Naturalistic Driving Data for the Analysis of Car-Following Models

John David Sangster

Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
in
Civil Engineering

Hesham A. Rakha, Chair
Jianhe Du
Kathleen L. Hancock

December 5th, 2011
Blacksburg, VA

Keywords: car-following modeling; traffic modeling; naturalistic driver behavior; traffic flow theory.

Naturalistic Driving Data for the Analysis of Car-Following Models

John David Sangster

Abstract

The driver-specific data from a naturalistic driving study provides car-following events in real-world driving situations, while additionally providing a wealth of information about the participating drivers. Reducing a naturalistic database into finite car-following events requires significant data reduction, validation, and calibration, often using manual procedures. The data collection performed herein included: the identification of commuting routes used by multiple drivers, the extraction of data along those routes, the identification of potential car-following events from the dataset, the visual validation of each car-following event, and the extraction of pertinent information from the database during each event identified.

This thesis applies the developed process to generate car-following events from the 100-Car Study database, and applies the dataset to analyze four car-following models. The Gipps model was found to perform best for drivers with greater amounts of data in congested driving conditions, while the Rakha-Pasumarthy-Adjerid (RPA) model was best for drivers in uncongested conditions. The Gipps model was found to generate the lowest error value in aggregate, with the RPA model error 21 percent greater, and the Gaxis-Herman-Rothery model (GHR) and the Intelligent Driver Model (IDM) errors 143 percent and 86 percent greater, respectively. Additionally, the RPA model provides the flexibility for a driver to change vehicles without the need to recalibrate parameter values for that driver, and can also capture changes in roadway surface type and condition. With the error values close between the RPA and Gipps models, the additional advantages of the RPA model make it the recommended choice for simulation.

Acknowledgements

To my adviser, Dr. Hesham Rakha, thank you for providing me the opportunity to work with you and your research team while pursing my graduate degrees. Your support, encouragement, and expectation of accomplishment have been instrumental in helping me to achieve my goals. It has been an honor to work for you on this research project, and I hope that it is just the beginning of a long and fruitful collaborative relationship.

To my committee member, Dr. Jianhe Du, thank you for your continued support and constructive criticism, which is always appreciated. Your GIS data extraction methods began this process over a year ago, and you've been instrumental in helping this project to overcome many speed bumps along the way.

To my committee member, Dr. Kathleen Hancock, thank you for providing oversight on this project as a member of my committee.

To my wife Kerry, my son Jonas, my daughter Emily, and our as-yet unnamed addition due to arrive in March, I thank you from the bottom of my heart. Every moment I spent on this thesis was a moment that I wasn't able to be there with you, or be there for you. You have shown your love for me by allowing me to follow my dreams, and supporting me the whole way through. My life with you is one of joy and wonder, and knowing that my academic success directly results in our collective success is the best motivation anyone could ask for. I believe that together, we will achieve great things.

To my parents, Dave and Dawn, thank you for giving your children wings, and showing us that it's okay to fly. We love you.

Table of Contents

Abstract ii
Acknowledgements iii
Table of Contents iv
List of Figures vi
List of Tables vii
List of Variables viii
Chapter 1 Introduction 1
1.1 Objectives of the Thesis 2
1.2 Thesis Structure 2
Chapter 2 Literature Review of Car-Following Models and Naturalistic Data 4
2.1 Introduction 4
2.2 Car-following Models 4
2.2.1 Notation for Model Formulation 5
2.2.2 Car-following Models Examined 5
2.2.3 Calibration of Model Parameters 13
2.2.4 Parameter Variation and Correlation 16
2.3 Naturalistic Data 16
Chapter 3 Processing Naturalistic Data for Car-following Model Calibration 18
3.1 Introduction to the Data Processing of a Naturalistic Study 18
3.2 Generating Data from a Naturalistic Driving Study 18
3.2.1 Data Collection 18
3.2.2 Data Management 19
3.2.3 Defining a Set of Goals for Data Processing 19
3.2.4 Data Selection 20
3.2.5 Event Processing 20
3.2.6 Data Validation 20
3.3 Case Study: Hundred Car Study 21
3.3.1 Use of GIS to Select Drivers and Roadway Sections 22
3.3.2 Process to Obtain Leading/Following Pairs 23
3.3.3 Selection of Leading/Following Vehicle Pairs 24
3.3.4 Smoothing of Data Elements 26
3.3.5 Interpolation of Data Elements 26

3.3.6 Summary of Data Analyzed 27
3.3.7 Visualization of the Dataset 28
3.4 Conclusions Regarding the Data Processing of a Naturalistic Study 31
3.4.1 Review of Data Generation Process 32
3.4.2 Review of the Case Study Procedures 33
Chapter 4 Analysis of Car-Following Models using Naturalistic Data 36
4.1 Introduction to the Analysis of Car-Following Models 36
4.2 Summary of Data Generated from Hundred Car Study 36
4.2.1 Tabular Representation of Data for Analysis 36
4.2.2 Visual Representation of Data for Analysis 37
4.3 Calculation of Optimized Model Parameters 38
4.3.1 Defining Parameter Boundary Conditions 38
4.3.2 Optimized Parameter Values 41
4.3.3 Visual Representations of the Solution Set 44
4.4 Analysis of Results 53
4.4.1 Comparison of Macroscopic Error Measures 53
4.4.2 Comparison of Microscopic Error Measures 54
4.4.3 Correlation of Parameters 58
4.5 Conclusions Regarding the Analysis of Car-following Models 60
4.5.1 Review of the Calculation of Optimized Model Parameters 60
4.5.2 Review of the Analysis of Results 62
Chapter 5 Thesis Conclusions and Recommendations for Further Research 65
5.1 Assess the Potential Benefits of Naturalistic Data for the Analysis of Car-Following Behavior 65
5.2 Develop a Process to Extract Car-Following Data from a Naturalistic Database 66
5.3 Generate a Sample Dataset from the Hundred Car Study 66
5.4 Perform Sample Car-Following Model Calibration 67
5.4.1 Review of the Calculation of Optimized Model Parameters 68
5.4.2 Review of the Analysis of Results 69
5.5 Identify Research Topics to Leverage Unique Naturalistic Data 70
APPENDIX A References 72
APPENDIX B IRB Permission Letter 77

List of Figures

Figure 2-1 Data source matrix linking bias and driver information. 17

Figure 3-1 Geographic Information System (GIS) data representation - example D304. 22
Figure 3-2 Process diagram for generation of car-following data from selected GIS data. 24
Figure 3-3 Example of DART visualization for data reduction. 25
Figure 3-4 Effect of interpolation on sample lead and follower vehicle speed profiles. 27
Figure 3-5 Combined macroscopic diagram for car-following event data. 29
Figure 3-6 Data expressed as aggregate results by event. 30
Figure 3-7 Graphical review of the data generation process. 33

Figure 4-1 Macroscopic diagrams of car-following event data: Driver 316. 38
Figure 4-2 Effect on error of changing the perception-reaction time in the Gipps model. 44
Figure 4-3 Macroscopic diagrams of car-following event data: Driver 124. 45
Figure 4-4 Macroscopic diagrams of car-following event data: Driver 304. 46
Figure 4-5 Macroscopic diagrams of car-following event data: Driver 316. 47
Figure 4-6 Macroscopic diagrams of car-following event data: Driver 350. 48
Figure 4-7 Macroscopic diagrams of car-following event data: Driver 358. 49
Figure 4-8 Macroscopic diagrams of car-following event data: Driver 363. 50
Figure 4-9 Macroscopic diagrams of car-following event data: Driver 367. 51
Figure 4-10 Macroscopic diagrams of car-following event data: Driver 462. 52
Figure 4-11 Examination of individual car-following event: Driver 367, 163614.29. 55
Figure 4-12 Examination of individual car-following event: Driver 304, 110803.01. 56
Figure 4-13 Examination of individual car-following event: Driver 124, 129453.02. 57
Figure 4-14 Distribution of event error for each of the simulation models. 58

List of Tables

Table 2-1	Most Reliable Estimates of Parameters within the GHR model.	7
Table 3-1	Quantity of data included in this study.	28
Table 4-1	Quantity of data included in this study.	37
Table 4-2	Method of parameter calibration, and associated car-following models.	39
Table 4-3	Parameters values calibrated from observed values.	40
Table 4-4	Parameter values calibrated from macroscopic optimization methods.	40
Table 4-5	Vehicle dynamics parameters.	41
Table 4-6	Parameter values calibrated from microscopic optimization methods.	41
Table 4-7	Results for calibrated parameters of the GHR model.	42
Table 4-8	Results for calibrated parameters of the Gipps model.	42
Table 4-9	Results for calibrated parameters of the IDM model.	43
Table 4-10	Results for calibrated parameters of the RPA model.	43
Table 4-11	Comparison of error function values.	53
Table 4-12	Linear correlation coefficient values between parameters.	59

List of Variables

A_f	*Vehicle frontal area (meters squared)*
C_d	*Drag coefficient (unitless)*
C_r	*Rolling coefficient (unitless)*
F_{n+1}	*Resultant force (Newtons)*
M_{ta}	*Mass of the vehicle on the tractive axle (kilograms), taken to be 55% of total vehicle mass*
P	*Vehicle power (kilowatts)*
R_{n+1}	*Total resistive force (Newtons)*
c_{r2}, c_{r3}	*Rolling resistance coefficients (unitless)*
c_1, c_2, c_3	*Parameters used for the Van Aerde steady-state model*
k_{jam}	*Density at jam density (veh/km/ln)*
l	*Calibration parameter for the GHR model (unitless)*
l_n	*Approximate length (m) of the leader vehicle, vehicle n, taken to be 4.5 m*
$mass_{n+1}$	*Vehicle mass (kg) of the follower vehicle, vehicle n + 1*
n	*Vehicle number n: the leading vehicle*
$n+1$	*Vehicle number* $n+1$*: the following vehicle*
q	*Flow rate (veh/hr/ln)*
q_c	*Flow rate at capacity (veh/hr/ln)*
q_j	*Flow rate at jam density (veh/hr/ln)*
t	*The current time step*
x	*Vehicle position (m)*
$\dot{x}$	*Vehicle velocity (m/s)*
$\dot{x}_d$	*Velocity desired by driver of vehicle* $n+1$ *(m/s)*
$\dot{x}_p$	*Speed at which vehicle attains its full power (m/s)*
$\ddot{x}$	*Vehicle acceleration (m/s^2)*
$\ddot{x}_{d-max}$	*Maximum acceleration rate desired by driver of vehicle* $n+1$ *(m/s^2)*
$\ddot{x}_{d-min}$	*Maximum deceleration rate desired by driver of vehicle* $n+1$ *(m/s^2)*

$\ddot{x}_{d-min,n}$	*The estimate made by the driver of vehicle* $n+1$ *of the maximum deceleration rate desired by the driver of vehicle* n *(m/s*^{2}*)*
w	*Vehicle weight-to-power ratio*
z	*Calibration parameter for the GHR model (unitless)*
Δ	*The change in value for a given variable at a given time step between the leading and following vehicle*
Δx_j	*Gross distance headway at standstill, or distance at jam density (meters)*
α	*Calibration parameter for the GHR model*
β	*Gear reduction factor (unitless)*
γ	*Driver throttle level input (unitless), taken from the maximum measured percentage throttle depression for each driver herein*
δ	*Acceleration exponent (unitless)*
η	*Driveline efficiency (unitless), taken as 0.7 herein*
μ	*Coefficient of roadway adhesion (unitless), taken as 0.6 herein*
τ	*The perception reaction time (seconds)*
τ_m	*Perception reaction time calculated from macroscopic measures (seconds)*
τ_S	*Safe time headway (seconds)*

Chapter 1 Introduction

Traffic-simulation tools make it possible to assess the results of alternative designs, without either the large cost of constructing each alternative, or the increased delay incurred by traffic during construction. The usefulness of the simulation is directly related to the accuracy of its predictions, a large part of which is attributed to the validity of the car-following model used at the core of the traffic simulation tool. An accurate prediction of the interaction between two adjacent vehicles extrapolates into an accurate prediction of overall traffic behavior and conditions that can be expected to occur in the real world.

The research in this thesis develops and tests a procedure for processing a large-scale naturalistic database for mobility analysis, specifically with application to the calibration of car-following models. To validate a given car-following model, observations must be made of real lead-follower vehicle pairs, with comparisons made between the simulated follower and the observed follower behavior. The data used to validate car-following models in the past has primarily come from test tracks, simulators, loop detectors, or aerial photography or videography. These studies either have driver-specific information and bias from lack of real-world driving situations, or accurate real-world driving data from anonymous drivers.

Naturalistic data involves the instrumentation of a vehicle to record driving behavior during normal every-day driving, with data collection generally lasting for a long period of time. Studies which collect naturalistic data have primarily been funded to conduct transportation safety research, though at the conclusion of these studies there remains a dataset available for further use, with a variety of potential applications, including the analysis of car-following models.

In this thesis, a process is developed to extract finite-length car-following events from a database of naturalistic driving data. A case study is used to apply the proposed procedure to the Hundred Car Study database, with a small fraction of the total database processed into car-following events. The processed events are then utilized to evaluate four car-following models including: the Gaxis-Herman-Rothery model (GHR); the Gipps model; the Intelligent Driver Model (IDM); and the Rakha-Pasumarthy-Adjerid model (RPA). The comparative analysis of these events is not a new contribution to the literature, but is instead an application of the unique dataset generated from the naturalistic driving data to these car-following models.

The analysis of parameter calibration results for the various models includes an examination of the degree to which any two parameters within a given model are linearly correlated with each other. The examination of parameter correlation has been done in previous literature studies, but not with naturalistic data. It is hypothesized that a larger pool of processed car-following events, from a broader variety of drivers, in combination with the personality profile information collected on each driver, may lead to a better understanding of the psychological underpinnings of driver behavior.

1.1 Objectives of the Thesis

The objectives of this thesis are: (1) to assess the potential benefits of using naturalistic driving data in the analysis of car-following behavior; (2) to develop a process for extracting car-following events from a naturalistic database; (3) to generate a sample dataset through the use of a data extraction case study on the Hundred Car Study; (4) to perform sample car-following model calibration using the data reduced from the case study; and (5) to identify research topics which can leverage the unique data generated from a naturalistic driving study. The procedures outlined in this thesis are intended to serve as a guide for researchers to perform data reduction on naturalistic datasets, including the completed Hundred Car Study, the on-going Strategic Highway Research Program 2 Study, and other future studies.

1.2 Thesis Structure

This thesis is organized into five chapters, beginning with this introduction as the first chapter. The second chapter is intended to provide a review of the relevant literature, including car-following model formulations and the sources of data available for analyzing car-following models. The third chapter defines the process of obtaining pertinent vehicle trajectory information for analyzing car-following models from a naturalistic driving dataset, and examines a data reduction case study of the Hundred Car Study. The fourth chapter analyzes a variety of car-following models using the data collected from chapter three. The fifth chapter provides the conclusions of the thesis and provides recommendations for future work.

The second chapter literature review covers both car-following models, and the data sources used to analyze car-following models. The literature review of car-following models includes an explanation of the notation used throughout the thesis, a discussion on the formation

of a variety of specific car-following models, a discussion on the calibration of model parameters, and a review of issues related to parameter variation and correlation. The specific car-following models discussed are limited to those which are investigated in later sections of this thesis, and include: the Gaxis-Rothery-Herman model (GHR); the Gipps' model; the Intelligent Driver Model (IDM); and the Rakha-Pasumarthy-Adjerid model (RPA). The discussion on the data sources used to analyze car-following models is less extensive than the discussion on the models themselves.

The third chapter examines the process by which data reduction can be performed on a naturalistic study dataset, for the purpose of application to the study of car-following models. The data reduction specifically focuses on the steps involved in selecting leader-follower vehicle pairs from the large dataset. As an overview, the process of generating the data generally includes: data collection; data management; a defined set of goals for data processing; the selection of a subset of data; the processing of individual events; and the validation of the data. Subsequently, the process developed to perform data reduction is applied as a case study to the Hundred Car Study, beginning with the selection of the data subset using of a Geographic Information Systems (GIS) application to identify routes and ending with the visualization of the collected data.

The fourth chapter utilizes the dataset generated in chapter three to perform a comparative analysis of the four car-following models discussed in the literature review. A summary is provided of the data used for the analysis, a justification is given for the parameter boundary conditions used, an optimization function is defined, the parameter values calibrated to minimize the objective function, and a comparison of errors is shown between the various models analyzed. A visual representation of the derived solutions is provided, and an analysis of the results is conducted. Specific analysis of the results includes an examination of linear correlation between the various parameters within a given model.

The fifth and final chapter provides summary conclusions for the thesis, reviewing the contributions of the work herein and presenting recommendations for future work.

Chapter 2 Literature Review of Car-Following Models and Naturalistic Data

2.1 Introduction

Funding for large-scale naturalistic driving data studies has thus far been primarily for the purpose of conducting safety analysis, with application of the dataset in a mobility context done as a secondary objective. The method of collecting data in a "naturalistic" way can provide a great deal of safety-related information including: detailed pre-crash/crash information; measures of driver distraction or drowsiness; instances of aggressive driving or driver error; and the incorporation of vehicle dynamics. Similar advantages can be found in the application of naturalistic data in the context of mobility applications, where naturalistic data is able to provide detailed driver information, while also providing real-world driving data with a minimal amount of bias.

This thesis seeks to develop a procedure for processing a large-scale naturalistic database for mobility applications, with a focus on the calibration of car-following models. To validate a given car-following model, observations must be made of real lead-follower vehicle pairs, with comparisons made between the simulated follower and the observed follower. The data used to validate car-following models in the past has primarily come from test tracks, simulators, or loop detectors. These studies either have driver-specific information and bias from lack of real-world driving situations, or accurate real-world driving data from anonymous drivers. Only recently has naturalistic driving data become available as an analysis data source.

2.2 Car-following Models

There are many explanations of the importance of car-following models in microscopic traffic-simulation in the literature, with an informative and succinct example being that of Panwai and Dia from 2005 [1], who explain that traffic-simulation tools provide the ability to evaluate and control different scenarios in an environment that doesn't disrupt real-world traffic conditions. The data available from naturalistic driving behavior studies can provide unique insight into the quality of a car-following model, particularly in its ability to accommodate diverse individual drivers.

Each car-following model predicts the time-space trajectory of a following vehicle when provided the time-space trajectory of a leading vehicle, and the original location and velocity of the following vehicle. The parameters of each model are calibrated such that the resulting simulated behavior matches the observed behavior as closely as possible. To arrive at these calibrated values it is necessary to generate observed and simulated vehicle trajectories, define limiting bounds for the parameter values, define an error objective function, and select an optimization methodology to minimize the defined error objective function.

2.2.1 Notation for Model Formulation

As presented in this thesis, the car-following models share a common notation, which includes:

n = *vehicle number n: the leading vehicle*

$n+1$ = *vehicle number $n+1$: the following vehicle*

x = *vehicle position*

$\dot{x}$ = *vehicle velocity*

$\ddot{x}$ = *vehicle acceleration*

$\ddot{x}_{d-max}$ = *maximum acceleration rate desired by driver of vehicle $n+1$*

$\ddot{x}_{d-min}$ = *maximum deceleration rate desired by driver of vehicle $n+1$*

$\dot{x}_d$ = *velocity desired by driver of vehicle $n+1$*

Δ = *the change in value for a given variable at a given time step between the leading and following vehicle*

Δx_{jam} = *gross distance headway at standstill*

t = *the current time step*

τ = *the perception reaction time*

An example of the notation is:

$\Delta\dot{x}_{n\to n+1}(t-\tau)$ = *the change in velocity between the leading and following vehicle that occurred at the ($t-\tau$) time-step.*

2.2.2 Car-following Models Examined

The GHR type of model, also known as the GM model, involves research spanning from the late 1950's until the middle of the 1960's [2], [3]. The CORSIM transportation simulation software

operates based on the PITT model, which has been shown to be analogous to the Pipes car-following model [4–6] under state-state conditions.

Gipps (1981) developed a model as a response to the existing proliferation of GHR type models [7]. The model seeks to: mimic the behavior of real traffic; use parameters that correspond to characteristics of drivers and vehicles so that most can be assigned without calibration procedures; and provide reasonable results if the interval of speed and position calculations is similar to that of the reaction time. The car-following models implemented by the AIMSUN traffic simulation software package are based on the Gipps model, though it has been noted that the software is continually being developed, and it is unknown how or if the underlying algorithms have been modified due to the proprietary nature of the software [1].

The Intelligent Driver Model (IDM) was developed as a high-fidelity model trying to reproduce traffic as realistically as possible, with a small number of parameters, and realistic breaking reactions [8]. The authors of the IDM cite the Gipps model as achieving these goals, though they are trying to overcome the limitations of the Gipps model, in that it shows no traffic instabilities or hysteresis effects for vanishing fluctuations.

Also similar to the Gipps model, the RPA model seeks to use parameters that correspond to characteristics of drivers and vehicles, enabling the calibration parameters to be assessed based on objective measures [9–11]. The RPA model imposes limitations on the steady-state Van Aerde model to ensure collision avoidance, and to account for vehicle dynamics constraints, as demonstrated by Park et al [12]. The RPA car-following model is implemented in the INTEGRATION traffic simulation software package.

All four models are described in detail in this chapter given that they are used in Chapter four of the thesis.

2.2.2.1 Gaxis-Rothery-Herman Model (GHR)

The GHR type of model, also known as the GM model, involves research spanning from the late 1950's until the middle of the 1960's [2], [3]. A discussion on the development of the GHR models, also known as the General Motors models, along with typical parameter values, can be found in May [13]. The fifth and final formulation of the model, sometimes referred to as the GM-5 model, generalizes the previous four forms and is seen in Equation (1).

$$\ddot{x}_{n+1}(t) = \left\{\frac{\alpha[\dot{x}_{n+1}(t)]^z}{[\Delta x_{n\to n+1}(t-\tau)]^l}\right\} \cdot [\Delta\dot{x}_{n\to n+1}(t-\tau)] \quad (1)$$

Where: α, z, l = *calibration parameters*

$\dot{x}_{n+1}(t)$, $\Delta x_{n\to n+1}(t-\tau)$, *and* $\Delta\dot{x}_{n\to n+1}(t-\tau)$ *are as defined in section 2.2.1*

The GHR model is particularly significant in the pantheon of car-following models as many other microscopic and macroscopic models can be represented as special cases of the GHR model. The first single-regime traffic stream model proposed by Greenshields in 1935 is one such case [14]. It can be expressed as a GHR model taking the *z* and *l* parameters to be 0 and 2, respectively, as seen in Equation (2).

$$\ddot{x}_{n+1}(t) = \left\{\frac{\alpha[\dot{x}_{n+1}(t)]^0}{[\Delta x_{n\to n+1}(t-\tau)]^2}\right\} \cdot [\Delta\dot{x}_{n\to n+1}(t-\tau)] \quad (2)$$

The CORSIM transportation simulation software operates based on the PITT model, which has been shown to be analogous to the Pipes car-following model [4–6] under steady-state conditions. The formulation of the Pipes model can be seen to be a special case of the GHR model with the *z* and *l* parameters set to zero, as seen in Equation (3).

$$\ddot{x}_{n+1}(t) = \alpha[\Delta\dot{x}_{n\to n+1}(t-\tau)] \quad (3)$$

Brackstone and McDonald (1999) compiled a list of parameter values used by previous research efforts, as summarized in **TABLE 2-1**. [15]

TABLE 2-1 Most Reliable Estimates of Parameters within the GHR model[a]

Source	*z*	*l*	Approach
Chandler et al. (1958) [2]	0	0	Micro
Herman and Potts (1959) [16]	0	1	Micro
Hoefs (1972) (dcn no brk/dcn brk/acn) [17]	1.5/0.2/0.6	0.9/0.9/3.2	Micro
Treiterer and Myers (1974) (dcn/acn) [18]	0.7/0.2	2.5/1.6	Micro
Ozaki (1993) (dcn/acn) [19]	0.9/-0.2	1/0.2	Micro

[a]Key: dcn/can: deceleration/acceleration; brk/no brk: deceleration with and without braking.

As the other models analyzed herein are representative of higher order models, the GM-5 formulation of the GHR models will be used in the analysis portion of this thesis. Additionally,

in line with previously documented calibration procedures for the GM-5 car following model, as seen in Table 2, the *z* and *l* parameters will be calibrated separately under acceleration and deceleration conditions, which provides a like number of calibration parameters among the models analyzed herein.

2.2.2.2 Gipps Model

Gipps (1981) developed a model as a response to the existing proliferation of GHR type models [7]. The Gipps model seeks to: mimic the behavior of real traffic; use parameters that correspond to characteristics of drivers and vehicles so that most can be assigned without calibration procedures; and provide reasonable results if the interval of speed and position calculations is similar to that of the reaction time. The model itself takes the form of Equation (4).

$$\dot{x}_{n+1}(t) = \min of:$$

$$\dot{x}_{n+1}(t-\tau) + 2.5\ddot{x}_{d-max} \cdot \tau \cdot \left[1 - \frac{\dot{x}_{n+1}(t-\tau)}{\dot{x}_d}\right] \cdot \sqrt{0.025 + \frac{\dot{x}_{n+1}(t-\tau)}{\dot{x}_d}} \quad , \tag{4}$$

$$\ddot{x}_{d-min} \cdot \tau + \sqrt{\left(\ddot{x}_{d-min}{}^2 \cdot (\tau)^2 - \ddot{x}_{d-min} \cdot \begin{Bmatrix} 2\left[\Delta x_{n\to n+1}(t-\tau) - \Delta x_j\right] - \\ \left[\dot{x}_{n+1}(t-\tau) \cdot \tau\right] - \frac{\dot{x}_n(t-\tau)^2}{\ddot{x}_{d-min,n}} \end{Bmatrix}\right)}$$

Where: $\ddot{x}_{d-min,n} = min\left(-3.0\, m/s^2, \frac{\ddot{x}_{d-min}-3.0}{2}\right)$, *the estimate made by the driver of vehicle* $n+1$ *of the maximum deceleration rate desired by the driver of vehicle* n

τ, $\Delta x_{n\to n+1}(t-\tau)$, Δx_j, $\dot{x}_d$, $\dot{x}_n(t-\tau)$, $\dot{x}_{n+1}(t)$, $\dot{x}_{n+1}(t-\tau)$, $\ddot{x}_{d-max}$ *, and* $\ddot{x}_{d-min}$ *are as defined in section 2.2.1*

The first component of the two-regime equation applies in traffic conditions with unconstrained flow, also known as steady-state flow conditions. The second component addresses conditions in constrained flow, where the velocity of the follower vehicle is based primarily on the time/space headway the must necessarily maintain a safe distance between it and the lead vehicle. Since the original publication of Gipps' model, research publications examining the calibration of the Gipps model have brought up critique in two areas: the feasible

range for the expected maximum deceleration of the leading vehicle; and the relationship between the perception-reaction time and the time-step [20–22].

Wilson primarily examined the wave-stability of the Gipps model, and specifically looks at cases where $\ddot{x}_{d-min} > \ddot{x}_{d-min,n}$ or $\ddot{x}_{d-min} < \ddot{x}_{d-min,n}$ [20]. Wilson determined that in some cases that the anticipated deceleration of the leading vehicle is expected to be less than the desired maximum deceleration of the following vehicle, the resulting time-space wave phenomenon becomes unstable. If the original formulation of the Gipps model is used, the implication would be that any follower vehicle with a desired deceleration in excess of -3.0 m/s^2 would fall into the category of potentially being infeasible. For the purposes of the model calibration performed herein, Gipps' original formulation for the anticipated leading vehicle maximum deceleration is used.

Rakha et. al, building on the work of Wilson, developed a method to express the steady-state portion of the Gipps model on the macroscopic fundamental diagram [21], [22]. The speed-flow relationship can be computed using Equation (5), and subsequently, the perception-reaction time can be computed using Equation (6).

$$q = \frac{1{,}000 * \dot{x}_{n+1}}{\Delta x_j + \frac{1}{2.4} * \tau * \dot{x}_{n+1} + \frac{1}{25.92 * \ddot{x}_{d-min}} * \left(1 - \frac{\ddot{x}_{d-min}}{\ddot{x}_{d-min,n}}\right) * \dot{x}_{n+1}{}^2} \quad (5)$$

$$\tau_m = 2.4\left(\frac{1{,}000}{q_c} - \frac{\Delta x_j}{\dot{x}_{capacity}} - \frac{\dot{x}_c}{25.92 * \ddot{x}_{d-min}}\left(1 - \frac{\ddot{x}_{d-min}}{\ddot{x}_{d-min,n}}\right)\right) \quad (6)$$

Where: q = *flow rate (veh/hr/ln)*

q_c = *flow rate at capacity (veh/hr/ln)*

$\dot{x}_c$ = *Speed at capacity (m/s)*

τ_m = *perception reaction time calculated from macroscopic measures (s)*

τ, Δx_j, $\ddot{x}_{d-max}$, *and* $\ddot{x}_{d-min}$ *are defined in section 2.2.1*

$\ddot{x}_{d-min,n}$ *is defined previously in this section*

This methodology may be verified by alternatively using the macroscopically optimized τ_m and by allowing τ to be optimized to minimize an error function during microsimulation.

2.2.2.3 Intelligent Driver Model (IDM)

The Intelligent Driver Model (IDM) was developed as a high-fidelity model trying to reproduce traffic as realistically as possible, with a small number of parameters, and realistic breaking reactions [8]. The IDM is seen in Equation (7).

$$\ddot{x}_{n+1}(t) = \ddot{x}_{d-max}\left\{1 - \left[\frac{\dot{x}_{n+1}(t)}{\dot{x}_d}\right]^{\delta} - \left[\frac{(\Delta x_j - l_n) + (\dot{x}_{n+1}(t)\cdot\tau_S) + \frac{\dot{x}_{n+1}(t)[\Delta\dot{x}_{n\to n+1}(t)]}{2\cdot\sqrt{\ddot{x}_{d-max}\cdot|\ddot{x}_{d-min}|}}}{\Delta x_{n\to n+1}(t) - l_n}\right]^2\right\} \quad (7)$$

Where: δ = *acceleration exponent*

τ_S = *safe time headway*

l_n = *approximate length of vehicle n, taken to be 4.5 meters*

$\Delta x_{n\to n+1}(t)$, Δx_j, $\dot{x}_d$, $\Delta\dot{x}_{n\to n+1}(t)$, $\dot{x}_{n+1}(t)$, $\ddot{x}_{d-max}$, *and* $\ddot{x}_{d-min}$ *are defined in section 2.2.1*

The IDM examines the maximum possible acceleration, and combines both acceleration and deceleration strategies into the formulation. The acceleration strategy takes the instantaneous ratio of the current velocity and desired velocity, and raises this value by the acceleration exponent, taken to be a default value of 4 in the original documentation. The deceleration strategy examines the desired spacing as a function of velocity of the following vehicle and the difference in velocity between the leading and following vehicles, and squares the result of this desired spacing divided by the current space headway. Effectively, this formulation balances the pull of the desired velocity against the push of the desired spacing.

2.2.2.4 Rakha-Pasumarthy-Adjerid Model (RPA)

Similar to the Gipps model, the RPA model seeks to use parameters that correspond to characteristics of drivers and vehicles, enabling the calibration parameters to be calibrated based on objective measures [9–11].

2.2.2.4.1 Van Aerde Steady-state Component

The RPA model uses the Van Aerde steady-state model first proposed in a nonlinear, single-regime functional form by M. Van Aerde and H. A. Rakha in 1995, as seen in Equation (8) [10].

$$\Delta x_{n\to n+1}(t) = c_1 + \frac{c_2}{\dot{x}_d - \dot{x}_{n+1}(t)} + c_3 \cdot \dot{x}_{n+1}(t) \tag{8}$$

Where: c_1, c_2, c_3 = *Parameters used for the Van Aerde steady-state model*

$\Delta x_{n\to n+1}(t)$, $\dot{x}_d$, $\dot{x}_{n+1}(t)$ *are defined in section 2.2.1*

This formulation provides the space headway of the leader-follower pair at time (t) using the velocity of the follower vehicle and four calibration parameters. The first calibration parameter, c_1, is the fixed distance headway constant (s), and is analogous to vehicle spacing at jam density. The second calibration parameter, c_2, is the first variable headway constant (m^2/s), and provides a measure of the driver's desire to return to their desired speed. The third calibration parameter, c_3, is the second variable headway constant (seconds), analogous to the driver sensitivity parameter in the Pipes model. It has been shown that constraints may be applied on the three calibration parameters by considering limitations experienced at jam conditions and capacity conditions, subscripts *j* and *c* respectively, as in Equation (9) [23], [24].

$$c_1 = \frac{\Delta x_j \cdot \dot{x}_d}{\dot{x}_c^{\,2}}(2\dot{x}_c - \dot{x}_d); \qquad c_2 = \frac{\Delta x_j \cdot \dot{x}_d}{\dot{x}_c^{\,2}}(\dot{x}_d - \dot{x}_c)^2; \qquad c_3 = \left(\frac{\Delta x_c}{\dot{x}_c} - \frac{\Delta x_j \cdot \dot{x}_d}{\dot{x}_c^{\,2}}\right) \tag{9}$$

Where: c_1, c_2, c_3 *are previously defined in this section*

Δx_j, *and* $\dot{x}_d$, *are defined in section 2.2.1*

$\dot{x}_c$ *is defined in section 2.2.2.2*

The application of these constraint conditions and the translation of the Van Aerde model into a speed formulation results in the Equation (10).

$$\dot{x}_{n+1}(t) = \frac{-c_1 + c_3 \cdot \dot{x}_d + \Delta x_{n\to n+1}(t) - \sqrt{[c_1 - c_3 \cdot \dot{x}_d - \Delta x_{n\to n+1}(t)]^2 - 4 \cdot c_3 [\Delta x_{n\to n+1}(t) \cdot \dot{x}_d - c_1 \cdot \dot{x}_d - c_2]}}{2 \cdot c_3} \tag{10}$$

2.2.2.4.2 Collision Avoidance Component

The RPA model imposes limitations on the steady-state Van Aerde model to ensure collision avoidance, and to account for vehicle dynamics constraints, as demonstrated by Park et al [12]. The resulting limiting conditions are shown in Equation (11) and Equation (12).

$$\dot{x}_{n+1}(t) = \sqrt{[\dot{x}_n(t)]^2 + 2 \cdot \left[\Delta x_{n \to n+1}(t) - \Delta x_j\right]} \quad (11)$$

2.2.2.4.3 Vehicle Dynamics Component

A detailed explanation of the application of the vehicle dynamics model is beyond the scope of this study, and is found in publication by Rakha et al [25]. The modeling of RPA car-following is produced by using the minimum velocity at time *t* from Equation (10), Equation (11), and Equation (12).

$$\dot{x}_{n+1}(t) = \dot{x}_{n+1}(t - \Delta t) + \Delta t \cdot \frac{F_{n+1}(t-\Delta t) - R_{n+1}(t-\Delta t)}{mass_{n+1}} \quad (12)$$

$$Where:\ F_{n+1}(t - \Delta t) = min\left[3{,}600 * \eta * \beta * \frac{\gamma * P}{\dot{x}_{n+1}(t-\Delta t)}, 9.8066 * M_{ta} * \mu\right]$$

$$\beta = \frac{1}{\dot{x}_{power}}\left\{1 + min\left[\dot{x}_{n+1}(t - \Delta t),\ \dot{x}_{power}\right] * \left(1 - \frac{1}{\dot{x}_{power}}\right)\right\}$$

$$\dot{x}_{power} = 1{,}164 * w^{-0.75}$$

$$R_{n+1}(t - \Delta t) = 0.047285 * C_d * A_f * \dot{x}_{n+1}(t - \Delta t)^2 + 9.8066 * mass_{n+1} * C_r[c_{r2} * \dot{x}_{n+1}(t - \Delta t) + c_{r3}]$$

The formulation included herein is nearly identical to that of the formulation provided in Rakha et al, with the exception of gamma (γ) being applied to engine power, where gamma is equal to the maximum percentage throttle observed to be used by a given driver.

The parameters included in the vehicle dynamics model not previously defined include:

F_{n+1} = *resultant force (N)*

R_{n+1} = *total resistive force (N)*

$mass_{n+1}$ = *vehicle mass (kg) of vehicle n + 1*

η = *driveline efficiency (unitless), taken as 0.7 herein*

β = *gear reduction factor (unitless)*

γ = *acceleration reduction factor (unitless), taken from the maximum measured percentage throttle depression for each driver herein*

P = *vehicle power (kW)*

M_{ta} = *mass of the vehicle on the tractive axle (kg), taken to be 55% of total vehicle mass herein*

μ = *coefficient of roadway adhesion (unitless), taken as 0.6 herein*

$\dot{x}_{power}$ = *speed at which vehicle attains its full power*

w = *vehicle weight-to-power ratio*

C_d = *drag coefficient (unitless)*

A_f = *vehicle frontal area (meters squared)*

C_r = *rolling coefficient (unitless)*

c_{r2}, c_{r3} = *rolling resistance coefficients (unitless)*

2.2.3 Calibration of Model Parameters

Each car-following model predicts the time-space trajectory of a following vehicle when provided the time-space trajectory of a leading vehicle, and the original location and velocity of the following vehicle. The parameters of each model are calibrated such that the resulting simulated behavior matches the observed behavior as closely as possible. To arrive at these calibrated values it is necessary to generate observed and simulated vehicle trajectories, define limiting bounds for the parameter values, define an error function, and select an optimization methodology to minimize the error objective function.

2.2.3.1 Discrete Time Generation of Vehicle Trajectories

In the case of the GHR model, the predicted acceleration of the following vehicle at time-step (t) is dependent on the vehicle trajectory information at time-step $(t - \tau)$, therefore all location (x), velocity $(\dot{x})$, and acceleration $(\ddot{x})$ information for the simulated vehicle trajectory that occurs during the initial time τ is set equal to the observed information during that time.

The GHR, and IDM methodologies each generate the anticipated acceleration of the following vehicle in a given time-step: $\ddot{x}_{n+1}(t)$. This information in itself does not provide the vehicle trajectory of the following vehicle, which must be extrapolated from the predicted acceleration value. The linear translations used are shown in Equation (13) and Equation (14) are used to solve the second-order ordinary differential equation.

$$\dot{x}_{n+1}(t) = \dot{x}_{n+1}(t - \Delta t) + \Delta t \cdot \ddot{x}_{n+1}(t - \Delta t) \quad (13)$$

$$x_{n+1}(t) = x_{n+1}(t - \Delta t) + \Delta t \cdot \dot{x}_{n+1}(t - \Delta t) \quad (14)$$

Alternatively, the Gipps and RPA models generate a predicted velocity value for a given time-step, with a translation to acceleration as shown in Equation (15).

$$\ddot{x}_{n+1}(t) = \Delta t^{-1} \cdot [\dot{x}_{n+1}(t + \Delta t) - \dot{x}_{n+1}(t)] \quad (15)$$

In cases where the prior information is unknowable at the beginning of a car-following event, the observed data is used for the simulation values. In cases where the posterior information is unknowable at the end of a car-following event, the simulation value is held constant from the simulation value in the previous time-step.

2.2.3.2 Boundary Conditions for Parameters

When performing model calibration it is necessary to set upper and lower limits for the values of each parameter. Many of the parameters can be fixed outright for each driver by examining both the raw dataset and by performing macroscopic optimization. The macroscopic data is analyzed using a heuristic automated tool (SPD-CAL), described by Rakha and Arafeh [26].

The desired velocity for each driver is taken from the values generated by SPD-CAL, and is used in the Gipps, IDM, and RPA models. The macroscopic data is additionally used in conjunction with Equation (7) to generate the coefficients for use with the steady-state portion of the RPA model, specific to each driver.

The minimum and maximum acceptable acceleration values for the follower vehicle are a result of taking the minimum and the maximum observed values from each driver. The anticipated maximum deceleration rate of the leading vehicle is determined from Gipps' assumptions, as previously shown in Equation (4). A mathematical analysis of the Gipps model by Wilson [20], and subsequently an analysis by Rakha and Wang [21] show that linear continuity may be lost when $\ddot{x}_{d-min} < \ddot{x}_{d-min,n}$, whereas the original formulation by Gipps [7] has $\ddot{x}_{d-min} < \ddot{x}_{d-min,n}$ only when $\ddot{x}_{d-min} < 3$, and $\ddot{x}_{d-min,n} = \frac{3+\ddot{x}_{d-min}}{2}$ otherwise. The original Gipps formulation of $\ddot{x}_{d-min,n}$ is held herein, but further study is warranted on the calculation of this parameter.

The bounds for the variables to be optimized during microscopic simulation are based on recommendations in the literature [8], [15]. Calibration is conducted using the evolutionary non-

linear approach applied in Microsoft Excel version 14.0, which utilizes a genetic algorithm to find local minima and a multi-start function to find global minima.

2.2.3.3 Optimization Functions to Match Simulated and Observed Results

The calibration of car-following models consists of determining the set of values for the variable parameters that result in the best match between the observed behavior of a following vehicle and the simulated behavior of the same following vehicle, using the observed behavior of a leading vehicle as an input. Punzo and Simonelli [27] examine a variety of calibration methods, citing that the error tests in common use are the root mean square error (RMSE), the root mean square percentage error (RMSPE), and Theil's inequality coefficient (U). It is noted that the choice of performance measure in the objective function may condition the results. As presented, the error formulations are shown in Equation (16), Equation (17), and Equation (18).

$$RMSe = \sqrt{\frac{1}{N}\sum_i \left(Y_i^{obs} - Y_i^{sim}\right)^2} \tag{16}$$

$$RMSPe = \sqrt{\frac{1}{N}\sum_i \left(\frac{Y_i^{obs} - Y_i^{sim}}{Y_i^{obs}}\right)^2} \tag{17}$$

$$U = \frac{\sqrt{\frac{1}{N}\sum_i \left(Y_i^{obs} - Y_i^{sim}\right)^2}}{\sqrt{\frac{1}{N}\sum_i \left(Y_i^{obs}\right)^2} + \sqrt{\frac{1}{N}\sum_i \left(Y_i^{sim}\right)^2}} \tag{18}$$

Examining recent thesis from five separate corresponding authors examining the calibration of car following parameters, it is found that there remains no consensus regarding the optimal formulation of parameter optimization [28–32]. If Theil's U coefficient is used, the number of observations is not included in the error function, which results in an inability to compare error results between the drivers included in the study. As a result, the proposed minimization objective function would be equal to a modified $RMSPe$ for a given car-following event, as a function of the goodness of fit for both the speed and the headway profiles of the following vehicle, taking the form of Equation (19).

$$F\left(\dot{x}_{n+1,sim}(t), \Delta x_{n\to n+1,sim}(t)\right) =$$

$$\sqrt{\frac{1}{N}} * \left(\sqrt{\frac{\sum\left(\dot{x}_{n+1,obs}(t) - \dot{x}_{n+1,sim}(t)\right)^2}{\sum\left(\dot{x}_{n+1,obs}(t)\right)^2}} + \sqrt{\frac{\sum\left(\Delta x_{n\to n+1,obs}(t) - \Delta x_{n\to n+1,sim}(t)\right)^2}{\sum\left(\Delta x_{n\to n+1,obs}(t)\right)^2}} \right) \tag{19}$$

2.2.4 Parameter Variation and Correlation

In discussing the method by which parameter values are assigned to various drivers in simulation, Kim and Mahmassani [31] cite that many simulators use Monte Carlo sampling of the individual parameter values to model heterogeneous drivers. Parameter values selected this way would require a marginal distribution for each parameter, through the use of a distribution function defined by a mean and standard deviation, or through the use of an empirical probability mass function. The resulting drivers are defined by a set of individual parameter values which are independently and randomly drawn for each driver from the separate marginal distributions. Kim and Mahmassani question this approach because it inherently assumes that the parameters are uncorrelated, whereas the theoretical expectation is that these parameters often reflect underlying behavioral factors and the various parameters associated with a given driver are expected to be correlated.

2.3 Naturalistic Data

When investigating the calibration of car-following models, the potential analysis and results are limited by the source of the data collected. Loop detector and aerial photography/videography data collection procedures provide accurate, in situ data for real traffic patterns, but measure only short-term or instantaneous events and provide no driver-specific data. Simulator and test-track collection procedures conversely can provide a wealth of driver-specific information, but may not reflect real-world driver behavior. The driver-specific data available from naturalistic driving studies provide variable length car-following events in real-world driving situations, while additionally providing a wealth of information about the participating drivers. The data bias and driver information associated with each data type is shown in **FIGURE 2-1**.

Data Source Matrix		Information on Drivers	
		Anonymous	Detailed
Data Bias	Artificial		Simulator Test Track
	Real-World	Loop Detector Aerial Photography	Naturalistic

FIGURE 2-1 Data source matrix linking bias and driver information.

A recent study by McLaughlin et al. conducted a summary investigation of the methods and applications relating to driver behavior [33]. As described in the summary thesis, naturalistic studies involve the inconspicuous instrumentation of participants' vehicles to collect data on multiple video views, forward radar, accelerations, speed, pedal actuation, and latitude and longitude, for the purpose of measuring driver behavior and performance over extended periods of time. Though naturalistic data collection provides data describing routine driving, these studies involve uncontrolled conditions and require extensive project setup and management compared to other formats of data collection. Though there have been a few studies to date involving naturalistic data, the study herein analyzes data obtained from the first such study, the Hundred Car Study performed by the Virginia Tech Transportation Institute [34–36].

Chapter 3 Processing Naturalistic Data for Car-following Model Calibration

3.1 Introduction to the Data Processing of a Naturalistic Study

This chapter examines the process by which data reduction can be performed on a naturalistic study dataset, for the purpose of application to the study of car-following models. The data reduction specifically focuses on the steps involved in selecting leader-follower vehicle pairs from the large dataset. As an overview, the process of generating the data generally includes: data collection; data management; a defined set of goals for data processing; the selection of a subset of data; the processing of individual events; and the validation of the data. Subsequently, the process developed to perform data reduction is applied as a case study to the Hundred Car Study, beginning with the selection of the data subset using of a Geographic Information Systems (GIS) application to identify routes and ending with the visualization of the collected data.

3.2 Generating Data from a Naturalistic Driving Study

The cost of generating a naturalistic driving database is extremely high, and a great deal of work and planning is done both before and after the data collection phase of the project to ensure that the information is organized and secured.

3.2.1 Data Collection

Within the naturalistic driving study, each of the sensing subsystems within the vehicle is kept independent of the others to constrain failures to a single sensor type [36]. Sensors may include: a box to obtain data from the vehicle network (on-board diagnostics); an accelerometer box for longitudinal and lateral acceleration; a system to provide information on the space headway between the leading and following vehicles; a system to detect conflicts with vehicles in adjacent lanes; an incident box to allow drivers to flag incidents for the research team; a video-based lane tracking system to measure lane keeping behavior; and video feeds for validation of the other sensor data. A video feed may include multiple camera views monitoring the driver's face and the driver's side view of the road, the forward road view, the rear road view, the passenger side view of the road, and a view of the driver's hands. Captured video is synchronized with the other sensor data to allow simultaneous display of archived information. A GPS positioning

system is used to collect information on the vehicle position, which can be used by technicians to locate the vehicles in the field for routine data collection or maintenance.

3.2.2 Data Management

Management of the data in a naturalistic driving study is of primary importance to ensure that that the data gathered is accessible and secured, maximizing its utility. These datasets may contain billions of time-steps worth of data, and are generally managed through a relational database.

3.2.3 Defining a Set of Goals for Data Processing

A naturalistic driving study conducted for safety research often has specific goals to be addressed as part of the initial data reduction, which may include: the characterization of crashes, near-crashes, and incidents; the quantification of near-crash events; the characterization of driver inattention; a measure of driver behavior over time; the causal factors of rear end conflicts; the causal factors of lane change behavior; a measure of inattention for rear end conflicts; the characterization of rear end scenarios; and a performance evaluation of the hardware, sensors, and data collection system [36]. One of the goals for a safety study may involve collecting as many crash and near-crash events as possible, which may skew the participant selection process, with higher mileage drivers and a larger sample of drivers below the age of 25. In the context of mobility applications, this initial bias can be overcome by selectively choosing drivers for analysis which are more representative of the driving public.

A note of caution for the application of naturalistic data to mobility applications: safety applications are not likely to share an interest with mobility applications in terms of the accuracy of certain data elements. For example, it is unlikely that a safety study would feel obligated to publish the accuracy of their geospatial location information or radar speed data, since these data elements are initially only intended to give a general idea of the field conditions immediately preceding an accident. In the absence of detailed specifications or accuracy documentation, all data elements from such a study are used in a given application, and any inherent inaccuracies are considered to be simply part of the background data bias.

3.2.4 Data Selection

Though the GPS data collected by naturalistic studies are often intended primarily for locating vehicles for routine data collection and vehicle maintenance, this dataset proves to be crucial for the application of naturalistic data to mobility research. By importing the GPS coordinates of the data points into a geographic information systems (GIS) application, it is possible to obtain a driver-by-driver holistic view of travel patterns. The density of the GPS points on the map indicate the frequency with which a driver travels a given route, providing visual identification of suitable corridors with large quantities of data for analysis of car-following behavior. By indexing the commuting routes of each of the drivers involved in a study, areas of overlap can be identified to allow for inter-driver comparison. Having homogeneity of roadway conditions when conducting cross-comparison between drivers is essential to ensure that variations in driver behavior are caused by the drivers themselves and not by the environment in which they are interacting.

3.2.5 Event Processing

In an ideal data environment, it is possible to use search criteria to automate the process of creating lead-follower car-following vehicle pairs. In the absence of an ideal data environment, it is necessary to conduct manual data reduction by visually verifying the presence of car-following for a given set of data. This decision should be made on a case-by-case basis, and can be influenced by the type of data to be gathered, whether macroscopic data is sufficient or whether specific vehicle pairings need to be simulated to calibrate model parameters.

3.2.6 Data Validation

The presence of error in the data may have a great effect on the accuracy of the calibrated parameters, due in part to the many parameter value combinations that yield near-optimal solutions. The speed and radar data are the only elements necessary for the generation of vehicle time-space trajectories for car-following analysis, so these elements are the most important to validate. Other data elements may be used for supplemental analysis such as the binning of trips.

The speed of the follower vehicle is obtained from either onboard diagnostics (OBD) or from an interpolation of the geospatial coordinates. Both sources of speed data are likely to update on a one second interval, while the database may be collected on a much smaller interval, such as one tenth of a second. The OBD data is generally more reliable than the speed data from

the GPS equipment, as the spatial coordinates of a moving object in GPS have a tendency to lag and then catch up, resulting in a recorded speed that oscillates about the actual speed traveled by the vehicle. The follower speed generated by the GPS should be used as a check for the OBD speed, which itself may regularly experience faults with incorrect speeds or no data recorded.

The radar data is an essential component of the naturalistic data for mobility applications, but is not as essential in conducting safety studies, so it may consequently suffer from either dependability or accuracy issues. The radar generally exhibits a capture area that is cone shaped, tracking multiple objects within the cone by assigning object identification numbers to each object as it is first acquired. If an object's speed exactly matches that of the instrumented vehicle, it has a tendency to disappear from the radar data, being assigned a new number when it is reacquired. The issues with radar data being dropping also apply to congested, stop and go traffic, causing data gaps to appear within a given event whenever vehicles come to a stop. Any automated procedure for the identification of car-following events must integrate multiple objects, correctly identifying the difference between a reacquired vehicle, and a vehicle which switches lanes and is subsequently replaced by a new vehicle.

3.3 Case Study: Hundred Car Study

The large size of the database generated by a naturalistic driving study is such that data reduction must be conducted using a carefully planned procedure. In the case of the Hundred Car Study, the initial dataset includes 108 individual drivers, with nearly 337,000 hours of data collected across 207,000 trips, which when sampled every 0.1 seconds results in more than 12 billion rows of data. Car-following trips analyzed herein are identified through a complex process which includes: selecting specific routes and drivers, utilizing a graphic information systems (GIS) application for visual identification; exporting pertinent data points from GIS to define trips of interest; selecting trips likely to have high incidence of car-following behavior based on average speed and variation of speed; validating leader/follower events through the review of video data; and separating the car-following events and overtaking events by comparing leader and follower speed profiles.

3.3.1 Use of GIS to Select Drivers and Roadway Sections

Through an application of SQL programming, database information is selected and saved as separate files differentiated by driver, with these files imported to a geographic information systems (GIS) application using the stored GPS coordinate data. A balance is necessary between the level of detail shown in GIS and the processing power needed to pan, zoom, and select the data. Retaining the associated file and frame identification numbers with each coordinate point, the information analyzed visually in GIS includes only one tenth of the data points, sampled once per second to reduce processing time. The file associated with each driver is opened in turn, with regular commuting routes identified where possible, maintaining a tabular list of all drivers and their corresponding commonly traveled routes. In this way specific routes regularly traveled by multiple drivers are identified. Determining that multiple drivers traversed the Dulles Airport Access Road (DAAR) on a regular basis, pertinent file and frame identification numbers are exported from the GIS application, again through the use of SQL programming. Fifteen drivers are identified through the GIS visual review, yielding a total of more than 500 hours' worth of data, more than 18 million rows of data, just over one percent of the total data available from the database. This roadway serves as a good example of highway driving behavior in an urban setting, with four through lanes throughout, and a constant speed limit of 55 mph (24.6 m/s). An example of the GIS visualization application is seen in **FIGURE 3-1**.

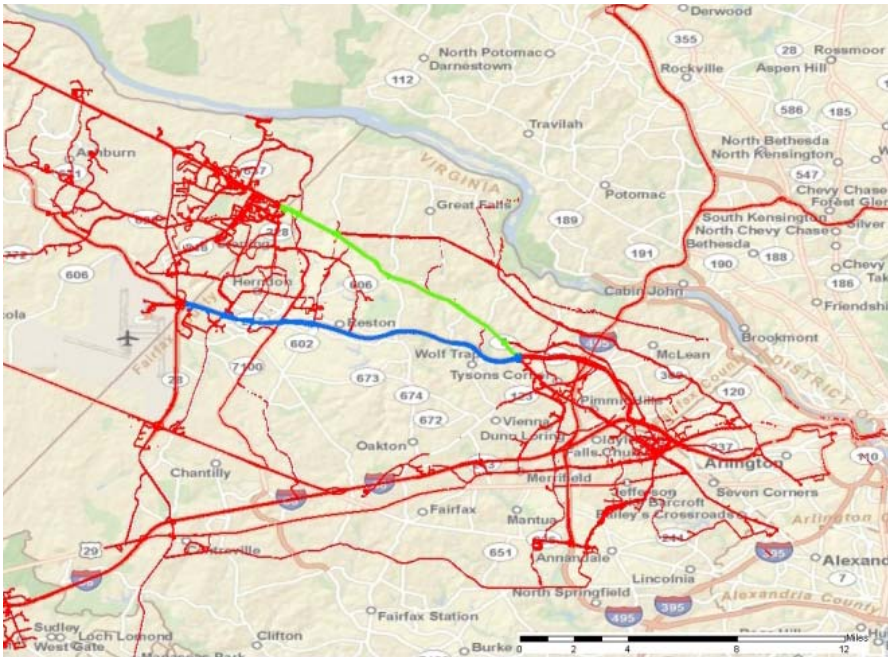


FIGURE 3-1 Geographic Information System (GIS) data representation - example D304.

The GIS data selection process proved to require a great deal of time, not in identifying the routes where were utilized for regular commutes, but in the time required for the computer to process the selection of data, the panning and zooming of the file, and the partitioning of the selected data into separate files. As computer processing power continues to advance this should become less of a concern when working with the data, but may have been a primary reason why more studies of this data were not pursued in the past.

Using naturalistic data for travel time studies becomes difficult due to the necessary start-up time of the GPS equipment as it connects with satellites and establishes an initial location. In the specific case of the Hundred Car Study, the time associated with each data point comes from the computer time, which is not a common time held between vehicles, nor is it continuous for a given vehicle, as it is often reset during the data collection process. This lack of an accurate time measure invalidates the potential for time of day dependent analysis, which should be a strong point for naturalistic. Future naturalistic driving behavior studies conducted as safety studies should take care to coordinate data file times among all vehicles for the benefit of use in mobility applications.

3.3.2 Process to Obtain Leading/Following Pairs

With specific drivers and routes identified through visual review, a preliminary data subset is identified of all data points along the given route belonging to drivers who frequently use the route. This data includes all travel traversing this route, and must be further reduced to include only data with lead-follower vehicle pairs, if the calibration of car-following models is to be performed. Careful planning of the data reduction process can result in a cost-effective, streamlined work effort, minimizing the need to duplicate work and maximizing the quality of the data generated. A sample process designed for application to the Hundred Car Study is seen in **FIGURE 3-2**, which may have application to other naturalistic data sets.

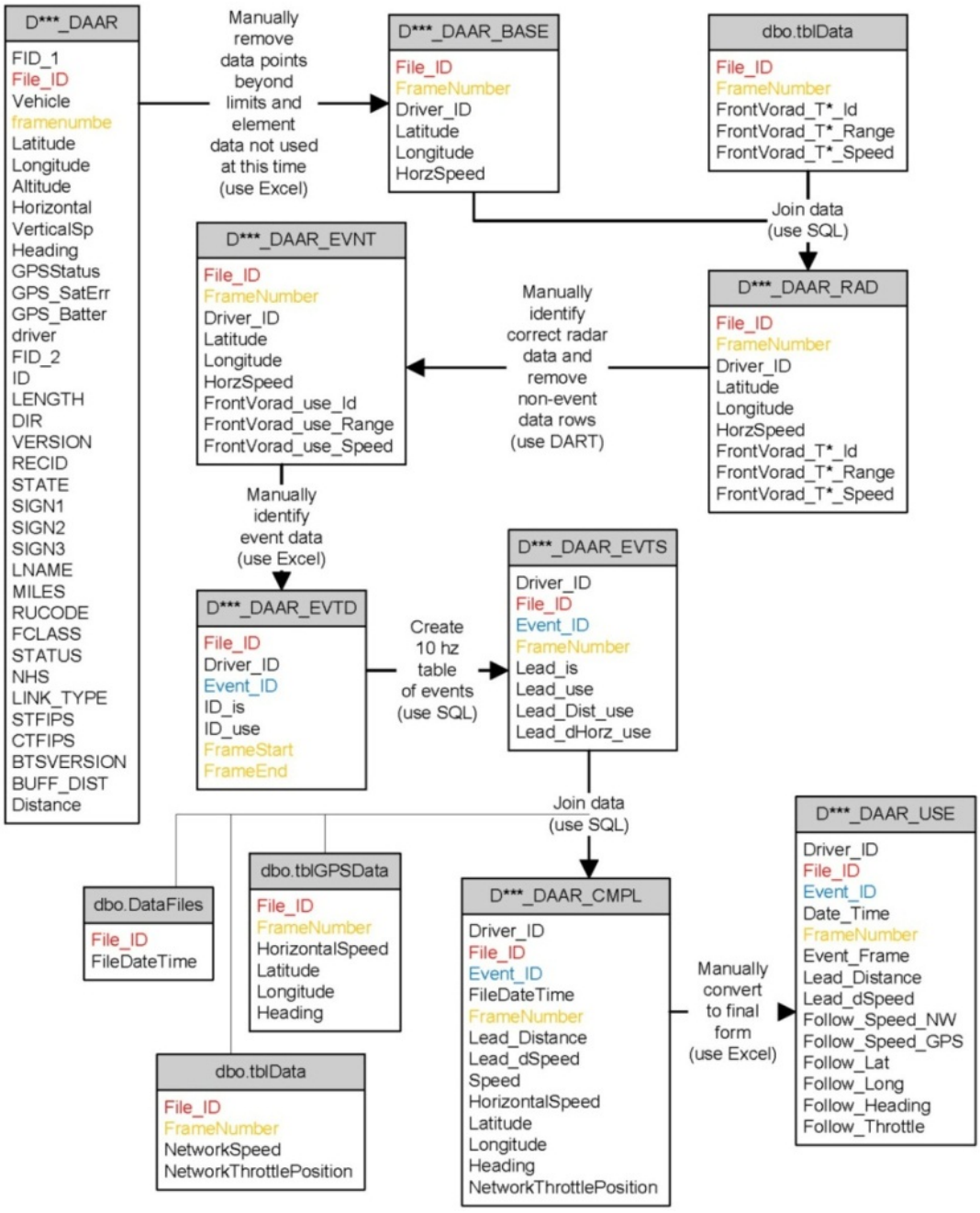


FIGURE 3-2 Process diagram for generation of car-following data from selected GIS data.

3.3.3 Selection of Leading/Following Vehicle Pairs

With the GIS application identifying more than 500 hours of data from 15 drivers along a stretch of highway, with a homogeneous speed limit and lane configuration, specific lead-follower vehicle pairs must then be identified. The forward facing radar used in the Hundred Car Study tracks up to seven individual objects at a time, assigning identification numbers sequentially

from 1 to 255 as each new object is observed. In reviewing the radar data it becomes apparent that the technology is not entirely reliable, with many objects unseen, and most objects dropped periodically and then reassigned new identification numbers as they are reacquired. Choosing to conduct data reduction manually using the DART software application developed by the Virginia Tech Transportation Institute, which graphically displays video and radar data linked by frame numbers, potential complications with data noise from the radar system is avoided. Manual reduction of data is not without its own issues, with many potential car-following events removed from the dataset due to a lack of ability to confirm the data. An example of the video imaging program used for identification of leader-follower pairs is seen in **FIGURE 3-3**.

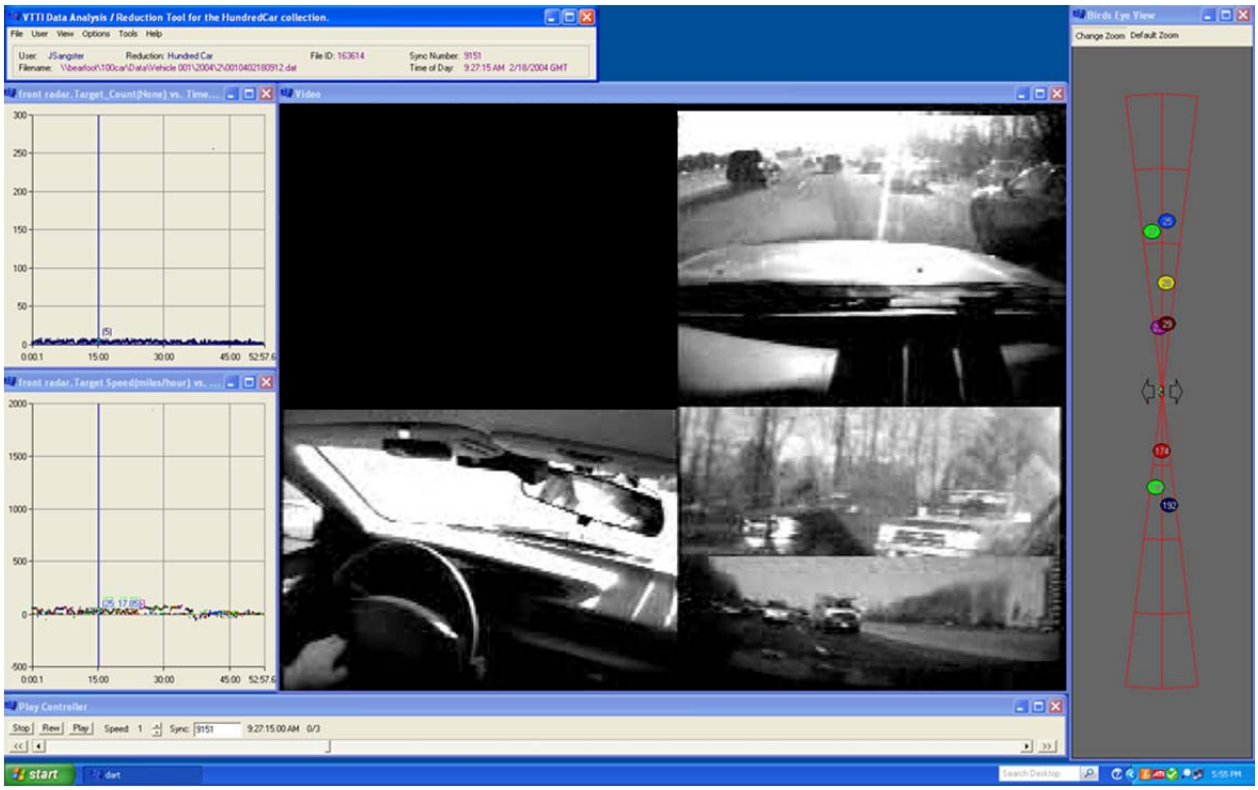


FIGURE 3-3 Example of DART visualization for data reduction.

The frontward looking video feed was found to be blank for the entirety of some drivers' data sets, and for a portion of the trips associated with other drivers. Additional occlusion was observed for a variety of reasons, including: a lack of contrast during nighttime driving; the blurring of the view from liquid precipitation; a covered view due to the build-up of frozen precipitation; and a lack of contrast due to glare at sunrise or sunset. The video feed is far from the only system to experience data outages, with one estimate from the Hundred Car Study

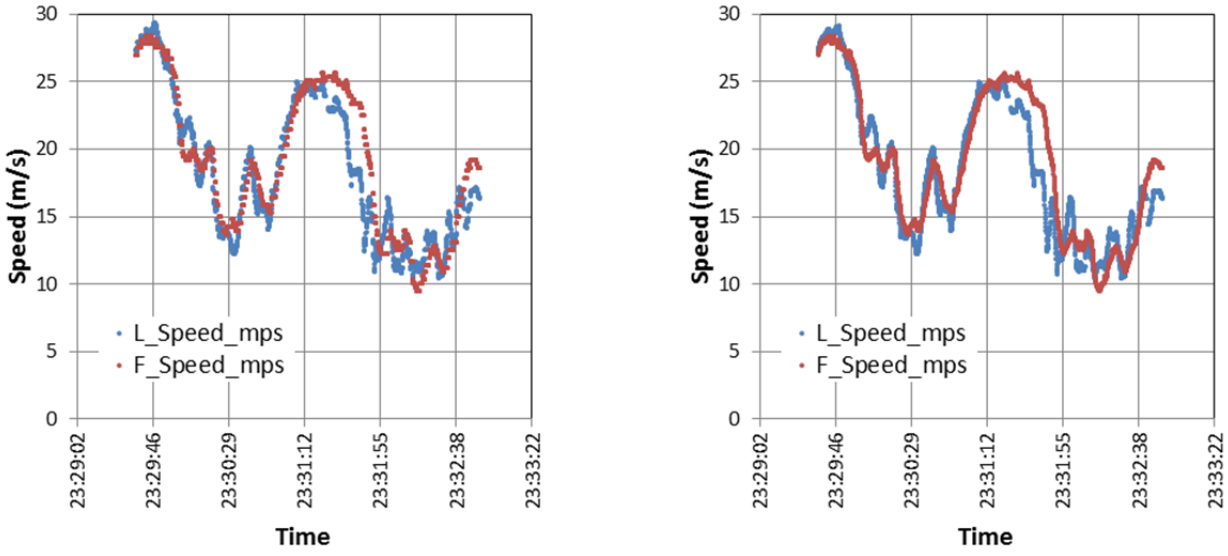
indicating that 1.37 million miles of data are complete out of an estimated total of 1.80 million miles of possible data, resulting in a 24% loss of data due to major or minor failures in hardware or software [36].

3.3.4 Smoothing of Data Elements

Outlier data is smoothed from the identified vehicle pairs by way of data validation against the anticipated physical speed/acceleration limits. As an example, the vehicle velocity at the 99.9th percentile shows the follower and leader speeds are 40.10 and 37.62 m/s, respectively before smoothing, and 32.04 and 35.16 m/s, respectively after smoothing. Without smoothing the leading and following vehicles have a maximum deceleration rate of 323.2 and 303.0 m/s^2, respectively, and a maximum acceleration rate of 283.9 and 303.6 m/s^2, respectively. With smoothing the leading and following vehicles have a maximum deceleration rate of 5.961 and 9.012 m/s^2, respectively, and a maximum acceleration rate of 5.948 and 8.893 m/s^2, respectively. While these values are still beyond the limits of the physical limitations of the vehicles, they are representative of the step-wise accuracy of the recorded velocity.

3.3.5 Interpolation of Data Elements

Along with the dropping of lead vehicle space headway and relative speed caused by the radar equipment, a number of other data elements were found to have gaps and/or update frequencies greater than one tenth of a second. The on board diagnostic (OBD) speeds recorded in the database are generally updated every one second, and held constant over the ten observations occurring each second. To provide vehicle trajectory profiles for analysis, the original data values are kept in the time-step when they are updated, with a linear interpolation applied to all time-steps occurring between observations. Interpolation is only allowed over gaps of two seconds or less, 20 time-steps, with larger gaps being removed from the dataset and new events being defined by the remaining data. Because a number of car-following models require an offset time, only events lasting longer than five seconds, 51 time-steps or greater, were kept in the dataset. This interpolation of data elements also serves as a final validation of the lead-follower pairs, identifying any mistakes made during the visual data reduction. An example of how the interpolation of data elements affects vehicle speed trajectories is seen in **FIGURE 3-4**.



(a) No Interpolation

(b) Interpolation

FIGURE 3-4 Effect of interpolation on sample lead and follower vehicle speed profiles.

3.3.6 Summary of Data Analyzed

A tabular summary of the quantity of data analyzed for this study and the data reduced in each step of the process is seen in **TABLE 3-1**. Note that seven of the initially identified fifteen drivers were missing data critical for the identification of car-following events, resulting in a total dataset of 999 minutes of car-following data from an initial 30,213 minutes exported from GIS.

TABLE 3-1 Quantity of data included in this study.

	Driver	Raw Data	Car-Following Pairings	
	Number	Minutes	Events	Minutes
Valid Data	124	636	188	79.7
	304	1,424	149	68.7
	316	4,280	332	190.9
	350	1,797	274	61.9
	358	1,712	443	185.6
	363	2,854	274	176.4
	367	2,991	507	211.2
	462	813	72	24.5
Invalid Data	102	152	Not Enough Leader-Follower Data	
	351	2,431	Missing Video or Speed Data	
	359	2,921	Missing Video or Speed Data	
	402	2,171	Missing Video or Speed Data	
	409	5,228	Missing Video or Speed Data	
	452	228	Missing Video or Speed Data	
	461	576	Missing Video or Speed Data	
	Total	**30,213**	**2,239**	**999**

3.3.7 Visualization of the Dataset

The macroscopic diagram provides a ready way to make observations and comparisons about the various drivers analyzed from a naturalistic driving behavior study. In **FIGURE 3-5**, the data from the eight drivers analyzed is aggregated into a single macroscopic interaction diagram, with color used to indicate steady-state conditions separately from acceleration and deceleration regimes. In **FIGURE 3-5**, the yellow-to-red range indicates data points in which the follower vehicle has a positive acceleration value, the blue-to-purple range indicates data points in which the follower vehicle has a negative acceleration value, and the green range indicates data points in which the follower vehicle is experiencing steady-state conditions.

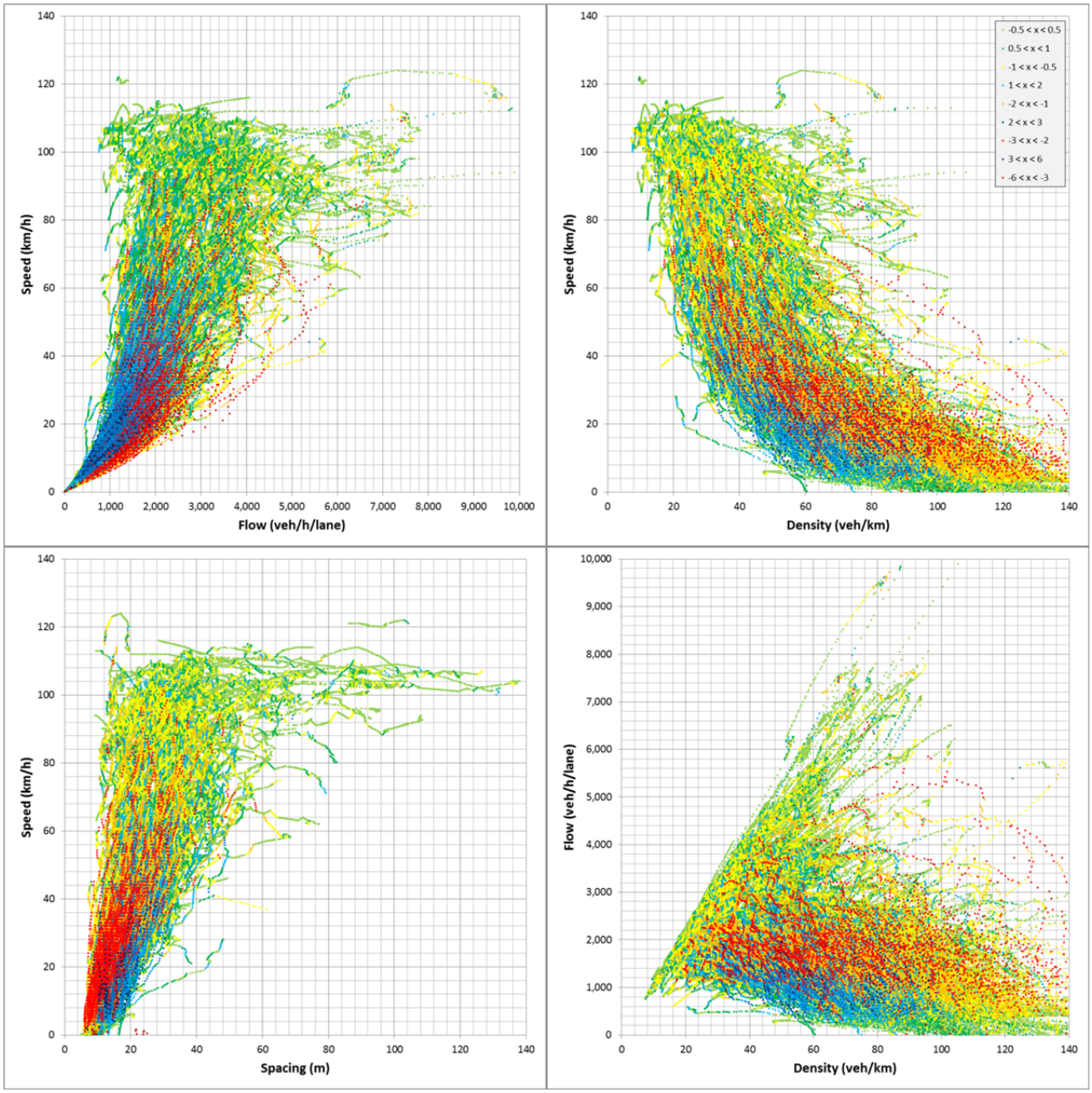
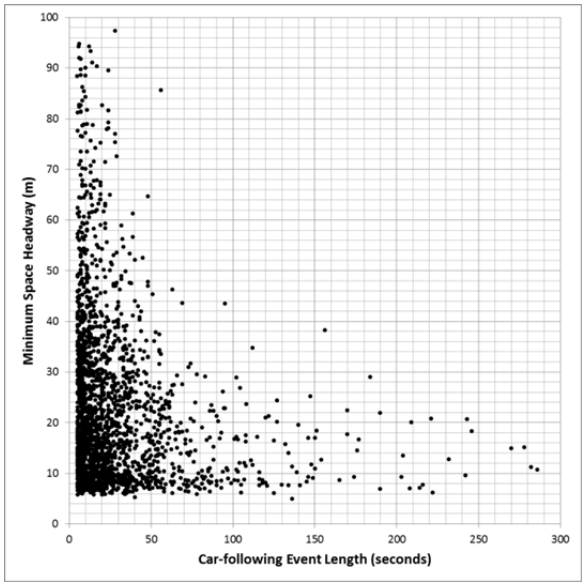
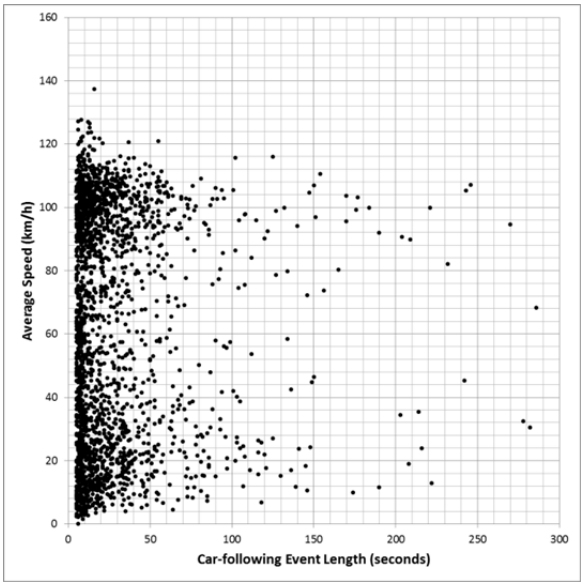


FIGURE 3-5 Combined macroscopic diagram for car-following event data.

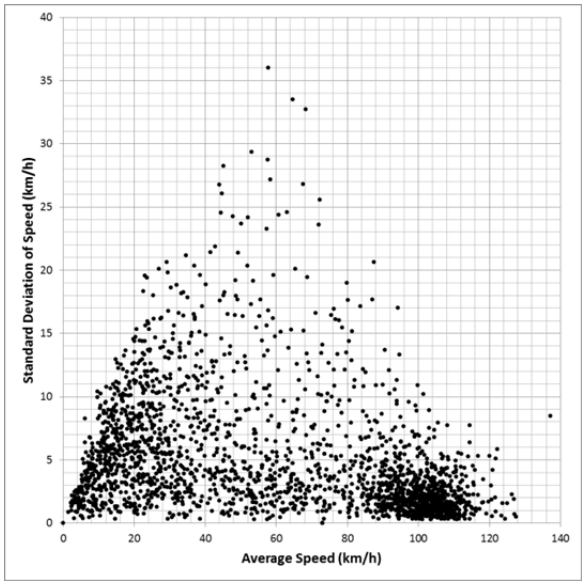
In addition to examining the car-following data for the purpose of parameter calibration, it is useful to examine the data from a metadata perspective, to identify how much data is available, and how the data is distributed across a number of different measures. The charts in **FIGURE 3-6** include: the minimum space headway versus the time length of each car-following event; the average speed versus the time length of each event; the standard deviation of speed versus the average speed of each event; and the min. space headway versus the average speed of each event.



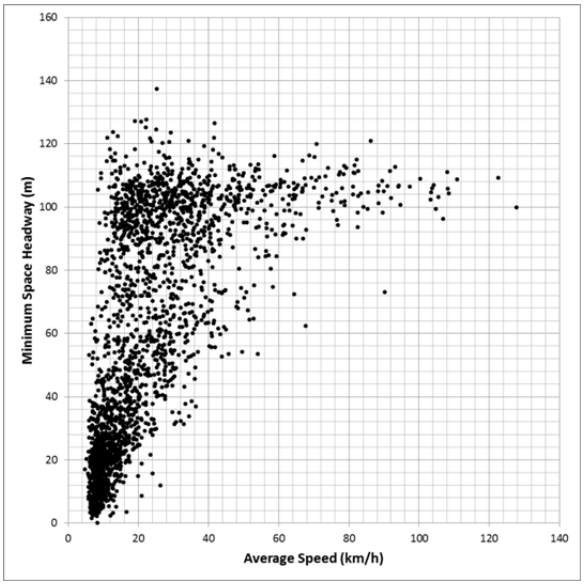
(a) Min. space headway vs. event length



(b) Average speed vs. event length



(c) Standard dev. of speed vs. average speed



(d) Min. space headway vs. average speed

FIGURE 3-6 Data expressed as aggregate results by event.

The interaction between the minimum space headway of each event and the average speed of each event indicates that the methodology used to define car-following interaction pairs by using a minimum error offset could be improved by coupling it with this type of metadata. The visualization figures serve to demonstrate the versatility of this dataset.

3.4 Conclusions Regarding the Data Processing of a Naturalistic Study

Although it is helpful to have a well-defined data processing plan when beginning to process the data from a naturalistic driving study, it is essential to maintain flexibility in the approach, to deal with unforeseen complications as they emerge. The process is essentially broken down into three steps: data selection, event processing, and data validation. For the analysis conducted herein it was important to select multiple drivers traveling the same section of highway, to ensure homogeneous roadway conditions between drivers, which was made easier by the fact that the Hundred Car Study used only participants located in the Northern Virginia area. Future studies will need to show that individual driver behaviors do not change from facility to facility, to increase the amount of data that can be selected and processed. The processing of events reduced the number of drivers selected by half, as essential information was missing from the dataset. This is important to note for future research, as the data reduction process is time-intensive, and time is spent more effectively when the completeness of the dataset for each driver is considered before selections are made. In the data validation portion of the study, a mismatch was discovered between the interval of data collected, and the interval of data updated by the on-board diagnostics computer, resulting in a need to post-process the data and smooth the discrete data. Problems with the dataset were identified by examining any velocity data recorded which resulted in an instantaneous acceleration beyond the limits of vehicle dynamics. One result of this smoothing effort was the discovery that on-board diagnostic velocity values were found to get "stuck" periodically at a given speed, from anywhere between a number of seconds of time to the remainder of the vehicle's trip. If the gap in verifiable data was larger than two seconds, the data was removed from the dataset; otherwise the information was replaced with a linear interpolation between data from before and after the perceived gap. As this dataset is meant to check whether a given simulation model accurately replicates realistic driving behavior, it is essential to have the observed data be as accurate as possible.

When all of the complications encountered in the data reduction process are taken together, it is seen that working with a naturalistic driving database for the purpose of understanding mobility applications requires a great deal of time investment. It is demonstrated, that this type of dataset would not be valuable for the purpose of calibrating a specific facility for a specific project, but is instead invaluable for its ability to investigate the disparity of behavior between and within drivers, by incorporating otherwise unobtainable data elements.

3.4.1 Review of Data Generation Process

In summary, the data generation process includes data collection, data management, the definition of a set of goals for the data reduction, selection of data, processing of events, and validation of the data. Data collection elements within the context of a naturalistic driving study may include the on-board diagnostics of a vehicle, an accelerometer box, front and rear radar detection and tracking of objects, a video lane-tracking system, GPS location data, and video feeds in multiple directions for future data validation. Data management is a critical issue for naturalistic datasets, and till now has been done through the use of relational database architecture. The goals defined for the data reduction will influence which elements are selected form the database, and may also influence the quality and quantity of data collected. In the context of mobility analysis, the selection of data can be difficult, as some studies may need homogeneity in roadway characteristics, and other studies may want carefully controlled heterogeneity. Event processing is necessary for the study of car-following models, relating data stored in individual time-steps into leader/follower events. The data validation procedures are the most important part of the process, as any potential errors or missing information not identified in the previous steps become apparent through validation. Information gathered in a naturalistic study which is not utilized specifically in formulating the event data can prove useful in data validation for cross-checking. **FIGURE 3-7** serves to illustrate the data generation steps to be followed.

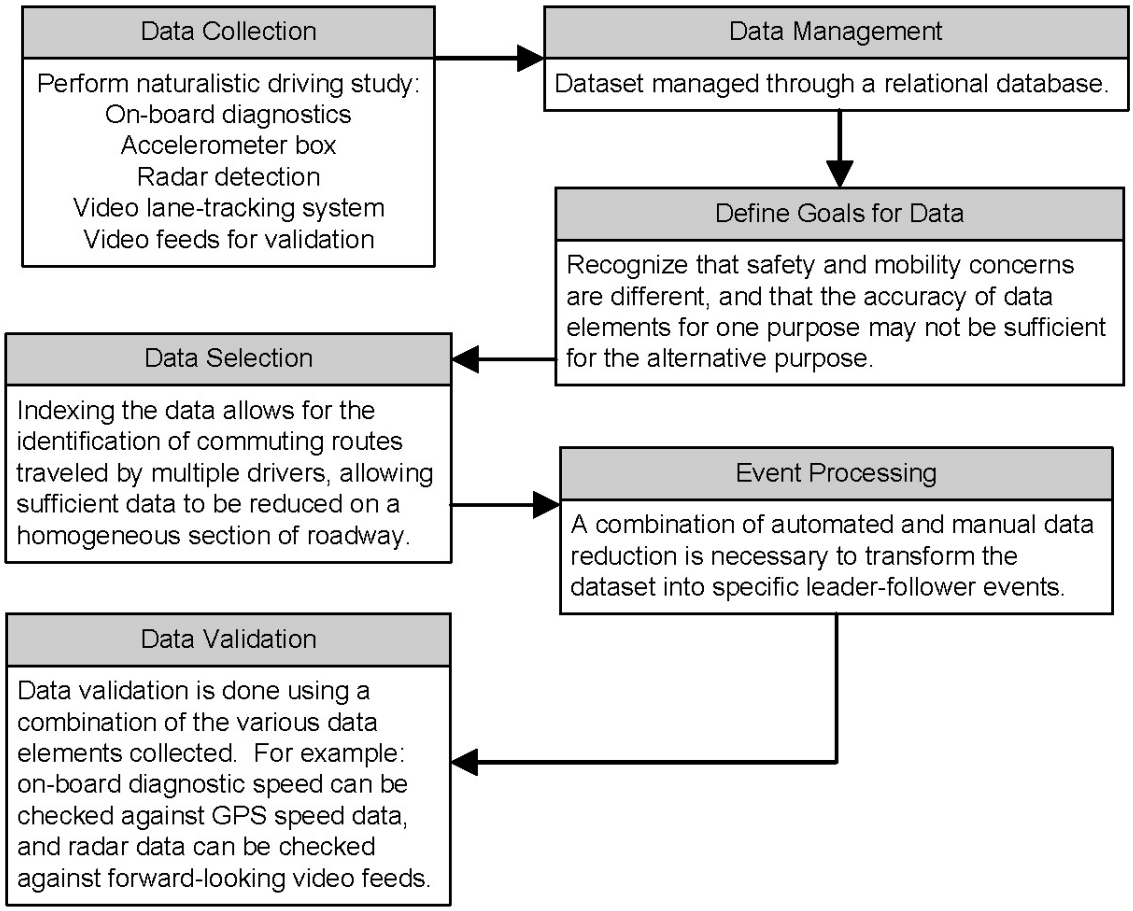


FIGURE 3-7 Graphical review of the data generation process.

3.4.2 Review of the Case Study Procedures

The data reduction case study of the Hundred Car Study includes: the use of GIS data to select drivers and roadway sections, the process of obtaining leading/following pairs, the selection of leading/following pairs for analysis, the smoothing of data elements, the interpolation of data elements, and methods for summarizing the data, visual, quantitative, and qualitative.

The GIS visualization of the data allows for specific roadway sections to be identified which exhibit high volumes of trip traffic, a necessity for choosing a homogeneous section of roadway on which to analyze multiple drivers' behavior. A balance is necessary between the level of detail shown in GIS and the processing power needed to pan, zoom, and select the data. Too much data included in the visualization results in a lengthy data selection process, but too little data makes it difficult to identify regular commuting routes. Computer processing power

may have prevented previous research efforts from working with naturalistic data, but this concern should be less of a hindrance in the future.

When the dataset is extracted from GIS, it contains many data elements necessary for categorizing the roadway location, geometry, and environment, which become superfluous when only one section of roadway is being analyzed. The extracted data points are merged with the radar data to identify object pairs, and segments of data without any associated radar data are removed. The radar data tracks up to seven objects at a given time, identified on the basis of having a different relative velocity than that of the vehicle. A complication of this process is that vehicles traveling precisely at the same speed are dropped from radar detection, as are stopped vehicles. By setting a minimum event length of five seconds, perception-reaction time can be fully implemented in simulation, and radar objects such as sign gantries which pass by quickly can be automatically removed from the dataset.

The most time-intensive part of the data reduction process occurs in the selection of leader/follower pairs, where it is necessary to validate the information obtained from merging the radar information into the dataset generated from GIS. The validation requires using a visualization program which links the file and frame number information between the radar data and the video feeds recorded. In the case of the Hundred Car Study, the video quality is generally low and is often occluded by precipitation or lighting conditions, such as glare or nighttime driving. By alternatively advancing and rewinding the data it is usually possible to use the combination of video and radar data to identify in-lane leading vehicles and select the appropriate object identification number from the radar data to define an event. A complication specific to the Hundred Car Study was the periodic loss of both video data and radar data due to equipment malfunctions during the data collection process. In some cases, the front video feed for a given driver was only valid for the first month of data collection and was not corrected at a later time, and in other cases the video feed failed to work for the duration of the data collection. All of the data processed up to this point with insufficient video or radar data to validate the event was required to be discarded. Unfortunately, additional data was also required to be removed from the dataset during the smoothing of data and the interpolation of data.

The interpolation and smoothing of data elements occurred concurrently as one process, and served to remove errors arising from the data collection process. The interpolation of data

elements became necessary as it was found that on-board diagnostics for the follower vehicle recorded velocity information every tenth of a second, but only updated the velocity information once per second. The result of these instantaneous velocity updates once per second was a time-step acceleration of 0 m/s^2 for nine tenths of a second, followed by an instantaneous acceleration equal to ten times the actual value for one tenth of a second. This problem led to the review of the dataset for similar instantaneous jumps, which then required a smoothing of the dataset. Problems with the dataset were identified by examining any velocity data recorded which resulted in an instantaneous acceleration beyond the limits of vehicle dynamics. One result of this smoothing effort was that the on-board diagnostic velocity values were found to get "stuck" periodically at a given speed, from anywhere between a number of seconds of time to the remainder of the vehicle's trip. By seeking out vehicles with OBD velocities that did not change over time, additional vehicles or segments of a vehicle's dataset were removed from the reduced dataset.

Of the fifteen drivers initially identified as commuting along the chosen route, only eight drivers resulted in viable datasets for analysis. The 500 hours of raw data pulled from GIS resulted in 2,200 individual car-following events, with just under 17 hours of data in the final analysis dataset. Additional data reduction should only be conducted after a thorough examination of the completeness of the dataset for each driver. This will add additional up-front time to the data reduction, but will prevent the effort spent to identify and process data which will subsequently not be suitable for analysis.

Chapter 4 Analysis of Car-Following Models using Naturalistic Data

4.1 Introduction to the Analysis of Car-Following Models

This chapter utilizes the dataset generated in chapter three to perform a comparative analysis of the four car-following models discussed in the literature review. A summary is provided of the data used for the analysis, a justification is given for the parameter boundary conditions used, an optimization function is defined, the parameter values calibrated to minimize the objective function, and a comparison of errors is shown between the various models analyzed. A visual representation of the derived solutions is provided, and an analysis of the results is conducted. Specific analysis of the results includes the macroscopic error measures for each driver resulting from each model, a discussion of sample events at the microscopic level of analysis, and an examination of linear correlation between the various parameters within a given model.

4.2 Summary of Data Generated from Hundred Car Study

Note that seven of the initially identified fifteen drivers were missing data critical for the identification of car-following events, resulting in a total dataset of 999 minutes of car-following data from an initial 30,213 minutes exported from the GIS database.

4.2.1 Tabular Representation of Data for Analysis

A tabular summary of the quantity of data analyzed for this study and the data reduced in each step of the process is summarized in **TABLE 4-1**.

TABLE 4-1 Quantity of data included in this study.

	Driver	Raw Data	Car-Following Pairings	
	Number	Minutes	Events	Minutes
Valid Data	124	636	188	79.7
	304	1,424	149	68.7
	316	4,280	332	190.9
	350	1,797	274	61.9
	358	1,712	443	185.6
	363	2,854	274	176.4
	367	2,991	507	211.2
	462	813	72	24.5
Invalid Data	102	152	Not Enough Leader-Follower Data	
	351	2,431	Missing Video or Speed Data	
	359	2,921	Missing Video or Speed Data	
	402	2,171	Missing Video or Speed Data	
	409	5,228	Missing Video or Speed Data	
	452	228	Missing Video or Speed Data	
	461	576	Missing Video or Speed Data	
	Total	**30,213**	**2,239**	**999**

4.2.2 Visual Representation of Data for Analysis

In all, data from eight drivers are available for analysis on the Dulles Airport Access Road (DAAR). Seven out of the fifteen drivers identified by GIS as commuting on the DAAR were found to be lacking sufficient data, such that verification of the time-space trajectories of both the leader and follower vehicle was not possible, with the primary cause being the absence of follower vehicle speed data from on-board diagnostics. In examining the dataset used for the current study, the data can be visually represented in the form of macroscopic diagrams for the eight drivers, with one example seen in **FIGURE 4-1**. All eight drivers are included in the results section of this chapter. In the current example shown, green points indicate steady-state conditions, while red points indicate deceleration and blue points indicate acceleration.

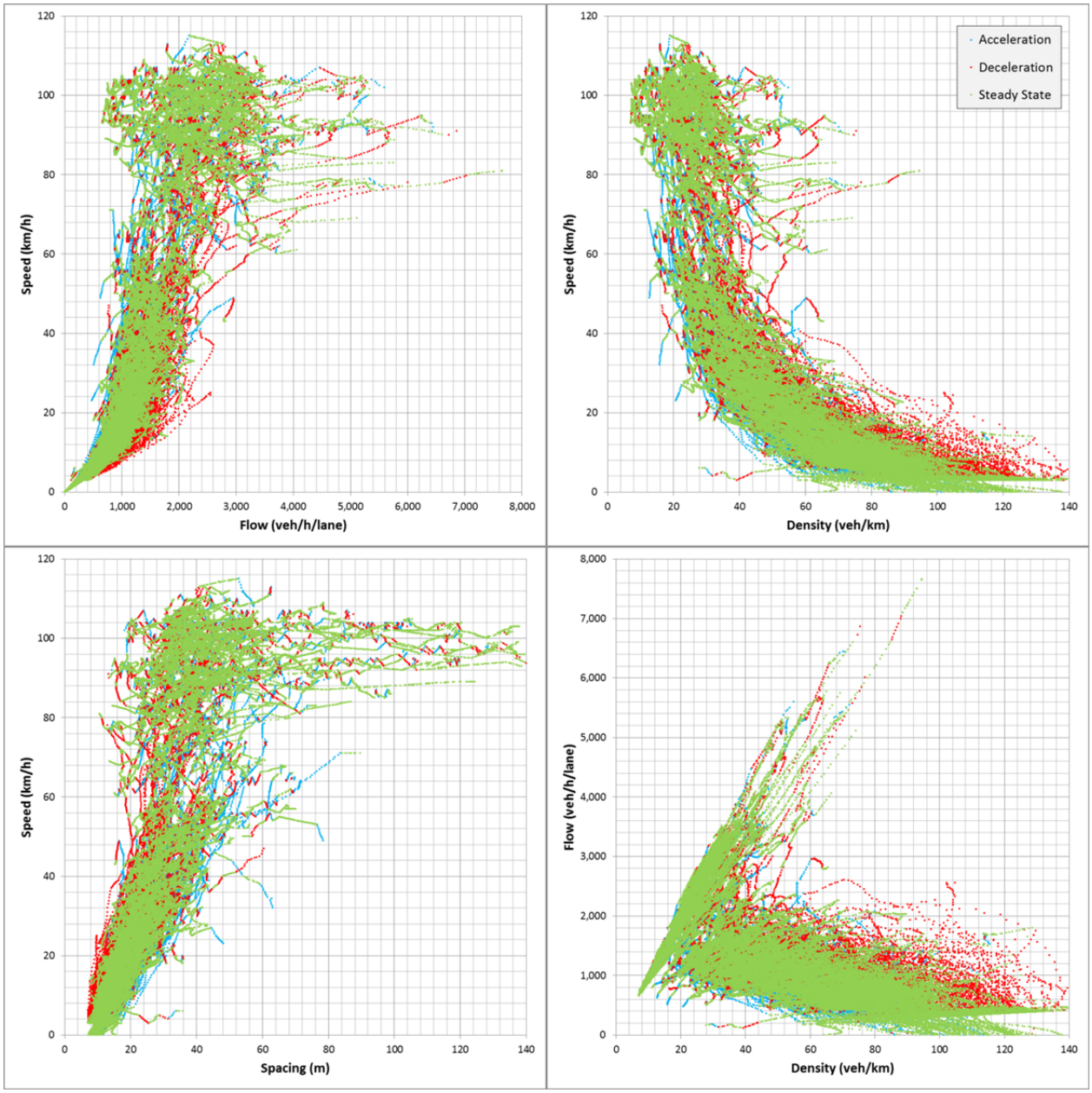


FIGURE 4-1 Macroscopic diagrams of car-following event data: Driver 316.

4.3 Calculation of Optimized Model Parameters

4.3.1 Defining Parameter Boundary Conditions

When performing model calibration it is necessary to set upper and lower limits for the values of each parameter. Many of the parameters can be fixed outright for each driver by examining both the raw dataset and by examining the macroscopic data. Though some of the parameters are

specific to a given car-following model, many of the calibration parameters have literal interpretations, like the desired free-flow speed of a driver, and are utilized by multiple models, as seen in **TABLE 4-2**. Some of the parameter values can be taken directly from observed values maximum and minimum values within the dataset, as shown in **TABLE 4-3**.

The macroscopic data is analyzed using a heuristic automated tool (SPD-CAL), described by Rakha and Azafeh [26]. The result of the macroscopic best-fit analysis provides certain parameter values, as shown in **TABLE 4-4**. The vehicle characteristics needed for the vehicle dynamics model are shown in **TABLE 4-5**.

The bounds for the variables to be optimized using the error function applied to the microscopic events are based on recommendations in the literature [8], [15]. Calibration is conducted using the evolutionary non-linear approach applied in Microsoft Excel version 14.0, which utilizes a combination of local minima and global minima through the use of a multi-start function to determine a heuristic solution. Both the variable boundary conditions and the solution sets are seen in T**ABLE 4-6**.

TABLE 4-2 Method of parameter calibration, and associated car-following models.

Parameter	Method of Optimization			Applies to Model			
	Observed	Macroscopic	Microscopic	GHR	Gipps	IDM	RPA
$\ddot{x}_{d-max}$	X				X	X	X
$\ddot{x}_{d-min}$	X				X	X	X
$\ddot{x}_{d-min,n}$	X				X		
γ	X						X
τ_m		X*			X		
$\dot{x}_d$		X			X	X	X
$\dot{x}_{capacity}$		X					X
$q_{capacity}$		X					X
k_{jam}		X		X	X	X	X
α			X	X			
z			X	X			
l			X	X			
τ_{GHR}			X	X			
τ_s			X			X	
δ			X			X	

* A discussion on the calculation of Gipps PRT is found in section 2.2.2.2.

TABLE 4-3 Parameter values calibrated from observed values.

Parameter	D124	D304	D316	D350	D358	D363	D367	D462	Aggregate
$\ddot{x}_{d-max}$	4.167	2.778	2.778	3.472	4.167	3.472	5.948	1.852	**5.948**
$\ddot{x}_{d-min}$	-4.63	-2.78	-3.47	-5.56	-5.56	-4.86	-5.96	-2.45	**-5.961**
$\ddot{x}_{d-min,n}$	-3.82	-3	-3.24	-4.28	-4.28	-3.93	-4.48	-3	**-4.481**
γ	0.631	0.6	0.71	0.514	0.576	0.529	0.764	0.639	**0.764**

The anticipated maximum deceleration rate of the leading vehicle is determined from Gipps' assumptions, as previously shown in Equation (4). This is not the currently recommended procedure; see section 2.2.3.2 for additional explanation on the work of Wilson [20], and Rakha and Wang [21].

TABLE 4-4 Parameter values calibrated from macroscopic optimization methods.

Parameter	D124	D304	D316	D350	D358	D363	D367	D462	Aggregate
τ_m	0.78	0.37	1.26	1.16	1.4	0.76	1.25	0.38	**1.1**
$\dot{x}_d$	104.9	115	96.5	93.8	116	115	94.4	115	**101.9**
$\dot{x}_{capacity}$	84.8	60	80.2	80.3	68.9	68.6	64.4	80	**82.2**
$q_{capacity}$	3,413	3,600	1,884	2,198	1,852	3,509	2,048	3,600	**2,383**
k_{jam}	149.3	153.7	123.6	137.5	161	150	169.5	131.2	**168.8**

TABLE 4-5 Vehicle dynamics parameters.*

Parameter	D124	D304	D316	D350	D358	D363	D367	D462	Aggregate
γ	0.631	0.6	0.71	0.514	0.576	0.529	0.764	0.639	**0.764**
$make$	Chevy	Toyota	Toyota	Toyota	Toyota	Toyota	?	Ford	-
$model$	?	Corolla	Corolla	Corolla	Camry	Camry	?	Explorer	-
$year$	2002	2002	2001	1997	1997	1997	?	1998	-
$\dot{x}_p$	179.3	179.3	179.3	179.3	215.4	215.4	179.3	145.7	**179.3**
P	90	90	90	90	145	145	90	119	**90**
M_{ta}	599.5	599.5	599.5	599.5	756.3	756.3	599.5	1045	**599.5**
C_d	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.5	**0.4**
A_f	2	2	2	2	2.18	2.18	2	2.94	**2**

*Blue highlighted areas indicate information assumed due to a lack of data.

TABLE 4-6 Parameter values calibrated from microscopic optimization methods.

Parameter	Min.	Max.	D124	D304	D316	D350	D358	D363	D367	D462	Aggregate
α	30	60	49.3	50.8	44.9	45.7	30	58.8	58.6	56.1	**30**
z_a	-0.5	1.0	0.299	-0.426	0.32	0.514	0.45	-0.07	0.208	0.28	**0.044**
z_d	0.0	1.0	2.224	1.539	1.838	2.188	1.874	1.878	2.143	2.262	**1.73**
l_a	1.5	3.0	0.002	0.62	0.15	0.33	0.136	0.46	0.336	0.011	**0**
l_d	1.5	3.0	2.017	2.523	1.729	2.026	1.613	2.465	2.204	2.181	**1.693**
τ_{GHR}	1.0	3.0	2.2	2.3	1.8	1.6	2	2	2.1	1.6	**2.4**
τ_s	0.1	5.0	1.323	3	2.522	2.093	2.365	2.371	1.904	2.332	**1.717**
δ	10	40	12.4	10	28.86	17.96	16.74	12.26	17.44	10.32	**16.79**

4.3.2 Optimized Parameter Values

A comparison of the calibrated parameters and corresponding error values for the eight drivers, as well as aggregate results, is given in table format. **TABLE 4-7** provides the calibrated parameter values and associated number of observations and error function values for the GHR model. **TABLE 4-8**, **TABLE 4-9**, and **TABLE 4-10** provide the same information for the Gipps model, the Intelligent Driver Model, and the RPA model, respectively.

TABLE 4-7 Results for calibrated parameters of the GHR model.

Parameter	D124	D304	D316	D350	D358	D363	D367	D462	Aggregate
τ	2.0	2.3	1.8	1.6	2.0	2.3	1.0	1.6	**2.3**
α	44.4	50.8	44.9	45.7	30.0	44.5	48.0	60.0	**34.3**
z_a	0.603	-0.426	0.320	0.514	0.450	-0.045	0.237	0.301	**0.030**
l_a	2.268	1.539	1.838	2.188	1.874	1.775	1.980	2.299	**1.696**
z_d	0.304	0.620	0.150	0.330	0.136	0.321	0.222	0.000	**0.024**
l_d	2.077	2.523	1.729	2.026	1.613	2.258	2.130	2.213	**1.733**
Δx_{jam}	6.698	6.506	8.091	7.273	6.211	6.667	5.900	7.622	**5.924**
No. Obs.	44,051	38,431	108,567	35,590	102,521	99,545	121,667	13,565	**551,406**
RMSPe	0.00082	0.00084	0.00102	0.00085	0.00118	0.00048	0.00110	0.00069	**0.00034**

TABLE 4-8 Results for calibrated parameters of the Gipps model.

Parameter	D124	D304	D316	D350	D358	D363	D367	D462	Aggregate
τ_m	0.8	0.4	1.3	1.2	1.4	0.8	1.3	0.4	**1.1**
Δx_{jam}	6.698	6.506	8.091	7.273	6.211	6.667	5.900	7.622	**5.924**
$\dot{x}_d$	29.14	31.94	26.81	26.06	32.22	31.94	26.22	31.94	**28.31**
$\ddot{x}_{d-max}$	4.167	2.778	2.778	3.472	4.167	3.472	5.948	1.852	**5.948**
$\ddot{x}_{d-min}$	-4.630	-2.778	-3.472	-5.556	-5.556	-4.861	-5.961	-2.451	**-5.961**
$\ddot{x}_{(n)-min}$	-3.815	-3.000	-3.236	-4.278	-4.278	-3.931	-4.481	-3.000	**-4.481**
No. Obs.	46,307	41,357	110,227	36,666	105,179	103,655	120,146	14,429	**578,274**
Error	0.00064	0.00033	0.00030	0.00094	0.00028	0.00019	0.00040	0.00055	**0.00014**

TABLE 4-9 Results for calibrated parameters of the IDM model.

Parameter	D124	D304	D316	D350	D358	D363	D367	D462	Aggregate
τ_S	1.323	3.000	2.522	2.093	2.365	2.371	1.904	2.200	**1.717**
Δx_j	6.698	6.506	8.091	7.273	6.211	6.667	5.900	7.622	**5.924**
$\dot{x}_d$	29.14	31.94	26.81	26.06	32.22	31.94	26.22	31.94	**28.31**
$\ddot{x}_{d-max}$	4.167	2.778	2.778	3.472	4.167	3.472	5.948	1.852	**5.948**
$\ddot{x}_{d-min}$	-4.630	-2.778	-3.472	-5.556	-5.556	-4.861	-5.961	-2.451	**-5.961**
δ	12.40	10.00	28.86	17.96	16.74	12.26	17.44	10.00	**16.79**
No. Obs.	45,367	37,507	106,243	34,514	101,192	99,545	117,104	13,133	**564,840**
Error	0.00142	0.00172	0.00101	0.00181	0.00069	0.00068	0.00057	0.00328	**0.00026**

TABLE 4-10 Results for calibrated parameters of the RPA model.

Parameter	D124	D304	D316	D350	D358	D363	D367	D462	Aggregate
c_3	0.703	0.252	1.474	1.257	1.398	0.439	1.274	0.507	**1.189**
Δx_j	6.698	6.506	8.091	7.273	6.211	6.667	5.900	7.622	**5.924**
c_1	6.322	1.039	7.756	7.067	3.309	3.617	4.619	6.163	**5.584**
$\dot{x}_d$	29.14	31.94	26.81	26.06	32.22	31.94	26.22	31.94	**28.31**
$\ddot{x}_{d-min}$	-4.630	-2.778	-3.472	-5.556	-5.556	-4.861	-5.961	-2.451	**-5.961**
c_2	10.97	174.64	8.96	5.36	93.53	97.43	33.57	46.60	**9.63**
γ	0.631	0.600	0.710	0.537	0.576	0.529	0.764	0.639	**0.764**
No. Obs.	47,623	41,819	114,211	39,625	110,938	105,573	126,260	14,645	**600,664**
Error	0.00068	0.00030	0.00040	0.00105	0.00042	0.00015	0.00049	0.00036	**0.00017**

One result of note within this data is the set error values produced for the Gipps model, if the perception-reaction time is optimized during microsimulation as a variable, or if the perception-reaction time is determined from the steady-state macroscopic formulation and held through to calculate errors. By varying the input value for the perception-reaction time in the Gipps model, and plotting the time against the resulting error generated by the simulation, is

shown in **FIGURE 4-2.** that the macroscopic optimization of the Gipps model yields results that are consistent with the results obtained by optimizing the .

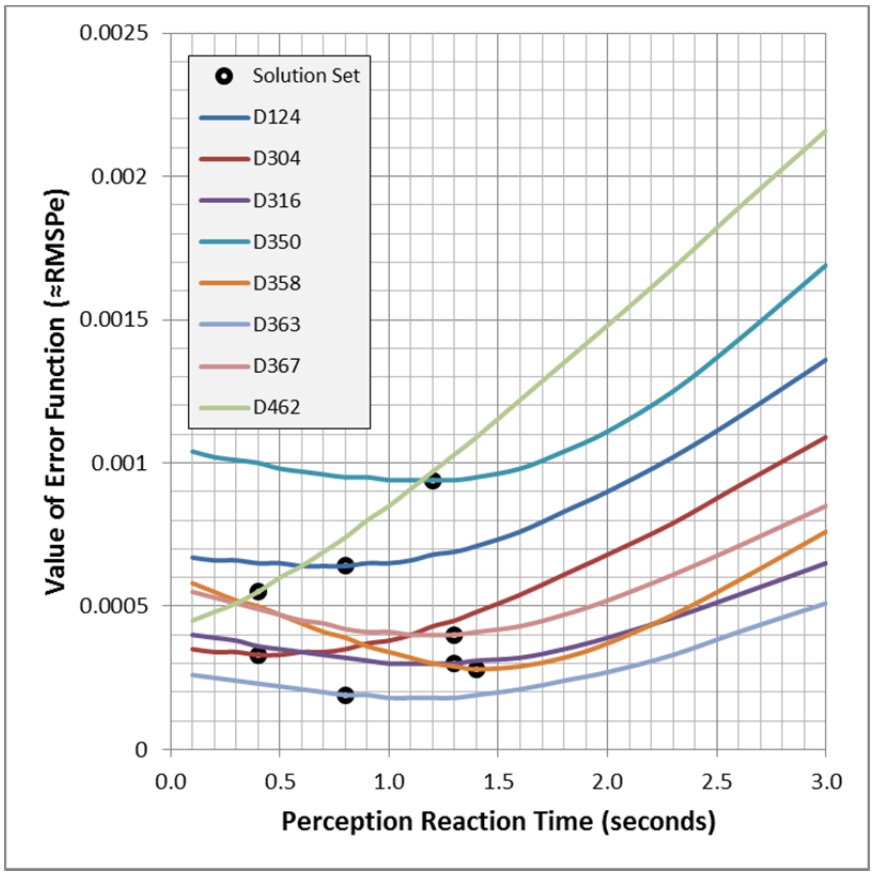


FIGURE 4-2 Effect on error of changing the perception-reaction time in the Gipps model.

The points labeled as "solution set" in **FIGURE 4-2** represent the values calibrated from macroscopic measures, and for six of the eight drivers are located at the optimized minimum value of the error function.

4.3.3 Visual Representations of the Solution Set

The macroscopic diagrams for the eight drivers included in the car-following model analysis herein are shown in **FIGURE 4-3** through **FIGURE 4-10**, along with the best-fit curve lines for the steady-state Van Aerde model, the steady-state portion of the RPA model, and the steady-state Gipps model, as calibrated with the PRT based upon the macroscopic formulation.

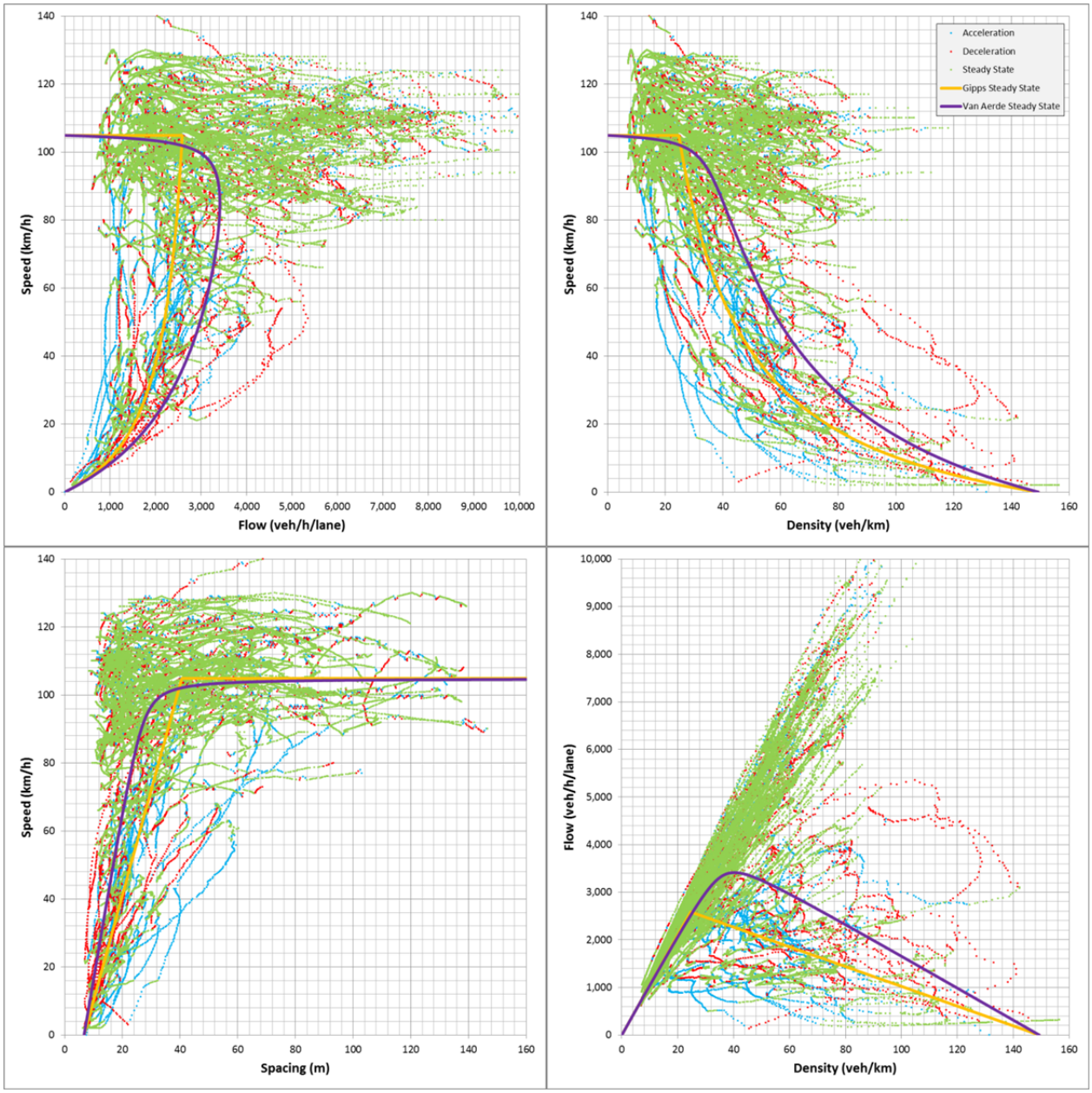


FIGURE 4-3 Macroscopic diagrams of car-following event data: Driver 124.

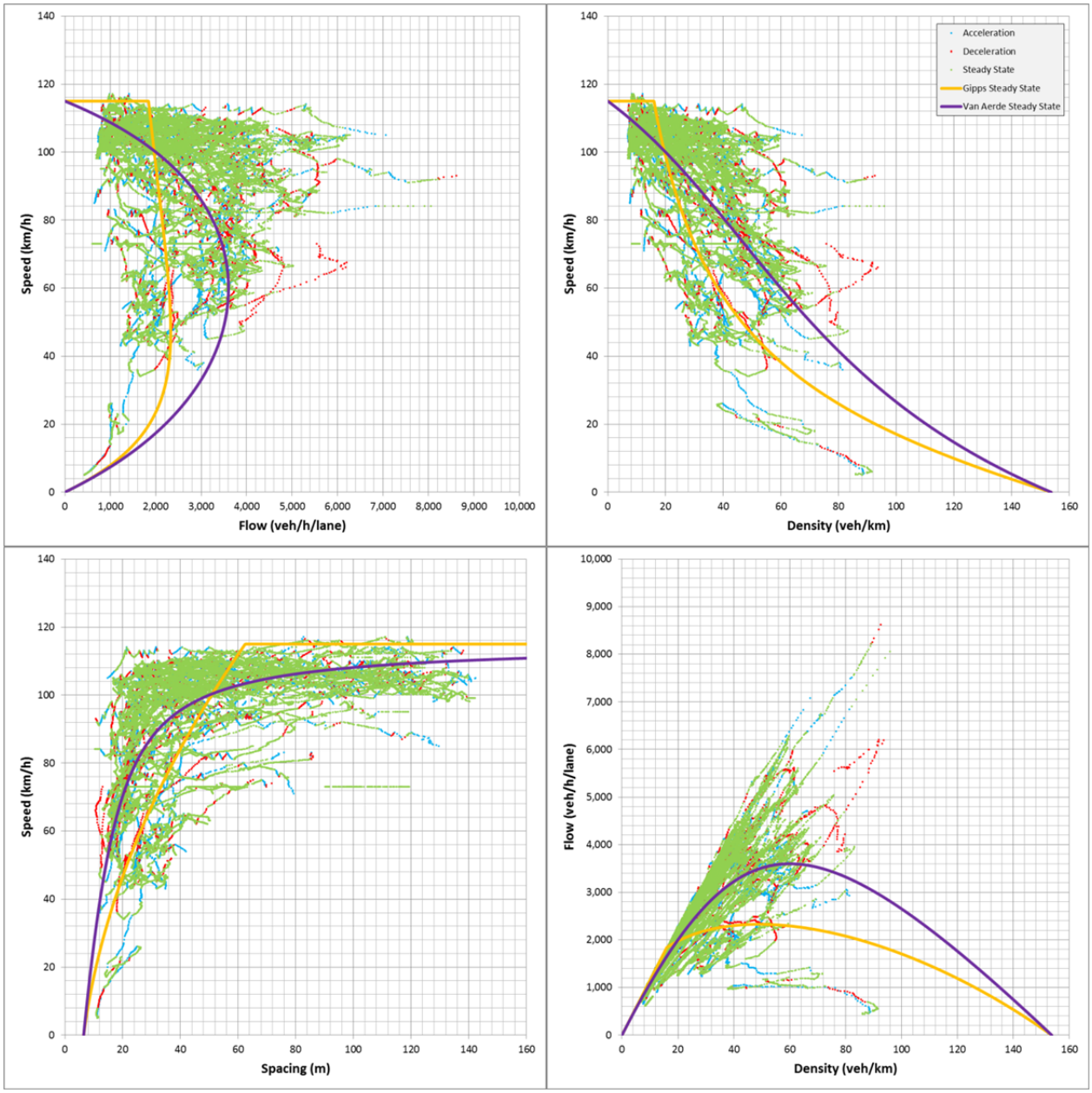


FIGURE 4-4 Macroscopic diagrams of car-following event data: Driver 304.

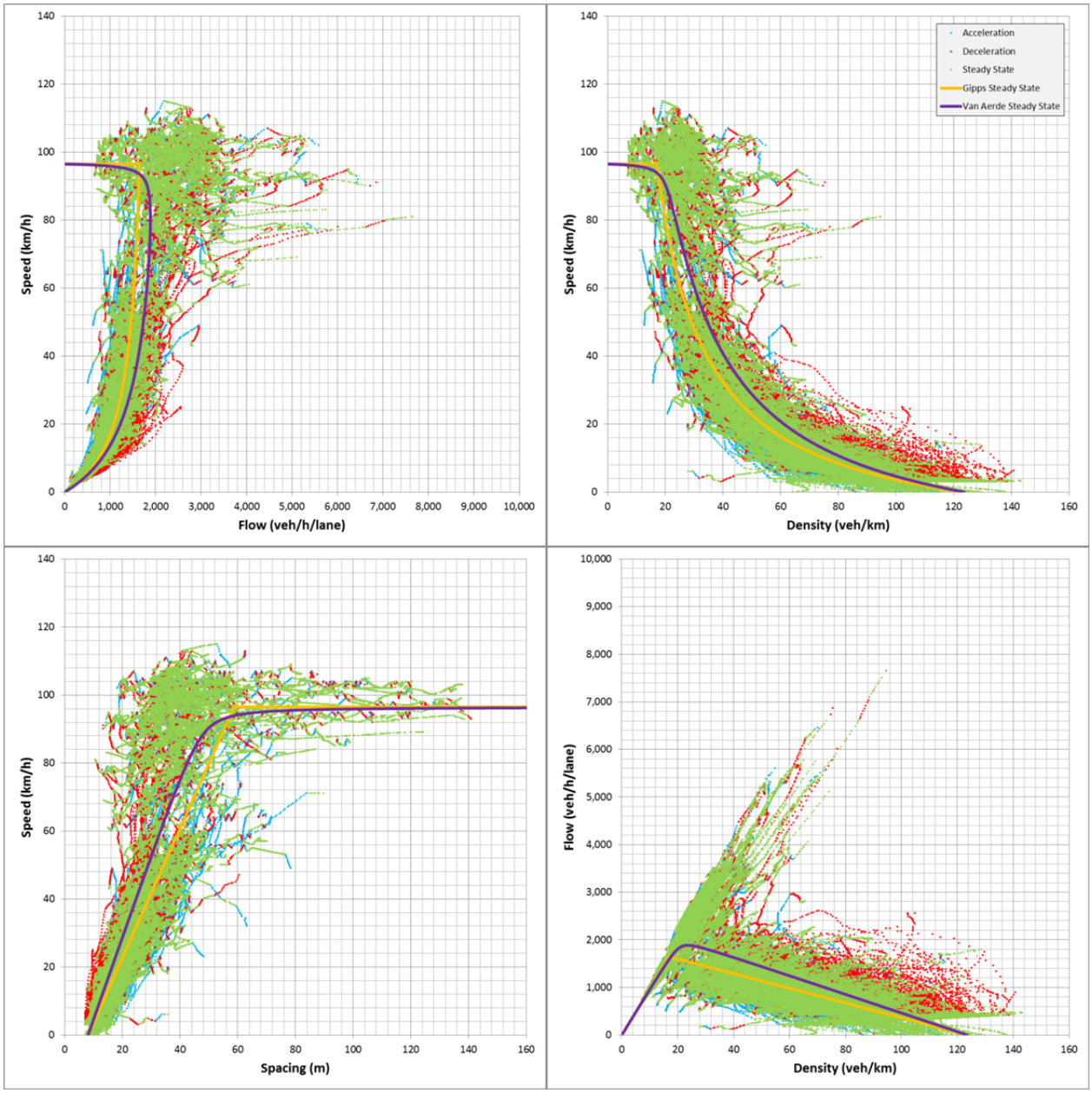


FIGURE 4-5 Macroscopic diagrams of car-following event data: Driver 316.

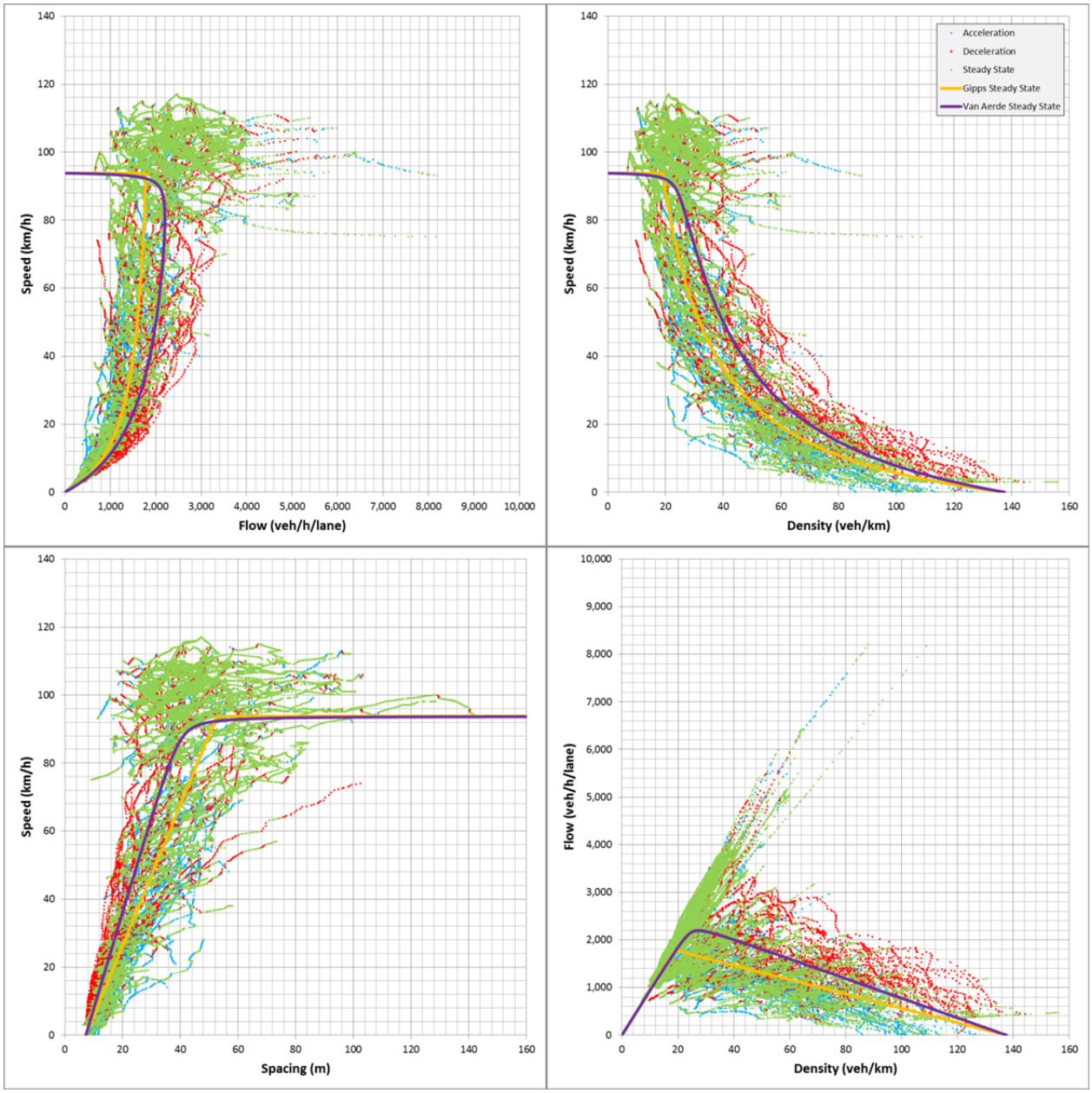


FIGURE 4-6 Macroscopic diagrams of car-following event data: Driver 350.

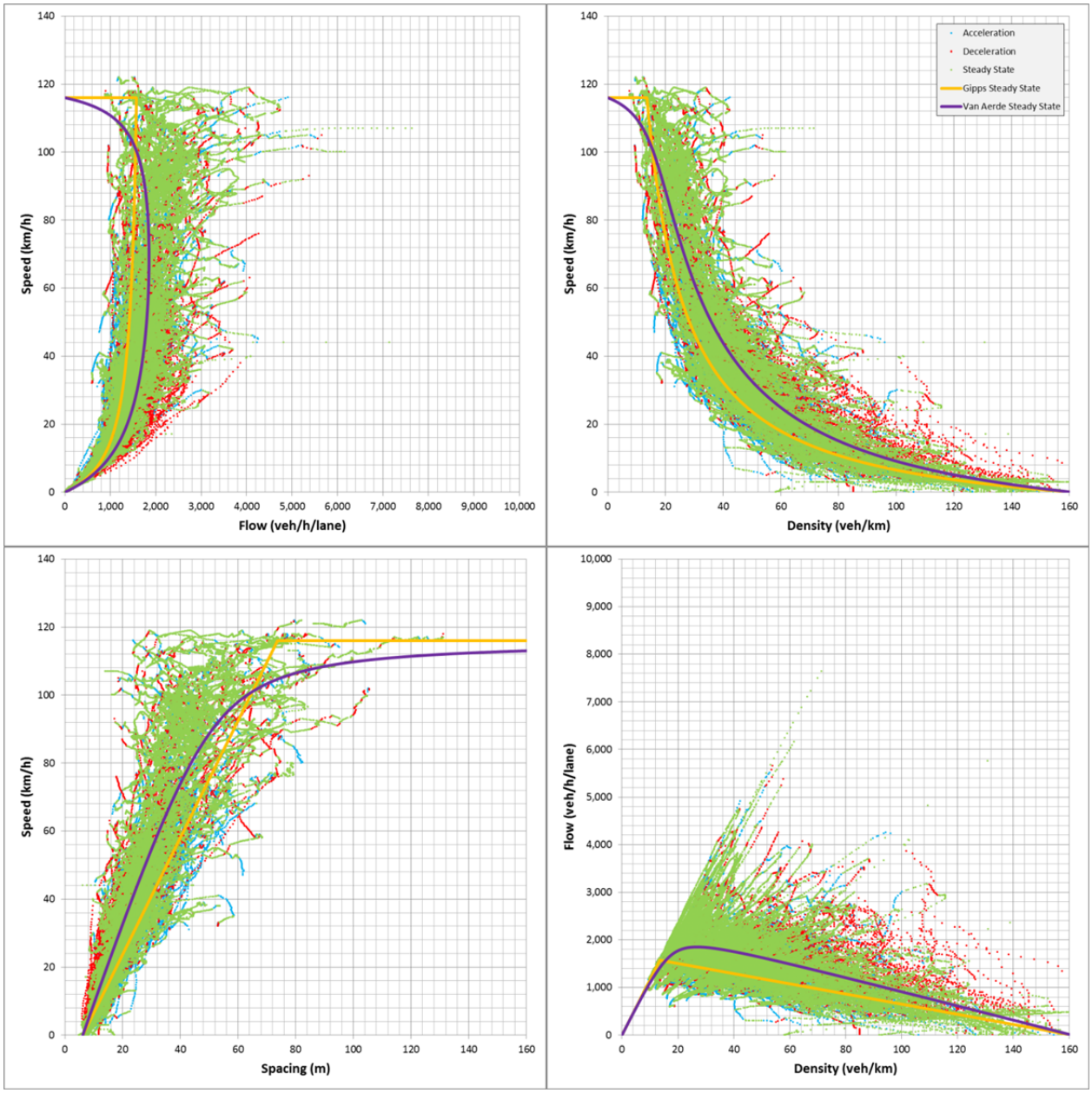


FIGURE 4-7 Macroscopic diagrams of car-following event data: Driver 358.

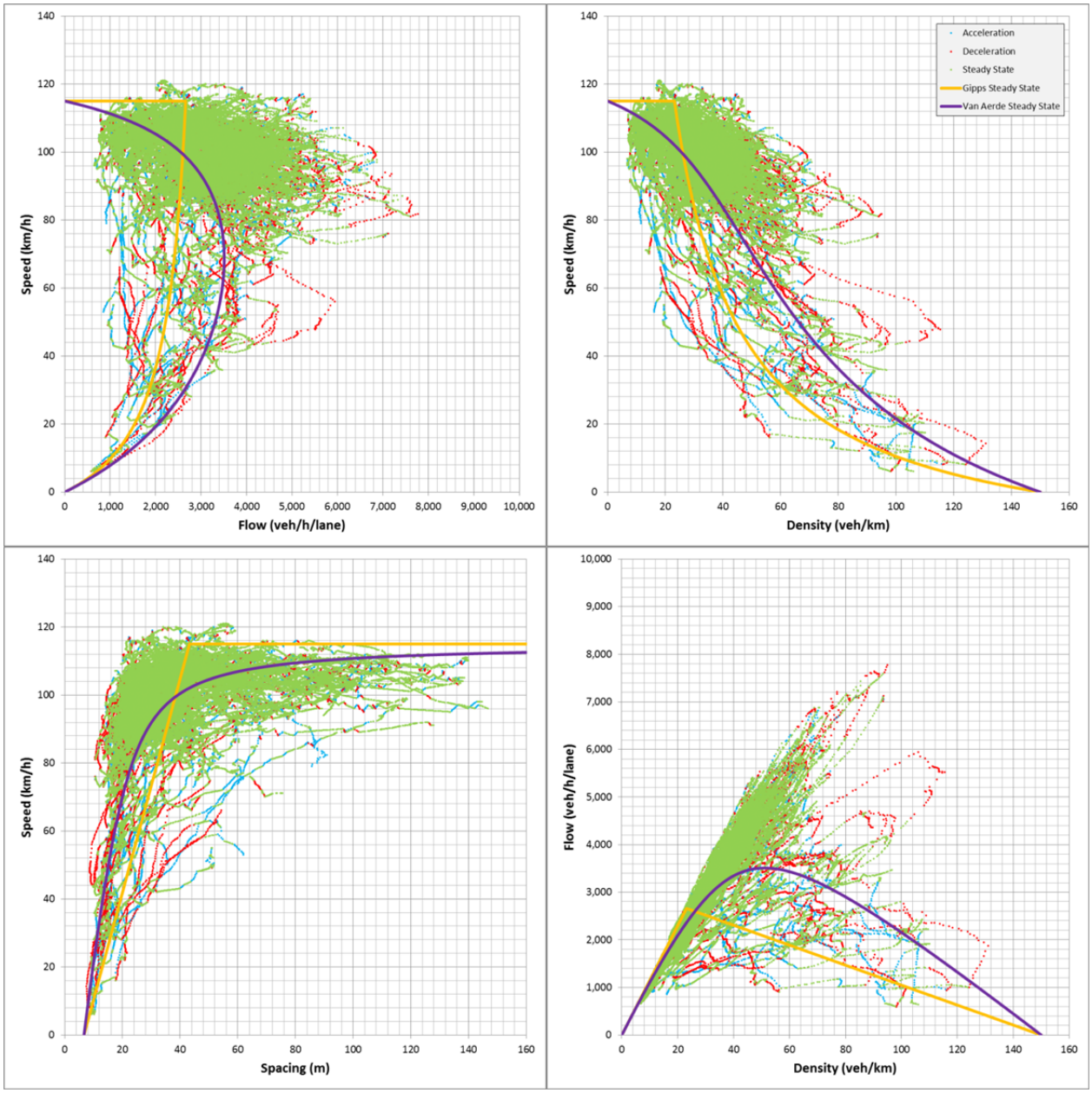


FIGURE 4-8 Macroscopic diagrams of car-following event data: Driver 363.

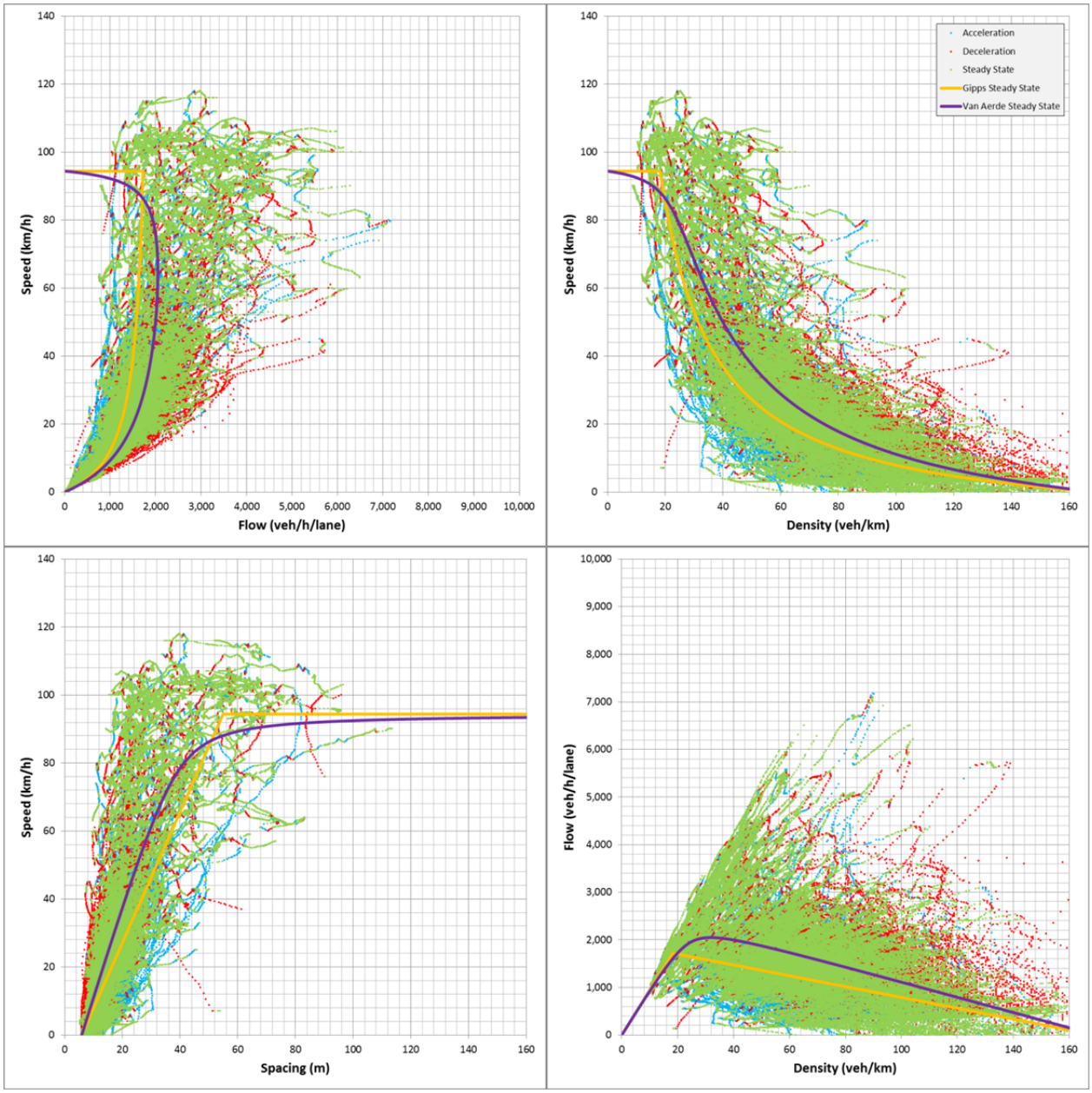


FIGURE 4-9 Macroscopic diagrams of car-following event data: Driver 367.

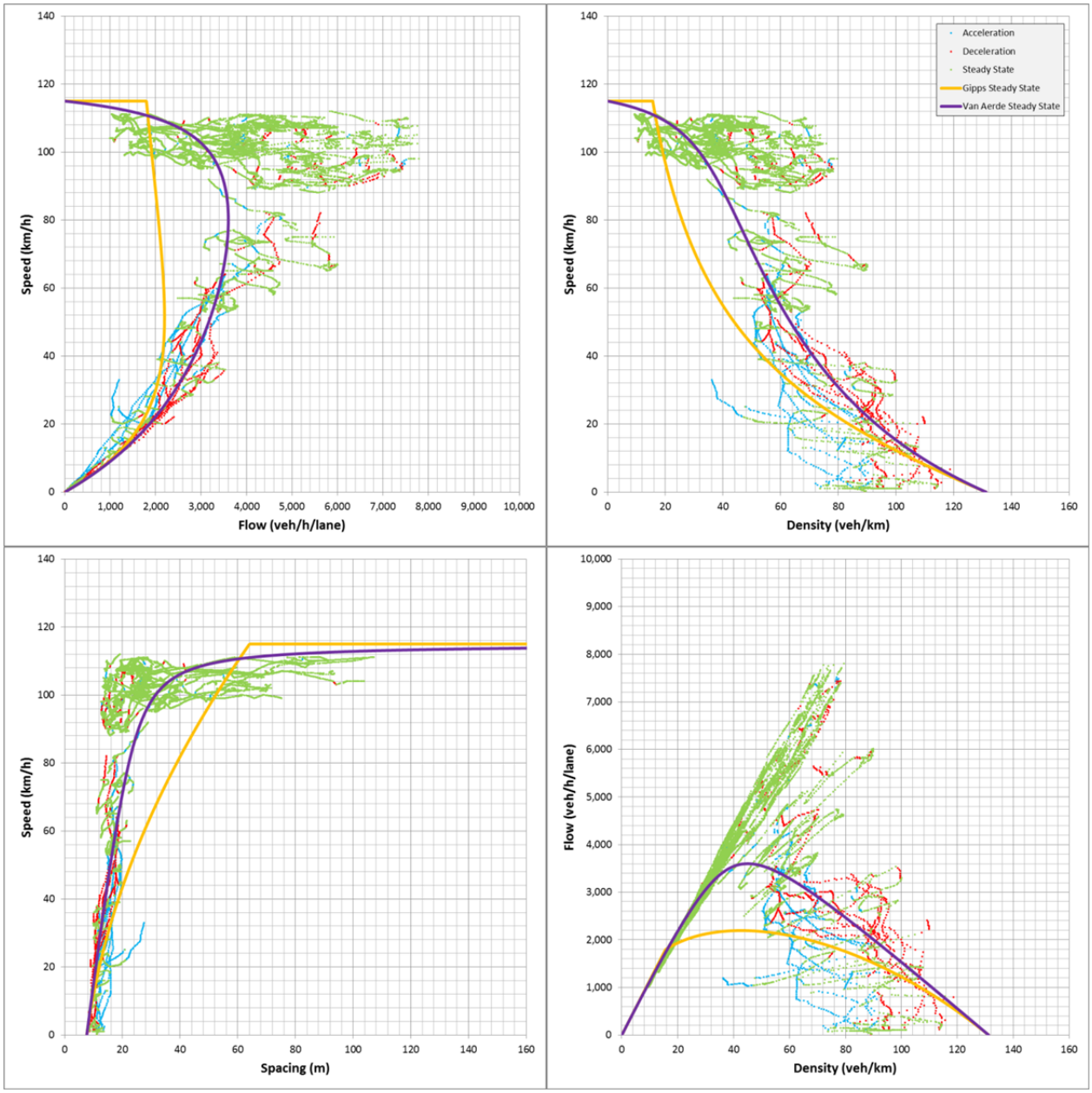


FIGURE 4-10 Macroscopic diagrams of car-following event data: Driver 462.

The visual representation of the solution set demonstrates how best-fit parameters may minimize the error for a given driver or a given facility, but are not representative of the behavior of each individual event. Drivers 316, 350, 358, and 367 appear to have excellent regime coverage for congested conditions, but may not have as much uncongested information as is desirable. In contrast, drivers 124, 304, 363, and 462 have very little data in the congested regime of traffic, but appear to have sufficient data collected in free-flow conditions. This may affect performance for various models, as some may perform better in one regime over another.

4.4 Analysis of Results

4.4.1 Comparison of Macroscopic Error Measures

As a Root Mean Square Percentage error value is used in the analysis, the error values may be compared directly both between models for a given driver, and between drivers for a given model, as shown in **TABLE 4-11**. Against the current dataset, the Gipps model performs the best among the models, performing better than the alternate models in the aggregate case, and in four of the eight individual drivers. The RPA model performed second best, with the best solution for three of the eight drivers, and the GHR model performed best in modeling the performance of one of the drivers.

TABLE 4-11 Comparison of error function values.

Error	D124	D304	D316	D350	D358	D363	D367	D462	Aggregate
GHR	0.00082	0.00084	0.00102	**0.00085**	0.00118	0.00048	0.00110	0.00069	0.00034
Gipps	**0.00064**	0.00033	**0.00030**	0.00094	**0.00028**	0.00019	**0.00040**	0.00055	**0.00014**
IDM	0.00142	0.00172	0.00101	0.00181	0.00069	0.00068	0.00057	0.00328	0.00026
RPA	0.00068	**0.00030**	0.00040	0.00105	0.00042	**0.00015**	0.00049	**0.00036**	0.00017

As observed from the macroscopic representation of the data solution, drivers 316, 350, 358, and 367 appear to have excellent regime coverage for congested conditions, while drivers 124, 304, 363, and 462 have very little data in the congested regime of traffic, but appear to have sufficient data collected in free-flow conditions. The comparison of error function values shows that in three of the four cases where data is heavily focused on the congested regime, the Gipps model performs best, while in three of the four cases where data is heavily focused on the free-flow regime, the RPA model performs best. An additional consideration is that the Gipps model relies on the calibrated maximum and minimum acceleration values as a limiting factor, whereas the RPA model relies on the vehicle dynamics. In the case of modeling individual drivers in future simulation, the RPA model provides the flexibility for a driver to change vehicles without the need to subsequently recalibrate parameter values for that driver. Furthermore, given that the

RPA model also captures the effect of pavement conditions, aerodynamic features of the vehicle, and the density of air, the model can be easily adjusted to reflect local conditions without the need to re-calibrate the parameters. All other models would require a re-calibration.

4.4.2 Comparison of Microscopic Error Measures

Though the RPA and Gipps models are calibrated herein entirely based on macroscopic data measures, and the IDM is calibrated largely on macroscopic data measures, all four models are assessed for error based on their ability to match a simulated following vehicle to an observed following vehicle, given the input of an observed leading vehicle, and a microscopic interaction. When examining the macroscopic diagrams for each driver shown in **FIGURE 4-3** through **FIGURE 4-10**, it becomes apparent that each model may predict the mean behavior, but it will necessarily not be able to consistently predict the individual trip behavior for a given driver. Individual trips are expected to exhibit higher or lower performance for each car-following model compared to the aggregate error calibrated for a given driver. The benefit of examining the individual events is to make observations regarding how the models are implemented. When examining individual car-following events there are a number of different graphical representations that can be used; since the error function used herein is a function of the space headway in meters and the velocity in meters per second, the graphical representation of the individual events will be presented using these measures.

The first microscopic sample examined is event number 163614.29 for Driver 367, as seen in **FIGURE 4-11**. For this event, the following errors were recorded: GHR = 0.01463; Gipps = 0.00578; IDM = 0.00576; RPA = 0.00634. In the case of this individual event, the GHR model, shown in orange, appears to be insensitive to large shifts in the velocity of the leader vehicle, which results in the minimum space headway governing at the beginning of the event, and an increasing space headway at the end of the event. The desired free-flow speed appears to be a limiting factor on the Gipps, IDM, and RPA models near the end of the event, as both the observed leading vehicle and following vehicle exceed the calibrated desired free flow speed, resulting in an increase in the space headway predicted by these models. In the time-velocity plot, the IDM model appears to be slightly over-sensitive to changes in speed, with spikes that exceed either the observed behavior or the behavior predicted by the other models, though in this case it does not appear to negatively impact the predicted space headway.

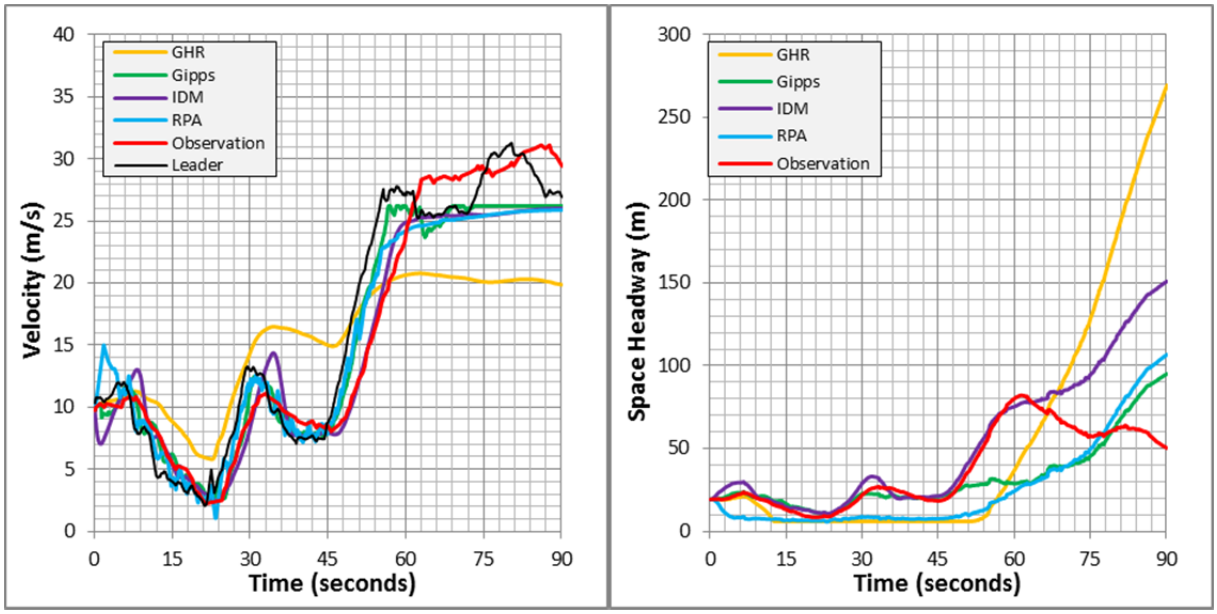


FIGURE 4-11 Examination of individual car-following event: Driver 367, 163614.29.

Event number 110803.01 for driver 304 is shown in **FIGURE 4-12**. For this event, the following errors were recorded: GHR = 0.00606; Gipps = 0.00263; IDM = 0.00895; RPA = 0.00267. As was also seen in the previous event, the GHR model, shown in orange, appears to be insensitive to large shifts in the velocity of the leader vehicle, which results in the minimum space headway governing at the beginning of the event, and an increasing space headway at the end of the event. The event shown is in the congested regime, so the limitation of the desired free-flow speed does not impact the Gipps, IDM, and RPA models as it did in the previous event. In the time-velocity plot, the IDM model is seen to be much more sensitive in this event than in the previous event, with large spikes that exceed either the observed behavior or the behavior predicted by the other models. The behavior predicted by both the RPA model and the Gipps model for this event appears to be more aggressive than the observed behavior of the following vehicle, as indicated by the small space-headway distances predicted.

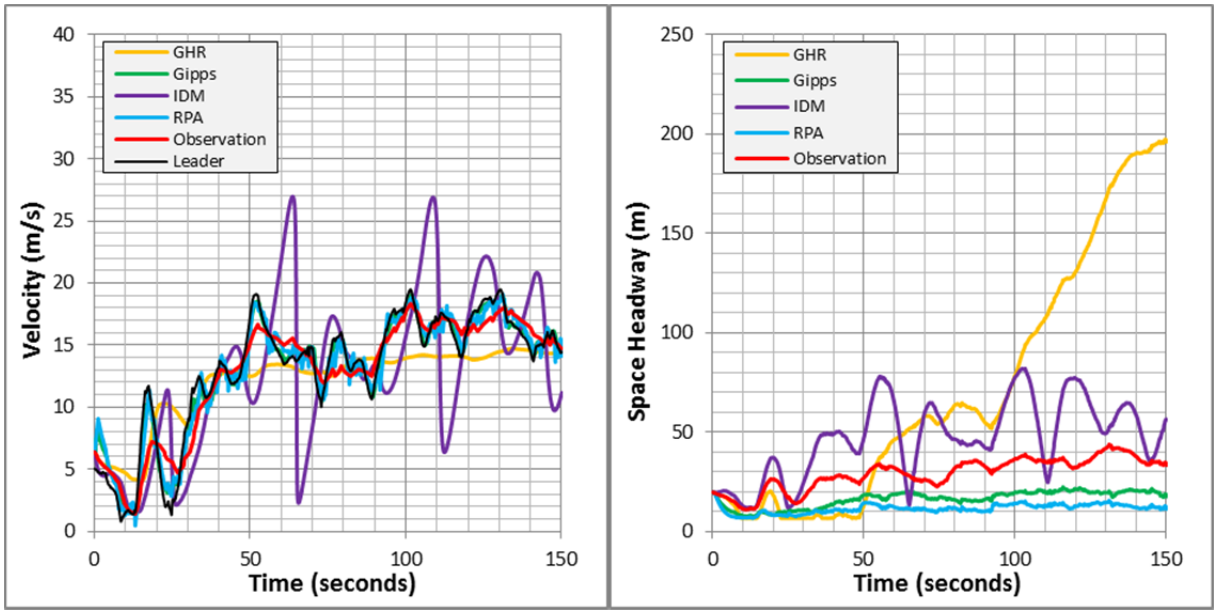


FIGURE 4-12 Examination of individual car-following event: Driver 304, 110803.01.

Event number 129453.02 for driver 124 is shown in **FIGURE 4-13**. For this event, the following errors were recorded: GHR = 0.00241; Gipps = 0.00295; IDM = 0.00363; RPA = 0.00268. The event shown, while mostly in the uncongested regime, exhibits behavior by the observed follower vehicle that does not exceed the desired free flow speed, so the limitation of the desired free-flow speed does not have the same impact the Gipps, IDM, and RPA models as was shown in **FIGURE 4-11**. In the time-velocity plot, the IDM model is seen to continue to have sensitivity issues, as seen in the previous events. The increased space headway caused by the IDM sensitivity turns into a benefit in the case of this individual event, as the following vehicle is seen to back off from the leading vehicle towards the end of the event, which is not predicted by any of the four simulation models. This is seen as an indication of external environmental factors which were not recorded as part of the current data reduction process; the observed follower vehicle may be slowing in anticipation of changing lanes or taking an off-ramp, or the follower vehicle may be reacting to a downstream traffic pattern in advance of the lead vehicle's reaction to the same pattern.

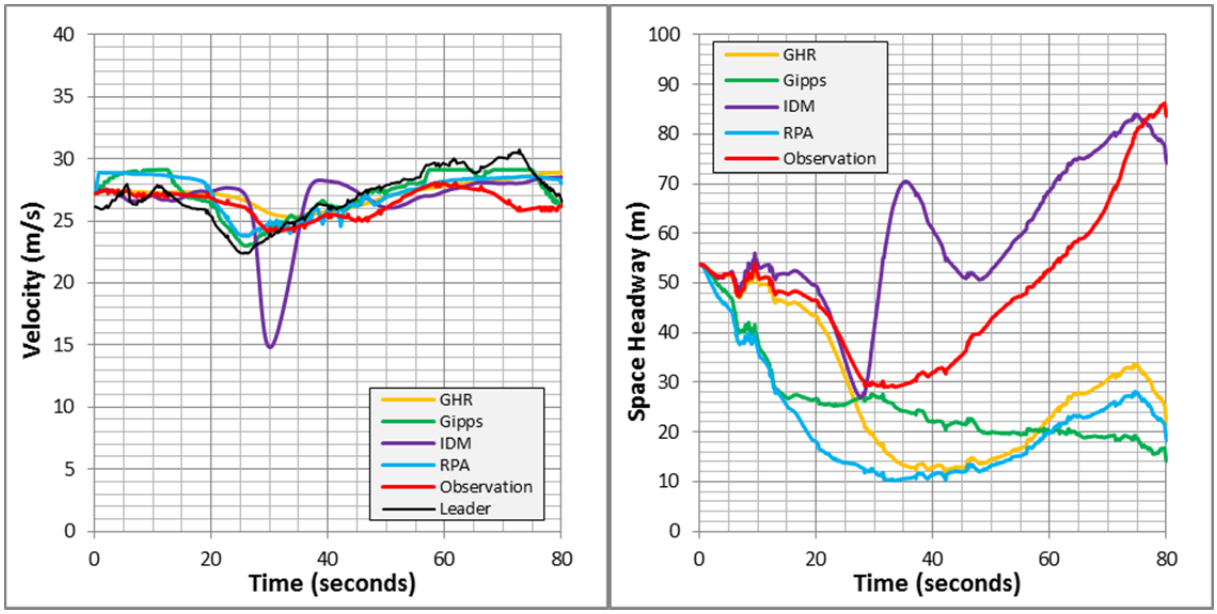


FIGURE 4-13 Examination of individual car-following event: Driver 124, 129453.02.

The individual trips display unique results when examining the simulated behavior of the following vehicle. Recognizing that no model is able to predict the behavior of every event due to the stochastic nature of driver's behavior, a quantitative approach to analyzing the distribution of the errors from a microscopic perspective would be to look at a plot of the error function value against the length of the event, as seen in **FIGURE 4-14**. Much of the data is for the GHR, Gipps, and IDM models is occluded in the figure due to the requirement of some data appearing on top, chosen in this case based on the order in which the models are presented throughout this paper. The event-specific error values shown are based on the parameter values calibrated to fit the aggregate datasets, and are not based on the parameter values calibrated to fit individual drivers. In general, the Gipps model is seen to generate lower individual event errors, the GHR model is seen to generate higher individual event errors, and the RPA model is seen to generate the most consistent level of error among all events.

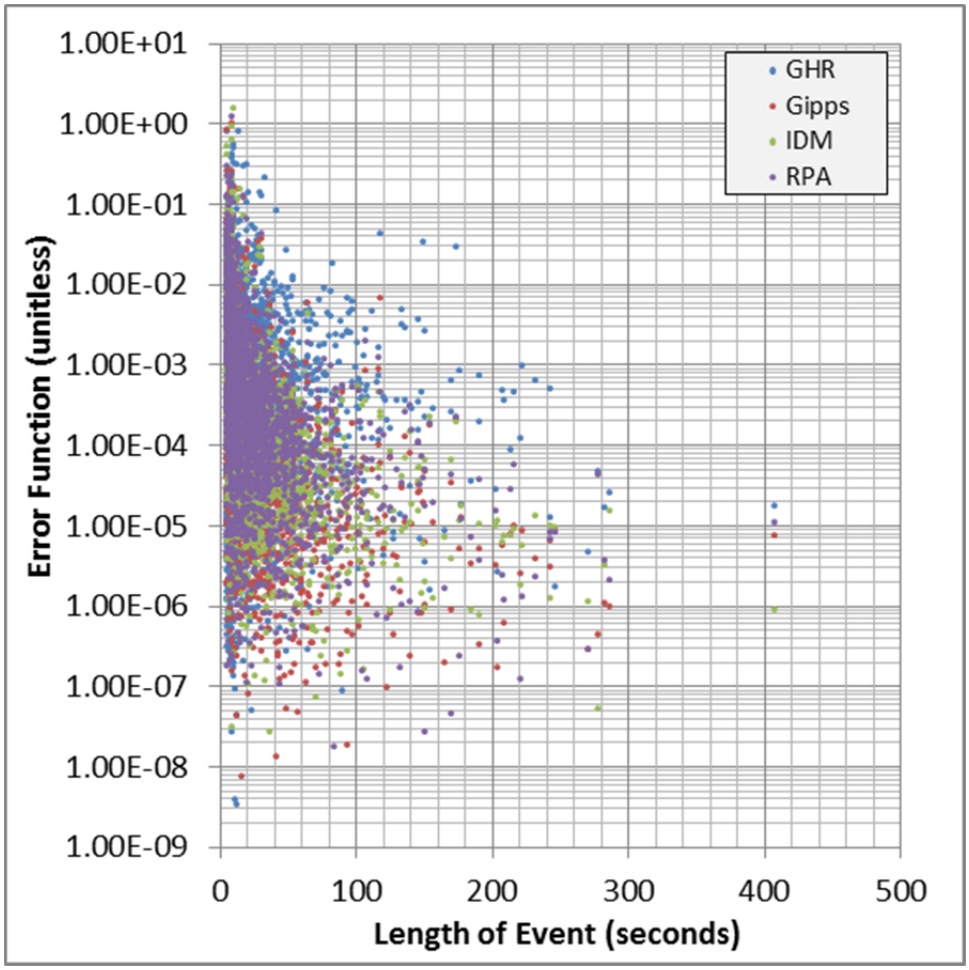


FIGURE 4-14 Distribution of event error for each of the simulation models.

4.4.3 Correlation of Parameters

As discussed within the literature review, the concept of correlation between parameter values has only recently been investigated, as datasets are becoming available which allow researchers to compare values across different drivers. In addition to the parameter values used for simulation in the car-following models, the category of age has been added to check for correlations between age and various parameters. The coefficient of correlation value provides a measure of linear correlation between the parameter listed in the row and the parameter listed in the column, as seen **in TABLE 4-12**. The more positive values indicate that as one variable increases the other variable also increases, and more negative values indicate that as one variable increases, the other variable shows a corresponding decrease in value. Cells with a blue hue indicate a positive linear correlation, while cells with a red hue indicate a negative linear correlation.

TABLE 4-12 Linear correlation coefficient values between parameters.

(a) First portion of the correlation coefficient values between parameters.

Parameter	*Age*	a(max)	a(min)	v(des)	v(cap)	q(cap)	k(jam)	*tao(ghr)*
Age								
a(max)	0.2132							
a(min)	-0.587	-0.841						
v(des)	-0.029	-0.425	0.4203					
v(cap)	-0.088	-0.303	0.0979	-0.322				
q(cap)	-0.442	-0.459	0.5767	0.6042	-0.023			
k(jam)	0.1269	0.7888	-0.581	0.147	-0.693	-0.076		
tao(ghr)	-0.509	0.4471	-0.09	0.2496	-0.568	0.2478	0.6935	
tao(s)	0.0165	-0.535	0.5151	0.4076	-0.559	0.0621	-0.173	-0.027
tao(m)	0.5134	0.5938	-0.751	-0.567	0.0708	-0.947	0.1825	-0.145
a(ghr)	-0.152	0.0077	0.2157	-0.082	-0.121	0.5606	0.0072	0.0558
z(a)	0.4502	0.1845	-0.455	-0.461	0.6779	-0.625	-0.248	-0.624
z(d)	0.2215	0.201	-0.286	-0.381	0.6608	-0.043	-0.134	-0.497
l(a)	0.0258	0.0828	-0.045	0.0749	-0.776	0.1495	0.3498	0.388
l(d)	-0.216	-0.109	0.2779	0.2769	-0.475	0.7417	0.2142	0.324
delta	0.1585	0.1173	-0.201	-0.662	0.2678	-0.81	-0.372	-0.315
throttle	-0.345	0.3449	0.0957	-0.415	-0.048	-0.289	0.0918	0.1758

(b) Second portion of the correlation coefficient values between parameters.

Parameter	*tao(s)*	*tao(m)*	*a(ghr)*	*z(a)*	*z(d)*	*l(a)*	*l(d)*	*delta*
tao(m)	-0.252							
a(ghr)	-0.1	-0.541						
z(a)	-0.552	0.6546	-0.461					
z(d)	-0.795	0.1077	0.2229	0.6924				
l(a)	0.5747	-0.184	0.2802	-0.738	-0.663			
l(d)	0.2404	-0.734	0.7962	-0.808	-0.204	0.6729		
delta	0.0319	0.7501	-0.361	0.4766	-0.063	-0.194	-0.666	
throttle	-0.165	0.1867	0.2481	0.0402	0.1288	-0.258	-0.158	0.3894

With data reduction completed for only eight drivers at this time, it is difficult to derive statistically significant solutions in regard to the correlation of parameters. It is recommended that this analysis be followed up upon as more driver data becomes available.

4.5 Conclusions Regarding the Analysis of Car-following Models

Based upon the data analysis conducted herein, the Gipps model is found to slightly outperform the RPA model in simulating the car-following behavior of a following vehicle, given the behavior of the leading vehicle as an input. Compared to the Gipps and RPA models, neither the Intelligent Driver Model nor the GHR model provided competitive errors. The application of the models with their respective calibrated parameters shows that while the optimization procedures have determined a best-fit line to match the observed behavior, the models do not account for the scatter observed within individual drivers, and thus more research is required to capture these differences. The naturalistic dataset allows for a breadth of analysis not previously available, but the results included herein are greatly constrained by the small amount of data thus far reduced into car-following events.

4.5.1 Review of the Calculation of Optimized Model Parameters

All of the parameter values fixed from observing the dataset are based on maximum or minimum values of a given value, and include the desired maximum acceleration of the follower vehicle, the desired maximum deceleration of the follower vehicle, and the maximum throttle depression of the follower vehicle. The anticipated desired maximum deceleration of the leader vehicle is directly derived from the maximum deceleration of the follower vehicle, and as such is included in the direct observation calibration category.

The vehicle dynamics component of the RPA model may also be considered to be observed, but in a different way than the acceleration and throttle information. All of the vehicle dynamics information must be obtained by examining the year, make, and model of the follower vehicle, and the parameter values are based on the specifications of the vehicle. In the case of driver 124 and driver 367, the specific make/model information of the vehicle was not available, so the vehicle dynamics were assumed to be equal to that of an early 2000's Toyota Corolla, the majority condition among other drivers, though not the most conservative case. Though the most conservative vehicle dynamics model would be that of the Ford Explorer driven by driver 462, the vehicle dynamics limitations for the aggregate data were also set equal to that of the Toyota Corolla, what was considered to be the average case.

The heuristic optimization program SPD-CAL determines a best fit line through macroscopic data points based on the desired free flow speed, the speed at capacity, the flow rate

at capacity, and the jam density. These four values can be used directly to calculate the space headway at jam density, the perception-reaction time for use with the Gipps model, and the three Van-Aerde steady state parameter values for the RPA model. The procedure of using macroscopic data to generate the perception-reaction time for the Gipps model was checked against a plot showing how the error value changes as perception-reaction time changes, and the method was validated.

Some parameter values cannot be calculated either from observation or from macroscopic measures, and must be optimized during microscopic simulation of the leader/follower pair. The delta value of the Intelligent Driver Model, and all six of the parameters used in the GHR model must be calibrated in this fashion. Each of these parameters was given upper and lower boundary conditions based on a literature review of previous solution sets. These boundary conditions become important in cases with a large number of calibration parameters and a small number of observations, as it is possible to obtain calibrated parameters which minimize the error function, but are not representative of real-world conditions.

When optimizing parameter values through microscopic simulation, many of the potential combinations of variable values result in behavior outside of the feasible region, including following vehicles which pass the leader vehicle and proceed to apply a negative velocity to return to the desired spade headway. In order to prevent crashing of the optimization procedure, the velocity of the follower vehicle was taken to be the maximum of the projected simulated velocity, and 0.1 meters/second, to prevent simulated vehicles from stopping and reversing direction while on the highway facility. An additional constraint was added to the GHR model setting the vehicle position to be equal to the maximum of the simulated position, and a position equal to the jam density distance behind the position of the leader vehicle.

The visual representation of the solution set serves to show how calibrated best-fit parameters may minimize the error for a given driver or a given facility, but these values are not representative of the behavior of each individual event. A potential hurdle for working with naturalistic data is exposed when examining the various macroscopic diagrams, as steady-state data is required from both congested and uncongested regimes in order to calibrate parameter values for individual drivers. Drivers 316, 350, 358, and 367 appear to have excellent regime coverage for congested conditions, but may not have as much uncongested information as is desirable. In contrast, drivers 124, 304, 363, and 462 have very little data in the congested

regime of traffic, but appear to have sufficient data collected in free-flow conditions. This difference in coverage of free-flow and congested regimes for the data proved to have an impact on model performance.

4.5.2 Review of the Analysis of Results

In the aggregate case, the Gipps model performed best with an error function value of 0.00014, while the RPA model was close behind with an error function value of 0.00017, an increase of 21 percent. The GHR model simulation resulted in an error function value of 0.00034, an increase of one 143 percent, and the IDM model simulation resulted in an error function value of 0.00026, an increase of 86 percent. In the aggregate, the Gipps and RPA models were found to produce competitive simulation results, while the GHR and IDM models were not.

It was observed from the macroscopic representation of the data solution, drivers 316, 350, 358, and 367 appear to have excellent regime coverage for congested conditions, while drivers 124, 304, 363, and 462 have very little data in the congested regime of traffic, but appear to have sufficient data collected in free-flow conditions. The comparison of error function values for individual drivers shows that in three of the four cases where data is heavily focused on the congested regime, the Gipps model performs best, while in three of the four cases where data is heavily focused on the free-flow regime, the RPA model performs best. It may be possible to improve upon the current results by modifying the SPD-CAL optimization of macroscopic data points by adding constraints based on observed free-flow speed and jam density values.

As an additional consideration, the Gipps model relies on the calibrated maximum and minimum acceleration values as a limiting factor, whereas the RPA model relies on the vehicle dynamics. In the case of modeling individual drivers in future simulation, the RPA model provides the flexibility for a driver to change vehicles without the need to subsequently recalibrate parameter values for that driver. One potential outcome of examining naturalistic data is that simulation of individual driver's behavior may be improved, and the ability to couple models to a specific driver, independent of the vehicle, may be an advantage. Furthermore, the RPA model can capture changes in roadway surface type (asphalt vs. rigid pavement), condition (dry, wet, or snowy), and location altitude on the vehicle acceleration performance without the need for re-calibration. The model can also capture the effect of roadway grades on vehicle performance which cannot be captured by the other models.

When examining specific lead-follower pairs on the microscopic level, observations can be made about the shortcomings or advantages of each model which may not be apparent when examining the results at the macroscopic level. Since the error function used herein is a function of the space headway in meters and the velocity in meters per second, the graphical representation of the individual events is presented using these measures. The observations made are necessarily not blanket statements which can be applied to every car-following event in the dataset, but are specific to a particular event and may extend to a general comment.

The success of each model is somewhat consistent between events within a given driver, but varies stochastically over time, and is greatly affected by the traffic regime in which it acts, and for many of the models is impacted by the degree to which the leading vehicle has large and/or quick shifts in velocity. The GHR model in particular is observed to not deal well with large changes in the velocity of the leading vehicle, and appears to level out above the observed value in the case of deceleration events, and below the observed value in the case of acceleration events. At high velocities, the Gipps, IDM, and RPA models all become asymptotic against the calculated desired free-flow speed, which may serve useful in the case of highway travel when not following another vehicle, but was observed to not be accurate in comparison to the observed follower vehicle. In the time-velocity plot, the IDM model is seen to have sensitivity issues which cause the velocity to oscillate around the observed value, leading to increased space-headway in comparison to the observed follower vehicle. In one particular case, the increased space headway caused by the IDM sensitivity turned into a benefit, as the following vehicle is seen to back off from the leading vehicle towards the end of the event, which is not predicted by any of the four simulation models. This example scenario raised the issue of external environmental factors which were not recorded as part of the current data reduction process; the observed follower vehicle may be slowing in anticipation of changing lanes or taking an off-ramp, or the follower vehicle may be reacting to a downstream traffic pattern in advance of the lead vehicle's reaction to the same pattern.

Recognizing that no model is able to predict the behavior of every event due to the stochastic nature of driver's behavior, the error function value was examined against the length of each individual event. In general, the Gipps model is seen to generate lower individual event errors, the GHR model is seen to generate higher individual event errors, and the RPA model is seen to generate the most consistent level of error among all events.

In examining the correlation between parameter values, the category of age was added to check for correlations between age and various parameters. The coefficient of correlation value provides a measure of linear correlation between each of the parameters, but no significance calculations are provided herein, so the analysis of these values is limited at this time.

Chapter 5 Thesis Conclusions and Recommendations for Further Research

The driver-specific data available from naturalistic driving studies provide variable length car-following events in real-world driving situations, while additionally providing a wealth of information about the participating drivers. The processing of a naturalistic driving study database into finite car-following events requires significant data reduction, validation, and calibration, often using manual procedures. If the data reduction procedure can be sufficiently streamlined, the wealth of information that can be learned from this unique dataset is worth pursuing. The conclusions and recommendations relating to the thesis are presented in the context of the objectives outlined for the thesis in the first chapter.

The objectives achieved by this thesis are: (1) to assess the potential benefits of using naturalistic driving data in the analysis of car-following behavior; (2) to develop a process for extracting car-following events from a naturalistic database; (3) to generate a sample dataset through the use of a data extraction case study on the Hundred Car Study; (4) to perform sample car-following model calibration using the data reduced from the case study; and (5) to identify research topics which can leverage the unique data generated from a naturalistic driving study. The procedures outlined in this thesis are intended to serve as a guide for researchers to perform data reduction on naturalistic datasets, including the completed Hundred Car Study, the on-going Strategic Highway Research Program 2 Study, and other future studies.

5.1 Assess the Potential Benefits of Naturalistic Data for the Analysis of Car-Following Behavior

The first objective, while addressed in section 2.3 of the thesis document, is intended to be an overarching theme throughout the paper. The potential analysis of car-following models is limited by the source of the data collected. Loop detector and aerial photography/videography data collection procedures provide accurate, in situ data for real traffic patterns, but measure only short-term or instantaneous events and provide no driver-specific data. Simulator and test-track collection procedures conversely can provide a wealth of driver-specific information, but may not reflect real-world driver behavior. The driver-specific data available from naturalistic driving studies provide variable length car-following events in real-world driving situations, while additionally providing a wealth of information about the participating drivers.

5.2 Develop a Process to Extract Car-Following Data from a Naturalistic Database

This objective is actively addressed by section **3.2** of the thesis document. The data generation process includes data collection, data management, the definition of a set of goals for the data reduction, selection of data, processing of events, and validation of the data. Data collection elements within the context of a naturalistic driving study may include the on-board diagnostics of a vehicle, an accelerometer box, front and rear radar detection and tracking of objects, a video lane-tracking system, GPS location data, and video feeds in multiple directions for future data validation. Data management is a critical issue for naturalistic datasets, and till now has been done through the use of relational database architecture. The goals defined for the data reduction will influence which elements are selected form the database, and may also influence the quality and quantity of data collected. In the context of mobility analysis, the selection of data can be difficult, as some studies may need homogeneity in roadway characteristics, and other studies may want carefully controlled heterogeneity. Event processing is necessary for the study of car-following models, relating data stored in individual time-steps into leader/follower events. The data validation procedures are the most important part of the process, as any potential errors or missing information not identified in the previous steps become apparent through validation. Information gathered in a naturalistic study which is not utilized specifically in formulating the event data can prove useful in data validation for cross-checking.

5.3 Generate a Sample Dataset from the Hundred Car Study

This objective is actively addressed by section 3.3 of the thesis document. The data reduction case study of the Hundred Car Study includes: the use of GIS data to select drivers and roadway sections, the process of obtaining leading/following pairs, the selection of leading/following pairs for analysis, the smoothing of data elements, the interpolation of data elements, and methods for summarizing the data, visual, quantitative, and qualitative.

The GIS visualization of the data allows for specific roadway sections to be identified which exhibit high volumes of trip traffic. Computer processing power may have prevented previous research efforts from working with naturalistic data, but this concern should be less of a hindrance in the future. The extracted data points are merged with the radar data to identify object pairs, and segments of data without any associated radar data are removed. A complication of this process is that vehicles traveling precisely at the same speed are dropped

from radar detection, as are stopped vehicles. The most time-intensive part of the data reduction process occurs in the selection of leader/follower pairs, where it is necessary to validate the information using a visualization program which links the radar data and the video feed records. A complication specific to the Hundred Car Study was the periodic loss of both video data and radar data due to equipment malfunctions during the data collection process. The interpolation and smoothing of data elements occurred concurrently as one process, and served to remove errors arising from the data collection process. On-board diagnostics for the follower vehicle recorded velocity information every tenth of a second, but only updated the velocity information once per second, causing an instantaneous acceleration equal to ten times the actual value for one tenth of a second. This problem led to the review of the dataset for similar instantaneous jumps, which then resulted in a smoothing of the dataset. Problems with the dataset were identified by examining any velocity data recorded which resulted in an instantaneous acceleration beyond the limits of vehicle dynamics. One result of this smoothing effort was the discovery that on-board diagnostic velocity values were found to get "stuck" periodically at a given speed, from anywhere between a number of seconds of time to the remainder of the vehicle's trip. Of the fifteen drivers initially identified as commuting along the chosen route, only eight drivers resulted in viable datasets for analysis.

Additional data reduction should only be conducted after a thorough examination of the completeness of the dataset for each driver. This will front-load the effort of data reduction, but will prevent wasted effort in identifying and processing data which is subsequently found to be invalid.

5.4 Perform Sample Car-Following Model Calibration

This objective is actively addressed by **Chapter 4** of the thesis document. A summary is provided of the data used for the analysis, a justification is given for the parameter boundary conditions used, an optimization function is defined, the parameter values calibrated to minimize the objective function, and a comparison of errors is shown between the various models analyzed. A visual representation of the derived solutions is provided, and an analysis of the results is conducted. Specific analysis of the results includes the macroscopic error measures for each driver resulting from each model, a discussion of sample events at the microscopic level of analysis, and an examination of linear correlation between the various parameters within a given

model. The car-following model calibration is broken into two separate parts, that of calibrating the parameter values, and that of analyzing the results of the calibrated models.

5.4.1 Review of the Calculation of Optimized Model Parameters

All of the parameter values fixed from observing the dataset are based on maximum or minimum values of a given value, and include the desired maximum acceleration of the follower vehicle, the desired maximum deceleration of the follower vehicle, and the maximum throttle depression of the follower vehicle. The anticipated desired maximum deceleration of the leader vehicle is directly derived from the maximum deceleration of the follower vehicle, and as such is included in the direct observation calibration category. The vehicle dynamics component of the RPA model may also be considered to be observed, by examining the year, make, and model of the follower vehicle, with the parameter values are based on the specifications of the vehicle. In the case of missing data, or in the case of aggregate data, the vehicle dynamics were assumed to be equal to that of an early 2000's Toyota Corolla, the majority condition among other drivers.

The heuristic optimization program SPD-CAL determines a best fit line through macroscopic data points based on the desired free flow speed, the speed at capacity, the flow rate at capacity, and the jam density. These four values can be used directly to calculate the space headway at jam density, the perception-reaction time for use with the Gipps model, and the three Van-Aerde steady state parameter values for the RPA model. The procedure of using macroscopic data to generate the perception-reaction time for the Gipps model was validated with a plot showing how the error value changes as perception-reaction time changes.

The delta value of the Intelligent Driver Model, and all six of the parameters used in the GHR model cannot be calculated either from observation or from macroscopic measures, and must be optimized during microscopic simulation of the leader/follower pair. Each of these parameters was given upper and lower boundary conditions based on a literature review of previous solution sets. When optimizing parameter values through microscopic simulation, many of the potential combinations of variable values result in behavior outside of the feasible region, including following vehicles which pass the leader vehicle and proceed to apply a negative velocity to return to the desired spade headway. In order to prevent crashing of the optimization procedure, the velocity of the follower vehicle was taken to be the maximum of the projected simulated velocity, and 0.1 meters/second, to prevent simulated vehicles from stopping and reversing direction while on the highway facility. An additional constraint was added to the

GHR model setting the vehicle position to be equal to the maximum of the simulated position, and a position equal to the jam density distance behind the position of the leader vehicle.

A potential hurdle for working with naturalistic data is exposed when examining the various macroscopic diagrams, as steady-state data is required from both congested and uncongested regimes in order to calibrate parameter values for individual drivers. Drivers 316, 350, 358, and 367 appear to have excellent regime coverage for congested conditions, but may not have as much uncongested information as is desirable. In contrast, drivers 124, 304, 363, and 462 have very little data in the congested regime of traffic, but appear to have sufficient data collected in free-flow conditions. This difference in coverage of free-flow and congested regimes for the data proved to have an impact on model performance.

5.4.2 Review of the Analysis of Results

In the aggregate case, the Gipps model performed best with an error function value of 0.00014, while the RPA model was close behind with an error function value of 0.00017, an increase of twenty-one percent. The GHR model simulation resulted in an error function value of 0.00034, an increase of one hundred forty-three percent, and the IDM model simulation resulted in an error function value of 0.00026, an increase of eighty-six percent. In the aggregate, the Gipps and RPA models were found to produce competitive simulation results, while the GHR and IDM models were not.

The comparison of error function values for individual drivers shows that in three of the four cases where data is heavily focused on the congested regime, the Gipps model performs best, while in three of the four cases where data is heavily focused on the free-flow regime, the RPA model performs best. It may be possible to improve upon the current results by modifying the SPD-CAL optimization of macroscopic data points by adding constraints based on observed free-flow speed and jam density values.

As an additional consideration, the Gipps model relies on the calibrated maximum and minimum acceleration values as a limiting factor, whereas the RPA model relies on vehicle dynamics. In the case of modeling individual drivers in future simulation, the RPA model provides the flexibility for a driver to change vehicles without the need to subsequently recalibrate parameter values for that driver.

The success of each model at the microscopic level is somewhat consistent between events within a given driver, but varies stochastically over time, and is greatly affected by the

traffic regime in which it acts, and for many of the models is impacted by the degree to which the leading vehicle has large and/or quick shifts in velocity. The GHR model in particular is observed to not deal well with large changes in the velocity of the leading vehicle, and appears to level out above the observed value in the case of deceleration events, and below the observed value in the case of acceleration events. At high velocities, the Gipps, IDM, and RPA models all become asymptotic against the calculated desired free-flow speed, which may serve useful in the case of highway travel when not following another vehicle, but was observed to be inaccurate in comparison to the observed follower vehicle. In the time-velocity plot, the IDM model is seen to have sensitivity issues which cause the velocity to oscillate around the observed value, leading to increased space-headway in comparison to the observed follower vehicle. In one particular case, the increased space headway caused by the IDM sensitivity turned into a benefit, as the following vehicle is seen to back off from the leading vehicle towards the end of the event, which is not predicted by any of the four simulation models. This example scenario raised the issue of external environmental factors which were not recorded as part of the current data reduction process; the observed follower vehicle may be slowing in anticipation of changing lanes or taking an off-ramp, or the follower vehicle may be reacting to a downstream traffic pattern in advance of the lead vehicle's reaction to the same pattern.

Recognizing that no model is able to predict the behavior of every event due to the stochastic nature of driver's behavior, the error function value was examined against the length of each individual event. In general, the Gipps model is seen to generate lower individual event errors, the GHR model is seen to generate higher individual event errors, and the RPA model is seen to generate the most consistent level of error among all events.

In examining the correlation between parameter values, the coefficient of correlation value provides a measure of linear correlation between each of the parameters, but no significance calculations are provided herein, so the analysis of these values is limited at this time.

5.5 Identify Research Topics to Leverage Unique Naturalistic Data

This objective relates to the recommendations for future work. In the immediate future, there are five separate tasks that are proposed to advance the work included herein. These research tasks include: identifying the environmental characteristics found during each car-following event and

analyzing the effect of the environment on model parameters; expanding the types of roadways on which data is reduced for the current drivers, assessing how behavior changes between local streets, arterials, and highways, expanding the number of drivers on which data is reduced, to broaden the dataset and increase statistical significance, compare and contrast the NGSIM dataset with the current dataset, and obtain pertinent personality and psychological information for drivers and conduct analysis on how psychology plays a role in parameter values.

While advancing the current track of research already begun, it is also advisable to examine the recent work being done my others in the field. Ossen et al. have recently published research on the heterogeneity of drivers and driving behavior, which may be well suited to analyze using a naturalistic driving dataset [29], [37–41]. The work of Kim and Mahmassani investigating correlation of parameter values is especially well suited for analysis with naturalistic driving data, and it would be a valuable addition to the literature to expand the parameter correlation discussion herein [31].

As measured purely in data, this study utilized less than 1.5% of the total data available from the Hundred Car Study. It is important to develop and refine the data reduction procedures now, using the Hundred Car Study, because the Second Strategic Highway Research Program (SHRP-2) dataset is projected to contain up to forty times the volume of raw data as was produced by the Hundred Car Study, from multiple locations around the country [42]. This new dataset will have vast potential to assess driver behavior in ways not possible before.

Though naturalistic data holds the promise to answer many of the questions that remain in the field of driver behavior, working with the data requires a great deal of planning, and rigorous data validation before the trajectory information is ready for use.

APPENDIX A References

[1] S. Panwai and H. Dia, "Comparative Evaluation of Microscopic Car-Following Behavior," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 3, pp. 314-325, Sep. 2005.

[2] R. E. Chandler, R. Herman, and E. W. Montroll, "Traffic Dynamics: Studies in Car Following," *Operations Research*, vol. 6, no. 2, pp. 165–184, Jan. 1958.

[3] R. W. R. Rothery, R. Silver, R. Herman, and C. Torner, "Analysis of Experiments on Single-lane Bus Flow," *Operations Research*, vol. 12, no. 6, pp. 913-933, 1964.

[4] A. Halati, H. Lieu, and S. Walker, "CORSIM - Corridor Traffic Simulation Model," in *Traffic Congestion and Traffic Safety in the 21st Century: Challenges, Innovations, and Opportunities*, 1997, pp. 570-576.

[5] H. A. Rakha and B. Crowther, "Comparison and Calibration of FRESIM and INTEGRATION Steady-state Car-following Behavior," *Transportation Research Part A: Policy and Practice*, vol. 37, pp. 1-27, Jan. 2003.

[6] L. A. Pipes, "An Operational Analysis of Traffic Dynamics," *Journal of Applied Physics*, vol. 24, no. 3, pp. 274–281, Jul. 1953.

[7] P. G. Gipps, "A Behavioural Car-following Model for Computer Simulation," *Transportation Research Part B: Methodological*, vol. 15, no. 2, pp. 105-111, Apr. 1981.

[8] M. Treiber, A. Hennecke, and D. Helbing, "Congested Traffic States in Empirical Observations and Microscopic Simulations," *Physical Review E*, vol. 62, no. 2, pp. 1805-24, Aug. 2000.

[9] H. A. Rakha, "Validation of Van Aerde's Simplified Steady-state Car-following and Traffic Stream Model," *Transportation Letters: The International Journal of Transportation Research*, vol. 1, no. 3, pp. 227–244, 2009.

[10] M. Van Aerde and H. A. Rakha, "Multivariate Calibration of Single Regime Speed-Flow-Density Relationships," *Vehicle Navigation and Information Systems, Proceedings of the 6th International Conference on*, pp. 334-341, 1995.

[11] M. Van Aerde, "Single Regime Speed-Flow-Density Relationship for Congested and Uncongested Highways," *Transportation Research Board, 74th Annual Meeting of the*, 1995.

[12] S. Park, H. A. Rakha, R. Alfelor, C. Y. D. Yang, and D. Krechmer, "Empirical Study of Impact of Icy Roadway Surface Condition on Driver Car- Following Behavior," in *Transportation Research Board, 90th Annual Meeting of the*, 2011, vol. 982, no. 617.

[13] A. D. May, *Traffic Flow Fundamentals*. Upper Saddle River, NJ: Prentice-Hall, Inc., 1990, pp. 1-464.

[14] B. D. Greenshields, J. Bibbins, W. Channing, and H. Miller, "A Study of Traffic Capacity," in *Highway Research Board, Proceedings of the*, 1935, pp. 448-477.

[15] M. Brackstone and M. Mcdonald, "Car-following: A Historical Review," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 2, no. 4, pp. 181–196, Dec. 1999.

[16] R. Herman and R. B. Potts, "Single Lane Traffic Theory and Experiment," in *Symposium on Theory of Traffic Flow, Proceedings of the*, 1959.

[17] D. H. Hoefs, "Entwicklung einer Messmethode uber den Bewegungsablauf des Kolonnenverrkehrs," 1972.

[18] J. Treiterer and J. A. Myers, "The Hysteresis Phenomenon in Traffic Flow," in *Symposium on Transportation and Traffic Theory, Proceedings of the Sixth International*, 1974, pp. 13-38.

[19] H. Ozaki, "Reaction and anticipation in the car-following behavior," in *International Symposium on Traffic and Transportation Theory, Proceedings of the 13th*, 1993, vol. 349, p. 366.

[20] R. E. Wilson, "An analysis of Gipps's car-following model of highway traffic," *IMA journal of applied mathematics*, vol. 66, no. 5, pp. 509–537, 2001.

[21] H. A. Rakha and W. Wang, "Procedure for Calibrating Gipps Car-Following Model," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2124, no. 1, pp. 113-124, Dec. 2009.

[22] H. A. Rakha, C. C. Pecker, and H. B. B. Cybis, "Calibration Procedure for Gipps Car-Following Model," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1999, pp. 115-127, Jan. 2007.

[23] S. Demarchi, "A New Formulation for Van Aerde's Speed-flow-density Relationship (in Portuguese)," in *Pequisa e Ensino em Transportes, XVI Congresso De*, 2002.

[24] H. A. Rakha, P. Pasumarthy, and S. Adjerid, "A simplified behavioral vehicle longitudinal motion model," *Transportation Letters: The International Journal of Transportation Research*, vol. 1, no. 2, pp. 95–110, 2009.

[25] H. A. Rakha, M. Snare, and F. Dion, "Vehicle Dynamics Model for Estimating Maximum Light-Duty Vehicle Acceleration Levels," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1883, pp. 40-49, Jan. 2004.

[26] H. A. Rakha and M. Azafeh, "Calibrating Steady-State Traffic Stream and Car-Following Models Using Loop Detector Data," *Transportation Science*, vol. 44, no. 2, pp. 151-168, Dec. 2010.

[27] V. Punzo and F. Simonelli, "Analysis and comparison of microscopic traffic flow models with real traffic microscopic data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1934, no. 1, pp. 53-63, Jan. 2005.

[28] H. A. Rakha and Y. Gao, "Calibration of Steady-state Car-following Models using Macroscopic Loop Detector Data," *Symposium on The Fundamental Diagram: 75 Years*, no. 540, 2008.

[29] S. Ossen and S. P. Hoogendoorn, "Heterogeneity in Car-Following Behavior: Theory and Empirics," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 2, pp. 182-195, Apr. 2011.

[30] L. Chong, M. M. Abbas, and A. Medina, "Driver Behavior Simulation Using Agent-based Backpropagation Neural Network," *Transportation Research Board, 90th Annual Meeting of the*, 2011.

[31] J.-W. Kim and H. S. Mahmassani, "Correlated Parameters In Driving Behavior Models : Car-Following Example and Implications For Traffic Microsimulation," *Transportation Research Board, 90th Annual Meeting of the*, pp. 1-24, 2011.

[32] N. Chiabaut, L. Leclercq, and C. Buisson, "From heterogeneous drivers to macroscopic patterns in congestion," *Transportation Research Part B: Methodological*, vol. 44, no. 2, pp. 299-308, Feb. 2010.

[33] S. Mclaughlin, J. Hankey, and T. A. Dingus, "Driver Measurement: Methods and Applications," *Engineering Psychology and Cognitive Ergonomics*, vol. 5639, pp. 404-413, 2009.

[34] M. M. Abbas, B. Higgs, Z. Adam, and A. Medina, "Comparison of Car-Following Models when Calibrated to Individual Drivers using Naturalistic Data," in *Transportation Research Board, 90th Annual Meeting of the*, 2011.

[35] T. Kim, D. Lovell, and Y. Park, "Empirical Analysis of Underlying Mechanisms and Variability in Car-Following Behavior," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1999, no. 1, pp. 170-179, Jan. 2007.

[36] T. A. Dingus et al., "The 100-Car Naturalistic Driving Study Phase II – Results of the 100-Car Field Experiment," 2006.

[37] S. Ossen and S. P. Hoogendoorn, "Driver Heterogeneity in Car Following and Its Impact on Modeling Traffic Dynamics," *Transportation Research Board, 86th Annual Meeting of the*, 2007.

[38] S. Ossen and S. P. Hoogendoorn, "Validity of Trajectory-Based Calibration Approach of Car-Following Models in Presence of Measurement Errors," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2088, no. 1, pp. 117-125, Dec. 2008.

[39] S. Ossen and S. P. Hoogendoorn, "Reliability of Parameter Values Estimated Using Trajectory Observations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1, no. 2124, pp. 36-44, Dec. 2009.

[40] S. P. Hoogendoorn and R. G. Hoogendoorn, "Generic Calibration Framework for Joint Estimation of Car-Following Models by Using Microscopic Data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2188, no. 1, pp. 37-45, Dec. 2010.

[41] R. G. Hoogendoorn, S. P. Hoogendoorn, K. A. Brookhuis, and W. Daamen, "Anticipation and Hysteresis: Parameter Value Changes and Model Performance in Simple and Multi-anticipative Car-following Models," in *Transportation Research Board, 90h Annual Meeting of the*, 2011.

[42] J. Antin, S. Lee, J. Hankey, and T. Dingus, "Design of the In-Vehicle Driving Behavior and Crash Risk Study," Washington, D.C., 2011.

APPENDIX B IRB Permission Letter

Office of Research Compliance
Institutional Review Board
2000 Kraft Drive, Suite 2000 (0497)
Blacksburg, Virginia 24060
540/231-4606 Fax 540/231-0959
e-mail irb@vt.edu
Website: www.irb.vt.edu

MEMORANDUM

DATE: July 7, 2011

TO: Hesham A. Rakha, Jianhe Du, Aly Tawfik Aly Ahmed Abdel Gal, Zhiqi Sha, Ismail Zohdy, Hao Chen

FROM: Virginia Tech Institutional Review Board (FWA00000572, expires May 31, 2014)

PROTOCOL TITLE: Mid-Atlantic Universities Transportation Center Study

IRB NUMBER: 10-677

Effective July 7, 2011, the Virginia Tech IRB Chair, Dr. David M. Moore, approved the amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report promptly to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at http://www.irb.vt.edu/pages/responsibilities.htm (please review before the commencement of your research).

PROTOCOL INFORMATION:
Approved as: **Expedited, under 45 CFR 46.110 category(ies) 5**
Protocol Approval Date: **11/4/2010**
Protocol Expiration Date: **11/3/2011**
Continuing Review Due Date*: **10/20/2011**
*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

FEDERALLY FUNDED RESEARCH REQUIREMENTS:
Per federally regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals / work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

*Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.

cc: File
Department Reviewer:Julie Cook