Report No. K-TRAN: KSU - 96 - 3
Final Report

PB99-175457

# SPEED ZONE GUIDELINES USING ROADWAY CHARACTERISTICS AND AREA DEVELOPMENT

Robert W. Stokes
Yacoub M. Najjar
Eugene R. Russell
Margaret J. Rys
Jahor L. Roy
Imad A. Basheer
Hossam E. Ali

Kansas State University
Manhattan, Kansas

August 1998

## K - TRAN

A COOPERATIVE TRANSPORTATION RESEARCH PROGRAM BETWEEN:
KANSAS DEPARTMENT OF TRANSPORTATION
THE KANSAS STATE UNIVERSITY
THE UNIVERSITY OF KANSAS

| | | |
|---|---|---|
| 1. **Report No.** K-TRAN: KSU-96-3 | 2. **Government Accession No.** | 3. **Recipient Catalog No.** |
| 4 **Title and Subtitle** SPEED ZONE GUIDELINES USING ROADWAY CHARACTERISTICS AND AREA DEVELOPMENT | | 5 **Report Date** August 1998<br>6 **Performing Organization Code** |
| 7. **Author(s)** Robert W. Stokes, Yacoub M. Najjar, Eugene R. Russell, Margaret J. Rys, Jahor L. Roy, Imad A. Basheer, and Hossam E. Ali | | 8 **Performing Organization Report No.** |
| 9 **Performing Organization Name and Address** Kansas State University, Department of Civil Engineering, Manhattan, Kansas 66506 | | 10 **Work Unit No. (TRAIS)**<br>11 **Contract or Grant No.** C-889 |
| 12 **Sponsoring Agency Name and Address** Kansas Department of Transportation, Docking State Office Bldg., Topeka, Kansas 66612 | | 13 **Type of Report and Period Covered** Final Report Aug.1995 to Aug.1998<br>14 **Sponsoring Agency Code** 106-RE-0076-01 |
| 15 **Supplementary Notes** | | |

16 **Abstract**


The objective of this study was to quantify effects that selected characteristics and adjacent development patterns have on roadway speeds. Based on the results of a literature search and the availability of data for Kansas highways, twenty seven variables were identified as possibly affecting speeds on rural state highways and thirty two variables were identified as possibly affecting speeds on urban state highways. Speed data and data for the potential explanatory variables were collected for a total of 539 sections of state highway (186 rural and 353 urban sections).

Two approaches were used to develop and test models to predict speeds on rural and urban state highways based on roadway characteristics and adjacent development patterns. The first approach was based on models in the form of multiple linear regression equations. The second approach employed artificial neural networks (ANN) to predict highway speeds. None of the regression models were entirely satisfactory in terms of their ability to predict the 85$^{th}$ percentile speeds on rural and urban highways within $\pm$ 5 mph. A number of regression models are presented, however, they should be used with caution.

Two independent databases were used to train two sets of ANN models of rural and urban speeds. The first set of ANN models (Stage One) was developed using the same data used in the regression analysis. The second set of ANN models (Stage 2) was developed using a database provided by the KDOT Bureau of Traffic Engineering. The second database contained only those variables that KDOT believed drivers consider in selecting a driving speed.

Overall, the Stage 2 ANN models developed in this study were found to perform much better than either the Stage 1 ANN models or the regression models in predicting rural and urban highway speeds.


| | | | |
|---|---|---|---|
| 17 **Key Words** Artificial, database, neural networks, urban, rural, regression analysis, speed zoning, variable, back propagation, modeling. | | 18 **Distribution Statement** No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161 | |
| 19 **Security Classification (of this report)** Unclassified | 20 **Security Classification (of this page)** Unclassified | 21 **No. of pages** 82 | 22 **Price** |

Form DOT F 1700.7 (8-72) r:\ktran\17KSU96.3doc

# Speed Zone Guidelines
# Using Roadway Characteristics and Area Development

by

Robert W. Stokes
Yacoub M. Najjar
Eugene R. Russell
Margaret J. Rys
Jahor L. Roy
Imad A. Basheer
Hossam E. Ali

Kansas State University
Department of Civil Engineering
Seaton Hall
Manhattan, KS 66506-2905

Prepared for

Kansas Department of Transportation
K-TRAN Project Number KSU-96-3

Lee Roadifer
KDOT Monitor

*FINAL REPORT*

June 1999

## PREFACE

This research project was funded by the Kansas Department of Transportation K-TRAN research program. The Kansas Transportation Research and New-Developments (K-TRAN) Research Program is an ongoing, cooperative and comprehensive research program addressing transportation needs of the State of Kansas utilizing academic and research resources from the Kansas Department of Transportation, Kansas State University and the University of Kansas. The projects included in the research program are jointly developed by professionals in KDOT and the universities.

## NOTICE



## DISCLAIMER

## ABSTRACT

The objective of this study was to quantify effects that selected characteristics and adjacent development patterns have on roadway speeds. Based on the results of a literature search and the availability of data for Kansas highways, twenty seven variables were identified as possibly affecting speeds on rural state highways and thirty two variables were identified as possibly affecting speeds on urban state highways. Speed data and data for the potential explanatory variables were collected for a total of 539 sections of state highway (186 rural and 353 urban sections). Two approaches were used to develop and test models to predict speeds on rural and urban state highways based on roadway characteristics and adjacent development patterns. The first approach was based on models in the form of multiple linear regression equations. The second approach employed artificial neural networks (ANN) to predict highway speeds. None of the regression models were entirely satisfactory in terms of their ability to predict the 85th percentile speeds on rural and urban highways within $\pm$ 5 mph. A number of regression models are presented, however, they should be used with caution. Two independent databases were used to train two sets of ANN models of rural and urban speeds. The first set of ANN models (Stage One) was developed using the same data used in the regression analysis. The second set of ANN models (Stage 2) was developed using a database provided by the KDOT Bureau of Traffic Engineering. The second database contained only those variables that KDOT believed drivers consider in selecting a driving speed. Overall, the Stage 2 ANN models developed in this study were found to perform much better than either the Stage 1 ANN models or the regression models in predicting rural and urban highway speeds.

# TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# *CHAPTER 1*

# INTRODUCTION

## 1.1 BACKGROUND

Speed zoning is the establishment of reasonable and safe speed limits on roadways based on an engineering study. Speed zoning in the United States is based on the principle of setting speed limits as near as practicable to the speed at or below which 85 percent of drivers are traveling. This reflects the safe speed as determined by a large majority of the drivers. This speed is subject to revision based upon such factors as accident experience, roadway geometry, parking, pedestrians, adjacent development and engineering judgment.

There are several problems associated with current procedures for establishing speed limits. One problem is that speed limits must be established on the basis of an engineering study. Specifically, speed studies must be conducted to determine the 85th percentile speed for a given section of roadway. These studies are essential in establishing realistic speed limits, but they are expensive and time consuming to conduct. The time and cost needed to perform speed studies becomes particularly significant in light of recent increases in the number of requests from citizens and neighborhood groups for state and local traffic engineering officials to implement actions to reduce "excessive" speeding on streets and highways. Frequently, these requests are for lower speed limits and/or the installation of traffic control devices, such as STOP signs. State and local transportation agencies simply do not have the resources to conduct the large numbers of engineering studies required to respond to all of these requests in a timely fashion.

Once the traffic engineer has collected the speed data needed to estimate the 85th percentile speed on a given roadway, the engineer must then rely upon professional judgment to determine if any other factors (e.g., accident experience, roadway geometry, adjacent development) warrant the establishment of a speed limit that is different from the observed 85th percentile speed. At the present time, the individual and cumulative quantitative effects of these "other factors" on roadway speeds have not been clearly defined.

The increasing pressure from the public to implement lower speed limits and/or other regulatory measures as a means of controlling speed and improving roadway safety reflects another problem, namely, some widely held public misconceptions concerning speed zoning. For example, the average citizen has the misconception that reducing the speed limit will slow the speed of traffic and increase roadway safety. Research indicates that a reasonable and prudent driver will drive the speed suggested by roadway and traffic conditions rather than relying on posted speed limits. Because accidents appear to depend less on absolute speed and more on the variation of speeds in the traffic stream, setting unrealistically low speed limits can actually lead to an increase in accidents. For example, a study conducted by the Florida Department of Transportation [1980] concluded that:

> "Motorists tend to pay little attention to speed regulations which they consider unreasonable unless there is an inordinate degree of enforcement. Unreasonable, low speed limits are commonly violated by a majority of motorists, making enforcement difficult, with resulting operating speeds somewhat higher than would exist with proper, realistic speed limits."

## 1.2 SIGNIFICANCE OF THE RESEARCH

The primary goal of this research effort was to quantify the effects that selected roadway characteristics and adjacent development patterns have on roadway speeds. Achievement of this goal could provide a cost-effective alternative to conducting individual, site-specific speed studies to determine and/or justify speed limits. Quantification of these factors could also improve the traffic engineer's ability to: 1) judge which, if any, of those factors warrant consideration in setting speed limits different from the 85th percentile speeds and 2) assess the magnitude of any such adjustments that may be required. Finally, this report contains a review of findings from previous research in the area of speed zoning, which, when used in conjunction with the findings of the present research effort, should be useful to traffic engineers in dispelling some of the widely held popular misconceptions concerning speed zoning.

## 1.3 SCOPE AND LIMITATIONS

This study focuses on identifying and quantifying the characteristics of the roadway and adjacent development patterns that affect roadway operating speeds. The results are based on data from 539 roadway sections on urban and rural state highways in Kansas. Though the study sites are representative of state highways in Kansas, the transferability of the results is not known.

## 1.4 ORGANIZATION OF THE STUDY

The research described in this report was conducted in several phases. In addition to this introductory chapter, this report consists of the following four chapters corresponding to the four basic phases of the research.

**Chapter 2: Literature Review**. Chapter 2 presents a review of the literature concerning speed zoning. Topics reviewed include definitions of the various types and measures of speed used by traffic engineers and roadway designers, the importance of reasonable speed limits, current procedures for establishing reasonable speed limits, evaluating the effectiveness of speed zones, and factors affecting roadway speeds.

**Chapter 3: Study Method**. Chapter 3 presents the details of the study design in terms of the problem investigated, study objectives, and the data collection and analysis procedures used to accomplish the objectives of the study.

**Chapter 4: Data Analysis**. The data analysis and model development phases of the research are documented in Chapter 4.

**Chapter 5: Conclusions, Recommendations and Implementation Plan**. The fifth and final chapter of the study contains a summary of research findings, puts forward the recommendations drawn from the research and outlines a plan for implementing the study findings.

*CHAPTER 2*

# LITERATURE REVIEW

---

## 2.1 INTRODUCTION

A speed zone is a section of street or highway where a speed limit different from the statutory speed limit has been established [Institute of Transportation Engineers (ITE), 1993]. The purpose of speed zoning, as stated in the Uniform Vehicle Code [National Committee on Uniform Traffic Laws and Ordinances, 1968], is to establish a speed limit that is "reasonable and safe for a given section of a roadway." This chapter presents a review of the literature pertaining to speed zoning. Topics reviewed include definitions of the various types and measures of speed used by traffic engineers and roadway designers, the importance of reasonable speed limits, current procedures for establishing reasonable speed limits, evaluating the effectiveness of speed zones, and factors affecting roadway speeds.

## 2.2 DEFINITIONS OF SPEED

Any discussion of speed limits and speed zoning must be based on a clear understanding of several key, speed-related terms. The following sections of this chapter provide an overview of the definitions and interpretations of the various types of speed that are considered by transportation engineers in establishing speed limits, as well as in roadway design. This information should be useful to transportation engineers in explaining speed-related concepts such as design speed, operating speed, 85th percentile speed, posted speed, and other measures of speed.

### 2.2.1 Design Speed

The American Association of State Highway and Transportation Officials (AASHTO) defines design speed as "the maximum safe speed that can be maintained over a specified section of highway when conditions are so favorable that the design features of the highway govern" [AASHTO, 1994]. According to AASHTO [1994]:

> "The speed selected for design should fit the travel desires and habits of nearly all drivers. In other words, the design should be nearly all inclusive of the typically desired speeds of drivers, where this is feasible."

In selecting an appropriate design speed for a section of highway, it is important to understand two qualifiers in the design speed policy cited above. First, it would not be practical to attempt to accommodate the desired speed of all drivers. Therefore, the qualification "nearly all drivers" is specified in the policy. Second, the qualification "where this is feasible" recognizes that many factors in addition to speed influence highway design. In some situations, terrain, adjacent development and other factors may make a design that would accommodate the desires of nearly all drivers either too costly and/or environmentally unacceptable [Krammes et al., 1996].

Once a design speed is selected, it influences several other important design decisions, such as the sharpness of horizontal and vertical curves, lane and shoulder widths, roadside clearances and drainage structures. The criteria for what are safe at a given design speed are based on comfort factors and near-worst case conditions (e.g., the performance of 1940s vehicles and locked-wheel braking on wet pavements) [Krammes et al., 1996]. These criteria include considerable margins of safety and may actually apply only to a small number of critical design features. As a result, the design speed of a highway is likely to underestimate the maximum safe speed along most of that highway [Krammes et al., 1996].

### 2.2.2 Operating Speed

The current AASHTO definition of operating speed is "the highest overall speed at which a driver can travel on a given highway under favorable weather conditions and under prevailing traffic conditions without at any time exceeding the safe speed as determined by the design speed on a section-by-section basis". This definition has little practical meaning. As a result, it is rarely used in practice [Krammes et al., 1996].

In current practice, operating speed on a roadway section is the speed at which drivers are observed operating their vehicles on that section. The 85th percentile speed is the most frequently used descriptive statistic for the operating speed associated with a particular location or geometric feature [Krammes et al., 1996].

### 2.2.3 85th Percentile Speed

The 85th percentile speed is the speed at or below which 85 percent of drivers are operating their vehicles. Researchers generally perform spot speed studies to obtain reliable estimates of 85th percentile speeds. In these studies, a speed measurement location is identified on a highway, and speeds are measured for an adequate sample of free-flowing vehicles (typically 100 - 125 vehicles).

### 2.2.4 Posted Speed

Posted speed refers to the maximum speed limit posted on a section of highway. Although speed zoning guidelines permit consideration of other factors (including roadside development, road and shoulder surface characteristics, and pedestrian and bicycle activity), basing posted speeds strictly upon measured 85th percentile speeds has been standard practice in this country for many years [Krammes et al., 1996]. The posted speed is generally obtained by rounding the 85th percentile speed to the nearest 5-mph increment. Using the 85th percentile speed in selecting posted speeds is based on the belief that the large majority of drivers are capable of judging appropriate speeds based upon roadway geometry, roadside development, etc., and that they will operate at speeds that are reasonable and prudent. Basing posted speeds on the 85th percentile speed also promotes uniformity among speeds at a given location. The benefit of a uniform speed is that vehicle collisions are less likely to occur if all drivers are traveling at about the same speed [Krammes et al., 1996].

There are two common exceptions to the 85th percentile speed procedure for establishing speed limits: 1) on sections of roadways with high accident experience, the posted speed may be as much as 7 mph lower than the 85th percentile speed and, 2) speed limits cannot be posted in excess of legislatively mandated maximum speed limits [Krammes et al., 1996].

The rationale for using the 85th percentile speed as the basis for establishing speed limits is discussed in greater detail in subsequent sections of this chapter.

### 2.2.5 Other Measures of Speed

Other measures of speed commonly used by traffic engineers are pace speed, space mean speed, time mean speed, and free flow speed. *Pace speed* is defined as the 5 or 10 mph band of travel speeds containing the largest number of observed vehicles [ITE, 1989]. *Time mean speed* is the average of spot speeds (e.g., speeds determined by radar) on a highway during an interval of time. *Space mean speed* is the harmonic mean of the speeds of vehicles passing a point on a highway during an interval of time. It is obtained by dividing the total distance traveled by two or more vehicles on a section of highway by the total time required by these vehicles to travel that distance. Time mean and space mean speeds can be calculated from the following equations:

$$\mu_t = (1/n)\sum\mu_i \qquad (1)$$

$$\mu_s = n/\sum(1/\mu_i = (nL)/\sum t_i \qquad (2)$$

where $\mu_t$ = time mean speed (ft/sec), $\mu_s$ = space mean speed (ft/sec), n = number of vehicles, $\mu_i$ = speed of the ith vehicle (ft/sec), $t_i$ = the time it takes the ith vehicle to travel across a section of highway (sec) and L = length of section of highway (ft).

In all instances (except the case of uniform speed), the space mean speed is lower than the time mean speed. The approximate relationship between these two mean speeds is:

$$\mu_t = \mu_s + \sigma^2_s / \mu_s \qquad (3)$$

where $\sigma^2_s$ = the variance of the space mean speed distribution.

*Free-flow speed* is defined as the theoretical speed of traffic when density is zero (that is, when no other vehicles are present) [Highway Capacity Manual (HCM), 1994].

## 2.3 IMPORTANCE OF REALISTIC SPEED LIMITS

Realistic speed limits are important for a variety of reasons:

1) They invite public compliance by conforming to the behavior of the majority.
2) They give a clear reminder of reasonable and prudent speeds to non-conforming drivers.
3) They offer an effective enforcement tool to the police.
4) They tend to minimize public antagonism toward police enforcement which results from obviously unreasonable regulations.
5) They encourage drivers to travel at the speed where the risk of accident involvement is the lowest.

Despite considerable evidence that supports the importance of realistic speed limits in achieving the basic principles cited above, the general public still clings to the following misconceptions concerning speed limits:

1) Reducing the speed limit will slow the speed of traffic.
2) Reducing speed limits will decrease the number of accidents and increase safety.
3) Raising the posted speed limit will cause an increase in the speed of traffic.
4) Any posted speed limit must be safer than an unposted speed limit, regardless of the prevailing traffic and roadway conditions.

5) Drivers will always drive 5 mph over the posted speed limit.

Contrary to popular belief, speed by itself is not a major cause of accidents. Accidents appear to depend less on absolute speed and more on the variation of speeds in the traffic stream. The following sections of this chapter provide a summary of previous research concerning the relationships between speed and accidents, speed and accident type, and the effectiveness of speed zoning in reducing accidents and speeds.

### 2.3.1 Speed and Safety

The relationship between speed and safety is characterized by the physics of vehicle motion. Transportation Research Board (TRB) Special Report 204 [TRB, 1984] identified the following four reasons why reduced speeds are likely to yield safer driving:

1) When traveling at a higher speed, a car moves a greater distance during the fixed period of time that it takes for the driver to react to a perceived problem.
2) On highways lacking adequate superelevation, a driver's ability to steer safely around curves diminishes with speed.
3) The distance required to stop a vehicle by braking increases with speed.
4) Crash severity increases disproportionately with speed at impact.

Another factor in the speed-safety relationship is that concerning variation in speeds on the same highway segment. As reported by Solomon [1964] and Cirillo [1968] a wider variability in speeds increases the frequency of motorists passing one another which, in turn, increases the opportunities for multi-vehicle accidents to occur.

In discussing this topic, the authors of TRB Special Report 204 [TRB, 1984] add:

> "speed variability contributes to the front-to-rear accidents prevalent on interstate highways. These accidents are most common near intersections as motorists who slow down and change lanes to exit mix with motorists traveling at much higher

speeds."

West and Dunn [1971] studied the relationship between speed and accidents on rural two lane highway segments. Their study was able to identify the exact speed of travel for a large proportion of vehicles involved in accidents on the study segments. A finding of West and Dunn [1971] was that slow drivers had higher accident involvement rates than fast drivers. It was also reported that when accidents involving turning vehicles were excluded, the involvement rates for both high and low speed drivers were very close and approximately six times higher than those of drivers traveling close to the mean traffic speed.

The speed at which it is safest for drivers to travel is within approximately one standard deviation of the mean travel speed of the traffic stream. Low speed and high speed-drivers are the most likely to become involved in accidents. The higher the speed of vehicles involved in accidents the greater is the likelihood of injury [West and Dunn, 1971].

### 2.3.2 Speed and Accident Types

The relationship between speed and accidents by accident type has been documented by Solomon [1964]. In his study, data were collected at 35 rural locations in eight different states. According to Solomon, at low speeds the predominant types of accidents are rear-end and angle accidents. Solomon also observed that the overall pattern for rear-end involvements at low speeds declines as speed increases. Rear-end accidents as a percentage of total accidents range between 40 percent and 50 percent at speeds between 20 mph and 50 mph. From 50 mph to 70 mph rear-end accidents decline as a percentage of total accidents to slightly above 20 percent [Coleman, 1995].

The substantially higher rear-end involvements at low to moderate speeds are explained by Solomon by comparing the speed difference between pairs of vehicles in normal traffic with that of the two colliding vehicles. Solomon [1964] states:

> "In summary, passenger car drivers involved in rear-end collisions were more likely to have been traveling at a speed difference much greater than that for pairs of vehicles in normal traffic."

Beatty [1972], studying similar urban and rural locations, found that 84 percent of two-vehicle accidents were reported as rear-end collisions or same-direction sideswipe. Beatty [1972] found that 40 percent of the two-vehicle accidents were in urban areas and only 27 percent of single-vehicle accidents occurred in these same areas. Beatty also confirmed Solomon's findings of speed difference and accident involvement. Beatty [1972] found:

> "the speed difference for two vehicles involved in an accident is, on the average 11.4 mph greater than for two randomly selected vehicles. The magnitude of the excess is almost twice the standard error even for single observations".

### 2.3.3 Evaluation of Speed Zone Effectiveness

A survey of highway officials on speed zoning practices was conducted in 1984 by the AASHTO Task Force on Speed Zoning and Control. The survey results, along with a review of the literature have been reported by Parker [1985]. The majority of the survey respondents (82 percent) identified the highest priority research area as the need to "determine the effects of altering speed limits on speed and accidents."

Kessler [1959], in an evaluation of 30 locations where the original speed limits were lower than the 85th percentile speed, found that when the speed limits were raised to the 85th percentile, the number of accidents decreased from 62 to 40. This suggests that raising the speed limit to reflect the 85th percentile speed reduced speed differentials and thereby reduced accident frequency. Avery [1960] examined speed compliance at 18 locations in St. Paul, Minnesota where the original speed limits were based on 85th percentile speeds. When speed limits were raised from 30 to 35 or 40 mph, no change in mean speed or 85th percentile speed was observed.

Dudek and Ullman [1987] examined six sites in rapidly developing urban fringe areas where speed limits previously posted at 55 mph (based on 85th percentile speed) were reduced to 45 mph. They found no significant changes in speeds, speed distribution, or accident rates.

Parker [1992] studied 99 experimental and comparison sites in 22 states where posted speed limits were based on 85th percentile speeds. In this study, the author found that where speed limits were raised, accidents were reduced by 6.7 percent after implementation. At sites where speed limits were lowered, accidents increased by 5.4 percent. In addition, Parker notes that "Lowering speed limits below the 50th percentile does not reduce accidents, but does significantly increase driver violations of the speed limit. Conversely, raising the posted speed limits did not increase speeds or accidents." Parker [1992] also found little change in the speed distribution as a result of raising or lowering the speed limits on urban and rural non-limited access highways.

McCoy et al. [1993] studied speed zoning by comparing the accident experience in speed zones with reasonable speed limits with the accident experience in speed zones with unreasonable speed limits. Reasonable speed limits were consistent with the speed limits determined by the Nebraska Department of Roads (NDOR) method (based on 85th percentile speeds) and unreasonable speed limits were lower than those determined by the NDOR method. They found that road user costs are minimized when the posted speed limits in speed zones are set equal to the reasonable speed limits. Speed zones with posted speed limits 5 and 10 mph lower than the reasonable speed limits have higher accident and travel time costs than zones with reasonable speeds.

## 2.4 DETERMINING REASONABLE SPEED LIMITS

During the period from 1950-1970, traffic engineers (often in conjunction with police enforcement agencies) gradually refined techniques to determine a safe speed limit. Accident frequency, severity, and accident type are the measures by which the safety of a roadway is determined. Accident statistics often serve as the variable of interest when a change is applied to a roadway (such as speed limits, signalization, or geometrics) [Coleman, 1995].

State and local transportation officials have practiced two similar procedures in determining a speed limit in a speed zone. The primary basis for both is the 85th percentile speed obtained in a spot speed sample. In the first approach, speed limits are based solely on the 85th percentile speed. In the second approach, "engineering judgement" is used to account for other existing roadway characteristics and area development factors to reduce the limit below the 85th percentile speed if deemed necessary. However, a small number of states and municipalities have practiced a policy which replaces engineering judgement with some type of quantification of roadway characteristics and area development variables to reduce the speed limit below the 85th percentile speed [Coleman, 1995]. The determination of which of these two policies of speed zoning leads to a safer driving condition for motorists is the subject of considerable debate within the traffic engineering profession [Coleman, 1995].

A review of current practices in establishing realistic speed limits is presented in the following sections of this chapter.

### 2.4.1 Current Practices

A speed considered reasonable by drivers, residents, legislators and enforcement officers is the appropriate speed limit to post in a speed zone under favorable weather conditions and under prevailing traffic conditions [AASHTO, 1994]. Speed zones should only be established on the basis of an engineering study [ITE, 1993]. The engineering study needs to consider roadside development, road and shoulder characteristics, pedestrian and bicycle activity, speed limits on adjoining road segments and accident experience or potential. Since the roadway as well as the speed limit might be changed, roadway speeds need to be restudied at a maximum interval of five years [ITE, 1993]. In addition, an engineering study needs to be conducted whenever there is a change in the roadway that would affect the prevailing speed. Such changes include elimination of parking, added lanes, signal coordination, changes in roadside development, traffic volume, turning movements and controls, and the number of commercial vehicles in the traffic stream. According to ITE [1993] Technical Council Committee 4M-25:

"The speed limit within a speed zone should be set at the nearest 5-mph increment to the 85th percentile speed or the upper limit of the 10 mph pace. In no case should the speed limit be set below the median speed or the 10 mph pace. No speed zone should be established in a location where the 85th percentile speed limit is within ±3 mph of the statutory speed limit."

Another approach used to set speed limits is the average test run speed method. Test-car speed runs are made by "driving as fast as it is comfortably safe" during off-peak periods when the test car will not be delayed by other traffic [McCoy et al., 1993]. McCoy et al. [1993] report that five states use the average test run speed to compute the prevailing speed, which is defined as the average of the 85th percentile speed, the upper limit of the 10-mph pace, and the average test run speed [McCoy et al., 1993]. The test-car speed method has been used by the Nebraska Department of Roads (NDOR) as a means to coordinate their speed zoning procedures. Together with local officials, NDOR engineers make a series of test runs over the section of roadway under consideration at speeds ranging from 30 to 50 mph and the participants indicate the level of comfort they experience at each speed [McCoy et al., 1993]. McCoy et al. [1993] note that there is no generally accepted procedure for conducting test-car speed runs as part of speed zoning studies. In addition, the results of test-car speed runs are usually highly variable unless several runs are made. Consequently, the establishment of speed zones based on the average test run speed are less consistent than those based solely on the 85th percentile speed and/or the 10-mph pace.

In 1985, Parker conducted a study of speed zoning practices for the U.S. Department of Transportation (USDOT). The purpose of the study was to review the principles and practices used to set speed limits. The primary factor used to set speed limits was the 85th percentile speed. The 85th percentile speed was used by all of the states and 86 percent of the local agencies surveyed. According to the USDOT study, most traffic officials generally agree that speed limits should reflect the speed of most drivers. The most commonly reported lower speed limit was 5 mph below the 85th percentile, with 10 mph below the 85th percentile speed representing the extreme [Parker, 1985]. The primary factors used in setting speed limits, as identified in the

USDOT study, are shown in Table 2-1.

**Table 2-1. Primary Factors Considered in Setting Speed Limits.**

| Factor | Agencies Reporting that Factor is Considered (Percent) | |
|---|---|---|
| | States | Locals |
| 85th percentile speed | 100 | 86 |
| Roadside development | 85 | 77 |
| Accident experience | 79 | 81 |
| 10 mph pace | 67 | 34 |
| Roadway geometrics | 67 | 57 |
| Average test run speed | 52 | 34 |
| Pedestrian volumes | 40 | 50 |

Source: Parker, 1985.

According to the USDOT study, the speed should be set at the speed driven by 85 to 90 percent of the free-flowing vehicles rounded up to the next 5-mph increment. This method results in speed limits that are not only acceptable to a large majority of motorists, but also fall within the speed range (85 to 90 percent of free moving vehicles) where the accident risk is lowest [Parker, 1985].

In 1989, ITE formed a Technical Committee on Speed Zoning Guidelines. The charge of the Committee was to address two questions: (1) criteria for speed zones, and (2) speed limits for speed zones. The Committee attempted to answer these questions based on: (1) a survey of professional groups involved in the establishment of speed zones, (2) a review of the literature, and (3) the expertise of the committee members.

The Committee's final report stressed the need for consistency in establishing speed zones and outlined a recommended practice to provide a consistent basis for the application of engineering principles to speed zoning. The ITE Committee recommends that speed zoning be based on an analysis of the current speed distribution of free-flowing vehicles and that the speed limit be set at the nearest 5-mph increment of the 85th percentile speed or the upper limit of the 10-mph pace [ITE, 1993]. The ITE Committee also recommends that the engineering study may consider other factors such as: (1) geometric features such as vertical and horizontal alignment, and sight distance, (2) roadside development, (3) road and shoulder surface characteristics, (4) pedestrian and bicycle activity, (5) speed limits on adjoining highway segments, and (6) accident experience or potential.

## 2.5 FACTORS AFFECTING ROADWAY OPERATING SPEEDS

This section presents a review of the literature concerning the roadway and area development factors that might have some effect on the speed of vehicles. The literature is discussed under the following areas: (1) effects of roadway geometry and roadside development on speeds, and (2) effects of traffic characteristics and traffic control devices on speeds.

### 2.5.1 Effects of Roadway Geometry and Roadside Development on Speeds

Geometric design decisions have a significant influence on the operating speed of vehicles on low-speed urban streets. The roadway geometry is determined by design speed but design speed is not always consistent with desired operating speed. Large curve radii and long tangent sections, for example, encourage speeds that may be higher than desirable in certain situations. Therefore, careful selection of design elements is essential in achieving desired operating speeds.

Poe and Mason [1995] reported that the following factors can affect the 85th percentile speed of vehicles: roadway characteristics (radius, grade and lane width), characteristics of the roadside (number of access points, lateral obstructions, sidewalks), median type and width, shoulder width, lateral clearances, horizontal and vertical clearance, and land use.

Poe and Mason [1995] reported that on low-speed (less than 40 mph) urban streets, degree of curvature, hazard rating, and grade were significant at the 95 percent confidence level in explaining speeds. The number of driveways, number of intersections, and lane width were found to be less significant. Equation 4 is the model developed by Poe and Mason [1995] to estimate the 85th percentile speed ($V_{85}$). As shown in Equation 4, increasing degree of curvature, grade, hazard rating, number of intersections, and number of driveways tend to lower the 85th percentile speed, while increasing lane width tends to increase the 85th percentile speed.

$$V_{85} = 61.7 - 0.23(DC) - 0.52(G) - 0.82(HR) - 2.66(IN) - 1.08(DR) + 0.15(LW) \quad .... (4)$$

$(R^2 = 0.67)$

where $V_{85}$ = 85th percentile speed, DC = degree of curvature, G = grade, HR = hazard rating (a measure of the number and severity of lateral obstructions within 1.5 meters of the roadway), IN = number of intersections, DR = number of driveways and LW = lane width.

According to the Highway Capacity Manual (HCM), lane and shoulder widths have significant impacts on speeds [HCM, 1994]. Narrow lanes cause vehicles to travel closer to each other (laterally) than most drivers would prefer. Motorists compensate by slowing down or observing larger longitudinal spacing for a given speed, which effectively reduces speeds. Lane widths less than 12 ft reduce travel speeds, but lane widths more than 12 ft are not considered to increase speed above the ideal situation [HCM, 1994].

Narrow shoulders and lateral obstructions are two important factors that affect the speed of vehicles. Many drivers will steer away from roadside or median objects they perceive to pose a hazard. This action brings them laterally closer to vehicles in adjacent lanes and causes the same reactions as those exhibited in narrow lanes [HCM, 1994].

Roadway grade, horizontal alignment, and traffic control devices have a significant effect on the operations of heavy vehicles [HCM, 1994]. The effects of positive grades are particularly significant at intersections, where vehicles must overcome both the grade and the inertia of starting from a stopped position at the same time.

Jarvis and Hoban [1989] developed an expert system advisor to assist in establishing reasonable speed limits. To develop guidelines for establishing speed zones Jarvis and Hoban [1989] suggest the following factors need to be considered: parking conditions, numbers of businesses with access, roadside developments, pedestrian and vehicle activity, land use and building setbacks. Other special roadside activities that effect speeds are school zones, playgrounds, frequent parking and unparking movements, and substantial crossing and turning traffic.

The Highway Capacity Manual [1994] notes that for every 10 access points per mile that affect a given direction of travel on a highway, travel speed may be reduced by 2.5 mph [HCM, 1994].

### 2.5.2 Effects of Traffic Characteristics and Control on Speeds

Traffic characteristics, conditions, and controls should be considered when establishing speed zones. The factors most commonly considered include [ITE, 1993]: traffic volumes, turning movements and controls, commercial vehicles/traffic, parking conditions, traffic control devices and extent and frequency of vehicle-pedestrian conflicts.

Some of the more commonly used traffic control devices and measures and their effects on volume, speed, and safety are shown in Table 2-2. As shown in Table 2-2, speed limits tend to improve traffic safety; however, speed limits have negligible effects on speeds and traffic volume.

*Table 2-2.* **Traffic Control Measures and their Effects on Volume, Speed and Safety.**

| Type of Device | Effects | | |
|---|---|---|---|
| | Volume | Speed | Safety |
| Speed Limit | little or no | little or no | improve |
| Access Regulation | reduce | unknown | unknown |
| Truck Restriction | unknown | unknown | unknown |
| Parking Control | increase | unknown | unknown |
| Turn Prohibition | reduce | increase | increase |
| Median Barrier | reduce | reduce | improve |
| Traffic Signal | reduce | reduce | improve |
| Stop Sign | reduce | reduce | mixed |
| Yield Sign | Unknown | reduce | improve |
| Pavement Undulation | little to reduce | reduce | mixed |

Source: ITE, 1993.

## 2.6 SUMMARY

This chapter has presented a review of the literature pertaining to speed zoning. Topics reviewed include definitions of the various types and measures of speed used by traffic engineers and roadway designers, the importance of reasonable speed limits, current procedures for establishing reasonable speed limits, evaluating the effectiveness of speed zones, and factors affecting roadway speeds.

Most state and local agencies use the 85th percentile speed as the basis for setting speed limits. Setting the speed limit at the 85th percentile speed appears to be the safest because it reduces speed differentials (the real cause of accidents). When the speed limit is below the 85th percentile, a few drivers will obey the posted speed. Most, however, will drive at a speed comfortable to them (i.e., ignore the posted speed and drive at the faster 85th percentile speed).

Due to the speed differential, an increase in accidents may result.

The literature review indicated that the following factors can be important in explaining speed. However, it appears that research directed at quantifying the effects that these factors have on speed has been extremely limited.

1. Roadway characteristics:
    - length of existing speed zones,
    - intersection spacing and geometry,
    - roadway surface condition,
    - presence and condition of shoulders,
    - presence and width of median,
    - number and width of lanes,
    - degree of curvature,
    - grade,
    - superelevation, and
    - design speed.

2. Roadside characteristics:
    - number of roadside businesses with access,
    - building setback and location,
    - sidewalk width and location,
    - adjacent land use,
    - density of adjacent development,
    - number of driveways,
    - lateral obstructions.

3. Traffic characteristics and control:
    - traffic volumes,

- turning movements and controls,
- commercial vehicles,
- parking conditions,
- on-street parking,
- traffic signals,
- road signs and their understandability,
- the extent and frequency of vehicle-pedestrian conflicts,
- pedestrian and bicycle activity,
- speed limits on adjoining highway segments, and
- posted speed limits.

4. Driver perception of roadway safety.
5. Level of enforcement.
6. Driver experience and prior knowledge of the road and roadside development.

*CHAPTER 3*

# STUDY METHOD

---

## 3.1 PROBLEM STATEMENT

Speed zoning in the United States is based on the principle of setting speed limits as near as practicable to the speed at or below which 85 percent of drivers are traveling. Use of the $85^{th}$ percentile speed as the basis for setting speed limits is subject to revision based upon such factors as accident experience, roadway geometry, parking, pedestrians, adjacent development and engineering judgment. There is a need to identify and quantify the individual and cumulative effects that these factors have on roadway operating speeds.

## 3.2 STUDY OBJECTIVES

The basic objective of this research effort was to quantify the effects that selected roadway characteristics and adjacent development patterns have on operating speeds on rural and urban state highways. In this study, operational speed is defined as the 85th percentile speed. Achievement of this objective could provide a cost-effective alternative to conducting individual, site-specific speed studies to determine and/or justify speed limits. Quantification of these factors could also improve the traffic engineer's ability to: 1) judge which, if any, of those factors warrant consideration in setting speed limits that are different from the 85th percentile speeds, and 2) assess the magnitude of any such adjustments that may be required. In addition, the results of this study should be useful to traffic engineers in dispelling some of the widely held popular misconceptions concerning speed zoning.

## 3.3 STUDY DESIGN

The study design addressed the following basic elements of the research: 1) identification of factors to be evaluated, 2) data collection and database development, and 3) data analysis, and model development and testing. The key components of these elements are outlined below.

### 3.3.1 Selection of Factors to be Evaluated

The selection of candidate factors to be evaluated was based on two considerations: 1) the results of the literature review, and 2) the availability of data from the Kansas Department of Transportation (KDOT) data files on roadway sections containing these factors and known roadway speeds. Based on the results of the literature review and data available from KDOT, the factors shown in Table 3-1 were selected as potentially important variables in terms of explaining the variability in 85th percentile speeds observed on rural and urban roadways on the Kansas state highway system.

### 3.3.2 Data Collection and Database Development

The roadway data needed for the analyses were extracted from the Control Section Analysis System (CANSYS) database maintained by KDOT [KDOT, 1994]. The control section is a unit of basic reporting, identification, and analysis. It is defined as a segment of roadway with reasonably uniform geometric, traffic, surface and base characteristics for its entire length. On divided facilities, each lane is considered to be a separate control section. All control sections are identified by a three-digit number plus a county identification number. When used together, these two numbers form a unique code for each individual control section in the state.

Tables 3-2 and 3-3 list the CANSYS database variables that correspond to the potential explanatory variables described in the previous section of this chapter (see Table 3-1). Table 3-2 lists potential explanatory variables for rural highways and Table 3-3 lists potential explanatory variables for urban highways. Note that the CANSYS variables include both continuous and categorical variables. The CANSYS database variables are defined in detail in the Appendix.

*Table 3-1.* **Factors Possibly Affecting Speeds on State Highways.**

| Rural Highways | Urban Highways |
|---|---|
| Lane Width | Lane Width |
| Surface Width | Surface Width |
| Shoulder Width | Shoulder Width |
| Surface Condition (smoothness - IRI[a]) | Surface Condition (smoothness - IRI) |
| Traffic Volume (ADT[b])<br>% of Trucks/Heavy Vehicles | Traffic Volume (ADT)<br>% of Trucks/Heavy Vehicles |
| Sight Distance | Sight Distance |
| Length of the Speed Zone | Length of the Speed Zone |
| Median Width | Median Width |
| Accidents | Accidents |
| Percent No Passing | |
| Access Control | Access Control |
| Presence of Shoulder:<br>Type of Shoulder (Paved, Gravel, mixed, etc.)<br>Severity of Road Ditch - Steepness of Side Slope, Depth, Presence of Trees, Rocks, etc. | Presence of Shoulder:<br>Type of Shoulder (Paved, Gravel, mixed etc.)<br>Severity of Road Ditch - Steepness of Side Slope, Depth, Presence of Trees, Rocks, etc. |
| Rideability | Rideability |
| Surface Type | Surface Type |
| Presence of Median Barrier (jersey barrier) | Presence of Median |
| Population Density | Population Density |
| Location (residential, commercial, business, etc.) | Location (residential, commercial, business etc.) |
| Land Use | Land Use |
| Street Classification: Link, Arterial, Collector, etc. | Street Classification: Link, Arterial, Collector, etc. |
| Lane Class | Lane Class |
| Presence of On-street Parking | Presence of On-street Parking |
| | Number of Grade Changes with Stop Signs |
| | Number of Grade Separated Interchanges |
| Side-street Frequency (per mile) | Side-street Frequency (per mile) |

[a] International Roughness Index.
[b] Average Daily Traffic.

*Table 3-2.* **Factors Possibly Affecting Speeds on Rural Highways.**

| Factors | CANSYS Data Name |
|---|---|
| Continuous Factors | |
| Width of Lanes | Lane Width |
| Surface Width | Surface Width |
| Width of Shoulder | Shoulder Width (right and left) |
| Surface Condition (smoothness) | International Roughness Index (IRI) |
| Traffic Volume (ADT)<br>% of Trucks/Heavy Vehicles | AADT[a], Current<br>Heavy Commercial Traffic, Current |
| Sight Distance | Substandard Stopping Sight Distance |
| Length of the Speed Zone | Subsection Length |
| Median Width | Average Median Width |
| Accidents | Number of Accidents (5-yr. total and current) |
| Percent No Passing | Restricted Passing |
| Categorical Factors | |
| Access Control | Access Control |
| Presence of Shoulder:<br>Type of Shoulder (Paved, Combination, Gravel, or Turf)<br>Severity of Road Ditch - Steepness of Side Slope, Depth, Presence of Trees, Rocks, etc. | Shoulder Type (right and left)<br>Side Slope |
| Surface Condition (Smoothness) | Rideability |
| Surface Type | Surface Type |
| Presence of Median Barrier (jersey barrier) | Median Type |
| Population Density | Population Density (right and left) |
| Location | Urban Location |
| Land Use | Land Use (right and left) |
| Street Classification | Functional Classification |
| Lane Class | Lane Class |
| Parking Type | Parking (right and left) |

[a] Annual Average Daily Traffic.

*Table 3-3.* **Factors Possibly Affecting Speeds on Urban Highways.**

| Factors | CANSYS Data Name |
|---|---|
| Continuous Factors | |
| Length of Speed Zone | Subsection Length |
| Width of Shoulder | Shoulder Width (right and left) |
| Surface Width | Surface Width |
| Surface Condition | International Roughness Index (IRI) |
| Traffic Volumes (ADT)<br>% Trucks/Heavy Vehicles | AADT, Current<br>Heavy Commercial Traffic, Current |
| Side-street Frequency (per mile) | Number of Intersections (signalized, other, etc.) |
| Sight Distance | Substandard Stopping Sight Distance |
| Accidents | Number of Accidents (5-yr. total and current) |
| Grade-changes with Stop Signs | Number of Interchanges with Stop Signs |
| Grade-separated Interchanges | Number of Grade-separated Interchanges |
| Width of Median | Average Median Width |
| Width of Lanes | Lane Width (calculated value) |
| Categorical Factors | |
| Access Control | Access Control |
| Street Classification (Connecting Link, Arterial, etc.) | Functional Classification |
| Land Use | Land Use (right and left) |
| Lane Class | Lane Class |
| Presence of Curb and Gutter/Shoulders | Shoulder Type (right and left) |
| Presence of On-street Parking<br>Type - Parallel, Diagonal or Other | Parking (right and left) |
| Surface Type | Surface Type |
| Surface Condition (smoothness) | Rideability |
| Presence of Median<br>Frequency of Median Openings<br>Presence of Turning Lanes in Median | Median Type |
| Population Density | Population Density (right and left) |
| Side Slope | Side Slope |
| Location | Urban Location |

The 85th percentile speed data needed in the study were manually extracted from KDOT speed study files and integrated with the CANSYS database variables as explained below. The KDOT files contain speed study data collected on 660 roadway sections for the period 1991 - 1994. Table 3-4 provides an example of the data collection form used to summarize the speed study files.

***Table 3-4.* Data Collection Form Used to Summarize KDOT Speed Study Data.**

| City | County Code | Location | Speed | | | | Control Section Number |
|---|---|---|---|---|---|---|---|
| | | | Limit (Posted) | Average Speed | Standard Deviation | 85th Percentile | |
| | | | | | | | |
| | | | | | | | |

Because data in the CANSYS database are referenced by control section number, it was necessary to identify the control section numbers for the speed study sites. This was accomplished by examining the KDOT 1995 "accident chain file." The "accident chain file" contains a listing of traffic accidents and roadway data by city, county, highway route number, and control section number. By manually searching the accident chain file, it was possible to associate a control section number with the speed study sites which had experienced at least one accident in 1995. Of course, those speed study sites which experienced no accidents in 1995, did not appear in the accident chain file. Using this procedure, it was possible to identify control section numbers for 539 of the 660 speed study sites in the KDOT data files. Once the control section numbers for the speed study sites were identified and recorded on the speed data sheet (Table 3-4), it was possible to merge the appropriate CANSYS roadway data into the file.

#### *3.3.2.1 Summary of the Database*

The final database consisted of 539 highway sections with speed data, roadway data and adjacent development characteristics. These 539 sites consisted of 186 rural highway sections and 353 urban highway sections. Table 3-5 provides a summary of the range of values for the continuous variables in the database.

***Table 3-5.* Range of Values for Continuous Variables.**

| Factors | Rural | Urban |
|---|---|---|
| Subsection Length (mile) | 0.003 - 6.117 | 0.006 - 1.068 |
| 85th Percentile Speed (mph) | 37.4 - 77 | 24.5 - 68 |
| Surface Width (ft) | 20 - 54 | 12 - 82 |
| Lane Width (ft) | 9 -12 | 9 - 12 |
| Shoulder Width, Inside (ft) | 0 - 10 | 0 - 10 |
| Shoulder width, Outside (ft) | 0 - 10 | 0 - 10 |
| IRI Current (inches/mile) | 0 - 182 | 0 - 308 |
| AADT | 290 - 13,500 | 550 - 66,055 |
| % Heavy Commercial Traffic (percent of total traffic) | 3 - 80 | 1 - 46 |
| Number of Locations with Substandard Stopping Sight Distance | 0 - 20 | 0 - 20 |
| Number of Accidents, 5-year Total | 0 - 90 | 0 - 805 |
| Number of Accidents, Current | 0 - 21 | 0 - 186 |
| Restricted Passing (percent of total section length) | 0 - 99 | |
| Average Median Width (ft) | 0 - 38 | 0 - 42 |
| Number of Intersections, Other | | 0 - 9 |
| Number of Signalized Intersections | | 0 - 3 |
| # of Grade Separated Intersections | | 0 - 2 |
| # of Interchanges with Stop Signs | | 0 - 1 |

In addition to the continuous variables listed in Table 3-5, the database included fourteen categorical variables for the rural highway sections and sixteen categorical variables for the urban highway sections. Table 3-6 summarizes the categorical variables in the database and shows the number of subcategories within each variable.

**Table 3-6. Categorical Variables in the Database.**

| Factors | Number of Subcategories | |
|---|---|---|
| | Rural | Urban |
| Access control | 3 | 3 |
| Functional class | 9 | 6 |
| Lane class | 3 | 6 |
| Landuse, left | 10 | 10 |
| Landuse, right | 10 | 9 |
| Median type | 4 | 7 |
| Population density, left | 4 | 6 |
| Population density, right | 5 | 6 |
| Parking, left | | 3 |
| Parking, right | | 3 |
| Rideability | 5 | 5 |
| Side slope | 5 | 5 |
| Surface type | 5 | 6 |
| Shoulder type, left | 3 | 3 |
| Shoulder type, right | 3 | 3 |
| Urban location | 4 | 6 |
| Total | 73 | 87 |

The combination of continuous and categorical variables represents a total of 27 potential explanatory variables for the rural models and a total of 32 potential explanatory variables for the urban models. The evaluation of these variables and the model development phases of the study are described in Chapter 4.

### 3.3.3 Data Analysis, Model Development and Testing

Two approaches were used to develop and test models to predict speeds on rural and urban state highways based on roadway characteristics and adjacent development patterns. The first approach was based on models in the form of multiple linear regression equations. The second approach employed artificial neural networks (ANNs) to predict highway speeds. The basic study design for each approach is presented below.

*3.3.3.1 Regression Analysis*

Multiple regression is an extension of simple, linear regression and can be used to account for the effects of several independent variables simultaneously. The basic form of the multiple regression model is expressed as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \qquad (5)$$

where Y = dependent variable (85th percentile speed), $\beta_0$ = equation constant, $\beta_1, \ldots, \beta_n$ = partial regression coefficients and $X_1, \ldots, X_n$ = independent variables.

The dependent variable (Y) in the regression models developed in this research is the 85th percentile speed on rural and urban state highways. The independent variables evaluated are described in Tables 3-2 and 3-3. The Statistical Analysis System (SAS) [SAS, 1989] was used to perform the regression analyses.

The SAS correlation procedure (PROC CORR) was used to identify those explanatory variables that were significantly correlated with the 85th percentile speed on rural and urban highways. The SAS plot and residual analysis procedures were also used in the data analysis and model development phases of the study. Each of the explanatory variables was plotted against the dependent variable (85th percentile speed) to evaluate the nature of the relationship between speed and each individual variable. Evaluation of the data plots provided indications of the need for possible data transformations and/or the need to consider the use of non-linear terms in the models.

The residual analysis procedure was used to identify "outliers" in the database and to evaluate certain basic assumptions concerning the use of regression analysis, such as the constancy of the variance of the error terms in the regression models.

The stepwise variable selection procedure (PROC STEPWISE) was used to select the independent variables in the speed models. An 85 percent confidence level was used in the stepwise procedure to select significant variables. The stepwise procedure provided candidate models consisting of one explanatory variable, two explanatory variables, ... , n explanatory variables; where n = the total number of explanatory variables considered for the rural and urban models. The General Linear Models procedure (GLM) was used to estimate parameters of the models, correlation coefficients ($R^2$), standard errors, and the probability values (p-values) for the model coefficients.

The selection of the "best" models from the candidate models provided by the stepwise procedure was based on: 1) the number of explanatory variables in the model (all other things being equal, the model with the least number of variables was preferred), 2) $R^2$ values, 3) standard errors, and 4) consistency in the signs of the coefficients of the explanatory variables.

#### *3.3.3.2 Neural Networks*

Artificial Neural Networks (ANNs) are empirical models designed to perform mapping of an input vector into an output vector. The architecture and operation of these networks is an over simplification of those of biological nervous systems. Therefore, ANNs are massively parallel systems that adapt according to stimulus induced by an external environment. In other words, ANNs (especially those based on supervised learning) are designed to learn incrementally from examples presented to them.

The architecture of a simple ANN based on a backpropagation training algorithm is a collection of nodes distributed over an input layer, hidden layer(s), and an output layer. The input variables of the problem are located in the input layer. The output layer contains the output variables, or what is being modeled. In statistical terms, the input layer contains the independent variables and the output layer contains the dependent variable. The nodes between successive layers are connected with links each carrying a weight that describes quantitatively the strength of that connection, thus denoting the strength of one node to affect the other node. For the backpropagation paradigm, no connections between nodes of the same layer is permitted and all connections proceed in the forward direction from input layer to hidden layer and then to output layer with no cyclic or backward connections.

In a backpropagation training algorithm, the first example (input and output vectors) is presented to the network whose connection weights have been initialized before presentation of the example. On each hidden node, the sum representing the scalar product of impinging nodes and their respective connection weights is computed. The sum is then converted to activation by using a typical transfer function such as the sigmoid. This procedure is repeated for each of the higher-level nodes until the output is computed. At this stage, an error function describing the difference between the computed output value and the target value is also calculated. All examples in the database are presented to the network in this forward fashion. Next, an average error function for all examples is determined which is used by the algorithm to adjust the connection weights on all the links starting from the output layer and down to the input layer. This procedure of forward presentation of examples and backward correction of links is repeated many times until the average error function is minimized.

It is essential when developing an ANN that at least two sub-databases from the original database be formed. A training database is used to train the network and the resulting network is then tested on the unseen examples from the testing sub-database. An optimal network is the one that has minimized a specific average error on the testing database. This procedure is conducted to prevent the network from memorizing the training data through an excessive, unnecessary number of training cycles or over-fitting that arises when large numbers of hidden nodes are attempted. The over-fitting phenomenon is very much like what happens when a large number of degrees of freedom is used in polynomial fitting in nonlinear regression. In these situations, the polynomial will be able to produce excellent predictions on those data points used for the regression but not on other data.

Currently, there are several learning paradigms available in the literature for training ANNs. Simpson (1990) listed more than 25 training algorithms or paradigms. The interested reader is referred to the many books and publications on ANNs, such as Simpson (1990), Zupan and Gastieger (1993), Hassoun (1995), Najjar et al. (1996a and 1996b), Basheer and Najjar (1996) and other references cited therein.

# *CHAPTER 4*

# DATA ANALYSIS

This chapter describes the statistical analyses conducted to identify those factors that may have a significant effect on 85th percentile speeds on rural and urban state highways in Kansas. The development and evaluation of regression and artificial neural network models (ANN) for estimating 85th percentile speeds based on these factors is also documented in this chapter.

## 4.1 REGRESSION ANALYSIS

The basic objective of this research effort was to identify factors affecting speeds and to develop models for estimating speeds on rural and urban state highways in Kansas based on roadway characteristics and area development patterns. Models in the form of multiple linear regression equations were developed in an effort to accomplish this objective. The results of the regression analysis phases of the study are documented in the following subsections of this chapter.

### 4.1.1 Rural Speed Models

The rural highway database consisted of 186 highway sections. Data from 180 of these sections were used in the model development and evaluation phases of the research. The six remaining sections were reserved for model validation purposes.

Twenty-seven independent variables were evaluated in the development of the rural highway speed models. The independent variables consisted of 14 categorical variables and 13 continuous variables. The categorical variables contained a total of 73 subcategories. The categorical variables were given a value of one when the variable was applicable to a highway section and a value of zero when the variable was not applicable to a given highway segment.

The initial data analysis activity consisted of a residual analysis of the 180 rural highway sections in the database. The purpose of these analyses was to identify any outliers in the data and to assess the suitability of the data for use in regression analysis. The results of the residual analyses indicated that the database did not contain any unduly influential data points (as measured by Cook's D [SAS, 1989]) and that the data did not violate any of the basic assumptions of regression analysis.

#### *4.1.1.1 Factors Affecting Speeds on Rural Highways*

The SAS GLM procedure was used to identify statistically significant correlations between 85th percentile speed and the 27 potential explanatory variables in the data set. A 95 percent confidence level was used in the correlation analyses. The statistically significant correlations identified in the analyses are summarized in Table 4-1.

The positive signs of the correlation coefficients for AADT, commercial traffic, and substandard stopping sight distance (SSD) appear to be inconsistent with expectations. One would normally expect roadway speeds to decrease with increasing AADT, commercial traffic, and substandard stopping sight distance. One possible explanation might be that the AADT variable, for example, reflects the effects of some other variable(s), such as functional classification. The positive sign for the AADT variable may reflect the fact that AADTs and operating speeds tend to increase from lower to higher functional classes of highways. Likewise, higher speed commercial traffic is more likely on the higher functional classes of highways where commercial traffic volumes tend to be higher. The data available for this research effort was not sufficient to evaluate the relationship between individual functional classifications and highway speed. The counter intuitive relationship between substandard sight distance and speed may reflect interaction with other variables that could be obscuring its true contribution. Variables of this nature should be used with caution.

*Table 4-1.* **Statistically Significant Correlations for Rural State Highways (95% Confidence).**

| Dependent Variable: 85th Percentile Speed | | |
|---|---|---|
| Independent Variables | Correlation Coefficient | p-value |
| Access control - Partial | 0.04 | 0.0001 |
| AADT | 0.01 | 0.003 |
| Accident 5-yr. total | -0.02 | 0.002 |
| Accident, current | -0.07 | 0.0001 |
| Commercial traffic | 0.11 | 0.0064 |
| Functional class - Major arterials | -0.35 | 0.0004 |
| Functional class - Minor arterials | 0.12 | 0.003 |
| Functional class - Other principal arterials | 0.01 | 0.0006 |
| International Roughness Index | -0.09 | 0.0363 |
| Land use (right) - Grassland | 0.02 | 0.0097 |
| Land use (right) - Public (park, game, reserve, etc.) | 0.05 | 0.0039 |
| Land use (right) - Irrigated | 0.15 | 0.023 |
| Population density (left) - Moderate density | 0.02 | 0.0087 |
| Population density (right) - Heavy density | -0.02 | 0.0018 |
| Rideability - Poor | 0.08 | 0.0001 |
| Rideability - Good | -0.05 | 0.035 |
| Restricted passing | -0.39 | 0.0001 |
| Side slope - 2 to 1 | -0.3 | 0.0001 |
| Side slope - 3 to 1 | 0.11 | 0.0081 |
| Substandard stopping sight distance | 0.19 | 0.0001 |
| Surface type - Asphaltic concrete overlay | -0.02 | 0.029 |
| Shoulder type (left) - Curb and Gutter | 0.1 | 0.035 |
| Surface width | 0.07 | 0.016 |

### *4.1.1.2 Model Development*

The SAS Stepwise Procedure was used to develop the candidate models. The stepwise procedure can be used to identify the best one-variable model, the best two-variable model, etc., where the best model is defined as the model with the highest $R^2$ within each set of n-variable models. The variable selection process was based on a significance level of 0.85. The Stepwise Procedure identified 24 candidate models with regression coefficients that were significant at the 0.85 significance level. The $R^2$ values and standard errors for the 24 candidate models are summarized in Table 4-2. As shown in Table 4-2, none of the models perform particularly well in terms of their ability to explain the variability observed in rural highway speeds.

*Table 4-2.* **Summary of R-Square Values and Standard Errors for Candidate Rural Speed Models.**

| Number of Variables | $R^2$ | Standard Error |
|---|---|---|
| 1 | 0.15 | 7.11 |
| 2 | 0.29 | 6.51 |
| 3 | 0.37 | 6.17 |
| 4 | 0.43 | 5.87 |
| 5 | 0.49 | 5.59 |
| 6 | 0.53 | 5.37 |
| 7 | 0.57 | 5.19 |
| 8 | 0.60 | 5.0 |
| 9 | 0.62 | 4.87 |
| 10 | 0.64 | 4.73 |
| 11 | 0.66 | 4.62 |
| 12 | 0.68 | 4.48 |
| 13 | 0.70 | 4.37 |
| 14 | 0.72 | 4.22 |
| 15 | 0.74 | 4.13 |
| 16 | 0.75 | 4.07 |
| 17 | 0.76 | 4.03 |
| 18 | 0.77 | 4.0 |
| 19 | 0.78 | 3.96 |
| 20 | 0.78 | 3.91 |
| 21 | 0.79 | 3.86 |
| 22 | 0.79 | 3.8 |
| 23 | 0.80 | 3.77 |
| 24 | 0.80 | 3.73 |

#### *4.1.1.3 Model Testing and Selection*

Current practice is to set highway speed limits at the nearest 5-mph increment of the 85th percentile speed. Therefore, it was considered desirable to develop a model capable of estimating 85th percentile speeds within ± 5 mph. Additional model selection criteria included a high R-square value and a low standard error.

Data from six randomly selected rural highway segments that were not used in the model estimation process were used in an attempt to identify a model that was capable of estimating observed 85th percentile speeds within ± 5 mph. Unfortunately, none of the candidate models possessed this level of accuracy at an acceptable level of confidence (see Table 4-3). As a result, no single model is put forth as the "best" model. As shown in Table 4-3, the incremental reductions in the widths of the prediction bands of the models are not substantial for the models with more than 10 variables. Therefore, only 10 models are offered for consideration (see Table 4-4). The models in Table 4-4 provide an indication of the relative importance of the various explanatory variables and can be used when initial estimates of rural highway speeds are needed. In such applications, it is suggested that the analyst select the model with the largest number of variables for which data are available.

*Table 4-3.* **Prediction Bands of Candidate Rural Speed Models (Observed vs. Predicted 85th Percentile Speed)**

| Number of Variables | Prediction Band (± mph) at Confidence Level Indicated | | |
|---|---|---|---|
| | 95 percent | 90 percent | 85 percent |
| 1 | 13.94 | 11.70 | 10.41 |
| 2 | 12.76 | 10.71 | 9.53 |
| 3 | 12.09 | 10.15 | 9.03 |
| 4 | 11.51 | 9.66 | 8.61 |
| 5 | 10.96 | 9.20 | 8.19 |
| 6 | 10.53 | 8.83 | 7.88 |
| 7 | 10.17 | 8.54 | 7.60 |
| 8 | 9.80 | 8.23 | 7.32 |
| 9 | 9.55 | 8.01 | 7.13 |
| 10 | 9.27 | 7.78 | 6.94 |
| 11 | 9.06 | 7.60 | 6.76 |
| 12 | 8.78 | 7.37 | 6.54 |
| 13 | 8.57 | 7.19 | 6.40 |
| 14 | 8.27 | 6.94 | 6.15 |
| 15 | 8.09 | 6.79 | 6.03 |
| 16 | 7.98 | 6.70 | 5.95 |
| 17 | 7.90 | 6.63 | 5.91 |
| 18 | 7.84 | 6.58 | 5.86 |
| 19 | 7.76 | 6.51 | 5.80 |
| 20 | 7.66 | 6.44 | 5.72 |
| 21 | 7.57 | 6.35 | 5.66 |
| 22 | 7.45 | 6.25 | 5.56 |
| 23 | 7.39 | 6.20 | 5.51 |
| 24 | 7.31 | 6.14 | 5.45 |

*Table 4-4.* **85th Percentile Speed Models for Rural State Highways.**

| Number of Variables | Models | p-values, $R^2$ and Standard Errors (S.E.) |
|---|---|---|
| 1-variable | $V_{85(R)} = 60.48 - 0.1(RP)$ | p-value = 0.0001, $R^2 = 0.15$, S.E. = 7.11 |
| 2-variable | $V_{85(R)} = 61.47 - 0.1(RP) - 10.91(ST2)$ | p-value = 0.0001, $R^2 = 0.29$, S.E. = 6.51 |
| 3-variable | $V_{85(R)} = 61.11 - 0.13(RP) - 9.84(ST2) + 0.03(SSD)^2$ | p-value = 0.0001, $R^2 = 0.37$, S.E. = 6.17 |
| 4-variable | $V_{85(R)} = 62.27 - 4.32(FC13) - 0.12(RP) - 7.06(ST2) + 0.04(SSD)^2$ | p-value = 0.0001, $R^2 = 0.43$, S.E. = 5.87 |
| 5-variable | $V_{85(R)} = 62.11 - 4.31(FC13) - 0.1(RP) - 6.83(SS2) - 7.49(ST2) + 0.04(SSD)^2$ | p-value = 0.0001, $R^2 = 0.49$, S.E. = 5.59 |
| 6-variable | $V_{85(R)} = 62.65 - 3.92(FC13) - 6.46(LUI3) - 0.11(RP) - 6.96(SS2) - 7.27(ST2) + 0.04(SSD)^2$ | p-value = 0.0001, $R^2 = 0.53$, S.E. = 5.37 |
| 7-variable | $V_{85(R)} = 62.7 - 3.72(FC13) - 6.54(LUI3) - 21.96(RID1) - 0.12(RP) - 8.17(SS2) - 7.31(ST2) + 0.04(SSD)^2$ | p-value = 0.0001, $R^2 = 0.57$, S.E. = 5.19 |
| 8-variable | $V_{85(R)} = 62.94 - 3.83(FC13) + 10.67(LC2) - 6.43(LUI3) + 22.61(RID1) - 0.13(RP) - 7.65(SS2) - 7.82(ST2) + 0.04(SSD)^2$ | p-value = 0.0001, $R^2 = 0.6$, S.E. =5.0 |
| 9-variable | $V_{85(R)} = 64.66 - 4.35(FC13) + 10.86(LC2) - 6.52(LUI3) + 20.58(RID1) - 0.13(RP) - 7.66(SS2) - 6.86(ST2) + 0.05(SSD)^2 - 0.0002(IRI)^2$ | p-value = 0.0001, $R^2 = 0.62$, S.E. = 4.87 |
| 10-variable | $V_{85(R)} = 61.23 - 3.94(FC13) + 11.33(LC2) - 6.93(LUI3) + 22.24(RID1) - 0.13(RP) - 6.57(SS2) - 8.02(ST2) + 0.39(SWI) + 0.05(SSD)^2 - 0.0002(IRI)^2$ | p-value = 0.0001, $R^2 = 0.64$, S.E. = 4.73 |

FC13 = Functional Class (minor arterial), LC2 = Lane Class (four lane undivided), LUI3 = Land Use Left (cultivated bottom land), RID1 = Rideability (poor), RP = Restricted Passing, SS2 = Side Slope (2 to 1), ST2 = Surface Type (mixed surface), SWI = Shoulder Width Left, $SSD^2$ = [Substandard Stopping Sight Distance]$^2$, $IRI^2$ = [International Roughness Index]$^2$. See Appendix for detailed description of variables.

## 4.1.2 URBAN SPEED MODELS

The urban highway database consisted of 353 highway sections. Data from 347 of these sections were used in the model development and evaluation phases of the research. The six remaining sections were reserved for model validation purposes.

Thirty-two independent variables were evaluated in the development of the urban highway speed models. The independent variables consisted of 16 categorical variables and 16 continuous variables. The categorical variables contained a total of 88 subcategories.

The initial data analysis activity consisted of a residual analysis of the 347 urban highway sections in the database. The purpose of these analyses was to identify any outliers in the data and to assess the suitability of the data for use in regression analysis. The results of the residual analyses indicated that the database did not contain any unduly influential data points (as measured by Cook's D [SAS, 1989]) and that the data did not violate any of the basic assumptions of regression analysis.

#### *4.1.2.1 Factors Affecting Speeds on Urban Highways*

The SAS GLM procedure was used to identify statistically significant correlations between 85th percentile speed and the 32 potential explanatory variables in the urban data set. A 95 percent confidence level was used in the correlation analyses. The statistically significant correlations identified in the analyses are summarized in Table 4-5.

The positive signs of the correlation coefficients for accident experience and commercial traffic appear to be inconsistent with expectations. One would normally expect roadway speeds to decrease with increasing accidents and commercial traffic. On the other hand, this relationship could be related to highway functional classification, and/or urban area type. For example, the higher classifications of urban arterials tend to have higher speeds, more commercial traffic and, therefore, possibly more accidents. As in the case of the rural highway analysis, the data available for this research effort was not sufficient to evaluate the relationship between highway speed and functional classifications or urban area type.

*Table 4-5.* **Statistically Significant Correlations for Urban State Highways (95% Confidence).**

| Dependent Variable: 85th Percentile Speed | | |
|---|---|---|
| Independent Variables | Correlations | p-values |
| Accident 5-year total | 0.06 | 0.044 |
| Commercial traffic | 0.34 | 0.0001 |
| Functional class - Freeways and expressways | 0.28 | 0.036 |
| Lane class - Six lane divided | 0.28 | 0.047 |
| Land use (right) - Irrigated | 0.15 | 0.011 |
| Land use (right) - Current business district | - 0.34 | 0.019 |
| Lane width | 0.14 | 0.016 |
| Rideability - Very good | 0.36 | 0.024 |
| Shoulder type (left) - Stabilized | 0.54 | 0.0005 |

### *4.1.2.2 Model Development*

The SAS Stepwise Procedure was used to develop the candidate models. The variable selection process was based on a significance level of 0.85. The Stepwise Procedure identified 30 candidate models with regression coefficients that were significant at the 0.85 significance level. The $R^2$ values and standard errors for the 30 candidate models are summarized in Table 4-6. As shown in Table 4-6, none of the models perform particularly well in terms of explaining the variability observed in urban highway speeds.

*Table 4-6.* **Summary of R-Square Values and Standard Errors for Candidate Urban Speed Models.**

| Number of Variables | $R^2$ | Standard Error |
|---|---|---|
| 1 | 0.3 | 8.42 |
| 2 | 0.36 | 8.07 |
| 3 | 0.4 | 7.8 |
| 4 | 0.48 | 7.31 |
| 5 | 0.51 | 7.11 |
| 6 | 0.53 | 6.96 |
| 7 | 0.55 | 6.83 |
| 8 | 0.56 | 6.73 |
| 9 | 0.58 | 6.64 |
| 10 | 0.59 | 6.54 |
| 11 | 0.61 | 6.45 |
| 12 | 0.63 | 6.37 |
| 13 | 0.64 | 6.24 |
| 14 | 0.65 | 6.15 |
| 15 | 0.66 | 6.08 |
| 16 | 0.67 | 6.0 |
| 17 | 0.67 | 5.95 |
| 18 | 0.68 | 5.92 |
| 19 | 0.68 | 5.88 |
| 20 | 0.69 | 5.82 |
| 21 | 0.7 | 5.79 |
| 22 | 0.7 | 5.76 |
| 23 | 0.71 | 5.73 |
| 24 | 0.71 | 5.71 |
| 25 | 0.71 | 5.68 |
| 26 | 0.72 | 5.66 |
| 27 | 0.72 | 5.63 |
| 28 | 0.72 | 5.61 |
| 29 | 0.73 | 5.58 |
| 30 | 0.73 | 5.56 |

### *4.1.2.3 Model Testing and Selection*

As in the case for rural highways, it was considered desirable to develop a model capable of estimating 85th percentile speeds within ± 5 mph. Additional model selection criteria included a high R-square value and a low standard error.

Data from six randomly selected urban highway segments that were not used in the model estimation process were used in an attempt to identify a model that was capable of estimating observed 85th percentile speeds within ± 5 mph. As in the case of the rural models, none of the candidate urban speed models exhibited this level of precision at an acceptable level of confidence (see Table 4-7). As a result, no single model can be identified as superior in terms of its overall accuracy. Table 4-8 summarizes 10 models that can be used to develop preliminary estimates of urban highway speeds. When using these models, it is suggested that the analyst select the largest model for which data are available.

*Table 4-7.* **Prediction Bands of Candidate Urban Speed Models (Observed vs. Predicted 85th Percentile Speed)**

| Number of Variables | Prediction Band (±mph) at Confidence Level Indicated | | |
|---|---|---|---|
| | 95 percent | 90 percent | 85 percent |
| 1 | 16.50 | 13.86 | 12.33 |
| 2 | 15.82 | 13.26 | 11.81 |
| 3 | 15.29 | 12.83 | 11.42 |
| 4 | 14.32 | 12.04 | 10.77 |
| 5 | 13.96 | 11.73 | 10.49 |
| 6 | 13.64 | 11.46 | 10.21 |
| 7 | 13.40 | 11.25 | 10.00 |
| 8 | 13.19 | 11.08 | 9.87 |
| 9 | 13.01 | 10.92 | 9.70 |
| 10 | 12.84 | 10.78 | 9.57 |
| 11 | 12.64 | 10.66 | 9.44 |
| 12 | 12.49 | 10.45 | 9.31 |
| 13 | 12.26 | 10.29 | 9.14 |
| 14 | 12.05 | 10.12 | 9.00 |
| 15 | 11.92 | 10.00 | 8.90 |
| 16 | 11.78 | 9.89 | 8.78 |
| 17 | 11.66 | 9.79 | 8.71 |
| 18 | 11.60 | 9.74 | 8.67 |
| 19 | 11.52 | 9.67 | 8.60 |
| 20 | 11.49 | 9.58 | 8.52 |
| 21 | 11.36 | 9.57 | 8.46 |
| 22 | 11.29 | 9.47 | 8.43 |
| 23 | 11.23 | 9.43 | 8.39 |
| 24 | 11.20 | 9.42 | 8.37 |
| 25 | 11.12 | 9.36 | 8.32 |
| 26 | 11.09 | 9.34 | 8.29 |
| 27 | 11.07 | 9.30 | 8.28 |
| 28 | 11.00 | 9.24 | 8.21 |
| 29 | 10.95 | 9.18 | 8.17 |
| 30 | 10.88 | 9.16 | 8.14 |

*Table 4-8.* **85th Percentile Speed Models for Urban State Highways.**

| Number of variables | Models | p-values, $R^2$ and Standard Errors (S.E.) |
|---|---|---|
| 1-variable | $V_{85(U)}$ = 38.06 + 1.32(SWO) | p-value = 0.0001, $R^2$ = 0.3, S.E. = 8.42 |
| 2-variable | $V_{85(U)}$ = 39.03 - 9.84(LUO6) + 1.22(SWO) | p-value = 0.0001, $R^2$ = 0.36, S.E. = 8.07 |
| 3-variable | $V_{85(U)}$ = 38 - 9.41(LUO6) + 4.81(RID4) + 1.08(SWO) | p-value = 0.0001, $R^2$ = 0.4, S.E. = 7.8 |
| 4-variable | $V_{85(U)}$ = 37.26 + 8.12(LUI11) - 8.66(LUO6) + 4.71(RID4) + 1.11(SWI) | p-value = 0.0001, $R^2$ = 0.48, S.E. = 7.31 |
| 5-variable | $V_{85(U)}$ = 36.04 + 7.72(LUI11) - 8.14(LUO6) + 3.69(PDI4) + 4.2(RID4) + 1.01(SWI) | p-value = 0.0001, $R^2$ = 0.51, S.E. = 7.11 |
| 6-variable | $V_{85(U)}$ = 35.88 + 7.94(LUI11) 16.28(LUO4) - 7.98(LUO6) + 3.54(PDI4) + 4.48(RID4) + 0.99(SWI) | p-value = 0.0001, $R^2$ = 0.53, S.E. = 6.96 |
| 7-variable | $V_{85(U)}$ = 35 + 0.013(ACC) + 7.59(LUI11) + 16.8(LUO4) - 7.86(LUO6) + 3.89(PDI4) + 4.67(RID4) + 1.02(SWI) | p-value = 0.0001, $R^2$ = 0.55, S.E. = 6.83 |
| 8-variable | $V_{85(U)}$ = 33.48 + 0.016(ACC) + 0.2(COM) + 8.52(LUI11) + 15.28(LUO4) - 7.23(LUO6) + 3.35(PDI4) + 4.03(RID4) + 0.91(SWI) | p-value = 0.0001, $R^2$ = 0.56, S.E. = 6.73 |
| 9-variable | $V_{85(U)}$ = 33.01 + 0.016(ACC) + 0.2(COM) + 8.53(LUI11) + 15.7(LUO4) - 6.98(LUO6) + 3.59(PDI4) + 4.47(RID4) + 6.57(RID5) + 0.86(SWI) | p-value = 0.0001, $R^2$ = 0.58, S.E. = 6.64 |
| 10-variable | $V_{85(U)}$ = 33.02 + 0.017(ACC) + 0.21(COM) + 10.62(LC5) + 7.59(LUI11) + 15.52(LUO4) - 7.05(LUO6) + 4(PDI4) + 4.17(RID4) + 6.92(RID5) + 0.77(SWI) | p-value = 0.0001, $R^2$ = 0.59, S.E. = 6.54 |

ACC = Accident (5-year total), COM = Commercial Traffic, LC5 = Lane Class (six lane divided), LUI11 = Land Use Left (highway median), LUO4 = Land Use Right (irrigated), LUO6 = Land Use Right (central business district), PDI4 = Population Density Left (light), RID4 = Rideability (very good), RID5 = Rideability (excellent), SWI = Shoulder Width Left, SWO = Shoulder Width Right. See Appendix for detailed description of variables.

### 4.1.3 Summary of Regression Analysis

The preceding sections of this chapter have documented the results of analyses directed at the development of regression models for estimating 85th percentile speeds on rural and urban state highways based on roadway and adjacent development characteristics. The models developed in the research suffer from a number of fundamental shortcomings. First, the R-square values are not particularly high. This is probably attributable to the high degree of variability observed in the CANSYS and highway speed databases. Secondly, the models have a large number of variables and are not as tractable as desired. However, because all the data needed to use the models are available in the

CANSYS database, the relatively large number of variables in the models was not considered to be a significant disadvantage. Finally, the models are " generic " models in that they represent all rural or urban highways. For example, the models are not capable of distinguishing between the different functional classes of highways. Unfortunately, the number of observations in the data available for this research effort was not large enough to develop and evaluate multi-variable, multiple linear regression models based on functional class, number of lanes and other subcategories of state highways. As a result, the regression models developed in this research effort should be used with caution. They should be adequate where preliminary, planning-level estimates of highway speeds are needed. It should be noted, however, that based on the results of this study, the neural network models appear to give better results with the available data. The neural network modeling effort is discussed below.

## 4.2 NEURAL NETWORK MODELS

Two independent databases were used to train two sets of Artificial Neural Network (ANN) models of rural and urban highway speeds. The first set of ANN models (the Stage 1 models) was developed using the same CANSYS-based data used in the regression analysis. The second set of ANN models (the Stage 2 models) was developed using a database provided by the KDOT Bureau of Traffic Engineering. The second database contained only those variables that KDOT engineers believed drivers consider in selecting a driving speed (i.e., those variables that drivers perceive and respond to as part of the driving task).

### 4.2.1 Stage 1 ANN Models

The data used to develop the regression models was also used to develop the Stage 1 Artificial Neural Networks (ANN) models. The reader may refer to the description of the data in previous chapters of this report. An ANN-Backpropagation computer program developed by Dr. Yacoub Najjar, Department of Civil Engineering, Kansas State University, was used to train and test the

ANN-based models. The program allows for simultaneous feedback during the entire training stage on the degree of network prediction accuracy for both the training and testing data sets. Accordingly, this allows the user to specify the optimum network in terms of the number of hidden nodes required and the maximum number of training iterations permitted. By using this program, two ANNs describing optimum performance for rural and urban highways were obtained. Issues related to the development of both networks are summarized below. Because each database contained a large number of input variables (86 continuous and categorical input variables for rural highways and 103 continuous and categorical input variables for urban highways), a trimming strategy was adopted to reduce the dimensionality of the models. Trimming was performed by initially designing a trial network which utilized the full number of input nodes (i.e., 86 for rural highways and 103 for urban highways). The relative importance of each input variable based on the strength of connection weights of the network was then evaluated. This procedure provided a ranking of the input variables according to their strength to affect the output parameter. A heuristic was then used to eliminate from the model those input parameters that were ranked below a threshold rank level. A new network with the reduced dimensionality was then developed and the procedure of ranking was repeated to eliminate additional input parameters below the threshold rank. While this was done repeatedly, the error on the testing and training data was monitored to insure that network performance had not deteriorated drastically between successive network trimmings. The criteria used to monitor model performance are the Mean of Absolute value of Relative Error (MARE), expressed in percent, and the coefficient of determination ($R^2$) [a measure of the agreement between the predicted and target values of the output (i.e., the 85% percentile speed, $V_{85}$)].

*4.2.1.1 Stage 1 Rural Highway Speed ANNs*

In the design of the rural ANNs, 165 data sets were used to train the networks. Twenty-one (21) sets were used for testing and cross validation. A list of the networks developed using the strategy described above is summarized in the Table 4-9. In this Table, IN-HN-ON denotes the number of nodes in input (IN), hidden (HN), and output layers (ON), respectively. Network (12-11-1) includes the following eight input variables: surface width, International Roughness Index (IRI), percent heavy commercial traffic, stopping sight distance, number of accidents (current), percent restricted passing, median width, and rideability (a categorical variable with five classes). These eight input variables constitute the 12 input nodes of the network. Network (7-4-1) is the same as network (12-11-1) without the rideability variable. Network (6-4-1) is the same as network (7-4-1) without the number of accidents (current). As shown in Table 4-9, network (12-11-1) is the best network to use for predicting $V_{85}$ on rural highways. This network has the highest overall $R^2$ and the lowest overall MARE. Therefore, if this network predicts $V_{85}$ = X mph, then the actual speed may lie between X(1 ± (overall MARE/100)) mph. For example, if this network has predicted $V_{85}$ = 60 mph, then the actual speed may lie between 58.77 mph [60(1 - 0.0205) = 58.77 mph] and 61.23 mph [60(1 + 0.0205) = 61.23 mph]. The agreement between the predicted and measured $V_{85}$ values using network (12-11-1) is shown in Figure 4-1a for the training and testing data combined. Figure 4-1b shows the agreement between the measured and predicted $V_{85}$ for the best regression equation.

***Table 4-9.*** **Stage 1 Rural Networks and Corresponding Prediction Accuracy Measures.**

| ANN | # Input parameters | Training | | Testing | | Overall | |
|---|---|---|---|---|---|---|---|
| IN-HN-ON | | MARE | $R^2$ | MARE | $R^2$ | MARE | $R^2$ |
| 86-2-1 | 26 | 2.81 | 0.939 | 3.96 | 0.846 | 2.94 | 0.928 |
| **12-11-1** | *8* | ***1.98*** | ***0.951*** | ***2.58*** | ***0.892*** | ***2.05*** | ***0.945*** |
| 7-4-1 | 7 | 6.01 | 0.698 | 5.31 | 0.726 | 5.93 | 0.702 |
| 6-4-1 | 6 | 5.53 | 0.641 | 7.05 | 0.591 | 5.70 | 0.636 |

IN-HH-ON denotes Number of nodes in input, hidden, and output layers, respectively.
MARE denotes Mean of Absolute value of Relative Error (percent).
$R^2$ denotes Coefficient of Determination.

### *4.2.1.2 Stage 1 Urban Highway Speed ANNs*

In designing the urban highway networks, 303 data sets were used in training the model. An

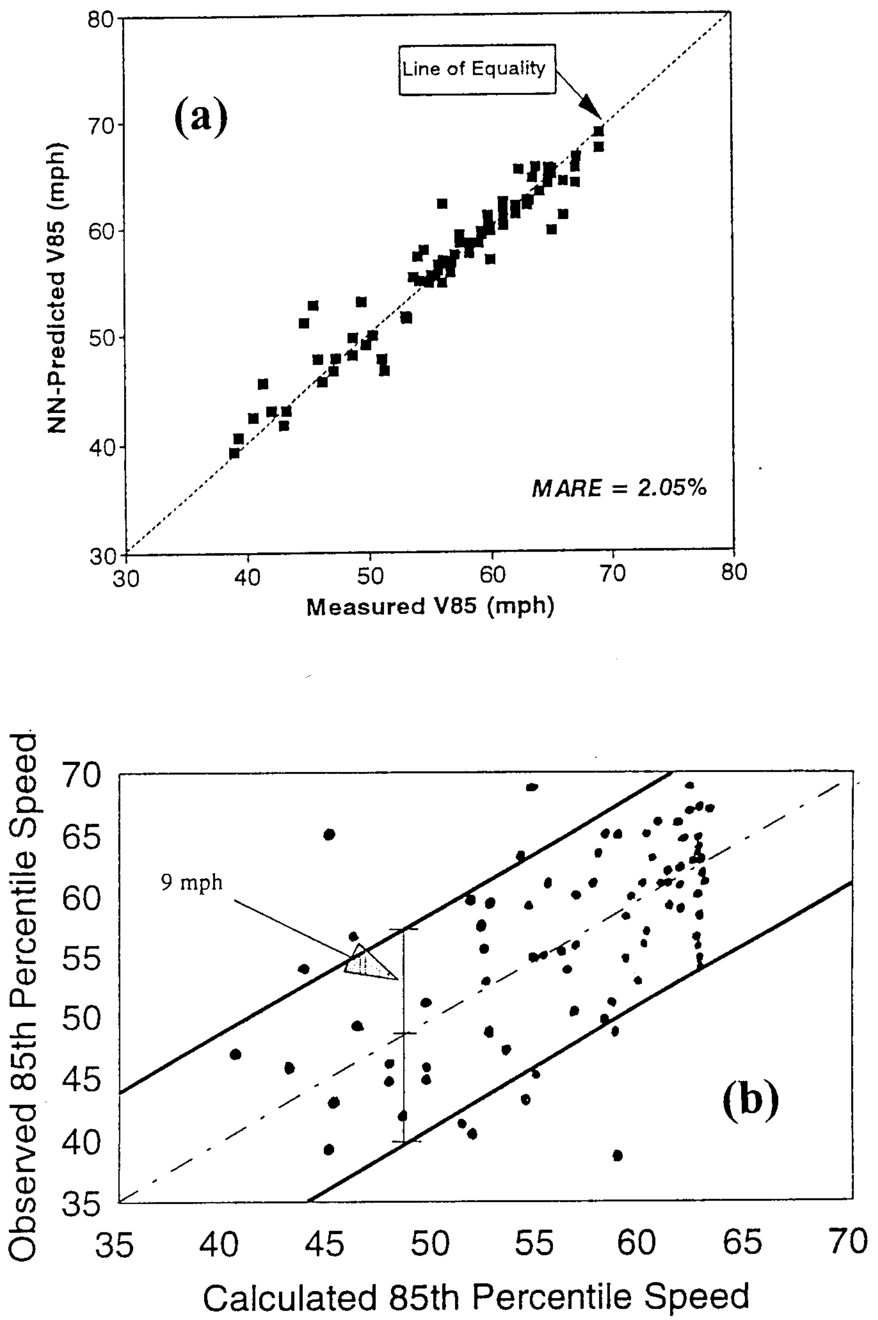



Figure 4-1 Agreement between measured 85th percentile speed and predictions using (a) neural networks, and (b) statistical regression for rural highways.

additional 50 data sets were used for testing. Table 4-10 summarizes the various networks designed and the subsequent evolution of these networks. In this table, network (30-2-1) contains seven continuous and four categorical variables with a total of 30 input nodes. The continuous variables are: AADT, percent heavy commercial traffic, stopping sight distance, number of accidents (5-years total), number of accidents (current), number of signalized intersections, and number of grade separated interchanges. The categorical variables are: rideability, land use (outside), urban location, and parking (outside). Network (12-10-1) includes the seven continuous variables of network (30-2-1) plus the rideability variable.

***Table 4.10*** **Stage 1 Urban Networks and Corresponding Prediction Accuracy Measures.**

| ANN | # Input Parameters | Training | | Testing | | Overall | |
|---|---|---|---|---|---|---|---|
| IN-HN-ON | | MARE | $R^2$ | MARE | $R^2$ | MARE | $R^2$ |
| 103-5-1 | 32 | 17.50 | 0.424 | 25.60 | 0.411 | 18.60 | 0.420 |
| 30-2-1 | 11 | 14.15 | 0.548 | 13.35 | 0.622 | 14.04 | 0.558 |
| 12-10-1 | 8 | 12.01 | 0.648 | 17.75 | 0.506 | 12.82 | 0.628 |

IN-HH-ON denotes number of nodes in input, hidden, and output layers, respectively.
MARE denotes Mean of Absolute value of Relative Error (percent).
$R^2$ denotes Coefficient of Determination.

As shown in Table 4-10, the initial overall MARE values were relatively high. The stepwise variable selection (SWVS) procedure described in previous sections of this chapter was used in an attempt to reduce this error. The 10 most important independent variables (three continuous and seven categorical variables) identified by the statistically-based SWVS procedure were used to design another network by training on the same data used previously. In this case, using the SWVS to identify the most important independent variables was believed to be more accurate than the network-based relative importance concept devised by studying the strengths of connection weights. This is because the SWVS procedure is based on analysis-of-variance while the relative importance of networks may suffer from excessive simplifying assumptions. Nevertheless, the network developed based on the SWVS regression model was also trimmed several times using the network relative importance procedure to determine if any improvement can be achieved. The networks obtained from the regression-based SWVS procedure are referred to as Hybrid Statistical Neural Networks (HSNNs). Table 4-11 summarizes the performance of the five HSNNs developed in this research effort.

***Table 4-11.*** **Hybrid Statistical Urban Networks and Corresponding Prediction Accuracy Measures.**

| HSNN | # Input Parameters | Training | | Testing | | Overall | |
|---|---|---|---|---|---|---|---|
| IN-HN-ON | | MARE | $R^2$ | MARE | $R^2$ | MARE | $R^2$ |
| ***48-5-1*** | ***10*** | ***6.12*** | ***0.893*** | ***12.64*** | ***0.512*** | ***7.04*** | ***0.840*** |
| 37-2-1 | 7 | 10.92 | 0.679 | 11.08 | 0.613 | 10.94 | 0.670 |
| 31-2-1 | 6 | 11.49 | 0.687 | 10.26 | 0.614 | 11.32 | 0.677 |
| 25-5-1 | 5 | 10.13 | 0.731 | 15.00 | 0.538 | 10.82 | 0.703 |
| 19-2-1 | 4 | 12.33 | 0.619 | 12.29 | 0.482 | 12.32 | 0.600 |

IN-HH-ON denotes number of nodes in input, hidden, and output layers, respectively.
MARE denotes Mean of Absolute value of Relative Error (percent).
$R^2$ denotes Coefficient of Determination.

The input parameters for the HSNN with 10 input parameters ordered based on the neural network-based relative importance concept are: number of accidents in five years, rideability, shoulder type inside, land use inside, functional class, lane class, population density inside, land use outside, percent heavy commercial traffic and lane width. Network (37-2-1) includes only the first seven of the parameters listed above, network (31-2-1) includes only the first six parameters, network (25-5-1) contains only the first five parameters, and network (19-2-1) contains only the first four parameters.

A comparison of Tables 4-10 and 4-11 shows that the HSNN with 10 parameters produces the most accurate predictions for $V_{85}$ on urban highways. Figure 4-2a shows the agreement between

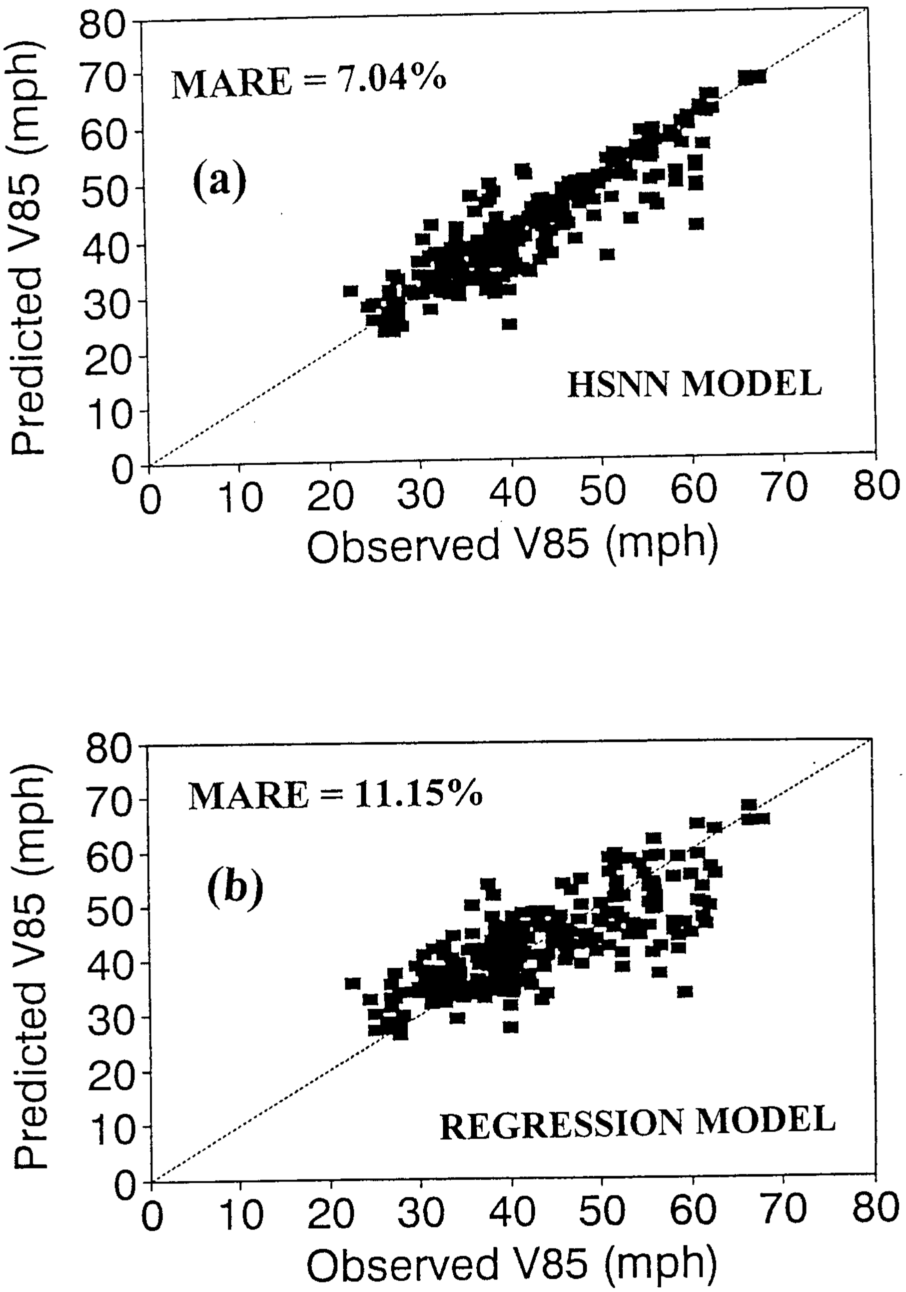



Figure 4-2 Agreement between measured 85th percentile speed and predictions using (a) hybrid statistical neural networks, and (b) statistical regression for urban highways.

the 85th percentile speeds ($V_{85}$) predicted using the HSNN with 10 parameters and the corresponding measured values. For comparison, Figure 4-2b also shows the agreement between measured and predicted $V_{85}$ values obtained from the best SWVS linear regression-based equation.

### 4.2.2 Stage 2 ANN Models

At the request of the KDOT project monitor, the Stage 1 urban and rural ANN-based models developed in this study were used to predict the 85th percentile speeds on all relevant Kansas highways. This was performed in order to obtain direct feedback from the KDOT project monitor and the Bureau of Traffic Engineering. Based on KDOT evaluation of the Stage 1 model results, it was determined that due to some apparent inconsistences/limitations in the available databases, additional neural network modeling should be pursued using new/revised databases. Also, it was recommended that the input parameters be limited to those variables that drivers perceive and respond to as part of the driving task. As a result, input variables such as accident frequencies were eliminated from the new databases. Modeling issues involved in developing the revised (Stage 2) ANN-based models are discussed in the following sections of this chapter.

#### *4.2.2.1 Stage 2 Rural Highway Speed ANNs*

In the design of the Stage 2 rural ANNs, the modeling effort was carried out in two consecutive phases. In the first phase, 56 data sets were provided by KDOT to characterize the 85th percentile (V85%) speed on two-lane rural Kansas highways. The 85th% speed was modeled by the neural network approach as a function of the following six variables: Average Daily Traffic (ADT), Shoulder Width, Lane Width, percent No Passing Zones, Route Class (B, C, D, and E) and Shoulder Type (surface, earth, combination or stabilized). The resulting ANN was then used to identify the most influential parameters that affect the 85th% speed. Based on ANN sensitivity analysis and consultation with KDOT personnel, it was decided to eliminate the Route Class and Lane Width variables. Also, the shoulder types were re-grouped into two categories: 1) Pavement/Combination (P/C); or 2) Gravel/Turf (G/T). As a result, it was desired to have the new ANN characterize the

85th% speed as a function of ADT, shoulder width, percent no passing zones and the shoulder type (P/C or G/T). Due to the relatively small number of data sets involved, no testing was performed in this phase of the analysis. It was decided to obtain more data sets in order to train the ANN on one part of the data sets and to test its prediction accuracy on the remaining data sets.

In the second phase of the Stage 2 analysis, 117 complete data sets (each representing one two-lane highway section with its five road-related variables and the corresponding measured 85th percentile speed) were provided by KDOT. The data sets were divided into two major training and testing sets. The training set containing 88 data sets was used to train the desired speed ANN model. At the same time the generalization capability (i.e., prediction accuracy) of the designed ANN was tested on the remaining 29 data sets which the network had never seen before. This process was repeated until the best performing net was achieved at 5 hidden nodes. This network can be designated as a 5-5-1 network to represent 1) 5 input nodes (shoulder width, shoulder type (P/C or G/T), ADT and percent no passing zone, 2) 5 hidden nodes, and 3) one output node denoting the 85th% speed. According to ANN methodology, it is the presence of the hidden nodes (5 in this case) that allows the neural network to accurately map (i.e., correlate) the 5-input road-related parameters to the desired output (the 85th% speed). A summary of key measures of accuracy for the Stage 2 rural speed ANN is given below:

- Overall average error = 4% (i.e., overall average accuracy in predicting V85% = 96%).
- Overall average 85th% speed (based on actual data sets) = 64.89 mph
- Overall average 85th% speed (predicted from ANN) = 64.86 mph

*4.2.2.2 Stage 2 Urban Highway Speed ANNs*

In designing the Stage 2 urban networks, 477 completely new/revised data sets were provided by KDOT for use in this study. The complete database was divided into three major groups. Data group 1 consisted of 246 sets which was used for network training purposes. Data group 2 consisted of another 116 sets which was used for on-line network testing purposes. Data group 3, which consisted of the remainder (115) sets, was reserved for model validation purposes. Each complete data set contained the following information:

1) The posted speed limit in MPH.
2) Average daily traffic (ADT) volumes.
3) Lane type category:
   a) 2-lane undivided (2LU).
   b) 2-lane with an extra center lane for left turns (2LW).
   c) 4-lane divided (4LD).
   d) 4-lane undivided (4LU).
   e) 4-lane with an extra center lane for left turns (4LW).
4) Parking type:
   a) Angular parking.
   b) Parallel parking.
   c) No parking.
5) Area density type:
   a) High business (commercial) zone.
   b) Low commercial (Business) zone.
   c) Low residential zone.
   d) Others (very low residential zone).
6) The field measured 85th percentile speed.

Based on both statistical and neural network analysis techniques, it was concluded that AADT has the least influence ($R^2$ of 0.002) on the 85th% speed. As a result, that variable was dropped from any further consideration. Moreover, the study revealed a strong correlation ($R^2$ of about 0.75) between the 85th% speed and the posted speed limit. Therefore, it was decided to develop two ANN models. One model will predict the 85th% speed for a given lane, parking and area density type, while the second model will utilize the same input variables as the first model plus the posted speed limit. These two models give the user the option of investigating situations with known or unknown posted speed limits. Prediction accuracy measures, in terms of $R^2$ and MARE for ANN models 1 and 2, are given in Table 4.12. As shown in Table 4.12, ANN model 2 (which utilizes the posted speed limit as an essential input parameter) predicts the actual 85th% speed with higher accuracy than ANN model 1. This reflects the strong correlation observed between the posted speed limit and the 85th% speed. Furthermore, a comparison of the predicted and the actual 85th% speeds of the validation data sets, shows that model 2 predicts the 85th% speed within ± 5 MPH of the actual values in 75% of the cases. For the case of ANN model 1, this accuracy measure drops to about 45% of the cases involved.

*Table 4.12.* **Stage 2 Urban Networks and Corresponding Prediction Accuracy Measures.**

| ANN | # Input | Training | | Testing | | Validation | | Overall | |
|---|---|---|---|---|---|---|---|---|---|
| IN-HN-ON | Parameters | MARE | $R^2$ | MARE | $R^2$ | MARE | $R^2$ | MARE | $R^2$ |
| 12-5-1 (model #1) | 3 | 14.1 | 0.55 | 13.3 | 0.476 | 19.6 | 0.365 | 15.2 | 0.488 |
| 13-2-1 (model #2) | 4 | 7.7 | 0.86 | 8.0 | 0.806 | 10.9 | 0.786 | 8.54 | 0.829 |

IN-HH-ON denote number of nodes in input, hidden, and output layers, respectively.
MARE denotes Mean of Absolute value of Relative Error, expressed in percent.
$R^2$ denotes the Coefficient of Determination.

### 4.2.3 Summary of Neural Network Models

Two independent databases were used to train two sets of Artificial Neural Network (ANN) models of rural and urban highway speeds. The first set of ANN models (the Stage 1 models) was developed using the same CANSYS-based data used in the regression analysis. The second set of ANN models (the Stage 2 models) was developed using a database provided by the KDOT Bureau of Traffic Engineering. The second database contained only those variables that KDOT engineers believed drivers consider in selecting a driving speed (i.e., those variables that drivers perceive and respond to as part of the driving task).

Based on the results of this study, the Stage 2 ANN models perform better than either the regression based models or the Stage 1 ANN models in predicting 85$^{th}$ percentile speeds on rural and urban state highways.

The Stage 2 ANN-based models for urban and rural state highways have been coded into $C^{++}$ software programs that can be run on any Pentium platform under Windows® 3.1, 95 or NT4.0 environments. Each program asks the user to enter the values of all required input variables for the specified model (urban or rural). The software program responds by giving the predicted 85th percentile speed which corresponds to the specified input parameters. It is to be noted that both models perform best when the input parameters are within the applicable ranges used in developing the models. Specifying input value(s) outside the specified applicable range(s) compels the models to extrapolate instead of interpolate. In this case, the reliability of model prediction may be questionable.

Both software programs and the corresponding instruction manuals have been provided to the KDOT project monitor in a separate document.

*CHAPTER 5*

# CONCLUSIONS, RECOMMENDATIONS AND IMPLEMENTATION PLAN

## 5.1 CONCLUSIONS

Speed zoning in the United States is based on the principle of setting reasonable and safe speed limits as close as practicable to the speed at or below which 85 percent of drivers are traveling. Speed studies, which are expensive and time consuming, must be conducted to determine the 85th percentile speeds. To implement speed limits that are different from the 85th percentile speeds, the engineer must rely upon professional judgement and consider other factors that may have some affect on vehicle speeds.

A comprehensive literature review and survey was conducted to identify those factors that may have some affect on vehicle speeds. Based on the results of the literature review and the availability of data for Kansas highways, 27 variables were identified as possibly affecting speeds on rural state highways. Thirty-two variables were identified as possibly affecting speeds on urban state highways. Speed data and data for the potential explanatory variables were collected for a total of 539 state highway sections (186 rural and 353 urban sections). These basic data sets were later supplemented by additional data provided by the KDOT Bureau of Traffic Engineering for use in developing refined neural network models of highway speeds.

Two approaches were used to develop and test models to predict speeds on rural and urban state highways based on roadway characteristics and adjacent development patterns. The first approach was based on models in the form of multiple linear regression equations. The second approach employed artificial neural networks (ANNs) to predict highway speeds. The general conclusions drawn from these two approaches are presented below.

### 5.1.1 Regression Analysis

The Statistical Analysis System (SAS) was used to analyze the CANSYS data used in this study. Correlation analyses were performed to identify any significant relationships between 85th percentile speeds and the potential explanatory variables. Twenty-three variables were found to be significantly correlated with the 85th percentile speed on rural highways and nine variables were found to be significantly correlated with the 85th percentile speed on urban highways.

The SAS stepwise regression procedure was used to develop models of 85th percentile speed as a function of the various explanatory variables. The Stepwise Procedure identified 24 candidate rural models and 30 candidate urban models with regression coefficients that were significant at the 0.85 significance level. None of the candidate models was entirely satisfactory in terms of predicting the 85th percentile speeds on rural and urban highways within ± 5 mph. However, a number of preliminary models are presented. The regression models should be used with caution. In cases where preliminary, planning-level estimates of highway speeds are needed, the regression models could be useful.

While none of the candidate regression models was entirely satisfactory in terms of accurately predicting the 85th percentile speeds on rural and urban highways, the regression analysis was beneficial in that it identified those factors that may be important in predicting 85th percentile speeds. The identification of these factors was an important contribution to the neural network modeling phases of the research effort.

The Stage 2 neural network models developed in this research effort do a better job than either the Stage 1 ANN models or the regression models in predicting speeds from the database available for this project and are recommended for use in those situations where more precise estimates of highway speeds are needed.

### 5.1.2. Neural Network Analysis

The rural highway ANN model recommended for use by the Department in predicting $85^{th}$ percentile speeds is the Stage 2 *5-5-1 model* that requires the following 4 input parameters: shoulder width, shoulder type (P/C or G/T), ADT and percent no passing zone. The model is expected to predict 85th percentile speeds with an average degree of accuracy of about 96% (i.e. ±4% average deviation from the actual value).

For urban highways, two Stage 2 ANN models can be used. Model 1 can be used when the posted speed limit in not known. Model 2 is recommended when the posted speed limit is known. Otherwise, both models utilize the same input parameters (i.e., parking type, lane type and area density type). Models 1 and 2 are expected to predict the 85th percentile speed with an average degree of accuracy of about 80% and 91%, respectively (i.e., ±20% and ±9% average deviation from the actual value for models 1 and 2, respectively).

## 5.2 RECOMMENDATIONS

This study identified a number of issues that suggest several potentially fruitful research directions, as outlined below.

The regression models developed in the research suffer from a number of fundamental shortcomings. First, the R-square values are not particularly high. This is probably attributable to the high degree of variability observed in the CANSYS and highway speed databases. Secondly, the models have a large number of variables and are not as tractable as desired. However, because all the data needed to use the models are available in the CANSYS database, the relatively large number of variables in the models was not considered to be a significant disadvantage. Finally, the models are " generic " models in that they represent all rural and urban state highways. For example, the models are not capable of distinguishing between the different functional classes of highways.

Unfortunately, the number of observations in the data available for this research effort was not large enough to develop and evaluate regression models based on functional class, number of lanes and other subcategories of state highways.

As a result, the regression models developed in this research should be used with caution. In cases where preliminary, planning-level estimates of highway speeds are needed, the regression models could be useful.

The Stage 2 neural network models developed in this research effort do a better job than either the Stage 1 ANN models or the regression models in predicting speeds from the database available for this project and are recommended for use in those situations where more precise estimates of highway speeds are needed.

Future efforts should focus on compiling a more extensive database so that models based on categories such as functional class and number of lanes can be evaluated. In this regard, KDOT may wish to consider incorporating the department's speed study data into the CANSYS database. This could be similar to the way current and historical traffic volume and accident data are included in CANSYS. As a minimum, KDOT should consider including control section identification numbers in their speed study data files. This would greatly facilitate future speed zone modeling efforts. In light of the recent changes in the speed limit on Kansas highways it is likely that additional speed studies will be needed.

In regard to the Stage 2 ANN-based models, it is highly recommended that these models be validated against new sets of data that have never been used in either training or testing the models. Accordingly, re-training of the network model(s) on previous and new data may be warranted if the prediction accuracy of either model significantly deviates from the accuracy measures reported in this study (i.e., 96% and 91% average degree of accuracy for rural and urban highways, respectively).

Further improvements and modifications to the urban network model may justify a more detailed and focused ANN research study. Also, periodic re-training of the developed networks (with updated input and output data) is recommended in order for these models to implicitly take into consideration any changes in driver behavior and/or traffic regulations that may occur over time.

## 5.3 IMPLEMENTATION PLAN

The Stage 2 ANN software programs developed in this study can be used to predict the $85^{th}$ percentile speeds on urban and rural state highways using a small number of readily available roadway- and traffic-related parameters. Use of these models could greatly reduce the need for costly and time consuming speed studies to determine the $85^{th}$ percentile speeds on rural and urban state highways. The recommended models can also be used to predict the potential effects of changes in certain roadway- and traffic-related parameters on $85^{th}$ percentile speeds.

In order to fully assess the validity of the ANN models developed in this study, the Department should continue its efforts to compile an extensive, historical database that can be used to test and re-train the Stage 2 models on a regular basis.

With regards to future data collection and data management efforts, the Bureau of Traffic Engineering should take an active role in KTRAN Project KSU-99-3 (*Long Range Plan to Improve Quality, Use and Understanding of the Traffic Databases Maintained by KDOT*). This KTRAN project is scheduled to begin in August 1998.

## REFERENCES


American Association of State Highway and Transportation Officials (AASHTO) (1994). *A Policy on Geometric Design of Highways and Streets*. AASHTO, Washington, D.C.

Avery, E.V. (1960). "Effect of Raising Speed Limits on Urban Arterial Streets." *Highway Research Board Bulletin 244*. pp. 88-97.

Basheer, Imad A. and Y.M. Najjar (1996). "A Neural Network-Based Distress Model for Kansas JPCP Longitudinal Joints." *Intelligent Engineering Systems Through Artificial Neural Networks*. Vol.6, pp.983-988.

Beatty, R.L. (1972). "Speed Analysis of Accidents on Interstate Highways." *Public Roads*. pp. 89-102.

Cirillo, J.A. (1968). "Interstate System Accident Research Study II. Interim Report II." *Public Roads*. pp 71-75.

Coleman, Fred, III. (1995). *Determination of a Discriminant Function as a Prediction Model for Effectiveness of Speed Zoning in Urban Areas*. Michigan State University.

Dudek, C.L. and G.L. Ullman (1987). *Speed Zoning and Control.* Research Report 334-2F. Texas Transportation Institute, College Station, TX.

Florida Department of Transportation (1980). *Speed Zoning for Highways, Roads and Streets in Florida*. Tallahassee, Florida.

Hassoun M.H. (1995). *Fundamentals of Artificial Neural Networks*. MIT Press, Cambridge, MA.

*Highway Capacity Manual* (HCM) (1994). Special Report No. 209. Transportation Research Board, Washington, D.C.

Institute of Transportation Engineers (ITE) (1993). *Speed Zone Guidelines: A Proposed Recommended Practice*. Publ. No. RP-024. ITE, Washington, D.C.

Jarvis, J.R. and C.J. Hoban (1989). "Development of a Speed Zoning Knowledge-Based System." *Civil Engineering Systems*. Vol. 6, No. 1-2, pp 71-79.

Kessler, W.L. (1959). "The Effect of the Speed Zone Modifications Occasioned by the Illinois Speed Law ." *Traffic Engineering*. Vol. 29.

KDOT (1994). CANSYS Data Collection and User Manual. Bureau of Transportation Planning, Topeka, KS.

Krammes, R.A., K. Fitzpatrick, J. D. Blaschke and D. B. Fambro (1996). *Speed - Understanding Design, Operating, and Posted Speed.* Research Report No. 1465-1. Texas Transportation Institute, College Station, TX.

McCoy, P.T., B.A. Moen, G. Pesti and M. Moussavi (1993). *Evaluation of Lower Speed Limits on Urban Highways.* Department of Civil Engineering, University of Nebraska-Lincoln.

Najjar, Y.M., I.A. Basheer and R. McReynolds (1996a). "Neural Modeling of Kansas Soil Swelling." *Transportation Research Record No. 1526*, pp. 14-19

Najjar, Y. M., I.A. Basheer and W.A. Naouss (1996b). "On the Identification of Compaction Characteristics by Neuronets." *Computers and Geotechnics* Vol. 18, No. 3, pp. 167-187.

National Committee on Uniform Traffic Laws and Ordinances (1968). *Uniform Vehicle Code and Model Traffic Ordinance.* Section 11 - 803.

Parker, M.R. (1985). *Guidelines for Establishing Speed Zones: Synthesis of Speed Zoning Practices.* Technical Report FHWA/RD-85/096. USDOT, FHWA, Washington, D.C.

Parker, M.R. (1992). *Effects of Raising and Lowering Speed Limits.* Technical Report FHWA-RD-92-084. USDOT, FHWA, Washington, D.C.

Poe, C.M. and J.M. Mason, Jr. (1995). "Geometric Design Guidelines to Achieve Desired Operating Speed on Urban Streets." In *1995 Compendium of Technical Papers.* 65th Annual Meeting, Institute of Transportation Engineers. pp 70-74.

SAS Institute Inc. (1989). *SAS/STAT User's Guide.* Release 6.03 Edition. Cary, NC.

Simpson P.K. (1990). *Artificial Neural Systems: Foundations, Paradigm, Applications, and Implementations.* Pergamon Press, New York, NY.

Solomon, D. (1964). *Accidents on Main Rural Highways Related to Speed, Driver and Vehicle.* Bureau of Public Roads, Traffic Systems Research Division. U.S. Government Printing Office, Washington, D.C.

Transportation Research Board (TRB) (1984). *A Decade of Experience.* Special Report 204. Transportation Research Board, Washington, D.C.

West, L.B., Jr. and J.W. Dunn (1971). "Accidents, Speed Deviation, and Speed Limits."

*Traffic Engineering*. pp 52-55.

Zupan J. and Gasteiger J. (1993). *Neural Networks for Chemists: An Introduction*. VCH, New York, NY

# APPENDIX
# DEFINITIONS OF CANSYS DATABASE VARIABLES

**AADT**: Current annual average daily traffic.

**AC: ACCESS CONTROL.**

| | |
|---|---|
| 0 | - No Access Control. |
| 1 | - Partial Access Control - Preference has been given to through traffic movement. In addition to interchanges, there may be some crossings at grade with public roads, but direct private driveway connections have been minimized. |
| 2 | - Full Access Control - Preference has been given to through traffic by use of interchanges with selected public roads. At grade and private connections are prohibited. |

**ACC: NUMBER OF ACCIDENT, 5-Yr.** This field contains the total number of accidents that have occurred in the control section during the five years period.

**CACC: NUMBER OF ACCIDENT CURRENT** This field contains the total number of accidents that have occurred in the control section during the current year.

**COM: PERCENT HEAVY COMMERCIAL TRAFFIC** The percentage (to the nearest whole percent) of the AADT that is buses and trucks. All 'two axle, four tire' pickups and panel trucks are excluded.

**FC: FUNCTIONAL CLASSIFICATION.**

| RURAL | | URBAN | |
|---|---|---|---|
| 11 | - Interstate | 51 | - Interstate |
| 12 | - Other Principal Arterials | 52 | - Freeways and Expressways |
| 13 | - Minor Arterials | 53 | - Other Principal Arterials |
| 21 | - Major Collectors | 54 | - Minor Arterials |
| 22 | - Minor Collectors | 61 | - Collectors |
| 31 | - Local roads | 71 | - Local Streets |

**GSI: # OF GRADE-SEPARATED INTERCHANGES**. The number of grade separated interchanges. Beginning interchange is included, ending is not included unless it is the route terminus.

**INTO: NUMBER OF INTERSECTIONS OTHER OR NO CONTROL** The number of at-grade intersections on the route being inventoried that are not controlled by either a signal or a stop sign. The beginning intersection is included in this count, but the ending intersection is excluded except when it is the route terminus. A flashing yellow signal shall be considered as "OTHER OR NO

CONTROL."

**INTS: NUMBER OF SIGNALIZED INTERSECTIONS** The number of intersections with signal lights of the subsection in question. The beginning intersection and all interior intersections with signals are included in this count. The ending intersection is excluded except when it is the route terminus.

**IRI: INTERNATIONAL ROUGHNESS INDEX CURRENT**: The South Dakota profilometer is used to collect wheel path profile data in inches/mile.

| | | |
|---|---|---|
| SMOOTH | < | 105 |
| MEDIUM | | 105 - 164 |
| ROUGH | > | 164 |

**ITOP: # OF INTERSECTIONS WITH STOP SIGNS**: The number of intersections with a stop sign controlling the route being inventoried. A continuous flashing red light shall be counted as a stop sign control. Beginning intersection is included, ending intersection is not included unless it is the route terminus.

**LC: LANE CLASS**

| | | | |
|---|---|---|---|
| 1 | - Two Lane Undivided | 8 | - Three Lane |
| 2 | - Four Lane Undivided | 9 | - Five Lane |
| 3 | - Four Lane Undivided | 10 | - One Lane, One Way |
| 4 | - Six Lane Undivided | 11 | - Two Lane, One Way |
| 5 | - Six Lane Divided | 12 | - Three Lane, One Way |
| 6 | - Eight Lane and Over Undivided | 13 | - Four Lane, One Way |
| 7 | - Eight Lane and Over Divided | 14 | - Two Lane Divided |

**LUO: LAND USE, OUTSIDE and LUI: LAND USE, INSIDE**

| | |
|---|---|
| 1 | - Grassland |
| 2 | - Cultivated upland |
| 3 | - Cultivated bottom land |
| 4 | - Irrigated |
| 5 | - Mineral development (oil, gas, etc.) |
| 6 | - Current business district |
| 7 | - Outlying business district |
| 8 | - Residential |
| 9 | - Industrial |
| 10 | - Public (park, game, reserve, etc.) |
| 11 | - Highway median |

"INSIDE/OUTSIDE" is "LEFT/RIGHT" on undivided facilities.

**LW: LANE WIDTH** Calculated value of lane width in feet.

**MT: MEDIAN TYPE**

| | |
|---|---|
| 0 | - None, double or single yellow stripes |
| 1 | - Painted for turning lanes |
| 2 | - Raised median (no curbs) |
| 3 | - Raised median with curbs |
| 4 | - Raised median with turning lanes |
| 5 | - Other raised median |
| 6 | - Depressed median |
| 7 | - Other |
| 8 | - Barrier |

**MW: AVERAGE MEDIAN WIDTH.** Coded with the average width of the median for the subjection. It is measured from the back of the curbs for raised medians and for the inside driving lane when the median is depressed. *The median includes inside shoulder.* For divided facilities the median width is halved. When the median is greater than 100 ft, Code '99'.

**PDI: POPULATION DENSITY, INSIDE and PDO: POPULATION DENSITY, OUTSIDE.** Density of population per square mile.

| | | | | |
|---|---|---|---|---|
| 1 | - Rural, Light density | 1 | thru | 9 |
| 2 | - Rural, Moderate density | 10 | thru | 39 |
| 3 | - Rural, Heavy density | 40 | thru | 999 |
| 4 | - City, Light density | 1 | thru | 2,499 |
| 5 | - City, Moderate density | 2,500 | thru | 7,499 |
| 6 | - City, Heavy density | 7,500 | thru | More |
| 7 | - Divided highway median. | | | |

**PI: PARKING, INSIDE and PO: PARKING, OUTSIDE** currently this field is only filled for city subsections. All rural sections are therefore coded zero which can be considered valid due to no parking lanes along rural highways.

| | |
|---|---|
| 0 | - No parking |
| 1 | - Parallel parking (approx. 8 ft) |
| 2 | - Diagonal parking (approx. 17 ft) |
| 3 | - Center parking |
| 4 | - Center plus parallel parking |
| 5 | - Center plus diagonal parking |
| 6 | - Width prohibits parking |

"INSIDE/OUTSIDE" is "LEFT/RIGHT" on undivided facilities.

**PSL: POSTED SPEED LIMIT.** This is the speed limit posted along the route.

**RID: RIDEABILITY**. This data element calculates the roughness of the roadway.

5 - Excellent, equivalent to a new roadway
4 - Very good
3 - Good
2 - Fair
1 - Poor
0 - Very poor

**RP: RESTRICTED PASSING.** Percent (to the nearest whole percent) of the control section with a sight distance less than 1,500 feet. Zero is coded for all sections on divided facilities. *Restricted Passing = 1.625* (passing restrictions, profile) + (passing restrictions, bridges).*

The value can not exceed '100'. If calculated value is greater than '100', it is set equal to '100'.

**SL: SUBSECTION LENGTH**. Length of the subsection in miles to thousandths.

**SS: SIDE SLOPE**. This field is coded with the change in width for a one foot change in height. Side slope is from the outside edge of the shoulder to the ditch. Side slope does not apply for subsection with curb and gutter, thus a zero is coded for this condition.

6 - 6 tol
5 - 5 tol
4 - 4 to 1
3 - 3 to 1
2 - 2 to 1
1 - 1 to 1 or greater
0 - Curb and gutter present

**SSD: SUBSTANDARD STOPPING SIGHT DISTANCE.** Each field holds the number of vertical curves with stopping sight distance in the given range. "Distance" is the nomogram stopping sight distance for a given speed.

| Design Speed (mph) | | | |
|---|---|---|---|
| Range | Minimum | Desirable | Distance (ft.) |
| 0 - 225.00 | 35 | | 225 |
| 225.01 - 250.00 | | 35 | 250 |
| 250.01 - 275.00 | 40 | | 275 |
| 275.01 - 325.00 | 45 | | 325 |
| 325.01 - 400.00 | 50 | | 400 |
| 400.01 - 450.00 | 55 | | 450 |
| 450.01 - 475.00 | | 50 | 475 |
| 475.01 - 525.00 | 60 | | 525 |
| 525.01 - 550.00 | 65 | | 550 |
| 550.01 - 625.00 | 70 | | 625 |
| 625.01 - 650.00 | | 60 | 650 |
| 650.01 - 725.00 | | 65 | 725 |
| 725.01 - 850.00 | | 70 | 850 |

(Derived from the KDOT design manual, 1984 and the 1990 AASHTO green book, P.120)

**ST: SURFACE TYPE**. The surface type in this field is converted from the research layer information. The table for the conversion that takes place follows the codes for this field.

0 - No surface, traffic riding on the base
1 - Bituminous surface treated
2 - Mixed surface
3 - Bituminous mixed overlay (composite)
4 - Asphaltic concrete
5 - Asphaltic concrete overlay
6 - Portland cement concrete
7 - Brick
8 - Gravel and/or stone
9 - Graded and drained earth

**STI: SHOULDER TYPE, INSIDE and STO: SHOULDER TYPE, OUTSIDE**. Shoulder stabilization codes are associated with each of shoulder types listed below. The values are:

1 - Unstabilized
2 - Stabilized
3 - Curb and Gutter

The values of stabilization codes in certain programs have different meanings than the "1", "2", and "3" listed above.

**SW: SURFACE WIDTH.** This width is measured in whole feet. It indicates additional width and parking. It does not include the shoulders. For roadways with curbs, it is the back-to-back width.

*Surface width = (# of thru lanes) (lane width) + (additional width) + (parking width)*

**SWI: SHOULDER WIDTH, INSIDE and SWO: SHOULDER WIDTH, OUTSIDE**
For roadways with shoulders, the shoulder width is coded in whole feet. When curbs are present, shoulder width does not apply and "0" is coded into these fields. "INSIDE/OUTSIDE" is "LEFT/RIGHT" on undivided facilities.

**UL: URBAN LOCATION.** For urban areas, enter the code that represents the predominant characteristics of the land area for the roadway section. If a roadway section is contained in two different areas then use the lower code for the section.

0 - Not applicable, not in urban area.

1 - Central Business District (CBD) - The traditional commercial and retail trade center in the center city of an urban area. An area having very high land value because of intense concentration of retail trade, office space, cultural, and service activities.

2 - High Density Business/Commercial Center (Excluding CBD) - One or more centers of business and/or commercial activities within the urban area (or a cluster of two or more adjacent smaller centers).

3 - Low Density Commercial - That portion of an urban area that is not the CBD or a high density business/commercial center and contains a lower density of business, industrial, warehousing, service, and strip development or a wide mixture/variety of such uses.

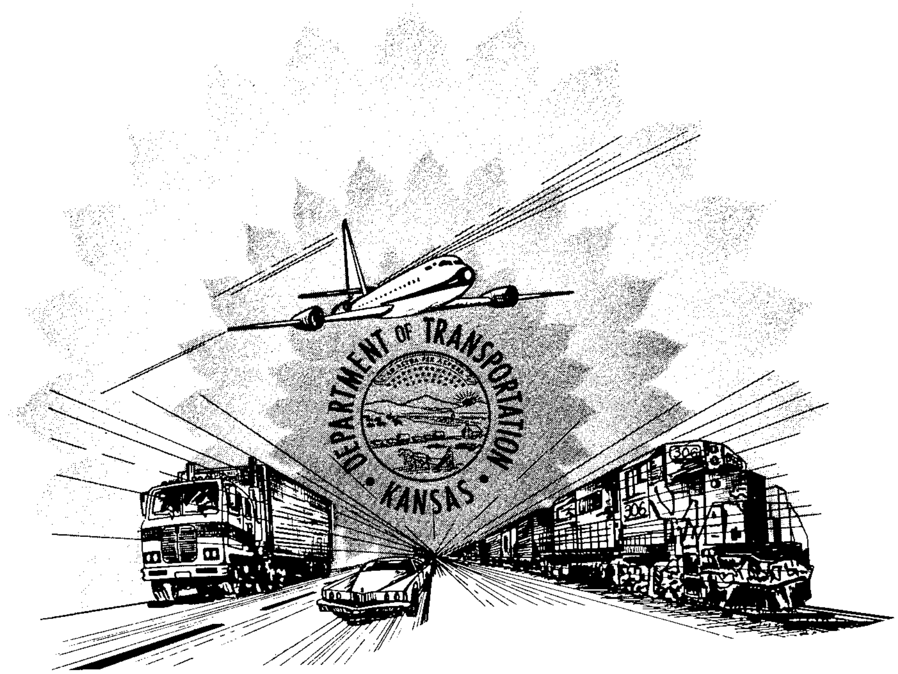
4 - High Density Residential - That portion of an urban area in which the major land use is residential and has a density of 5,000 or more persons per square mile (2,000 persons/$KM^2$).

5 - Low Density Residential - That portion of an urbanized area in which the major land use is residential and has a density less than 5,000 persons per square mile (2,000 persons/$KM^2$). The development density is greater than or equal to one dwelling unit per acre (250 dwellings/$KM^2$).

6 - Other, including undeveloped land and residential areas having a density of less than one dwelling unit per acre.

# K - TRAN

**KANSAS TRANSPORTATION RESEARCH
AND
NEW - DEVELOPMENTS PROGRAM**




A COOPERATIVE TRANSPORTATION RESEARCH PROGRAM BETWEEN:

KANSAS DEPARTMENT OF TRANSPORTATION

THE KANSAS STATE UNIVERSITY

THE UNIVERSITY OF KANSAS