UMTRI-94-23-2

ANALYSIS OF OPERATIONS AND SAFETY OF Y-INTERSECTIONS

VOLUME II: APPENDICES

Lidia P. Kostyniuk

and

Dawn L. Massie

The University of Michigan
Transportation Research Institute
Ann Arbor, Michigan 48109-2150

November 1994

FINAL REPORT

Contract Number
MDOT 93-0863

Prepared for the
Michigan Department of Transportation
P.O. Box 30050
Lansing, Michigan 48909

UMTRI The University of Michigan Transportation Research Institute

This work was supported by the Michigan Department of Transportation. The opinions, findings, and conclusions expressed in this publication are those of the authors and do not necessarily reflect the views of the Michigan Department of Transportation.

Technical Report Documentation Page

1. Report No. UMTRI-94-23-2	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Analysis of Operations and Safety of Y-Intersections Volume II: Appendices		5. Report Date November 23, 1994 6. Performing Organization Code
7. Author(s) Lidia P. Kostyniuk, Dawn L. Massie		8. Performing Organization Report No. UMTRI-94-23-2
9. Performing Organization Name and Address University of Michigan Transportation Research Institute 2901 Baxter Road Ann Arbor, Michigan 48109-2150		10. Work Unit No. 11. Contract or Grant No. MDOT 93-0863
12. Sponsoring Agency Name and Address Michigan Department of Transportation P.O. Box 30050 Lansing, Michigan 48909		13. Type of Report and Period Covered Final report 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Accident experience at unsignalized, two-way intersections, where the major road curves to the right and the side road continues on a tangent to the curve, was examined with the goal of determining whether this type of intersection has accident problems more severe than other types of three-legged intersections.

The study included a review of case studies of severe accidents at these sites, analysis of data from a Y-intersection improvement program from Washtenaw County, Michigan, and analysis of accident data from three-legged intersections on the Michigan state trunkline. Field observations were made at 53 sites. A survey was conducted of state Departments of Transportation concerning their experiences with this type of intersection.

The analyses show that this special type of Y-intersection does not pose a unique risk relative to other three-legged intersections. Its accident patterns are similar to that of all curved Y-intersections. The criteria used for selecting curved Y-sites for safety improvements are sufficient for identifying problem sites among the special Y-intersections also.

17. Key Words Y-intersection, curve, three-legged, accident rates	18. Distribution Statement		
19. Security Classif. (of this report) None	20. Security Classif. (of this page) None	21. No. of Pages 182	22. Price

Acknowledgments

We gratefully acknowledge the help of the Washtenaw County Road Commission in providing us with the original Y-intersection data file and in giving us access to their accident and volume information.

We wish to thank several members of the Traffic and Safety Division of the Michigan Department of Transportation (MDOT). Dr. Dale Lighthizer served as the technical project manager and facilitated the progression of this research. Brett Kach sorted, matched, and merged the various MDOT data files that we used to develop the final analysis sample. Mary McClain provided us with the more recent accident records for the Washtenaw County Y sites.

Many members of the University of Michigan Transportation Research Institute staff contributed to this project. Dr. Paul E. Green developed the method of Bayesian estimation of rates, used in the statistical modelling of this research, as part of his Ph.D. dissertation. He also performed the statistical tests required in the analyses. Cecil Lockard and Raymond Masters performed the field studies, and Professor Leland Quackenbush carried out the interviews of the State Departments of Transportation. Dr. Kenneth Campbell served as project director and offered his insights throughout the project. We gratefully acknowledge their contributions.

CONTENTS

APPENDIX A - WASHTENAW COUNTY MATERIAL

Table A-1 - List of Washtenaw County Y Sites

	TOWNSHIP	MAJOR ROAD	MINOR ROAD
1	Saline	Macon	Mooreville
2		Jordan	Arkona
3		Goodrich	Hack
4		Jordan	Macon
5		Goodrich	Arkona
6	York	Stony Creek	Willow
7		Petersburg	Day
8		Saline	Jewell
9	Manchester	Lemm	Schleeweis
10		Fahey	Lemm
11		Burtless	Henzie
12	Sharon	Heim	Hayes
13		Grass Lake	Sharon Hollow
14		Pleasant Lake	Grass Lake
15		Sharon Hollow	Easudes
16	Freedom	Bethel Church	Eisman
17		Waters	Loeffler
18		Waters	Lima Center
19		Bethel Church	Esch
20	Superior	Prospect	Frains Lake
21	Ann Arbor	Woodland	Blakeway
22	Bridgewater	Clinton	Fisk
23		Clinton	Allen
24		Austin	Eisman
25		Lima Center	Hoelzer
26		Burmeister	Schellenberger
27		Hogan	Wilber
28		Clinton	Hoelzer
29	Northfield	Whitmore Lake	Kearney
30		7 Mile	Earhart
31		7 Mile	Donna Lane
32	Scio	Huron River Dr.	Maple
33		Daleview	Bylington
34		Huron River Dr.	Tubbs
35	Dexter	Dexter-Pinckney	Wylie
36	Lima	Dexter-Chelsea	Wylie
37		Fletcher	Sager
38		Dancer	Jerusalem
39		Jerusalem	Guenther
40	Lyndon	Waterloo	Lingane
41		Farnsworth	Jaycox
42	Webster	Mast	Daly
43	Salem	Angle	Tower
44		6 Mile	Dixboro
45		7 Mile	Tower

Table A-2 - Example of Washtenaw County Site Data Form

WASHTENAW Y INTERSECTIONS

CASE NUMBER 6
TOWNSHIP York
NAME Stony Creek/Willow

ENVIRONMENT Rural
TYPE OF INTERSECTION Regular Y

ISLAND PRESENT <u>YES</u> NO

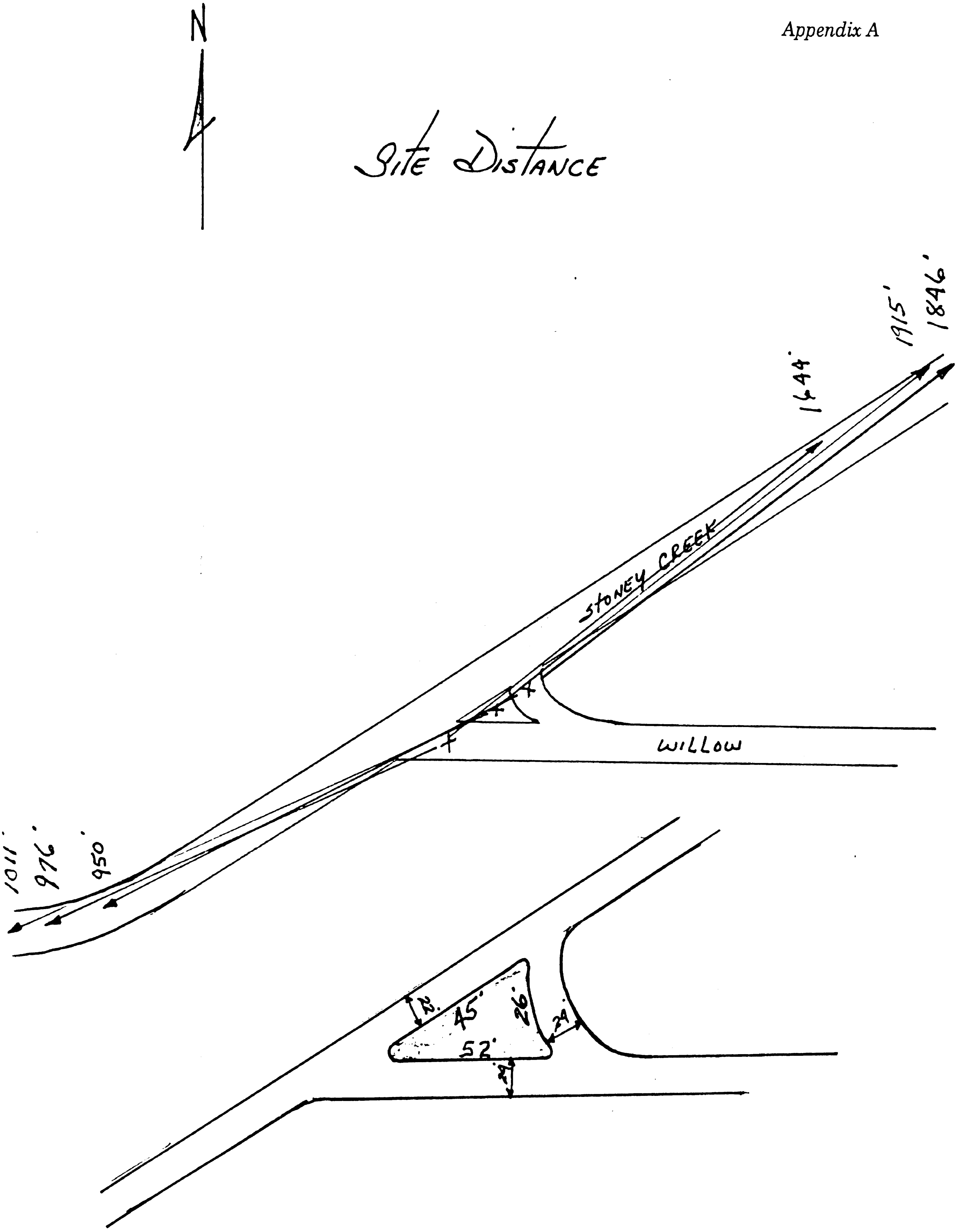
IF YES, SIZE ~ 600 ft.

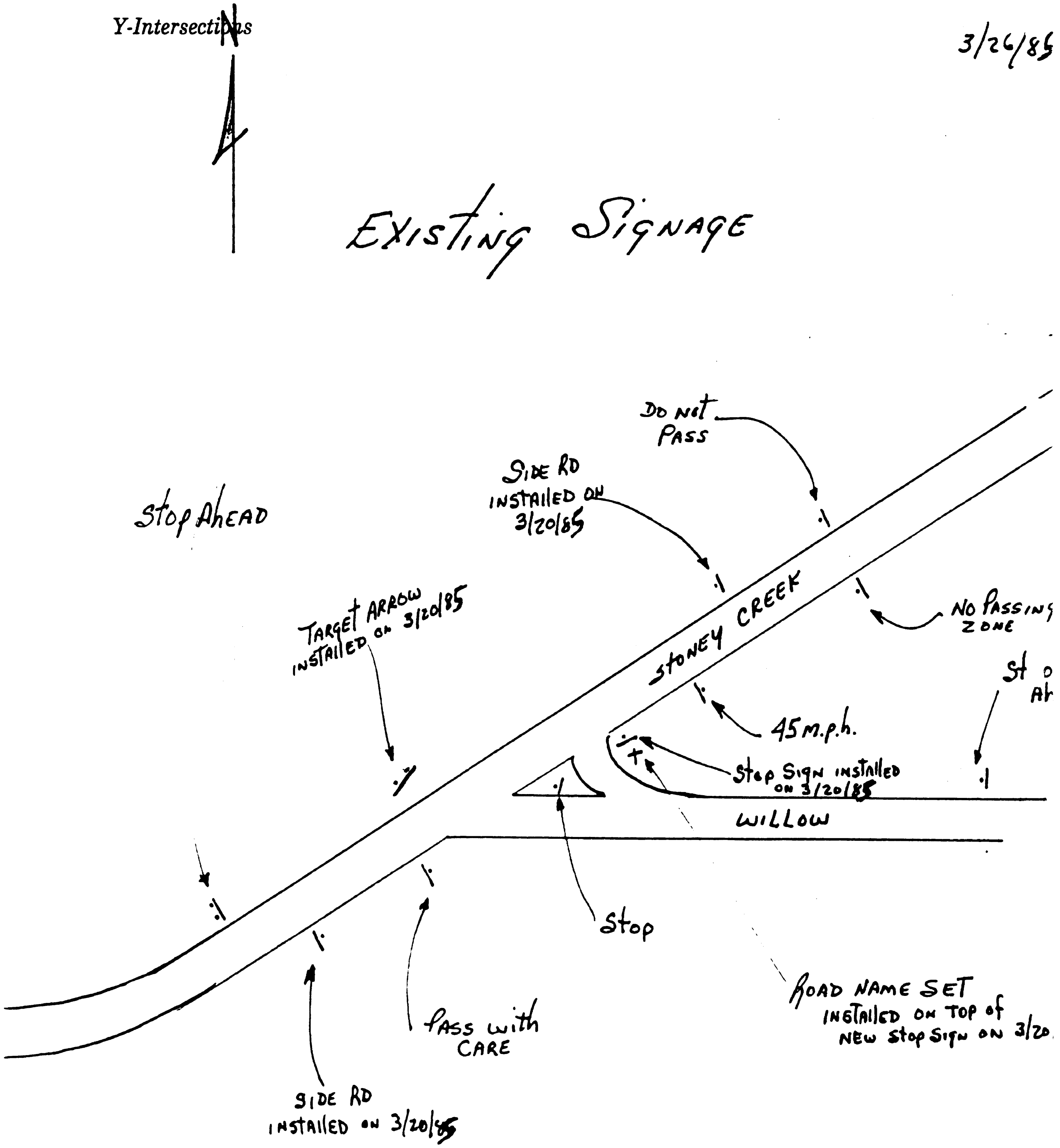
MAJOR ROAD Stony Creek <u>PAVED</u> GRAVEL
SIGNS No passing, 45 mph
VOLUME 2400 vpd
VOLUME CHANGES OVER TIME <u>NOT MUCH CHANGE</u>
SMALL INCREASE
LARGE INCREASE
SMALL DECREASE
LARGE DECREASE

MINOR ROAD Willow PAVED <u>GRAVEL</u>
SIGNS Stop
VOLUME 300 vpd
VOLUME CHANGES OVER TIME <u>NOT MUCH CHANGE</u>
SMALL INCREASE
LARGE INCREASE
SMALL DECREASE
LARGE DECREASE

ACCIDENTS	TYPE	SEVERITY
1980 - 0		
1981 - 1	Fixed Object	Injury
1982 - 0		
1983 - 0		
1984 - 0		
1986 - 0		
1987 - 0		
1988 - 0		
1989 - 0		
1990 - 0		

CHANGES MADE
Target arrow installed, side road advanced warning signs installed
Road name set installed

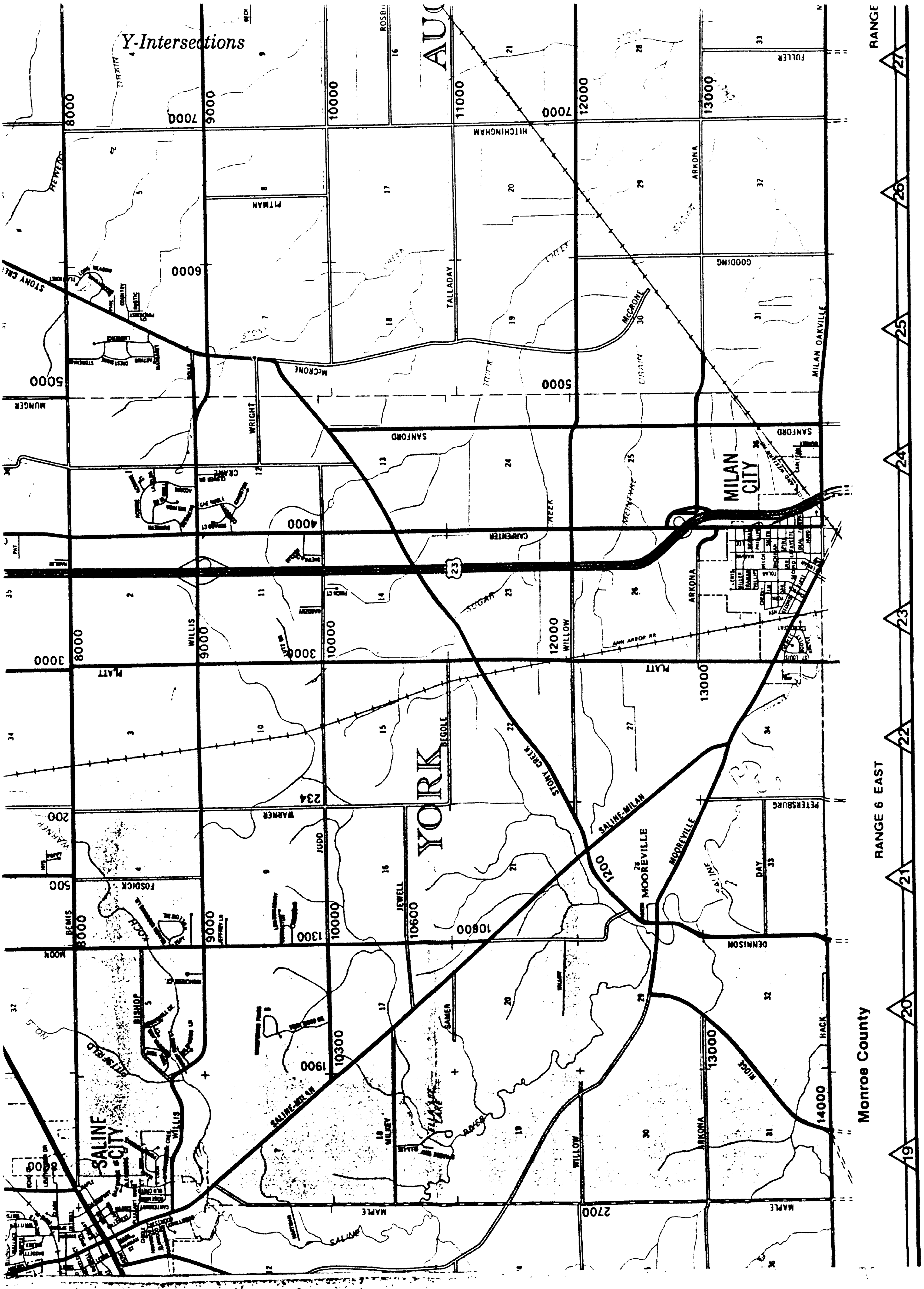

N
Site Distance
1915'
1846'
1644'
STONEY CREEK
WILLOW
1011
976'
950'
22'
45'
26'
24
52'
24


N
3/26/85
Existing Signage
Do not Pass
Side RD installed on 3/20/85
No Passing Zone
Stop Ahead
Target Arrow installed on 3/20/85
Stoney Creek
St o
Ah
45 m.p.h.
Stop Sign installed on 3/20/85
Willow
Stop
Road Name Set installed on top of New Stop Sign on 3/20
Pass with Care
Side RD installed on 3/20/85

'ORK

NAME	1986	1987	1988	1989	1990	1991	1992	1993
SALINE MILAN RD S OF SALINE CITY LIMITS	4505	-	5078	-	3960	-	3658	-
SALINE MILAN RD S OF STONY CREEK RD	3189	-	4354	-	4090	-	2654	-
SANFORD RD S OF STONY CREEK RD	583	-	631	-	481	-	625	-
SANFORD RD (NB) S OF OAKVILLE MILAN RD	731	-	-	-	-	-	-	-
SANFORD RD (SB) N OF OAKVILLE MILAN RD	532	-	-	-	-	-	-	-
STONY CREEK RD E OF CARPENTER RD	2248	-	3470	-	2399	-	2500	-
STONY CREEK RD E OF SALINE MILAN RD	2348	-	-	-	-	-	-	-
STONY CREEK RD W OF CARPENTER RD	1821	-	2559	-	2488	-	2733	-
STONY CREEK RD (EB) W OF CARPENTER RD	-	-	1284	-	-	-	-	-
STONY CREEK RD (EB) W OF PLATT RD	-	-	-	-	-	-	1080	-
STONY CREEK RD (WB) E OF CARPENTER RD	-	-	1461	-	-	-	-	-
STONY CREEK RD (WB) E OF PLATT RD	-	-	-	-	-	-	1202	-
WARNER RD S OF JUDD RD	-	-	-	161	-	-	-	-
WILLIS RD E OF CARPENTER RD	4957	-	5698	-	7039	-	5313	-
WILLIS RD E OF SALINE CITY LIMITS	1888	-	2450	-	1963	-	2008	-
WILLIS RD E OF MOON RD	-	-	-	-	-	-	2646	-
WILLIS RD W OF MOON RD	-	-	-	-	-	-	1981	-
WILLIS RD W OF PLATT RD R X R CROSSING	-	-	-	2338	-	-	-	-
WILLIS RD (EB) W OF PLATT RD	1005	795	-	-	-	-	-	-
WILLIS RD (WB) E OF PLATT RD	1895	2200	-	-	-	-	-	-
WILLOW RD E OF CARPENTER RD	1481	-	2203	-	2486	-	1617	-
WILLOW RD W OF CARPENTER RD	350	-	505	-	481	-	299	-

Y-Intersections



APPENDIX B - MDOT TRUNKLINE FILE MATERIAL

Table B-1 - Variable List for UMTRI Version of MDOT Trunkline Data File

Variable Number	SAS Name	Description	Type	NDEC	Length (bytes)	Position
1	CELL	Sampling Stratum	Num	0	3	0
2	TNAME	Trunkline English Name	Char	0	12	3
3	XNAME	Crossroad English Name	Char	0	18	15
4	BMP	Beginning Milepoint	Num	3	8	33
5	EMP	Ending Milepoint	Num	3	8	41
6	POPNUM	Population Code	Num	0	3	49
7	RD	Development Code	Num	0	3	52
8	PAVE	Pavement Type Code	Num	0	3	55
9	GRADE	Grade	Num	0	3	58
10	TERRAIN	Terrain	Num	0	3	61
11	CURVE	Curve Type Code	Num	0	3	64
12	DEGREE	Degree of Curve	Num	2	8	67
13	DISTRICT	District	Num	0	3	75
14	CONSECT	Control Section	Num	0	4	78
15	MILEPT	Milepoint	Num	3	8	82
16	LANES	Number of Lanes	Num	0	3	90
17	LANES_R	Extra Lanes - Right	Num	0	3	93
18	LANES_L	Extra Lanes - Left	Num	0	3	96
19	PARK_R	On-Street Parking - Right	Num	0	3	99
20	PARK_L	On-Street Parking - Left	Num	0	3	102
21	LANEWID	Lane Width - Plus	Num	0	3	105
22	SWPR	Shoulder Width - Plus Right	Num	0	3	108
23	SWPL	Shoulder Width - Plus Left	Num	0	3	111
24	NOPASS	No Passing Zone	Num	0	3	114
25	LAND_USE	Roadside Development	Num	0	3	117
26	SPEEDLIM	Speed Limit	Num	0	3	120
27	INT_TYPE	Intersection Type	Num	0	3	123
28	SIGNAL	Signal Control Type	Num	0	3	126
29	ADT	Average Daily Traffic	Num	0	4	129
30	TOT	Total Accidents	Num	0	3	133
31	INJURY	Injury Accidents	Num	0	3	136
32	FATAL	Fatal Accidents	Num	0	3	139
33	WET	Wet Accidents	Num	0	3	142
34	ICY	Icy Accidents	Num	0	3	145
35	DARK	Dark Accidents	Num	0	3	148
36	MISC_SV	Miscellaneous Single Vehicle	Num	0	3	151
37	ROLLOVER	Rollover	Num	0	3	154
38	TRAIN	Hit Train	Num	0	3	157
39	PARKVEH	Hit Parked Vehicle	Num	0	3	160
40	BACKING	Backing	Num	0	3	163
41	PARKING	Parking	Num	0	3	166
42	PED	Pedestrian	Num	0	3	169
43	FOBJ	Fixed Object	Num	0	3	172
44	OOBJ	Other Object	Num	0	3	175
45	ANIMAL	Animal	Num	0	3	178
46	BIKE	Bicycle	Num	0	3	181
47	HDON	Head-on	Num	0	3	184

48	ANGST	Angle Straight	Num	0	3	187
49	RE	Rear-end	Num	0	3	190
50	ANGTN	Angle Turn	Num	0	3	193
51	SSPASS	Sideswipe - Passing	Num	0	3	196
52	RELT	Rear-end Left Turn	Num	0	3	199
53	RERT	Rear-end Right Turn	Num	0	3	202
54	OTH_DR	Other Driveway	Num	0	3	205
55	ANGDR	Angle at Driveway	Num	0	3	208
56	REDR	Rear-end at Driveway	Num	0	3	211
57	SSMEET	Sideswipe - Meeting	Num	0	3	214
58	HDONLT	Head-on Left Turn	Num	0	3	217
59	DUALLT	Dual Left Turn	Num	0	3	220
60	DUALRT	Dual Right Turn	Num	0	3	223
61	CONSECMP	Control Section Milepoint	Num	3	8	226
62	HWAYTYPE	Highway Area Type	Num	0	3	234
63	HWAYCODE	Highway Area Code	Num	0	3	237
64	WEEKDAY	Weekday	Num	0	3	240
65	HOUR	Hour of Occurrence	Num	0	3	243
66	MONTH	Month of Accident	Num	0	3	246
67	DATE	Day of Month	Num	0	3	249
68	YEAR	Year of Accident	Num	0	3	252
69	WEATHER	Weather Condition	Num	0	3	255
70	LIGHTING	Lighting	Num	0	3	258
71	SURFACE	Road Surface Condition	Num	0	3	261
72	DEFECT	Road Defect	Num	0	3	264
73	A_INJS	A Injuries	Num	0	3	267
74	B_INJS	B Injuries	Num	0	3	270
75	C_INJS	C Injuries	Num	0	3	273
76	ALIGN	Road Alignment	Num	0	3	276
77	SPECTAGS	Special Accident Tags	Num	0	3	279
78	ACC_TYPE	Accident Type	Num	0	3	282
79	NUM_VEHS	Number of Vehicles	Num	0	3	285
80	DISTXRD	Distance from Crossroad	Num	0	4	288
81	DIRXRD	Direction from Crossroad	Num	0	3	292
82	NUMUNINJ	Number of Persons Uninjured	Num	0	3	295
83	VTSUB1	Vehicle Type Subscript, Veh 1	Num	0	3	298
84	DRINT1	Driver Intent, Driver 1	Num	0	3	301
85	VIOL1	Violation, Driver 1	Num	0	3	304
86	CONCIR1	Contributing Circumstance, Veh 1	Num	0	3	307
87	DIRTRAV1	Direction of Travel, Veh 1	Num	0	3	310
88	OBHIT1	Object Hit, Veh 1	Num	0	3	313
89	IMPCODE1	Impact Code, Veh 1	Num	0	3	316
90	VTSUB2	Vehicle Type Subscript, Veh 2	Num	0	3	319
91	DRINT2	Driver Intent, Driver 2	Num	0	3	322
92	VIOL2	Violation, Driver 2	Num	0	3	325
93	CONCIR2	Contributing Circumstance, Veh 2	Num	0	3	328
94	DIRTRAV2	Direction of Travel, Veh 2	Num	0	3	331
95	OBHIT2	Object Hit, Veh 2	Num	0	3	334
96	IMPCODE2	Impact Code, Veh 2	Num	0	3	337
97	VTSUB3	Vehicle Type Subscript, Veh 3	Num	0	3	340
98	DRINT3	Driver Intent, Driver 3	Num	0	3	343
99	VIOL3	Violation, Driver 3	Num	0	3	346
100	CONCIR3	Contributing Circumstance, Veh 3	Num	0	3	349
101	DIRTRAV3	Direction of Travel, Veh 3	Num	0	3	352

102	OBHIT3	Object Hit, Veh 3	Num	0	3	355
103	IMPCODE3	Impact Code, Veh 3	Num	0	3	358
104	NUM_KILL	Number of Persons Killed	Num	0	3	361
105	NUM_INJ	Number of Persons Injured	Num	0	3	364
106	ACCSEV	Accident Severity	Num	0	3	367
107	ARNUM	Accident Report Number	Num	0	5	370
108	CURVECAT	Curve Category	Num	0	3	375
109	SPECIAL	Special Y Flag	Num	0	3	378
110	SIDE_ADT	Side Road Avg. Daily Traffic	Num	0	4	381
111	SIDEPAVE	Side Road Pavement Type	Num	0	3	385
112	CONTRAST	Contrast Between Roads	Num	0	3	388

Table B-2 - Codebook for UMTRI Version of MDOT Trunkline Data File

Variable 1	CELL	Acc Freq	Percent	CELL	Site Freq	Percent
Curve/Rural/Right/T	1	294	3.0	1	131	6.3
Curve/Rural/Right/Y	2	693	7.0	2	227	11.0
Curve/Rural/Left/T	3	373	3.8	3	131	6.3
Curve/Rural/Left/Y	4	680	6.9	4	223	10.8
Curve/Urban/Right/T	5	857	8.7	5	131	6.3
Curve/Urban/Right/Y	6	417	4.2	6	74	3.6
Curve/Urban/Left/T	7	998	10.1	7	128	6.2
Curve/Urban/Left/Y	8	454	4.6	8	78	3.8
Tangent/Rural/T	9	638	6.5	9	250	12.1
Tangent/Rural/Y	10	835	8.5	10	250	12.1
Tangent/Urban/T	11	1978	20.1	11	250	12.1
Tangent/Urban/Y	12	1648	16.7	12	194	9.4

Variable 2 TNAME (alpha)

Variable 3 XNAME (alpha)

Variable 4	BMP	Acc Freq	Percent	BMP	Site Freq	Percent
	0	130	1.3	0	24	1.2
	0.01	19	0.2	0.01	2	0.1
		.			.	
		.			.	
		.			.	
	33.381	1	0.0	33.381	1	0.0

Variable 5	EMP	Acc Freq	Percent	EMP	Site Freq	Percent
	0.01	1	0.0	0.01	1	0.0
		.			.	
		.			.	
		.			.	
	33.965	1	0.0	33.965	1	0.0

Variable 6	POPNUM	Acc Freq	Percent	POPNUM	Site Freq	Percent
50,000 and over	1	622	6.4	1	47	2.3
40,000-49,999	2	62	0.6	2	9	0.4
25,000-39,999	3	131	1.3	3	12	0.6
5,000-24,999	4	2334	23.7	4	185	9.0
Under 5,000/incorporated	5	1748	17.7	5	339	16.4
Unincorporated	9	4968	50.4	9	1475	71.4

Variable 7	RD	Acc Freq	Percent	RD	Site Freq	Percent
Rural	1	3101	31.4	1	1173	56.7
Rural dense/small cities	2	2193	22.2	2	474	22.9
Small city boundaries	3	4	0.0	3	3	0.1
Rural in character	4	270	2.7	4	44	2.1
Residential	5	1576	16.0	5	170	8.2
Outlying business dist.	6	2399	24.3	6	184	8.9
Fringe area	7	218	2.2	7	14	0.7
Central business dist.	8	104	1.1	8	5	0.2

NOTE: Values 1-3 defined as "rural and under 5000 population."
Values 4-8 defined as "urban and over 5000 population."

Variable 8	PAVE	Acc Freq	Percent	PAVE	Site Freq	Percent
Surf. trtmt over bitum on flex base	1	87	0.9	1	44	2.1
Surf. trtmt over bitum on rigid base	2	47	0.5	2	8	0.4
Bitum. over flex base	4	2247	22.8	4	838	40.5
Bitum. over concrete/brick	5	6715	68.1	5	1056	51.1
Concrete (jointed)	7	769	7.8	7	121	5.9

Variable 9	GRADE	Acc Freq	Percent	GRADE	Site Freq	Percent
No	0	8796	89.2	0	1704	82.4
Yes	1	1069	10.8	1	363	17.6

Variable 10	TERRAIN	Acc Freq	Percent	TERRAIN	Site Freq	Percent
Level	1	7247	73.5	1	1220	59.0
Rolling	2	2618	26.5	2	847	41.0

Variable 11	CURVE	Acc Freq	Percent	CURVE	Site Freq	Percent
Right Curve	1	2413	24.5	1	592	28.6
Left Curve	2	2604	26.4	2	579	28.0
North/East	3	2007	20.3	3	416	20.1
North/West	4	914	9.3	4	200	9.7
South/East	5	1782	18.1	5	246	11.9
South/West	6	145	1.5	6	34	1.6

Variable 12	DEGREE	Acc Freq	Percent	DEGREE	Site Freq	Percent
	0	128	2.6	0	24	2.0
	0.02	23	0.5	0.02	2	0.2
		.			.	
		.			.	
		.			.	
	90.27	3	0.1	90.27	1	0.1

Frequency Missing = 4848 Frequency Missing = 896
(only missing for tangent intersections)

Y-Intersections

Variable 13

	DISTRICT	Acc Freq	Percent	DISTRICT	Site Freq	Percent
Crystal Falls	1	918	9.3	1	283	13.7
Newberry	2	349	3.5	2	134	6.5
Cadillac	3	1129	11.4	3	335	16.2
Alpena	4	793	8.0	4	239	11.6
Grand Rapids	5	950	9.6	5	163	7.9
Saginaw	6	1199	12.2	6	252	12.2
Kalamazoo	7	1992	20.2	7	321	15.5
Jackson	8	1585	16.1	8	241	11.7
Southfield	9	950	9.6	9	99	4.8

Variable 14

CONSECT	Acc Freq	Percent	CONSECT	Site Freq	Percent
1011	8	0.1	1011	4	0.2
	.			.	
	.			.	
	.			.	
83053	17	0.2	83053	1	0.0

Variable 15

MILEPT	Site Freq	Percent	MILEPT	Site Freq	Percent
0	68	0.7	0	8	0.4
0.01	2	0.0	0.01	2	0.1
	.			.	
	.			.	
	.			.	
33.42	1	0.0	33.42	1	0.0

Variable 16

LANES	Acc Freq	Percent	LANES	Site Freq	Percent
2	5934	60.2	2	1725	83.5
3	259	2.6	3	22	1.1
4	3413	34.6	4	304	14.7
5	259	2.6	5	16	0.8

Variable 17

	LANES_R	Acc Freq	Percent	LANES_R	Site Freq	Percent
No auxiliary lane	0	207	89.6	0	92	86.8
Truck climbing lane	1	24	10.4	1	14	13.2

Frequency Missing = 9634   Frequency Missing = 1961

Variable 18

	LANES_L	Acc Freq	Percent	LANES_L	Site Freq	Percent
No auxiliary lane	0	183	65.8	0	57	75.0
Truck climbing lane	1	95	34.2	1	19	25.0

Frequency Missing = 9587   Frequency Missing = 1991

Variable 19	PARK_R	Acc Freq	Percent	PARK_R	Site Freq	Percent
No on-street parking	0	2388	86.2	0	507	87.7
Restricted parking	1	24	0.9	1	5	0.9
Exclusive parking lane	3	359	13.0	3	66	11.4

Frequency Missing = 7094 Frequency Missing = 1489

Variable 20	PARK_L	Site Freq	Percent	PARK_L	Site Freq	Percent
No on-street parking	0	2383	84.4	0	511	87.2
Restricted parking	1	26	0.9	1	6	1.0
Angle parking	2	16	0.6	2	5	0.9
Exclusive parking lane	3	398	14.1	3	64	10.9

Frequency Missing = 7042 Frequency Missing = 1481

Variable 21	LANEWID	Site Freq	Percent	LANEWID	Site Freq	Percent
8 feet or less	8	2	0.0	8	2	0.1
9 feet	9	26	0.3	9	11	0.5
10 feet	10	1737	17.6	10	360	17.4
11 feet	11	3031	30.7	11	747	36.1
12 feet	12	5039	51.1	12	940	45.5
14 feet	14	7	0.1	14	1	0.0
15 feet or more	15	23	0.2	15	6	0.3

Variable 22	SWPR	Acc Freq	Percent	SWPR	Site Freq	Percent
	0	4133	41.9	0	423	20.5
	1	3	0.0	1	2	0.1
	2	3	0.0	2	3	0.1
	3	45	0.5	3	12	0.6
	4	25	0.3	4	15	0.7
	5	11	0.1	5	7	0.3
	6	111	1.1	6	47	2.3
	7	22	0.2	7	12	0.6
	8	1207	12.2	8	345	16.7
	9	73	0.7	9	31	1.5
	10	4117	41.7	10	1136	55.0
	11	10	0.1	11	6	0.3
	12	105	1.1	12	28	1.4

Variable 23	SWPL	Acc Freq	Percent	SWPL	Site Freq	Percent
	0	4068	41.2	0	414	20.0
	1	3	0.0	1	2	0.1
	2	41	0.4	2	6	0.3
	3	40	0.4	3	9	0.4
	4	47	0.5	4	20	1.0
	5	18	0.2	5	10	0.5
	6	101	1.0	6	42	2.0
	7	74	0.8	7	14	0.7
	8	1116	11.3	8	345	16.7
	9	59	0.6	9	28	1.4
	10	4042	41.0	10	1136	55.0
	11	25	0.3	11	5	0.2
	12	231	2.3	12	36	1.7

Variable 24

	NOPASS	Acc Freq	Percent	NOPASS	Site Freq	Percent
Passing allowed (both dirs.)	0	3585	50.6	0	829	48.1
No passing zone (ascend. dir.)	1	822	11.6	1	206	12.0
No passing zone (descend. dir.)	2	893	12.6	2	235	13.6
No passing zone (both dirs.)	3	1790	25.2	3	452	26.2

Frequency Missing = 2775   Frequency Missing = 345

Variable 25

	LAND_USE	Acc Freq	Percent	LAND_USE	Site Freq	Percent
Rural	1	3513	35.6	1	1212	58.6
Urban	3	6352	64.4	3	855	41.4

Variable 26

SPEEDLIM	Acc Freq	Percent	SPEEDLIM	Site Freq	Percent
25	298	3.0	25	34	1.6
30	919	9.3	30	106	5.1
35	1805	18.3	35	212	10.3
40	737	7.5	40	138	6.7
45	930	9.4	45	167	8.1
50	932	9.4	50	118	5.7
55	4244	43.0	55	1292	62.5

Variable 27

	INT_TYPE	Acc Freq	Percent	INT_TYPE	Site Freq	Percent
Tee Left	10	2472	25.1	10	507	24.5
Tee Right	11	2659	27.0	11	511	24.7
Terminal Tee	12	7	0.1	12	3	0.1
Obtuse Wye Left	20	1125	11.4	20	215	10.4
Acute Wye Left	22	1356	13.7	22	287	13.9
Obtuse Wye Right	24	1128	11.4	24	275	13.3
Acute Wye Right	26	1118	11.3	26	269	13.0

Variable 28

	SIGNAL	Acc Freq	Percent	SIGNAL	Site Freq	Percent
No Signal	0	9482	96.1	0	2035	98.5
Flasher	4	383	3.9	4	32	1.5

Variable 29

ADT	Acc Freq	Percent	ADT	Site Freq	Percent
30	1	0.0	30	1	0.0
	.			.	
	.			.	
	.			.	
3739	124	1.3	3739	2	0.1

Variable 30

TOT	Acc Freq	Percent	TOT	Site Freq	Percent
0	834	8.5	0	627	30.3
	.			.	
	.			.	
	.			.	
112	111	1.1	112	1	0.0

Variable 31

INJURY	Acc Freq	Percent	INJURY	Site Freq	Percent
0	2027	20.5	0	1157	56.0
1	1470	14.9	1	408	19.7
2	1261	12.8	2	210	10.2
3	1005	10.2	3	107	5.2
4	926	9.4	4	72	3.5
5	232	2.4	5	17	0.8
6	473	4.8	6	26	1.3
7	399	4.0	7	18	0.9
8	281	2.8	8	13	0.6
9	252	2.6	9	10	0.5
10	59	0.6	10	3	0.1
11	229	2.3	11	6	0.3
12	89	0.9	12	2	0.1
13	63	0.6	13	2	0.1
14	260	2.6	14	5	0.2
15	82	0.8	15	2	0.1
18	65	0.7	18	1	0.0
20	50	0.5	20	1	0.0
22	69	0.7	22	1	0.0
23	99	1.0	23	1	0.0
26	97	1.0	26	1	0.0
28	48	0.5	28	1	0.0
31	111	1.1	31	1	0.0
32	107	1.1	32	1	0.0
39	111	1.1	39	1	0.0

Variable 32

FATAL	Acc Freq	Percent	FATAL	Site Freq	Percent
0	9447	95.8	0	2034	98.4
1	401	4.1	1	32	1.5
2	17	0.2	2	1	0.0

Variable 33

WET	Acc Freq	Percent	WET	Site Freq	Percent
0	2686	27.2	0	1329	64.3
1	1615	16.4	1	370	17.9
2	1167	11.8	2	146	7.1
3	824	8.4	3	82	4.0
4	734	7.4	4	49	2.4
5	317	3.2	5	18	0.9
6	285	2.9	6	15	0.7
7	232	2.4	7	12	0.6
8	275	2.8	8	9	0.4
9	222	2.3	9	9	0.4
10	69	0.7	10	3	0.1
11	155	1.6	11	4	0.2
12	132	1.3	12	4	0.2
13	128	1.3	13	3	0.1
14	88	0.9	14	2	0.1
16	112	1.1	16	2	0.1
17	50	0.5	17	1	0.0
18	133	1.3	18	2	0.1
19	116	1.2	19	2	0.1
20	99	1.0	20	1	0.0
22	107	1.1	22	1	0.0
33	111	1.1	33	1	0.0
45	97	1.0	45	1	0.0
50	111	1.1	50	1	0.0

Variable 34

ICY	Acc Freq	Percent	ICY	Site Freq	Percent
0	3075	31.2	0	1359	65.7
1	1931	19.6	1	382	18.5
2	1358	13.8	2	146	7.1
3	1069	10.8	3	81	3.9
4	510	5.2	4	38	1.8
5	487	4.9	5	20	1.0
6	393	4.0	6	13	0.6
7	250	2.5	7	13	0.6
8	240	2.4	8	3	0.1
9	39	0.4	9	2	0.1
10	18	0.2	10	1	0.0
12	24	0.2	12	1	0.0
13	133	1.3	13	2	0.1
15	51	0.5	15	1	0.0
16	253	2.6	16	4	0.2
18	34	0.3	18	1	0.0

Variable 35

DARK	Acc Freq	Percent	DARK	Site Freq	Percent
0	1999	20.3	0	1070	51.8
1	1535	15.6	1	447	21.6
2	1391	14.1	2	229	11.1
3	1045	10.6	3	124	6.0
4	746	7.6	4	76	3.7
5	607	6.2	5	40	1.9
6	420	4.3	6	24	1.2
7	287	2.9	7	11	0.5
8	210	2.1	8	8	0.4
9	279	2.8	9	8	0.4
10	413	4.2	10	11	0.5
11	110	1.1	11	4	0.2
12	81	0.8	12	3	0.1
13	118	1.2	13	3	0.1
14	91	0.9	14	2	0.1
15	39	0.4	15	1	0.0
16	36	0.4	16	1	0.0
20	111	1.1	20	1	0.0
21	60	0.6	21	1	0.0
22	180	1.8	22	2	0.1
23	107	1.1	23	1	0.0

Variable 36

MISC_SV	Acc Freq	Percent	MISC_SV	Site Freq	Percent
0	9107	92.3	0	2018	97.6
1	531	5.4	1	43	2.1
2	180	1.8	2	5	0.2
3	47	0.5	3	1	0.0

Variable 37

ROLLOVER	Acc Freq	Percent	ROLLOVER	Site Freq	Percent
0	7928	80.4	0	1861	90.0
1	1320	13.4	1	163	7.9
2	449	4.6	2	32	1.5
3	118	1.2	3	7	0.3
4	50	0.5	4	4	0.2

Variable 38

TRAIN	Site Freq	Percent
0	9841	99.8
1	20	0.2
2	4	0.0

TRAIN	Site Freq	Percent
0	2064	99.9
1	2	0.1
2	1	0.0

Variable 39

PARKVEH	Acc Freq	Percent
0	8238	83.5
1	1278	13.0
2	208	2.1
3	35	0.4
4	4	0.0
5	30	0.3
6	72	0.7

PARKVEH	Site Freq	Percent
0	1922	93.0
1	121	5.9
2	15	0.7
3	3	0.1
4	1	0.0
5	2	0.1
6	3	0.1

Variable 40

BACKING	Acc Freq	Percent
0	8079	81.9
1	1570	15.9
2	180	1.8
4	36	0.4

BACKING	Site Freq	Percent
0	1976	95.6
1	87	4.2
2	3	0.1
4	1	0.0

Variable 41

PARKING	Acc Freq	Percent
0	9724	98.6
1	141	1.4

PARKING	Site Freq	Percent
0	2060	99.7
1	7	0.3

Variable 42

PED	Acc Freq	Percent
0	8985	91.1
1	751	7.6
2	129	1.3

PED	Site Freq	Percent
0	2021	97.8
1	43	2.1
2	3	0.1

Variable 43

FOBJ	Acc Freq	Percent
0	3211	32.5
1	2500	25.3
2	1324	13.4
3	878	8.9
4	500	5.1
5	548	5.6
6	144	1.5
7	97	1.0
8	88	0.9
9	111	1.1
10	151	1.5
11	32	0.3
12	118	1.2
14	103	1.0
17	60	0.6

FOBJ	Site Freq	Percent
0	1325	64.1
1	420	20.3
2	160	7.7
3	63	3.0
4	36	1.7
5	27	1.3
6	11	0.5
7	6	0.3
8	4	0.2
9	4	0.2
10	3	0.1
11	2	0.1
12	3	0.1
14	2	0.1
17	1	0.0

Variable 44

OOBJ	Acc Freq	Percent
0	9406	95.3
1	459	4.7

OOBJ	Site Freq	Percent
0	2030	98.2
1	37	1.8

Variable 45

ANIMAL	Acc Freq	Percent	ANIMAL	Site Freq	Percent
0	7739	78.4	0	1665	80.6
1	1313	13.3	1	267	12.9
2	442	4.5	2	87	4.2
3	118	1.2	3	22	1.1
4	131	1.3	4	16	0.8
5	87	0.9	5	6	0.3
6	23	0.2	6	3	0.1
9	12	0.1	9	1	0.0

Variable 46

BIKE	Acc Freq	Percent	BIKE	Site Freq	Percent
0	8943	90.7	0	2007	97.1
1	727	7.4	1	54	2.6
2	17	0.2	2	2	0.1
3	178	1.8	3	4	0.2

Variable 47

HDON	Acc Freq	Percent	HDON	Site Freq	Percent
0	6723	68.2	0	1818	88.0
1	1717	17.4	1	183	8.9
2	932	9.4	2	44	2.1
3	298	3.0	3	15	0.7
4	115	1.2	4	4	0.2
5	44	0.4	5	2	0.1
6	36	0.4	6	1	0.0

Variable 48

ANGST	Acc Freq	Percent	ANGST	Site Freq	Percent
0	6458	65.5	0	1830	88.5
1	1722	17.5	1	168	8.1
2	581	5.9	2	38	1.8
3	181	1.8	3	9	0.4
4	311	3.2	4	10	0.5
5	262	2.7	5	5	0.2
6	83	0.8	6	2	0.1
7	30	0.3	7	1	0.0
8	25	0.3	8	1	0.0
11	212	2.1	11	3	0.1

Variable 49

RE	Acc Freq	Percent	RE	Site Freq	Percent
0	2806	28.4	0	1377	66.6
1	1279	13.0	1	302	14.6
2	926	9.4	2	131	6.3
3	676	6.9	3	72	3.5
4	615	6.2	4	51	2.5
5	490	5.0	5	32	1.5
6	418	4.2	6	23	1.1
7	286	2.9	7	15	0.7
8	271	2.7	8	14	0.7
9	141	1.4	9	7	0.3
10	144	1.5	10	5	0.2
11	24	0.2	11	1	0.0
12	112	1.1	12	3	0.1
13	69	0.7	13	3	0.1
14	318	3.2	14	8	0.4
15	97	1.0	15	3	0.1
16	30	0.3	16	1	0.0
17	118	1.2	17	3	0.1
18	69	0.7	18	1	0.0
19	30	0.3	19	1	0.0
20	146	1.5	20	3	0.1
21	48	0.5	21	1	0.0
23	61	0.6	23	2	0.1
27	111	1.1	27	1	0.0
33	51	0.5	33	1	0.0
36	50	0.5	36	1	0.0
40	204	2.1	40	2	0.1
43	65	0.7	43	1	0.0
66	111	1.1	66	1	0.0
78	99	1.0	78	1	0.0

Variable 50

ANGTN	Acc Freq	Percent	ANGTN	Site Freq	Percent
0	4847	49.1	0	1679	81.2
1	1881	19.1	1	250	12.1
2	713	7.2	2	54	2.6
3	662	6.7	3	32	1.5
4	455	4.6	4	16	0.8
5	101	1.0	5	6	0.3
6	109	1.1	6	5	0.2
7	165	1.7	7	5	0.2
8	92	0.9	8	3	0.1
9	98	1.0	9	2	0.1
10	130	1.3	10	4	0.2
11	132	1.3	11	2	0.1
12	74	0.8	12	3	0.1
13	46	0.5	13	1	0.0
17	69	0.7	17	1	0.0
19	64	0.6	19	1	0.0
22	111	1.1	22	1	0.0
24	65	0.7	24	1	0.0
25	51	0.5	25	1	0.0

Variable 51

SSPASS	Acc Freq	Percent	SSPASS	Site Freq	Percent
0	9446	95.8	0	2044	98.9
1	406	4.1	1	22	1.1
2	13	0.1	2	1	0.0

Y-Intersections

Variable 52

RELT	Acc Freq	Percent	RELT	Site Freq	Percent
0	6876	69.7	0	1820	88.1
1	1786	18.1	1	194	9.4
2	486	4.9	2	33	1.6
3	298	3.0	3	10	0.5
4	419	4.2	4	10	0.5

Variable 53

RERT	Acc Freq	Percent	RERT	Site Freq	Percent
0	7742	78.5	0	1934	93.6
1	1737	17.6	1	117	5.7
2	386	3.9	2	16	0.8

Variable 54

OTH_DR	Acc Freq	Percent	OTH_DR	Site Freq	Percent
0	7579	76.8	0	1937	93.7
1	1237	12.5	1	102	4.9
2	311	3.2	2	14	0.7
3	213	2.2	3	7	0.3
5	99	1.0	5	3	0.1
6	111	1.1	6	1	0.0
8	97	1.0	8	1	0.0
9	111	1.1	9	1	0.0
10	107	1.1	10	1	0.0

Variable 55

ANGDR	Acc Freq	Percent	ANGDR	Site Freq	Percent
0	7342	74.4	0	1906	92.2
1	1330	13.5	1	125	6.0
2	335	3.4	2	16	0.8
3	226	2.3	3	7	0.3
4	278	2.8	4	9	0.4
5	222	2.3	5	2	0.1
8	25	0.3	8	1	0.0
9	107	1.1	9	1	0.0

Variable 56

REDR	Acc Freq	Percent	REDR	Site Freq	Percent
0	6151	62.4	0	1788	86.5
1	1590	16.1	1	189	9.1
2	564	5.7	2	42	2.0
3	391	4.0	3	18	0.9
4	175	1.8	4	9	0.4
5	102	1.0	5	4	0.2
6	122	1.2	6	4	0.2
7	81	0.8	7	3	0.1
8	117	1.2	8	3	0.1
10	96	1.0	10	2	0.1
11	218	2.2	11	2	0.1
15	147	1.5	15	2	0.1
26	111	1.1	26	1	0.0

Variable 57

SSMEET	Acc Freq	Percent	SSMEET	Site Freq	Percent
0	9079	92.0	0	2022	97.8
1	724	7.3	1	43	2.1
2	62	0.6	2	2	0.1

Variable 58

HDONLT	Acc Freq	Percent	HDONLT	Site Freq	Percent
0	6521	66.1	0	1862	90.1
1	1653	16.8	1	145	7.0
2	731	7.4	2	37	1.8
3	234	2.4	3	10	0.5
4	95	1.0	4	2	0.1
5	121	1.2	5	2	0.1
6	53	0.5	6	2	0.1
7	112	1.1	7	2	0.1
8	137	1.4	8	2	0.1
9	47	0.5	9	1	0.0
10	97	1.0	10	1	0.0
12	64	0.6	12	1	0.0

Variable 59

DUALLT	Acc Freq	Percent	DUALLT	Site Freq	Percent
0	9701	98.3	0	2059	99.6
1	164	1.7	1	8	0.4

Variable 60

DUALRT	Acc Freq	Percent	DUALRT	Site Freq	Percent
0	9593	97.2	0	2050	99.2
1	267	2.7	1	16	0.8
2	5	0.1	2	1	0.0

Variable 61

CONSECMP	Acc Freq	Percent
0	44	0.5
0.01	34	0.4
	.	
	.	
	.	
26.74	1	0.0

Frequency Missing = 587

Variable 62

	HWAYTYPE	Acc Freq	Percent
Interchange area	1	50	0.5
Intersection area	2	9228	99.5

Frequency Missing = 587

Y-Intersections

Variable 63	HWAYCODE	Acc Freq	Percent
W/in confines of intersection	0	2702	29.1
W/in 100' N. of intersection	1	561	6.0
Off ramps near main road	21	4	0.0
Off ramps b/t roads	22	1	0.0
Leave ramp or on crossroad	23	25	0.3
Enter ramp from crossroad	24	11	0.1
On ramps near main road	26	6	0.1
	51	462	5.0
Crossings commercial d/w	56	1072	11.6
Railroad crossing (at grade)	59	35	0.4
Crossings school/church d/w	60	1	0.0
Crossings leave crossover	73	3	0.0
	75	1	0.0
Right turn flare (slot)	79	1	0.0
Turn channel at crossroad	83	1	0.0
Turn channel b/t roadways	84	1	0.0
Turn channel at trunkline	85	5	0.1
	86	1	0.0
Crossings other or not known	99	4385	47.3

Frequency Missing = 587

Variable 64	WEEKDAY	Acc Freq	Percent
Sunday	1	997	10.7
Monday	2	1211	13.1
Tuesday	3	1256	13.5
Wednesday	4	1284	13.8
Thursday	5	1348	14.5
Friday	6	1722	18.6
Saturday	7	1460	15.7

Frequency Missing = 587

Variable 65	HOUR	Acc Freq	Percent
Mid-1am	1	213	2.3
1am-2am	2	174	1.9
2am-3am	3	186	2.0
3am-4am	4	103	1.1
4am-5am	5	78	0.8
5am-6am	6	103	1.1
6am-7am	7	191	2.1
7am-8am	8	372	4.0
8am-9am	9	374	4.0
9am-10am	10	310	3.3
10am-11am	11	368	4.0
11am-noon	12	455	4.9
Noon-1pm	13	580	6.3
1pm-2pm	14	557	6.0
2pm-3pm	15	602	6.5
3pm-4pm	16	841	9.1
4pm-5pm	17	772	8.3
5pm-6pm	18	735	7.9
6pm-7pm	19	538	5.8
7pm-8pm	20	428	4.6
8pm-9pm	21	359	3.9
9pm-10pm	22	327	3.5
10pm-11pm	23	310	3.3
11pm-mid	24	276	3.0
Not known	25	26	0.3

Frequency Missing = 587

Variable 66	MONTH	Acc Freq	Percent
January	1	818	8.8
February	2	733	7.9
March	3	640	6.9
April	4	585	6.3
May	5	722	7.8
June	6	737	7.9
July	7	736	7.9
August	8	790	8.5
September	9	796	8.6
October	10	912	9.8
November	11	807	8.7
December	12	1002	10.8

Frequency Missing = 587

Variable 67

DATE	Acc Freq	Percent
1	296	3.2
2	364	3.9
3	328	3.5
4	321	3.5
5	334	3.6
6	304	3.3
7	265	2.9
8	295	3.2
9	326	3.5
10	303	3.3
11	294	3.2
12	326	3.5
13	264	2.8
14	376	4.1
15	307	3.3
16	311	3.4
17	287	3.1
18	283	3.1
19	301	3.2
20	315	3.4
21	333	3.6
22	308	3.3
23	299	3.2
24	275	3.0
25	302	3.3
26	303	3.3
27	284	3.1
28	272	2.9
29	273	2.9
30	270	2.9
31	159	1.7

Frequency Missing = 587

Variable 68

YEAR	Acc Freq	Percent
87	1852	20.0
88	1862	20.1
89	2013	21.7
90	1861	20.1
91	1690	18.2

Frequency Missing = 587

Variable 69

	WEATHER	Acc Freq	Percent
Clear or cloudy	1	6877	74.1
Fog	2	104	1.1
Raining	3	1205	13.0
Snowing	4	1081	11.7
Not known	5	11	0.1

Frequency Missing = 587

Variable 70	LIGHTING	Acc Freq	Percent
Daylight	1	6204	66.9
Dawn or dusk	2	427	4.6
Darkness/street lights	3	804	8.7
Darkness/no street lights	4	1824	19.7
Not known	5	19	0.2

Frequency Missing = 587

Variable 71	SURFACE	Acc Freq	Percent
Dry	1	5564	60.0
Wet	2	2095	22.6
Snowy or icy	3	1584	17.1
Other or not known	4	35	0.4

Frequency Missing = 587

Variable 72	DEFECT	Acc Freq	Percent
None	1	9243	99.6
Loose mat. on surface	3	17	0.2
Holes, ruts, bumps	4	11	0.1
Drifting snow	6	2	0.0
Slippery when wet	8	1	0.0
Other or not known	9	4	0.0

Frequency Missing = 587

Variable 73	A_INJS	Acc Freq	Percent
	0	8769	94.5
	1	378	4.1
	2	90	1.0
	3	28	0.3
	4	10	0.1
	5	1	0.0
	6	1	0.0
	7	1	0.0

Frequency Missing = 587

Variable 74	B_INJS	Acc Freq	Percent
	0	8353	90.0
	1	739	8.0
	2	145	1.6
	3	27	0.3
	4	8	0.1
	5	2	0.0
	6	4	0.0

Frequency Missing = 587

Y-Intersections

Variable 75

C_INJS	Acc Freq	Percent
0	7722	83.2
1	1179	12.7
2	283	3.1
3	62	0.7
4	23	0.2
5	6	0.1
6	1	0.0
7	1	0.0
8	1	0.0

Frequency Missing = 587

Variable 76

	ALIGN	Acc Freq	Percent
Straight	1	7926	85.4
Curve	2	1320	14.2
Transition area	3	21	0.2
Not known	4	11	0.1

Frequency Missing = 587

Variable 77

	SPECTAGS	Acc Freq	Percent
School bus involved	1	39	0.4
School bus associated	2	1	0.0
School bus other associated	3	14	0.2
Deer involved	4	614	6.6
Deer associated	5	30	0.3
Emergency or pursuit	6	14	0.2
Construction zone	8	112	1.2
None of the above	10	8454	91.1

Frequency Missing = 587

Variable 78	ACC_TYPE	Acc Freq	Percent
Misc. single vehicle	0	56	0.6
Overturn	10	272	2.9
Hit train	20	4	0.0
Hit parked vehicle	30	198	2.1
Backing	48	99	1.1
Parking	49	8	0.1
Pedestrian	50	53	0.6
Fixed object	60	1560	16.8
Other object	70	43	0.5
Animal	80	640	6.9
Bicycle	90	70	0.8
Head-on	141	373	4.0
Angle straight	144	430	4.6
Rear-end	147	2558	27.6
Angle turn	244	907	9.8
Sideswipe same dir.	342	24	0.3
Rear-end left turn	345	344	3.7
Rear-end right turn	346	159	1.7
Other driveway	440	204	2.2
Angle driveway	444	251	2.7
Rear-end driveway	447	579	6.2
Sideswipe opp. direction	543	49	0.5
Head-on left turn	545	366	3.9
Dual left turn	645	11	0.1
Dual right turn	646	20	0.2

Frequency Missing = 587

Variable 79	NUM_VEHS	Acc Freq	Percent
	1	2881	31.1
	2	5946	64.1
	3	407	4.4
	4	39	0.4
	5	3	0.0
	6	1	0.0
	10	1	0.0

Frequency Missing = 587

Variable 80	DISTXRD	Acc Freq	Percent (in feet)
	0	1514	16.3
	1	10	0.1
		.	
		.	
		.	
	47520	1	0.0

Frequency Missing = 587

Y-Intersections

Variable 81	DIRXRD	Acc Freq	Percent
At intersection	0	1514	16.3
North	1	1365	14.7
Northeast	2	570	6.1
East	3	1233	13.3
Southeast	4	603	6.5
South	5	1454	15.7
Southwest	6	619	6.7
West	7	1325	14.3
Northwest	8	586	6.3
Unknown	9	9	0.1

Frequency Missing = 587

Variable 82	NUMUNINJ	Acc Freq	Percent
	0	799	8.6
	1	2302	24.8
	2	2939	31.7
	3	1662	17.9
	4	798	8.6
	5	415	4.5
	6	189	2.0
	7	77	0.8
	8	45	0.5
	9	34	0.4
	10	10	0.1
	11	5	0.1
	12	2	0.0
	16	1	0.0

Frequency Missing = 587

Variable 83	VTSUB1	Acc Freq	Percent
	0	2	0.0
Passenger car	1	6973	75.2
Truck	2	1853	20.0
Motorcycle, moped, etc.	3	106	1.1
School bus	4	20	0.2
Commercial bus	5	8	0.1
Farm equipment	6	2	0.0
Construction equipment	7	7	0.1
Emer veh, snowmobile, etc.	8	305	3.3
Other road vehicle	11	2	0.0

Frequency Missing = 587

Variable 84	DRINT1	Acc Freq	Percent
Go straight ahead	1	5528	59.6
Overtaking or passing	2	264	2.8
Change lanes	3	245	2.6
Make right turn	4	476	5.1
Make left turn	5	1574	17.0
Make U turn	6	31	0.3
Slowing or stopping	7	9	0.1
Starting up on road	8	69	0.7
Entering parking	9	8	0.1
Leaving parking	10	22	0.2
Backing	11	241	2.6
Stopped on road	12	241	2.6
Pursuing/being pursued	13	10	0.1
Avoid object	14	2	0.0
Avoid animal	15	51	0.5
Avoid pedestrian	16	14	0.2
Lost load from vehicle	17	20	0.2
Avoid veh same/opp. dir.	18	376	4.1
Avoid veh at an angle	19	76	0.8
Other or not known	20	21	0.2

Frequency Missing = 587

Variable 85	VIOL1	Acc Freq	Percent
No hazardous action	1	1943	20.9
Speed too fast	2	1020	11.0
Speed too slow	3	5	0.1
Failed to yield ROW	4	1574	17.0
Wrong way	5	4	0.0
Drove left of center	6	673	7.3
Improper turn/signal	7	289	3.1
Improper backing/start	8	227	2.4
Followed too close	9	3376	36.4
Other or not known	10	167	1.8

Frequency Missing = 587

Variable 86	CONCIR1	Acc Freq	Percent
DUI	1	369	4.0
Reckless/careless driving	2	225	2.4
Ill, fatigued, inattention	3	113	1.2
Failed comply w/ lic. res.	4	1	0.0
Obscured vision	5	156	1.7
Defective equipment	6	129	1.4
Lost control	7	7	0.1
None	8	1678	18.1
Skidding	9	648	7.0
Other or not known	10	5952	64.2

Frequency Missing = 587

Variable 87	DIRTRAV1	Acc Freq	Percent
North	1	2123	22.9
Northeast	2	369	4.0
East	3	1937	20.9
Southeast	4	388	4.2
South	5	1865	20.1
Southwest	6	321	3.5
West	7	1866	20.1
Northwest	8	362	3.9
Unknown	9	47	0.5

Frequency Missing = 587

Variable 88	OBHIT1	Acc Freq	Percent
No object hit	1	7565	81.5
Guardrail, guard post	2	150	1.6
Highway sign	3	352	3.8
Street light, utility pole	4	260	2.8
Culvert	5	17	0.2
Ditch, embankment, stream	6	325	3.5
Bridge pier or abutment	7	3	0.0
Bridge rail or deck	8	7	0.1
Tree	9	175	1.9
Highway or r/r signal	10	12	0.1
Building	11	31	0.3
Mailbox	12	95	1.0
Fence	13	48	0.5
Traffic island or curb	14	72	0.8
Concrete median barrier	15	1	0.0
Other on t-way object	16	90	1.0
Other off t-way object	17	68	0.7
Overhead fixed object	18	6	0.1
Unknown or non-motor veh	19	1	0.0

Frequency Missing = 587

Variable 89	IMPCODE1	Acc Freq	Percent
Rollover	0	273	2.9
Center front	1	3633	39.2
Right front	2	1219	13.1
Right side	3	613	6.6
Right rear	4	443	4.8
Center rear	5	516	5.6
Left rear	6	399	4.3
Left side	7	571	6.2
Left front	8	1339	14.4
Other impact or misc.	9	272	2.9

Frequency Missing = 587

Variable 90	VTSUB2	Acc Freq	Percent
No vehicle 2	0	2756	29.7
Passenger car	1	4928	53.1
Truck	2	1233	13.3
Motorcycle, moped, etc.	3	38	0.4
School bus	4	19	0.2
Commercial bus	5	8	0.1
Farm equipment	6	1	0.0
Construction equipment	7	7	0.1
Emer veh, snowmobile, etc.	8	162	1.7
Pedestrian	9	55	0.6
Pedalcycle	10	70	0.8
Other road vehicle	11	1	0.0

Frequency Missing = 587

Variable 91	DRINT2	Acc Freq	Percent
No vehicle 2	0	2756	29.7
Go straight ahead	1	3180	34.3
Overtaking or passing	2	98	1.1
Change lanes	3	33	0.4
Make right turn	4	320	3.4
Make left turn	5	1222	13.2
Make U turn	6	8	0.1
Slowing or stopping	7	26	0.3
Starting up on road	8	15	0.2
Entering parking	9	5	0.1
Leaving parking	10	8	0.1
Backing	11	27	0.3
Stopped on road	12	1332	14.4
Pursuing/being pursued	13	1	0.0
Avoid animal	15	5	0.1
Avoid pedestrian	16	3	0.0
Lost load from vehicle	17	3	0.0
Avoid veh same/opp. dir.	18	152	1.6
Avoid veh at an angle	19	84	0.9

Frequency Missing = 587

Variable 92	VIOL2	Acc Freq	Percent
No vehicle 2	0	2756	29.7
No hazardous action	1	5612	60.5
Speed too fast	2	52	0.6
Speed too slow	3	1	0.0
Failed to yield ROW	4	234	2.5
Wrong way	5	1	0.0
Drove left of center	6	98	1.1
Improper turn/signal	7	66	0.7
Improper backing/start	8	15	0.2
Followed too close	9	405	4.4
Other or not known	10	38	0.4

Frequency Missing = 587

Variable 93	CONCIR2	Acc Freq	Percent
No vehicle 2	0	2756	29.7
DUI	1	21	0.2
Reckless/careless driving	2	9	0.1
Ill, fatigued, inattention	3	7	0.1
Failed comply w/ lic. res.	4	2	0.0
Obscured vision	5	44	0.5
Defective equipment	6	13	0.1
None	8	5486	59.1
Skidding	9	87	0.9
Other or not known	10	853	9.2

Frequency Missing = 587

Variable 94	DIRTRAV2	Acc Freq	Percent
No vehicle 2	0	2756	29.7
North	1	1442	15.5
Northeast	2	272	2.9
East	3	1466	15.8
Southeast	4	267	2.9
South	5	1201	12.9
Southwest	6	228	2.5
West	7	1387	14.9
Northwest	8	223	2.4
Unknown	9	36	0.4

Frequency Missing = 587

Variable 95	OBHIT2	Acc Freq	Percent
No vehicle 2	0	2756	29.7
No object hit	1	6350	68.4
Guardrail, guard post	2	2	0.0
Highway sign	3	8	0.1
Street light, utility pole	4	12	0.1
Ditch, embankment, stream	6	3	0.0
Bridge rail or deck	8	2	0.0
Tree	9	2	0.0
Highway or r/r signal	10	1	0.0
Building	11	3	0.0
Mailbox	12	5	0.1
Fence	13	3	0.0
Traffic island or curb	14	1	0.0
Other on t-way object	16	2	0.0
Other off t-way object	17	2	0.0
Unknown or non-motor veh	19	126	1.4

Frequency Missing = 587

Variable 96	IMPCODE2	Acc Freq	Percent
No vehicle 2/Rollover	0	2762	29.8
Center front	1	1274	13.7
Right front	2	618	6.7
Right side	3	365	3.9
Right rear	4	471	5.1
Center rear	5	1853	20.0
Left rear	6	429	4.6
Left side	7	504	5.4
Left front	8	774	8.3
Other impact or misc.	9	228	2.5

Frequency Missing = 587

Variable 97	VTSUB3	Acc Freq	Percent
No vehicle 3	0	8820	95.1
Passenger car	1	363	3.9
Truck	2	72	0.8
Motorcycle, moped, etc.	3	1	0.0
Commercial bus	5	1	0.0
Emer veh, snowmobile, etc.	8	14	0.2
Pedestrian	9	6	0.1
Other road vehicle	11	1	0.0

Frequency Missing = 587

Variable 98	DRINT3	Acc Freq	Percent
No vehicle 3	0	8820	95.1
Go straight ahead	1	190	2.0
Make right turn	4	6	0.1
Make left turn	5	80	0.9
Make U turn	6	2	0.0
Slowing or stopping	7	3	0.0
Leaving parking	10	1	0.0
Backing	11	2	0.0
Stopped on road	12	166	1.8
Avoid object	14	1	0.0
Avoid veh same/opp. dir.	18	7	0.1

Frequency Missing = 587

Variable 99	VIOL3	Acc Freq	Percent
No vehicle 3	0	8820	95.1
No hazardous action	1	400	4.3
Speed too fast	2	4	0.0
Failed to yield ROW	4	2	0.0
Improper backing/start	8	1	0.0
Followed too close	9	51	0.5

Frequency Missing = 587

Y-Intersections

Variable 100	CONCIR3	Acc Freq	Percent
No vehicle 3	0	8820	95.1
Ill, fatigued, inattention	3	2	0.0
Obscured vision	5	4	0.0
None	8	395	4.3
Skidding	9	6	0.1
Other or not known	10	51	0.5

Frequency Missing = 587

Variable 101	DIRTRAV3	Acc Freq	Percent
No vehicle 3	0	8820	95.1
North	1	113	1.2
Northeast	2	15	0.2
East	3	106	1.1
Southeast	4	22	0.2
South	5	73	0.8
Southwest	6	15	0.2
West	7	95	1.0
Northwest	8	14	0.2
Unknown	9	5	0.1

Frequency Missing = 587

Variable 102	OBHIT3	Acc Freq	Percent
No vehicle 3	0	8820	95.1
No object hit	1	452	4.9
Unknown or non-motor veh	19	6	0.1

Frequency Missing = 587

Variable 103	IMPCODE3	Acc Freq	Percent
No vehicle 3	0	8820	95.1
Center front	1	109	1.2
Right front	2	12	0.1
Right side	3	6	0.1
Right rear	4	10	0.1
Center rear	5	212	2.3
Left rear	6	29	0.3
Left side	7	25	0.3
Left front	8	40	0.4
Other impact or misc.	9	15	0.2

Frequency Missing = 587

Variable 104	NUM_KILL	Acc Freq	Percent
	0	9243	99.6
	1	32	0.3
	2	3	0.0

Frequency Missing = 587

Variable 105	NUM_INJ	Acc Freq	Percent
	0	6679	72.0
	1	1734	18.7
	2	562	6.1
	3	190	2.0
	4	73	0.8
	5	24	0.3
	6	7	0.1
	7	5	0.1
	8	3	0.0
	9	1	0.0

Frequency Missing = 587

Variable 106	ACCSEV	Acc Freq	Percent
Fatal	1	35	0.4
Personal injury	2	2580	27.8
Property damage only	3	6663	71.8

Frequency Missing = 587

Variable 107	ARNUM	Acc Freq	Percent
	20	1	0.0
		.	
		.	
		.	
	917001	1	0.0

Frequency Missing = 587

Variable 108	CURVECAT	Acc Freq	Percent
Obtuse Y on left side of right curve or acute Y on right side of left curve	1	824	8.4
Acute Y on left side of right curve or obtuse Y on right side of left curve	2	1006	10.2
Acute Y on right side of right curve or obtuse Y on left side of left curve	3	181	1.8
Obtuse Y on right side of right curve or acute Y on left side of left curve	4	233	2.4
Tangent, obtuse Y on left side or acute Y on right side	5	1238	12.5
Tangent, acute Y on left side or obtuse Y on right side	6	1245	12.6
T-intersection	7	5138	52.1

	CURVECAT	Site Freq	Percent
Obtuse Y on left side of right curve or acute Y on right side of left curve	1	238	11.5
Acute Y on left side of right curve or obtuse Y on right side of left curve	2	239	11.6
Acute Y on right side of right curve or obtuse Y on left side of left curve	3	53	2.6
Obtuse Y on right side of right curve or acute Y on left side of left curve	4	72	3.5
Tangent, obtuse Y on left side or acute Y on right side	5	193	9.3
Tangent, acute Y on left side or obtuse Y on right side	6	251	12.1
T-intersection	7	1021	49.4

Variable 109	SPECIAL	Acc Freq	Percent	SPECIAL	Site Freq	Percent
Special right	1	300	3.0	1	88	4.3
Special left	2	367	3.7	2	78	3.8
Unknown	8	3	0.0	8	3	0.1
Not special	9	9195	93.2	9	1898	91.8

Variable 110	SIDE_ADT	Acc Freq	Percent	SIDE_ADT	Site Freq	Percent
	19	1	0.0	19	1	0.0
		.			.	
		.			.	
		.			.	
	6800	1	0.0	6800	1	0.0
Unknown	9998	268	2.7	9998	90	4.4
Not applicable	9999	9198	93.2	9999	1901	92.0

Variable 111	SIDEPAVE	Acc Freq	Percent	SIDEPAVE	Site Freq	Percent
Gravel	1	77	0.8	1	13	0.6
Paved	2	412	4.2	2	40	1.9
Unknown	9	9376	95.0	9	2014	97.4

Variable 112	CONTRAST	Acc Freq	Percent	CONTRAST	Site Freq	Percent
Contrast	1	318	3.2	1	39	1.9
No contrast	2	171	1.7	2	14	0.7
Unknown	9	9376	95.0	9	2014	97.4

APPENDIX C

BAYESIAN ESTIMATION OF RATES USING A HIERARCHICAL LOG-LINEAR MODEL

by
Paul E. Green

C.1 Introduction

Suppose rates y_i/t_i , $i = 1, \ldots, N$ are observed where y_i is in the form of a count and $t_i > 0$ is a known continuous measure of exposure. Poisson models are widely used in the regression analysis of rates (see Holford 1980, Frome 1983, Breslow and Day 1987, and McCullagh and Nelder 1989). Consider the problem of estimating the true unknown rates λ_i, and assessing the goodness of fit of the Poisson model.

C.1.1 The Classical Poisson Model for the Analysis of Rates

Assuming the random observations Y_i are independent and Poisson distributed with mean $t_i\lambda_i$, a frequentist approach to this problem is often through the classical log-linear model

$$\log \lambda_i = \mathbf{x}_i^T \boldsymbol{\beta} \qquad i = 1, \ldots, N$$

where $\mathbf{x}_i$ is a p-dimensional vector of known explanatory variables, and $\boldsymbol{\beta}$ is a p-dimensional vector of unknown parameters. Maximum likelihood estimates (MLE's) of $\boldsymbol{\beta}$ are most easily obtained by using adjusted dependent variable regression, which is a form of iteratively reweighted least squares (IRLS). In this iterative scheme let the linear predictor $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = \log \lambda_i$ and form the adjusted dependent variable

$$z_i = \eta_i + \frac{1}{\lambda_i}\left(\frac{y_i}{t_i} - \lambda_i\right) \qquad i = 1, \ldots, N.$$

Let $\mathbf{X}$ be the matrix with $\mathbf{x}_i$ as its rows and regress $\mathbf{z} = (z_1, \ldots, z_N)$ on $\mathbf{X}$ with weights $t_i\lambda_i$ to obtain a new estimate of $\boldsymbol{\beta}$. This estimate yields a new linear predictor and hence a new adjusted dependent variable $\mathbf{z}$ for the next iteration. This process continues until changes in $\boldsymbol{\beta}$ between successive iterations are sufficiently small. The initial step may use the data y_i/t_i as the first estimate of λ_i, except for small modifications when y_i is zero. The MLE at convergence, call it $\hat{\boldsymbol{\beta}}$, gives estimates $\hat{\lambda}_i = \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})$.

C.1.2 Assessing Goodness of Fit for the Poisson Model

Typically, goodness of fit for the Poisson model is assessed with a likelihood ratio statistic called the *deviance.* A saturated model has as many parameters as observations, giving a perfect fit. Excluding any scale parameter, the deviance equals twice the log likelihood under the saturated model, minus twice the log likelihood under the reduced model. For the Poisson model the deviance, sometimes labelled G^2, is

$$2\sum\left[y_i \log\left(\frac{y_i}{t_i\hat{\lambda}_i}\right) - (y_i - t_i\hat{\lambda}_i)\right].$$

When this statistic is divided by the scale parameter ϕ it is called the *scaled deviance.* As an example, consider the Normal theory linear model where Y_i is assumed $N(\mu_i, \sigma^2)$. For this case, ϕ equals σ^2 and the deviance is equivalent to the residual sum of squares. For Poisson models ϕ is usually set to one when the model holds.

Another measure of goodness of fit commonly used is the Pearson X^2 statistic which takes the form

$$\sum\frac{(y_i - t_i\hat{\lambda}_i)^2}{t_i\hat{\lambda}_i}$$

and can be derived by expanding G^2 in a second order Taylor series about $t_i\hat{\lambda}_i$. For the Poisson models considered here it is generally argued that both G^2 and X^2 have asymptotic chi squared distributions: however, G^2 is usually preferred for several reasons. First, G^2 constitutes a natural choice because it is likelihood-based. Second, it is additive for nested sets of models when MLE's are used, whereas X^2 in general is not. Although the deviance is preferred for these reasons, there are some drawbacks. For large samples G^2 is likely to reject even good models since it will detect small differences and report them as significant. On the other hand, for sparse data G^2 must be used with caution since it is likely to be a liberal test.

C.1.3 Overdispersed Poisson Counts

Unfortunately, a phenomenon which occurs occasionally with Poisson data is extra-Poisson variation or overdispersion. When analyzing rates by fitting Poisson regression models by the method of maximum likelihood, the variation in the data is often greater than assumed by Poisson sampling theory, resulting in large deviances. In many observational studies the numerators of the rates are counts of rare occurrences, and much exposure may accumulate before any considerable number of events occur. For some cases, due to physical or geographical constraints, it is feasible to observe only limited amounts of exposure. While parameter estimates are often not seriously affected, standard errors are too small and tests of hypotheses are inadequate because the mean-variance relationship is misspecified.

Users of Poisson regression models have long been aware of the phenomenon of overdispersion and methods have been developed to deal with this problem. Some frequentist methods involve the addition of a scale parameter into the likelihood or variance function; the scale parameter can then be estimated from the data. Breslow (1984) adapted a scheme introduced by Williams (1982) for handling extra-Binomial variation in logistic regression models. Efron (1986) considered a class of regression families that model overdispersion while carrying out the usual regression analyses for the mean. They are called *double exponential families* because they enjoy exponential family properties simultaneously for the mean and scale parameters.

C.1.4 Hierarchical Poisson Models

Bayesian methodology provides a convenient framework for the analysis of rates through the use of hierarchical models. The idea is to develop a model that accommodates extra-Poisson variation, yet reduces to the classical model when the model holds. This can be accomplished by expressing prior belief in a log-linear model and incorporating a scale parameter in the prior distribution. Leonard and Novick (1986) and Albert (1988 b) developed Bayesian hierarchical models for the computation of posterior densities when cell frequencies satisfy a log-linear model a priori. Here, two hierarchical Poisson models are proposed that additionally incorporate varying exposures for rates. For both models assume the random counts $Y_1, \ldots, Y_N$ Poisson distributed with mean $t_i\lambda_i$ and consider a Bayesian two-stage prior distribution. At the first stage assume $\lambda_1, \ldots, \lambda_N$ independent, with λ_i distributed according to the conjugate Gamma density such that $E[\lambda_i] = \mu_i$. Furthermore, to reflect belief in the log-linear model let the prior mean μ_i satisfy

$$\log \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$$

noting that μ_i is a function of the known $\mathbf{x}_i$ and unknown $\boldsymbol{\beta}$. Additionally, the first stage Gamma prior includes a parameter comparable to the scale parameter introduced in frequentist approaches for accommodating extra-Poisson variation. Therefore, the second stage prior has $p+1$ unknown hyperparameters: the parameter vector $\boldsymbol{\beta}$, and a scale parameter. The two models differ with regard to the scale parameter. In the first model, called Model A, the posterior estimate of $0 < \gamma < 1$ can be used to assess the adequacy of the log-linear model. As $\gamma \rightarrow 1$, Model A reduces to the classical log-linear model. In fact, it will be shown that the posterior density of γ can be written as a function of the deviance statistic. In the second model, called Model B, the posterior estimate of the scale parameter $\alpha > 0$ has the interpretation of a fixed exposure. As α gets large relative to the t_i, Model B reduces to the classical log-linear model. Later, it will be shown that γ is a transformation of α. For Models A and B the posterior estimate of λ_i compromises between y_i/t_i and the posterior estimate of μ_i, depending on the posterior estimate of the respective scale parameter.

The Bayesian paradigm is intuitively appealing since inference can be made by inspecting the marginal posterior densities of λ_i, $\boldsymbol{\beta}$, γ and α. Unfortunately, this is often an arduous task due to high dimensional integrations in the posterior calculations. Some progress has been made in this area that involves analytic or numerical approximations (see Tierney and Kadane 1986). Recently, Markov chain Monte Carlo (MCMC) methods such as Gibbs sampling have been developed that alleviate the need to calculate difficult integrals; however, MCMC methods can be computer intensive (see Gelfand and Smith 1990, Zeger and Karim 1991, and Dellaportas and Smith 1993 for examples). The focus here is on obtaining the posterior moments of the λ_i and full posterior densities of the hyperparameters using approximate methods in the posterior calculations. Two approximations are crucial to the analyses. In particular, Laplace's method for integrals is used with respect to the parameter vector $\boldsymbol{\beta}$, and $p+1$-dimensional integration is reduced to a much easier one-dimensional numerical integration. In addition, the marginal distribution of Y_i conditional on the hyperparameters is Negative Binomial, and an approximation to this distribution due to Albert (1988) simplifies the posterior calculations where only standard output from a Poisson model is required.

Model A is described in Section 2 where the Poisson hierarchical model is presented in three steps. Step 1 represents the likelihood, and Steps 2 and 3 correspond to the two stages of the prior distribution. In Section 3 approximate methods are given that lead to two important results for making posterior inference. Section 4 presents approximations to the posterior densities of the parameters of interest using the results from Section 3. Section 5 describes Model B in three steps and Sections 6 and 7 present the posterior calculations and posterior inference procedures, respectively. In Section 8 the methods are illustrated with a data example. Models A and B are fit to accident data from this research where the number of head-on accidents observed per 100 million vehicles (exposure) is cross-classified by two factors.

C.2 Poisson Hierarchical Model A

Consider a Poisson hierarchical model with a two-stage prior distribution. Model A can be written in three steps.

Step 1

The first step assumes that conditional on λ_i, the Y_i are independent and
$Y_i|\lambda_i \sim \text{Poisson}(t_i\lambda_i)$.

$$f(Y_i|\lambda_i) = \frac{e^{-t_i\lambda_i}(t_i\lambda_i)^{y_i}}{y_i!} \qquad Y_i = 0, 1, 2 \ldots$$

$$E[Y_i|\lambda_i] = t_i\lambda_i \qquad Var(Y_i|\lambda_i) = t_i\lambda_i$$

The exposures $t_i > 0$ are always known and fixed.

Step 2

The second step represents the first stage of the prior specification. Conditional on the hyperparameters α_i and $\boldsymbol{\beta}$, the λ_i are independent and
$\lambda_i|\alpha_i, \boldsymbol{\beta} \sim \text{Gamma}(\alpha_i\mu_i, \alpha_i)$.

$$\pi(\lambda_i|\alpha_i, \boldsymbol{\beta}) = \frac{\alpha_i^{\alpha_i\mu_i}}{\Gamma(\alpha_i\mu_i)} \lambda_i^{\alpha_i\mu_i - 1} e^{-\alpha_i\lambda_i} \qquad \lambda_i > 0, \;\; \alpha_i > 0$$

$$E[\lambda_i|\alpha_i, \boldsymbol{\beta}] = \mu_i \qquad Var(\lambda_i|\alpha_i, \boldsymbol{\beta}) = \frac{\mu_i}{\alpha_i}$$

At this step prior belief in the log-linear model is expressed and the prior mean is assumed to satisfy

$$\log \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}.$$

The Gamma prior is parameterized so that expectation of λ_i is μ_i where μ_i is a function of the known $\mathbf{x}_i$ and unknown $\boldsymbol{\beta}$. A scale parameter α_i is also incorporated into the prior that affects the variance function. All of the parameters have meaningful interpretations. Noting the similarity between the observed and unobserved rates, the hyperparameter α_i can be interpreted as an a priori exposure. As $\alpha_i \to \infty$, the density of λ_i becomes concentrated about its mean μ_i.

Observed rate	Unobserved rate
$E\left[\frac{Y_i}{t_i} \mid \lambda_i\right] = \lambda_i$	$E[\lambda_i \vert \alpha_i, \beta] = \mu_i$
$Var\left(\frac{Y_i}{t_i} \vert \lambda_i\right) = \frac{\lambda_i}{t_i}$	$Var(\lambda_i \vert \alpha_i, \beta) = \frac{\mu_i}{\alpha_i}$

Step 3

The third step represents the second stage of the prior specification.

$$\pi(\alpha_i, \beta) = \frac{t_i}{(t_i + \alpha_i)^2} \qquad \alpha_i > 0, \;\; \beta \in R^p$$

In this noninformative prior α_i and β are assumed independent. This density is proper with respect to α_i, however, its moments do not exist. Although it resembles an exponential density in appearance, Leonard and Novick (1986) called it a *Cauchy-tail* prior due to its longer tail. The hyperparameter β is given the prior $\pi(\beta) = 1$ which is an improper flat prior commonly used for regression coefficients because of its good frequentist properties. Consider now the change of variable

$$\gamma = \frac{\alpha_i}{\alpha_i + t_i}$$

so that the second stage prior in terms of γ is $\pi(\gamma, \beta) = 1$. Motivation for this prior and the useful transformation to the fixed γ will be given when considering the posterior calculations. A distinguishing feature of Model A is that γ is fixed while α_i and t_i vary. In Section 5 Model B is considered where α is fixed and γ_i and t_i vary.

C.3 Posterior Calculations for Model A

The primary goal of the analysis is to obtain the posterior density of λ_i. Reparameterize α_i to γ and

$$\begin{aligned} f(\lambda_i \vert \mathbf{y}) &= \int_{R^p} \int_0^1 f(\lambda_i, \gamma, \beta \,\vert \mathbf{y}) \, d\gamma \, d\beta \\ &= \int_{R^p} \int_0^1 f(\lambda_i \vert \mathbf{y}, \gamma, \beta) \, f(\gamma, \beta \,\vert \mathbf{y}) \, d\gamma \, d\beta \end{aligned} \tag{C.1}$$

which unfortunately is largely intractable due to the p-dimensional $\boldsymbol{\beta}$. The secondary goal will be to obtain the posterior densities of γ and the individual β_j, $j = 1, \ldots, p$. Using approximate methods in the posterior calculations the secondary goal can be accomplished, and the primary goal can also be accomplished, at least up to the first two moments.

First, consider the two posterior densities in the integrand of C.1. The posterior density $f(\lambda_i|\mathbf{y}, \gamma, \boldsymbol{\beta})$ is tractable, and inspection of its mean will reveal an interpretation for the hyperparameter γ. The posterior density $f(\gamma, \boldsymbol{\beta}|\mathbf{y})$ in the integrand of C.1 can also be derived exactly; however, it will be approximated so that the posterior calculations only depend on standard output from a Poisson log-linear model. The posterior density $f(\gamma|\mathbf{y})$ will be obtained by integrating $\boldsymbol{\beta}$ out of the approximate $f(\gamma, \boldsymbol{\beta}|\mathbf{y})$ using Laplace's method for integrals. As a byproduct of Laplace's method it will be shown that $f(\boldsymbol{\beta}|\mathbf{y}, \gamma)$ is approximately p-variate Normal, and the posterior densities of the individual β_j can then be derived using one-dimensional numerical integrations over γ. Finally, using the above results, approximate methods for computing the posterior moments of λ_i are presented.

C.3.1 The Posterior Density of λ_i Conditional on γ and β

Since the Gamma density is the conjugate prior for the Poisson, it is clear that the first density in the integrand of C.1 is also Gamma distributed. By multiplying the Poisson likelihood by the Gamma prior in steps 1 and 2 of the hierarchical model, $\lambda_i|\mathbf{y}, \gamma, \boldsymbol{\beta} \sim \text{Gamma}(y_i + \alpha_i\mu_i, t_i + \alpha_i)$.

$$f(\lambda_i|\mathbf{y}, \gamma, \boldsymbol{\beta}) = \frac{(t_i + \alpha_i)^{y_i+\alpha_i\mu_i}}{\Gamma(y_i + \alpha_i\mu_i)} \lambda_i^{y_i+\alpha_i\mu_i-1} e^{-(t_i+\alpha_i)\lambda_i}$$

$$E[\lambda_i|\mathbf{y}, \gamma, \boldsymbol{\beta}] = \frac{y_i + \alpha_i\mu_i}{t_i + \alpha_i} \qquad Var(\lambda_i|\mathbf{y}, \gamma, \boldsymbol{\beta}) = \frac{y_i + \alpha_i\mu_i}{(t_i + \alpha_i)^2}$$

A more revealing interpretation of this expectation is given by writing

$$E[\lambda_i|\mathbf{y}, \gamma, \boldsymbol{\beta}] = \left(\frac{t_i}{t_i + \alpha_i}\right)\frac{y_i}{t_i} + \left(\frac{\alpha_i}{t_i + \alpha_i}\right)\mu_i = (1 - \gamma)\frac{y_i}{t_i} + \gamma\mu_i$$

showing that the posterior expectation of λ_i conditional on γ and $\boldsymbol{\beta}$ is a weighted average of the observed rate and the prior mean μ_i, with weight $\gamma = \alpha_i/(t_i + \alpha_i)$. The denominator $t_i + \alpha_i$ can be interpreted as an effective sample size, while the numerator α_i represents exposure in the prior. As $\gamma \to 1$, $E[\lambda_i] \to \mu_i$, supporting belief in the log-linear model.

With this interpretation of γ, motivation for the second stage prior can be given. If γ is assumed Uniform(0,1) a priori, then a change of variable shows that the density of $\alpha_i = \gamma\, t_i/(1-\gamma)$ is

$$\pi(\alpha_i) = \frac{t_i}{(t_i + \alpha_i)^2} \qquad \alpha_i > 0$$

which is the second stage prior, and this explains its vague, noninformative nature. Although this density is proper, its moments are infinite. Note that α_i is the same multiple of t_i for each i.

C.3.2 Approximating the Joint Posterior Density of γ and β

The second density in the integrand of $f(\lambda_i|\mathbf{y})$ is

$$f(\gamma, \beta|\mathbf{y}) = C(\mathbf{y})\, m(\mathbf{Y}|\gamma, \beta)\, \pi(\gamma, \beta)$$

where $C(\mathbf{y})$ is a constant of proportionality, $m(\mathbf{Y}|\gamma,\beta)$ is the joint marginal mass function of $\mathbf{Y}$ conditional on γ and β, and $\pi(\gamma,\beta) = 1$ is the second stage prior. The only real contribution comes from m. For a single Y_i

$$\begin{aligned} m(Y_i|\gamma,\beta) &= \int_0^\infty f(Y_i|\lambda_i)\, \pi(\lambda_i|\gamma,\beta)\, d\lambda_i \\ &= \frac{\Gamma(y_i + \alpha_i\mu_i)}{\Gamma(\alpha_i\mu_i)\, y_i\,!} (1-\gamma)^{y_i}\, \gamma^{\alpha_i\mu_i} \end{aligned}$$

$$E[Y_i|\gamma,\beta] = t_i\mu_i \qquad Var(Y_i|\gamma,\beta) = t_i\mu_i + \frac{t_i^2\mu_i}{\alpha_i} = \frac{t_i\mu_i}{\gamma}$$

and $Y_i|\gamma,\beta$ is Negative Binomial. Since this distribution is conditional on γ, and because γ is a one-to-one transformation of α_i, these two terms are used interchangeably for notational convenience. It can be shown using moment generating functions that as $\gamma \to 1$, or similarly as $\alpha_i \to \infty$,

$$m(Y_i|\gamma,\beta) \xrightarrow{d} \frac{e^{-t_i\mu_i}(t_i\mu_i)^{y_i}}{y_i\,!}.$$

Hence the Negative Binomial converges in distribution to a Poisson distribution with mean $t_i\mu_i$ as $\gamma \to 1$. Moreover, since the Y_i are assumed independent

$$m(\mathbf{Y}|\gamma,\beta) = \prod_{i=1}^{N} m(Y_i|\gamma,\beta).$$

From a frequentist point of view, the joint marginal mass function m can be considered a likelihood function in a model for overdispersed counts. The location parameter $\boldsymbol{\beta}$ and the scale parameter γ can be estimated by maximum likelihood giving estimates $\hat{\mu}_i$ and $\hat{\gamma}$. In the Bayesian context, however, the focus is on the posterior density

$$f(\gamma, \boldsymbol{\beta} | \mathbf{y}) = C(\mathbf{y}) \prod_{i=1}^{N} m(Y_i | \gamma, \boldsymbol{\beta}), \tag{C.2}$$

and in particular $f(\gamma|\mathbf{y})$ and $f(\boldsymbol{\beta}|\mathbf{y})$. Inference concerning γ will be made through $f(\gamma|\mathbf{y})$ by integrating $\boldsymbol{\beta}$ out of C.2 using Laplace's method for integrals. In order to accomplish this it is necessary to find the $\boldsymbol{\beta}$ that maximizes C.2; however, using the exact posterior is problematic for two reasons. First, the $\boldsymbol{\beta}$ that maximizes C.2 depends on γ. Second, a maximization routine such as the Newton Raphson algorithm is needed which does not take advantage of the IRLS algorithm. In addition, $m(Y_i|\gamma, \boldsymbol{\beta})$ is not an exponential family and experience has shown it to be unstable for small values of γ.

Instead of working with C.2 directly, it is convenient to use an approximation to $m(Y_i|\gamma, \boldsymbol{\beta})$ given by Albert (1988 b) for the case when all $t_i = 1$, and adapted here for varying t_i. Consider the random variable $W_i = \gamma Y_i$, which is a simple scale transformation of Y_i. Then

$$E[W_i|\gamma, \boldsymbol{\beta}] = \gamma t_i \mu_i \qquad Var(W_i|\gamma, \boldsymbol{\beta}) = \gamma t_i \mu_i$$

and the mean equals the variance. Although W_i is not discrete, suppose its density function is given by the Poisson form

$$f(W_i|\gamma, \boldsymbol{\beta}) = \frac{e^{-\gamma t_i \mu_i} (\gamma t_i \mu_i)^{w_i}}{\Gamma(w_i + 1)} \qquad w_i > 0.$$

Transforming back to Y_i

$$m(Y_i|\gamma, \boldsymbol{\beta}) \approx \gamma f_{w_i}(\gamma Y_i|\gamma, \boldsymbol{\beta}) = \frac{\gamma\, e^{-\gamma t_i \mu_i} (\gamma t_i \mu_i)^{\gamma y_i}}{\Gamma(\gamma y_i + 1)}. \tag{C.3}$$

It should be noted that this distribution, sometimes called a *quasi-likelihood*, does not sum exactly to one. In addition, the expectation and variance are approximately $t_i \mu_i$ and $t_i \mu_i / \gamma$, respectively. A feature of this approximation, however, is that it differs from the double Poisson family used by Efron (1986) only by the Stirling's formula

$$\log \Gamma(\gamma y_i + 1) = (\gamma y_i + \frac{1}{2}) \log(\gamma y_i) - \gamma y_i + \frac{1}{2} \log(2\pi) + O[(\gamma y_i)^{-1}]. \tag{C.4}$$

Also, as $\gamma \to 1$ in approximation C.3

$$m(Y_i|\gamma, \boldsymbol{\beta}) \xrightarrow{d} \frac{e^{-t_i \mu_i} (t_i \mu_i)^{y_i}}{y_i\,!}$$

showing that, like the exact Negative Binomial, the quasi-likelihood also reduces to a Poisson distribution in the limit. Therefore, C.3 should be accurate as $\gamma \to 1$. Finally, the approximation to the joint posterior density of γ and $\boldsymbol{\beta}$ is

$$f(\gamma, \boldsymbol{\beta}\,|\mathbf{y}) \approx C(\mathbf{y})\,\gamma^N \prod_{i=1}^{N} \frac{e^{-\gamma t_i \mu_i}(\gamma t_i \mu_i)^{\gamma y_i}}{\Gamma(\gamma y_i + 1)}. \tag{C.5}$$

It will be shown that the $\boldsymbol{\beta}$ that maximizes C.5 does not depend on γ. Furthermore, standard output from a Poisson model can be used that takes advantage of the IRLS algorithm.

C.3.3 Approximating $f(\gamma|\mathbf{y})$ Using Laplace's Method

Obtaining $f(\gamma|\mathbf{y})$ requires integrating $\boldsymbol{\beta}$ out of $f(\gamma, \boldsymbol{\beta}\,|\mathbf{y})$. Following methods proposed by Leonard and Novick (1986) and Tierney and Kadane (1986) this difficult p-dimensional integration is replaced by a much easier p-dimensional maximization using Laplace's method. Excluding some constants the log of density C.5 is

$$\log f(\gamma, \boldsymbol{\beta}\,|\mathbf{y}) \approx N \log \gamma + \gamma \log \gamma \sum y_i - \sum \log \Gamma(\gamma y_i + 1) + \gamma\,\ell(\mathbf{y}, \boldsymbol{\beta}) \tag{C.6}$$

where $\ell(\mathbf{y}, \boldsymbol{\beta}) = \sum y_i \log(t_i \mu_i) - t_i \mu_i$. Consider a second order Taylor series expansion about the value $\hat{\boldsymbol{\beta}}$ that maximizes ℓ. Then

$$\ell(\mathbf{y}, \boldsymbol{\beta}) \approx \ell(\mathbf{y}, \hat{\boldsymbol{\beta}}) - \frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{R} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \tag{C.7}$$

where

$$\mathbf{R} = -\left.\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}}$$

denotes the negative Hessian matrix evaluated at the maximum. The function ℓ is a standard Poisson log likelihood; as a result, the quantities $\hat{\boldsymbol{\beta}}$ and $\mathbf{R}$ are readily available as output from a standard Poisson model using IRLS. Furthermore, estimation of $\boldsymbol{\beta}$ does not depend on γ. Putting C.7 into C.6 and exponentiating gives, up to a multiplicative constant, the approximation

$$\begin{aligned} f(\gamma, \boldsymbol{\beta}\,|\mathbf{y}) \;&\approx\; \exp\left[\gamma\,\ell(\mathbf{y}, \hat{\boldsymbol{\beta}}) + N \log \gamma + \gamma \log \gamma \sum y_i - \sum \log \Gamma(\gamma y_i + 1)\right] \\ &\quad \times \exp\left[-\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \gamma \mathbf{R} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]. \end{aligned} \tag{C.8}$$

From C.8 two results can be deduced:

Result 1

Integrating $\boldsymbol{\beta}$ out of C.8, the approximate posterior density of γ is

$$\begin{aligned} f(\gamma|\mathbf{y}) &\approx C_1\,(2\pi)^{p/2}\,\gamma^{-p/2}\,|\mathbf{R}|^{-1/2}\,f(\gamma,\hat{\boldsymbol{\beta}}\,|\mathbf{y}) \\ &= C_2\,\exp\left[-\frac{p}{2}\log\gamma + \gamma\,\ell(\mathbf{y},\hat{\boldsymbol{\beta}})\right] \\ &\quad \times \exp\left[N\log\gamma + \gamma\log\gamma\sum y_i - \sum\log\Gamma(\gamma y_i + 1)\right] \end{aligned}$$

noting that $|\gamma\mathbf{R}|^{-1/2} = \gamma^{-p/2}|\mathbf{R}|^{-1/2}$. Moreover, $(2\pi)^{p/2}$ and $|\mathbf{R}|^{-1/2}$ are constants since they do not depend on γ, and C_2 is the constant of integration.

Result 2

Hold γ fixed in C.8 and the posterior density of $\boldsymbol{\beta}$ is approximately p-variate Normal.

$$\boldsymbol{\beta}\,|\mathbf{y},\gamma \;\dot{\sim}\; N_p(\hat{\boldsymbol{\beta}},(\gamma\mathbf{R})^{-1})$$

These two results are extremely useful in the posterior inferences that follow.

C.4 Posterior Inference For Model A

C.4.1 Posterior Inference About the Hyperparameters

The posterior density of γ is approximated using Result 1. Since Model A reduces to the classical Poisson model as $\gamma \to 1$, examination of $f(\gamma|\mathbf{y})$ provides a way of assessing the adequacy of the log-linear model. Other quantities such as posterior moments of γ can be calculated using one-dimensional numerical integrations. In addition, evidence of a relationship between $f(\gamma|\mathbf{y})$ and the classical deviance statistic exists. Replace $\log\Gamma(\gamma y_i + 1)$ in Result 1 by the Stirling's formula approximation in C.4 and it can be shown that

$$f(\gamma|\mathbf{y}) \propto \gamma^{(N-p)/2}\,\exp\left[-\frac{\gamma}{2}\sum D(y_i\,;t_i\hat{\mu}_i)\right]$$

where $\sum D(y_i\,;t_i\hat{\mu}_i)$ is the deviance statistic. This approximation has the form of a *truncated* Gamma density because γ is restricted to the open interval (0,1). Nevertheless, this demonstrates that as the deviance gets small relative to $N-p$, the posterior density of γ tends to a degenerate point mass at one.

Using Result 2, $\beta_j|\mathbf{y},\gamma \sim N(\hat{\beta}_j, \gamma^{-1}r_j)$ where r_j is the jth diagonal element of $\mathbf{R}^{-1}$. Approximations to the full posterior densities of the individual β_j may be obtained by writing

$$\begin{aligned} f(\beta_j|\mathbf{y}) &= \int_0^1 f(\beta_j,\gamma|\mathbf{y})\,d\gamma \\ &= \int_0^1 f(\beta_j|\mathbf{y},\gamma)\,f(\gamma|\mathbf{y})\,d\gamma. \end{aligned} \tag{C.9}$$

The posterior moments of $\boldsymbol{\beta}$ can be calculated and written in closed form

$$\begin{aligned} E[\boldsymbol{\beta}|\mathbf{y}] &= E_{\gamma|y}[E[\boldsymbol{\beta}|\mathbf{y},\gamma]] \;\approx\; \hat{\boldsymbol{\beta}} \\ Var(\boldsymbol{\beta}|\mathbf{y}) &= E_{\gamma|y}[Var(\boldsymbol{\beta}|\mathbf{y},\gamma)] + Var_{\gamma|y}(E[\boldsymbol{\beta}|\mathbf{y},\gamma]) \\ &\approx E_{\gamma|y}[\gamma^{-1}]\,\mathbf{R}^{-1} \end{aligned}$$

where subscripts denote distributions over which expectations are taken. Here, the posterior mean of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}$, the usual MLE from a standard Poisson model. However, the usual covariance matrix $\mathbf{R}^{-1}$ is inflated by the expected value of γ^{-1}. In practice, $f(\beta_j|\mathbf{y})$ is plotted by evaluating the one-dimensional integral in C.9 over a range of probable β_j values.

C.4.2 Posterior Inference About the Rates

Making inference about the hyperparameters only involves Results 1 and 2 and the ability to perform one-dimensional numerical integration. Although the integral in C.1 is largely intractable, Results 1 and 2 can be exploited here also to approximate the first two posterior moments of $f(\lambda_i|\mathbf{y})$. In particular,

$$\begin{aligned} E[\lambda_i|\mathbf{y}] &= E_{\gamma,\beta|y}[E[\lambda_i|\mathbf{y},\gamma,\boldsymbol{\beta}]] \\ &= E_{\gamma,\beta|y}[(1-\gamma)\frac{y_i}{t_i} + \gamma\mu_i] \\ &= \frac{y_i}{t_i}[1 - E_{\gamma,\beta|y}[\gamma]] + E_{\gamma,\beta|y}[\gamma\exp(\mathbf{x}_i^T\boldsymbol{\beta})] \end{aligned}$$

and it remains to calculate two expectations with respect to $f(\gamma,\boldsymbol{\beta}|\mathbf{y})$. However, the p-dimensional integrations over $\boldsymbol{\beta}$ in these expectations drop out from Result 2 since

$$\begin{aligned} E_{\gamma,\beta|y}[\gamma] &= \int_0^1\int_{R^p} \gamma\, f(\gamma,\boldsymbol{\beta}|\mathbf{y})\,d\boldsymbol{\beta}\,d\gamma \\ &= \int_0^1 \gamma\left[\int_{R^p} f(\boldsymbol{\beta}|\mathbf{y},\gamma)\,d\boldsymbol{\beta}\right] f(\gamma|\mathbf{y})\,d\gamma \\ &= \int_0^1 \gamma\, f(\gamma|\mathbf{y})\,d\gamma \;=\; E_{\gamma|y}[\gamma] \end{aligned}$$

and

$$\begin{aligned}
E_{\gamma,\beta|y}[\gamma \exp(\mathbf{x}_i^T\boldsymbol{\beta})] &= \int_0^1 \int_{R^P} \gamma \exp(\mathbf{x}_i^T\boldsymbol{\beta})\, f(\gamma,\boldsymbol{\beta}\,|\mathbf{y})\, d\boldsymbol{\beta}\, d\gamma \\
&= \int_0^1 \gamma \left[\int_{R^P} \exp(\mathbf{x}_i^T\boldsymbol{\beta})\, f(\boldsymbol{\beta}\,|\mathbf{y},\gamma)\, d\boldsymbol{\beta}\right] f(\gamma|\mathbf{y})\, d\gamma \qquad \text{(C.10)} \\
&\approx \int_0^1 \gamma \exp\left[\mathbf{x}_i^T\hat{\boldsymbol{\beta}} + \frac{1}{2}\mathbf{x}_i^T(\gamma\mathbf{R})^{-1}\mathbf{x}_i\right] f(\gamma|\mathbf{y})\, d\gamma \\
&= \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}}) E_{\gamma|y}\left[\gamma \exp\left(\frac{1}{2\gamma}\mathbf{x}_i^T\mathbf{R}^{-1}\mathbf{x}_i\right)\right]
\end{aligned}$$

where the integral in brackets in C.10 is the moment generating function of a p-variate Normal. As a result, the posterior mean of λ_i can be written in terms of standard output from a Poisson model

$$E[\lambda_i|\mathbf{y}] \approx \frac{y_i}{t_i}\left[1 - E_{\gamma|y}[\gamma]\right] + \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})\, E_{\gamma|y}\left[\gamma \exp\left(\frac{1}{2\gamma}\mathbf{x}_i^T\mathbf{R}^{-1}\mathbf{x}_i\right)\right]$$

where expectations with respect to $f(\gamma|\mathbf{y})$ only require one-dimensional numerical integrations. To approximate the variance consider the second posterior moment

$$\begin{aligned}
E[\lambda_i^2|\mathbf{y}] &= E_{\gamma,\beta|y}[E[\lambda_i^2|\mathbf{y},\gamma,\boldsymbol{\beta}]] \\
&= E_{\gamma,\beta|y}\left[\frac{y_i + \alpha_i\mu_i + (y_i + \alpha_i\mu_i)^2}{(t_i + \alpha_i)^2}\right]
\end{aligned}$$

and after writing in terms of γ, the calculations are similar to those in C.10. After some simplifications

$$\begin{aligned}
Var(\lambda_i|\mathbf{y}) &\approx \frac{y_i(y_i+1)}{t_i^2}\, E_{\gamma|y}[(1-\gamma)^2] \\
&+ \frac{(2y_i+1)}{t_i} \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})\, E_{\gamma|y}\left[\gamma\,(1-\gamma) \exp\left(\frac{1}{2\gamma}\mathbf{x}_i^T\mathbf{R}^{-1}\mathbf{x}_i\right)\right] \\
&+ \exp(2\mathbf{x}_i^T\hat{\boldsymbol{\beta}})\, E_{\gamma|y}\left[\gamma^2 \exp\left(\frac{2}{\gamma}\mathbf{x}_i^T\mathbf{R}^{-1}\mathbf{x}_i\right)\right] \\
&- E[\lambda_i|\mathbf{y}]^2.
\end{aligned}$$

Using the same procedures

$$\begin{aligned}
E[\mu_i|\mathbf{y}] &= E[\exp(\mathbf{x}_i^T\boldsymbol{\beta})|\mathbf{y}] = E_{\gamma|y}[E[\exp(\mathbf{x}_i^T\boldsymbol{\beta})|\mathbf{y},\gamma]] \\
&\approx \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}})\, E_{\gamma|y}\left[\exp\left(\frac{1}{2\gamma}\mathbf{x}_i^T\mathbf{R}^{-1}\mathbf{x}_i\right)\right] \\
&> \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}}).
\end{aligned}$$

Note that the posterior rates obtained under the log-linear model are always greater than those using the usual frequentist inferences.

C.5 Poisson Hierarchical Model B

Model B is very similar to Model A except for one important difference: in Model B, α is held fixed and γ_i is allowed to vary. The transformation from α to γ_i is the same as in Model A; however, the posterior estimate of α has the interpretation of a *fixed* exposure. As in Model A, the goal is to obtain posterior moments of the λ_i and full posterior densities of the hyperparameters using approximate methods in the posterior calculations. Again, this is accomplished using standard output from Poisson models where many of the calculations in Model A carry over to Model B with only slight modifications. The differences will be highlighted. Consider Model B written in three steps with a two-stage prior distribution.

Step 1

The first step is the same as Model A. Conditional on λ_i, the Y_i are independent and $Y_i|\lambda_i \sim \text{Poisson}(t_i\lambda_i)$.

Step 2

The second step is the same as Model A except replace α_i by α. Conditional on the hyperparameters α and $\boldsymbol{\beta}$, the λ_i are independent and $\lambda_i|\alpha, \boldsymbol{\beta} \sim \text{Gamma}(\alpha\mu_i, \alpha)$.

Step 3

The second stage hyperprior for Model B is also noninformative.

$$\pi(\alpha, \boldsymbol{\beta}) = \frac{\nu}{(\nu + \alpha)^2} \qquad \alpha > 0, \;\; \boldsymbol{\beta} \in R^p$$

As in Model A this is a proper prior with respect to α and the moments do not exist. The parameter ν could be regarded as unknown and assigned a prior; however, that will not be pursued here and ν is considered known. In practice, a value of ν needs to be chosen. Leonard and Novick (1986) used this prior when $\nu = t_i = 1$. Estimation of α depends on the t_i since α is interpreted as a fixed exposure, and it makes sense to somehow incorporate the t_i into the prior. Note, however, that ν cannot equal t_i as in Model A because ν must be fixed. Some justification for this prior will be given when considering the posterior calculations. Note also that unlike Model A, no transformation to γ_i is made here since $f(\alpha|\mathbf{y})$ will be desired when making posterior inference.

C.6 Posterior Calculations for Model B

C.6.1 The Posterior Density of λ_i Conditional on α and β

For any distribution considered in Model A that is conditional on α_i, it is only necessary to replace α_i by α and $\lambda_i|\mathbf{y},\alpha,\boldsymbol{\beta} \sim \text{Gamma}(y_i+\alpha\mu_i, t_i+\alpha)$. The expectation is

$$E[\lambda_i|\mathbf{y},\alpha,\boldsymbol{\beta}] = \left(\frac{t_i}{t_i+\alpha}\right)\frac{y_i}{t_i} + \left(\frac{\alpha}{t_i+\alpha}\right)\mu_i = (1-\gamma_i)\frac{y_i}{t_i} + \gamma_i\mu_i$$

and the posterior mean conditional on α and $\boldsymbol{\beta}$ is a weighted average of the observed rate and the prior mean with weight γ_i. Here, γ_i is a monotone decreasing function of t_i and as $t_i \to \infty$, $E[\lambda_i] \to y_i/t_i$, which is intuitively appealing since the observed rate is consistent in t_i. On the other hand, when the data are well explained by the Poisson model, $\alpha \to \infty$ and $E[\lambda_i] \to \mu_i$.

Consider motivation for the second stage prior by assuming that $\gamma_i \sim \text{Uniform}(0,1)$. A change of variable shows that the distribution of $\alpha = \gamma_i t_i/(1-\gamma_i)$ is

$$\pi(\alpha) = \frac{t_i}{(t_i+\alpha)^2} \qquad \alpha > 0.$$

Since α is fixed, t_i cannot vary and replacing t_i by ν gives the second stage prior. Estimation of α depends on the exposures, however, and ν should be some function of the t_i. Christiansen (1992) chose $\nu = \min t_i$ and this works well in practice, yet other choices are possible. Larger choices of ν will encourage more shrinkage towards the posterior estimates of μ_i. Even so, this prior is sufficiently vague so that the data should dominate the posterior density. In addition, other priors including improper ones are viable in this situation; however, care must be taken to ensure that the posterior mode of α is finite. For example, the flat prior $\pi(\alpha,\boldsymbol{\beta}) = 1$ is innappropriate in cases when the Poisson model holds since the posterior mode of α may not exist. In practice, it is useful to divide the t_i by some constant when the exposures are large so that the posterior estimate of α is reasonable. This will only affect the intercept term of the regression parameter $\boldsymbol{\beta}$.

C.6.2 Approximating the Joint Posterior Density of α and β

The joint posterior distribution of α and $\boldsymbol{\beta}$ is

$$f(\alpha,\boldsymbol{\beta}|\mathbf{y}) = C(\mathbf{y})\,m(\mathbf{Y}|\alpha,\boldsymbol{\beta})\,\pi(\alpha,\boldsymbol{\beta})$$

where $C(\mathbf{y})$ is a constant term, $m(\mathbf{Y}|\alpha,\boldsymbol{\beta})$ is the marginal distribution of $\mathbf{Y}$ conditional on α and $\boldsymbol{\beta}$, and $\pi(\alpha,\boldsymbol{\beta})$ is the second stage prior. Since m is conditional on α, it is the same as in Model A with α_i replaced by α, and for a single Y_i

$$m(Y_i|\alpha,\boldsymbol{\beta}) = \frac{\Gamma(y_i+\alpha\mu_i)}{\Gamma(\alpha\mu_i)\,y_i\,!}(1-\gamma_i)^{y_i}\,\gamma_i^{\alpha\mu_i}$$

$$E[Y_i|\alpha,\boldsymbol{\beta}] = t_i\mu_i \qquad Var(Y_i|\alpha,\boldsymbol{\beta}) = t_i\mu_i + \frac{t_i^2\mu_i}{\alpha} = \frac{t_i\mu_i}{\gamma_i}.$$

To facilitate the use of standard output from Poisson models in the posterior calculations the Negative Binomial distribution can be approximated by

$$m(Y_i|\alpha,\boldsymbol{\beta}) \approx \frac{\gamma_i\,e^{-\gamma_i t_i\mu_i}\,(\gamma_i t_i\mu_i)^{\gamma_i y_i}}{\Gamma(\gamma_i y_i+1)}$$

and an approximation to the joint posterior distribution of α and $\boldsymbol{\beta}$ is

$$f(\alpha,\boldsymbol{\beta}|\mathbf{y}) \approx C(\mathbf{y})\left[\prod_{i=1}^{N}\frac{\gamma_i\,e^{-\gamma_i t_i\mu_i}\,(\gamma_i t_i\mu_i)^{\gamma_i y_i}}{\Gamma(\gamma_i y_i+1)}\right]\frac{\nu}{(\nu+\alpha)^2}. \tag{C.11}$$

Note that $f(\alpha,\boldsymbol{\beta}|\mathbf{y})$ is written in terms of γ_i only where it depends on m and a change of variable is not necessary since m is conditional on α. However, $\pi(\alpha,\boldsymbol{\beta})$ is written in terms of α.

C.6.3 Approximating $f(\alpha|\mathbf{y})$ Using Laplace's Method

Laplace's method for integrals can be used to obtain $f(\alpha|\mathbf{y})$ by integrating $\boldsymbol{\beta}$ out of $f(\alpha,\boldsymbol{\beta}|\mathbf{y})$. Up to an additive constant the log of density C.11 is

$$\begin{aligned}\log f(\alpha,\boldsymbol{\beta}|\mathbf{y}) &\approx -2\log(\nu+\alpha) + \ell(\boldsymbol{\gamma},\mathbf{y},\boldsymbol{\beta}) \\ &+ \sum \log\gamma_i + \gamma_i y_i\log\gamma_i - \log\Gamma(\gamma_i y_i+1)\end{aligned} \tag{C.12}$$

where $\ell(\boldsymbol{\gamma},\mathbf{y},\boldsymbol{\beta}) = \sum\gamma_i[y_i\log(t_i\mu_i) - t_i\mu_i]$. Since γ_i is not fixed, it cannot be taken outside the summation as in Model A and ℓ is a weighted Poisson log likelihood. This is a crucial difference between Models A and B because the $\boldsymbol{\beta}$ that maximizes ℓ now depends on α. Expand ℓ in a second order Taylor series about the value $\tilde{\boldsymbol{\beta}}_\alpha$ that maximizes ℓ for α fixed and

$$\ell(\boldsymbol{\gamma},\mathbf{y},\boldsymbol{\beta}) \approx \ell(\boldsymbol{\gamma},\mathbf{y},\tilde{\boldsymbol{\beta}}_\alpha) - \frac{1}{2}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_\alpha)^T\mathbf{R}_\alpha(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_\alpha) \tag{C.13}$$

where

$$\mathbf{R}_\alpha = -\left.\frac{\partial^2\ell}{\partial\boldsymbol{\beta}\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}_\alpha}$$

denotes the negative Hessian matrix evaluated at the maximum. While estimation of β depends on α, the quantities $\tilde{\beta}_\alpha$ and $\mathbf{R}_\alpha$ are easily obtained as output from a Poisson model with γ_i declared as prior weights. Putting C.13 into C.12 and exponentiating gives an approximation to C.11 up to a multiplicative constant as

$$\begin{aligned} f(\alpha,\beta|\mathbf{y}) &\approx \exp\left[-2\log(\nu+\alpha)+\ell(\boldsymbol{\gamma},\mathbf{y},\tilde{\beta}_\alpha)\right] \\ &\quad \times \exp\left[\sum \log\gamma_i + \gamma_i y_i \log\gamma_i - \log\Gamma(\gamma_i y_i + 1)\right] \\ &\quad \times \exp\left[-\frac{1}{2}(\beta-\tilde{\beta}_\alpha)^T \mathbf{R}_\alpha (\beta-\tilde{\beta}_\alpha)\right]. \end{aligned} \tag{C.14}$$

From C.14 two results can be deduced:

Result 3

Integrating β out of C.14, the approximate posterior density of α is

$$\begin{aligned} f(\alpha|\mathbf{y}) &\approx C_1\,(2\pi)^{p/2}\,|\mathbf{R}_\alpha|^{-1/2}\,f(\alpha,\tilde{\beta}_\alpha|\mathbf{y}) \\ &= C_2\,\exp\left[-\frac{1}{2}\log|\mathbf{R}_\alpha| - 2\log(\nu+\alpha)+\ell(\boldsymbol{\gamma},\mathbf{y},\tilde{\beta}_\alpha)\right] \\ &\quad \times \exp\left[\sum \log\gamma_i + \gamma_i y_i \log\gamma_i - \log\Gamma(\gamma_i y_i + 1)\right]. \end{aligned}$$

Note that unlike Model A, the determinant $|\mathbf{R}_\alpha|$ must be calculated since it depends on α.

Result 4

Hold α fixed in C.14 and the posterior density of β is approximately p-variate Normal.

$$\beta|\mathbf{y},\alpha \mathrel{\dot\sim} N_p(\tilde{\beta}_\alpha, \mathbf{R}_\alpha^{-1})$$

C.7 Posterior Inference For Model B

C.7.1 Posterior Inference About the Hyperparameters

Posterior inference concerning α is made using Result 3. As $\alpha \to \infty$, Model B reduces to the classical Poisson model. Note that $f(\alpha|\mathbf{y})$ is defined for $\alpha > 0$; however, in practice a grid is constructed over probable values of α and a weighted Poisson model is fit to obtain $\tilde{\boldsymbol{\beta}}_\alpha$ and $\mathbf{R}_\alpha$ for each α. Therefore, quantities such as posterior moments of α are calculated using one-dimensional summations.

Using Result 4, $\beta_j|\mathbf{y},\alpha \sim N(\tilde{\beta}_{j\alpha}, r_{j\alpha})$ where $r_{j\alpha}$ is the jth diagonal element of $\mathbf{R}_\alpha^{-1}$, and approximations to the posterior densities of the individual β_j are given by

$$\begin{aligned} f(\beta_j|\mathbf{y}) &= \int f(\beta_j,\alpha|\mathbf{y})\,d\alpha \\ &= \int f(\beta_j|\mathbf{y},\alpha)\,f(\alpha|\mathbf{y})\,d\alpha \\ &\approx \sum_\alpha f(\beta_j|\mathbf{y},\alpha)\,f(\alpha|\mathbf{y}). \end{aligned}$$

The posterior moments of $\boldsymbol{\beta}$ can be written as

$$\begin{aligned} E[\boldsymbol{\beta}|\mathbf{y}] &= E_{\alpha|y}[E[\boldsymbol{\beta}|\mathbf{y},\alpha]] \;\approx\; E_{\alpha|y}[\tilde{\boldsymbol{\beta}}_\alpha] \\ Var(\boldsymbol{\beta}|\mathbf{y}) &= E_{\alpha|y}[Var(\boldsymbol{\beta}|\mathbf{y},\alpha)] + Var_{\alpha|y}(E[\boldsymbol{\beta}|\mathbf{y},\alpha]) \\ &\approx E_{\alpha|y}[\mathbf{R}_\alpha^{-1}] + Var_{\alpha|y}(\tilde{\boldsymbol{\beta}}_\alpha) \end{aligned}$$

showing that the posterior mean of $\boldsymbol{\beta}$ is a weighted average of the $\tilde{\boldsymbol{\beta}}_\alpha$ with the posterior density of α as the weighting density. Unlike Model A, the second term in the posterior variance of $\boldsymbol{\beta}$ does not vanish, and it is most easily calculated by

$$Var_{\alpha|y}(\tilde{\boldsymbol{\beta}}_\alpha) = E_{\alpha|y}[\tilde{\boldsymbol{\beta}}_\alpha\tilde{\boldsymbol{\beta}}_\alpha^T] - \left(E_{\alpha|y}[\tilde{\boldsymbol{\beta}}_\alpha]\right)\left(E_{\alpha|y}[\tilde{\boldsymbol{\beta}}_\alpha]\right)^T.$$

C.7.2 Posterior Inference About the Rates

The first two posterior moments of λ_i are found as in Model A, with slight modifications. In particular

$$\begin{aligned} E[\lambda_i|\mathbf{y}] &= E_{\alpha,\beta|y}[E[\lambda_i|\mathbf{y},\alpha,\boldsymbol{\beta}]] \\ &= E_{\alpha,\beta|y}\left[\left(\frac{t_i}{t_i+\alpha}\right)\frac{y_i}{t_i} + \left(\frac{\alpha}{t_i+\alpha}\right)\mu_i\right] \\ &\approx y_i\,E_{\alpha|y}\left[\frac{1}{t_i+\alpha}\right] + E_{\alpha|y}\left[\left(\frac{\alpha}{t_i+\alpha}\right)\exp\left(\mathbf{x}_i^T\tilde{\boldsymbol{\beta}}_\alpha + \frac{1}{2}\mathbf{x}_i^T\mathbf{R}_\alpha^{-1}\mathbf{x}_i\right)\right] \end{aligned}$$

which depends on Results 3 and 4. The second posterior moment can be written

$$\begin{aligned} E[\lambda_i^2|\mathbf{y}] &= E_{\alpha,\beta|y}[E[\lambda_i^2|\mathbf{y},\alpha,\boldsymbol{\beta}]] \\ &= E_{\alpha,\beta|y}\left[\frac{y_i+\alpha\mu_i+(y_i+\alpha\mu_i)^2}{(t_i+\alpha)^2}\right] \end{aligned}$$

and after some simplifications

$$\begin{aligned} Var(\lambda_i|\mathbf{y}) &\approx y_i(1+y_i)\, E_{\alpha|y}\left[\frac{1}{(t_i+\alpha)^2}\right] \\ &+ (2y_i+1)\, E_{\alpha|y}\left[\frac{\alpha}{(t_i+\alpha)^2}\exp\left(\mathbf{x}_i^T\tilde{\boldsymbol{\beta}}_\alpha + \frac{1}{2}\mathbf{x}_i^T\mathbf{R}_\alpha^{-1}\mathbf{x}_i\right)\right] \\ &+ E_{\alpha|y}\left[\left(\frac{\alpha}{t_i+\alpha}\right)^2 \exp\left(2\mathbf{x}_i^T\tilde{\boldsymbol{\beta}}_\alpha + 2\mathbf{x}_i^T\mathbf{R}_\alpha^{-1}\mathbf{x}_i\right)\right] \\ &- E[\lambda_i|\mathbf{y}]^2. \end{aligned}$$

The same methods yield the approximation

$$\begin{aligned} E[\mu_i|\mathbf{y}] &= E[\exp(\mathbf{x}_i^T\boldsymbol{\beta}\,|\mathbf{y})] = E_{\alpha|y}[E[\exp(\mathbf{x}_i^T\boldsymbol{\beta})|\mathbf{y},\alpha]] \\ &\approx E_{\alpha|y}\left[\exp\left(\mathbf{x}_i^T\tilde{\boldsymbol{\beta}}_\alpha + \frac{1}{2}\mathbf{x}_i^T\mathbf{R}_\alpha^{-1}\mathbf{x}_i\right)\right]. \end{aligned}$$

The calculations for these expectations only require one-dimensional summations over $f(\alpha|\mathbf{y})$. For example, the posterior mean of α is calculated as

$$E_{\alpha|y}[\alpha] \approx \sum_\alpha \alpha\, f(\alpha|\mathbf{y}).$$

C.8 Data Example

C.8.1 Accident Data Example

In this example the methods of Models A and B are illustrated with the assumption that the nonexchangeable means μ_i satisfy a log-linear model a priori. Consider the data presented in the left panel of Table 4.19. The y_i are the accident counts, the t_i are the exposures, and y_i/t_i are the observed rates. It is standard practice to divide the t_i by a constant to make the magnitude of the rates reasonable for analysis. The data are cross-classified by the two factors CT and TY which represent the configurations curve-tangent (Curve=1, Tangent=2), and type of intersection (T=1, Y=2), respectively.

Suppose inference is to be drawn about the accident data. The right portion of Table 4.19 compares Models A and B where $E[\mu_i|\mathbf{y}]$ is the smoothed rate under the log-linear model, $E[\lambda_i|\mathbf{y}]$ is the adjusted posterior rate used for inference, and $se(\lambda_i|\mathbf{y})$

Table : 4.19 Accident Data Analysis (Models A and B)

Data					Model A			Model B		
y_i	t_i	CT	TY	y_i/t_i	$E[\mu_i\|\mathbf{y}]$	$E[\lambda_i\|\mathbf{y}]$	se$(\lambda_i\|\mathbf{y})$	$E[\mu_i\|\mathbf{y}]$	$E[\lambda_i\|\mathbf{y}]$	se$(\lambda_i\|\mathbf{y})$
12	20.61	1	1	0.582	0.934	0.811	0.162	0.908	0.791	0.163
39	32.55	1	2	1.198	1.282	1.252	0.161	1.286	1.244	0.166
43	39.59	1	1	1.086	0.934	0.985	0.135	0.908	0.997	0.139
34	24.60	1	2	1.382	1.282	1.315	0.178	1.286	1.322	0.185
14	21.92	2	1	0.639	0.607	0.617	0.119	0.609	0.619	0.124
20	18.80	2	2	1.064	0.836	0.913	0.167	0.863	0.929	0.171
26	44.17	2	1	0.589	0.607	0.600	0.094	0.609	0.597	0.099
22	32.81	2	2	0.671	0.836	0.778	0.129	0.863	0.771	0.133

Table : C.1 Comparison of Hyperparameter Estimates

	Model A		Model B	
Parameter	Mean	(se)	Mean	(se)
β_0	0.031	(0.413)	-0.062	(0.455)
CT	-0.433	(0.191)	-0.403	(0.208)
TY	0.320	(0.187)	0.352	(0.207)
γ	0.656	(0.220)		
α			62.48	(61.12)

is its standard error. Although the assumptions of the two models are quite different the posterior means of λ_i are quite similar. Note that the posterior rates of λ_i always compromise between y_i/t_i and the posterior means of μ_i.

The adequacy of the log-linear model can be assessed by inspecting the marginal posterior densities of the hyperparameters. Table C.1 gives hyperparameter estimates for Models A and B where β_0 is an intercept term. In general, γ close to 1 and α greater than the largest exposure indicate that the log-linear model is consistent with the data. The posterior mean of γ is 0.656 giving some support to the log-linear model. In addition, the posterior estimate of α is 62.48 and no observations have smaller exposures. The classical test statistic G^2 for this model is 6.7 on 5 degrees of freedom and can be compared to tables of chi-square. Approximate relative risks can be calculated by exponentiating parameter estimates. For example, the risk of the curve configuration is approximately $e^{0.4} \approx 1.5$ times greater than the tangent. Furthermore, the risk of the Y intersection is estimated to be $e^{0.32} \approx 1.4$ times that of the T.

References

Albert, J.H., 1988a, "Computational Methods Using a Bayesian Hierarchical Generalized Linear Model," *Journal American Statistical Association*, Vol. 83, pp. 1037-1044.

Albert, J.H., 1988b, "Bayesian Estimation of Poisson Means Using a Hierarchical Log-Linear Model," *Bayesian Statistics*, Vol. 3, Bernardo, J.M., et al. eds., pp. 519-531.

Breslow, N.E., 1984, "Extra-Poisson Variation in Log-Linear Models", *Applied Statistics*, Vol. 33, pp. 38-44.

Breslow, N.E. and Day, N.E., 1987, *Statistical Methods in Cancer Research, Vol. 2: The Design and Analysis of Cohort Studies*, I.A.R.C., Lyon.

Christiansen, C.L., 1992, *Hierarchical Poisson Modeling*, unpublished Ph.D. thesis, Center for Statistical Science, University of Texas, Austin.

Dellaportas, P. and Smith, A.F.M., 1993, "Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling," *Applied Statistics*, Vol. 42, pp. 443-459.

Efron, B., 1986, "Double Exponential Families and Their Use in Generalized Linear Regression," *Journal of the American Statistical Association*, Vol. 81, pp. 709-721.

Frome, E.L., 1983, "The Analysis of Rates Using Poisson Regression Models," *Biometrics*, Vol. 39, pp. 665-674.

Gaver, D.P. and O'Muircheartaigh, I.G., 1987, "Robust Empirical Bayes Analyses of Event Rates," *Technometrics*, Vol. 29, pp. 1-15.

Gelfand, A.E. and Smith, A.F.M., 1990, "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, Vol. 85, pp. 398-409.

Holford, T.R., 1980, "The Analysis of Rates and Survivorship Using Log-Linear Models," *Biometrics*, Vol. 36, pp. 299-305.

Leonard, T. and Novick, M.R., 1986, "Bayesian Full Rank Marginalization for Two-Way Contingency Tables," *Journal of Educational Statistics*, Vol. 11, pp. 33-56.

McCullah, P. and Nelder, J.A., 1989, *Generalized Linear Models*, 2nd ed., Chapman and Hall.

Tierney, L. and Kadane, J., 1986, "Accurate Approximations for Posterior Moments and Marginals," *Journal of American Statistical Association*, Vol. 81, pp. 82-86.

Williams, D.A., 1982, "Extra-Binomial Variation in Logistic Linear Models," *Applied Statistics*, Vol. 31, pp. 144-148.

Zeger, S.L. and Karim, R., 1991, "Generalized Linear Models with Random Effects: A Gibbs Sampling Approach," *Journal of the American Statistical Association*, Vol. 86, pp. 79-86.

APPENDIX D - FIELD STUDY MATERIAL

Table D-1 - Field Site Identifiers

Site No.	District	Control Section	Trunkline Name	Crossroad Name	Beginning Milepoint	Ending Milepoint
1	2	2051	M77	ALGER AVE	12.300	12.485
2	2	21031	M35	LAKE SHORE DR	15.395	15.620
3	3	18032	US27BR	BASS LAKE ROAD	2.630	2.800
4	3	45021	M72	CARTER ROAD	20.845	21.210
5	3	5031	M88	CAYUGA ST	11.235	11.300
6	3	43021	US10	M37	9.537	9.573
7	3	53031	OLD US31	MEYERS RD	6.220	6.525
8	3	43011	M37	MICHIGAN AVE	1.090	1.305
9	3	51031	M22	MILLER ROAD	2.585	3.075
10	3	15071	M75	NORTH SHORE DR	12.025	12.200
11	4	69022	M32	MCCOY RD	1.980	3.005
12	5	59044	M46,M66	ALMY RD	1.880	2.430
13	5	34032	M66	BELLEVIEW DR	6.005	6.225
14	5	62032	M37	JACKSON ST	3.755	3.870
15	5	70081	M104	KRUEGER AVE	2.750	2.980
16	5	34021	M50	NASH HWY	2.720	3.940
17	5	19061	M21	PEWAMO RD	0.000	0.495
18	5	62012	M20	SIX MILE RD	5.230	6.205
19	5	62012	M20	SIX MILE RD	6.205	7.555
20	5	61131	M37	WHITE RD	1.810	2.150
21	6	26012	M18	BARD RD	11.930	12.430
22	6	79081	M25	BAY PARK RD	10.335	10.940
23	6	73021	M57	CORUNNA RD	13.735	13.910
24	6	73051	M13	EAST RD	6.565	6.975
25	6	74071	M25	ST CLAIR RD	2.820	2.910
26	7	80072	M40	32ND ST	12.845	13.570
27	7	13032	M66	CAPITAL AVE	0.425	0.480
28	7	13061	I94BL	COLUMBIA AVE	8.645	8.770
29	7	78012	US131	DEPOT ST	2.430	2.535
30	7	39082	M43,M89	GULL LAKE DR	11.770	11.970
31	7	8011	M43	GUN LAKE RD	18.400	18.550
32	7	3023	M89	JEFFERSON ST	8.915	9.035
33	7	11074	M140	MAPLE GROVE RD	8.590	9.140
34	7	14042	US12	MASON ST	4.650	4.910
35	7	78042	M60	MICHIGAN AVE	1.565	1.645
36	7	14061	M60	PINE LAKE ST	3.795	4.180
37	7	11074	M140	POKAGON RD	5.580	6.095
38	7	3072	M40,M89	SHERMAN ST	0.580	0.655
39	7	14032	M62	TWIN LAKES RD	4.030	4.195
40	7	8011	M43	YECKLEY RD	17.035	17.205
41	8	46041	M34	BENNER HWY	11.807	12.312
42	8	38071	M50	BROOKLYN RD	15.525	16.240
43	8	58051	US24	CRABB RD	0.485	0.915
44	8	33091	M52,M106	GREEN RD	0.000	1.160
45	8	38061	M60	HOMER RD	2.015	2.554
46	8	30062	US12	LITCHFIELD RD	7.025	7.470
47	8	38082	I94BL	MAIN ST	1.720	1.995
48	8	47041	M36	SPICER RD	21.315	21.480
49	8	38051	M106	TERRITORIAL RD	16.050	16.210
50	9	77033	M25	24TH ST	0.000	0.135
51	9	77041	M136	AVOCA RD	1.985	2.530
52	9	77012	M19	WILKES RD	7.240	7.725
53	9	77012	M19	WILKES RD	7.725	8.215

Table D-2 - Instructions for Site Visits

April, 1994

SITE VISITS - SPECIAL Y-INTERSECTIONS

OBJECTIVE:

To examine a set of specially configured Y-type intersections for physical characteristics that may contribute to safety (or lack thereof).

The characteristics that we are looking at:

The geometry, pavement type, signing, pavement markings, site distance, optical illusions (i.e., tree-lines, utility poles, fence lines, or anything that may suggest that the main path is along the minor road).

We also wish to know how drivers behave when they proceed onto the minor road. Do they signal (which way)? Do they appear to check for oncoming traffic? Do they cut across the opposing lane diagonally on the left turns and cut over the shoulder on the right turns?

How do drivers coming out of the minor road behave? Do they signal? Do they look? How difficult is it for them to see to the right, to the left?

INSTRUCTIONS

You will visit a set of intersections where:

1. The main road turns to the right and the minor road continues straight (or very close to it) ahead.

2. The main road turns to the left and the minor road continues straight (or very close to it) ahead.

You are to:

1. Fill out a checklist about the site.

2. Observe the traffic and record your observations about the behavior of the drivers making turns.

3. Photograph the intersection and the approaches to the Y on the main route and on the minor route.

4. Note anything unique or unusual about the site.

PHOTOGRAPHS

Use a wide angle lens - 35 mm is good for this. Show the road, not the sky. For point and shoot cameras, use fast film—400 ASA, flash mode will do nothing for these pictures.

1. First photo - identify the intersection, show road names.

2. Approach to Y-intersection from major road from right shoulder and from cross-road sign, if there is one; if no cross-road sign, about 100 feet before the intersection.

3. Closer to intersection, show the intersection - want to see tree lines, utility poles, pavement markings, stop sign on minor road.

   If the road curves to the right, take this photo from the left shoulder.

   If the road curves to the left, take this photo from the right shoulder.

4. Approach to the Y-intersection from the minor road. Should show what the driver sees approaching the intersection. Take from about 20 feet back of the stop line.

5. Anything else of interest about this intersection. For example, are vehicles making their own path by cutting across the shoulder?

IMPORTANT

You have to keep track of the sites and photos. After a while they all look the same. The identifier photo is very critical. Keep notes on order of sites and number of photos taken at each site. Tag each roll of film with tape giving date and sites.

Table D-3 - Data Form for Site Visits

SITE CHECKLIST

Initials

Site - Trunkline Road
Cross Road
Curve Right or Left

Date -
Time - from to

Trunkline Road

Number of lanes

Pavement type

Pavement markings near intersection

Condition of pavement marks

Double yellow centerline

White edgelines

Other - specify

Signs on Trunkline Approach to Intersection
Curve ahead warning
Turn ahead warning
Speed advisory
No passing
Crossroad name sign
Route guide signs

Other (specify or sketch)

Minor Road

Number of lanes

Pavement type

Pavement markings
centerline
edgelines
other - specify

Signs on Minor Road
Stop
Stop ahead
Intersection ahead warning
Route signs
Other (specify or sketch)

INTERSECTION

Road names

Make a simple sketch of intersection - show the extent of the curve on the main road and the angle at which the minor road intersects with the major road. Show any bump-outs, posts intended to prevent cutting through the shoulder for right turns, additional lanes, etc.

Any signs of recent construction or recent realignment of the intersection? If yes, specify.

<u>Optical Illusions</u>
Does the minor road appear to be continuation of the main route?

Why?	Why not?
It is straight on	
Tree line	
Utility poles	
Fence lines	
Other (specify)	

Comments

DRIVER BEHAVIOR

Road names

<u>For Curve Right Sites</u>

Vehicles proceeding onto the minor road (left turns)

Number of turns observed

Do drivers signal left?

Do drivers slow down for turn?

Do they appear to check for oncoming traffic (if you can tell)?

Where do they start crossing the opposing lane of traffic?

Start far back how far back?

Start right at the intersection

If there is oncoming traffic where do the turning vehicles wait?

<u>For Vehicles coming from minor road onto the major road</u>

How many observed?

Do they signal?

Do they stop?

Do they appear to check for on-coming traffic?

DRIVER BEHAVIOR

<u>Road Names</u>

<u>For Curve Left Sites</u>

Vehicles proceeding onto the minor road (right turns)

Number of turns observed

Do drivers signal right?

Do drivers slow down for the (right) turn?

Where do they start making the turn?
It's important here to note if there is a bump-out - even a small one.
If no bump-out
    Proceed in a straight path from main road to the minor road
    Other, explain

If bump-out
    Turn at the bump-out
    Ignore bump-out and go straight through

<u>For Vehicles coming from minor road onto the major road</u>

How many observed?

Do they signal?

Do they stop?

Do they appear to check for oncoming traffic?

Your opinion: How difficult is it to see to the right and to the left for these drivers?

Comments

APPENDIX E - PHOTOGRAPHIC RECORD

Site No.	District	Control Section	Trunkline Name	Crossroad Name	Beginning Milepoint	Ending Milepoint
1	2	2051	M77	ALGER AVE	12.300	12.485
2	2	21031	M35	LAKE SHORE DR	15.395	15.620
3	3	18032	US27BR	BASS LAKE ROAD	2.630	2.800
4	3	45021	M72	CARTER ROAD	20.845	21.210
5	3	5031	M88	CAYUGA ST	11.235	11.300
6	3	43021	US10	M37	9.537	9.573
7	3	53031	OLD US31	MEYERS RD	6.220	6.525
8	3	43011	M37	MICHIGAN AVE	1.090	1.305
9	3	51031	M22	MILLER ROAD	2.585	3.075
10	3	15071	M75	NORTH SHORE DR	12.025	12.200
11	4	69022	M32	MCCOY RD	1.980	3.005
12	5	59044	M46,M66	ALMY RD	1.880	2.430
13	5	34032	M66	BELLEVIEW DR	6.005	6.225
14	5	62032	M37	JACKSON ST	3.755	3.870
15	5	70081	M104	KRUEGER AVE	2.750	2.980
16	5	34021	M50	NASH HWY	2.720	3.940
17	5	19061	M21	PEWAMO RD	0.000	0.495
18	5	62012	M20	SIX MILE RD	5.230	6.205
19	5	62012	M20	SIX MILE RD	6.205	7.555
20	5	61131	M37	WHITE RD	1.810	2.150
21	6	26012	M18	BARD RD	11.930	12.430
22	6	79081	M25	BAY PARK RD	10.335	10.940
23	6	73021	M57	CORUNNA RD	13.735	13.910
24	6	73051	M13	EAST RD	6.565	6.975
25	6	74071	M25	ST CLAIR RD	2.820	2.910
26	7	80072	M40	32ND ST	12.845	13.570
27	7	13032	M66	CAPITAL AVE	0.425	0.480
28	7	13061	I94BL	COLUMBIA AVE	8.645	8.770
29	7	78012	US131	DEPOT ST	2.430	2.535
30	7	39082	M43,M89	GULL LAKE DR	11.770	11.970
31	7	8011	M43	GUN LAKE RD	18.400	18.550
32	7	3023	M89	JEFFERSON ST	8.915	9.035
33	7	11074	M140	MAPLE GROVE RD	8.590	9.140
34	7	14042	US12	MASON ST	4.650	4.910
35	7	78042	M60	MICHIGAN AVE	1.565	1.645
36	7	14061	M60	PINE LAKE ST	3.795	4.180
37	7	11074	M140	POKAGON RD	5.580	6.095
38	7	3072	M40,M89	SHERMAN ST	0.580	0.655
39	7	14032	M62	TWIN LAKES RD	4.030	4.195
40	7	8011	M43	YECKLEY RD	17.035	17.205
41	8	46041	M34	BENNER HWY	11.807	12.312
42	8	38071	M50	BROOKLYN RD	15.525	16.240
43	8	58051	US24	CRABB RD	0.485	0.915
44	8	33091	M52,M106	GREEN RD	0.000	1.160
45	8	38061	M60	HOMER RD	2.015	2.554
46	8	30062	US12	LITCHFIELD RD	7.025	7.470
47	8	38082	I94BL	MAIN ST	1.720	1.995
48	8	47041	M36	SPICER RD	21.315	21.480
49	8	38051	M106	TERRITORIAL RD	16.050	16.210
50	9	77033	M25	24TH ST	0.000	0.135
51	9	77041	M136	AVOCA RD	1.985	2.530
52	9	77012	M19	WILKES RD	7.240	7.725
53	9	77012	M19	WILKES RD	7.725	8.215



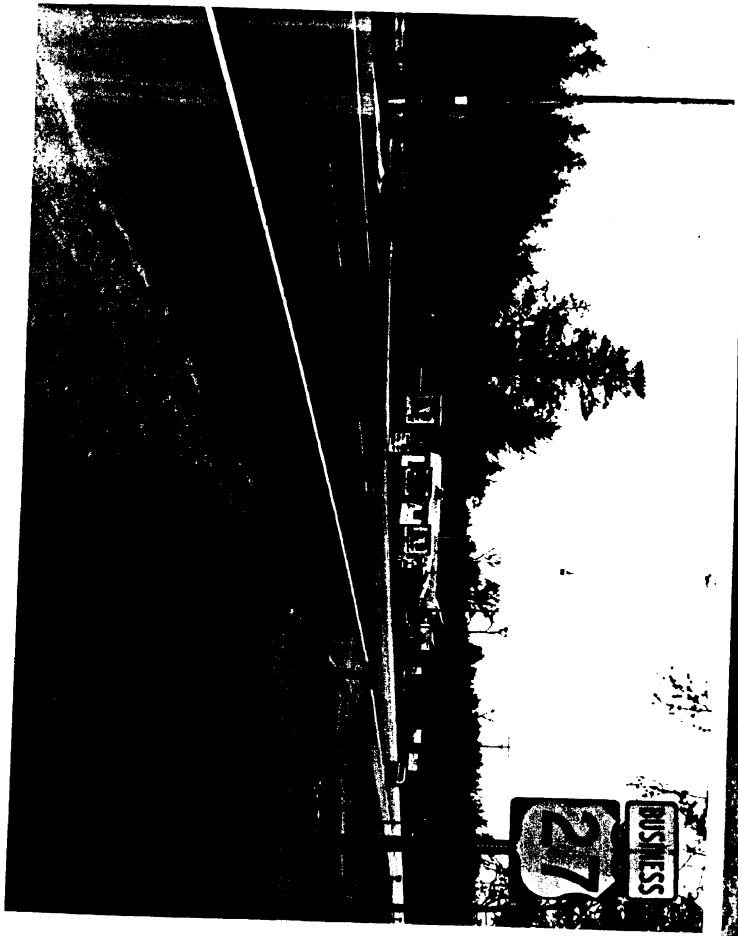
Site 1

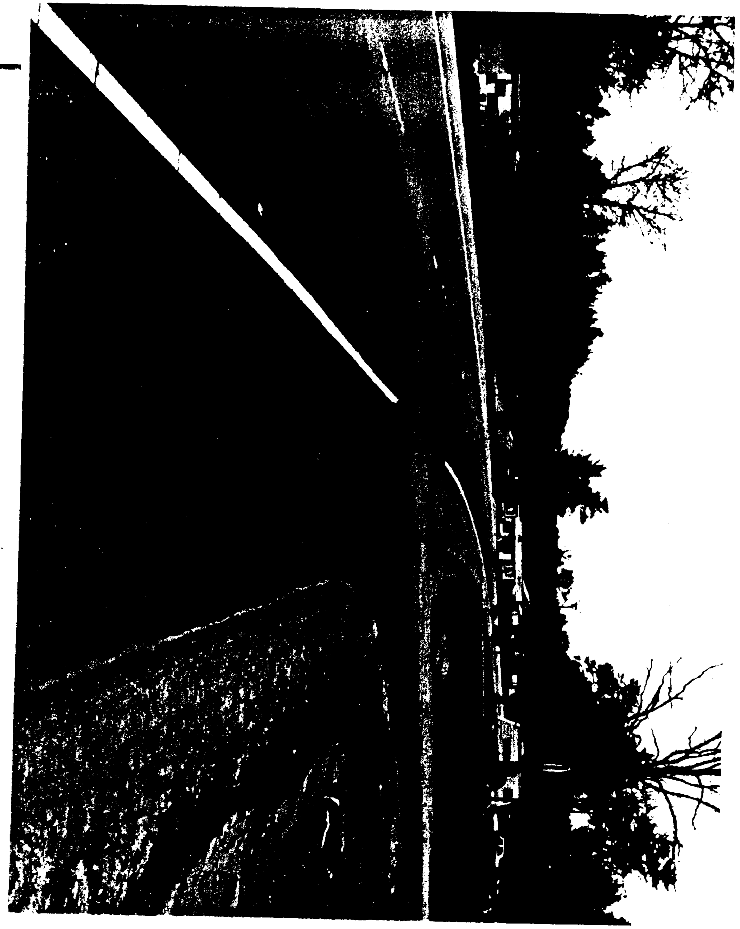

STOP

Site 2


SHORE DR


STOP



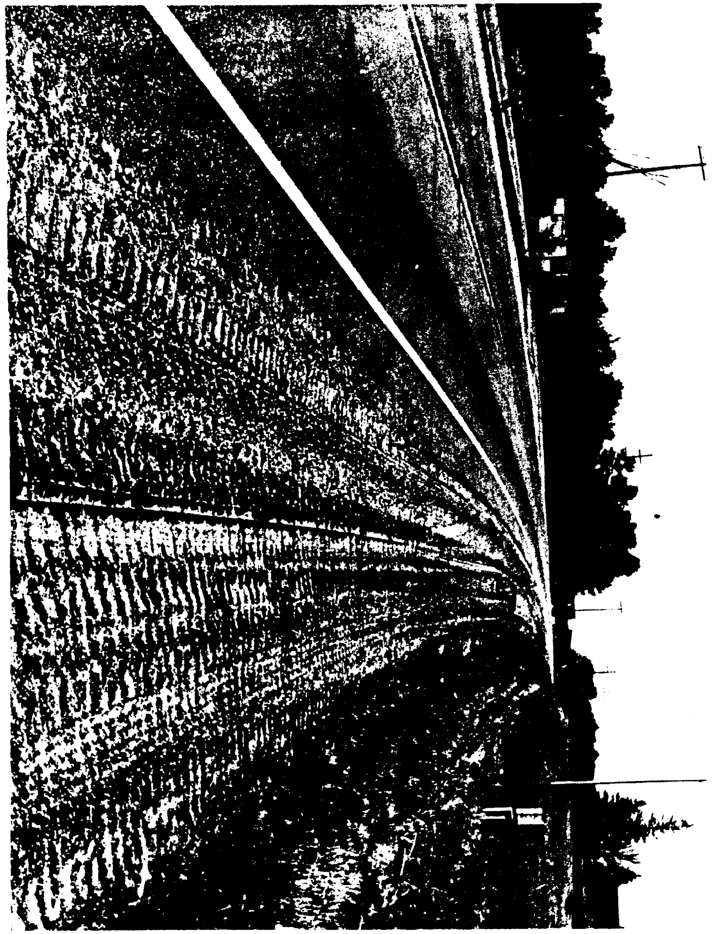



Site 3

Site 5

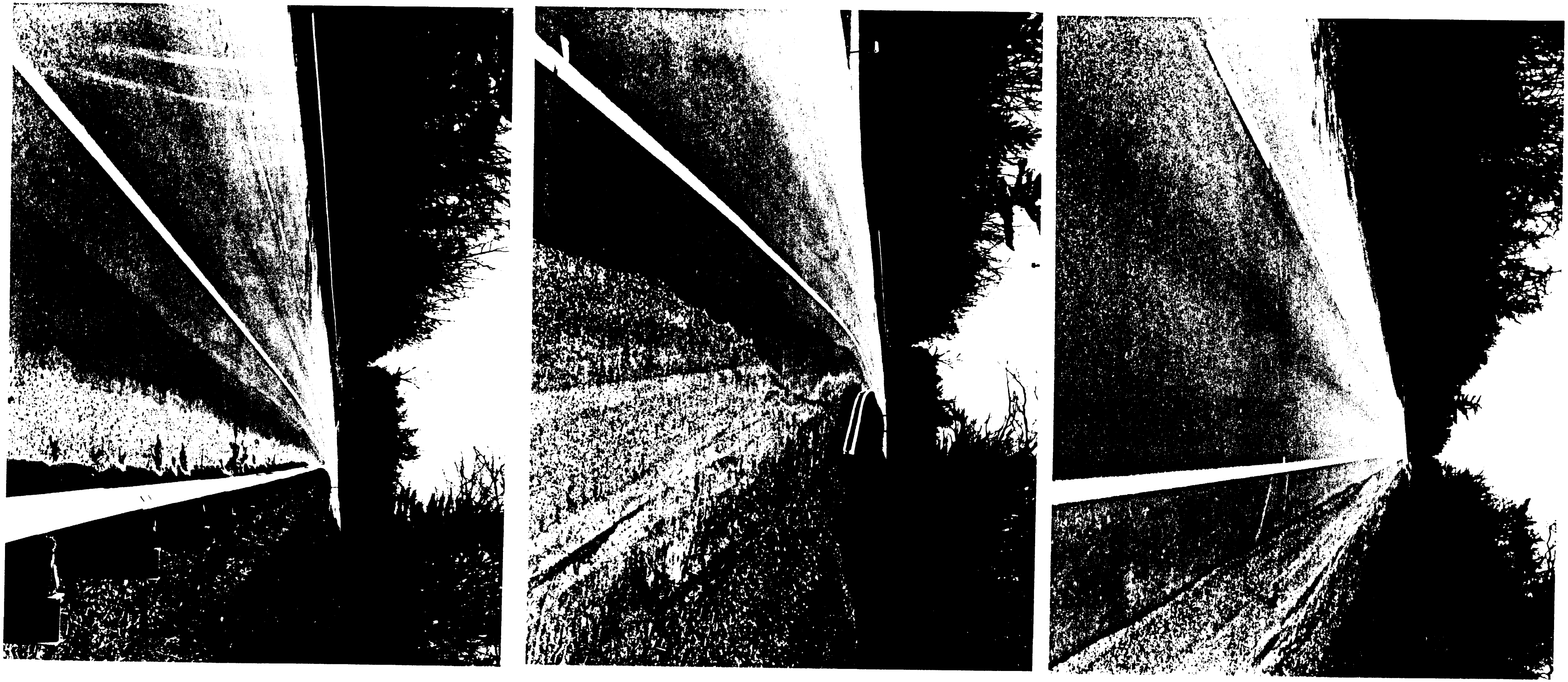

STOP

Site 6



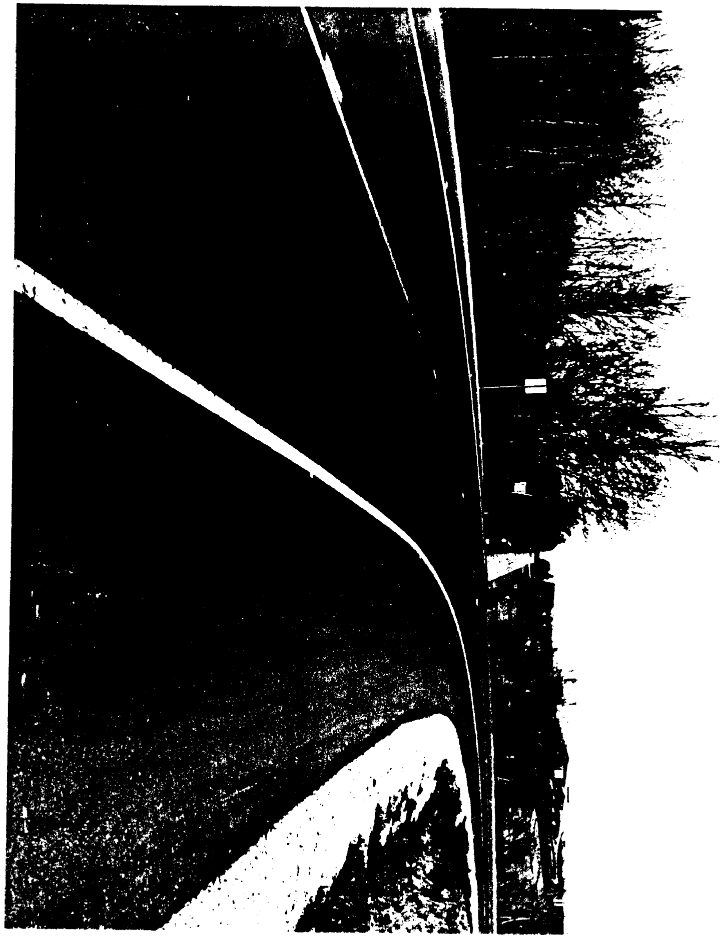
Site 7

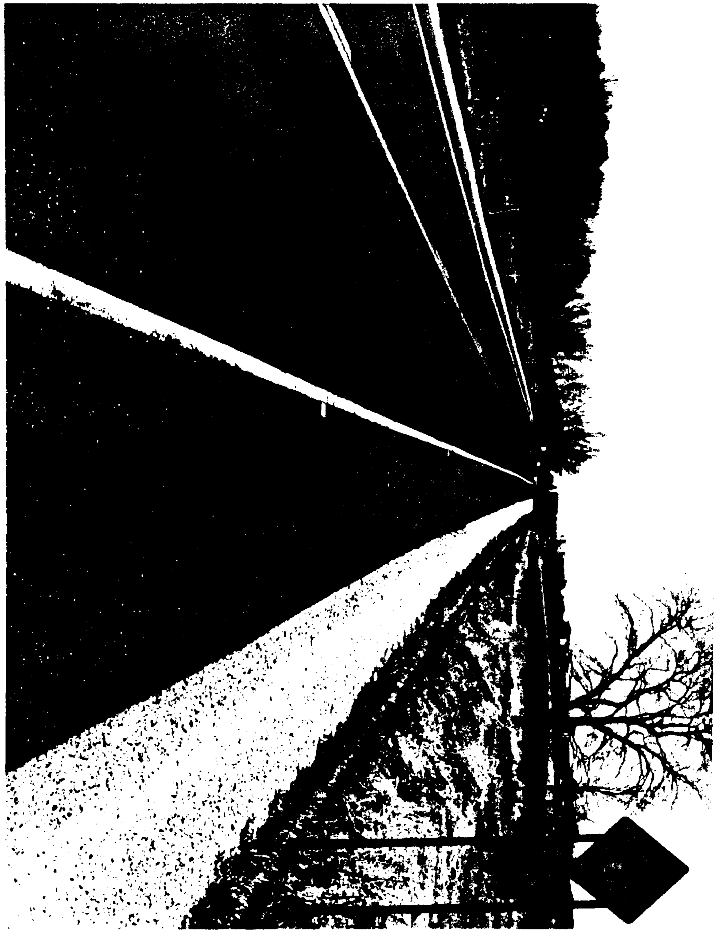

STOP



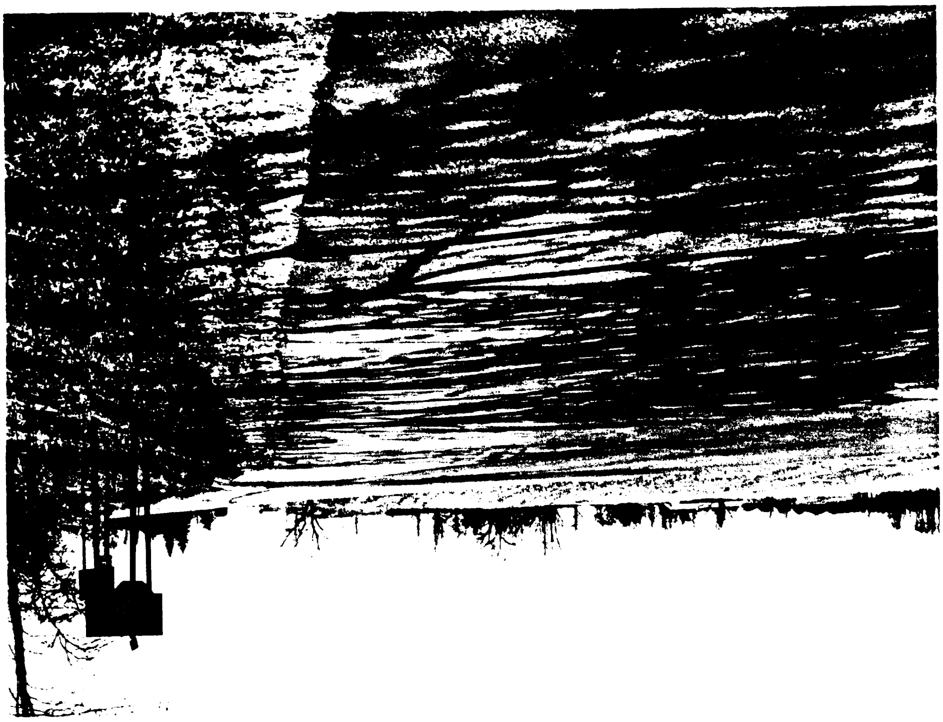
Site 8

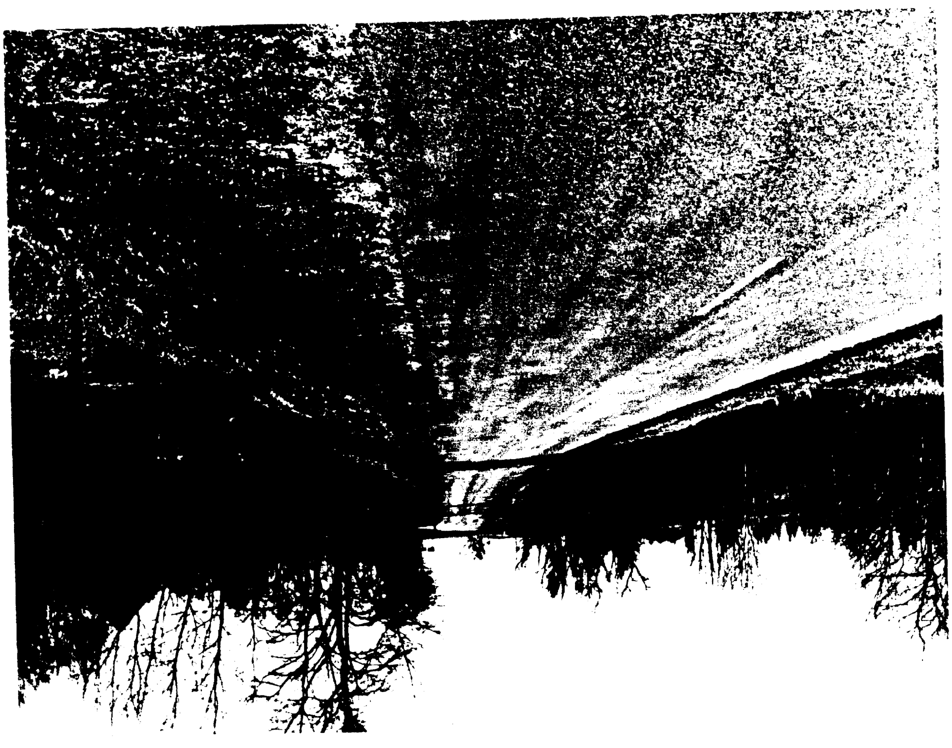

STOP

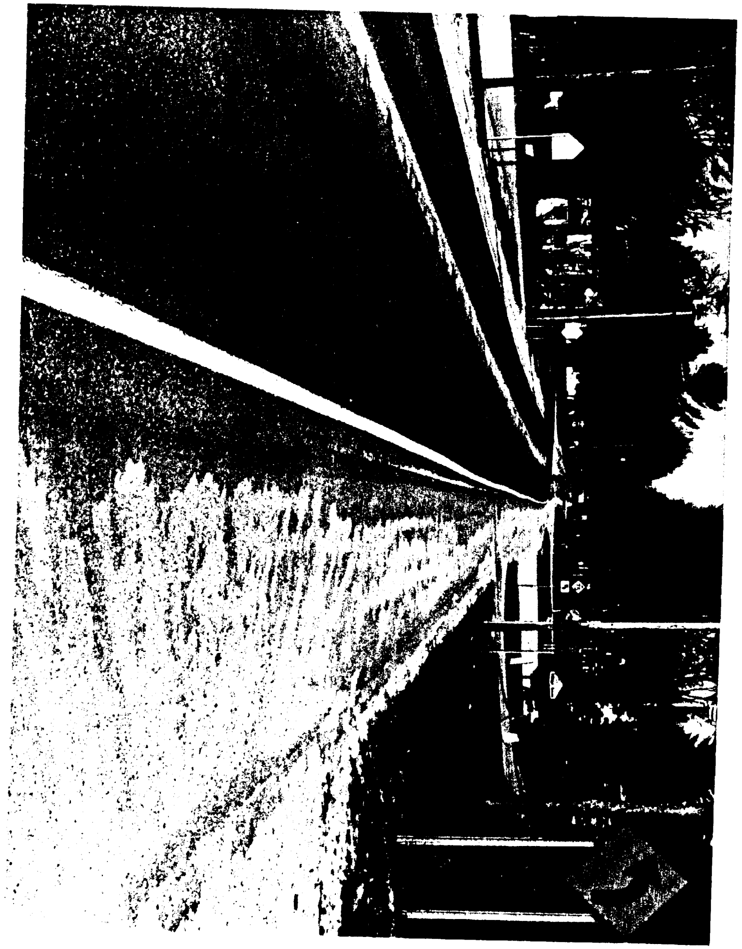




Site 9

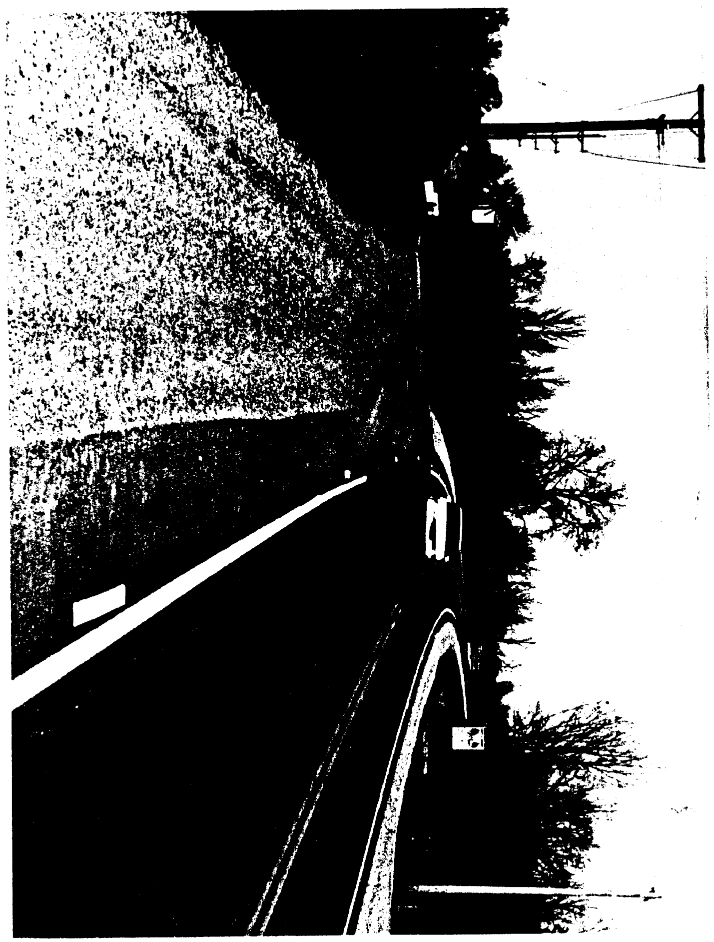
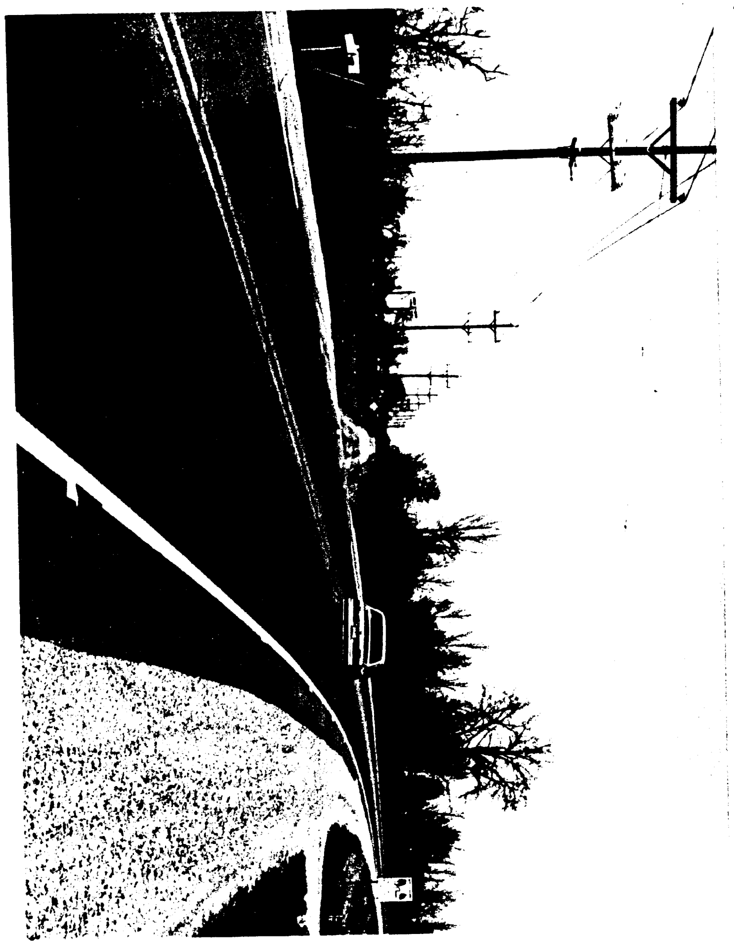
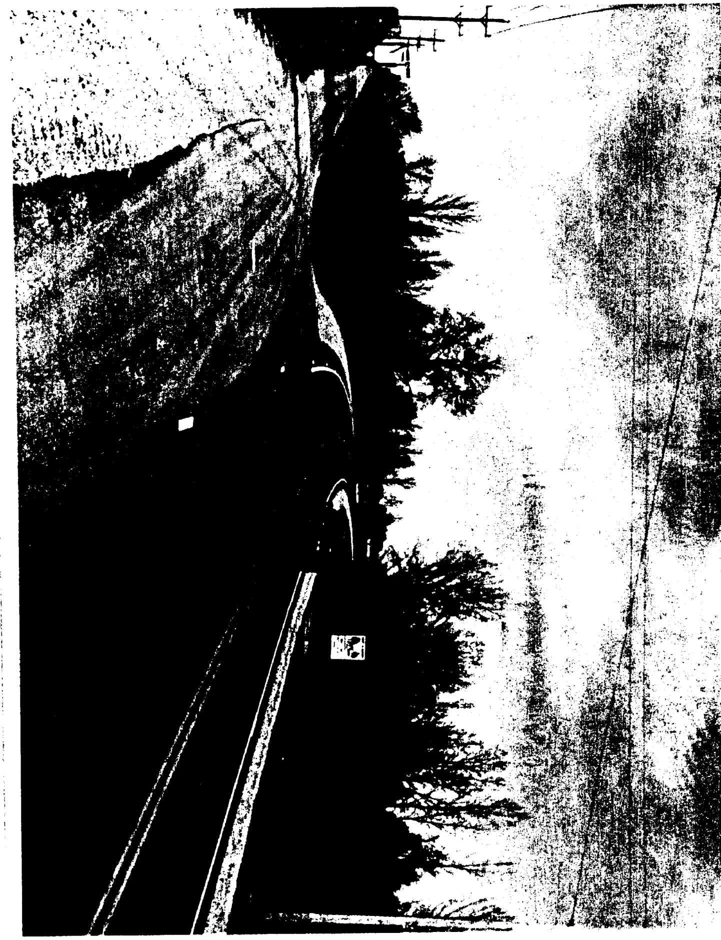




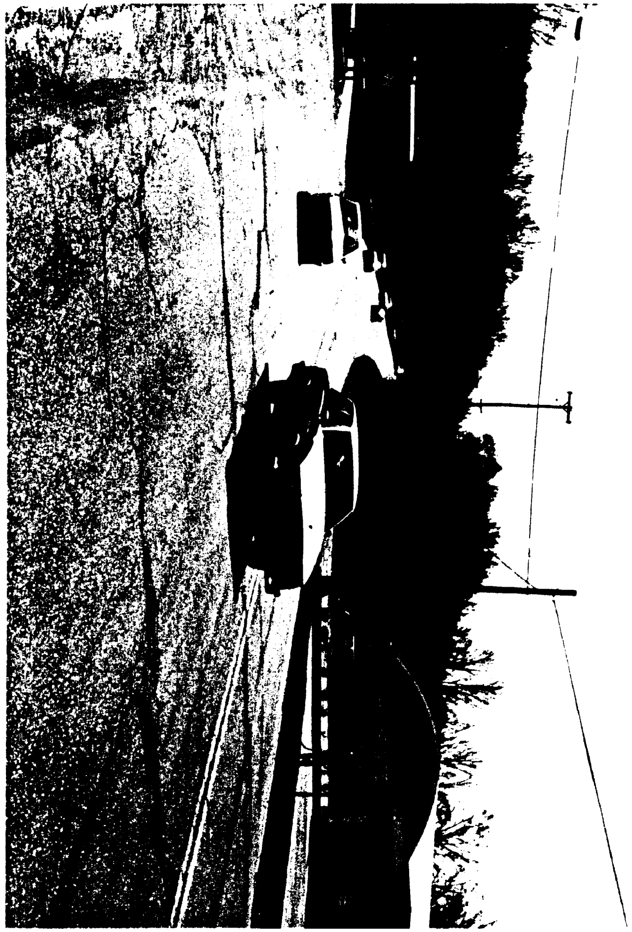


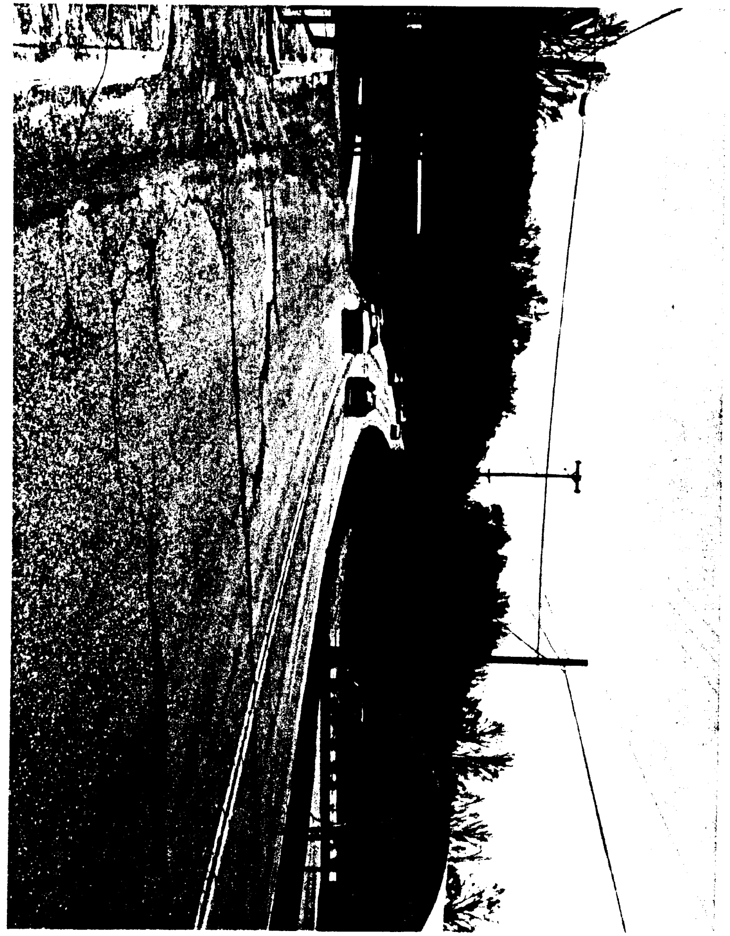
Site 10

Site 11

Site 12

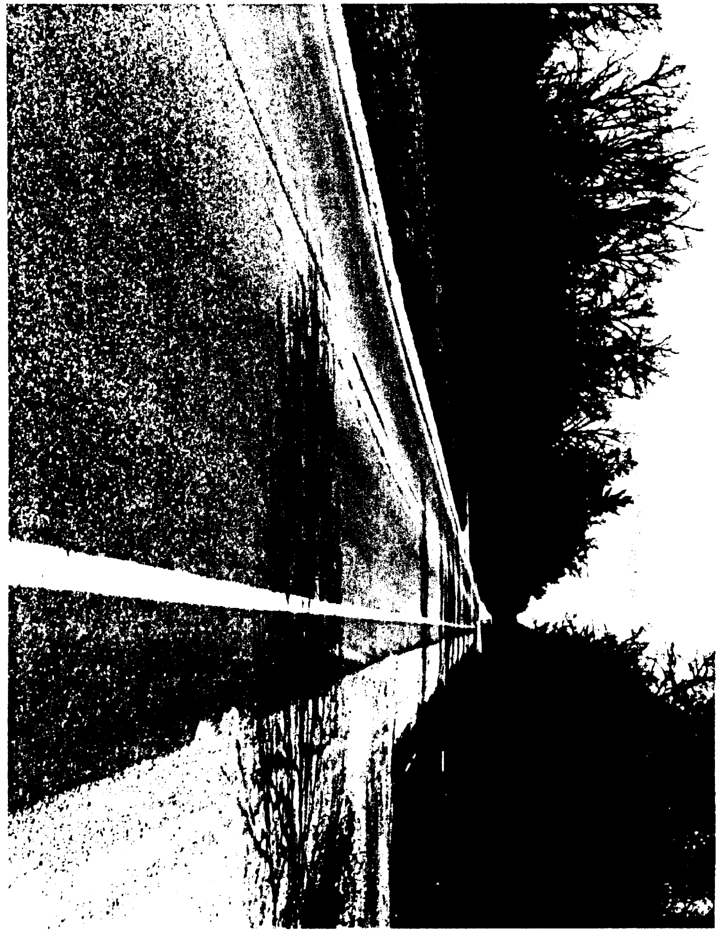




Site 13


STOP


STOP

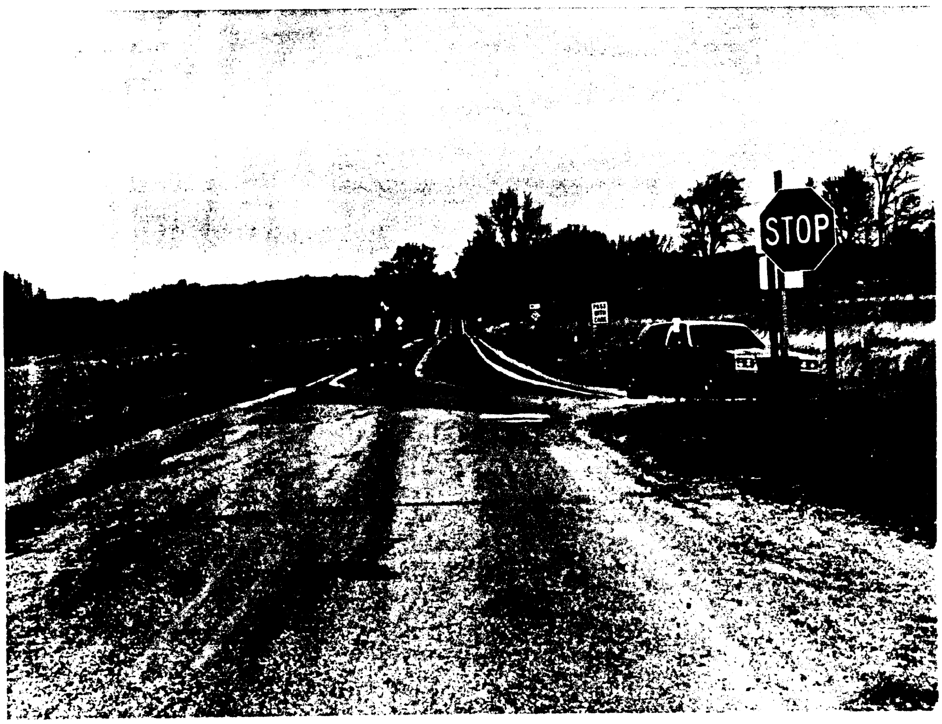


Site 14

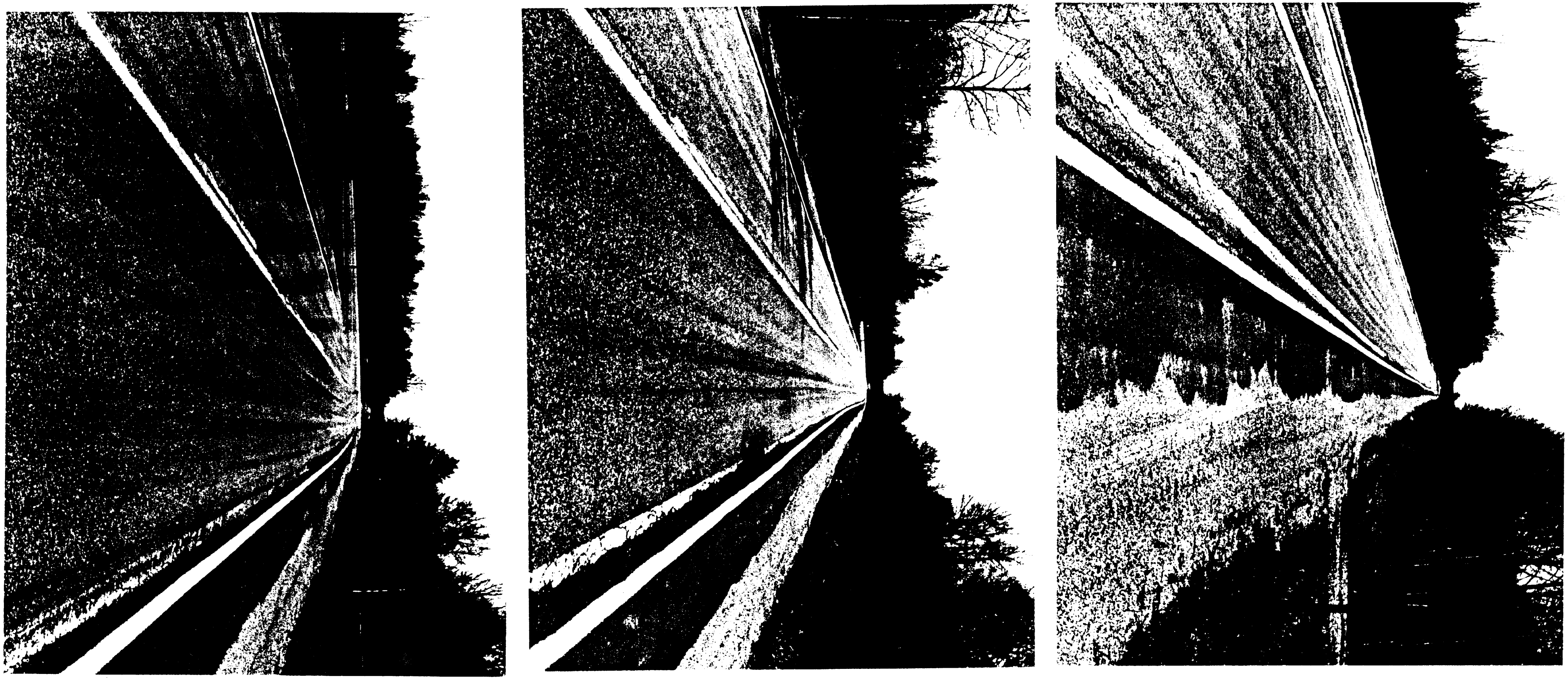
Site 15

Site 16

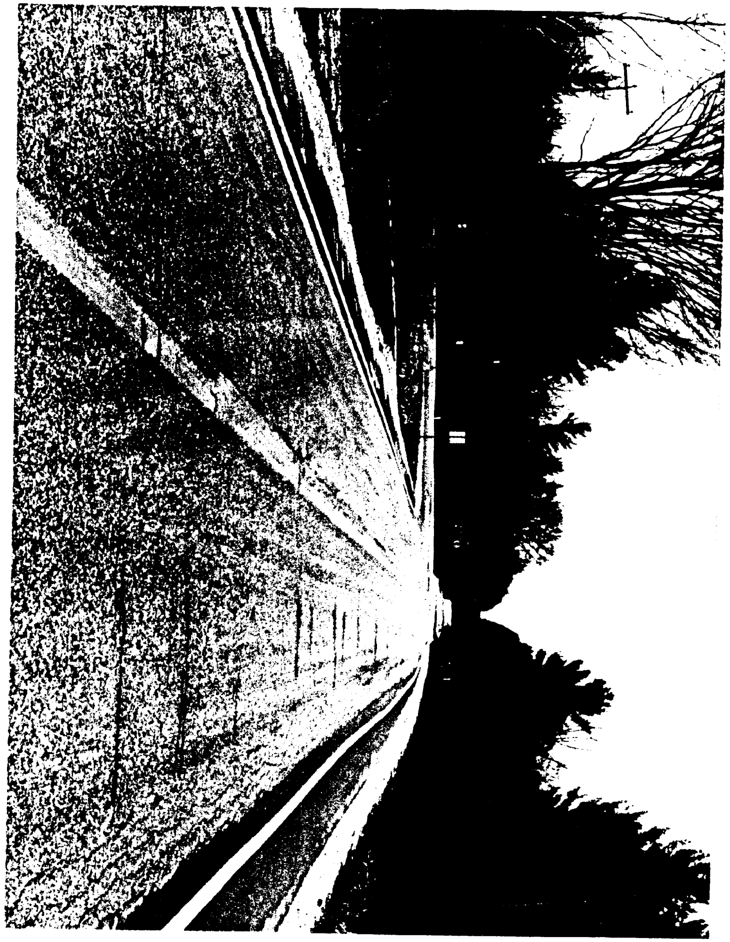

STOP


STOP

Site 17

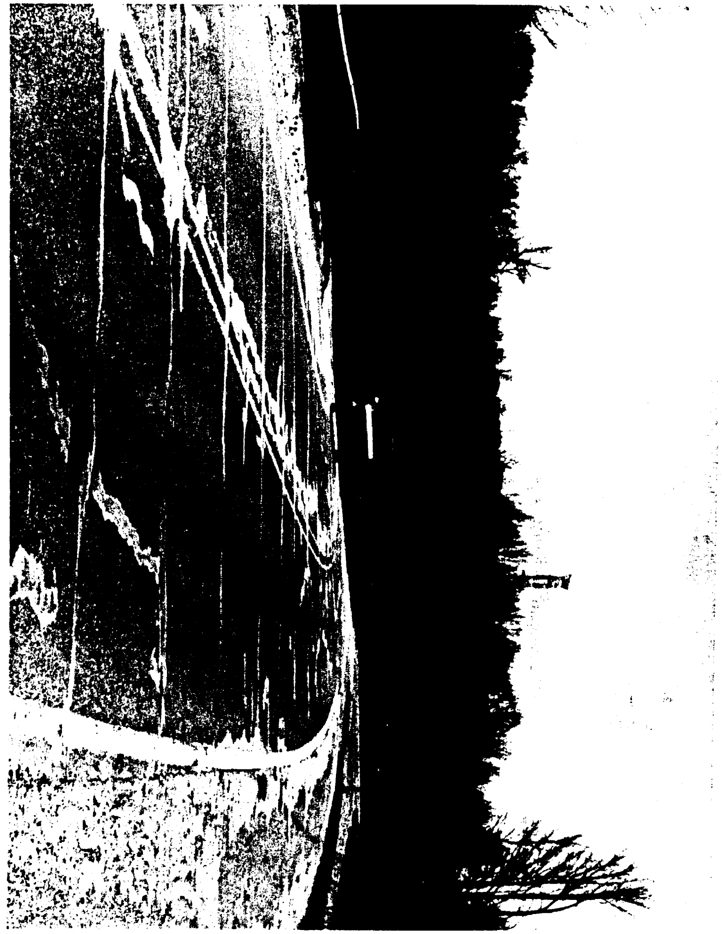


Site 18

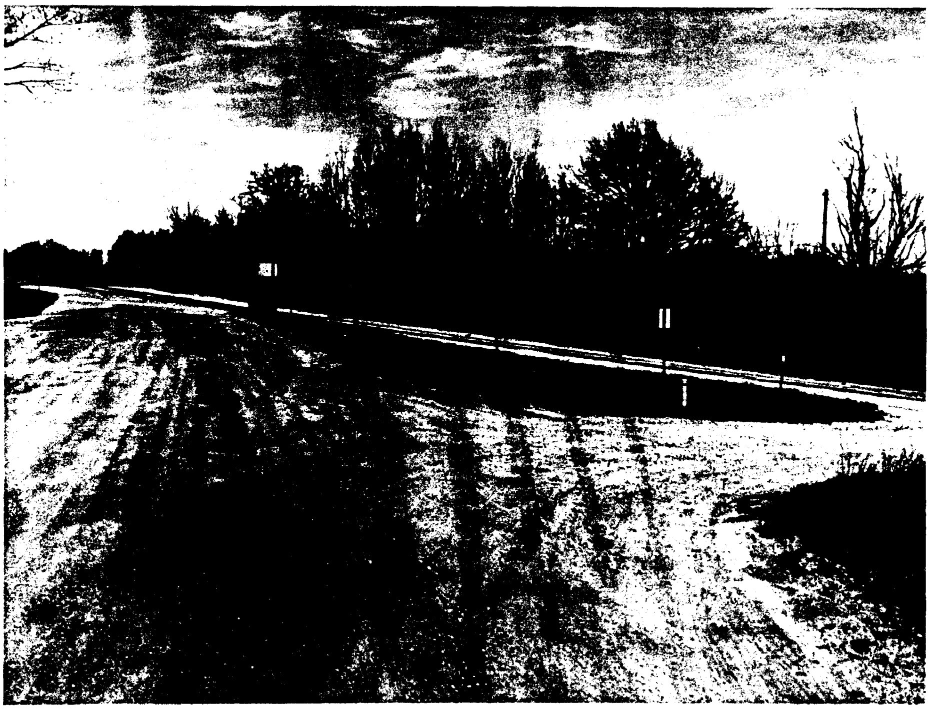


Site 19

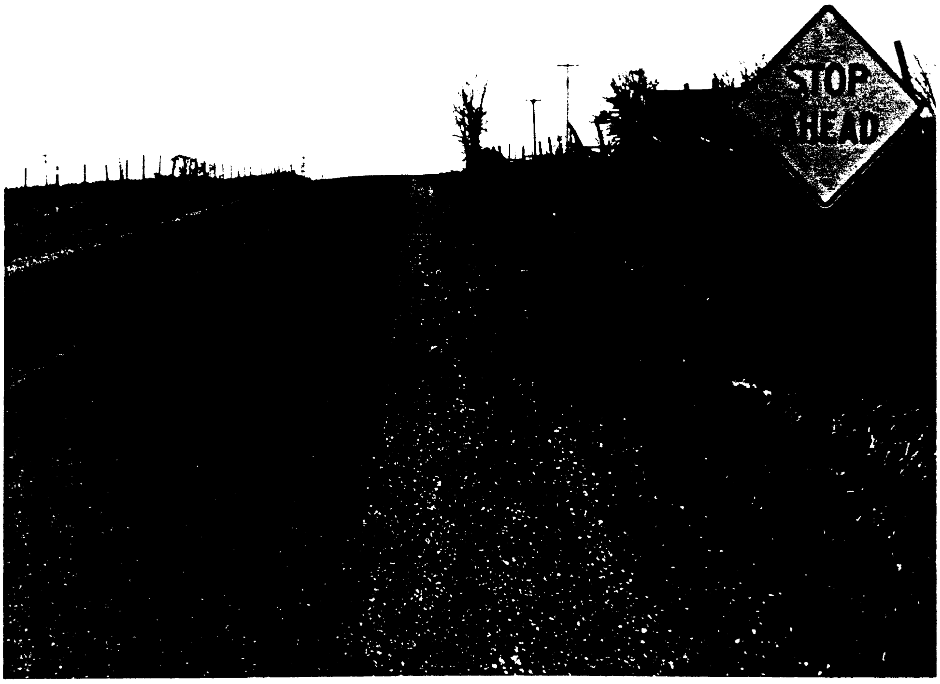


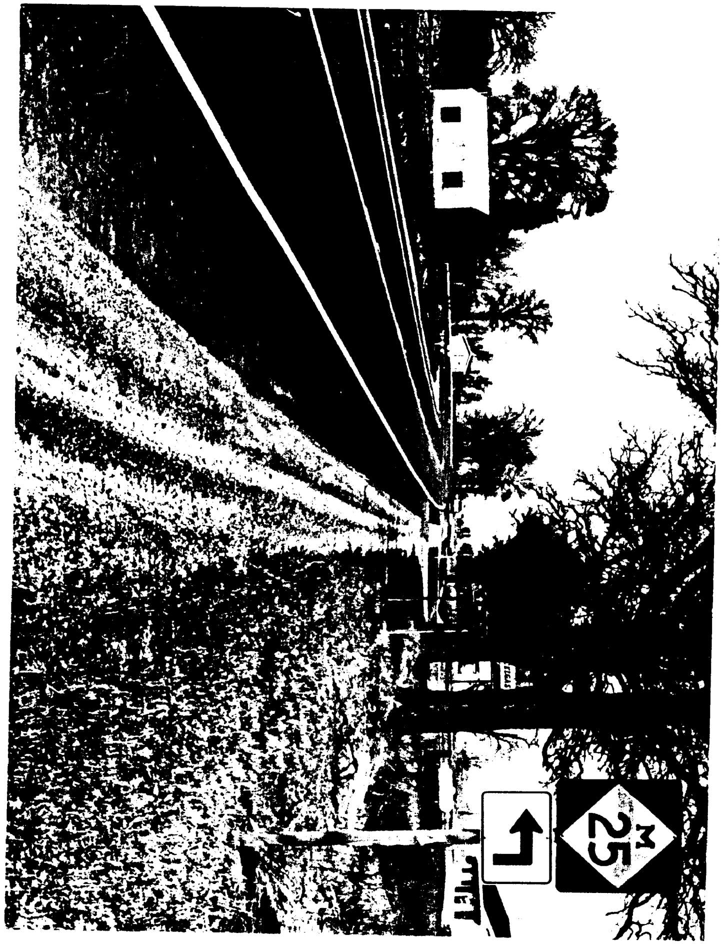


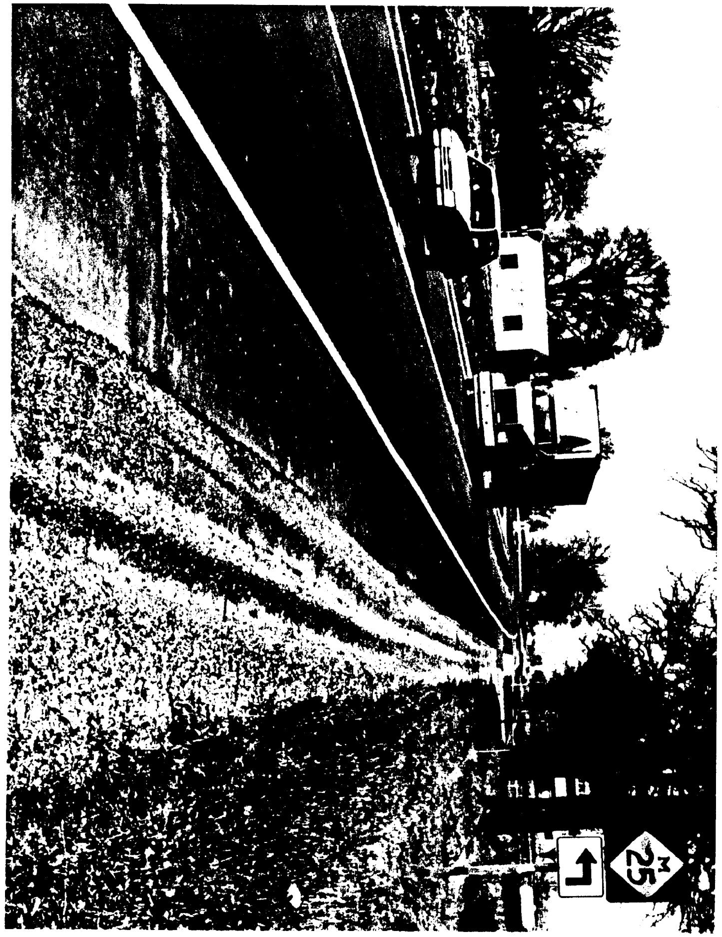
Site 20



Site 21


STOP
AHEAD

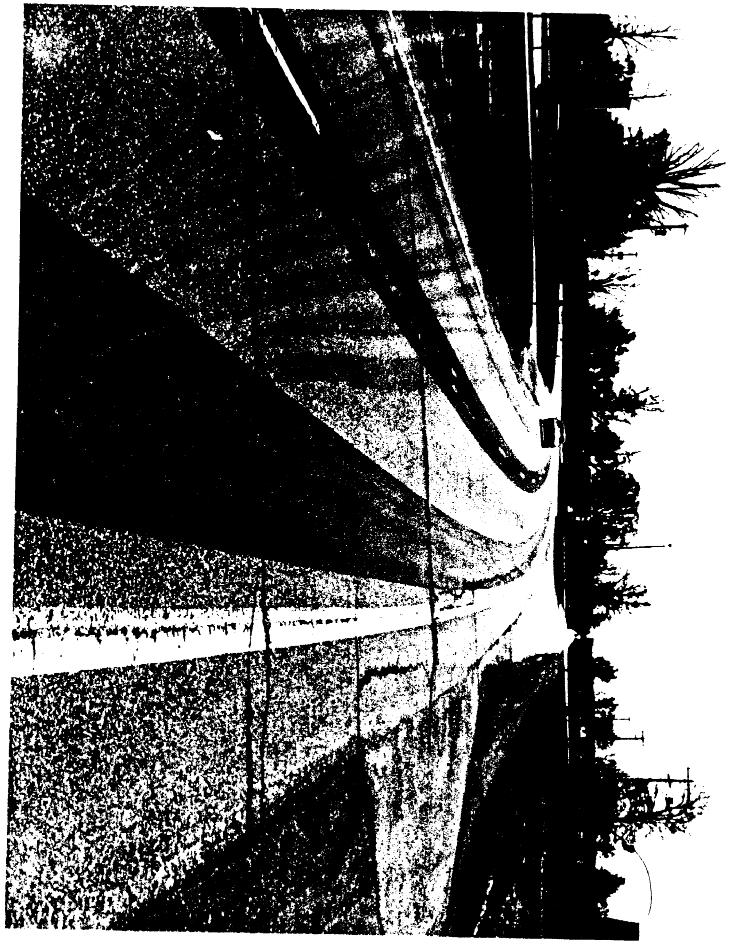


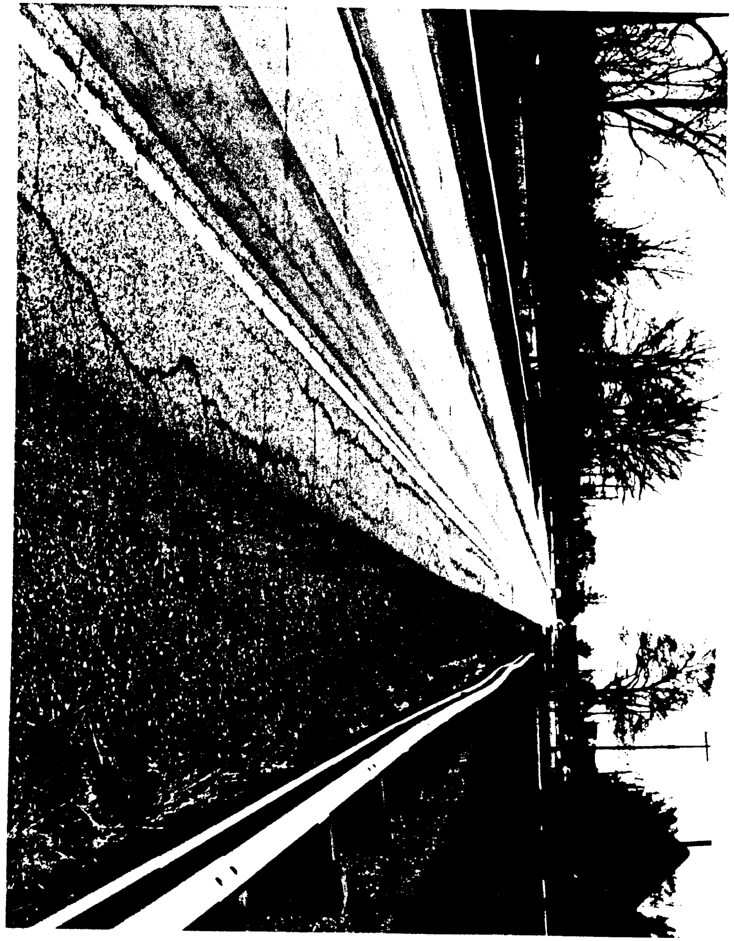


Site 22

Site 23


STOP



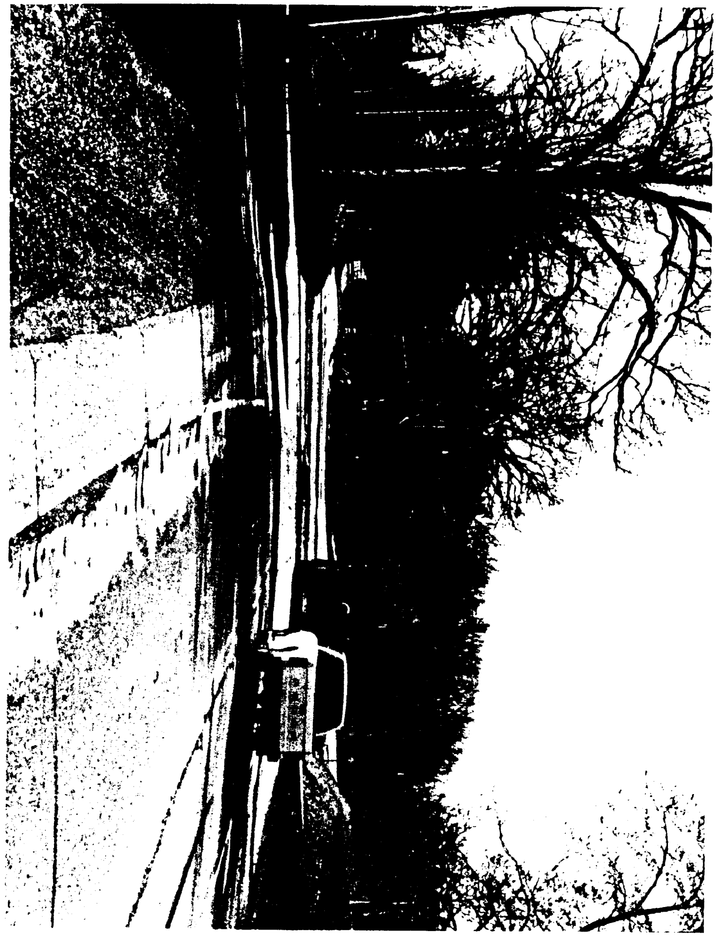


Site 24

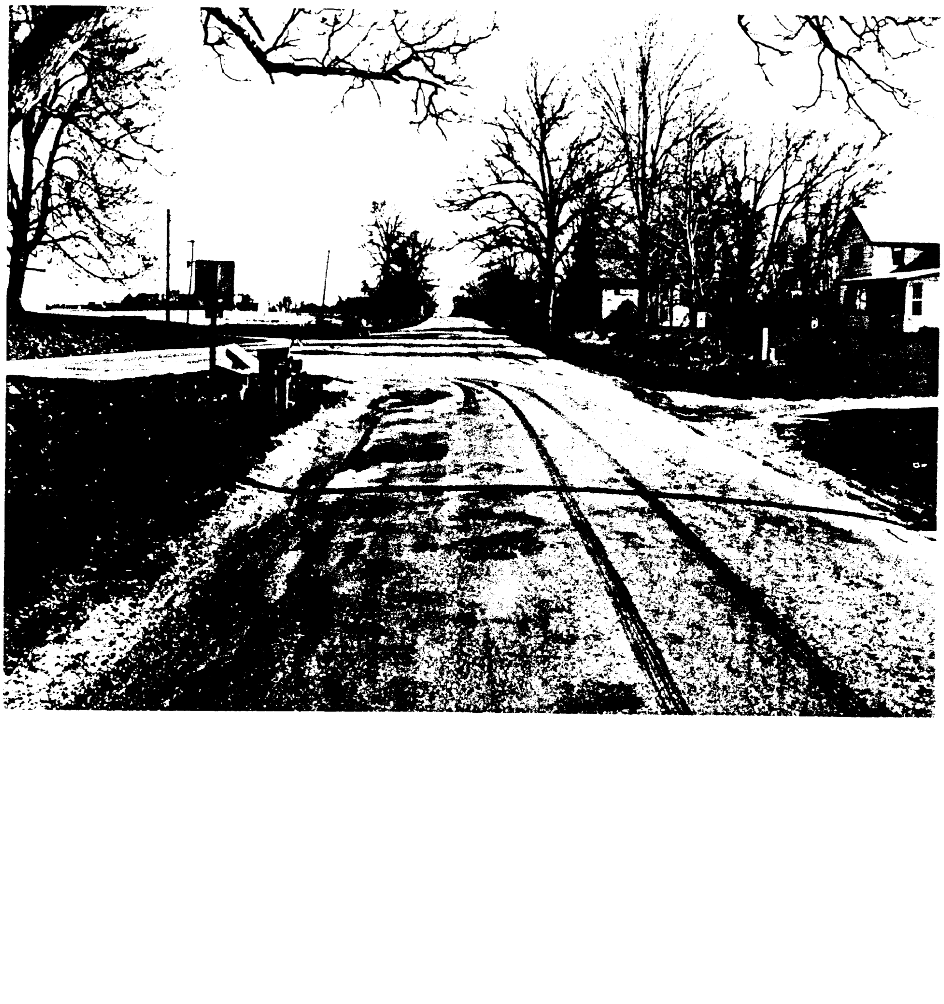

MILLER


STOP

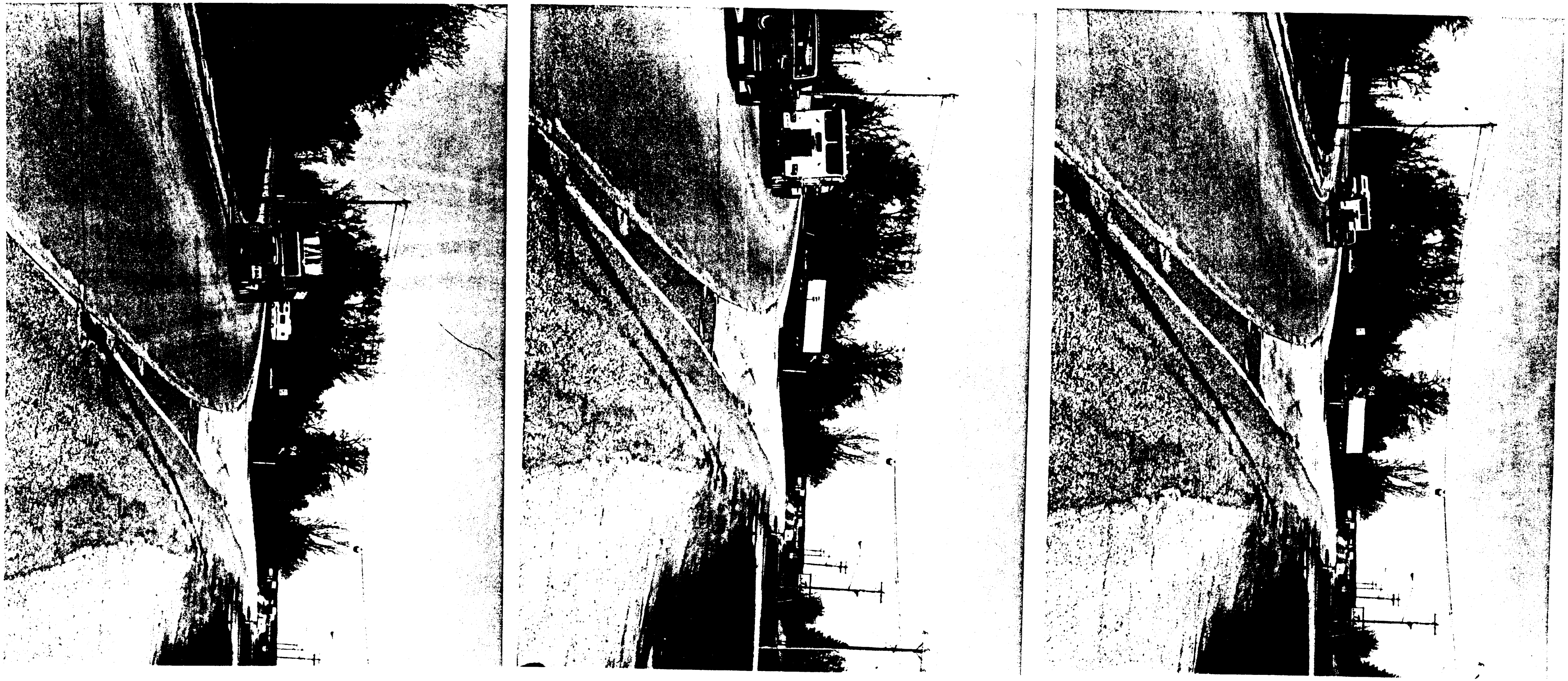
Site 25



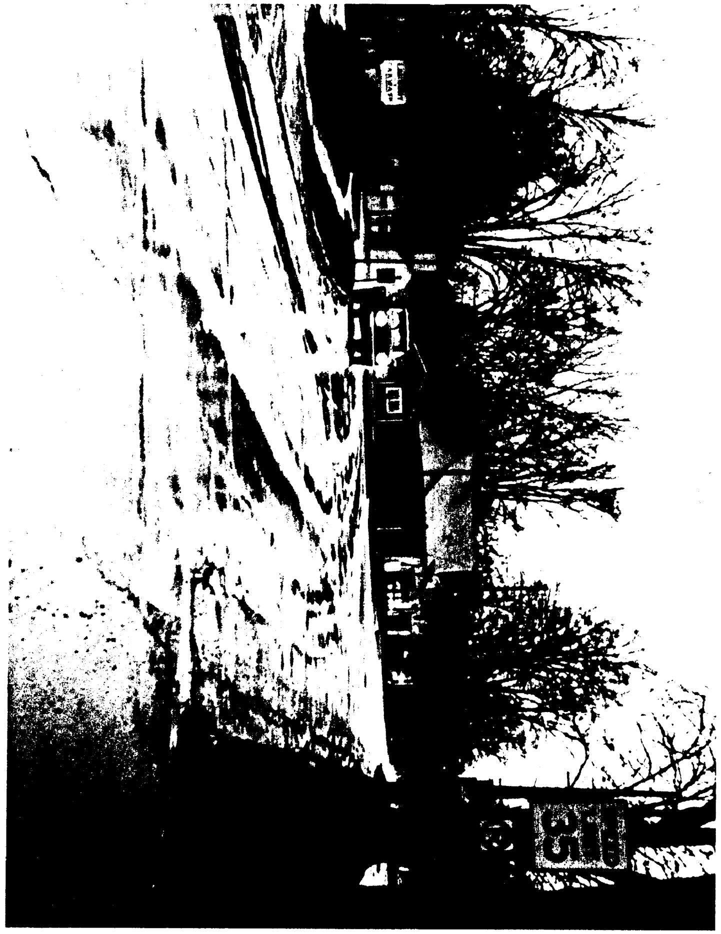
Site 26



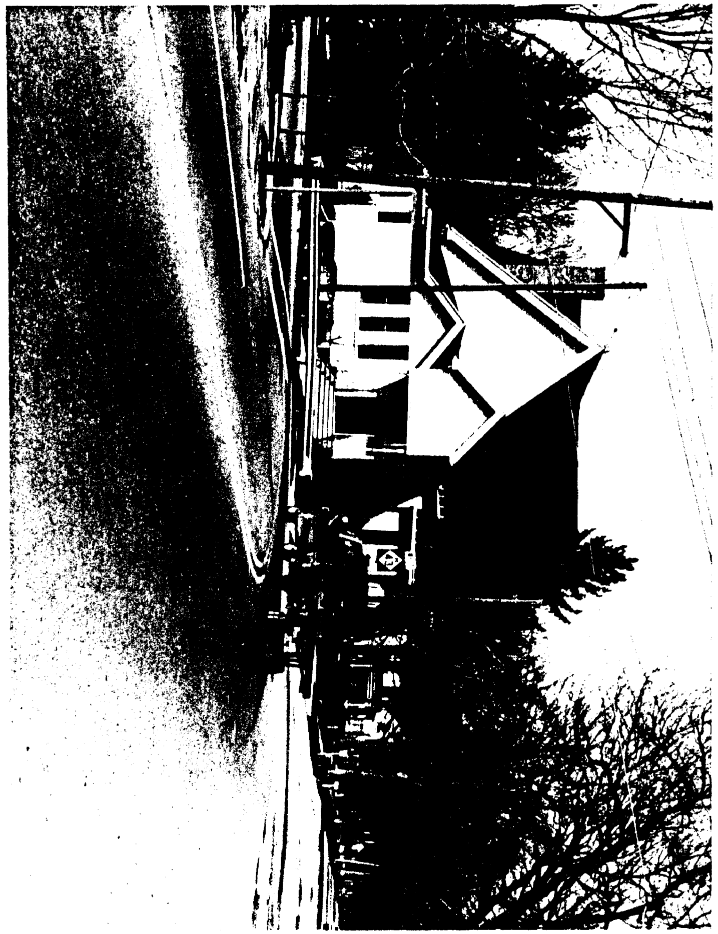
Site 27



Site 28



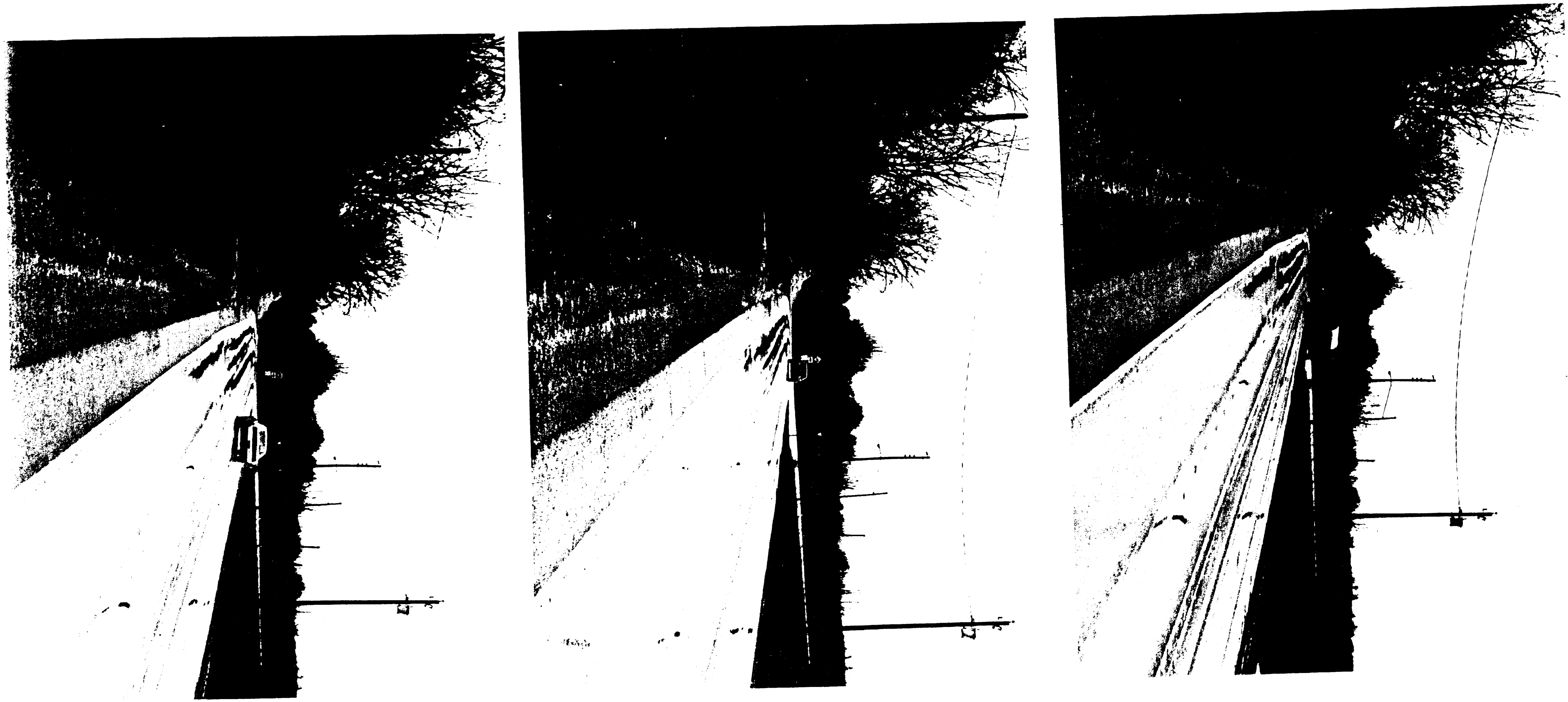
Site 29



Site 30

Site 31

Site 32



Site 33


STOP

Site 34


STOP


STOP

Site 35


STOP


STOP

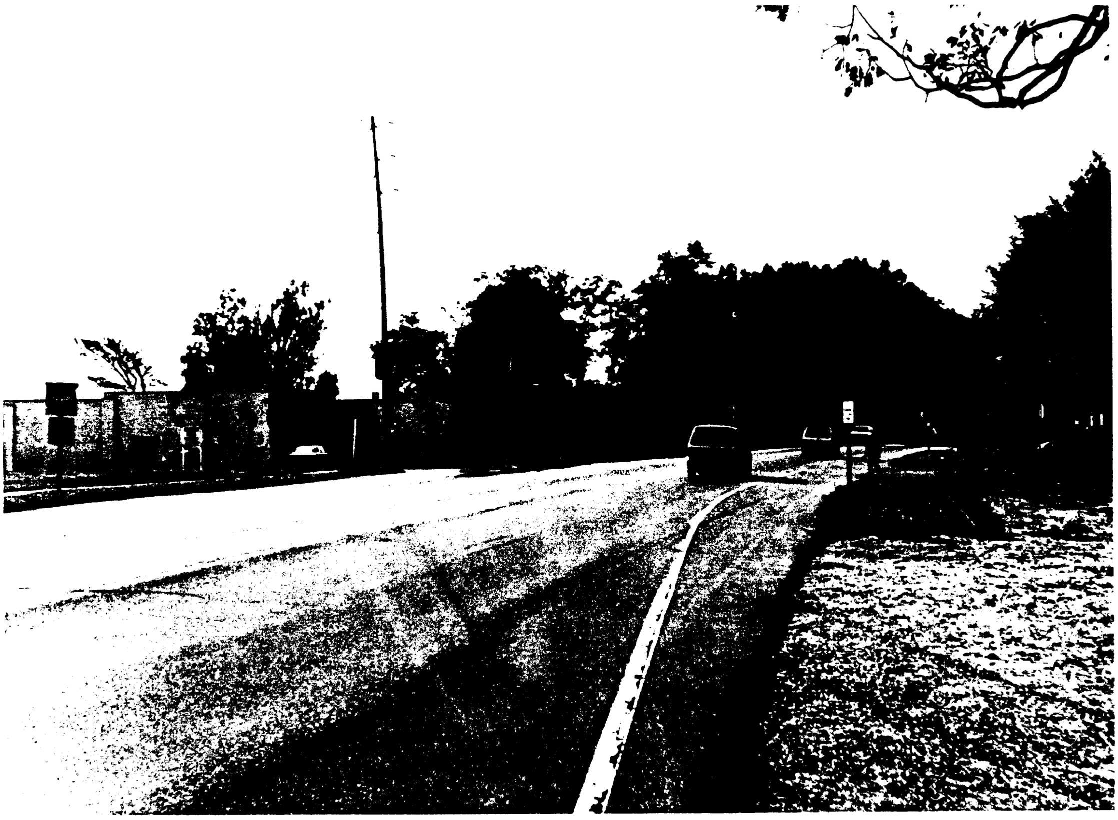
Site 36


STOP

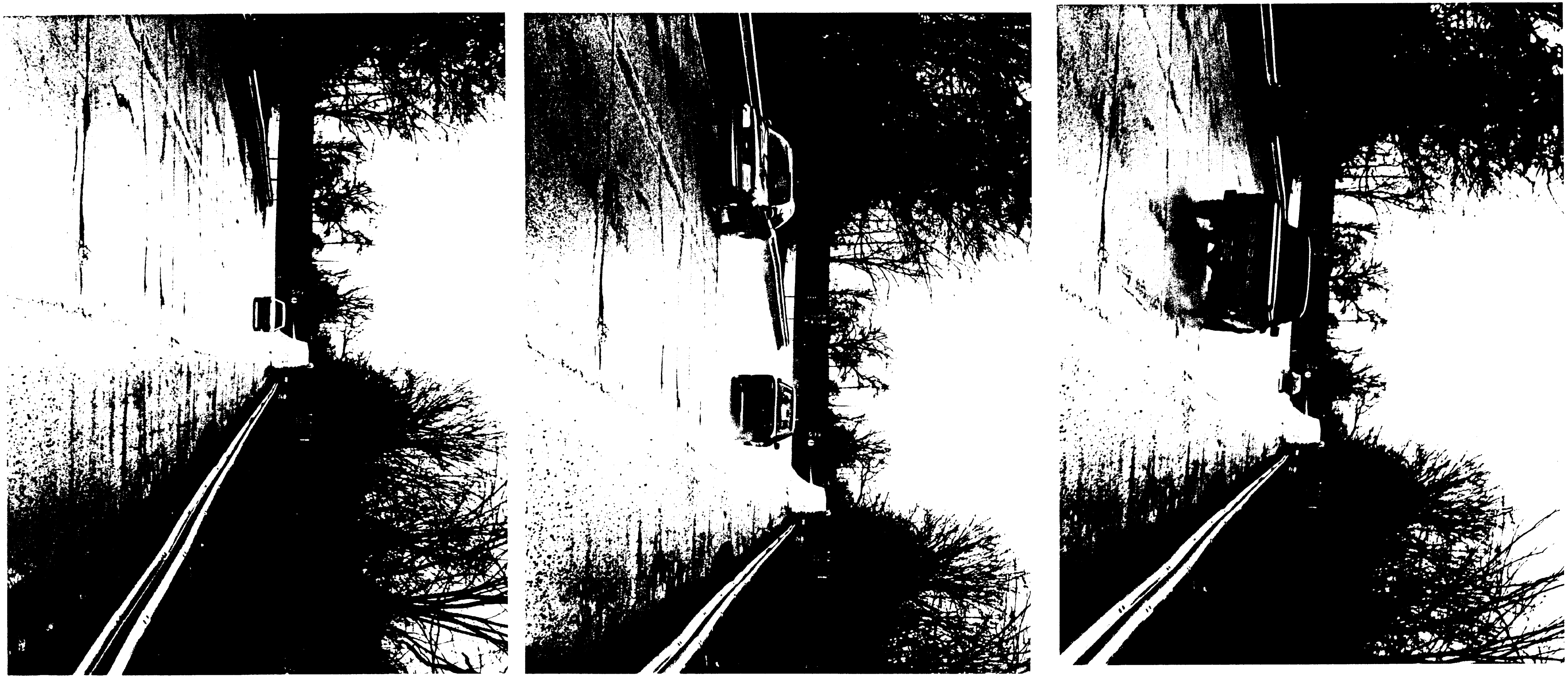

STOP

Site 37

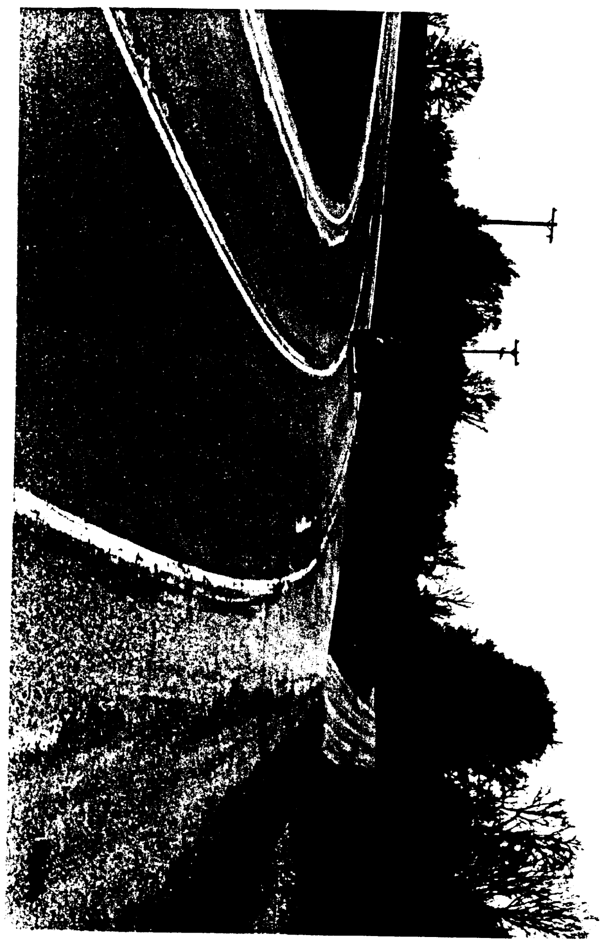
Site 38

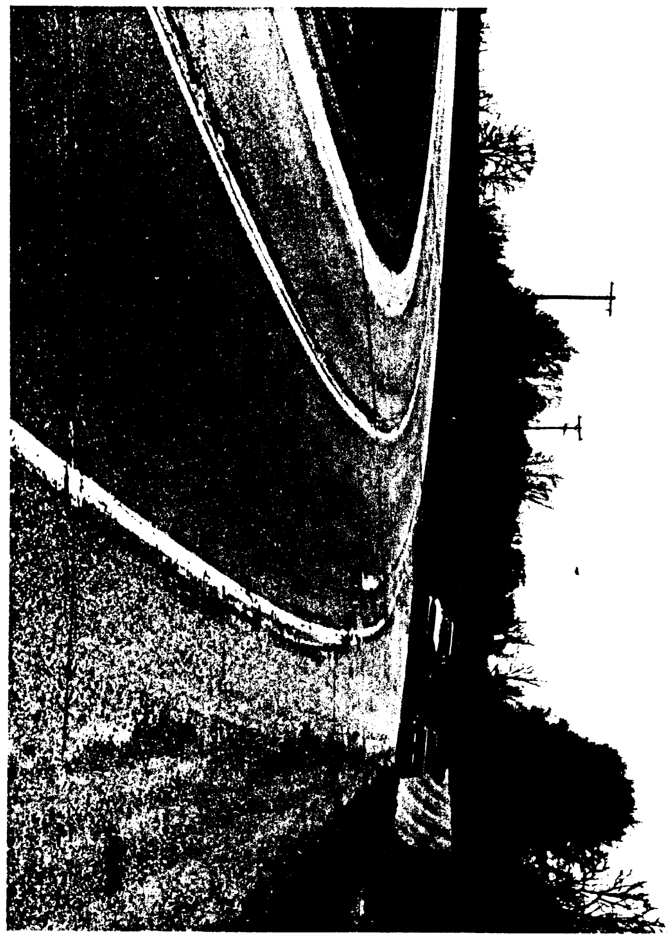


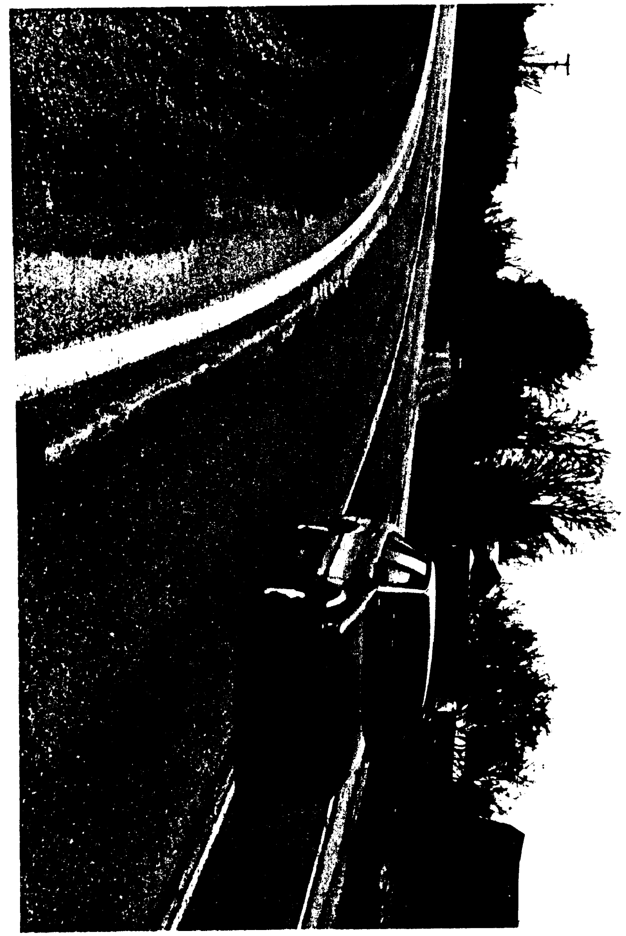

STOP



Site 39



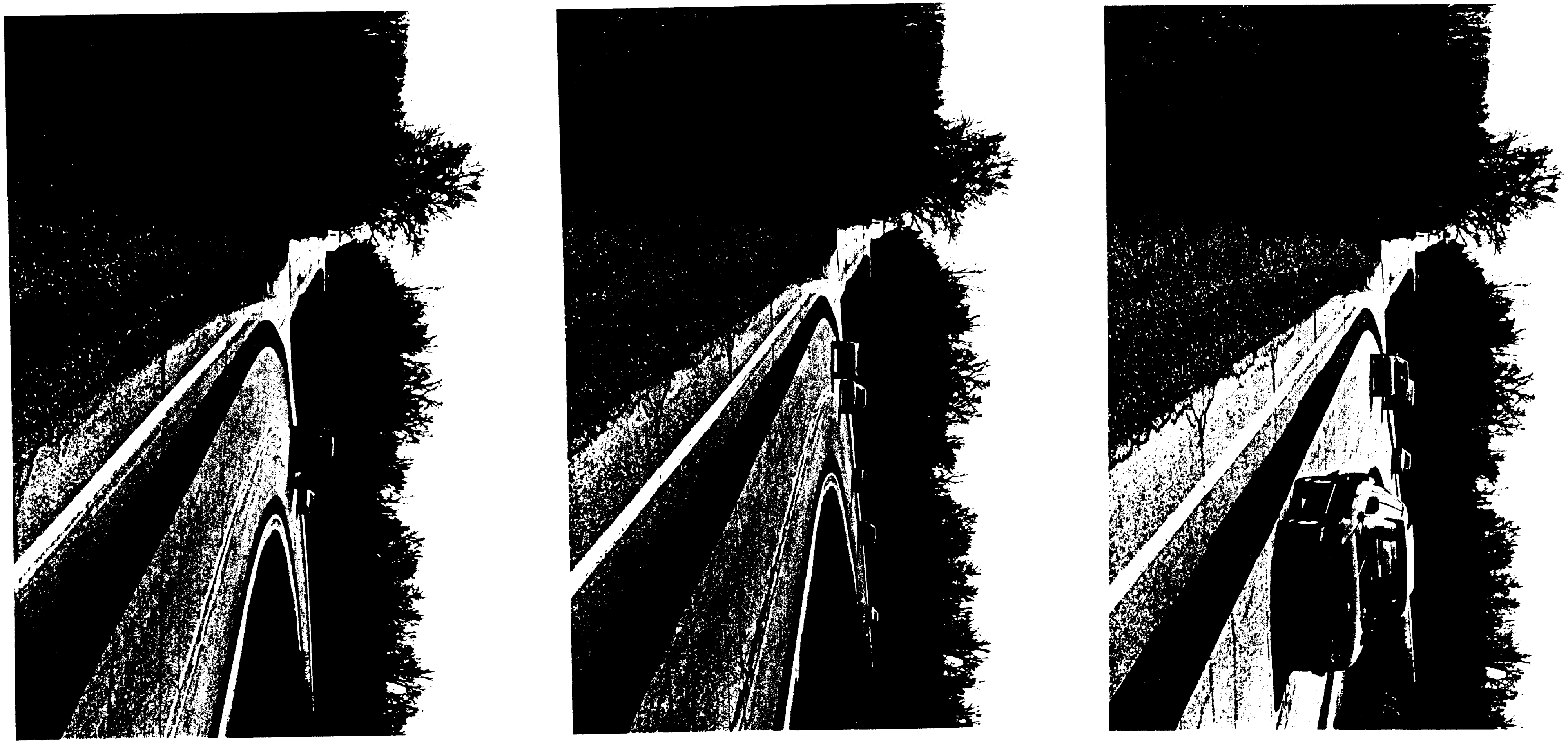




Site 40


STOP


STOP

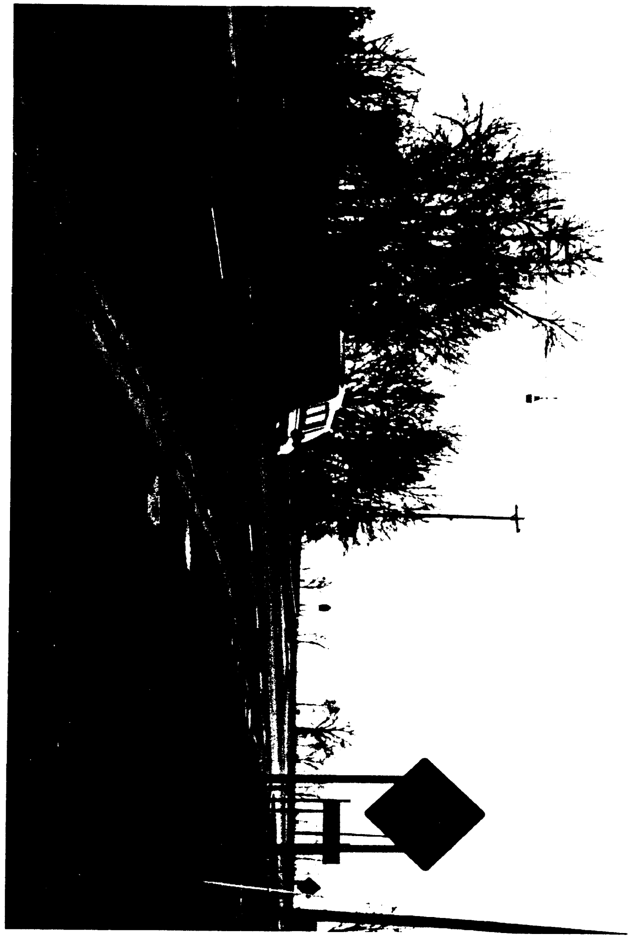


Site 41


STOP


M 34
STOP

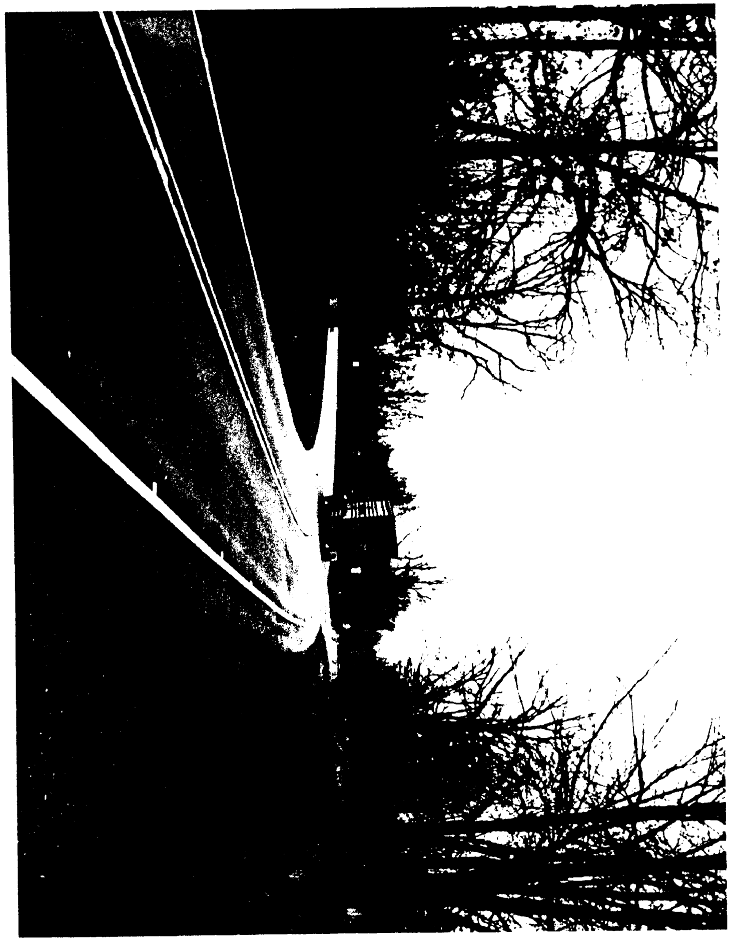
Site 42



Site 43


ONLY

Site 44

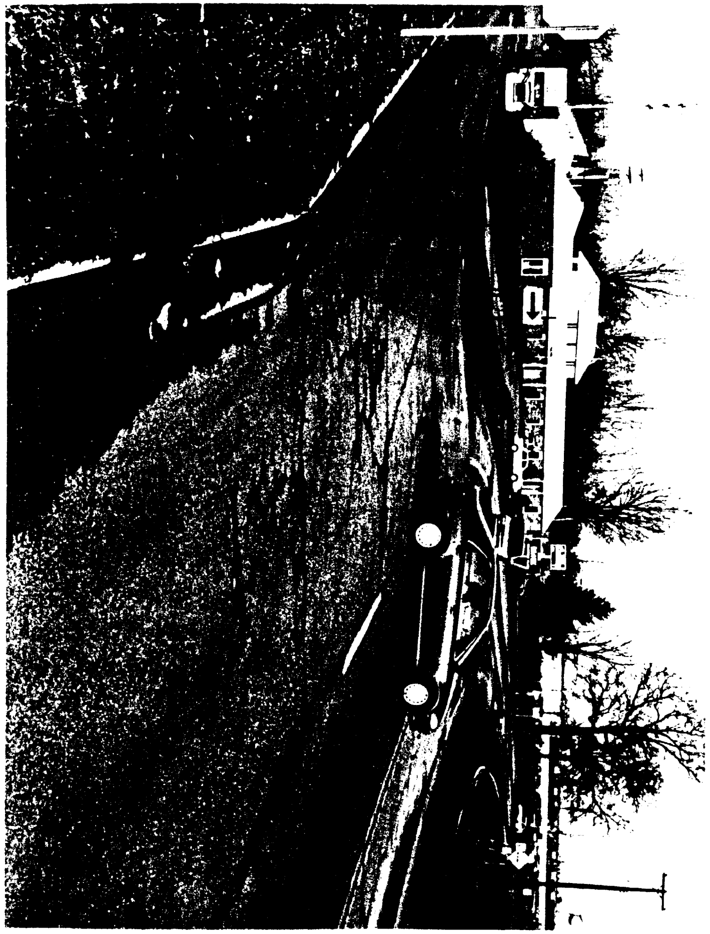


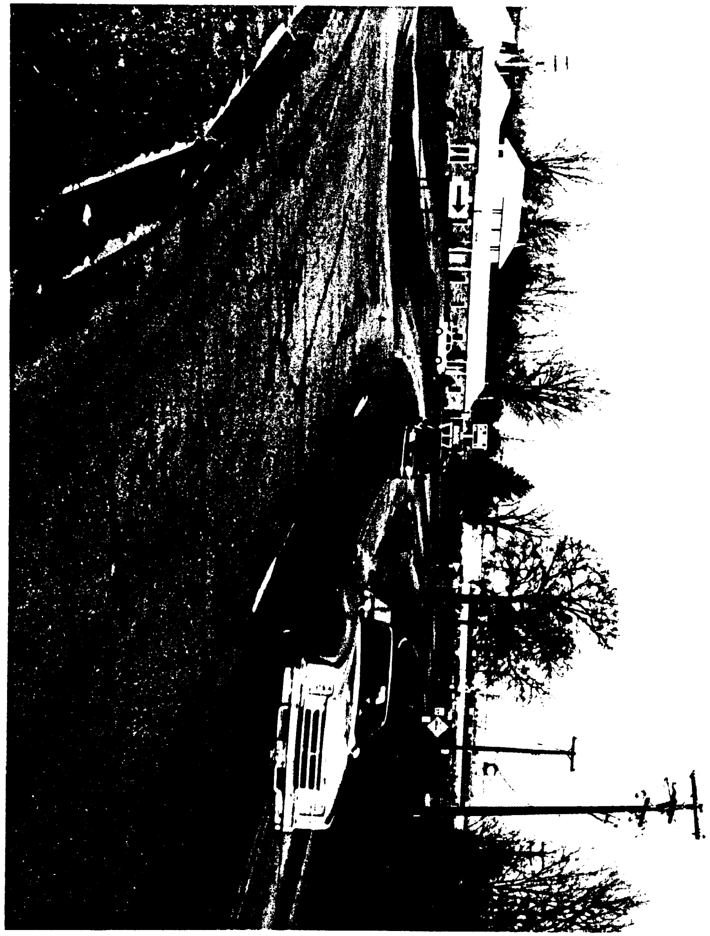
Site 45

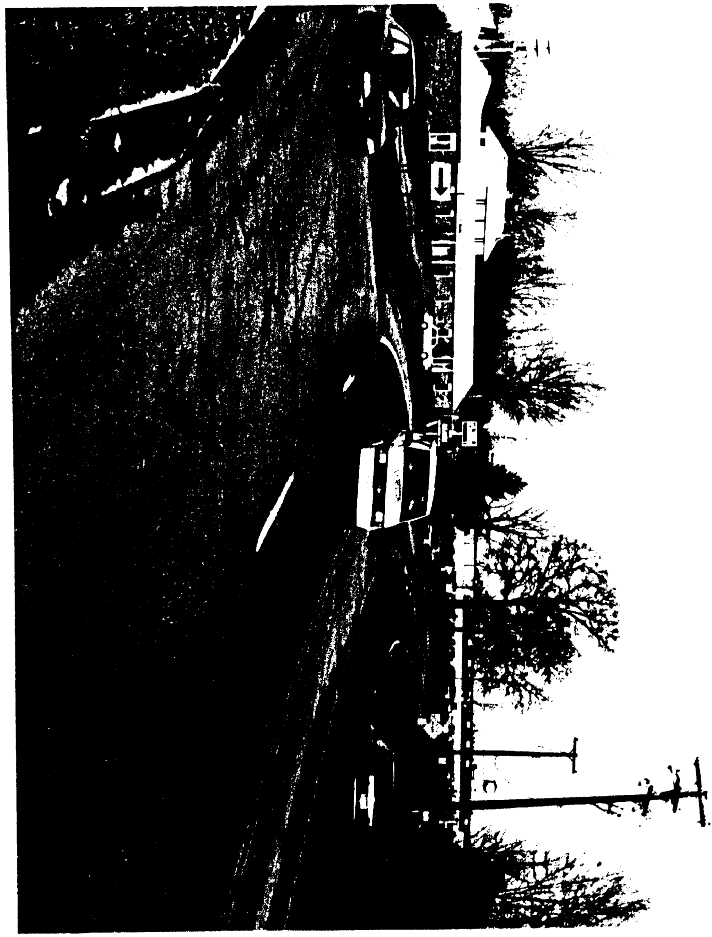

STOP


STOP

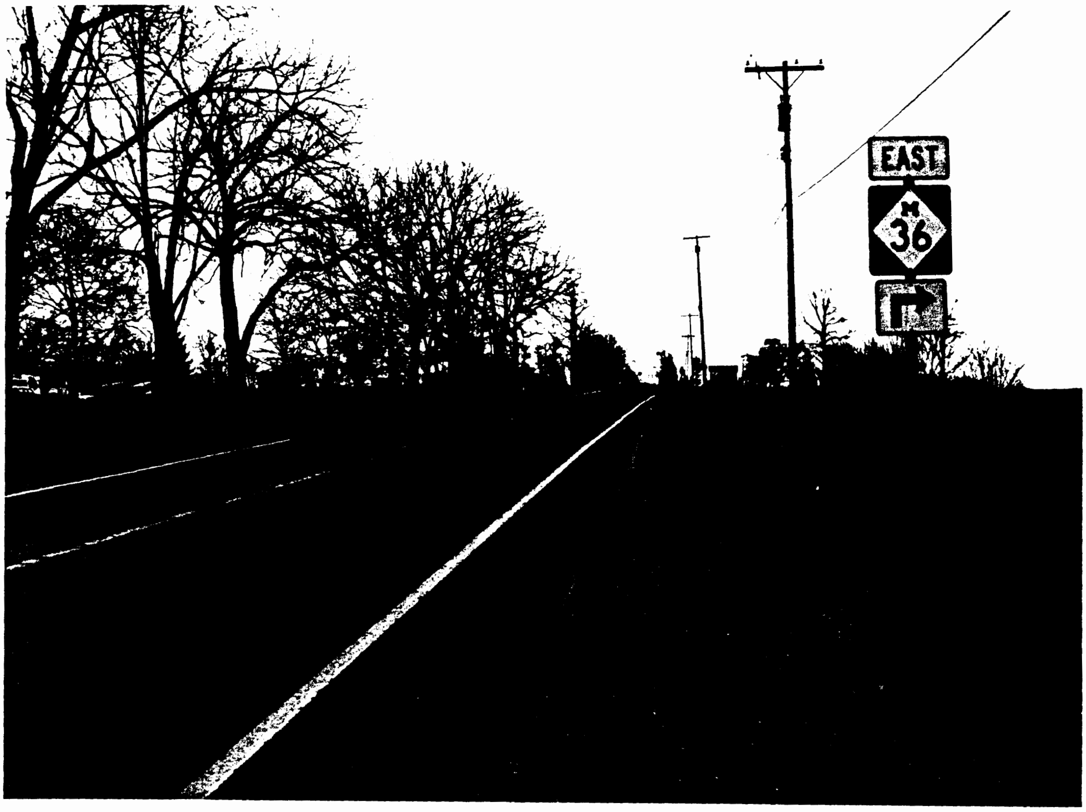
Site 46





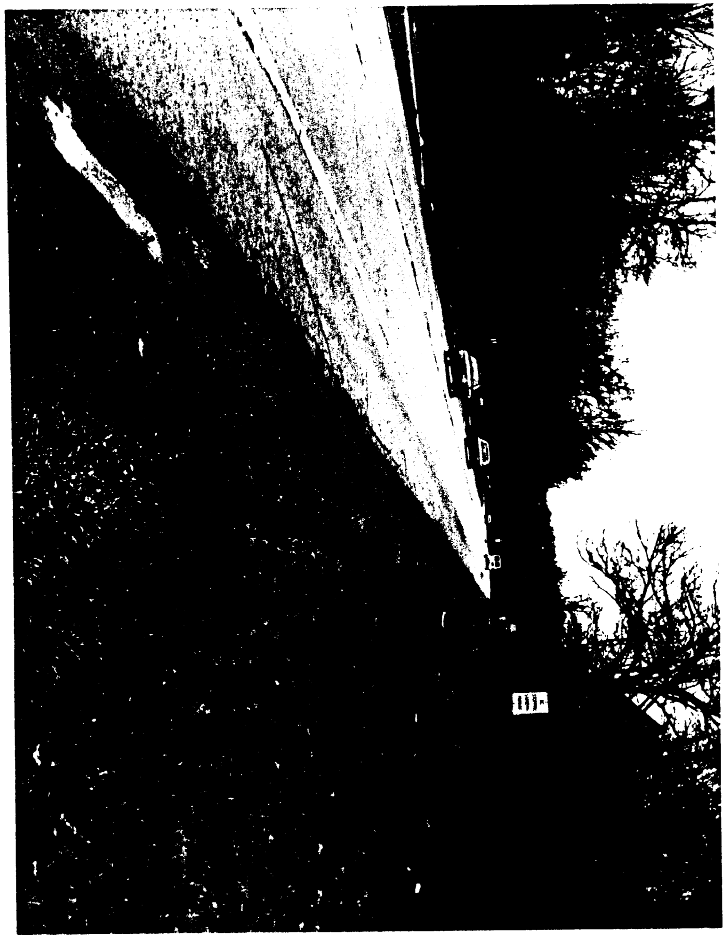


Site 47

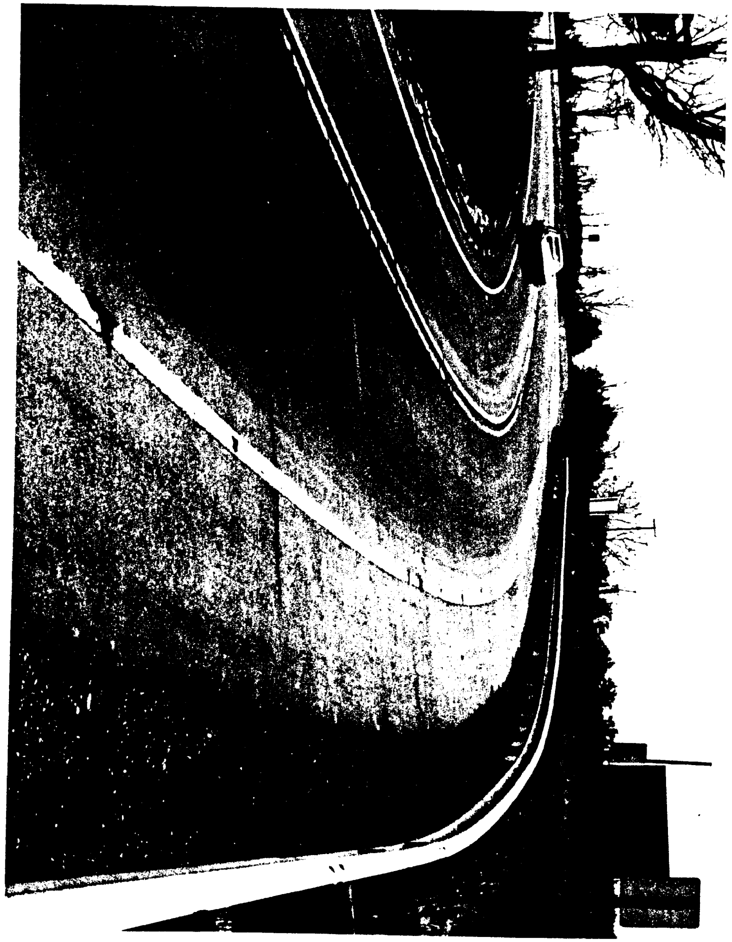


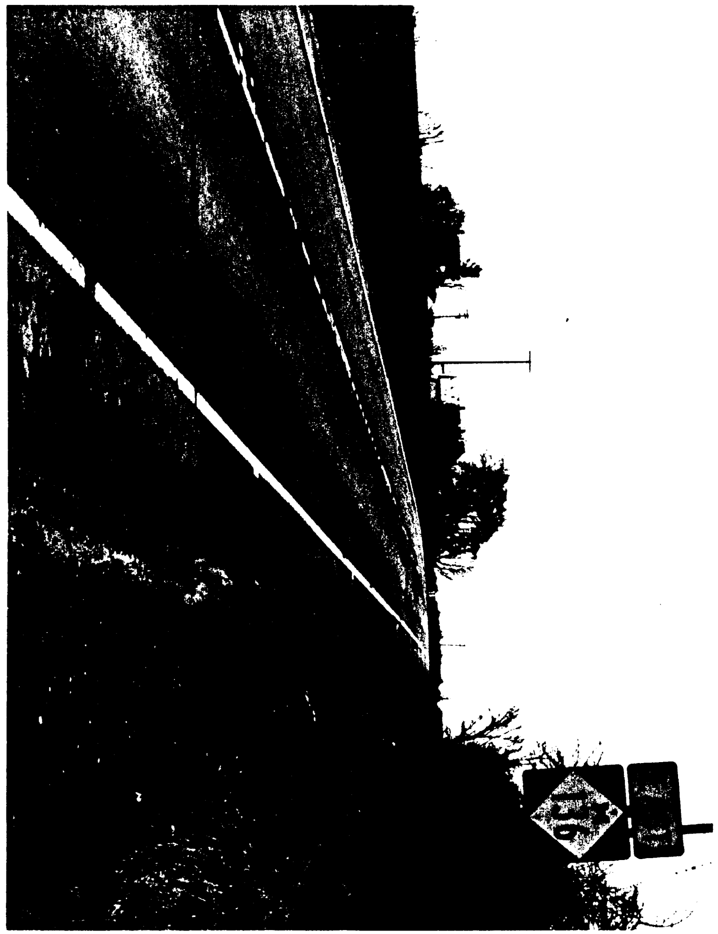
Site 48

Site 49

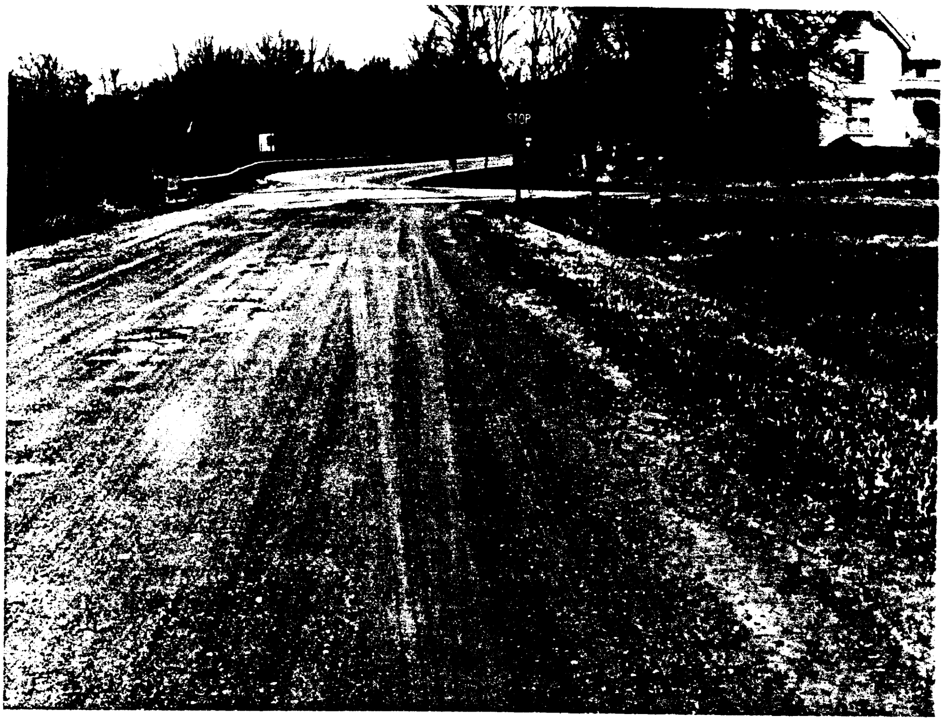


Site 50

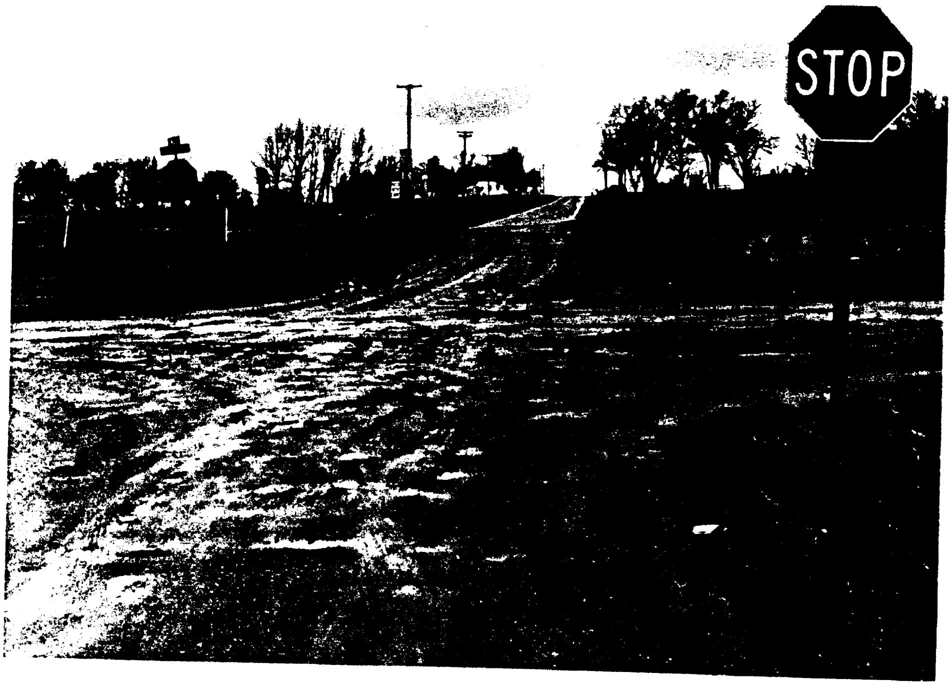




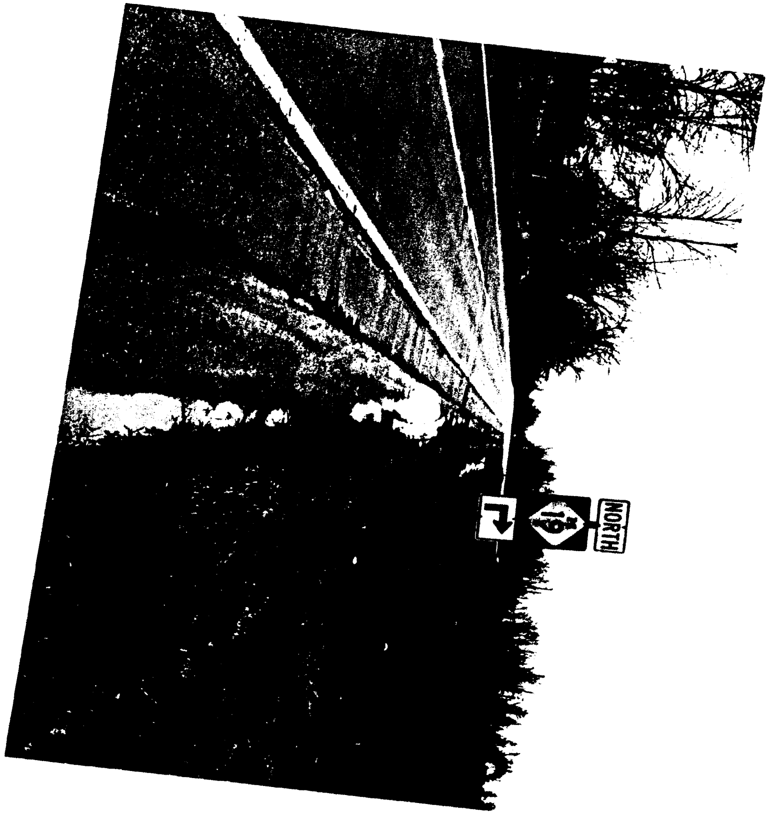
Site 51


STOP

Site 52


STOP


DO
NOT
PASS



Site 53