Normal acceleration characteristics of passenger cars provide valuable information for various aspects of transportation engineering, including use in the geometric designs of intersections, freeway ramps, passing lanes, acceleration lanes, turning bays, or in the development of traffic simulation or fuel consumption models.

However, the data used in previous studies are either outdated or characterized by limitations from data collecting technologies and methodologies. Therefore, the study results may not represent the accelerating behavior of current drivers or modern vehicles in terms of time frame, prediction accuracy and study scopes.
This thesis presents an acceleration process study based on a current database collected by in-vehicle Global Positioning System (GPS) technology over an extended period of time. The acceleration trip data used in this study provides a better representation of real world natural driving behaviors. The author of this thesis developed a third-order polynomial model to describe speed profiles of the leading acceleration process at signalized intersections on arterial roadways. The model also investigates roadway physical feature effects and driver-vehicle effects on the acceleration process. The roadway physical features include horizontal alignments, intersection layouts, and number of through travel lanes. The driver-vehicle effect is modeled as the random effect while the study interest is on the driver population instead of the sample drivers in the data.
Normal Acceleration Characteristics of the Leading Vehicle in A Queue at Signalized Intersections on Arterial Streets

by
Hong Zhu

A THESIS

submitted to

Oregon State University

in partial fulfillment of the requirements for the degree of Master of Science

Presented January 9, 2007
Commencement June 2008

APPROVED:

Major Professor, representing Civil Engineering

Head of the Department of Civil, Construction, and Environmental Engineering

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

___________________________________________
Hong Zhu, Author
ACKNOWLEDGEMENTS

I would like to express my deepest thanks to all of my committee members and everyone who helped make this thesis succeed.

Dr. Katharine Hunter-Zaworski, thank you for your continuous encouraging, cheering, guiding, and believing in me to achieve my goal from the day one, with that I can go as far as I could. I am also grateful to Dr. Virginia Lesser for giving me very valuable insights and guidance of statistical modeling techniques while I was confused. I express sincere appreciation to Dr. Karen Dixon for her mentorship, kind support, trenchant critiques and remarkable patience while doing this project.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Overview</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Study Scope and Study Objectives</td>
<td>2</td>
</tr>
<tr>
<td>2  LITERATURE REVIEW OF ACCELERATION MODELS</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Overview</td>
<td>4</td>
</tr>
<tr>
<td>2.1.1 Kinematics Acceleration Models</td>
<td>6</td>
</tr>
<tr>
<td>2.1.2 Dynamics Acceleration Models</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Acceleration Characteristics at Intersections</td>
<td>21</td>
</tr>
<tr>
<td>2.2.1 Stop Controlled Intersections</td>
<td>21</td>
</tr>
<tr>
<td>2.2.2 Signalized Intersections</td>
<td>21</td>
</tr>
<tr>
<td>2.3 Summary</td>
<td>23</td>
</tr>
<tr>
<td>3  DATA COLLECTION AND EXTRACTION METHODOLOGY</td>
<td>24</td>
</tr>
<tr>
<td>3.1 Data Collection and Sampling Method</td>
<td>24</td>
</tr>
<tr>
<td>3.2 Comparison of Data Collection Methods</td>
<td>27</td>
</tr>
<tr>
<td>3.3 Data Extraction Criteria</td>
<td>28</td>
</tr>
<tr>
<td>3.4 Method for Separating Leading and Following Acceleration Trips</td>
<td>30</td>
</tr>
<tr>
<td>3.5 Sensitivity Study of Different Speed Limit Groups</td>
<td>36</td>
</tr>
<tr>
<td>4  MODEL DEVELOPMENT</td>
<td>38</td>
</tr>
<tr>
<td>4.1 Data Structure of Speed Profiles in the Acceleration Process</td>
<td>38</td>
</tr>
<tr>
<td>4.2 Linear Mixed-Effects Modeling Method</td>
<td>42</td>
</tr>
<tr>
<td>4.3 Model Specification</td>
<td>44</td>
</tr>
<tr>
<td>4.4 Model Estimation and Interpretation</td>
<td>48</td>
</tr>
<tr>
<td>4.4.1 Model Estimation Results</td>
<td>48</td>
</tr>
<tr>
<td>4.4.2 Model Interpretation</td>
<td>50</td>
</tr>
<tr>
<td>4.5 Model Comparison</td>
<td>53</td>
</tr>
<tr>
<td>4.5.1 Introduction of Three Ordinary Regression Models</td>
<td>53</td>
</tr>
<tr>
<td>4.5.2 Model Comparison</td>
<td>57</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Constant Acceleration Model</td>
</tr>
<tr>
<td>2</td>
<td>Two Phase Acceleration Model</td>
</tr>
<tr>
<td>3</td>
<td>Linearly-Decreasing Acceleration Model (Acceleration vs. Speed)</td>
</tr>
<tr>
<td>4</td>
<td>Linearly-Decreasing Acceleration Model (Acceleration vs. Time)</td>
</tr>
<tr>
<td>5</td>
<td>Nonlinear Polynomial Acceleration Model</td>
</tr>
<tr>
<td>6</td>
<td>Chowdhury’s Two-Phase Linear-Decreasing Acceleration Model</td>
</tr>
<tr>
<td>7</td>
<td>Bearing Ranges of Driving Directions</td>
</tr>
<tr>
<td>8</td>
<td>Schematics of Distinguishing Queue Leaders</td>
</tr>
<tr>
<td>9</td>
<td>Average Speed Profiles Comparison</td>
</tr>
<tr>
<td>10</td>
<td>Scatter Plot Of Speed Profiles</td>
</tr>
<tr>
<td>11</td>
<td>Speed Profiles Scatter Plot vs. Fitted Model</td>
</tr>
<tr>
<td>12</td>
<td>Speed Profiles: Straight Segment vs. Horizontal Curve</td>
</tr>
<tr>
<td>13</td>
<td>Driver-Vehicle Effect</td>
</tr>
<tr>
<td>14</td>
<td>Normal Probability Plot Of Fixed-Effects</td>
</tr>
<tr>
<td>15</td>
<td>Normal Probability Plot of the Random-Effect</td>
</tr>
<tr>
<td>16</td>
<td>Example of Predicted Speed Profile and Acceleration Profile</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Acceleration Models Overview ................................................................. 4</td>
</tr>
<tr>
<td>2</td>
<td>Study Driver and Vehicle Profile ............................................................... 27</td>
</tr>
<tr>
<td>3</td>
<td>Data Collection Methods Comparison ............................................................ 28</td>
</tr>
<tr>
<td>4</td>
<td>Acceleration Trip Data Example ................................................................. 29</td>
</tr>
<tr>
<td>5</td>
<td>Bearing Ranges of Driving Directions .......................................................... 31</td>
</tr>
<tr>
<td>6</td>
<td>Longitudinal Data Structure .......................................................................... 40</td>
</tr>
<tr>
<td>7</td>
<td>Longitudinal Data Summary ............................................................................ 40</td>
</tr>
<tr>
<td>8</td>
<td>Comparison: Fixed vs. Random Driver-Vehicle Effects ....................................... 41</td>
</tr>
<tr>
<td>9</td>
<td>Regression Results of the Mixed-Effects Model ............................................... 49</td>
</tr>
<tr>
<td>10</td>
<td>95% Confidence Intervals of Coefficients ....................................................... 50</td>
</tr>
<tr>
<td>11</td>
<td>Model Comparison: Model I, II III, and IV ....................................................... 57</td>
</tr>
<tr>
<td>12</td>
<td>Example of Predicted Speed Profile Data ....................................................... 64</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

1.1 Overview

Vehicle acceleration characteristics provide critical information for the geometric design of intersections, freeway ramps, passing lanes, acceleration lanes, and turning bays. Researchers also use acceleration models in traffic simulation and fuel consumption models.

The acceleration process only occurs for short time periods throughout the natural driving process. This makes it difficult to capture natural driving behaviors and to investigate the influence of road physical features on the acceleration process. Historically, researchers performed various vehicle acceleration characteristics studies dating back decades, and developed various types of models to describe the acceleration process. However, the data used in previous studies are either outdated or characterized by limited data collection techniques and methodologies. Therefore, the study results may not be able to represent the natural driving behavior of current drivers and modern vehicles in terms of time frame, prediction accuracy, and study scopes.

This thesis presents an acceleration process study based on a current database produced by in-vehicle Global Positioning System (GPS) technology over an extended period of time. Speed profiles developed from second-by-second speed data extracted from the database are able to accurately represent the entire acceleration process. The author developed a third-order polynomial model to describe speed profiles of the acceleration process at signalized intersections on arterial roadways. By incorporating variables representing the effects of physical roadway
features, this model also reveals the influence on the speed profiles during acceleration for roadway intersection features, including horizontal alignments, intersection layouts, and number of lanes. Moreover, this study provides the first effort to explore the driver-vehicle effect on speed profiles for the acceleration process at signalized intersection locations.

1.2 Study Scope and Study Objectives

It is reported that vehicles may accelerate differently while being queue leaders compared to following vehicles in a queue (Long, 2000). The queue leader is also called as the leading vehicle in a queue. The acceleration trips of queue leaders are called as leading acceleration trips. The scope of this study is focused on normal acceleration processes of leading vehicles while accelerating from a stopped condition at signalized intersections on arterial roadways with a speed limit of 72 km/h (45 mph). It covers one of the major and important types of acceleration processes for our roadway system. This study utilizes passenger-vehicle data only and does not include public vehicles, trucks, or similar vehicle configurations. The passenger car can be expected to accelerate differently than other types of vehicles such as recreational vehicles, buses, and trucks.

Generally, there are four major variables involved in the acceleration process: speed, time, rate of acceleration, and distance. In this study, speed and time are the original information collected from the field. As a result, these two variables exhibit better accuracy than secondary data, such as the rate of acceleration, which is ultimately calculated using the speed and time. Distance, as acquired using in-vehicle GPS technology, is an accumulative variable, and can therefore result in confounding errors thereby decreasing data accuracy. So, this thesis focuses on using the speed vs. time relationship, which is the speed profile, to model the acceleration process.
The objectives of this study are:

1. Develop data extraction methodologies by incorporating different information resources in order to distinguish the data representing the normal acceleration process of leading vehicles from that of following vehicles in queues.

2. Investigate the hypothesis that the acceleration process can be modeled with good fit by the third-order polynomial relationship between the speed and time.

3. Explore what influences physical roadway features can have on the acceleration process.

4. Evaluate the influences caused by the driver population in general by modeling driver-vehicle subjects as the random effect.

The contribution of this study is to provide the initial effort exploring the potential influence on speed profiles for vehicles accelerating from a stopped condition at signalized intersections. The study specifically investigates the influence of roadway physical features, such as horizontal curvature alignments, intersection skewness, and number of through travel lanes. A second contribution of this effort is to investigate the driver-vehicle effect on speed profiles, and to test the theory of applying the polynomial model with relatively simple functional form to describe this relationship. The real world data set used in this study provides the opportunity to capture and model the actual acceleration behavior of known drivers in their natural driving environment.
2 LITERATURE REVIEW OF ACCELERATION MODELS

2.1 Overview

Historical research regarding vehicle acceleration analysis has focused on two types of acceleration characteristics: the acceleration capability and normal acceleration process. According to the categorization by Rakha et al. (2004), previous research has resulted in two types of models that describe acceleration characteristics—vehicle kinematics acceleration models and vehicle dynamics acceleration models. Table 1 shows the acceleration models introduced in this study.

<table>
<thead>
<tr>
<th>Category</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinematics Acceleration Models</td>
<td>Constant Acceleration Model</td>
</tr>
<tr>
<td></td>
<td>Two Phase Acceleration Model</td>
</tr>
<tr>
<td></td>
<td>Linearly-Decreasing Acceleration Model</td>
</tr>
<tr>
<td></td>
<td>Polynomial Acceleration Model</td>
</tr>
<tr>
<td>Dynamics Acceleration Models</td>
<td></td>
</tr>
</tbody>
</table>

In a vehicle acceleration capability study, the vehicle is assumed to accelerate at a maximum acceleration rate throughout the acceleration process. The acceleration is only restricted by the vehicle’s dynamic capabilities. This type of study is important for the automobile manufactures as automobile magazines publish the vehicle performance specifications for different types of vehicles based on the auto road test. The acceleration capability is also useful in accident investigation analysis and for understanding situations that require drivers to apply maximum acceleration, such as executing a passing maneuver.
Searle (1999) and Rakha et al. (2004) developed dynamics acceleration models to study acceleration performance. Vehicle dynamics acceleration models predict the acceleration rate based on equations that reflect the vehicle’s physical motion characteristics. This type of model takes into account a vehicle’s engine power output and power loss from various resistance factors. Vehicle dynamics models provide a better prediction of a vehicle’s maximum acceleration capability than they do for normal acceleration characteristics (Searle, 1999).

For typical driving conditions, vehicles operate in normal acceleration conditions rather than at maximum acceleration. The Institute of Transportation Engineers (ITE) *Traffic Engineering Handbook* (1999) illustrates examples of normal acceleration behavior as a vehicle starts after a traffic signal turns green and a vehicle performs a passing maneuver on a four-lane highway. The handbook also indicates that the maximum acceleration is seldom used in normal driving. There are many examples where normal acceleration estimates are needed. These examples include intersection design, acceleration lane design, and traffic simulation (Akcelik & Besley, 2002).

Vehicle kinematics acceleration models generally predict the acceleration rate based on mathematical relationships derived through statistical analysis of empirical data. Generally, the kinematics acceleration models include:

- Constant acceleration;
- Two phase acceleration;
- Linearly-decreasing acceleration; and
- Polynomial acceleration.

Typically, a full normal acceleration process begins with the vehicle at a complete stop with zero speed and zero acceleration. After the vehicle accelerates through time and distance, the process ends when the vehicle reaches the relative constant free flow speed. This free flow speed is often
referred to as the equilibrium speed or the desired speed. Based on the ITE Traffic Engineering Handbook (1999), in the normal acceleration condition, drivers apply approximately 65 percent of the maximum acceleration capability.

Field data collection for the normal acceleration process varies significantly from the data collection of the maximum acceleration, since there are more variations in normal driving conditions. A few examples of these variations are:

- Vehicle acceleration behavior of queue leaders after the signal changes to green;
- Vehicle acceleration in a platoon;
- Vehicle acceleration with through maneuvers;
- Vehicle acceleration during a turning maneuver;
- Vehicle acceleration during a passing maneuver;
- Vehicle acceleration from speed of zero; and
- Vehicle acceleration from a non-zero speed.

Acceleration characteristics vary in the above scenarios. This study focuses on the first item, the acceleration of queue leaders after the signal changes to green.

2.1.1 Kinematics Acceleration Models

2.1.1.1 Constant Acceleration Model
The constant acceleration model (see Figure 1) assumes that vehicles keep accelerating with the same acceleration rate through the entire acceleration process (Akcelik & Biggs, 1987; Bham & Benekohal, 2002; Long, 2000). This model also reflects a linearly increasing relationship between speed and time, and the slope of the line represents the rate of acceleration.
Bham and Benekohal (2002) indicated two methods to calculate the value of the acceleration rate, including:

1. The fundamental physics formula of motion under constant acceleration,

\[ v_f = v_i + at \]  \hfill (2-1)  
\[ x = v_it + \frac{1}{2}at^2 \]  \hfill (2-2)

where,

\( t \) = acceleration time (second),
\( v_f \) = speed at time \( t \) (m/s),
\( v_i \) = initial speed (m/s),
\( a \) = constant acceleration (m/s\(^2\)), and
\( x \) = travel distance at time \( t \) (m).
2. Linear regression analysis based on field data by defining time as the independent variable and speed as the dependent variable.

The constant acceleration model has been widely used in various design processes (AASHTO, 2004) and other study purposes due to the extreme simplicity and convenience of the practical application of the model. However, several studies (Bham & Benekohal, 2002; Long, 2000; Searle, 1999) demonstrated that the assumption of a constant acceleration rate is unrealistic for through gear acceleration. This type of acceleration includes most of the acceleration processes for both manual transmission and automatic transmission vehicles. Searle (1999) concluded that the constant acceleration assumption is more realistic while vehicles accelerate with the maximum acceleration rate with a given gear.

2.1.1.2 Two Phase Acceleration Model (Dual-Regime Model)

The two phase acceleration model (see Figure 2) assumes that a high acceleration rate associates with lower speed and a lower acceleration rate associates with higher speed. The two phase model applies two acceleration rates to describe the acceleration process instead of one. This model modification improves the constant acceleration model and still maintains simplicity at the same time. However, the two phase acceleration model provides a discontinuous acceleration profile, and so does not represent the real acceleration process which has a continuous acceleration and speed profile.
Bham and Benekohal (2001) studied the acceleration characteristics of a platoon based on two sets of field data. The two phase model provides a better fit to the data than the constant acceleration model does. This study also proposed the use of 13 m/s (29.5 mph) as the speed breaking point that separates the two phases, see the following equations (2-3, 2-4):

\[
\begin{align*}
\text{Phase I for } v' < 13 \text{ m/s} : & \quad v_{f1} = v_{i1} + a_1 t \\
\text{Phase II for } v' \geq 13 \text{ m/s} : & \quad v_{f2} = v_{i2} + a_2 t
\end{align*}
\]

where,

- \( v_{i1} \) = initial speed of phase I (m/s),
- \( v_{f1} \) = final speed of phase I (m/s),
- \( v_{i2} \) = initial speed of phase II (m/s), \( v_{f2} = v_{f1} \),
- \( v_{i2} \) = final speed of phase II (m/s), and
- \( a_1, a_2 \) = average acceleration of phase I and phase II respectively (m/s\(^2\)).
2.1.1.3 Linearly-Decreasing Acceleration Model

The literature review identified two types of linearly-decreasing acceleration models. One represents the linearly-decreasing relationship between acceleration and speed, and the other represents the relationship between acceleration and time.

**Acceleration and Speed Relationship**

The linearly-decreasing acceleration model is also called a non-uniform acceleration model (Bham & Benekohal, 2002) or linear decay model (Rakha et al., 2004). This model assumes that during the acceleration process the acceleration rate linearly decreases as speed increases (see equation 2-5 and Figure 3).

\[
\frac{dv}{dt} = \alpha - \beta * v
\]  

(2-5)

where,

- \(v\) = speed (m/s),
- \(\frac{dv}{dt}\) = derivative of speed with respect to acceleration time, which is the acceleration rate \((m/s^2)\),
- \(\alpha\) = a constant, \(\alpha = a_{max}\), and
- \(\beta\) = a constant, the slope of the linearly decreasing line.

![Figure 3  Linearly-Decreasing Acceleration Model (Acceleration vs. Speed)](image_url)
This model also assumes that the maximum acceleration rate \( a_{\text{max}} \) occurs at the beginning of the acceleration process, and the acceleration rate reaches zero when speed approaches the desired speed or free flow speed \( (v_{\text{ff}}) \). Speeds will remain approximately constant at the end of the acceleration process. The constant \( \alpha \) and \( \beta \) can be estimated through linear regression analysis based on field data or can be calculated from general motion functions.

Long (2000) performed an extensive review of linearly-decreasing acceleration models for leading vehicle acceleration characteristics. Long concluded that the linearly-decreasing acceleration model provides better accuracy and consistency within different ranges of speed than the constant acceleration model.

Chowdhury and Rao (1989) reported a linearly-decreasing, two-phase acceleration model with a similar form. The researchers used the model to represent acceleration characteristics of the leading passenger car accelerating at signalized intersections when the traffic signal changes to green. For each phase, different sets of the constant \( \alpha \) and \( \beta \) can be estimated using field data. The authors explained that the dual acceleration phases were most likely due to the gear shifting from the first to the second gear. The next section will exhibit Chowdhury and Rao’s study in more detail.

**Acceleration vs. Time Relationship**

Akcelik and Biggs (1987) proposed a model to represent the linearly-decreasing relationship between the acceleration rate and time (see equation 2-6, 2-7 and Figure 4).

\[
a(t) = 2 \times (1 - \theta)(v_f - v_i) / t_a
\]

\[
v(t) = v_i + (2 - \theta)\theta(v_f - v_i)
\]

\[(2-6)\]

\[(2-7)\]
where,

- $t$ = the acceleration time (second),
- $a(t) = \text{the acceleration at time } t \text{ (m/s}^2\text{)},$
- $v(t) = \text{the speed at time } t \text{ (m/s)},$
- $v_i = \text{the initial speed (m/s)},$
- $v_f = \text{the final speed (m/s)},$
- $t_a = \text{the total acceleration time to reach the final speed (second), and}$
- $\theta = \frac{t}{t_a}, \text{ time ratio}.$

Lee and Rioux (1977) proposed and applied the linearly-decreasing acceleration model with the relationship depicted in equation 2-8 (see Figure 4) for the TEXAS Model—a microscopic traffic simulation package.

$$a_f = a_i + \beta \times t \quad (2-8)$$

where,

- $a_f = \text{final acceleration (m/s}^2\text{)},$
- $a_i = \text{initial acceleration (m/s}^2\text{), } a_i = a_{\max}, \text{ and,}$
- $\beta = \text{constant, acceleration slope, negative value.}$

![Figure 4](image)

**Figure 4**  Linearly-Decreasing Acceleration Model (Acceleration vs. Time)
The linearly-decreasing acceleration model assumes that the maximum acceleration rate occurs at the beginning of the acceleration process while vehicle speed is zero, and decreases to zero at the end of the acceleration process \((t = t_a)\) while vehicle speed reaches free flow speed. Research shows that the linearly-decreasing acceleration model provides a good representation of normal acceleration characteristics except for the unrealistic assumption of high initial acceleration rate at the beginning of the acceleration process (Akcelik & Biggs, 1987; Bham & Benekohal, 2002; Long, 2000).

2.1.1.4 Polynomial Acceleration Model

Akcelik and Biggs (1987) and Akcelic and Besley (2002) proposed a nonlinear polynomial acceleration model to describe the relationship between acceleration rate and time (see Equation 2-9, and Figure 5).

\[
a(t) = r a_m \theta (1 - \theta^m)^2
\]

(2-9)

where,

\(a(t)\) = the acceleration at time \(t\) \((m/s^2)\),

\(t_a\) = the total acceleration time to reach the final speed (second),

\(t\) = time since the start of acceleration (second),

\(\theta\) = \(t/t_a\), time ratio,

\(m\) = model calibration parameter, and

\(r\) = model parameter given by:

\[
r = \frac{\left[ (1 + 2m)^{\frac{1}{m}} \right]^2}{4m^2}
\]

(2-10)
The researchers collected field data from an instrumented vehicle through urban and non-urban area chase-car studies of 1037 vehicles in Sydney Australia. The authors did not publish the information for the experimental design, nor did they describe how to control spacing during the car chasing process in the report. The study team selected acceleration records by satisfying all of the criteria listed below:

Initial speed \( (v_i) \) : \( 0 \leq v_i \leq 1 \text{ km/h} \ (0.63 \text{ mph}) \)

The time at the end of acceleration process \( (t_a) \) : When the speed stops increasing or does not increase greatly during the next 5 seconds

Final speed \( (v_f) \) : \( v_f \geq 20 \text{ km/h} \ (12.5 \text{ mph}) \)

These criteria increase the likelihood that the data closely represent the full acceleration process. Studies (Akcelik and Biggs, 1987; Wang et al., 2004) reported that the polynomial model
provided good fit to the field data. However, the function form of this model is relatively more complicated than others which makes it not very convenient to use.

Wang et al. (2004) also proposed a polynomial acceleration model (see equation 2-11) to describe the normal acceleration characteristics at stop-controlled intersections,

\[ \sqrt{a} = \alpha - \beta v \]  

(2-11)

where,

- \( a \) = acceleration rate (m/s\(^2\)),
- \( v \) = speed (m/s), and,
- \( \alpha \) and \( \beta \) = constants, estimated based on empirical data.

This study analyzed the field data from over 100 GPS-instrumented vehicles, including 656 acceleration trips and 76 drivers. These acceleration trip data were collected from various road functional classifications, including principal arterial, minor arterial, major collector, and minor collector. These 656 acceleration trips all met the following criteria:

Initial speed (\( v_i \)) : \( v_i = 0 \) km/h (0 mph)

The time at the end of acceleration process (\( t_a \)) : When speed increases less than 0.16 km/h (0.1 mph) between two successive one-second intervals.

Final speed (\( v_f \)) : \( v_f \geq 32 \) km/h (20 mph)

The research team performed linear regression analysis based on part of the data set to derive the polynomial model, equation 2-11. The study also showed that polynomial models proposed by
Akcelic & Biggs (1987) and Wang et al. (2004) provide a good fit based on the reserved acceleration trip data. The polynomial model developed by Wang et al. also provides the first effort to apply regression analysis to the acceleration process for through maneuvers and turning maneuvers respectively.

- Through maneuvers:
  \[ \sqrt{a} = 1.381 - 0.011v \]
  \[ R^2 = 0.424 \]
- Turning maneuvers:
  \[ \sqrt{a} = 1.289 - 0.009v \]
  \[ R^2 = 0.359 \]

2.1.2 Dynamics Acceleration Models

Researchers developed dynamics acceleration models mainly for vehicle acceleration capability studies. These studies are very valuable for automotive manufacturing, vehicle performance comparisons and auto performance testing. The basic assumption of this type of model is that vehicles accelerate with the maximum acceleration rate through entire acceleration process. Searle (1999) proposed a dynamics acceleration model to describe the through-gear acceleration characteristics with the maximum acceleration rate by revealing physical mechanisms of the acceleration process based on the engine power losses. Generally, the maximum engine power of a vehicle represents the vehicle performance capability. However, the actual engine power used for acceleration is lower than the peak engine power due to power loss from various sources. Therefore, Searle assumed that vehicles accelerate with a constant percentage of the maximum engine power throughout the acceleration process, represented with the acceleration efficiency parameter \( \eta \). Searle recommended \( \eta = 50\% \) acceleration efficiency for most of the vehicles.
Searle introduced another parameter, power constant, $k$, to establish the equations for predicting vehicle speed, time, and distance in the acceleration processes (see equation group 2-12).

\[
k = \frac{7.9 \times \eta \times P_{\text{max}}}{M}
\]

\[
\begin{align*}
    v_f^2 &= v_o^2 + 2kt \\
    v_f^3 &= v_o^3 + 3ks \\
    s^2 &= \frac{8}{9}kt^3
\end{align*}
\]  

where,

- $k$ = power constants (bhp/ton, kilowatt/ton),
- $\eta$ = acceleration efficiency,
- $P_{\text{max}}$ = maximum engine power (bhp, kilowatts),
- $M$ = mass of vehicle (tons),
- $v_o$ = initial speed (km/h),
- $v_f$ = final speed (km/h),
- $s$ = travel distance (km), and
- $t$ = acceleration time (second).

Searle used data from *Autocar Magazine* to validate the model. This data represent the acceleration capability of vehicles to accelerate from rest to various speeds. The validation is based on the comparison of acceleration times for reaching the corresponding speeds. The validation indicated similar results between predicted values and the measured values. Searle (1999) indicated that the normal acceleration characteristics cannot be deduced from engineering principles; therefore, these characteristics must be acquired from field observation and measurement.
Rakha et al. (2004) performed another acceleration study based on the data collected by GPS instrumented vehicles. Subjects performed test drives on a 1.6 km (1 mile) testing roadway. The study included 13 GPS instrumented vehicles. The test vehicles included light-duty vehicles, such as subcompact, compact, midsize, full-size, minivan, pickup, and sports utility vehicles. Test drivers maneuvered their vehicles from a full stop and accelerated at a maximum acceleration rate to three target speed thresholds:

- Accelerate from 0 to 56 km/h (0 to 35mph)
- Accelerate from 0 to 88 km/h (0 to 55mph)
- Accelerate from zero to maximum attainable speed within the test road section

Rakha et al. (2004) established the model (see equation group 2-13) based on the Newton’s Second Law: the force equals the product of the vehicle mass and the vehicle acceleration:

$$a = \frac{F - R}{M}$$  

(2-13)

$$F = \min(F_r, F_{\text{max}})$$

$$F_r = 3600\eta\beta \frac{P}{v}$$

$$F_{\text{max}} = 9.8066M_{\text{ia}}\mu$$

$$R = R_u + R_r + R_g$$

$$R_u = c_1C_dC_hA_fv^2$$

$$R_r = 9.8066C_r(c_2v + c_3)\frac{M}{1000}$$
\[ R_g = 9.8066MG \]

where,

\[ a = \text{acceleration (m/s}^2) \],
\[ F = \text{tractive force (N)} \],
\[ R = \text{resistance force (N)} \],
\[ M = \text{mass of vehicle (kg)} \],
\[ F_t = \text{engine tractive force (N)} \],
\[ F_{\text{max}} = \text{maximum force which can be sustained between tire and pavement (N)} \],
\[ R_a = \text{aerodynamic resistance force (N)} \],
\[ R_r = \text{rolling resistance force (N)} \],
\[ R_g = \text{grade resistance force (N)} \],
\[ \beta = \text{variable power factor} \],
\[ \eta = \text{transmission efficiency} \],
\[ P = \text{engine power (kW)} \],
\[ v = \text{vehicle speed (km/h)} \],
\[ M_{\text{tn}} = \text{mass of vehicle on tractive axle (kg)} \],
\[ \mu = \text{coefficient of friction tired and pavement} \],
\[ c_1 = \text{constant accounting for density of air at sea level} \],
\[ C_d = \text{vehicle drag coefficient} \],
\[ C_h = \text{altitude coefficient} \],
\[ A_f = \text{vehicle frontal area (m}^2) \],
\[ C_r = \text{rolling coefficient} \],
\[ c_2, c_3 = \text{rolling resistant coefficient, and} \]
\[ G = \text{percent grade (m/100m)} \].
This model provides equations to calculate the effective tractive force and various resistance forces respectively. Rakha et al. (2004) introduced the variable power efficiency factor $\beta$ in order to calculate the effective tractive force. Unlike the acceleration efficiency factor $\eta$ mentioned in the dynamics acceleration model developed by Searle, the power efficiency factor $\beta$ is a function of speed. Therefore Rakha et al. established this acceleration model without the assumption of constant percentage output of maximum engine power for the accelerating maneuver. The resistance force term takes into account various resistance sources, including aerodynamic, rolling, and grade resistance.

This model still assumes that vehicles accelerate with maximum acceleration. However, Rakha et al. introduced the concept of applying a user defined acceleration reduction factor $\gamma$ into the model in order to model the vehicle’s normal acceleration characteristics. This factor provides a certain percentage reduction of the maximum acceleration.

This dynamics acceleration model provides a better prediction of vehicle maximum acceleration characteristics, and takes into account various vehicle types. The model’s input parameters do not need to be estimated from field data, since they can be found from the vehicle manuals, specifications, or auto magazines. Based on the field data collected as described above in Rakha’s study (Rakha et al., 2004), which reflects the vehicle maximum acceleration characteristics, this acceleration model shows a good fit in all of the five relationship domains:

- Speed vs. distance
- Speed vs. time
- Acceleration vs. distance
- Acceleration vs. time
- Acceleration vs. speed
2.2 Acceleration Characteristics at Intersections

Generally, intersections are locations where drivers need to perform stop-and-go maneuvers, according to different types of traffic control devices, such as stop signs and traffic signals. Acceleration characteristics at intersections also represent most of the normal acceleration behaviors indicated in ITE Traffic Engineers Hand Book (1999). Therefore this report provides an overview of acceleration studies related to intersections with stop control and signal control devices.

2.2.1 Stop Controlled Intersections

As introduced in the previous section, Wang et al. (2004) performed a passenger car acceleration study for both through and turning maneuvers at stop controlled intersections. The results show that there is no significant difference between acceleration characteristics for turning and through movements. The relationship between acceleration and speed still can be formulated with a similar polynomial format, see equation 2-11.

2.2.2 Signalized Intersections

Briefly mentioned previously, Chowdhury & Rao (1989) proposed a two-phase linearly-decreasing acceleration model (see equation 2-14, 2-15 and Figure 6) based on field data collected from a signalized intersection. This study focused on the leading passenger car in a queue which starts accelerating when the signal turns green with through-movements. The observed data were collected from spot studies. Researchers collected the travel time and travel distance in order to calculate the average speed and average acceleration in the acceleration process. All of the data came from the same intersection, and, researchers picked the leading passenger car randomly.
Phase I: \[
\frac{dv}{dt} = \alpha_1 - \beta_1 \cdot v \\
\tag{2-14}
\]

Phase II: \[
\frac{dv}{dt} = \alpha_2 - \beta_2 \cdot v \\
\tag{2-15}
\]

where,

\( v \) = speed (m/s),

\( \frac{dv}{dt} \) = derivative of speed respect to time, the acceleration for each phase (m/s\(^2\)),

\( \alpha_1, \alpha_2 \) = constants, see Figure 6, and

\( \beta_1, \beta_2 \) = constants, acceleration slope of phase I and phase II respectively.

![Figure 6](image)

Figure 6  Chowdhury’s Two-Phase Linear-Decreasing Acceleration Model

The data used in this study have low accuracy since the speed and acceleration rate need to be calculated based on the assumption of constant speed and constant acceleration during each travel time and travel distance. Therefore, the model may not be able to represent the real situation with better predictions.
Long (2000) made an extensive discussion and study of acceleration characteristics of starting vehicles. He mentioned previous research that was carried out at signalized intersections decades ago using data from spot studies. For example, Dockerty (1966) performed his study at two intersection sites, and focused on leading vehicles in queues. This study is an uncontrolled test study. The linearly-decreasing acceleration model provided a better fit to the field data without pedestrian interference. Dockerty also reported that leading vehicles tend to accelerate slower in the outside lane than in the middle lane.

2.3 Summary

Various acceleration process studies and modeling efforts have been carried out for about a half-century. The restrictions and limitations of previous studies include low data accuracy, secondary data calculation assumptions, lack of data points to represent a full acceleration process, and small sample size. Also, most of the models were developed based on various assumptions which are unrealistic. Researchers still need a better understanding of the acceleration process of queue leaders at signalized intersections based on high quality data and appropriate modeling methods.

This thesis presents improvements of the acceleration process study in the following aspects:

- Data quality and quantity
- Less assumptions for model developing
- Initial Investigation of road physical features effects and driver-vehicle effects
3 DATA COLLECTION AND EXTRACTION METHODOLOGY

3.1 Data Collection and Sampling Method

The author performed this study using data generated from a real-time data collection activity over an extended period of time. The unique database included in-vehicle GPS technology records. Members of the research team installed the GPS in-vehicle data collection equipment in private vehicles, and each device included a CPU, power system, cellular transceiver, and other sensors. The equipment turned on and off automatically with the vehicle ignition. Recorded data were automatically transferred to a data server over wireless connections every week.

The participating drivers and vehicles were randomly selected in the Atlanta urban area (Wang, 2006). The study roadways are the actual driver-self-selected routes without instructions of any kind from researchers. The drivers selected routes as appropriate for their personal purposes. The database includes second-by-second information including date, time, speed, and vehicle location.

The recruitment survey and sampling method of this study has been destroyed due to the contractual issue beyond our control, so the author of this thesis is not able to obtain the original and detailed information of the random sampling method. Therefore, this thesis presents a similar approach that was used by the same research team in a different study, the Commute Atlanta Value Pricing Program: Recruitment Methods and Travel Diary Response Rate (Ogle et al, 2005).

In the Commute Atlanta research program, the sample selection method started from the 2001 8,000-household Atlanta travel survey study. The 8,000-household travel survey study was prepared for the Strategies for Metropolitan Atlanta Transportation and Air Quality
(SMARTRAQ) study. The following section introduces the related information of the 8,000-household travel survey study and the Commute Atlanta study.

The survey research and consulting firm, NuStats, stratified households with the Net Residential Density (NRD), which is the number of dwelling units divided by the land area within residential area for a given area measured in acres (Chapman et al, 2003). Five NRD levels were established, including 0-1.999, 2-3.999, 4-5.999, 6-7.999, and 8+ dwelling units per net residential acre. The goal was to have 20% samples falling within each of five NRD levels. The household recruitment process used a random digit dialing method. The households were selected at random from a database of telephone numbers, which included all households with telephones in the study regions. The study regions, 13 counties in the Atlanta metro area, are the same as the study area of this acceleration study. The household recruitment and data collection process includes the following major steps:

- Advance mailing;
- Recruitment call and interview: Collect information such as household income category, household size, ethnicity, vehicle ownership, etc;
- Geocoding of important addresses;
- Mailing travel diaries to households;
- Placing reminder calls; and
- Data retrieval interview.

About 8000 households participated in the travel diary study. The research team summarized the distributions of household income category, household size, and vehicle ownership for the sampled household data.
The Commute Atlanta study used a standard phone list for the Atlanta region as the sampling framework. The research team developed the sampling strata by first examining the cross-classification matrices (information of household income, household size, and vehicle ownership) of the 8,000-household study. Based on that, the research team established the groupings of income, household size, and vehicle ownership that represent the major distribution of households in the Atlanta region. The recruitment firm randomly selected households from the phone list to receive advance letters and the subsequent phone calls to recruit participants. For each phone call, the recruitment company collected household social-demographic information, decided which sampling strata this household fell into, and scheduled the equipment installation. The recruitment company also kept tracking the sampling distributions in order to satisfy the sampling plan with the targeted percentage samples in each stratum.

In this acceleration study, the research team used a random stratified sampling approach to recruit households from the 8,000-household which participated in the 2001 Atlanta travel survey study. The research team used the 8,000-household database as the sampling framework to send advance letters, make preceding phone calls, and schedule the instrument installation. The sampling and data collection result are shown in Table 2 and present the driver and vehicle profile (Wang, 2006).
Table 2     Study Driver and Vehicle Profile

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of drivers</td>
<td>145</td>
</tr>
<tr>
<td>Received GPS data records</td>
<td>25,096,786</td>
</tr>
<tr>
<td>Female</td>
<td>61%</td>
</tr>
<tr>
<td>Male</td>
<td>39%</td>
</tr>
<tr>
<td>Age less than 18</td>
<td>5%</td>
</tr>
<tr>
<td>Age between 18 and 45</td>
<td>44%</td>
</tr>
<tr>
<td>Age between 45 and 60</td>
<td>35%</td>
</tr>
<tr>
<td>Age larger than 60</td>
<td>17%</td>
</tr>
<tr>
<td>Passenger car</td>
<td>58%</td>
</tr>
<tr>
<td>Minivan</td>
<td>17%</td>
</tr>
<tr>
<td>SUV</td>
<td>17%</td>
</tr>
<tr>
<td>Pickup</td>
<td>8%</td>
</tr>
</tbody>
</table>

3.2 Comparison of Data Collection Methods

The GPS data collection technique used in this study is quite innovative. Historically, acceleration study researchers have been restricted by data collection techniques which affected data quantity and quality. Most of previous studies collected data using spot acceleration data at static locations within a certain distance downstream of one or two intersections (Beakey, 1938; Chowdhury and Rao, 1989). Researchers were only able to collect limited data points for each acceleration event. Typically, the data information included travel time and travel distance. They then used this information to calculate the corresponding average speed and acceleration rate. These data are not able to represent the entire continuous acceleration process, and data accuracy can be confounded due to the secondary data calculation procedures.

GPS-equipped vehicles provide the ability to collect continuous driving records at various time intervals (for this study this interval is one-second), and as a result, can more accurately represent the complete acceleration process. The real world data also capture natural driving behaviors. This additional information is a benefit not available from test driver experiments. Table 3 shows
a summary of advantages and disadvantages of four types of data collection methods commonly used for acceleration studies.

<table>
<thead>
<tr>
<th>Data Collection Methods</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| **Spot Study**          | $\checkmark$ Convenience of equipment setting  
$\checkmark$ Better control of road situations | $\checkmark$ Limited data points  
$\checkmark$ Secondary data lacking of accuracy |
| **Controlled Test Study** | $\checkmark$ Better control of driver-vehicle and road conditions | $\checkmark$ Driving behaviors are as natural as in real world situations  
$\checkmark$ Limited driver-vehicle participations |
| **Chase-Car Study**      | $\checkmark$ Flexibility of picking chasing targets | $\checkmark$ Consistency of chasing spacing control  
$\checkmark$ Influences from following behavior  
$\checkmark$ Lacking of driver-vehicle information |
| **GPS Uncontrolled Experiments** | $\checkmark$ Longer time frame  
$\checkmark$ Continuous information and entire coverage of acceleration processes  
$\checkmark$ Natural driving behavior  
$\checkmark$ Randomization sampling of driver-vehicle subjects | $\checkmark$ Less information control of driving situations and road conditions  
$\checkmark$ Data processing and modeling challenge |

### 3.3 Data Extraction Criteria

The author applied the data processing methodology developed from a previous operating speed study (Wang, 2006) in order to extract the trip data segments which represent acceleration events at signalized intersections. The author executed these procedures by using PERL (Practical Extraction and Report Language) as a tool in this study.

Referring to the criteria used in previous studies (Akcelik and Biggs, 1987; Wang et al., 2004), the acceleration data extraction method in this thesis uses the following criteria to increase the likelihood of extracting full normal acceleration trips:
• The initial speed is zero;
• The speed continuously increases, which represents a general feature of the normal acceleration process;
• The final speed is greater than 75% of the speed limit of the roadway;
• The acceleration trip only includes through-movement maneuvers.

The study vehicles for these acceleration trips traversed several different types of roadways (local streets, collectors, minor arterials, and major arterials) under four speed limit groups (48, 56, 64, and 72 km/h (30, 35, 40, and 45mph)). Table 4 shows an example of the acceleration trip data records.

<table>
<thead>
<tr>
<th>Driver_NO.</th>
<th>TIME (sec)</th>
<th>LATITUDE (degree)</th>
<th>LONGITUDE (degree)</th>
<th>SPEED (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>0</td>
<td>33.969236</td>
<td>-84.489115</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>33.969236</td>
<td>-84.489115</td>
<td>1.4476</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>33.969264</td>
<td>-84.4891</td>
<td>3.74</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>33.969304</td>
<td>-84.489072</td>
<td>5.9752</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>33.969361</td>
<td>-84.489035</td>
<td>8.1576</td>
</tr>
<tr>
<td>27</td>
<td>5</td>
<td>33.969427</td>
<td>-84.488982</td>
<td>9.6272</td>
</tr>
<tr>
<td>27</td>
<td>6</td>
<td>33.96951</td>
<td>-84.488924</td>
<td>10.934</td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>33.969603</td>
<td>-84.488856</td>
<td>12.43</td>
</tr>
<tr>
<td>27</td>
<td>8</td>
<td>33.969707</td>
<td>-84.488782</td>
<td>13.7104</td>
</tr>
<tr>
<td>27</td>
<td>9</td>
<td>33.969817</td>
<td>-84.4887</td>
<td>14.5156</td>
</tr>
<tr>
<td>27</td>
<td>10</td>
<td>33.969933</td>
<td>-84.488613</td>
<td>15.2636</td>
</tr>
<tr>
<td>27</td>
<td>11</td>
<td>33.970056</td>
<td>-84.488523</td>
<td>16.2404</td>
</tr>
<tr>
<td>27</td>
<td>12</td>
<td>33.970185</td>
<td>-84.488425</td>
<td>17.0192</td>
</tr>
<tr>
<td>27</td>
<td>13</td>
<td>33.970319</td>
<td>-84.488323</td>
<td>17.6396</td>
</tr>
<tr>
<td>27</td>
<td>14</td>
<td>33.970456</td>
<td>-84.488219</td>
<td>18.0004</td>
</tr>
<tr>
<td>27</td>
<td>15</td>
<td>33.970595</td>
<td>-84.488112</td>
<td>18.2072</td>
</tr>
<tr>
<td>27</td>
<td>16</td>
<td>33.970738</td>
<td>-84.488005</td>
<td>18.4052</td>
</tr>
</tbody>
</table>
3.4 Method for Separating Leading and Following Acceleration Trips

The extracted acceleration trips include two types of acceleration:

- Acceleration trips of leading vehicles in a queue, and
- Acceleration trips of following vehicles in a queue.

For this research effort, the leading vehicle acceleration process is the topic of interest. The author developed a methodology to distinguish leading from following acceleration trips by incorporating geometric information of study intersections. The author located study intersections through the internet tool Google Earth based on the longitude and latitude information of the starting location of each acceleration trip. The aerial photo provided the opportunities to identify the geometric layouts of study intersections, including lane configurations, stop line locations, pavement markings, etc.

The average space occupied by one vehicle in a queue is about 8.1m (25ft), which was estimated by previous study (Bonneson, 1992). Therefore, for this study the author used the criteria of vehicles being queue leaders in acceleration trips as those that have starting locations within the upstream distance of 8.1m (25ft) from stop lines. After taking into account the potential measurement error of GPS data at about ±5m (±16ft), the author considered the distance of 8.1m (25ft) to be a practical criteria to use in this study for distinguishing leading and following acceleration trips. The next section describes the five-step procedure used for identifying the leading vehicle acceleration based on this criterion.

**The Five-Step Procedure**

1. Define driving directions based on bearing ranges.
Table 5 and Figure 7 present the schematic based on bearing ranges to define driving directions of acceleration trips, including North Bound (NB), West Bound (WB), South Bound (SB), and East Bound (EB).

<table>
<thead>
<tr>
<th>Directions</th>
<th>Bearing Ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>N45°E and N45°W</td>
</tr>
<tr>
<td>WB</td>
<td>N45°W and S45°W</td>
</tr>
<tr>
<td>SB</td>
<td>S45°W and S45°E</td>
</tr>
<tr>
<td>EB</td>
<td>S45°E and N45°E</td>
</tr>
</tbody>
</table>

Since the fundamental methodologies are similar for each driving direction, the procedure description in this section will use north bound trips as the example to illustrate the general algorithm. The simplified sketches (a), (c), and (e) in Figure 8 demonstrate three possible situations for the northbound driving direction with the reference point A in the second quadrant,
on the Y-axis, and in the first quadrant respectively. The starting locations from queue leaders should fall within the shaded area in these three sketches. This area represents the space bounded by the stop bar, the backup line of 8.1m (25ft) from the stop bar, and the travel lane pavement marking on both sides.
Figure 8  Schematics of Distinguishing Queue Leaders
-- North Bound Driver Direction
(Not to scale)
2. Find the longitude and latitude of the reference point A and O.

The reference point A is defined as the midpoint of the stop line for the through travel lanes (see Figure 8). The reference point O is defined as the point with a perpendicular distance of 8.1m (25ft) to the stop bar, and the line OA is perpendicular to the stop bar (see Figure 8). The longitude and latitude of reference point A and O can be obtained manually based on the aerial photos of each study intersection as depicted in Google Earth.

3. Convert the longitude and latitude of the reference point A and O to coordinates in the local Cartesian coordinate system.

The reference point O is taken as the origin of the local Cartesian coordinate system (right-hand coordinate system) with coordinates as (0, 0). The coordinates of the reference point A, $x_a$ and $y_a$ are calculated using equations 3-1 and 3-2 respectively.

$$x_a = (\text{longitude}_A - \text{longitude}_O) \frac{\pi}{180} R_{earth} \quad (3-1)$$

$$y_a = (\text{latitude}_A - \text{latitude}_O) \frac{\pi}{180} R_{earth} \quad (3-2)$$

Where,

$R_{earth}$ = the radius of earth; $R_{earth} = 20925524.9$ ft.

The earth radius is based on the assumption that the earth is a sphere. The latitude and longitude coordinate system are based on the assumption that the earth is an ellipsoid. However, the differences between sphere radius and ellipsoid radius of the earth at the given latitude will not lead to significant differences for this study purpose considering the marginal changes of latitude and longitude and the measure errors of GPS data.
4. Calculate the reference angle θ.

In Figure 8, plots (b), (d), and (f) illustrate that the reference angle θ can be calculated based on the equation 3-3.

\[ \theta = \arctg \left( \frac{x_a}{y_a} \right) \]  

(3-3)

5. Check the starting location.

In Figure 8, the points D and E in the plots (b), (d), and (f) represent the starting locations of two acceleration trip examples. The acceleration trips will be the leading vehicle acceleration from queue leaders by meeting the following criteria in the corresponding situations:

• For situations as in plot (a):

  The starting location is in the first quadrant with \( \alpha > \theta \);

  The starting location is in the second quadrant; and

  The starting location is in the third quadrant with \( \alpha < \theta \).

• For situations as in plot (c):

  The starting location is in the first and second quadrant.

• For situations as in plot (e):

  The starting location is in the first quadrant;

  The starting location is in the second quadrant with \( \alpha > \theta \); and

  The starting location is in the fourth quadrant with \( \alpha < \theta \).

The angle \( \alpha \) can be calculated by repeating the step 3 and the step 4. The longitude and latitude information of point D or E are used instead of the longitude and latitude of reference point A. The author separated leading and following vehicle acceleration trips based this methodology.
3.5 Sensitivity Study of Different Speed Limit Groups

The author calculated the average speed profiles for each speed limit category by averaging speeds based on the sample size for each second. Figure 9 shows differences and similarities in speed profiles for the various speed limit groups.

![Figure 9](image.png)

Figure 9  Average Speed Profiles Comparison

Generally, by comparing the speed profiles for the acceleration process in the different speed limit categories, it shows that the speed profile in the higher speed limit group, 72 km/h (45mph), tends to be associated with a higher speed increasing rate at the early stage of the acceleration (0s to 10s). The higher speed category also exhibits longer acceleration times (ta=24s) and higher final speeds. The speed profile in the lower speed limit group, 48 km/h (30mph), tends to be associated with a lower speed increasing rate, shorter acceleration time (ta=15s), and lower final speed. The speed profiles are very similar for the speed limit group of 56 km/h (35mph) and 64 km/h (40mph). These two speed profiles have similar speed increasing rates, acceleration times (ta=18s), and final speeds. The speed profiles in all four groups have similar monotonically increasing curvature patterns, which indicate a possible polynomial relationship between speed and time. And the final speed of each average speed profile reaches the corresponding speed
limit. Figure 9 does not show an obvious S-shaped speed profile, reported by previous studies (Akcelik and Biggs, 1987; Wang et al. 2004).

Based on the sensitivity study of speed profiles for the various speed limit categories, the results indicate that statistical models should be separately developed for the various speed limit groups. This thesis presents statistical modeling and analysis results for the speed limit group 72 km/h (45mph) on principle and minor arterial roadways in the next chapter. This speed limit category represents the majority of the collected acceleration trip data of this study.
4 MODEL DEVELOPMENT

This chapter presents statistical modeling methodologies and analysis results for speed profiles of the acceleration process on arterial roadways with a speed limit of 72km/h (45mph). Section 4.1 introduces data structures and general considerations of variables used in the model based on the questions of interest. From Section 4.2 to Section 4.4, the thesis presents the mixed-effects model and analysis results. In addition to adoption of the mixed-effects model to study acceleration process, the thesis also analyzes three ordinary regression models, discusses the overall significance of regression results, and compares them with the results from the mixed-effects model in Section 4.5. Section 4.6 discusses the assumption diagnostics issues of the mixed-effects model. Finally, Section 4.7 demonstrates an application example of the fitted mixed-effects model.

4.1 Data Structure of Speed Profiles in the Acceleration Process

Figure 10 presents the scatter plot of speed profiles for acceleration trips. One-second interval speed records represent vehicle acceleration processes at signalized intersections on arterial streets with a speed limit of 72km/h (45mph). The data points showed in Figure 10 do not include the data points with a speed of zero at the beginning of each acceleration trip. Since the data frequency occurs on one-second intervals, the actual beginning of a trip is unknown, but the author can determine that a driver initiated a trip during a one-second interval by examining the speed at one-second. The time indexes are reset from zero seconds for the convenience of modeling the intercept term in regression analysis.
Table 6 and Table 7 illustrate and summarize the longitudinal acceleration trip data structure. This data structure is graphically depicted for the same type of information in Figure 10. Longitudinal data generally indicates that repeated measures are made on subjects over time. Therefore, responses within subjects are more likely correlated, which will violate the independent assumption required of regression analysis. In the acceleration study, repeated measures of speeds were recorded from the same driver either at the same location or at different locations over time. Speeds from the same driver are more likely correlated after accounting for the influence of time. Therefore, the driver-vehicle effect should be taken into account in the data analysis. Speeds recorded at the same location also have the tendency of correlation due to the similar road features that are expected to have similar influence on speeds. The model should take into account influences of locations as well.
Table 6  Longitudinal Data Structure

<table>
<thead>
<tr>
<th>Driver-vehicle (i)</th>
<th>Locations (j)</th>
<th>Time</th>
<th>0 1 ………………………………T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>y_{110}</td>
<td>y_{11p}</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>M_1</td>
<td>y_{1M_1 0}</td>
<td>y_{1M_1 j}</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>y_{210}</td>
<td>y_{21j}</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>M_2</td>
<td>y_{2M_2 0}</td>
<td>y_{2M_2 j}</td>
</tr>
<tr>
<td>s</td>
<td>1</td>
<td>y_{s10}</td>
<td>y_{s1j}</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>M_s</td>
<td>y_{sM_s 0}</td>
<td>y_{sM_s j}</td>
</tr>
</tbody>
</table>

Where,

Y_{ijk}: the response (speed) for driver i at location j at k-th second

Table 7  Longitudinal Data Summary

<table>
<thead>
<tr>
<th>Total Number of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drivers</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>36</td>
</tr>
</tbody>
</table>

The author chose to use roadway physical features to represent influences from different locations. The roadway features are the essential information for various aspects of the roadway geometric design. Intersection locations can be represented by a wide variety of physical road feature combinations. Based on the available information, the roadway features discussed in this study include horizontal alignments, intersection layouts, and number of travel lanes. The model treats these roadway features as fixed effects, which means that the model investigates effects on speed profiles for each individual roadway features.
In this study, the research team randomly selected driver-vehicle subjects from the general driver population of the Atlanta urban area. The driver-vehicle effect can be treated as either a fixed effect or a random effect. However, as illustrated in Table 8, fixed-effects and random-effects modeling strategies address different questions of interest and lead to different statistical inferential scopes. The main interest of the study is to estimate the between-driver variability in the driver population. Hence, this study treats the driver-vehicle effect as the random effect.

<table>
<thead>
<tr>
<th>Study Interests</th>
<th>Driver-Vehicle Effect Types</th>
<th>Inferential Scopes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed-Effect</td>
<td>Effect on speed profiles from each specific individual driver-vehicle subject in the sample</td>
<td>Only on driver-vehicle subjects in the sample</td>
</tr>
<tr>
<td>Random-Effect</td>
<td>Effect on speed profiles from the driver population</td>
<td>On the population where the sample was drawn</td>
</tr>
</tbody>
</table>

The previous operating speed study (Wang, 2006) applied the mixed-effects modeling method to investigate the driver-vehicle random effect and roadway physical feature fixed effects on operating speeds for low-speed urban environments. This thesis adopts the mixed-effects modeling method to model speed profiles of the acceleration process. This model is able to predict what influences roadway physical features and driver-vehicle subjects can have on speed profiles. This is the first large-scale modeling effort to study the road physical feature effects and driver-vehicle random effects on speed profiles of the acceleration process.

The next section will introduce the mixed-effects model which incorporates the fixed-effects of physical road features and the random-effect of driver-vehicle subjects after accounting for the
influence of time. Furthermore, the author also developed three ordinary regression models, discussed the overall significance of regression, and compared them with the mixed-effects model.

4.2 Linear Mixed-Effects Modeling Method

A linear mixed-effects model is a linear regression model incorporating both fixed-effects and random-effects. The general formulation of the linear mixed-effects model is (Pinheiro and Bates, 2000):

\[ y_i = x_i\beta + z_i b_i + \varepsilon_i \]  

(4-1)

Where,

- \( y_i \): the vector of responses for subject i (driver-vehicle i),
- \( x_i \): the known fixed effect regressor matrix for subject i,
- \( \beta \): the vector of fixed effects,
- \( z_i \): the known random effect regressor matrix for subject i,
- \( b_i \): the vector of random effects. Assume \( b_i \sim N(0, \Psi) \), \( \Psi \) is the variance-covariance matrix of random effects, and
- \( \varepsilon_i \): the vector of random error. Assume \( \varepsilon_i \sim N(0, \sigma^2 I_{n_i}) \).

Assume \( b_i \) and \( \varepsilon_i \) to be independent for different subjects and independent of each other for the same subject. Assume \( b_i \) and \( \varepsilon_i \) to be independent with regressors.

The Maximum Likelihood (ML) estimation is one of the commonly used parameter estimation methods for mixed-effects models (Pinheiro and Bates, 2000). Unlike the Ordinary Least Square (OSL) estimation, which estimates parameters in order to minimize the sum of squared errors, the
Maximum Likelihood estimation procedure estimates parameters in order to maximize the likelihood of sample data occurring. This section introduces the major principle of the Maximum Likelihood estimation method by using a simple example.

Assume a regression model:

\[
y_i = \beta_0 + \beta_1 x_i + \epsilon_i
\]

(4-2)

where,

\[y_i: \text{ the dependent variable,}\]
\[x_i: \text{ the independent variable,}\]
\[\beta_0, \beta_1: \text{ the parameters}\]
\[\epsilon_i: \text{ the random error, assume } \epsilon_i \sim N(0, \sigma^2).\]

Therefore, \( y_i \) is a random variable and follow normal distribution as \( y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \).

The likelihood function for the model form depicted by equation 4-2 is the probability density function for the data given the parameters,

\[
L(\beta_0, \beta_1, \sigma^2 | y_i)
= \prod_{i=1}^{n} p(y_i | \beta_0, \beta_1, \sigma^2)
= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right\}
\]

(4-3)

Equation 4-3 can be re-written as shown in equation 4-4,

\[
\ln(L(\beta_0, \beta_1, \sigma^2))
= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2
\]

(4-4)
Since the log function is monotonically increasing with an easier working functional form, instead of maximizing $L(\beta_0, \beta_1, \sigma^2)$, $\ln\left( L(\beta_0, \beta_1, \sigma^2) \right)$ is maximized following an optimization methodology in order to estimate the parameters. The detailed optimization procedures have been extensively introduced in the studies by Pinheiro and Bates (2000) and Verbeke et al. (1997).

The Restricted Maximum Likelihood (REML) estimation is another commonly used estimation method. The Restricted Maximum Likelihood estimation and the Maximum Likelihood estimation methods are both based on the likelihood principle. Generally they perform similarly. However, REML estimates the variances components (between group and within group variance) by accounting for the loss in degrees of freedom associated with an estimation of the fixed effect while ML does not (Verbeke et al., 1997). This study applies REML method to estimate parameters of the mixed-effects model.

### 4.3 Model Specification

The scatter plot of speed profiles (see Figure 10) showed a clear curvature pattern of the acceleration process. It shows speeds increasing rapidly at the beginning with a decreasing rate as time increases and ending at a relatively stable speed level commonly described as desired speed or cruising speed. Therefore, the author established the mixed-effects model (4-5) to describe speed profiles of the acceleration process by incorporating a third-order polynomial relationship of time, road feature fixed effects, and the driver-vehicle random effect.

The mixed-effects model is specified as:

$$ y_{ijk} = \beta_0 + \beta_1 x_{ijk} + \beta_2 x_{ijk}^2 + \beta_3 x_{ijk}^3 + \beta_4 d_{1j} + \beta_5 d_{2j} + \beta_6 d_{3j} + (b_{0i} + \epsilon_{ijk}) $$

(4-5)
Where,

\( y_{ijk} \): speeds (m/s) recorded from driver i at location j on the k-th second of an acceleration trip,

\( x_{ijk} \): k-th (sec) time count in one-second interval for driver i at location j,

\( d_{ij} \): two level factor, represents the road horizontal alignment feature along the acceleration process at location j,

<table>
<thead>
<tr>
<th>( d_{ij} )</th>
<th>Horizontal Alignment</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tangent</td>
<td><img src="image" alt="Tangent Illustration" /></td>
</tr>
<tr>
<td>0</td>
<td>Smooth Curve</td>
<td><img src="image" alt="Smooth Curve Illustration" /></td>
</tr>
</tbody>
</table>

\( d_{2j} \): two level factor, represents the intersection layout feature at location j,

<table>
<thead>
<tr>
<th>( d_{2j} )</th>
<th>Intersection Layout</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Skewed (( \theta &lt; 85^\circ ))</td>
<td><img src="image" alt="Skewed Illustration" /></td>
</tr>
<tr>
<td>0</td>
<td>Un-skewed (( 85^\circ \leq \theta \leq 90^\circ ))</td>
<td><img src="image" alt="Un-skewed Illustration" /></td>
</tr>
</tbody>
</table>

\( d_{3j} \): two level factor, represents the through traffic lane numbers at location j,

<table>
<thead>
<tr>
<th>( d_{3j} )</th>
<th>Number of Through Traffic Lanes</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td><img src="image" alt="2 Lanes Illustration" /></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td><img src="image" alt="3 Lanes Illustration" /></td>
</tr>
</tbody>
</table>
\[ \begin{align*} \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6 : & \quad \text{parameters for independent variables,} \\
 b_{0i} : & \quad \text{the random effect, which is a random variable that represents the deviation from the population mean of speed for subject } i. \ \text{Assume } b_{0i} \sim N(0, \sigma_b^2). \ \sigma_b^2 \text{ is the between-group(subject) variance.} \\
 \epsilon_{ijk} : & \quad \text{the random error of the recorded speeds of subject } i \text{ at location } j \text{ for } k\text{-th second.} \ \\
 & \quad \text{Assume } \epsilon_{ijk} \sim N(0, \sigma^2), \ \sigma^2 \text{ is the within-group (subject) variance.} \\
 & \quad \text{Assume } \epsilon_{ijk} \text{ and } \epsilon_{ij'k'}, \ \forall k, k' \in (0,...,T); j, j' \in (1,...,J), \text{ independent with each other, } E[\epsilon \epsilon'] = \sigma^2 I. \\
 & \quad \text{Assume } b_{0i}, \ \forall i, i \in (1,...,N), \text{ independent with } \epsilon_{ijk}, \ \forall i, i \in (1,...,N); \forall j, j \in (1,...,J); \forall k, k \in (0,...,T);. \\
 & \quad \text{Assume } b_{0i} \text{ and } \epsilon_{ijk} \text{ to be independent with regressors.} \\
 & \quad \text{Based on the above assumptions of the random effect and the random errors, then} \\
 \text{Var}[b_{0i}] = \sigma_b^2 \\
 \text{Cov}[b_{0i}, b_{0i}] = 0 \quad (i \neq i) \\
 \text{Var}[\epsilon_{ijk}] = \sigma^2 \\
 \text{Cov}[\epsilon_{ijk}, \epsilon_{ij'k'}] = 0 \quad (i \neq i) \end{align*} \]

Let \( n_i \) be the number of the observations of subject \( i \), the variance-covariance matrix of driver-vehicle subject \( i \) is a symmetric matrix with the dimension of \( n_i \times n_i \):

\[
\sum_{n_i=1}^{n} = \begin{pmatrix} \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \end{pmatrix}_{n_i \times n_i} \] (a)
$b_{bi}$ and $\varepsilon_{ijk}$ are random variables, and represent the unexplained variations of the speeds. Both of them are estimated by the variance of errors. Based on these similarities, the term $(b_{bi} + \varepsilon_{ijk})$ is treated as the compound “error” in the model. The compound errors for subject $i$ are correlated with each other that is caused by the common term $b_{bi}$.

- $Var\left[(b_{bi} + \varepsilon_{ijk})\right] = \sigma^2 + \sigma_b^2$.

$\sigma^2 + \sigma_b^2$ is the variance of the compound “error” for the same observation (speed) from subject $i$ at location $j$ for $k$-th second.

- $Cov\left[(b_{bi} + \varepsilon_{ijk}), (b_{bi} + \varepsilon_{ij'k'})\right] = \sigma_b^2 \cdot \left( j \neq j'; k \neq k' \right)$

The covariance of the errors for two different observations (speeds) from the same driver-vehicle subject $i$ is $\sigma_b^2$. The errors of two observations from the same driver are expected to be correlated after accounting for the influence of road features and time.

The variance-covariance matrix of the compound error for all $N$ driver-vehicle subjects in the data is

$$
\begin{pmatrix}
\sum_{(n_i, n_j)} & 0 & \cdots & 0 \\
0 & \sum_{(n_i, n_j)} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sum_{(n_i, n_j)}
\end{pmatrix}_{N \times N}
$$

$M = \sum_{i=1}^{N} n_i$, $M$ is the total number of observations.

- The diagonal matrices are the variance-covariance matrices for each driver-vehicle subject.

- $Cov\left[(b_{bi} + \varepsilon_{ijk}), (b_{bi} + \varepsilon_{ij'k'})\right] = 0 \cdot \left( i \neq i' \right)$

The errors of observations (speeds) from different driver-vehicle subjects \( i \) and \( i' \) are expected to be independent after accounting for the influence of road physical features and time. The off-diagonal terms are matrices with all zero elements.

This mixed-effects model (4-5) indicates that speeds are influenced by time, road features, and characteristics of driver-vehicle subjects. Random error, \( \epsilon_{ij} \), represents the unexplained variation of speeds after accounting for time, road features and driving behavior. Each driver-vehicle subject has an individual influence on speed profiles in the acceleration process. The random effect, \( b_{0i} \), represents the deviation of the group mean of driver-vehicle subject \( i \) from the population mean.

### 4.4 Model Estimation and Interpretation

#### 4.4.1 Model Estimation Results

By applying linear mixed-effects modeling technique with the REML estimation method, the fitted model and results are presented by equation 4-6 and Table 9. All of the estimates in the model are statistically significant. The fitted response variable, speed (\( \hat{y} \)), is represented by \( v \) in the regression equations for the rest of the thesis.
\[ v = 2.24 + 2.46t - 0.12t^2 + 0.002t^3 - 0.79\text{Curve} - 0.32\text{Skew} + 0.42\text{Lane} \] (4-6)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.24</td>
<td>0.276</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Time</td>
<td>2.46</td>
<td>0.038</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Time^2</td>
<td>-0.12</td>
<td>0.004</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Time^3</td>
<td>0.002</td>
<td>0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Curve</td>
<td>-0.79</td>
<td>0.097</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Skew</td>
<td>-0.32</td>
<td>0.095</td>
<td>0.0007</td>
</tr>
<tr>
<td>Lane</td>
<td>0.42</td>
<td>0.099</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

| Between-group residual standard deviation | \( \sigma_b = 1.448 \) |

| Within-group residual standard deviation | \( \sigma = 1.356 \) |

| Coefficient of determination (R^2) | 0.95 |

The coefficient of determination \( (R^2) \) is a statistic that measures goodness-of-fit of a model. Appendix A presents details for the coefficient of determination. The \( R^2 = 0.95 \) can be interpreted that 95% of the speed variation can be explained by the mixed-effects model.

Table 10 presents the 95% confidence intervals of the estimators. The 95% confidence interval is interpreted as: we are 95% confident that the interval contains the parameter.
### Table 10: 95% Confidence Intervals of Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Lower</th>
<th>Coefficients</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.70</td>
<td>2.24</td>
<td>2.78</td>
</tr>
<tr>
<td>Time</td>
<td>2.39</td>
<td>2.46</td>
<td>2.54</td>
</tr>
<tr>
<td>Time^2</td>
<td>-0.129</td>
<td>-0.12</td>
<td>-0.11</td>
</tr>
<tr>
<td>Time^3</td>
<td>0.0019</td>
<td>0.002</td>
<td>0.0024</td>
</tr>
<tr>
<td>Curve</td>
<td>-0.98</td>
<td>-0.79</td>
<td>-0.60</td>
</tr>
<tr>
<td>Skew</td>
<td>-0.51</td>
<td>-0.32</td>
<td>-0.14</td>
</tr>
<tr>
<td>Lane</td>
<td>0.22</td>
<td>0.42</td>
<td>0.61</td>
</tr>
</tbody>
</table>

#### 4.4.2 Model Interpretation

**4.4.2.1 Third-Order Polynomial Relationship between Speed and Time**

The fitted mixed-effects model provides a good fit to the data with $R^2 = 0.95$. Figure 11 presents an example graphically by superimposing the data scatter plot with the fitted model curve under the road physical features as straight road segments, un-skewed intersections with three through travel lanes. The fitted speed profile represents the similar pattern with that observed from the data. In Figure 11, the final speed of the fitted curve is 19.6 m/s (44.5 mph). This is a reasonable and expected result according to the speed limit of 72 km/h (45mph) on arterial streets where the data were collected. Overall, the significant result shows that speeds are highly correlated with time, hence the third-order polynomial model is appropriate for this modeling effort.
4.4.2.2 Road Physical Feature Effects

With remaining all other variables constant, drivers tend to drive at lower speeds on horizontal curves than on straight road segments (see Figure 12). However, only three out of the twenty-eight study intersections were characterized by horizontal curvature. Even though the model shows reasonable and significant results, this horizontal curvature variable merits further investigation with a larger sample size in order to provide stronger evidence.
Drivers tend to drive at lower speeds at skewed intersections than at un-skewed intersections. Fourteen out of the twenty-eight study intersections were characterized by skewed intersections.

The horizontal curvature has a stronger influence on reducing speeds than the skewed intersection as depicted by comparing the coefficients of ‘Curve’ and ‘Skew’ in the model as - 0.79 m/s (1.80 mph) < - 0.32 m/s (0.73 mph) (see Table 9).

Drivers tend to drive at higher speeds on road segments with two through travel lanes than with three through travel lanes. More chances for lane changing behavior on three-lane road segments may relate to this phenomenon. There are seven out of the twenty-eight study intersections characterized by three-through travel lanes.

The physical road effects on speed profiles are statistically significant, even though the actual magnitude of the influence on speeds is not very critical. For example, the speed reduction from curved horizontal alignments compared to straight horizontal alignments is about 0.79 m/s (1.80 mph). The speed reduction due to skewed intersections compared to un-skewed intersections is about 0.32 m/s (1.73 mph). The speed increase at locations with two through traffic lanes compared to locations with three through lanes is about 0.42 m/s (0.95 mph). The combination of curved and skewed intersections will lower the speed at about 1.11 m/s (2.52 mph) at the intercept. This combined influence has the largest physical road feature effect.

4.4.2.3 The Driver-Vehicle Random Effect

The fitted mixed-effect model (4-6) treats the driver effect as the random effect, which is represented by a random variable. This random variable is assumed to follow the normal distribution with the mean value of zero and variance, $\sigma^2_h$ ($\sigma^2_h = 2.097$), which is also called as
the between-group variance. The within-group variance estimated by the model is \( \sigma^2 = 1.839 \).

The intra-class correlation coefficient (ICC) can be calculated as:

\[
ICC = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} = 0.53
\]

This ICC value can be interpreted as the driver-vehicle random effect accounts for 53% of the unexplained variance of speed (Pinheiro and Bates, 2000). ICC is also an indicator of the significance of the random effect. If the ICC is equal to zero, there is no random effect from driver-vehicle subjects. The larger value of ICC shows that more proportions of the unexplained variance are caused by random effects. In this case driver-vehicle random-effect is significant and the mixed-effects model is the appropriate method to use in order to answer the question of interest pertaining to the driver-vehicle effect at the population level.

### 4.5 Model Comparison

The author also estimates another three ordinary regression models following Ordinary Least Squares (OLS) estimation method. This section presents these three models and compares the results with that from the mixed-effects model. The mixed-effects model is labeled as model I. The three ordinary regression models are labeled as model II, III, and IV respectively.

#### 4.5.1 Introduction of Three Ordinary Regression Models

**Model II: Regression Model — Driver-Vehicle Fixed Effect**

As discussed in section 4.1, the driver-vehicle effect can also be treated as the fixed effect instead of the random effect when studying the various influences on speed profiles by individual driver-
vehicle subjects, with the inferential scope to be only on the driver-vehicle subjects included in the sample. The fixed-effect model assumes differences on speeds across driver-vehicle subjects, while the random-effect model explores the difference in error variance.

Model specification:

\[ y_{ijk} = \beta_0 + \beta_1 x_{ijk} + \beta_2 x_{ijk}^2 + \beta_3 x_{ijk}^3 + \beta_4 d_{ij} + \beta_5 d_{2j} + \beta_6 d_{3j} + \beta_7 \text{DRIVER} + \varepsilon_{ijk} \]  

(4-7)

Where,

\text{DRIVER}: This 36-level factor represents main effects for individual driver-vehicle subject in the data, which includes 35 dummy variables (the dropped dummy variable is the reference level),

Other variables and parameters are the same as those introduced in the mixed-effects model.

Fitted model:

\[ v = 3.02 + 2.47t - 0.12t^2 + 0.002t^3 - 0.78\text{Curve} - 0.34\text{Lane} + \hat{\beta}_i \text{DRIVER} \]  

(0.197) (0.038) (0.004) (0.0001) (0.098) (0.097) (0.101)

\[ \sigma = 1.356 \]

\[ R^2 = 0.95 \]

Where,

\( \hat{\beta}_i \) represent coefficients for each driver-vehicle subject, which measures the deviations of each driver-vehicle subject from the referenced driver-vehicle level. Appendix B presents the detailed regression results.

The parenthesized numbers beneath the coefficients are their standard error.
The coefficient of determination, $R^2$, shows that about 95% of the variation in speed can be explained by the regression model. The estimates are all significant in the model. Appendix B presents the detailed regression results.

**Model III: Pooled Regression Model — Pooled Driver-Vehicle Subjects**

Model III assumes that there are no different influences on speeds across driver-vehicle subjects after accounting for the influence of other independent variables.

Model specification:

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk} + \beta_2 x_{ijk}^2 + \beta_3 x_{ijk}^3 + \beta_4 d_{1j} + \beta_5 d_{2j} + \beta_6 d_{3j} + \epsilon_{ijk} \quad (4-9)$$

Where,

Variables and parameters are the same as those introduced in the mixed-effects model.

Fitted model:

$$v = 1.82 + 2.47t - 0.12t^2 + 0.002t^3 - 0.52Curve - 0.54Skew + 0.84Lane$$

$$\sigma = 1.751$$

$$R^2 = 0.91$$

Where,

The parenthesized numbers beneath the coefficients are their standard error.

The coefficient of determination, $R^2$, shows that about 91% of the speed variation can be explained by the model III. The estimates are all significant. Appendix B presents the detailed regression analysis results.
However, the speed profiles tend to vary across driver-vehicle subjects after accounting for other independent variables. The repeated measures from the same driver-vehicle subject are more likely correlated with each other as compared to the measures from different driver-vehicle subjects. Therefore model III may violate the independence assumption required for ordinary linear regression for this regard.

**Model IV: Pooled Regression Model — Pooled Driver-Vehicle and Road Features**

The model IV is another pooled regression model. It assumes that after accounting for the influence of time, there are no different influences on speeds across driver-vehicle subjects and across all locations. This model does not count for the potential influences from driver-vehicle subjects and road physical features. As discussed in Chapter 2, most of previous studies applied a similar analysis approach to develop various acceleration models by only accounting for the relationships between speed vs. time, acceleration vs. time, or acceleration vs. speed.

**Model specification:**

\[
y_{ijk} = \beta_0 + \beta_1 x_{ijk} + \beta_2 x_{ijk}^2 + \beta_3 x_{ijk}^3 + \epsilon_{ijk} \tag{4-11}
\]

Where,

Variables and parameters are the same as those introduced in the mixed-effects model.

**Fitted model:**

\[
v = 1.96 + 2.48t - 0.12t^2 + 0.002t^3 \tag{4-12}
\]

\[
(0.1181) \ (0.051) \ (0.006) \ (0.0002)
\]

\[
\sigma = 1.837
\]

\[
R^2 = 0.90
\]
The coefficient of determination, $R^2$, shows that about 90% of the speed variation can be explained by model IV. The estimates are all significant. Appendix B presents the detailed regression results.

After accounting for time, speed profiles still tend to vary across driver-vehicle subjects and locations. The repeated measures from the same driver-vehicle subject or from the same location are most likely correlated. Therefore model IV may also violate the independent assumption required for ordinary linear regression.

### 4.5.2 Model Comparison

Table 11 presents the estimates of the four fitted models:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Variable</th>
<th>Coefficient</th>
<th>Variable</th>
<th>Coefficient</th>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.24</td>
<td>Intercept</td>
<td>3.02</td>
<td>Intercept</td>
<td>1.82</td>
<td>Intercept</td>
<td>1.96</td>
</tr>
<tr>
<td>Time</td>
<td>2.46</td>
<td>Time</td>
<td>2.47</td>
<td>Time</td>
<td>2.47</td>
<td>Time</td>
<td>2.48</td>
</tr>
<tr>
<td>$\text{Time}^2$</td>
<td>-0.12</td>
<td>$\text{Time}^2$</td>
<td>-0.12</td>
<td>$\text{Time}^2$</td>
<td>-0.12</td>
<td>$\text{Time}^2$</td>
<td>-0.12</td>
</tr>
<tr>
<td>$\text{Time}^3$</td>
<td>0.002</td>
<td>$\text{Time}^3$</td>
<td>0.002</td>
<td>$\text{Time}^3$</td>
<td>0.002</td>
<td>$\text{Time}^3$</td>
<td>0.002</td>
</tr>
<tr>
<td>Curve</td>
<td>-0.79</td>
<td>Curve</td>
<td>-0.78</td>
<td>Curve</td>
<td>-0.52</td>
<td>Curve</td>
<td>-0.52</td>
</tr>
<tr>
<td>Skew</td>
<td>-0.32</td>
<td>Skew</td>
<td>-0.34</td>
<td>Skew</td>
<td>-0.54</td>
<td>Skew</td>
<td>-0.54</td>
</tr>
<tr>
<td>Lane</td>
<td>0.42</td>
<td>Lane</td>
<td>0.44</td>
<td>Lane</td>
<td>0.84</td>
<td>Lane</td>
<td>0.84</td>
</tr>
<tr>
<td>Between-subject variance</td>
<td>$\sigma_p^2 = 2.097$</td>
<td>DRIVER</td>
<td>$\hat{\beta}_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-subject variance</td>
<td>$\sigma^2 = 1.839$</td>
<td>Error variance</td>
<td>$\sigma^2 = 1.839$</td>
<td>Error variance</td>
<td>$\sigma^2 = 3.066$</td>
<td>Error variance</td>
<td>$\sigma^2 = 3.375$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.95</td>
<td>$R^2$</td>
<td>0.95</td>
<td>$R^2$</td>
<td>0.91</td>
<td>$R^2$</td>
<td>0.90</td>
</tr>
</tbody>
</table>

* $\hat{\beta}_i$: estimate for factor “DRIVER”, which measures how far the driver-vehicle subject $i$ is away from the referenced driver-vehicle level (Pinheiro and Bates, 2000).
The author performed an Extra-Sum-of-Squares F-test to evaluate the overall significance of the regression in order to find the best model among the three ordinary regression models. Appendix C presents the model comparison methods and results. Model II is the full model and also is the best one among the three models. This result indicates that there is strong evidence that, after accounting for time, speeds vary for different driver-vehicle subjects and physical road features. Therefore, model I and model II are the best fitting models to predict the speed profiles of the acceleration process by accounting for time, physical road features, and driver-vehicle subjects.

Meanwhile, Figure 13 illustrates the differences between the driver-vehicle subject random effect and the fixed effect graphically. In model I, driver-vehicle subject $i$ shifts the mean speed profile under the corresponding road physical features with a random variable $b_{0i}$. The $b_{0i}$ is assumed to follow the normal distribution with the mean value of zero and constant variance of $\sigma_b^2$. The $\sigma_b^2$ estimates the between driver variability, which is the only parameter requiring estimation in order to describe this random effect.

![Figure 13](image-url)
In model II, each driver-vehicle shifts the speed profile of the referenced driver level up or down by a fixed amount on the intercept term. As we can see, the number of parameters increases with the number of drivers included in the data. For example, 36 driver-vehicle subjects in the data result in 35 parameters that need to be estimated in model II. One must use more degrees of freedom to estimate these parameters in model II than in model I. As we can see, model II only models the specific driver-vehicle subjects in the data. The main interest of this study is the driving behaviors of the driver population where the samples were drawn instead of that of the 36 drivers in the sample. Therefore, model I is a more appropriate model for this study purpose.

The advantage of the mixed-effect model is its ability to interpret driver-vehicle effects as random effects by estimating between-group variance. Furthermore, the results can be inferred to the driver population. This allows the model to be practically applied in a real world context.

The mixed-effects model is a sophisticated statistical modeling method compared to the simpler and more commonly used ordinary regression method. As one of the first attempts to apply the mixed-effect model in speed profile modeling of the acceleration process at signalized intersection locations, this thesis focuses on strengths and weaknesses of different modeling methods. More extensive study in the future may reveal more advantages of applying this powerful and flexible technique in acceleration process modeling.
4.6 Model Assumption Diagnostics for the Mixed-Effects Model

There are two basic assumptions of the mixed-effects model for the error distribution:

1. The within-group errors are normally distributed.

The normality assumption assumes that the random error terms, $e$, follow the normal distribution. The normal probability plot (see Figure 14) presents ordered residuals versus the expected value from a standard normal distribution. In this case the plot shows a relatively straight line. This behavior indicates a reasonable result for the assumption of normality of within-group errors.

![Normal Probability Plot Of Fixed-Effects](image)

Figure 14 Normal Probability Plot Of Fixed-Effects

2. The random effects are normally distributed.

In Figure 15, there appears to be some asymmetry and skew of the distribution of the random effect with a few outliers presented in the plot. The normal probability plot shows somewhat less reasonable results for the assumption of normality of the random effect. However, generally the estimation results are robust for normality, especially the valid inferences which can still be obtained for fixed effects (Verbeke et al., 1997).
4.7 Application Example of the Mixed-Effects Model

The Mixed-effects model introduced in the previous sections can be used to predict speed and acceleration profiles of the acceleration process. This prediction capability can be represented by the following example.

Assume a signalized intersection with the following physical roadway features:

- Straight horizontal alignment: Curve = 0;
- Unskewed intersection layout: Skew = 0;
- Two through traffic lanes: Lane = 1;

These variable values can then be substituted into equation 4-6 to determine expected resulting speed values.
\[ v = 2.24 + 2.46t - 0.12t^2 + 0.002t^3 - 0.79\text{Curve} - 0.32\text{Skew} + 0.42\text{Lane} \]

\[ = 2.24 + 2.46t - 0.12t^2 + 0.002t^3 - 0.79 \times 0 - 0.32 \times 0 + 0.42 \times 1 \]

\[ v = 2.66 + 2.46t - 0.12t^2 + 0.002t^3 \]  \hspace{1cm} (4-13)

The acceleration profile can be calculated by differentiating the equation 4-13 with respect to time \( t \):

\[ a = \frac{dv}{dt} = 2.46 - 0.24t + 0.006t^2 \]  \hspace{1cm} (4-14)

Figure 16 and Table 12 depict the predicted speed profile plot, acceleration profile plot and data for the acceleration process at the example signalized intersection with a straight horizontal alignment, unskewed intersection layout and two through traffic lanes. Traffic simulation and fuel consumption models can apply the predicted speed and acceleration information to replicate the leading vehicle acceleration behavior. For example, it can be used to model the ‘free driving’ behavior of a car-following model in traffic simulation programs. Free driving behavior simulates the leading vehicle accelerating to the desired free flow speed (VISSIM User Manuel 4.00, 2004). This will help the simulation model provide better predictions of the real world scenarios.
Figure 16  Example of Predicted Speed Profile and Acceleration Profile
<table>
<thead>
<tr>
<th>Time (sec)</th>
<th>Speed (m/s)</th>
<th>Speed (mph)</th>
<th>Acceleration Rate (m/s/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.7</td>
<td>6.0</td>
<td>2.46</td>
</tr>
<tr>
<td>1</td>
<td>5.0</td>
<td>11.4</td>
<td>2.226</td>
</tr>
<tr>
<td>2</td>
<td>7.1</td>
<td>16.2</td>
<td>2.004</td>
</tr>
<tr>
<td>3</td>
<td>9.0</td>
<td>20.5</td>
<td>1.794</td>
</tr>
<tr>
<td>4</td>
<td>10.7</td>
<td>24.3</td>
<td>1.596</td>
</tr>
<tr>
<td>5</td>
<td>12.2</td>
<td>27.8</td>
<td>1.41</td>
</tr>
<tr>
<td>6</td>
<td>13.5</td>
<td>30.8</td>
<td>1.236</td>
</tr>
<tr>
<td>7</td>
<td>14.7</td>
<td>33.4</td>
<td>1.074</td>
</tr>
<tr>
<td>8</td>
<td>15.7</td>
<td>35.6</td>
<td>0.924</td>
</tr>
<tr>
<td>9</td>
<td>16.5</td>
<td>37.6</td>
<td>0.786</td>
</tr>
<tr>
<td>10</td>
<td>17.3</td>
<td>39.2</td>
<td>0.66</td>
</tr>
<tr>
<td>11</td>
<td>17.9</td>
<td>40.6</td>
<td>0.546</td>
</tr>
<tr>
<td>12</td>
<td>18.4</td>
<td>41.7</td>
<td>0.444</td>
</tr>
<tr>
<td>13</td>
<td>18.8</td>
<td>42.6</td>
<td>0.354</td>
</tr>
<tr>
<td>14</td>
<td>19.1</td>
<td>43.3</td>
<td>0.276</td>
</tr>
<tr>
<td>15</td>
<td>19.3</td>
<td>43.9</td>
<td>0.21</td>
</tr>
<tr>
<td>16</td>
<td>19.5</td>
<td>44.3</td>
<td>0.156</td>
</tr>
<tr>
<td>17</td>
<td>19.6</td>
<td>44.6</td>
<td>0.114</td>
</tr>
<tr>
<td>18</td>
<td>19.7</td>
<td>44.8</td>
<td>0.084</td>
</tr>
<tr>
<td>19</td>
<td>19.8</td>
<td>45.0</td>
<td>0.066</td>
</tr>
<tr>
<td>20</td>
<td>19.9</td>
<td>45.1</td>
<td>0.06</td>
</tr>
<tr>
<td>21</td>
<td>19.9</td>
<td>45.3</td>
<td>0.066</td>
</tr>
<tr>
<td>22</td>
<td>20.0</td>
<td>45.4</td>
<td>0.084</td>
</tr>
</tbody>
</table>
5 CONCLUSIONS

5.1 Conclusions

This study is based on a current database with one-second interval speed information collected by in-vehicle GPS technology. The author applied the existing data processing approach from previous research (Wang, 2006), and developed a methodology to distinguish speed profiles of the acceleration process from leading and following vehicles in a queue at signalized intersections.

Based on the sensitivity study, comparing average speed profiles for various speed limit groups: 48, 56, 64, 72 km/h (30, 35, 40, 45 mph), the results indicate a difference in speed profiles for each studied speed limit group except for the 56 km/h (35mph) and 64 km/h (30mph) speed limit categories. Therefore, the author recommends that each speed limit group should be modeled separately. This unique modeling approach will also account for functional classifications.

The study generated a third-order polynomial model, with a simplified functional form and a reasonable fit to the study data. This model helps describe speed profiles of the leading vehicle acceleration at signalized intersections on arterial streets with a speed limit of 72km/h (45mph). The author also provided the first large-scale effort to explore the potential effects on speed profiles due to various physical roadway features and driver-vehicle subjects. The mixed-effects model revealed that 53% of unexplained speed variations are caused by different driver-vehicle subjects. It indicates the significant result of the driver-vehicle random effect on speed profiles of the acceleration process. Considering the driver-vehicle effect as a random effect provides the opportunity to make statistical inferences to the driver’s population where the samples were drawn.
The physical roadway features discussed in this study had significant influences in speed profiles of the acceleration process. Drivers tend to choose lower speeds at locations with skewed intersections, alignments with horizontal curvature, or road segments with three through lanes. However, the sample size of horizontal curve features (three locations) is too small. Thus, the author recommends further study on the possible influence of horizontal curvature based on a larger sample size.

Overall, the driver-vehicle effect and roadway physical feature effects must be taken into account in the model, since the model comparison shows significant results and convincing evidence that the speeds differ for different vehicle-driver subjects and different physical roadway features in the acceleration process. The random driver-vehicle influences on speeds indicate significant results. This observation suggests that the random driving behavior acted as a significant and uncontrollable factor in participating driver’s activities. The physical roadway features act as an important controllable factor in participating driver’s activities. However, the roadway design features are only partially able to influence the way people drive. A considerable part of driver’s natural behavior is simply a random effect due to the various natural driving behaviors observed.

5.2 Recommendations for Future Studies

While observations (speeds) are collected over time, the adjacent and near adjacent observations will more likely relate to each other. Therefore, the assumption that response variables are independent with each other may no longer be. In this situation, serial correlation may be a problem in both mixed-effects models and ordinary regression models. Future study on time series analysis will help to address this issue. The correlation structures need to be reconstructed to model dependence of within-group errors in the context of mixed-effects models.
Future studies can also help to answer the questions:

- What are the influences of various road functional classes and speed limit groups on speed profiles of the acceleration process?
- What are the influences of driver gender, driver age and vehicle type on speed profiles of acceleration process?

The author also recommends validating the proposed model based on a reserved or different data set in order to provide more convincing evaluation of prediction results.
BIBLIOGRAPHY


APPENDICES

Appendix A  \( R^2 \): The Coefficient of Determination

The coefficient of determination can be interpreted as the proportion of variability of the dependent variable that can be explained by the model. \( R^2 \) can be calculated by the following formula:

\[
R^2 = \frac{\text{SST} - \text{SSE}}{\text{SST}}
\]

Where,

- \( R^2 \). Coefficient of determination,
- \( \text{SST} \): Total sum of squares, the sum of squares of the difference of the dependent variable \((y_i)\) and its grand mean \((\bar{y})\), \( \text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \), and
- \( \text{SSE} \): Error sum of squares, the sum of squares of the difference of the dependent variable \((y_i)\) and the fitted value \(\hat{y}_i\), \( \text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y})^2 \).
Appendix B  Regression Analysis Results of Ordinary Regression Models

Model II

Fitted model:

\[ v = 3.02 + 2.47t - 0.12t^2 + 0.002t^3 - 0.78\text{Curve} - 0.34\text{Skew} + 0.44\text{Lane} + \beta_1\text{DRV}_{1} \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.02</td>
<td>0.197</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time</td>
<td>2.47</td>
<td>0.038</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time^2</td>
<td>-0.12</td>
<td>0.004</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time^3</td>
<td>0.002</td>
<td>0.0001</td>
<td>0.0000</td>
</tr>
<tr>
<td>Curve</td>
<td>-0.78</td>
<td>0.098</td>
<td>0.0000</td>
</tr>
<tr>
<td>Skew</td>
<td>-0.34</td>
<td>0.097</td>
<td>0.0004</td>
</tr>
<tr>
<td>Lane</td>
<td>0.44</td>
<td>0.101</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv1</td>
<td>1.4598</td>
<td>0.2363</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv2</td>
<td>1.7406</td>
<td>0.2705</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv3</td>
<td>3.4295</td>
<td>0.2689</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv4</td>
<td>1.8120</td>
<td>0.3551</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv5</td>
<td>-0.8778</td>
<td>0.4564</td>
<td>0.0546</td>
</tr>
<tr>
<td>Drv6</td>
<td>-1.8329</td>
<td>0.2738</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv7</td>
<td>-2.3286</td>
<td>0.2419</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv8</td>
<td>-1.8952</td>
<td>0.3400</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv9</td>
<td>-1.5552</td>
<td>0.3329</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv10</td>
<td>-0.2230</td>
<td>0.3226</td>
<td>0.4895</td>
</tr>
<tr>
<td>Drv11</td>
<td>-1.4747</td>
<td>0.3861</td>
<td>0.0001</td>
</tr>
<tr>
<td>Drv12</td>
<td>-1.2227</td>
<td>0.3404</td>
<td>0.0003</td>
</tr>
<tr>
<td>Drv13</td>
<td>-1.6660</td>
<td>0.3670</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv14</td>
<td>1.1341</td>
<td>0.2726</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv15</td>
<td>2.7972</td>
<td>0.3959</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv16</td>
<td>-1.6756</td>
<td>0.2337</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv17</td>
<td>-1.2391</td>
<td>0.3149</td>
<td>0.0001</td>
</tr>
<tr>
<td>Drv18</td>
<td>1.4026</td>
<td>0.2193</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv19</td>
<td>-1.4473</td>
<td>0.2857</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv20</td>
<td>-1.6742</td>
<td>0.3938</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv21</td>
<td>-1.2167</td>
<td>0.3841</td>
<td>0.0016</td>
</tr>
<tr>
<td>Drv22</td>
<td>-0.0655</td>
<td>0.3942</td>
<td>0.8680</td>
</tr>
<tr>
<td>Drv23</td>
<td>-3.5101</td>
<td>0.3107</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv24</td>
<td>1.2699</td>
<td>0.1706</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv25</td>
<td>2.3470</td>
<td>0.4069</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv26</td>
<td>-1.6076</td>
<td>0.2332</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv27</td>
<td>-1.5980</td>
<td>0.1858</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv28</td>
<td>-2.0069</td>
<td>0.1744</td>
<td>0.0000</td>
</tr>
<tr>
<td>Variable</td>
<td>Coefficient</td>
<td>Standard Error</td>
<td>P-value</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Drv29</td>
<td>0.8211</td>
<td>0.3431</td>
<td>0.0168</td>
</tr>
<tr>
<td>Drv30</td>
<td>-0.1203</td>
<td>0.1878</td>
<td>0.5217</td>
</tr>
<tr>
<td>Drv31</td>
<td>1.5321</td>
<td>0.3670</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv32</td>
<td>-0.7213</td>
<td>0.1679</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv33</td>
<td>-0.5573</td>
<td>0.1921</td>
<td>0.0038</td>
</tr>
<tr>
<td>Drv34</td>
<td>-1.5358</td>
<td>0.1895</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drv35</td>
<td>0.3523</td>
<td>0.3404</td>
<td>0.3008</td>
</tr>
</tbody>
</table>

Residual standard error 1.356 on 2260 degree of freedom

R² 0.95

**Model III**

Fitted model:

\[ v = 1.82 + 2.47t - 0.12t^2 + 0.002t^3 - 0.52\text{Curve} - 0.54\text{Skew} + 0.84\text{Lane} \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.82</td>
<td>0.141</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time</td>
<td>2.48</td>
<td>0.049</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time^2</td>
<td>-0.12</td>
<td>0.006</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time^3</td>
<td>0.002</td>
<td>0.0002</td>
<td>0.0000</td>
</tr>
<tr>
<td>Curve</td>
<td>-0.52</td>
<td>0.093</td>
<td>0.0000</td>
</tr>
<tr>
<td>Skew</td>
<td>-0.54</td>
<td>0.079</td>
<td>0.0000</td>
</tr>
<tr>
<td>Lane</td>
<td>0.84</td>
<td>0.080</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Residual standard error 1.751 on 2295 degree of freedom

R² 0.91
Model IV

Fitted model:

\[ v = 1.96 + 2.48t - 0.12t^2 + 0.002t^3 \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.96</td>
<td>0.118</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time</td>
<td>2.48</td>
<td>0.051</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time^2</td>
<td>-0.12</td>
<td>0.006</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time^3</td>
<td>0.002</td>
<td>0.0002</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Residual standard error 1.837 on 2298 degree of freedom

R^2 0.90
Appendix C  Model Comparison Results

Model III vs. Model IV

Full model: Model III

\[ v = 1.82 + 2.47t - 0.12r^2 + 0.002r^3 - 0.52\text{Curve} - 0.54\text{Skew} + 0.84\text{Lane} \]

Reduced model: Model IV

\[ v = 1.96 + 2.48t - 0.12r^2 + 0.002r^3 \]

The Null hypothesis can be tested by F-test:

Ho: \( \beta_4 = \beta_5 = \beta_6 = 0 \)

(3 parameters are tested)

The alternative hypothesis:

Ho: at least one of the parameters is not equal to zero

The formula is used to calculate F-statistic is:

\[
F - \text{statistic} = \frac{\text{extra sum of squares}}{\text{number of test parameter}} \left(\frac{\sigma^2_{\text{full}}}{\sigma^2_{\text{full}}}\right)
\]

Extra sum of squares = sum of squared residuals from reduced model – sum of squared residuals from full model

<table>
<thead>
<tr>
<th>Sum-of-Squares F-test Results: Model III vs. Model IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of squared residual</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Full model</td>
</tr>
<tr>
<td>Reduced model</td>
</tr>
</tbody>
</table>
The very small p-value \( (p-value = \text{pr}(F_{3,2295} > 78.27) = 0.000) \) indicates a strong evidence that at least one of the coefficients differs from zero. The test result shows strong evidence to reject the null hypothesis. Therefore, the full model, model III, provides a better fit to the data.

**Model II vs. Model III**

**Full model: Model II**

\[
v = 3.02 + 2.47t - 0.12t^2 + 0.002t^3 - 0.78\text{Curve} - 0.34\text{Skew} + 0.44\text{Lane} + \hat{\beta}_i \text{ DRIVER}
\]

**Reduced model: Model III**

\[
v = 1.82 + 2.47t - 0.12t^2 + 0.002t^3 - 0.52\text{Curve} - 0.54\text{Skew} + 0.84\text{Lane}
\]

The Null hypothesis can be tested by F - test:

Ho: \( \beta_7 = \beta_8 = \ldots = \beta_{44} = 0 \)

(35 parameters are tested)

The alternative hypothesis:

Ho: at least one of the parameters is not equal to zero

The formula is used to calculate \( F\text{-statistic} \) is:

\[
F\text{-statistic} = \frac{\text{extra \, sum \, of \, squares}}{\frac{\text{number \, of \, test \, parameter}}{\hat{\sigma}_\text{full}}^2}
\]

Extra sum of squares = sum of squared residuals from reduced model – sum of squared residuals from full model

| Sum-of-Squares F-test Result: Model II vs. Model III |
|-----------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                  | Sum of squared residual | Degree of Freedom | Extra-Sum of Squares | Residual SE | F-statistic | p-value |
| Full model                      | 4157.6            | 2260            | 2881.2          | 1.356         | 44.7         | 0.0000          |
| Reduced model                   | 7038.8            | 2295            |                 |               |              |                 |
The very small p-value (p-value = pr (F_{35,2260} > 44.7) = 0.000) indicates a strong
evidence that at least one of the coefficients differs from zero. The test result shows
strong evidence to reject the null hypothesis. The full model, model II, provides a better
fit to the data.