ANALYSIS OF FREEWAY INCIDENT DURATION FOR ATIS APPLICATIONS

Woon Kim, M.S.
Research Assistant, Civil and Environmental Department
University of Maryland at College Park
1173 G. Martin Hall, Dept. of CEE, University of Maryland at College Park, MD 20740 USA
+1-301-405-2638, wkim2@umd.edu

Gang-Len Chang, Ph.D.
Professor, Civil and Environmental Department
University of Maryland at College Park
1173 G. Martin Hall, Dept. of CEE, University of Maryland at College Park, MD 20740 USA
+1-301-405-1953, gang@umd.edu

Steven M. Rochon, P.E.
Office of CHART & ITS Development
Maryland State Highway Administration
7491 Connelley Drive, Hanover, MD 21076 USA
+1-410-787-5880, srochon@sha.state.md.us

This paper presents a methodology for developing a model for estimating and predicting incident duration and identifying variables influencing the incident duration in the state of Maryland. The incident information from years 2003 to 2005 from the Maryland State Highway (MDSHA) database is used for model development, and year 2006 for the model validation. Classification Trees (CT) were used for a preliminary analysis to understand the influence of the variables associated with an incident. Based on the findings from CT, this study employed the Rule-Based Tree Model (RBTM) to develop the primary prediction model. The overall confidence for the estimated model was over 80% with several remarkable findings regarding the associations between factors and incident duration. Although the estimated results from RBTM are quite acceptable, supplemental models along with better quality database are required to improve the prediction accuracy for the duration of a detected incident.

Key Words: incident duration, prediction model, tree model, incident management

INTRODUCTION

Incidents on congested highways, regardless of involving personal fatalities, injuries, or property damages will all cause considerable reduction in capacity (1), followed by heavy congestion and delay, and thus give birth to the enormous socioeconomic loss. With reliably predicted incident duration, responsible agencies can convey the information to travelers by updating the Variable Message Signs, estimate the resulting queue length and assess the need to implement detour operations and any other control strategies. Thus, an effective model to predict incident duration can be a valuable tool to mitigate congestion due to incidents.

The objective of this paper is to present a methodology for estimating and forecasting incident duration and identifying variables influencing the incident duration in the state of Maryland using the MDSHA (the Maryland State Highway) incident database collected from year 2003 to year 2006. Note that in this study, the prediction of incident duration is based on the range of time interval it may fall, such as between 15~30 minutes.

LITERATURE REVIEW

Incident duration has been studied by numerous researchers for several decades with various methodologies. The most representative approaches are [1] Probabilistic Distributions (2), [2] Conditional Probabilities (3), [3] Linear Regression Models (4), [4] Time Sequential Models (5), [5] Decision Trees and Classification Trees (6), and [6] Discrete Choice Models (7). Although there are a variety of existing techniques with acceptable results, they cannot be directly applied to incidents that occurred at any other locations. Each model was developed with different incident data sources and descriptive variables, and thus yields somewhat different results. Therefore, for any target application, it is necessary to develop a new model for different traffic conditions and available data sources.

PRELIMINARY ANALYSIS

Data Description

For this study, highway incident data extracted from CHART-II Database (Coordinated Highways Action Response Team) were used. The CHART II database maintained by MDSHA contains information about the details of each incident occurring in Maryland, including nature of incident, time of detection/occurrence, response time, vehicles involved, types of involved vehicles, number of lanes/shoulders blocked, responded unit, pavement conditions, etc. The model was developed based on the data from year 2003 to year 2005 and validated based on the year 2006 database.

Average Incident Duration

Before starting the model development, the average incident duration under different classifications is computed to investigate its relationships with explanatory variables. The results show that the incident duration increases with the number of heavy vehicles (e.g. tractor-trailers, single unit trucks, or pickup vans) involved or the number of blocked lanes. The incident durations on weekends and at night are generally longer than the durations on weekdays and in the daytime due to the longer response and clearance times.

It is noticeable that incidents occurred in the four major freeways, I-495, I-95, I-695, and I-270, have relatively shorter duration than others. It can be explained by the fact that most of operations centers are located near those four major roads, which results in shortening response time, therefore shortening incident duration.

It is also found that the incident duration exhibits remarkable differences between different incident types. On the basis of the distribution of incident duration frequency by each incident nature, it was found that incidents involving disabled vehicles and property damage are likely to have a shorter duration, while incidents causing personal injuries and fatalities are more likely to have longer duration. These results are consistent with the observations that the distribution of incident durations varies with its nature. Therefore, incident nature emerges as one of the most significant factors for classifying incidents of different durations.

Classification and Regression Tree (CART)

Using CART, which is a type of decision tree technique introduced and popularized by Breiman et al. in 1984 (8), the preliminary model was developed. Overall, CART performed quite well for short (5~20 minutest) or middle (20~70 minutes) ranges of incident duration, especially, for these between 5 to 10 minutes. However, it did not provide satisfactory results for incidents of long duration (e.g., longer than 1 hour). Although the results from CART were not satisfactory enough to be recommended for the incident management, they confirmed that the incident nature is the most significant variable for classification of incident duration.

METHODOLOGY AND ANALYSIS

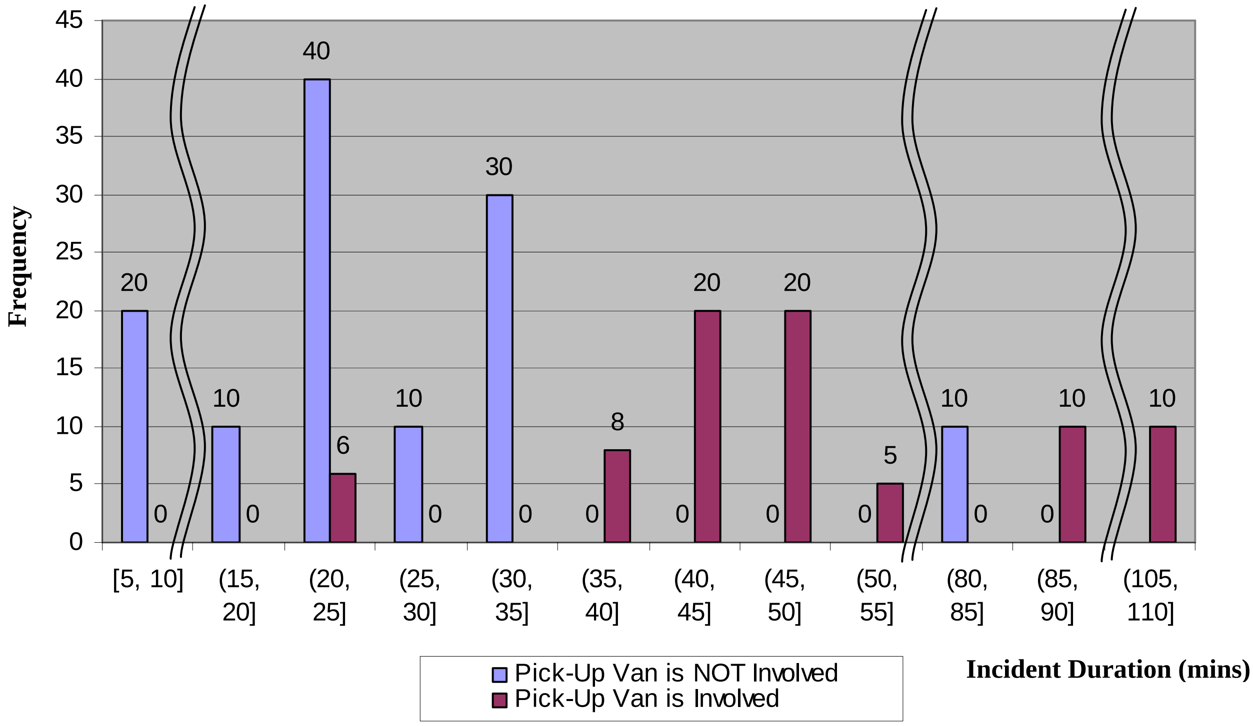
Rule-Based Tree Model (RBTM)

Based on the findings from preliminary analysis, this study has redesigned a classification tree, named a Rule-Based Tree Model (RBTM), using the following procedures. Note that incident duration, which was grouped into 5-minute intervals, is used to develop a model in this approach.

Step 1: Set the incident nature as the first splitter.
In this research, incident nature is classified as collision-fatality, collision-personal injury, collision-property damage, disabled vehicles or others which include all other kinds of incidents.

Step 2: Determine the next splitter for each node.
This step is to generate a crosstabulation table (9) to determine the next splitter for each node. A crosstabulation table can display the number of cases in each category defined by two or more specified variables. For each independent and dependent variable (i.e., incident duration), this step shall create a crosstabulation table along with a bar chart to show the distribution of frequency for different categories of the independent variable that is potentially associated with the incident duration. Then, the independent variable that exhibits a most different kind of distribution in different categories shall be selected as the next splitter.

Figure 1 is created to assist better understanding of this step. The cases used for this figure is taken from some part of the real dataset used for this research but not directly related to this stage. Assume that this bar chart is created based on the incidents whose nature is collision-personal injury from Step 1. As shown in the figure, a frequency bar chart is created for each category of selected independent variable, that is, whether any pick-up van is involved with incidents or not. In this instance, two categories display obviously different distributions. Incidents without pick-up van involvement are highly likely to be cleared within 30 minutes, while incidents involved with pick-up vans are more likely to last longer. To see if there are any other independent variables showing more different distributions for their categories, we create this kind of bar charts for every available independent variable. After comparing their results, we select the independent variable which is displaying the most different distribution for its categories as a next splitter.



*Note: some of incident duration ranges are omitted since their frequencies are zero, and they are replace with a symbol .

Figure 1. Distributions of Categories for a Pick-Up Van Involved Indicator

Step 3: Split the node based on the determined splitter in each category.
The focus of this step is to convert each splitting node into *If-then; Else-then* statement, which will constitute the set of rules for determining the incident duration for the node.

For instance, consider an example presented in the previous steps whose nature is collision-personal injury (CPI). Also, assume that the pick-up van involvement indicator is selected as the next splitter. Then, the split nodes on the basis of categories of this splitter can be presented as *If-then; Else-then* statements such as *If Incident Nature is CPI & Pick-up Van is Not involved, then Incident Duration is ...; If Incident Nature is CPI & Pick-up Van is involved, then Incident Duration is ….*

Step 4: Assign the estimated/predicted incident duration range for each split node.
This is to determine the best representative range of incident duration for each node. To achieve this, one shall first search the interval less than or equal to 30 minutes which covers at least 70% of all cases within a node. If no such interval exists within the node, then one can assign the shortest interval covering at least 60% of all cases within the node as the predicted incident duration for that node.

In Figure 1, most of the incidents that do not involve pick-up vans are distributed in the range of 5~30 minutes. Since this incident duration range is less than 30 minutes and covers about 92% (110/120) of all cases within given conditions (within this node), 5~30 minutes will be the most plausible incident duration interval in this node. On the other hand, incidents involved with pick-up vans are more widely distributed, and thus it is more complicated to determine the incident duration range representing the given conditions. In this instance, first, we need to find any 30 minute-interval covering the most of incidents, and it turns out to be 20~50 minutes covering approximately 68% of incidents in the node. Since it does not exceed 70%, secondly, we find any shortest duration interval covering at least 60% of all cases. In the figure above, it appears to be 35~50 minutes with about 61% covering.

Step 5: <u>Repeat Step 2 to Step 4 for all nodes until satisfy the predetermined criteria for stopping the tree growth.</u>

When a node satisfies one of the following criteria, one can stop the tree at that node.

1. No independent variable is available as a splitter.
2. There is only one observation left in a node.

Repeating Step 2 to Step 4 will help improve the model performance. Considering the instance presented previously, adding another splitter can either narrow the range of incident durations estimated/predicted or increase the proportion of incidents covered by that range of incident duration. This procedure can be repeated until no independent variable shows different distributions of its categories. This implies that categories of remaining independent variables barely contribute on discriminating incident durations or there is only one case left in given conditions. As an additional splitter is added, the number of cases falling in that node decreases for sure, and it usually faces the first stopping criterion before only one case remains in a node.

Figure 2 describes the structure of the Rule-Based Tree Model (RBTM). The developed RBTM starts with the first splitter, and for incidents falling in each category of the first splitter the second splitter is determined. In Figure 2, categories of the first splitter are expressed with a subscript 1, i.e., *Category* A_1 or *Category* B_1. Note that the second splitters can be different for each subset of incidents for each category of the first splitter depending on its characteristics and distributions. Figure 2 reflects this feature with numbering the 2nd splitters such as *2nd Splitter-1* or *2nd Splitter-2*. Likewise, in the figure the categories of second splitters are subscribed with numbers, e.g., 2-1 or 2-2, to distinguish multiple splitters at the same level. The tree model continues to extend with third splitters and so on. In this study, the first splitter turns out to be incident nature and it has five categories which creates five branches extending from the first splitter unlike the figure.

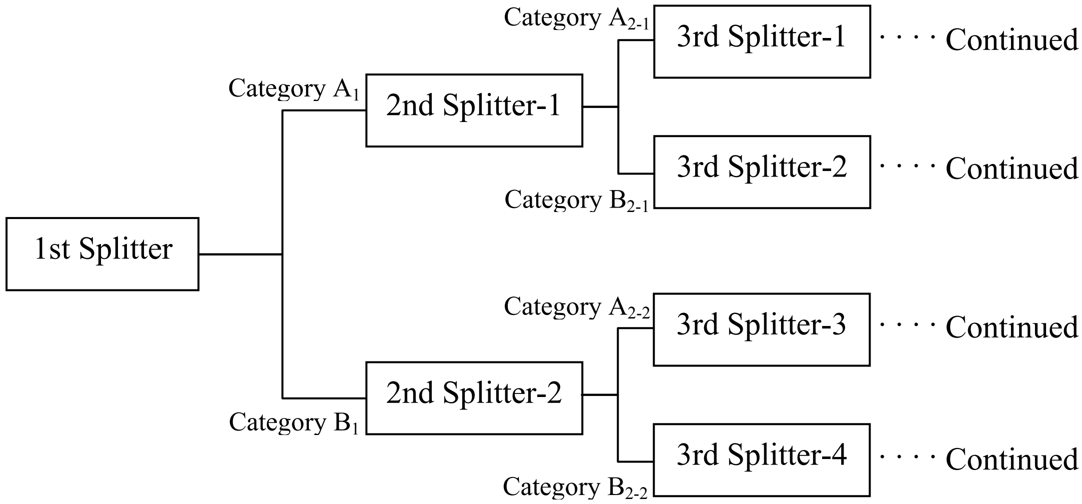


Figure 2. Illustration of the Structure of the Rule Based Tree Model

To evaluate the performance of the rules for each node, this study adopts the concepts of ***support*** and ***confidence*** developed for Associate Rules (9). The ***support*** for the rule refers to the number of cases that satisfies the *If-Then* rule. The ***confidence*** of the rule is defined as the ratio of the number of cases satisfying the *If-Then* rule (i.e., the ***support***) to the number of cases satisfying the *If* statement only. The indicator of c***onfidence*** is conceptually the same as the conditional probability of the *Then* statement given the *If* statement of the rule.

Results and Overall Findings

The developed RBTM shows that the *County* factor is the next splitter for incident natures of *Collision-Personal Injury*, *Collision-Property Damage*, *Disabled Vehicle* and *Others*. It implies that the duration for the same type of incidents varies significantly among different jurisdictions. On the other hand, the next splitter of *Collision-Fatality* is the *Weekend* factor. The reason why the *County* factor does not emerge as the next splitter can be found in the sample size. Since the total number of fatality incidents is usually much smaller than any other natures', it is hard to acquire any significant statistical results with a variable having multiple categories, e.g., *County*. For example, the total number of fatality incidents used to develop the RBTM is only 75 and the number of counties in Maryland is more than 10 so that some of counties may include only a few incidents. Thus, once more data for fatality incidents are collected, *County* would emerge as the next splitter like other incident natures.

Due to the time constraint, the RBTMs are first developed for Montgomery County in Maryland for each incident nature except for *Collision-Fatality*. This paper presents only a part of the developed RBTM in Figure 3 due to the constraint of the paper length. As shown in Figure 3, the estimated incident duration range is narrow as additional splitters are included. Table 1 summarizes the estimation and validation results. As shown in Table 1, the overall estimated results from the developed model are quite satisfactory, while the overall validated results are not as satisfactory as the estimated one. However, there are still remarkable findings through the model development.

Table 1. Summary of Overall Model Results

Overall Model Results		Incident Nature				
		Collision – Fatality	Collision – Personal Injury[1]	Collision –Property Damage[1]	Disabled Vehicles[1]	Others[1]
Number of Terminal Rules		21	32	37	5	12
Average Confidence	Estimated	98.23%	81.04%	89.71%	87.62%	92.93%
	Validated	12.19%	45.45%	39.36%	66.50%	63.33%

*Note: [1] Results only for Montgomery County

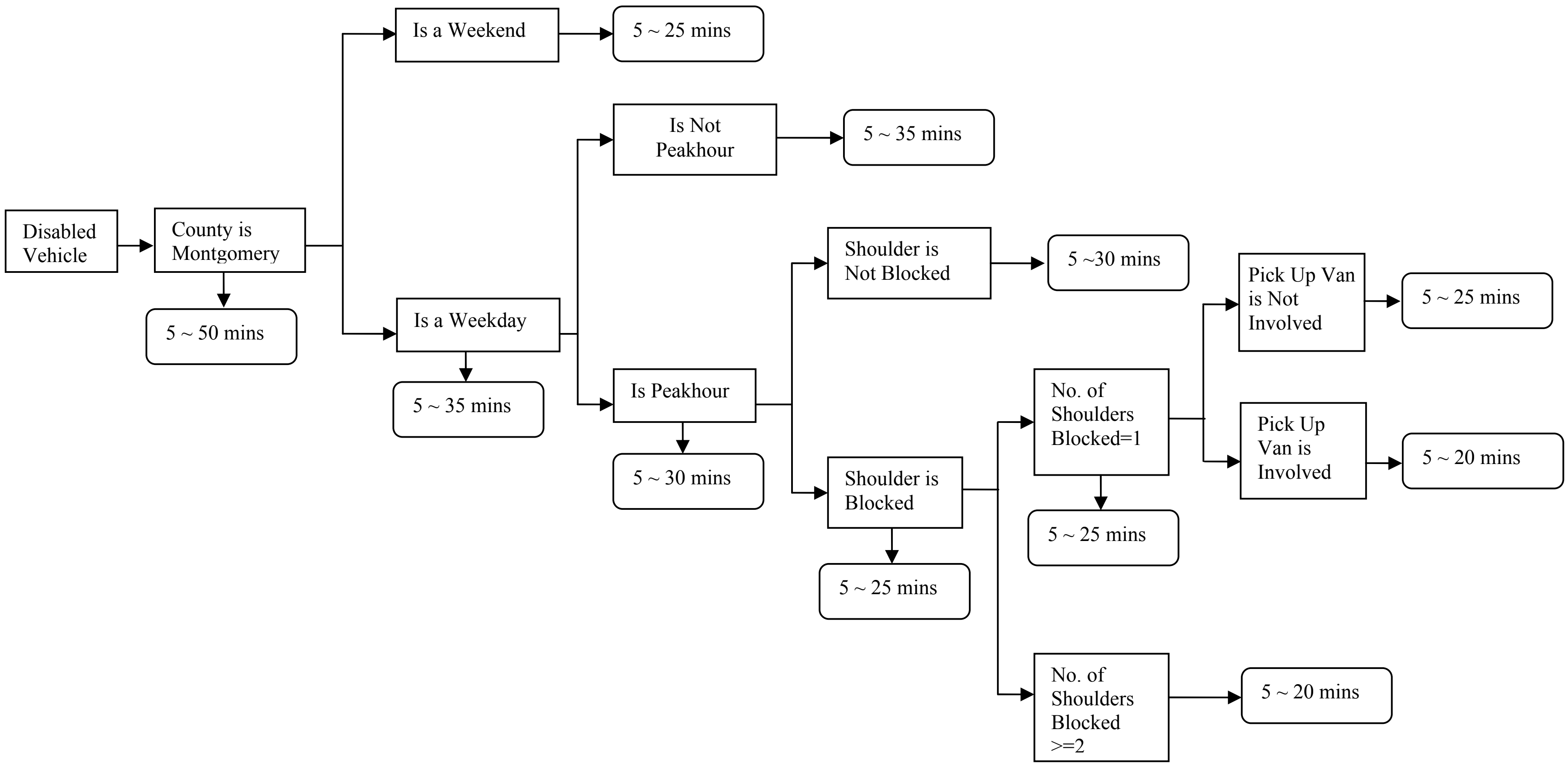


Figure 3. Rule Based Tree Model for Disabled Vehicles in Montgomery County in Maryland

1. The sequence of splitters varies significantly among different categories of incidents, since different natures have different characteristics and are associated with different contributing factors. For instance, although the second splitter for *Collision-Personal Injury (CPI)* and *Collision-Property Damage (CPD)* is same, *County*, the third splitter for *CPI* is *Total Number of Blocked Lanes*, while the third splitter for *CPD* is *Tractor-Trailer Involvement Indicator*.
2. Rule-Based Tree Models are more flexible for assigning an appropriate estimated incident duration range in given conditions (sub-dataset or node) than Classification and Regression Tree Models (CART). In RBTM, the assigned ranges of incident duration are different from one another unlike CART in terms of the length of range. Considering the instance used in the previous section, the estimated incident duration range for incidents not involved with pick-up vans is 5~30 minutes which is a 25 minute-interval. On the other hand, the estimated incident duration range for incidents involved with pick-up vans is 35~50 minutes which is a 15 minute-interval. This flexibility contributes on searching the range of incident duration representing the given incidents the best.
3. Incidents occurring at night time or during off-peak hours generally take a longer duration than those in daytime due to the lack of sufficient response units for incident clearance operations.
4. When incidents result in *Collision-Fatality,* or *Property Damage*, the clearance operation is generally more efficient in the shoulder-lane blocked scenarios than those leaving it open. In the developed RBTM for fatality incidents, the estimated incident duration range for incidents without shoulder blockage is 120~200 minutes under certain conditions, while the estimated incident duration for incidents with shoulder blockage is 100~180 minutes under the same conditions. This finding implies that shoulder lane blockage helps reduce the duration of severe accidents as it provides a wider space for emergency response units to do the work.
5. Similarly, during the *Collision-Fatality* incidents, if the emergency response unit can close more lanes in the same direction, it generally results in a shorter duration. For instance, the developed RBTM shows that under particular conditions incidents managed with more than two lanes blocked last 55~80 minutes, whereas incidents managed with less than or equal to two lanes blocked are cleared in 120~160 minutes under the same conditions for other variables.
6. The impact of wet pavement, a proxy variable for rainy days, on the efficiency of incident response operations is not definitive for the existing data records. It shows a positive correlation with the incident duration for those resulting in *Collision-Property Damage*, since but a reverse relation for the category of *Collision-Fatality* incidents. For all other types of incidents, its impacts on the resulting incident duration are not statistically significant.

CONCLUSIONS AND FUTURE RESEARCH

Although the estimated results from the Rule-Based Tree Model (RBTM) are quite acceptable, one shall not expect the above RBTM to capture all embedded relations and provide the operationally acceptable performance for real-world applications due to the complex nature of incidents and response operations. Hence, grounded on the promising information generated from the RBTM, supplemental models are required to improve the prediction accuracy for the duration of a detected incident. In addition, it is essential to integrate the CHART database with police accident records to construct a dataset with better quality for calibrating models of incidents, especially involved with fatalities.

REFERENCES

(1) Transportation Research Board (TRB), *Special Report 209: Highway Capacity Manual,* National Research Council, Washington, D.C., 1994

(2) T. F. Golob, W. W. Recker, and J. D. Leonard, "An Analysis of the Severity and Incident Duration of Truck-Involved Freeway Accidents", *Accident Analysis and Prevention,* Vol. 19, No. 4, August 1987, pp. 375-395

(3) D. Nam, and F. Mannering, "An Exploratory Hazard-Based Analysis of Highway Incident Duration", *Transportation Research – A*, Vol. 34A, No. 2, 2000, pp. 85-102

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

(5) A. J. Khattak, J. L. Schofer, and M-H. Wang, "A Simple Time Sequential Procedure for Predicting Freeway Incident Duration", *IVHS Journal*, Vol. 2, No. 2, 1995, pp. 113-138

(6) K. Smith, and B. Smith, "Forecasting the Clearance Time of Freeway Accidents", Research Report, STL-2001-01, Center for Transportation Studies, University of Virginia, Charlottesville, VA, 2001

(7) P-W. Lin, N. Zou, and G-L. Chang, "Integration of a Discrete Choice Model and a Rule-Based System for Estimation of Incident Duration: a Case Study in Maryland", CD-ROM of Proceedings of the 83rd TRB Annual Meeting, Washington, D.C., 2004

(8) L. Breiman, J. H. Friedman, R. A. Olshen., and C. J. Stone, Classification and regression trees

(9) T. Hill, and P. Lewicki, Statistics: Methods and Applications