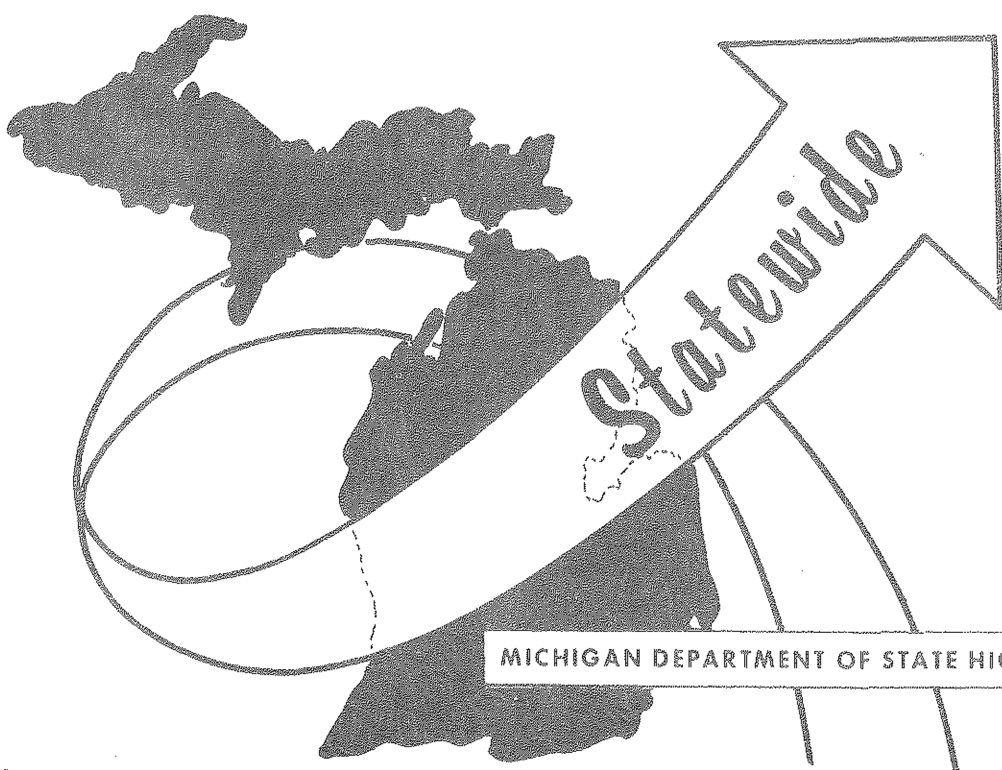
Statewide Transportation Analysis & Research

MICHIGAN'S
STATEWIDE TRANSPORTATION
MODELING SYSTEM

**PRELIMINARY INVESTIGATION:
A TECHNIQUE FOR THE PRO-
JECTION OF ACCIDENT RATES**

Report no. 11
APRIL 1975
STATEWIDE
RESEARCH AND DEVELOPMENT



MICHIGAN DEPARTMENT OF STATE HIGHWAYS AND TRANSPORTATION

MICHIGAN DEPARTMENT

OF

STATE HIGHWAYS AND TRANSPORTATION

BUREAU OF TRANSPORTATION PLANNING

MICHIGAN'S
STATEWIDE TRANSPORTATION
MODELING SYSTEM

**PRELIMINARY INVESTIGATION:
A TECHNIQUE FOR THE PRO-
JECTION OF ACCIDENT RATES**

Report no. 11
APRIL 1975
STATEWIDE
RESEARCH AND DEVELOPMENT

STATE OF MICHIGAN

WILLIAM G. MILLIKEN, GOVERNOR

DEPARTMENT OF STATE HIGHWAYS AND TRANSPORTATION

STATE HIGHWAYS BUILDING — POST OFFICE DRAWER K — LANSING, MICHIGAN 48904

JOHN P. WOODFORD, DIRECTOR

April 9, 1975

Mr. Sam F. Cryderman, Deputy Director
Bureau of Transportation Planning
Michigan Department of State Highways
and Transportation
P.O. Drawer K
Lansing, Michigan 48906

Dear Mr. Cryderman:

The following report was written by the Statewide Planning Procedures and Development Section to document their initial efforts to devise a more logical means of projecting accident rates on future year highway links. Such a development would prove most valuable in the safety analysis of regional transportation plans.

This report was prepared by Mr. Mark D. DuBay of the Statewide Section under the supervision of Richard E. Esch.

Sincerely,

R. J. Lilly

R. J. Lilly, Administrator
Highway Planning Division

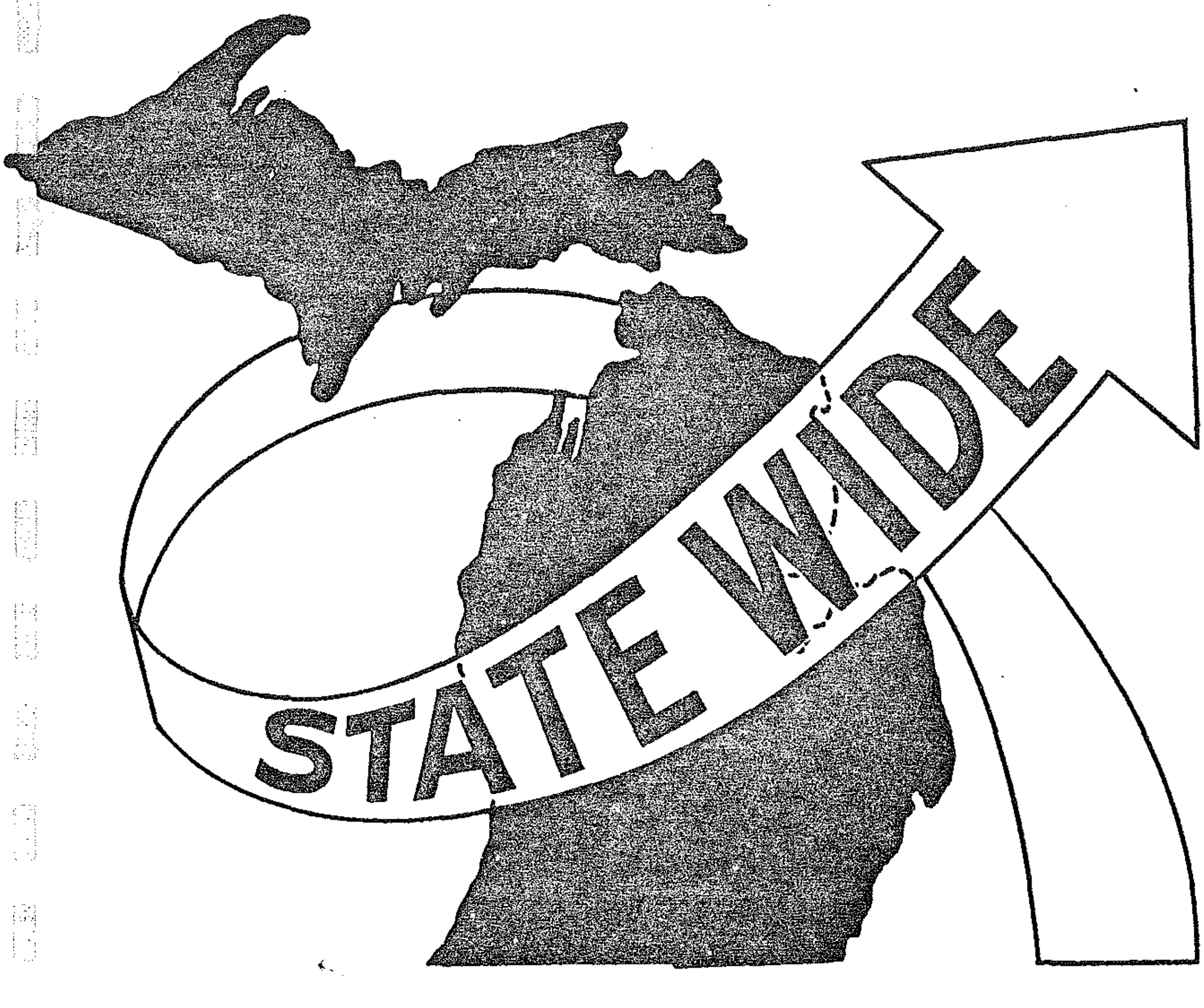
TABLE OF CONTENTS

PRELIMINARY INVESTIGATION:
A TECHNIQUE FOR THE PRO-
JECTION OF ACCIDENT RATES

BY

MARK D. DUBAY

INTRODUCTION

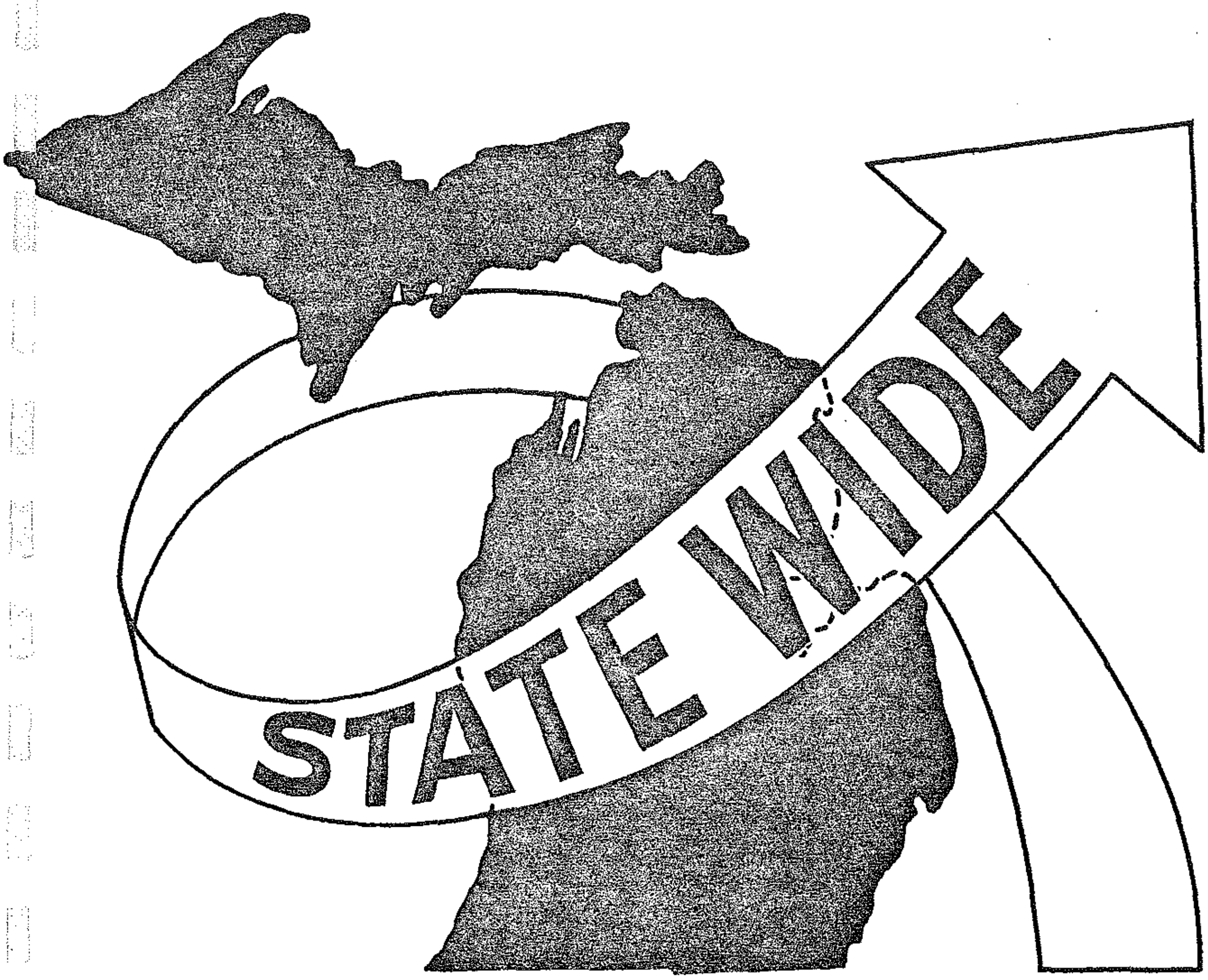


INTRODUCTION

On July 30, 1974, the Statewide Research and Development Section published a brief report* explaining how base accident rates for old and newly proposed routes are calculated within the Statewide Transportation Modeling System. Since no technique was then available for the projection of these base year accident rates, they, out of necessity, were assumed to remain constant. Although an obvious flaw in this type of reasoning was apparent, it nevertheless gave the Department a means of evaluating alternate transportation plans until a better method could be devised. This report documents a serious attempt to find a more logical means of calculating base and future year accident rates on a link-by-link basis from variables which "physically" describe a roadway - e.g. its right-of-way, sight distance and/or surface condition. Although the data presented here does not substantiate our initial belief that it is possible to project accident rates with the proper combination of such variables, it does not, by the same token, dissuade us from believing that there is merit in this approach. Those who wish to contine our cursory investigation in this area should find this information valuable in that it frees them to investigate other possible avenues without fear of duplicating past efforts.

* Accident Rates: 547 Zone System
By Alan R. Friend

PROBLEM DEFINITION



PROBLEM DEFINITION

The following formula is presently used to calculate accident rates within the Statewide Transportation Modeling System.

$$\text{Accident Rate} = \frac{\text{Number of Accidents} \times 100{,}000{,}000}{\text{Distance} \times \text{AADT} \times 365}$$

Notice that the two "key" variables used in this formula are the number of accidents which have historically occurred on a road (expressed in terms of 100 million vehicle miles) and the observed AADT. To project future accident rates on a road, the Department must then be able to also project these key variables. The calculation of probably future AADT has been a fairly routine process for quite some time now but, as suggested in the introduction, no procedure is currently available to forecast link-specific future accident rates. If the Department is to evaluate alternate highway plans in terms of safety, which is one of its prime responsibilities, it is forced to assume that accident rates remain constant through time. Although this assumption is unreasonable (e.g. the impact of technology will surely change these rates), it allows the Department to choose a "safest" highway plan from a series of alternate proposals.

The reader who is familiar with the procedures utilized in a transportation modeling system* will recognize the fact that alternate highway plans shift minimum paths between zones and therefore the flow of traffic between them. Although the link-specific accident rates do not change from plan-to-plan, their projectd traffic volumes do. Logically, then, because the above accident rate formula is tied to these varying traffic volumes, a certain plan will emerge as superior (from a safety perspective) when the historical accident

*STATEWIDE TRAVEL IMPACT ANALYSIS PROCEDURES (VOL X-A) By Mark D. DuBay rate is multiplied by the projected number of trips passing through each system link and there rates are summed for all paths within the system. The higher the traffic volumes assigned by the model to the more hazardous highway paths, obviously, the less safe the over-all plan will be in the final analysis.

This type of safety evaluation ignores many crucial factors which may, in some way, be responsible for the number of accidents experienced on a certain section of the highway system. Some variables are too complex to be included or even discussed here (e.g. the role of human behavior) but others are fairly simple and, in fact, already exist on the modeling system's network file. Figure 1, taken from this report's predecessor (see footnote on previous page), indicates that there is a definite difference in the accident rates between urban and rural links. As the density of traffic flow increases on urban links, can we be confident enough to say that accidents will in turn consistently rise? From the information in this figure, the design of a road (expressway vs. non-expressway) would seem to emerge as a significant variable. Does the number and width of road lanes significantly influence the occurance of accidents on a certain highway link? The point here is simply this - there is evidence to suggest that if a road segment possesses certain physical qualities it may be the site of an undue number of accidents when the proper combination of truly "causal" variables is also present. The question becomes whether such physical traits have a pervasive effect; whether they are sufficiently related to accident rates to allow the Department to use this relationship to predict accidents. If such a relationship were found to exist, a more realistic, easily updated accident rate could be assigned to each system link - as its physical characteristics change (are projected to change) so too would its assigned

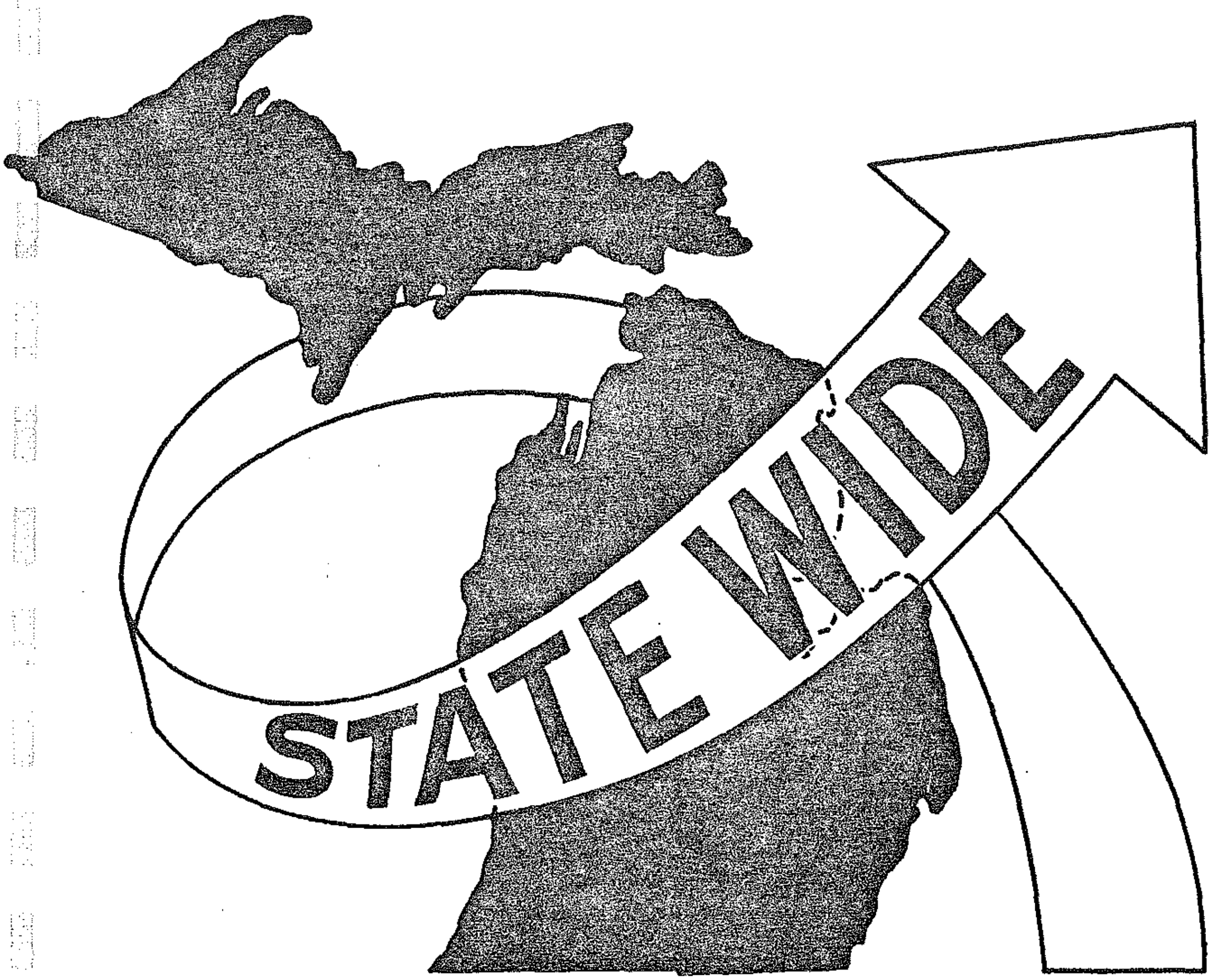
JURISDICTION 1 IS (RUR)		
# LANES=4	147.9	TOTAL MILES= 697.68
# LANES=GREATER THAN 4	145.2	TOTAL MILES= 85.92
JURISDICTION 2 IS (URB)		
# LANES=4	243.7	TOTAL MILES= 56.55
# LANES=GREATER THAN 4	170.5	TOTAL MILES= 98.68
JURISDICTION 3 FAP (RUR)		
# LANES=4 AND GREATER EXPRESSWAY	156.3	TOTAL MILES= 101.17
# LANES=4 AND GREATER NON—EXPRESSWAY	352.4	TOTAL MILES= 695.62
# LANES=LESS THAN 4	386.3	TOTAL MILES= 4303.38
JURISDICTION 4 FAP (URB)		
# LANES=4 AND GREATER EXPRESSWAY	307.0	TOTAL MILES= 133.89
# LANES=4 AND GREATER NON—EXPRESSWAY	648.2	TOTAL MILES= 391.00
# LANES=LESS THAN 4	864.9	TOTAL MILES= 210.67
JURISDICTION 5 FAS (RUR)		
# LANES=4 AND GREATER	1174.4	TOTAL MILES= 10.43
# LANES=LESS THAN 4	418.2	TOTAL MILES= 2239.22
JURISDICTION 6 FAS (URB)		
# LANES=4 AND GREATER	816.6	TOTAL MILES= 19.23
# LANES=LESS THAN 4	1449.1	TOTAL MILES= 60.75

FIGURE 1

SUMMARIZED FROM 1970 HIGHWAY NETWORK

accident rate. Such a relationship would be invaluable in assigning accident rates to proposed highway routes which currently are assigned rates which reflect those experienced on "similar" roads. A technique of this type would lead to a better safety evaluation of proposed highway plans.

ANALYSIS



Anaylsis

To quickly test our hypothesis that indeed a road's physical traits could be used as a predictor of accident rates, it was decided to use regression analysis to determine the strength of the statistical relationship. The physical "descriptors" which were to be utilized as the independent variables in this analysis currently reside on the modeling system's network file. They have been obtained from various Divisions throughout the Department and have undergone a conversion process only when necessary. A 1970 volume-to-capacity ratio was used in both its daily and hourly forms as a possible predictor of accident rates - abbreviated as DVCR and HVCR in the following discussion. Other descriptive variables and hopeful predictors included the number of lanes (NLAN), the lane width (LANW), the right-of-way (ROW), surface condition (SURF) and the sight distance (SITE).

The reader is assumed to be at least vaguely familiar with the concepts involved in regression analysis. No attempt is made to even briefly explain the statistics produced. The first and more lengthly portion of the investigation involved the use of the simple regression technique to determine if any relationship exists between the above mentioned variables and the accident rates on the major state trunklines. These trunklines included only those roads of the first six jurisdictional types.*

---

*The 1970 network has approximately 1900 links of these types within the system.

Jurisdiction 1 - Rural Interstate
Jurisdiction 2 - Urban Interstate
Jurisdiction 3 - Rural FAP
Jurisdiction 4 - Urban FAP
Jurisdiction 5 - Rural FAS
Jurisdiction 6 - Urban FAS

Figures 2 through 7 show the statistics calculated when each of the physical descriptors are regressed against accidents on each of the six road types. Although the value of R^2 (Fraction of Removed Variance) is, of course, not the only indicator used to determine the strength of a relationship between two variables, it is one of the first statistics generally considered. If a valid and usable relationship were found to exist between the dependent variable (accident rate) and one of the independent variables (physical descriptors), this R^2 value would approach either a +1 or a-1. Of the forty values presented in these first six figures none exceed .07. No relationship is therefore assumed to exist. (From the scatter plots provided in the original output, no relationship whatever seems to exist - i.e., neither one of a linear nor of a curvilinear nature.)

To be completely confident with our investigation it was decided to re-run the data using the simple and the multiple regression techniques. In both runs the dependent and independent variables remained the same. But in the former (simple regression) approach all six road types were "lumped" together - they were not stratified into groups as they had for the previous six runs. The relevant output appears in Figure 8. The R^2 value for all seven variables used is again extremely low. The multiple regression technique allows its user to emply a series of independent variables as a means of

PROC DEF= JOB1
DATA DEF= DESCRP HIGHWAY PHYSICAL DESCRIPTORS BASIS
REGRESSION; Y=AX+B

X VARIABLE	Y VARIABLE	N	REGRESSION COEFF A	B	S.ERR	R(X,Y)	FRACTION OF REMOVED VAR
HVCR	ACC	183	34.1691	137.3038	97.021	0.075	5.67e-03
DVCR	ACC	183	55.9082	133.3945	96.523	0.126	0.02
NLAN	ACC	183	-3.8376	171.8596	97.263	-0.026	6.91e-04
LANW	ACC	183	6.7512	75.5000	96.927	0.087	7.58e-03
SURF	ACC	183	4.6790	145.4267	97.194	0.046	2.11e-03
ROW	ACC	183	-0.1651	203.7058	96.755	-0.105	0.01

*** VARIANCE OF X OR Y VAR=0, ANALYSIS OF SITE DELETED.

FIGURE 2

JOB DEF= ****** PREDICT ACCIDENTS AS A FUNCTION OF ROAD DESCRIPTORS FOR JUR=2
PROC DEF= JOB2
DATA DEF= DESCRP HIGHWAY PHYSICAL DESCRIPTORS BASIS
REGRESSION; Y=AX+B

X VARIABLE	Y VARIABLE	N	REGRESSION COEFF A	REGRESSION COEFF B	S.ERR	R(X,Y)	FRACTION OF REMOVED VAR
HVCR	ACC	64	109.2556	167.8571	194.456	0.182	0.03
DVCR	ACC	64	52.5654	200.6433	196.538	0.110	0.01
NLAN	ACC	67	-14.2652	306.3368	193.947	-0.118	0.01
LANW	ACC	67	1.2826	217.6337	195.294	0.017	3.05e-04
SURF	ACC	67	48.0094	136.7275	188.017	0.271	0.07
ROW	ACC	67	0.1855	178.5335	193.903	0.120	0.01

*** VARIANCE OF X OR Y VAR=0, ANALYSIS OF SITE DELETED.

FIGURE 3

JOB DEF= ****** PREDICT ACCIDENTS AS A FUNCTION OF ROAD DESCRIPTORS FOR JUN=3

PROC DEF= JOB3
DATA DEF= DESCRP HIGHWAY PHYSICAL DESCRIPTORS

BASIS

REGRESSION: Y=AX+B

X VARIABLE	Y VARIABLE	N	REGRESSION COEFF A	REGRESSION COEFF B	S.ERR	R(X,Y)	FRACTION OF REMOVED VAR.
HVCR	ACC	931	101.6766	297.5433	280.743	0.168	0.03
DVCR	ACC	931	151.8305	290.5275	277.654	0.222	0.05
NLAN	ACC	935	4.9538	364.3358	285.198	0.016	2.60e-04
LANW	ACC	935	-1.1943	389.9947	285.228	-0.007	5.01e-05
SURF	ACC	935	19.7330	320.5688	284.473	0.073	5.34e-03
ROW	ACC	935	-0.7413	447.3722	279.050	-0.207	0.04
SITE	ACC	935	1.1838	365.6498	284.720	0.060	3.61e-03

FIGURE 4

JOB DEF= ****** PREDICT ACCIDENTS AS A FUNCTION OF ROAD DESCRIPTORS FOR JUR=4
PROC DEF= JOB4
DATA DEF= DESCRP HIGHWAY PHYSICAL DESCRIPTORS

BASIS

REGRESSION: Y=AX+B

X VARIABLE	Y VARIABLE	N	REGRESSION COEFF A	REGRESSION COEFF B	S.ERR	R(X,Y)	FRACTION OF REMOVED VAR.
HVCR	ACC	387	45.8859	813.0772	742.726	0.042	1.73E-03
DVCR	ACC	387	21.2135	841.3198	743.233	0.019	3.69E-04
NLAN	ACC	392	-31.8406	987.4959	742.647	-0.079	6.23E-03
LANW	ACC	392	14.9321	700.5102	743.521	0.062	3.89E-03
SURF	ACC	392	45.5059	736.5856	742.950	0.074	5.41E-03
ROW	ACC	392	-2.1728	1061.5845	718.518	-0.264	0.07
SITE	ACC	392	-4.7461	860.0369	744.485	-0.036	1.30E-03

FIGURE 5

JOB DEF= ****** PREDICT ACCIDENTS AS A FUNCTION OF ROAD DESCRIPTORS FOR JUR=5 PAGE 1
PROC DEF= JOB5
DATA DEF= DESCRP HIGHWAY PHYSICAL DESCRIPTORS BASIS
REGRESSION: Y=AX+B

X VARIABLE	Y VARIABLE	N	REGRESSION COEFF A	REGRESSION COEFF B	S.ERR	R(X,Y)	FRACTION OF REMOVED VAR.
HVCR	ACC	338	27.4073	385.3824	286.285	0.033	1.08e-03
DVCR	ACC	338	62.4668	378.7492	285.751	0.069	4.80e-03
NLAN	ACC	338	130.7929	133.6819	283.799	0.135	0.02
LANW	ACC	338	12.0054	274.0322	285.684	0.073	5.27e-03
SURF	ACC	338	-4.6087	413.2464	286.388	-0.019	3.58e-04
ROW	ACC	338	-1.0926	458.4940	285.368	-0.086	7.47e-03
SITE	ACC	338	1.1826	379.3880	285.544	0.079	6.24e-03

FIGURE 6

JOB DEF= ******   PREDICT ACCIDENTS AS A FUNCTION OF ROAD DESCRIPTORS FOR JUR=6   PAGE 1
PROC DEF= JOB6
DATA DEF= DESCRP   HIGHWAY PHYSICAL DESCRIPTORS   BASIS
REGRESSION: Y=AX+B

X VARIABLE	Y VARIABLE	N	REGRESSION COEFF A	B	S.ERR	R(X,Y)	FRACTION OF REMOVED VAR
HVCR	ACC	42	103.0656	1292.9957	3313.599	0.031	9.45e-04
DVCR	ACC	42	199.4998	1215.5415	3310.209	0.055	2.99e-03
NLAN	ACC	44	-150.0514	1745.7743	3238.834	-0.064	4.05e-03
LANW	ACC	44	29.4714	1058.6970	3244.053	0.029	8.36e-04
SURF	ACC	44	79.4976	1110.2910	3243.954	0.030	8.97e-04
ROW	ACC	44	-9.5004	1800.8945	3227.704	-0.104	0.01
SITE	ACC	44	-38.4211	1427.1915	3236.949	-0.072	5.21e-03

FIGURE 7

JOB DEF= ****** PROC DEF= JOB7 DATA DEF= DESCRP REGRESSION: Y=AX+B

PREDICT ACCIDENTS AS A FUNCTION OF ROAD DESCRIPTORS FOR JUR=1-6

HIGHWAY PHYSICAL DESCRIPTORS

BASIS

X VARIABLE	Y VARIABLE	N	REGRESSION COEFF A	B	S.ERR	R(X,Y)	FRACTION OF REMOVED VAR
HVCR	ACC	1945	211.8089	307.9980	666.024	0.168	0.03
DVCR	ACC	1945	262.6994	308.5275	660.883	0.208	0.04
NLAN	ACC	1959	18.5180	416.2311	673.603	0.040	1.62e-03
LANW	ACC	1959	-13.7305	621.0484	673.499	-0.044	1.93e-03
SURF	ACC	1959	26.8136	399.5986	673.466	0.045	2.02e-03
ROW	ACC	1959	-1.3076	618.7307	660.561	-0.200	0.04
SITE	ACC	1959	-1.8226	486.5849	673.645	-0.039	1.49e-03

FIGURE 8

more accurately predicting a change in the dependent variable. Typically, when simple regression analysis is used and no relationship between variables is found, use of the multiple regression approach will yield no better results. But in some cases data values "interact" causing the importance of certain variables to be significantly improved. Unfortunately, the combination of independent variables which were at our disposal did not display this phenomenon. As can be seen in Figure 9, regressing all seven independent variables against the accident rates for all six road types simultaneously resulted in a relationship only slightly better than the previous R^2 values (Coefficient of Determination = .0854).

JOB DEF= ******
PROC DEF= JOB8
DATA DEF= DESCRP

PREDICT ACCIDENTS AS A FUNCTION OF ROAD DESCRIPTORS FOR JUR=1-6

HIGHWAY PHYSICAL DESCRIPTORS

BASIS

MULTIPLE LINEAR REGRESSION ANALYSIS

GROUP NO 1
NO OF INDEPENDENT VARIABLES 6

EPS= 5.0000E-11 FOR SINGULAR MATRIX TEST

DEPENDENT VARIABLE = ACC

MEANS AND STANDARD DEVIATIONS

VARIABLE	MEAN	ST.DEVIATION
HVCR	0.7897	0.5347
NLAN	3.0494	1.4475
LANW	10.9090	1.9554
SURF	2.7321	1.1110
ROW	112.4416	102.7958
SITE	7.8756	14.3161
ACC	475.2704	675.4144

NORMAL MATRIX=

	1	2	3	4	5	6
ROW 1	555.7171	80.3778	-124.7997	173.2598	-23000.0961	891.2466
ROW 2	0.0000	4073.2617	206.7362	-722.2848	158315.6021	-15446.0555
ROW 3	0.0000	0.0000	7432.8925	996.5877	117454.1712	-4182.0226
ROW 4	0.0000	0.0000	0.0000	2399.4416	-56035.9028	6826.1763
ROW 5	0.0000	0.0000	0.0000	0.0000	**********.***	-888039.1218
ROW 6	0.0000	0.0000	0.0000	0.0000	0.0000	398425.8900

FIGURE 9-A

CORRELATION MATRIX-

	1	2	3	4	5	6	
ROW 1	1.0000	0.0534	-0.0614	0.1500	-0.2153	0.0599	0.1677
ROW 2	0.0534	1.0000	0.0376	-0.2310	0.5473	-0.3834	0.0322
ROW 3	-0.0614	0.0376	1.0000	0.2360	0.3006	-0.0768	-0.0729
ROW 4	0.1500	-0.2310	0.2360	1.0000	-0.2524	0.2208	0.0353
ROW 5	-0.2153	0.5473	0.3006	-0.2524	1.0000	-0.3104	-0.2048
ROW 6	0.0599	-0.3834	-0.0768	0.2208	-0.3104	1.0000	-0.0411
ROW 7	0.1677	0.0322	-0.0729	0.0353	-0.2048	-0.0411	1.0000

FIGURE 9-B

REGRESSION COEFFICIENTS-

VARIABLE	COEFFICIENT	BETA COEFFICIENT
CONSTANT	357.8772	
HVCR	127.2919	0.1008
NLAN	74.4824	0.1596
LANW	4.2699	0.0124
SURF	-2.6590	-0.0044
ROW	-1.9655	-0.2991
SITE	-3.6272	-0.0769

STANDARD ERROR OF ESTIMATE............. = 646.9269
COEFFICIENT OF DETERMINATION........... = 0.0854
COEFFICIENT OF DETERMINATION (ADJ).... = 0.0826
MULTIPLE CORRELATION COEFFICIENT...... = 0.2922
MULTIPLE CORRELATION COEFFICIENT (ADJ) = 0.2874

STANDARD DEVIATIONS AND T VALUES OF
REGRESSION COEFFICIENTS-

VARIABLE	STD DEVIATION	T	STD DEV BETA	F
HVCR	29.0002	4.39	0.0230	19.27
NLAN	13.0693	5.70	0.0280	32.48
LANW	8.4516	0.51	0.0245	0.26
SURF	14.7925	-0.18	0.0243	0.03
ROW	0.1942	-10.12	0.0296	102.46
SITE	1.1324	-3.20	0.0240	10.26

HIGH ORDER PARTIAL CORRELATION COEFF AND R2-DELETE

VARIABLE	PARTIAL CORR COEF	R2-DELETE
HVCR	0.0992	0.0763
NLAN	0.1284	0.0701
LANW	0.0115	0.0853
SURF	-0.0041	0.0854
ROW	-0.2241	0.0371
SITE	-0.0726	0.0806

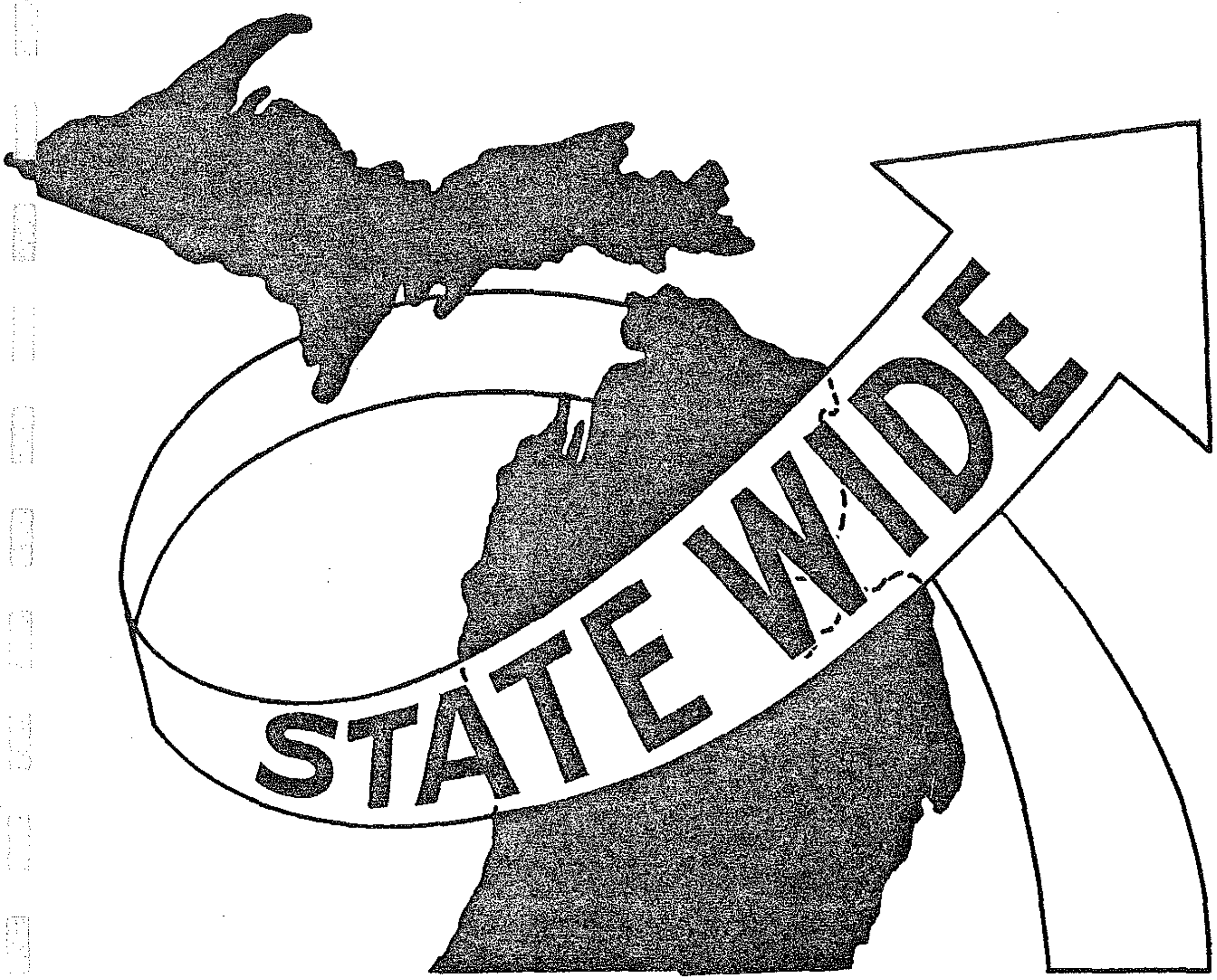
ANALYSIS OF VARIANCE TABLE

SOURCE	D.F.	SUM SQUARES	MEAN SQUARES	F
REGRESSION	6	75741849.9320	12623641.6554	30.16
ERROR	1938	811080969.8200	418514.4323	
TOTAL	1944	886822819.7500		

CONFIDENCE LEVEL OF F(6, 1938) = 100.00%

FIGURE 9-C

CONCLUSION



CONCLUSION

It was hoped at the beginning of this study that the seven independent variables utilized in the regression equation would have sufficient explanatory power to readily permit the projection of accident rates without further investigation or expensive data collection. Since the data for these variables already existed within the highway link file, the finding of a relationship between a road's accident rate and "descriptors" of its physical features would have, of course, made the cost of model development extremely low. The results of this preliminary investigation indicated that these seven independent variables do not in themselves possess the necessary "power of prediction". This is not to say that these variables should be discounted in any future study but rather other variables which describe a road's physical qualities should be added to them. If our original hypothesis is indeed correct, there is a proper combination of explanatory variables - the cost of determining exactly what this combination might be is, at the moment however, prohibitive. Other persons interested in continuing this study may contact the Statewide Planning Procedures and Development Section for any link-specific data they may wish to obtain.