

**Final report of ITS Center project: Accident management using wireless networks**

A Research Project Report

**For the Center for ITS Implementation Research**

**A U.S. DOT University Transportation Center**

**ACCIDENT MANAGEMENT USING WIRELESS NETWORKS:  
Estimating Incident Related Congestion on Freeways Based on Incident  
Severity**

Principal Investigator  
Dr. William Scherer

Center for Transportation Studies  
University of Virginia  
351 McCormick Road,  
P.O. Box 400742  
Charlottesville, VA 22904-4742

July 2007

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.



Research Report No. UVACTS-15-0-113  
July 2007

**ACCIDENT MANAGEMENT  
USING WIRELESS NETWORKS:**

**Estimating Incident Related Congestion on Freeways  
Based on Incident Severity**

By:  
Avi Kripalani  
William Scherer

**A Research Project Report  
For the Center for ITS Implementation Research (ITS)  
A U.S. DOT University Transportation Center**

Avi Kripalani and William Scherer  
Department of Systems and Information Engineering  
University of Virginia

**Center for Transportation Studies** at the University of Virginia produces outstanding transportation professionals, innovative research results and provides important public service. The Center for Transportation Studies is committed to academic excellence, multi-disciplinary research and to developing state-of-the-art facilities. Through a partnership with the Virginia Department of Transportation's (VDOT) Research Council (VTRC), CTS faculty hold joint appointments, VTRC research scientists teach specialized courses, and graduate student work is supported through a Graduate Research Assistantship Program. CTS receives substantial financial support from two federal University Transportation Center Grants: the Mid-Atlantic Universities Transportation Center (MAUTC), and through the National ITS Implementation Research Center (ITS Center). Other related research activities of the faculty include funding through FHWA, NSF, US Department of Transportation, VDOT, other governmental agencies and private companies.

**Disclaimer:** The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

CTS Website  
<http://cts.virginia.edu>

Center for Transportation Studies  
University of Virginia  
351 McCormick Road, P.O. Box 400742  
Charlottesville, VA 22904-4742  
434.924.6362

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle ACCIDENT MANAGEMENT USING WIRELESS NETWORKS: ESTIMATING INCIDENT RELATED CONGESTION ON FREEWAYS BASED ON INCIDENT SEVERITY		5. Report Date July 2007
		6. Performing Organization Code
7. Author(s) Avi Kripalani, William Scherer		8. Performing Organization Report No.
9. Performing Organization and Address Center for Transportation Studies University of Virginia PO Box 400742 Charlottesville, VA 22904-7472		10. Work Unit No. (TRAIIS)
		11. Contract or Grant No.
12. Sponsoring Agencies' Name and Address Office of University Programs, Research and Special Programs Administration US Department of Transportation 400 Seventh Street, SW Washington DC 20590-0001		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code
15. Supplementary Notes		
16. Abstract <p>The effects of traffic incidents on metropolitan freeways extend beyond causing congestion and delays. Immediate impacts include decreased productivity, increased pollution and reduced safety on highways. State and local governments spend billions of dollars annually on construction projects and Intelligent Transportation Systems (ITS) in an effort to curb the adverse consequences of incidents and incident related delays.</p> <p>Effective identification and response on highways is one key to reducing the costs associated with traffic incidents. Within the context of a prototype incident identification and response system developed by the University of Virginia's Systems Technology Integration Laboratory (STIL), this research aims to develop a statistical approach to modeling congestion associated with freeway incidents. The ability to predict congestion will provide more information and greater situational awareness to emergency responders and traffic managers, and will allow travelers to make more informed route selection decisions.</p> <p>By combining data from multiple sources, it is possible to match incident severity estimates for freeway incidents with associated traffic flow counts at the time of the accident. Four metrics for freeway congestion were derived from the traffic flow data, and these metrics were then modeled as functions of the incident severity estimates. The results of multiple linear regression analysis showed that quantifiable relationships exist between the congestion metrics and incident severity data such as the number of vehicles involved in an incident as well as the number of serious injuries reported at the scene.</p>		
17 Key Words Incidents, congestion, data integration, VII		18. Distribution Statement No restrictions. This document is available to the public.

## Table of Contents

<b>Abstract</b>	<b>2</b>
<b>Introduction</b>	<b>3</b>
Problem Statement	3
Intelligent Highway System	6
<b>Literature Review</b>	<b>9</b>
Freeway Incident Management	9
Event Data Recorders	12
Injury Prediction	14
Freeway Concepts	15
Traffic Prediction	17
Injury Severity	20
<b>Methodology</b>	<b>21</b>
Data	21
State Data System	21
Virginia Transportation Research Council (VTRC)	23
Data Integration	23
Traffic Flow Data	24
Metrics	25
Modeling	27
Assumptions	28
Study Limitations	28
Analysis of Complete Set	29
Reduced Data Set	41
Classification Trees	41
<b>Results</b>	<b>43</b>
Accident Volume Model	43
Maximum Difference Model	46
Vehicle Hours Lost Model	48
Percent Vehicle Hours Lost Model	50
<b>Conclusions</b>	<b>53</b>
Contributions	53
Future Work	55
<b>References</b>	<b>57</b>

## **Abstract**

The effects of traffic incidents on metropolitan freeways extend beyond causing congestion and delays. Immediate impacts include decreased productivity, increased pollution and reduced safety on highways. State and local governments spend billions of dollars annually on construction projects and Intelligent Transportation Systems (ITS) in an effort to curb the adverse consequences of incidents and incident related delays.

Effective identification and response on highways is one key to reducing the costs associated with traffic incidents. Within the context of a prototype incident identification and response system developed by the University of Virginia's Systems Technology Integration Laboratory (STIL), this research aims to develop a statistical approach to modeling congestion associated with freeway incidents. The ability to predict congestion will provide more information and greater situational awareness to emergency responders and traffic managers, and will allow travelers to make more informed route selection decisions.

By combining data from multiple sources, it is possible to match incident severity estimates for freeway incidents with associated traffic flow counts at the time of the accident. Four metrics for freeway congestion were derived from the traffic flow data, and these metrics were then modeled as functions of the incident severity estimates. The results of multiple linear regression analysis showed that quantifiable relationships exist between the congestion metrics and incident severity data such as the number of vehicles involved in an incident as well as the number of serious injuries reported at the scene.

# Introduction

## Problem Statement

Situational awareness in the critical early stages of a traffic incident is the key to alleviating a great deal of the congestion that plagues America's major metropolitan regions (VDOT). Being able to detect, respond to and clear an incident faster and more efficiently can reduce congestion related delays from traffic accidents by up to 45% according to the Virginia Department of Transportation (VDOT). Within the scope of a comprehensive incident detection and response system, this research aims to model freeway traffic flow metrics by using measures for incident severity.

Developed by the University of Virginia's Systems Technology Integration Laboratory (STIL), the Intelligent Highway System (IHS) aims to automate and streamline the process of incident detection and response. By leveraging the increased sensing and communications capabilities of motor vehicles, the IHS is able to automatically detect car accidents and instantly broadcast pertinent data to appropriate authorities in order to begin the process of response.

One of the goals of the IHS is to be able to predict the traffic implications of an incident. An incident is any event that temporarily reduces roadway capacity, such as accidents, debris, disabled vehicles, and hazardous material spills (HCM). Because this work is part of a larger context of an emergency response system, this study will focus on accidents only.

It is intuitively clear that the congestion due to an incident that occupies two of three freeway lanes will cause more congestion than an incident that occupies only one (Smith and Qin 362). What is not clear is the extent to which congestion due to an

incident is also a function of the severity of the incident. Severity can be defined either as the damage done to the vehicles themselves, or as the degree of injury sustained by the passengers involved in the accident. This research aims to model the relationships between incident severity and the congestion associated with that incident. Measures of incident severity include vehicle damage, the number of vehicles involved, the number of injuries and the severity of the injuries, among others.

A widely accepted metric for measuring the effect of incidents on traffic flow is capacity reduction (Smith and Qin). Capacity is defined by the Transportation Research Board as the maximum hourly rate at which persons or vehicles can reasonably be expected to transverse a point or uniform section of a lane or roadway during a given time period under prevailing roadway, traffic, and control conditions (TRB). In order to truly measure the capacity reduction on a roadway, the roadway of interest must be at its capacity before an incident occurs. For freeways, this capacity is estimated to be 2200 vehicles per hour per lane (HCM). When the capacity of a roadway is reduced below the demand on that roadway, perhaps in the event of an accident, the volume counts decrease for the duration of incidents. This is a reflection of the decreased speed at which drivers are able to move along the roadway. The data set used for this research was limited in size and contained only a small percentage of incidents where flow on the roadway was at or near capacity before an incident, therefore alternative metrics for quantifying the effects of an incident were developed.

The obvious benefits of increased situational awareness are those of better route guidance and more efficient and accurate re-routing of traffic via Variable Message Signs on highways (Hobeika). The information could also be disseminated over the internet to

allow travelers to make more informed decisions regarding their route selection based on predicted traffic delays.

As described in the Literature Review, it is apparent that much study has been done regarding the prediction of traffic conditions following incidents. Significant research has also been conducted to predict accident severity based on pertinent data. In the past, however, these two issues have been treated as separate problems and not addressed together in one body of work. There has been little or no research to bridge the gap between two areas of study that intuitively seem to be connected. This research aims to close that gap and determine whether including incident severity estimates is valuable in the modeling of freeway congestion.

## **Intelligent Highway System (IHS)**

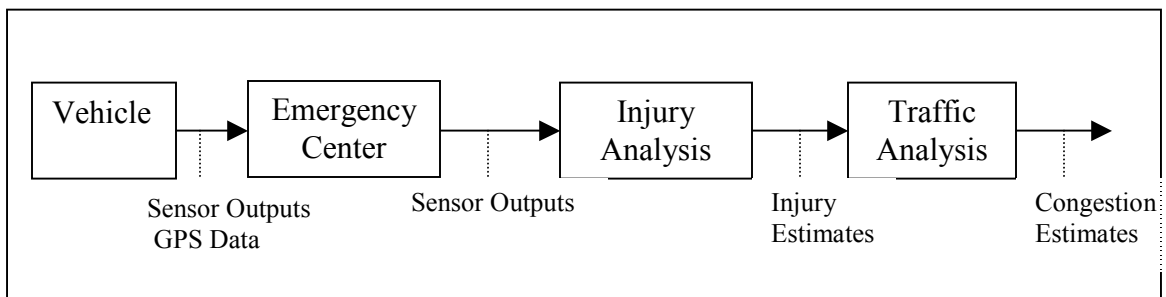
The broader context for the proposed prediction method is the Intelligent Highway System (IHS). Developed as a prototype system, the IHS aims to automate and streamline the process of incident identification and response. The system leverages the increased sensing and communications capabilities of new vehicles. With the addition of acceleration sensors, a Global Positioning System (GPS) receiver and a modem connected to the cellular network, a vehicle can become a source of useful data at the point of incident.

The data that can be sent from the vehicle is limited only by the sensing capabilities of that vehicle. In the near future, automobiles will be equipped with a wide range of sensors including seat belt sensors, video cameras and biometric sensors to determine the condition of the passengers with ever increasing accuracy.

When an incident occurs, and the accelerometers experience a g-force above a pre-determined threshold, the vehicle's computer recognizes an accident event. This process is similar to airbag control processes available in automobiles today (German). The computer then broadcasts relevant data including GPS position and accelerometer outputs to an Emergency Center using Hypercast, a Java based communications protocol. The Emergency Center can be thought of as an information broker that broadcasts appropriate messages to appropriate players in the event of an accident, also using Hypercast. The Hypercast protocol's flexibility in terms of enabling message transmission between wired and wireless users is well documented (Hunter et. al), and was the logical protocol choice for the application development.

One of the players that receive data from the Emergency Center is the Crash Simulation Lab. The Crash Lab runs a finite element simulation, with the accelerometer data as input, to estimate the severity of the injuries sustained by the passengers in the crash. The Crash lab then broadcasts the severity estimates back to the Emergency Center and to area hospitals as well as the Traffic Analysis Center. The Traffic Analysis Center is where congestion metrics would be estimated. Currently, the predictive models that exist do not use severity measures to estimate metrics for congestion

The IHS is similar to automatic accident notification systems currently available in vehicles manufactured by General Motors, among others, as a service available through the manufacturer. Incident response centers that are a part of the service are automatically notified in the event of air bag deployment in a subscribing vehicle, at which point emergency responders can be notified by service staff. The IHS, however, is a fully automated, distributed system, that can be modified and upgraded rapidly and easily without any disruption to the system. Individual localities can implement an IHS for a given jurisdiction and stand alone from similar national or even state wide systems. The diagram below shows the flow of information in the event of an incident.



**Figure 1 Information Flow over the HIS**

## **Literature Review**

### **Freeway Incident Management**

When attempting to improve the large scale system that includes the freeways of a metropolitan area and its supporting emergency response agencies, it is important to address certain issues relating to incident identification, response and management in the current system. The proposed work aims to model metrics for congestion in terms of incident severity, but the interaction between the type of response and the capacity reduction/duration of an incident must be noted. It is also necessary to examine the emergency response perspective and traffic management perspective and note their similarities and differences in the event of an accident on a major freeway. Effective freeway management can also lead to fewer incidents. According to the Federal Highway Administration's Traffic Management Handbook, 13 percent of all peak period crashes were a direct result of a previous incident (FHWA).

There is an interaction between the scale of emergency response to a situation and the congestion as well as the duration of an incident. It can be argued that with all other factors being equal, an incident with a larger emergency response will cause greater congestion and result in longer delays than an incident with a smaller response (Saunders). Additionally, an incident with inadequate response will have a longer duration, or time to clear, than an incident with appropriate response because of the lag in arrival of the right equipment and personnel (Nathanail).. Therefore, it is in the best interest of motorists that an appropriate response to an incident is deployed. The decisions regarding the type of response to an emergency can be made based on better information with the implementation of a system such as the IHS that would provide

information from the scene such as injury severity and data that could allow for immediate traffic analysis.

There are two perspectives of an incident that have potentially competing objectives. The Emergency response perspective has the goal of minimizing the time it takes to respond to an incident with the appropriate equipment and personnel, and the traffic management perspective has the goal of restoring the normal traffic flow on the affected roadway (Zografos, 536). These somewhat competing objectives have an effect on the scale of the response and the time it takes to clear an incident, and therefore play a role in determining the congestion associated with an incident.

The motivation for improvements in response to freeway incidents is clear when we realize that approximately 650,000 Americans suffer serious injuries in vehicle crashes every year, yet these injuries occur in less than one percent of all motor vehicle crashes (Champion). Thus, it is important to be able to identify the crashes that have the potential for causing life threatening injuries. By accurately identifying the small percentage of accidents that may cause serious injuries, better resource allocation can be achieved, increasing the likelihood of timely response. It has been established that the first 60 minutes following an incident, or the “golden hour”, determines whether a patient will live or die (Champion). Therefore, quantifiable benefits can be achieved by reducing response time.

There are three key stages that make up the response time of a vehicular accident (Evanco):

- Accident notification time, or the time between the crash and emergency medical service (EMS) notification. This includes detection time

- The time between EMS notification and EMS arrival at the scene of the crash
- The time between EMS arrival at the scene and arrival at the hospital

In a statistical model based study, Evanco et. al . found that the average notification time was 5.2 minutes and an average of 43.8 deaths per year per state in the year 1990. The study examined the effects of variables such as Vehicle Miles Traveled (VMT), Mean Vehicle Speed (MVS) and alcohol consumption. The conclusions of the work emphasize the significant impact of reducing the notification time after an accident, noting that a reduction in the average notification time 5.2 minutes to 3 minute would result in an 11% reduction in fatalities, and a reduction to 2 minutes would result in a 15% reduction in fatalities (Evanco). An automatic notification system such as The Intelligent Highway System (IHS) could conceivably reduce the notification time after an accident to a few seconds, resulting in an immediate reduction in lives lost.

## **Event Data Recorders**

A central motivation for determining the impact of injury severity on freeway capacity is the increasing availability of Event Data Recorders (EDRs) in many late model cars and light trucks (German). EDRs are the motor vehicle equivalent to the ‘black box’ found in airplanes. While the importance of black boxes in the event of an aviation incident is well documented, the fact that their capabilities extend to automobiles is not as widely known.

As an adjunct to air bag sensing and control systems, EDRs offer a new source of useful data that can further the understanding of on-road traffic safety issues. Combined with the widespread use of mobile communications capabilities, EDRs can provide immediate data for analysis in the event of an accident. Considerable roadblocks are present, however, that are slowing the path to widespread use of EDRs. These roadblocks include a lack of standardization of the data recorded, storage formats and means of retrieval. Privacy concerns must also be addressed before universal usage of EDRs becomes a reality as well as legal and liability issues.

Widely used in systems developed for vehicles in North America and Europe, EDRs are made up of vehicle mounted accelerometers used to monitor the crash pulse of an accident. An onboard microprocessor analyzes the vehicle’s acceleration and determines air-bag deployment based on pre-programmed decision logic. The EDR makes possible the capture of relevant data for post-crash analysis. Recorded data includes vehicle speed, engine RPM, throttle position, brake-switch status and seatbelt use. All of these parameters add valuable insights into the potential for injuries in the event of an accident. The data can be sent for analysis immediately and the information

returned by the analysis can assist in emergency response decisions and traffic planning strategies.

Post accident, the data recorded by EDRs can be used by safety researchers, law enforcement and insurance companies. If widespread use is adopted, insurance companies will no doubt want access to the data to assign fault and support legal actions. Such stakeholder interests raise questions of ownership, accessibility and the use of data in individual cases for purposes other than safety analysis.

Currently, General Motors offers a proprietary GM Crash Data Retrieval (CDR) system. Available in GM vehicles, the device is used to obtain data from vehicles involved in collisions. Use of EDRs is more widespread in research institutions where data from staged collisions is collected by EDRs as part of crash testing programs. The EDRs used currently record change in velocity ( $\Delta V$ ) slightly – but consistently – higher than the actual  $\Delta V$  because of the rebounding velocity recorded by the device. The EDRs work very accurately when recording the crash pulse, which is not affected by the rebound velocity (German).

The increased availability of EDRs in commercial vehicles makes it necessary to be able to analyze the data recorded by the devices. By modeling congestion as a function of incident severity, it would be possible to predict the traffic implications of an accident that involves a vehicle equipped with an EDR.

## Injury Prediction

The ability of Event Data Recorders (EDRs) to record deceleration in vehicles in the event of a collision has been utilized to investigate the correlation between pre-determined acceleration threshold limits for injury and the potential for passenger injury in a crash (Gabauer). Models for determining injury severity, or the potential for injury, however, are still imperfect at best and inadequate in the case of side-swipe and other non-frontal collisions. One method for determining the potential for injury is mapping peak acceleration values in the three dimensions to the Acceleration Severity Index (ASI) (Gabauer). The ASI combines the three dimensional acceleration components to provide a single numerical value that indicates the severity of an incident. The ASI equation is described below (Gabauer):

$$ASI(t) = \left[ \left( \frac{\bar{a}_x}{\hat{a}_x} \right)^2 + \left( \frac{\bar{a}_y}{\hat{a}_y} \right)^2 + \left( \frac{\bar{a}_z}{\hat{a}_z} \right)^2 \right]^{\frac{1}{2}}$$

where  $\bar{a}_x$ ,  $\bar{a}_y$  and  $\bar{a}_z$  are the 50-ms average component accelerations, and  $\hat{a}_x$ ,  $\hat{a}_y$  and  $\hat{a}_z$  are threshold accelerations for each component. In previous crash studies it has been determined that ASI is generally indicative of occupant injury in frontal collisions (Gabauer). There has also been a high correlation between analysis of EDR results in calculating delta-V and other, more traditional models for calculating the accepted metric for injury severity.

While the methods for analyzing data from EDRs have not yet been perfected, the devices are capable of providing immediate data that has otherwise never been available in the event of an accident. This fact, combined with the increasing availability of EDRs

in consumer automobiles makes it important to be able to make use of the newly available data.

### **Freeway Concepts**

Freeways are segments of road that provide uninterrupted flow (TRB). They have no intersections, and access to and from the freeway is limited to ramps. This study focuses on freeways only, as they are the least complicated roadway facility to model when attempting to determine congestion. There are no interactions with intersections, traffic signals, and other roadway conditions that exist on other types of facilities. Freeway capacity is described as “the maximum sustained 15-min flow rate, expressed in passenger cars per hour per lane, that can be accommodated by a uniform freeway segment under prevailing traffic and roadway conditions in one direction of flow” (TRB 13-1).

For any given segment of roadway there is a capacity that can be determined. In the event of an accident or excess demand, the capacity of a freeway segment is reduced temporarily due to the blockage of lanes. According to the *Highway Capacity Manual*, an incident does not have to block a travel lane to create a bottleneck, or temporarily reduce capacity.

At any given time, a freeway segment is experiencing one of three conditions, or flow types. The segment is either “unsaturated”, in a state of “queue discharge” or “oversaturated flow”. An unsaturated segment is unaffected by upstream or downstream conditions, with speed ranges of 55 mph to 75 mph. The capacity of the highway is determined under this condition because it allows for the maximum speed and flow rate. The queue discharge regime is characterized by traffic that has just passed through a

bottleneck and is accelerating back to its free-flow speed. This acceleration generally takes 0.5 to 1 mile downstream from the bottleneck.

Finally, the oversaturated flow regime is one that is influenced by the effects of a downstream bottleneck – a traffic accident, for example. The traffic flow in the congested regime can vary based on flows and speeds depending on the characteristics of the bottleneck itself, and queues can grow to several thousand feet upstream of the bottleneck. These queues differ from queues at intersections in that vehicles move slowly through the queue with periods of stopping. The reduced capacity of a freeway segment is calculated under the oversaturated flow conditions. By comparing the capacity, or vehicles per hour per lane under prevailing conditions, to the capacity during an incident, it is possible to quantify the effect of a traffic incident on the freeway system.

Ideally, we would conduct such a comparison for this study as a primary metric. The limited data set that was available for use, caused by the difficulty in accurately matching incident severity factors with location data, prevented the use of capacity reduction as a metric. Instead, five other metrics associated with traffic flow ( $v/h/l$ ) and their relationships with several independent variables were examined.

## **Traffic Prediction**

There is a large body of work devoted to traffic prediction using various inputs and predicting various traffic conditions, including volume and speed. Smith and Demetsky show that nonparametric regression can be used to forecast freeway volumes. Using historical data for a given site, a k-nearest neighbor (KNN) approach is used to predict traffic volumes over 15 minute intervals. The method proved to be accurate and portable, with some calibration required to fit the model to specific sites.

Nonparametric regression has also proven useful in estimating accident duration. Qi and Smith used a KNN model to forecast the duration of an incident based on time of day, day of week, incident type, blocked lanes, weather and the number of vehicles involved. One issue addressed by this work was the use of categorical variables in nonparametric regression. The data set used for the study included many categorical variables, including incident type and weather conditions. This issue had to be addressed before a distance measure, or objective function, was developed. As a result, distance matrices were developed for each categorical variable to allow for a quantitative measure of how 'far apart' each categorical value was from the others, with the values of the elements in the distance matrix indicating the impact of the independent variables on the duration.

Alternative methods for forecasting traffic conditions include time-series modeling, neural nets, decision trees and linear regression. The work of Yang et. al. uses auto-regressive models to predict traffic speed. Using historical freeway flow data, future speed is forecasted by assigning coefficients of regression to previous time intervals in an effort to minimize the prediction error. This method may be difficult to adapt to the

proposed work because time-series modeling does not account for external factors including incidents and their associated severity.

Smith and Qin found that the capacity reduction due to accidents was greater than the physical blockage on roadways. As the number of lanes blocked increases, so too does the capacity reduction. The study also stated a case for the argument that capacity reduction can be modeled as a random variable.

## **Injury Severity**

The work by Qi and Smith is an effort to predict accident duration based on available data. The proposed work aims to use a similar modeling approach to forecast congestion due to incidents based on incident severity. By using sensor data to first gain insight into the severity of the crash itself, emergency responders can be better prepared in the event of an accident. Here, we examine some of the analytical techniques for predicting crash severity based on roadway conditions such as weather, speed, guardrail configurations and roadway geometry.

Kim et. al. used logistic regression to predict the severity of injury based on crash type and location. The study found that the odds of a fatal incident increase greatly for those involved in head on and roll over crashes. Being rear ended had the highest likelihood of injury, but the lowest odds of being killed. Seat belt use was also analyzed as well as interaction effects between the examined variables. The limitation of a logistic model is the fact that the response variable is binary. In this case, fatal/nonfatal crash was the response. While this is a good starting point for analysis, it would be ideal to have more levels of injury represented in the response variable.

Many statistical approaches exist to model crash severity as a function of roadway, roadside, operational, environmental and other descriptive variables. Regression methods can be used to test the contribution of these factors to crash severity. Donnell and Mason used pavement surface conditions, use of drugs or alcohol, presence of an interchange entrance ramp, crash type and traffic volume to predict the severity of median-related crashes. The method chosen was a logistic regression model. Similarly, there exist many crash-type specific models to estimate crash severity, with logistic

regression being the modeling method of choice for most studies. Studies have been conducted to predict the severity of truck-related incidents, teen-driver related incidents and other specific categories of crashes. As described by the discussion of the Intelligent Highway System, it may be possible in the near future to determine, based on data sent by vehicles involved in accidents, the type of crash that has occurred. It would then be feasible for a specific model to be applied to the situation, and greater accuracy in severity estimation could be achieved. For the purposes of this work, however, incident severity data, including injury levels, gathered from police reports will be used to model congestion, as a data set with accelerometer data readings from accidents is not readily available.

It is clear that research has been conducted to estimate traffic conditions based on explanatory variables such as volume in previous time intervals, time of day and incident occurrence. Much investigation has also been conducted to predict the severity of injuries in a car accident based on state variables such as seatbelt usage, and roadway conditions. A missing link in current research exists, however, in predicting the traffic implications of an incident based on measures of incident severity.

## Methodology

### Data

The data necessary to conduct this study was consolidated from multiple sources, each filling a specific need of the study. Incident data, including date, time and location along with injury severity and other measures of incident severity were not available in a single data set for reasons concerning the privacy of individuals involved in the incidents. As a result, data from the National Highway Traffic Safety Administration (NHTSA) was merged with data available through the Virginia Transportation Research Council (VTRC).

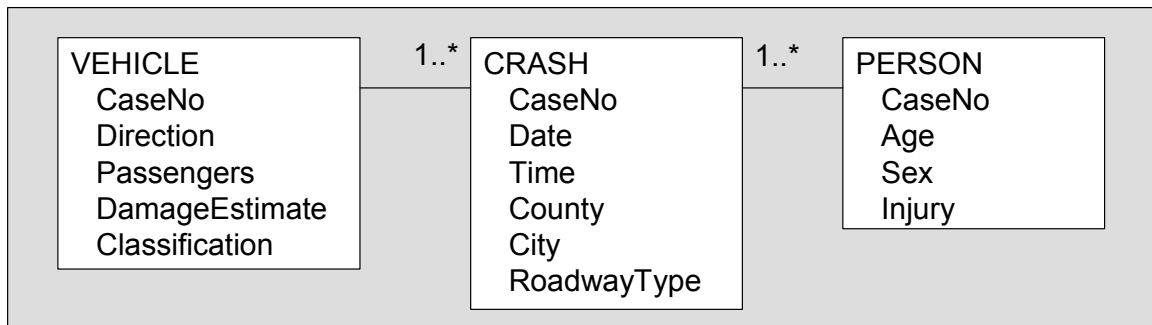
#### *State Data System*

The State Data System, a data collection program sponsored by NHTSA, records every roadway accident in Virginia, among other states. The data, collected from Police Accident Reports (PARs), includes separate files for the accident itself, the vehicles involved, and the people involved. For this study, the SDS data set for Virginia in the year 2003 was made available by NHTSA. The *crash* data file consists of information such as date, time, county and roadway type.

The *vehicle* data file describes the damage sustained by each vehicle in a given accident as well as characteristics of the vehicle such as direction of travel, number of passengers and whether the vehicle is classified as private or commercial. Finally, the *person* file records the age, injury severity and sex of each person involved in the incident. Injury severity is classified on a scale from zero to six, as shown below.

Injury Severity (INJ)		Format: C
Indicates the severity of the injury suffered in the crash.		
Element Values:	Meaning:	
0	Not injured	
1	Dead before report made	
2	Visible signs of injury, as bleeding wound or distorted member, or had to be carried from scene	
3	Other visible injury, as bruises, abrasions, swelling, limping, etc.	
4	No visible injury, but complaint of pain or momentary unconsciousness	
5	Died later	
6	Not stated	

Figure 2 shows a simplified relational database view of the SDS data. Incidents are identified across the three files through a common case number (CASENO).



**Figure 2: SDS Relational View**

While the SDS contains a wealth of valuable information regarding traffic incidents, exact time and location data is not available. Specifically, time data is limited to the hour of the incident and location data is limited to the city in which the incident occurred. For the purposes of this study, however, exact time and location data is necessary in order to conduct an investigation into traffic volumes and how they are affected by an incident. To bridge this gap, the SDS data were merged with another data set, the incident data archive provided by the VTRC.

### *Virginia Transportation Research Council (VTRC)*

The VTRC collects data pertaining to incidents on interstates in the Northern Virginia and Hampton Roads areas. Because of the higher quality of data available in Hampton Roads, this area was chosen to investigate for the study. The incident data archive records the date, location and time (to the minute) of accident events, along with other descriptive data pertaining to the incident. The location data available includes city, roadway type and roadway name, direction of travel and a location code, which maps to a one to three mile section of roadway.

### *Data Integration*

In order to make a positive match between incidents across the two data sets, the following criteria were used: date, hour, city and roadway type. The set was narrowed to include only the cities of Norfolk, Chesapeake and Virginia Beach because of the limited availability of traffic flow data. This also narrowed that number of incidents to a more manageable size for analysis purposes. Additionally, this study focused on incidents occurring on weekdays as a starting point for the analysis.

The lack of precise time data in the SDS set further narrowed the number of incidents available for investigation. To reduce the possibility of incorrectly matching incidents across data sets, SDS incidents that occurred within one hour of each other were eliminated from the set. This helped ensure accurate matching of data points even if incident data was rounded to the nearest hour or incorrectly entered.

Once completed, the data integration process resulted in 76 usable data points. Incidents for which a positive match was made across the two accident data sets and for

which traffic flow data was available were considered usable. Observing the 76 data points reveals that they are a representative sample of the larger population, and have similar characteristics in terms of number and type of injuries.

### *Traffic Flow Data*

The final piece of data necessary for the study was the traffic flow data available through the Virginia Transportation Research Council (VTRC). For each incident, volume counts were collected from the appropriate inductive loop detectors based on the location of the incident. Loop detectors were selected by matching the location of the incident with the location of loop detectors recorded in the VTRC database. The volume counts were collected beginning thirty minutes before the time of the incident and ending two hours following cleanup. These volume counts were collected on one minute intervals, and for analysis purposes 10-minute averages were calculated to account for the high variability in individual one minute counts (Smith and Qin). In addition to the volume counts before, during and after the incident, volume counts were collected for the same time period one and two weeks prior to the incident and averaged to determine the historical volume.

Future work should consider other data including speeds and occupancy counts to gain better understanding about traffic conditions post-incident. Once the data were collected and the set reduced to incidents for which good data was available, six metrics for analysis were calculated for each incident. The following section details these metrics.

## Metrics

To complete the data set for analysis of the effects of incidents on traffic flow, five metrics – one independent variable and four dependent variables - were calculated for each of the 76 incidents. Figure 3 is a sample chart for one incident included in the study.

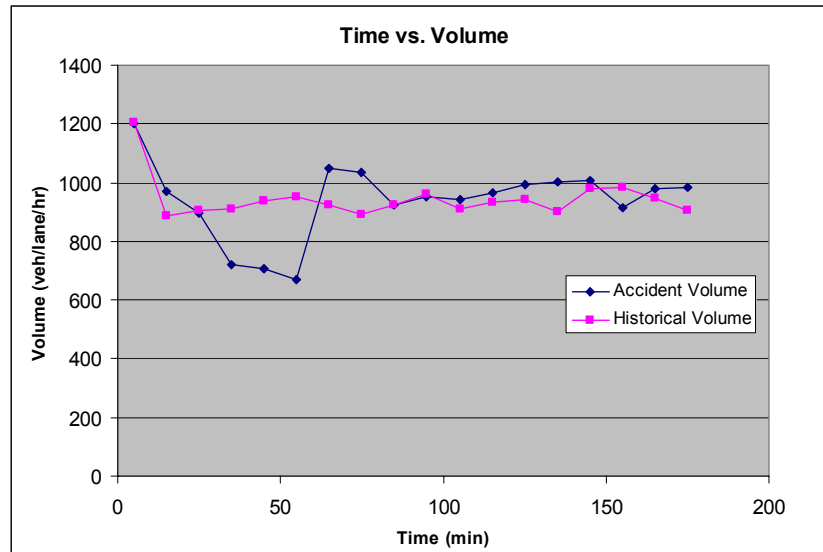


Figure 3

This figure shows the VTRC query output for a typical incident in chart form. We see that the historical volume remains relatively steady throughout the 180 minute period of data collection. At about the thirty minute mark, the accident volume curve dips well below the historical volume curve, illustrating what is assumed to be the impact of the incident, which occurs at  $t=30$ . An interesting characteristic of this and many of the other Time vs. Volume curves is that once the incident is cleared at  $t=70$ , the volume exceeds the historical volume counts for some time, and later returns to a level close to that which was observed in the prior weeks. This is the queue discharge phase that occurs after the incident is cleared, and the characteristics of this phase may lend further

insight into the incident response in terms of how many vehicles were re-routed. For the purposes of this work, however, the following values were recorded for each incident:

*Accident Volume* – Represented by length  $B$  in Figure 3, the minimum volume during the time of the accident is an indicator of the number of vehicles that were able to pass through the loop detector in one minute during accident conditions.

*Historical Volume* – Represented by the sum of lengths  $A$  and  $B$ , this is an indicator of the historical demand on the freeway segment of interest at the time of the incident under normal conditions.

*Maximum Difference* – Represented by length  $A$ , this is an indicator of unsatisfied demand that is assumed to be a result of the incident.

*Vehicle Hours Lost* – The light shaded region in the diagram, the area of which was calculated by subtracting the area under the accident volume curve from the area under the historical volume curve. The vehicle hours that we assume would have been supplied under normal conditions were lost as a result of the accident. This may be interpreted as a measure of aggregated delay.

*Percent Vehicle Hours Lost* – The light shaded region was also calculated as a percent of the total vehicle hours under normal conditions. This dependent variable was investigated to normalize for the fact that some of the incidents occurred on roadways with low historical volumes.

The historical volume was used as an independent variable, along with the data collected from the SDS and the VTRC, to model the remaining five metrics described above. The independent variables and their correlation coefficients are described below

Variable Name	Description	Variable Type
hist_vphpl	Historical volume (v/h/l)	Discrete
numinj	Number of injuries reported	Discrete
numveh	Number of vehicles involved	Discrete
numpeople	Number of people involved	Discrete
vehdam2	Number of vehicles with No Damage	Discrete
vehdam3	Number of overturned vehicles	Discrete
vehdam4	Number of vehicles with motor damage	Discrete
vehdam5	Number of vehicles with undercarriage damage	Discrete
vehdam6	Number of totaled vehicles	Discrete
vehdam7	Number of vehicles damaged by fire	Discrete
vehdam8	Number of vehicles with other damage	Discrete
inj0	Number of people reporting no injuries	Discrete
inj1	Number dead before report made	Discrete
inj2	Number of people bleeding or that had to be carried from scene	Discrete
inj3	Number of people with other visible injuries and not carried from the scene	Discrete
inj4	No visible injury, but complained of pain or momentary unconsciousness	Discrete
inj5	Number of people that died later (within 30 days)	Discrete

	hist_vphpl	numinj	numveh	numpeople	vehdam2	vehdam3	vehdam6	vehdam8	inj0	inj2	inj3	inj4
hist_vphpl	-	0.174	0.275	0.287	0.246	-0.205	0.056	0.211	0.125	0.073	0.113	0.124
numinj	-	-	0.124	0.473	-0.028	0.148	-0.066	0.081	-0.487	0.711	0.501	0.468
numveh	-	-	-	0.831	0.186	-0.283	0.449	0.732	0.7736	-0.143	0.189	0.229
numpeople	-	-	-	-	0.107	-0.205	0.328	0.609	0.516	0.107	0.414	0.441
vehdam2	-	-	-	-	-	0.054	0.154	-0.159	0.147	-0.11	0.082	0.007
vehdam3	-	-	-	-	-	-	-0.049	-0.457	-0.343	0.341	-0.041	-0.112
vehdam6	-	-	-	-	-	-	-	-0.134	0.416	-0.101	0.229	-0.135
vehdam8	-	-	-	-	-	-	-	-	0.533	-0.175	0.062	0.287
inj0	-	-	-	-	-	-	-	-	-	-0.552	-0.033	-0.119
inj2	-	-	-	-	-	-	-	-	-	-	0.235	-0.109
inj3	-	-	-	-	-	-	-	-	-	-	-	-0.112
inj4	-	-	-	-	-	-	-	-	-	-	-	-

The use of these independent variables in exploring the relationship between incident severity and the related congestion is described in the next section.

## **Modeling**

The objective of the data analysis portion of this study was to determine the relationships between the independent variables and the previously developed metrics for congestion. By developing multiple linear regression models, it is possible to gain insight into the relationships that exist among the available data. Linear models were developed because pair-wise observation of independent/dependent variable pairs did not reveal any potential logarithmic or higher order relationships within the data. Several models are developed for varying subsets of the data and these models are then tested for their predictive performance on test data points not used in model development. Analysis was conducted for the complete set as well as a subset of 30 points with steady and high historical volumes on the day/time of the accident.

### *Assumptions*

Before conducting analysis of the data and associated models, a key assumption must be made regarding the behavior of traffic on the freeway sections studied for this work. Because the data comes from multiple cities and multiple loop detectors, we must assume that the historical volume collected represents the demand for that detector on the given day of the week at the given time of day. We can justify this assumption for the context of this study because the goal is to examine relationships rather than derive a rule or theorem regarding traffic behavior after an accident as a function of incident severity.

### *Study Limitations*

Certain types of freeway accidents were not included in this study. Namely, there were no fatalities or crashes involving commercial vehicles and vehicles with hazardous materials. Because these cases make up a relatively small percentage (less than 0.5%) of

all accidents, and likely have vastly different characteristics from other types of accidents, it would not be realistic to include them in a study that examines only 76 data points. These special cases likely deserve individual study rather than being included in a small scale analysis such as this. Additionally, this study did not account explicitly for the time of day when an incident occurred. The lack of data prevented such stratification, but high and low volumes (i.e. rush hour) are accounted for by including historical volume as an independent variable

#### *Analysis of the Complete Set*

The first set of models were developed using a training set that was randomly selected from the complete set of 76 points. The training set contained 51 points, and the testing – or validation – set contained the remaining 25 points. For each of the five proposed metrics, the following steps were followed:

- a. Created a multiple linear regression model using all available independent variables
- b. Conducted a backwards step-wise refinement process to optimize the adjusted  $R^2$  value of the model and create a ‘reduced’ model.
- c. Observed the significant variables
- d. Validated model assumptions
- e. Removed data points that were outliers according to the Cook’s Distance metric for regression models (if necessary)
- f. Repeated steps (a) and (b) to create a new reduced model.

#### *Maximum Difference Metric*

We first attempt to model the behavior of the maximum difference metric. Recall that this metric measures the maximum difference between the volume (v/l/h) observed during the accident and the historical volume observed by the same loop detector at the same time of day and day of week as the accident. After conducting backwards stepwise refinement, Table A.1 shows the best model that could be created for this metric:

lm(formula = maxdiff_vphpl ~ hist_vphpl + numinj, data = injuries.training)				
Residuals:				
Min	1Q	Median	3Q	Max
-267.31	-123.59	-21.25	70.09	577.94
Coefficients:				
	<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>Pr(&gt; t )</b>
<b>(Intercept)</b>	-143.4662	65.4341	-2.193	0.033222 *
<b>hist_vphpl</b>	0.2934	0.0498	5.891	3.68e-07 ***
<b>numinj</b>	119.8875	31.4845	3.808	0.000398 ***
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 193.1 on 48 degrees of freedom				
Multiple R-Squared: 0.536, Adjusted R-squared: 0.5167				
F-statistic: 27.73 on 2 and 48 DF, p-value: 9.907e-09				

Table A.1

The adjusted  $R^2$  value for this model is 0.51, showing some correlation between the independent variables and the congestion metric. We see that the historical volume variable is very significant, and shows that that the maximum difference between historical volume and accident volume is about 29.34% of the historical volume. The other variable – also significant – is the number of injuries in an incident. The presence of this variable makes intuitive sense because it has a positive sign and indicates a positive relationship between volume reduction and the number of injuries in an incident. Next, we verify that the model assumptions are met. Figure 4 shows model diagnostic charts:

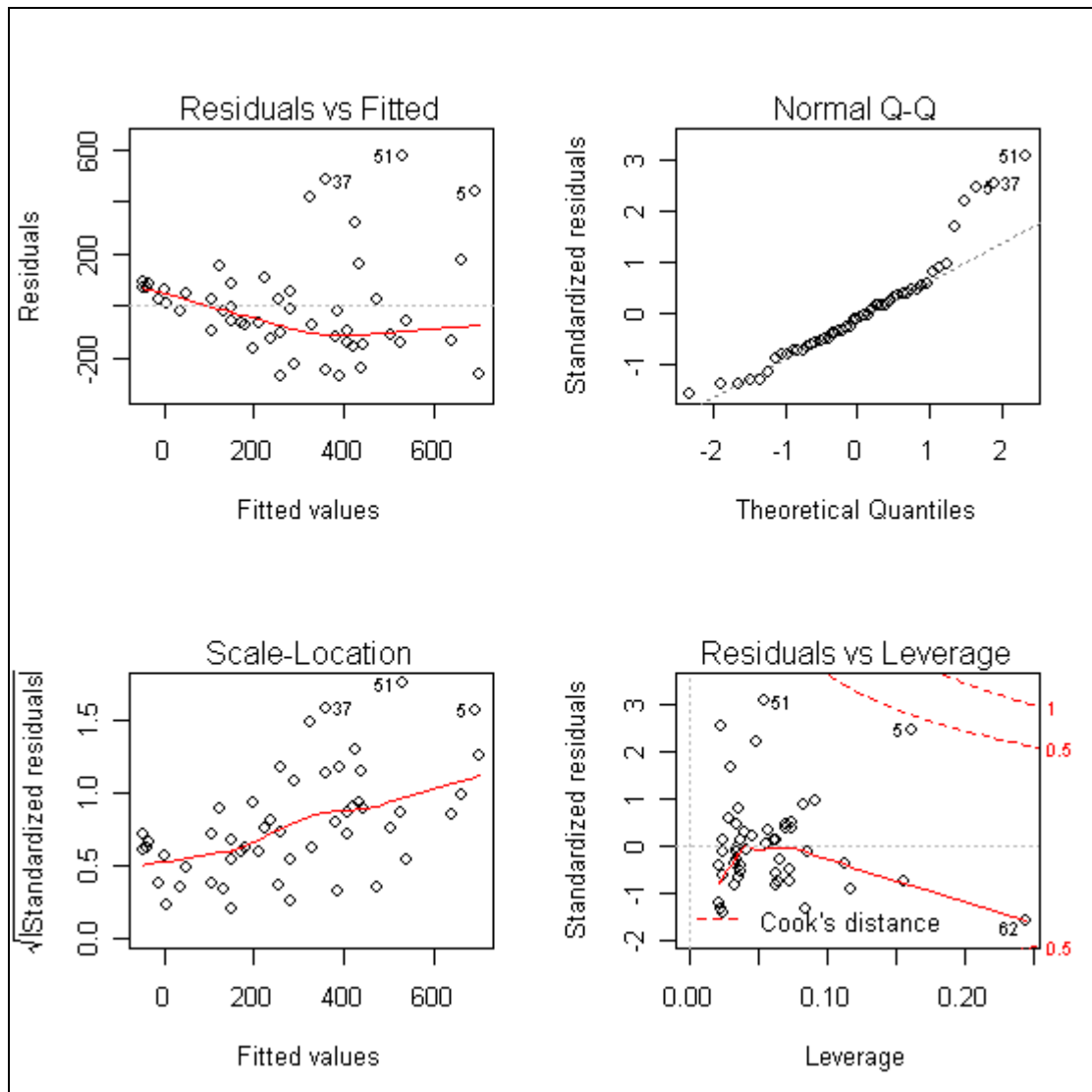


Figure 4

We see no strong trend in the Fitted vs. Residual plot and a Normal Q-Q plot that shows normally distributed residuals. The Cook's distance plot, a measure of each point's influence on the overall model, shows three points that have a high influence (data points 5, 51 and 62). The widely accepted rule of thumb is that if a point has a Cook's Distance value of 1.0 or higher, we can remove the point and create the regression model without that point. While none of the points have Cook's Distance values as high as 1.0, it may be of interest to observe the effect of reporting the results of removing

these points. Table A.2 shows the best model that was developed after removing the three points described above:

```
lm(formula = maxdiff_vphpl ~ hist_vphpl + numinj, data = injuries.training.reduced)
Residuals:
  Min    1Q  Median    3Q   Max
-250.00 -93.58 -22.57  72.43 509.63
Coefficients:
              Estimate      Std. Error    t value    Pr(>|t|)
(Intercept)  -136.81798     43.95692    -3.113    0.0027 **
hist_vphpl    0.26475       0.03735     7.088    9.22e-10 ***
numinj       124.67311     20.31491     6.137    4.68e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 154.4 on 69 degrees of freedom
Multiple R-Squared:  0.6115,    Adjusted R-squared:  0.6002
F-statistic:  54.3 on 2 and 69 DF,  p-value: 6.82e-15
```

**Table A.2**

We see that the adjusted  $R^2$  increases slightly with little change in the coefficient estimates for the variables. The important points to note from the analysis of the maximum difference metric are the relatively high adjusted  $R^2$  along with the positive relationship with the number of injuries and the positive, fractional relationship with the historical volume. All these are intuitive results that support the hypothesis that freeway congestion is a function of incident severity, assuming that historical volume is an indication of demand during the accident.

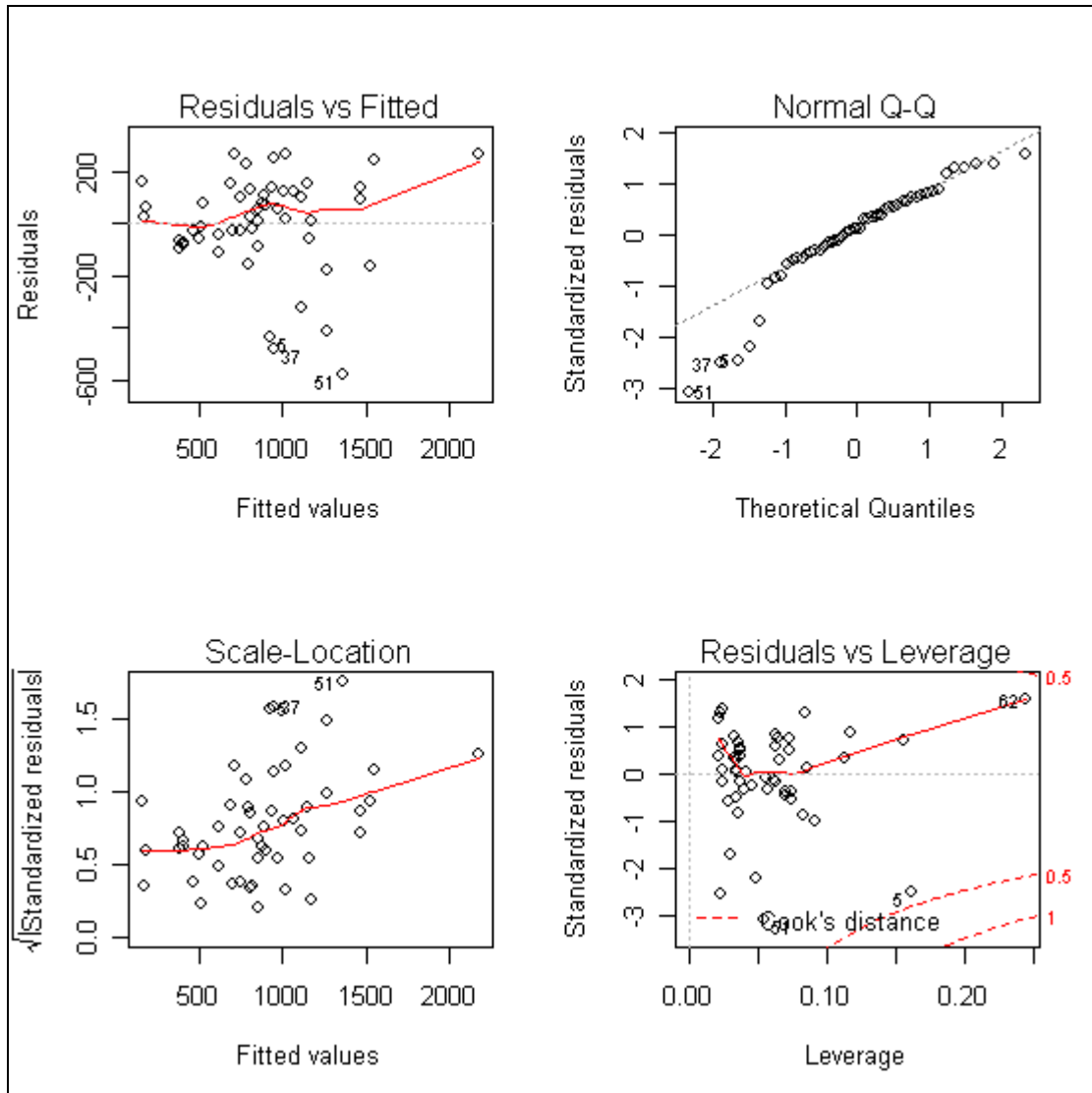
*Accident Volume Metric*

The second metric we examine is the minimum volume along the freeway during the time needed to clear the accident (incident duration is recorded in the VTRC database along with the time of incident). Table A.3 shows the initial results of multiple regression modeling with an adjusted  $R^2$  of 0.803:

```
lm(formula = acc_vphpl ~ hist_vphpl + numinj, data = injuries.training)
Residuals:
  Min    1Q  Median    3Q   Max
-577.94 -70.09  21.25 123.59 267.31
Coefficients:
              Estimate      Std. Error    t value    Pr(>|t|)
(Intercept)   143.4662     65.4341     2.193     0.033222 *
hist_vphpl     0.7067      0.0498    14.190    < 2e-16 ***
numinj        -119.8875     31.4845    -3.808    0.000398 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 193.1 on 48 degrees of freedom
Multiple R-Squared:  0.8109,    Adjusted R-squared:  0.803
F-statistic: 102.9 on 2 and 48 DF,  p-value: < 2.2e-16
```

**Table A.3**

It is interesting to note that the coefficient estimates are in line with the estimates for the maximum difference model. The historical volume has a positive, fractional relationship with a coefficient of 0.71. The coefficient for the maximum difference model is 0.29. The coefficients sum to one, resulting in the intuitively obvious relationship that the sum of the minimum volume and the maximum difference is equal to the historical volume. This relationship exists because the maximum difference is a metric derived from the historical volume and accident volume measurements, and was included in the analysis for the purposes of investigating the characteristics of traffic behavior from as many different perspectives as possible. Figure 5 shows the model diagnostic charts:



**Figure 5**

We see no pattern in the Fitted vs. Residuals charts and the residuals are normally distributed, with some deviation at the tails. The Cook's Distance chart, however, shows once again that data points 5, 51 and 62 have high Cook's Distance values. Again, we report modeling results after removing these points from the training set in Table A.4. Once again, the adjusted  $R^2$  sees a slight increase along with slight adjustments in the coefficient estimates

```

Call:
lm(formula = acc_vphpl ~ hist_vphpl + numinj, data = injuries.training.reduced)
Residuals:
  Min    1Q  Median    3Q   Max
-509.63 -72.43  22.57  93.58 250.00
Coefficients:
              Estimate      Std. Error    t value    Pr(>|t|)
(Intercept)  136.81798    43.95692     3.113    0.0027 **
hist_vphpl    0.73525     0.03735    19.685   < 2e-16 ***
numinj       -124.67311    20.31491    -6.137   4.68e-08 ***
Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 154.4 on 69 degrees of freedom
Multiple R-Squared: 0.8507,    Adjusted R-squared: 0.8464
F-statistic: 196.6 on 2 and 69 DF,  p-value: < 2.2e-16

```

**Table A.4**

Here we have a very high adjusted  $R^2$  value and coefficients that show a negative relationship between the number of injuries and the volume at the time of the accident, along with a positive, fractional relationship between the accident volume and the historical volume. While we cannot assume causation, the relationships that exist are of interest for the purposes of this study.

#### *Vehicle Hours Lost Metric*

Next, we examine the metric of Vehicle Hours Lost. While the method of calculating Vehicle Hours Lost employed in this study is not an accepted method of measuring delays, it was included to fully investigate the characteristics of the Time vs. Volume chart. Table A.5 shows the best attainable model for the Vehicle Hours Lost metric.

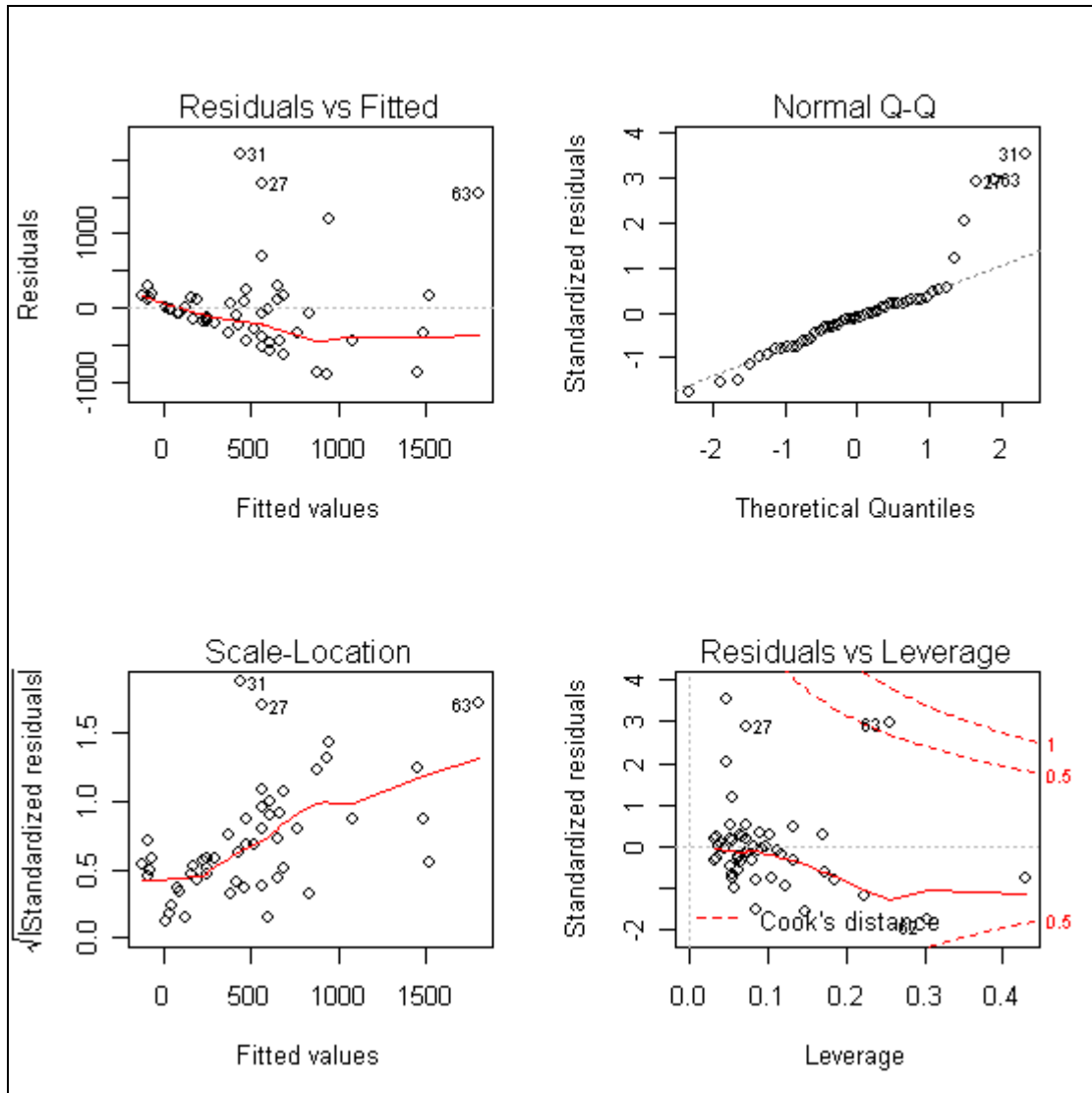
```

Call:
lm(formula = arealost ~ hist_vphpl + numveh + inj0 + inj2, data = injuries.training)
Residuals:
  Min    1Q  Median    3Q   Max
-882.06 -340.30 -63.96  136.06 2071.48
Coefficients:
              Estimate      Std. Error    t value    Pr(>|t|)
(Intercept)  -336.5537      297.8530    -1.130     0.26436
hist_vphpl    0.4236         0.1566      .705     0.00956 **
numveh       293.2180      160.5999     1.826     0.07438 .
inj0        -241.5870      155.0468    -1.558     0.12605
inj2         290.9864      158.6546     1.834     0.07311 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 604.4 on 46 degrees of freedom
Multiple R-Squared: 0.3621,    Adjusted R-squared: 0.3066
F-statistic: 6.527 on 4 and 46 DF,  p-value: 0.0003018

```

**Table A.5**

We see that the Vehicle Hours Lost can be modeled as a function of the historical volume, number of vehicles, number of uninjured and number of people with visible injuries. All of these coefficients are intuitive in their relationship to the dependent variable, but the adjusted  $R^2$  is low compared to the other metrics. Examination of the model diagnostics in Figure 6 reveals the existence of some outliers in the Cook's Distance metric.



**Figure 6**

When the potential outliers are removed (points 63 and 62), we see little change in the adjusted  $R^2$ , but the metric can now be modeled best using only the historical volume and the number of severely injured people as shown in Table A.6. Additionally, the model assumptions are met, as shown in Figure 7

```

Call:
lm(formula = arealost ~ hist_vphpl + inj2, data = injuries.reduced.training)
Residuals:
  Min   1Q   Median   3Q   Max
-714.49 -239.55 -57.61  90.41 1555.69
Coefficients:
              Estimate      Std. Error    t value    Pr(>|t|)
(Intercept)  -194.1899     157.5037    -1.233    0.224007
hist_vphpl    0.4712       0.1296     3.634    0.000713 ***
inj2          287.7378     95.8450     3.002    0.004364 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 442.9 on 45 degrees of freedom
Multiple R-Squared:  0.3483,    Adjusted R-squared:  0.3193
F-statistic: 12.02 on 2 and 45 DF,  p-value: 6.555e-05

```

Table A.6

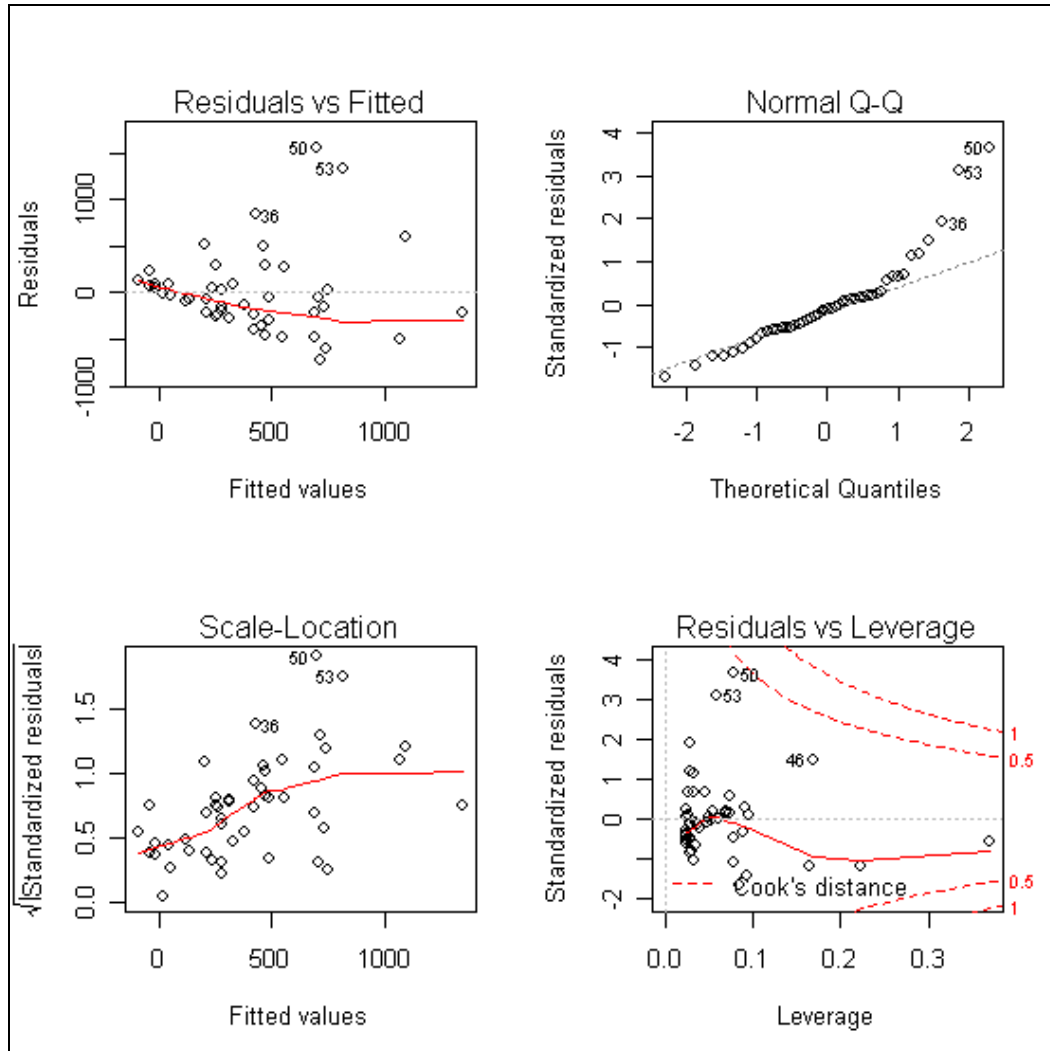


Figure 7

*Percent Vehicle Hours Lost Metric*

In further exploration of the Vehicle Hours Lost metric, we examine the absolute number as a percentage of the total vehicle hours during normal conditions on the given roadway (See previous section for complete explanation of this metric). Table A.7 shows the modeling results for this metric

Call:				
lm(formula = PercentAreaLost ~ hist_vphpl + inj0 + inj3 + vehdam3, data = injuries.training)				
Residuals:				
Min	1Q	Median	3Q	Max
-0.13695	-0.04569	-0.01367	0.03808	0.23207
Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.1290534	0.0329257	3.920	0.000293 ***
hist_vphpl	0.0000343	0.0000203	1.690	0.097885 .
inj0	-0.0291254	0.0117225	2.485	0.016672 *
inj3	0.2401116	0.0798890	3.006	0.004283 **
vehdam3	0.6658071	0.0821506	8.105	2.07e-10 ***
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 0.0772 on 46 degrees of freedom				
Multiple R-Squared: 0.665, Adjusted R-squared: 0.6358				
F-statistic: 22.82 on 4 and 46 DF, p-value: 1.947e-10				

**Table A.7**

We see that historical volume, the number of uninjured people, the number of people with moderate injuries, and the number of overturned vehicles are all significant indicators of the Percent Vehicle Hours Lost metric with intuitive coefficients that suggest a positive relationship between incident severity and the congestion metric. The adjusted R<sup>2</sup> of 0.64 is much higher than that attained when modeling Vehicle Hours Lost, indicating that the demand under normal conditions is an important factor to consider when modeling the behavior of these roadways.

*Summary of Complete Set Analysis*

By modeling the various potential metrics for freeway congestion, we were able to see the relationships between our independent variables of incident severity and the

dependent indicators of freeway congestion. The table below is a brief summary of the modeling results for the training data selected from the complete data set:

<b>Complete Data Set - Modeling Coefficients</b>							
	Adjusted R squared	Historical Volume	Number of Injuries	Number Uninjured	Number of Severe Injuries	Number of Minor Injuries	Overtuned Vehicles
Max Difference	0.6	0.265	124.67	-	-	-	-
Accident Volume	0.85	0.735	-124.67	-	-	-	-
Vehicle Hours Lost	0.32	0.47	-	-	287.74	-	-
Percent Vehicle Hours Lost	0.64	3.43E-05	-	-0.29	-	0.24	0.67

We see that the historical volume is a significant variable in all of the models, and various measures of incident severity are used for the different dependent variables. The number of injuries proves to be a strong indicator of the impact on volume, while injury measurements are prevalent in the models relating to Vehicle Hours Lost.

The complete data set represents a wide range of pre-accident conditions, and further reduction of the data set was conducted in an attempt to strengthen the modeling capabilities of the independent variables. The set was reduced to include incidents that were characterized by relatively high volumes (above 1000 v/h/l) and steady historical volumes (less than 15% deviation throughout the duration of the incident). This reduced data set was made up of 30 incidents, and the analysis of this set follows

### *Reduced Data Set*

The set of 30 points was partitioned once again into training and test sets. This time, the training set contained 25 points with five test set points. While this set of data is very small, the goal was to investigate whether stratifying the incidents based on certain characteristics would make the models more robust. All of the dependent variables were modeled, and it was found that all of them produced models that had lower adjusted  $R^2$  values than when modeling was conducted with the entire set. Additionally, there were several cases of counter-intuitive variable coefficients that contradicted the findings in the previous section. Several different partitions of training/test sets were made, all leading to different confounding outcomes. These results, combined with the fact that the models had, on average, 19 degrees of freedom, made it difficult to make conclusions based on this set. While there may be value in stratifying based on pre-incident conditions and applying appropriate models for those conditions, more data and a more thorough investigation of such a method is necessary, but beyond the scope of this study.

### *Classification Trees*

Another modeling method that was attempted for the complete data set was the use of classification trees. Specifically, the Classification and Regression Trees (CART) algorithm was implemented by assigning an index (A thru F) based on pre-determined partitions to the minimum volume dependent variable. The CART algorithm had a 28% misclassification rate, and only partitioned the classification tree based on historical volume. While we cannot say that classification trees are

not a good method for modeling accident volume, the results of the algorithm on this particular data set were not strong.

In the next section, the performance of selected models on their training sets is reported and analyzed.

## Results

### Accident Volume Model

We begin by examining the performance of the Accident Volume model on the 25 point testing data set. The best model (shown below) was used to predict the Accident Volume metric for the training set points based on the independent variables, and the predicted volumes were compared to the observed volumes.

```
lm(formula = acc_vphpl ~ hist_vphpl + numinj, data = injuries.training)
Residuals:
  Min    1Q  Median    3Q   Max
-577.94 -70.09  21.25 123.59 267.31
Coefficients:
              Estimate      Std. Error    t value    Pr(>|t|)
(Intercept)   143.4662      65.4341     2.193    0.033222 *
hist_vphpl     0.7067       0.0498    14.190   < 2e-16 ***
numinj        -119.8875     31.4845    -3.808   0.000398 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 193.1 on 48 degrees of freedom
Multiple R-Squared:  0.8109,    Adjusted R-squared:  0.803
F-statistic: 102.9 on 2 and 48 DF,  p-value: < 2.2e-16
```

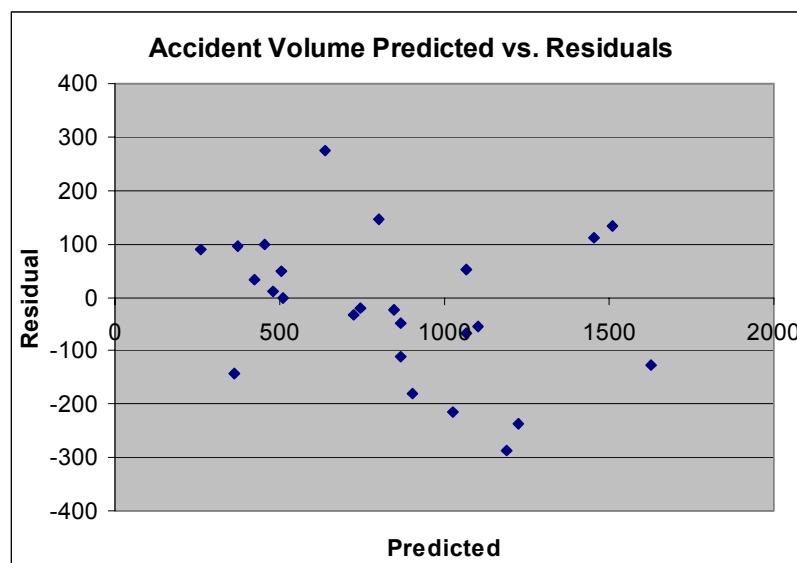
**Table R.1**

Table R.2 shows the historical volume and number of injuries for each of the 25 test set points along with the predicted and actual values for Accident Volume as well as the residuals (prediction error) for each point.

Injuries	Accident Volume			Residual
	Historical	Predicted	Observed	
0	2036.4	1627.6	1754.4	-126.8
0	1784.4	1512.4	1378.8	133.6
0	1798.2	1451.5	1341.6	109.9
0	1493.3	1226.1	1464.0	-237.9
1	1517.4	1186.9	1474.8	-287.9
1	1381.5	1101.5	1155.0	-53.5
0	1249.5	1066.3	1132.5	-66.2
1	1482.0	1066.1	1014.0	52.1
1	1373.0	1026.0	1242.0	-216.0
1	1128.0	900.7	1081.2	-180.5
0	1006.2	866.1	915.6	-49.5
0	1019.4	865.6	975.6	-110.0
1	965.3	844.4	868.5	-24.1
2	1327.5	800.0	654.0	146.0
1	967.5	742.9	762.0	-19.1
0	813.0	723.3	757.5	-34.2
5	1604.4	638.4	362.4	276.0
0	522.0	508.2	510.0	-1.8
0	517.5	504.9	456.0	48.9
0	499.5	481.3	471.0	10.3
1	438.0	454.7	356.4	98.3
0	420.6	423.0	388.8	34.2
0	339.0	372.9	278.0	94.9
1	519.0	362.4	504.0	-141.6
0	199.5	259.6	171.0	88.6

**Table R.2**

Figure 8 shows no trend in the residuals vs. fitted plots.



**Figure 8**

The error values appear to be normally distributed, with a mean of -18.3 and a standard deviation of 133.97. To further quantify the predictive value of this model, the observed Accident Volume metric was given a classification (A thru F) based on the following values:

Volume (v/h/l)	Classification	Number Classified
>1700	A	3
1400-1700	B	4
1100-1400	C	14
800-1100	D	21
500-800	E	15
<500	F	19
<b>Total</b>		<b>76</b>

**Table R.3**

The model's predicted values for Accident Volume were then assigned a classification based on the same ranges as above, and the predicted classification was compared to the observed classification. The model had a 60% classification rate, correctly identifying the classification of 15 out of 25 training set points. Additionally, 100% of the predicted classifications were within one classification of the observed value.

Next, the ranges for each classification were adjusted to allow the borders to overlap. This was done to prevent misclassifications near the borders and "soften" the ranges for each classification. The new borders are described in Table R.3.

Volume (v/h/l)	Classification
>1675	A
1375 – 1725	B
1075 – 1425	C
775 – 1125	D
475 – 825	E
<525	F

**Table R.4**

When the classification ranges were modified, the model accurately predicted 17 out of 25 classifications.

## Maximum Difference Model

The performance of the Maximum Difference model (R.5) on the test set was examined next. Table R.6 shows the number of injuries, the historical volumes as well as the predicted and observed maximum differences between historical volume and volume during the accident.

Lm(formula = maxdiff_vphpl ~ hist_vphpl + numinj, data = injuries.training)				
Residuals:				
Min	1Q	Median	3Q	Max
-267.31	-123.59	-21.25	70.09	577.94
Coefficients:				
	<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>Pr(&gt; t )</b>
<b>(Intercept)</b>	-143.4662	65.4341	-2.193	0.033222 *
<b>hist_vphpl</b>	0.2934	0.0498	5.891	3.68e-07 ***
<b>numinj</b>	119.8875	31.4845	3.808	0.000398 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 193.1 on 48 degrees of freedom				
Multiple R-Squared: 0.536, Adjusted R-squared: 0.5167				
F-statistic: 27.73 on 2 and 48 DF, p-value: 9.907e-09				

Table R.5

Injuries	Historical Volume	Maximum Difference		Residual
		Predicted	Observed	
0	1784.4	259.09	405.6	-146.5
0	499.5	9.43	28.5	-19.1
0	1798.2	390.82	456.6	-65.8
0	1493.3	300.75	29.3	271.5
1	522.0	13.88	12.0	1.9
1	199.5	-79.18	28.5	-107.7
0	1373.0	386.6	131.0	255.6
1	1381.5	284.86	226.5	58.4
1	1006.2	156.89	90.6	66.3
1	438.0	-46.8	81.6	-128.4
0	1482.0	380.3	468.0	-87.7
0	813.0	99.83	55.5	44.3
1	965.3	108.92	96.8	12.2
2	1604.4	930.54	1242.0	-311.5
1	967.5	309.24	205.5	103.7
0	519.0	145.21	15.0	130.2
5	1128.0	184.73	46.8	137.9
0	2036.4	461.16	282.0	179.2
0	1517.4	455.99	42.6	413.4
0	420.6	-13.87	31.8	-45.7
1	1327.5	445.68	673.5	-227.8
0	517.5	12.55	61.5	-49.0
0	1019.4	162.99	43.8	119.2
1	1249.5	234.65	117.0	117.7
0	339.0	-40.17	61.0	-101.2

**Table R.6**

The plot of fitted vs. residuals for this model shows a potential upward trend, as shown in Figure 9. The model performs better for lower predicted values of the dependent variable, with the range of error terms increasing as the predicted values increase. The residuals have a mean of 24.85 and a standard deviation of 162.9. It should be noted that the model predicts negative values for several points, indicating an increase in volume after an accident occurs. This makes the model difficult to interpret in some cases, and indicates that the intercept value for the model may need to be adjusted, or forced to go through the origin.

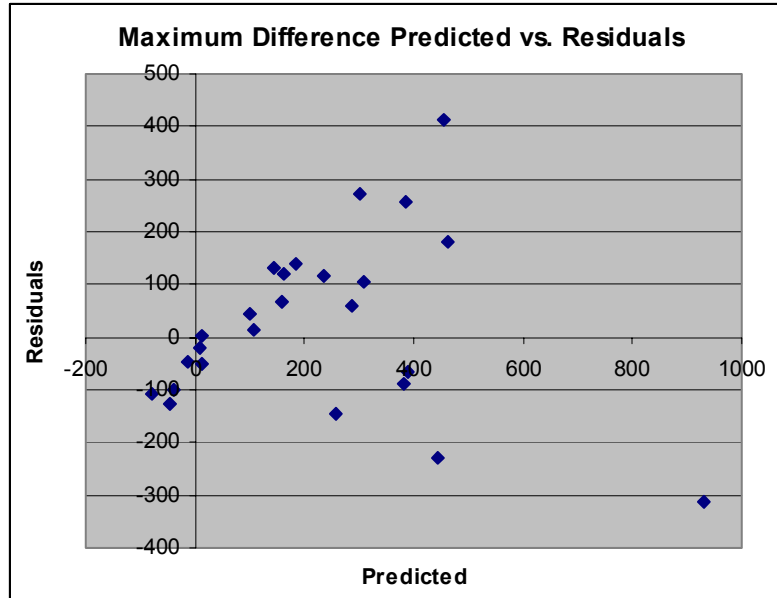


Figure 9

### Vehicle Hours Lost Model

The results of the Vehicle Hours Lost model (Table R.7) were similar to those of the Maximum Difference Model.

```

Call:
lm(formula = arealost ~ hist_vphpl + numveh + inj0 + inj2, data = injuries.training)
Residuals:
  Min   1Q   Median   3Q   Max
-882.06 -340.30 -63.96  136.06 2071.48
Coefficients:
              Estimate      Std. Error    t value      Pr(>|t|)
(Intercept)  -336.5537     297.8530     -1.130     0.26436
hist_vphpl    0.4236       0.1566       2.705     0.00956 **
numveh       293.2180     160.5999     1.826     0.07438 .
inj0        -241.5870     155.0468     -1.558     0.12605
inj2         290.9864     158.6546     1.834     0.07311 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 604.4 on 46 degrees of freedom
Multiple R-Squared:  0.3621,    Adjusted R-squared:  0.3066
F-statistic: 6.527 on 4 and 46 DF,  p-value: 0.0003018

```

Table R.7

The strength of the model lies in predicting the lower values, with some negative predicted values. Again, it may be advantageous to force the linear model through the origin for interpretive purposes. Table R.8 shows the historical volume, number of vehicles, number uninjured and number seriously injured along with the predicted and

observed vehicle hours lost for each of the test set points. Figure 10 shows no trend in the plot of predicted vs. residual values. The residuals had a mean of 24.85 and a standard deviation of 162.89.

Historical Volume	Vehicles	Uninjured	Seriously Injured	Vehicle Hours Lost		Residual
				Predicted	Observed	
199.47	1	1	0	-169.53	34.47	-204
339	2	2	0	-80.07	72	-152.07
438	1	0	1	-27.92	236.5	-264.42
1604.4	2	0	2	1634.57	5708	-4073.43
420.6	1	1	0	-73.65	18.5	-92.15
517.5	2	2	0	-2.68	16.25	-18.93
499.5	1	1	0	-39.44	3.75	-43.19
519	2	2	0	11.73	13.5	-1.77
522	2	2	0	-0.72	5	-5.72
1327.5	3	2	0	413.28	1832.5	-1419.22
813	2	2	0	125.45	15	110.45
967.5	2	1	1	1025.25	151	874.25
965.25	1	0	1	200.69	62.75	137.94
1006.2	2	2	0	209.22	203.75	5.47
1019.4	1	1	0	185.98	16.75	169.23
1482	3	2	0	631.07	81.5	549.57
1128	3	2	0	90.81	19	71.81
1249.5	3	3	0	336.04	45.6	290.44
1381.5	3	2	0	510.26	159.5	350.76
1373	2	1	0	583.49	436.7	146.79
1798.2	2	2	0	552.62	608.75	-56.13
1784.4	4	4	0	185.32	348.75	-163.43
1493.25	2	2	0	420.4	45.5	374.9
1517.4	5	4	0	1034.88	3140.2	-2105.32
2036.4	2	2	0	655.9	724.5	-68.6

Table R.8

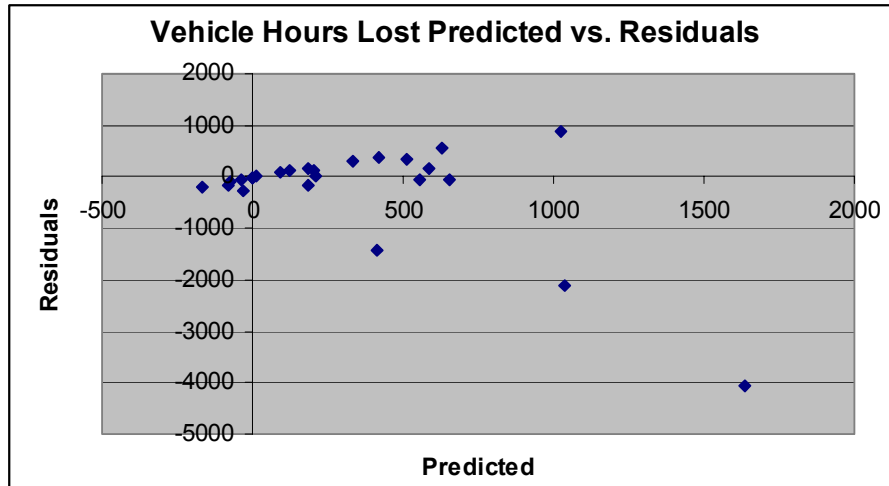


Figure 10

### Percent Vehicle Hours Lost Model

Finally, we examine the model for Percent Vehicle Hours Lost. Table R.9 shows the model used for analysis.

```

Call:
lm(formula = PercentAreaLost ~ hist_vphpl + inj0 + inj3 + vehdam3, data = injuries.training)
Residuals:
  Min    1Q  Median    3Q   Max
-0.13695 -0.04569 -0.01367  0.03808  0.23207
Coefficients:
              Estimate      Std. Error  t value      Pr(>|t|)
(Intercept)  0.1290534    0.0329257    3.920    0.000293 ***
hist_vphpl   0.0000343    0.0000203    1.690    0.097885 .
inj0         -0.0291254    0.0117225   -2.485    0.016672 *
inj3         0.2401116    0.0798890    3.006    0.004283 **
vehdam3      0.6658071    0.0821506    8.105    2.07e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.0772 on 46 degrees of freedom
Multiple R-Squared:  0.665,    Adjusted R-squared:  0.6358
F-statistic: 22.82 on 4 and 46 DF, p-value: 1.947e-10

```

Table R.9

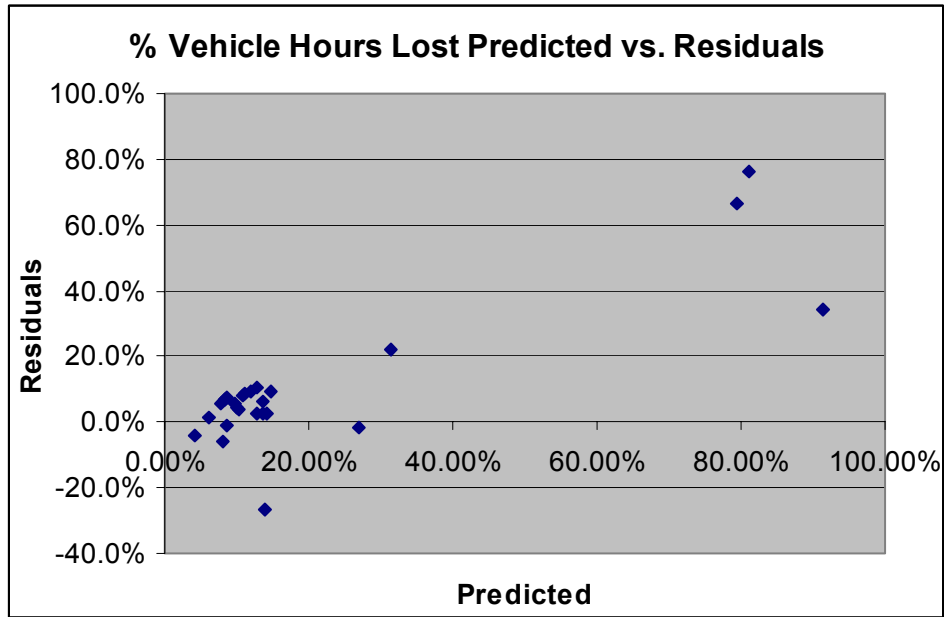
We see that historical volume, minor injuries and overturned vehicles contribute to higher loss of vehicle hours. Additionally, the number of uninjured people reduced the amount of lost vehicle hours, which may be a counter-intuitive finding because it suggests that an accident involving a large number of uninjured people would have a very low loss of vehicle hours. Further examination of the data,

however, shows that the range of this variable is from zero to five, which makes the coefficient more intuitive for the purposes of this model. Table R.10 shows the historical volumes, number uninjured, number with minor injuries, number of overturned vehicles along with predicted and observed values for percent vehicle hours lost and associated residuals for each of the 25 test data points.

Historical Volume	Uninjured	Minor Injuries	Overturned Vehicles	Percent Vehicle Hours Lost		Residual
				Predicted	Observed	
199.47	1	0	0	10.06%	5.77%	4.3%
339	2	0	0	8.09%	13.90%	-5.8%
438	0	0	1	79.42%	12.96%	66.5%
1604.4	0	3	0	91.44%	57.46%	34.0%
420.6	1	0	0	10.79%	2.67%	8.1%
517.5	2	0	0	8.68%	9.63%	-0.9%
499.5	1	0	0	11.05%	2.26%	8.8%
519	2	0	0	8.35%	1.75%	6.6%
522	2	0	0	8.69%	1.32%	7.4%
1327.5	2	0	0	13.91%	40.23%	-26.3%
813	2	0	0	9.65%	4.29%	5.4%
967.5	1	0	0	14.13%	11.42%	2.7%
965.25	0	0	1	81.16%	5.00%	76.2%
1006.2	2	0	0	10.28%	6.28%	4.0%
1019.4	1	0	0	12.75%	1.96%	10.8%
1482	2	0	0	14.75%	5.30%	9.5%
1128	2	0	0	7.85%	2.08%	5.8%
1249.5	3	0	0	6.01%	4.41%	1.6%
1381.5	2	1	0	31.44%	9.06%	22.4%
1373	1	0	0	13.58%	7.18%	6.4%
1798.2	2	0	0	12.88%	10.28%	2.6%
1784.4	4	0	0	4.17%	8.47%	-4.3%
1493.25	2	0	0	11.88%	2.34%	9.5%
1517.4	4	1	0	26.89%	28.27%	-1.4%
2036.4	2	0	0	13.67%	11.01%	2.7%

**Table R.10**

The plot of predicted vs. residuals shows the possibility of the presence of outliers within the data set.



## **Conclusions**

### **Contributions**

The contributions of this study begin with the data itself. Extensive research was conducted to find data that provided measures of incident severity along with any kind of date and location information. The State Data System was the best available option in that it provided data for the incident itself, along with each person and vehicle involved in the incident. The data set was coded in such a way that it was possible to distinguish between freeway incidents and incidents that occurred on other types of roadways. Additionally, city, date and hour data was provided.

Matching the State Data System incidents that occurred on Virginia freeways that were equipped with loop detectors was the main challenge of the data set development phase of this study. The VTRC's incident archives were queried based on the available data from the State Data System to create a set of 214 positive matches across the two data sets. The next step was an attempt to collect volume and speed data for appropriate times and sensors that matched the incidents. Volume count data at the one minute aggregate level was available for only 76 of the 214 data points, while speed data was available for only a fraction of the incidents. As a result, the study concentrated on analysis of the volume counts as part of this preliminary investigation.

Once the data was compiled, the result was a database of 76 incidents with incident severity data for each, including vehicle damage, the number of injuries, and the severity of each injury. Additionally, each incident was associated with a table of volume count data that detailed the traffic flow over the loop detector of interest beginning thirty minutes before the incident and ending two hours after the incident was reported to have

been cleared. To allow for data mining of these incidents, metrics had to be developed to quantify the effect of the incident on the flow of traffic.

The development of two of these metrics is the second contribution of this study, the Vehicle Hours Lost and the Percent Vehicle Hours Lost. While the metric of capacity reduction would have been the ideal dependent variable for this study, the scarcity of reliable matches between the two incident data sets dictated that alternative metrics be chosen. Four congestion metrics were derived from the volume count data available through the VTRC database. While some metrics proved to be better subjects of multiple regression analysis than others, it was instructive to observe the relationships between these dependent variables and the incident severity data.

Finally, the development of the models to examine the relationships between the independent variables and the derived metrics was the third contribution. We saw that it is possible to model the minimum level of traffic flow experienced during an incident using the historical volume for that roadway and the number of injuries reported from the incident with an adjusted  $R^2$  of 0.803. While we cannot assume causation between the independent and dependent variables, the existence of this strong relationship certainly suggests that incident severity may play a role in determining the traffic implications of a freeway incident. We also noted that historical volume is a key piece of information in modeling all of the congestion metrics, which is a well documented phenomenon. What was newly discovered in this study, however, was that various measures of incident severity – number of vehicles, number of injuries, severity of injuries, vehicle damage – can be used to model various measures of congestion.

## **Future Work**

A significant limitation of this work is the site-specific nature of the data collected. Freeway characteristics in Northern Virginia near Tyson's Corner would be very different from those found in the Hampton Roads area. For this reason, narrowing the region of study to a small number of stretches of freeway would help the predictive power of the study, along with studying incidents that occurred only when the freeway was at or near its capacity. Future investigations of the effect of incident severity on freeway flow should also include speed and occupancy data as part of the study. The analysis of volumes alone does not tell the full story of the implications of an incident, and using only volume data forced the inclusion of an important assumption regarding historical data and freeway demand during an incident. By including speed and occupancy data, this assumption can be avoided, and the study can be more robust as a result.

All of these improvements require more data, namely more detailed time data from the State Data System. If the SDS incident archive had time data accurate to the minute rather than the hour, many more matches could have been made across the two data sets, which would have allowed for more complete analysis.

A larger data set would also allow for stratification of model development based on pre-crash conditions as well as the type of crash. While stratification of the data in this study did not prove so, the hypothesis is that stratifying based on pre-crash traffic conditions may yield better predictive models. Modeling based on prior knowledge of crash type may also prove useful.

Another factor that may be useful to know is the level of emergency response for these incidents. It would be of great interest to emergency responders as well as traffic managers to fully investigate the interaction between traffic flow and the level of response from police, fire and rescue departments.

The primary motivation for this study was to have predictive modeling of traffic conditions as part of a larger emergency response system. By integrating such models into a system such as General Motors' OnStar system, the models could not only be immediately useful in decision making, but could be improved upon as more incident data is collected. The ever-increasing sensor capabilities of new vehicles are a promising trend for analytical tools such as these.

## References

- Abdel-Aty, Mohammed, M. Abdalla. "Linking Roadway Geometrics and Real-Time Traffic Characteristics to Model Daytime Freeway Crashes". *Transportation Research Record* V. 106 p. 106.
- Abdel-Aty, Mohamed, N. Uddin, A. Pande, M. Abdalla, L. Hsia. "Predicting Freeway Crashes from Loop Detector Data by Matched Case-Control Logistic Regression". *Transportation Research Record* V. 1897 p. 88
- Al-Ghamdi, Ali S. "Using Logistic Regression to Estimate the Influence of Accident Factors on Injury Severity". *Accident Analysis and Prevention*. V. 34 p. 729
- Baek, Seungki, Hyunho Chang, Jaebeom Park. "Development of Dynamic Freeway Simulation Model for Incident Management." *Proceedings of the Eastern Asia Society for Transportation Studies*. V. 4, October 2003.
- Bremmer, Daniela, Keith C. Cotton, Dan Cotey, Charles E. Prestrud, Gary Westby. "Measuring Congestion: Learning from Operational Data" *Washington State Department of Transportation*
- Cassidy, Michael J., Robert L. Bertini. "Observations at a Freeway Bottleneck." *Working Paper, Department of Civil and Environmental Engineering, University of California at Berkeley*
- Champion, Howard R. "New Tools to Reduce Deaths and Disabilities by Improving Emergency Care: URGENCY Software, Occult Injury Warnings, and Air Medical Services Database." *Uniformed Services University of the Health Sciences*
- Chen, Chao, K. Petty, A. Skabardonis, P. Varaiya, Z. Jia. "Freeway Performance Measurement System – Mining Loop Detector Data". *Transportation Research Record* V. 1748 p. 96.
- Chong, Miao, Ajith Abraham, Mercin Paprzycki. "Traffic Accident Analysis Using Machine Learning Paradigms." *Informatica* 29 p. 89.
- Cohan, Harry, Frank Southwork. "On the Measurement and Valuation of Travel Time Variability Due to Incidents on Freeways." *Journal of Transportation and Statistics*. December 1999
- Davis, Gary A. "Nonparametric Regression and Short-Term Freeway Traffic Forecasting". *Journal of Transportation Engineering* V. 117 p. 178.

- Donnell, Eric, J.M Mason, Jr. "Predicting the Severity of Median-Related Crashes in Pennsylvania by Using Logistic Regression". *Transportation Research Record* V. 1897 p. 55.
- Evanco, William M. "The Impact of Rapid Incident Detection on Freeway Accident Fatalities." *Mitretek Center for Information Systems*. June 1996
- Federal Highway Administration "Traffic Incident Management Handbook" November, 2000.
- Gabauer, Douglas, Hampton C. Gabler. "Evaluation of the Acceleration Severity Index Threshold Values Utilizing Event Data Recorder Technology." *Working Paper, Department of Mechanical Engineering, Rowan University*
- German, Alan, Jean-Louis Comeau, Brian Monk. "The Use of Event Data Recorders in the Analysis of Real-World Crashes" *Proceedings of the Canadian Multidisciplinary Road Safety Conference XII. June 10-13 2001*.
- Hill, John D. Linda Ng Boyle. "Assessing the Relative Risk of Severe Injury in Automotive Crashes for Older Female Occupants." *Accident Analysis and Prevention*. V. 38 p. 148.
- Hobeika, A. G. "A Route Guidance System Under Modified Network Conditions." *Center for Transportation Research, Virginia Polytechnic Institute and State University*.
- Hunter, Jason M., Timothy Matlack, Keven Schweer, Myong S. Sung, Thomas E. Ward, Altaf S. Bahora, Steven C. Davis. "Applications of Networking Capabilities to Assist in Situational Awareness." *Proceedings of the 2004 Systems and Information Engineering Design Symposium*.
- Jia, Zhanfeng, Pravin Varaiya, Chao Chen, Karl Petty, Alex Skabardonis. "Congestion, Excess Demand, and Effective Capacity in California Freeways." *PeMS Development Group. University of California, Berkeley*
- Kim, Karl, N. Larence, J. Richardson, L. Li. "Analyzing the Relationship Between Crash Types and Injuries in Motor Vehicle Collisions in Hawaii". *Transportation Research Record* V. 1467 p. 9
- Lorenz, Matt, Lily Elefteriadou. "A Probabilistic Approach to Defining Freeway Capacity and Breakdown." *Transportation Research Circular E-C018: 4<sup>th</sup> International Symposium on Highway Capacity*.
- Mussa, Renatus, J. E. Upchurch. "Monitoring Urban Freeway Incidents by Wireless Communications". *Transportation Research Record*. V. 1748 p. 153.

- Nathanail, Teti, Kostas Zografos. "Simulation of Freeway Incident Restoration Operations." *1994 Vehicle Navigation & Information Systems Conference Proceedings*.
- Qi, Y., Brian L. Smith. "Identifying nearest neighbors in a large-scale incident data archive". *Transportation Research Record*. V. 1879 p. 89.
- Ridella, Stephen A., Matthew J. McCann, Roy C. Turnbull. "Development of Restraint Systems with Considerations for Equality of Injury Risk." *TRW Automotive, Occupant Safety Systems*. P. 05-0403.
- Saunders, G., A. Gough. "Emergency Ambulances on the Public Highway Linked with Inconvenience and Potential Danger to Road Users." *Emergency Medical Journal*
- Shimamura, Munemasa, Minoru Yamazaki, Goro Fujita. "Method to Evaluate the Effect of Safety Belt Use by Rear Seat Passengers on the Injury Severity of Front Seat Occupants." *Accident Analysis and Prevention*. V. 37 p 5.
- Shojaati, M. "Correlation Between Injury Risk and Impact Severity Index ASI." *3<sup>rd</sup> Swiss Transport Research conference*. March 19, 2003
- Sladover, Steven E. "Potential Contributions of Intelligent Vehicle/Highway Systems (IVHS) to Reducing Transportation's Greenhouse Gas Production." *Conference on Transportation and Global Climate Change: Long-Run Options*. August, 1991.
- Shu, Yantai, Zhigang Jin, Lianfang Zhang, Lei Wang. "Traffic Prediction using FARIMA Models." *IEEE 1999*
- Skabardonis, Alexander, K. Petty, H. Noeimi, D. Rydzewski, P. Varaiya. "I-880 Field Experiment: Data-Base Development and Incident Delay Estimation Procedures". *Transportation Research Record* V. 1554 p. 204.
- Smith, Brian L., M.J. Demetsky. "Multiple-Interval Freeway Traffic Flow Forecasting". *Transportation Research Record* V. 1554 p. 137.
- Smith, Brian L., Jared M. Ulmer. "Freeway Traffic Flow Rate Measurement: Investigation into Impact of Measurement Time Interval". *Journal of Transportation Engineering* V. 129 p. 223-229.
- Smith, Brian L., Ling Qin and Ramkumar Venkatanarayana. "Characterization of Freeway Capacity Reduction Resulting from Traffic Accidents". *Journal of Transportation Engineering* V. 129 p. 362.

Sullivan, Edward. "New Model for Predicting Freeway Incidents and Incident Delays". *Journal of Transportation Engineering* V. 123 p. 267.

Varaiya, Pravin. "Freeway Performance Measurement System: Final Report" *California Partners for Advanced Transit and Highways*. Working Paper UCB-ITS-PWP-2001-1

Virginia Department of Transportation (VDOT)  
[www.vdot.virginia.gov/infoservice/smart-default.asp](http://www.vdot.virginia.gov/infoservice/smart-default.asp)

Yang, Fan, Z. Yin, H. X. Liu, B. Ran. "Online Recursive Algorithm for Short-Term Traffic Prediction". *Transportation Research Record* V. 1879 p. 1.

Zografos, Kostas G., Teti Nathanail and Panos Michalopoulos. "Analytical Framework for Minimizing Freeway-Incident Response Time". *Journal of Transportation Engineering* V. 119 p. 535.