1. Report No. FHWA/TX-05/0-4073-1	2. Government Accession No.	Recipient's Catalog No.
4. Title and Subtitle Probability Generation of Frequency and Severity of Nonrecurring Congestion Due to Accidents to Improve Emissions Analysis		5. Report Date December, 2003
		6. Performing Organization Code
7. Author(s) Lei Yu, Fengxiang Qiao, and Linhua Li		8. Performing Organization Report No. 0-4073-1
9. Performing Organization Name and Address Department of Transportation Studies Texas Southern University 3100 Cleburne Avenue Houston, TX 77004		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 0-4073
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Transfer Office P.O. Box 5080 Austin, Texas 78763-5080		13. Type of Report and Period Covered Technical Report 9/1/02-8/31/03
		14. Sponsoring Agency Code

15. Supplementary Notes

Project conducted in cooperation with Federal Highway Administration
Research project title: Probability Generation of Frequency and Severity of Nonrecurring Congestion Due to Accidents to Improve Emissions Analysis

16. Abstract

Accidents may cause nonrecurring congestion and result in additional delays, vehicle emissions, and other negative effects. Studies have been conducted on the modeling of the accident frequency, duration and even lane blockage. However, there has been less attention paid to the timely updating of the distributions of these variables. This report presents an application of the Bayesian approach to continuously update the distribution of the accident frequency, duration, and affected lanes based on new, available information. Data for the accident frequency in Harris County, Texas, from 1992 through mid-2000 and data for the accident duration and lanes affected in Houston, Texas, from 2000 to mid-2002 were used to construct the historical distributions, while information in recent 6 to 7months were used to update the distributions. Poisson, lognormal, and binomial distributions were modeled for the accident frequency (including variations for time of day and day of week), duration, and lanes affected, respectively. The updating was realized through renewing the conjugate distributions of these modeled distributions, therefore updating key parameters inside the original modeled distributions. The entire process, though incorporating complex probability analysis, was implemented in Microsoft Excel in a rather simple form that practical engineers and researchers can use easily. The methods of microscopic and macroscopic estimation of emissions caused by nonrecurring congestion due to accidents were proposed and case studies in the Houston area show that extra emissions could be generated that would impact emissions in the entire area.

17. Key Words Accident Analysis, Nonrecurring Congestion, Bayesian Approach, Statistical Analysis, Information Updating, Emission Estimation	18. Distribution Statement No restriction. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161 www.ntis.gov

19. Security Classify (of this report) Unclassified	20. Security Classify (of this page) Unclassified	21. No. of Pages 100	22. Price

Probability Generation of Frequency and Severity of Nonrecurring Congestion Due to Accidents to Improve Emissions Analysis

By

Lei Yu, Ph.D., P.E., Fengxiang Qiao, Ph.D., Linhua Li

Report Number 4073-1

Research Project Number 0-4073

Project Title: Probability Generation of Frequency and Severity of Nonrecurring Congestion Due to Accidents to Improve Emissions Analysis

Sponsored by the Texas Department of Transportation

in Cooperation with

U.S. Department of Transportation

Federal Highway Administration

December 2003

Texas Southern University
3100 Cleburne Avenue
Houston, Texas 77004

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Texas Department of Transportation. This report does not constitute a standard, specification, or regulation.

Notice

The United States Government and the State of Texas do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the subject of this report.

Acknowledgments

The authors would like to express their sincere gratitude for the support and valuable comments they received from Project Director William Jordan, Texas Department of Transportation (TxDOT), through the course in conducting this project. Appreciation is also extended to David E. Fink and his coworkers at Houston TranStar, Dr. Ned Levine of the Houston-Galveston Area Council, Carol Nixon of TxDOT's Houston Planning Office, and many others for providing the accident data, traffic information, and valuable comments.

The authors also express their sincere thanks to those who responded to the e-mail survey on emission estimation methods under congestion cases. The authors also thank the research assistants and student helpers in the Department of Transportation Studies at Texas Southern University for collecting the data. Finally, the authors would like to express thanks to all other personnel who provided assistance and contributed directly and indirectly to this project.

Credits for Sponsor

Research performed in cooperation with the Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.

Table of Contents

List of Figures

List of Tables

Summary

Studies focusing on a comprehensive quantification of the traffic delay and mobile emission impact of nonrecurring congestion are rare in the U.S., and currently there is no methodology that can be readily used to measure nonrecurring traffic delay and mobile emissions. Even though incident data sets have been analyzed, it is difficult to generalize the results from one city to another because factors influencing the occurrence of incidents vary among cities.

In this project, we examined and identified the data, methods, and tools necessary to integrate nonrecurring congestion from accidents with transportation and air quality modeling. These include analyzing Texas accident data, probability generation of frequency and severity of accidents, and emission estimation of nonrecurring congestion due to accidents.

Texas accident data from the Texas Department of Transportation System Accident Data File (or TRF accident file) and Houston TranStar were analyzed. TRF data for the analysis were from 1992 through 2000, covering 9 years. Accident duration and lanes affected information was collected from TranStar. The data span from January 2000 to early January 2003.

Accident data analyses were conducted by using the software @Risk, which is embedded into Microsoft Excel. Probability properties of accident frequency, duration, and lanes affected were identified. By using @Risk, the distributions of accident data were identified. It was found that accident frequencies, including accidents by time of day and by day of week, follow either the negative binomial distribution or the Poisson distribution. Most of the distributions of

accident duration followed the lognormal distribution (LD). However, the numbers of lanes affected by accident followed the binomial distribution.

Based on the accident information and corresponding traffic volumes, accidents per million vehicle miles traveled (VMT) by facility and county were also obtained. Basically, the accidents per million VMT decreased in both counties in recent years.

In probability analysis, the Bayesian approach was introduced into the updating of probability parameters of accident frequency and severity based on the new accident information. The proposed approach can tell, for example, whether the accident information from last month is more valuable than the accident data for previous years. The whole process was implemented in a Microsoft Excel spreadsheet and tests were conducted in Harris County.

The updating practice contained two scenarios. The first scenario used the most-recent data from Houston TranStar's Web site to update the historical parameter distribution. New accident information captured from the Houston TranStar Web site was recorded for 5 weeks (April 11 – May 2, 2003, and June 2–16, 2003).

The second scenario used the "pseudo new information" for multiple months to continuously update the historical parameter distributions. Accident frequency for 6 months (July 2000 to December 2000), and duration and lane blockage for 7 months (July 2002 to January 2003) was used to update the prior distribution of historical parameter six and seven times (i.e., month by month), respectively.

The estimations of emission of the nonrecurring congestion because of accidents were conducted including the microscopic calculation and the macroscopic evaluation. The microscopic calculation estimates the emissions caused by each individual accident. The macroscopic evaluation evaluates the impact to the whole area of emissions caused by nonrecurring congestion due to accidents.

In the microscopic evaluation of emissions, the Microsoft Excel worksheet was designed to calculate the emissions caused by the individual accident. Lane affected and accident duration were two of the multiple inputs on the worksheet. Three days' real accident information, which was obtained from Houston TranStar, was used to test the calculation of the worksheet.

The macroscopic evaluation was based on the required Environmental Protection Agency emission estimation model MOBILE6. Accident frequency, duration, and lane affected result in the changes of several input variables of MOBILE6. These variables are VMT by facility, VMT by hour, speed VMT, etc. All these changes will result in the change of emission estimations by MOBILE6.

By simulation and estimation, it is found that the nonrecurring congestion due to accidents affects emissions. For 3 typical days, the extra emissions caused by accidents in the Houston area varied from 143.69 lbs to 488.66 lbs for volatile organic compounds (VOC), and from 53.41 lbs to 181.66 lbs for oxides of nitrogen (NO_x).

The nonrecurring accidents also impact the emissions in the entire area. Because a small change of emission factors will result in a large amount of the change of total emissions, the impact of nonrecurring congestion to total emissions cannot be neglected.

Further processing and testing of the accident data for other cities/counties in Texas is recommended. The Microsoft Excel worksheet developed for this research can be used in the implementation stage with necessary improvements. The corresponding emissions impact of nonrecurring congestion due to accidents in other areas also can be estimated.

CHAPTER 1

INTRODUCTION

1.1 Background of Research

Traffic congestion is a critical problem in urban areas. Between 1976 and 1996, the number of vehicle miles traveled (VMT) in the United States increased by 77 percent, while the mileage of roads and streets increased only by 2 percent (FHWA 1998). Over the years, the percentage of the peak-hour VMT that occurs under congested conditions has increased steadily, although at a slower pace in recent years. Congestion usually results in time delays, increased fuel consumption, pollution, stress, health hazards, and added vehicle wear.

Congestion is called recurring when it is triggered by (1) a daily event, or (2) a periodic event, such as a baseball game at the local stadium (McShane 1990). Given recurring congestion of either type, a response can be preprogrammed for a certain time interval, or the system can "discover" the event and invoke a response. The response can be fixed or adaptive.

Congestion may be nonrecurring when (1) the actual congestion occurs only on certain days, or only periodically within a day, but at well-known problem locations; or (2) the actual congestion occurs rarely at the specific site, and due to a truly unusual event (McShane 1990). An example of the first case is the congestion that occurs on an arterial because double-parked trucks at certain locations temporarily eliminate a lane on the arterial or on a side street, causing the spillback. An example of the second is an accident on a street that normally has no problems.

In estimating the mobile source emissions on roadway networks, it is a common practice to consider the emissions caused in general cases of transportation operations including the recurring congestion, with little consideration for the impact of nonrecurring congestion due to accidents. Studies focusing on a comprehensive quantification of the traffic delays and mobile emissions due to nonrecurring congestion are rare and there are currently few methodologies that can be readily used to measure nonrecurring traffic delays and mobile emissions. However, it is difficult to generalize the results from one city to another because factors influencing the occurrence of incidents vary across cities.

The nonrecurring congestion in urban areas due to vehicle accidents has become one of the major causes of increased delays, road-user costs, and mobile source emissions. Motor vehicle emission budgets and project analysis in nonattainment areas rely on regional travel-demand model results and subsequent mobile source emissions estimates. Regional mobile source emission estimates could be improved when impact from nonrecurring congestion due to accidents is incorporated. The nonrecurring congestion caused increased mobile source emissions in two ways: (1) by increasing the idle times of stopped vehicles, and (2) by increasing the acceleration and deceleration activities of slowed-down vehicles. The frequency and severity of nonrecurring incidents on urban roadway sections could be developed and associated with the estimated travel and emissions results. These results could be used to improve mobile source emissions inventories, motor vehicle emissions budgets, and a better assessment of projects and programs that mitigate nonrecurring congestion incidents and their effects.

1.2 Objectives of Research

The aim of this research is to examine and identify the data, methods, and tools necessary to integrate nonrecurring congestion due to accidents into transportation and air quality modeling. Specific objectives are to:

- collect and analyze Texas accident data;
- generate and update the probability parameters of frequency, duration, and lane blockage of accidents;
- estimate the extra emissions of nonrecurring congestion due to accidents.

1.3 Outline of Report

The next chapter of this report presents the literature review of the state of the art and state of the practice in the modeling of accident frequency, duration, and lane blockage. Chapter 3 describes the property of accident data. Chapter 4 introduces the method of how to update frequency, duration, and lanes affected based on new information. Chapter 5 presents the practice of updating probability properties based on new information in Harris County and Houston. Chapter 6 provides the emission estimations with congestion due to accidents. Finally, Chapter 7 gives conclusions for this report.

CHAPTER 2

REVIEW OF STATE OF THE ART AND STATE OF THE PRACTICE

This chapter reviews state of the art and state of the practice on nonrecurring congestion, accident probability models, Bayesian approach on accident analysis, and emission estimation models.

2.1 Characteristics of Congestion

Congestion is a peculiar phenomenon. It now results in 5.7 billion annual person hours of delay in the United States (Congestion Mitigation 2003). The resulting traffic slowdowns can have a wide range of negative effects on people and on the business and economy, including impact on air quality (due to additional vehicle emissions), quality of life (due to personal time delays), and business activity (due to the additional costs and reduced service areas for workforce, supplier, and customer markets).

Schrank and Lomax (2003) estimated that the total cost of congestion in seventy-five urban areas was $69.5 billion in 2001, or an average of $520 per person each year in the U.S. Twenty urban areas had a total annual congestion cost of at least $1 billion each, and the areas with populations over 3 million account for more than 60 percent of the congestion cost. They also estimated that 90 percent of total congestion costs in major urban areas is attributed to travel delays, with the other 10 percent attributed to fuel costs.

Of the transportation characteristics most frequently covered in the press, congestion is the source of the bitterest complaints. In several metropolitan areas, voters ranked traffic congestion as the number one problem superseding the traditional concerns of crime, housing, and unemployment (Al-Deek H. 2003). Only in recent years has coping with congestion taken its rank as a major, if not dominant, mission of most state transportation departments.

Though unpleasant, congestion is a symbol of robust economic activity and prosperity. Many politicians and members of the business community fear that some efforts to mitigate congestion, particularly those oriented toward the demand side, may damage the local economy (Hensing 1996).

Congestion is inevitable. It is growing everywhere, evidenced in longer and longer peak periods and it has spread to locations and times heretofore congestion free. Thus, we are cautious in the verbs we choose in dealing with congestion. We can "fight" it, "reduce" it, "cope with" it, and "ease" it. But we cannot "eliminate," "erase," or "destroy" it.

There is a certain amount of imprecision in the definition of congestion. In its rulemaking concerning Congestion Management Systems, the Federal Highway Administration (FHWA) defines congestion as "the level at which transportation system performance is no longer acceptable due to traffic interference" (Congestion Mitigation 2003). The key word, of course, is "acceptable." What is acceptable to someone in Chicago may not be acceptable in Houston, San Francisco, or Des Moines, and so on down the population scale. Complaints about congestion may be heard in yet smaller cities, and event-based or seasonal congestion in rural areas can be severe. Thus, there is a relative sense to the word "congestion," which is strengthened by the continuation of the FHWA definition: "The level of system performance may vary by type of transportation facility, geographic location (metropolitan area or sub-area, rural area), and/or time of day." However, regardless of location or timing, the frequency and severity of congestion delays increase as traffic volumes grow to exceed the road system capacity.

A popular definition of congestion is a condition of traffic delay when the flow of traffic is slowed below reasonable speeds because the number of vehicles trying to use the road exceeds the traffic network capacity to handle them.

Congestion may be a "recurring" form of congestion, which is repeating as to location or time of day. Or, it may be a "nonrecurring" form of congestion, which is due to an abnormal occurrence such as a traffic accident or maintenance activities.

Recurring congestion is primarily a product of when the demand to use a particular roadway exceeds the roadway's capacity, and it tends to be concentrated into short time periods, such as "rush hours." Thus, recurring congestion is used to refer to repetitive, predictable, peak-hour congestion. Recurring congestion is commonly addressed by the use of policy options such as transit, growth management, traffic operational improvements, and transportation demand measures.

Nonrecurring congestion, however, has been estimated to account for as much as 60 percent of the congestion experienced in a transportation system (Office of Operations 2003). In the next section, nonrecurring congestion is discussed.

2.2 Nonrecurring Congestion

Nonrecurring congestion refers to congestion that occurs on an irregular basis. In contrast to the recurring congestion, which is due to daily morning and afternoon commute travel, nonrecurring congestion often arises from events that occur separately from daily commuter-related congestion. It occurs at random locations and times.

Nonrecurring congestion is the result of traffic accidents (crashes); vehicle breakdowns (stalled vehicles, spilled loads, hazardous material spills); maintenance/construction activities (work zones); police activity; adverse weather conditions; rubbernecking; and other unforeseen events. The causes of nonrecurring congestion can be divided into two categories: incidents and special events. Incidents include accidents, distractions, and vehicle breakdowns that reduce traffic flow. Special events include bad weather conditions, police activity, etc.

National statistics indicate that more than 60 percent of urban freeway congestion in the U.S. is related to incidents. It is expected to exceed 70 percent by 2005, so it is necessary to develop methods for estimating highway incident congestion.

Much of the reason for urban freeway congestion is that when the flow of traffic is impeded or stopped, delay increases exponentially as the number of vehicles and occupants back up along the route.

Nonrecurring congestion also affects alternative routes by forcing unanticipated traffic volumes onto lesser-used facilities, thus increasing the congestion on the alternate routes. These effects will continue for extended time periods and on additional routes following an event as travelers seek alternate routes with fewer delays.

2.3 Accident Frequency, Duration, and Land Blockage Models

Incidents are different from accidents, which are only one part of incidents. However, most of the analytical methods for frequency, duration, and lanes affected by incidents and accidents are essentially the same. Therefore, this literature review includes models and statistical methods for both incidents and accidents.

2.3.1 Frequency Analysis

Incident or accident frequency has been studied from different perspectives and using varied methodologies. For example, it has been studied either for varying levels of severity—such as property damage, injury, or fatality—or for the different areas where the incidents happened—such as urban versus rural, freeway versus street, intersection versus middle block, etc. Most of the studies focused on identifying the influencing factors of accident frequency.

Hamerslag et al. (1982) studied how the expected number of accidents depends on road and traffic characteristics. In other work, Okamoto and Koshi (1989) used linear regression to model relationships between accident rates and the geometric design of roads. Miaou et al. (1992) established empirical relationships between truck accidents and key highway geometric design components by using a Poisson regression approach. Hadi et al. (1995) used a negative binomial regression analysis to estimate the effects of cross-sectional design elements and found that increasing lane width, shoulder width, center shoulder width, and median width were significant in reducing the number of accidents. Karlaftis and Tarko (1998) also developed separate negative binomial models for accidents for separate groups: urban suburban and rural counties. Lee and Mannering (1999) investigated the relationships among roadway geometry, roadside characteristics, and run-off-roadway accident frequency and severity. Teng and Qi (2002) used Poisson and negative binomial models and their zero-inflated models to analyze incident frequency and found that rain is the only factor that significantly influenced incident frequency.

From the methodological perspectives, many applications of the accident frequency statistical modeling have been undertaken. Miaou and Lum (1993) demonstrated that conventional linear regression models were not appropriate for modeling vehicle accident events on roadways, and the test statistics from these models were often erroneous. They concluded that Poisson and negative binomial regression models were a more appropriate tool in accident modeling.

Shankar et al. (1997) argued that the traditional application of Poisson and negative binomial models did not address the possibility of zero-inflated counting processes. Lambert (1992) distinguished the truly safe road section (zero accident state) from the unsafe section (non-zero accident state, but with the possibility of having zero observed accidents) to show that a zero-inflated model structure is often appropriate for estimating the accident frequency of road sections. Zero-inflated probability processes, such as the zero-inflated Poisson and zero-inflated negative binomial regression models, allow one to better isolate independent variables that determine the relative accident likelihoods of safe versus unsafe roadways. In other work, Miaou (1994) evaluated the statistical performance of three types of models (Poisson regression, negative binomial regression, and zero-inflated Poisson regression) in studying the relationship between truck accidents and the geometric design of road sections. Miaou recommended that the Poisson regression model be an appropriate model for developing the relationship when the mean and variance of the accident frequencies were approximately equal. If the overdispersion was moderate or high, the use of both the negative binomial regression and zero-inflated Poisson regression were found to be more appropriate. However, in general, the zero-inflated Poisson regression model seemed a justified model when accident data exhibited a high zero-frequency state.

2.3.2 Duration Analysis

The length of duration from the time an incident happens to the time the traffic recovers to normal is defined as the duration of accident. The duration of an accident is composed of four important and distinct components: detection, response, clearance, and recovery (TRB 1994). Figure 1 illustrates the four phases of a freeway accident. It is obvious that accident occurred, detected, responded, cleared, and traffic flow restored to normal are the distinct points of the phases.

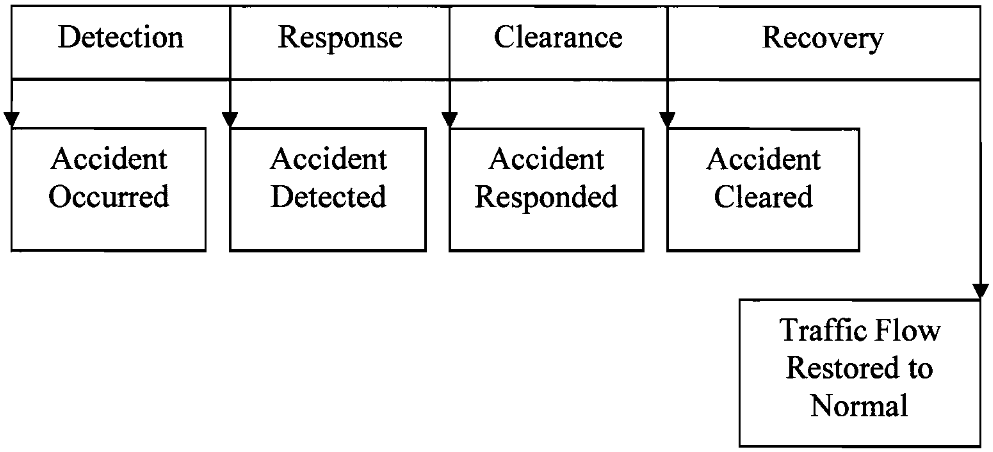


FIGURE 1 Four phases of a freeway accident.

Golob et al. (1987) analyzed freeway accidents that involved trucks. Based on the hypothesis that the length of each phase was influenced by the length of the preceding phases, they were able to conclude that the total duration of an accident is modeled according to a lognormal distribution.

Other studies by Giuliano (1989), Garib et al. (1997), and Sullivan (1997) have supported the use of a lognormal distribution to describe the freeway incident duration. Jones et al. (1991) used a similar distribution, the log-logistic distribution to fit a specific data set from the Seattle area. Nam and Mannering (2000) found that the Weibull distribution also could be used to describe certain incident data.

Ozbay and Kachroo (1999) studied the incident clearance by using a linear regression function. Garib et al. (1997) also developed a linear regression model to predict the incident duration.

Another use of probability in incident duration is to develop conditional probabilities. Traffic managers may be interested in the probability of an incident lasting 5 minutes given that it has already been active for 15 minutes, or other similar cases. Jones et al. (1991) reported on conditional probabilities. Nam and Mannering (2000) also used conditional probabilities in hazard-based models to find the likelihood that an incident would end in the next short time period given its continuing duration.

Khattak et al. (1995) developed the time sequential model. The authors identified ten distinct stages of the incident duration based on the availability of information. The length of time for each stage differed for each incident, but it was truncated after a maximum of 10 minutes. Each stage had a separate truncated regression model, and the models progressively add more variables. The time sequential model was not tested or validated in the study due to a small sample size of 109 freeway accidents.

2.3.3 Lane Blockage Analysis

There have not been many studies in which the number of lanes blocked by an incident or accident has been discussed. The reason may be that defining the blockage is difficult. Furthermore, few former studies have been devoted to identifying the influencing factors for the number of lanes blocked by an incident. The most recent study, by Teng and Qi (2002), used an ordered probit model to better capture the order inherent in the number of lanes blocked during an incident. Because this model can recognize the ordinal nature of a dependent variable, it is especially appropriate for modeling the number of lanes blocked by an incident, which are ordered in nature.

2.3.4 Comments on Current Methods

The current methodologies in analyzing the accident frequency, duration, and number of lanes blocked are all based on historical data. Therefore, the research results are basically the summary of historical information. To predict these three variables in the future, the latest accident information should be incorporated into the model. For example, for prediction purposes, the accident information from last year should be more valuable than the accident data from two years ago. The Bayesian approach can incorporate historical and current information together and update old estimations with new ones, giving different weight to historical data and new data based on the important underlying structure of the data (Ang and Tang 1975).

2.4 Bayesian Approach on Accident Analysis

Bayesian analysis is not very new for accident analysis. Heydecker and Wu (1991) examined proportions of accidents in England that occurred at a site with a given feature, depending on the distributional assumptions as a binomial distribution with its mean θ

postulated to be beta distribution. This work was extended by Bolduc and Bonin (1998) by involving the use of the multinomial distribution for accident data in Canada and the Dirichlet distribution for the parameter θ. Nembhard and Yong (1995) found that Bayes estimators behave rationally in estimating truck accident rates. In the research in this report, the Bayesian approach will be used further in the analysis and prediction of accident frequency, duration, and lanes affected.

The classical statistical approach assumes that the parameter θ (e.g., average accident occurrence in the past years) from mass field data is constant (but unknown) and the sample statistics are used as the estimator of θ. Different from the classical methods, the Bayesian approach holds that the unknown parameter θ with a distribution also is assumed (or modeled) to be a random variable.

By Bayesian approach, possible values of parameter θ can be assumed to be a data set $\{\theta_i\}$ ($i = 0, 1, 2, ..., n$) with relative likelihoods $\{p_i = P'(\theta = \theta_i)\}$, which is the *prior* distribution of θ. Then, if additional information ε (e.g., average accident occurrence in the recent month) becomes available, the *prior* distribution on θ may be modified formally to the *posterior* distribution $P''(\theta = \theta_i)$ as follows:

$$P''(\theta = \theta_i) = \frac{P(\varepsilon|\theta = \theta_i)P'(\theta = \theta_i)}{\sum_{i=1}^{n} P(\varepsilon|\theta = \theta_i)P'(\theta = \theta_i)} . \quad \text{(Eq 1)}$$

In Equation 1, $P(\varepsilon|\theta = \theta_i)$ is the conditional probability or likelihood of the experimental outcome ε assuming that the parameter is θ_i.

The denominator is proportionally constant. Let $L(\theta)$ represent the likelihood function, Equation 1 is expressed as

$$P''(\theta = \theta_i) = kL(\theta) \cdot P'(\theta = \theta_i) . \quad \text{(Eq 2)}$$

The expected value of θ is then commonly used as the Bayesian estimator of the parameter as

$$\theta'' = E(\theta|\varepsilon) = \sum_{i=1}^{n} \theta_i P''(\theta = \theta_i). \quad \text{(Eq 3)}$$

For example (Ang and Tang 1975), suppose from previous experience with similar conditions, the average accident rate ν at an improved intersection would be between one and three per year with an average of two. The probabilities for one, two, and three accidents per year are 0.3, 0.4, and 0.3, respectively. The accident occurrence is assumed to be a Poisson process. During the first month after construction, one accident occurred. This is the new information denoted as ε. Then, the posterior probabilities are

$$P''(\nu=1)=\frac{e^{-1/12}(1/12)(0.3)}{e^{-1/12}(1/12)(0.3)+e^{-1/6}(2/12)(0.4)+e^{-1/4}(3/12)(0.3)}=0.166.$$

Similarly, $P''(\nu=2)=0.411$ and $P''(\nu=3)=0.423$.

In the above analysis, the possible values of parameter θ were a discrete set of values. In many situations, however, the value of a parameter could be a continuum of possible values. For example, the average length of accident duration could be assumed as a continuous random variable. In this case, the corresponding results would be as follows, being analogous to Equations 1–3.

Let θ be the parameter with a continuous prior density function $f'(\theta)$. The prior distribution $f'(\theta)$ can be revised in light of the experimental outcome ε using the Bayesian theorem. The posterior probability for θ will be

$$f''(\theta)=\frac{P(\varepsilon|\theta)f'(\theta)}{\int_{-\infty}^{\infty}P(\varepsilon|\theta)f'(\theta)d\theta}. \qquad \text{(Eq 4)}$$

The term $P(\varepsilon|\theta)$ is the conditional probability or likelihood of observing the experimental outcome ε assuming the value of the parameter to be θ. Hence $P(\varepsilon|\theta)$ is a function of θ and is commonly referred to as the *likelihood* function of θ and denoted as $L(\theta)$. The denominator is independent of θ, which is simply a normalizing constant required to make $f''(\theta)$ a proper density function. Then, Equation 4 can be expressed as

$$f''(\theta)=kL(\theta)f'(\theta). \qquad \text{(Eq 5)}$$

The updated estimate of the parameter θ is given by

$$\theta'' = E(\theta|\varepsilon) = \int_{-\infty}^{\infty} \theta f''(\theta) d\theta . \qquad \text{(Eq 6)}$$

The discrete case of Bayesian approach can be used in the updating of accident frequency and lanes affected, while the continuous one is suitable for accident duration updating. Before using the Bayesian approach, the properties of accident data (including the distribution of the sample variables) should be identified beforehand.

2.5. Air Quality Models Reflecting the Impact of Congestion

Over the past decades, numerous emission models have been developed for estimating mobile source emissions, which include the Environmental Protection Agency (EPA) emission factor model MOBILE (for all states except California) and EMFAC (California only), comprehensive modal emission models CMEM, traffic simulation based emission models such as INTEGRATION and CORSIM, and numerous other microscopic emission models.

MOBILE is the EPA standard, which was developed to estimate overall emissions levels, trends over time, and the effectiveness of mobile-source emission-control strategies. It deals with the on-road vehicle emission factors of carbon monoxide (CO), volatile organic compounds (VOC), and oxides of nitrogen (NOx) at different speed values. Even the latest version, MOBILE6, does not allow users to get additional values in the descriptive output for the emission rate of vehicles idling in grams per hour, although it has included vehicle idling in proportion to normal driving. However, MOBILE6 requires input of many mobile source-related traffic indicators such as VMT BY FACILITY, VMT BY HOUR, speed VMT, and VMT Fraction. It is hoped that the effective use of these indicators could reflect a variety of scenarios under either recurring or nonrecurring congestion.

CHAPTER 3

ANALYZING TEXAS ACCIDENT DATA

In this chapter, the accident data from the Department of Public Safety (DPS) and Houston TranStar will be analyzed. Theoretically, accidents are not equal to vehicle crashes. For example, the crashes caused purposely are not accidents. However, in this research, the term "accident" is used rather than "crash."

3.1 Data Preparation

Based on the data available, the accident frequency analysis is focused on freeways in Harris County, Texas. The duration and lanes affected analyses are for freeways in the Houston area. The main reason for selecting Harris County and Houston for analysis is TranStar.

Accident frequency data were obtained from the Texas Department of Transportation (TxDOT) System Accidents Data File (TRF accident files), which has been processed based on data from DPS. Completed records of accidents for the entire Texas freeway system are stored in this file, including accident location, severity, weather, time of accidents, etc. Data used for the later analysis are for Harris County and Travis County, covering 9 years—from 1992 through 2000. It should be noted that at the time of this report's preparation, the 2000 accident data file was the most recent to be released to the public. Information such as duration and lanes affected is not included in the TRF accident file.

Accident duration and lanes affected data were collected from Houston TranStar. The data spans from January 2000 to early January 2003. Similarly, accident duration and lanes affected information after this time period has not been entered into the database. However, TranStar provides instant accident information to the public on the Internet. The online information includes accident location, accident detected time and/or verified time, moved time and/or cleared time, accident vehicle type, the number of vehicles involved, the number of lanes affected, and even the speed on each freeway segment from automatic vehicle identification (AVI) data records. However, this instant information, which will be cleared after each accident is gone, is a source of "new information." It should be noted that Houston TranStar uses "lanes affected" instead of "lanes blocked" to describe the lanes that have no through capacity or low through capacity. In order to align with TranStar's terminology, this report will refer to "lanes affected."

3.2 Characteristics Analysis

The TRF accident files were obtained from TxDOT for 1992 through 2000. All the data were used for the analysis of accident characteristics. Data processing was conducted in Microsoft Excel with proper data arrangements among different worksheets.

The total number of accidents in Texas obtained from the TRF accident file varies from year to year, as shown in Figure 2. However, the figure shows that the total accident quantity has a trend of decreasing after 1994. It is perhaps due to the development of incident management systems in Texas.

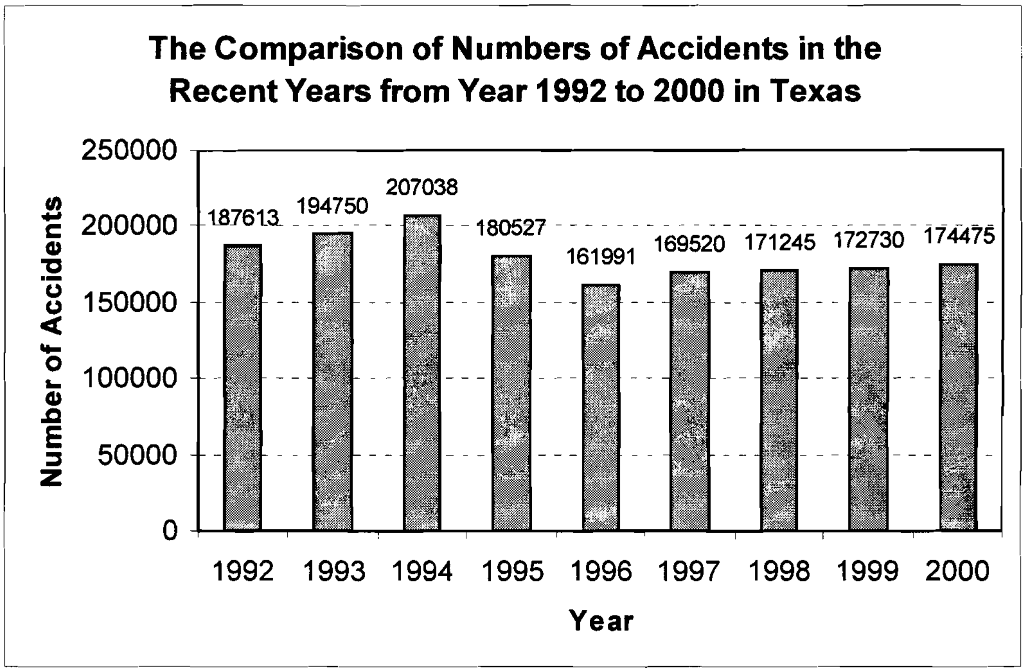


FIGURE 2 Comparison of number of accidents from 1992 to 2000 in Texas.

The research team chose the nonattainment area of Harris County in the Houston area and the near-nonattainment area of Travis County in the Austin area as the research counties. The overall accident comparison of Harris County to Travis County covers 9 years (1992–2000) and is listed in Table 1.

TABLE 1 Overall Comparison of Accidents in Harris County and Travis County

	Harris County (Nonattainment)	Travis County (Near Nonattainment)
Area (Square Miles)	1778	989
Average Number of Accidents Per Square Mile Per Year	15.45	7.24

From Table 1, it is seen that there are fewer accidents per square mile in the near-nonattainment county (Travis) than in the nonattainment county (Harris). The accidents that happened contribute more idling emissions.

The comparison of accident frequencies can be found from Figure 3 and Figure 4. Figure 3 compares the accident frequencies between the two counties by day of week, while Figure 4 compares the accident frequencies between the two counties by time of day. Basically, there are similar trends between the two counties although the absolute values of the frequencies are different.

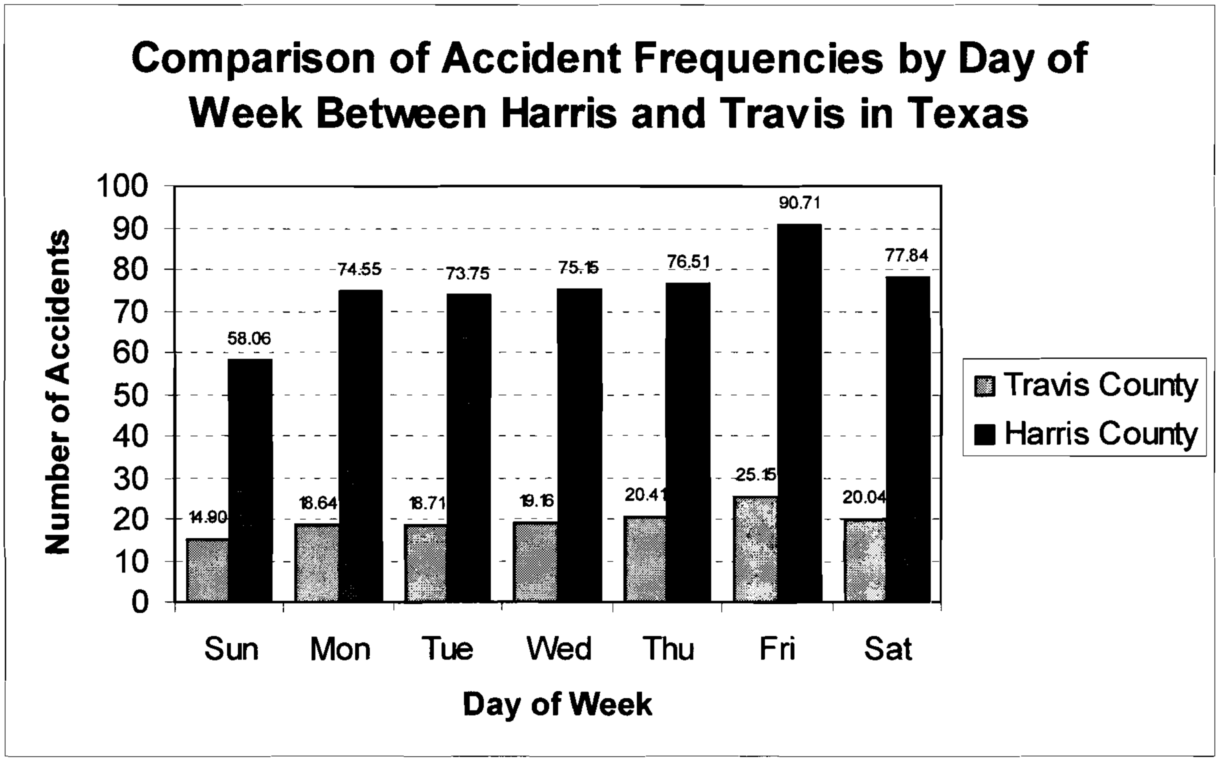


FIGURE 3 Comparison of accident frequency in Harris County and Travis County by day of week.

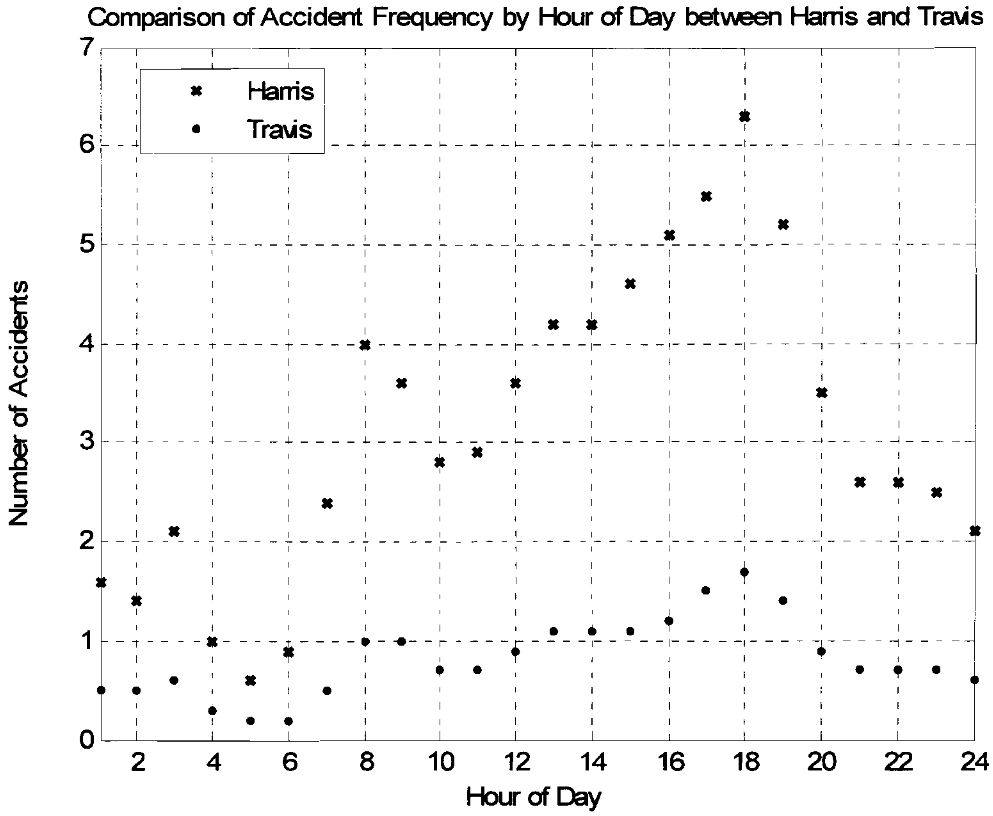


FIGURE 4 Comparison of accident frequency in Harris County and in Travis County by time of day.

Figure 3 shows that Friday has the highest accident frequency. A possible reason is that people are eager to drive home for the weekend and, therefore, pay less attention to driving safety on Fridays. Saturday also has a high frequency in both counties, although the traffic volumes are much smaller than on workdays.

From Figure 4, the accident frequencies have two peaks in one day. One peak is from 7:00 a.m. to 9:00 a.m. The other peak is from 3:00 p.m. to 7:00 p.m. Both of these peaks happen during the high traffic volume times—the morning peak and afternoon peak. This means that the accident is proportional to the traffic volume. The accident, which happens during the peak hours, causes the nonrecurring congestion.

The data on the accident duration and lanes affected are obtained from Houston TranStar incident files. There are four categories among accidents, which are classified according to the severity of the accidents and the types of vehicles involved. The four categories are non-heavy truck minor accident, heavy truck minor accident, non-heavy truck major accident, and heavy truck major accident. Obviously, different categories have different impacts on the duration and lanes affected. Therefore, the following analyses are divided into these four categories. It is noted that only 13,842 accidents in the database can be used in the analyses of the duration and lanes affected due to the absence of some part of the data such as cleared time or occurred time of accidents.

Table 2 shows the average duration for the four categories. From the TranStar data, the standard deviations were calculated for each category. The confidence level of 90 percent was used in calculating intervals. The error rate was calculated by the division of interval and average duration.

TABLE 2 Comparison of Accident Duration by Different Categories

	Sample Size	Percentage	Average Duration (minutes)	Standard Deviation	Confidence Interval (minutes)	Error Rate
Non-Heavy Truck Minor Accidents	7930	57.29%	33.67	39.37	±0.73	2.16%
Heavy Truck Minor Accidents	539	3.89%	87.68	101.35	±7.18	8.19%
Non-Heavy Truck Major Accidents	5002	36.14%	43.85	45.93	±1.07	2.44%
Heavy Truck Major Accidents	371	2.68%	120.03	124.88	±10.67	8.88%
Total	13842	100.00%	41.77	52.46	±0.73	1.76%

From Table 2, heavy truck related accidents have much longer duration than non-heavy truck related accidents. This means that non-heavy truck accidents and heavy truck accidents are two totally different kinds of accidents. The impact to the freeway can also be significantly different. The reason is that heavy trucks are more difficult to move and clear than non-heavy trucks. However, heavy truck accidents make up only 6.5 percent of all accidents. The average duration of non-heavy truck major accidents is 10 minutes longer than that of minor ones, while the duration of heavy truck accidents is over 30 minutes longer than heavy truck minor accidents.

Table 3 shows the durations by different lanes affected. The confidence rate of 90 percent was also used to calculate intervals.

TABLE 3 Comparison of Duration by Different Lanes Affected

Lanes Affected	Sample Size	Average Duration (minutes)	Standard Deviation	Confidence Interval (minutes)	Error rate
0 lane	990	44.48	49.81	±2.60	5.85%
1 shoulder	1704	42.56	37.56	±1.50	3.52%
2 shoulders	35	40.67	22.41	±6.40	15.74%
1 lane	3954	30.56	38.09	±1.00	3.26%
1 lane + 1 shoulder	1760	38.01	43.11	±1.69	4.45%
1 lane + 2 shoulders	38	42.66	37.54	±10.27	24.08%
2 lanes	2122	37.38	39.34	±1.40	3.76%
2 lanes+ 1 shoulder	918	45.86	49.62	±2.69	5.87%
2 lanes+ 2 shoulders	29	41.66	35.73	±11.29	27.09%
3 lanes	457	46.77	45.84	±3.53	7.54%
3 lanes+ 1 shoulder	225	51.31	43.19	±4.74	9.23%
3 lanes+ 2 shoulders	10	42.72	36.82	±21.34	49.96%
4 lanes	44	65.38	82.34	±20.87	31.92%
4 lanes+ 1 shoulder	36	60.96	39.41	±11.09	18.20%
4 lanes+ 2 shoulders	6	62.51	41.92	±34.48	55.17%
5 lanes	1	41	N/A	N/A	N/A
5 lanes+ 1 shoulder	4	36.5	23.21	±27.31	74.81%
5 lanes+ 2 shoulders	4	76.2	90.94	±106.99	140.41%
6 lanes	0	N/A	N/A	N/A	N/A
6 lanes+ 1 shoulder	1	119.75	N/A	N/A	N/A
6 lanes+ 2 shoulders	0	N/A	N/A	N/A	N/A
All lanes	370	100.48	117.13	±10.03	9.97%
HOV	39	57.81	61.85	±16.69	28.87%
Ramp	1094	63.34	91.44	±4.55	7.18%

From the above table, the relative durations increase as the lanes affected increase. It is seen that the duration for zero-main-lane-affected accidents (including 0 lane affected, 1 shoulder affected, and 2 shoulders affected accidents) is longer than 1 lane affected accidents. That is because there was no hurry to clear an accident that did not affect the main lane traffic.

Table 4 shows the lanes affected by different accident categories. It is shown that the average number of lanes affected by minor accidents is less than the number affected by major accidents, and the average number of lanes affected by non-heavy truck accidents is less than the number affected by heavy truck accidents.

TABLE 4 Number of Lanes Affected by Different Categories

	Sample Size	Percentage	Average Number of Lanes Affected	Standard Deviation	Interval	Error Rate
Non-Heavy Truck Minor	7930	57.29%	1.08	0.83	0.02	1.43%
Heavy Truck Minor	539	3.89%	1.32	1.09	0.08	5.86%
Non-Heavy Truck Major	5002	36.14%	1.44	1.05	0.02	1.69%
Heavy Truck Major	371	2.68%	1.69	1.42	0.12	7.19%
Total	13842	100.00%	1.23	0.97	0.01	1.09%

3.3 Data Distribution Analysis

The main task of the data distribution analysis is to identify the proper distributions of the accident frequency, duration, and lanes affected for the prepared data.

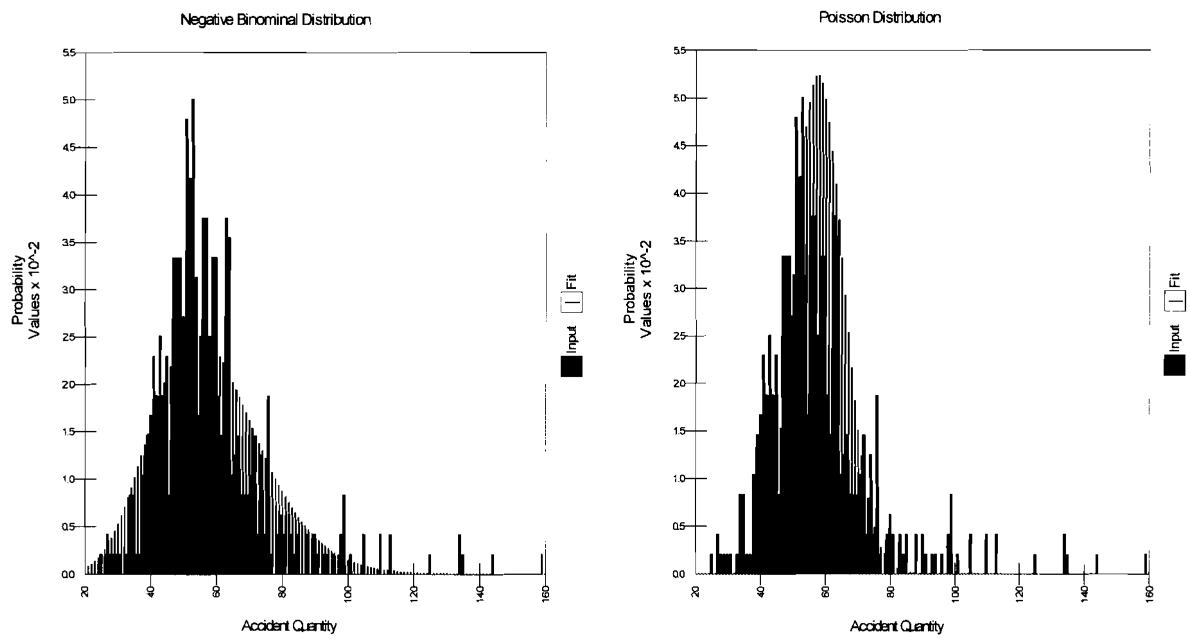
3.3.1 Identify Distribution for Frequency

Data used for identifying the frequency distribution of Harris County freeway accidents are from January 1992 through June 2000. The remainder of the frequency data (July 2000 through December 2000) was left for updating the parameters of the distribution. In other words, this part of the data was reserved as the "new information."

The software @Risk 4.5, which is embedded in Microsoft Excel, was used for identifying the distributions. Accident frequencies classified by day of week and by time of day were analyzed. The results showed that negative binomial distribution (NBD) and Poisson distribution

(PD) are the best two distributions for accident frequencies for both day of week and time of day. Both of the distributions passed the goodness-of-fit tests. Poisson distribution is used for the later analysis (updating parameter based on Bayesian approach) because it is relatively simple.

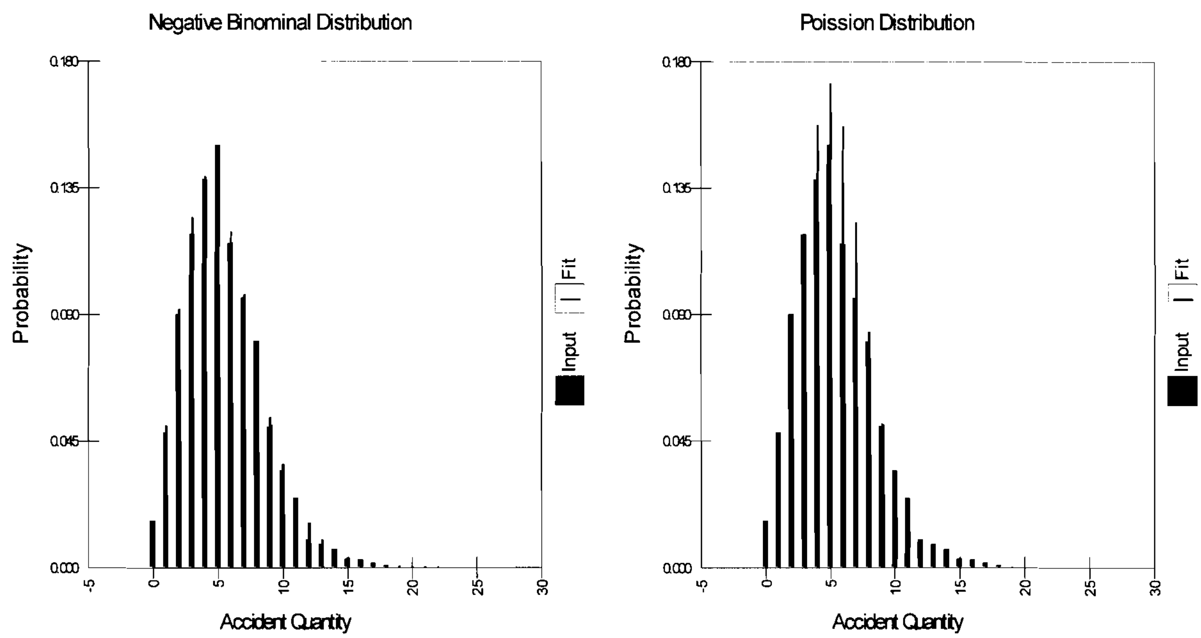
Figure 5 shows the accident frequency distribution on Sundays in Harris County.



(a) By NBD (b) By PD

FIGURE 5 Accident frequency distribution on Sundays in Harris County.

Figure 6 shows the accident frequency distribution estimation during the time period from 5:00 p.m. to 6:00 p.m. in Harris County. By the chi-square test, both NBD and PD are acceptable.



(a) By NBD (b) By PD

FIGURE 6 Accident frequency distribution from 5:00 p.m. to 6:00 p.m. in Harris County.

Table 5 and Table 6 list the density functions for distributions by day of week and time of day in Harris County.

TABLE 5 Distribution of Accidents by Day of Week in Harris County

	Mean	Negative Binomial Distribution	Poisson Distribution
Monday	74.5544	$P(x)=\binom{21+x}{x}0.2279^{22}*0.7721^{x}$	$P(x)=\frac{74.5544^{x}}{x!}e^{-74.5544}$
Tuesday	73.7490	$P(x)=\binom{35+x}{x}0.3280^{36}*0.6720^{x}$	$P(x)=\frac{73.7490^{x}}{x!}e^{-73.7490}$
Wednesday	75.1503	$P(x)=\binom{31+x}{x}0.2986^{32}*0.7014^{x}$	$P(x)=\frac{75.1503^{x}}{x!}e^{-75.1503}$
Thursday	76.5073	$P(x)=\binom{26+x}{x}0.2609^{27}*0.7391^{x}$	$P(x)=\frac{76.5073^{x}}{x!}e^{-76.5073}$
Friday	90.7077	$P(x)=\binom{30+x}{x}0.2547^{31}*0.7453^{x}$	$P(x)=\frac{90.7077^{x}}{x!}e^{-90.7077}$
Saturday	77.8434	$P(x)=\binom{18+x}{x}0.2547^{19}*0.7453^{x}$	$P(x)=\frac{77.8434^{x}}{x!}e^{-77.8434}$
Sunday	58.0585	$P(x)=\binom{15+x}{x}0.2160^{16}*0.7840^{x}$	$P(x)=\frac{58.0585^{x}}{x!}e^{-58.0585}$

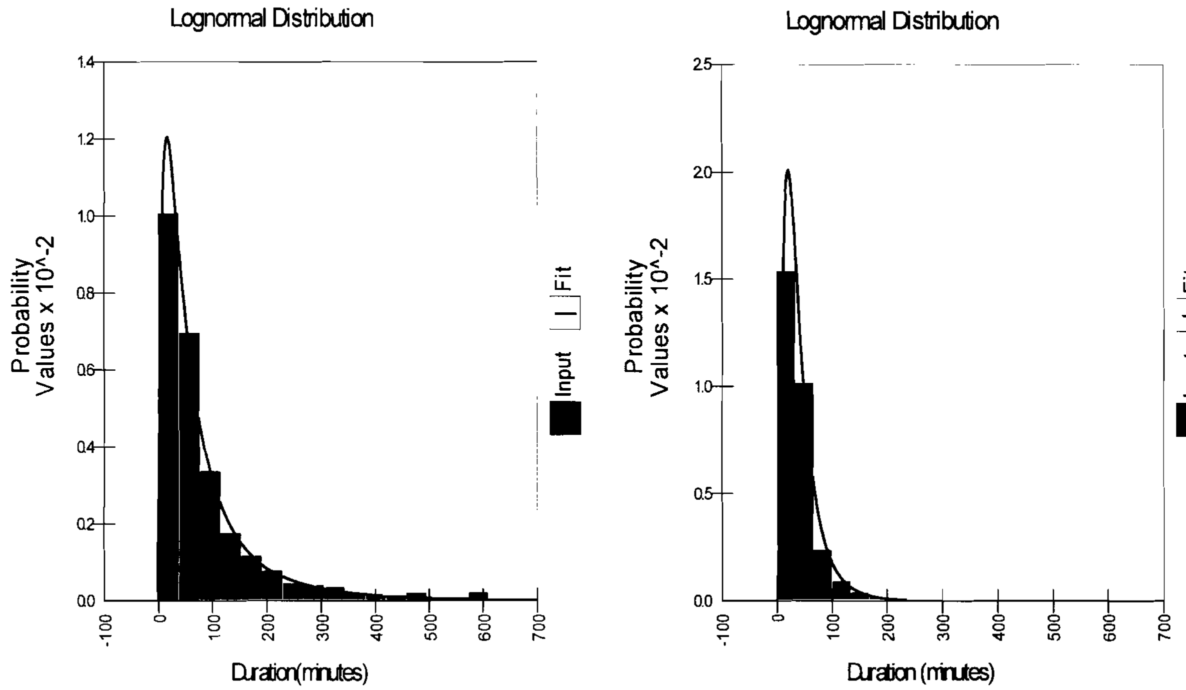
TABLE 6 Distribution of Accidents by Time of Day in Harris County

	Mean	Negative Binomial Distribution $P(x)=\binom{s+x-1}{x}p^s(1-p)^x$		Poisson Distribution $P(x)=\frac{\lambda^x}{x!}e^{-\lambda}$
		s	p	λ
Midnight—12:59 a.m.	1.5638	2	0.5612	1.5638
1:00 a.m.—1:59 a.m.	1.4250	2	0.5839	1.4250
2:00 a.m.—2:59 a.m.	2.0542	2	0.4933	2.0542
3:00 a.m.—3:59 a.m.	0.9881	2	0.6693	0.9881
4:00 a.m.—4:59 a.m.	0.5994	2	0.7694	0.5994
5:00 a.m.—5:59 a.m.	0.8767	2	0.6952	0.8767
6:00 a.m.—6:59 a.m.	2.4021	3	0.5553	2.4021
7:00 a.m.—7:59 a.m.	4.0100	2	0.3328	4.0100
8:00 a.m.—8:59 a.m.	3.6496	3	0.4512	3.6496
9:00 a.m.—9:59 a.m.	2.8186	4	0.5866	2.8186
10:00 a.m.—10:59 a.m.	2.9135	5	0.6318	2.9135
11:00 a.m.—11:59 a.m.	3.5580	6	0.6277	3.5580
Noon—12:59 p.m.	4.2094	6	0.5877	4.2094
1:00 p.m.—1:59 p.m.	4.1997	7	0.6250	4.1997
2:00 p.m.—2:59 p.m.	4.6414	8	0.6328	4.6414
3:00 p.m.—3:59 p.m.	5.0651	8	0.6123	5.0651
4:00 p.m.—4:59 p.m.	5.4731	8	0.5938	5.4731
5:00 p.m.—5:59 p.m.	6.2682	8	0.5607	6.2682
6:00 p.m.—6:59 p.m.	5.1833	8	0.6068	5.1833
7:00 p.m.—7:59 p.m.	3.4904	7	0.6673	3.4904
8:00 p.m.—8:59 p.m.	2.5957	7	0.7295	2.5957
9:00 p.m.—9:59 p.m.	2.6274	6	0.6955	2.6274
10:00 p.m.—10:59 p.m.	2.4776	5	0.6687	2.4776
11:00 p.m.—11:59 p.m.	2.1163	3	0.5864	2.1163

3.3.2 Identify Distribution for Duration

As mentioned in Chapter 2, the accident duration should include four parts: detection time, response time, clearance time, and recovery time. Because it is hard to determine the recovery time, the prepared database contains only the first three parts of the accident duration. Therefore, the following duration-related analyses are all based on these three parts.

The software used for the duration probability analysis was @Risk4.5. Results show that there are three types of distributions that can fit for the provided duration data. They are lognormal distribution (LND), loglogistic distribution, and Pearson5 distribution. By A-D (Anderson-Darling Statistic) and K-S (Kolmogorov-Smirinov Statistic) tests, lognormal distribution is the best one for the duration of non-heavy truck minor accidents, heavy truck minor accidents, and heavy truck major accidents, and it is the third-best one for non-heavy truck major accidents. Figure 7 illustrates lognormal distribution as fitting for heavy truck minor accidents and non-heavy truck major accidents. The distributions for heavy truck major accidents and non-heavy truck minor accidents are similar.



(a) LND for heavy truck minor accidents (b) LND for non-heavy truck major accidents

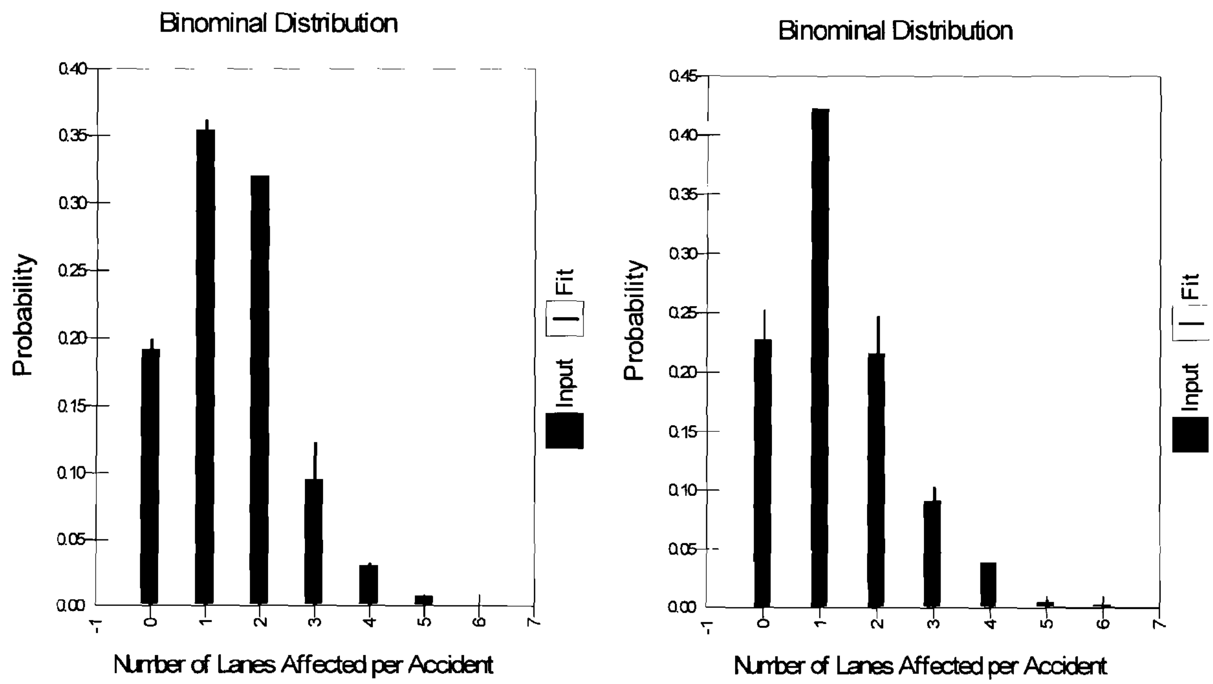
FIGURE 7 Lognormal distribution for accident duration.

3.3.3 Identify Distribution for Lanes Affected

There are two special problems for lanes affected. First, the number of accidents that affected ramps is 6.8 percent of the total accidents. There should be differences among the impact of affected off-ramps, on-ramps, and interchange ramps. However, the database does not provide this type of difference. The second problem is that accidents that only affected shoulders should not have significant impact on congestion and delay, except rubbernecking. Therefore, this analysis for lanes affected concentrated on the main lanes only. Accidents that only affected shoulders are counted as zero main lane affected. It is believed that this will not affect the illustration of updating parameter distribution by the Bayesian approach.

Similar to the accident duration data, the lanes-affected data were also divided into four categories. From @Risk analysis, results show that the binomial distribution and Poisson distribution are the best two for all four categories. Based on the chi-square test, binomial distribution is good for non-heavy truck minor accidents, heavy truck minor accidents, and non-heavy truck major accidents, and Poisson distribution is good for heavy truck major accidents.

Figure 8 illustrates binomial distribution as fitting for non-heavy truck major accidents and heavy truck minor accidents.



(a) Non-heavy truck major accidents (b) Heavy truck minor accidents

FIGURE 8 Probability analyses of lanes affected by accidents.

3.3.4 Estimation of Accidents per Million Vehicles Miles Traveled by Facility and County

The vehicle miles traveled (VMT) are an indicator of the travel levels on the roadway system and are restricted to motor vehicles. VMT is an estimated value of the number of miles traveled by automobiles in a given time period. It is the product of roadway segment length (miles) and roadway segment volume (number of vehicles).

For the estimation of accidents per million VMT, it is necessary to obtain the traffic volume for each highway. Traffic maps for Harris County and Travis County from 1992 through 2000 were obtained from TxDOT's Houston Planning Office and Traffic Map Office in the Austin district.

Based on the traffic volume information from the traffic maps, freeway section length information from maps, and milepoint information from TRF accident files, VMTs were calculated for freeways and freeway sections in Harris County and Travis County.

Then, based on the freeway facility information and the corresponding accident information from TRF accident files and VMTs, accident rates per million VMT for a particular year in Harris County and Travis County were calculated. The calculation is shown in the following equation.

$$Accident/MillionVMT = \frac{Number\ \ of\ \ Accidents}{Traffic\ \ Volume * Roadway\ \ Section\ \ Length} * 10^6 \qquad \text{(Eq 7)}$$

Figure 9 and Table 7 show the overall accident rates per million VMT in Harris County and Travis County. Table 8 and 9 show the accident rates per million VMT by facility types in Harris County and Travis County, respectively.

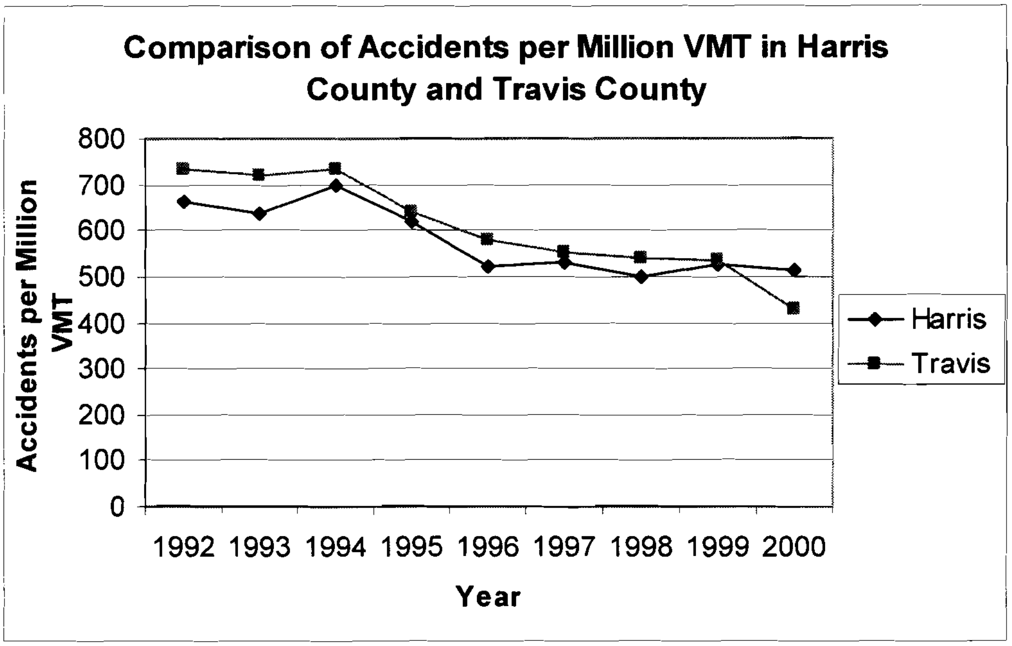


FIGURE 9 Comparison of accidents per million VMT in Harris County and Travis County.

TABLE 7 Accident Rates per Million VMT in Harris County and Travis County

Year	Harris County	Travis County
2000	513.11	429.93
1999	527.05	535.33
1998	501.06	540.93
1997	530.63	553.84
1996	522.47	580.04
1995	619.60	641.52
1994	697.87	732.21
1993	637.80	719.38
1992	663.46	734.12

Figure 9 and Tables 7–9 show that accidents per million VMT in both counties decreased from 1992 to 2000. Travis County had a faster rate of decrease—41.4 percent over the 9-year period. Accidents per million VMT in Harris County also had a decrease of 22.7 percent during this period. During the 9-year boom of vehicles and infrastructure, VMT also had a big increase. However, the total number of accidents had only a moderate decrease. Therefore, both counties experienced a large decrease in accidents per million VMT. This is due to the development of infrastructure, intelligent transportation systems (ITS), incident management, accident preventive measures, and law enforcement.

TABLE 8 Number of Accidents per Million VMT by Facility Type in Harris County

Freeway	2000	1999	1998	1997	1996	1995	1994	1993	1992
Interstate	450.56	438.25	449.27	453.38	442.80	500.79	565.71	517.87	547.42
Urban Freeway	455.13	503.51	455.85	505.71	518.61	620.84	705.26	645.63	668.73
Principal Arterial	769.92	805.73	796.46	920.41	832.09	1178.49	1230.17	1209.82	1123.45
Minor Arterial	596.44	684.18	659.79	598.76	623.42	688.83	1105.90	1183.10	1215.45
Rural Major Collector or Urban Collector	364.01	391.05	702.10	679.71	306.62	516.23	690.61	473.60	615.86
Rural Minor Collector	N/A	N/A	N/A	N/A	N/A	N/A	309.28	508.93	783.85

TABLE 9 Number of Accidents per Million VMT by facility in Travis County

Freeway	2000	1999	1998	1997	1996	1995	1994	1993	1992
Interstate	475.89	580.04	598.66	653.79	594.08	635.75	738.16	721.62	713.45
Urban Freeway	303.69	417.44	422.03	388.66	473.00	532.77	595.45	609.91	685.59
Principal Arterial	613.97	746.03	751.50	861.95	863.08	958.42	842.33	838.12	778.71
Minor Arterial	424.99	437.74	394.34	424.97	412.77	447.77	1042.83	925.93	977.96
Rural Major Collector or Urban Collector	562.66	503.70	474.72	512.04	519.96	580.48	750.26	707.21	735.44
Rural Minor Collector	307.69	313.97	231.35	1014.37	1037.17	512.82	389.48	1139.60	699.30
Local road or street	N/A	N/A	N/A	N/A	N/A	N/A	943.40	1690.23	1997.10

The trend is the same for the accidents based on the facility type. During this 9-year period, the accidents per million VMT decreased on all facility types.

It is noted that the results on the rural minor collector and local road or street have a big difference from year to year. The reason is that there were very few accidents on these roads and these roads are very short. For example, only three accidents happened on rural minor collector facility types with a VMT of 2632.5 in 1993, and only one accident happened on this facility type in 1994. The result for 1994 is only one-third of that in 1993.

There are gaps between the results in 1994 and 1995, and the results in 1997 and 1998. The reason is that the classification of facility types in TRF accident files changed twice in 1995 and 1998, respectively. The same roadway may belong to minor arterial this year, but to principle arterial next year. This made the results less convincing and comparable.

It is shown that the interstate highways, urban freeways, and rural major/urban collectors have fewer accidents per million VMT than the others. Relatively, they are safer than other roadway types, but they have different reasons for safety. Interstate highways and urban freeways have large traffic volumes, fast speeds, and long lengths. The traffic flows are uninterrupted. Therefore, the denominator (million VMT) is large, which makes the results small. On the other hand, rural major/urban collectors in Harris County are interrupted by other traffic, and have small volumes and slow speeds. Therefore, the numerator (number of accidents) is small, which also makes results small.

CHAPTER 4

METHODOLOGY OF UPDATING PARAMETERS OF ACCIDENT PROBABILITY MODELS BASED ON NEW INFORMATION

This chapter provides the methodology of updating parameters of accident probability models based on new information. The accident probability models include the models for accident frequency, duration, and lanes affected. The key technique is to apply the Bayesian approach to update the relevant parameters.

4.1 General Description

When deriving the posterior distribution of a parameter, considerable mathematical simplification can be achieved if the parameter distribution is appropriately chosen. The pair of random variables' distribution and its parameter's distribution are known as the conjugate pair. The commonly used conjugate distribution can be found in Table 8.1 of the probability book by Ang and Tang (1975). For example, if the basic random variables are Poisson, lognormal, and binomial distributions, their corresponding parameter's conjugate distributions are gamma, normal, and beta distributions, respectively. The following will model the updating process of the accident frequency, duration, and number of lanes affected by one accident, respectively, where the conjugate distribution will help to simplify the updating calculations.

4.2 Updating Accident Frequency Based on New Information

As has been identified, the distribution of the accident frequency for the prepared data sets follows Poisson distribution. Let x be the number of accidents happening within a time period t, the accident frequency with the average occurrence of μ can be written as:

$$p_X(x) = \frac{(\mu \cdot t)^x}{x!} e^{-\mu \cdot t}. \qquad \text{(Eq 8)}$$

From the book by Ang and Tang (1975), the distribution μ is a gamma distribution with parameters ν and κ.

$$f_M(\mu) = \frac{\nu(\nu\mu)^{\kappa-1} e^{-\nu\mu}}{\Gamma(\kappa)} \qquad \text{(Eq 9)}$$

The mean and variance of μ are $E(\mu) = \kappa/\nu$ and $Var(\mu) = \kappa/\nu^2$, respectively.

Parameters ν'' and κ'' of the posterior distribution of μ is updated by new information (t, x) in the following form:

$$\nu'' = \nu' + t, \qquad \text{(Eq 10)}$$

$$\kappa'' = \kappa' + x. \qquad \text{(Eq 11)}$$

where ν'': posterior parameter for ν,

ν': prior parameter for ν,

κ'': posterior parameter for κ,

κ': prior parameter for κ,

t: time period (days or hours) for new information, and

x: number of new accidents that happened in time period t.

The posterior mean for parameter μ can be calculated as: $E''(\mu) = \kappa''/\nu''$, which is also the average occurrence of accidents for the case that the basic random variable (now it is the accident frequency) is Poisson distribution.

The above updating process can be further explained by the following example. Based on the experiences and historical data in the past 5 years, the mean occurrence (μ) of accidents is 60 per day with the variance of 1. Suppose the occurrence of accidents in this area is Poisson distribution, with gamma distribution being its parameter μ's conjugate distribution. Therefore, $E'(\mu) = \kappa'/\nu' = 60$ and $Var'(\mu) = \kappa'/\nu'^2 = 1$. So the two parameters can be solved as $\kappa' = 3600$ and $\nu' = 60$, if in the most recent month, the average occurrence of accidents is forty per day. Applying Equation 10 and Equation 11, the posterior parameters become $\kappa'' = 3600 + 40 = 3640$ and $\nu'' = 61$. Hence, the updated average occurrence of accidents is $E''(\mu) = \kappa''/\nu'' = 59.67$ per day. Obviously, the new information changed the historical information somewhat.

4.3 Updating Accident Duration Based on New Information

As tested before, lognormal distribution of accident duration is valid for the test data. Let x be the duration time, then the density function for the duration is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\zeta x} \cdot \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\zeta}\right)^2\right]. \qquad \text{(Eq 12)}$$

By definition, $\ln x$ is normal with the known standard deviation ζ, while λ is the mean of $\ln x$. The conjugate distribution for λ is normal distribution with parameters μ and σ (Ang and Tang 1975).

$$f_\Lambda(\lambda) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{\lambda - \mu}{\sigma}\right)^2\right] \qquad \text{(Eq 13)}$$

The mean and variance of normal distribution are $E(\lambda) = \mu$ and $Var(\lambda) = \sigma^2$, respectively.

The posterior distribution is also a normal distribution with parameters $\dfrac{\mu(\zeta^2/n) + \sigma^2\overline{\ln(x)}}{\zeta^2/n + \sigma^2}$ and $\sqrt{\dfrac{\sigma^2(\zeta^2/n)}{\sigma^2 + (\zeta^2/n)}}$. Therefore, the posterior parameters can be updated by:

$$\mu'' = \frac{\mu'(\zeta^2/n) + \sigma^2\overline{\ln(x)}}{\zeta^2/n + \sigma^2}, \qquad \text{(Eq 14)}$$

$$\sigma'' = \sqrt{\frac{\sigma^2(\zeta^2/n)}{\sigma^2 + (\zeta^2/n)}}. \qquad \text{(Eq 15)}$$

where μ'': posterior parameter for μ,

μ': prior parameter for μ,

σ'': posterior parameter for σ,

n: sample size of new duration data,

ζ: standard deviation of logarithm of new duration data, and

x: duration of new accidents.

The mean value of accident duration $E''(x)$ is calculated as

$$E''(x) = \sqrt{\frac{e^{2\mu''} + \sqrt{e^{4\mu''} + 4e^{2\mu''} \cdot \zeta^2}}{2}}. \qquad \text{(Eq 16)}$$

where $E''(x)$: posterior mean of duration,

μ'': $E''(\lambda)$, posterior mean of the normal distribution for λ, and

ζ: parameter in lognormal distribution function, assumed to be known.

For example, based on the historical information, the prior distribution of parameter λ follows normal (3, 0.05). This means $\mu' = 3$ and $\sigma' = 0.05$. By observation, the recent three accidents lasted 30, 35, and 40 minutes, respectively. The natural logarithms of these durations are 3.40, 3.56, and 3.69, respectively. So, the sample mean $\overline{\ln(x)} = 3.55$ and the sample standard deviation ζ=0.14. From the relationships given in Equations (14) and (15), the parameters of posterior distribution $f''(\lambda)$ are:

$$\mu'' = \frac{3(0.14^2/3) + 3.55(0.05)^2}{(0.14^2/3) + (0.05)^2} = 3.15 \text{ and } \sigma'' = \sqrt{\frac{(0.14^2/3)(0.05)^2}{(0.14^2/3) + (0.05)^2}} = 0.043.$$

Assuming $\zeta = 36$ from historical data, the updated average duration is $E(x)'' = \sqrt{\frac{e^{6.3} + \sqrt{e^{12.6} + 4e^{6.3} \cdot 36^2}}{2}}$ =33 (minutes per accident). The final results modified the historical distribution.

4.4 Updating Lanes Affected by Accidents Based on New Information

Based on the previous distribution identification, binomial distribution is the best distribution for number of main lanes affected under most of the cases. Let x be the average number of main lanes affected in one accident, n be the maximum number of lanes affected in one accident, and θ be the probability that x lanes are affected. Then, the distribution of lanes affected by accident is:

$$p_X(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}. \qquad \text{(Eq 17)}$$

The parameter n is probably the maximum number of lanes. The conjugated distribution is a beta distribution with parameters q and r. (Ang and Tang 1975)

$$f_\Theta(\theta) = \frac{\Gamma(q+r)}{\Gamma(q)\Gamma(r)} \theta^{q-1} (1-\theta)^{r-1} \qquad \text{(Eq 18)}$$

The mean and variance of this beta distribution are $E(\theta) = \frac{q}{q+r}$ and $Var(\theta) = \frac{qr}{(q+r)^2(q+r+1)}$.

With new information (n, x), the parameters of the posterior distribution become:

$$q'' = q' + x, \qquad \text{(Eq 19)}$$

$$r'' = r' + n - x. \qquad \text{(Eq 20)}$$

where q'': posterior parameter for q,

q': prior parameter for q,

r'': posterior parameter for r,

r': prior parameter r,

n: maximum number of main lanes affected by one accident, and

x: average number of main lanes affected by one accident.

The average number of lanes affected is calculated by:

$$E''(x) = n \cdot \theta'' \qquad \text{(Eq 21)}$$

where $E''(x)$ is the posterior mean of number of lanes affected in one accident, while θ'' is the posterior mean of the beta distribution for θ.

For example, based on historical experiences, the number of lanes affected follows the binomial distribution. The mean and variance of its parameter θ are 0.2 and 0.001, respectively. According to the formula for mean and variance of beta distribution, it is easy to solve $q' = 31.8$ and $r' = 127.2$. By recent observation, the average number of lanes affected by accident is 2 and the total number of main lanes is 6. From Equation 19 and Equation 20, $q'' = 31.8 + 2 = 33.8$ and $r'' = 127.2 + 6 - 2 = 131.2$. Hence, the updated mean is $E''(\theta) = 33.8/(33.8 + 131.2) = 0.205$. Therefore, the number of lanes affected per accident is $E(x)'' = 6 * 0.205 = 1.23$.

To better understand and compare the previous modeling process for updating frequency, duration, and lanes affected analysis, accident variables' distributions and parameter conjugate distributions, together with the mean, variance, and posterior statistics are listed in Table 10. For more detailed information about conjugate distributions and posterior statistics, readers may refer to probability books (e.g., *Optimal Statistical Decisions*, DeGroot 1970).

TABLE 10 Identified Accident Variable Distributions and Conjugate Distribution of Parameters

	FREQUENCY	DURATION	LANES AFFECTED
Variable Distribution	$p_X(x) = \frac{(\mu \cdot t)^x}{x!} e^{-\mu \cdot t}$	$f_X(x) = \frac{1}{\sqrt{2\pi}\zeta x} \cdot \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\zeta}\right)^2\right]$	$p_X(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$
Conjugate Distribution	$f_{\mathrm{M}}(\mu) = \frac{\nu(\nu\mu)^{\kappa-1} e^{-\nu\mu}}{\Gamma(\kappa)}$	$f_{\Lambda}(\lambda) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{\lambda-\mu}{\sigma}\right)^2\right]$	$f_{\Theta}(\theta) = \frac{\Gamma(q+r)}{\Gamma(q)\Gamma(r)} \theta^{q-1}(1-\theta)^{r-1}$
Parameter Mean	$E(\mu) = \kappa/\nu$	$E(\lambda) = \mu$	$E(\theta) = \frac{q}{q+r}$
Parameter Variance	$Var(\mu) = \kappa/\nu^2$	$Var(\lambda) = \sigma^2$	$Var(\theta) = \frac{qr}{(q+r)^2(q+r+1)}$
Posterior Statistics	$\nu'' = \nu' + t$ $\kappa'' = \kappa' + x$	$\mu'' = \frac{\mu'(\zeta^2/n) + \sigma^2 \overline{\ln(x)}}{\zeta^2/n + \sigma^2}$ $\sigma'' = \sqrt{\frac{\sigma^2(\zeta^2/n)}{\sigma^2 + (\zeta^2/n)}}$	$q'' = q' + x$ $r'' = r' + n - x$

Note: Conjugate distributions and posterior statistics were adopted from Table 8.1 (Ang and Tang 1975)

CHAPTER 5

PRACTICE OF UPDATING PARAMETERS OF ACCIDENT PROBABILITY MODELS BASED ON NEW INFORMATION

This chapter illustrates the practice of updating accident frequency, duration, and lanes affected based on new information. The test beds are Harris County and the Houston area.

In order to apply the methodology in Chapter 4 in practice, the entire updating process has been implemented in Microsoft Excel. As long as the new information is entered, the posterior distribution of parameters and the accident variables after updating will be calculated automatically. It is very convenient for engineers to use.

The updating practice contains two scenarios. The first scenario is a onetime updating based on the information from Houston TranStar's most-recent data to update the historical parameter distribution. The second scenario is a multiple-time updating using "pseudo new information" for several months to continuously update the historical parameter distributions. Before the two scenarios were exercised, the prior distribution was generated from historical data.

5.1 Generating Prior Distribution

For scenario one, the prior distribution for frequency, duration, and lanes affected came from the Houston TranStar database from January 2000 through January 2002. For scenario two, the prior distribution for accident frequency came from the Texas Department of Transportation System Accident Data File from January 1992 through June 2000; prior distribution for duration and lanes affected came from the same source as in scenario one. This means that the frequency analysis for scenario two focused on Harris County, while the other analyses were for the Houston area.

The historical data were analyzed in Chapter 3. The results showed that accident frequency follows a Poisson distribution, accident duration follows a lognormal distribution, and lanes affected follow a binomial distribution. The parameters for these three distributions were calculated for later updating.

5.2 Onetime Updating Based on Recent New Information

In scenario one, new accident information was captured from the Houston TranStar Web site, where the accident information on Houston freeways was published dynamically on the Internet (http://traffic.tamu.edu/incidents). Accident data were recorded for 5 weeks (April 11–May 2, 2003, and June 2–16, 2003). Accident frequency, duration, and lanes affected were retrieved.

Table 11 lists the updating results. It can be seen from the table that prior frequency 73.66 per day was updated to 57.53 per day based on the new information (200 accidents in 4 days.) The average duration per accident was updated from 36.08 minutes to 36.079 minutes with the new information being 36.034 minutes. Here, there are few changes between prior and posterior because the new information is close to the prior one. For lanes affected, the prior lanes affected were 1.23 lanes per accident, which was updated to 1.21 lanes per accident with an average of new information of 0.67 lanes affected.

TABLE 11 Parameters and Sample Variables Updating Summary for Scenario One

	PARAMETER θ			SAMPLE VARIABLE x		
		prior	posterior	prior	new data	posterior
Frequency	ν	1.8294	5.8294	73.66 (accidents/day)	200 accidents in 4 days	57.43 (accidents/day)
	κ	134.75	334.75			
Duration	μ	3.2464	3.2463	36.08 (minutes /accident)	36.03 (minutes /accident)	36.079 (minutes /accident)
	σ	0.008	0.0079			
Lanes affected	q	39.902	40.568	1.23 (lanes /accident)	0.67 (lanes /accident)	1.21 (lanes /accident)
	r	154.5	159.83			

5.3 Multiple-Time Updating Based on Pseudo New Information

In scenario two, the accident frequency for 6 months (July 2000 through December 2000), and the duration and lanes affected for 7 months (July 2002 through January 2003) were used to update the prior distribution of historical parameters six and seven times (i.e., month by month), respectively. Because this information is not virtually new, it is a kind of "pseudo new information."

5.3.1 Updating Frequency by Day of Week

As stated earlier, the prior distribution of frequency for scenario two came from January 1992 through June 2000 in Harris County. At first, the pseudo new information was applied to the prior distribution by using Equations 10 and 11. Figure 10 illustrates the progress of the updating process. An example is Sunday, located on the top left of Figure 10. The historical prior mean for the past 8 1/2 years was about 57.7 accidents per day. In July 2000, it was observed that the average occurrence of accidents (pseudo new information) was about 61.6 per day. By applying Equation 10, Equation 11, and other formulas, the posterior occurrence of accidents was updated to 60.2 per day. The same process proceeded with more pseudo new information. Until the end of December 2000, the posterior occurrence of accidents after six updates was 62.6 per day. Please note that from July 2000 to December 2000, the posterior values always followed the direction demonstrated by the pseudo new data, though did not take exactly that value. This magical effect is what the Bayesian approach was supposed to be.

Similar processes proceeded to the other 6 days of the week. It can be seen from Figure 10 that whether the pseudo new data kept going down (e.g., Monday, Thursday), or varied up and down (e.g., Tuesday, Wednesday), the posterior could trace the trend and incorporate the pseudo new information by making its own modifications. Furthermore, this type of modification is slight, though in many cases the variances of pseudo new data are large. This is reasonable because the posterior value contains more information, including both historical and new information. Relatively, there is much less information from the new data than from the historical data. The new data should not take much higher weights in the updating process.

In the bottom right of Figure 10, there is a plot illustrating the prior and posterior accidents by day of week after the 6-month updating. The curve for the posterior between Monday and Friday is below the prior, which implies that the accidents on the weekdays decreased during the last half of 2000. However, accidents might increase on the weekends during the same period. This reminds us to pay more attention to seeking countermeasures for the accident increase on Sunday and Saturday. Also refer to Table 12 for the detailed updating process and results for accident frequency by day of week.

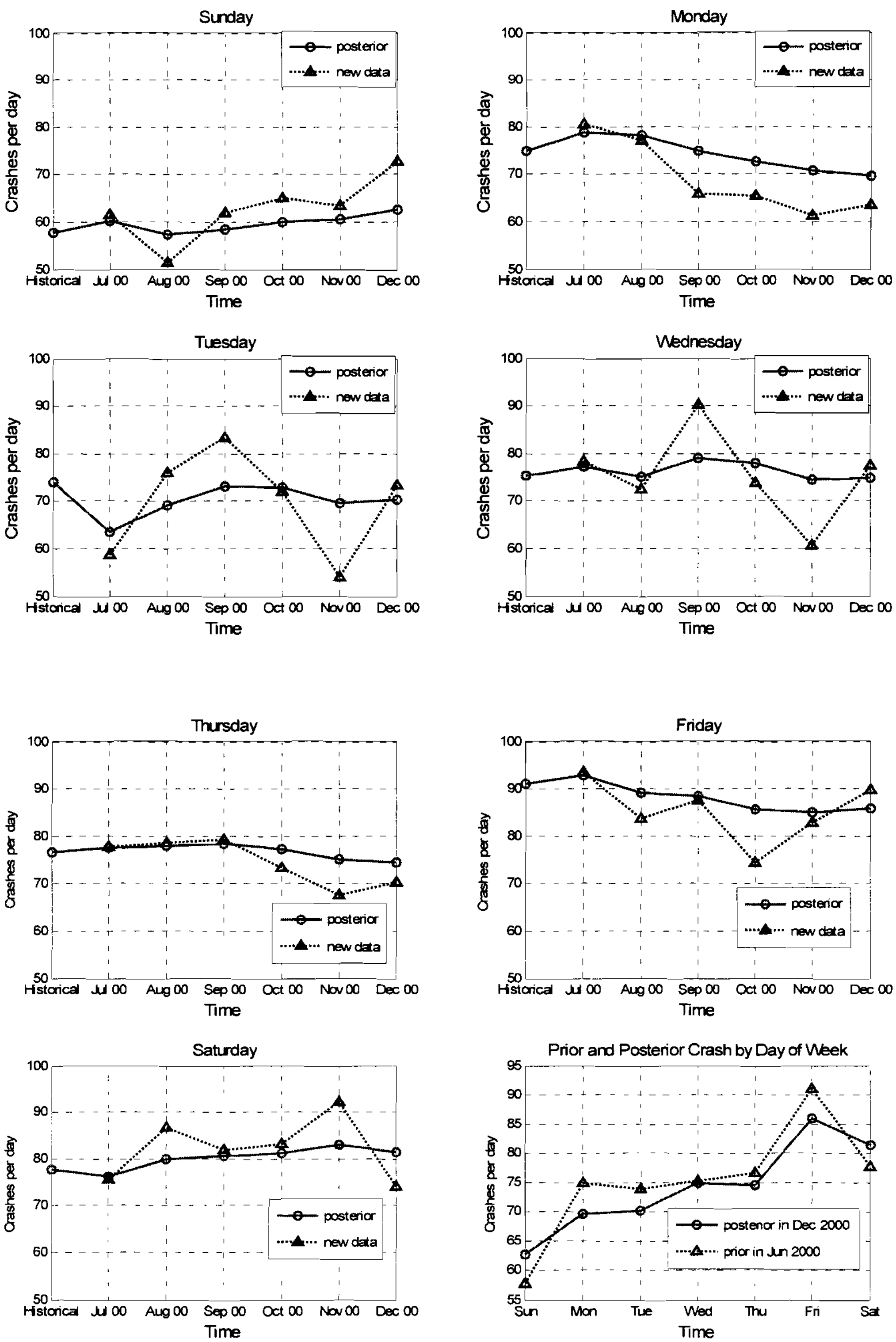


FIGURE 10 Updating parameters for accident frequency by day of week based on new data.

TABLE 12 Comparison of Prior and Posterior Means of Accident Frequency by Day of Week

	Monday			Tuesday			Wednesday		
	prior mean	new data(/day)	posterior mean	prior mean	new data(/day)	posterior mean	prior mean	new data(/day)	posterior mean
Jul 2000	74.842	80.600	78.833	73.939	58.750	63.529	75.243	78.500	77.364
Aug 2000	78.833	77.000	78.179	63.529	76.000	69.283	77.364	72.400	75.137
Sep 2000	78.179	66.000	74.977	69.283	83.500	73.116	75.137	90.250	79.129
Oct 2000	74.977	65.400	72.608	73.116	72.000	72.835	79.129	73.750	78.005
Nov 2000	72.608	61.250	70.732	72.835	54.000	69.674	78.005	60.600	74.400
Dec 2000	70.732	63.500	69.706	69.674	73.250	70.188	74.400	77.500	74.841

	Thursday			Friday			Saturday		
	prior mean	new data(/day)	posterior mean	prior mean	new data(/day)	posterior mean	prior mean	new data(/day)	posterior mean
Jul 2000	76.678	77.750	77.450	91.126	93.500	92.813	77.661	75.600	76.204
Aug 2000	77.450	78.600	77.995	92.813	83.750	89.048	76.204	86.750	80.014
Sep 2000	77.995	79.250	78.340	89.048	87.600	88.553	80.014	82.000	80.632
Oct 2000	78.340	73.250	77.243	88.553	74.500	85.536	80.632	83.250	81.153
Nov 2000	77.243	67.600	75.196	85.536	82.750	85.043	81.153	92.250	82.997
Dec 2000	75.196	70.250	74.478	85.043	89.800	85.904	82.997	74.000	81.450

	Sunday		
	prior mean	new data(/day)	posterior mean
Jul 2000	57.685	61.600	60.235
Aug 2000	60.235	51.500	57.243
Sep 2000	57.243	62.000	58.457
Oct 2000	58.457	65.000	60.039
Nov 2000	60.039	63.500	60.600
Dec 2000	60.600	72.600	62.622

Data Source: TxDOT
Test Bed: Harris County

5.3.2 Updating Frequency by Time of Day

The updating process for frequency by time of day followed the same procedure as for frequency by day of week. Figure 11 plots part of the updating results for frequency by time of day. The bottom right of Figure 11 shows the general picture of the prior and posterior accident by time of day for Harris County. From this figure, it is seen that the first seven plots correspond to different time periods including peak hour (e.g., 8:00 a.m.–8:59 a.m., 6:00 p.m.–6:59 p.m.) and nonpeak hour (e.g. 3:00 a.m.–3:59 a.m., 5:00 a.m.–5:59 a.m.). Still, no matter how the pseudo new information data varied, the posterior could always capture the trends.

From the time of day analysis, the accident frequency has two peaks. One is during 7:00 a.m. to 9:00 a.m., and the other one is from 3:00 p.m. to 7:00 p.m. Both peaks happen during the high-traffic volume time periods. It seems that the accidents are proportional to traffic volumes. The nadir of the accident frequency comes during 4:00 a.m. to 5:00 a.m. when the traffic volume is low (see the last plot of Figure 11). During low accident periods, such as 3:00 a.m. to 3:59 a.m., 5:00 a.m. to 5:59 a.m., and 9:00 p.m. to 9:59 p.m., the new accident data are very stable and very close to the prior frequencies, so the posterior frequencies are keeping straight lines. During the high accident frequency periods, such as 8:00 a.m. to 8:59 a.m., 4:00 p.m. to 4:59 p.m., and 6:00 p.m. to 6:59 p.m., the new accident data have big variances, so the posterior frequencies also vibrate. After Bayesian updating, the results show that the accident frequency has a tendency to decrease during the relatively high-frequency period (7:00 a.m. to 10:00 p.m.), while increasing slightly during the relatively low-frequency period (11:00 p.m. to 7:00 a.m.). See Table 13 for the detailed updating process and results for accident frequency by time of day.

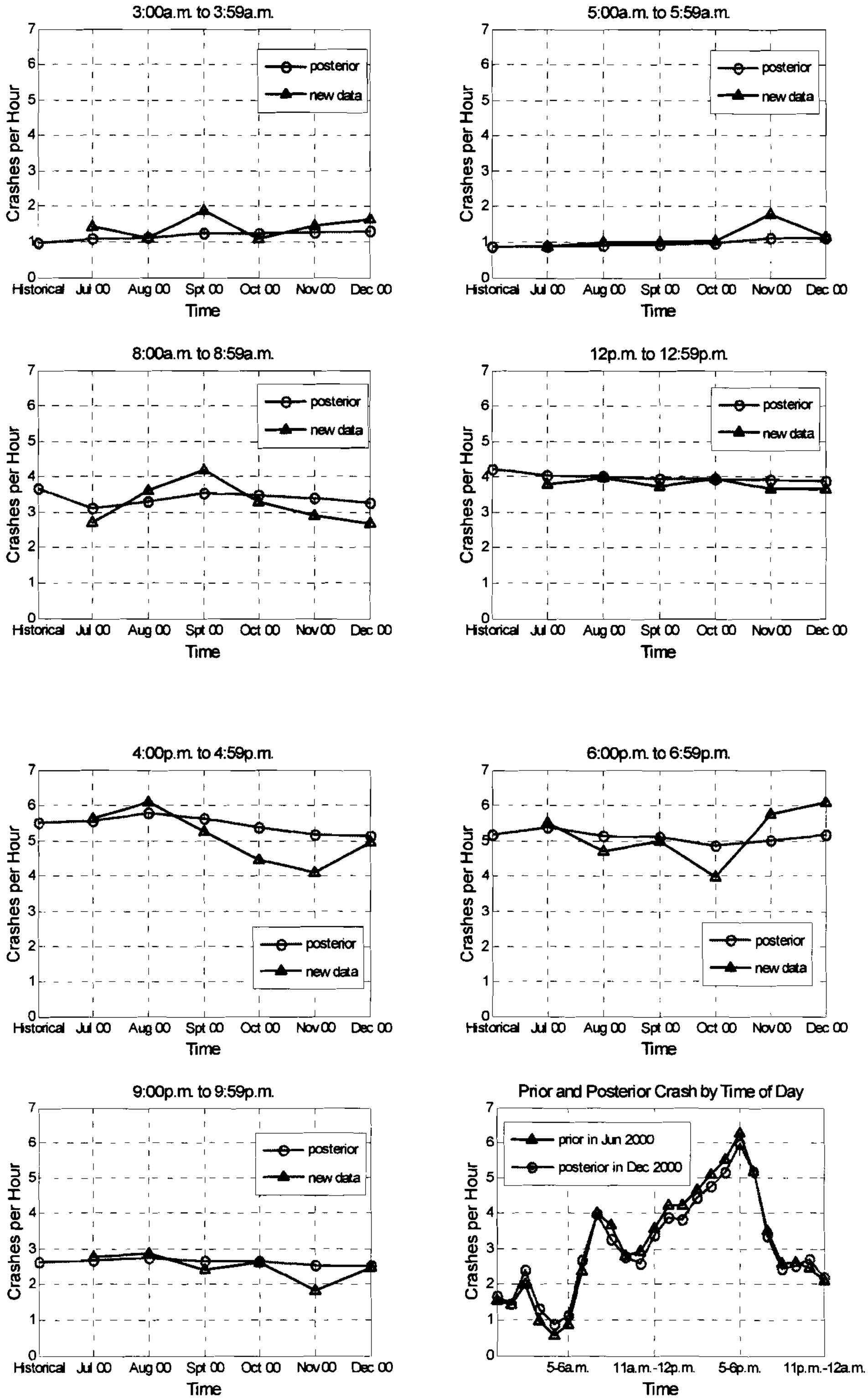


FIGURE 11 Updating parameters for accident frequency by time of day based on new data.

TABLE 13 Comparison of Prior and Posterior Means of Crash Frequency by Time of Day for Harris County

	Midnight—12:59 a.m.			1:00 a.m.—1:59 a.m.			2:00 a.m.—2:59 a.m.		
	prior	new data	posterior	prior	new data	posterior	prior	new data	posterior
Jul 2000	1.555	1.516	1.544	1.420	1.613	1.475	2.016	2.516	2.112
Aug 2000	1.544	1.645	1.566	1.475	0.871	1.342	2.112	2.290	2.141
Sep 2000	1.566	1.767	1.601	1.342	1.467	1.364	2.141	3.000	2.257
Oct 2000	1.601	2.258	1.700	1.364	1.871	1.442	2.257	2.194	2.249
Nov 2000	1.700	1.500	1.675	1.442	1.467	1.445	2.249	3.133	2.343
Dec 2000	1.675	1.645	1.671	1.445	1.677	1.472	2.343	3.032	2.411

	3:00 a.m.—3:59 a.m.			4:00 a.m.—4:59 a.m.			5:00 a.m.—5:59 a.m.		
	prior	new data	posterior	prior	new data	posterior	prior	new data	posterior
Jul 2000	0.961	1.419	1.103	0.580	0.903	0.734	0.861	0.871	0.867
Aug 2000	1.103	1.129	1.109	0.734	0.516	0.664	0.867	1.000	0.918
Sep 2000	1.109	1.900	1.256	0.664	1.000	0.744	0.918	1.000	0.940
Oct 2000	1.256	1.097	1.230	0.744	0.871	0.769	0.940	1.032	0.961
Nov 2000	1.230	1.467	1.262	0.769	1.067	0.817	0.961	1.800	1.107
Dec 2000	1.262	1.645	1.309	0.817	1.161	0.866	1.107	1.161	1.115

	6:00 a.m.—6:59 a.m.			7:00 a.m.—7:59 a.m.			8:00 a.m.—8:59 a.m.		
	prior	new data	posterior	prior	new data	posterior	prior	new data	posterior
Jul 2000	2.384	2.000	2.194	4.014	3.194	3.638	3.676	2.710	3.122
Aug 2000	2.194	2.548	2.311	3.638	4.129	3.792	3.122	3.613	3.301
Sep 2000	2.311	3.500	2.600	3.792	4.733	4.012	3.301	4.200	3.535
Oct 2000	2.600	3.226	2.725	4.012	4.226	4.053	3.535	3.290	3.483
Nov 2000	2.725	3.033	2.776	4.053	4.267	4.087	3.483	2.900	3.384
Dec 2000	2.776	2.097	2.678	4.087	3.258	3.971	3.384	2.677	3.278

	9:00 a.m.—9:59 a.m.			10:00 a.m.—10:59 a.m.			11:00 a.m.—11:59 a.m.		
	prior	new data	posterior	prior	new data	posterior	prior	new data	posterior
Jul 2000	2.822	2.710	2.768	2.936	2.645	2.755	3.572	3.323	3.455
Aug 2000	2.768	3.194	2.906	2.755	2.516	2.663	3.455	3.806	3.567
Sep 2000	2.906	2.600	2.833	2.663	2.533	2.628	3.567	3.967	3.662
Oct 2000	2.833	2.645	2.796	2.628	2.194	2.533	3.662	2.645	3.462
Nov 2000	2.796	2.733	2.786	2.533	2.667	2.556	3.462	2.800	3.357
Dec 2000	2.786	2.806	2.789	2.556	2.806	2.595	3.357	3.677	3.402

TABLE 13 Comparison of Prior and Posterior Means of Crash Frequency by Time of Day for Harris County (Cont'd)

	Midnight—12:59a.m.			1:00a.m.—1:59a.m.			2:00a.m.—2:59a.m.		
	prior	new data	posterior	prior	new data	posterior	prior	new data	posterior
Jul 2000	1.555	1.516	1.544	1.420	1.613	1.475	2.016	2.516	2.112
Aug 2000	1.544	1.645	1.566	1.475	0.871	1.342	2.112	2.290	2.141
Sep 2000	1.566	1.767	1.601	1.342	1.467	1.364	2.141	3.000	2.257
Oct 2000	1.601	2.258	1.700	1.364	1.871	1.442	2.257	2.194	2.249
Nov 2000	1.700	1.500	1.675	1.442	1.467	1.445	2.249	3.133	2.343
Dec 2000	1.675	1.645	1.671	1.445	1.677	1.472	2.343	3.032	2.411

	3:00 a.m.—3:59 a.m.			4:00 a.m.—4:59 a.m.			5:00 a.m.—5:59 a.m.		
	prior	new data	posterior	prior	new data	posterior	prior	new data	posterior
Jul 2000	0.961	1.419	1.103	0.580	0.903	0.734	0.861	0.871	0.867
Aug 2000	1.103	1.129	1.109	0.734	0.516	0.664	0.867	1.000	0.918
Sep 2000	1.109	1.900	1.256	0.664	1.000	0.744	0.918	1.000	0.940
Oct 2000	1.256	1.097	1.230	0.744	0.871	0.769	0.940	1.032	0.961
Nov 2000	1.230	1.467	1.262	0.769	1.067	0.817	0.961	1.800	1.107
Dec 2000	1.262	1.645	1.309	0.817	1.161	0.866	1.107	1.161	1.115

	6:00 a.m.—6:59 a.m.			7:00 a.m.—7:59 a.m.			8:00 a.m.—8:59 a.m.		
	prior	new data	posterior	prior	new data	posterior	prior	new data	posterior
Jul 2000	2.384	2.000	2.194	4.014	3.194	3.638	3.676	2.710	3.122
Aug 2000	2.194	2.548	2.311	3.638	4.129	3.792	3.122	3.613	3.301
Sep 2000	2.311	3.500	2.600	3.792	4.733	4.012	3.301	4.200	3.535
Oct 2000	2.600	3.226	2.725	4.012	4.226	4.053	3.535	3.290	3.483
Nov 2000	2.725	3.033	2.776	4.053	4.267	4.087	3.483	2.900	3.384
Dec 2000	2.776	2.097	2.678	4.087	3.258	3.971	3.384	2.677	3.278

	9:00 a.m.—9:59 a.m.			10:00 a.m.—10:59 a.m.			11:00 a.m.—11:59 a.m.		
	prior	new data	posterior	prior	new data	posterior	prior	new data	posterior
Jul 2000	2.822	2.710	2.768	2.936	2.645	2.755	3.572	3.323	3.455
Aug 2000	2.768	3.194	2.906	2.755	2.516	2.663	3.455	3.806	3.567
Sep 2000	2.906	2.600	2.833	2.663	2.533	2.628	3.567	3.967	3.662
Oct 2000	2.833	2.645	2.796	2.628	2.194	2.533	3.662	2.645	3.462
Nov 2000	2.796	2.733	2.786	2.533	2.667	2.556	3.462	2.800	3.357
Dec 2000	2.786	2.806	2.789	2.556	2.806	2.595	3.357	3.677	3.402

5.3.3 Updating Duration

The prior information for the duration analysis came from Houston TranStar and the study area is the Houston freeway system. The updating results for the duration are illustrated in Figure 12, where the updating for the four categories is separated. Compared with situations of frequency updating, the posterior of duration varies slightly even though it still captures the trend of the new data. Only when the new data had a relatively large variance—as in the case of heavy truck minor accident—the posterior duration changed more obviously than the others. The reason is that the distribution of duration is lognormal, while its parameter is normal-normal, which is different from the frequency' distribution. There are weights associated with the variance of historical data and the new data. If the historical data vary slightly, they may get a higher weight. On the other hand, if the new data get a higher variance, a smaller weight would be assigned to the new data.

From July 2002 through January 2003, the heavy truck related accidents (the right two plots) decreased. However, for non heavy trucks, the average duration remained almost the same after 7 months elapsed. The updating process of parameters of a normal-normal distribution pair and the updated results are listed in Table 14.

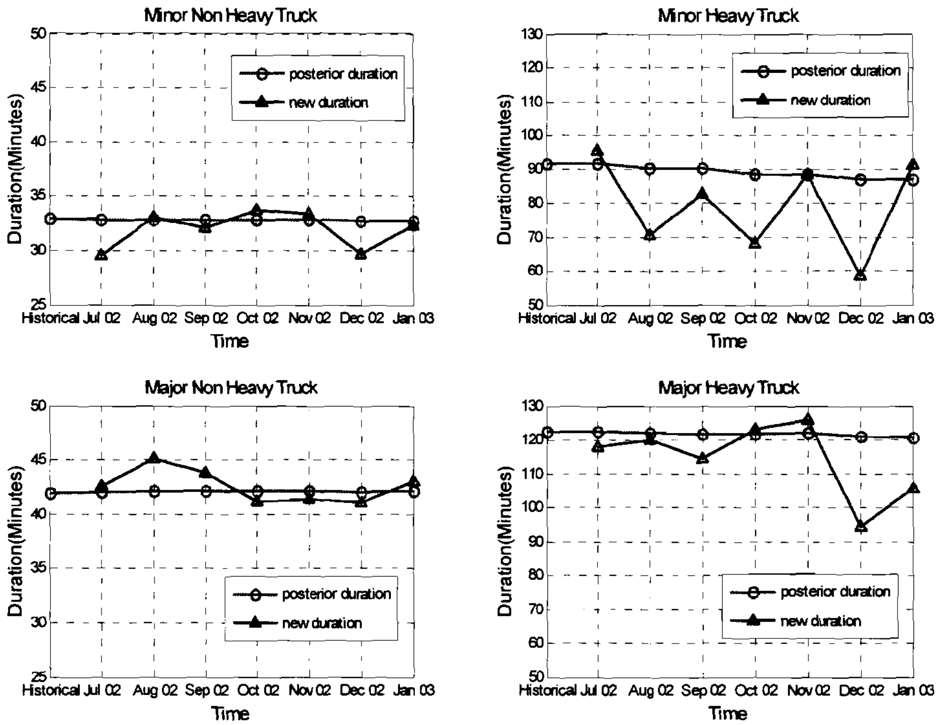


FIGURE 12 Updating accident duration based on new data.

TABLE 14 Comparison of Prior and Posterior Means of Parameter μ in Accident Duration Estimation

Comparison of Prior and Posterior Variables for Duration of Non Heavy Truck Minor Accidents				
	Prior Mean	New Mean	Posterior Mean	Posterior Duration(minutes)
Jul 2002	3.110535326	2.940534193	3.104602594	32.85691581
Aug 2002	3.104602594	3.115367555	3.105083165	32.86717555
Sep 2002	3.105083165	3.068922556	3.103813264	32.84007356
Oct 2002	3.103813264	3.143461358	3.105339755	32.87265526
Nov 2002	3.105339755	3.127241928	3.105779679	32.88205301
Dec 2002	3.105779679	2.946518167	3.101071985	32.78167036
Jan 2003	3.101071985	3.079194606	3.100762926	32.77509444

Comparison of Prior and Posterior Variables for Duration of Heavy Truck Minor Accidents				
	Prior Mean	New Mean	Posterior Mean	Posterior Duration(minutes)
Jul 2002	3.976225382	4.04761343	3.981527745	91.72624453
Aug 2002	3.981527745	3.533531432	3.954688096	90.25700328
Sep 2002	3.954688096	3.811597948	3.951198369	90.06816962
Oct 2002	3.951198369	3.464148685	3.92014907	88.40986494
Nov 2002	3.92014907	3.923476108	3.920206183	88.41287969
Dec 2002	3.920206183	3.200922929	3.893012199	86.99196269
Jan 2003	3.893012199	3.972864049	3.893431948	87.01367562

Comparison of Prior and Posterior Variables for Duration of Non Heavy Truck Major Accidents				
	Prior Mean	New Mean	Posterior Mean	Posterior Duration(minutes)
Jul 2002	3.473733422	3.494517279	3.474516192	42.00426608
Aug 2002	3.474516192	3.573838381	3.478958529	42.13694745
Sep 2002	3.478958529	3.533655202	3.480908273	42.19535915
Oct 2002	3.480908273	3.447042241	3.479739637	42.16033534
Nov 2002	3.479739637	3.452474449	3.478587213	42.12583563
Dec 2002	3.478587213	3.443854571	3.477558094	42.09505931
Jan 2003	3.477558094	3.507629088	3.477990144	42.10797631

Comparison of Prior and Posterior Variables for Duration of Heavy Truck Major Accidents				
	Prior Mean	New Mean	Posterior Mean	Posterior Duration(minutes)
Jul 2002	4.339028956	4.278110637	4.337517316	122.3253394
Aug 2002	4.337517316	4.306032935	4.336727319	122.265288
Sep 2002	4.336727319	4.227559607	4.330966398	121.8285263
Oct 2002	4.330966398	4.349186017	4.331335828	121.8564738
Nov 2002	4.331335828	4.385668673	4.33428121	122.0795899
Dec 2002	4.33428121	3.907617967	4.321779106	121.1361673
Jan 2003	4.321779106	4.097395459	4.315391381	120.6577836

Data Source: Houston TranStar; Test Bed: Houston, TX

5.3.4 Updating Lanes Affected

Data for updating lanes affected are from the same source as for duration. The study area is also the Houston freeway system. Figure 13 shows the lanes updating results. Generally, there is not much change in the posterior values for lanes affected. Only for heavy truck major accidents is the posterior visibly decreased. The new data for heavy truck accidents are not very stable, probably because of the small sample size of heavy truck accidents. The updating process of parameters of binomial-beta distribution pair and the updated results are listed in Table 15.

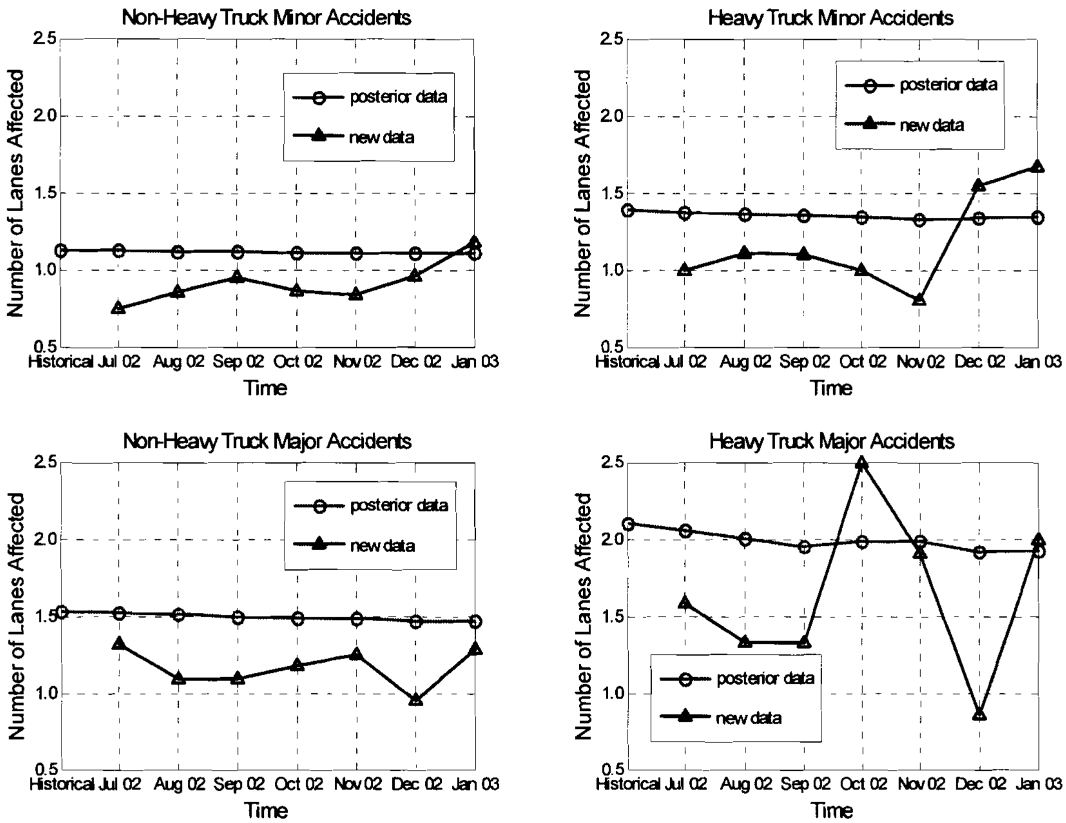


FIGURE 13 Updating lanes affected based on new data.

TABLE 15 Comparison of Prior and Posterior Means of Parameter θ for Number of Lanes Affected Estimation

	Prior θ	New θ	Posterior θ	Posterior lanes affected(lanes)
Jul 2002	0.188135138	0.125203252	0.18713773	1.122826381
Aug 2002	0.18713773	0.141951838	0.186432752	1.11859651
Sep 2002	0.186432752	0.158950617	0.186010569	1.116063417
Oct 2002	0.186010569	0.144402036	0.185381047	1.112286284
Nov 2002	0.185381047	0.139583333	0.184698472	1.108190829
Dec 2002	0.184698472	0.159848485	0.184333543	1.106001256
Jan 2003	0.184333543	0.196078431	0.184503523	1.10702114

Comparison of Prior and Posterior Variables for Lanes Affected Heavy Truck Minor Accident

	Prior θ	New θ	Posterior θ	Posterior lanes affected(lanes)
Jul 2002	0.231604629	0.166666667	0.229393666	1.376361998
Aug 2002	0.229393666	0.185185185	0.227938046	1.367628278
Sep 2002	0.227938046	0.183333333	0.226516196	1.359097176
Oct 2002	0.226516196	0.166666667	0.224667327	1.348003964
Nov 2002	0.224667327	0.133333333	0.221930392	1.331582349
Dec 2002	0.221930392	0.257575758	0.222967472	1.337804829
Jan 2003	0.222967472	0.277777778	0.224517059	1.347102356

Comparison of Prior and Posterior Variables for Lanes Affected Non Heavy Truck Major Accident

	Prior θ	New θ	Posterior θ	Posterior lanes affected(lanes)
Jul 2002	0.254448022	0.220750552	0.25351975	1.521118503
Aug 2002	0.25351975	0.182598039	0.25161843	1.509710582
Sep 2002	0.25161843	0.182336182	0.249809556	1.498857335
Oct 2002	0.249809556	0.196551724	0.24845444	1.490726638
Nov 2002	0.24845444	0.208333333	0.247458911	1.484753465
Dec 2002	0.247458911	0.158854167	0.245313585	1.47188151
Jan 2003	0.245313585	0.213888889	0.244570707	1.467424243

Comparison of Prior and Posterior Variables for Lanes Affected Heavy Truck Major Accident

	Prior θ	New θ	Posterior θ	Posterior lanes affected(lanes)
Jul 2002	0.349933135	0.263888889	0.342863943	2.057183658
Aug 2002	0.342863943	0.222222222	0.333704798	2.002228788
Sep 2002	0.333704798	0.222222222	0.325838247	1.955029484
Oct 2002	0.325838247	0.416666667	0.331824938	1.990949631
Nov 2002	0.331824938	0.318181818	0.330981298	1.985887788
Dec 2002	0.330981298	0.142857143	0.320025836	1.920155017
Jan 2003	0.320025836	0.333333333	0.320758155	1.92454893

Source of Data to Yield the Parameters in this Table: Houston TranStar; Test Bed: Houston, TX

The Bayesian approach can make accident prediction more accurate and is able to incorporate new information. Latest accident prediction is always needed in order to assign the proper number of patrol and police officers. It is also the prerequisite for the air quality analysis.

CHAPTER 6

EMISSION ESTIMATION WITH CONGESTION DUE TO ACCIDENTS

Chapter 6 calculates how many additional emissions the nonrecurring congestion due to accidents can generate. The emissions caused by congestion due to accidents are calculated by two methods. One is the microscopic calculation based on the idle emission rates. The other one is the macroscopic estimation through MOBILE6. The first method can calculate how many extra emissions will be generated by the congestion caused by accidents. The second method will estimate the overall effects of congestion on the whole area due to accidents.

6.1 Nonattainment Counties and Near-Nonattainment Counties in Texas

In response to the Clean Air Act of 1970, the U.S. Environmental Protection Agency (EPA) established National Ambient Air Quality Standards (NAAQS) for various pollutants known as "criteria" pollutants that adversely affect human health and welfare. Three major transportation-related criteria pollutants include:

- Ozone (O_3) and its precursors, volatile organic compounds (VOC), and oxides of nitrogen (NO_X),

- Particulate Matter (PM), and
- Carbon Monoxide (CO).

Significant progress has been made in reducing criteria pollutant emissions from motor vehicles and improving air quality since the 1970s, even as vehicle travel has increased rapidly. However, challenges still remain. In 1997, EPA issued revised standards for ozone and particulate matter that reflect improved understanding of the health effects of these pollutants. Based on monitored data, in 2000, approximately 121 million people in the United States resided in counties that did not meet the air quality standards for at least one NAAQS pollutant. (FHWA 2002)

NAAQS are set by EPA to protect public health and welfare. An area is a nonattainment area if it exceeds the concentration level for the specified form of the standard and evaluation time frame. Table 16 shows the primary air quality standards for transportation-related pollutants.

TABLE 16 Primary Air Quality Standards for Transportation-related Pollutants

Pollutant	Type of Average	Concentration
CO	8-hour 1-hour	9 ppm (10mg/m^3) 35 ppm (40 mg/m^3)
O_3	8-hour 1-hour	0.08 ppm (157 $\mu g/m^3$) 0.125 ppm (235 $\mu g/m^3$)
$PM_{2.5}$	Annual 24-hour	15 $\mu g/m^3$ 65 $\mu g/m^3$
PM_{10}	Annual 24-hour	50 $\mu g/m^3$ 150 $\mu g/m^3$

ppm = parts per million
mg/m^3 = milligrams per meter cubed
$\mu g/m^3$ = micrograms per meter cubed

(Source: Transportation Air Quality Selected Facts and Figures 2002)

Ozone in the upper atmosphere filters the sun's ultraviolet radiation and protects us from harmful effects such as skin cancer. However, at ground level where we can breathe it, ozone levels above the national standard may aggravate lung and respiratory disorders and may cause injury to plants and damage to certain man-made materials. There is also concern that long-term exposure to levels above the standard may cause lung damage in healthy individuals. High levels of ozone—including levels above the 0.125 parts per million standard—can have serious human health effects.

Texas has sixteen nonattainment counties for ozone, one nonattainment county (El Paso) for PM_{10}, and one nonattainment county (El Paso) for CO. Texas also has twenty-one near-nonattainment counties for ozone whose ozone levels are close to the federal standard. Tables 17 and 18 show the nonattainment counties and near-nonattainment counties in Texas.

TABLE 17 Nonattainment Areas in Texas

Area	Counties
Houston/Galveston	Brazoria, Chambers, Fort Bend, Galveston, Harris, Liberty, Montgomery, Waller
Beaumont/ Port Arthur	Hardin, Jefferson, Orange
Dallas/Fort Worth	Collin, Dallas, Denton, Tarrant
El Paso	El Paso

(Source: TNRCC and EPA Web Pages Sep 20, 2001)

TABLE 18 Near-Nonattainment Counties in Texas

Bastrop	Bexar	Caldwell	Comal
Ellis	Gregg	Guadalupe	Harrison
Hays	Johnson	Kaufman	Nueces
Parker	Rockwall	San Patricio	Smith
Travis	Upshur	Victoria	Williamson
Wilson			

(Source: TNRCC and EPA Web Pages Sep 20, 2001)

The ground-level ozone is not directly emitted, but is formed by the reaction of NO_x and VOC in the presence of sunlight. Strategies to reduce ozone must focus on reducing emissions of ozone-forming pollutants. The microscopic calculation concentrates only on the two ozone-forming pollutants NO_x and VOC, while the macroscopic evaluation includes three major transportation-related pollutants.

Because the mobile source emissions are a major source of the emissions, controlling the emission of NO_x and VOC from automobiles is urgent for Texas nonattainment areas. Idling emissions caused by nonrecurring congestion due to accidents are a major part of the idle emissions, so it is necessary to evaluate the idling emissions by nonrecurring accidents in order to develop effective emission control strategies.

6.2 Idling Emission Rates

Idling emission rates are a special case for the vehicle emission modeling. Idling emission rates can be calculated directly from the exhaust emission factors from MOBILE5 and MOBILE6 run with an input speed of 2.5 miles per hour. The predicted exhaust emission factors will be in units of "gram per mile." To convert these "grams per mile" emissions to the idling emission factors in the unit of "grams per hour," multiply the emission factors by the speed (2.5 mph). The speed of 2.5 mph is chosen because it contains a conservative measure of the idling time in the MOBILE model.

*Idling emission rate (g/hr) = Emissions factors at 2.5 mph (g/mi) * Average speed (2.5 mph)* (Eq 22)

Idling emission rates, in some cases, may include effects from an engine start. However, accounting for the effects of engine starts on exhaust emissions during the vehicle operation (such as at intersections) is very difficult. In this research, the calculation of idling emissions does not include any effects from engine starts based on the assumption that the vehicles blocked on freeways do not stop engines.

In order to better evaluate the idling emission caused by the nonrecurring congestion, two types of the emission calculation were carried out in this research. One is the microscopic calculation, and the other is the macroscopic evaluation. In the microscopic calculation, @Risk-embedded Microsoft Excel worksheets were built for each accident. The idling emission was calculated for each minute after the occurrence of accidents. The accumulated idling emission was the sum of every minute's idling emission. In the macroscopic evaluation, MOBILE6 was used to evaluate the comprehensive air quality effects by nonrecurring congestion due to accidents. The default values in input files such as speed vehicle miles traveled (VMT) and VMT by hour were modified because of the nonrecurring congestion. The MOBILE6 was run by new input values and results were shown in the output file.

6.3 Microscopic Calculation

In order to calculate the idling emissions, the information about the traffic condition and accidents should be identified first, including accident location, occurred time, cleared time, total number of lanes on freeway, the number of lanes affected (blocked), the traffic volume where and when the accident happens, and so on. Based on this information, the incoming volume, clearing volume, idling status, as well as the caused emission, can be estimated.

In order to simulate the whole process minute by minute, a Microsoft Excel worksheet was developed with the format as shown in Table 19. Table 19 illustrates the microscopic estimation of idling emissions due to an accident on August 11, 2003, at I-45 northbound at Allen Parkway in Houston.

Some of items in the worksheet in Table 19, like the roadway section, time occurred, time cleared, lane affected, etc., are self-explanatory. However, there are other items that need further description and are explained in the next several subsections.

TABLE 19 Microscopic Emission Calculation Worksheet

Roadway Section=	I-45 North Northbound at ALLEN PARKWAY						
Time occured=	15	52	**Duration=**	9 minutes			
Time Cleared=	16	1				VOC Lbs	1.64
Number of Lanes =	4		**Freeway capactity available=**	0.58		NOx Lbs	0.61
Lanes Affected=	1						
ADT=	**221,778**	**VOC**		**NOX**			
Direction Split=	0.5	**9.82 2.5 mph**		**3.65 2.5 mph**			
Hour Fraction=	0.0660042	24.55 **grams/hr**		9.125 **grams/hr**			

	Minutes	Direction Split	Hour Fraction	Incoming Vol	Open Lanes	Clearing Vol	New Idlers	Cum Idlers	Idle Hours		VOC (Lbs)	NOx (Lbs)
3:52 p.m.	1	0.5	0.066004	122	3	97	25	25	0.21	0.21	0.01	0.00
3:53 p.m.	2	0.5	0.066004	122	3	97	25	51	0.63	0.63	0.03	0.01
3:54 p.m.	3	0.5	0.066004	122	3	97	25	76	1.06	1.06	0.06	0.02
3:55 p.m.	4	0.5	0.066004	122	3	97	25	101	1.48	1.48	0.08	0.03
3:56 p.m.	5	0.5	0.066004	122	3	97	25	127	1.90	1.90	0.10	0.04
3:57 p.m.	6	0.5	0.066004	122	3	97	25	152	2.32	2.32	0.13	0.05
3:58 p.m.	7	0.5	0.066004	122	3	97	25	177	2.74	2.74	0.15	0.06
3:59 p.m.	8	0.5	0.066004	122	3	97	25	203	3.17	3.17	0.17	0.06
4:00 p.m.	9	0.5	0.070818	131	3	97	34	237	3.66	3.66	0.20	0.07
4:01 p.m.	10	0.5	0.070818	131	4	167	-36	201	3.65	3.65	0.20	0.07
4:02 p.m.	11	0.5	0.070818	131	4	167	-36	166	3.06	3.06	0.17	0.06
4:03 p.m.	12	0.5	0.070818	131	4	167	-36	130	2.46	2.46	0.13	0.05
4:04 p.m.	13	0.5	0.070818	131	4	167	-36	94	1.87	1.87	0.10	0.04
4:05 p.m.	14	0.5	0.070818	131	4	167	-36	59	1.28	1.28	0.07	0.03
4:06 p.m.	15	0.5	0.070818	131	4	167	-36	23	0.68	0.68	0.04	0.01
4:07 p.m.	16	0.5	0.070818	131	4	167	-36	-13	0.09	0.09	0.00	0.00
4:08 p.m.	17	0.5	0.070818	131	4	167	-36	-48	-0.51	-	-	-
4:09 p.m.	18	0.5	0.070818	131	4	167	-36	-84	-1.10	-	-	-
4:10 p.m.	19	0.5	0.070818	131	4	167	-36	-120	-1.70	-	-	-
4:11 p.m.	20	0.5	0.070818	131	4	167	-36	-155	-2.29	-	-	-
4:12 p.m.	21	0.5	0.070818	131	4	167	-36	-191	-2.89	-	-	-

6.3.1 Explanation of the First Row of the Worksheet

Several items are listed on the left part of the first row of the worksheet in Table 19. The item "Roadway section" is the name of the road section. Here it is "I-45 northbound at ALLEN PARKWAY." "Time Occurred" and "Time Cleared" provide the occurring time and clearing time of the accidents. "Number of Lanes" and "Lanes Affected" are self-explanatory. Average daily traffic (ADT) will be discussed later in this section. "Direction Split" and "Hour Fraction" are determined based on the local traffic survey and also will be discussed later. In Table 19, "Direction Split" was chosen as 0.5 and the "Hour Fraction" for the time period was chosen as 0.0660042 for the analyzed road.

In the middle of the first row, there are some other items. "Duration" allows the user to input the accident duration, which is reported by Transportation Management Center (for real time simulation) or predicted by the proposed approaches in Chapter 4 and Chapter 5 (for simulation of future emission estimations). "Freeway capacity available" will be discussed in the next part of this section. The numbers under VOC and NO_x are the idle emission rates, which will be discussed later.

The small matrix on the right part of the worksheet lists the total estimated extra emissions in lbs. Next, the calculations of several important items are discussed.

Average daily traffic

In the worksheet, ADT in 2003 is predicted based on the AADT (annual average daily traffic) at this accident location from 1992 to 2001 by Microsoft Excel. The historical AADT can be obtained from the historical traffic database.

Idling emission rate for VOC and NO_x

For the idle emission rate, Equation 22 is used.

In this formula, the "Emissions at 2.5 mph" are weighted arithmetic average emissions of all kinds of vehicles at 2.5 mph.

$$\text{Emission at 2.5 mph (g/mi)} = \sum_{i=1}^{28}(E_i p_i) \qquad \text{(Eq 23)}$$

where, E_i : the emission factor (g/mi) of vehicle type i at 2.5 mph. (MOBILE6 divides all vehicles into 28 types; i stands for vehicle type.), and

p_i : the percentage that vehicle type i occupies.

In the emission calculation, 9.82 g/mi is used for VOC and 3.65g/mi is used for NOx. These two values came from the Texas Department of Transportation (TxDOT). Therefore, the emission rates at 2.5mph for VOC and NOx are: 24.55 g/hr and 9.125 g/hr, respectively, in Texas. These factors are already adjusted for fuel, temperature, and hot and cold starts.

Directional Split and Hour Fraction

In this emission calculation, $a_{direction}$ is assumed to be 0.5, which means that the inbound volume is assumed to be the same as the outbound volume. a_{hour} has different values for each one-hour period within one day. The values were calculated from hourly volumes at seven different locations on freeways in Houston in 2000. They are the average percentages of hourly volumes over the daily total traffic volumes. Equation 24 is used to calculate the average hourly fraction. Table 20 shows the results of a_{hour} for different hour periods. The assumption is that these average hourly fractions will be suitable to the traffic volume distribution for the whole Houston area.

$$a_{hour,ith} = \frac{\sum_{j=1}^{7} q_{i,j}}{\sum_{i=1}^{24}\sum_{j=1}^{7} q_{i,j}} \qquad \text{(Eq 24)}$$

where $q_{i,j}$ is the traffic volume at i^{th} hour period at location j.

TABLE 20 Hourly Fractions in Houston Area

Time Period	Hour Fraction	Time Period	Hourly Fraction
Midnight—12:59 a.m.	0.0119	Noon—12:59 p.m.	0.0562
1:00 a.m.—1:59 a.m.	0.0079	1:00 p.m.—1:59 p.m.	0.0577
2:00 a.m.—2:59 a.m.	0.0072	2:00 p.m.—2:59 p.m.	0.0606
3:00 a.m.—3:59 a.m.	0.0060	3:00 p.m.—3:59 p.m.	0.0660
4:00 a.m.—4:59 a.m.	0.0090	4:00 p.m.—4:59 p.m.	0.0708

5:00 a.m.—5:59 a.m.	0.0248	5:00 p.m.—5:59 p.m.	0.0738
6:00 a.m.—6:59 a.m.	0.0522	6:00 p.m.—6:59 p.m.	0.0619
7:00 a.m.—7:59 a.m.	0.0631	7:00 p.m.—7:59 p.m.	0.0478
8:00 a.m.—8:59 a.m.	0.0542	8:00 p.m.—8:59 p.m.	0.0373
9:00 a.m.—9:59 a.m.	0.0495	9:00 p.m.—9:59 p.m.	0.0330
10:00 a.m.—10:59 a.m.	0.0506	10:00 p.m.—10:59 p.m.	0.0265
11:00 a.m.—11:59 a.m.	0.0534	11:00 p.m.—11:59 p.m.	0.0184

Freeway Capacity Available

According to the study conducted by Sullivan (1997), about 40 percent of the accidents occurred in travel lanes, 10 percent on the median shoulder, and the rest on the right shoulder. In 60 percent of accidents, the drivers were able to move their vehicles onto the shoulder. During congested periods, an accident can induce 500 to 1,000 vehicle hours of delay. The impact of accidents on traffic flow can be considerable because the presence of police cars, freeway service patrols, tow trucks, ambulances, and fire trucks reduces the freeway capacity. The freeway capacity is reduced based on the total number of freeway lanes and the number of lanes blocked because of the accident.

Table 21 shows the overall freeway capacity available based on the total number of lanes and the number of lanes blocked. For example, if an accident happens on a freeway that has two main lanes and blocks one main lane, instead of 50 percent reduction in capacity due to the one lane blocked, a 65 percent reduction is observed. An extra 15 percent is clearly due to motorists rubbernecking as they pass the accident site.

TABLE 21 Percentage of Freeway Section Capacity Available under Incident Conditions (Ozbay and Kachroo 1999) (Original reference: Owen and Urbanek 1978)

Number of Freeway Lanes in Each Direction	Shoulder Disablement	Shoulder Accident	Lanes Blocked One	Lanes Blocked Two	Lanes Blocked Three
2	0.95	0.81	0.35	0	N/A
3	0.99	0.83	0.49	0.17	0
4	0.99	0.85	0.58	0.25	0.13
5	0.99	0.87	0.65	0.40	0.20
6	0.99	0.89	0.71	0.50	0.25
7	0.99	0.91	0.75	0.57	0.36
8	0.99	0.93	0.78	0.63	0.41

In the worksheet, the right value for the available capacity according to the number of lanes blocked (affected) and the total number of lanes is calculated as 0.58.

6.3.2 Explanation of the Worksheet Columns

Below the first row of the worksheet (Table 19), there are several columns. The first two are time indicators. “Direction Split” and Hour Fraction” are transferred from the first row of the worksheet. “Incoming Vol” and “Clearing Vol” means incoming volume and clearing volume. “Open Lane” is the number of lanes that is not affected by the accident. “New Idlers” and “Cum Idlers” are the newly generated and cumulated number of vehicles that are idling. “Idle Hours” is the idling time in hours. The last two columns are the estimated number of emissions for VOC and NO_X, respectively. Next, the calculations for some of the items are presented.

Incoming Volume (*Incoming VOL*)

In the “incoming volume” column of Table 19, the following formula is used to calculate the incoming volume for each minute.

$$\text{Income Volume} = ADT \times a_{direction} \times a_{hour} / 60 \qquad \text{(Eq 25)}$$

where $a_{direction}$: Direction split parameter, representing the percentage of traffic volume that is in the same traffic direction at the accident location, and

a_{hour}: Hour fraction, representing the percentage of traffic volume that exists during the hour when the accident happens.

Clearing Traffic Volume (*Clearing Vol*)

The clearing traffic volume calculation is based on whether the accident is cleared. After the accident is cleared, it is possible for the clearing volume to reach the freeway's capacity. It follows this formula:

$$\text{Clearing volume by minute} = C_p \times n / 60. \qquad \text{(Eq 26)}$$

where, C_p: The planning capacity for each lane on freeways per hour. Department of Planning, TxDOT Houston uses 2,500 veh/hr in its practice;

n: The number of total lanes on the freeway.

Before the accident is cleared, the clearing volume follows the following:

$$\text{Clearing volume by minute} = n \times C_a / 60. \qquad \text{(Eq 27)}$$

where, C_a: Freeway capacity available after accident occurred.

New Idlers and Cumulative Idlers (*Cum Idlers*)

In the column of "New Idlers," the new idling volumes are the difference between incoming volumes and clearing volumes. Cumulative idlers are the sum of the new idlers in current minute and the cumulative idlers in the last minute.

In order to calculate the idling hours, Equation 28 was used based on the assumption that vehicles are arriving evenly within one minute. Therefore, the average waiting time for the new idlers is 0.5 minute per minute. However, for the cumulative idlers, the waiting time is 1 minute.

$$\text{Idle Hours} = C_{i-1} / 60 + N_i \times 0.5 / 60 \qquad \text{(Eq 28)}$$

where, C_{i-1}: Cumulative idlers at the $(i\text{-}1)^{\text{th}}$ minute, and

N_i: New idlers at the i^{th} minute.

Estimation of Microscope Emission

In the last two columns of the worksheet in Table 19, idling emissions of VOC and NO_x are calculated for each minute as long as the idling hours are positive.

$$\text{Idle Emissions (lbs)} = (\text{Idle Hours} \times f_{idle})/454 \qquad \text{(Eq 29)}$$

where, f_{idle}: Idle emission rates, which are 24.55 g/hr and 9.125 g/hr for VOC and NO_x, respectively.

The total emissions are listed on the right-hand side of the table. They are the sum of the idle emission for each minute.

In this emission example, the total idle emission is 1.64 lbs for VOC, and 0.61 lbs for NO_x. The accident was cleared in 9 minutes. The new idler becomes negative in the tenth minute because the clearing volume is bigger than incoming volume. It took an extra 6 minutes for the freeway to recover to normal conditions. During the recovery period, there were still some vehicles waiting. Therefore, idle emissions exist during the recovery time until the cumulative idlers are cleared.

The worksheet in Table 19 can also be used to calculate idle emission caused by recurring congestion. When incoming volume is bigger than clearing volume, idle emissions occur. The new idlers and cumulative idlers can be calculated and idle emissions can be shown automatically.

6.3.3 Microscopic Emission Estimation in Houston

In order to estimate the microscope emissions in the Houston area, the research team recorded accidents from the Houston TranStar Web site for 3 days (Aug 11, 14, and 22, 2003). The recorded information included location, occurred/verified time, cleared time, description, vehicle type, lanes affected, and even the speed charts on the freeway. Based on the recorded real time accident data, 3 entire days of idle emissions were calculated and simulated. The means of total idle emission of nonrecurring congestion due to accidents on August 11, 2003, for VOC and NOx are about 488.6597 lbs (ranging from 465.2926 lbs to 518.1751 lbs) and 181.6301 lbs (ranging from 172.9448 lbs to 192.6007 lbs). The detailed emission results are shown in Table 22.

TABLE 22 Microscopic Emission Calculation Results

		Mean(lbs)	Minimum(lbs)	Maximum(lbs)
Aug 11, 2003	VOC	488.6597	465.2926	518.1751
	NO_x	181.6301	172.9448	192.6007
Aug 14, 2003	VOC	229.7247	195.8251	268.2263
	NO_x	154.5760	152.0323	157.4581
Aug 22, 2003	VOC	143.6901	137.2534	150.1267
	NO_x	53.40822	51.01578	55.80066

From the results, the idle emission varies a lot. That is because the accident frequency, occurred time, accident sites, number of lanes affected, and traffic conditions, etc. are varying greatly. Many more accidents happened on Aug 22 than on Aug 11, but the idle emissions on Aug 11 were much more than on Aug 22. The reason is that the majority of accidents on Aug 22 only affected shoulders. During the analyzed days, the available freeway capacity was still 99 percent. No idle emissions were generated.

6.4 Macroscopic Estimation

The macroscopic estimation is to estimate the extent to which the nonrecurring congestion because of accidents happened in the entire area. The air quality model used for emission estimation is the EPA required model MOBILE6.

MOBILE6 emission rates include vehicle idling in proportion to normal driving. Because the goal of the research is to evaluate the idle emissions caused by nonrecurring congestion due to accidents, it is necessary to estimate the emissions from idling explicitly.

The impact of nonrecurring congestion on air emissions may be reflected by the mobile source emission-related travel indicators: VMT BY SPEED, VMT BY FACILITY, VMT BY HOUR, VMT FRACTION (VMT MIX), etc., as shown in Figure 14. By changing these inputs, the emission estimation by MOBILE6 will also be changed.

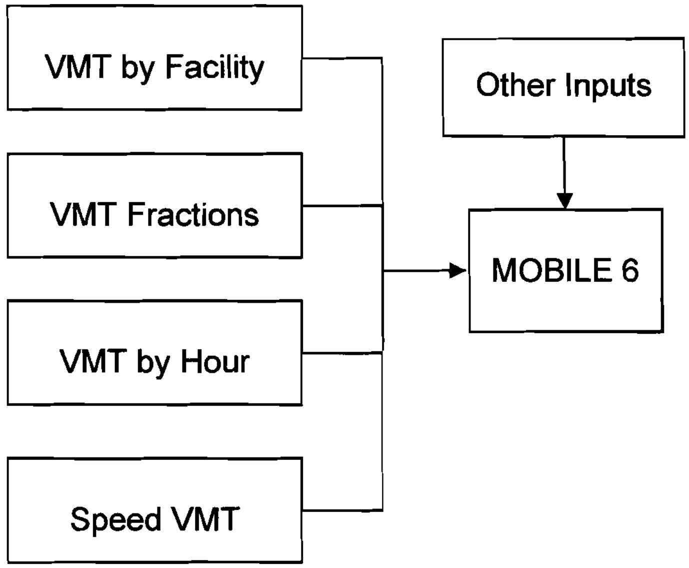


FIGURE 14 Four VMT-related inputs for MOBILE6.

Nonrecurring congestion will result in the detour of part of the vehicles from one facility type (e.g., the freeway) to the other types (e.g., the arterial roads or the local roads). Because this relates to the traffic assignment in the entire area, the change of VMT facility type was not considered in this research. For a similar reason, the fractional VMT was also not considered. In the following subsections, focus is placed on the estimation of the change emission factors due to the change of speed VMT and VMT BY HOUR.

6.4.1 Change of Speed Vehicle Miles Traveled

Figure 15 illustrates the factors that would affect the speed VMT. From Figure 15, it is shown that apart from the traffic volume-related factors, accident duration and number of lanes affected are also two important factors that will affect the speed VMT. As discussed in Chapter 4 and Chapter 5, these two factors—together with the accident frequency—can be frequently updated by using the Bayesian approach.

In the following part of this subsection, the important factors, especially the traffic volume-related factors that appear in Figure 15, are described with the necessary equations listed.

The incoming traffic volume can be simulated based on Equation 24. If the total number of lanes and number of lanes affected are obtained, the total traffic volume blocked can be calculated by Equation 30.

$$\textit{Volume Blocked} = (\textit{Incoming volume} - \textit{Clearing volume}) * \textit{Duration} \qquad \text{(Eq 30)}$$

The congestion length can be surveyed from the field or simulated by programs. In the Houston area, it can be obtained from TranStar's Web site. With this factor, the affected VMT can be calculated by Equation 31.

$$\textit{Affected VMT} = \textit{Volume blocked} * \textit{Congestion length} \qquad \text{(Eq 31)}$$

The total VMT was calculated based on the traffic volume on each road and the length of the road section. VMT hourly split can be obtained by either the local empirical practice or from the MOBILE 6 VMT default. The total hourly VMT is calculated by multiplying VMT hourly split by the total VMT as in Equation 32.

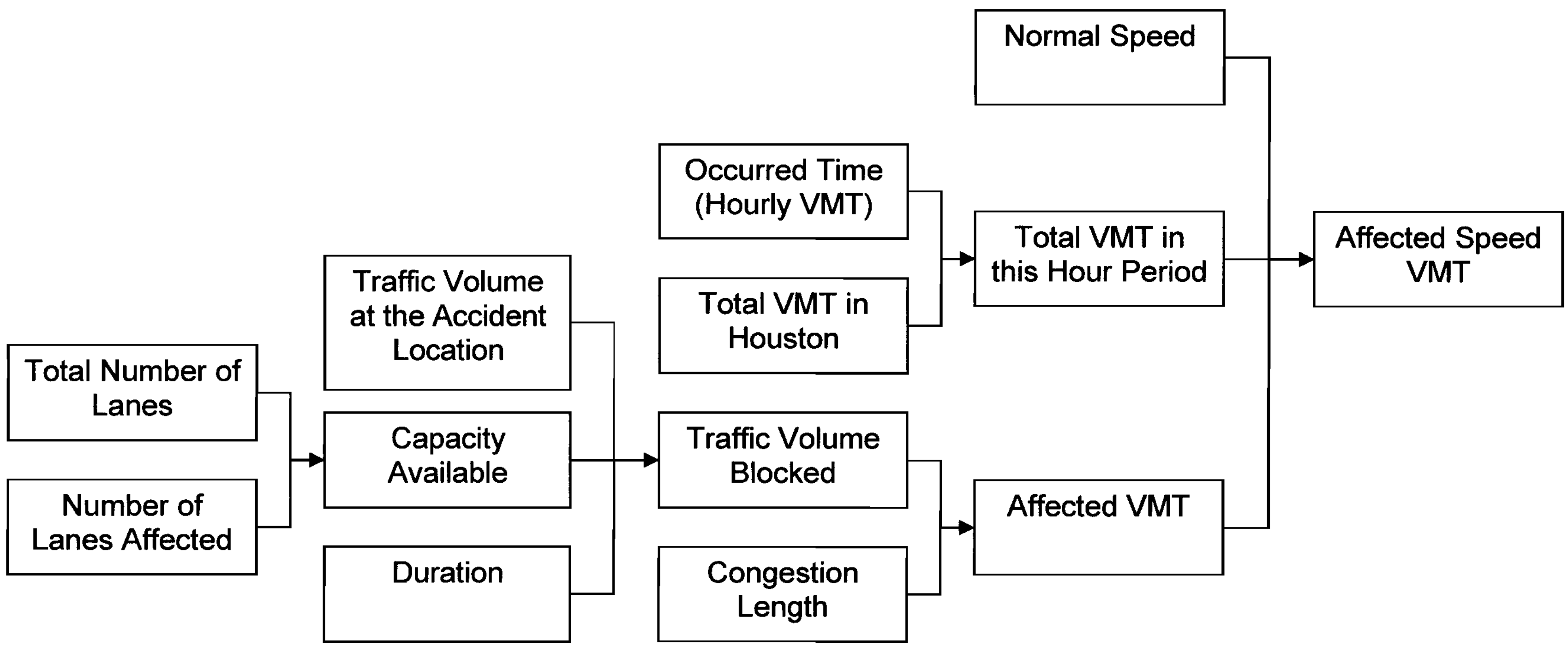


FIGURE 15 The factors that influence speed VMT.

$$Total\ hourly\ VMT = Total\ VMT * VMT\ hourly\ split \quad \text{(Eq 32)}$$

The normal speed in Figure 15 means the average speed before the accident happened. This can be obtained from the dynamic speed charts on TranStar's Web site.

Last, according to the calculating results mentioned before, the influential rate is shown below:

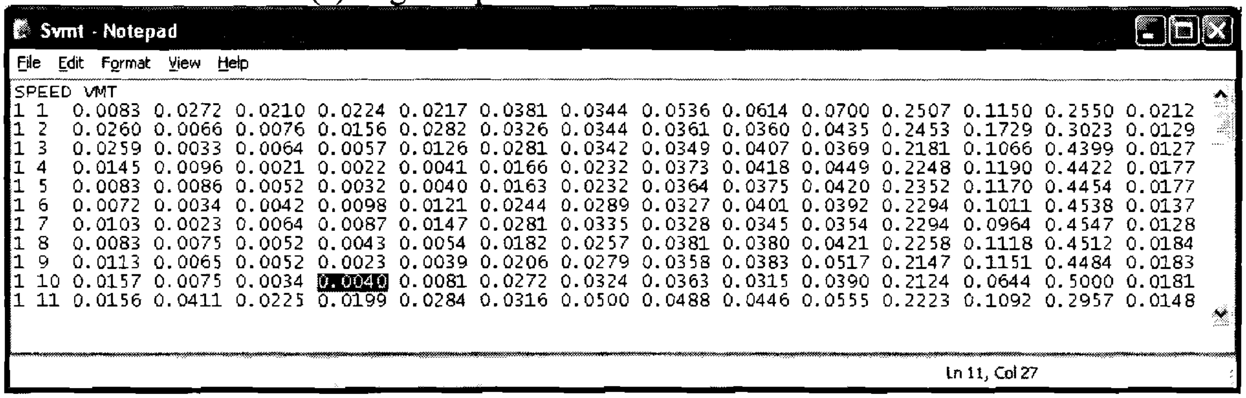
$$VMT\ influential\ rate = affected\ VMT / Total\ hourly\ VMT \quad \text{(Eq 33)}$$

After obtaining the affected speed VMT, the next step is to change the speed distribution in the corresponding MOBILE6 input file (speed VMT). The method is to deduct the VMT influential rate from the column with the speed range that includes the normal speed in the right hour period, and add this rate to the column with 2.5 mph in this same hour period.

As for the example of the case shown in Table 19, the entire process was conducted like this. Because the traffic volume blocked is 237 vehicles and congestion length is 2.3 miles, the affected VMT is 237*2.3=545.1vehmi. By calculation, the total VMT on the Houston freeways monitored by TranStar is calculated as $41{,}157{,}354\ \text{veh} \cdot \text{mi}$. According to the total VMT and the hourly VMT coefficient (0.0636 for 3:00 p.m. to 4:00 p.m.), the total VMT in this hour period is $41{,}157{,}354 * 0.0636 = 2{,}617{,}608\ \text{veh} \cdot \text{mi}$. Therefore, the VMT influential rate is 545.1/2,617,608 = 0.0002. The distribution changing method is to find freeway facility type1 in the first column (because the accident happened on a freeway, not an arterial), right hour of the day in the second column (row 10 represents the 3:00 p.m. to 3:59 p.m. period), and the normal speed on freeway (15 mph obtained from TranStar's Web site), then deduct the influential rate 0.0002 from 0.0042 in the eleventh row and sixth column and add 0.002 to 0.0155 in the eleventh row and third column. Figure 17 shows the changing process.

```
Svmt - Notepad
File  Edit  Format  View  Help
SPEED VMT
1 1  0.0083 0.0272 0.0210 0.0224 0.0217 0.0381 0.0344 0.0536 0.0614 0.0700 0.2507 0.1150 0.2550 0.0212
1 2  0.0260 0.0066 0.0076 0.0156 0.0282 0.0326 0.0344 0.0361 0.0360 0.0435 0.2453 0.1729 0.3023 0.0129
1 3  0.0259 0.0033 0.0064 0.0057 0.0126 0.0281 0.0342 0.0349 0.0407 0.0369 0.2181 0.1066 0.4399 0.0127
1 4  0.0145 0.0096 0.0021 0.0022 0.0041 0.0166 0.0232 0.0373 0.0418 0.0449 0.2248 0.1190 0.4422 0.0177
1 5  0.0083 0.0086 0.0052 0.0032 0.0040 0.0163 0.0232 0.0364 0.0375 0.0420 0.2352 0.1170 0.4454 0.0177
1 6  0.0072 0.0034 0.0042 0.0098 0.0121 0.0244 0.0289 0.0327 0.0401 0.0392 0.2294 0.1011 0.4538 0.0137
1 7  0.0103 0.0023 0.0064 0.0087 0.0147 0.0281 0.0335 0.0328 0.0345 0.0354 0.2294 0.0964 0.4547 0.0128
1 8  0.0083 0.0075 0.0052 0.0043 0.0054 0.0182 0.0257 0.0381 0.0380 0.0421 0.2258 0.1118 0.4512 0.0184
1 9  0.0113 0.0065 0.0052 0.0023 0.0039 0.0206 0.0279 0.0358 0.0383 0.0517 0.2147 0.1151 0.4484 0.0183
1 10 0.0155 0.0075 0.0034 0.0042 0.0081 0.0272 0.0324 0.0363 0.0315 0.0390 0.2124 0.0644 0.5000 0.0181
1 11 0.0156 0.0411 0.0225 0.0199 0.0284 0.0316 0.0500 0.0488 0.0446 0.0555 0.2223 0.1092 0.2957 0.0148
Ln 11, Col 27
```

(a) original speed VMT file



Svmt - Notepad

File Edit Format View Help

```
SPEED VMT
1 1  0.0083 0.0272 0.0210 0.0224 0.0217 0.0381 0.0344 0.0536 0.0614 0.0700 0.2507 0.1150 0.2550 0.0212
1 2  0.0260 0.0066 0.0076 0.0156 0.0282 0.0326 0.0344 0.0361 0.0360 0.0435 0.2453 0.1729 0.3023 0.0129
1 3  0.0259 0.0033 0.0064 0.0057 0.0126 0.0281 0.0342 0.0349 0.0407 0.0369 0.2181 0.1066 0.4399 0.0127
1 4  0.0145 0.0096 0.0021 0.0022 0.0041 0.0166 0.0232 0.0373 0.0418 0.0449 0.2248 0.1190 0.4422 0.0177
1 5  0.0083 0.0086 0.0052 0.0032 0.0040 0.0163 0.0232 0.0364 0.0375 0.0420 0.2352 0.1170 0.4454 0.0177
1 6  0.0072 0.0034 0.0042 0.0098 0.0121 0.0244 0.0289 0.0327 0.0401 0.0392 0.2294 0.1011 0.4538 0.0137
1 7  0.0103 0.0023 0.0064 0.0087 0.0147 0.0281 0.0335 0.0328 0.0345 0.0354 0.2294 0.0964 0.4547 0.0128
1 8  0.0083 0.0075 0.0052 0.0043 0.0054 0.0182 0.0257 0.0381 0.0380 0.0421 0.2258 0.1118 0.4512 0.0184
1 9  0.0113 0.0065 0.0052 0.0023 0.0039 0.0206 0.0279 0.0358 0.0383 0.0517 0.2147 0.1151 0.4484 0.0183
1 10 0.0157 0.0075 0.0034 0.0040 0.0081 0.0272 0.0324 0.0363 0.0315 0.0390 0.2124 0.0644 0.5000 0.0181
1 11 0.0156 0.0411 0.0225 0.0199 0.0284 0.0316 0.0500 0.0488 0.0446 0.0555 0.2223 0.1092 0.2957 0.0148
```

Ln 11, Col 27

(b) changed speed VMT file

FIGURE 16 Changing default value of speed VMT.

6.4.2 Change of VEHICLE MILES TRAVELED BY HOUR

The command VMT BY HOUR permits the user to allocate total VMT among the 24 hours of each day. The values for VMT BY HOUR are independent of the facility type; that is, the VMT fraction covers all facility types. MOBILE6 uses national data for the default distribution of VMT BY HOUR. If an accident happens, and the recovered time is located in a different hour other than its occurring time, the distribution of VMT BY HOUR will be changed accordingly. As shown in Figure 18, VMT BY HOUR is related to three factors: occurring time, total VMT, and affected VMT.

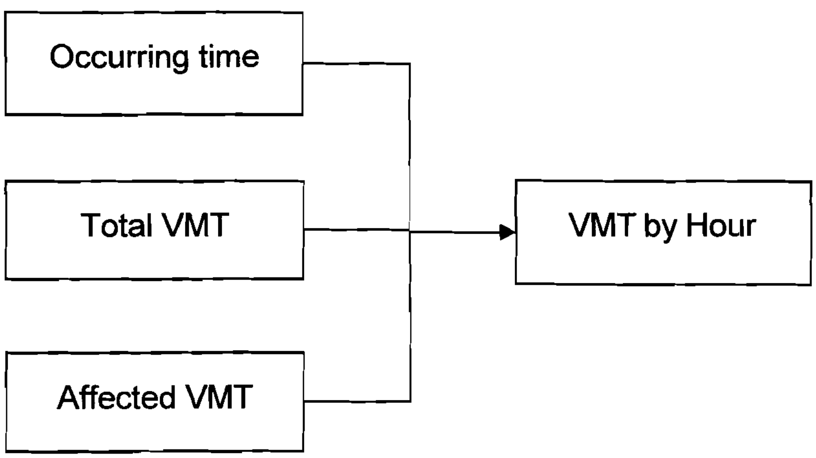


FIGURE 17 Factors related to VMT by hour.

As in the last case, the traffic volume blocked is 203 vehicles before 4 p.m.; congestion length is 2.3 mile. So, the affected VMT is 203*2.3=466.9vehmi. The influential rate for VMT BY HOUR is 466.9/41157354=1.1×10^{-5}. Because the smallest change of distribution in the VMT BY HOUR files is$1.0*10^{-4}$, this influential rate (1.1×10^{-5}) cannot be reflected. Therefore, it is not necessary to change the default values of VMT BY HOUR in this case.

6.4.3 Macroscopic Estimation of Emission Factors

By this method, after changing the default values in VMT BY HOUR and speed VMT, the estimated emission factors can be obtained after running MOBILE6 with the changed input. The results are shown in Table 23.

TABLE 23 Macroscopic Idle Emission Evaluation Results

		Default Emission Factors(g/mi)	Composite Emission Factors(g/mi)
Aug 11, 2003	VOC	1.027	1.030
	CO	11.807	11.817
	NO_x	2.117	2.117
Aug 14, 2003	VOC	1.027	1.028
	CO	11.807	11.810
	NO_x	2.117	2.117
Aug 22, 2003	VOC	1.027	1.027
	CO	11.807	11.809
	NO_x	2.117	2.117

From the results, it appears that the impact of idle emissions on the air is not so significant. However, multiplying the difference of emission rates by total VMT will result in the total emissions increase, which is not small.

In all three days, the emission rates for NO_x did not change. This does not mean that nonrecurring congestion has no impact on emission factors. It only means that the impact of emission factors for NO_x is less than 0.001g/mile per vehicle. Because a small change in emission factors will result in a large change in total emission, the impact of nonrecurring congestion on total emissions cannot be neglected.

CHAPTER 7

CONCLUSION

In this research report, the literature review focused on the state of the art and the state of the practice of nonrecurring congestion, the probability generation of accident frequency, duration, and lane blockage, as well as the estimation of the caused emissions.

The accident data from the Texas Department of Public Safety and Houston TranStar were analyzed. The probability distribution of Texas accident data was obtained and discussed, including the distribution of accident frequency (accident by day of week and by time of day from January 1992 through June 2000); the distribution of accident duration; the distribution of lanes affected; and the estimation of accidents per million vehicle miles traveled by facility and county.

Next, the methodology of updating parameters of accident probability models based on new information was proposed. The famous Bayesian approach was used in the updating process and a Microsoft Excel worksheet embedded with @Risk software was developed. The proposed method used the accident information in Harris County and in the Houston area. Case studies show that this method is implemented easily and the updating process is able to capture up-to-date accident information in the traditional probability analysis.

Finally, the estimations of emissions caused by nonrecurring congestion due to accidents were conducted using two methods. The first method is microscopic and simulation-based, focusing on the extra emissions caused by each individual accident. The second method is macroscopic, trying to estimate the change of emission factors in the entire area when accidents happen. It uses the Environmental Protection Agency approved emission estimation software MOBILE6 by changing some of its input files to meet the changing local environment (the accident). By tracking case studies in the Houston area with the two methods, it is concluded that the individual accidents would cause extra emissions, which vary based on the traffic environment and accident severity. As for the entire Houston area, the nonrecurring congestion due to accidents will result in a slight change of emission factors in this region, and therefore extra emissions are produced.

The researchers recommend further processing and testing of accident data for other cities/counties in Texas. The Microsoft Excel worksheet developed for this project can be used in the implementation stage with necessary improvements made. The testing of emissions impact of nonrecurring congestion due to accidents in other areas is also necessary before more general conclusions can be made.

REFERENCES

Al-Deek, H. Traffic Congestion: Problem and Possible Solutions, available: http://132.170.164.161/courses/ASP/master.asp (Accessed: Dec. 22, 2002)

Ang, A. H.-S. and W. H. Tang. *Probability Concepts in Engineering Planning and Design*, John Wiley & Sons Inc., New York, 1975.

Bolduc, D. and S. Bonin. Bayesian Analysis of Road Accidents: A General Framework for Multinominal Case," Paper number 9802, Université Laval, Ste-Foy, Canada, G1K7P4, 1998.

Congestion Mitigation, (FHWA.dot.gov), available: http://www.fhwa.dot.gov/congestion/congest2.htm. (Accessed: July 29, 2003)

DeGroot, M. H. *Optimal Statistical Decisions*, McGraw-Hill, Inc., 1970.

Garib, A., A. E. Radwan. and H. Al-Deek. Estimating Magnitude and Duration of Incident Delays," *Journal of Transportation Engineering,* Vol. 123, No. 6, pp. 459–466, 1997.

Giuliano, G. Incident Characteristics, Frequency, and Duration on a High Volume Urban Freeway," *Transportation Research – A*, Vol. 23A, No. 5, pp. 387–396, 1989.

Golob, T. F., W. W. Recker, and J. D. Leonard. An Analysis of Truck Involved Freeway Accident," *Accident Analysis and Prevention*, Vol. 19, No.5, pp. 375–395, 1987.

Guide to using @risk: Risk Analysis and Simulation Version 4.5. Palisade Corporation, New York, 2002.

Hadi, M. A., J. Aruldhas, L. F. Chow, and J. A. Wattleworth. Estimating Safety Efforts of Cross-Section Design for Various Highway Types Using Negative Binomial Regression. *Transportation Research Record 1500*, pp. 169–177, 1995.

Hamerslag, R., J. P. Roos, and M. Kwakernaak. Analysis of Accidents in Traffic Situations by Means of Multiproportional Weighted Poisson Model. *Transportation Research Record 847*, pp. 29–36, 1982.

Hensing, D. J. Some Perspectives on Congestion: The Emerging Mission of Modern Transportation Departments. (Iowa State University), available: http://www.ctre.iastate.edu/pubs/semisesq/session3/hensing/ (Accessed: Jan. 25, 2003), 1996.

Heydecker, B. and J. Wu. Using the information in road accident records. *The 19th Summer Annual Meeting of PTRC*, University of Sussex, England, 1991.

Jones, B., L. Janssen, and F. Mannering. Analysis of the Frequency and Duration of Freeway Accidents in Seattle, *Accident Analysis and Prevention*, Vol. 23, No.4, pp. 239–255, 1991.

Karlaftis, M. G. and A. P. Tarko. Heterogeneity Considerations in Accident Modeling, *Accident Analysis and Prevention*, Vol. 30, No.4, pp. 425-433, 1998.

Khattak, A. J., J. L. Schofer, and M.-H. Wang. A Simple Time Sequential Procedure for Predicting Freeway Incident Duration, *IVHS Journal*, Vol. 2(2), pp. 113–138, 1995.

Kottegoda, N. T. and R. Rosso. *Statistics, Probability, and Reliability for Civil and Environmental Engineers,* McGraw-Hill Companies, Inc., 1997.

Lambert, D. Zero-inflated Poisson regression, with an Application to Defects in Manufacturing, *Technometrics*, Vol. 34, No.1, pp. 1–14, 1992.

Lee, J. and F. Mannering. *Analysis of Roadside Accident Frequency and Severity and Roadside Safety Management*, Final Research report, Prepared for Washington State Transportation Commission, 1999.

Lindley, J. A. A Methodology for Quantifying Urban Freeway Congestion, *Transportation Research Record 1132*, Washington D.C., Transportation Research Board, 1987.

Lord D. Issues Related to the Application of Accident Prediction Models for the Computation of Accident Risk on Transportation Networks, *81st Annual Meeting of Transportation Research Board*, CD-ROM, Washington D.C., 2002.

Miaou, S.-P. and H. Lum. Modeling Vehicle Accidents and Highway Geometric Design relationships, *Accident Analysis and Prevention*, Vol. 25, No. 6, pp. 689–709, 1993.

Miaou, S.-P., P. S. Hu, T. Wright, A. K. Rathi, and S. C. Davis. Relationships between Truck Accidents and Highway Geometric Design: A Poisson Regression Approach, *Transportation Research Record 1376*, pp. 10–18, 1992.

Miaou, S.-P. The Relationship between Truck Accidents and Geometric Design of Road Sections: Poisson versus Negative Binominal Regressions, *Accident Analysis and Prevention*, Vol. 26, No. 4, pp. 471–482, 1994.

Nam, D. and F. Mannering. An Exploratory Hazard-based Analysis of Highway Incident Duration, *Transportation Research - A*, Vol.34, pp. 85–102, 2000.

Nembhard, D. A. and M. R. Young. Parametric Empirical Bayes Estimates of Truck Accident Rates, *Journal of Transportation Engineering*, Vol. 121, No. 4, pp. 359–363, 1995.

"Nonattainment and Near-Nonattainment Counties" (cuc.org), available: http://www.cuc.org/tcawg/Resources/nonattainment_counties.htm (Accessed: May 29, 2003)

Office of Operations, FHWA. Congestion Mitigation.

http://www.fhwa.dot.gov/congestion/congest2.htm. Accessed July 29, 2003.

Okamoto, H. and M. Koshi. A Method to Cope with the Random Errors of Observed Accident Rates in Regression Analysis, *Accident Analysis and Prevention*, Vol. 21, No. 4, pp. 317–332, 1989.

Ouyang, Y., V. Shankar, and T. Yamamoto. Modeling the simultaneity in Injury Causation in Multi-vehicle Collisions, *81st Annual Meeting of Transportation Research Board*, CD-ROM, Washington D.C., 2002.

Ozbay, K. and P. Kachroo. *Incident Management in Intelligent Transportation Systems*, Artech House, Boston, MA, 1999.

Schrank, D. and T. Lomax. *The 2003 Annual Urban Mobility Report*, Texas Transportation Institute, College Station, TX, 2003.

Shankar, V., J. Milton, and F. Mannering. Modeling Accident Frequencies as Zero-Altered Probability Process: An Empirical Inquiry, *Accident Analysis and Prevention*, Vol. 29, No. 6, pp. 829–837, 1997.

Skabardonis, A., T. Chira-Chavala, and D. Rydzewski. *The I-880 Field Experiment: Effectiveness of Incident Detection Using Cellular Phones*, California PATH Research Report, UCB-ITS-PRR-98-1, 1998.

Smith, K. and B. L. Smith. *Forecasting the Clearance Time of Freeway Accidents*, Research Project Report for the National ITS Implementation Research Center, 2001.

Sullivan, E. C. New Model for Predicting Incidents and Incident Delay, *Journal of Transportation Engineering*, Vol. 123, pp. 267–275, 1997.

Tang, T., M. Roberts, and C. Ho. "Sensitivity Analysis of Mobile6" FHWA Resource Center-Atlanta, FHWA-RC-Atlanta-03-0007, 2002.

Technical Guidance on the Use of MOBILE6 for Emission Inventory Preparation, US Environmental Protection Agency, 2002.

Teng, H. and Y. Qi. An Investigation of Incident Frequency, Duration and Lanes Blocked for Determining Traffic Delay, *81st Annual Meeting of Transportation Research Board*, CD-ROM, Washington D.C., 2002.

Transportation Air Quality Selected Facts and Figures, US Department of Transportation, Federal Highway Administration, 2002.

Environmental Protection Agency. *User's Guide to Mobile6.1 and Mobile6.2 Mobile Source Emission Factor Model*, EPA420-R-02-028, 2002.