Development of Analysis Methods Using Recent Data

DETAILS

0 pages | | PAPERBACK
ISBN 978-0-309-43525-3 | DOI 10.17226/22850

AUTHORS

The Second
STRATEGIC HIGHWAY RESEARCH PROGRAM

SHRP 2 REPORT S2-S01A-RW-1

Development of Analysis Methods Using Recent Data

GARY DAVIS AND JOHN HOURDOS
University of Minnesota

TRANSPORTATION RESEARCH BOARD

WASHINGTON, D.C.
2012
www.TRB.org

Subscriber Categories

Highways
Safety and Human Factors

The Second Strategic Highway Research Program

America's highway system is critical to meeting the mobility and economic needs of local communities, regions, and the nation. Developments in research and technology—such as advanced materials, communications technology, new data collection technologies, and human factors science—offer a new opportunity to improve the safety and reliability of this important national resource. Breakthrough resolution of significant transportation problems, however, requires concentrated resources over a short time frame. Reflecting this need, the second Strategic Highway Research Program (SHRP 2) has an intense, large-scale focus, integrates multiple fields of research and technology, and is fundamentally different from the broad, mission-oriented, discipline-based research programs that have been the mainstay of the highway research industry for half a century.

The need for SHRP 2 was identified in *TRB Special Report 260: Strategic Highway Research: Saving Lives, Reducing Congestion, Improving Quality of Life,* published in 2001 and based on a study sponsored by Congress through the Transportation Equity Act for the 21st Century (TEA-21). SHRP 2, modeled after the first Strategic Highway Research Program, is a focused, time-constrained, management-driven program designed to complement existing highway research programs. SHRP 2 focuses on applied research in four areas: Safety, to prevent or reduce the severity of highway crashes by understanding driver behavior; Renewal, to address the aging infrastructure through rapid design and construction methods that cause minimal disruptions and produce lasting facilities; Reliability, to reduce congestion through incident reduction, management, response, and mitigation; and Capacity, to integrate mobility, economic, environmental, and community needs in the planning and designing of new transportation capacity.

SHRP 2 was authorized in August 2005 as part of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU). The program is managed by the Transportation Research Board (TRB) on behalf of the National Research Council (NRC). SHRP 2 is conducted under a memorandum of understanding among the American Association of State Highway and Transportation Officials (AASHTO), the Federal Highway Administration (FHWA), and the National Academy of Sciences, parent organization of TRB and NRC. The program provides for competitive, merit-based selection of research contractors; independent research project oversight; and dissemination of research results.

SHRP 2 Report S2-S01A-RW-1

ISBN: 978-0-309-12889-6

Copyright Information

Note: SHRP 2 report numbers convey the program, focus area, project number, and publication format. Report numbers ending in "w" are published as web documents only.

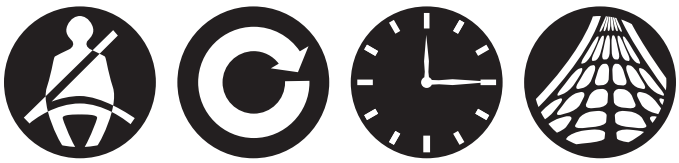
Notice

The project that is the subject of this report was a part of the second Strategic Highway Research Program, conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for their special competencies and with regard for appropriate balance. The report was reviewed by the technical committee and accepted for publication according to procedures established and overseen by the Transportation Research Board and approved by the Governing Board of the National Research Council.

The opinions and conclusions expressed or implied in this report are those of the researchers who performed the research and are not necessarily those of the Transportation Research Board, the National Research Council, or the program sponsors.

The Transportation Research Board of the National Academies, the National Research Council, and the sponsors of the second Strategic Highway Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of the report.



SHRP 2 Reports

Available by subscription and through the TRB online bookstore:
www.TRB.org/bookstore

Contact the TRB Business Office:
202-334-3213

More information about SHRP 2:
www.TRB.org/SHRP2

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

SHRP 2 STAFF

Ann M. Brach, *Director*
Stephen Andrle, *Deputy Director*
Kizzy Anderson, *Senior Program Assistant, Implementation*
James Bryant, *Senior Program Officer, Renewal*
Mark Bush, *Senior Program Officer, Renewal*
Kenneth Campbell, *Chief Program Officer, Safety*
JoAnn Coleman, *Senior Program Assistant, Capacity*
Eduardo Cusicanqui, *Finance Officer*
Walter Diewald, *Senior Program Officer, Safety*
Jerry DiMaggio, *Implementation Coordinator*
Charles Fay, *Senior Program Officer, Safety*
Carol Ford, *Senior Program Assistant, Safety*
Elizabeth Forney, *Assistant Editor*
Jo Allen Gause, *Senior Program Officer, Capacity*
Abdelmename Hedhli, *Visiting Professional*
James Hedlund, *Special Consultant, Safety Coordination*
Ralph Hessian, *Visiting Professional*
Andy Horosko, *Special Consultant, Safety Field Data Collection*
William Hyman, *Senior Program Officer, Reliability*
Linda Mason, *Communications Officer*
Michael Miller, *Senior Program Assistant, Reliability*
Gummada Murthy, *Senior Program Officer, Reliability*
David Plazak, *Senior Program Officer, Capacity*
Monica Starnes, *Senior Program Officer, Renewal*
Noreen Stevenson-Fenwick, *Senior Program Assistant, Renewal*
Charles Taylor, *Special Consultant, Renewal*
Onno Tool, *Visiting Professional*
Dean Trackman, *Managing Editor*
Pat Williams, *Administrative Assistant*
Connie Woldu, *Administrative Coordinator*
Patrick Zelinski, *Communications Specialist*

ACKNOWLEDGMENTS

This work was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials. It was conducted in the second Strategic Highway Research Program (SHRP 2), which is administered by the Transportation Research Board of the National Academies. The project was managed by Walter Diewald, Senior Program Officer for SHRP 2 Safety.

The research described in this report was performed under SHRP 2 Safety Project S01A by the Center for Transportation Studies, the Minnesota Traffic Observatory, and the Department of Civil Engineering of the University of Minnesota. Gary A. Davis, Professor of Civil Engineering, was the principal investigator. Other authors of this report are John Hourdos, Director of the Minnesota Traffic Observatory and co-principal investigator; Ted Morris, Manager of the Minnesota Traffic Observatory; and Hui Xiong and Indrajit Chatterjee, Graduate Assistants in the Department of Civil Engineering.

FOREWORD

Walter Diewald, PhD, *SHRP 2 Senior Program Officer, Safety*

A large component of the safety research undertaken in the second Strategic Highway Research Program (SHRP 2) is aimed at reducing injuries and fatalities that result from highway crashes. Through a naturalistic driving study (NDS) involving more than 3,000 volunteer drivers, SHRP 2 expects to learn more about how individual driver behavior interacts with vehicle and roadway characteristics. In anticipation of the large volume of data to be collected during the SHRP 2 NDS, several projects were undertaken to demonstrate that it is possible to use existing NDS data and data from other sources to further the understanding of the risk factors associated with road crashes. More specifically, the four projects conducted under the title Development of Analysis Methods Using Recent Data examined the statistical relationship between surrogate measures of collisions (conflicts, critical incidents, near collisions, or roadside encroachment) and actual collisions. This report presents the results of one of these projects, undertaken by the Center for Transportation Studies at the University of Minnesota.

This report documents the second phase of a two-phase project under SHRP 2 Safety Project S01A. The primary objective of this work was to establish an analytic foundation for using conflicts and near crashes as surrogate measures. The project introduced a counterfactual analytic approach suggesting that a traffic event qualifies as a crash cause under two conditions: (a) both the event and the crash occurred and (b) had the event in question not occurred, then the crash also would not have occurred. Data from site-based field studies and vehicle studies were used to extend these ideas from a trajectory model to more complicated scenarios.

The report introduces an approach to microscopic (i.e., individual event) modeling of crash-related events, where driver actions, initial speeds, and vehicle locations are treated as inputs to a physical model describing vehicle motion. This choice of modeling strategy reflects a need for such models if realistic crash processes are to be included in microscopic traffic simulation models. The simple trajectory model can be used to estimate features of crash and near-crash events—such as driver reaction times, following headways, and deceleration rates—from trajectory data produced from a site-based field study. Given sufficiently large samples of crash and near-crash events, this method can be used to compile distributions for these inputs for use in traffic simulation models. Finally, the report illustrates how a trajectory model, together with estimates of input variables, can quantify the degree to which a non-crash event could have been a crash event.

The report describes how these ideas were extended to more complicated scenarios by using data from both vehicle- and site-based field studies, including data obtained from the 100-car vehicle-based field study, data from site-based video on Interstate 94 from the Minnesota Traffic Observatory, and site-based radar data from the Cooperative Intersection Collision Avoidance Systems (CICAS) intersection in North Carolina.

CONTENTS

Executive Summary

This report describes work done during Phase 2 of SHRP 2 Safety Project S01A, Development of Analysis Methods Using Recent Data. In the report submitted at the end of Phase 1, three research problems were identified on which progress was needed. The first of these was identification of an appropriate class of structural models describing how crash and near-crash events developed, together with analytic tools for fitting these models to data expected from the vehicle-based and site-based field studies. The second problem involved counterfactual screening of supposed near-crash events to determine their similarity to crashes. The third problem involved developing plausible models of how drivers select evasive actions as functions of the situations in which they find themselves. Solutions to the second and third research problems are contingent on solution of the first, so the bulk of the project team's effort during Phase 2 has been devoted to structural modeling of crash and near-crash events, using data from the 100-car vehicle-based field study, from site-based video data collected by the Minnesota Traffic Observatory (MTO), and from site-based Doppler shift data obtained from the Cooperative Intersection Collision Avoidance Systems (CICAS) project. The team's focus has been on crashes and near crashes involving more than one vehicle, of the sort that occurs at intersections.

Background

Chapter 1 of this report outlines the context within which this research has taken place. In the United States, there are two substantial national efforts related to road traffic safety. The first is the development of the first edition of a *Highway Safety Manual* (*HSM*). The second is the design and execution of the SHRP 2 Safety field studies. The project team points out that it may be possible for the data collected in the SHRP 2 field studies to support the development of microscopic (i.e., individual event) crash models, which can be incorporated into traffic simulation models to supplement or replace the macroscopic regression methods used in the first edition of the *HSM.* For this to occur, though, it will be necessary to identify and fit plausible microscopic models of crash-related events using the SHRP 2 field data. The team illustrates how this might be accomplished using a simple braking-to-stop model applied to trajectory data extracted from site-based video and then illustrates how once a fitted model is at hand, it is possible to quantify the expected number of crashes in a set of noncrash events.

Chapter 2 takes up the problem of extending these ideas to more complicated situations, and the project team proposes a modeling strategy where driver behavior is treated as a piecewise constant sequence of acceleration changes. Given such an acceleration history and initial values for a vehicle's location and speed, it is logical to move toward a system of ordinary differential equations to get predicted histories of the vehicle's speed and position. Fitting such a model then involves identifying the appropriate break points in the acceleration profile, the corresponding acceleration levels, and the initial conditions that best fit observed trajectory data. The team

illustrates model identification and estimation using speedometer, radar range, and radar range-rate data for a near-crash event from the 100-car vehicle-based study. The team also illustrates a what-if counterfactual analysis where the final deceleration of the following vehicle is varied over a range of values, for each of which, other things being equal, the probability that a collision would have resulted is computed.

Findings

Chapter 3 describes the project team's work with data from the 100-car study. Data were obtained for 33 crash and near-crash events, each consisting of approximately 30 s of forward video and 30 s of instrumentation measures at 10 Hz. The instrumentation measures included speeds from the instrumented vehicle's speedometer, range and range rate for objects ahead of the instrumented vehicle from its forward radar, accelerometer measures, GPS positions, heading, yaw measures, and indicators of brake, accelerator, and turn signal use. After reviewing the data, seven of the 33 events were determined to have data sufficient to attempt modeling, and model identification, estimation, and goodness-of-fit evaluation using speedometer, range, and range-rate data are described. Six of the events involved a leading vehicle and a following vehicle decelerating in the same lane of traffic, and it was possible to identify plausible acceleration profiles for the leading and the following drivers. From these measures, it was then possible to estimate the following driver's reaction time, together with measures describing the situation at the time his or her reaction phase began. The team also reconstructed one event involving a swerving maneuver by the following driver, where the forward radar data were limited.

Chapter 4 describes work using site-based video data from the University of Minnesota's Beholder system. This is a set of video cameras, computers, and wireless communication equipment positioned to overlook a section of Interstate 94 in downtown Minneapolis. Vehicle positions, extracted from video recordings and rectified for camera position effects, provided the raw data for the analyses. Four rear-end crash events in July–October 2008 were analyzed, with two of these events also including trajectories for vehicles not involved in the crash. In each case, it was possible to identify plausible acceleration profiles for each of the involved drivers, which in turn provided information on reaction times and conditions at the start of the reaction phase.

Chapter 5 describes pilot work using site-based Doppler shift data collected by the University of Minnesota's Intelligent Vehicles Laboratory as part of the CICAS project. This configuration consists of a coordinated set of radar units collecting information on the positions and speeds of major approach vehicles approaching a two-way-stop controlled intersection, and LIDAR units collecting information on the positions of minor approach vehicles. Since the instrumentation configuration was designed to support a prototype driver-information system and not to collect data on vehicle trajectories, data acquisition and preparation were not as straightforward as with the other data sets; but with some data mining and postprocessing, it was possible to identify one event suggesting braking on the part of a major approach vehicle in response to minor approach crossing. It was then possible to identify and fit plausible acceleration profiles for both drivers.

Chapter 6 presents the conclusions and recommendations for further research.

Conclusions

1. At least for situations where direction of travel is roughly constant, trajectory-based reconstruction of crash-related events, where trajectory data are used to fit parsimonious models of driver behavior, is feasible using both vehicle-based and site-based data.
2. It is possible to extend the methods of counterfactual analysis to more complicated structural models involving differential equations.
3. At least for rear-ending events, there is some limited evidence that the distributions of evasive actions for crashes and near crashes share some overlap, so that it should be possible to find near-crash events that are similar in other respects to crashes.

4. Although the CICAS system as currently configured was not designed to collect and process crash and near-crash trajectory data, with technical modifications it could support site-based field research, at least at lower-volume intersections.
5. The usefulness of the data produced by the SHRP 2 vehicle-based field study will be strongly dependent on the ability to calibrate and maintain the data-collection systems.

Recommendations

1. The modeling methods presented in this report should be extended to handle two-directional trajectories.
2. When attempting to include crash events in a microscopic traffic simulation, plausible models that close the feedback loop between existing conditions and driver actions will be necessary. This issue should be pursued using the data from the SHRP 2 field studies.
3. The trajectory modeling methods described in the report should be enhanced to allow for possible serial correlation in trajectory data.
4. Compiling data on gap-selection and other intersection-related events will require a data setup different from the SHRP 2 vehicle-based field study, so the SHRP 2 vehicle-based field study should be complemented with site-based research.
5. Clear descriptions of data collection and processing and associated metadata should be required in future major data-collection efforts.

CHAPTER 1

Background and Project Objectives

Background

In its essence, rational planning involves using predicted consequences to guide selection from among a set of possible actions. In road safety engineering, this requires being able to predict the frequency and severity of crashes that are expected to result from a given design or operational configuration. Hauer (*1*) has argued persuasively for developing a scientifically justifiable methodology for making these predictions, and the last 10 years have seen two major initiatives in the United States related to this issue. One is the development of the *Highway Safety Manual* (*HSM*), the first edition of which was released in 2010 by the American Association of State Highway and Transportation Officials (AASHTO). The other is the safety focus of the SHRP 2 research program.

Highway Safety Manual

The goal of the *HSM* is to provide highway professionals with tools for explicitly considering the safety impacts of engineering actions. The dominant methodology used to develop the *HSM* is statistical analysis of crash-frequency data, where the basic units of analysis are either sections of highway or intersections and where the dependent variables are crash frequencies observed over one or more years. In some cases, these frequencies can be broken down by crash type or severity. Generalized linear models are used to describe baseline associations between crash frequency and observable road features, while the effects of changes from these baseline conditions are captured through empirically determined crash-modification factors. The effect of an intervention is then predicted by first using the base model to predict crash frequency under the prevailing conditions, and then multiplying this expected frequency by a crash-modification factor that reflects the effect of the change of interest. Ideally, the crash-modification factor was estimated from a well-conducted before/after study that controlled for selection bias effects. The strong reliance of the *HSM*'s first edition on this type of statistical modeling is in large part the result of historical developments, where the integrated crash and roadway databases maintained by the Highway Safety Information System (HSIS) and several individual states had attained a useful degree of completeness and the development of empirical Bayes and hierarchical Bayes methods during the 1980s and 1990s brought the supporting statistical tools to a useful degree of maturity.

Regression analyses of aggregated observational data have well-understood limits to their ability to discover and describe underlying causal processes (*2*). In 2006, a daylong workshop was held during the annual meeting of the Transportation Research Board (TRB), which focused on elucidating these limits and discussing alternative methods. At that workshop, Bonneson and Lord (*3*) pointed out an interesting analogy with the development of the *Highway Capacity Manual,* where first-generation regression models for predicting traffic signal delay using naïve specifications of independent variables were later replaced by regression models where the form of the independent variables was justified theoretically. These in turn were replaced by models where the functional forms relating traffic flow, capacity, and signal timing to delay were justified theoretically. The macroscopic methods developed for the *Highway Capacity Manual* were later supplemented and in some instances replaced by microscopic traffic simulation models. Hope was expressed that a similar evolution might occur for the *HSM,* with first-edition regression models being supplemented or replaced by structural models that explicitly describe the mechanisms' underlying crash occurrence. A better understanding of crash mechanisms could also support the use of microscopic traffic simulation models to predict the safety consequences of engineering decisions, similar to how microscopic models are now used to predict operational consequences. As researchers working in this area are acutely aware, however, a major obstacle to progress is the lack of good microscopic data regarding crash occurrence and driver behavior more generally.

SHRP 2 Safety Program

As originally conceived, the safety focus area of the SHRP 2 research program comprised two major field studies: a vehicle-based study involving "volunteer drivers and a sophisticated instrumentation package installed in the volunteers' vehicles" and a site-based study involving video recording of vehicle movements at specific locations. These studies were "intended to support a comprehensive safety assessment of how driver behavior and performance interact with roadway, environmental, vehicular, and human factors and the influence of these factors and their interactions on collision risk" (*4*). At present, the vehicle-based study is going forward while the site-based study has been limited to preliminary design, with further work dependent on the availability of additional funding. The SHRP 2 S01 projects develop and apply analytic methods relevant to these field studies by identifying salient research questions and then attempting to answer them using existing data of the type expected from the field studies. The S01 request for proposal explicitly identified as important "application of crash surrogates" and "the formulation of analytic methods to quantify the relationship of human factors, driver behavior, vehicle, roadway, and environmental factors to collision risk." Special attention was directed to roadway departure and intersection crashes.

There is little doubt that the SHRP 2 vehicle-based study should produce a rich and unprecedented source of information concerning driver behavior in normal and crash situations and that this should support development and evaluation of vehicle-based safety technologies. The study should also support traditional statistical investigations seeking to identify associations between roadway conditions and crash occurrence. In the project team's view, the SHRP 2 field studies could also provide data supporting the development and application of microscopic crash models, similar to how existing crash record and roadway databases support the development and application of regression-based approaches. For this to occur, however, analytic tools are needed that can fit and test microscopic models using field study data and can extract the sort of measurements needed to quantify driver behavior in crash-related conditions. The S01 project thus has two interrelated objectives. The first is to develop analytic tools and demonstrate how these can be used to conduct structural model development, using the sort of data expected from the SHRP 2 field studies. The second is to develop a rigorous method for characterizing near crashes so that observations of near crashes might serve as useful surrogates for actual crashes. The approach taken by the project team can be called trajectory-based reconstruction of crash-related events. That is, time history data of vehicle positions, speeds, or both are used to estimate values for variables describing drivers' actions and characterizing the conditions leading to crash-related events. The focus of the present project is on crashes involving more than one vehicle, resulting from car-following or gap-selection behavior of the types often occurring at intersections.

Structural Modeling of Crash-Related Events

The starting point is Pearl's (*5*) notion of a causal model, which in the abstract consists of a set of exogenous variables, a set of endogenous variables, and, for each endogenous variable, a structural equation describing how the variable responds to changes in other model variables. A causal model can be represented qualitatively using a directed graph, with the nodes of the graph representing variables and directed arrows indicating direct causal dependencies.

Figure 1.1 displays a simple graphical model for a generic crash. The node *u,* possibly vector-valued, denotes variables describing background conditions. The node *x* denotes the variable describing an evasive action, and the node *y* denotes the crash condition. The crash condition *y* is assumed to be a *deterministic* function of *u* and *x,* such that

$y(u,x) = 0$, if the values for u and x do not produce a crash

$y(u,x) = 1$, if u and x produce a crash

To make these ideas more concrete, consider a simple two-vehicle, rear-ending collision model (*6, 7*). Such an event might be observed in the field, or it could arise within a microscopic traffic simulation. The initial speed and braking deceleration of the leading vehicle are denoted by v_1 and a_1, respectively; the initial speed and braking deceleration of the following vehicle by v_2 and a_2, respectively; and the following driver's headway and reaction time by h_2 and r_2, respectively. A collision occurs when the stopping distance available to the following driver is less than that needed to stop without colliding with the lead vehicle. Using simple physics, this can be expressed as

$$v_2 r_2 - \frac{v_2^2}{2a_2} > h_2 v_2 - \frac{v_1^2}{2a_1} \tag{1.1}$$

If the following driver's deceleration is taken as the avoidance action, then for the rear-end collision, the variables

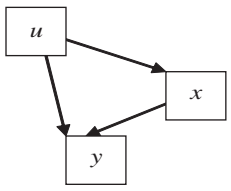


Figure 1.1. Graphical model of a crash-related event.

(v_1, a_1, v_2, r_2, h_2) are components of u, a_2 is the evasive action, and the collision function is

$$y(u,x) = \begin{cases} 0, \text{ if } v_2 r_2 - \dfrac{v_2^2}{2a_2} \le h_2 v_2 - \dfrac{v_1^2}{2a_1} \\ 1, \text{ if } v_2 r_2 - \dfrac{v_2^2}{2a_2} > h_2 v_2 - \dfrac{v_1^2}{2a_1} \end{cases} \quad (1.2)$$

In this treatment, an event consists of a specification of values for each of the model variables. This specification, plus Structural Equation 1.2, is sufficient to determine whether that event leads to a crash. As they are currently implemented, microscopic simulation models are restricted to combinations of values that do not lead to crashes, and an open research question involves determining realistic relative frequencies for those combinations that do.

To help illustrate the usefulness of this approach, an example originally presented by Davis and Swenson (*7*) is used. Figure 1.2 shows trajectories for a platoon of seven vehicles successively braking to stops while traveling in the same lane of a freeway. The leftmost vehicle was the first vehicle in the platoon, the rightmost vehicle was the last, and a collision occurred between the two rightmost vehicles. These trajectories were constructed from a video recording of the event by first digitizing each vehicle's position on successive video frames and then using standard photogrammetry methods to convert from image coordinates to ground coordinates. Applying the simple braking model described, each of these trajectories can be described by the physical model

$$\begin{aligned} & v_k t, \quad t \le t0_k \\ & y_k(t) = v_k t0_k + a_k (t - t0_k)^2 / 2, \quad t0_k < t \le t0_k + v_k/(-a_k) \\ & v_k t0_k - v_k^2 / 2a_k, \quad t > t0_k + v_k/(-a_k) \end{aligned} \quad (1.3)$$

where $y_k(t)$ denotes the (one-dimensional) position of Vehicle k at time t, v_k denotes the initial speed of Driver k, a_k denotes his or her braking acceleration, and $t0_k$ denotes the time at which braking began. This model can be connected to the rear-end collision model described by noting that the reaction time of Driver k is simply

$$r_k = t0_k - t0_{k-1} \quad (1.4)$$

while the initial following headway between Vehicles k and k–1 when Driver k–1 began braking is

$$h_k = \left(y_k(t0_{k-1}) - y_{k-1}(t0_{k-1})\right)/v_k \quad (1.5)$$

At least for simple event types like this, it is then possible to generate estimates of initial speeds, braking rates, and times of braking initiation by fitting Equation 1.3 to observed trajectories. Davis and Swenson (*7*) describe how Bayes estimates could be computed from trajectory data of the type displayed in Figure 1.2. Table 1.1 summarizes the resulting estimates for these data.

In principle, then, trajectory-based reconstruction can be used to fit structural models and estimate important features of crash-related events. One of the project team's objectives is to extend these methods to handle more complicated situations and to exploit the type of data expected from the SHRP 2 in-vehicle field study.

Crash Surrogates

The second objective of this project is to develop a quantitative method for characterizing and identifying events that can serve as useful crash surrogates. This is because crashes, especially severe crashes, tend to be rare, so that if one could identify near crashes or other surrogate events that carry information about

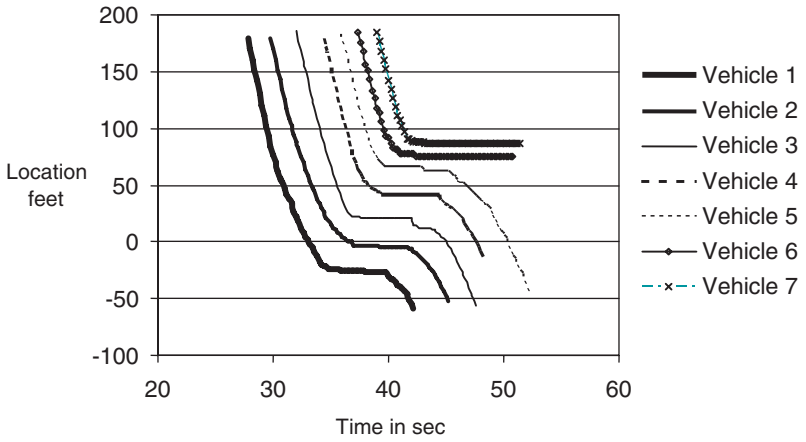


Figure 1.2. Trajectories of seven vehicles braking successively to stops.

Table 1.1. Estimates of Speed, Headway, Reaction Time, Braking Deceleration, and Minimal Successful Deceleration for Seven Vehicles

Vehicle (k)[a]	v_k (ft/s)[b]	h_k (s)[c]	r_k (s)[d]	a_k (ft/s^2)[e]	a_{ko} (ft/s^2)[f]
1	50.0	—	—	–6.8	—
2	46.7	1.69	1.91	–6.5	–6.2
3	41.8	2.00	4.21	–12.6	–11.4
4	42.3	1.87	1.86	–14.2	–12.8
5	39.3	1.21	1.44	–16.0	–14.4
6	42.3	1.17	1.07	–17.3	–17.1
7	41.7	1.24	1.65	–20.3	–24.8

[a] Number of vehicles
[b] Speed of the k vehicle
[c] Headway of the k vehicle
[d] Reaction time of the k vehicle
[e] Braking acceleration
[f] Minimum deceleration needed to avoid collision

how crashes occur, the value of both in-vehicle and site-based studies would increase.

Roughly speaking there are two ways that near-crash events might be used as crash surrogates. On the one hand, one might carry out an intensive study of how individual near crashes occurred, with the goal of identifying causal factors for each event. This would be similar to using investigation and reconstruction of actual crashes to gain insight into how and why crashes occur. On the other hand, one might use counts of near crashes as a dependent variable and then look to see how these are associated with roadway or driver characteristics. This would be similar to carrying out a statistical study of crash frequency. In either case, though, the starting point is a set of noncrash events and the need to determine the extent to which each could be regarded as a near crash.

Returning to the literature, it is possible to find two related but different approaches to defining crash surrogates. One is the definition of conflict as put forward by the *International Calibration Study of Traffic Conflict Techniques* (*ICSTCT*): "A traffic conflict is an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged" (*8*). However, it turned out that when attempting to find empirical associations between conflict and crash frequencies, it was helpful if conflicts could be graded as to their seriousness or severity. This distinction is included in the definition of near crash used in the 100-car study, which can be regarded as a pilot for the SHRP 2 vehicle-based field study: "Any circumstance that requires a rapid, evasive maneuver by the subject vehicle, or any other vehicle, pedestrian, cyclist, or animal to avoid a crash. A rapid, evasive maneuver is defined as a steering, braking, accelerating, or any combination of control inputs that approaches the limit of the vehicle capabilities" (*9*).

Both these definitions assume a counterfactual definition of the surrogate event, where a crash would have occurred had an evasive action not been performed. The 100-car study definition places an additional condition, however, that the magnitude of the evasive action should be in some sense extreme. To see the value of this additional condition, it is helpful to return to Table 1.1. The table gives estimates of initial speeds, headways, reaction times, and the actual and minimum successful decelerations estimated from the vehicle trajectories. First, consider the interaction between Vehicles 1 and 2. The minimum deceleration by Vehicle 2 needed to avoid collision was about –6.2 ft/s^2, while the actual deceleration was about –6.5 ft/s^2. Had the deceleration been slightly less, other things being equal, a crash would have occurred. So, arguably, this event satisfies the *ICSTCT* definition of a conflict. Most would agree, though, that it does not satisfy the 100-car study condition that the evasive action be extreme. For the interaction between Vehicles 5 and 6, the minimum successful deceleration was about –17.1 ft/s^2 and the actual deceleration was about –17.3 ft/s^2. This event also satisfies the *ICSTCT* definition of conflict but comes closer to satisfying the 100-car study condition as well. The project team uses causal models to construct a quantitative measure that captures this difference.

To start, Figure 1.3 shows the probability of collision as a function of the following vehicle's braking deceleration for both vehicle pairs. The analysis is probabilistic because the values of the important event variables are not known with certainty but rather only up to their posterior distributions given the trajectory data. Figure 1.3 was prepared by setting the follower's deceleration to each of a set of target values and then using Monte Carlo simulation to compute the probability of a crash.

Figure 1.3 shows that for Vehicles 1 and 2, decelerations greater than about –7 ft/s^2 prevent a crash with high probability, while for Vehicles 5 and 6, decelerations greater than about –20 ft/s^2 are needed for a similar degree of certainty. The latter seems qualitatively close to the definition of near crash used in the 100-car study, but to quantify this degree of closeness, it is necessary to specify what is meant by an evasive action that approaches the limit of the vehicle's capabilities. One, but not necessarily the only, way to do this is to apply the results of the emergency braking study carried out by Fambro et al. (*10*), where the distribution of braking decelerations used by drivers confronted with a surprise braking situation had a mean of about –20.3 ft/s^2 and a standard deviation of about 2.6 ft/s^2.

Figure 1.4 adds to Figure 1.3 a normal distribution with the given mean and standard deviation. Roughly speaking, the degree to which a conflict qualifies as a near crash is

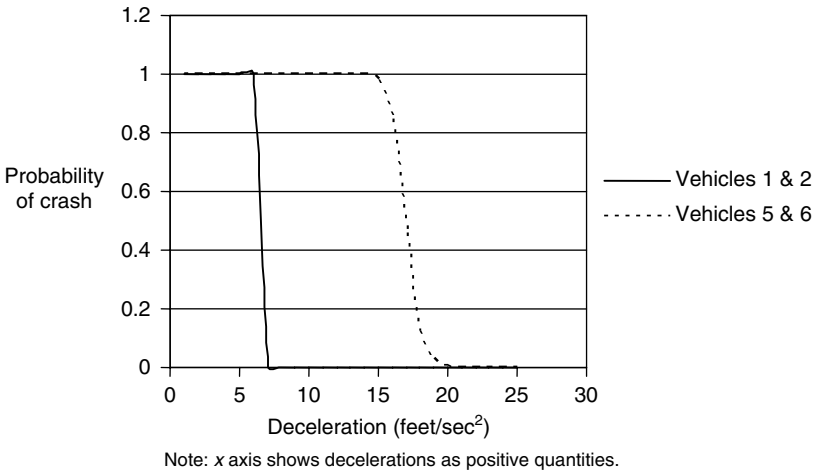


Figure 1.3. Crash probability as a function of counterfactual values for following vehicle's deceleration.

determined by how much of this extreme braking distribution lies to the left of the crash probability curve. More formally, the probability that a conflict could have been a crash is found by integrating the crash probability curve with respect to the extreme braking distribution. Although analytically intractable, this computation is readily carried out using Monte Carlo methods. Table 1.2 gives these results for each of the noncolliding vehicle pairs from Table 1.1, and the sum of these probabilities can be taken as the expected number of crashes in this set of conflict events had the evading drivers taken their decelerations from the given distribution.

Summary

To summarize, the project team has introduced an approach to microscopic modeling of crash-related events, where driver actions, together with initial speeds and vehicle locations, are treated as inputs to a physical model describing vehicle motion. The team's choice of this modeling strategy is rooted in the fact that models of this sort are needed if realistic crash processes are to be included in microscopic traffic simulation models. The team has illustrated how a simple version of a trajectory model can be used to estimate features of crash and near-crash events, such as driver reaction times, following headways, and deceleration rates, from trajectory data of the sort produced from a site-based field study. Given sufficiently large samples of crash and near-crash events, this method could be used to compile distributions for these inputs, which could, in turn, be used in traffic simulation models. Finally, the team has also illustrated how a trajectory model, together with estimates of input variables, can quantify the degree to which a noncrash event could have been a crash. One potential application of this technique would be to process a set of noncrash events produced either

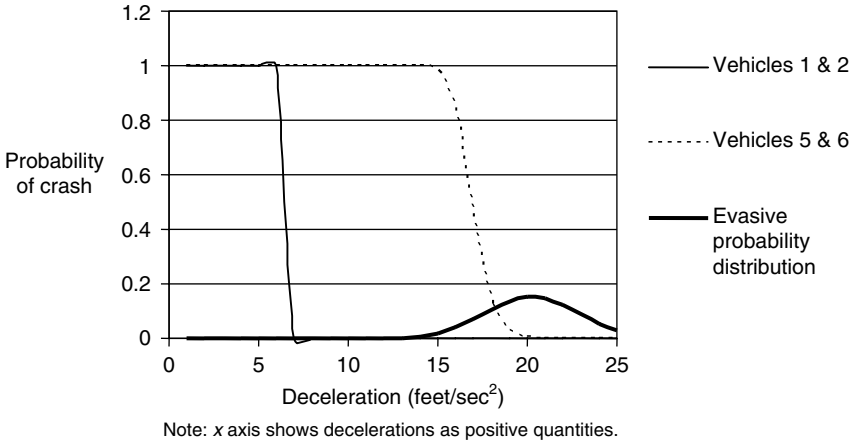


Figure 1.4. Crash probability versus following vehicle deceleration and probability density function for emergency decelerations.

Table 1.2. Crash Probabilities for Each Vehicle Pair, Obtained Using the Emergency Braking Distribution

Lead Vehicle	Following Vehicle	*P* (crash)[a]
1	2	0
2	3	0
3	4	0
4	5	.004
5	6	.138
Sum		**.142**

[a] Probability of crash

by a single driver or at a single location to produce an expected number of crashes in this set. This expected number of crashes could then serve as a dependent variable in a study of driver or site features believed to be related to safety.

The remainder of this report describes the effort to extend these ideas to more complicated scenarios using data produced by both vehicle-based and site-based field studies. Chapter 2 outlines the analytic procedures and tools developed for this project and illustrates their use. Chapter 3 presents analyses of data obtained from the 100-car vehicle-based field study. Chapter 4 describes analyses of data from site-based video on Interstate 94, while Chapter 5 describes work with site-based radar data from the Cooperative Intersection Collision Avoidance Systems (CICAS) intersection in North Carolina. Chapter 6 presents the study's conclusions and recommendations.

References

1. Hauer, E. The Engineering of Safety and the Safety of Engineering. In *Challenging the Old Order: Towards New Directions in Traffic Safety Theory* (J. Rothe, ed.), Transaction Publishers, London, 1990, pp. 29–71.
2. Freedman, D. From Association to Causation via Regression. In *Causality in Crisis?* (V. McKim and S. Turner, eds.), University of Notre Dame Press, Notre Dame, Ind., 1997, pp. 113–162.
3. Bonneson, J., and D. Lord. Theory, Explanation, and Prediction in Road Safety. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.
4. SHRP 2 Request for Proposals: Development of Analysis Methods Using Recent Data. Transportation Research Board of the National Academies, Washington, D.C., September 11, 2006.
5. Pearl, J. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, Cambridge, United Kingdom, 2000.
6. Brill, E. A Car-Following Model Relating Reaction Times and Temporal Headways to Accident Frequency. *Transportation Science,* Vol. 6, No. 4, 1972, pp. 343–353.
7. Davis, G., and T. Swenson. Collective Responsibility for Freeway Rear-Ending Accidents? An Application of Probabilistic Causal Models. *Accident Analysis and Prevention,* Vol. 38, 2006, pp. 728–736.
8. Guttinger, V. Conflict Observation in Theory and Practice. In *International Calibration Study of Traffic Conflict Techniques* (E. Asmussen, ed.), Springer-Verlag, Berlin, 1984, pp. 17–24.
9. Dingus, T., S. Klauer, V. Neale, A. Petersen, S. Lee, J. Sudweeks, M. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. Doerzaph, J. Jermeland, and R. Knipling. *The 100-Car Naturalistic Driving Study: Phase II—Results of the 100-Car Field Experiment.* Report DOT HS 810 593. National Highway Traffic Safety Administration, U.S. Department of Transportation, 2006.
10. Fambro, D., K. Fitzpatrick, and R. Koppa. *NCHRP Report 400: Determination of Stopping Sight Distances.* Transportation Research Board, National Research Council, Washington, D.C., 1997.

CHAPTER 2

Overview of Analytic Method

State-Space Model

Chapter 1 illustrated the trajectory-based approach with simple braking-to-stop models estimated from data extracted from site-based video. An important objective of this project was to develop a common analytic framework that could be applied to data from either site- or vehicle-based sensor configurations and could accommodate events more complicated than constant speed followed by simple braking-to-stop. Ultimately, the goal is to model vehicle trajectories in two dimensions, where both braking and steering could serve as evasive actions. This report, however, focuses on the simpler problem of modeling vehicle motion in one direction.

The basic idea is to model driver behavior as a piecewise constant series of acceleration changes, which are then treated as inputs into a dynamic trajectory model. The vehicle's state at a given time is its location and velocity, and the trajectory model takes the acceleration input sequence and numerically integrates the associated differential or difference equations to produce time histories of vehicle locations and speeds. For discrete-time data, the trajectory model can be conveniently represented using the generic linear state-space form

$$\begin{aligned} x(t+1) &= Ax(t) + Ba(t) \\ y(t) &= Cx(t) \end{aligned} \tag{2.1}$$

where $x(t)$ is a vector of state variables (position and velocity), $a(t)$ is a vector of input variables (accelerations), and $y(t)$ is the vector of observed variables. A, B, and C stand for matrices of coefficients.

The nature of A, B, C, $x(t)$, and $y(t)$ will vary depending on the class of events being modeled and the sort of data available. For two vehicles following on a straight road, the simplest trajectory model consists of two state variables for each vehicle, location and speed, with linear acceleration values as inputs. That is, if Δ denotes the basic time interval of the data, then the deterministic progression for a leading and following vehicle can be captured by the linear equation

$$\begin{bmatrix} x1(t+1) \\ x2(t+1) \\ v1(t+1) \\ v2(t+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x1(t) \\ x2(t) \\ v1(t) \\ v2(t) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \Delta & 0 \\ 0 & \Delta \end{bmatrix} \begin{bmatrix} a1(t) \\ a2(t) \end{bmatrix} \tag{2.2}$$

Here, $x1(t)$ and $x2(t)$ give the locations of the leading and following vehicles at time t, $v1(t)$ and $v2(t)$ are the corresponding speeds, and $a1(t)$ and $a2(t)$ are the accelerations. For motion in two directions, a similar structure can be used but with state and input variables for each direction. Given initial values for the state variables and the time history for the inputs, the trajectories of both vehicles can be replicated.

Since Equation 2.2 describes vehicle motion irrespective of the data-collection scenario, the primary difference between the site- and vehicle-based scenarios will be the observation equation. For vehicle trajectories extracted from video, the observations consist of measurements of position for each vehicle, leading to an observation equation of the form

$$\begin{bmatrix} y1(t) \\ y2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x1(t) \\ x2(t) \\ v1(t) \\ v2(t) \end{bmatrix} \tag{2.3}$$

A standard calculation shows that this system is observable, and the two-vehicle system decomposes into two one-vehicle systems.

For the vehicle-based data from the 100-car study, the primary observations are the speed of the following vehicle obtained from its speedometer, and the range and range rate

for the lead vehicle relative to the follower, obtained from the forward radar. This leads to an observation equation of the form

$$\begin{bmatrix} y1(t) \\ y2(t) \\ y3(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} x1(t) \\ x2(t) \\ v1(t) \\ v2(t) \end{bmatrix} \tag{2.4}$$

This system is not observable. For instance, it is not possible to obtain estimates of absolute position for each vehicle. It also does not decompose into two independent subsystems, so that estimation and inference will generally require working with both vehicles.

For site-based data from the CICAS radar system, observations consist of both position and speed for individual vehicles, leading to an observation equation of the form

$$\begin{bmatrix} y1(t) \\ y2(t) \\ y3(t) \\ y4(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x1(t) \\ x2(t) \\ v1(t) \\ v2(t) \end{bmatrix} \tag{2.5}$$

This also decomposes into two separate subsystems, one for each vehicle.

Given estimates of a driver's initial speed, the times at which he or she changed acceleration, and the corresponding accelerations, the differential equations can be solved to give predicted time histories of that vehicle's position and speed and predicted values for the observations. Fitting a trajectory model then involves searching plausible combinations of values for these input quantities to find those that best account for the data. The counterfactual simulation needed to assess the degree to which a near crash might have been a crash, in turn, involves using probability distributions characterizing the residual uncertainty in the model parameters as input to a Monte Carlo simulation, where the state equation is integrated using random draws from this distribution and the occurrence or nonoccurrence of a collision recorded. In the early stages of this study, several different approaches to implementing these steps were experimented with, including modeling the differential equation model using response-surface approximations (*1*), nonlinear least-squares estimation using asymptotic normal approximations to characterize posterior uncertainty (*2*), and Bayesian analysis using Markov Chain Monte Carlo (MCMC) simulation. The WinBUGS software (*3*) can be used to implement MCMC estimation for a variety of relatively complicated models, but the project team's initial attempts to implement this by directly coding the differential equations in WinBUGS led to estimation runs with excessively long time demands. This problem was circumvented by using the WinBUGS differential equation interface. This provides compiled procedures that can be included in a WinBUGS model specification, which numerically solves ordinary differential equations using Runge-Kutta methods. As is the standard practice in working the MCMC estimation, exploratory analyses were first conducted using frequentist methods, in this case nonlinear least-squares, implemented using either MATLAB (*4*) or R (*5*). This was done to understand the complexity of the acceleration model suggested by a given data set and to obtain reasonable starting values for the MCMC simulation. Bayes estimates for model parameters were then computed using WinBUGS, and counterfactual simulation was carried out using the MCMC sample of the posterior distribution for these parameters.

Illustrative Example

The general approach is illustrated here using case 104119 from the 100-car study. This event was a potential near crash that involved a lead vehicle and a following vehicle successively braking to a stop, with the follower stopping short of collision. This was a vehicle-based study, with the following vehicle being the instrumented vehicle. The data employed in this analysis were the speedometer-measured speeds for the instrumented vehicle and range and range rate for the leading vehicle, obtained from the follower's forward radar.

Step 1: Graphical Inspection of Data

Figure 2.1 shows the time history of the speedometer-measured speeds of the following vehicle converted to units of ft/s. The piecewise constant nature to this relation was characteristic of all speedometer data obtained for 100-car study cases. Inspection of Figure 2.1 suggested a two-phase model where the following driver was initially traveling at about 21 ft/s and accelerating until about 11 s from the start of the data series. He or she then decelerated at a roughly constant rate until coming to a stop.

Figure 2.2 shows the time history of the range and range-rate data provided by the instrumented vehicle's forward radar. The discontinuities in the time-series result from periods of missing data, when the forward radar apparently lost the leading vehicle as a target. To get an initial sense of the leading driver's actions, an approximate speed profile for the leading vehicle was often used, obtained by adding the range rate to the following vehicle's speedometer data. A plot of these approximate speeds is shown in Figure 2.3. This

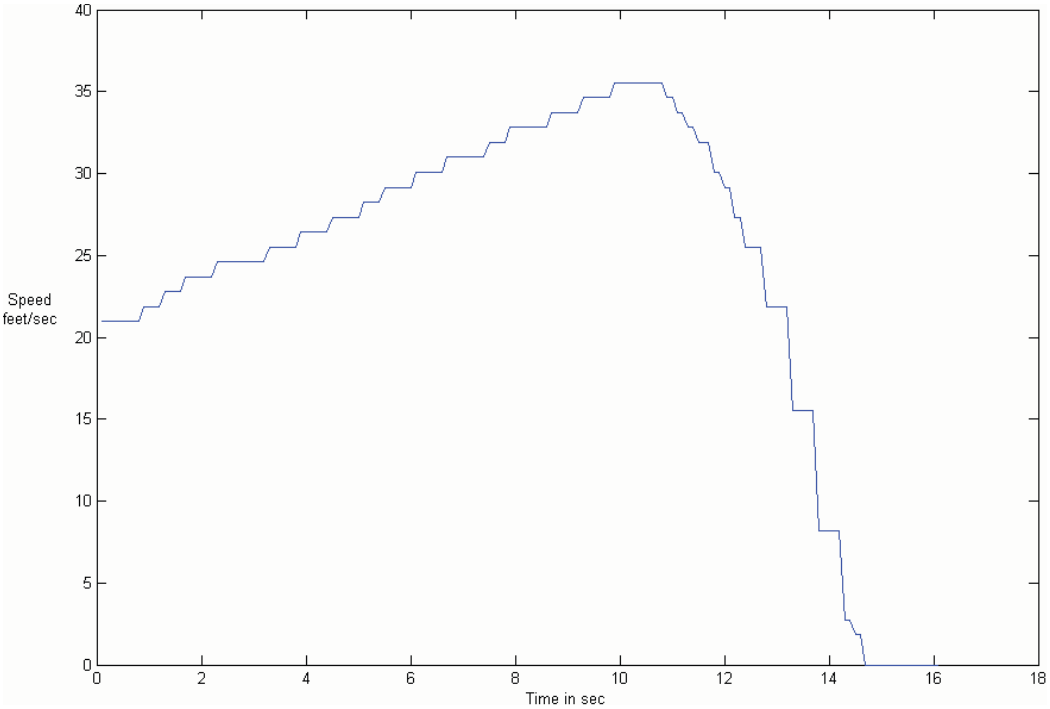
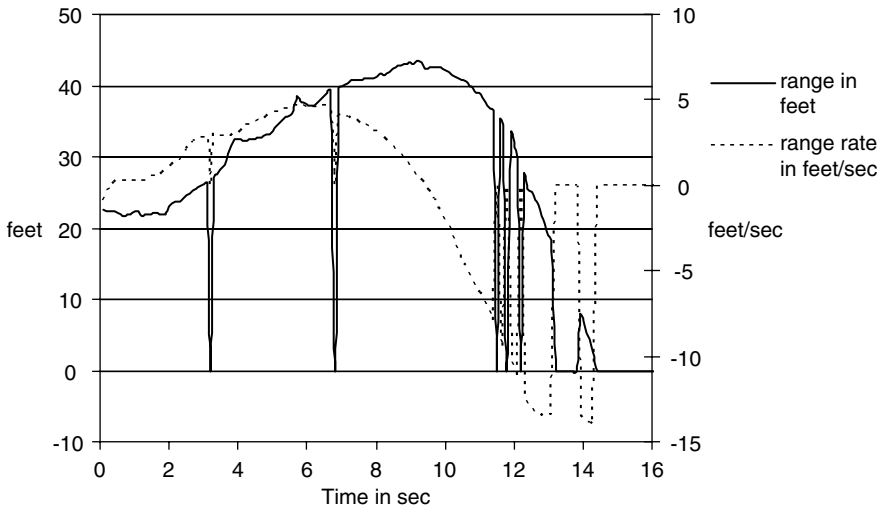


Figure 2.1. Speedometer speeds of following vehicle.

suggests a three-phase model where a period of constant acceleration is followed by a period of roughly constant speed, which in turn is followed by a period of constant deceleration leading to a stop.

Step 2: Nonlinear Least-Squares Estimation of Proposed Models

It is necessary to estimate the following vehicle's initial speed, its acceleration during the first phase, the time at which deceleration began, and the deceleration characteristic of the second phase. To estimate these, a MATLAB script was written that took trial values for these parameters as inputs and simulated the following vehicle's position and speed over time by solving the differential equations using a simple Euler's method. The difference between the simulated speeds and the speedometer speeds was computed for each time interval, and the squares of these differences were summed to produce a measure of fit between the model and the speedometer data. This script was then embedded in a numerical



Note: Discontinuities are due to missing data.

Figure 2.2. Range and range-rate data for the leading vehicle.

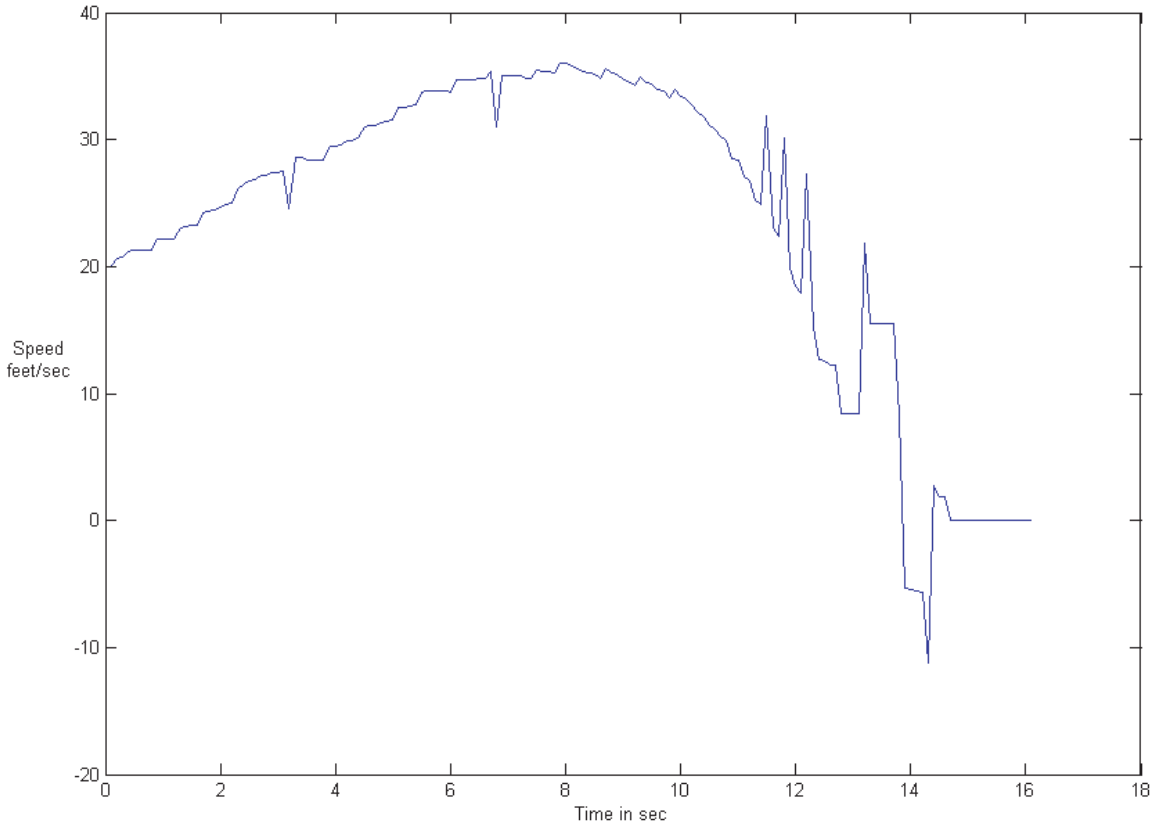


Figure 2.3. Approximate speed of lead vehicle determined from follower speed and range rate.

search procedure to find the parameter values that minimized the sum-of-squares. The resulting estimates and initial approximate standard errors are shown in Table 2.1.

For the lead vehicle, the three-phase model was fit to the approximate speed data shown in Figure 2.3. The resulting nonlinear least-squares estimates and approximate standard errors are given in Table 2.2.

Step 3: Bayes Estimation of Vehicle Models

Final Bayes estimates were computed using the MCMC software WinBUGS. In essence, WinBUGS generates a simulated realization of a Markov chain whose stationary distribution is the same as the Bayesian posterior distribution of the model parameters given the data. For this, the data were the speed profile for the follower and the range and range-rate profiles for the leader. Model parameters consisted of acceleration profiles and initial speeds for each vehicle, and the parameters for both vehicles were estimated in combination simultaneously. In the WinBUGS model, predicted values for the follower's speed and the range and range rate for the

Table 2.1. Nonlinear Least-Squares Estimates for the Following Vehicle's Acceleration Model (Case 104119)

Parameter	Estimate	Standard Error
Initial acceleration (ft/s^2)	1.45	.02
Final acceleration (ft/s^2)	–9.9	.17
Transition time (seconds from start)	11.2	.04

Table 2.2. Nonlinear Least-Squares Estimates for Lead Vehicle Acceleration Model (Case 104119)

Parameter	Estimate	Standard Error
First acceleration (ft/s^2)	2.36	.05
Second acceleration (ft/s^2)	–0.84	.23
Third acceleration (ft/s^2)	–9.86	.40
First change point (s)	6.75	.06
Second change point (s)	10.4	.09

Table 2.3. Posterior Summary for Trajectory Model Parameters

Variable	Mean	Standard Deviation	2.5%ile	97.5%ile
Following Vehicle				
Initial speed (ft/s)	20.6	0.24	20.1	21.0
First acceleration (ft/s^2)	1.45	0.04	1.38	1.54
Second acceleration (ft/s^2)	–9.47	0.21	–9.89	–9.05
First change (s)	11.15	0.06	11.04	11.26
Leading Vehicle				
Initial speed (ft/s)	18.92	0.31	18.3	19.52
First acceleration (ft/s^2)	3.34	0.14	3.08	3.62
Second acceleration (ft/s^2)	0.62	0.05	0.51	0.72
Third acceleration (ft/s^2)	–11.94	0.32	–12.58	–11.34
First change (s)	3.47	0.13	3.22	3.73
Second change (s)	10.61	0.05	10.51	10.72

leader were computed by numerically solving the differential equations using ordinary differential equation interface. For this case, a 10,000-iteration burn-in period followed by a 70,000-iteration MCMC sample produced acceptable convergence. Table 2.3 summarizes the Bayes estimates for both leader and follower.

Figure 2.4 shows the speedometer speeds of the following vehicle, together with the posterior means estimated from the MCMC sample. Figure 2.5 shows similar results for the range data from the forward radar. Both cases provide reasonable approximations to the observed information.

Finally, Figure 2.6 shows the probability of a rear-end collision between the two vehicles as a function of counterfactual final decelerations by the following driver. These probabilities were computed using Balke and Pearl's (*6*) Twin Network method, where the follower's deceleration is set at a target value and then, for each outcome of the MCMC, sample values for the remaining parameters are used as inputs to solve the differential equations. This describes what would have happened had the event involved those parameter values and the counterfactual follower's deceleration. Simulated range values less than zero are taken to indicate a collision, and the fraction of the MCMC sample values that lead to collision is an estimate of the collision probability. In this case, had the follower braked at less than about –10 ft/s^2, a collision would probably have resulted.

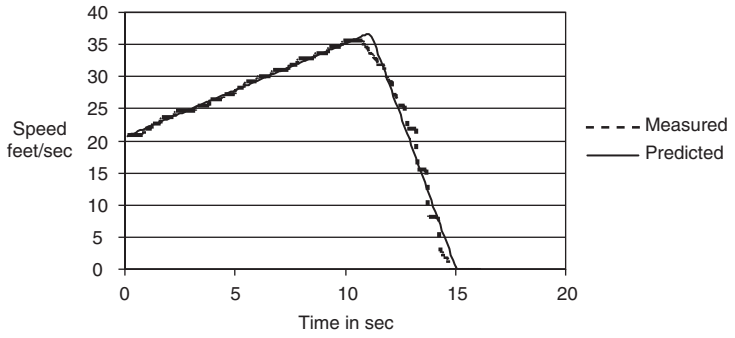


Figure 2.4. Measured following vehicle speeds and posterior mean predicted speeds from MCMC sample.

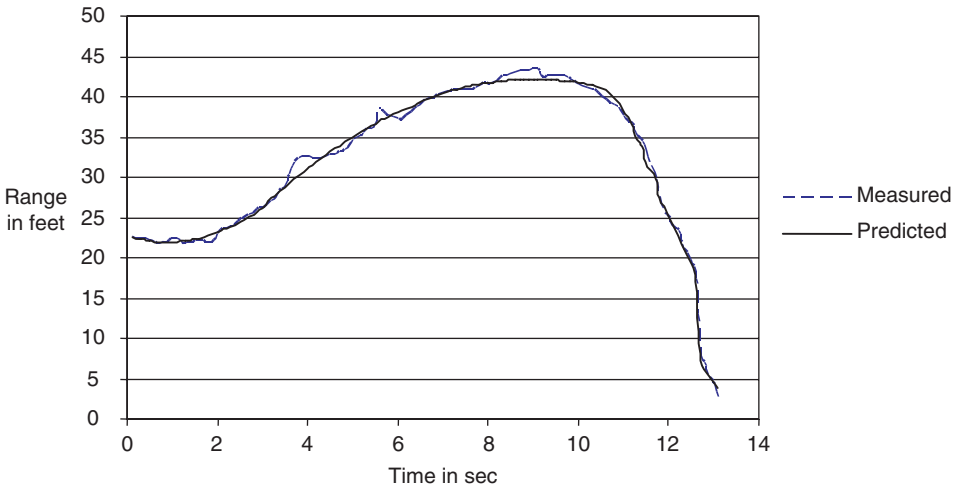


Figure 2.5. Measured range and posterior mean predicted range from MCMC sample.

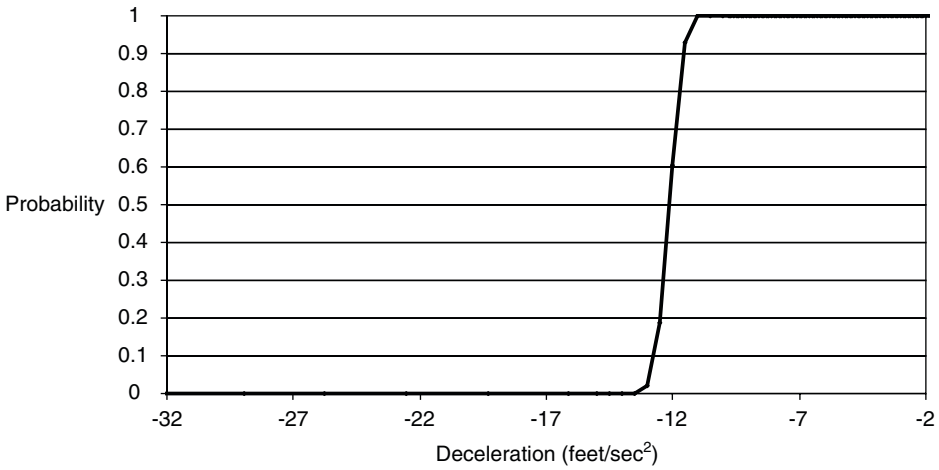


Figure 2.6. Collision probability as a function of counterfactual values for follower's final deceleration.

References

1. Bayarri, M., J. Berger, D. Higdon, M. Kennedy, A. Kottas, R. Paulo, J. Sacks, J. Cafeo, J. Cavendish, C. H. Lin, and J. Tu. *A Framework for Validation of Computer Models.* Technical Report 128. National Institute of Statistical Sciences, Research Triangle Park, N.C., 2002.
2. Seber, G. A. F., and C. J. Wild. *Nonlinear Regression,* Wiley, New York, 1989.
3. Lunn, D., A. Thomas, N. Best, and D. Spiegelhalter. WinBUGS—A Bayesian Modelling Framework: Concepts, Structure, and Extensibility. *Statistics and Computing,* Vol. 10, No. 4, 2000, pp. 325–337.
4. *Using MATLAB (Version 6).* The MathWorks, Inc., Natick, Mass., 2002.
5. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2008. www.R-project.org.
6. Balke, A., and J. Pearl. Probabilistic Evaluation of Counterfactual Queries. Presented at 12th National Conference on Artificial Intelligence (AAAI-94), Seattle, Wash., 1994.

CHAPTER 3

Analyses Using Vehicle-Based Data

Data Acquisition and Preparation

In May and June 2008, the Virginia Tech Transportation Institute (VTTI) provided the project team with data for 33 events observed during the 100-car in-vehicle study. These data comprised time series of measurements from the in-vehicle sensors, along with the videos from the forward camera, for about 30 s preceding and including the crash or near-crash event. Table 3.1 lists the variables obtained.

The first task after receiving these data was to review them for completeness and accuracy and flag those events for which the data were either incomplete or problematic. For this task, a MATLAB-based graphical browsing tool was developed that loaded the data from the VTTI-provided text file, allowing the analyst to plot the history of selected measurements and to perform two initial simulations of the trajectory of the instrumented vehicle, using either speed- or acceleration-based measurements as input. Inconsistencies between the outputs of these two models would then indicate problems with either the speed or the acceleration data (or possibly both) for that event.

For five of the events, measurements from the forward radar were not available, so for these it was not possible to reconstruct the actions of the leading vehicle. For several more of the events, there was a clear discrepancy between the trajectory of the instrumented vehicle as indicated by the speedometer, heading, and yaw measurements, and the trajectory reconstructed from the acceleration measurements, indicating that caution should be used when using these data.

For the instrumented vehicle, the data available consisted of speedometer output, lateral and longitudinal accelerations, yaw, heading, and indications of the status of the turn signal, the brake, and the accelerator, recorded at 10 Hz. For the lead vehicle, the available data consisted of range, range rate, and azimuth obtained from the forward-viewing radar, also recorded at 10 Hz. Latitude and longitude values from the GPS receiver were available, but in all cases these values were essentially constant throughout the event.

After initial examination of the data obtained for each of the 33 cases, a summary was prepared (see Table 3.2) identifying the potential cases that could be examined for this research, based on the quality and the completeness of the available data. Several of the events involved lane-changing, swerving, or merging on the part of one of the involved vehicles, so that the forward radar of the instrumented vehicle provided only limited information about the leading vehicle. In total, seven of the 33 events were analyzed. Those seven cases are 99540, 104119, 73082, 104851, 104283, 60289, and 92660.

In the following sections, the analysis of each of these seven events is described. The accuracy of the conclusions, though, depends on the accuracy of the data on which they are based. Although the project team is confident that the results are consistent with the data provided, these analyses are presented as examples of what could potentially be accomplished using vehicle-based field data rather than as final determinations of what truly happened in these events.

Case 99540

Description from video: Figures 3.1–3.3 illustrate that in this event, the instrumented (i.e., following) vehicle is traveling in the right-hand lane of a multilane highway and exits this highway to the right. The exit ramp connects to another multilane highway, and the leading vehicle slows and then comes to a stop at the merge point. The following vehicle collides with the lead vehicle, which moves forward and then stops on the roadway's shoulder.

Approximately 35 s of data were available from the instrumented vehicle at 10 Hz. These data included speeds from the instrumented vehicle's speedometer, and range and range rate from its forward radar, as shown in Figure 3.4.

In addition to the original data, approximate speeds for the leading vehicle were computed by adding the instrumented

Table 3.1. Data Obtained from 100-Car In-Vehicle Study

Data Requested	Data Delivered	Notes	Units
	Sync		frames
Speed	Speed		mile/hr
GPS location	lat, lon		degrees/s
Throttle	Throttle		unitless
Brake	brake_onoff	0 = off, 1 = on	
Lateral acceleration	accel_x		*g*
Longitudinal acceleration	accel_y		*g*
Target ID	fwd_ID_n	n = 1,2, . . . 7	cycles 1-255
Range	fwd_range_n	n = 1,2, . . . 7	ft
Range-rate	fwd_range_rate_n	n = 1,2, . . . 7	ft/s
Azimuth	fwd_azimuth_n	n = 1,2, . . . 7	degrees
Yaw	Yaw		degrees/s
Turn signal state	turn_signal	0 = off; 1 = left; 2 = right; 3 = both	
Video forward	Video forward	avi format. Video is not synchronized with parametric data.	

Table 3.2. Summary of the Data Obtained from 33 Events

Event ID	Event Type	Speedometer	Forward Radar
04	Rear-end crash		Cuts out ~4 seconds before crash
20	Rear-end crash	Nonzero after video indicates vehicle stopped	OK
24	Swerve to pole crash	Possible	Only prior to swerve
104119	**Rear-end conflict**	**OK**	**OK**
104283	**Stationary lead vehicle**	**OK**	**OK**
104851	**Rear-end conflict**	**OK**	**OK**
113846	Lane change conflict	Reads constant	
118405	Rear-end conflict	Reads constant	
135941	Merging conflict		No near lead vehicle
139130	Multiple crash		
167847	Evasive swerve		
179442	Rear-end conflict		Corrupted
180462	Rear-end conflict		Corrupted
183129	Merging conflict		
189066	Rear-end conflict		Corrupted
20993	Rear-end conflict	Reads constant	
24523	Lane-change conflict		After start of evasive action
38499	Rear-end crash		Positive range during/after crash
40003	Rear-end crash	Reads constant 0	
52243	Rear-end conflict		Corrupted

(continued on next page)

Table 3.2. Summary of the Data Obtained from 33 Events (continued)

Event ID	Event Type	Speedometer	Forward Radar
60289	**Rear-end conflict**	**OK**	**OK**
73082	**Rear-end conflict**	**OK**	**OK**
86319	Rear-end conflict	Reads constant	corrupted
86535	Evasive swerve conflict		
87089	Rear-end crash		Available only ~2 seconds before crash
92660	**Evasive swerve conflict**	**OK, done** (without counterfactual)	**Limited**
99129	Evasive swerve conflict		
99540	**Rear-end crash**	**OK**	**OK**
122474	Rear-end conflict		Missing
151062	Merging conflict		Missing
1984	Rear-end crash		Missing
3188	Merging conflict		Missing
3336	Merging conflict		Missing

Boldface: Analyzed events.

vehicle's speedometer speed to the forward radar's range rate. The speeds for the two vehicles are displayed in Figure 3.5.

Exploratory modeling for both vehicles was conducted using MATLAB. For the following (instrumented) vehicle, a three-stage model was fit, where a period of initial deceleration lasting about 2 s was followed by a period of gentler deceleration, which was followed by a short period of much stronger deceleration starting less than 1 s before the collision. For the leading vehicle, a two-stage model was fit, where roughly 2 s of deceleration was followed by a period of being stopped, which lasted until the collision. Bayes estimates for each vehicle's initial speed, the time points at which each driver changed acceleration and the corresponding accelerations were computed using WinBUGS, and the results are displayed in Table 3.3. For this case, because the data acquisition began after the leading driver had begun his or her final deceleration, the follower's reaction time was computed as the difference between when the lead vehicle came to a stop and when the following driver initiated the final deceleration.

At the time the forward radar acquired the leading vehicle, the driver of the instrumented vehicle was traveling at about

Figure 3.1. An initial view from instrumented vehicle's forward camera (Case 99540).

Figure 3.2. View as the instrumented vehicle exits (Case 99540). The leading vehicle is visible on the exit ramp.

Figure 3.3. View at the time of collision with the leading vehicle (Case 99540).

14.75 ft/s, while the driver of the leading vehicle was traveling at about 8.9 ft/s. The leading driver was decelerating at about −3.7 ft/s^2 and came to a stop about 2.4 s after the acquisition. The driver of the following vehicle was initially decelerating at about −2.9 ft/s^2, but after about 1.9 s eased up to about −0.5 ft/s^2. About 2.37 s after the leading vehicle came to a stop, the driver of the following vehicle began braking at about −12.6 ft/s^2, but this was not sufficient to prevent a collision. At the time the lead vehicle came to a stop, the following vehicle was about 20.7 ft behind and traveling at about 9 ft/s.

Figure 3.6 shows the speedometer speeds for the instrumented vehicle, along with the speed history predicted by the model. Allowing for the piecewise constant nature of the recorded speedometer output, the model gives a reasonable representation of the follower's speed profile.

Figure 3.7 shows the radar-measured range between the leading and following vehicle, along with the range predicted by the model. Again, the model gives a reasonable representation of the data.

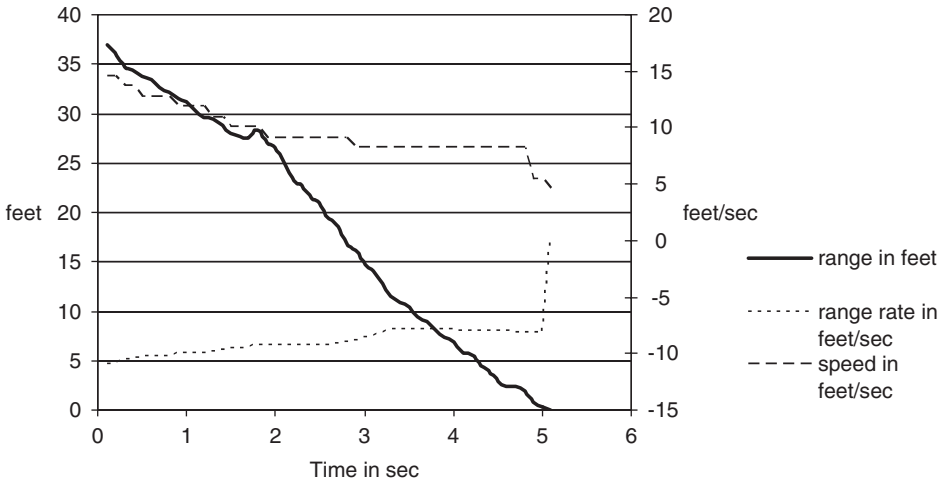


Figure 3.4. Speed, range, and range-rate data (Case 99540).

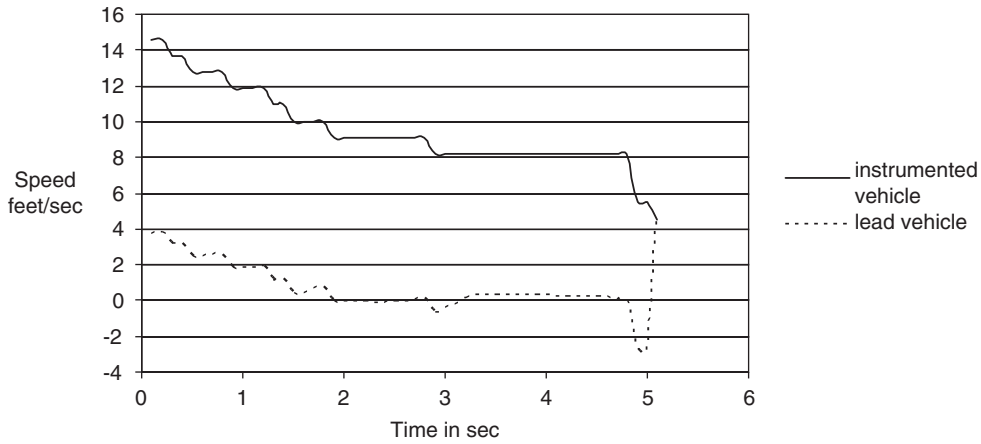


Figure 3.5. Speeds for leading and following vehicles (Case 99540).

Table 3.3. WinBUGS Estimates for the Model Parameters for Case 99540

Variable	Mean	Standard Deviation	2.5%ile	97.5%ile
Following Vehicle				
Initial speed (ft/s)	14.75	0.16	14.45	15.08
First acceleration (ft/s^2)	–2.87	0.14	–3.18	–2.61
Second acceleration (ft/s^2)	–0.48	0.08	-0.63	–0.32
Third acceleration (ft/s^2)	–12.57	2.31	–17.25	–8.10
First change (s)	1.94	0.09	1.75	2.12
Second change (s)	4.78	0.03	4.71	4.83
Reaction time (s)	2.37	0.12	2.13	2.59
Speed start reaction (ft/s)	8.96	0.11	8.75	9.18
Separation start reaction (t)	20.74	1.01	18.81	22.64
Leading Vehicle				
Initial speed (ft/s)	8.94	0.44	8.11	9.81
First acceleration (ft/s^2)	–3.72	0.35	–4.45	–3.085
Stop time (s)	2.41	0.12	2.20	2.59

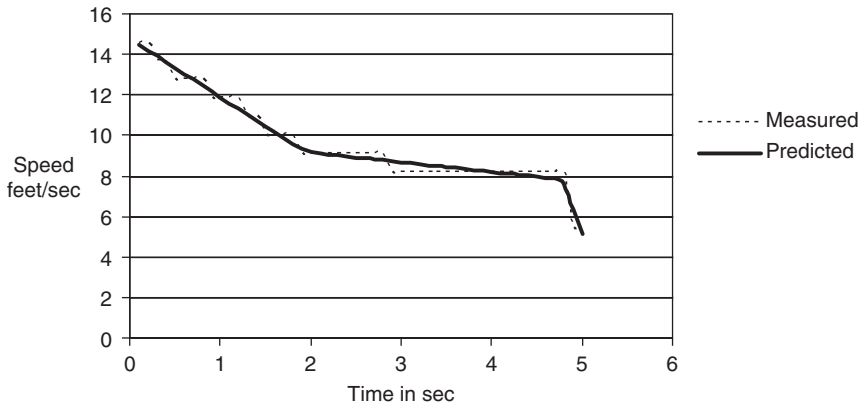


Figure 3.6. Measured and modeled following vehicle speeds (Case 99540).

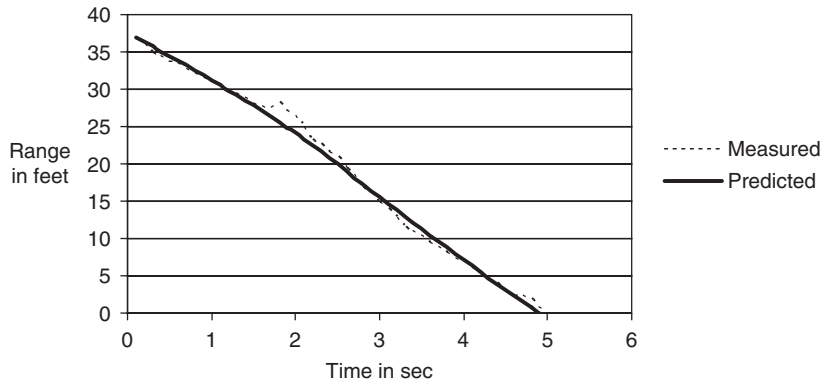


Figure 3.7. Measured and modeled range data (Case 99540).

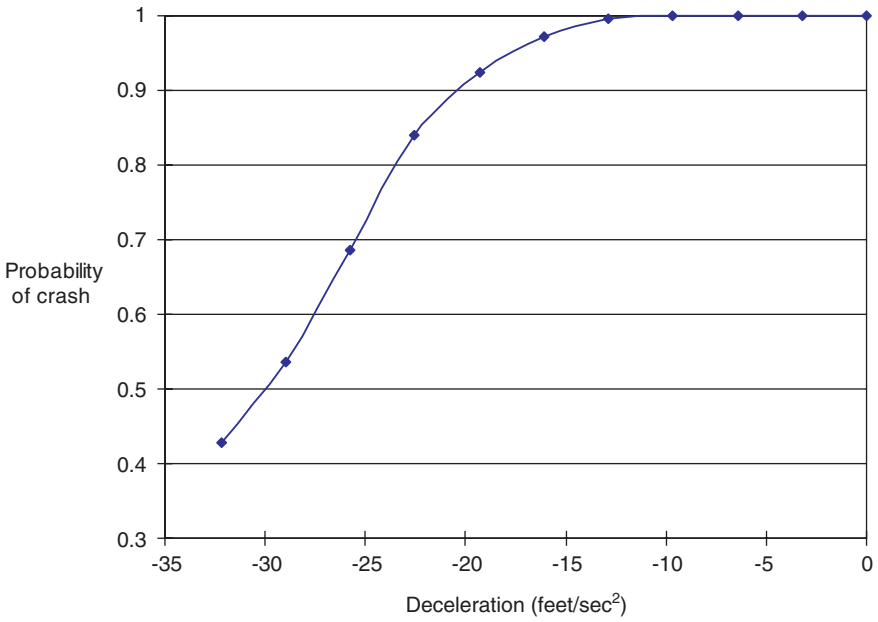


Figure 3.8. Counterfactual model (Case 99540).

To assess the avoidability of this crash, probabilities of collision were computed as a function of counterfactual final decelerations on the part of the following driver, as displayed in Figure 3.8. For this event, because the following driver did not initiate evasive action until close to collision, even fairly high counterfactual decelerations are not sufficient to prevent the collision.

Case 104119

Description from video: In this event, the instrumented (i.e., following) vehicle was traveling on a signalized roadway and turned left at an intersection following the lead vehicle. The leader braked to a stop, as did the follower, without colliding (Figures 3.9–3.11).

Approximately 35 s of data were available at 10 Hz, including speedometer-measured speeds for the instrumented vehicle and range and range rate from the follower's forward radar. It was possible to reliably identify the lead vehicle in the radar data for about 16 s, and the speed, range, and range rate for the period are displayed in Figure 3.12.

Figure 3.13 compares the speed trajectories of the leading and following vehicles.

Exploratory modeling for both vehicles was conducted using MATLAB. For the following (instrumented) vehicle, a two-stage model was fit, where a period of initial acceleration was

Figure 3.9. View at end of follower's left turn (Case 104119).

Figure 3.10. View at approximately when leader began stopping (Case 104119).

Figure 3.11. View when follower stopped (Case 104119).

followed by deceleration to a stop. For the leading vehicle, a three-stage model was fit, where a first stage of acceleration was followed by a stage of gentler acceleration, which was then followed by decelerating to a stop. Bayes estimates for each vehicle's initial speed, the time points at which each driver changed acceleration, and the corresponding accelerations were computed using WinBUGS. The results are displayed in Table 3.4. The following driver's reaction time was determined as the difference between when the follower began final deceleration and when the leader began final deceleration.

In this case, at the time of the start of the data series, the leading vehicle was traveling at about 18.9 ft/s and accelerating at about 3.34 ft/s^2, while the follower was traveling at about 20.6 ft/s and accelerating at about 1.45 ft/s^2. After about 3.46 s, the leader eased his or her acceleration to about 0.62 ft/s^2, and after about 10.6 s from the start of the data series, the leader began decelerating at about −11.94 ft/s^2, which continued until the vehicle stopped. About 11.15 s from the start of the data series, the follower began decelerating at about −9.47 ft/s^2. At the time the leader began final deceleration, the follower was about 20.7 ft behind and traveling at about 36 ft/s. The follower's reaction time was fairly quick, about 0.53 s.

Figure 3.14 compares the following vehicle's speed as given by its speedometer and as predicted by the fitted model. Figure 3.15 shows a similar comparison of the range between the follower and leader as given by the forward radar and the range as predicted by the trajectory models. In both cases, there is a plausible reconstruction of the data series.

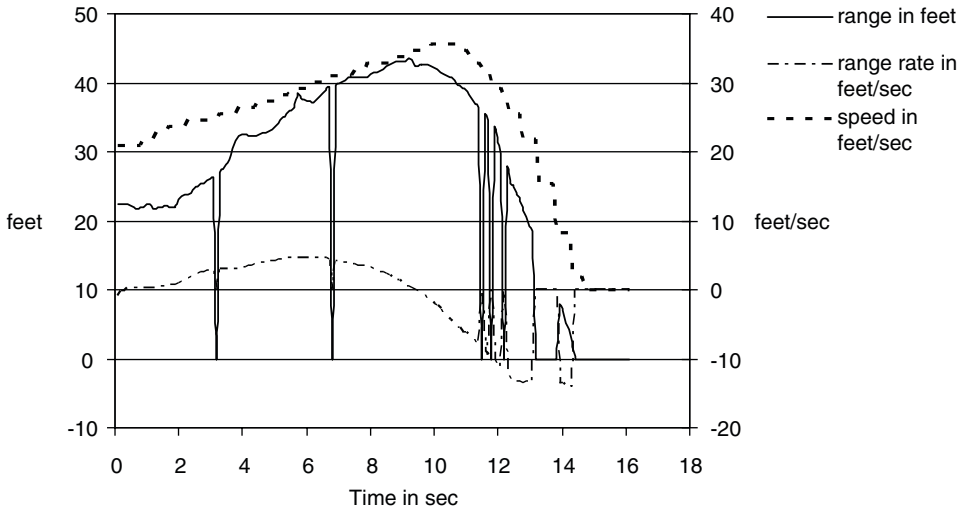
Finally, Figure 3.16 displays the probability of a rear-end crash occurring as a function of counterfactual final decelerations on the part of the following driver.

Case 73082

Description from video: In this event, both the leading and the following vehicles are initially stopped at a signalized intersection. The lead vehicle accelerates, then the follower accelerates and the leader pulls away from the follower. The leader then brakes to a stop, as does the follower, and the follower stops short of the leader without colliding (Figures 3.17–3.19).

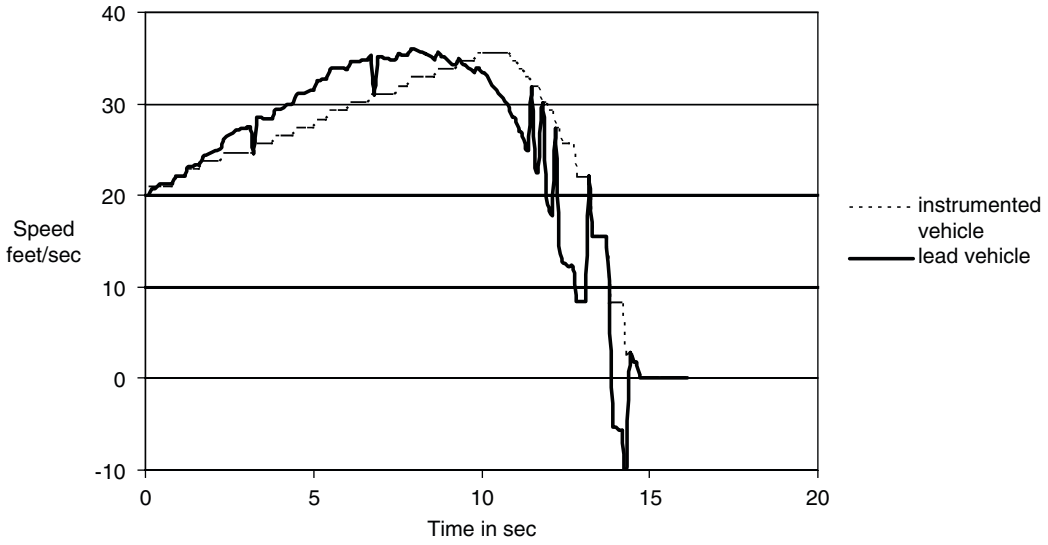
Figures 3.20 and 3.21 show the speed, range, and range-rate plots of the following and leading vehicles.

Exploratory modeling suggested a three-stage model for the following vehicle and a four-stage model for the leader. Bayes estimates for each vehicle's initial speed, the time points at which each driver changed acceleration, and the corresponding accelerations were computed using WinBUGS.



Note: Discontinuities are due to missing data.

Figure 3.12. Speed, range, and range-rate data (Case 104119).



Note: Discontinuities are due to missing data.

Figure 3.13. Speeds of leading and following vehicles (Case 104119).

These results are displayed in Table 3.5. The following driver's reaction time was determined as the difference between the time when the follower's final deceleration began and the time when the leader's final deceleration began.

For this case, the follower's initial speed was taken to be known and equal to 0 ft/s, while at the start of the data sequence the leader was traveling at about 2.15 ft/s and accelerating at 10.1 ft/s^2. After about 1.8 s, the leader reduced acceleration to about 4.1 ft/s^2. After about 8.8 s the leader began decelerating at −3.9 ft/s^2, and after about 10.2 s, the leader increased the deceleration rate to about −13.5 ft/s^2. The following driver initially accelerated at about 5.4 ft/s^2, and after about 6.65 s, eased to about 2.0 ft/s^2. After about 12 s, the follower began braking at the relatively high rate of −18.3 ft/s^2. The follower did not appear to have responded to the leader's initial deceleration.

Figure 3.22 compares the following vehicle's speed as given by its speedometer and the speed as predicted by the fitted model. Figure 3.23 shows a similar comparison of the range between the follower and leader as given by the

Table 3.4. WinBUGS Estimation Summary for Case 104119

Variable	Mean	Standard Deviation	2.5%ile	97.5%ile
Following Vehicle				
Initial speed (ft/s)	20.6	0.24	20.1	21.0
First acceleration (ft/s^2)	1.45	0.04	1.38	1.54
Second acceleration (ft/s^2)	–9.47	0.21	–9.89	–9.05
First change (s)	11.15	0.06	11.04	11.26
Reaction time (s)	0.53	0.04	0.45	0.62
Speed start reaction (ft/s)	36.04	0.23	35.6	36.5
Separation start reaction (ft)	20.74	1.01	18.81	22.64
Leading Vehicle				
Initial speed (ft/s)	18.92	0.31	18.3	19.52
First acceleration (ft/s^2)	3.34	0.14	3.08	3.62
Second acceleration (ft/s^2)	0.62	0.05	0.51	0.72
Third acceleration (ft/s^2)	–11.94	0.32	–12.58	–11.34
First change (s)	3.47	0.13	3.22	3.73
Second change (s)	10.61	0.05	10.51	10.72

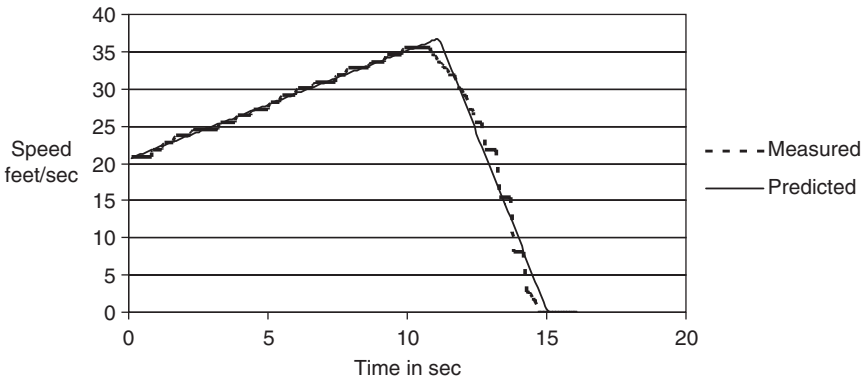


Figure 3.14. Measured and modeled following vehicle speeds (Case 104119).

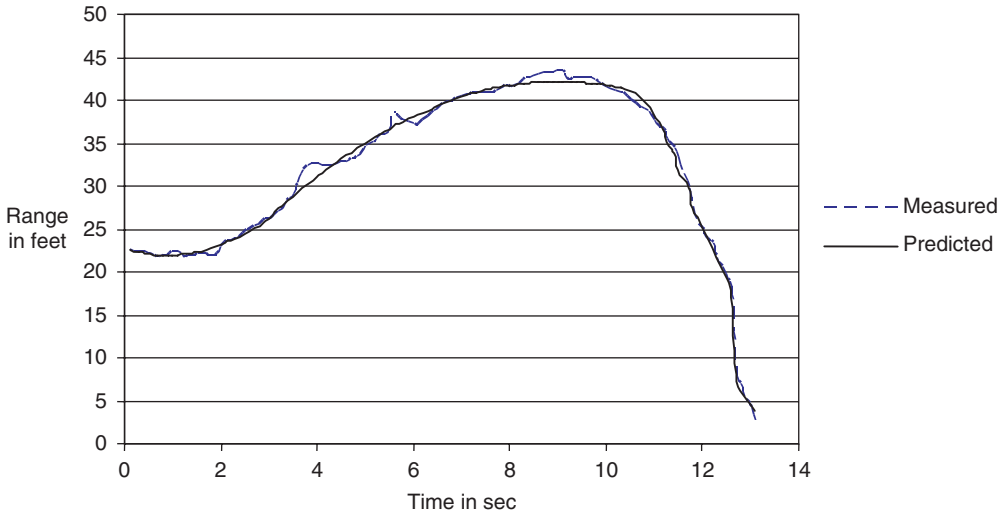


Figure 3.15. Measured and modeled range data (Case 104119).

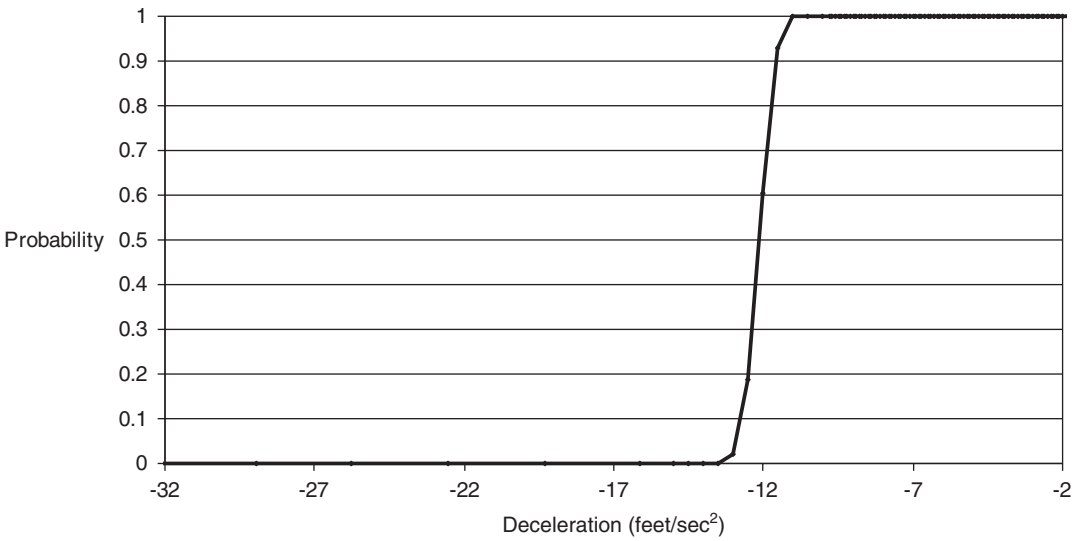


Figure 3.16. Counterfactual model (Case 104119).

Figure 3.17. Both vehicles in initial stopped position (Case 73082).

forward radar and the range as predicted by the trajectory models. In both cases, there is a plausible reconstruction of the data series.

Case 104851

In this case, the instrumented vehicle (i.e., the following vehicle) at the start of the video (see Figure 3.24) took a right turn and continued to follow the leading vehicle. But then the leading vehicle decelerated and came to a complete stop. This forced the following vehicle also to decelerate, resulting in a near crash. However, the following driver's deceleration was sufficient to enable the vehicle to come to a complete stop without any collision. Although the total length of the video was 19 s, the event actually happened within the first 8 s. In the remaining period of time, both vehicles were stopped.

Figure 3.18. Lead vehicle begins braking (Case 73082).

Figure 3.19. Following vehicle stopped behind leader (Case 73082).

The leading vehicle came to a complete stop, indicated by the brake lights, which almost resulted in a rear-end collision.

Figure 3.25 shows the speed trajectories of the leading and following vehicles. The blue and red lines indicate the speed of the following and leading vehicles in ft/s.

The speed of the following vehicle was obtained directly from the speedometer of the instrumented car. The approximate speed of the leading vehicle was calculated by adding the speed of the following vehicle and the range-rate data obtained from radar. A similar approach was adopted for all the remaining cases discussed in this section. Figure 3.26 shows the range and range-rate data obtained for this event.

After initial estimates of the change points and accelerations were obtained from MATLAB, the trajectory model was fitted in WinBUGS for final estimates. In this case, a three-stage model was developed for the following vehicle, with initial acceleration followed by two different deceleration stages. Table 3.6 gives the final MCMC simulation estimates of the parameters.

When the radar acquired the leading vehicle, the initial speeds of the following and leading vehicles were 25.66 ft/s and 26.07 ft/s, respectively. The leading vehicle decelerated in three different stages. The first two deceleration stages were characterized by mild deceleration followed by a very steep deceleration (–24.29 ft/s^2), bringing the leading vehicle to a complete stop. Subsequently, the following vehicle initially was accelerating for 2.626 s, and then it decelerated at –21.76 ft/s^2 followed by a third deceleration of –2.87 ft/s^2. The predicted piecewise acceleration model was compared by fitting the observed data. The range and speed of the following vehicle was fitted as shown Figures 3.27 and 3.28. Table 3.6 lists the estimates obtained from the WinBUGS output.

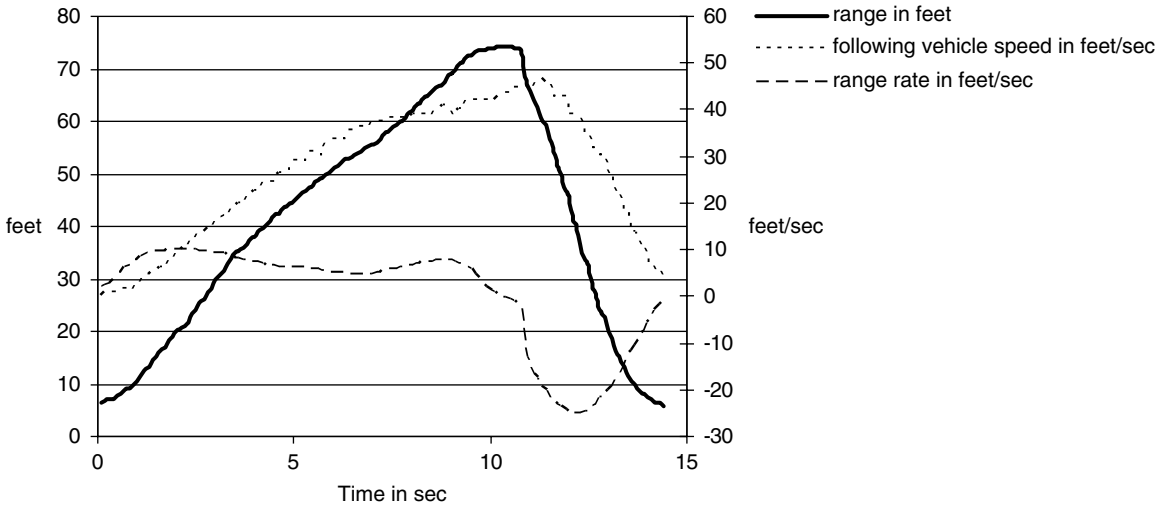


Figure 3.20. Speed, range, and range-rate data (Case 73082).

Similar to the previous cases, a counterfactual model based on different deceleration rates for the following vehicle was simulated, and for each deceleration, a probability of crash was computed. Figure 3.29 shows how the chance of crash varies over different counterfactual deceleration values.

Case 104283

This event, as shown in Figures 3.30–3.32, is a near crash. The instrumented vehicle was traveling in the rightmost lane of an arterial and continued to travel until it was forced to a complete stop to avoid a rear-end collision with the leading vehicle, which was waiting for a gap to change lanes at a merging section of the arterial. The total duration of the video was 35 s, and the event occurred at about 23 s.

Respective speed trajectories for the leading and following vehicles were plotted in Figure 3.33. The radar could only manage to capture the leading vehicle's information for about 5 s. Also, the range and range-rate data were collected as shown in Figure 3.34.

Initial speed of the following vehicle was 60 ft/s, compared to the initial speed of 30.93 ft/s for the leading vehicle (Table 3.7). This speed is the estimated speed of the leading vehicle when the radar captured information about the leading vehicle for the first time. A two-stage model was proposed for the following vehicle where in the first stage the vehicle decelerated at −7.57 ft/s^2 until 3.92 s, then shifted to a stronger deceleration rate of −16.16 ft/s^2. The leading vehicle's trajectory was also fitted with a two-stage model, with −10 ft/s^2 of deceleration in the first stage for 1.714 s, and −4.332 ft/s^2 deceleration in the second stage.

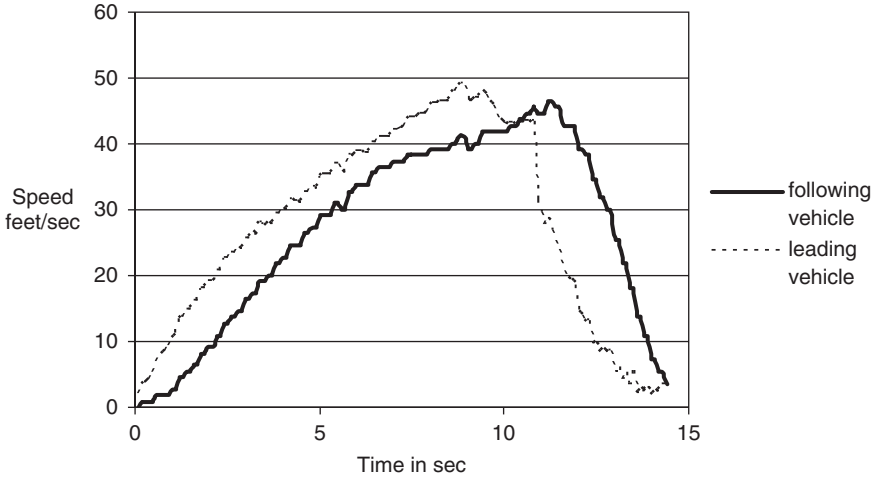


Figure 3.21. Speeds of leading and following vehicles (Case 73082).

Table 3.5. Bayes Estimation Summary for Case 73082

Variable	Mean	Standard Deviation	2.5%ile	97.5%ile
Following Vehicle				
Initial speed (ft/s)	0	—	—	—
First acceleration (ft/s^2)	5.41	0.03	5.34	5.47
Second acceleration (ft/s^2)	2.0	0.12	1.74	2.23
Third acceleration (ft/s^2)	–18.33	0.41	–19.16	–17.51
First change (s)	6.65	0.09	6.485	6.84
Second change (s)	11.96	0.05	11.86	12.05
Reaction time (s)	1.75	0.12	1.50	2.0
Speed start reaction (ft/s)	43.08	0.37	42.37	43.78
Separation start reaction (ft)	74.05	0.19	73.61	74.37
Leading Vehicle				
Initial speed (ft/s)	2.15	0.23	1.715	2.56
First acceleration (ft/s^2)	10.1	0.26	9.64	10.59
Second acceleration (ft/s^2)	4.07	0.05	3.97	4.18
Third acceleration (ft/s^2)	–3.87	0.74	–5.36	–2.46
Fourth acceleration (ft/s^2)	–13.52	0.41	–14.41	–12.76
First change (s)	1.82	0.06	1.69	1.93
Second change (s)	8.77	0.09	8.61	8.96
Third change (s)	10.21	0.09	10.03	10.39

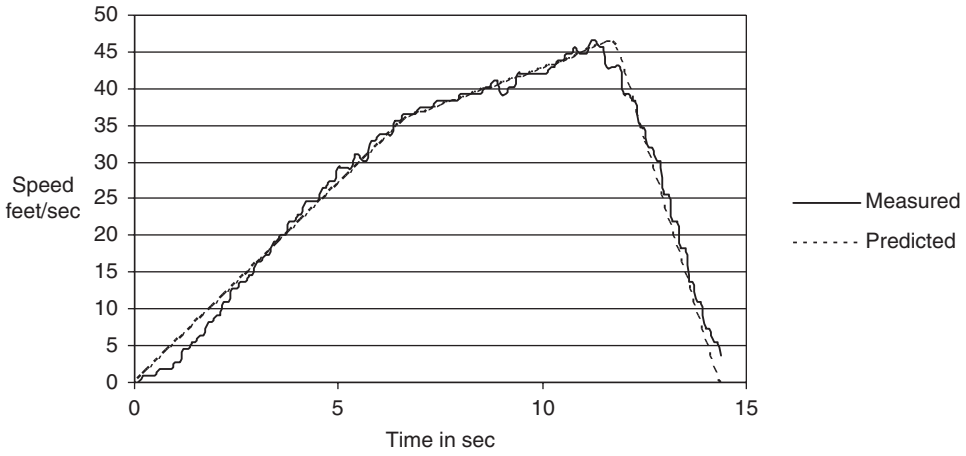


Figure 3.22. Measured and modeled speedometer speeds from following vehicle (Case 73082).

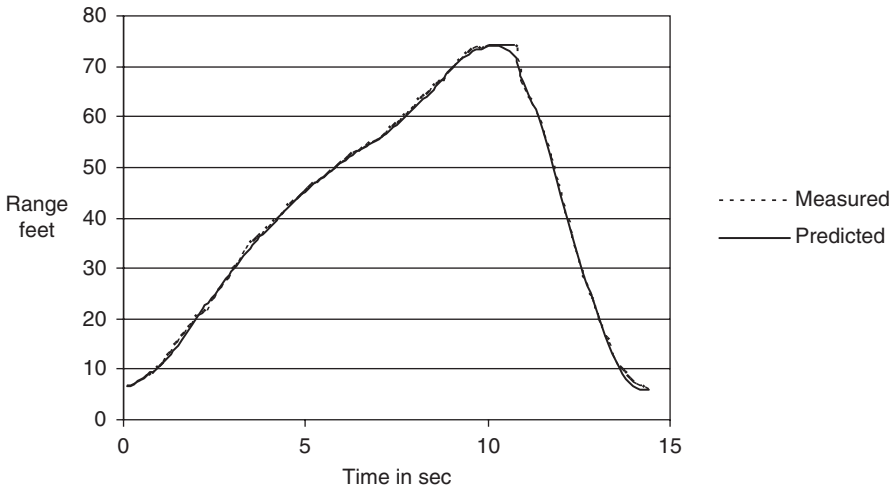


Figure 3.23. Measured and modeled range data for leading vehicle (Case 73082).

The piecewise model appears plausible as shown in Figures 3.35 and 3.36 fitting the observed speed of the following vehicle and range data.

Also, a counterfactual model (see Figure 3.37) was developed that shows that if everything else remained constant but the deceleration of the following vehicle in the last stage was stronger than −5.7 ft/s^2, then the probability of a crash is essentially zero; on the other hand, if the deceleration was weaker than about −4.2 ft/s^2, a crash is nearly certain.

Case 60289

As shown in Figure 3.38, in this event the two vehicles were closely following each other. The leading vehicle accelerated and then traveled at uniform speed before it decelerated to almost zero speed. The following vehicle kept to the same pattern as the leading vehicle, shown in Figure 3.39. Range and range-rate data obtained from the radar is shown in Figure 3.40.

Figure 3.24. View of the leading vehicle (Case 104851).

A four-stage model was constructed to fit the following vehicle's speed trajectory. Initially, the following vehicle was traveling at 11.21 ft/s and then accelerated at 3.15 ft/s^2 for 5.36 s. Then it traveled at almost constant speed for another 5 s before decelerating at −2.419 ft/s^2 for 4.3 s, followed by a strong deceleration of −10.74 ft/s^2, and finally came to a stop at 16.47 s. A similar pattern was observed for the leading vehicle, which had an initial acceleration stage of 8.16 ft/s^2 for 1.764 s, followed by a period of 8.277 s of almost constant speed, and then two deceleration stages, with the final deceleration rate as high as −9.502 ft/s^2. The similar speed profile of the two vehicles seems reasonable because they were following each other closely in this case. Table 3.8 shows the WinBUGS estimates.

Predicted versus observed speed and range values (see Figures 3.41 and 3.42) were plotted and look quite reasonably represented by the four-stage model.

A similar counterfactual model (see Figure 3.43) was developed to show the probability of a crash for different deceleration values in the final stage of the following vehicle.

Case 92660

In this event, a vehicle was closely following another vehicle on a two-lane, two-way highway as shown in Figure 3.44. The leading vehicle suddenly stopped, and the following vehicle had to swerve to avoid a collision, as shown in Figure 3.45.

The speed trajectory (see Figure 3.46) shows that the leading vehicle initially traveled at mild acceleration for a considerable period of time followed by a steep deceleration, which intensified in the last or final phase. On the other hand, the following (instrumented) vehicle started with a gentle acceleration and

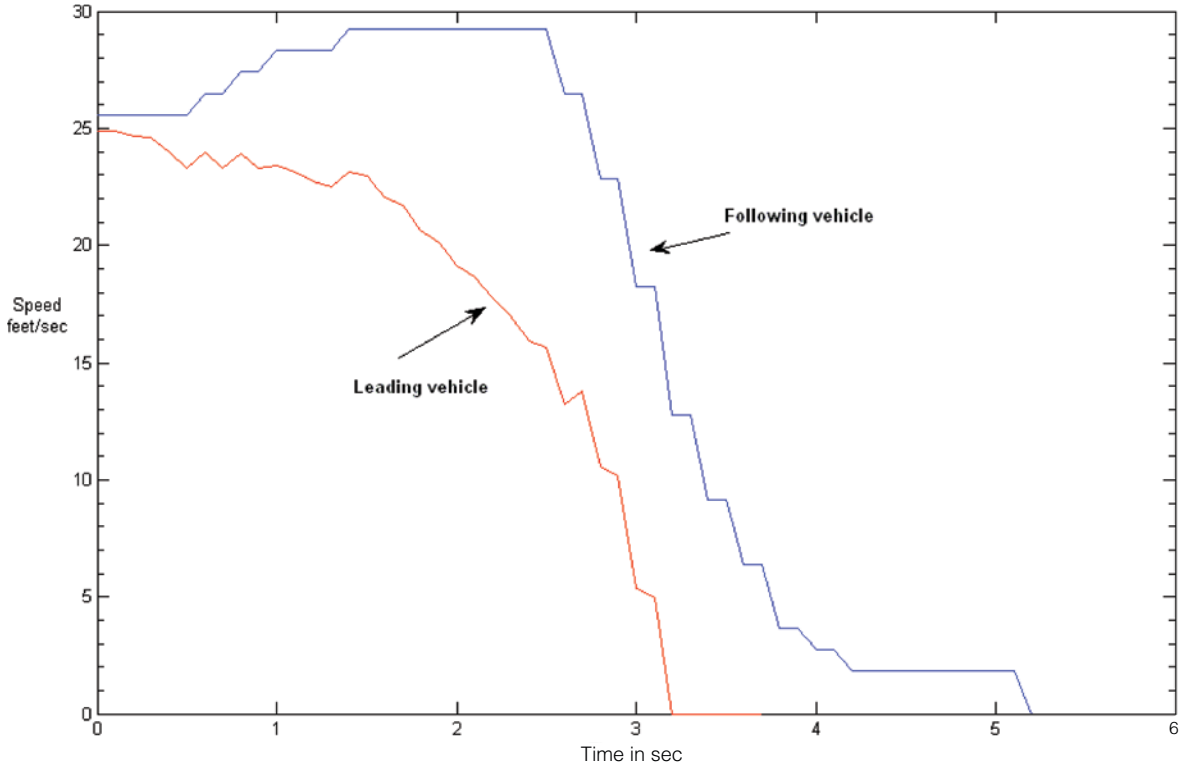


Figure 3.25. Speed trajectories of the leading and following vehicles (Case 104851).

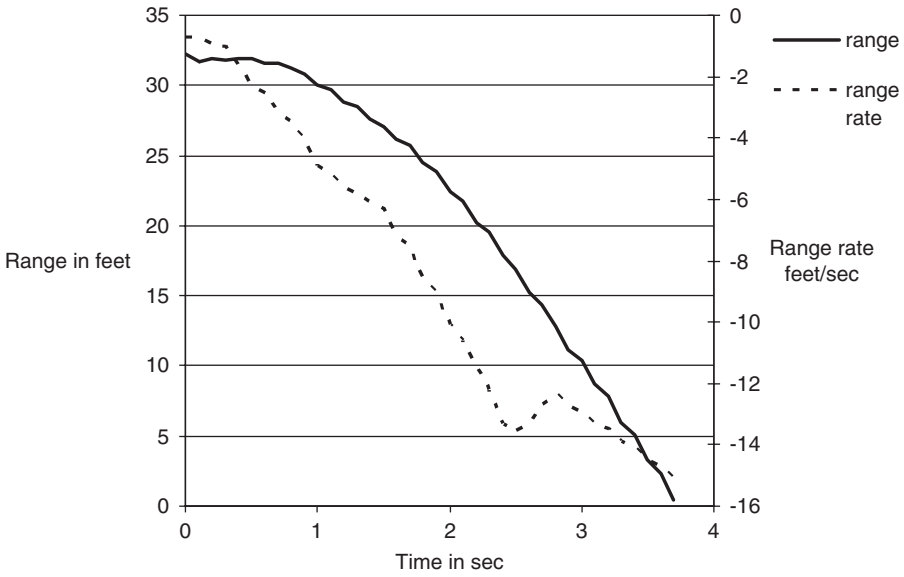


Figure 3.26. Range and range-rate data (Case 104851).

Table 3.6. WinBUGS Estimates for the Model Parameters for Case 104851

Variable	Mean	Standard Deviation	2.50%	Median	97.50%
Following Vehicle					
Initial speed (ft/s)	25.66	0.3869	24.92	25.66	26.45
First acceleration (ft/s^2)	1.567	0.2429	1.072	1.566	2.027
Second acceleration (ft/s^2)	–21.76	0.7101	–23.06	–21.79	–20.3
Third acceleration (ft/s^2)	–2.876	0.3197	–3.461	–2.891	–2.212
First change (s)	2.626	0.02057	2.572	2.629	2.657
Second change (s)	3.811	0.03753	3.755	3.806	3.891
Reaction time (s)	1.059	0.04373	0.9932	1.052	1.161
Leading Vehicle					
Initial speed (ft/s)	26.07	0.4079	25.27	26.07	26.88
First acceleration (ft/s^2)	–2.973	0.2788	–3.54	–2.971	–2.434
Second acceleration (ft/s^2)	–6.172	0.4611	–7.155	–6.153	–5.319
Third acceleration (ft/s^2)	–24.29	0.7389	–25.61	–24.34	–22.71
First change (s)	2.752	0.02641	2.674	2.755	2.793
Second change (s)	3.362	0.01218	3.339	3.361	3.387

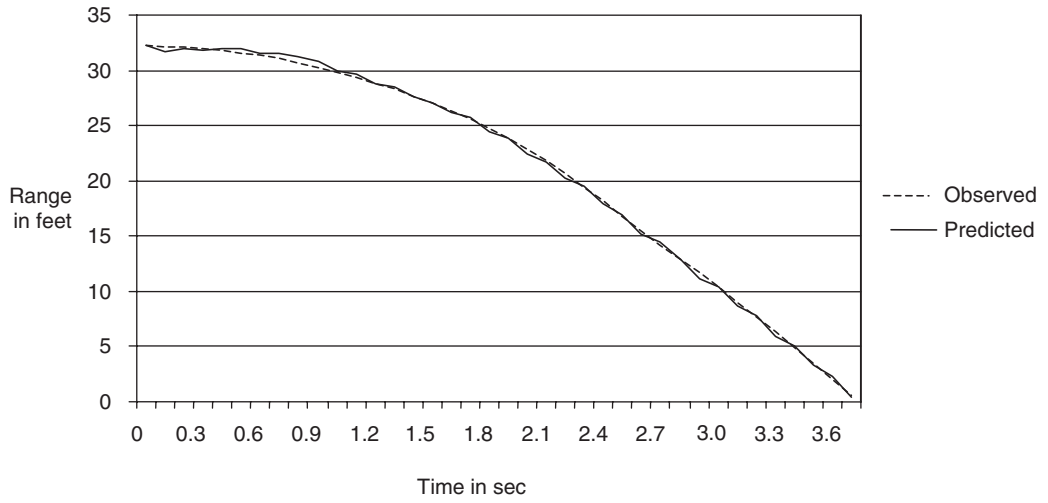


Figure 3.27. Predicted and observed range data (Case 104851).

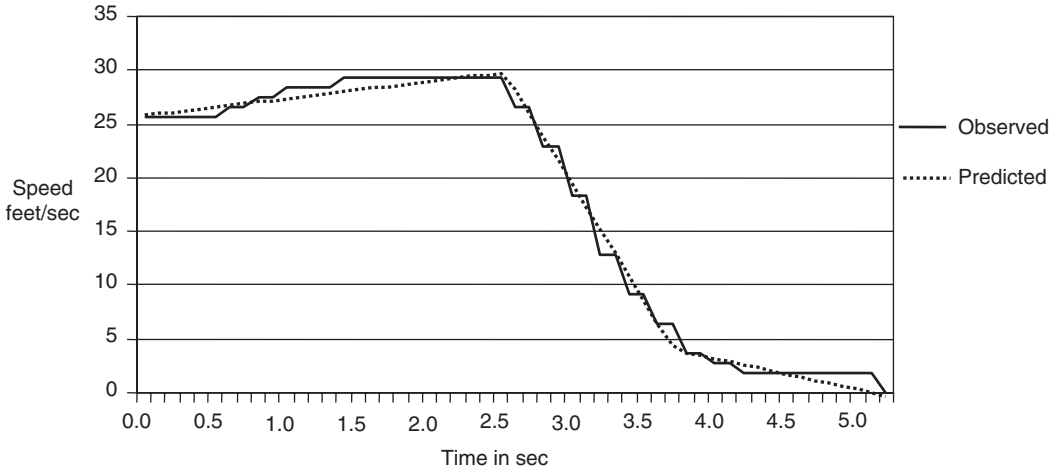


Figure 3.28. Predicted and observed instrumented speeds (Case 104851).

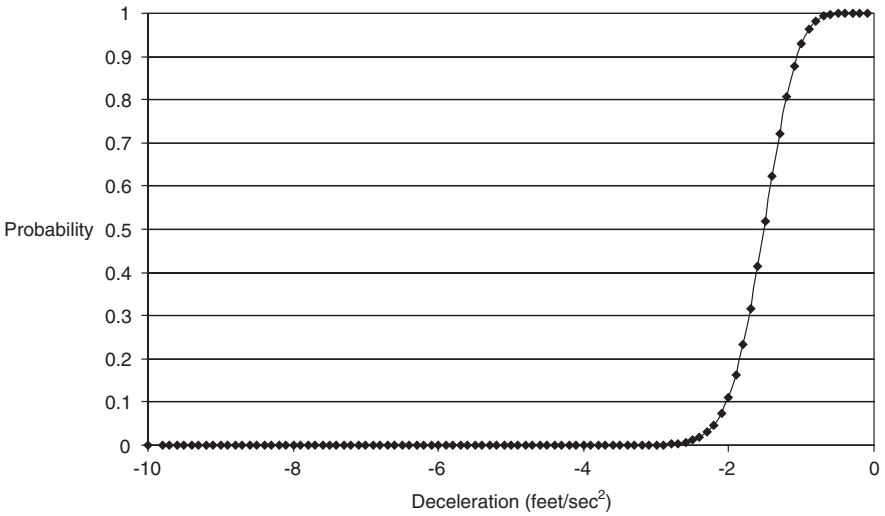


Figure 3.29. Counterfactual model (Case 104851).

then moved at almost constant speed for some time, followed by the final deceleration as a reaction to the leading vehicle's behavior. Range and range-rate data are shown in Figure 3.47.

The video shows that the following vehicle had to swerve around the lead vehicle to avoid a crash, which suggests a two-dimensional analysis. The focus here, however, is to extract as much information as possible from a one-dimensional approach.

Table 3.9 lists the WinBUGS estimates for the parameters. A three-stage model was developed for both the leading and the following vehicles. The most highlighted result that can be seen from the estimates is the very strong deceleration in the final stage. The leading vehicle had an initial speed of 40 ft/s and accelerated gently at 0.5 ft/s^2 for 7 s before decelerating at −14.76 ft/s^2 for approximately 2.3 s, followed by a more intense deceleration of −23 ft/s^2 and finally stopping. The following vehicle, with an initial speed of 33.75 ft/s, accelerated at 2.48 ft/s^2 for 4.23 s and then moved at almost zero acceleration for another 4.5 s before finally decelerating at −15.57 ft/s^2. Figure 3.48 shows the predicted and observed range values.

The counterfactual model for this case study is not shown because the following vehicle, aside from decelerating, swerved to avoid the crash.

Figure 3.30. Initial view from the instrumented vehicle traveling in the rightmost lane (Case 104283).

Figure 3.31. View near the merging section, with leading vehicle waiting for a gap (Case 104283).

Figure 3.32. Following vehicle just managed to stop, resulting in a near-crash scenario (Case 104283).

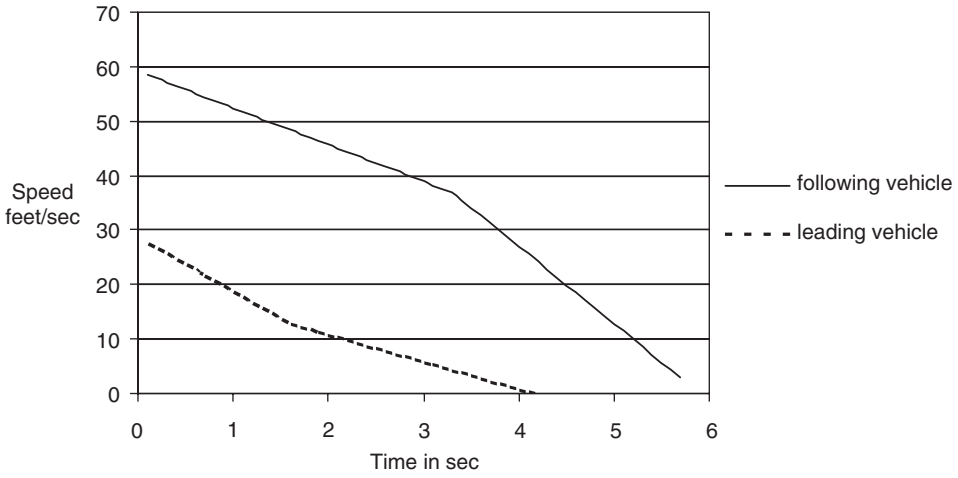


Figure 3.33. Speed trajectories for the leading and following vehicles (Case 104283).

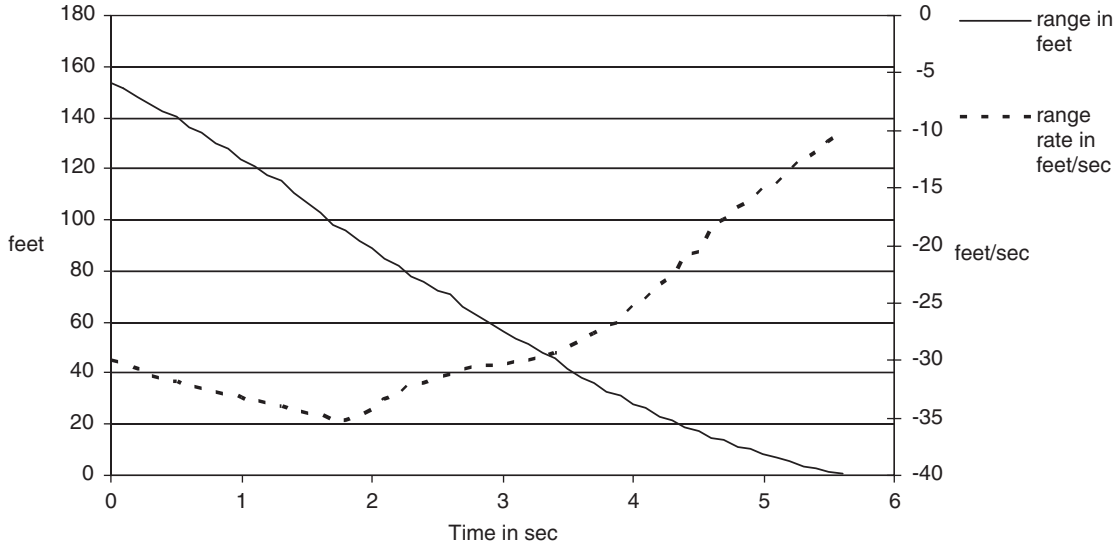


Figure 3.34. Range and range-rate data (Case 104283).

Table 3.7. WinBUGS Estimates for Case 104283

Variable	Mean	Standard Deviation	2.50%	Median	97.50%
Following Vehicle					
Initial speed (ft/s)	60.12	0.3074	59.54	60.12	60.74
First acceleration (ft/s^2)	–7.574	0.1257	–7.821	–7.573	–7.333
Second acceleration (ft/s^2)	–16.16	0.372	–16.86	–16.17	–15.41
First change (s)	3.922	0.04541	3.823	3.929	3.991
Time when stop (s)	5.805	0.027	5.754	5.804	5.861
Reaction time (s)	2.207	0.0607	2.083	2.209	2.322
Leading Vehicle					
Initial speed (ft/s)	30.93	0.437	30.12	30.91	31.81
First acceleration (ft/s^2)	–10.86	0.3323	–11.57	–10.84	–10.26
Second acceleration (ft/s^2)	–4.332	0.1572	–4.648	–4.331	–4.03
First change (s)	1.714	0.05528	1.602	1.716	1.816
Time when stop (s)	4.559	0.07132	4.424	4.557	4.704

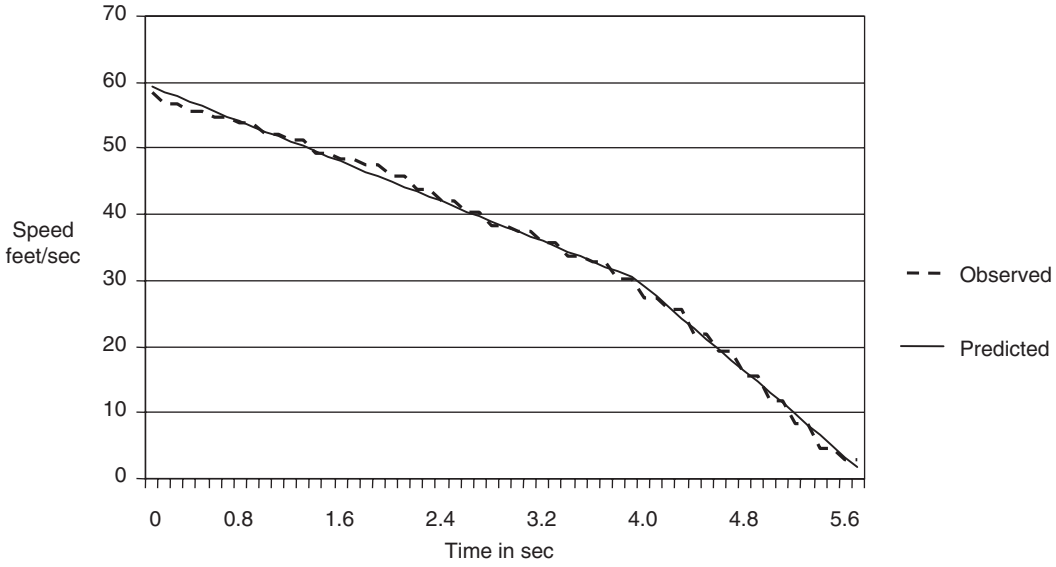


Figure 3.35. Predicted and observed instrumented speeds (Case 104283).

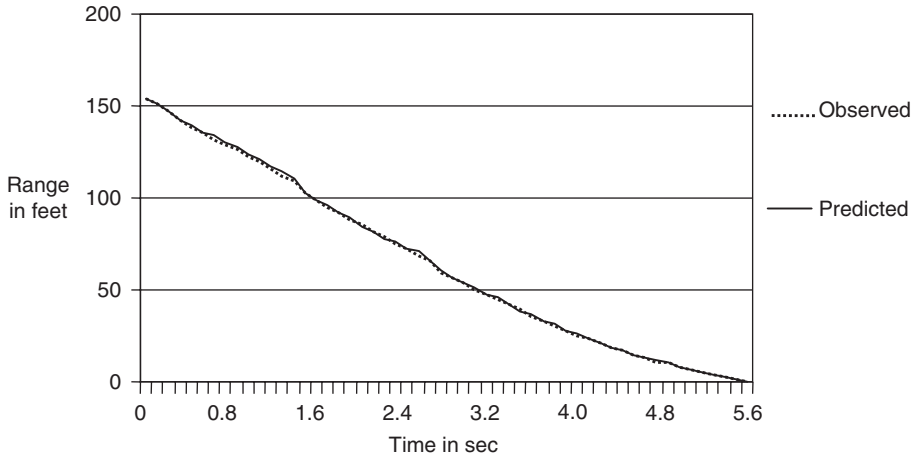


Figure 3.36. Predicted and observed range values (Case 104283).

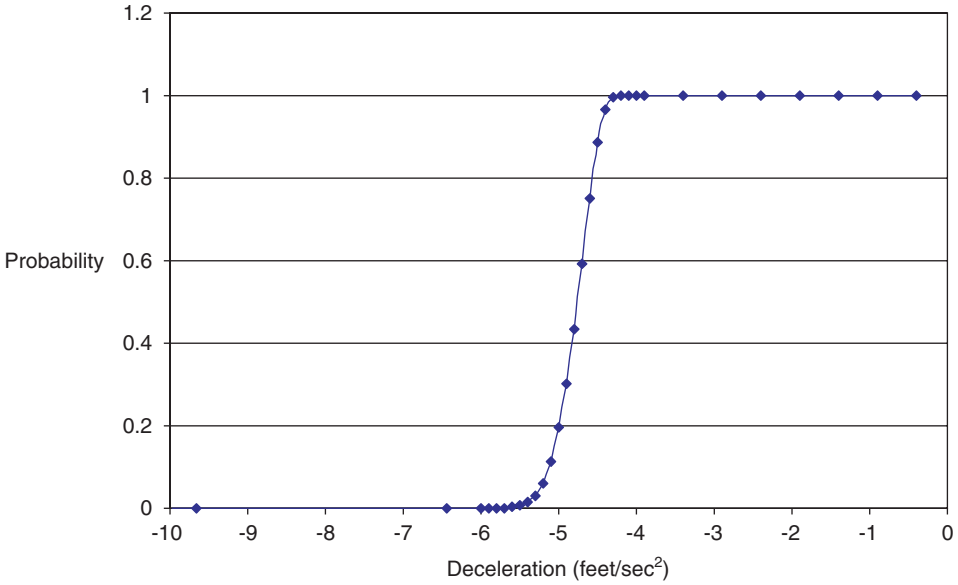


Figure 3.37. Counterfactual model (Case 104283).

Figure 3.38. Initial view of the leading vehicle (Case 60289).

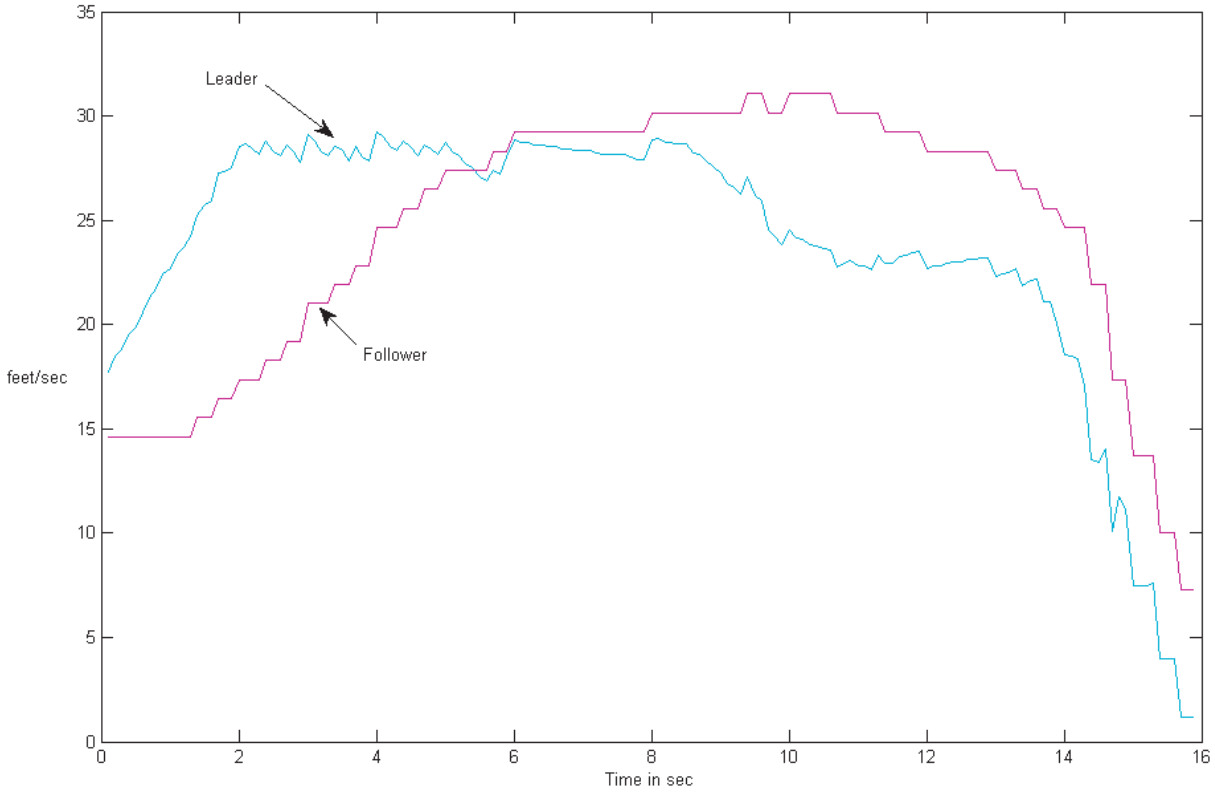


Figure 3.39. Speed trajectories of the leading and following vehicles (Case 60289).

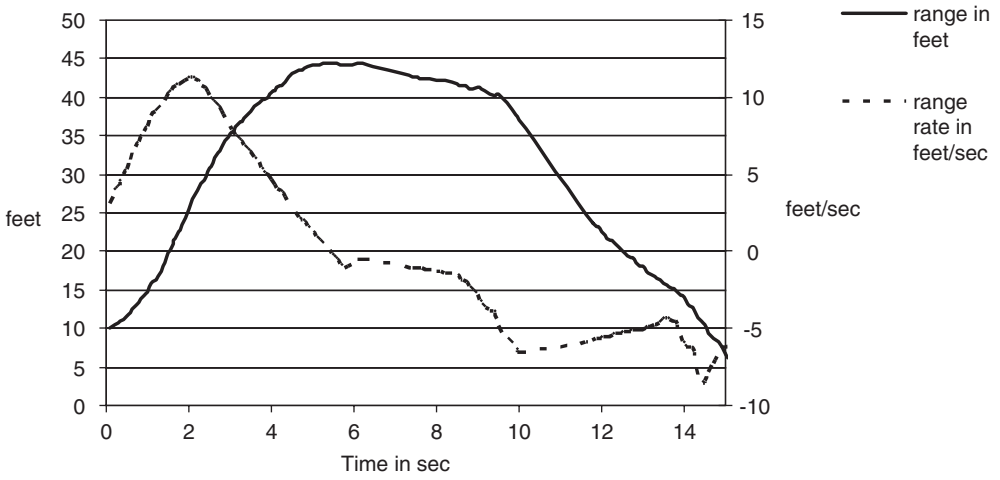


Figure 3.40. Range and range-rate data (Case 60289).

Table 3.8. WinBUGS Estimates for Case 60289

Variable	Mean	Standard Deviation	2.50%	Median	97.50%
Following Vehicle					
Initial speed (ft/s)	11.21	0.2399	10.74	11.21	11.67
First acceleration (ft/s^2)	3.149	0.06246	3.034	3.146	3.275
Second acceleration (ft/s^2)	0.7217	0.03928	0.6431	0.7219	0.7981
Third acceleration (ft/s^2)	–2.419	0.1075	–2.635	–2.415	–2.22
Fourth acceleration (ft/s^2)	–10.74	0.3316	–11.38	–10.74	–10.07
First change (s)	5.364	0.06755	5.236	5.362	5.495
Second change (s)	10.71	0.05586	10.61	10.71	10.83
Third change (s)	14.3	0.03184	14.23	14.3	14.36
Time when stopped (s)	16.47	0.05112	16.37	16.47	16.57
Reaction time (s)	0.4847	0.04604	0.3996	0.4829	0.5788
Leading Vehicle					
Initial speed (ft/s)	14.07	0.3193	13.46	14.07	14.71
First acceleration (ft/s^2)	8.16	0.2434	7.679	8.164	8.631
Second acceleration (ft/s^2)	–0.00214	0.04673	–0.09826	–0.00121	0.08024
Third acceleration (ft/s^2)	–1.45	0.07088	–1.591	–1.448	–1.312
Fourth acceleration (ft/s^2)	–9.502	0.3231	–10.12	–9.506	–8.857
First change (s)	1.764	0.03767	1.69	1.763	1.841
Second change (s)	8.277	0.1001	8.073	8.28	8.461
Third change (s)	13.81	0.03837	13.74	13.81	13.88

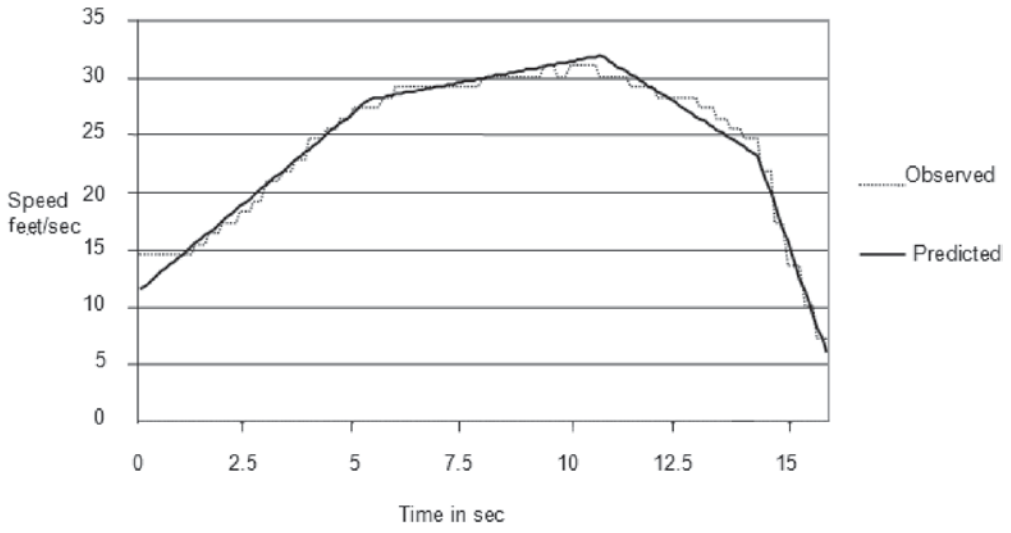


Figure 3.41. Predicted and observed instrumented speeds (Case 60289).

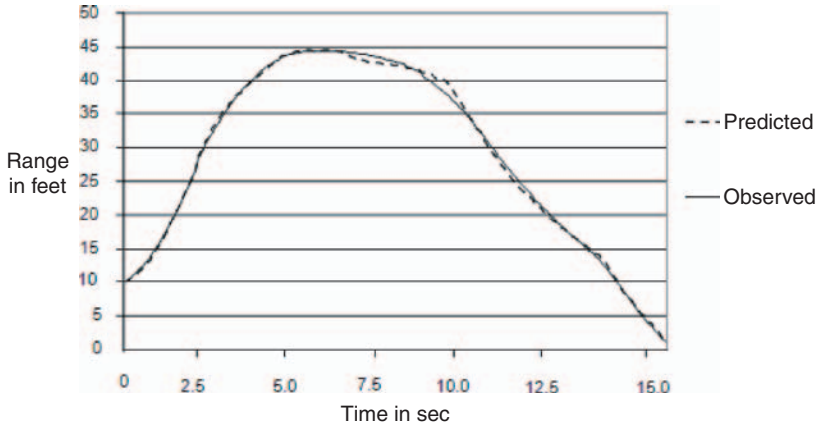


Figure 3.42. Predicted and observed range values (Case 60289).

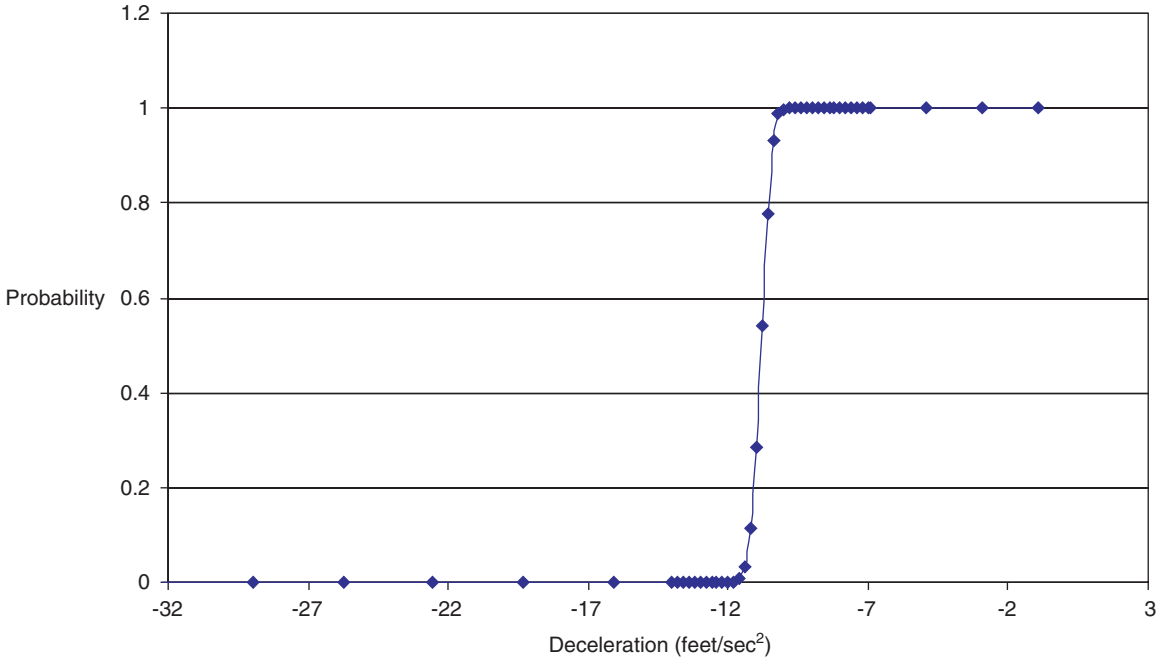


Figure 3.43. Counterfactual model (Case 60289).

Figure 3.44. Initial view of the leading vehicle on a two-lane, two-way highway (Case 92660).

Figure 3.45. Instrumented vehicle swerved to avoid collision (Case 92660).

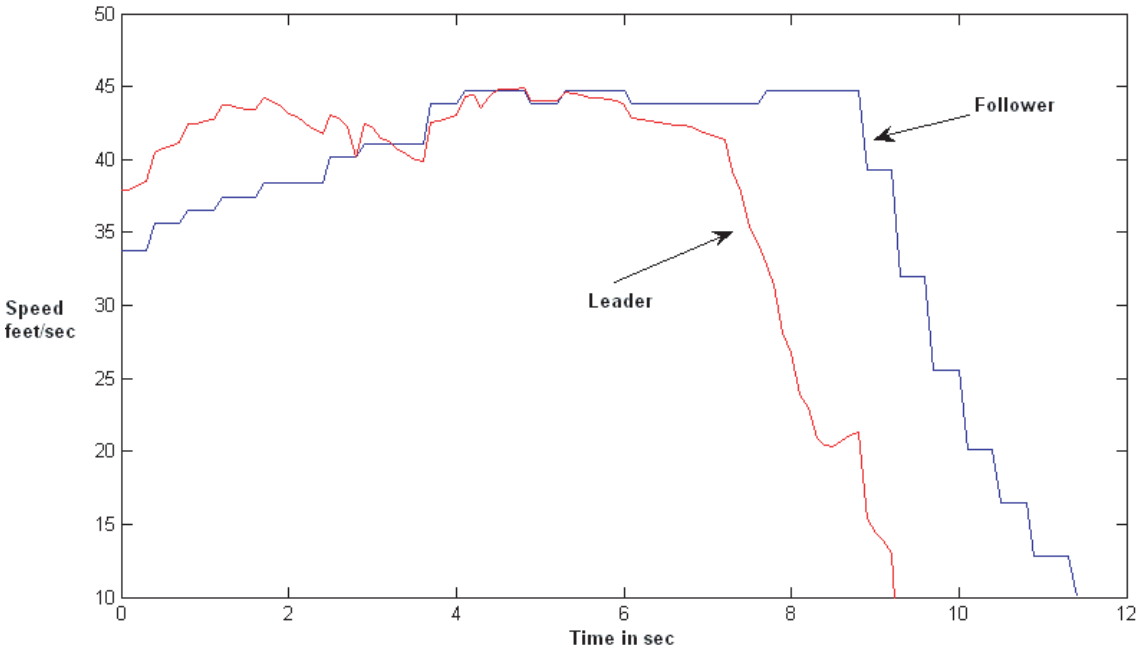


Figure 3.46. Speed trajectories of the leading and following vehicles (Case 92660).

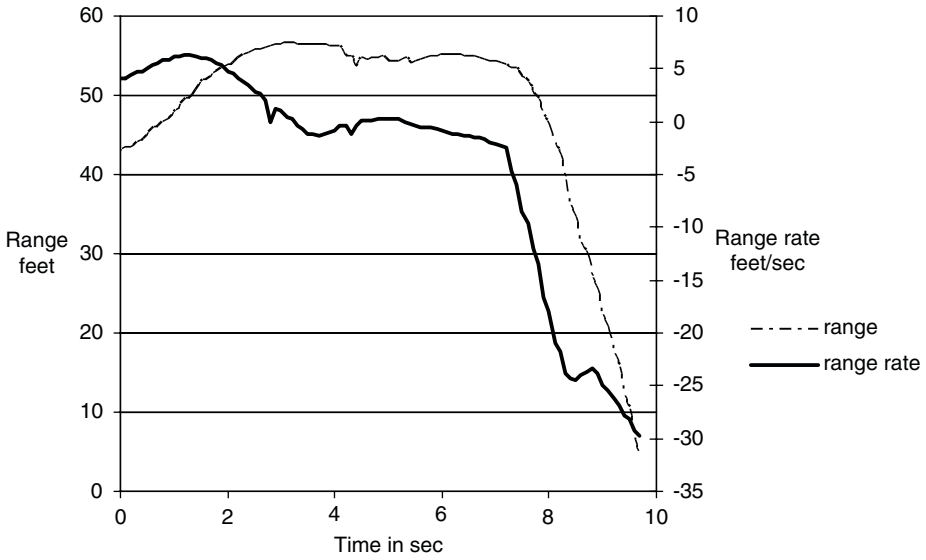


Figure 3.47. Range and range-rate data (Case 92660).

Table 3.9. WinBUGS Estimates for the Parameters for Case 92660

Variable	Mean	Standard Deviation	2.50%	Median	97.50%
Following Vehicle					
Initial speed (ft/s)	33.75	0.2246	33.33	33.74	34.2
First acceleration (ft/s^2)	2.479	0.07467	2.332	2.48	2.625
Second acceleration (ft/s^2)	0.1819	0.08929	0.01293	0.1802	0.3597
Third acceleration (ft/s^2)	–15.57	0.8826	–17.29	–15.57	–13.83
First change (s)	4.229	0.09666	4.036	4.232	4.407
Second change (s)	8.73	0.04538	8.635	8.734	8.812
Time when stop (s)	11.62	0.1311	11.38	11.61	11.9
Leading Vehicle					
Initial speed (ft/s)	40.98	0.2145	40.58	40.98	41.41
First acceleration (ft/s^2)	0.504	0.05306	0.3998	0.506	0.6038
Second acceleration (ft/s^2)	–14.76	0.4033	–15.55	–14.76	–13.98
Third acceleration (ft/s^2)	–23.62	3.443	–31.91	–23.15	–18.31
First change (s)	7.004	0.03893	6.928	7.005	7.077
Second change (s)	9.269	0.1305	9.003	9.267	9.51

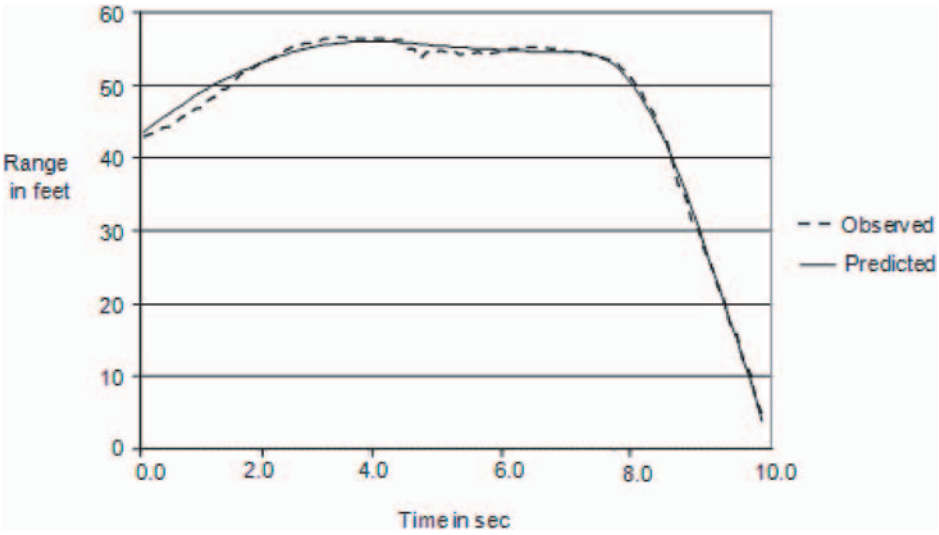


Figure 3.48. Predicted and observed range values (Case 92660).

CHAPTER 4

Analyses Using Site-Based Video Data

Data Acquisition and Preparation

The locations selected for analyses of site-based video data were two freeway segments on westbound Interstate 94 near downtown Minneapolis. One is approximately 500 ft long, located between Minnesota State Highway 65 and Portland Avenue, and the other is approximately 600 ft long, located between Portland Avenue and Park Avenue South. Figure 4.1 provides an overhead view of these two sites.

The traffic traveling on these two segments was recorded using two cameras installed by the Minnesota Traffic Observatory (MTO) on the roof of a 121-ft-high building near 3rd Avenue. The videos were transferred to the MTO and saved at a resolution of 640 × 480 pixels and 10 frames/s, from 11 a.m. to 8 p.m. daily. Given a target vehicle, its time-series of positions expressed in screen coordinates were manually extracted frame by frame using VideoPoint software (see Figure 4.2).

The screen coordinates obtained from VideoPoint were then converted to ground coordinates by first matching several reference points on the video images to corresponding points on high-resolution satellite photos and then applying standard photogrammetric methods.

I-94 Case 1

Description from the video: In this event, the following vehicle (Vehicle 2) and leading vehicle (Vehicle 1) were traveling in the middle lane (Figures 4.3 and 4.4). Vehicle 2 collided with Vehicle 1.

Trajectory data after the collision were not collected. Approximately 15 s of data were available from the video. Inspection of these data indicated that both vehicles were traveling at constant speeds approximately 9 s before the collision. Only the trajectory data from the last 6 s were used, and these are displayed in Figure 4.5.

Exploratory modeling for both vehicles was conducted using R software. For Vehicle 1, a two-stage model was fit, where a gentle acceleration lasting about 2.7 s was followed by a stronger deceleration. For Vehicle 2, a one-stage model was fit, where a gentle acceleration lasted until the collision. Bayes estimates for each vehicle's initial speed, the time points at which the accelerations changed, and the accelerations in all stages were computed using WinBUGS. These results are displayed in Table 4.1.

At the beginning of the study period, Driver 2 was traveling at 24.21 ft/s, while Driver 1 was traveling at 27.5 ft/s. Driver 1 accelerated at 1.6 ft/s^2 for about 2.7 s and then decelerated at −8.779 ft/s^2 for about 3.1 s. Driver 2 was accelerating at about 1.16 ft/s^2 until the collision. No evidence from video or trajectory data showed that Driver 2 decelerated to prevent a collision. Figures 4.6 and 4.7 show the distance trajectories of Vehicles 1 and 2, along with the distance trajectories predicted by the models.

I-94 Case 2

Description from the video: In this event, the following vehicle (Vehicle 2) and leading vehicle (Vehicle 1) were traveling in the middle lane (Figures 4.8 and 4.9). Vehicle 2 collided with Vehicle 1.

Approximately 4 s of data were available from the video. The trajectory data used for analysis are displayed in Figure 4.10.

Exploratory modeling for both vehicles was conducted using R software. For Vehicle 1, a two-stage model was fit, where a strong deceleration lasted about 1.5 s and was followed by a less strong deceleration. For Vehicle 2, a one-stage model was fit, where a gentle acceleration lasted until the collision. Bayes estimates are displayed in Table 4.2.

At the beginning of the study period, Driver 2 was traveling at 33.02 ft/s, while Driver 1 was traveling at 33.12 ft/s.

(text continues on page 44)

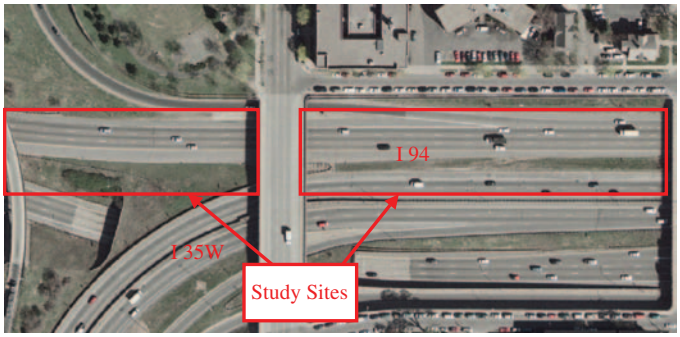


Figure 4.1. Satellite view of two distance-based trajectory-data-collection sites.

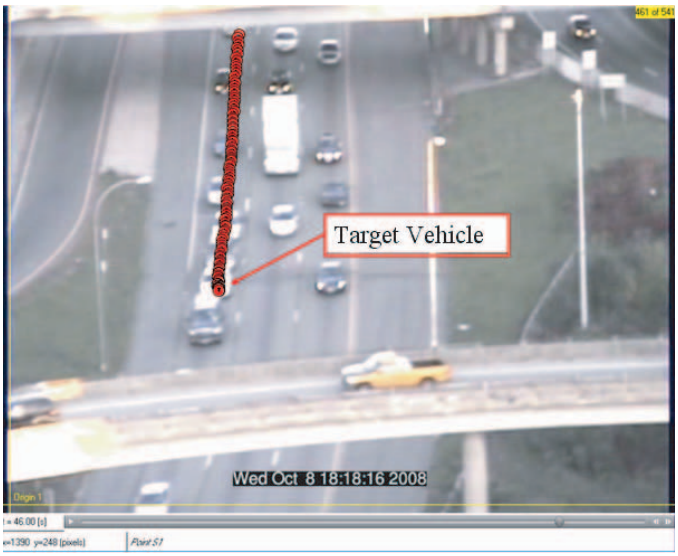


Figure 4.2. Illustration of VideoPoint trajectory extraction.

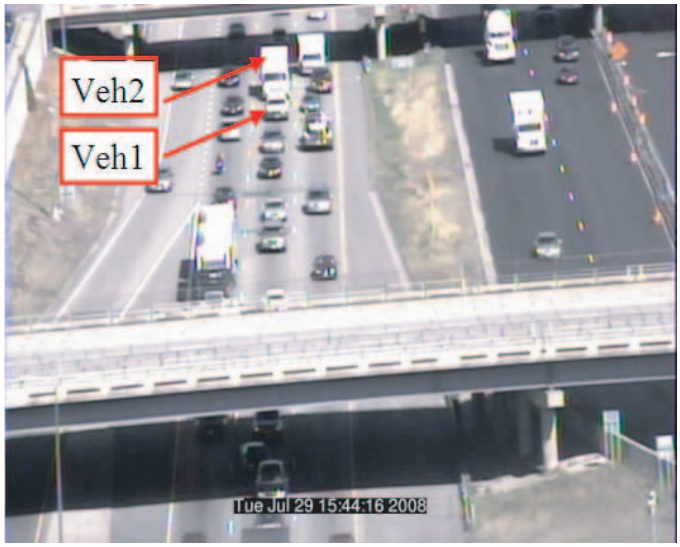


Figure 4.3. View at the time the two vehicles enter the study segment (I-94 Case 1).

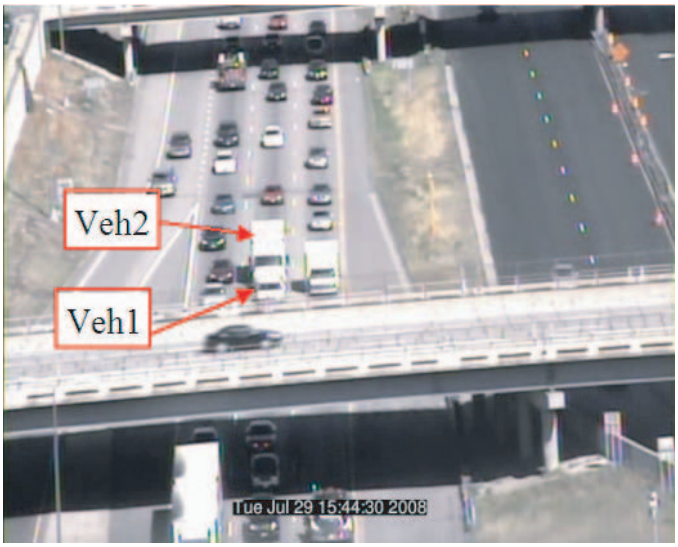


Figure 4.4. View at the time of collision (I-94 Case 1).

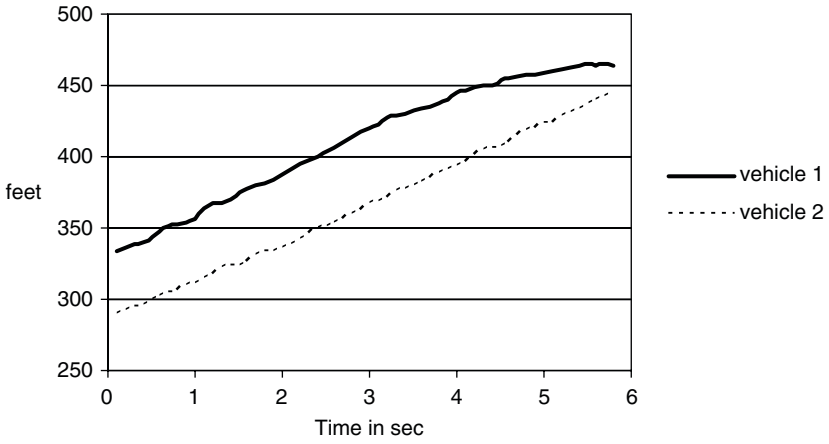


Figure 4.5. Distance trajectory data (I-94 Case 1).

Table 4.1. WinBUGS Estimates for I-94 Case 1

Variable	Mean	Standard Deviation	2.5%ile	97.5%ile
Vehicle 2 (following)				
Initial speed (ft/s)	24.21	0.1731	23.87	24.56
First acceleration (ft/s^2)	1.155	0.07787	0.9997	1.306
Vehicle 1 (leading)				
Initial speed (ft/s)	27.5	0.5015	26.46	28.45
First acceleration (ft/s^2)	1.643	0.4585	0.8183	2.615
Second acceleration (ft/s^2)	–8.779	0.4278	–9.656	–7.983
First change (s)	2.663	0.1136	2.433	2.883

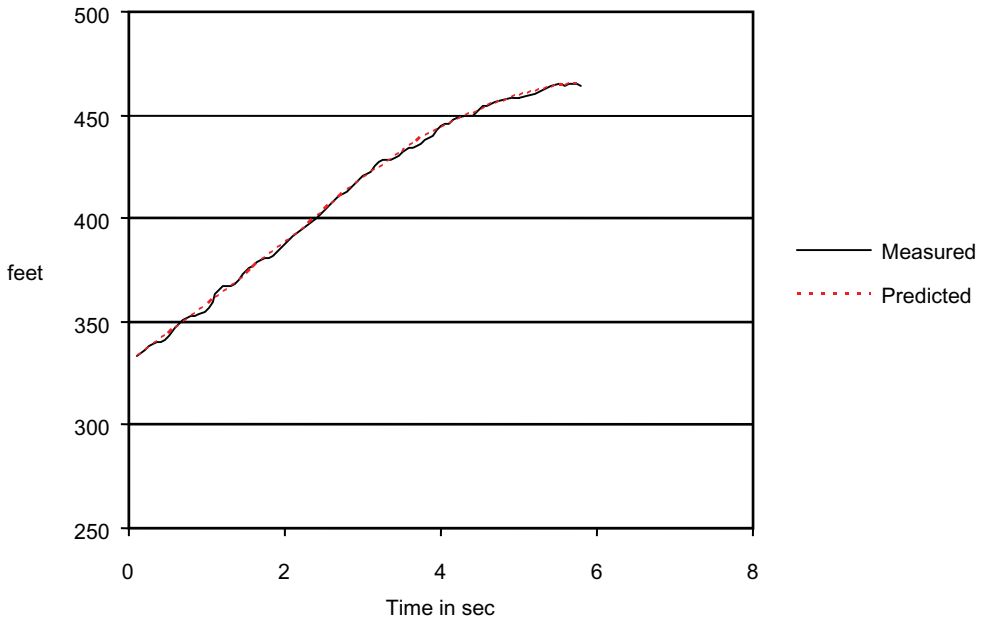


Figure 4.6. Measured and modeled Vehicle 1 distance trajectories (I-94 Case 1).

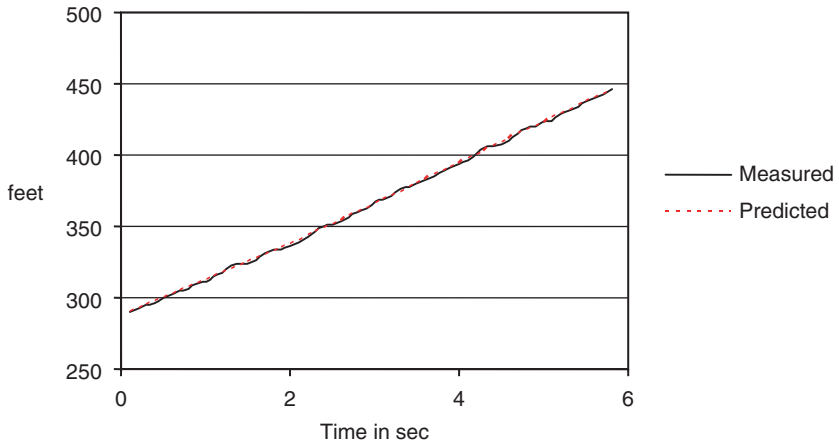


Figure 4.7. Measured and modeled Vehicle 2 distance trajectories (I-94 Case 1).

Figure 4.8. View at the time the two vehicles enter the study segment (I-94 Case 2).

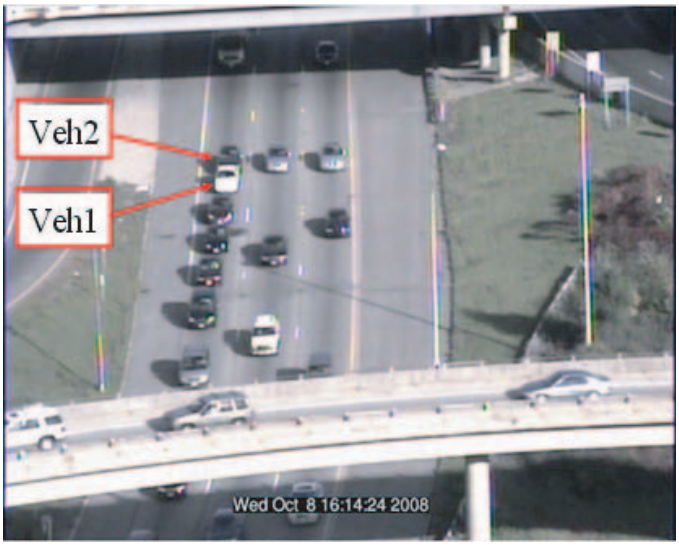


Figure 4.9. View at the time of collision (I-94 Case 2).

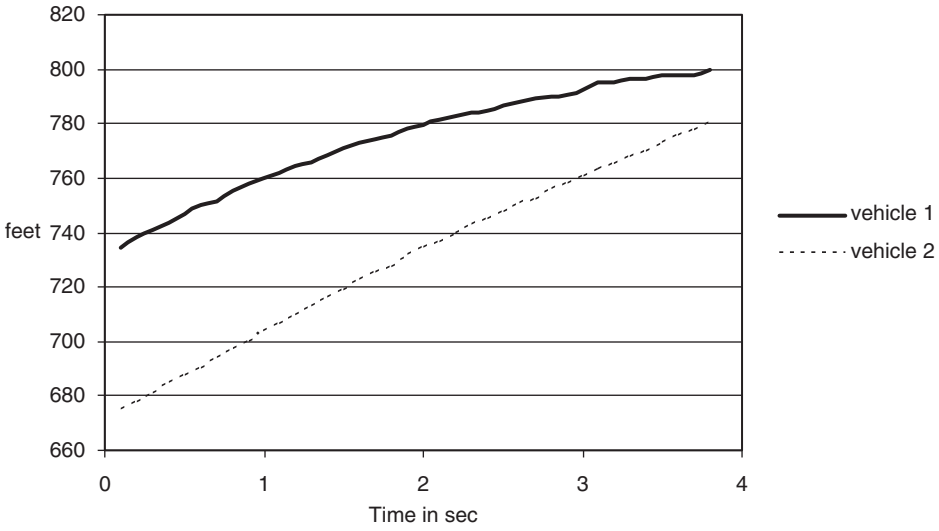


Figure 4.10. Distance trajectory data (I-94 Case 2).

Table 4.2. WinBUGS Estimates for I-94 Case 2

Variable	Mean	Standard Deviation	2.5%ile	97.5%ile
Vehicle 2 (following)				
Initial speed (ft/s)	33.02	0.2069	32.62	33.43
First acceleration (ft/s^2)	–2.512	0.1425	–2.793	2.231
Vehicle 1 (leading)				
Initial speed (ft/s)	33.12	0.8424	31.89	35.13
First acceleration (ft/s^2)	–10.52	1.911	–15.11	–8.507
Second acceleration (ft/s^2)	–5.436	0.5798	–6.377	–4.099
First change (s)	1.505	0.3309	0.8608	2.134

(continued from page 40)

Driver 1 was decelerating at –10.52 ft/s^2 for about 1.5 s and then decelerated at –5.436 ft/s^2 for about 2.3 s. Driver 2 was decelerating at –2.512 ft/s^2 but was not able to prevent the collision. Figures 4.11 and 4.12 show the distance trajectories of Vehicles 1 and 2, along with the distance trajectory predicted by the models.

I-94 Case 3

Description from the video: In this event, three vehicles are traveling in the right lane of the study segment (Figures 4.13 and 4.14). Vehicle 3 collides with Vehicle 2.

Approximately 7 s of data were available from the video. The trajectory data used for analysis are displayed in Figure 4.15.

Exploratory modeling for each vehicle was conducted using R software. For Vehicle 1, a three-stage model was fit, where a gentle acceleration lasting for about 2.3 s was followed by a 2.4-s stronger deceleration and then by a 2.4-s gentler deceleration. For Vehicle 2, a two-stage model was fit, where a gentle deceleration lasting for about 3.5 s was followed by a 3.2-s stronger deceleration. For Vehicle 3, a three-stage model was fit. The trajectory modeling showed that Driver 3 decelerated in all three stages. Bayes estimates are displayed in Table 4.3. The reaction time of Driver 2 was calculated as the time difference between Driver 2's first change point and Driver 1's first change point. The reaction time of Drivers 3 was calculated as the time difference between Driver 3's second change point and Driver 2's first change point.

At the beginning of study period, Drivers 1, 2, and 3 were traveling at 21.33 ft/s, 33.42 ft/s, and 51.29 ft/s, respectively. Driver 1 accelerated at 1.048 ft/s^2 for about 2.2 s at first, then decelerated at –8.8 ft/s^2 for about 2.3 s, and then decelerated at –0.8968 ft/s^2 for about 2.4 s. Driver 2 was decelerating at –0.8968 ft/s^2 for about 3.5 s and then decelerated at –8.0 ft/s^2 for about 3.2 s. Vehicle 3 was apparently initially traveling at a much higher speed (51.29 ft/s) than the other two vehicles.

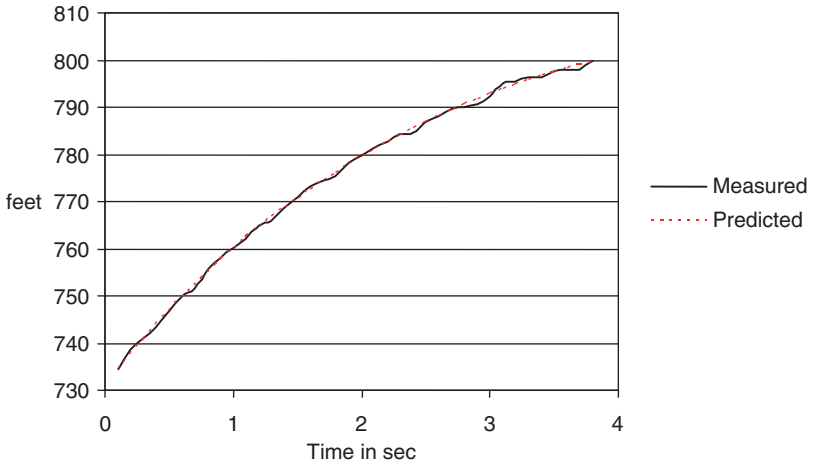


Figure 4.11. Measured and modeled Vehicle 1 distance trajectories (I-94 Case 2).

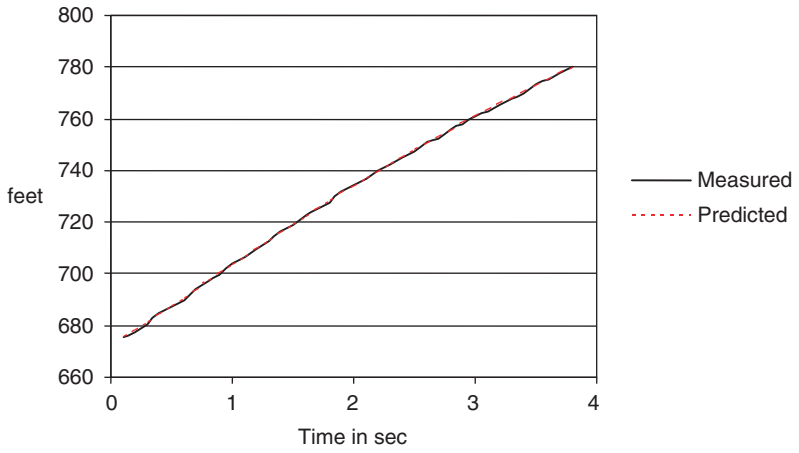


Figure 4.12. Measured and modeled Vehicle 2 distance trajectories (I-94 Case 2).

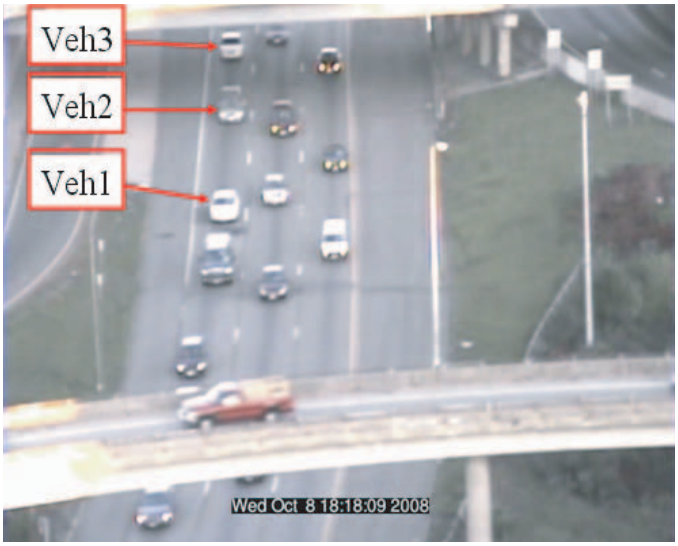


Figure 4.13. View at the time the three vehicles enter the study segment (I-94 Case 3).

Figure 4.14. View at the time of collision (I-94 Case 3).

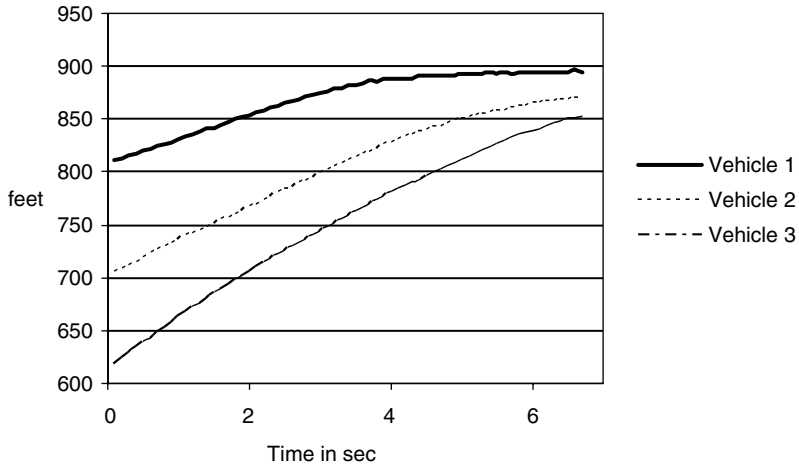


Figure 4.15. Distance trajectory data (I-94 Case 3).

Table 4.3. WinBUGS Estimates for I-94 Case 3

Variable	Mean	Standard Deviation	2.5%ile	97.5%ile
Vehicle 3				
Initial speed (ft/s)	51.29	0.3508	50.66	52.03
First acceleration (ft/s^2)	–5.618	0.3798	–6.475	–4.992
Second acceleration (ft/s^2)	–3.206	0.2493	–3.598	–2.617
Third acceleration (ft/s^2)	–10.24	2.002	–14.88	–7.215
First change (s)	2.295	0.3079	1.726	2.926
Second change (s)	5.375	0.2048	4.924	5.743
Reaction time (s)	1.886	0.2197	1.412	2.285
Vehicle 2				
Initial speed (ft/s)	33.42	0.1933	33.03	33.80
First acceleration (ft/s^2)	–0.8968	0.1368	–1.156	–0.6174
Second acceleration (ft/s^2)	–8.004	0.2241	–8.463	–7.576
First change (s)	3.489	0.0792	3.333	3.645
Reaction time (s)	1.231	0.1336	0.9572	1.489
Vehicle 1				
Initial speed (ft/s)	21.33	0.3953	20.53	22.07
First acceleration (ft/s^2)	1.048	1.048	0.2296	1.984
Second acceleration (ft/s^2)	–8.8	0.5076	–9.974	–7.967
Third acceleration (ft/s^2)	–0.9801	0.3604	–1.535	–0.1625
First change (s)	2.258	0.1116	2.046	2.49
Second change (s)	4.595	0.1188	4.353	4.819

Although Driver 3 noticed that the traffic in front was slowing down and started decelerating before entering the data-collection segment, Driver 3 was still not able to avoid collision with Vehicle 2 after decelerating at –10.24 ft/s^2 for about 2.3 s in its last stage.

Figures 4.16, 4.17, and 4.18 show the distance trajectories of Vehicles 1, 2, and 3, along with the distance trajectory predicted by the models.

In this case, to assess the avoidability of collision between Vehicles 2 and 3, probabilities of collision between these vehicles were computed as a function of counterfactual final decelerations (the third stage) of Vehicle 3. This relationship is displayed in Figure 4.19.

The probability of collision between Vehicles 1 and 2 was also evaluated, assuming that Driver 2 was decelerating at different rates in the last stage. This relationship is displayed in Figure 4.20. In this case, since the relative speed between Vehicles 1 and 2 in the last stages is small, for most counterfactual deceleration rates of Vehicle 2, the corresponding probabilities of collision are either 1 or 0.

I-94 Case 4

Description from the video: In this event, three vehicles were traveling in the right lane of the study segment (Figures 4.21 and 4.22). Vehicle 3 collided with Vehicle 2.

Approximately 10 s of data were available from the video. The trajectory data used for analysis are displayed in Figure 4.23.

Exploratory modeling for both vehicles was conducted using R software. A three-stage model was fit to each vehicle trajectory. Vehicle 1 accelerated for about 2.6 s, which was followed first by a 4.1-s strong deceleration and then by a 3.7-s gentle deceleration. The behavior of Drivers 2 and 3 was almost the same as that of Driver 1 but with stronger deceleration in their last stages. The WinBUGS estimates are shown in Table 4.4.

(text continues on page 51)

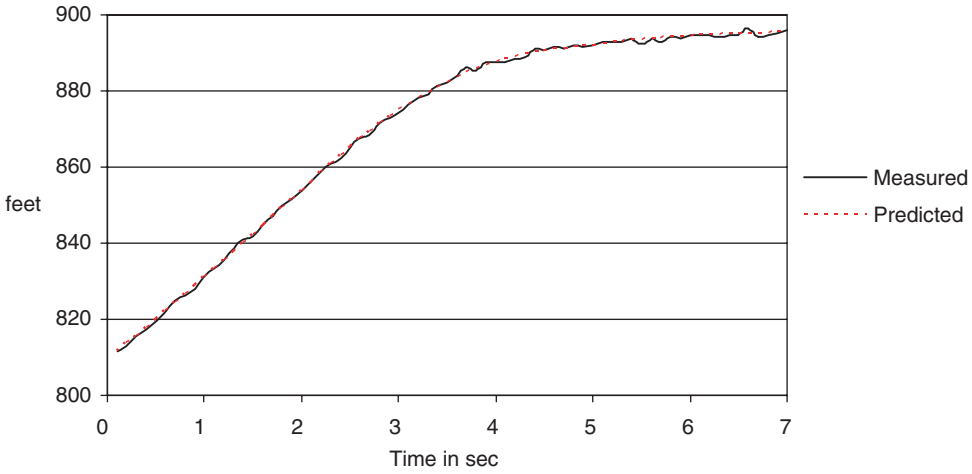


Figure 4.16. Measured and modeled Vehicle 1 distance trajectories (I-94 Case 3).

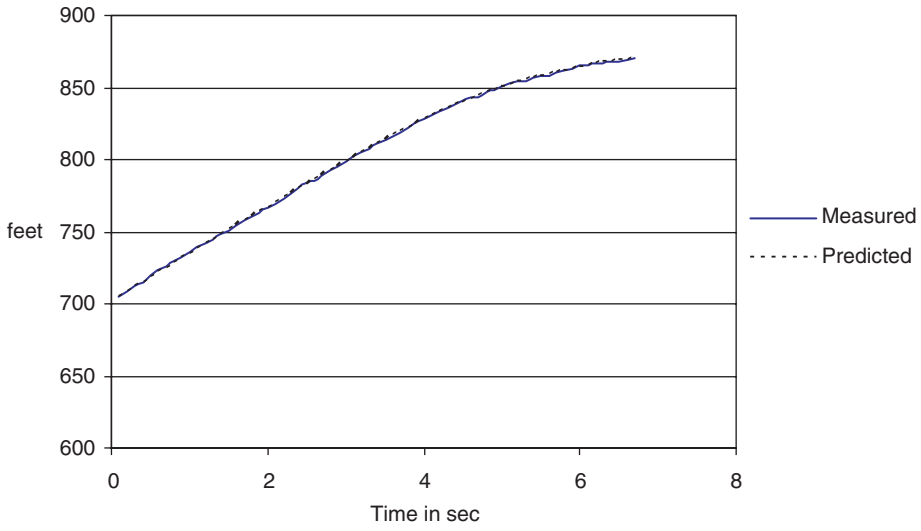


Figure 4.17. Measured and modeled Vehicle 2 distance trajectories (I-94 Case 3).

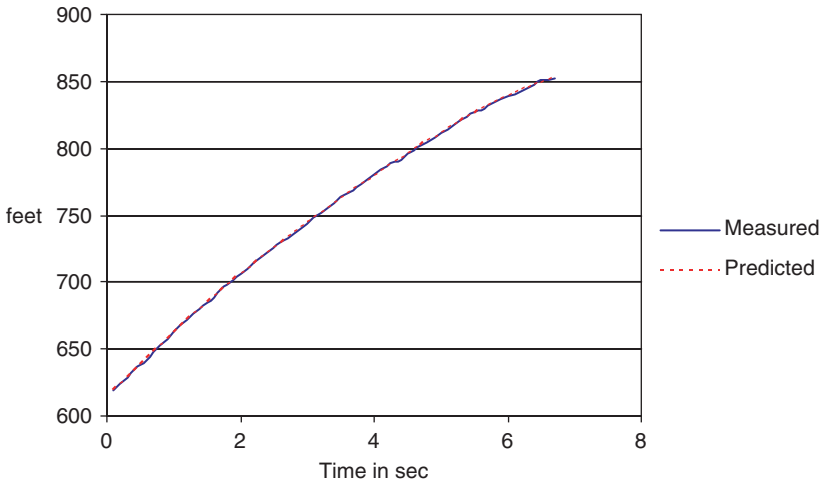


Figure 4.18. Measured and modeled Vehicle 3 distance trajectories (I-94 Case 3).

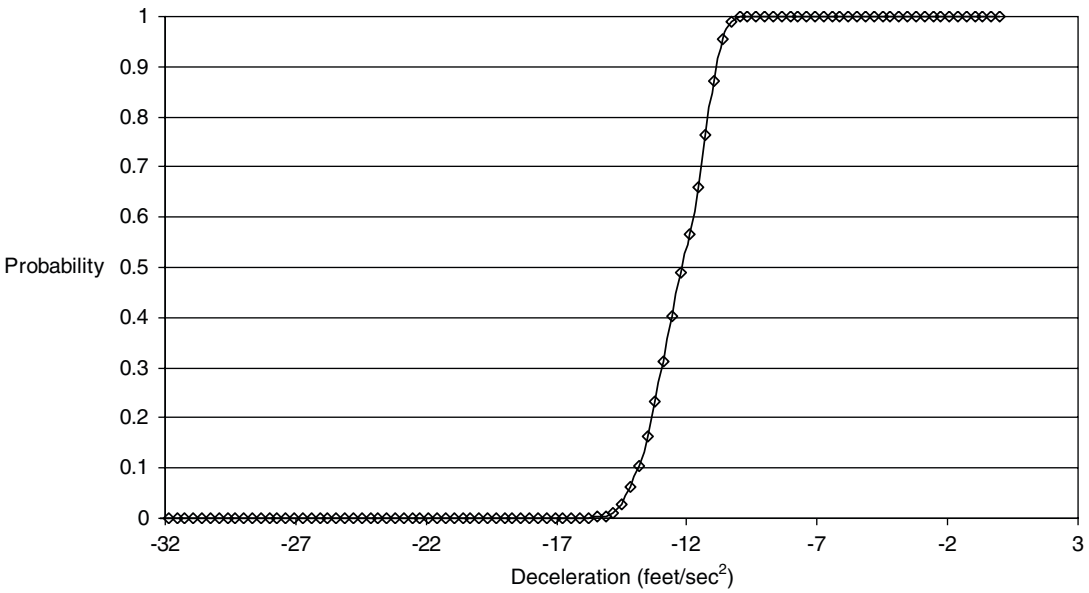


Figure 4.19. Probability of collision as a function of counterfactual final deceleration by Driver 3 (I-94 Case 3).

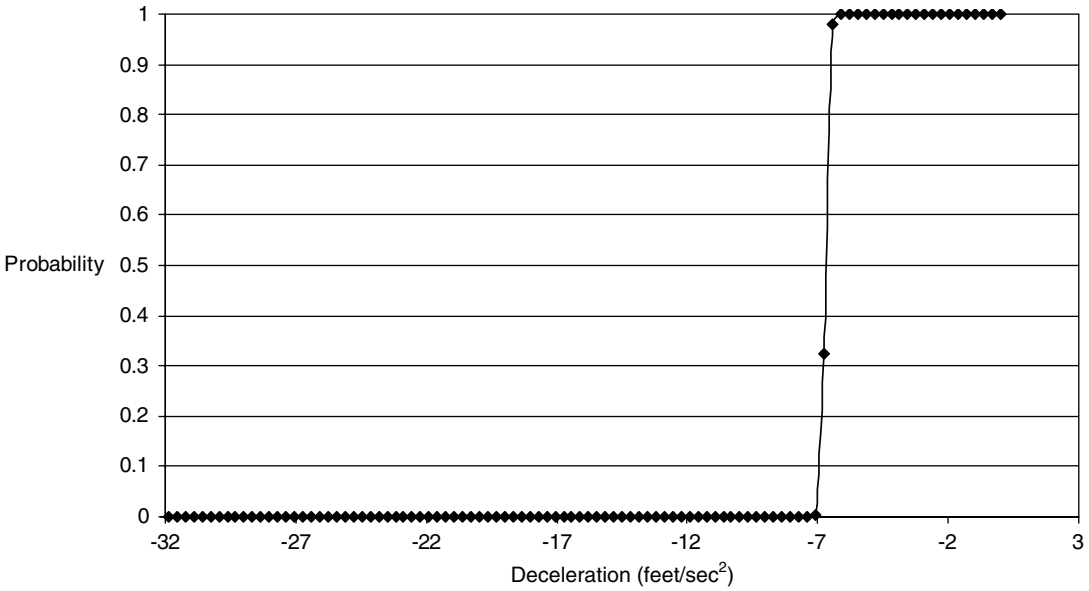


Figure 4.20. Probability of collision as a function of counterfactual final deceleration by Driver 2 (I-94 Case 3).

Figure 4.21. View at the time the three vehicles enter the study segment (I-94 Case 4).

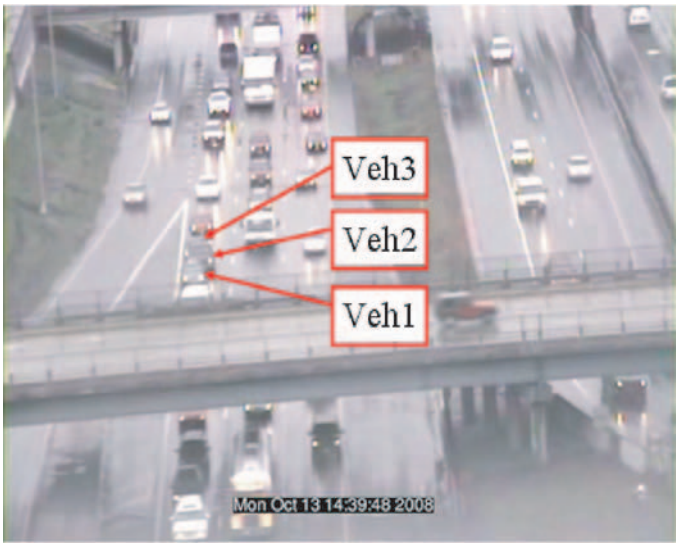


Figure 4.22. View at the time of collision between Vehicles 2 and 3 (I-94 Case 4).

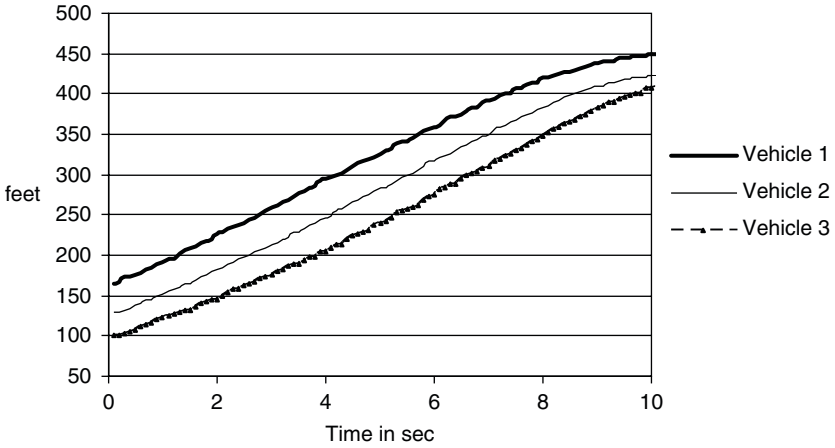


Figure 4.23. Distance trajectory data (I-94 Case 4).

Table 4.4. WinBUGS Estimates for I-94 Case 4

Variable	Mean	Standard Deviation	2.5%ile	97.5%ile
Vehicle 3				
Initial speed (ft/s)	21.94	0.164	21.61	22.25
First acceleration (ft/s^2)	2.591	0.06958	2.46	2.735
Second acceleration (ft/s^2)	–0.8846	0.5106	–1.874	0.1521
Third acceleration (ft/s^2)	–12.89	3.821	–21.83	–7.752
First change (s)	5.9	0.2449	5.336	6.334
Second change (s)	8.723	0.2545	8.171	9.163
Reaction time (s)	1.112	0.2855	0.5099	1.62
Vehicle 2				
Initial speed (ft/s)	25.67	0.3143	25.0	26.21
First acceleration (ft/s^2)	2.514	0.2232	2.164	3.012
Second acceleration (ft/s^2)	–0.2509	0.2551	–0.7828	0.2243
Third acceleration (ft/s^2)	–10.48	0.6891	–11.95	–9.231
First change (s)	3.769	0.3356	3.092	4.372
Second change (s)	7.612	0.1186	7.382	7.847
Reaction time (s)	0.8966	0.1646	0.5863	1.223
Vehicle 1				
Initial speed (ft/s)	29.41	0.5065	28.27	30.27
First acceleration (ft/s^2)	2.247	0.5051	1.52	3.485
Second acceleration (ft/s^2)	–0.7111	0.1872	–1.073	–0.3352
Third acceleration (ft/s^2)	–7.312	0.306	–7.923	–6.729
First change (s)	2.625	0.3319	1.913	3.216
Second change (s)	6.715	0.1183	6.475	6.938

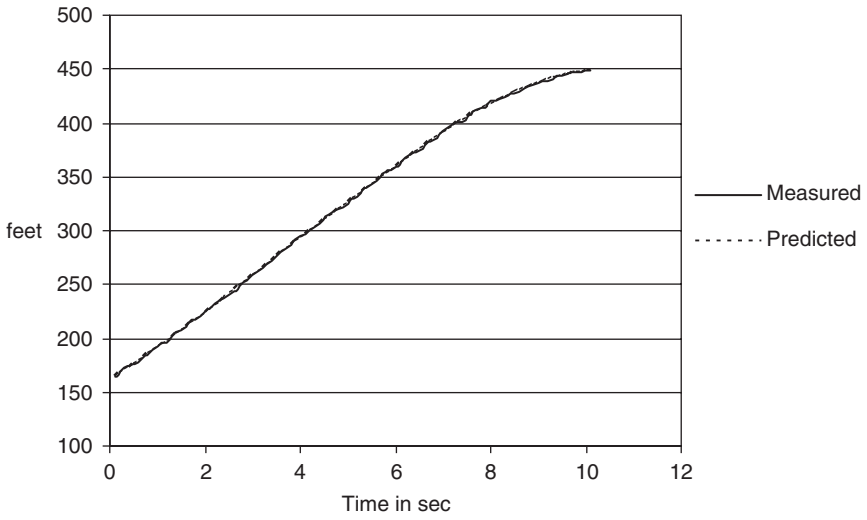


Figure 4.24. Measured and modeled Vehicle 1 distance trajectories (I-94 Case 4).

(continued from page 46)

The reaction times of Drivers 2 and 3 were calculated as the time difference of last change points between Vehicles 2 and 1, and between Vehicles 3 and 2.

At the beginning of the study period, Drivers 1, 2, and 3 were traveling at about 29.41 ft/s, 25.67 ft/s, and 21.94 ft/s, respectively. Driver 1 first accelerated at 2.247 ft/s^2 for about 2.6 s, then decelerated at –0.7111 ft/s^2 for about 4.1 s, and then decelerated at –7.312 ft/s^2 for about 3.7 s. Driver 2 was accelerating at 2.514 ft/s^2 for about 3.8 s, decelerated at –0.2509 ft/s^2 for about 3.8 s, and then decelerated at –10.48 ft/s^2 to avoid colliding with Vehicle 1. Driver 3 noticed the strong deceleration of Vehicle 2 in its last stage. Although Driver 3 had the strongest final deceleration (at –12.89 ft/s^2) for about 1.3 s, this was not sufficient to prevent the collision.

Figures 4.24, 4.25, and 4.26 show the distance trajectories of Vehicles 1, 2, and 3, respectively, along with the distance trajectories predicted by the models.

In this case, the relationship between a counterfactual final deceleration of Vehicle 3 and the probability of collision with Vehicle 2 is displayed in Figure 4.27. The probability of collision between Vehicle 1 and 2 was also evaluated, assuming Driver 2 was decelerating at a different rate in the last stage. This relationship is displayed in Figure 4.28.

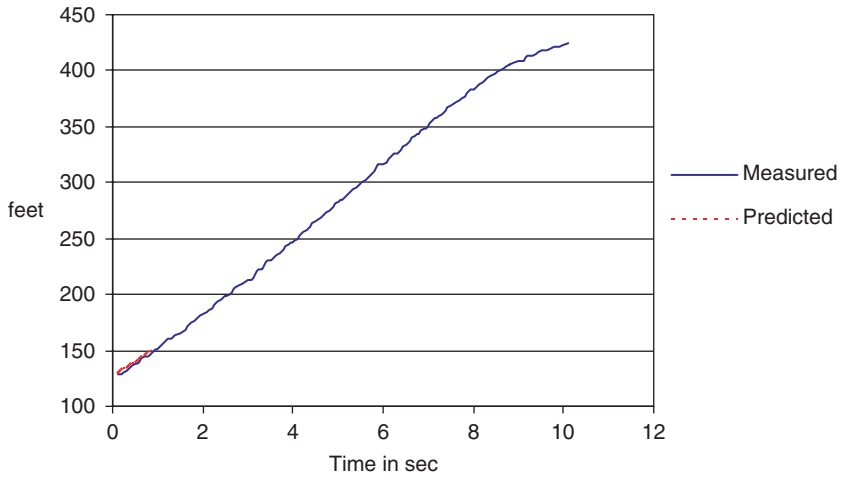


Figure 4.25. Measured and modeled Vehicle 2 distance trajectories (I-94 Case 4).

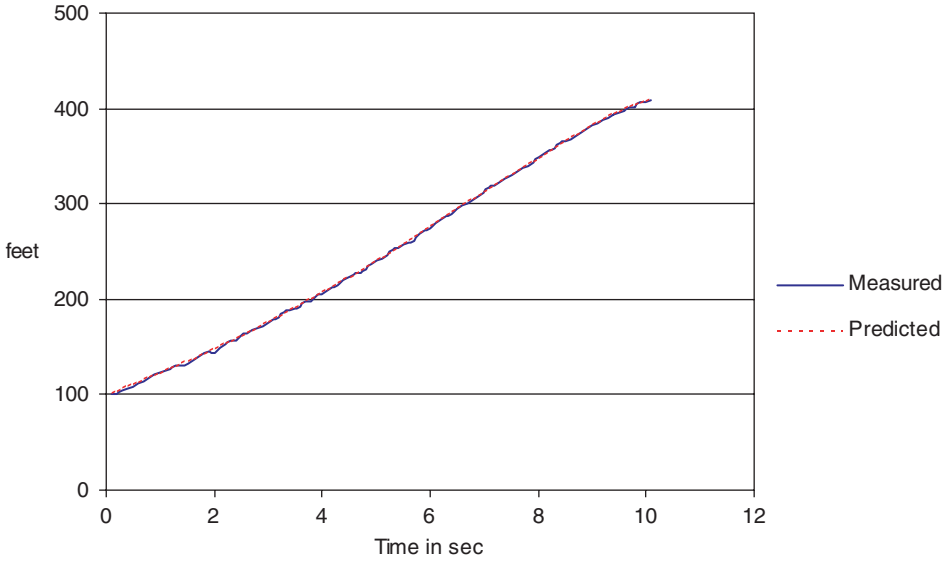


Figure 4.26. Measured and modeled Vehicle 3 distance trajectories (I-94 Case 4).

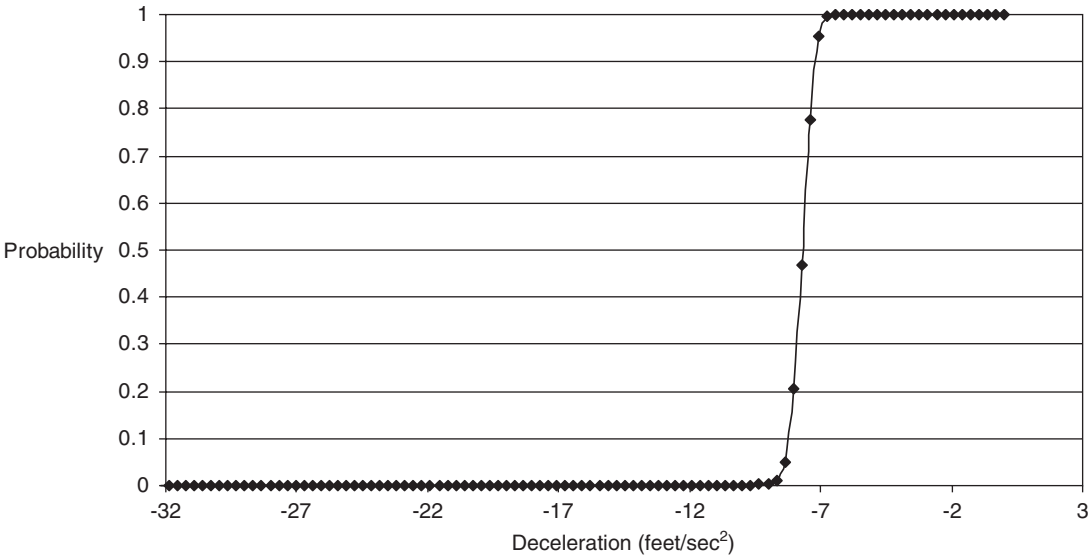


Figure 4.27. Probability of collision as a function of counterfactual final deceleration by Driver 3 (I-94 Case 4).

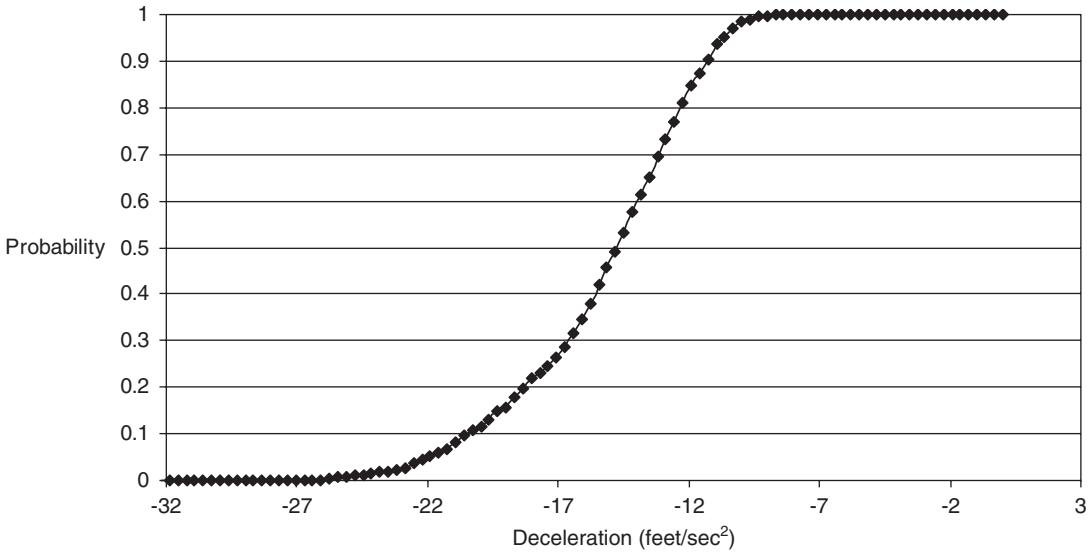


Figure 4.28. Probability of collision as a function of counterfactual final deceleration by Driver 2 (I-94 Case 4).

CHAPTER 5

Analyses Using CICAS Site-Based System

General Methodology

One of the goals for this project was to explore the utility and usability of existing data in traffic safety analysis, specifically in the investigation of surrogate measures of safety. The approach requires vehicle trajectories. The data collected from the Cooperative Intersection Collision Avoidance Systems (CICAS) project intersections, at least at first glance, looked like a promising source of information. This infrastructure was designed for the purposes of the CICAS project, however, and certain features constrain the usefulness of these data for more general research purposes. In adherence to the principal objective of the SHRP 2 Safety Project S01, effort was devoted to assessing the potential utility of this extensive data source. This was done both to test the project's core methodology and to identify possible future improvements to the data-collection infrastructure. This chapter summarizes the findings of the analysis and the problems encountered, along with their proposed solutions, and concludes with an exploratory analysis of one near-crash event.

During the course of this project, all CICAS project sites were explored as potential data sources. The Minnesota Highway 52 site, although it has more recorded crash cases, deploys older technology; the North Carolina US-74E site has better sensors and corrects some errors in the collection and postprocessing of data. These were the main reasons for focusing the project's efforts on the data originating from the North Carolina site. The site is illustrated in Figure 5.1; US-74 is the main road and Strawberry Boulevard is the side road. The goal was to attempt to isolate near-crash events and use the structural-model methodology to investigate the background and input conditions characterizing a crash or near-crash event and their relation to the occurrence or nonoccurrence of a collision. The data harvesting procedure involved the extraction of possible cases from the CICAS database, use of filters for the removal of false positives, and visual inspection of the resulting trajectories. It is important to note that in contrast to the other data sources used in this project, the magnitude of the data available from the CICAS site made necessary the development of automated methods for event data extraction and the creation of utilities that accelerate manual inspection of the extracted information.

The CICAS vehicle trajectory database structure and available information is described in Appendix B. In summary, the information available for each vehicle includes x, y coordinates in the state plane system, speed, and acceleration measurements, all at 10 Hz. It is vital to understand that these are not the raw collected measurements from the site but are derived from the raw measurements provided by the instrumentation during a data-reduction stage. From the detection hardware used, it is known that the sensor radar-based measurements for main-approach vehicles are range, range rate, and azimuth at 10 Hz, with similar information collected from the LIDAR sensors for the side-road vehicle (although it is unclear at what interval). The postprocessing of the raw data includes filtering as well as projections to known roadway features that introduce some difficulties for this study. For example, the position of the vehicle is always projected onto the lane centerline, resulting in some loss of information regarding the transition during lane changes, but more importantly during evasive actions taken by the drivers to avoid collisions. In addition, after closer inspection of the data, the project team concluded that the side-road vehicle position is not updated at 10 Hz, since the provided data generate a stepwise trajectory profile. Speed and acceleration for the side-road vehicle display apparently unrealistic behavior, so they were excluded from this investigation.

Figure 5.1 shows the North Carolina site for the CICAS project. The CICAS database codes events based on side-road vehicle maneuvers, along with the gap duration from the main-road vehicle movements. This database postprocessing makes possible the harvesting of completed cross-mainline maneuvers, where the accepted gaps of the side-road crossing vehicle may have been small. This involves finding a plausible

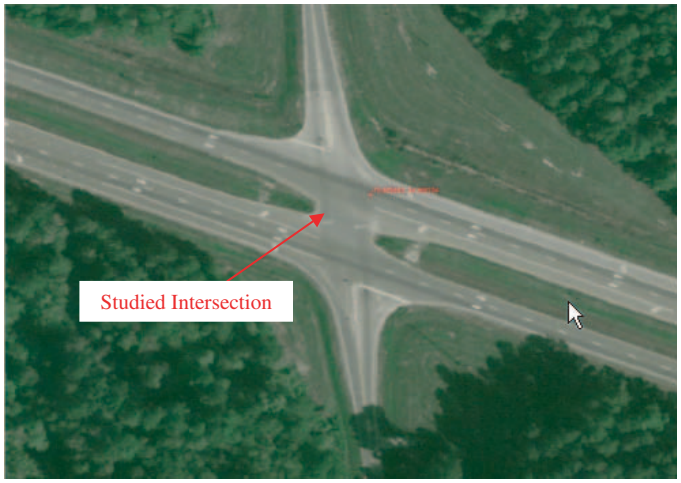


Figure 5.1. CICAS project at North Carolina site.

time envelope of small, accepted lag times during the cross-mainline maneuver over the complete lifetime of the tracked vehicle. The search was separated by direction for the two cross-main-road maneuvers (one for each direction). The technique is illustrated for one of the maneuvers. Essentially, it involves two inner join steps. The first step is the inner join of all tracked targets associated with small gap events for the given maneuver within the observation lifetime of the tracked target that completed the maneuver (i.e., when the target was first and last seen by the system):

```
CREATE VIEW closecalls AS SELECT targetid, maneuver,
    veh_class, sub1_acpt_lag_zone_exit_date, sub1_acpt_lag_
    zone_exit_time
FROM vehicle_acpt_lag
WHERE (maneuver = 1) AND ((3.0 > ANY(sub1_acpt_lags))
    OR (3.0 > ANY(sub2_acpt_lags)));

CREATE VIEW lifetime AS SELECT vt.targetid, vt.zone_
    first_seen, vt.date_first_seen, vt.time_first_seen, vt.zone_
    last_seen, vt.date_last_seen, vt.time_last_seen
FROM vehicle_time AS vt
INNER JOIN closecalls AS cc ON (vt.targetid = cc.targetid);
```

The second step uses the lifetime of this target to harvest all tracked targets present during the evolution of the accepted lag times for the maneuver.

```
SELECT *
FROM tracked_targets AS tt
INNER JOIN lifetime AS ltm ON
(tt.date = ltm.date_first_seen) AND
(tt.time >= ltm.time_first_seen) AND
(tt.time <= ltm.time_last_seen);
```

The preceding procedure returned close to 3,000 cases. These are possible near-crash events. To try to reduce the false positives, a filtering program was developed.

The first step in the filtering program is to determine the predominate direction of travel for all the tracked vehicles. Essentially, this can be achieved by keeping track of the lane number as the target progresses in time. Second, since the primary interest is in events representing near-crash cases between a side-road vehicle and main-road vehicle, after the first sort, the side-road vehicles are examined. For each side-road vehicle, main-road vehicles are scanned at the same time step and run through a rejection process. In this process, if any test fails to return at least one possible conflicting main-road vehicle, the time step is advanced. The process is summarized in the following steps:

Step 1: If no mainline vehicles are present for the given time step, then the time step is advanced.
Step 2: A distance between the main-road vehicles and the side-road vehicle must lie below a specified threshold.
Step 3: The speed of the side-road vehicle must be above a specified minimum threshold.
Step 4: The speed of the main-road vehicle must be above a specified minimum threshold.

The rejection at each step can be stored in a table that contains the date/time and target IDs. The level of elimination depends on the specified speed/distance parameters of the trajectories. The result of the filtering program contained 297 cases of near crashes similar to a recorded crash case.

A sample trajectory of a main-road tracked vehicle can be seen in Figure 5.2 for one of the resulting maneuvers. Note that as long as the lateral position relative to the center lane is within a certain margin, the target is snapped to the lane. Generally, what has been determined thus far is that the main-road tracked vehicle trajectories seem plausible. Unfortunately, this is not the case with the trajectories of the side-road vehicle. Although when the complete maneuver is plotted in space, the trajectories seem plausible (Figure 5.3), when animated in time, it becomes clear that peculiar tracking errors exist at critical locations during the maneuver. Figure 5.4 illustrates one such case, which has been typical of what has been observed thus far.

The side-road tracked vehicle completed a northbound cross maneuver over US-74 (aligned east-west in Figure 5.3). It can be reasonably assumed that the vehicle never reverses course during the maneuver. Under such circumstances, the displacement curve from the start of the maneuver to the last tracked location should always monotonically increase. As seen in Figure 5.4, particularly at points *A*, *B*, and *C*, this is not the case.

The side-road vehicle seems to make jumps backward, which is unrealistic. More important, these jumps do not seem to be

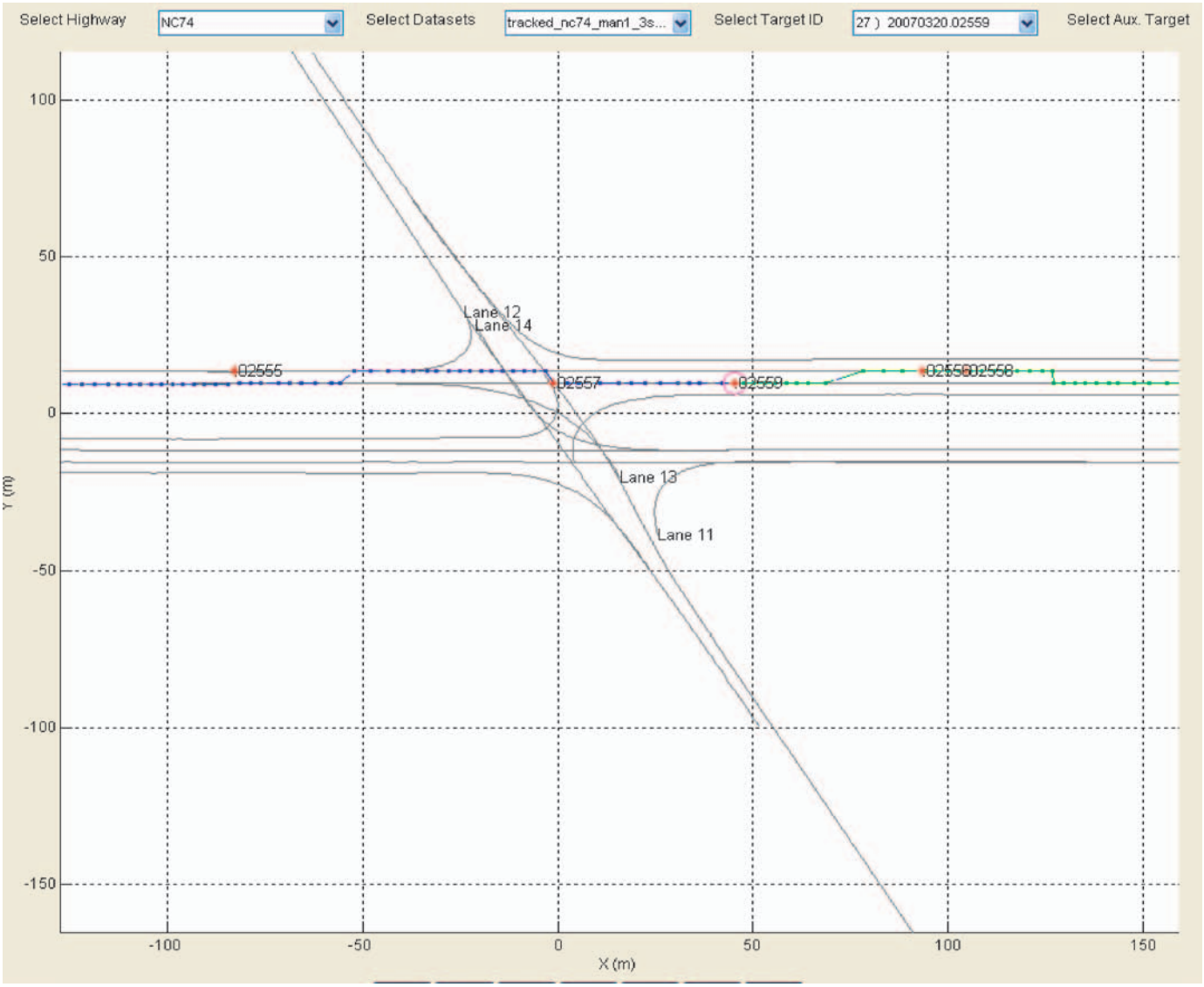


Figure 5.2. Tracked target interface showing trajectories of a main-road vehicle.

isolated points on the overall trajectory, since the progression seems realistic following these course corrections. This last element renders difficult the correction of these trajectories even manually. After close inspection of many cases, numerous hypotheses have been formed on the nature of these discrepancies. Specifically, the errors seem to concentrate in areas where the side-road vehicle is stationary or a main-road vehicle is nearby. For the portable site deployments such as North Carolina, side-road vehicles are essentially tracked through the intersection using horizontal LIDAR units placed at each approach of the side road and at a location in the median. The height of the LIDAR horizontal scan is set to approximate the height of the vehicle bumper. The LIDAR scanner sweep could have missed the bumper, occlusions from other vehicles, sensor alignment (as the tracked vehicle passes from one sensor data-collection field to another), or other algorithm processing factors may also contribute to the error. Of course, such anecdotal evidence is not enough to pinpoint the source of the difficulty and certainly not enough to allow the creation of automated correction methodologies, which, in any case, are outside the scope of this project. The aforementioned issues constrained the project's ability to process large numbers of possible near crashes. Regardless, the rest of this chapter includes the analysis, lessons learned, and modeling results for one near-crash event at the North Carolina CICAS site.

CICAS North Carolina US-74E Near-Crash Case: 12:11 p.m., April 25, 2007

Lacking video for this selected event, the sequence of events can only be deduced from the animated trajectory information for the vehicles involved. Figure 5.5 shows a frame captured from the visualization tool.

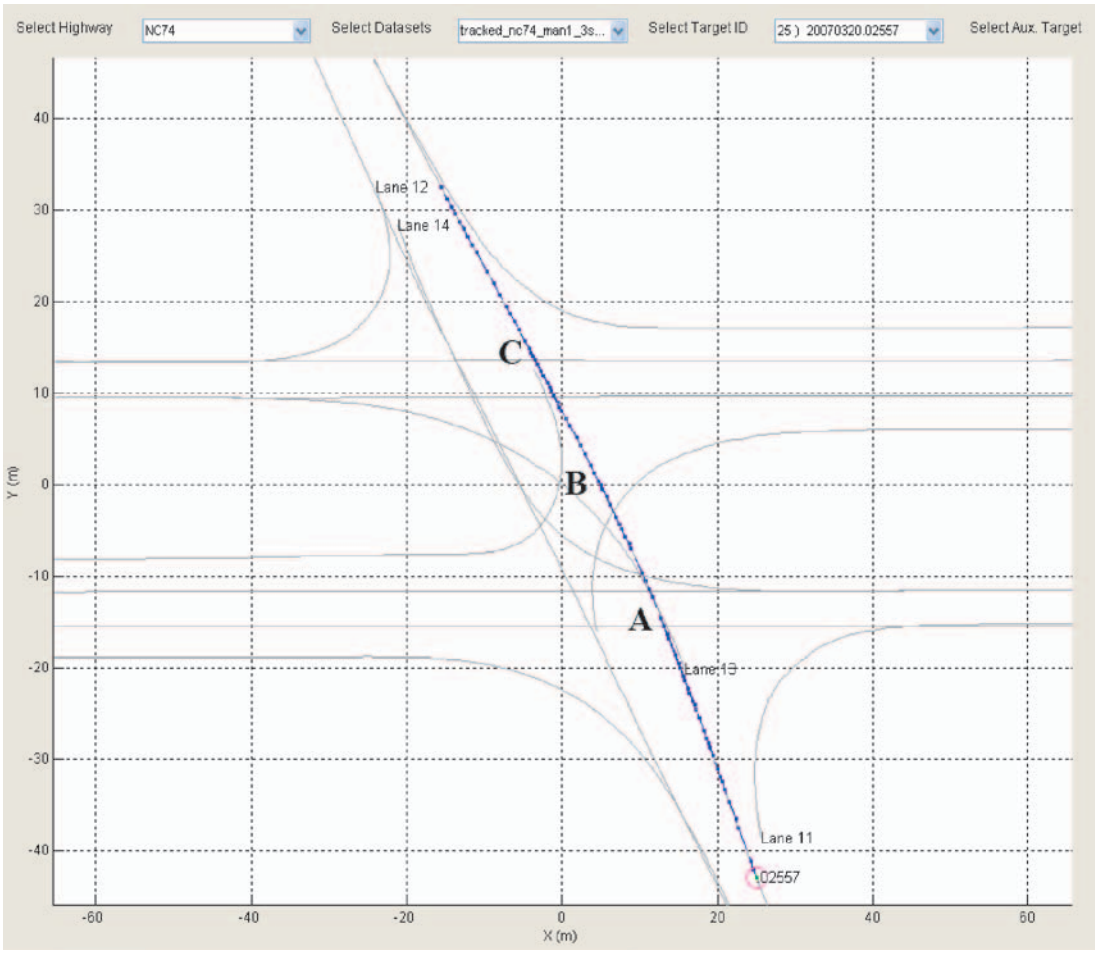


Figure 5.3. Tracked target interface showing trajectories of a side-road vehicle.

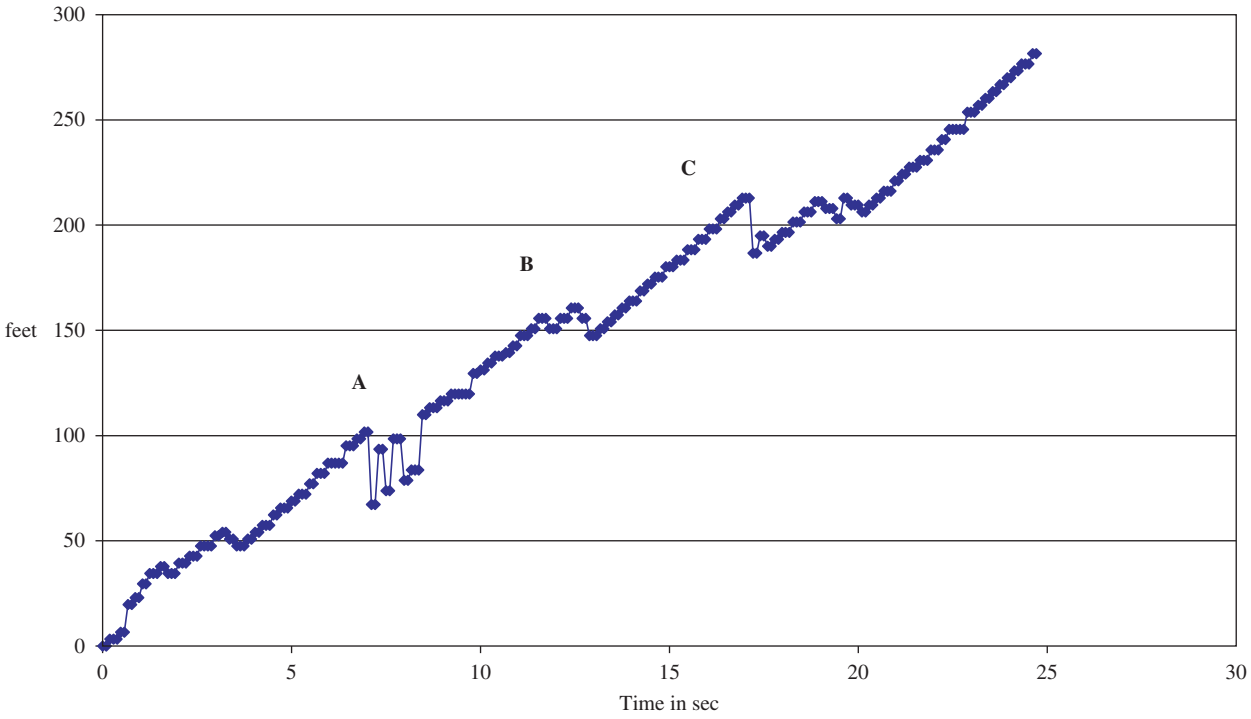


Figure 5.4. Side-road vehicle displacements.

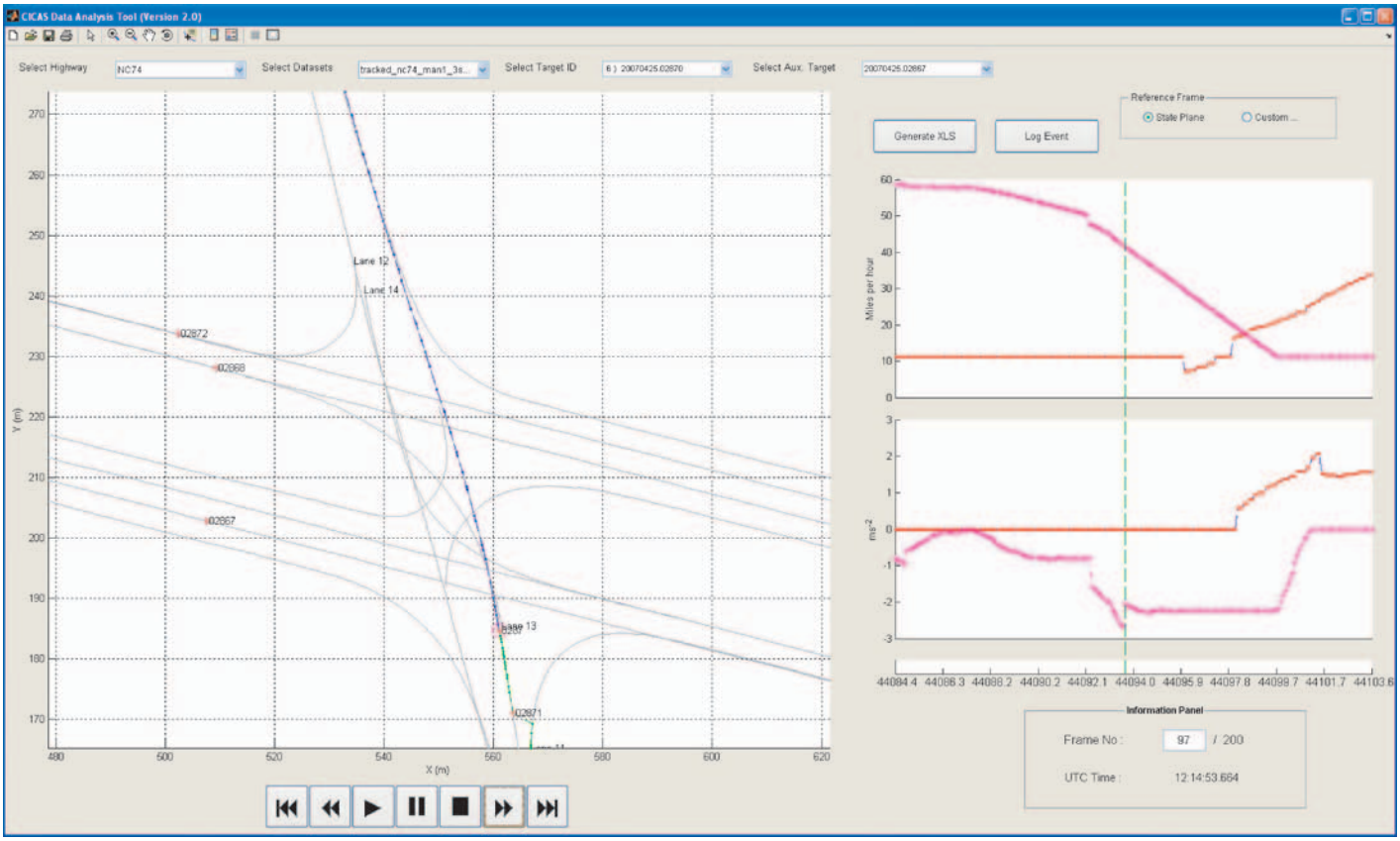


Figure 5.5. US-74 near-crash event, 12:11 p.m., April 25, 2007.

On the left side of Figure 5.5, there is a schematic of the intersection with the lines representing lane and turning centerlines, while on the right side the speed and acceleration graphs for the mainline and side-road vehicles are provided. A solid red line indicates the side-road vehicle, while magenta points indicate the mainline vehicle. On the graph side, the green vertical line indicates the values at the current frame displayed. In Figure 5.6, a zoomed view of the trajectories is presented.

In Figure 5.6, the red circle indicates the mainline vehicle involved in the near crash. The vehicle moves toward the intersection in subsequent frames. The green (short) part of the side-vehicle trajectory is the movement already performed, and the blue (longer) part is the trajectory in future frames. The conflict point for these trajectories is also indicated. For the purposes of developing a model describing the vehicle trajectories, all coordinates were translated to a system that has its center at the conflict point and the *x* axis parallel to the mainline vehicle trajectory. This simplified the kinematic equations of the vehicles by using only the distance to the conflict point. The speed and acceleration values provided in the CICAS database already represent the projected vectors on the trajectory of each vehicle. From the animation of the event, as well as the graphs of speed and acceleration, a description of the event was deduced. Specifically, the side-road vehicle is performing a northbound crossing of the intersection beginning from a standing position approximately 11 m (36 ft) from the conflict point. Considering this trajectory and the movement rate, the mainline vehicle, which initially is moving with an approximate speed of 26.8 m/s (60 mph), will eventually collide with the side-road vehicle, unless a change of attitude is implemented. The main-road driver, realizing this, performs a sequence of decelerations of up to more than −2 m/s^2 (−6.6 ft/s^2) to prevent the crash. The maneuver is performed successfully with room to spare. It is important to note that the described behavior does not consider any lateral movements of the vehicles, signifying types of evasive action other than deceleration, because if such movements did happen, they were smaller than the threshold used in the database for projecting the vehicle to the closest lane centerline. As discussed earlier, the animation of the side-road vehicle, which depended on the *x, y* coordinates, indicated that the reported speeds and accelerations do not agree. As seen in Figure 5.6, the side-road vehicle apparently never stops but appears to move with a constant speed of approximately 11 mph (4.9 m/s) even while the acceleration increases from zero until well after the vehicle is clearly moving. Therefore, for the rest of the analysis, only the *x, y* coordinates of

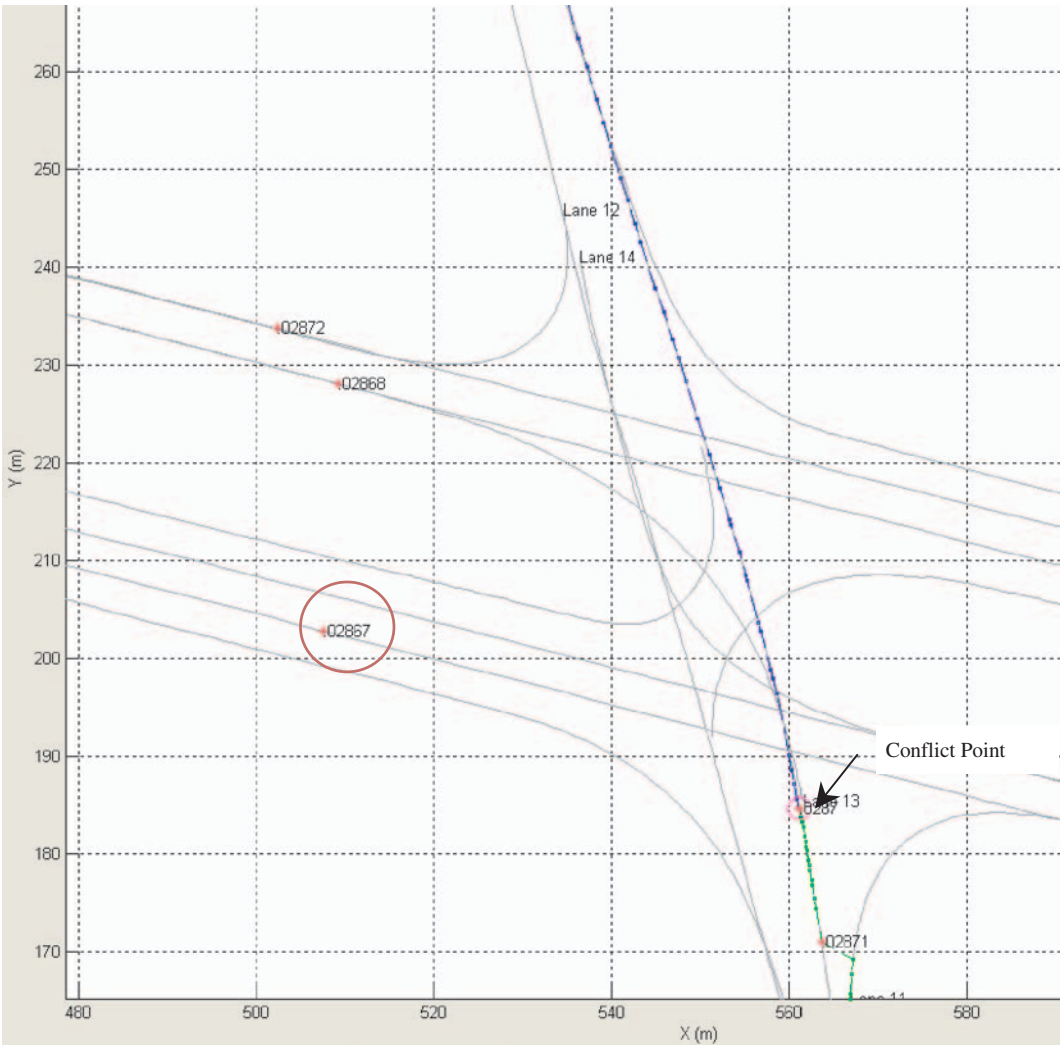


Figure 5.6. Lane centerline trajectories and selected targets.

the side-road vehicle are discussed, and speed and acceleration are inferred from them.

The analysis of the data followed two routes. An empirical review of the information was performed to determine the concurrence of the provided information, followed by a more in-depth analysis producing Bayes estimates for each vehicle's initial speed, the time points at which accelerations changed, and the accelerations in all stages. The reason for the empirical review was based on the quick realization of errors and discrepancies in the information provided, some of which was described earlier in this chapter and are illustrated in the following sections specifically for this event. For the first part of the analysis, approximately 19 s of data were used for the mainline vehicle; 12 s of these were spent approaching the conflict point from a starting point approximately 276 m away. Figure 5.7 presents a subset of these for clarity.

From this figure, certain problematic measurements, both in speed and distance, can immediately be observed. Specifically, speed exhibits a single point drop of ~1.7 kph (1.0 mph), which, apart from being physically unrealistic, is not evident in the distance measurements. To explore the extent of this phenomenon, the project team proceeded to use the available data of speed and acceleration to estimate displacement per interval and compared this to the displacement produced by the *x, y* coordinates. This comparison is seen in Figure 5.8. For the purpose of the estimation, it was assumed that the starting values were known and the process progressed from there. The estimates of distance based on speed and acceleration, although not identical, show better agreement with each other than they do with the provided distance from the database. This can be seen as an indication that displacement uses the sensor range measurement, while speed and acceleration use range rate. Still, it is evident that significant filtering has been introduced.

In Figure 5.9, a similar comparison is made on speed values, one being provided and the other calculated based on

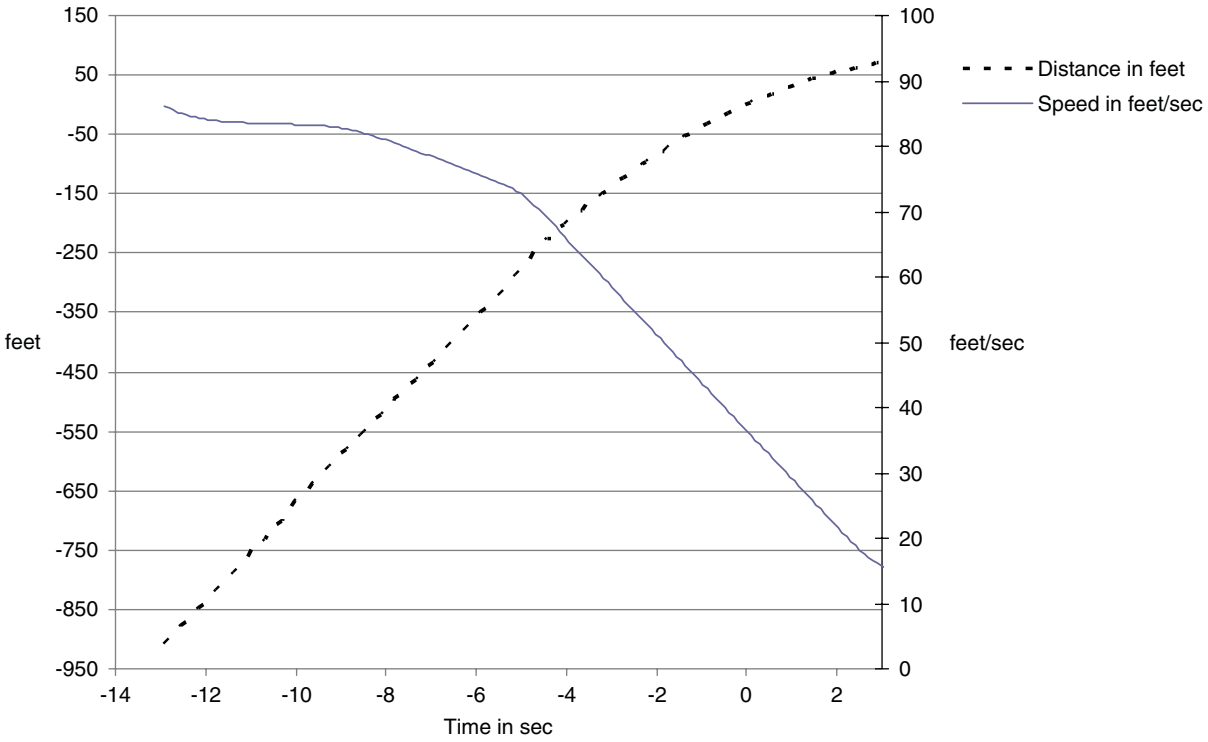


Figure 5.7. Mainline vehicle distance and speed.

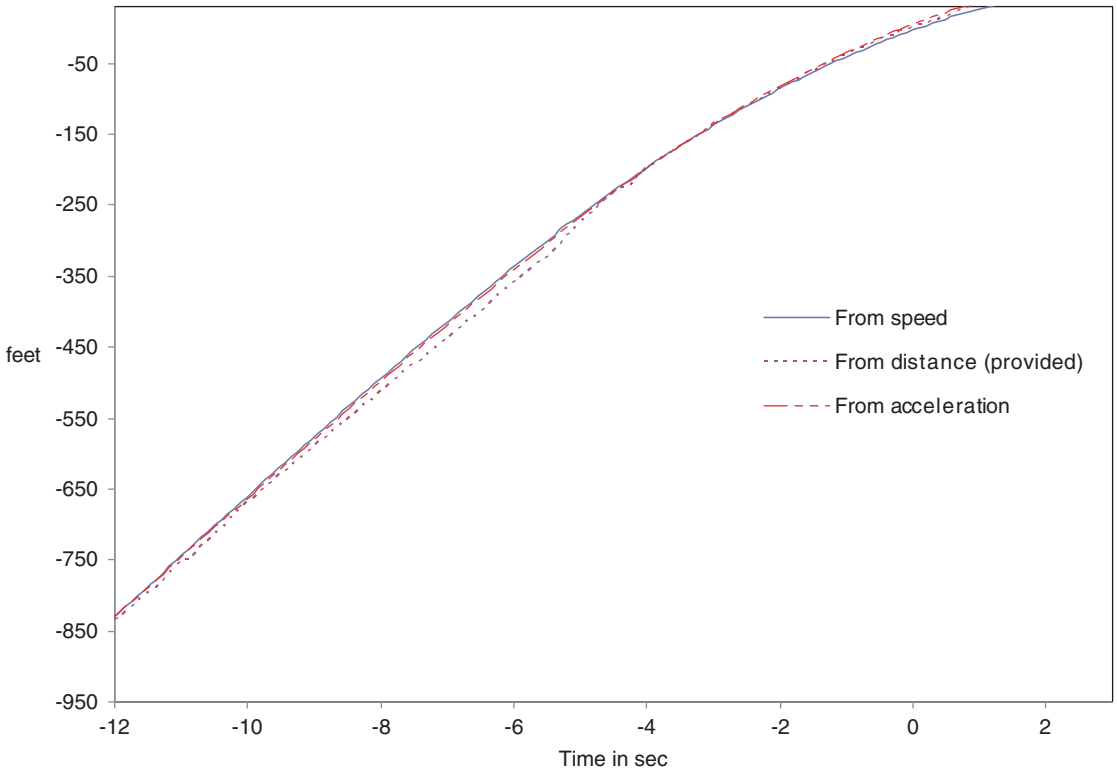


Figure 5.8. Mainline vehicle distance comparison.

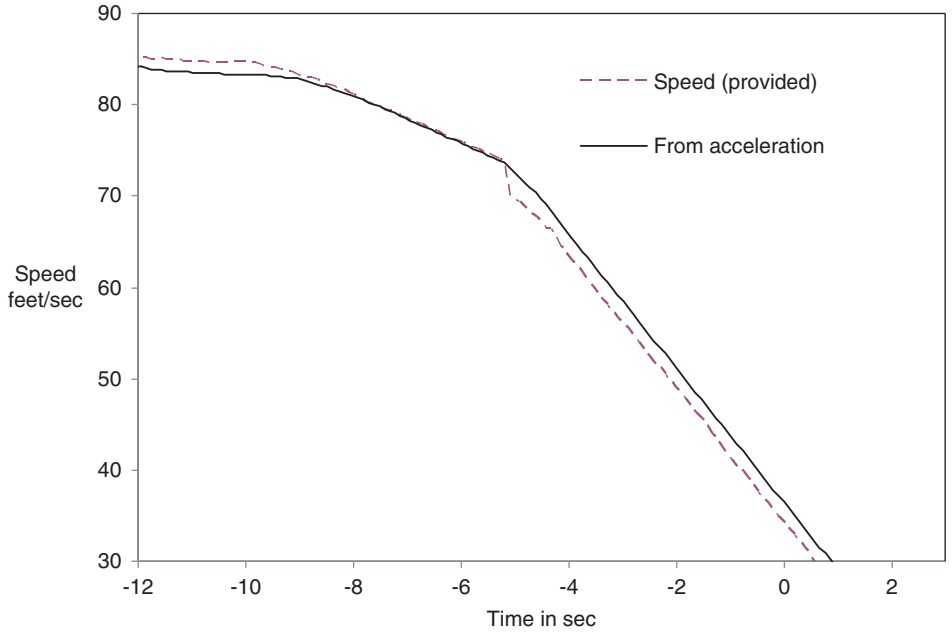


Figure 5.9. Mainline vehicle speed comparison.

acceleration. For the speed estimates calculated from acceleration, the abnormal speed drop discussed earlier is not observed, but this can be either due to the introduction of an error in postprocessing or an artifact of the sampling methodology that will not allow it to propagate the acceleration measurements. That is, if accelerations were originally estimated as finite differences of speed, then a step change in speed will generally not be present in the acceleration estimates. Although in this case this single error can be easily corrected, further investigation of the sensor characteristics and data reduction methodology would be required to facilitate better use of CICAS-collected mainline data with the trajectory-based methodology.

Figure 5.10 presents the best information regarding the side-road vehicle's distance-to-conflict point calculated from *x, y* coordinates. From the figure, it is evident that the measurements have both noise and errors and that these are not produced at the 10-Hz interval. The profile shows an unnatural stepwise progress for the vehicle. In this figure, back-and-forth movements of the vehicle can be observed, as discussed earlier in this chapter. This case of a near crash was selected for analysis after a large number of events were reviewed to locate a relevant case with the fewest discrepancies, so that manual postprocessing with a minimal amount of abductive inference was possible. Specifically, in this case the problems can be found mainly in time periods that are not important for the analysis (before the vehicle comes to rest at the stop line and after it has sufficiently cleared the conflict point). Hence, further investigation used only information from time −6 s to time 1 s, approximately.

Exploratory modeling for both vehicles was conducted empirically in Excel to identify a plausible piecewise acceleration model. For the mainline vehicle, a three-stage model was fit, where a gentle deceleration lasting for about 3 s was followed by a stronger deceleration for 1 s and then by an even stronger one until the vehicle passed the conflict point. For the side-road vehicle, a one-stage model was fit, where a constant mild acceleration carried it over the first part of the intersection. To avoid some of the errors and data discrepancies discussed earlier, further analysis used a subset of the data starting at the point 8.2 s before the arrival of the mainline vehicle at the conflict point and ending at that point.

Bayes estimates for each vehicle's initial speed, the time points at which accelerations changed, and the accelerations in all stages were computed using WinBUGS. For this analysis, it was deemed necessary to remove some of the most obvious problems exhibited in the provided data. Specifically, in the case of the mainline vehicle, the single point speed drop was corrected by adding a fixed value to all subsequent data points. For the side-road vehicle, some abnormal back-and-forth data points at the beginning of the time period were removed. The WinBUGS results are displayed in Table 5.1. Figures 5.11 to 5.13 display the comparison between model estimates and provided measurements.

One interesting observation resulting from the Bayes estimates of the driver behavior parameters involves the hypothetical mainline vehicle driver reaction time. Although the data are not sufficient to draw precise conclusions about the actions of the mainline driver, it can reasonably be assumed that the reaction to the side-road vehicle's encroachment on the mainline

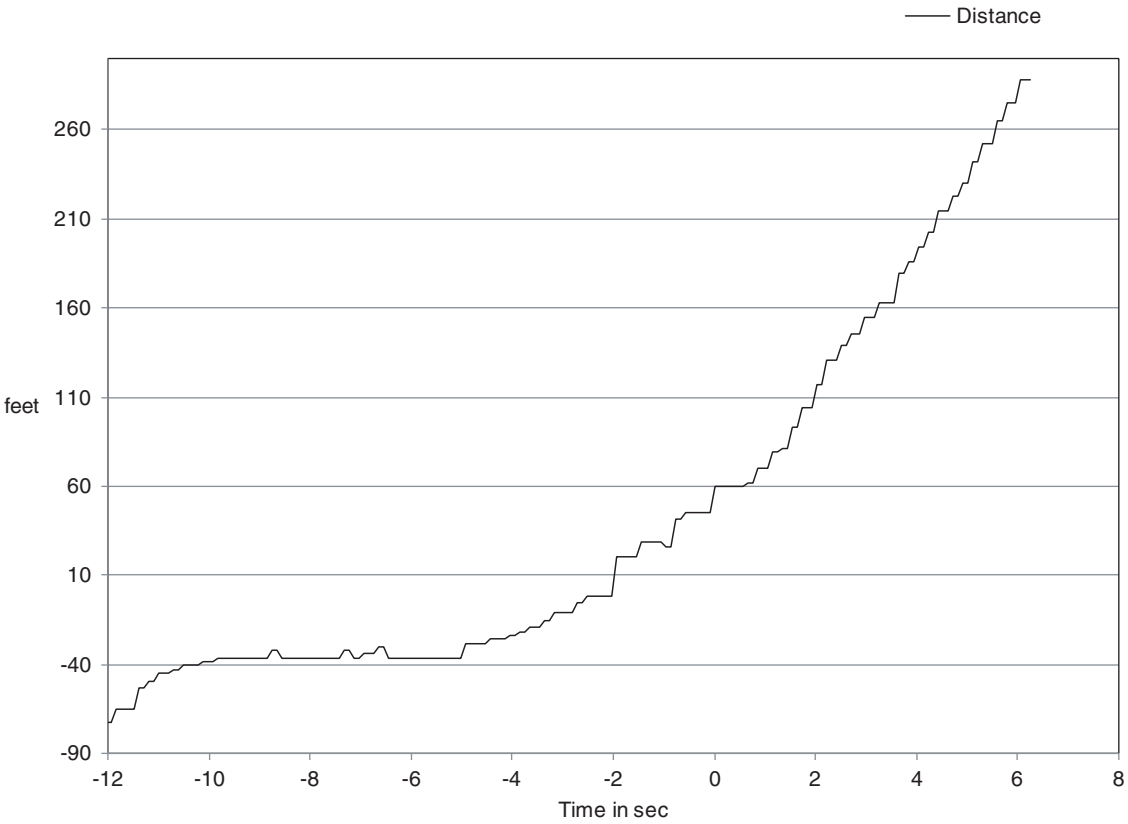


Figure 5.10. Side-road vehicle distance.

Table 5.1. WinBUGS Estimates for CICAS Near-Crash Case

Variable	Mean	Standard Deviation	2.5%tile	97.5%tile
Side-Road Vehicle (initial speed zero)				
First acceleration (ft/s^2)	4.74616	0.2188416	4.32632	5.18896
Movement start (s)	–6.1422	0.102	–6.339	–5.95308
Main-Road Vehicle				
Initial Speed (ft/s)	82.0328	0.0498888	81.9344	82.1312
First acceleration (ft/s^2)	–2.4987	0.0270666	–2.551184	–2.44426
Second acceleration (ft/s^2)	–5.32016	0.0070487	–5.61536	–4.92984
Third acceleration (ft/s^2)	–7.04872	0.0201326	–7.0848	–7.00608
First change (s)	–5.23692	0.03635	–5.31564	–5.17452
Second change (s)	–4.323	0.06229	–4.45548	–4.22988

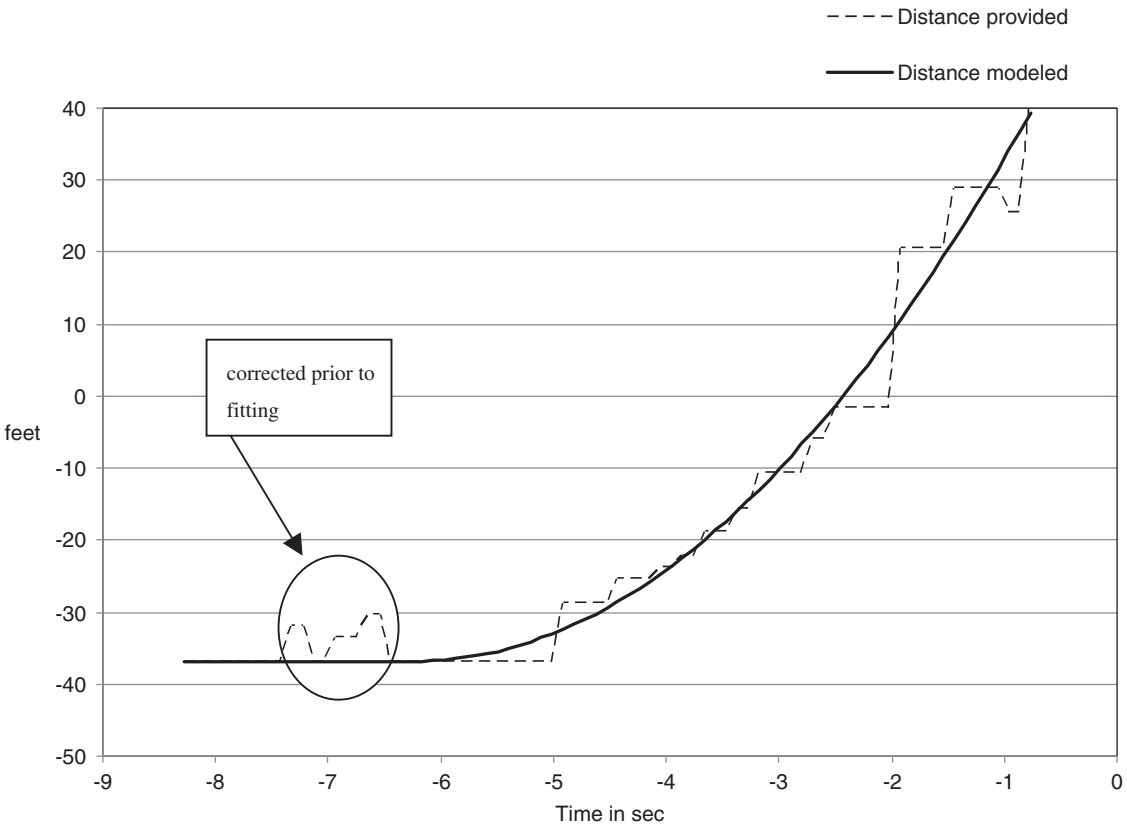


Figure 5.11. Measured and modeled side-road vehicle distances.

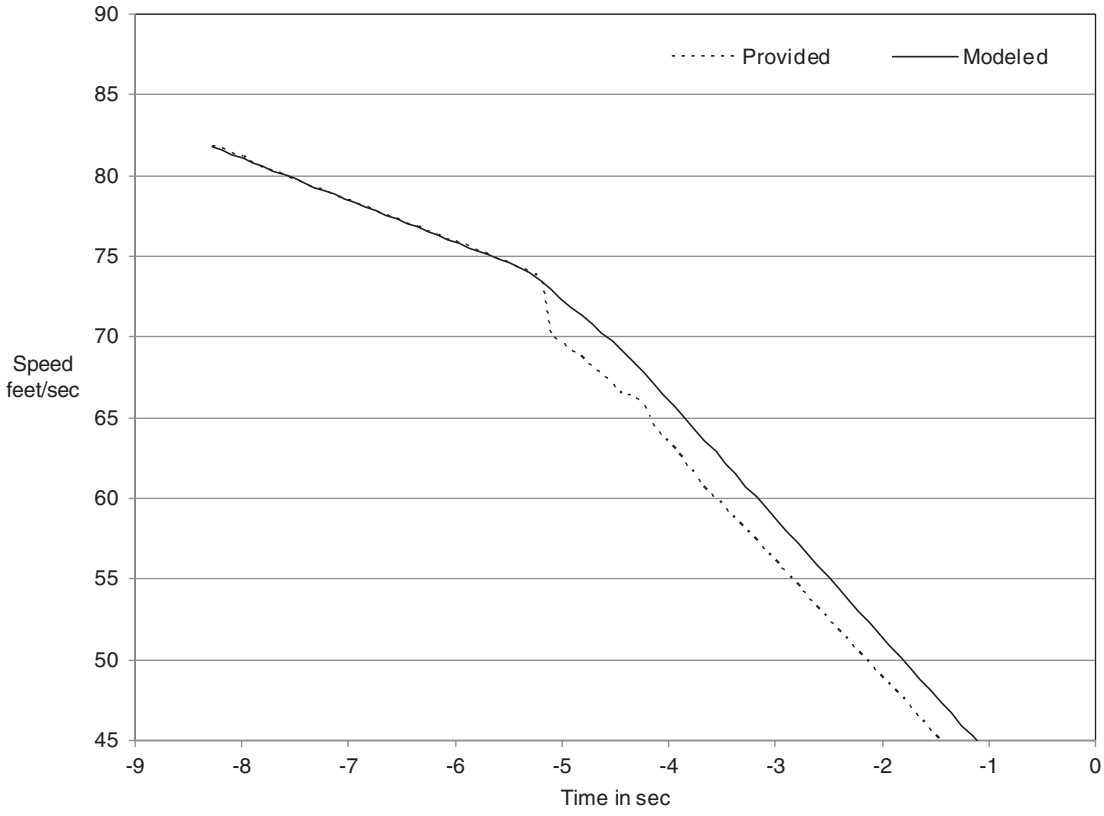


Figure 5.12. Measured and modeled main-road vehicle speeds.

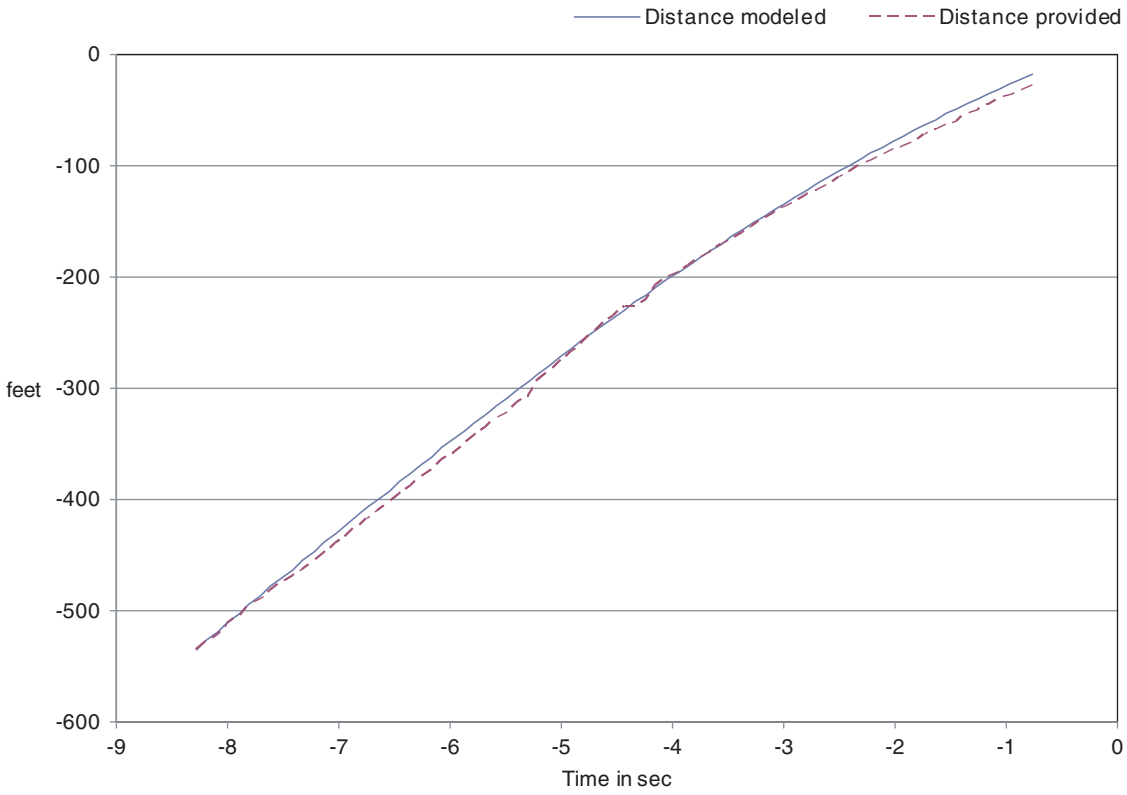


Figure 5.13. Measured and modeled main-road vehicle distances.

must be somewhere between the first and second change time points, −5.24 to −4.323. If the Bayes estimate for the side vehicles starting time of encroachment is ignored and the data provided are followed literally, the encroachment may be said to start at around time point −5.012. This indicates a possible reaction time zone of −0.224 to 0.689 s. This is a tight reaction time for a driver noticing movement at a distance of more than 70 m (~230 ft). Alternatively, though, if the Bayes estimate for the side vehicle time of encroachment is accepted, then the reaction time zone is 0.9 to 1.8 s, which is much more reasonable.

In conclusion, Chapter 4 described how the trajectory-based reconstruction method could be applied to site-based video data to estimate values of physical and behavioral variables characterizing crash and near-crash events. Reduction of the video data was conducted manually; issues concerning automatic reduction of video data, although important, were treated as outside the scope of this project. The main objective of the work described in Chapter 5 was to assess the ability of an alternative technology, based on using Doppler shift methods, to support the analytic approach. One advantage of this alternative technology is that the problem of automatically extracting vehicle trajectories from video is avoided and that a database of potentially useful events currently exists. The project team's experience indicated that its structural modeling and counterfactual screening methods can be applied to these data, but that to support this research, the existing data-collection and storage methods would require some technical modifications.

CHAPTER 6

Conclusions and Recommendations

Summary

It may be helpful to recall the primary goal of this project: to develop and test analytic methods that could then be applied to data produced by the SHRP 2 Safety field studies. The general approach was first to identify interesting research problems and then attempt to solve them using data similar to what is expected from the field studies. Chapters 3 to 5 illustrated how Bayesian statistical methods can be used to fit models to trajectory data, the parameters of these models being related to interesting event features. Although the empirical results presented in Chapters 3 to 5 offer some tantalizing suggestions regarding the relationships between crashes and conflicts, and the degree to which noncrash events provide information about crashes, the project team's position is that the number of events analyzed is too small to justify drawing general conclusions. Such conclusions, of course, are what one hopes will result from the field studies. In this chapter, then, the discussion is restricted to methodological issues.

In the work plan submitted at the end of Phase 1, three research problems were identified on which progress was needed. The first of these was identification of an appropriate class of structural models describing how crash and near-crash events developed, together with analytic tools for fitting these models to data expected from the vehicle- and site-based field studies. The second problem involved counterfactual screening of supposed near-crash events to determine their similarity to crashes, and in the Phase 1 report, a theoretical result was developed, which indicated that those near-crash events that are most like crashes would have evasive actions more extreme than those used in crashes. The third problem involved developing plausible models of how drivers select evasive actions as functions of the situations in which they find themselves. In the Phase 1 report, it was indicated that given an adequately large sample of crash and near-crash events for which estimates of background conditions and driver actions were available, it should be possible to conduct exploratory modeling of evasive action selection using regression-type models.

Solutions to the second and third research problems are contingent on a solution to the first, so the bulk of the effort during Phase 2 was devoted to structural modeling of crash and near-crash events. Chapter 2 of this report described a class of models that characterized driver behavior as a sequence of discrete changes in acceleration and illustrated how an ordinary differential equation taking this acceleration sequence as an input, together with initial speeds and positions, could be integrated to give a predicted trajectory for the vehicle's motion. It was then illustrated how the parameters describing a driver's acceleration sequence and the initial conditions could be estimated from vehicle trajectory data. It was also shown how the identified model and estimates could be used to address the second research problem by computing the probability a crash would have resulted, other things being equal, as a function of a range of counterfactual evasive actions. Chapter 3 of this report described the application of the project team's methods to seven rear-ending crash and near-crash events obtained from the 100-car vehicle-based field study. Chapter 4 described application to six rear-ending crash and near-crash events using site-based video data obtained from the MTO, while Chapter 5 described a pilot application to an intersection angle conflict using site-based Doppler shift data obtained from the CICAS system.

Conclusions

Statistical analyses of crash frequency data can identify reliable associations between crash experience and roadway or driver features but are of limited value in discovering how crashes occur. The first conclusion is that in situations where the direction of travel is roughly constant, trajectory-based reconstruction of crash-related events, where trajectory data are used to fit parsimonious models of driver behavior, is feasible using both vehicle-based and site-based data. The

product of such a reconstruction is a set of estimates of when and by how much drivers changed their acceleration and the background conditions associated with these changes. These estimates can in turn be used to produce estimates of driver reaction times, following distances, and selected/rejected gaps. Bayes estimates, especially estimates of posterior probability distributions, can be obtained using Markov Chain Monte Carlo simulation. This approach is especially helpful in studying crash-related events involving two or more vehicles, where information on the behavior of drivers in noninstrumented vehicles is required.

One goal of traffic safety research is identifying causes of traffic crashes. The notion of cause has historically been rather difficult to pin down, but a good case can be made that a common core exists in what is meant by cause in statistical estimation of crash-reduction factors, in reconstruction of individual crashes, and in simulation of crash events. This common core is a counterfactual definition of cause (*1*). As indicated in Chapter 1, a counterfactual component is also present in working definitions of traffic conflict and near crash. Given a simple structural model of an event along with Bayes estimates of the posterior distribution for the model's parameters, it is possible to quantify the degree to which a near crash could have been a crash by perturbing a driver's evasive action and computing the probability that a crash results. The second conclusion is that it is possible to extend the methods of counterfactual analysis to more complicated structural models involving differential equations.

The Phase 1 report presented a theoretical argument to the effect that for a near-crash event to be similar to a crash event, the near-crash event should have an evasive action more extreme than that in the crash event. If crashes tend to involve extreme evasive actions, then this would imply that those near crashes that are most similar to crashes would tend to be less frequent than crashes. Although a conclusive test of this prediction would require more data than were available in this study, Table 6.1 hints that this might not be as big a problem as feared.

The maximum evasive deceleration observed for the crash events was about −12.9 ft/s^2, while the maximum evasive deceleration observed for a near-crash event was about −21.8 ft/s^2, with the second-most extreme successful evasive deceleration being about −18.3 ft/s^2. The third conclusion is that, at least for rear-ending events, there is some limited evidence that the distributions of evasive actions for crashes and near crashes share some overlap, so that it should be possible to find near-crash events that are similar in other respects to crashes.

Vehicle-based data configurations have definite limits regarding information provided about noninstrumented vehicles involved in multivehicle crashes or near crashes. Chapter 4 illustrated how site-based video with manual extraction of vehicle trajectory data can support structural modeling and counterfactual analysis of multivehicle events, but clearly this is not feasible for processing very large numbers of events that might be expected from a longitudinal study. The CICAS data-collection system, based on radar and LIDAR units, collects and processes large amounts of vehicle trajectory data at intersections. The current CICAS configuration and architecture is designed to provide available gap information to minor-approach drivers rather than to process and analyze data on crashes and conflicts. With some technical modifications,

Table 6.1. Estimated Maximum Evasive Deceleration for Following Vehicles in 10 Events Analyzed in Chapters 3 and 4

Event ID	Deceleration (ft/s^2)	Event Type
100-Car Vehicle-Based Data		
99540	–12.6	Rear-end crash
104119	–9.5	Rear-end near crash
73082	–18.3	Rear-end near crash
104851	–21.8	Rear-end near crash
104283	–16.2	Rear-end near crash
60289	–10.7	Rear-end near crash
I-94 Site-Based Data		
Oct 8 1600: vehicles 1 and 2	–8.0	Rear-end near crash
Oct 8 1600: vehicles 2 and 3	–10.2	Rear-end crash
Oct 13: vehicles 1 and 2	–10.5	Rear-end near crash
Oct 13: vehicles 2 and 3	–12.9	Rear-end crash

however, this system has the potential to provide data needed for structural modeling of crash and near-crash events, at least at lower-volume intersections.

Conclusions are only as reliable as the data upon which they are based, and as described in Chapter 3, in many cases, the vehicle-based data from the 100-car study showed incompleteness, inconsistencies, or errors that limited the ability to use them. These included cases where the forward radar data were missing or corrupted, where the speedometer data were clearly in error, and where there were marked differences in the vehicle trajectory as implied by the accelerometer measures and as implied by the speedometer and heading measures. It was also true for all cases studied that the GPS coordinates were not sufficiently refined to determine vehicle trajectories. It is recognized that the 100-car study was a pilot effort and that detailed quantitative reconstruction of events was probably not one of the study's objectives. Nonetheless, the final conclusion here is that the usefulness of the data produced by the SHRP 2 vehicle-based field study will be strongly dependent on the ability to calibrate and maintain the data-collection systems.

Recommendations for Future Work

All but one of the crash and near-crash events analyzed so far show essentially straight-line trajectories for the involved vehicles. Although the project team has developed a preliminary model code that allows for two-directional trajectories, it was not possible within the time and resource constraints to bring this to the degree of maturity achieved for straight-line events. Since two-directional motion occurs when a driver uses swerving as an evasive action and when a left-turning driver selects a gap, the first recommendation is for developing and testing trajectory estimation tools that handle two-directional trajectories.

In the reconstruction of crash and near-crash events, driver inputs such as acceleration rates, reaction times, and following distances can be treated as exogenous quantities to be estimated, and the issue of how drivers select evasive actions does not arise. When attempting to include crash events in a microscopic traffic simulation, however, plausible models that close the feedback loop between existing conditions and driver actions are necessary. The second recommendation is for conducting research on this issue using the data from the SHRP 2 field studies.

The Phase 1 report pointed out that in some cases the residuals obtained after fitting a trajectory model showed serial correlation. When serial correlation is present but unaccounted for, the standard errors and confidence intervals associated with parameter estimates suggest more precision for those estimates than is justified. That is, although in all the analyzed cases the trajectory models that were fit give reasonable descriptions of data, there may be greater uncertainty in the parameter estimates than have been so far acknowledged. The third recommendation is that this project's model estimation methods be enhanced to allow for possible serial correlation.

The project's experience with vehicle-based data indicated that except for car-following events where both vehicles remain primarily in the same lane, the quantitative information available about driver behavior in noninstrumented vehicles is essentially nonexistent. Compiling data on gap-selection and other intersection-related events will then require a different data setup. The fourth recommendation, then, is that the vehicle-based study be complemented with site-based research.

During the course of this project, a recurring factor was the unknown or uncertain influence of the measurement method or hardware, as well as the postcollection filtering, on the available data. In all future studies, consistency and transparency in data collection and processing methodologies are paramount. Past efforts naturally were affected by the objectives and priorities of the projects funding, so that future dissemination of data may have received less emphasis. Considering that the primary objective of SHRP 2 Safety Project S01 was to examine the usability of existing data, it is clear that the broader utility of data in a project be recognized. Therefore, the final recommendation is that, beginning with SHRP 2 and similar federally funded projects, a clause be added to each RFP to make sure that the data-collection setup, postcollection processing, and storage and availability of information are clearly described in the final report.

Reference

1. Davis, G. Towards a Unified Approach to Causal Analysis in Traffic Safety Using Structural Causal Models. In *Transportation and Traffic Theory in the 21st Century* (M. A. P. Taylor, ed.), Elsevier Science, Ltd., Oxford, United Kingdom, 2002, pp. 247–266.

APPENDIX A

Analysis Tools Developed in This Project

MATLAB GUIs

The SHRP 2 Safety Project S01A graphical user interface (GUI) for data analysis has been implemented in the MATLAB GUI Development Environment. It displays a map of the highway and the trajectories of the tracked targets overlaid on the highway. Playback can be controlled as desired, and there are side plots displaying other information regarding the acceleration and velocity of the tracked targets. Reference frames can be changed as required, and an output text file can be generated, which contains the trajectory of a specific target vehicle for postprocessing.

The browsing tool can accommodate data from the CICAS-instrumented sites (tested with data from the Minnesota and North Carolina sites), as well as process data from the VTTI 100-car study. Figures A.1 and A.2 are screenshots of the site-based data GUI and the vehicle-based data GUI, respectively.

In the case of the site-based data, the browser generates the intersection geometry based on the available survey information provided by the research team. The coordinate systems of the vehicle trajectories are all in state plane, but the browser allows the user to set any arbitrary coordinate system and performs the conversion automatically. Considering that the site-based data contain trajectories from multiple vehicles, the browser allows the user to select which vehicles to observe in order to avoid visual clutter. Selecting the Play button (Figure A.1) plays the animation. The blue trajectory of the selected target changes into green, referencing past periods. All the vehicles' locations are plotted on the highway, and the mantissa of the target IDs are shown next to each vehicle. The magenta circle around one of the vehicles indicates that this is the selected target. Other playback controls are as follows:

- The Pause button pauses the animation. Clicking Play again resumes the animation from the current frame.
- The Stop button stops the animation and returns to the first frame.
- The Forward (Rewind) button moves forward (backward) one frame at a time.
- The Fast Forward (Fast Rewind) button moves forward (backward) 10 frames at a time.
- The scroll bar above the playback controls can also be used to navigate through the animation frames.

When the animation is playing, clicking any of the forward/rewind controls or the scroll bar will immediately pause the animation.

The vehicle-based data browser interface is similar in nature, but since there is no specific map to be displayed, additional measurement charts are presented instead.

MATLAB Estimation Scripts

Straight-Line Trajectory Model

```
function simout=traj1d(beta)
% function simout=traject1d(beta)
% input:
% x0=beta(1)
% vx0=beta(2)
% ax0=beta(3)
% ax1=beta(4)
% ax2=beta(5)
% tc1=beta(6)
% tc2=beta(7)
% output:
% simout=[x' vx' ax']
% global n delta

global n delta
x0=beta(1)
vx0=beta(2)
ax0=beta(3)
ax1=beta(4)
ax2=beta(5)
tc1=beta(6)
tc2=beta(7)
```

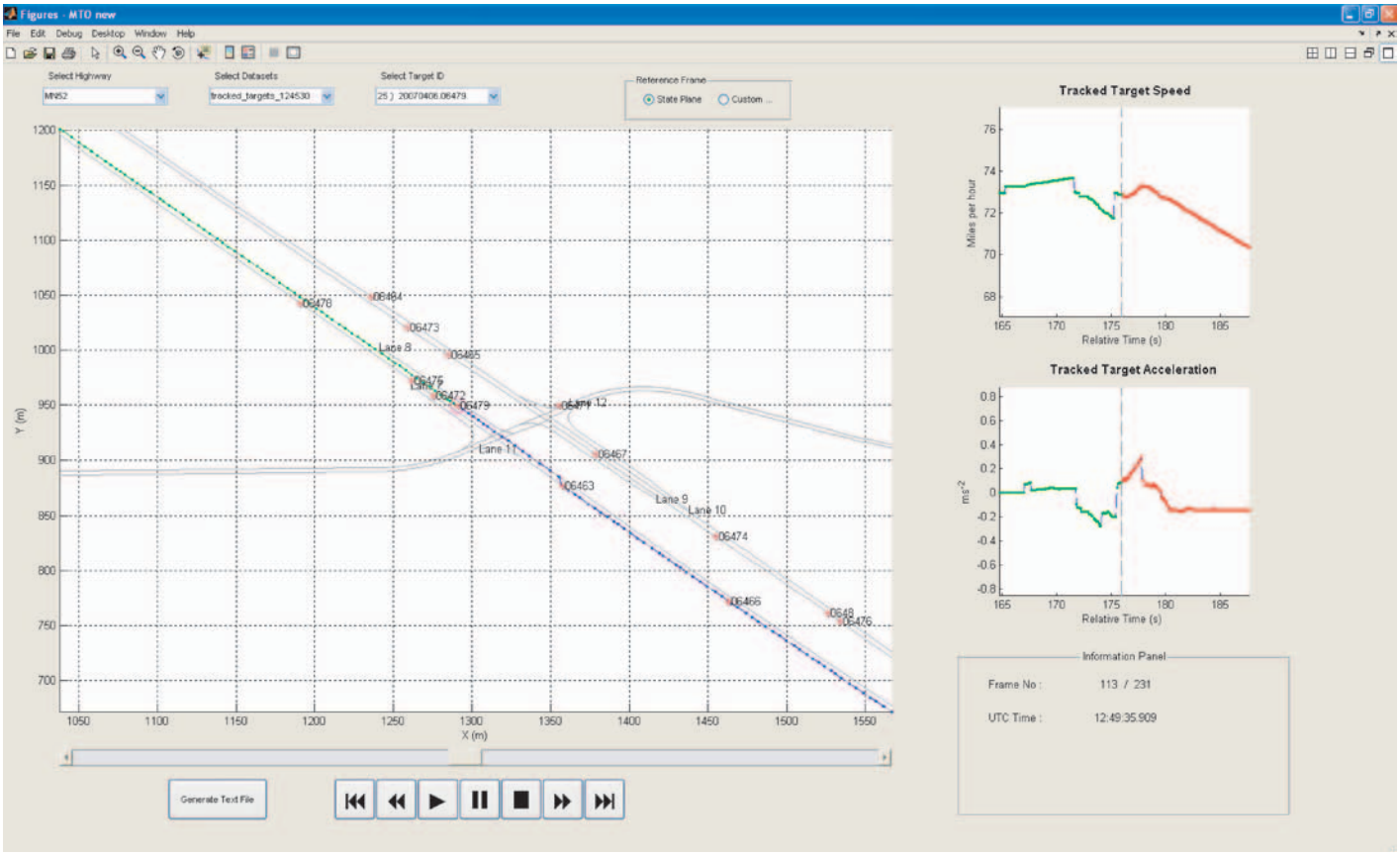


Figure A.1. Site-based data browser.

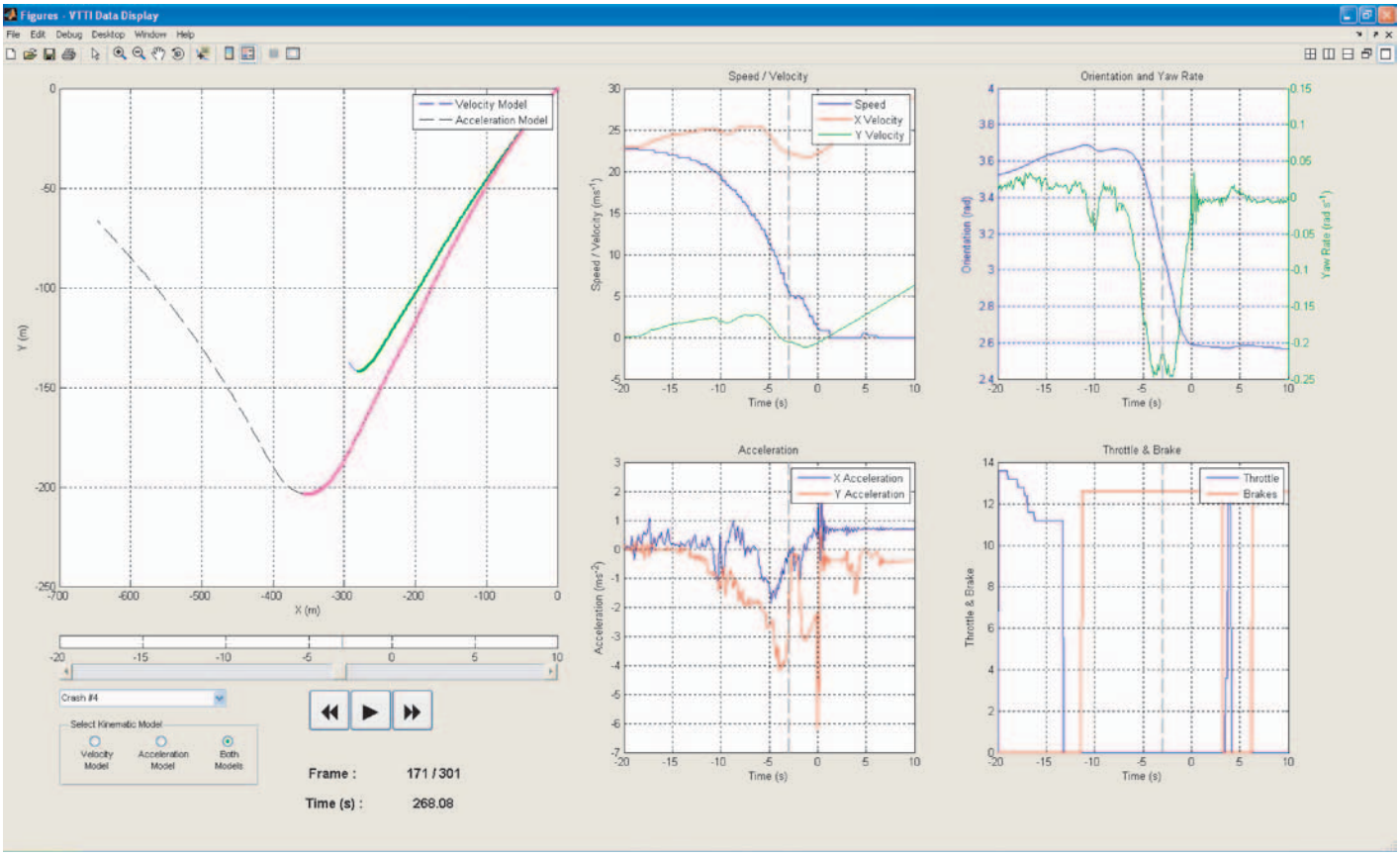


Figure A.2. Vehicle-based data browser.

```
x(1)=x0;
vx(1)=vx0;
ax(1)= ax0;
for t=2:n
    x(t)=x(t-1)+vx(t-1)*delta;
    vx(t)=vx(t-1)+ax(t-1)*delta;
    if vx(t) < 0
      vx(t)=0;
    end
    ax(t)=ax0;
    if (t*delta > tc1)
        ax(t)=ax1;
    end
    if(t*delta > tc2)
        ax(t)=ax2;
    end
end
simout=[x' vx' ax'];
```

Compute Sum-of-Squares

```
function ss=sumsqr0(alpha)
global n delta tc1 tc2 ydat1
betax=[alpha tc1 tc2];
simoutx=traj1d(betax);
err=ydat1-simoutx(:,1);
ss=err'*err;
```

Grid Search of Change Points

```
% grid search over critical time points
for lead vehicle
clear ssx;
nn=size(t0grid);
nt0=nn1(2);
nn1=size(t1grid);
nt1=nn1(2);
for i=1:nt0
for j=1:nt1
t0=t0grid(i);
t1=t1grid(j);
alpha1=fminsearch(@sumsqr,alpha0,options);
betax=[alpha1 t0 t1];
ssx(i,j)=ss2(beta2x);
end
end
```

Example of WinBUGS Code

Model:

```
# 100-CAR case 104119
# 3-block model for follower (vehicle 1)
# 4-block model for leader (vehicle 2)
# separate calls to ode.block for leader and follower
# includes counterfactual simulation
# lagged initialization
# speed reparameterization

{

for (k in 1:n.grid) {grid[k] <- k*delta}

for (k in 1:n.good1) {
v1[k] ~ dnorm(vhat[k],tau1)
vhat[k] <- state1[index1[k],2] }

for (k in 1:n.good2) {
range[k] ~ dnorm(rangehat[k], tau2)
rangehat[k] <- state2[index2[k],1]-state1[index2[k],1]
rrate[k] ~ dnorm(ratehat[k],tau3)
ratehat[k] <- state2[index2[k],2]-state1[index2[k],2]
}

state1[1:n.grid,1:dim] <- ode.block(inits1[1:n.block1,1:dim],
grid[1:n.grid],D(A1[1:dim],t),origins1[1:n.block1],tol)

D(A1[1],t) <- A1[2]
D(A1[2],t) <- acc1

acc1 <- piecewise(vec.acc1[1:n.block1])

vec.acc1[1] <- a11
vec.acc1[2] <- a12
vec.acc1[3] <- 0

origins1[1] <- 0
origins1[2] <- t11
# v1x <- v10+a11*t11
# origins1[3] <- t11+v1x/(-a12)
origins1[3] <- t12

inits1[1,1] <- 0-v10*delta
inits1[1,2] <- v10
inits1[2,1] <- 0
inits1[2,2] <- 0
inits1[3,1] <- 0
inits1[3,2] <- 0

state2[1:n.grid,1:dim] <- ode.block(inits2[1:n.
block2,1:dim],grid[1:n.grid],D(A2[1:dim],t),origins2[1:n.
block2],tol)

D(A2[1],t) <- A2[2]
D(A2[2],t) <- acc2

acc2 <- piecewise(vec.acc2[1:n.block2])
vec.acc2[1] <- a21
vec.acc2[2] <- a22
vec.acc2[3] <- a23
vec.acc2[4] <- 0

origins2[1] <- 0
origins2[2] <- t21
origins2[3] <- t22
# v2x <- v20+a21*t21+a22*(t22-t21)
# origins2[4] <- t22+v2x/(-a23)
origins2[4] <- t23
# follower reaction time
r <- t11-t22
yindex0 <- round(t22/delta)
yy0 <- rangehat[yindex0]
vv0 <- vhat[yindex0]

inits2[1,1] <- range[1]-v20*delta
inits2[1,2] <- v20
```

```
inits2[2,1] <- 0
inits2[2,2] <- 0
inits2[3,1] <- 0
inits2[3,2] <- 0
inits2[4,1] <- 0
inits2[4,2] <- 0

v10 ~ dnorm(0, 1.0E-06)I(0,)
v11 ~ dnorm(0,1.0E-06)I(0,)

# final speed is stopped
v12 <- 0
a11 <- (v11-v10)/t11
a12 <- (v12-v11)/(t12-t11)
v20 ~ dnorm(0,1.0E-06)I(0,)
v21 ~ dnorm(0, 1.0E-06)I(0,)
v22 ~ dnorm(0, 1.0E-06)I(0,)

# final speed is stopped
v23 <- 0
a21 <- (v21-v20)/t21
a22 <- (v22-v21)/(t22-t21)
a23 <- (v23-v22)/(t23-t22)

t11 ~ dunif(10,12)
t12 ~ dunif(13,16)
t21 ~ dunif(3,5)
t22 ~ dunif(9,11.5)
t23 ~ dunif(13,15)

tau1 ~ dgamma(.001,.001)
tau2 ~ dgamma(.001,.001)
tau3 ~ dgamma(.001,.001)
sig21 <- 1/tau1
sig22 <- 1/tau2
sig23 <- 1/tau3

# counterfactual simulation

# follower model

state1.star[1:n.grid,1:dim] <- ode.block(inits1.star[1:n.
block1,1:dim], grid[1:n.grid],D(A1.star[1:dim],t),origins1.
star[1:n.block1],tol)

D(A1.star[1],t) <- A1.star[2]
D(A1.star[2],t) <- acc1.star

acc1.star <- piecewise(vec.acc1.star[1:n.block1])

vec.acc1.star[1] <- a11.star
vec.acc1.star[2] <- a12.star
vec.acc1.star[3] <- 0

a11.star <- a11

origins1.star[1] <- 0
origins1.star[2] <- t11
v1x.star <- v10+a11.star*t11
origins1.star[3] <- t11+v1x.star/(-a12.star)
inits1.star[1,1] <- 0-v10*delta
inits1.star[1,2] <- v10
inits1.star[2,1] <- 0
inits1.star[2,2] <- 0
inits1.star[3,1] <- 0
inits1.star[3,2] <- 0

# leader model

state2.star[1:n.grid,1:dim] <- ode.block(inits2.star[1:n.
block2,1:dim],grid[1:n.grid],D(A2.star[1:dim],t),origins2.star
[1:n.block2],tol)

D(A2.star[1],t) <- A2.star[2]
D(A2.star[2],t) <- acc2.star

acc2.star <- piecewise(vec.acc2.star[1:n.block2])
vec.acc2.star[1] <- a21.star
vec.acc2.star[2] <- a22.star
vec.acc2.star[3] <- a23.star
vec.acc2.star[4] <- 0

a21.star <- a21
a22.star <- a22
a23.star <- a23

origins2.star[1] <- 0
origins2.star[2] <- t21
origins2.star[3] <- t22
v2x.star <- v20+a21.star*t21+a22.star*(t22-t21)
origins2.star[4] <- t22+v2x.star/(-a23.star)

inits2.star[1,1] <- range[1]-v20*delta
inits2.star[1,2] <- v20
inits2.star[2,1] <- 0
inits2.star[2,2] <- 0
inits2.star[3,1] <- 0
inits2.star[3,2] <- 0
inits2.star[4,1] <- 0
inits2.star[4,2] <- 0

for (k in 1:n.grid) {
range.star[k] <- state2.star[k,1]-state1.star[k,1]
hitcheck[k] <- 1-step(range.star[k]) }
hitsum <- sum(hitcheck[])
hit <- step(hitsum-0.5) }

}
```

Data click on one of the arrows to open the data

Inits

```
list(v10=20.7,v11=20,tau1=1,v20=20,v21=20,v22=20,tau2=1,
tau3=1,t11=11.2,t12=14.5,t21=4,t22=10.4,t23=14)
```

APPENDIX B

The CICAS Site-Based System

Database Architecture

This section gives a detailed description of the database schema and metadata used for each field site. The metadata are described first, followed by the associated schema that enables extraction of target tracking and raw sensor point tracking data.

Metadata

The metadata are contained within several XML text files, which describe the range sensor locations, the zone-to-region mapping, the intersection center, and the maneuver topology. Of key importance for the report herein are the definitions of various maneuver events, which consist of tracked targets passing through a defined combination of three zones (Figure B.1). The project team is interested in the side-road vehicle maneuver events that correspond with the crash event occurring at the site: the side-road vehicle attempts to cross the median zone to the other side, requiring the driver of that vehicle to judge and accept gaps from each mainline direction during the crossing maneuver. This maneuver is defined when the vehicle enters and exits from zones 2 and 1, respectively, and vice versa, entering and exiting zones 1 and 2; i.e., the maneuver attribute within the vehicle_acpt_lag relational table (Table B.1). For a given maneuver, other zones define stop events (stop bar), pullout after the stop (accepted lag), and progression after pullout (transition) events. A maneuver type is defined only if a vehicle (e.g., a tracked target) has been determined to pass through all zones constituting the maneuver. Thus, continuing with the example, and referring to Figure B.1, the straight-through zone1-zone2 maneuver passes through zones 9, 10, 11, 7, 15, 16, and 17. Note that the recorded crash for this site did not have the straight-straight maneuver assigned because the maneuver was never completed. A piece of the metadata file defining maneuver types, maneuver_config_<database name>.xml, follows, with the zones referenced in Figure B.1.

```
<?xml version= “1.0” encoding= “UTF-8”?>

<!-- follow xml rules -->
<!-- this is a comment -->
<!-- comments can span
more than one line -->
<maneuver_config>
  <config>
    <num_maneuver ctype=“int”>14</num_maneuver>
  </config>
  <maneuver1>
    <!-- maneuverID -->
    <maneuverID ctype=“int”>1</maneuverID>
    <!-- enter_zone -->
    <enter_zone ctype= “int”>1</enter_zone>
    <!-- exit_zone -->
    <exit_zone ctype=“int”>2</exit_zone>
    <!-- description of the maneuver -->
    <description ctype=“char”>Strawberry Blvd N to
      Strawberry Blvd N</description>
    <!-- type of maneuver, straight,right,left -->
    <type ctype=“char”>straight,straight</type>
    <roadType ctype=“char”>minor</roadType>
  </maneuver1>
  <maneuver2>
    <!-- maneuverID -->
    <maneuverID ctype=“int”>2</maneuverID>
    <!-- enter_zone -->
    <enter_zone ctype=“int”>2</enter_zone>
    <!-- exit_zone -->
    <exit_zone ctype=“int”>1</exit_zone>
    <!-- description of the maneuver -->
    <description ctype=“char”>Strawberry Blvd S to
      Strawberry Blvd S</description>
    <!-- type of maneuver, straight,right,left -->
    <type ctype=“char”>straight,straight</type>
    <roadType ctype=“char”>minor</roadType>
  </maneuver2>
```

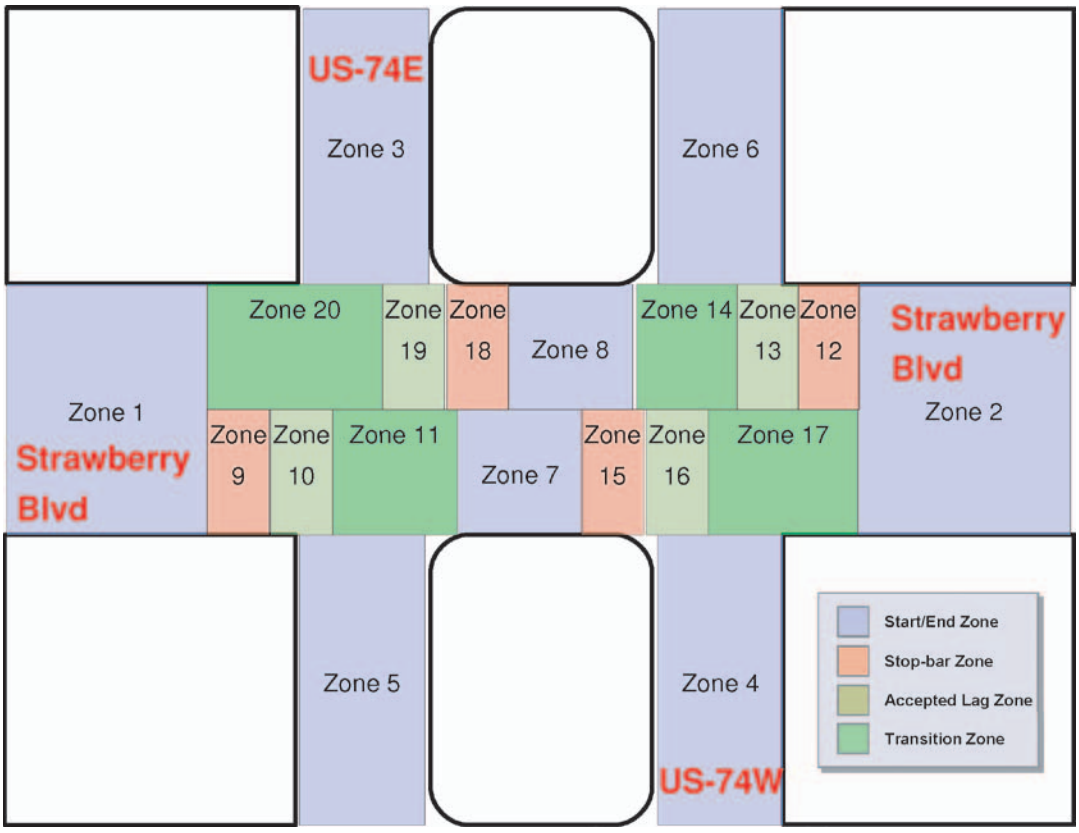


Figure B.1. Zone convention for the North Carolina site.

Schema

Presently, nine tables make up the database schema for each site: (1) gap_data, (2) sensor_collector, (3) sensor_status, (4) tracked_targets, (5) vehicle_acpt-lag, (6) vehicle_avail_gap, (7) vehicle_region, (8) vehicle_time, (9) vehicle_zone. Of most importance to the extraction of near-crash information are the relational tables: vehicle_acpt_lag, tracked_targets, sensor_collector, and vehicle_zone (Tables B.1–B.4). Other tables—vehicle_time and vehicle_region—can be used to "follow" a known tracked target and qualify time ranges of interest and so forth. The simplest way to mine interesting traffic events is to query the dates and times of specific events within the vehicle_acpt_lag table. The returned date-time stamps can then be used to develop secondary queries to extract actual tracked-target trajectories within the tracked_targets table.

Regarding the vehicle classification attribute veh_class: Although nine classification categories are detected, they are grouped into four categories that can be reliably discriminated (within 5% accuracy). They are as follows:

1. Motorcycles, sedans, and small SUVs (Classes 1 and 2);
2. Large SUVs and pickup trucks (Class 3);
3. Single-unit trucks (Class 4–Class 6); and
4. Semitrucks and other large vehicles (Class 7 and higher).

The optimal decision boundaries of the class-discriminate function are based on using vehicle height *and* length.

For most of the intersection decision support (IDS) sites to date, there are 14 maneuver types; most contain two sub-maneuver actions.

A close cousin to the vehicle_zone table is the vehicle_region table. The attributes are similar except "region" is substituted for "zone." The primary keys are therefore targetid and region. There is also the zone attribute, which can be used for relational outer-joins with vehicle_zone to obtain all region-based events of the tracked target, as follows:

```
SELECT zone, region, enter_region_date, enter_region_time,
exit_region_date, exit_region_time, enter_region_speed, exit_
region_speed FROM vehicle_region AS vr OUTER JOIN
vehicle_zone AS vz ON (vz.targetid = vr.targetid, vz.zone =
vr.zone);
```

The vehicle_time table contains relationships defining when a tracked target was first and last observed, as well as other information about the tracked target. Table B.5 lists the attributes.

Table B.1. vehicle_acpt_lag

Attribute	Type	Description/Comments
targetid	Double	Primary key, target vehicle ID of tracked maneuver.
maneuver	Integer	Primary key, maneuver type, if not defined, equals a blank space.
veh_class	Integer	Vehicle type (1–9 categories) by experimentally determined length/max-height ranges.
sub1_acpt_lag_zone	Integer	Submaneuver accepted lag zone number.
sub1_acpt_lag_zone_exit_date	Date	Date front of vehicle crossed over the lag zone.
sub1_acpt_lag_zone_exit_time	Time without time zone	Time front of vehicle crossed over the lag zone.
sub1_stopbar_zone	Integer	Submaneuver.
sub1_stopbar_zone_exit_date	Double var length array	Date front of vehicle crossed over the stop-bar zone.
sub1_stopbar_zone_exit_time	Double var length array	Time front of vehicle crossed over the stop-bar zone.
sub1_acpt_lags	Double var length array	Accepted lag times in seconds of mainline vehicles for each conflicting lane (front of side vehicle has crossed over lag zone). Typically, length = 2 for two lanes w.r.t. intersection center reference.
sub1_acpt_gaps	Double var length array	Accepted gap times of mainline vehicles in seconds for each conflicting lane (front of vehicle has crossed over stop-bar zone [not pulled onto mainline]) w.r.t. intersection center reference.
sub1_rej_gaps	Double var length array	Same convention as accepted gaps but rejected gaps (vehicle did *not* proceed with a maneuver).
sub1_stopbar_acpt_lags	Double var length array	Accepted lag times in seconds of mainline vehicles for each conflicting lane (front of side vehicle has crossed over stop-bar zone). Typically, length = 2 for two lanes w.r.t. intersection center reference.
sub1_stopbar_rej_lags	Integer var length array	Accepted lag times in seconds of mainline vehicles for each conflicting lane (front of side vehicle has crossed over stop-bar zone). Typically, length = 2 for two lanes w.r.t. intersection center reference.
sub2_acpt_lag_zone	Integer	Second submaneuver portion of maneuver = n.
sub2_acpt_lag_zone_exit_date	Date	Date front of vehicle crossed over the second lag zone.
sub2_stopbar_zone	Integer	Same convention as sub1 attributes but for second submaneuver.
sub2_acpt_lag_zone_exit_time	Time without time zone	Same convention as sub1 attributes but for second submaneuver.
sub2_stopbar_zone	Integer	Same convention as sub1 attributes but for second submaneuver.
sub2_stopbar_zone_exit_date	Double var length array	Same convention as sub1 attributes but for second submaneuver.
sub2_stopbar_zone_exit_time	Double var length array	Same convention as sub1 attributes but for second submaneuver.
sub2_acpt_lags	Double var length array	Same convention as sub1 attributes but for second submaneuver.
sub2_acpt_gaps	Double var length array	Same convention as sub1 attributes but for second submaneuver.
sub2_rej_gaps	Double var length array	Same convention as sub1 attributes but for second submaneuver.
sub2_stopbar_acpt_lags	Double var length array	Same convention as sub1 attributes but for second submaneuver.
sub2_stopbar_rej_lags	Integer var length array	Same convention as sub1 attributes but for second submaneuver.

Table B.2. tracked_targets

Attribute	Type	Description/Comments
date	Date	Primary key.
time	Time without time zone	Primary key.
num_targets	Integer	Defines number of elements in remaining table attributes.
targetid	Double var length array	Unique target ID, based on time-date encoded number in the mantissa and a unique identifier in the abscissa.
x_state	Double var length array	NAD83 state plane coordinate, meters.
x_state	Double var length array	NAD83 state plane coordinate, meters.
x_dot_state	Double var length array	Estimate of velocity, w.r.t. NAD83 state plane coordinate system.
y_dot_state	Double var length array	Estimate of velocity, w.r.t. NAD83 state plane coordinate system.
speed	Double var length array	Speed, m/s.
heading	Double var length array	Vehicle heading, radians.
acceleration	Double var length array	m/s^2
distance_to_intersection	Double var length array	w.r.t. intersection center.
time_to_intersection	Double var length array	In seconds, if not calculated, equals 1,000.
cur_lane	Integer var length array	Mainline lanes; lane numbers defined in map file for each site.
region	Integer var length array	Refer to appropriate region map and datafile.
sensor_type	Integer var length array	Refer to metadata description for sensor type, equals –1 for undefined.
sensor_index	Integer var length array	Refer to metadata description for sensor number and associated state plane coordinate. Location, equals –1 for undefined.
length	Double var length array	Estimate of vehicle length, if not defined equals 0.
max_height	Double var length array	Estimate of vehicle height, if not defined equals 0.

Other Relational Tables

There are two other tables within each site database that have not been considered for extraction of near-misses or crash data. Essentially, they can be used to derive traffic characteristics (related to lag times) and sensor operation over desired time periods. The gap_data table (Table B.6) contains more global information about primary, secondary, and tertiary lags for every time stamp during field operation. The size of the variable length array is based on the number of lanes used for each lag attribute (table name does not reflect that it represents mainline lag times). If there is no measurable lag, the lag times are set to 1,000 (sec).

Lastly, the sensor_status table, summarized in Table B.7, contains the status of all sensors at the site for each time stamp recorded during operation. It can be used to ensure that all the roadside data collection sensors were operational over the time period of the near-miss event candidates under study. This is achieved by querying for any values equal to zero in the sensor_status attribute.

Currently Available Crash Data

The crashes involve two events from the Minnesota Highway 52 fixed site and one from the North Carolina site. The project team obtained four videos for the three crashes that were date-time stamped (one of the crashes had video from two camera perspectives). The status of these crash data sets is summarized in Table B.8.

The project team harvested target trajectories of the main-road and side-road vehicles for one of the crashes (date stamped: 4-06-2007, which occurred at the Highway 52 site). In this case, the side-road vehicle did not contain a valid tracked-target trajectory. Rather, the team filtered the raw sensor collector data, which contained the front-vehicle targets of the side-road vehicle. Targets through the center zones of the intersections are missing.

In terms of near crashes, the general strategy to harvest data has been to search for occurrences of side-road vehicles accepting short lags for certain maneuver types and then to use MATLAB to visually infer potential "near crashes" for

Table B.3. sensor_collector

Attribute	Type	Description/Comments
date	Date	Primary key, date stamp.
time	Time without time zone	Primary key, time stamp.
num_targets	Integer	Number of raw sensor target points.
sensortimestamp	Double var length array	Internal sensor time stamp.
x_local	Double var length array	Sensor target point w.r.t. sensor coordinate system.
y_local	Double var length array	
x_dot_local	Double var length array	Sensor estimate of velocity.
y_dot_local	Double var length array	Sensor estimate of velocity.
x_state	Double var length array	Sensor target point w.r.t. NAD83 state plane system.
y_state	Double var length array	Sensor target point w.r.t. NAD83 state plane system.
x_dot_state	Double var length array	Sensor target point w.r.t. NAD83 state plane system.
y_dot_state	Double var length array	Sensor target point w.r.t. NAD83 state plane system.
height	Double var length array	Vehicle max. height.
length	Double var length array	Vehicle length.
num_sensors	Integer	Number of entries for the following three attributes: sensor_index, sensor_type, and sensor_status.
sensor_index	Integer var length array	The sensor number as defined in the XML metadata file.
sensor_type	Integer var length array	The sensor type as defined in the XML metadata file.
sensor_status	Integer var length array	Sensor status.

Table B.4. vehicle_zone

Attribute	Type	Description/ Comments
targetid	Date	Primary key, date stamp.
zone	Integer	Primary key.
region_first_seen	Integer	
enter_zone_date	Date	
enter_zone_time	Time without time zone	
region_last_seen	Integer	
exit_zone_date	Date	
exit_zone_time	Time without time zone	
enter_zone_speed	Double	
enter_zone_inter_dist	Double	
veh_time_in_zone	Double	
min_speed_zone	Double	
max_speed_zone	Double	
ave_speed_zone	Double	

Table B.5. vehicle_time

Attribute	Type	Description/ Comments
targetid	Double	Primary key, tracked target.
date_first_seen	Date	
time_first_seen	Time without time zone	
region_first_seen	Integer	
date_last_time	Date	
time_last_seen	Integer	
region_last_seen	Integer	
zone_first_seen	Integer	
zone_last_seen	Integer	
maneuver	Double	
veh_time_tracked	Double	
veh_length	Double	
veh_height	Double	
veh_min_speed	Double	
veh_max_speed	Double	
veh_avg_speed	Double	

Table B.6. gap_data

Attribute	Type	Description/Comments
date	Date	Primary key, date stamp.
time	Time without time zone	Primary key, every time step during field operation.
num_lanes	Integer	
primary_gap	Double var array	First vehicle lag times in each lane; defaults to 1,000 if there is no measurable lag.
secondary_gap	Double var array	Second vehicle lag times in each lane; defaults to 1,000 if there is no measurable lag.
tertiary_gap	Integer	Third vehicle lag times in each lane; defaults to 1,000 if there is no measurable lag.

Table B.7. sensor_status

Attribute	Type	Description/Comments
date	Date	Primary key, date stamp.
time	Time without time zone	Primary key, time stamp.
num_sensors	Integer	Total number of sensors used at the site.
sensor_status	Integer var array	Stores "1" for good, "0" for faulty/not operational.
sensor_index	Integer var array	Sensor index ID numbers.
sensor_type	Integer var array	List of the type of sensor (LIDAR, RADAR, and so on) corresponding to each sensor index.

further verification and analysis. For example, there may be points in the mainline trajectories where the tracked target changed lanes well enough ahead of the side-road vehicle to "easily" avoid a collision. Or perhaps the mainline vehicle was not in a conflicting lane at all. This is not possible to detect based on a query alone.

Initially the attempt at drilling down to find this data was not successful. This was primarily because of a "bug" in the accepted lag/gap calculations stored within the site databases. However, this bug has now been resolved and the database of particular interest (North Carolina) has been rebuilt and reinstalled.

The database will essentially be queried for straight-line cross-intersection maneuvers that are similar to the attempted maneuver by the side vehicle involved in the crash at the North Carolina site. In essence, the side-road vehicle must accept two lags, one from each mainline direction, during its mainline crossing maneuver. The criteria of the accepted lag calculation is arbitrary, of course, but times of less than 3 s have been chosen; that is, as the vehicle crosses the road, the first submaneuver of the vehicle crossing into the median *or* the second, last submaneuver of the vehicle crossing over the remainder of the mainline road accepted lags of 3 s.

Table B.8. Current Crash Data Sets

Site, Date of Crash	Camera Data	Database Trajectories	Accident Reports	Comments
HW52/CSAH9, 4/6/2007	1 camera	Y	Y	Side vehicle trajectory available as raw sensor data from camera tracking system.
HW52/CSAH9, 09/21/2007	2 cameras	N	Y	Side vehicle trajectory available as raw sensor collector (SC) data from camera tracking system.
North Carolina	1 camera	Y	N	Side vehicle tracked-targets file is not correct; advice is to go back to raw, sensor collector data for visual observation of potential "matches" of moving side vehicle target.

```
SELECT targetid, maneuver, veh_class, sub1_acpt_lag_zone_
exit_date, sub1_acpt_lag_zone_exit_time FROM vehicle_
acpt_lag WHERE
   (maneuver = 1) AND
   ((sub1_acpt_lags < 3.0) OR (sub2_acpt_lags < 3.0));
```

The returned date-time stamps and tracked-target ID of the vehicles that completed the maneuvers will be used as part of the criteria to harvest tracked-target and raw sensor collector data. The other part of the search criteria, particularly the amount of time before the accepted lag event times occurred, needs to be considered. It is desired to observe the side vehicle trajectory starting immediately before the vehicle entered the stop-bar region of the side road. The vehicle_zone relational table will be used to obtain the date-time when the particular target ID entered the first stop-bar zone. In this case, the desired zone number is specified in the following query:

```
SELECT enter_zone_date, enter_zone_time FROM vehicle_
zone WHERE
   (zone == 9) AND (targetid == < returned targetid from
      vehicle_acpt_lag >);
```

Optionally, the completion time for the near-miss candidate to be observed could be defined using the same vehicle_zone table as follows:

```
SELECT exit_zone_date, exit_zone_time FROM vehicle_zone
WHERE
   (zone == 17) AND (targetid == < returned targetid from
      vehicle_acpt_lag >);
```

Note that zones 9 and 17 represent the stop-bar and transition zone, respectively, for the given maneuver type. For the side-road maneuver crossing the other direction, maneuver=2, the stop-bar and exiting transition zones are 12 and 20, respectively. Yet another method to obtain a more "broad" envelope is to query the vehicle_time relational table as follows:

```
SELECT zone_first_seen, date_first_seen, time_first_seen,
zone_last_seen, date_last_seen, time_last_seen FROM vehicle_
time WHERE (targetid == < returned targetid from vehicle_
acpt_lag >);
```

In either case, the time ranges are then used to harvest the tracked_targets relational table within the desired envelope. Note that possible mainline lane changing maneuvers by the potentially conflicting vehicle are not precisely characterized within the tracked-targets data. That is, after the lateral movement of the vehicle exceeds an experimentally determined threshold, the tracked-target attribute lane will be updated accordingly. Thus, in addition to the tracked-targets data, the raw sensor collector data will be harvested over the same period for later analysis. In order to use the raw sensor collector data, however, a completely new predictive filter will need to be developed to extract the desired vehicles to "track," in addition to creating optimal lateral paths for the vehicle. Note that the lateral error of the radar sensors used to obtain the mainline vehicle trajectories incur large lateral errors because of surface ambiguity of the vehicle's front or rear bumper and surfaces.

APPENDIX C

Outline of a Causal Theory of Traffic Conflicts and Collisions

Gary A. Davis, John Hourdos, and Hui Xiong

Department of Civil Engineering, Minnesota Traffic Observatory, University of Minnesota

Using recent developments in causal analysis, a minimal model capable of rigorously representing traffic conflicts and crashes is constructed. This model is then used to derive relationships between these types of events. The first result indicates that the magnitude of the minimum sufficient evasive action can be used to partition the space of background conditions, leading to a natural scale for ranking the severity of conflicts. The second result indicates that crashes can possibly arise from any region of the space of background conditions with the contributions of different regions of the background space to the crash population being weighted by evasive action. The third result gives a counterfactual definition of a type of conflict called a close encounter and relates the relative frequency of close encounters to that of crashes, with the evasive action determining the crash-to-conflict ratio. It is then illustrated how trajectory-based reconstruction can be used to classify close encounters with respect to seriousness and to estimate the potential number of crashes in a set of close encounters.

Introduction

It is a commonplace observation that certain drivers or locations experience a higher frequency of traffic crashes than do other comparable entities. Hauer (1997) has identified the expected number of crashes of a stated type over a stated time interval as an appropriate measure of the safety of an entity; other things being equal, safety can be estimated by counting the crashes that have occurred. The relative rarity of crashes, however, means that even for entities and crash types of relatively high frequency, such as multiple vehicle crashes at busy urban intersections, observation periods of several years may be needed to produce estimates with acceptable statistical properties. This has led to a search for surrogate events that occur at more easily observable rates and whose frequency or nature are indicative of the frequency or nature of crashes.

The hypothesis that noncatastrophic events might provide information concerning the nature of catastrophic events is not new and can be found, for example, in the use of "critical incidents" in evaluating the proficiency of air traffic controllers and assessing pilot errors (Flanagan 1954). In traffic engineering, the first use of traffic conflicts as crash surrogates is usually attributed to Perkins and Harris (1968), but during the 1970s several versions of the traffic conflict technique were developed and employed in both North America and Europe (Asmussen 1984). The Association for International Cooperation in Traffic Conflict Techniques proposed the following definition of traffic conflict:

> A traffic conflict is an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged. (Guttinger 1984, p. 18)

In what follows, we will first briefly review the literature on conflicts and surrogate events, with a focus on three interrelated issues. The first issue concerns Heinrich's Triangle as a model for traffic safety. The second issue concerns the grading of conflicts or surrogate events with respect to seriousness or severity. The third issue concerns the relation between crash frequencies and conflict frequencies.

Heinrich's Triangle

Heinrich's Triangle is a name given to the hypothesis that events can be ranked in order of increasing severity but decreasing frequency. This idea is often represented graphically with

a pyramid divided into horizontal sections, each section representing a class of events. The height of a section from the pyramid's base represents the severity of the corresponding class of events, while the volume of the section represents the relative frequency for that class. For example, fatal accidents are least frequent, so they would occupy the tip of the pyramid. Below fatal accidents might be injury accidents and then noninjury accidents, close calls, and so forth. Heinrich (1959) developed this hypothesis with regard to industrial accidents and conjectured that stable relative frequencies could be found between the levels of a pyramid. In traffic safety, qualitative versions of Heinrich's Triangle have been invoked for conceptual illustration (e.g., Hauer 1997; Svensson and Hyden 2006), while Dingus et al. (1999) pointed out that establishing predictable relationships between the levels of a Heinrich's Triangle could justify using observations of near-crash events to evaluate the potential safety benefits of crash-avoidance technologies. Dingus et al. (2006) then provided empirical support for such a triangular relationship between conflicts, near crashes, and crashes observed in the 100-Car Naturalistic Driving Study.

Grading Conflict Severity

Interest in grading the severity of conflicts arose relatively early in the development of traffic conflict methods, in connection with difficulties encountered in establishing empirical relationships between conflict counts and observed crash frequencies. For example, Guttinger (1984) and Baguley (1984) presented selected graphs showing crashes as increasing functions of conflicts, but more comprehensive studies by Cooper (1984) and Migletz et al. (1985) reported instances where crash experience appeared to increase, appeared to have no relation, or appeared to decrease as measures of conflicts increased. If a Heinrich's Triangle-like relationship is accurate, then it would be expected that establishing a relationship between events in neighboring levels of the pyramid would be easier than between events at one level and events aggregated over several lower levels. This in turn means that objective criteria would be needed for assigning events to their appropriate severity levels. As an example, in a study of vehicle and pedestrian interactions, Guttinger reported a clear association between crash frequency and the frequency of serious conflicts, which were defined as follows:

> Serious conflict: a sudden motor reaction by a party or both of the parties involved in a traffic situation towards the other to avoid a collision, with a distance of about one metre or less between those involved. (Guttinger 1984, p. 19)

Here, a serious conflict is defined using both the physical separation between the involved entities and the "suddenness" of evasive action. For vehicle-to-vehicle interactions, Hayward (1972) introduced the use of time to collision (TTC) as a measure of conflict severity. At a given instant, TTC is the time at which two road-using entities would collide if each persisted on its present course. Hayward illustrated how, when plotting TTC versus time, TTC would first decrease to a minimal value and then increase after one or more of the involved entities initiated evasive action. The minimum value of TTC for an encounter was taken as a measure of how close the encounter was to an actual collision.

In the Swedish traffic conflict technique, TTC at the instant a road user initiates evasive action, rather than the minimum TTC over a time interval, was taken as the measure of conflict severity. In its initial implementations at an urban intersection, the Swedish TTC defined serious conflicts as those with TTC < 1.5 s. Using this definition, estimates of the crashes per conflict with orders of magnitude of 10^{-5} were reported (Hyden and Linderholm 1984). When attempting to extend this idea to rural intersections, however, it was found that vehicle speed also needed to be considered, ultimately leading Svensson and Hyden (2006) to grade severity as a function of the ratio of TTC to closing speed (CS),

$$\text{Severity} = f(\text{TTC}/\text{CS}) \tag{1}$$

with higher values of TTC/CS indicating conflicts of lower severity. This characterization is especially interesting because TTC/CS is inversely proportional to the minimum deceleration needed to bring the vehicle to a stop in the TTC interval. Svensson and Hyden also advanced the intriguing hypothesis that severe conflicts result from a mismatch between a road user's expectations and actual events.

A similar approach to grading the severity of vehicle interactions has been developed in the United States to support evaluation of collision warning systems. In Smith et al. (2002), the situation between leading and following vehicles was characterized in terms of the current separation between the two vehicles, called the range (R), and the speed difference between the vehicles, called the range rate (RR). Using driver behavior data observed in driving simulators and test-track experiments, range versus range-rate curves were then used to partition the set of possible range and range-rate combinations into subsets reflecting crash, near-crash, conflict, and low-risk situations. For example, the relationship

$$R = RR^2/(2(.65g)) \tag{2}$$

where g denotes the gravitational acceleration, was suggested as roughly dividing the crash and near-crash conditions.

In Najm et al. (2006), a TTC was defined via

$$TTC = \frac{R}{RR} \tag{3}$$

and combinations of TTC versus RR were used to define boundaries between near-crash and conflict situations. Since range rate can also be interpreted as closing speed (CS), there is a clear similarity between this approach to grading severity and that developed by Svensson and Hyden.

Finally, in the 100-car study, near-crash events were defined explicitly in terms of the magnitude of a required evasive action, as

> Any circumstance that requires a rapid, evasive maneuver by the subject vehicle, or any other vehicle, pedestrian, cyclist, or animal to avoid a crash. A rapid, evasive maneuver is defined as a steering, braking, accelerating, or any combination of control inputs that approaches the limit of the vehicle capabilities. (Dingus et al. 2006)

Since, as pointed out, the ratio TTC/CS is inversely proportional to a stopping deceleration, we have an additional overlap among the proposed methods for grading conflict severity.

Relationship Between Crashes and Conflicts

The third issue identified involves the relationship, if any, between crash frequency and conflict frequency. Early efforts focused on attempting to establish correlations between conflict and crash frequencies, but Hauer (1982) noted that if conflicts are treated as crash opportunities, some of which actually result in crashes, then the relationship between conflicts and crash frequency should take the form

$$\text{Expected number of crashes} = (\text{Number of conflicts}) \times (\text{crash-to-conflict ratio}) \quad (4)$$

That is, the number of conflicts is a measure of crash opportunities, while the crash-to-conflict ratio reflects the probability that a given crash opportunity results in a crash. In this case, the theoretical correlation between the random variables generating crash and conflict frequencies will depend both on the magnitude of the crash-to-conflict ratio and on the variability in the number of conflicts, and so could be low even when these quantities are known perfectly. Hauer and Garder (1986) gave a more extensive treatment of this approach, pointing out that the usefulness of Equation 4 as a predictive tool depended first on the ability to estimate the crash-to-conflict ratio with acceptable precision and second on the stability of this ratio across different entities or times.

If conflicts of different degrees of severity have different probabilities of resulting in crashes, then a study that mixes together conflicts of varying severity may find it difficult to identify a stable crash-to-conflict ratio. This then leads to the problem of identifying groupings of conflicts having stable ratios. One possible solution is to interpret the crash-to-conflict ratio as the probability that a conflict results in a crash and then allow this probability to vary continuously with a measure of conflict severity. For example, Saunier and Sayed (2008), drawing on work by Hu et al. (2004), have conjectured that, when conflict severity is measured using TTC, a relationship of the form

$$P(\text{crash}|TTC) = \exp\left[-(TTC/\sigma)^2\right] \quad (5)$$

may prove useful.

An alternative approach to relating observed conflict frequencies to crash likelihood is to treat crashes and conflicts as events varying continuously on some dimension, with crashes being at an extreme of this dimension. It might then be possible to apply extreme-value statistics to estimate the probability of these extreme events. This approach was initially proposed by Campbell et al. (1996), who applied it to three measures of car-following obtained from a study of drivers in instrumented vehicles: range, TTC, and available reaction time. For each of these variables, a value of zero defines the point at which a collision occurs; and for each of these variables, the fitted extreme-value distribution assigned a probability of zero to crash events. The authors pointed out that for this method to be valid, crash and noncrash events should result as random outcomes from the same underlying probability distribution, and that if crash events and normal-driving events come from different populations, then the standard assumptions of extreme-event statistics may not be satisfied.

This is an important point, and it will be helpful to give it a more formal restatement. Suppose driving events can be divided into two subsets, "normal driving" and "close calls." Let y denote a numerical variable characterizing these events, and let yc denote a critical value for y such that when $y < yc$, a crash occurs. Then the probability of a crash is given by the appropriate mixture of the two sets of conditions

$$P(\text{crash}) = P(y < yc|\text{normal driving})P(\text{normal driving}) + P(y < yc|\text{close call})P(\text{close call}) \quad (6)$$

Given an adequate sample of normal driving conditions, standard extreme-value methods could be used to approximate $P(y < yc|\text{normal driving})$, and given a sample of close calls, extreme-value methods could be used to approximate $P(y < yc|\text{close call})$. However, if $P(y < yc|\text{normal driving}) \neq P(y < yc|\text{close call})$, extreme-value methods applied to observations of normal driving conditions will generally not be sufficient to estimate crash probabilities.

Analytic results relevant to this issue have been given by Al-Hussaini and El-Adll (2004). One of the key features of extreme-value statistics is that the distribution functions of extreme values from random samples tend to converge, after suitable normalization, to one of only three limiting distributions. This means that if it is possible to identify which of these three distributions is appropriate for a given problem, then predictions of extreme events can be accomplished even when detailed knowledge of the underlying probability model is lacking. Al-Hussaini and El-Adll (2004) show that the classic convergence results do not necessarily apply to data generated by finite mixtures of distributions, such as Equation 6. In this case, the limiting distribution for the extreme values can depend both on the type of mixture components and on their relative weights.

One approach to solving this problem would be to identify an observable variable that could serve as a proxy for the different driving conditions and then allow one's extreme-value model to depend on this. Apparently independently of the aforementioned work, Songchitruksa and Tarko (2006) applied extreme-event methods to observations of postencroachment time (PET) obtained from video recordings of vehicle encounters at intersections. By allowing their extreme-value distributions to vary with measures of traffic flow, the authors were able to generate predicted crash frequencies that were, at least for some intersections, consistent with observed crash frequencies.

To summarize, there has been a long-standing interest in using surrogate events such as traffic conflicts as a proxy for crashes, and a long-standing belief that it should be possible to relate the expected number of crashes to a measure of conflict frequency. Initial difficulties in establishing stable crash-to-conflict relationships led to an interest in grading the severity of conflicts, with the idea being that severe conflicts would be more reflective of crashes. Currently in the literature there are descriptions of several methods for assessing severity of a noncrash event, which rely on measures of the kinematic variables, and show some overlap.

One point to emphasize is that the working definitions of observable conflicts cited previously all have counterfactual components. That is, a conflict is defined as a noncrash event where, had things been different, a crash would have resulted. Counterfactuals are somewhat difficult to address rigorously (Lewis 1973), but recently Pearl (2000) has described formal tools that allow one to assess probabilities assigned to both factual and counterfactual statements. In what follows, we will use some of these tools to give a rigorous counterfactual definition of traffic conflict and then use that definition to derive a relationship between the relative frequency of crashes and that of conflicts of differing severity, where it turns out that the crash-to-conflict ratios are governed by drivers' ability to achieve successful evasive action. We will then illustrate how these formal results might be applied to actual crash and near-crash events.

Causal Model of Crashes and Conflicts

Our starting point is Pearl's (2000) notion of a causal model, which consists of a set of exogenous variables, a set of endogenous variables and, for each endogenous variable, a structural equation describing how that variable responds to changes in other model variables. A causal model can be represented qualitatively using a directed graph, with the nodes of the graph representing variables and directed arrows indicating direct causal dependencies. Figure C.1 displays the simple graphical model that underlies our treatment. The node u, which may be vector-valued, denotes variables describing background conditions. The node x denotes the variable describing evasive or avoidance action, and the node y is a crash-related outcome. Associated with y is a critical value yc such that if $y(u,x) < yc$, then a crash results.

In what follows, we will assume that the evasive action x takes on values from a discrete set $X = \{x_1, x_2, \ldots, x_n\}$, with $x_1 < x_2 < \cdots < x_n$, and the background conditions take one value from a denumerable set U. Since continuous ranges of possible evasive action values or background conditions can be approximated arbitrarily well by discrete sets, this is a relatively weak constraint. It will, however, allow us to derive results using elementary mathematics and avoid some of the technicalities that arise when treating general sets X.

To complete the model specification, we need a probability distribution over the values taken on the background variables, denoted by $P(u)$, and a structural equation describing how evasive actions are selected. However, the results we will develop require only the weaker condition that we have a conditional probability distribution for the evasive action, denoted by $P(x|u)$. The conditional independence structure implied by the graphical model in Figure C.1 then means that the full joint distribution for our model factors into one with the form

$$P(y,x,u) = P(y|x,u)P(x|u)P(u) \tag{7}$$

The nature of u and x and the functional form for $y(u,x)$ will, of course, vary depending on the type of crash under

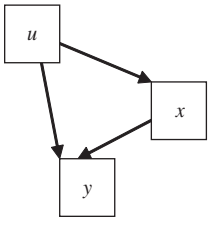


Figure C.1. Simple graphical model of crash events depicting conditional independence structure.

consideration. For example, in Brill's (1972) simple two-vehicle rear-ending collision model, the initial speed and braking deceleration of the leading vehicle can be denoted by v_1 and a_1, respectively, while v_2 and a_2 denote the initial speed and braking deceleration of the following vehicle, and h_2 and r_2 denote the following headway and reaction time of the following driver. A collision occurs when the stopping distance available to the following driver is less than that needed to stop without colliding with the lead vehicle. That is, when

$$v_2 h_2 + \frac{v_1^2}{2a_1} - \left(v_2 r_2 + \frac{v_2^2}{2a_2} \right) < 0 \tag{8}$$

If we take the following driver's deceleration as the avoidance action, then the variables (v_1, a_1, v_2, r_2, h_2) would be components of u, the evasive action x would be a_2, and the outcome function would be

$$y(u,x) = v_2 h_2 + \frac{v_1^2}{2a_1} - r_2 v_2 - \frac{v_2^2}{2a_2} \tag{9}$$

Placing a probability distribution over the values taken on by the model's exogenous variables produces what Pearl (2000) calls a probabilistic causal model. This distribution, together with the structural equations for the endogenous variables, is then sufficient to compute probabilities assigned to both factual and counterfactual events.

As another example, consider the interaction of major and minor approach vehicles at a two-way-stop controlled intersection, where the minor approach vehicle crosses in front of the major approach vehicle.

Let:

y_2 = initial distance of major approach vehicle from collision point
v_2 = initial speed of major approach vehicle
r_2 = reaction time of major approach driver
a_2 = deceleration of major approach driver
y_1 = initial distance of minor approach vehicle from collision point
v_1 = initial speed of minor approach vehicle
a_1 = acceleration of minor approach driver

Here, the deceleration of the major approach driver is taken as the evasive action and the remaining variables describe background conditions. The time needed by the minor approach driver to clear the collision point is

$$\hat{t}_1 = \frac{\sqrt{v_1^2 + 2a_1 y_1}}{a_1} \tag{10}$$

The distance separating the major approach from the collision point at this time is

$$y(u,x) = \begin{cases} y_2 - v_2 \hat{t}_1, & r_2 > \hat{t}_1 \\ y_2 - v_2 \hat{t}_1 + \dfrac{a_2 (\hat{t}_1 - r_2)^2}{2}, & r \le \hat{t}_1 \end{cases} \tag{11}$$

A collision is taken to occur when $y(u,x) < 0$.

In what follows, we will assume that the function $y(u,x)$ is monotonic in x, in the sense that, for each u, $x_1 > x_2$ implies that $y(u,x_1) > y(u,x_2)$. That is, for a given set of background conditions, increasing the magnitude of the evasive action does not produce a crash where one would not have occurred otherwise. It is straightforward to verify that, given reasonable constraints of the background conditions, this monotonicity condition is satisfied by both the above models.

Model Properties

The first consequence of our model is that set X, together with the monotonic structural equation $y(u,x)$, can be used to partition the set of background conditions. That is, the sets

$U_1 = \{u: y(u,x_j) \ge yc, j = 1, \ldots, n\}$
$U_2 = \{u: y(u,x_1) < yc \;\&\; y(u,x_j) \ge yc, j = 2, \ldots, n\}$
. . .
$U_j = \{u: y(u,x_i) < yc, i = 1, \ldots, j-1 \;\&\; y(u,x_i) \ge yc, i = j, \ldots, n\}$
. . .
$U_{n+1} = \{u: y(u,x_i) < yc, j = 1, \ldots, x_n\}$

form a partition of the set of background conditions. A proof of this claim is given in the appendix to this paper.

One way to interpret the sets $U_1, \ldots, U_{n+1}$ is as follows. Define the function

$x\min(u)$ = the smallest value x_j such that $y(u,x_j) \ge yc$.

For each possible background condition u, $x\min(u)$ gives the lowest value of the evasive action that prevents a crash. U_1 is then the set of conditions where a crash never occurs, U_j is the set of all conditions for which $x\min(u) = x_j$, $j = 1, \ldots n$, and U_{n+1} is the set of conditions where a crash is unavoidable. The background conditions belonging to a given set in our partition share the same value for $x\min$, so each set in our partition is identified by its characteristic minimum successful evasive action. If $x\min(u)$ is taken as a measure of the severity of an event, then we have a partition of events with respect to severity.

The second consequence is that probability of a crash is a mixture of the crash probabilities from the different sets in our partition.

$$P(\text{crash}) = P(y < yc) = \sum_{j=2}^{n+1} P\left(x \le x_{j-1} \middle| u \in U_j\right) P\left(u \in U_j\right) \qquad (12)$$

(A proof of this claim is given in the appendix to this paper.)

This relationship states that, in principle, crashes can result from most regions of the space of background conditions, and it is the evasive action model that converts the possibility of crash into whether a crash actually occurs. Other things being equal, those regions where it is difficult to achieve a successful evasive action should contribute proportionally more to the set of crashes.

Our third result concerns the relationship between conflicts and crashes. As noted, working definitions of conflict and near crash have a counterfactual component: an observed conflict or near crash is a noncrash event that, had things been different, could have been a crash. A rigorous definition of conflict will thus require a method for formally stating this counterfactual condition, and one of the strengths of Pearl's theory is that it provides just such a method. Pearl (2000) has used this formal language to explicate different ideas regarding causal effects and to identify conditions that allow causal effects to be estimated from nonexperimental studies. Davis (2002) has argued that the concepts of causal effect used in statistical safety studies, crash simulation studies, and crash reconstruction can be treated as instances of what Pearl calls probability of necessity.

Following Pearl's treatment, the notation $y_{x=x0}$ will stand for the value taken on by the variable y when the variable x is set, counterfactually, to the value $x0$. Probabilities regarding counterfactual claims are then evaluated using the modified graphical model depicted in Figure C.2. In order to sidestep the surplus meaning attached to notions like conflict and near crash, we will define a close encounter as an event with $y \ge yc$ (crash does not occur), $x = x_j$ (evasive action of level x_j was observed), but $y_{x=x_{j-1}} < yc$ (had the next weakest evasive action x_{j-1} been employed, a crash would have occurred.) It can then be shown that

$$\begin{aligned} P(CE) &= P\left(y \ge yc \,\&\, x = x_j \,\&\, y_{x=x_{j-1}} < yc\right) \\ &= P\left(x = x_j \middle| U_j\right) P\left(U_j\right) \end{aligned} \qquad (13)$$

(A proof of this claim is given in the appendix to this paper.)

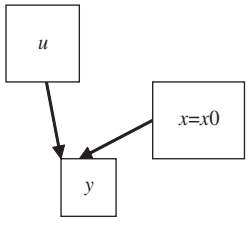


Figure C.2. Modified graphical model of crash for assessing counterfactual event x = x0.

Here the notation $P(U_j)$ has been used as shorthand for $P(u \in U_j)$. This result can then be used to eliminate $P(U_j)$ in Equation 13, giving the probability of a crash in terms of close encounters and evasive actions

$$P(y < yc) = \sum_{j=2}^{n+1} P\left(y \ge yc \,\&\, x = x_j \,\&\, y_{x=x_{j-1}} < yc\right) \left(\frac{P\left(x \le x_{j-1} \middle| U_j\right)}{P\left(x = x_j \middle| U_j\right)}\right) \qquad (14)$$

For background events in set U_j, a crash occurs if $x \le x_{j-1}$, while a close encounter occurs if $x = x_j$. The ratio $P(x \le x_{j-1} | U_j)/P(x = x_j | U_j)$ can thus be interpreted as a crash-to-conflict ratio, and Equation 14 states that crashes can result from close encounters of varying severity, with possibly different crash-to-conflict ratios. Equation 14 is a generalization of the relationship proposed in Equation 4, with an explicit representation of the crash-to-conflict ratio.

To summarize, Equation 12 states that, for this model, the probability of crash can be expressed in terms of the probability distribution for background conditions together with knowledge of how evasive actions are chosen. Equation 14 states that the probability distributions for the background conditions can be dispensed with if it is possible to find probability distributions for a type of traffic conflict we call a close encounter. In either case, though, it is the evasive action model that determines the relative contribution of a region in the space of background conditions to the population of crashes.

Applying the Results

Since the definition of a close encounter has both factual and counterfactual components, determining whether an observed event is a close encounter requires conducting a counterfactual test. This is most easily done when one has at hand the structural equation $y(u,x)$ and plausible estimates of the values for the event's background variables. For the simple rear-ending collision model described, Davis and Swenson (2006) have discussed how Bayes estimates of the background conditions and the evasive action can be obtained by fitting a model to a vehicle's trajectory data extracted from a video recording of the event. Given posterior distributions characterizing the uncertainty regarding the event's background conditions, it is then straightforward to compute the probability of a crash as a function of different values of the evasive action, using Monte Carlo simulation. Figure C.3 shows plots of the crash probability versus braking deceleration for two noncollision events observed on Interstate 94 by (the predecessor of) the Minnesota Traffic Observatory. For the encounter

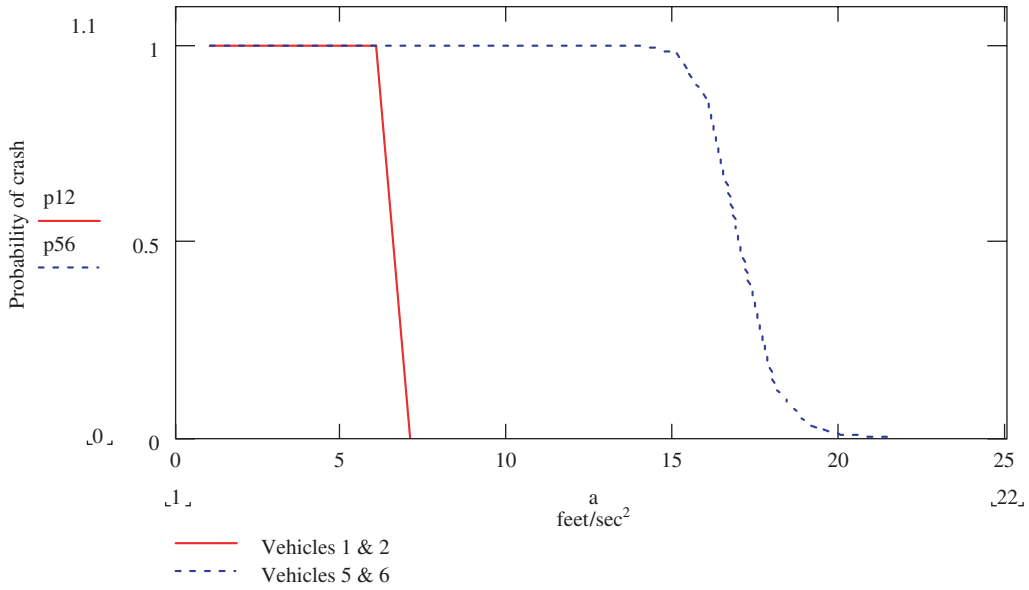


Figure C.3. Crash probability versus braking deceleration for two noncrash events. The event involving Vehicles 1 and 2 is arguably less severe than the event involving Vehicles 5 and 6.

between Vehicles 1 and 2, braking decelerations on the part of Driver 2 below about 7 ft/s^2 would probably have resulted in a collision, while for the encounter between Vehicles 5 and 6, braking decelerations by Driver 6 below about 15 ft/s^2 would probably have resulted in a collision.

The results shown in Figure C.3 make no reference to the actual braking decelerations used and so by themselves are not sufficient to determine if the event qualifies as a close encounter. Figure C.4, however, displays point estimates of actual and minimal braking decelerations for vehicles involved in three freeway rear-ending crash events, also taken from Davis and Swenson (2006), and including the two events illustrated in Figure C.3. The three rightmost points on Figure C.4 represent decelerations for colliding vehicles, while the points in the cluster to the left represent, with one exception, decelerations of successful stops before colliding. The first interesting observation from Figure C.4 is that the noncolliding drivers appeared to select decelerations close to the minima needed to stop without colliding. Since for each noncolliding vehicle a relatively small decrease in deceleration would have led, other things being equal, to a collision, each of the successful stops arguably qualifies as a close encounter. The second interesting observation is that the decelerations used by colliding drivers tend to be substantially greater than those used in the observed close encounters. Referring to Equation 14, the crash-to-conflict ratio for near-crash events for subset U_j takes the form

$$P\left(x \le x_{j-1} \middle| U_j\right) \Big/ P\left(x = x_j \middle| U_j\right) \tag{15}$$

This ratio will tend to be large for subsets where the minimum successful evasive action comes from the right-hand tail of the distribution of evasive actions. That is, other things being equal, crashes should tend to have extreme values of the evasive action, and this is what Figure C.4 shows. This gives us a possible interpretation of the empirical findings that the frequency of serious conflicts tends to be more reliably associated with the frequency of crashes.

The results in Figure C.4 also suggest a means for approximating the expected number of crashes from a set of noncrash events, where detailed knowledge of the evasive action model may not be necessary. If the limited data shown in Figure C.4 are typical of rear-ending events, the space of background conditions can be roughly divided into two subsets, one where crash occurrence tends to be negligible, characterized by $P(x < x\min(u)) \approx 0$, and one subset from which crashes tend to be generated—that is, where $P(x < x\min(u)) > 0$. If, in addition, the evasive action model can be taken as roughly constant in this crash-producing region, then an estimate of the number of crashes in a set of noncrash events can be computed as the sum of the probabilities that each of these events could have been a crash. That is, if $P(x = x_j | u = u_i) = P(x = x_j)$, then

$$P(\text{crash}) = P(y < yc) = \sum_j P\left(x\min > x_j\right) P\left(x = x_j\right) \tag{16}$$

Figure C.5 shows the two relationships depicted in Figure C.3, along with a proposed probability distribution for

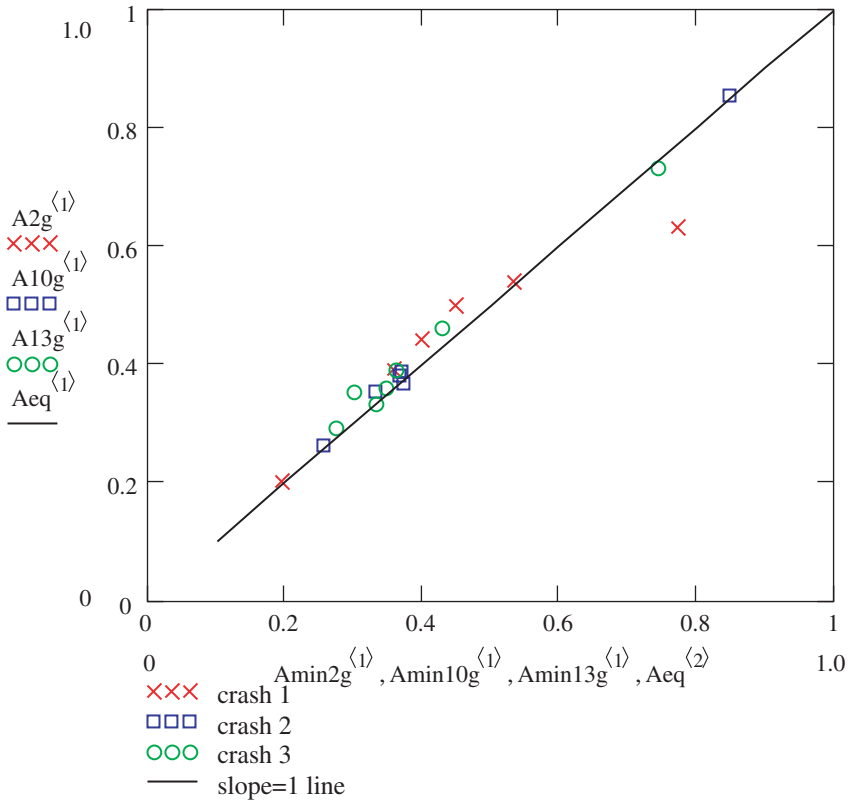


Figure C.4. Minimum versus observed decelerations (in g units) for vehicles stopping on a congested freeway. The x axis is minimum deceleration in g units, and the y axis is observed deceleration in g units. The three rightmost points are from colliding vehicles.

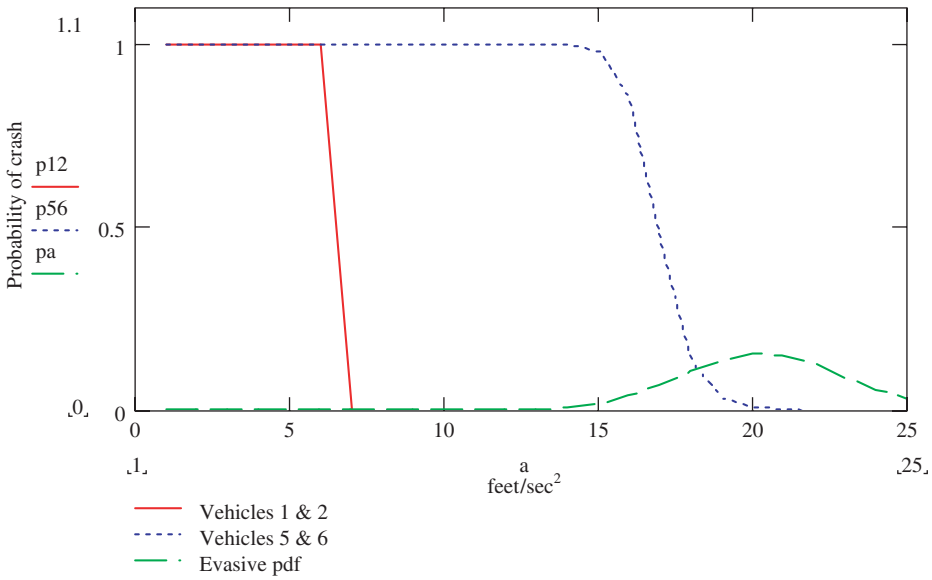


Figure C.5. Crash probability versus deceleration curves for two noncrash events, together with a proposed probability density for decelerations.

Table C.1. Crash Probabilities Computed for Five Leader–Follower Pairs of Noncolliding Vehicles

Vehicle Number		Estimated Follower Deceleration (ft/s²)[a]		
Leader	**Follower**	**Actual**	**Minimum**	***P*(crash)**
1	2	6.5 (.06)	6.2 (.06)	0
2	3	12.6 (.99)	11.4 (.66)	0
3	4	14.2 (.51)	12.8 (.43)	0
4	5	16.0 (.91)	14.4 (.63)	.004
5	6	17.3 (1.6)	17.1 (1.5)	.138
Sum				**.142**

[a] Standard deviations for deceleration estimate are given in parentheses.

emergency braking decelerations. This proposed distribution is based on a normal distribution, with mean and variance taken from the surprise braking tests reported by Fambro et al. (1997). Since for monotonic relationships $y(x,u) < yc$ only if $x\min(u) > x$, the function giving crash probability versus braking deceleration is equivalent to the function giving probability of failing to achieve the minimum safe deceleration versus braking deceleration. "Integrating" this function with respect to the braking deceleration distribution then gives the probability that the event could have been a crash had the following driver's deceleration been similar to that observed by Fambro et al. (1997). Table C.1 displays these counterfactual crash probabilities computed for five leader–follower pairs of noncolliding vehicles observed on Interstate 94 using the Fambro et al. (1997) statistics. Pairs 1–2 and 5–6 are the same, as shown in Figures C.3 and C.5. The expected number of crashes in this set, which is simply the sum of these probabilities, would be 0.142.

Discussion

Traffic collisions tend to be relatively rare but of nontrivial financial and personal consequence when they happen. It is not surprising then that substantial effort has been invested in finding useful surrogates for collisions, which are easier to observe but that still provide information on the frequency and nature of collisions. Our review of the literature on traffic conflicts and near crashes identified two important trends. The first is that conflicts can be ranked with respect to severity and that the more severe conflicts tend to show reliable associations with crash experience. The second is that working definitions of observed conflicts contain a counterfactual component, where a conflict is an event in which, had some action not occurred, a crash would have occurred.

In this paper, we postulated a minimal theoretical structure that incorporates crashes, near crashes, and conflicts; defined rigorously what we mean by a surrogate event; and then derived a relationship between crash propensity and the surrogate. Our treatment produced a natural, one-dimensional method for grading the severity of noncrash events in terms of their minimum successful evasive actions. Our treatment also produced a generalization of Hauer's proposed relationship between expected crashes and conflicts, where the crash-to-conflict ratio is governed by the evasive action model, suggesting that a better understanding of crash-to-conflict ratios could be had through a better understanding of evasive action. Finally, combining our treatment with some preliminary empirical work on rear-ending conflicts and crashes led to a suggested alternative method for estimating the expected number of crashes from a set of conflicts, where each conflict is assigned a probability of (counterfactually) being a crash.

It may occasionally happen that a theoretical treatment can substitute for empirical work, but more often the best a theoretical treatment can offer is an interpretation of past work and a guide for future work. For situations where crash outcome is a monotonic function of an evasive action, our treatment indicated that converting a set of conflict events to an estimate of crash tendency requires (1) counterfactual testing to correctly classify the conflict events, and (2) understanding how evasive actions are selected over a range of conflict severities, so that correct crash-to-conflict ratios can be determined. These requirements, in turn, suggest directions for future work. First, one important focus would be on developing defensible models of how drivers select evasive actions as functions of background conditions. Second, we indicated previous work where counterfactual testing was accomplished through trajectory-based reconstruction of rear-ending events, based on fairly simple algebraic trajectory models. This initial work should be extended to support more complicated models, possibly based on differential equations. Third, it may happen that the structural features of a particular problem allow Steps 1 and 2 to be simplified. We illustrated one such simplification where detailed knowledge of evasive action mechanisms was replaced by an empirically derived probability density, and highlighted the assumptions needed to justify this simplification. We suspect that the majority, if not all, of proposed traffic conflict techniques can be regarded as simplifications of Steps 1 and 2, but where the simplifying assumptions are to greater or lesser degrees unclear. If this is the case, then the credibility and usefulness of economical traffic conflict techniques could be enhanced substantially by unpacking these assumptions and determining the conditions under which they are valid.

References

Al-Hussaini, E., and M. El-Adll. 2004. Asymptotic Distribution of Normalized Maximum Under Finite Mixture Models. *Statistics and Probability Letters,* Vol. 70, pp. 109–117.

Asmussen, E. (ed.). 1984. *International Study of Traffic Conflict Techniques.* Springer-Verlag, Berlin.

Baguley, C. 1984. The British Traffic Conflict Technique. In *International Study of Traffic Conflict Techniques* (E. Asmussen, ed.), Springer-Verlag, Berlin, pp. 59–73.

Brill, E. 1972. A Car-Following Model Relating Reaction Times and Temporal Headways to Accident Frequency. *Transportation Science,* Vol. 6, No. 4, pp. 343–353.

Campbell, K., H. Joksch, and P. Green. 1996. *A Bridging Analysis for Estimating the Benefits of Active Safety Technologies.* Report UMTRI-96-18. University of Michigan Transportation Research Institute, Ann Arbor, Mich.

Cooper, P. 1984. Experience with Traffic Conflicts in Canada with Emphasis on Post Encroachment Time Techniques. In *International Study of Traffic Conflict Techniques* (E. Asmussen, ed.), Springer-Verlag, Berlin, pp. 75–96.

Davis, G. 2002. Towards a Unified Approach to Causal Analysis in Traffic Safety Using Structural Causal Models. In *Transportation and Traffic Theory in the 21st Century* (M. A. P. Taylor, ed.), Elsevier Science, Ltd., Oxford, United Kingdom, pp. 247–266.

Davis, G., and T. Swenson. 2006. Collective Responsibility for Freeway Rear-Ending Accidents? An Application of Probabilistic Causal Models. *Accident Analysis and Prevention,* Vol. 38, No. 4, pp. 728–736.

Dingus, T., S. Hetrick, and M. Mollenhauer. 1999. Empirical Methods in Support of Crash Avoidance Model Building and Benefits Estimation. *ITS Journal,* Vol. 5, pp. 93–125.

Dingus, T., S. Klauer, V. Neale, A. Petersen, S. Lee, J. Sudweeks, M. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. Doerzaph, J. Jermeland, and R. Knipling. 2006. *The 100-Car Naturalistic Driving Study: Phase II—Results of the 100-Car Field Experiment.* Report DOT HS 810 593. National Highway Traffic Safety Administration, U.S. Department of Transportation.

Fambro, D., K. Fitzpatrick, and R. Koppa. 1997. *NCHRP Report 400: Determination of Stopping Sight Distances.* Transportation Research Board, National Research Council, Washington, D.C., 1997.

Flanagan, J. 1954. The Critical Incident Technique. *Psychological Bulletin,* Vol. 51, pp. 327–358.

Guttinger, V. 1984. Conflict Observation in Theory and Practice. In *International Study of Traffic Conflict Techniques* (E. Asmussen, ed.), Springer-Verlag, Berlin, pp. 17–24.

Hauer, E. 1982. Traffic Conflicts and Exposure. *Accident Analysis and Prevention,* Vol. 14, No. 5, pp. 359–364.

Hauer, E. 1997. *Observational Before-After Studies in Road Safety.* Pergamon, Oxford, United Kingdom.

Hauer, E., and P. Garder. 1986. Research into the Validity of the Traffic Conflicts Technique. *Accident Analysis and Prevention,* Vol. 18, No. 6, pp. 471–481.

Hayward, J. 1972. Near-Miss Determination Through Use of a Scale of Danger. In *Highway Research Record 384,* HRB, National Research Council, Washington, D.C., pp. 24–34.

Heinrich, H. 1959. *Industrial Accident Prevention: A Scientific Approach.* McGraw-Hill, New York.

Hu, W., X. Xiao, D. Xie, T. Tan, and S. Maybank. 2004. Traffic Accident Prediction Using 3-D Model-Based Vehicle Tracking. *IEEE Transactions on Vehicular Technology,* Vol. 53, No. 3, pp. 677–694.

Hyden, C., and L. Linderholm. 1984. The Swedish Traffic-Conflicts Technique. In *International Study of Traffic Conflict Techniques* (E. Asmussen, ed.), Springer-Verlag, Berlin, pp. 133–139.

Lewis, D. 1973. *Counterfactuals.* Harvard University Press, Cambridge, Mass.

Migletz, D., W. Glauz, and K. Bauer. 1985. *Relationships Between Traffic Conflicts and Accidents.* Report FHWA-RD-84-042. Federal Highway Administration, U.S. Department of Transportation.

Najm, W., M. Stearns, H. Howarth, J. Koopman, and J. Hitz. 2006. *Evaluation of an Automotive Rear-End Collision Avoidance System.* Report DOT HS 810 569. National Highway Traffic Safety Administration, U.S. Department of Transportation.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, Cambridge, United Kingdom.

Perkins, S., and J. Harris. 1968. Traffic Conflict Characteristics: Accident Potential at Intersections. *Highway Research Record 225,* HRB, National Research Council, Washington, D.C., pp. 35–43.

Saunier, N., and T. Sayed. 2008. Probabilistic Framework for Automated Analysis of Exposure to Road Collision. *Transportation Research Record: Journal of the Transportation Research Board, No. 2083,* Transportation Research Board of the National Academies, Washington, D.C., 96–104.

Smith, D., W. Najm, and R. Glassco. 2002. Feasibility of Driver Judgment as Basis for a Crash Avoidance Database. *Transportation Research Record: Journal of the Transportation Research Board, No. 1784,* Transportation Research Board of the National Academies, Washington, D.C., pp. 9–16.

Songchitruksa, P., and A. Tarko. 2006. The Extreme Value Theory Approach to Safety Estimation. *Accident Analysis and Prevention,* Vol. 38, No. 4, pp. 811–822.

Svensson, A., and C. Hyden. 2006. Estimating the Severity of Safety Related Behaviour. *Accident Analysis and Prevention,* Vol. 38, No. 2, pp. 379–385.

Appendix to "Outline of a Causal Theory of Traffic Conflicts and Collisions"

Derivations of Main Results

Let U denote the set of possible values for the background variables u and X denote the set of possible values for the evasive action x. For simplicity, we will assume that U is countable, and X is finite,

$$X = \{x_1, x_2, .., x_n\}, \text{ with } x_1 < x_2 < .. < x_n.$$

Proposition 1. Consider the sets

$U_1 = \{u \in U : y(u,x_j) \geq yc, j = 1, . . , n\}$

$U_2 = \{u \in U: y(u,x_1) < yc, y(u,x_j) \geq yc, j = 2, . . , n\}$

. . .

$U_j = \{u \in U : y(u,x_i) < yc, i = 1, . . , j-1 \ \& \ y(u,x_i) \geq yc, i = j, . . , n\}$

. . .

$U_{n+1} = \{u \in U : y(u,x_j) < yc, j = 1, . . , n\}$

The sets $U_1, U_2, . . , U_{n+1}$ form a partition of U.

Proof: We need to show that $U_1 \cup U_2 \cup \ldots U_{n+1} = U$, and $U_i \cap U_j = \phi$ whenever $i \neq j$. Let $u \in U_j$. Then $u \in U$. Now let $u \in U$. Then $y(u,x_1), y(u,x_2), . . y(u,x_n)$ are all well defined. If $y(u,x_1) \geq yc$, $u \in U_1$, while if $y(u,x_n) < yc$, then $u \in U_{n+1}$. Now, suppose $y(u,x_1) < yc$ and $y(u,x_n) \geq yc$. By the monotonicity property of $y(u,x)$, there exists value k such that $y(u, x) < yc$ for all $x < x_k$ and $y(u,x) \geq yc$ for all $x \geq x_k$. Hence, $u \in U_k$, and we have shown that $U_1 \cup U_2 \cup \ldots U_{n+1} = U$. Next, suppose $i \neq j$ and there exists u belonging to both U_i and U_j. If $i < j$, $u \in U_i$ implies $y(u,x_i) \geq yc$ but $u \in U_j$ implies $y(u,x_i) < yc$, a contradiction. The case $i > j$ is handled similarly, giving us $U_i \cap U_j = \phi$ when $i \neq j$.

Proposition 2. $P(y < yc) = \sum_{j=2}^{n+1} P(x \leq x_{j-1} | u \in U_j) P(u \in U_j)$

Proof:

Let a crash indicator function be defined as

$$\hat{y}(u_i, x_j) = \begin{cases} 1, \text{if } y(u_i, x_j) < yc \\ 0, \text{if } y(u_i, x_j) \geq yc \end{cases}$$

Then,

$$P(y < yc) = \sum_i \sum_{j=1}^{n} \hat{y}(u_i, x_j) P(x = x_j | u_i) P(u_i)$$

$$= \sum_{u \in U_1} \left(\sum_{j=1}^{n} \hat{y}(u, x_j) P(x = x_j | u) P(u) \right)$$

$$+ \sum_{u \in U_2} \left(\sum_{j=1}^{n} \hat{y}(u, x_j) P(x = x_j | u) P(u) \right) + \cdots$$

$$+ \sum_{u \in U_{n+1}} \left(\sum_{j=1}^{n} \hat{y}(u, x_j) P(x = x_j | u) P(u) \right)$$

But

$$\sum_{u \in U_1} \left(\sum_{j=1}^{n} \hat{y}(u, x_j) P(x = x_j | u) P(u) \right) = \sum_{u \in U_1} \left(\sum_{j=1}^{n} (0) P(x = x_j | u) P(u) \right)$$

$$= 0$$

and for $k = 2, . . , n + 1$

$$\sum_{u \in U_k} \left(\sum_{j=1}^{n} \hat{y}(u, x_j) P(x = x_j | u) P(u) \right)$$

$$= \sum_{u \in U_k} \left((1) P(x = x_1 | u) P(u) + \cdots + (1) P(x = x_{k-1} | u) P(u) \right.$$

$$\left. + (0) P(x = x_k | u) P(u) + \cdots + (0) P(x = x_n | u) P(u) \right)$$

$$= \sum_{u \in U_k} P(x \leq x_{k-1} | u) P(u) = P(x \leq x_{k-1} | u \in U_k) P(u \in U_k)$$

Substituting appropriately then gives us the desired result.

Proposition 3. $P(y \geq yc \ \& \ x = x_j \ \& \ y_{x=x_{j-1}} < yc) = P(x = x_j | u \in U_j) P(u \in U_j)$

Proof:

$$P(y \geq yc \ \& \ x = x_j \ \& \ y_{x=x_{j-1}} < yc)$$

$$= \sum_i (1 - \hat{y}(u_i, x_j)) \hat{y}(u_i, x_{j-1}) P(x = x_j | u_i) P(u_i)$$

$$= \sum_i \hat{y}(u_i, x_{j-1}) P(x = x_j | u_i) P(u_i) - \sum_i \hat{y}(u_i, x_j) \hat{y}(u_i, x_{j-1})$$

$$P(x = x_j | u_i) P(u_i)$$

Now,

$$\sum_i \hat{y}(u_i, x_{j-1}) P(x = x_j | u_i) P(u_i)$$

$$= \sum_{u \in U_1} \hat{y}(u_i, x_{j-1}) P(x = x_j | u_i) P(u_i) + \cdots + \sum_{u \in U_{n+1}} \hat{y}(u_i, x_{j-1})$$

$$P(x = x_j | u_i) P(u_i)$$

Referring to the partition of U defined above, for any $i = 1,2, . . , n$

$$\hat{y}(u, x_i) = \begin{cases} 1, \text{if } u \in U_k, k \geq i+1 \\ 0, \text{if } u \in U_k, k \leq i \end{cases}$$

This then implies

$$\sum_{u\in U_k} \hat{y}(u,x_i)\,P(x=x_j|u)\,P(u)$$
$$= \begin{cases} P(x=x_j|u\in U_k)\,P(u\in U_k), \text{ if } k\geq i+1 \\ 0, \text{ if } k\leq i \end{cases}$$

so that

$$\sum_i \hat{y}(u_i,x_{j-1})P(x=x_j|u_i)P(u_i) = P(x=x_j|u\in U_j)\,P(u\in U_j)+\cdots + P(x=x_j|u\in U_{n+1})P(u\in U_{n+1})$$

Similarly,

$$\sum_i \hat{y}(u_i,x_j)\,\hat{y}(u_i,x_{j-1})P(x=x_j|u_i)P(u_i)$$
$$= \sum_{u\in U_1} \hat{y}(u_i,x_{j-1})\,\hat{y}(u,x_j)P(x=x_j|u_i)P(u_i)+\cdots + \sum_{u\in U_{n+1}} \hat{y}(u_i,x_{j-1})\,\hat{y}(u,x_j)P(x=x_j|u_i)P(u_i)$$

and since

$$\hat{y}(u,x_{j-1})\,\hat{y}(u,x_j) = \begin{cases} 1, \text{ if } u\in U_k, k\geq j+1 \\ 0, \text{ if } u\in U_k, k\leq j \end{cases}$$

we get

$$\sum_i \hat{y}(u_i,x_{j-1})\hat{y}(u_i,x_j)\,P(x=x_j|u_i)P(u_i)$$
$$= P(x=x_j|u\in U_{j+1})P(u\in U_{j+1})+\cdots + P(x=x_j|u\in U_{n+1})P(u\in U_{n+1})$$

Putting these together we have

$$P(y\geq yc \,\&\, x=x_j \,\&\, y_{x=x_{j-1}} < yc)$$
$$= \big(P(x=x_j|u\in U_j)P(u\in U_j)+\cdots+P(x=x_j|u\in U_{n+1})\,P(u\in U_{n+1})\big) - \big(P(x=x_j|u\in U_{j+1})P(u\in U_{j+1})+\cdots + P(x=x_j|u\in U_{n+1})P(u\in U_{n+1})\big)$$
$$= P(x=x_j|u\in U_j)P(u\in U_j)$$

Related SHRP 2 Research

Development of Analysis Methods Using Recent Data (S01B, S01C, S01E)

Design of the In-Vehicle Driving Behavior and Crash Risk Study (S05)

Technical Coordination and Independent Quality Assurance for Field Study (SO6)

Analysis of In-Vehicle Field Study Data and Countermeasure Implications (S08)