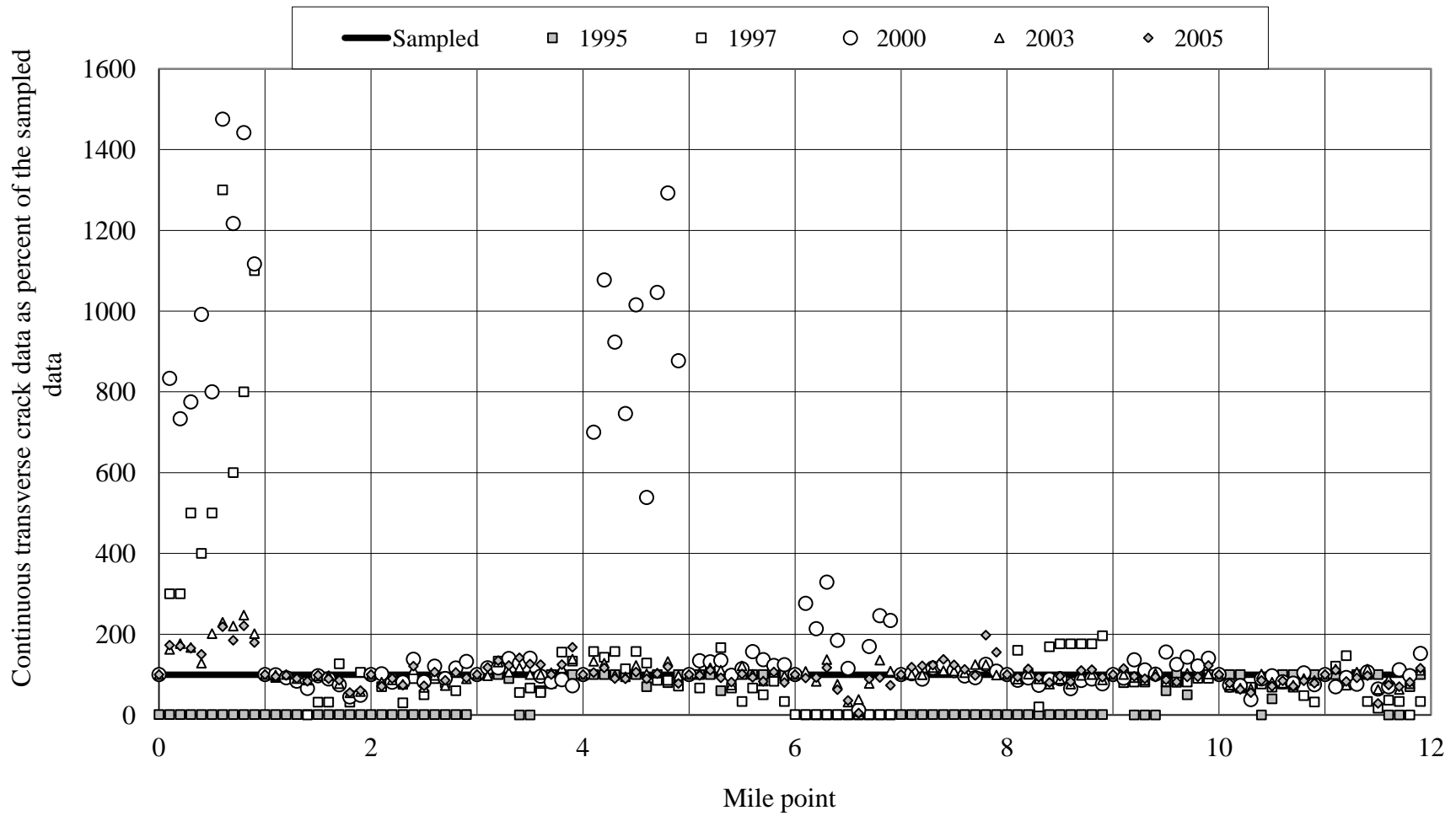


Figure B.39 Continuous time series transverse crack data as percent of the sampled data along LA-1 (composite) in Louisiana

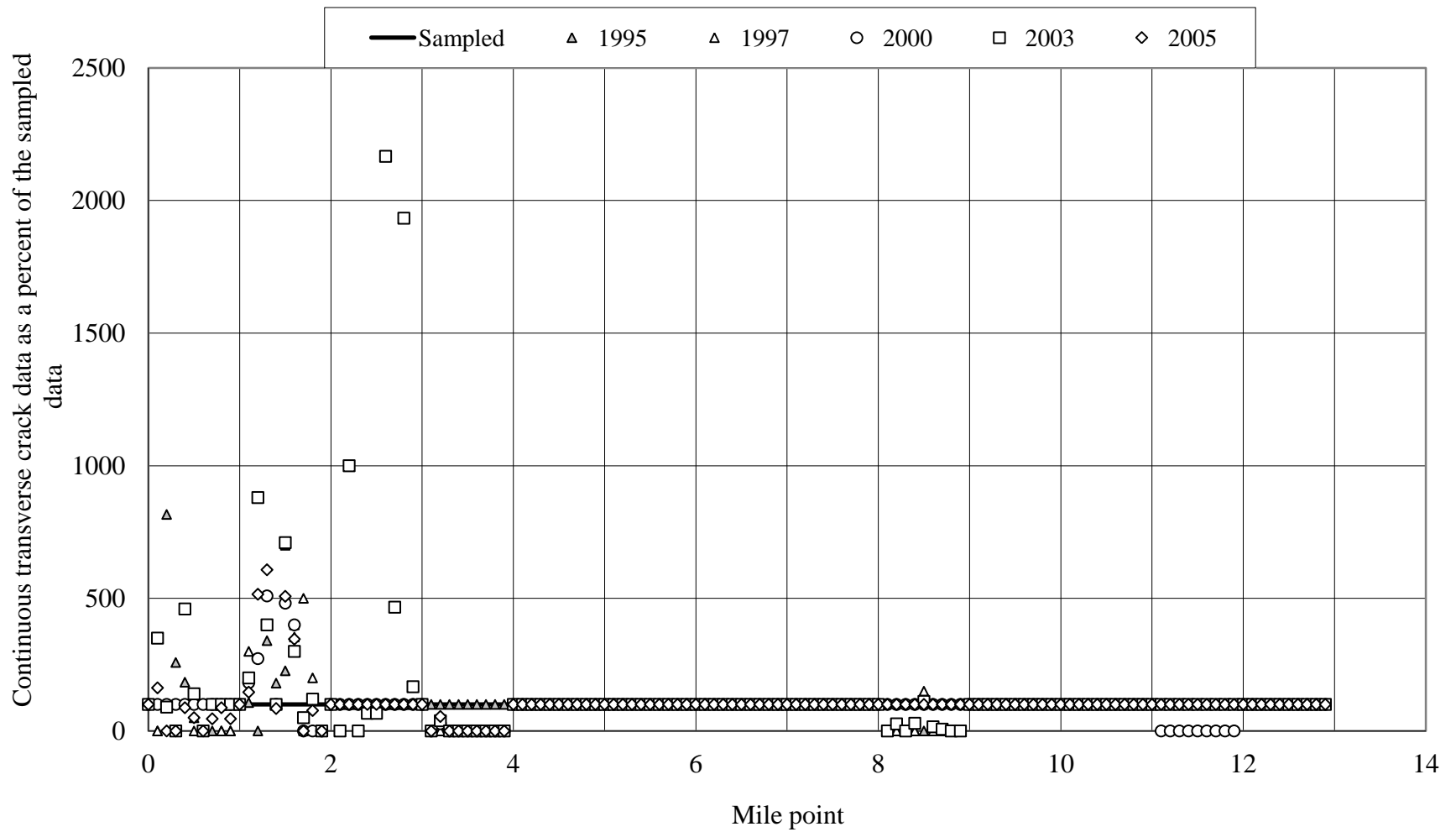


Figure B.40 Continuous time series transverse crack data as percent of the sampled data along LA-526 (rigid) in Louisiana

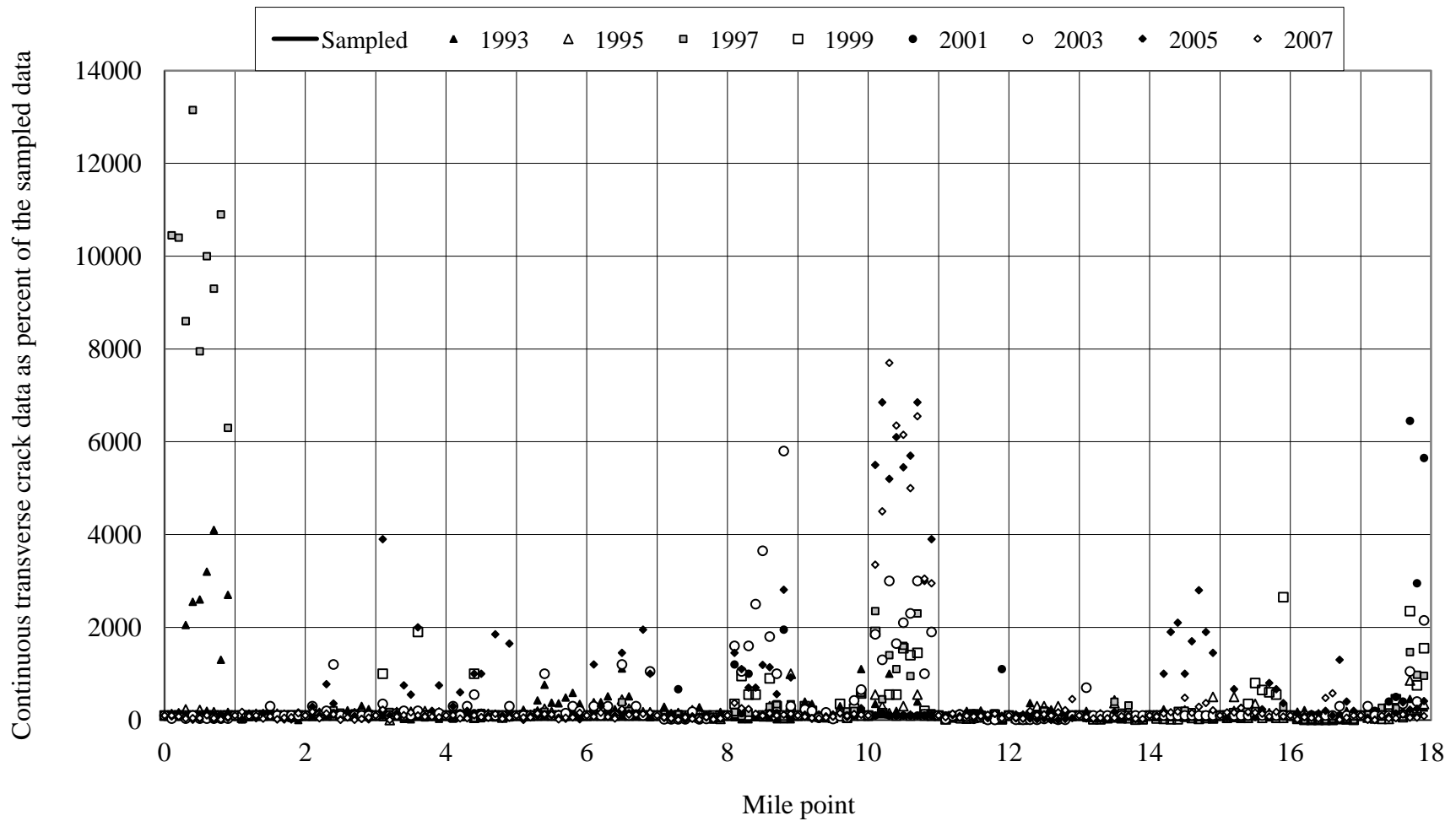


Figure B.41 Continuous time series transverse crack data as percent of the sampled data along I-94 (composite) in Michigan

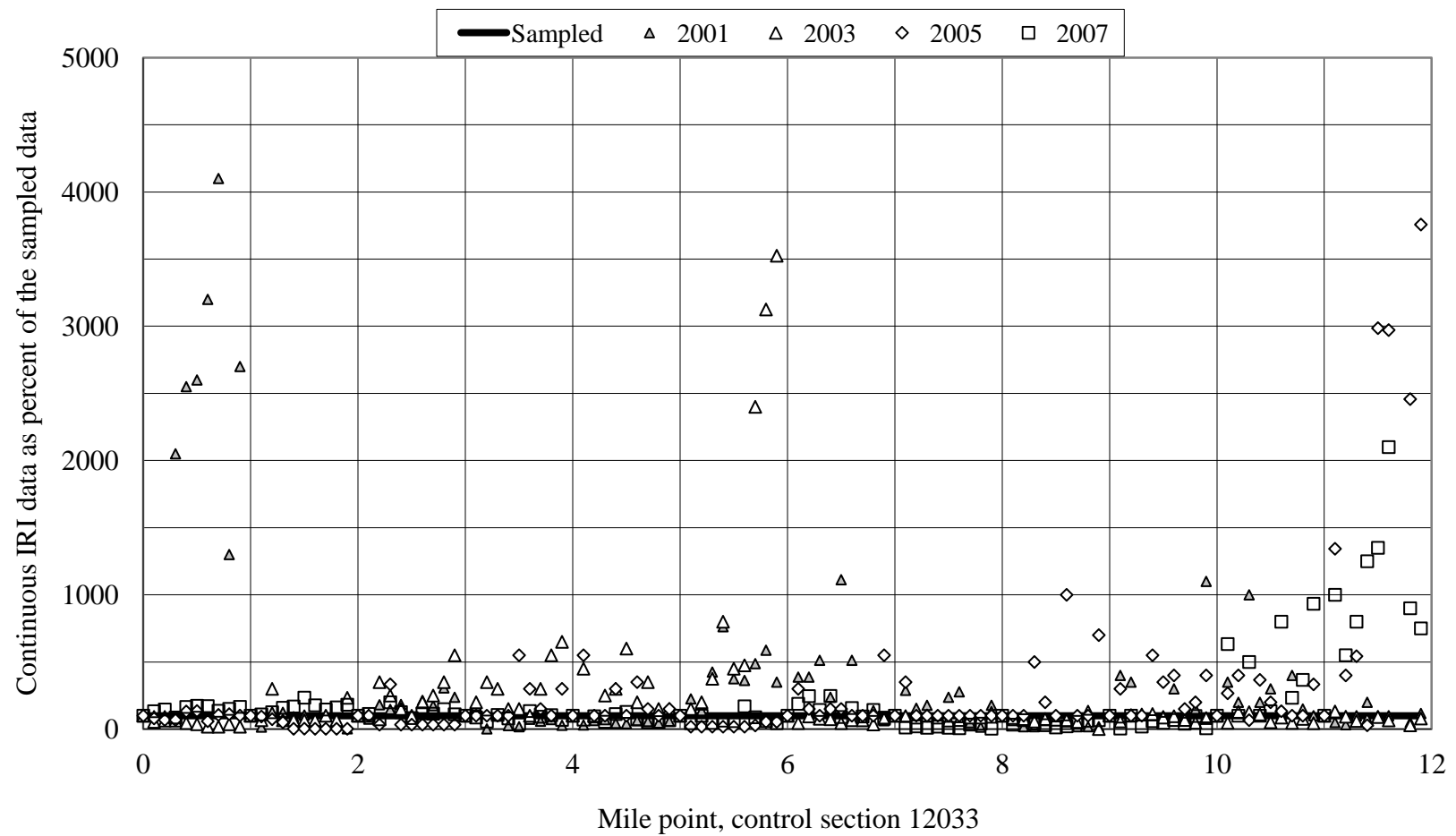


Figure B.42 Continuous time series transverse crack data as percent of the sampled data long I-69 (CS 12033, rigid) in Michigan

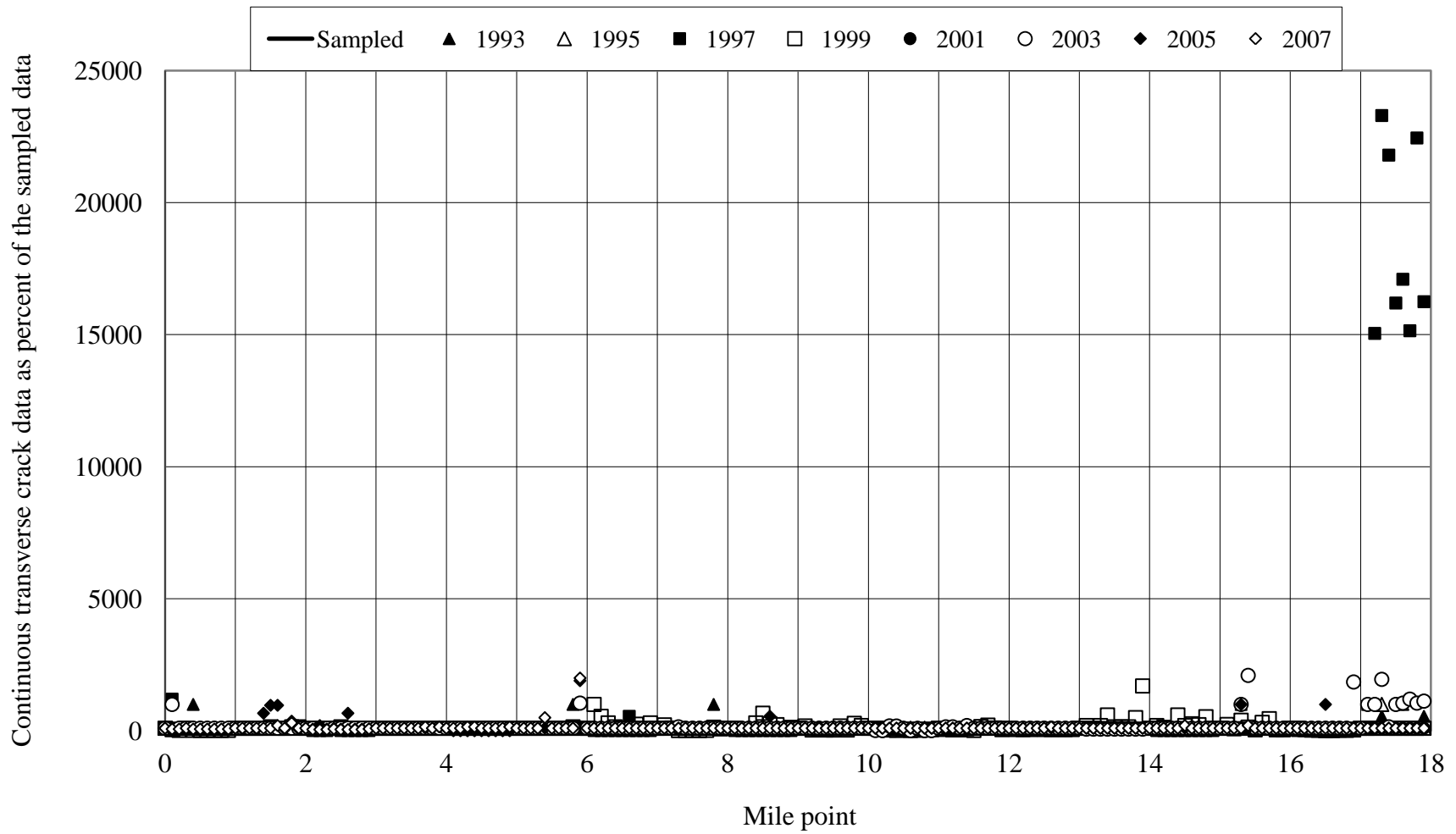


Figure B.43 Continuous time series transverse crack data as percent of the sampled data along US-31 (flexible) in Michigan

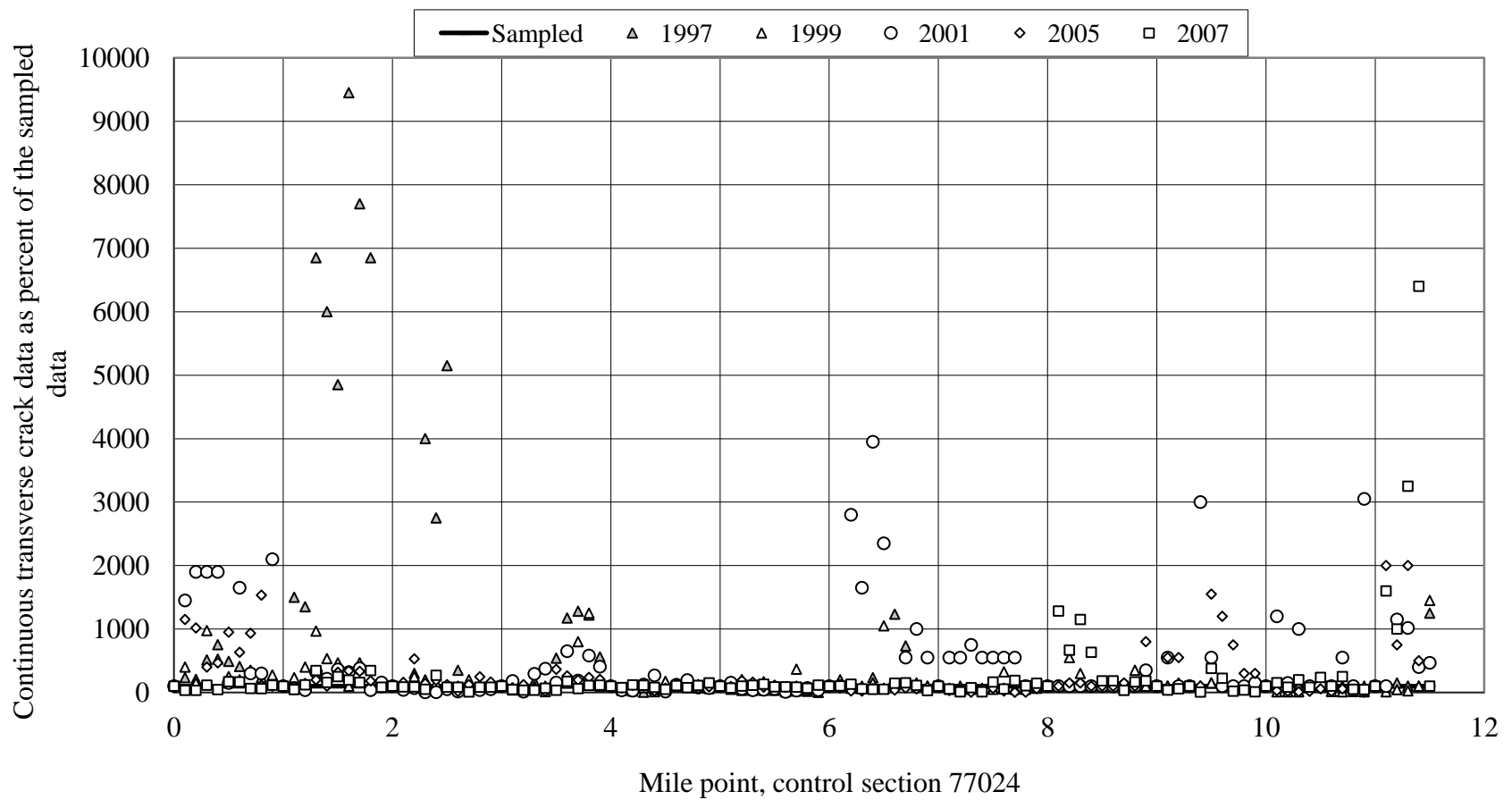


Figure B.44 Continuous time series transverse crack data as percent of the sampled data along I-69 (CS 77024, rigid) in Michigan

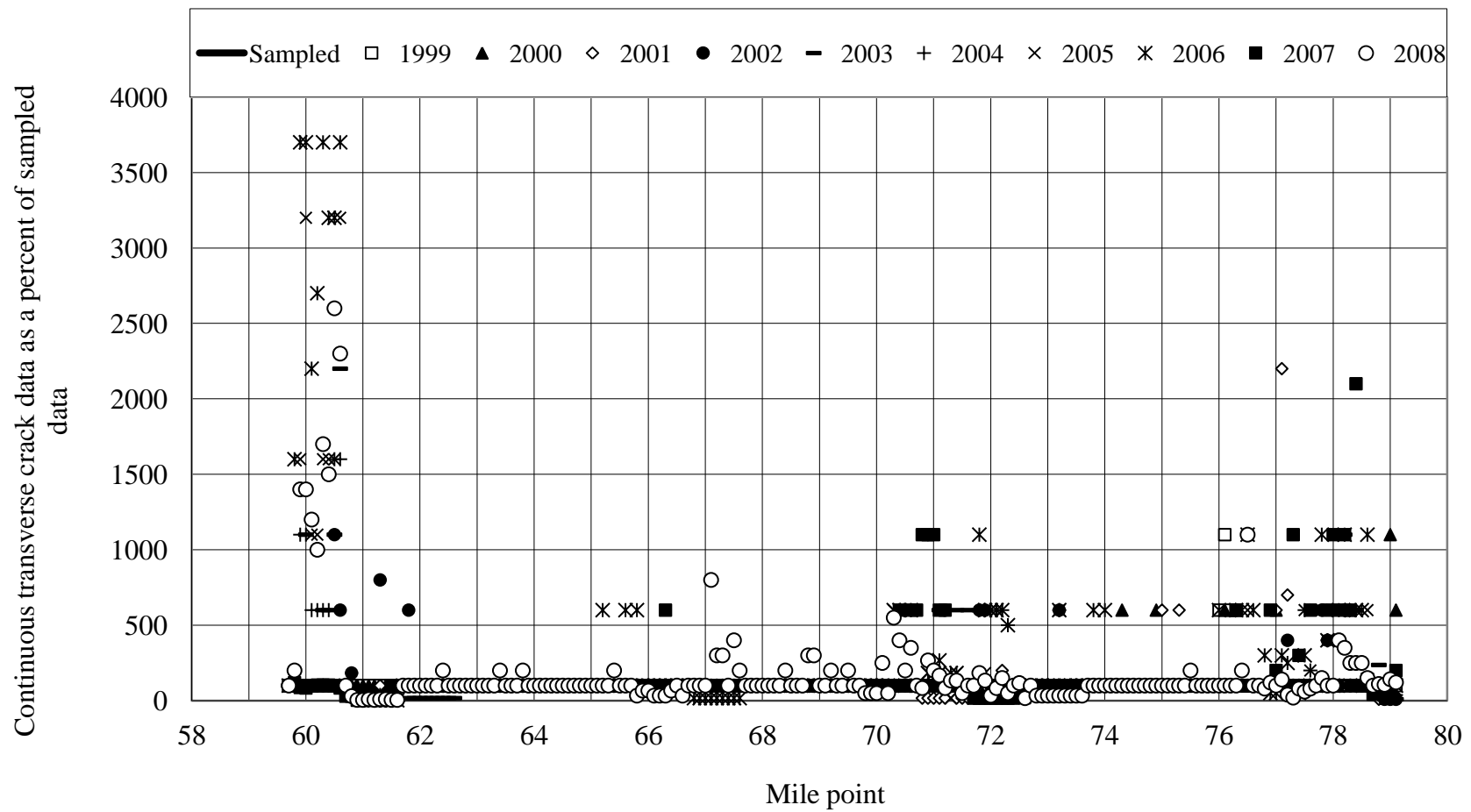


Figure B.45 Continuous time series transverse crack data as percent of the sampled data along SRID 005 (composite) in Washington

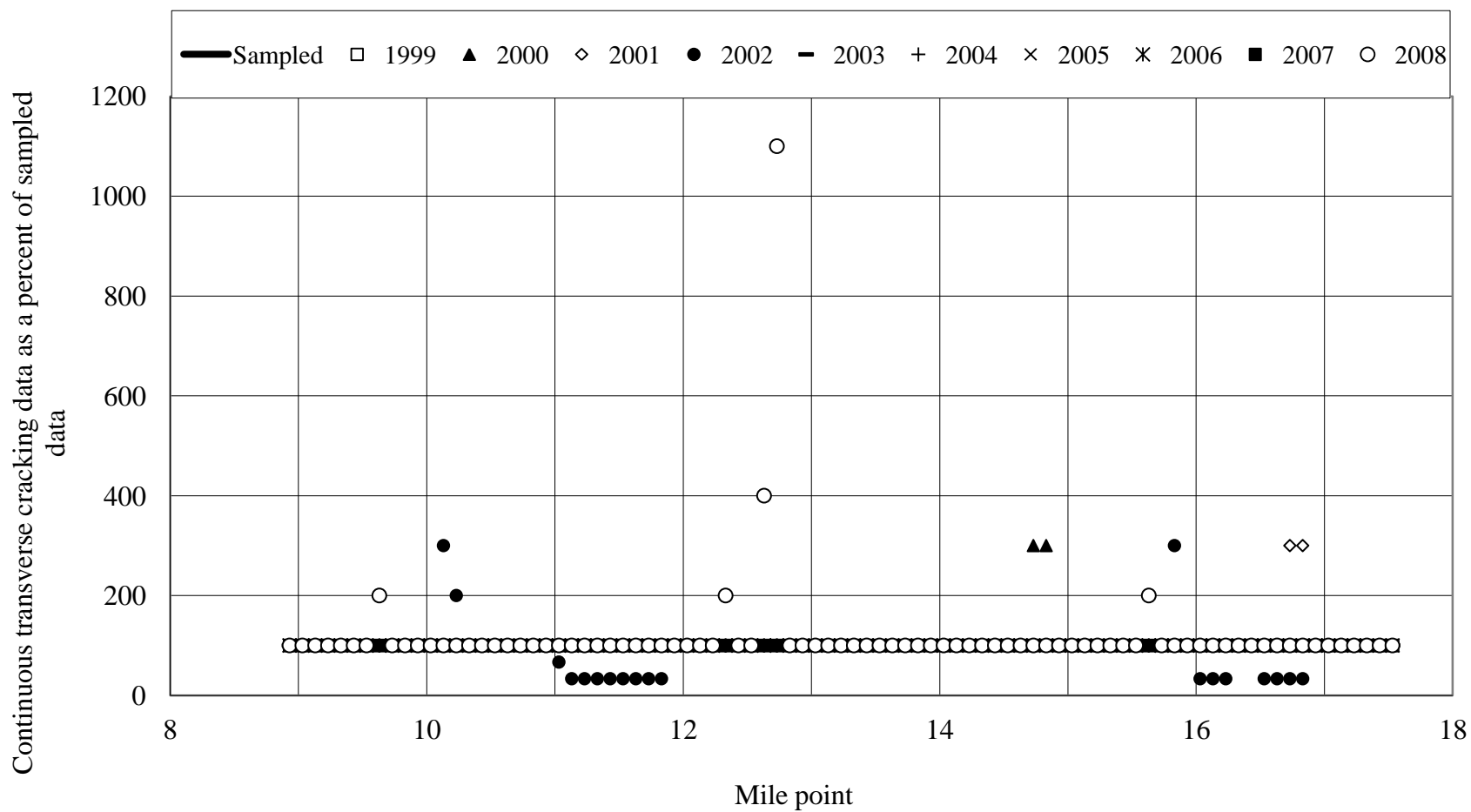


Figure B.46 Continuous time series transverse crack data as percent of the sampled data along SRID 161 (composite) in Washington

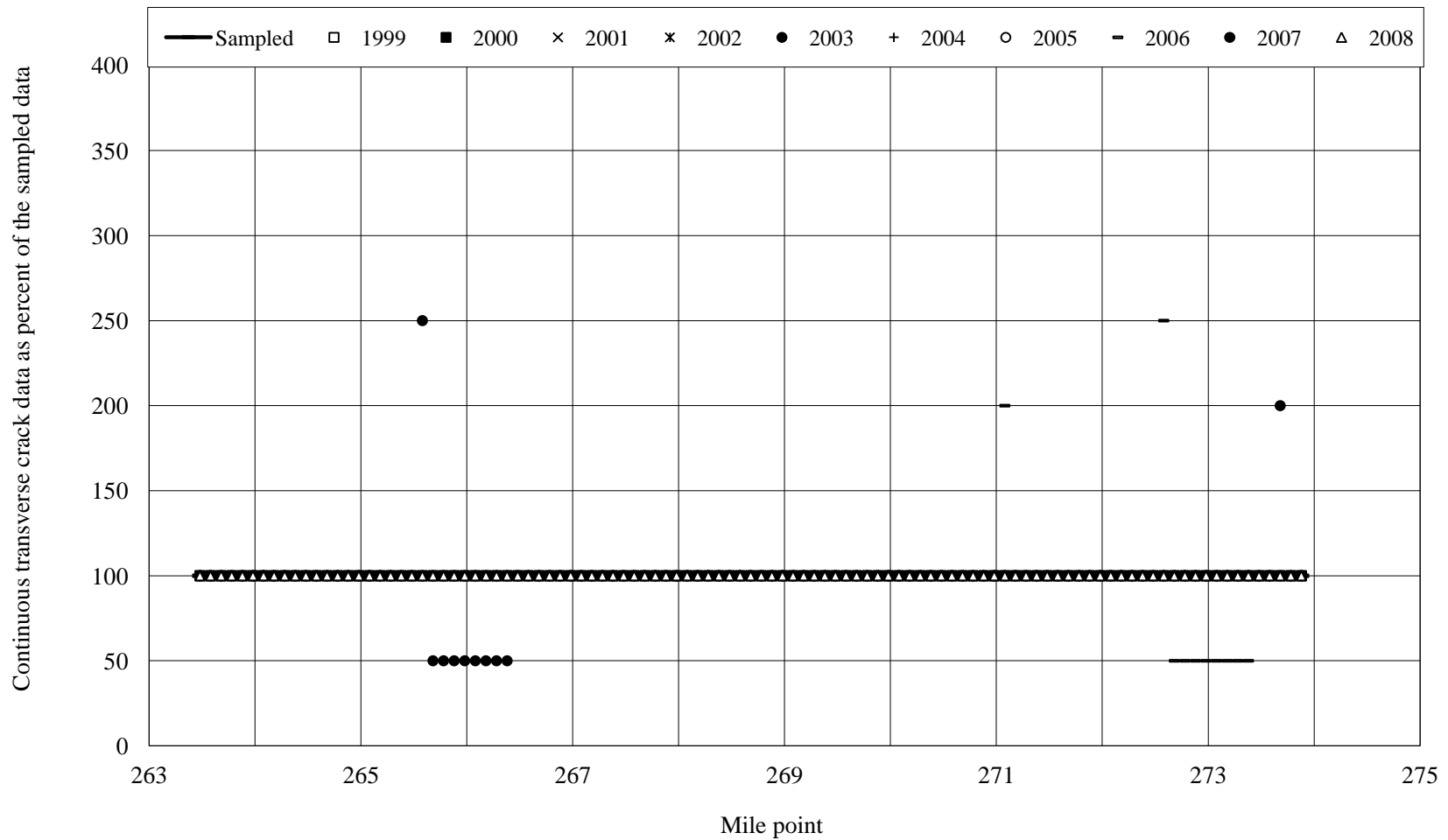


Figure B.47 Continuous time series transverse crack data as percent of the sampled data along SRID 005 (flexible) in Washington

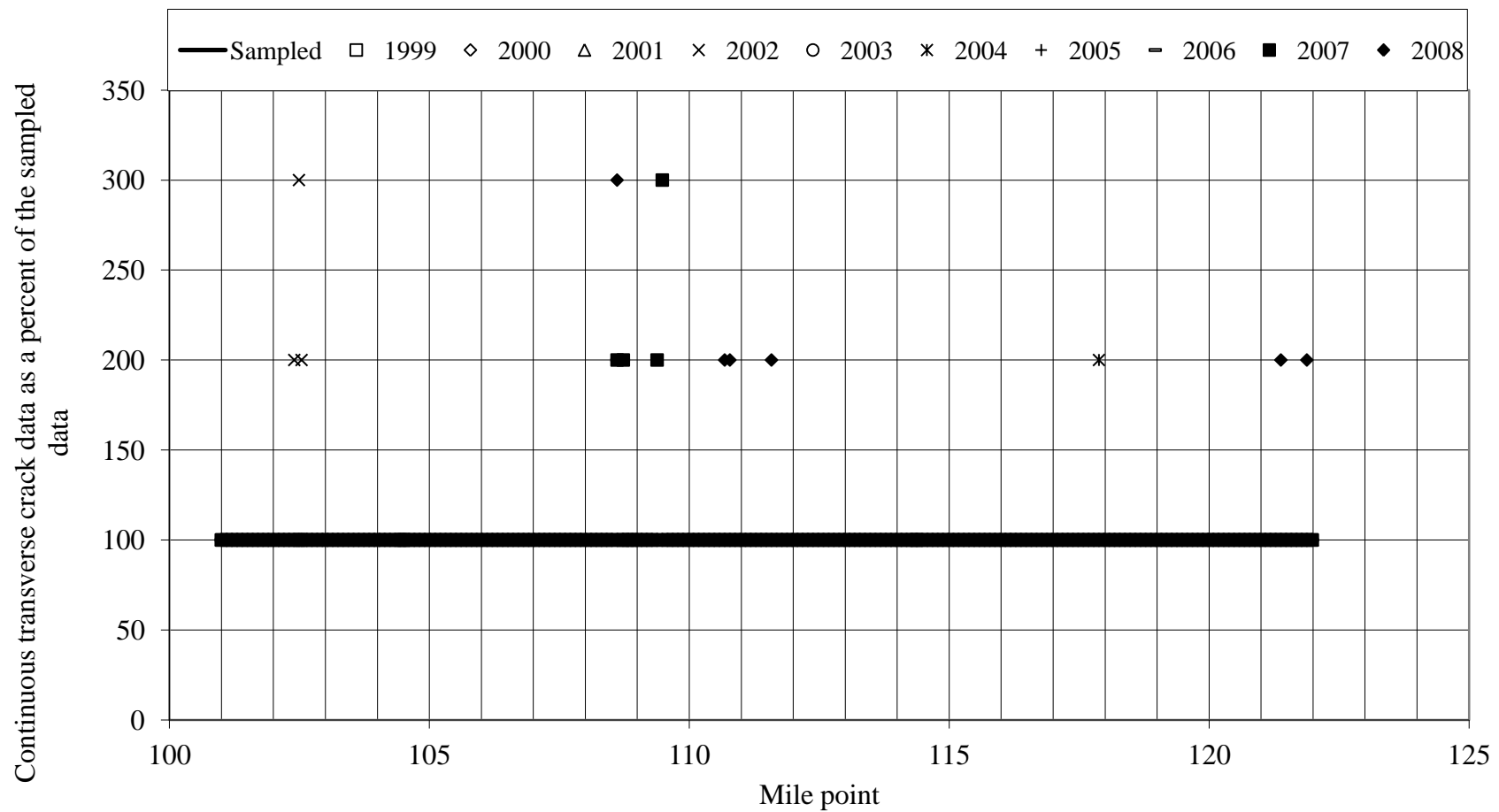
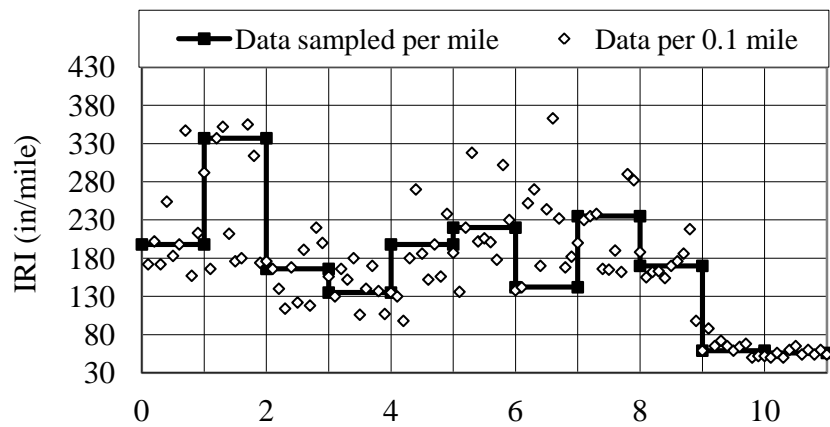
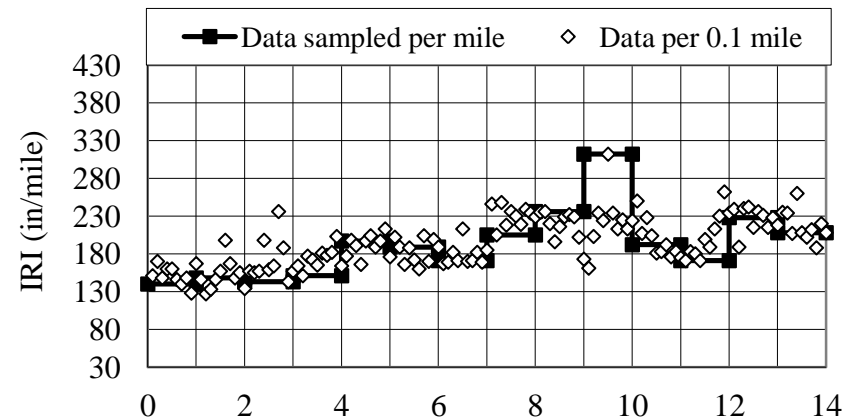


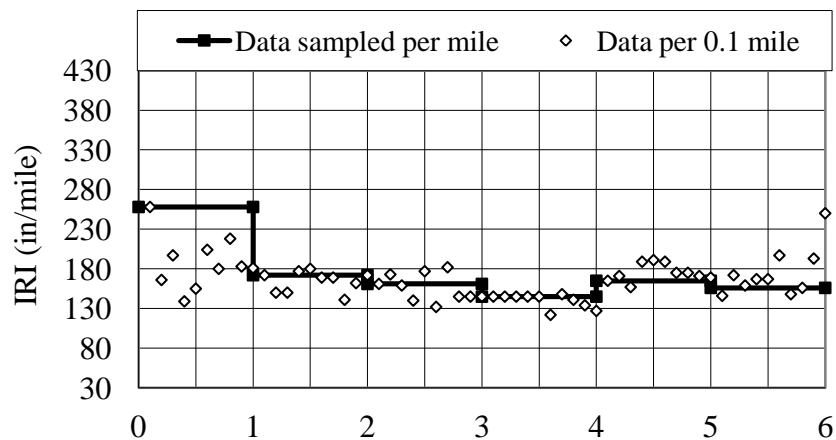
Figure B.48 Continuous time series transverse crack data as percent of the sampled data along SRID 082 (rigid) in Washington



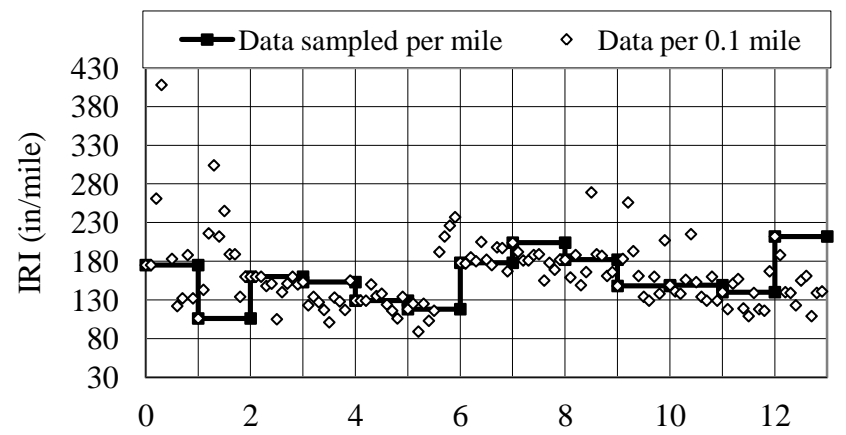
a) Mile point along LA-34 (flexible)
Year 1999



b) Mile point along LA-83 (rigid)
Year 1995



c) Mile point along LA-1 (composite)
Year 1995



d) Mile point along LA-526 (rigid)
Year 1999

Figure B.49 Continuous and randomly sampled IRI data per mile along four roads in Louisiana

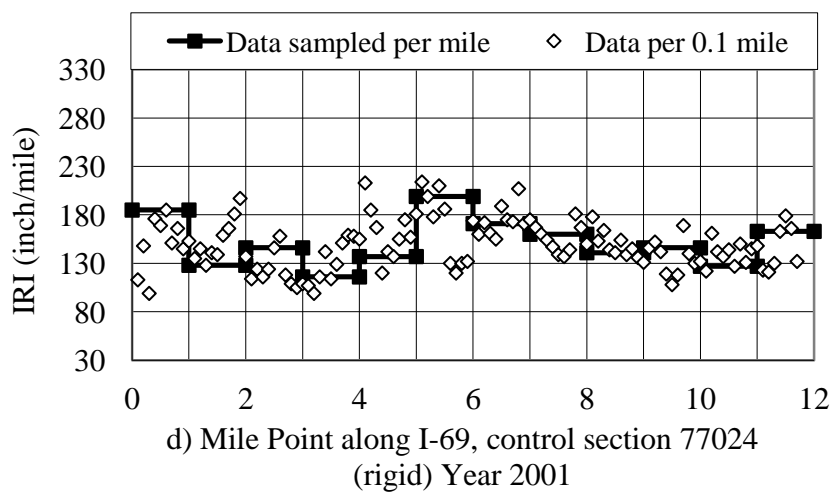
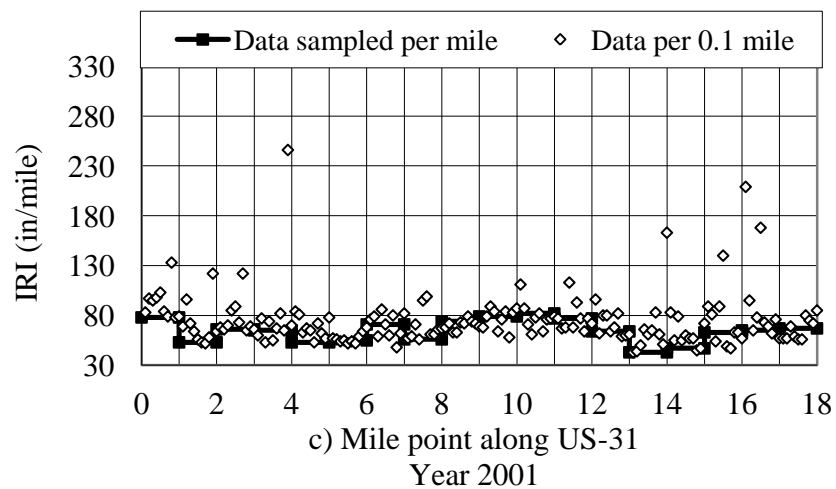
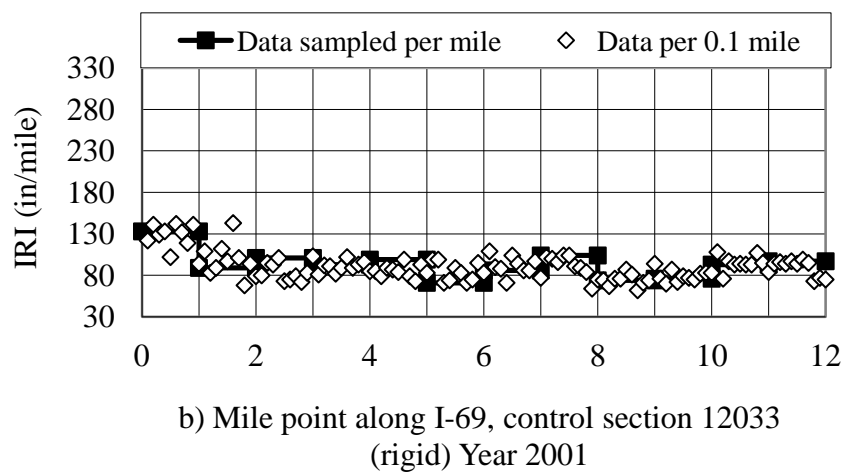
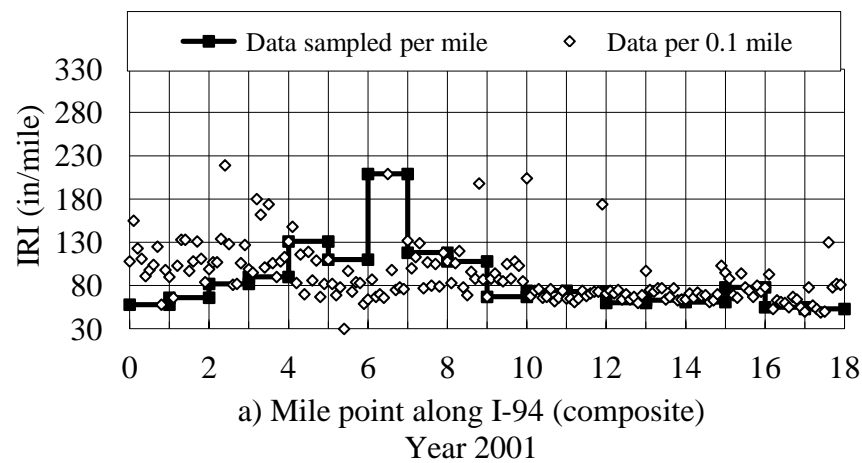
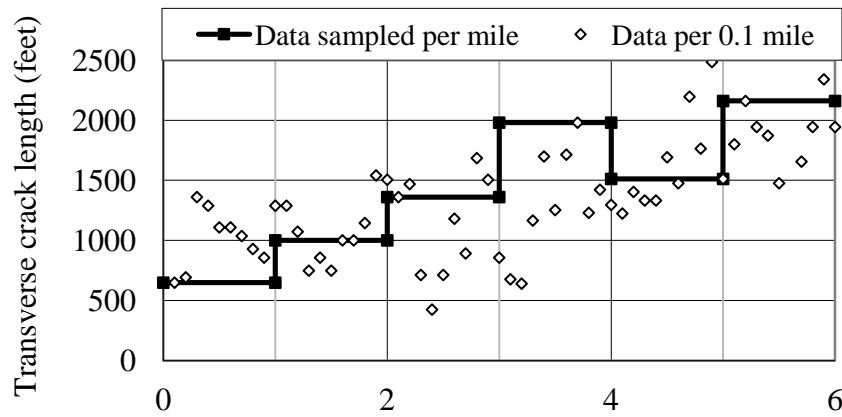
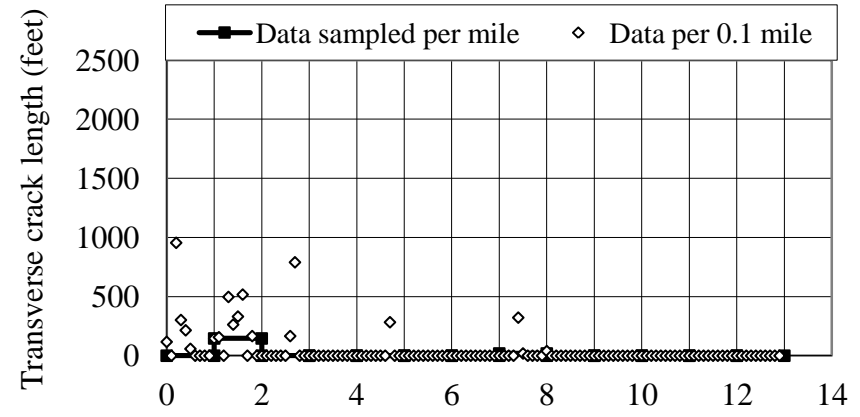


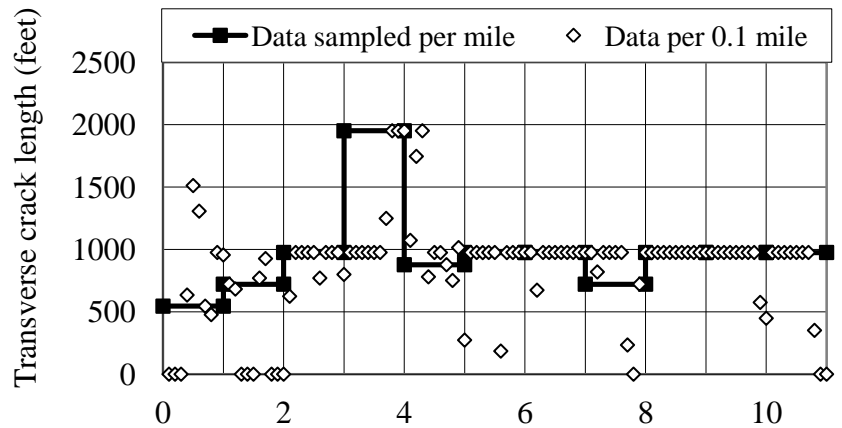
Figure B.50 Continuous and randomly sampled IRI data per mile along four roads in Michigan



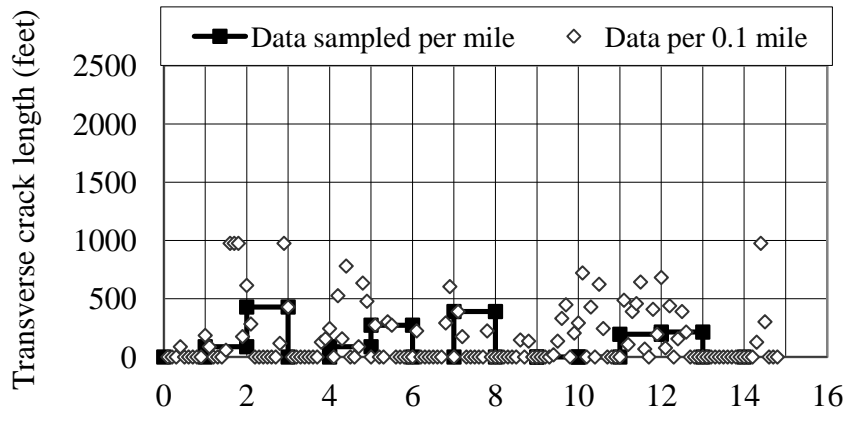
c) Mile point along LA-1 (composite)
Year 1997



d) Mile point along LA-526 (rigid)
Year 1995



a) Mile point along LA-34 (flexible)
Year 1995



b) Mile point along LA-83 (rigid)
Year 1995

Figure B.51 Continuous and randomly sampled transverse crack data along four roads in Louisiana

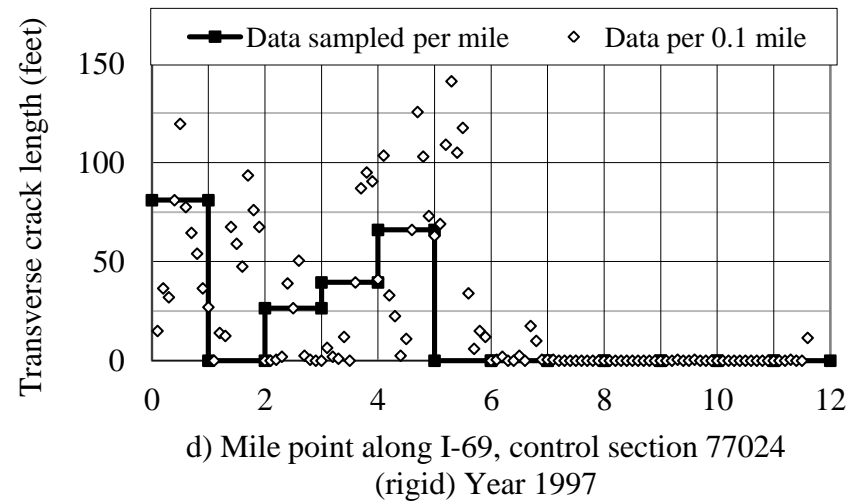
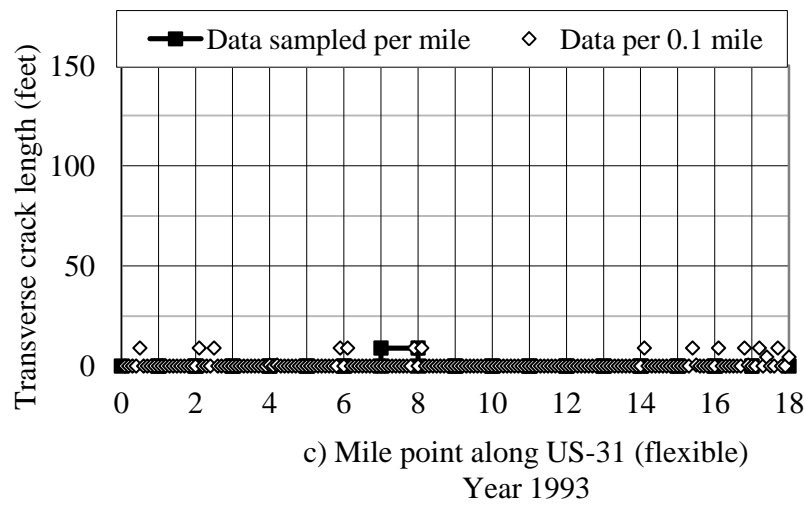
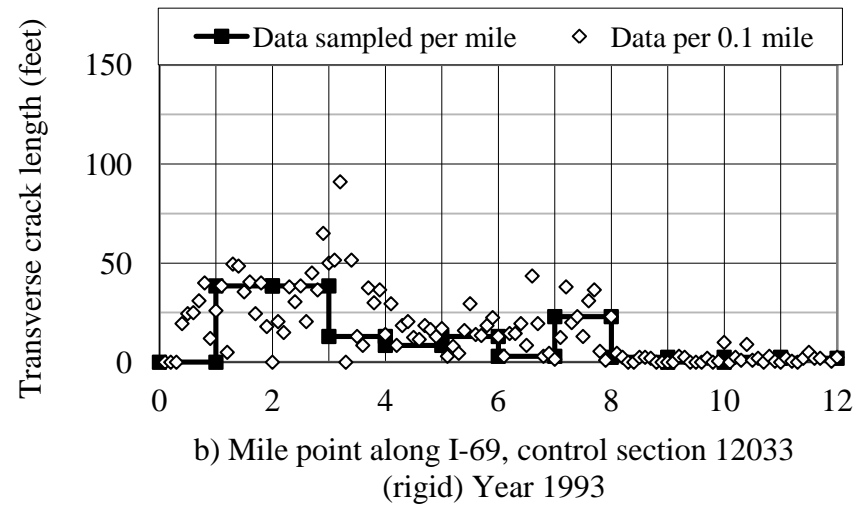
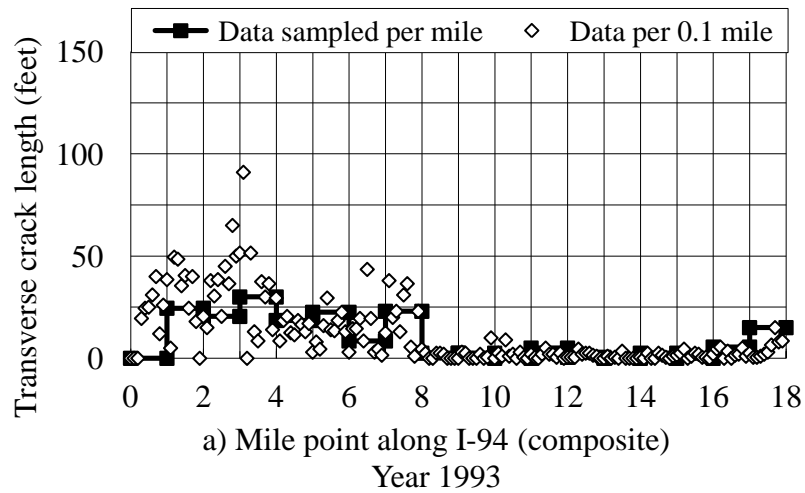


Figure B.52 Continuous and randomly sampled transverse crack data along four roads in Michigan

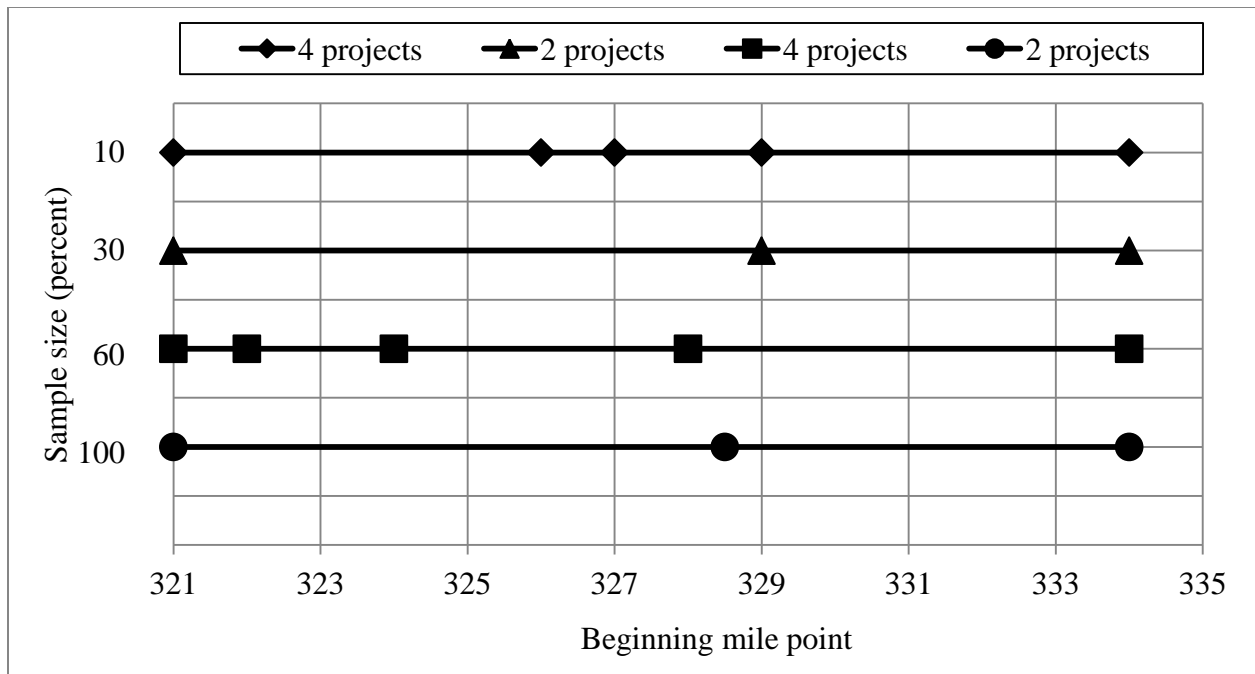


Figure B.53 Project boundaries based on the continuous (100 percent sampled) and the sampled rut depth data along a section of the flexible HWY 24 in Colorado

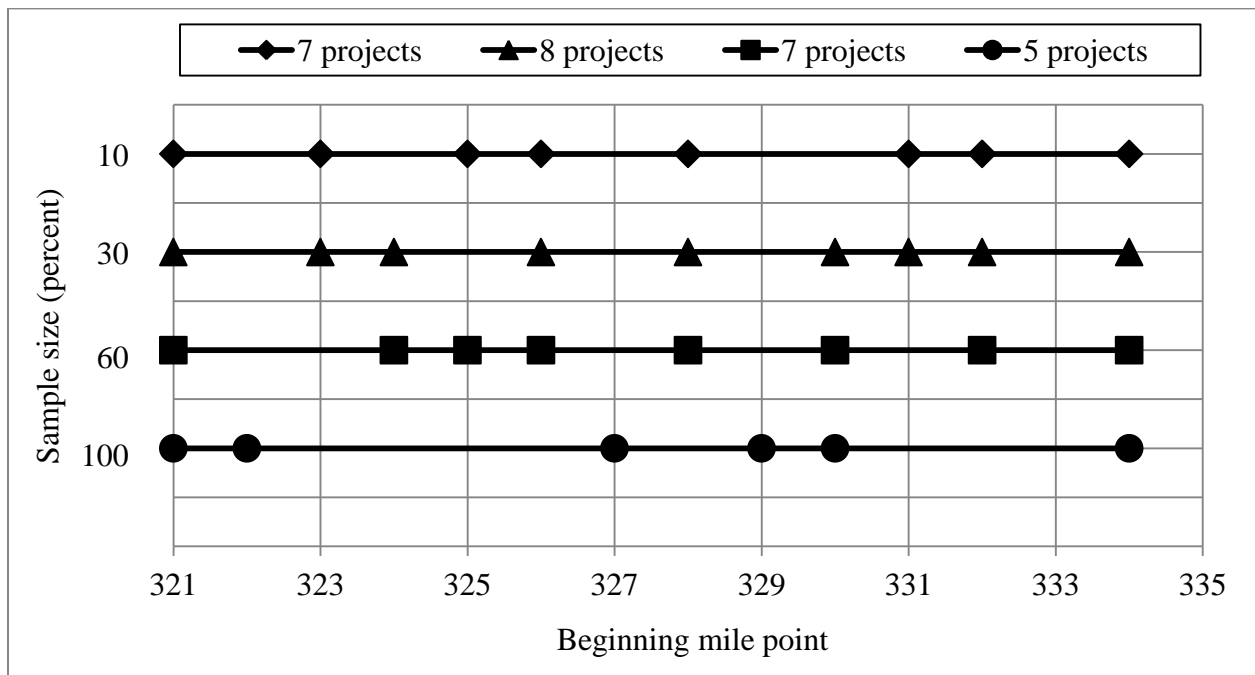


Figure B.54 Project boundaries based on the continuous (100 percent sampled) and the sampled transverse cracking data along a section of the flexible HWY 24 in Colorado

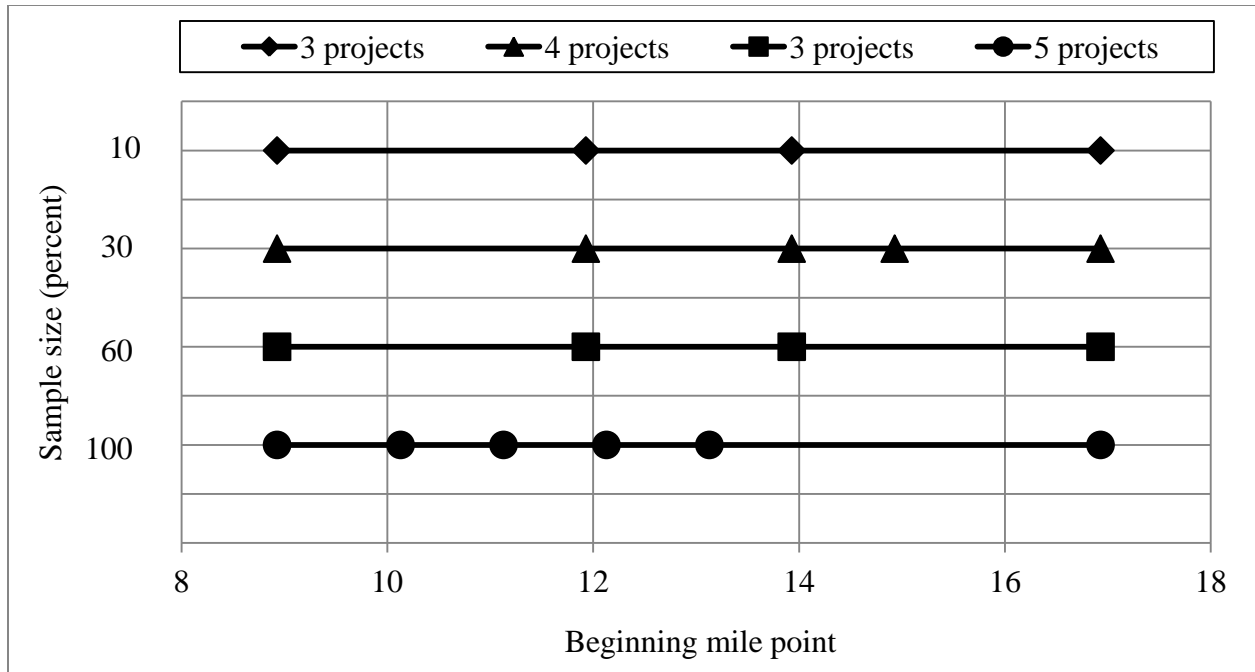


Figure B.55 Project boundaries based on the continuous (100 percent sampled) and the sampled rut depth data along a section of the composite SRID 161 in Washington

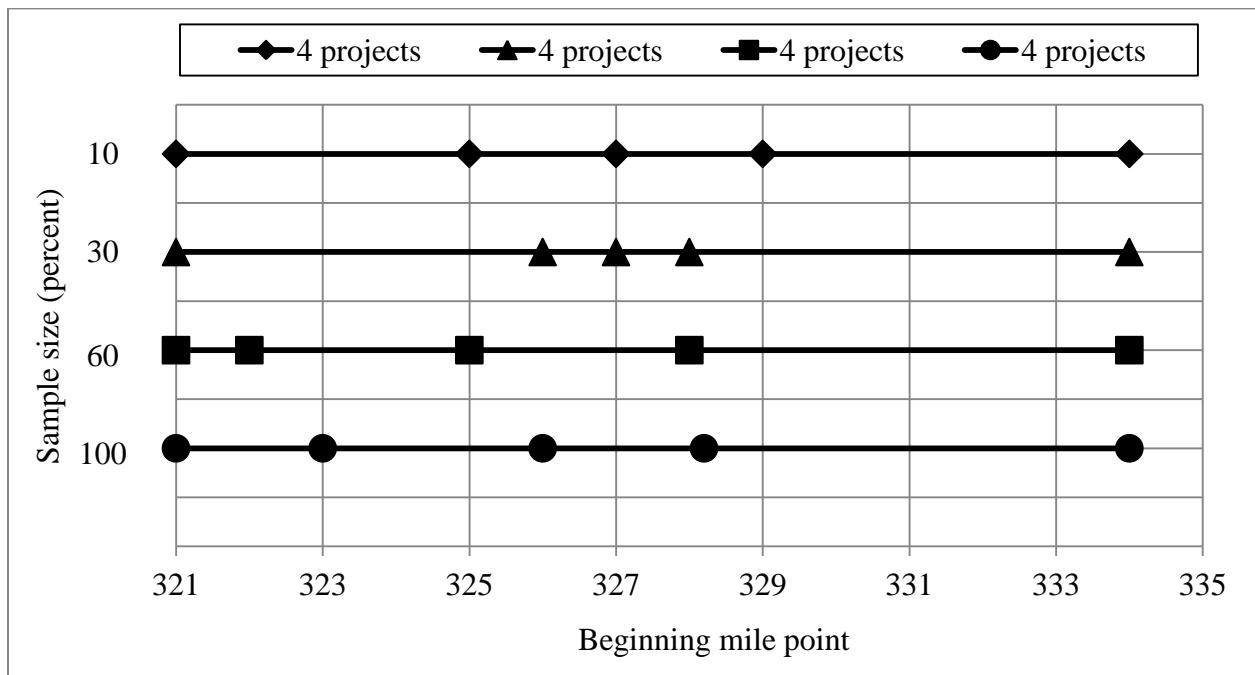


Figure B.56 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (rut depth) data along a section of the flexible HWY 24 in Colorado

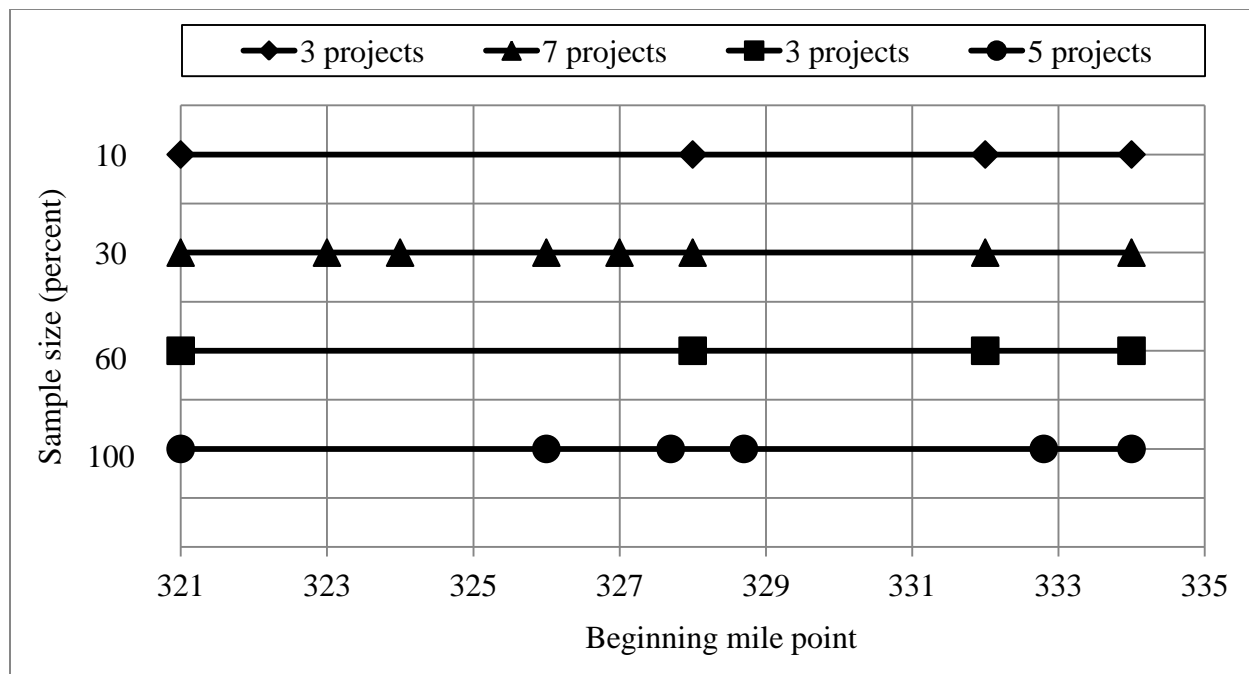


Figure B.57 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (transverse cracking) data along a section of the flexible HWY 24 in Colorado

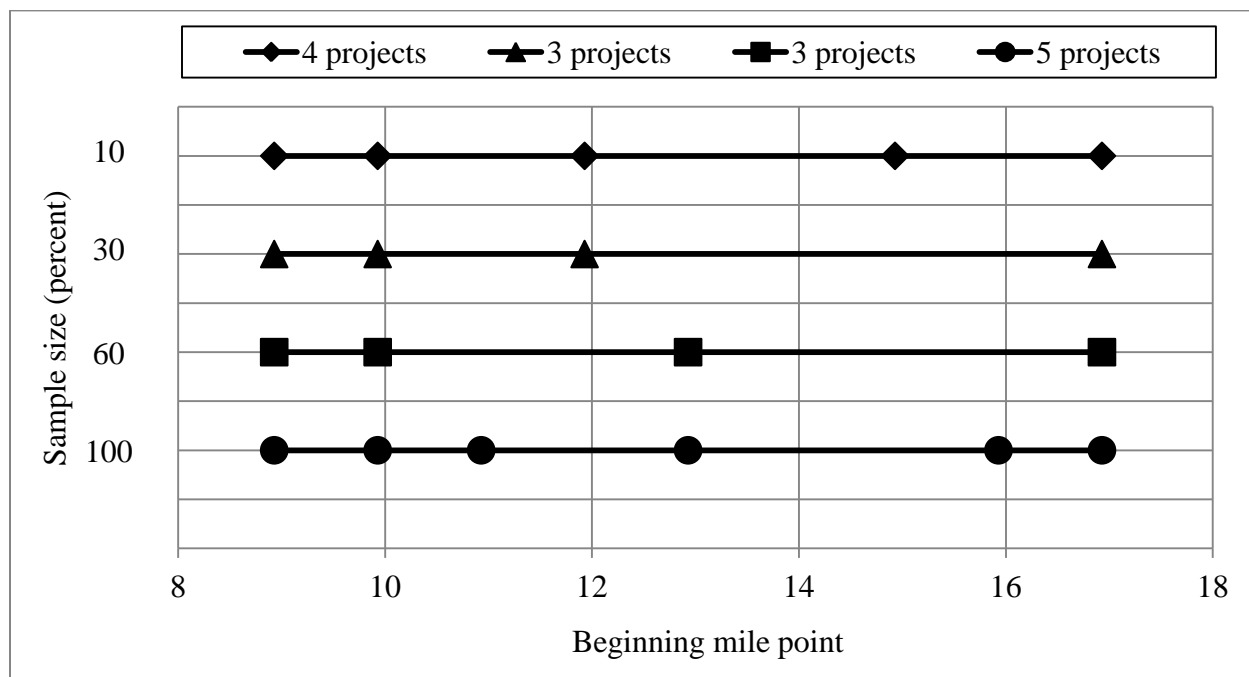


Figure B.58 Project boundaries based on the continuous (100 percent sampled) and sampled RSL (rut depth) data along a section of the composite SRID 161 in Washington

APPENDIX C

PAVEMENT ANALYSIS LENGTH FIGURES

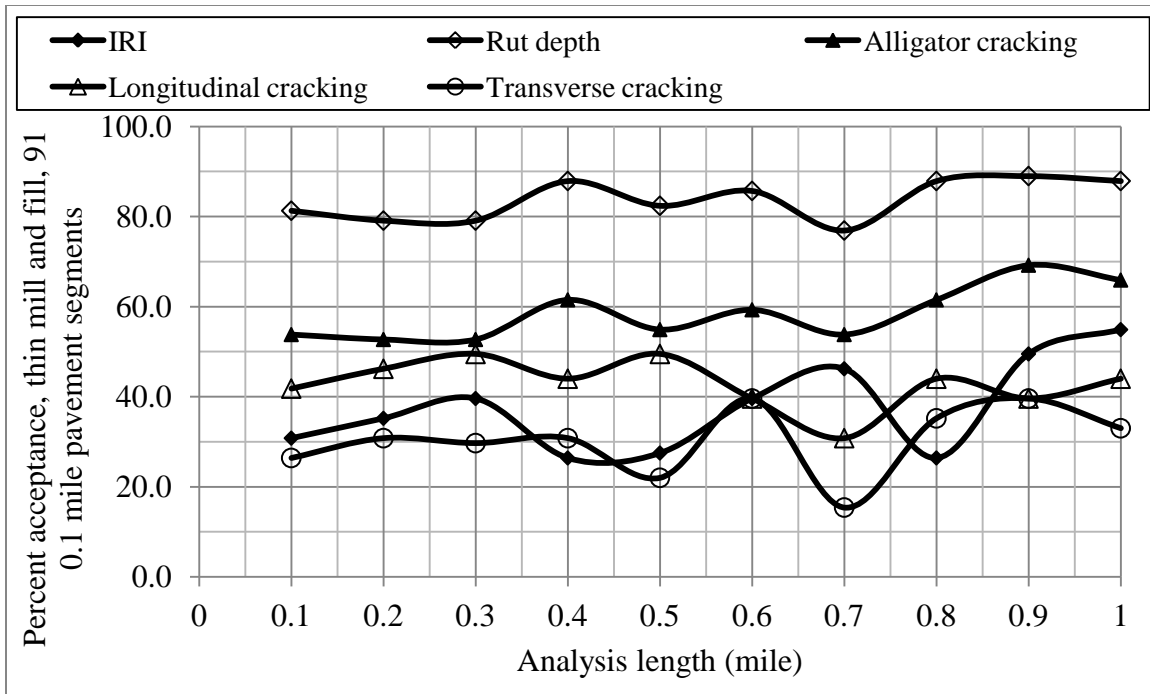


Figure C.1 Percent acceptance versus pavement analysis length, thin mill and fill, Colorado

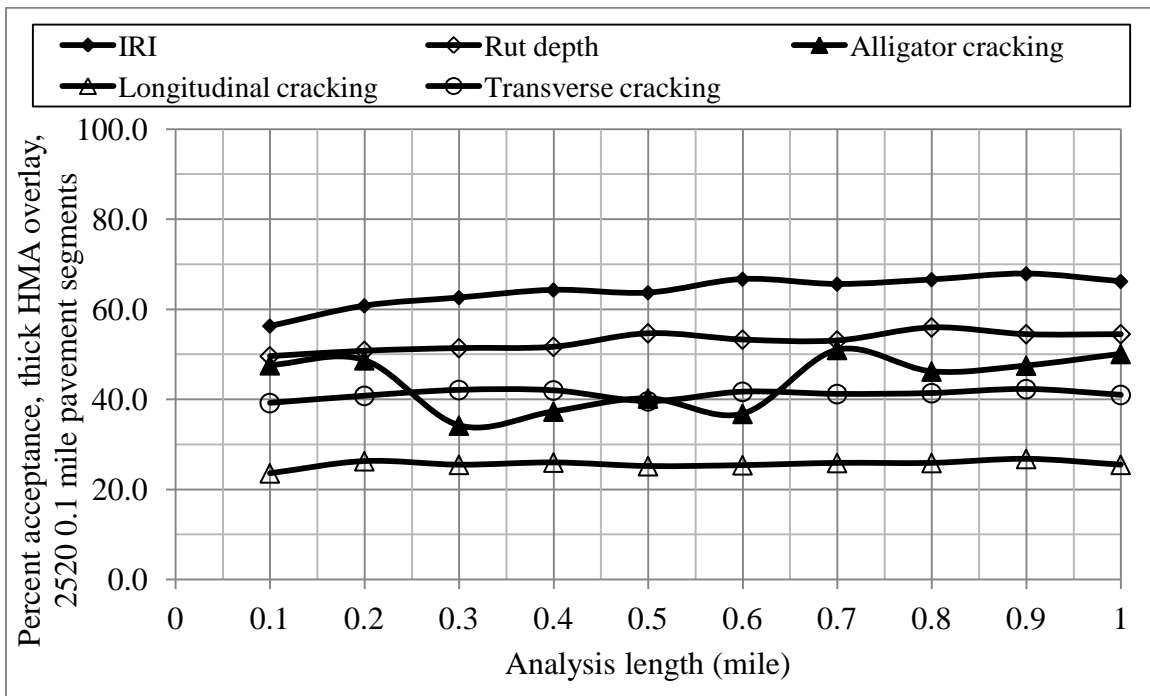


Figure C.2 Percent acceptance versus pavement analysis length, thick HMA overlay, Louisiana

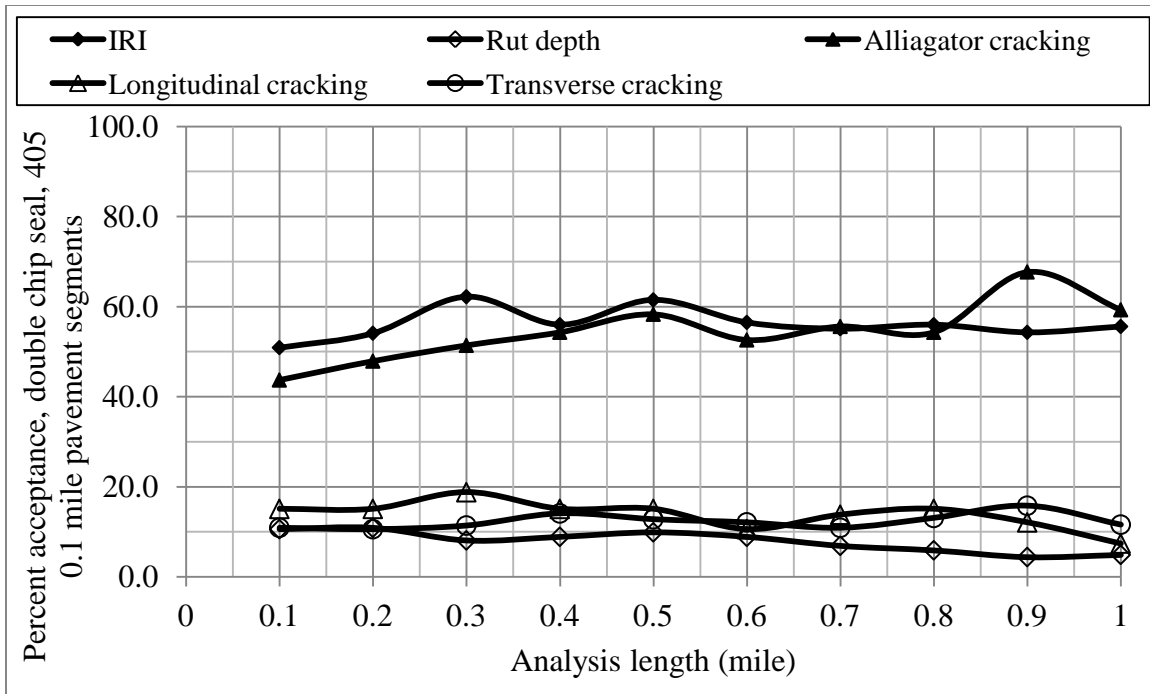


Figure C.3 Percent acceptance versus pavement analysis length, double chip seal, Louisiana

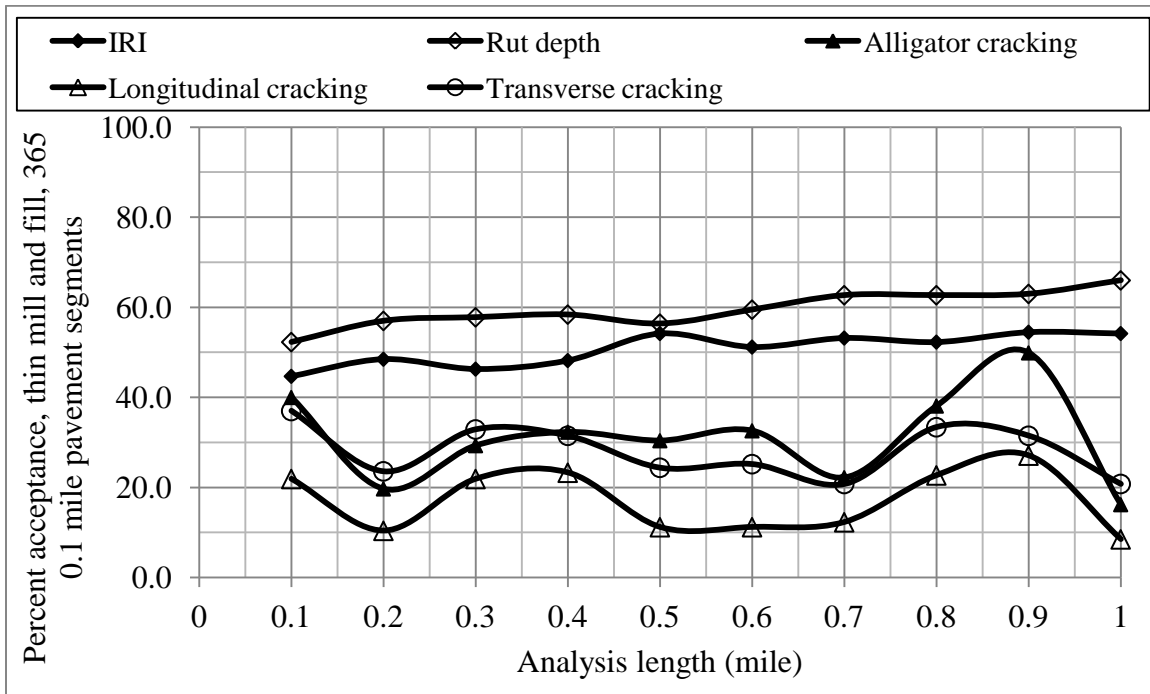


Figure C.4 Percent acceptance versus pavement analysis length, thin mill and fill, Louisiana

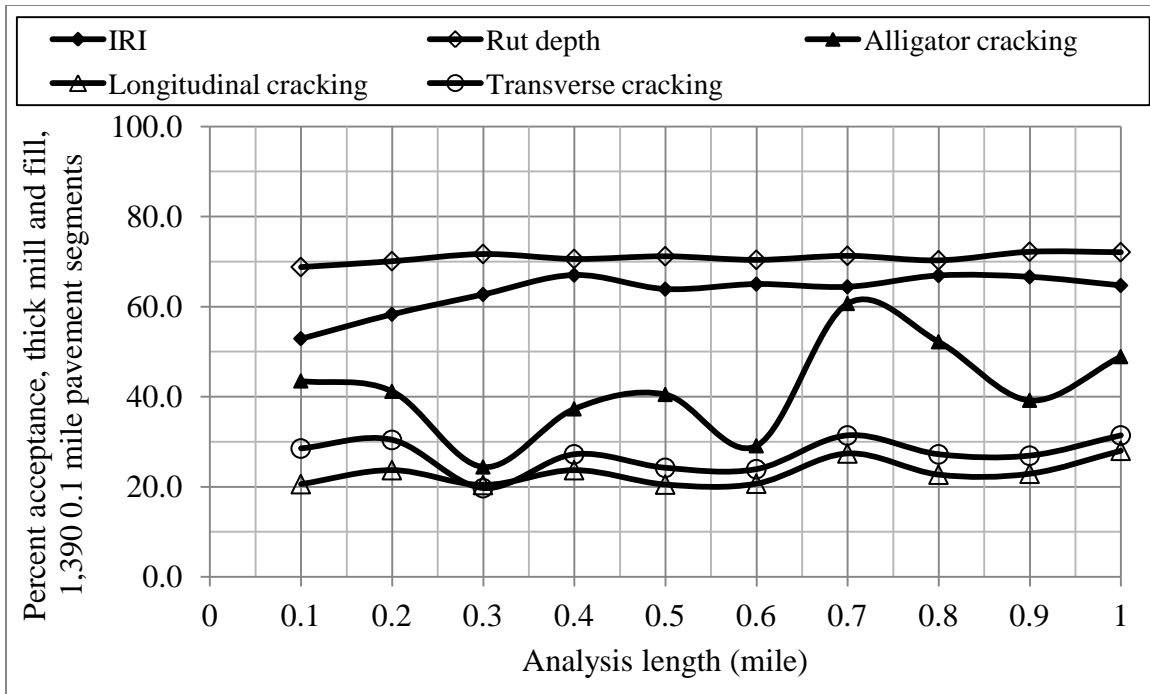


Figure C.5 Percent acceptance versus pavement analysis length, thick mill and fill, Louisiana

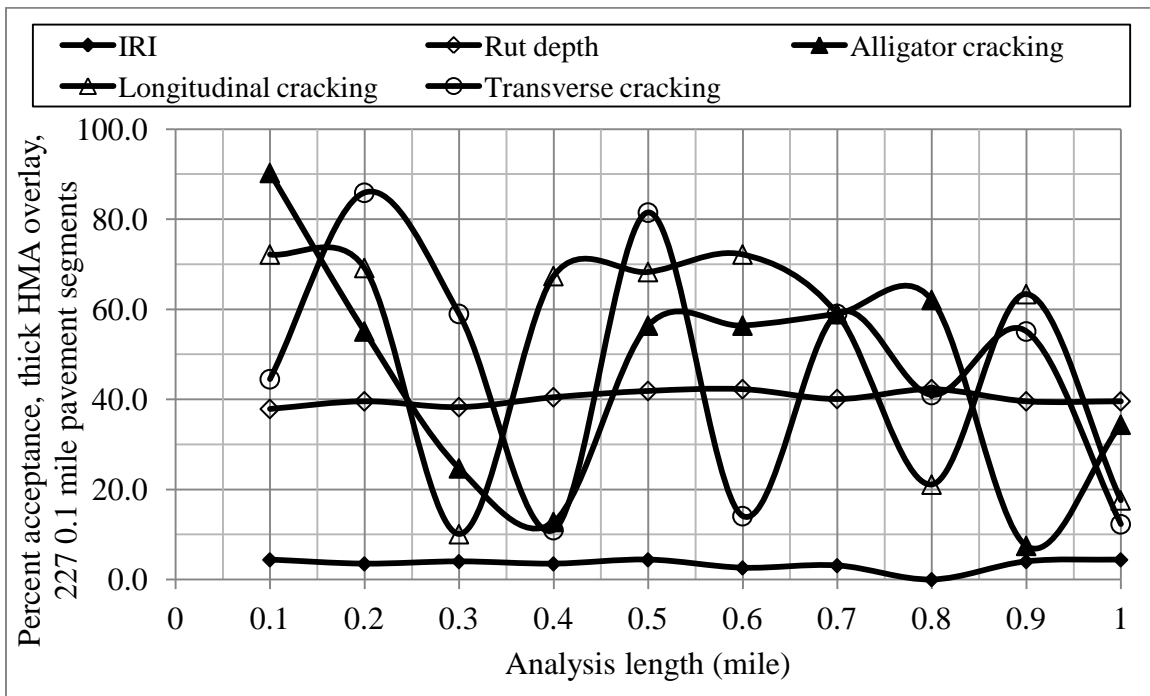


Figure C.6 Percent acceptance versus pavement analysis length, thick HMA overlay, Washington

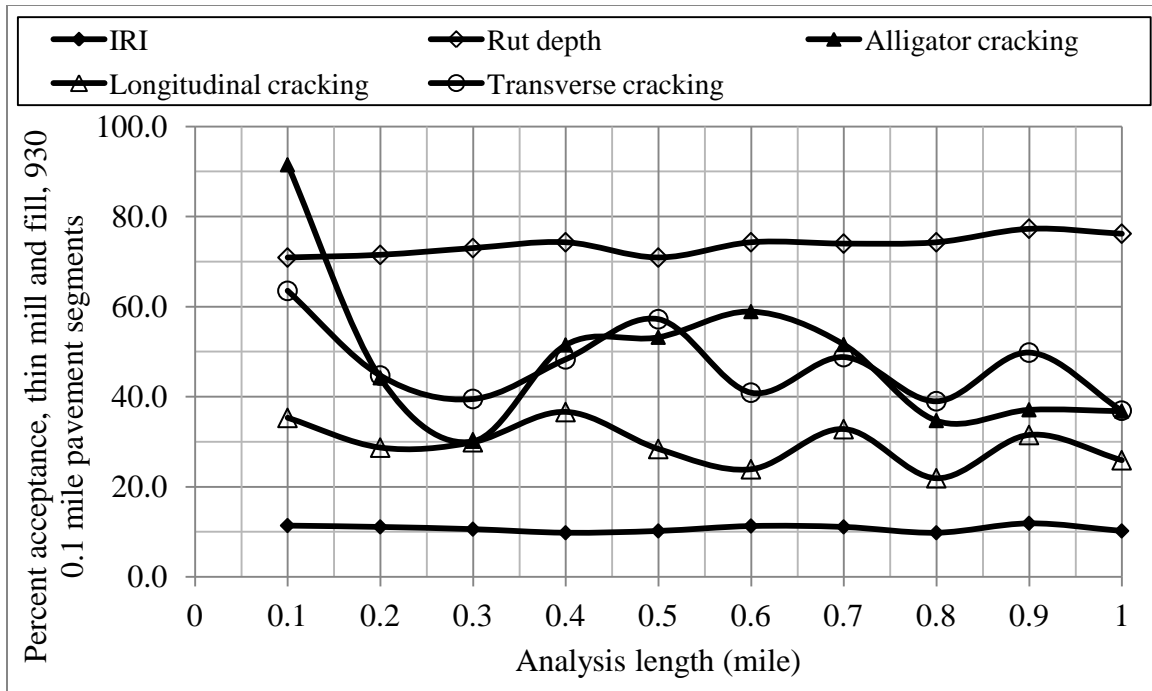


Figure C.7 Percent acceptance versus pavement analysis length, thin mill and fill, Washington

APPENDIX D

PAVEMENT PERFORMANCE DATA IMPUTATION FIGURES

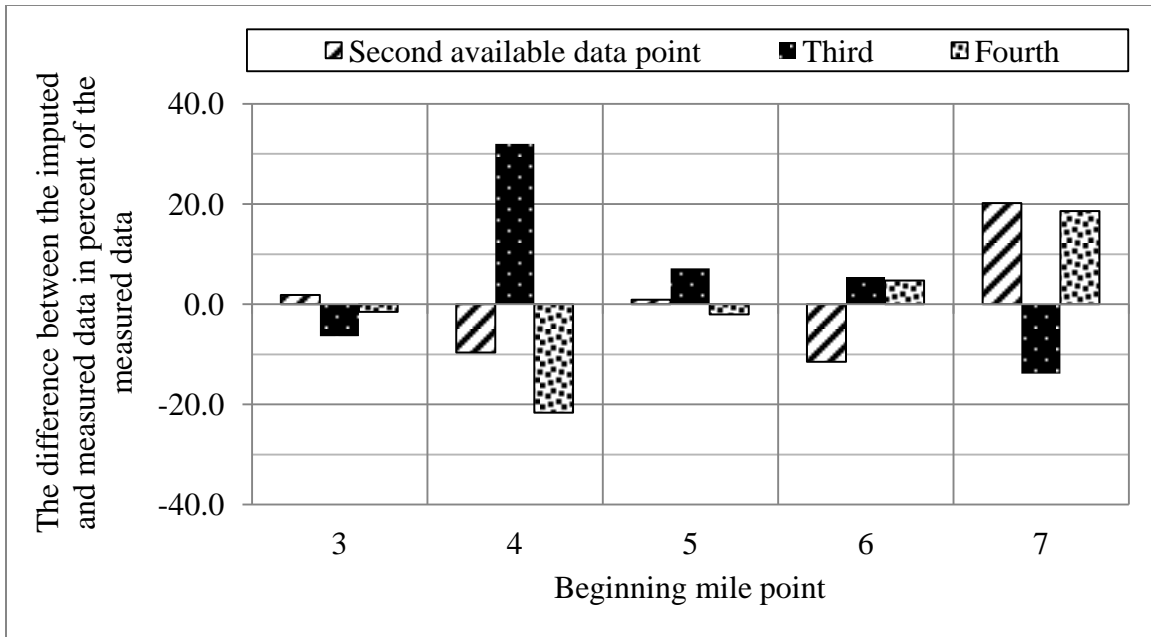


Figure D.1 The difference between the imputed and measured data in percent of the measured data versus the BMP along LA 21 (control section 0.02902) in Louisiana, three data points imputed using the linear interpolation method

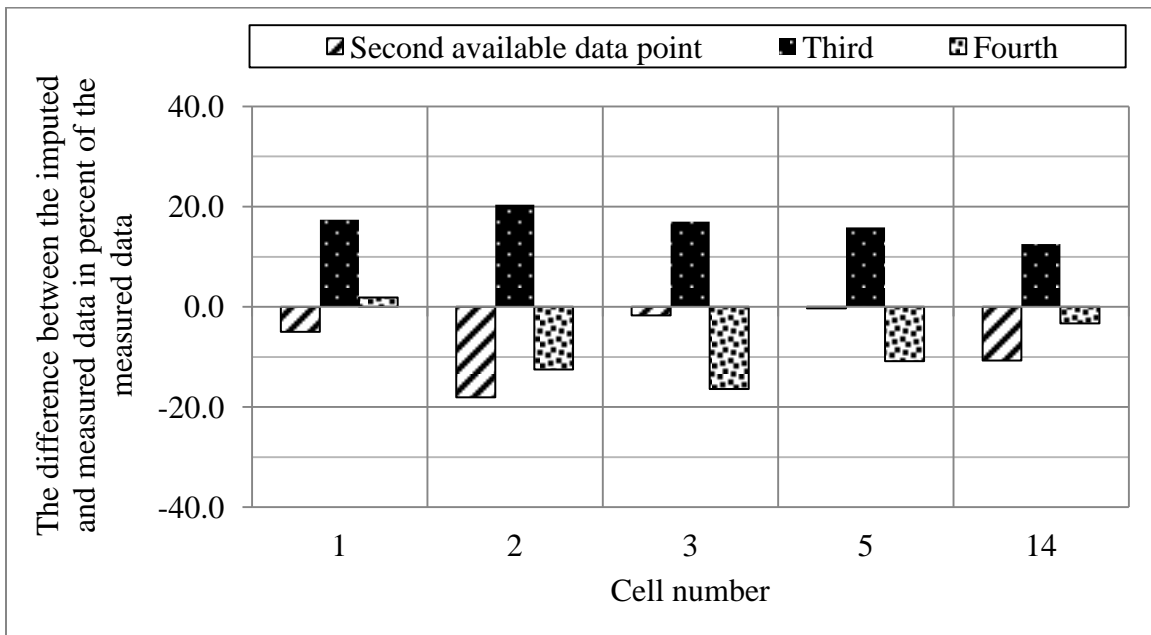


Figure D.2 The difference between the imputed and measured data in percent of the measured data versus cells 1, 2, 3, 5, and 14 at MnROAD, three data points imputed using the linear interpolation method

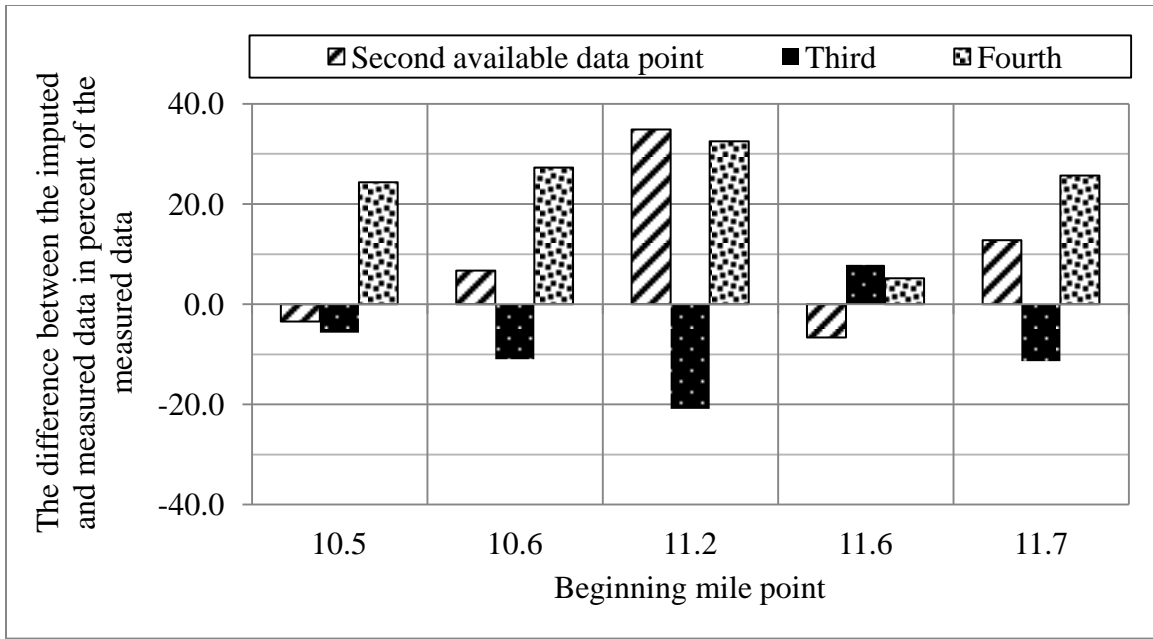


Figure D.3 The difference between the imputed and measured data in percent of the measured data versus the BMP along SRID 099 in Washington, three data points imputed using the linear interpolation method

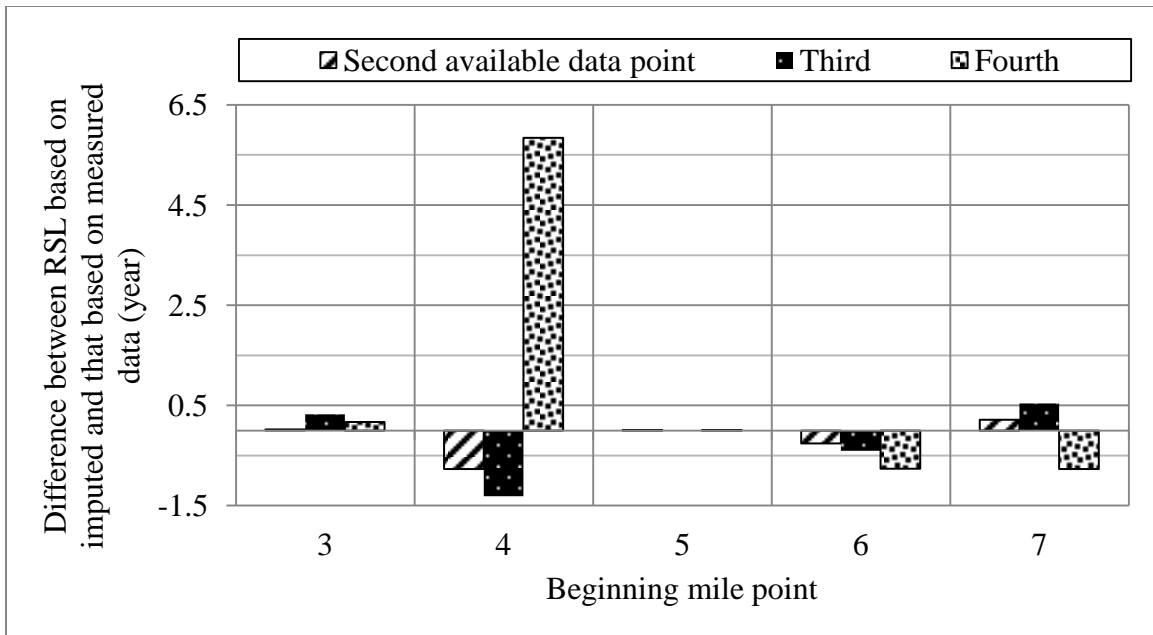


Figure D.4 The difference between the RSL values based on imputed and measured data versus the BMP along LA 21 (CS 0.02902) in Louisiana, three data points imputed using the linear interpolation method

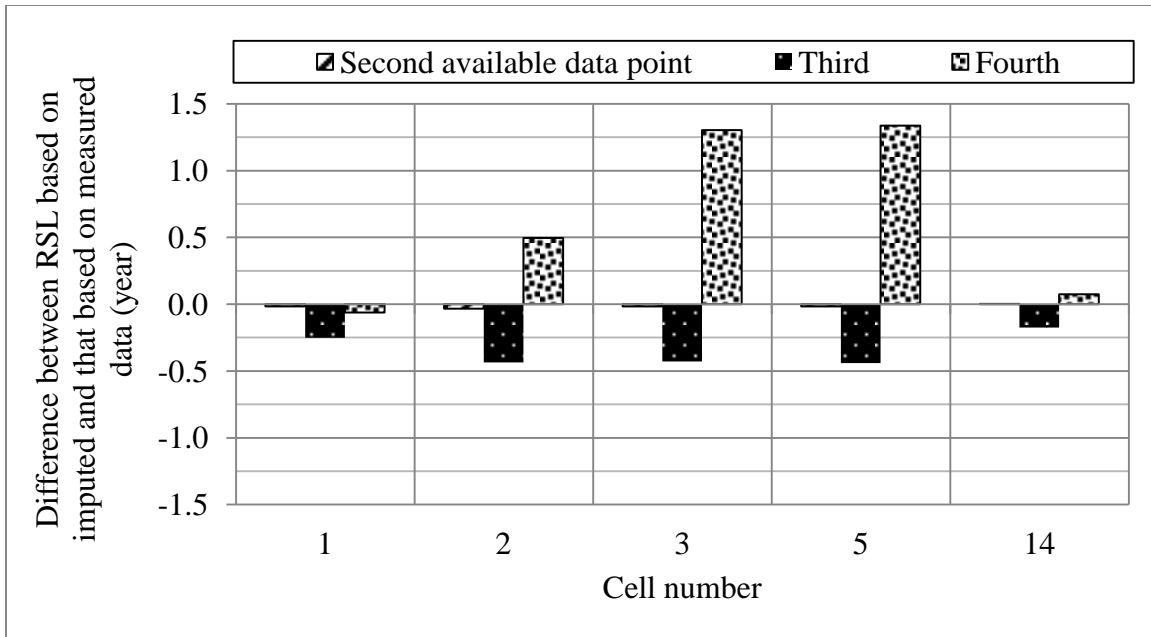


Figure D.5 The difference between the RSL values based on imputed and measured data versus the cell number at MnROAD, three data points imputed using the linear interpolation method

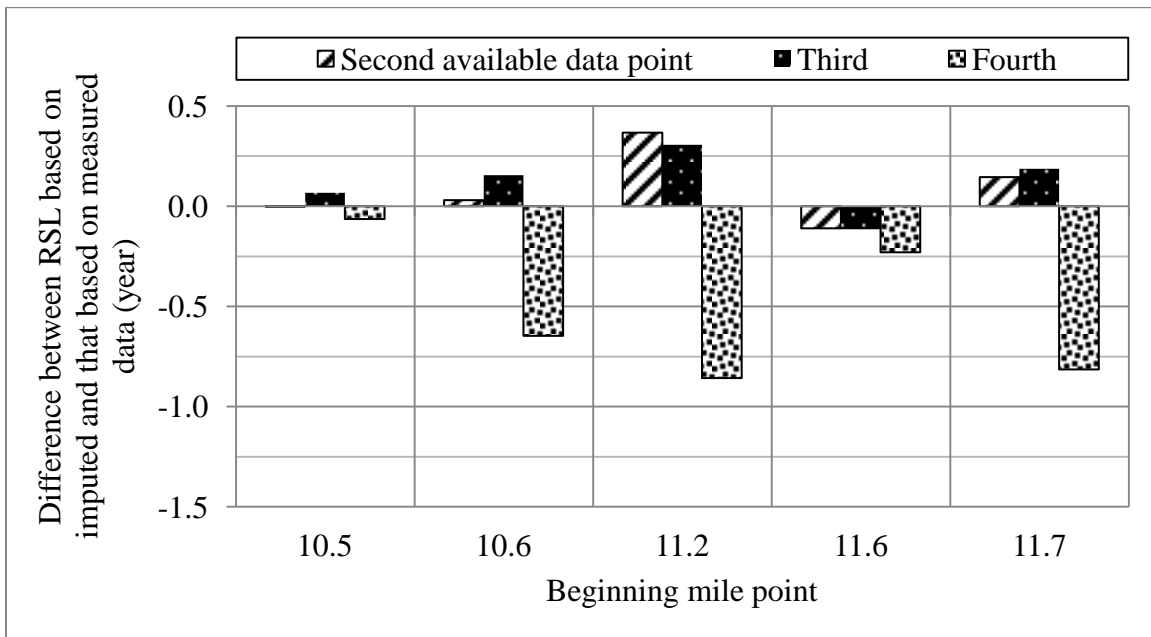


Figure D.6 The difference between the RSL values based on imputed and measured data versus the BMP along SRID 099 in Washington, three data points imputed using the linear interpolation method

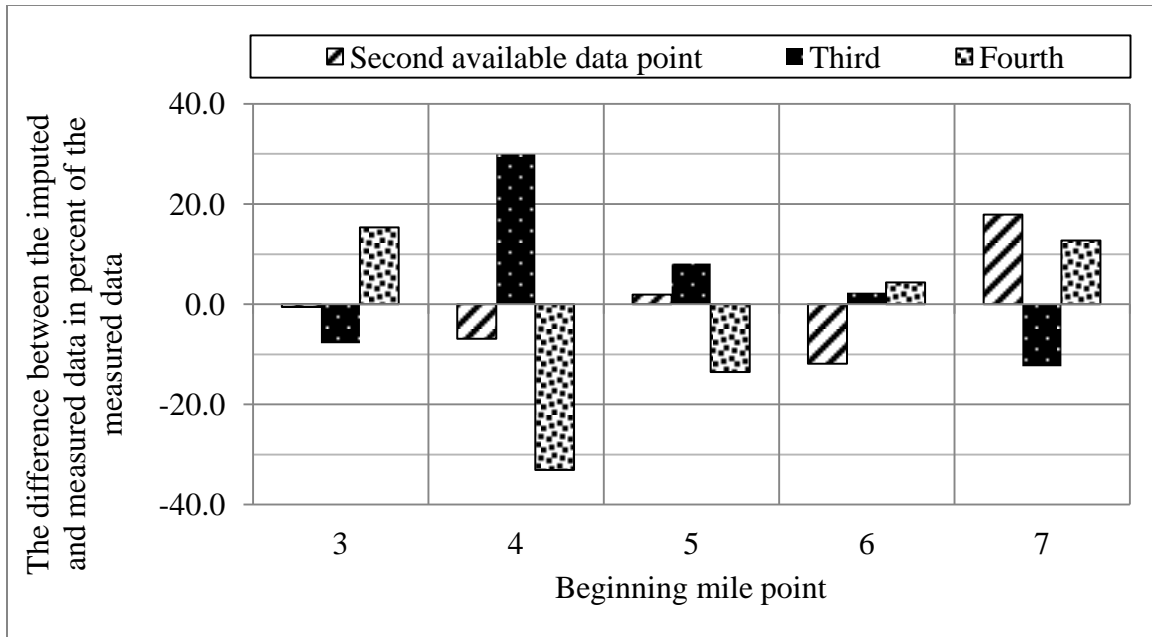


Figure D.7 The difference between the imputed and measured data in percent of the measured data versus the BMP along LA 21 (control section 0.02902) in Louisiana, three data points imputed using the regression method

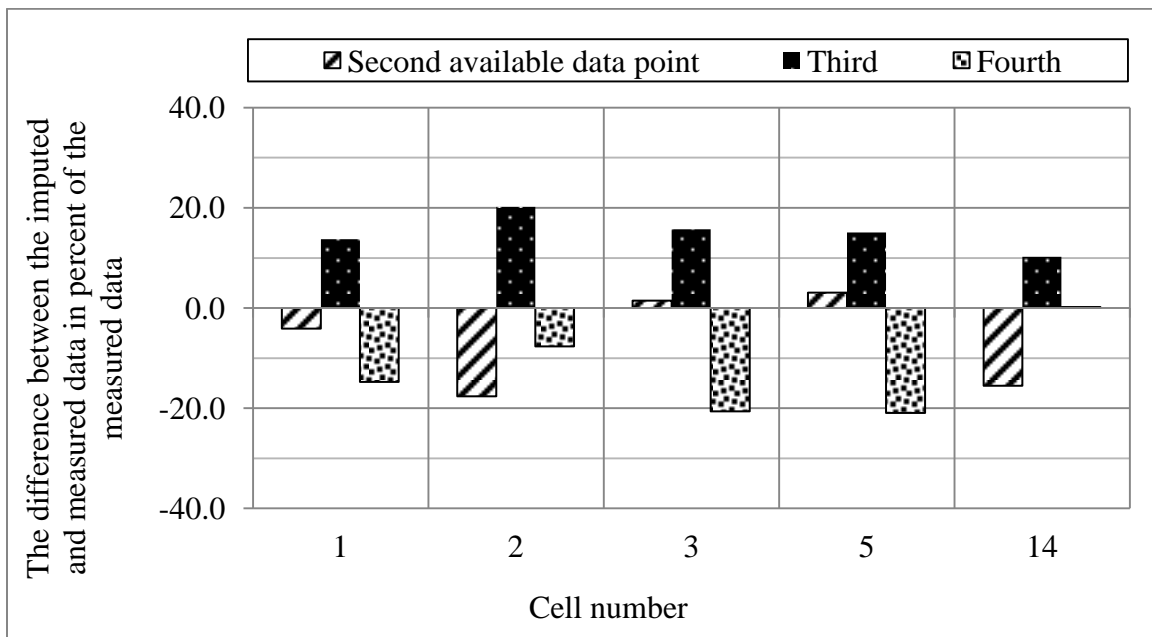


Figure D.8 The difference between the imputed and measured data in percent of the measured data versus cells 1, 2, 3, 5, and 14 at MnROAD, three data points imputed using the regression method

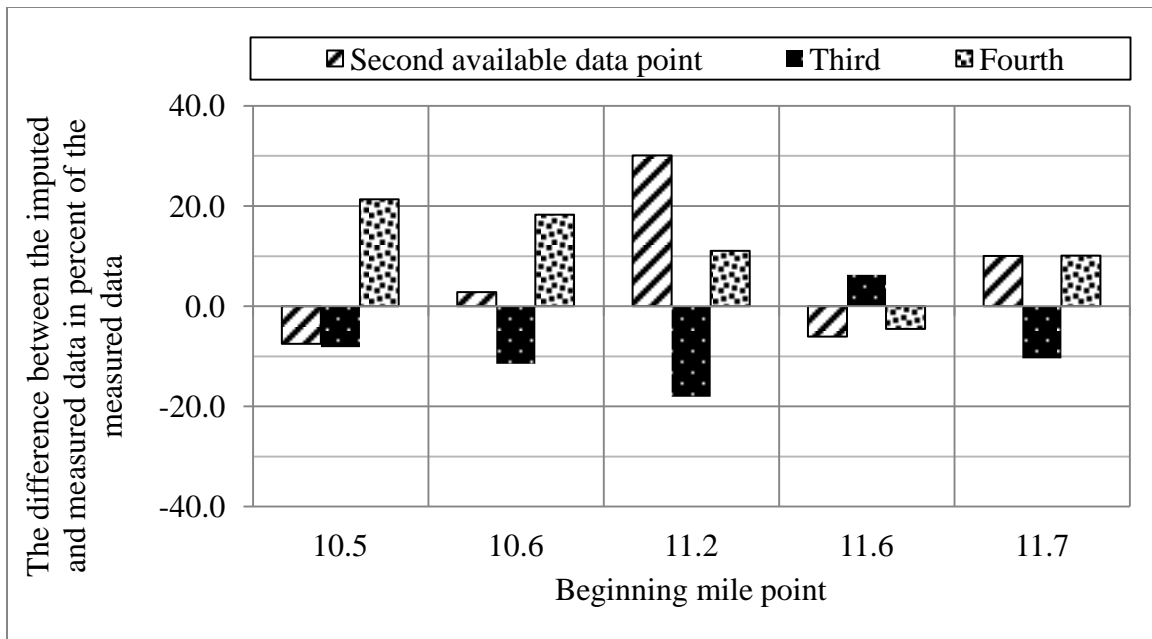


Figure D.9 The difference between the imputed and measured data in percent of the measured data versus the BMP along SRID 099 in Washington, three data points imputed using the regression method

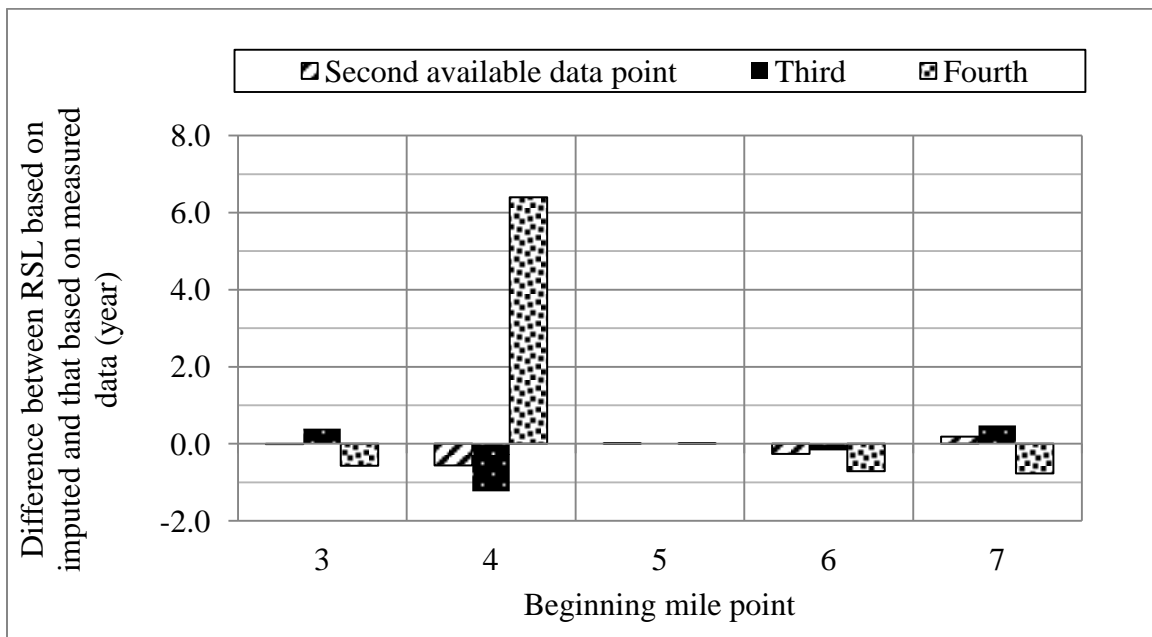


Figure D.10 The difference between the RSL values based on imputed and measured data versus the BMP along LA 21 (CS 0.02902) in Louisiana, three data points imputed using the regression method

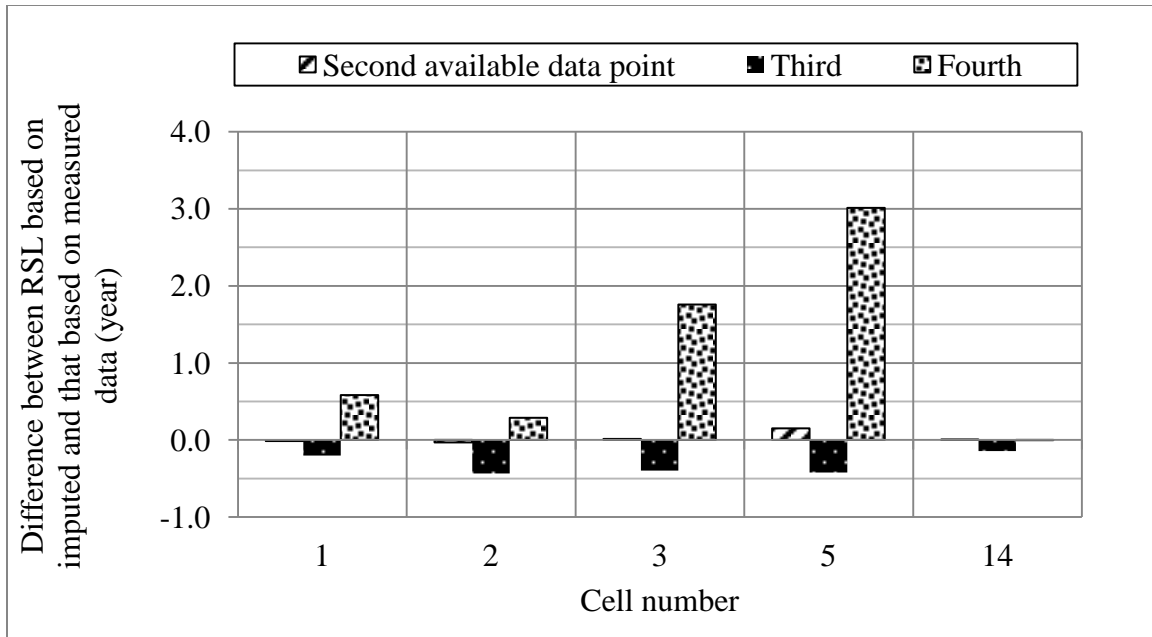


Figure D.11 The difference between the RSL values based on imputed and measured data versus the cell number at MnROAD, three data points imputed using the regression method

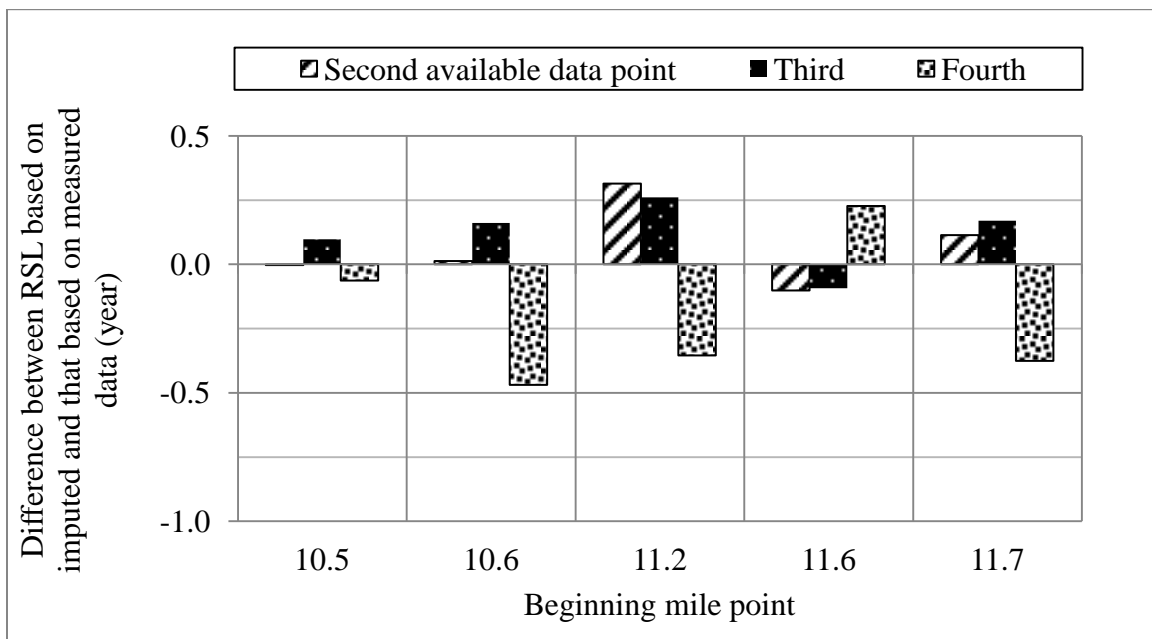


Figure D.12 The difference between the RSL values based on imputed and measured data versus the BMP along SRID 099 in Washington, three data points imputed using the regression method

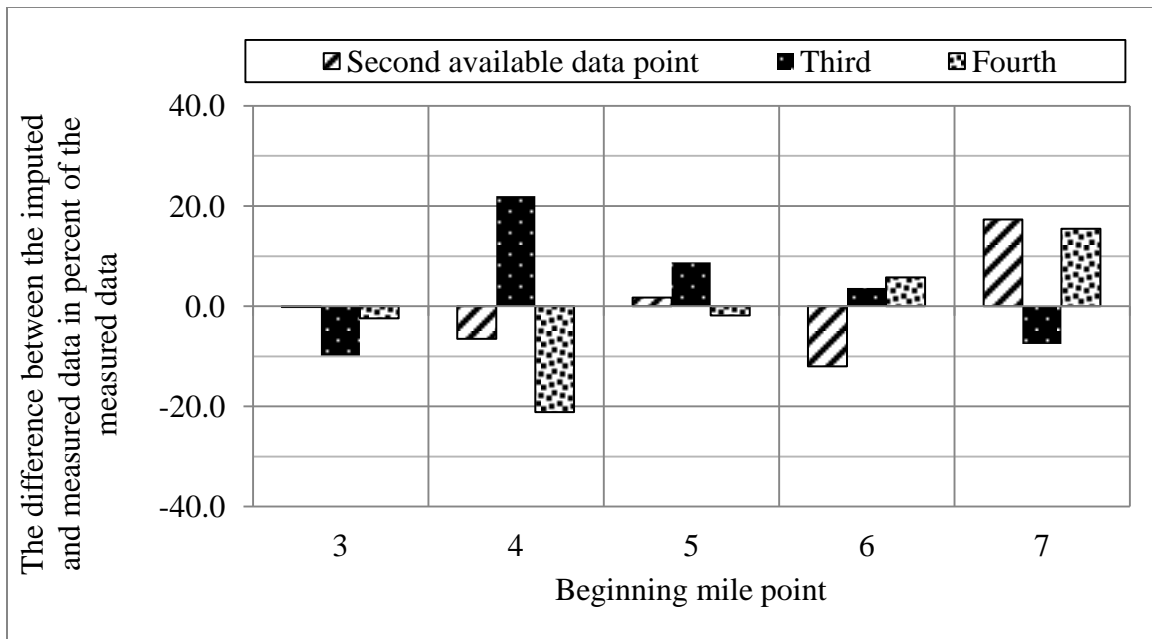


Figure D.13 The difference between the imputed and measured data in percent of the measured data versus the BMP along LA 21 (control section 0.02902) in Louisiana, three data points imputed using the moving regression method

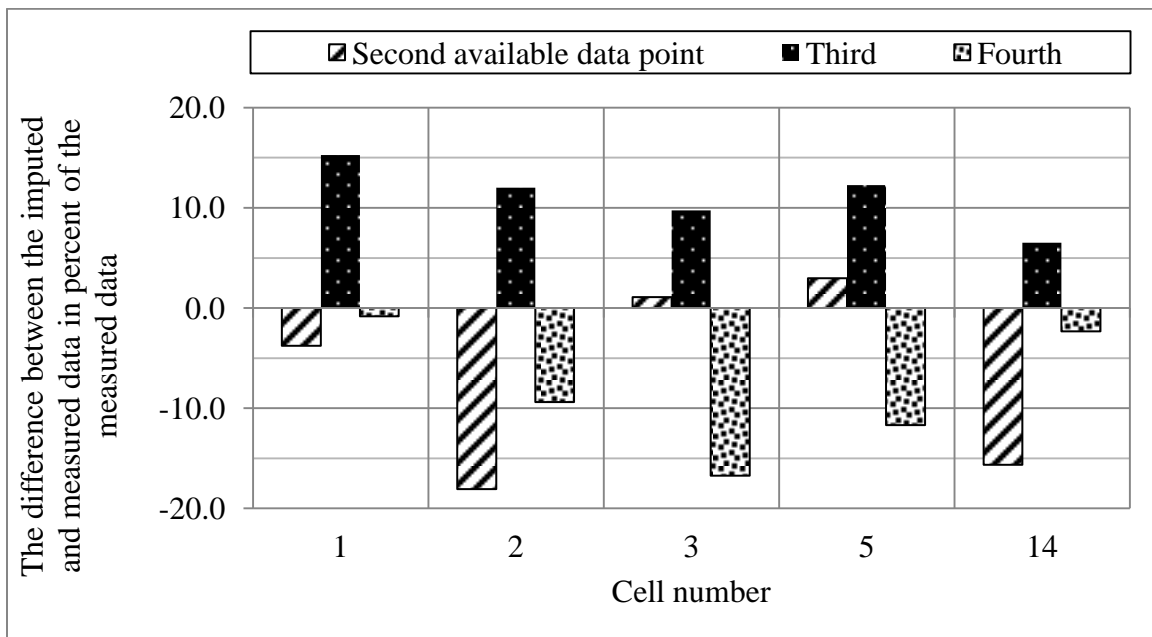


Figure D.14 The difference between the imputed and measured data in percent of the measured data versus cells 1, 2, 3, 5, and 14 at MnROAD, three data points imputed using the moving regression method

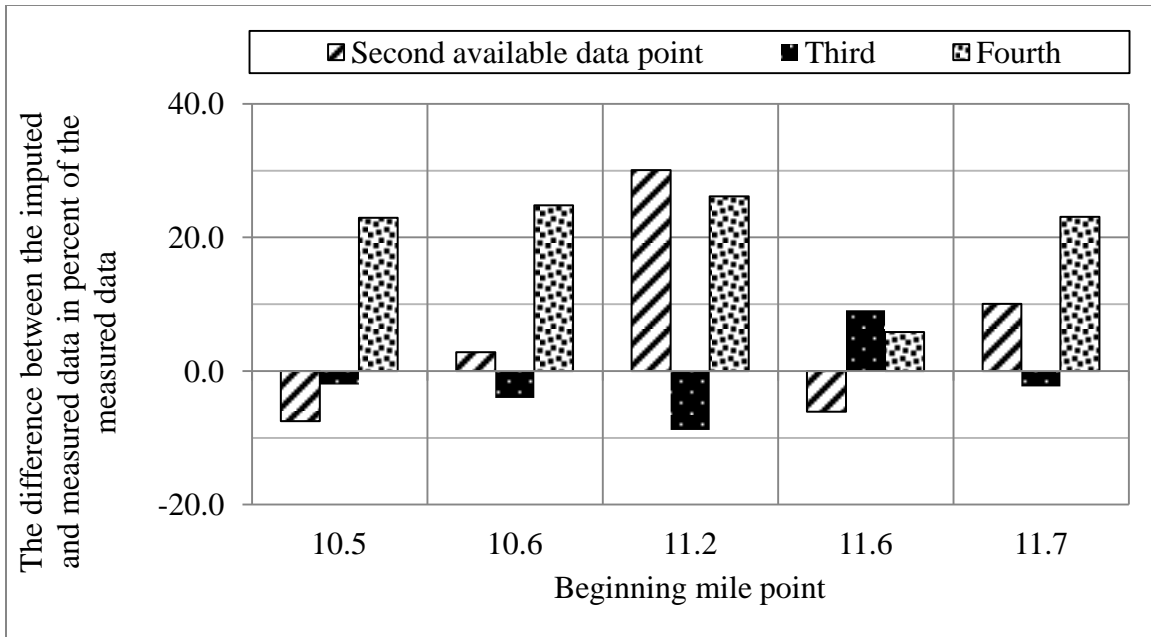


Figure D.15 The difference between the imputed and measured data in percent of the measured data versus the BMP along SRID 099 in Washington, three data points imputed using the moving regression method

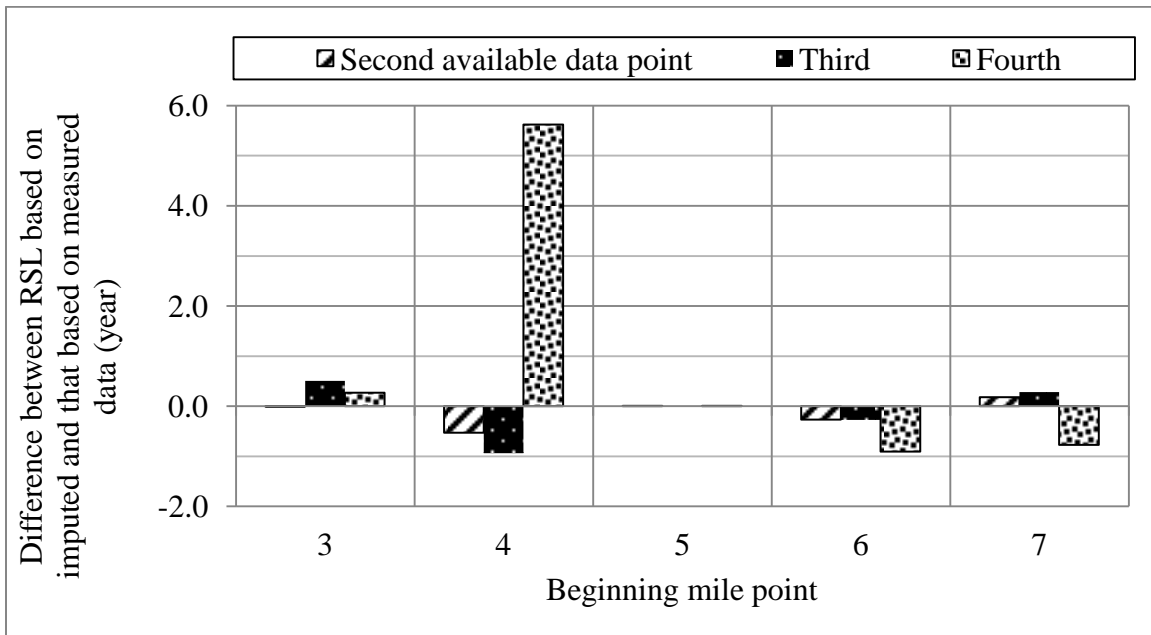


Figure D.16 The difference between the RSL values based on imputed and measured data versus the BMP along LA 21 (CS 0.02902) in Louisiana, three data points imputed using the moving regression method

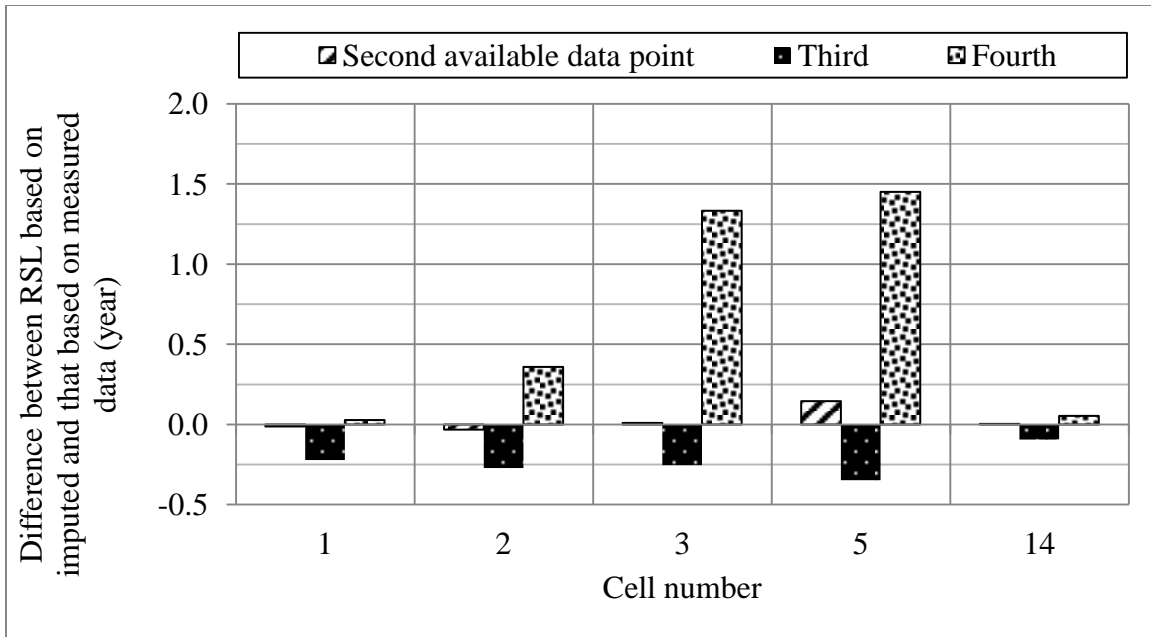


Figure D.17 The difference between the RSL values based on imputed and measured data versus the cell number at MnROAD, three data points imputed using the moving regression method

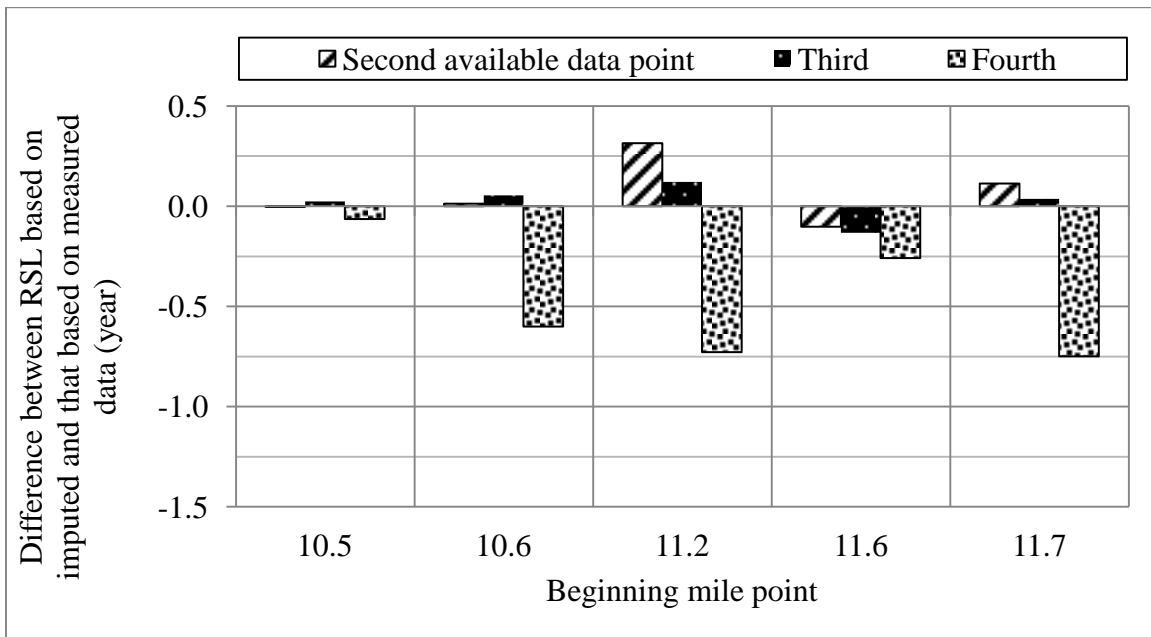


Figure D.18 The difference between the RSL values based on imputed and measured data versus the BMP along SRID 099 in Washington, three data points imputed using the moving regression method

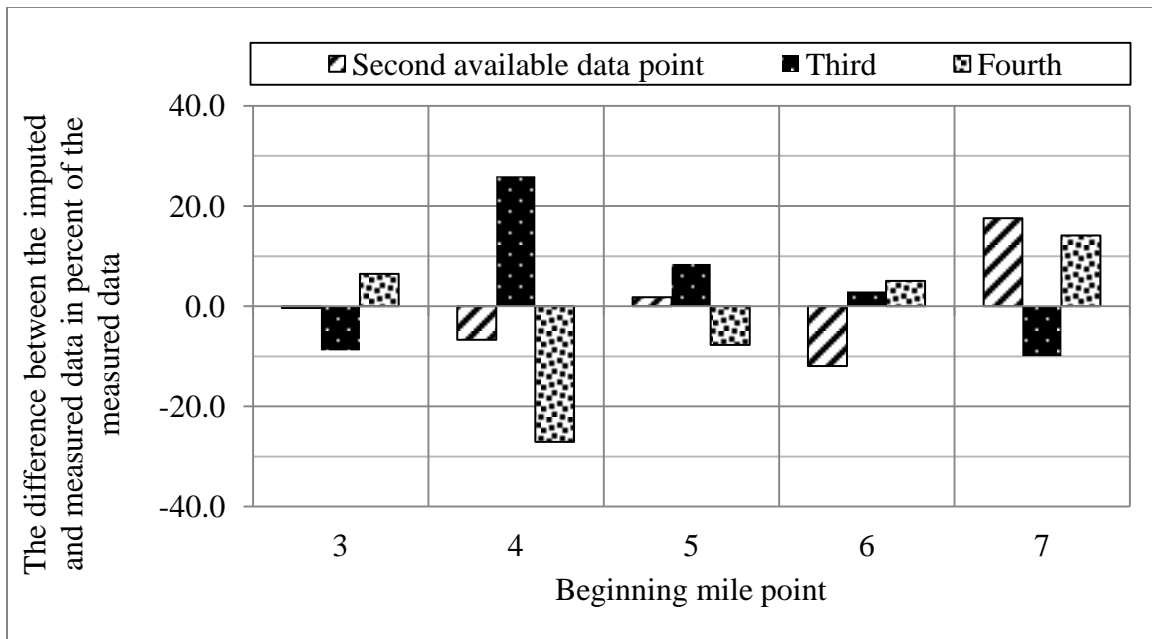


Figure D.19 The difference between the imputed and measured data in percent of the measured data versus the BMP along LA 21 (control section 0.02902) in Louisiana, three data points imputed using the multiple regressions method

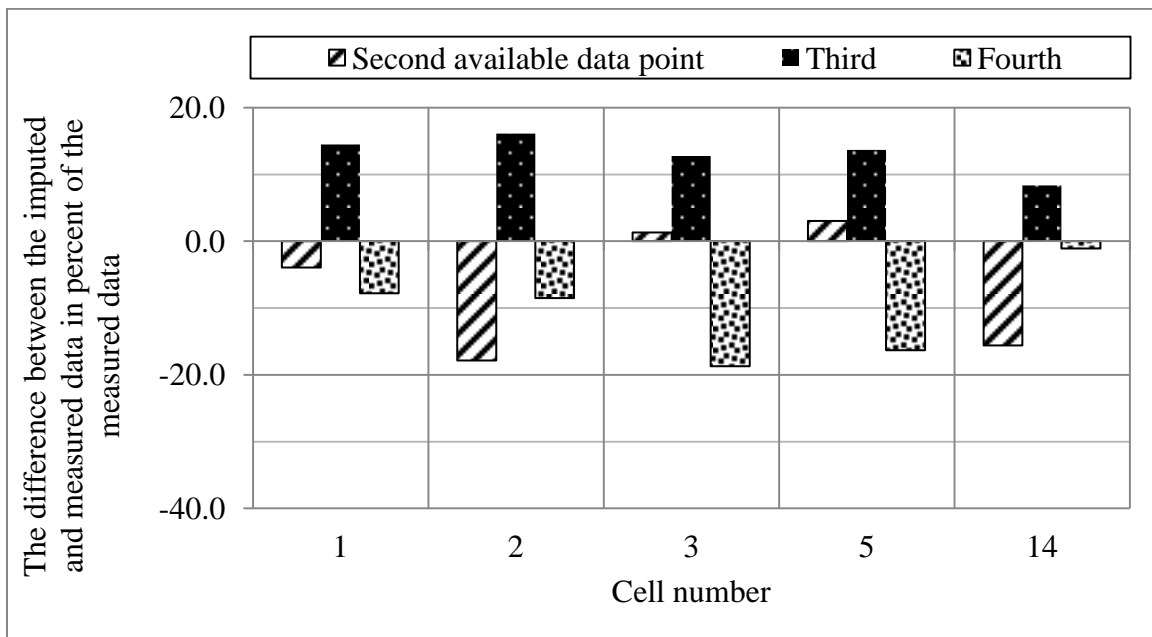


Figure D.20 The difference between the imputed and measured data in percent of the measured data versus cells 1, 2, 3, 5, and 14 at MnROAD, three data points imputed using the multiple regressions method

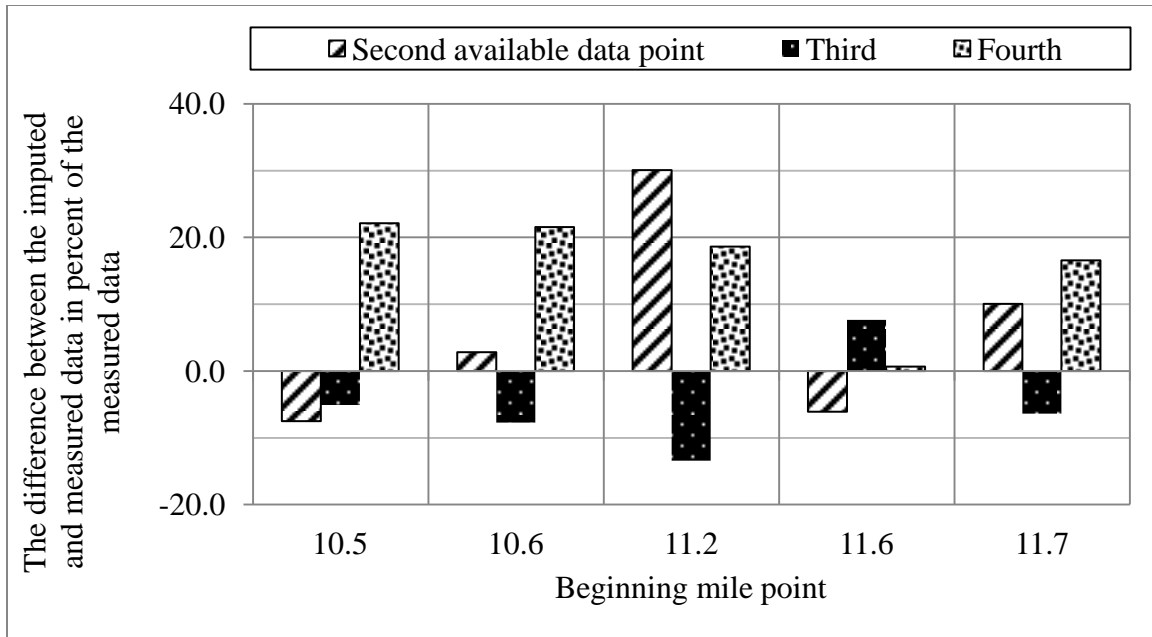


Figure D.21 The difference between the imputed and measured data in percent of the measured data versus the BMP along SRID 099 in Washington, three data points imputed using the multiple regressions method

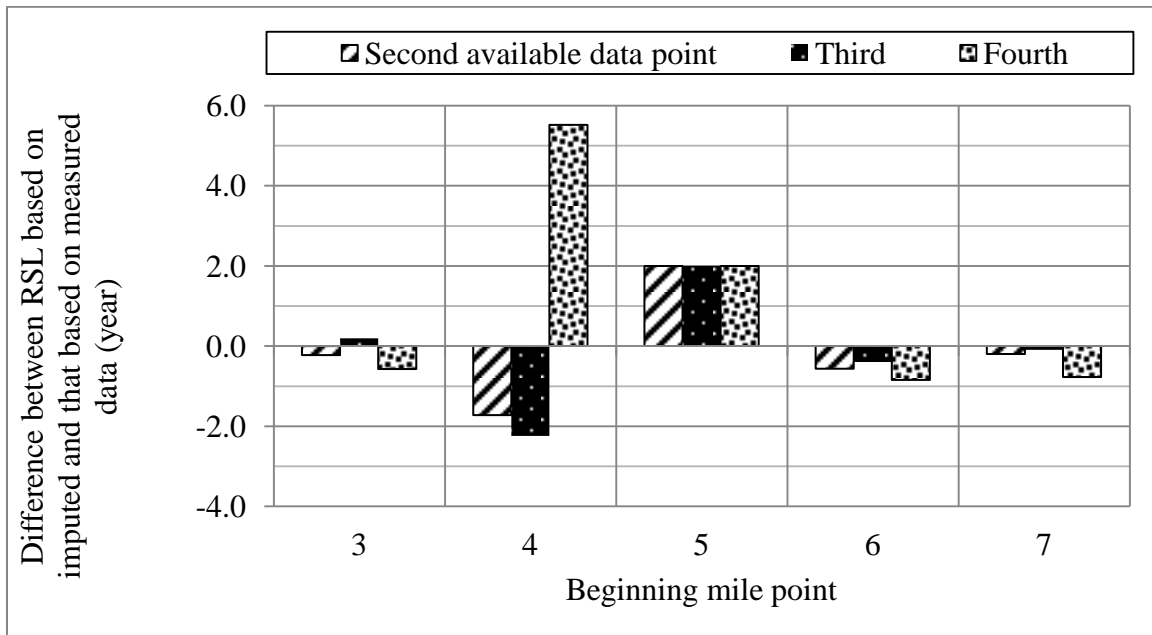


Figure D.22 The difference between the RSL values based on imputed and measured data versus the BMP along LA 21 (CS 0.02902) in Louisiana, three data points imputed using the multiple regressions method

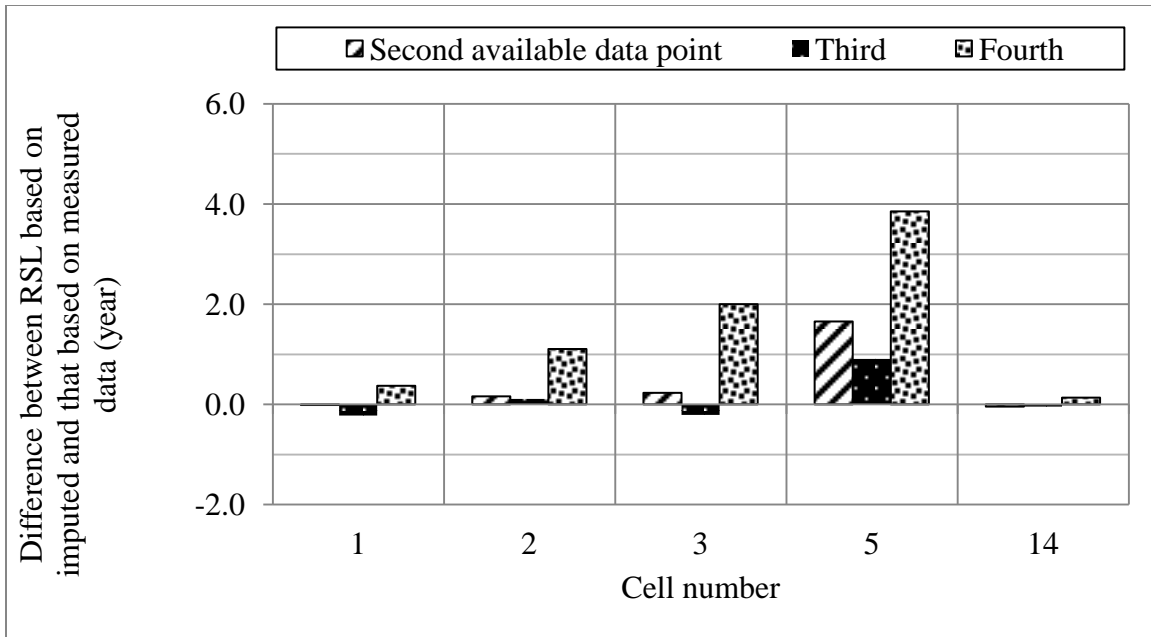


Figure D.23 The difference between the RSL values based on imputed and measured data versus the cell number at MnROAD, three data points imputed using the multiple regressions method

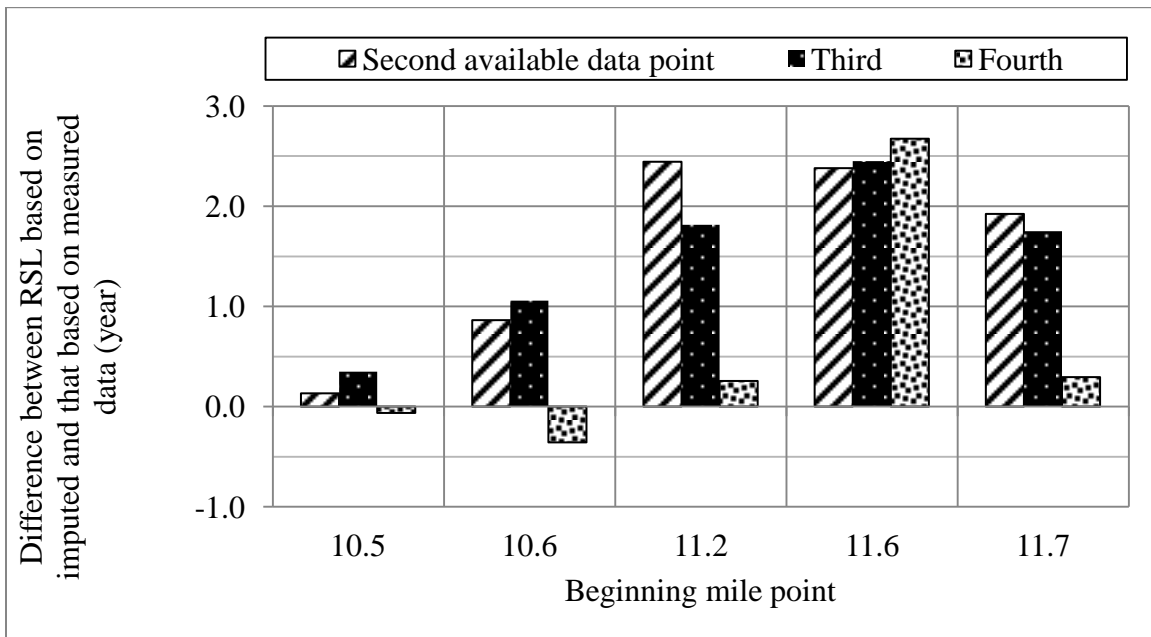


Figure D.24 The difference between the RSL values based on imputed and measured data versus the BMP along SRID 099 in Washington, three data points imputed using the multiple regressions method

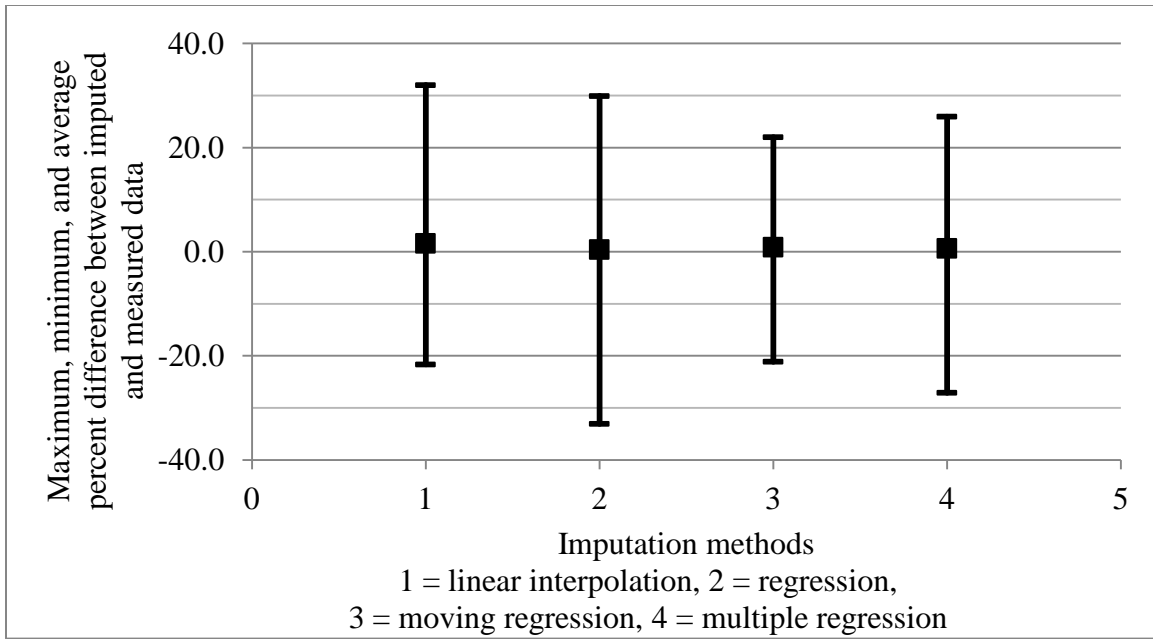


Figure D.25 Maximum, minimum, and average percent difference between imputed and measured data at BMPs 3, 4, 5, 6, and 7 along LA 21 (CS 0.02902) in Louisiana, using four imputation methods

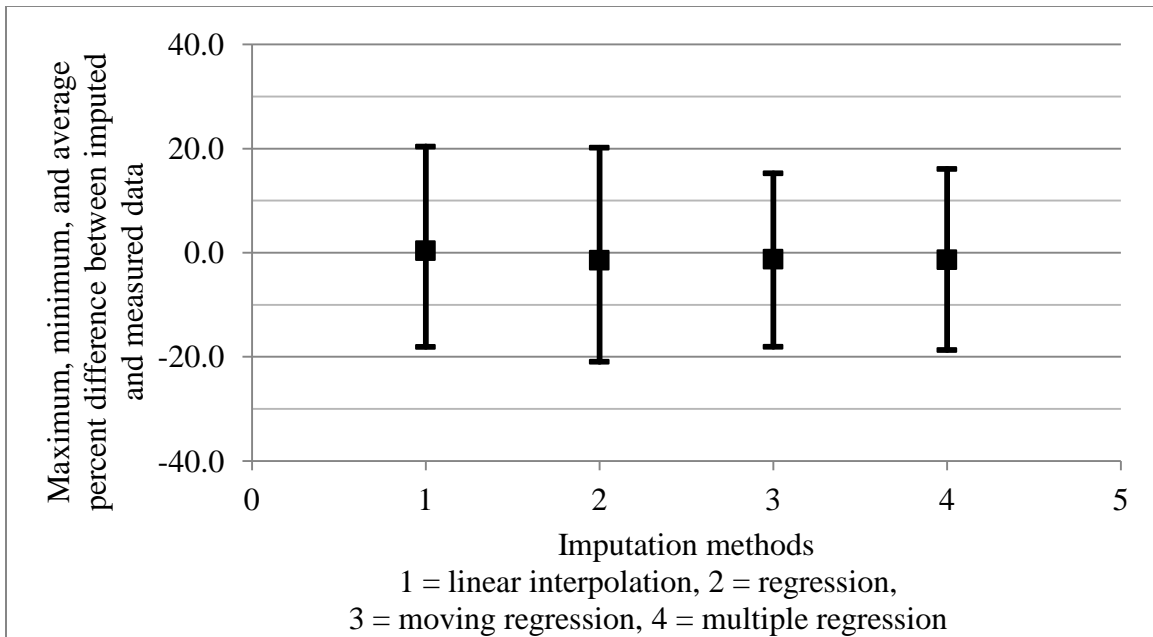


Figure D.26 Maximum, minimum, and average percent difference between imputed and measured data at cells 1, 2, 3, 5, and 14 at MnROAD, using four imputation methods

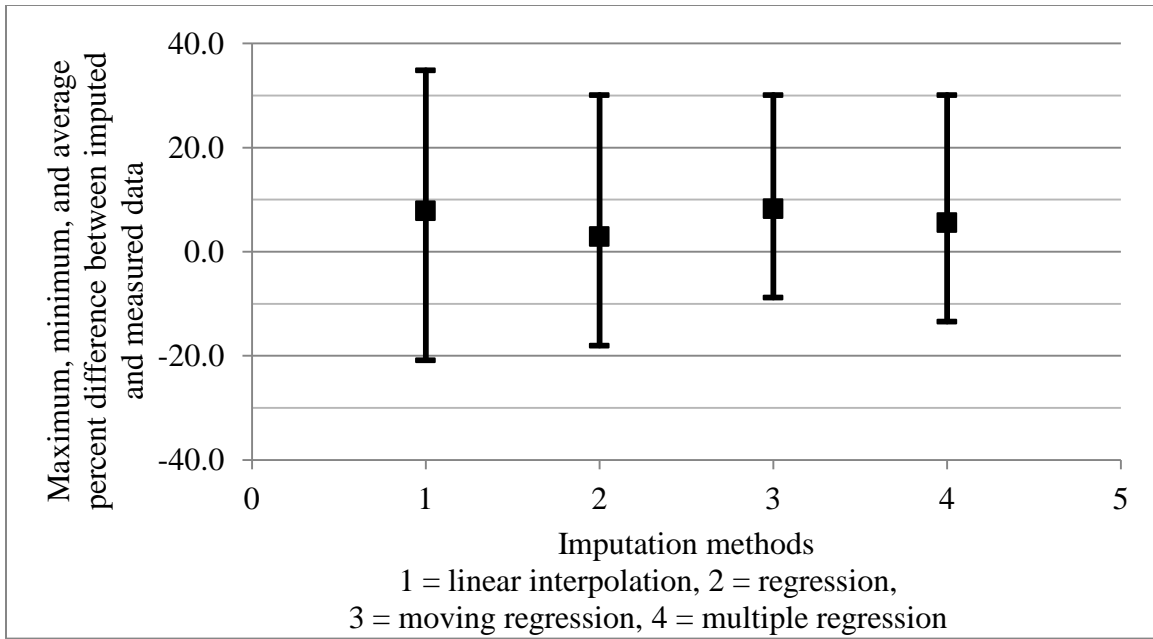


Figure D.27 Maximum, minimum, and average percent difference between imputed and measured data at BMPs 10.5, 10.6, 11.2, 11.6, and 11.7 along SRID 099 in Washington, using four imputation methods

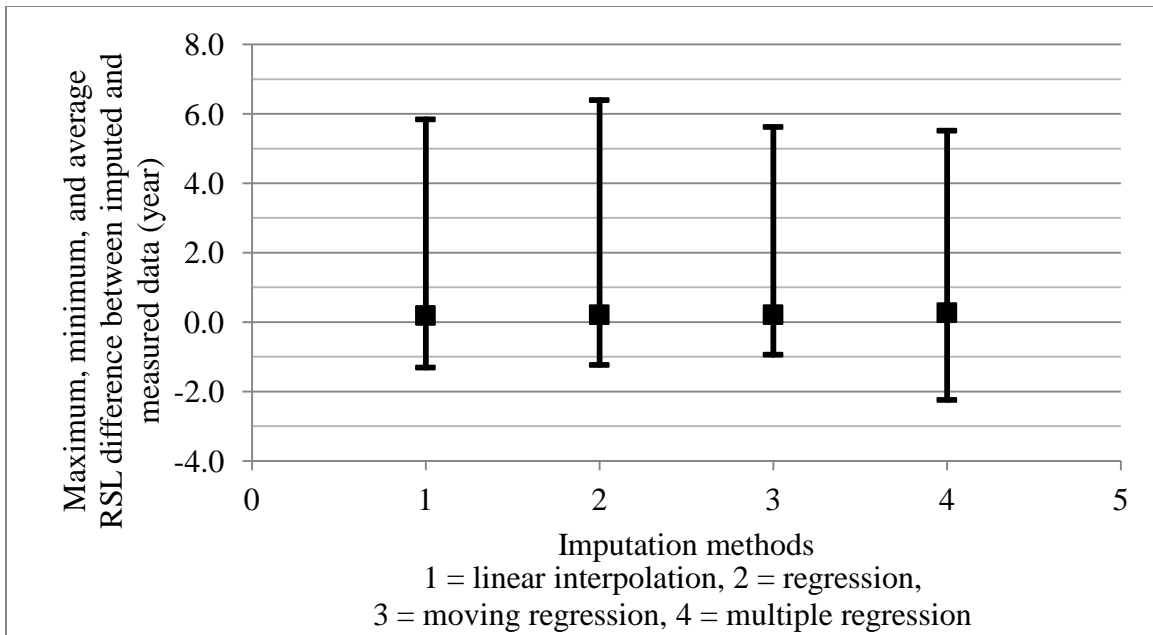


Figure D.28 Maximum, minimum, and average difference between RSL values based on imputed and measured data at BMPs 3, 4, 5, 6, and 7 along LA 21 (CS 0.02902) in Louisiana, using four imputation methods

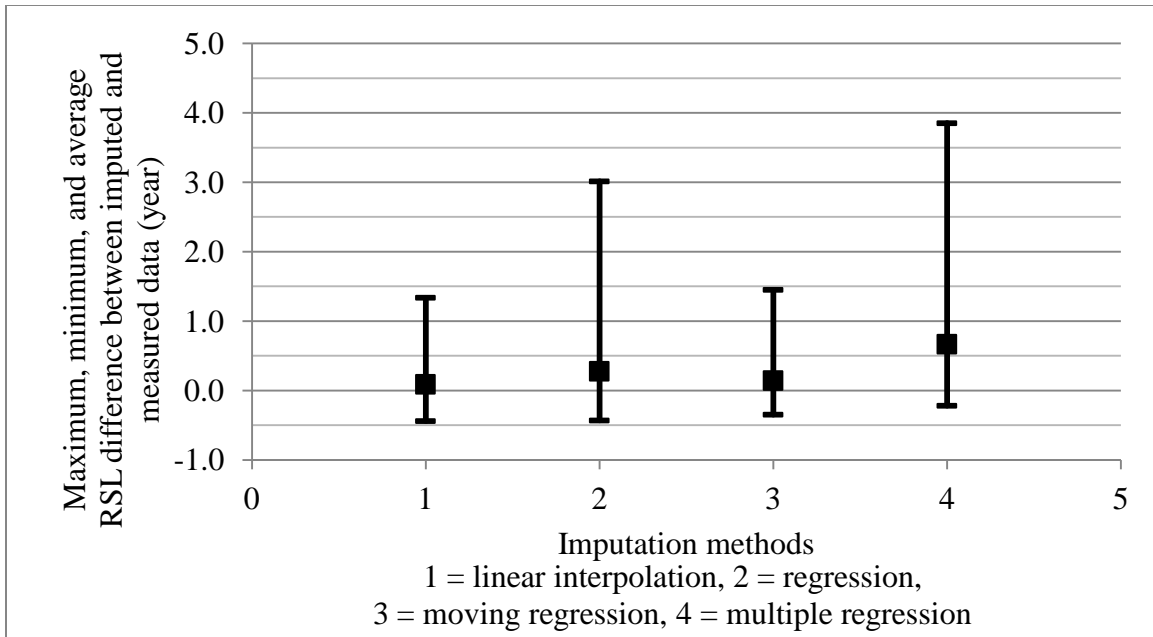


Figure D.29 Maximum, minimum, and average difference between RSL values based on imputed and measured data at cells 1, 2, 3, 5, and 14 at MnROAD, using four imputation methods

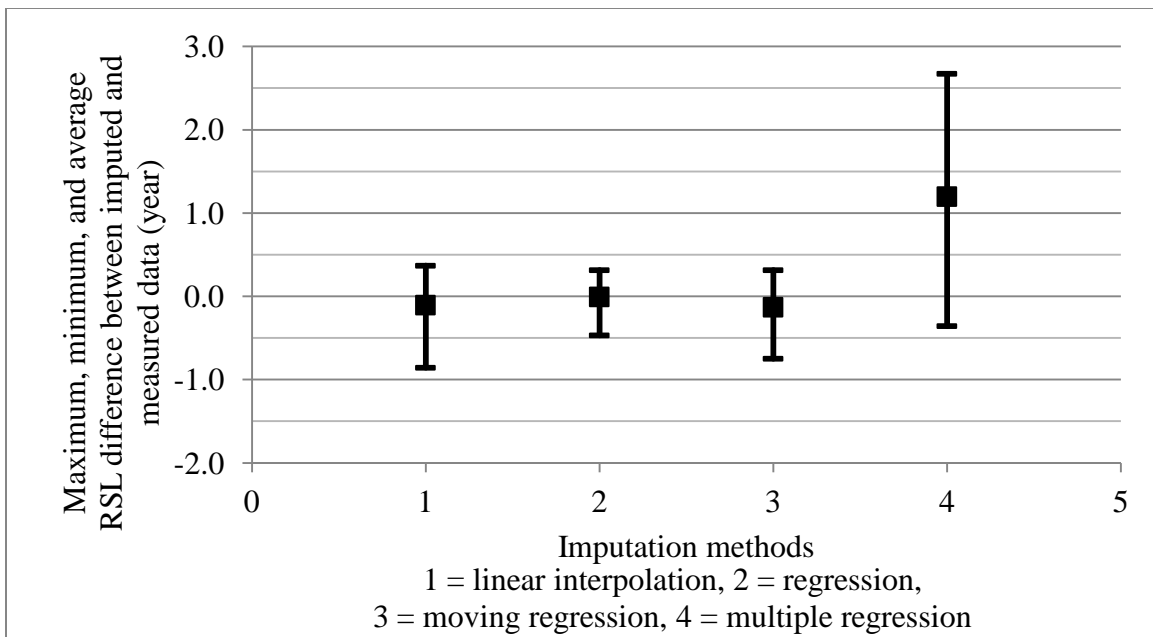


Figure D.30 Maximum, minimum, and average difference between RSL values based on imputed and measured data at BMPs 10.5, 10.6, 11.2, 11.6, and 11.7 along SRID 099 in Washington, using four imputation methods

REFERENCES

REFERENCES

1. American Association of State Highway and Transportation Officials (AASHTO). *AASHTO Guide for Design of Pavement Structures*. Washington, D.C. 1993.
2. American Association of State Highway and Transportation Officials (AASHTO). *Pavement Management Guide*. Washington, D.C. 2001.
3. American Society for Testing and Materials (ASTM). *Standard Guide for Network Level Pavement Management*. ASTM Designation E 1166-00. West Conshohocken, Pennsylvania. 2000.
4. ARA, Inc., ERES Consultants Division. *Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures Final Report. Part 2: Design Inputs*. Prepared for the National Cooperative Highway Research Program (NCHRP). Champaign, Illinois. 2004.
5. Baladi, G.Y., E.C. Novak, Jr., and W.H. Kuo. *Pavement Condition Index – Remaining Service Life*. Pavement Management Implementation: ASTM STP 1121. American Society for Testing and Materials. Atlantic City, New Jersey. 1991.
6. Baladi, G.Y., S.W. Haider, T.A. Dawson, and C.M. Dean. *Optimization of and Maximizing the Benefits from Pavement Management Data Collection: Final Report*. DTFH61-07-R-00137. Department of Civil and Environmental Engineering, Michigan State University. East Lansing, Michigan. 2012.
7. Baladi, G. Y., M. Snyder, and F. McKelvey. *Highway Pavements*. National Highway Institute (NHI). Volumes I, II and III. Washington, D.C. 1992.
8. Baladi, G.Y., S.W. Haider, K. Chatti, L. Galehouse, T. Dawson, C. Dean, R. Muscott, N. Tecca, and M. McCloskey. *Optimization of and Maximizing the Benefits from Pavement Management Data Collection: Interim Report Task 1*. Department of Civil and Environmental Engineering, Michigan State University. East Lansing, Michigan. 2009.
9. Bennett, C.R., A. Chamorro, C. Chen, H. de Solminihac, and G.W. Flintsch. *Data Collection Technologies for Road Management*. The World Bank East Asia Pacific Transport Unit. Washington, D.C. 2007.
10. Bennett, C.R. *Sectioning of Road Data for Pavement*. Proceedings of the 6th International Conference on Management Pavements. Queensland, Australia. 2004.
11. Cafisco, S., A. Di Graziano, H.J. Kerali, and J.B. Odoki. *Multicriteria Analysis for Pavement Maintenance Management*. Transportation Research Record: Journal of the Transportation Research Board. No. 1816. Washington, D.C. 2002.

12. Chien, S., Y. Tang, and P. Schonfeld. *Optimizing Work Zones for Four Lane Highways*. Journal of Transportation Engineering, Volume 127, No.2. 2002.
13. Colorado Department of Transportation (CDOT). *About CDOT*. <http://www.coloradodot.info/about>. Denver, Colorado. 2010.
14. Botelho, F. *National Perspective on Pavement Management*. "Transportation Research News." Issue Number 173. Transportation Research Board. Washington, D.C. 1994.
15. Daniels, G., W.R. Stockton, and R. Hundley. *Estimating Road User Costs Associated with Highway Construction Projects*. Transportation Research Record: Journal of the Transportation Research Board. No. 1732. Washington, D.C. 2000.
16. Dawson, T.A. *Evaluation of Pavement Management Data and Analysis of Treatment Effectiveness Using Multi-Level Treatment Transition Matrices*. Dissertation for the Degree of Ph. D., Department of Civil and Environmental Engineering, Michigan State University. East Lansing, Michigan. 2012.
17. Dawson, T.A. *Optimization of and Maximizing the Benefits from Pavement Management Data Collection, PowerPoint Presentation*. Federal Highway Administration (FHWA) Research Meeting. Washington, D.C. January 24th, 2012.
18. Dean, C.M., G.Y. Baladi, T.A. Dawson, A.C. Beach, S.W. Haider, and K. Chatti. *Impacts of Pavement Condition Analysis Length on Treatment Transition Matrices and Treatment Benefits*. Conference proceedings of the 91st Transportation Research Board Annual Meeting. Washington, D.C. 2012.
19. Dean, C.M., T.A. Dawson, G.Y. Baladi, S.W. Haider, and S. Nadarajah. *Effect of Sampling on Pavement Management Decisions*. Submitted for presentation and publication at the 90th Transportation Research Board Annual Meeting. Washington, D.C. 2011.
20. Deighton, R.A. and D.G. Blake. *Improvement to Utah's Location Referencing System to Allow Data Integration*. Deighton Paper Library. Whitby, Ontario, Canada. 1993.
21. Dudek, C.I., S.H. Richards, J.L. Buffington. *Some Effects of Traffic Control on Four Lane Divided Highways*. Transportation Research Record: Journal of the Transportation Research Board. No. 1086. Washington, D.C. 1986.
22. Elhadi, H.M.A. *GIS, A Tool for Pavement Management*. Royal Institute of Technology (KTH). Stockholm, Sweden. 2009.
23. Farhan, J. and T.F. Fwa. *Pavement Performance Data Quality Assurance: Analysis on Missing Data Imputation*. Proceedings of the Transportation Research Board 91st Annual Meeting. Paper No.12-1004. Washington, D.C. 2012.

24. Federal Highway Administration (FHWA). *Distress Identification Manual for the Long-Term Pavement Performance Program*. United States Department of Transportation. Washington, D.C. 2003.
25. Federal Highway Administration (FHWA). *Economic Analysis Primer: Life Cycle Cost Analysis*. United States Department of Transportation. Washington, D.C. 2011.
26. Federal Highway Administration (FHWA). *Highway Performance Monitoring System Field Manual for the Continuing Analytical and Statistical Database*. United States Department of Transportation Office of Highway Policy Information. Washington, D.C. 2005.
27. Federal Highway Administration (FHWA). *Pavement Management Primer*. United States Department of Transportation. Washington, D.C. 2011.
28. Federal Highway Administration (FHWA). *Pavement and Road Surface Management for Local Agencies*. Participant's Manual. United States Department of Transportation. Washington, D.C. 1995.
29. Fillastre, C. *Louisiana Department of Transportation and Development (LADOTD) Research Meeting with Michigan State University*. Meeting conducted by Gilbert Y. Baladi. Baton Rouge, Louisiana. 2011.
30. Finn, F. *Pavement Management Systems – Past, Present, and Future, Public Roads: 80 Years Old, But the Best Is Yet to Come*. Public Roads Publication Volume 62, Number 1. United States Department of Transportation. Washington, D.C. 1998.
31. Flintsch, G. and K. K. McGhee. *NCHRP Synthesis 401: Quality Management of Pavement Condition Data Collection*. National Cooperative Highway Research Program (NCHRP). Washington, D.C. 2009.
32. Flintsch, G., R. Dymond, and J. Collura. *NCHRP Synthesis 335: Pavement Management Applications Using Geographic Information Systems*. National Cooperative Highway Research Program (NCHRP). Washington, D.C. 2004.
33. Foote, K.E. and D.J. Huebner. *Error, Accuracy, and Precision*. University of Texas at Austin. Department of Geography. Austin, Texas. 1995.
34. Freeman, T.J. and J.E. Ragsdale. *Development of Certification Equipment for TxDOT Automated Pavement Distress Equipment*. FHWA/TX-03/4204-1. United States Department of Transportation Federal Highway Administration. Washington, D.C. 2003.
35. Fwa, T.F. *The Handbook of Highway Engineering*. CRC Press, Taylor & Francis Group. Boca Raton, Florida. 2006.
36. Galehouse, L. *SHA Survey*. National Center for Pavement Preservation (NCPPI). East Lansing, Michigan. 2010.

37. Garmin Ltd. *What is GPS? How it Works*. www.garmin.com/aboutGPS/. Accessed February 3, 2012.
38. Graham, J.W., P.E. Cumsille, and E. Elek-Fisk. *Methods for Handling Missing Data*. Research Methods in Psychology. Handbook of Psychology Volume 2. p 87-114. John Wiley and Sons. New York, New York. 2003.
39. Haider, S.W., G.Y. Baladi, K. Chatti, and C.M. Dean. *Effect of Pavement Condition Data Collection Frequency on Performance Prediction*. Transportation Research Record: Journal of the Transportation Research Board. Pavement Management 2010 Volume 1. p 67-80. Washington, D.C. 2010.
40. Haider, S.W., H.K. Salama, N. Buch, and K. Chatti. *Influence of Traffic Inputs on Rigid Pavement Performance Using M-E PDG in the State of Michigan*. Fifth International Conference on Maintenance and Rehabilitation of Pavements and Technological Control. p 85-90. Park City, Utah. 2007.
41. Hall, K.T., C.E. Correa, S.H. Carpenter, and R.P. Elliot. *Guidelines for Life Cycle Cost Analysis of Pavement Rehabilitation Strategies*. Conference proceedings of the 82nd Transportation Research Board Annual Meeting. Washington, D.C. 2003.
42. Hartgen, D.T., R.K. Karanam, and M.G. Fields. *18th Annual Report on the Performance of State Highway Systems 1984-2007/8*. Reason Foundation (RF). Los Angeles, California. 2009.
43. Holt, F.B. and W.L. Gramling, eds. *Pavement Management Implementation*. American Society for Testing and Materials. STP 1121. Philadelphia, Pennsylvania. 1991.
44. Hudson, W.R, R. Haas, and R.D. Pedigo. *NCHRP Report 215: Pavement Management System Development*. National Cooperative Highway Research Program (NCHRP). Washington, D.C. 1979.
45. Iowa Department of Transportation (Iowa DOT). *Location Referencing System Team Report*. Ames, Iowa. 1998.
46. Lenz, R.W. *Pavement Design Guide Manual*. Texas Department of Transportation (TxDOT) Construction Division, Materials and Pavements Section. Austin, Texas. 2011.
47. Lewis, D.L. *NCHRP Synthesis 269: Road User and Mitigation Costs in Highway Pavement Projects*. National Cooperative Highway Research Program (NCHRP). Washington, D.C. 1999.
48. Little, R.A. and D.B. Rubin. *Statistical Analysis with Missing Data*. J. Wiley and Sons. New York, New York. 1987.

49. Louisiana Department of Transportation and Development (LADOTD). *2006 Annual Report*. Baton Rouge, Louisiana. 2007.
50. Lytton, R.L. *Concepts of Pavement Performance Prediction and Modeling*. Proceedings of the Second North American Pavement Management Conference. Toronto, Canada. 1987.
51. McGhee, K.H. *NCHRP Synthesis 334: Automated Pavement Distress Collection Techniques*. National Cooperative Highway Research Program (NCHRP). Washington, D.C. 2004.
52. McNerney, M.T. and T. Rioux. *Geographic Information System (GIS) Needs Assessment for TxDOT Pavement Management Information Systems*. University of Texas at Austin Center for Transportation Research. Research Report No. 0-1747-2. Austin, Texas. 2000.
53. Michigan Department of Transportation (MDOT). *A Citizen's Guide to MDOT*. Lansing, Michigan. 2010.
54. Minnesota Department of Transportation (MnDOT). *MnROAD Research – Improving the Way We Construct and Maintain Our Highways in Cold Weather Climates*. <http://www.dot.state.mn.us/mnroad/index.html>. Minneapolis, Minnesota. 2011.
55. Minnesota Road Research Project (MnROAD). *National Research and Technology Center, Providing Research Insight for a Safe, Efficient, and Cost-Effective Transportation System*. The MnROAD Test Facility. Monticello, Minnesota. 2008.
56. Morgado, J. and J. Neves. *Accounting for User Costs When Planning Pavement Maintenance and Rehabilitation Activities*. Technical University of Lisbon. Lisbon, Portugal. 2009.
57. Naiel, A.K.A. *Flexible Pavement Rut Depth Modeling for Different Climate Zones*. Dissertation for the Degree of Ph. D., Department of Civil Engineering, Wayne State University. Detroit, Michigan. 2010.
58. Nostrand, W. A. *The History of Pavement Management in the Federal Highway Administration*” *Federal Highway Administration (FHWA)*. Special Technical Publication No. 1121. Philadelphia, Pennsylvania. 1992.
59. Odermatt, N., V. Janoo, and R. Magnusson. *Analysis of Permanent Deformation in Subgrade Materials Using a Heavy Vehicle Simulator*. Proceedings of the International Conference in Accelerated Pavement Testing. Reno, Nevada. 1999.
60. Pavement Interactive. *Project Level Approach to Pavement Management*. www.pavementinteractive.org. July 23, 2008.
61. Peterson, D.E. *NCHRP Synthesis 135: Pavement Management Practices*. National Cooperative Highway Research Program (NCHRP). Washington, D.C. 1987.

62. Reigle, J.A. *Development of an Integrated Project-Level Pavement Management Model Using Risk Analysis*. Dissertation for the Degree of Ph. D., Department of Civil and Environmental Engineering, West Virginia University. Morgantown, West Virginia. 2000.
63. Robertson, N.F., D.P. Latimer, and A. Nata-Atmadja. *A Study of the Impact of Road Segmentation Schemas on Predicted Maintenance Investment Outcomes*. Sixth International Conference on Managing Pavements. Brisbane, Australia. 2004.
64. Smith, K.L., L. Titus-Glover, L.D. Evans. *Pavement Smoothness Index Relationships, Final Report*. ERES Division of Applied Research Associates, Inc. Report No. FHWA-RD-02-057. Washington, D.C. 2002.
65. Smith, R.E., C.M. Change-Albitres. *The Impact of Semi-Automated Distress Collection Methods on Pavement Management Network-Level Analysis Using the MTC Streetsaver*. Texas Transportation Institute Project 476290. College Station, Texas. August 2007.
66. Smith, R. E., T. J. Freeman, and O. J. Pendleton. *Evaluation of Automated Pavement Distress Data Collection Procedures for Local Agency Pavement Management*. Texas Transportation Institute (TTI). Texas A & M University. College Station, Texas. 1998.
67. Titus-Glover, L., C. Fang, M. Alam, K. O'Toole, and M.I. Darter. *Pavement Health Track (PHT), Remaining Service Life (RSL) Forecasting Models, Technical Information*. Federal Highway Administration (FHWA) Office of Asset Management. Washington, D.C. 2010.
68. Tremblay, G., M. Grondin, D. Leroux, and J. Carrier. *Efficient Way to Measure Pavement Distress Manually: Ministère des Transports du Quebec Experience*. Sixth International Conference on Managing Pavements. Brisbane, Australia. 2004.
69. Zhang, L. and O. Smadi. *What is Missing in Quality Control of Contracted Pavement Distress Data Collection?* Proceedings of the 88th Annual Transportation Research Board (TRB) Meeting. Paper #09-0062. Washington, D.C. 2009.
70. Zhang, S., K. Liao, and X. Zhu. *A SAS Macro for Single Imputation*. Merck and Co. Inc. Upper Gwynedd, Pennsylvania. 2008.
71. Zimmerman, K.A. *NCHRP Synthesis 222: Pavement Management Methodologies to Select Projects and Recommend Preservation Treatment*. National Cooperative Highway Research Program (NCHRP). Washington, D.C. 1995.
72. Zimmerman, K.A. and D. G. Peshkin. *Supporting Preventive Maintenance Programs with Pavement Management*. Proceedings of the 6th International Conference on Management Pavements. Queensland, Australia. 2004.
73. Wang, K.C.P. and O. Smadi. *Automated Imaging Technologies for Pavement Distress Surveys*. Pavement Monitoring and Evaluation Committee, Transportation Research Circular E-C156. Transportation Research Board. Washington, D.C. 2011.

74. Wang, K.C.P. and W. Gong. *Automated Pavement Distress Survey: A Review and A New Direction*. Department of Civil Engineering, University of Arkansas. Fayetteville, Arkansas. 2002.
75. Washington State Department of Transportation (WSDOT). *WSDOT Pavement Guide – Pavement Types*. Olympia, Washington. 2010.
76. Washington State Department of Transportation (WSDOT). *Training Module 4: Pavement Management Systems*. www.training.ce.washington.edu/WSDOT. Accessed May 31st, 2012.
77. Wayman, J.C. *Multiple Imputation for Missing Data: What Is It and How Can I Use It?* Proceedings of the Annual Meeting of the American Education Research Association. Chicago, Illinois. 2003.
78. Yang, J. *Road Crack Condition Performance Modeling Using Recurrent Markov Chains and Artificial Neural Networks*. Dissertation for the Degree of Ph. D., Department of Civil and Environmental Engineering, University of South Florida. Tampa, Florida. 2004.