Validation of WAIMSS Incident Duration Estimation Model

Wei Wu
Virginia Polytechnic Institute and State University

Pushkin Kachroo
University of Nevada,Las Vegas, pushkin@unlv.edu

Kaan Ozbay
Rutgers University - New Brunswick/Piscataway, kaan@rci.rutgers.edu

Validation of WAIMSS Incident Duration Estimation Model

Wei Wu
Center for Transportation Research
Virginia Tech, Blacksburg, VA 24061

Pushkin Kachroo
Bradley Department of Electrical and Computer Engineering,
Virginia Tech, Blacksburg, VA 24061
Kaan Özbay
Department of Civil & Environmental Engineering
Rutgers, the State University of New Jersey
Piscataway, NJ 08855-0909

ABSTRACT

This paper presents an effort to validate the incident duration estimation model of WAIMSS – Wide Area Incident Management Support System (WAIMSS). Duration estimation model of WAIMSS predicts the incident duration based on an estimation tree which was calibrated using incident data collected in Northern Virginia.

The validation process started with collection of new incident data which was conducted by video taping incident management processes in Northern Virginia and keeping detailed incident logs of actual incidents at Northern Virginia Traffic Control Center. The collected incident data was then partitioned into a number of subsets according to the structure of the original estimation tree. Due to the limited sample size, a full scale test of the distribution, mean and variance of incident durations was performed only for the root node of the estimation tree, while only mean tests were executed at all other nodes whenever a data subset was available. Further studies were also conducted on the model error and tree structure issues especially related to complex incidents – incidents with multiple major discriminating incident characteristics.

The statistical analyses in general, strongly supported WAIMSS estimations of incident duration distribution, mean and variance. The error analysis provided encouraging results based on the distribution of estimation errors and estimation error percentages.

A major structural deficiency of the current model was also revealed. While WAIMSS duration estimation model is effective in dealing with incidents with one important characteristic, for complex incidents with more than one discriminating characteristic, only the most significant incident characteristic is used by the current model and others are simply ignored. Thus, data analysis shows a trend of under-estimation for complex incidents. Alternatives to improve this shortcoming of the current model along with several other possible improvements are also recommended.

1. INTRODUCTION

Recent research efforts at Virginia Tech Center for Transportation Research for developing a real-time incident management system that is capable of managing traffic flow and roadway infrastructure in the event an incident resulted with an hybrid expert-GIS Wide-Area Incident Management Support System (WAIMSS) [1]. WAIMSS combines the powerful spatial data handling capabilities of a Geographic Information System (GIS) with the rule-based logic of an expert system in a fully integrated Expert-GIS framework to provide interactive content and group process support for incident management operations.

One of the major functional components of WAIMSS [2], is the incident duration estimation module. The core of this module is a tree-structured incident duration estimation model (referred to as WAIMSS model later). The derivation of the WAIMSS model was based on a good understanding of incident characteristics and the factors affecting the incident clearance times. It was found that the incident duration was determined mainly by the incident type, severity, and clearance characteristics. Those factors were then used to develop estimation / decision trees for each incident type using a variant of the tree-structured regression method. Statistical analysis of data had also shown that the duration of incidents was normally distributed for homogenous sub-groups of incidents that were categorized according to the most significant affecting factors. This property of incidents was then used to estimate the likelihood of an incident to last more than a specific period of time.

WAIMSS model was developed by analyzing historical incident data (referred to as C-data as it was used for model calibration) which was collected by the Northern Virginia (NOVA) incident management personnel. B-data contained over 5000 incident cases occurred in Northern Virginia, provided a sound basis for the development and calibration WAIMSS model. As the incident duration is a vital decision-making factor in a wide range of incident response and traffic control operations, it is of critical importance to conduct a further study on WAIMMS model in order to implement it on-line in a real world scenario. The research presented here is an effort to validate the existing WAIMSS model and identify some of the potential improvements.

2. WAIMMS DURATION ESTIMATION MODEL

Several other studies on incident duration estimation were conducted in the past. Jones et al. [3] developed a multivariate statistical model for incident frequency and duration based on State Police Dispatch records over a two-year period. Golob et al. [4][5] developed a log-normal incident duration model based on theoretical considerations of the incident process. Effort by Wang et al. [6] at Northwestern University attempted to develop an initial capability to provide incident duration estimation which improves as the incident process proceeds. Another effort at Northern University by Sethi et al. [7] developed an incident duration estimation diagram based on detailed study of incident data from Northwest Central Dispatch.

WAIMSS model was developed based on the concept of tree Structured regression, where the tree is constructed by partitioning the data set into a sequence of gradually more homogeneous subsets, called node. Each node produces a response a variable or a prediction based on the node data subset. In WAIMSS' model, this prediction is the mean of the node data subset, i.e.:

$$\hat{y}_k = \sum_{i \in Node\ k} y_i / N_k$$

where N_k is the number of incidents in node k, and y_i is the duration of ith incident in node k and $\hat{y}$ is the predicted incident duration. Approximating incident duration in node k by $\hat{y}$ minimizes the squared errors in node k as defined by:

$$Cost = \sum_{i \in Node\ k} (y_i - \hat{y}_k)^2$$

At the root of the estimation tree, level 0, is the category of all incidents without any classification. From this root node, Several sub-nodes branch out, such as Road Hazard, Property Damage, Personal Injury, Disabled Truck, Vehicle Fire, Hazmat, Weather Related, and Car Breakdown. These first level sub-nodes are further branched to next level sub-nodes.

Property Damage incident type is divided into Truck Involved and Cars Only types; Cars Only type is further divided into 1-3 Cars Involved and 4 and More Cars involved types. As this process goes on, more information of a incident is used in the estimation, and a more accurate prediction is expected.

WAIMSSestimation tree can be used as part of a complete incident management support system folowingtwodistint two steps are involved. The first one is to find the most appropriate node using the current available incident information; then y_k is used as the duration prediction for that incident. As an incident proceeds, more information about the incident may be obtained, therefore the prediction precision can be refined over the incident process.

3. STUDY APPROACH

As the goal of the study is to test the validity of the existing model and identify likely improvements, this research is conducted in three basic steps:

- Data collection
- Statistical testing of assertions made by WAIMSS' model
- Identification potential improvements based on both analytical and statistical studies.

These steps are discussed in more detail in the following sections.

4. DATA COLLECTION

As a result of a meeting with Virginia Department of Transportation (VDOT) and Federal Highway Administration (FHWA) officials, it was decided that new real world data should be used for validation purposes. Since VDOT's NOVA Traffic Management System Control Center has a CCTV traffic monitoring system which covers portions of I-395, and I-66 and other road segments in NOVA, the Center was chosen as the best location for data collection.

The first phase of data collection which produced a data set of 46 traffic incidents was conducted from March 3, 1997 to March 22, 1997. For each traffic incident, data gathered include:

- Videotape recording – By remote maneuvering, field cameras were zoomed to the incident scene whenever an incident was reported or identified. The camera then was used to keep track of all the activities at the incident scene, and the traffic queuing process whenever possible. All these video clips were captured on a VCR.
- Incident log – Traffic controllers at the NOVA Traffic Management System Control Center also kept a written description of the incidents, which outline the details of incidents and corresponding emergency responses and traffic control operations.

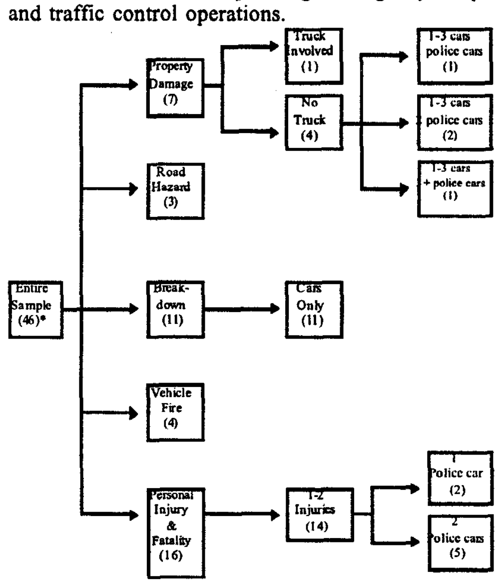

Figure 1. Classification Tree of V-Data (* Number of cases)

The dual information sources are complementary to each other and provided a complete description of each incident. The data synthesis was conducted by compiling information contained in video tape recordings and in the written logs, which provided us with us the validation data set (referred to as V-data or simply sample). Before statistical tests were performed, V-data were partitioned into smaller subsets based on the structure of the original WAIMSS estimation tree. Each of these subsets is called a node sample. Figure 1 displays the resultant data classification tree.

The entire validation data sample has 46 incident cases, which are classified into fifteen node samples at four different layers. Note that some cases are dropped out during the classification process because:

- they do not fit in the classification adopted by WAIMSS, or
- there is not enough information for further classification.

In addition, it is seen that the sample data set does not cover the entire estimation tree. This obviously presents a need for more data to be able to study the missing groups.

5. STATISTICAL TESTING

Basically, three types of statistical assertions are made by the original WAIMSS model:

- Distribution of incident duration – WAIMSS model assumes that, except for the overall data set, the incident durations of homogeneous subsets of incidents conform to normal distribution.
- Means of incident duration – WAIMSS model uses the means of individual duration groups as the estimated incident duration. This is actually the model prediction, which will be tested.
- Variance of incident duration – Variance of incident duration is used in computing the probability of incident duration exceeding a certain threshold.

Accordingly, our statistical testing should be performed to test the validity of these statistical assertions. Due to the limited size of node samples, however, testing on distribution and variance is performed only on the overall data set, i.e., the root node, while only mean testing is performed on all node samples. In general, tests on mean follow the strategy shown in Figure 2.

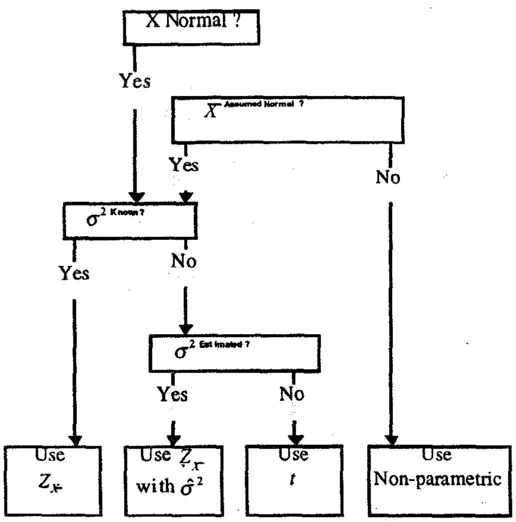

Figure 2. Strategy for Test of Mean. Source [8].

5.1. Entire Sample

WAIMSS model assertions were based on the statistical results of the calibration data set. Therefore, the test of WAIMSS model is basically a test of the statistical consistency of the calibration data set and the validation data set. Table 1 shows a brief listing of statistics of the validation and calibration samples.

V-data set is also shown to be non-normal and this result supports the first assertion of WAIMSS model, namely the non-normality of the overall incident data. The mean of both data sets are also shown to be very close and this result supports the use of mean as the prediction value by the WAIMSS model. However, there is a large difference between the standard deviations and this is most likely due to the small size of V-data set.

5.1.1. Test of Distribution

To test the normal distribution assumption, a χ^2 Goodness-of-Fit test is performed. The entire V- data (46 cases) were used, and Table 2 listed the data grouped into time intervals.

Table 1. List of Sample Statistics and Model Assertions

	V-data	C-data
Observation	46	650
Mean	42.96	45.03
Mode	44	30
Maximum	256	417
Minimum	3	7
Standard Deviation	43.24	33.85
Distribution	Non-Normal	Non-Normal

Table 2. Observed versus Expected Incident Duration

Intervals (min)	Observed Frequency O_i	Expected Frequency E_i
0-10	6	10.18483026
10-20	8	3.447964734
20-30	8	3.90249721
30-50	10	8.438993811
50-60	6	4.096246202
60-	8	15.92945742
Total	46	46

The test statistic is

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 17.15 > \chi^2_{0.005,5} = 16.75$$

So the normal distribution hypothesis is rejected even at 0.5% significance level. A histogram of the V-data , Figure 3, also shows a clear pattern different from a normal distribution. However, the histogram shows a very similar pattern of C-data, as is shown in Figure 4. The following points are observed from the two histograms:

- While most incidents last less than one hour, the spread of the incident duration is much wider.

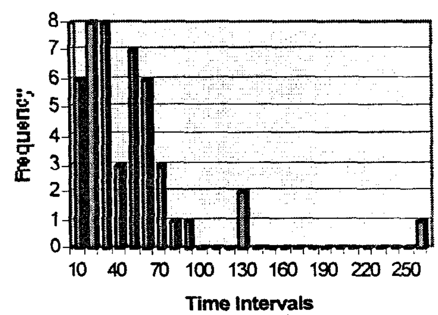

Figure 3. Incident Duration Distribution of the V-Sample

- It shows two significant peaks in incident duration distribution. The global maximum appears around 30 minutes, which is corresponding to the largest cluster of minor incidents; and a secondary local maximum appears around 60 minutes, which corresponds to the cluster of moderate incidents

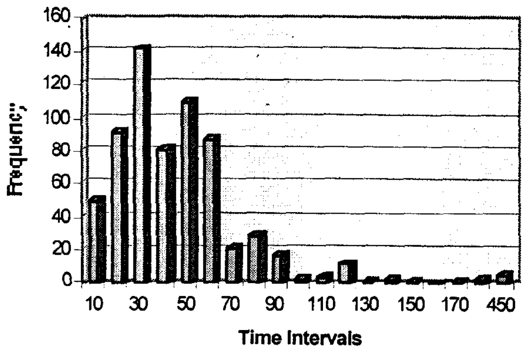

Figure 4. Incident Duration Distribution of the C-Sample

5.1.2. Test of Mean

As verified before, the incident duration of the entire data set does not follow the normal distribution. However, according to Central Limit Theorem, we can still assume normal distribution to test the mean because of our large sample size. The test statistic used is

$$z_0 = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{N}}$$

where $\overline{X}$ and N are the mean and size of the validation sample, and μ_0 equals 45 and σ equals 33.85 as asserted by the WAIMSS model. The test resulted in a p-value of 67.78% , a strong indication of the correctness of the original assumption.

5.1.3. Test of Variance

F test is used in the testing because that the incident duration is not normally distributed. The test statistic is

$$f_0 = \frac{S_1^2}{S_2^2} = 1.63 < f_{0.01,45,649} = 1.75$$

So, we can not reject the null hypothesis that the variances of the calibration and the validation samples are the same at the significant level of 1%. This result supports the fact that the calibration and the validation samples are coming from the same population.

5.2. Summary of Statistical Tests

Statistical tests using the V-data set in general supports WAIMSS incident duration estimation model, and whenever data allows, WAIMSS assertions on incident duration distributions, means and variances are verified to be statistically acceptable at large margins in most of the scenarios. For example, results of mean test of Property Damage incident duration that are partitioned into five different nodes at three different levels, provided useful insights. Except for one node – Property Damage with 1~3 cars involved and 2 police cars responded, all other node data show very strong consistency with the WAIMSS model, with p-values varying from 92.8% to 25.6%. In general, the average p-value is 42.8%. A summary of the mean tests is shown in Table 3.

Table 3. Summary of the Mean Tests

Node Description	Model Prediction (min)	Sample Mean (min)	P-Value (%)	Sample Size
Entire Sample	43	45.03	67.8	46
Property damage	42	36.2	60.4	7
Truck involved	42	44.00	92.8	1
No truck	N/A			4
1-3 cars and 1 police car	38	13.00	25.6	1
1-3 cars and 2 police cars	46	25.00	1.3	2
1-3 cars and 3+ police cars	65	74.00	33.8	1
Road hazard	27	6.00	1.2	3
Breakdown	N/A			11
Cars only	27	26.82	97.5	11
Vehicle fire	43	57	18.4	4
Personal injury	51	49.4	82.6	14
1-2 injuries	49	49.36	96.1	14
1 police car	42	67	12.4	2
2 police car	52	46.20	64.3	5

6. OTHER ANALYSIS

6.1. Error Analysis

Even though the previous statistical test results have shown a good agreement with WAIMSS model estimations, especially from the perspective of a traffic controller the error of an individual prediction is of a great importance. Since an ormal distribution assumption was used by WAIMSS' model for most of the the homogeneous incident categories, the risk of having a particular error can be calculated using normal distribution and its associated parameters. However, due to the limited V-data size, as an alternative approach, we directly use the differences between the predicted duration and the observed duration in our error analysis.

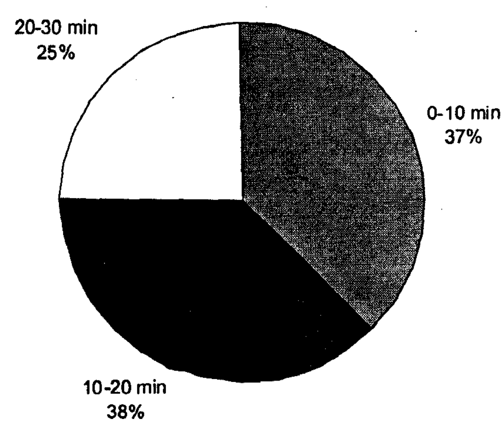

Figure 5. Distribution of Estimation Errors

For incident cases which fit different nodes at different levels, the prediction from the most terminal node is used. We define the *Estimation Error* as :

(Observed Duration) - (Predicted Duration)

and *Estimation Error Percentage* as:

100% (Estimation Error)/ (Predicted Duration)

The average absolute *Estimation Error* is 14.2 minutes, and the overall sum of *Estimation Error* is -134 minutes, which is an indication of under-estimation. The distribution of the *Estimation Error* which is shown in Figure 5 categorizes the *Estimation Errors* into three main time intervals.

It is shown that three-eighths of the prediction errors are less 10 minutes, and three-fourths are less than 20 minutes. And all the errors are less than half an hour. However, the *Estimation Error Percentage* does not look as good as the *Estimation Error*. Figure 6 shows the *Estimation Error Percentage* for each incident used in this analysis.

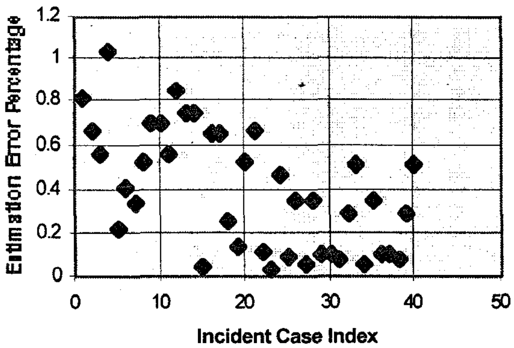

Figure 6. Estimation Error Percentage

A cumulative distribution of *Estimation Error Percentage* is shown in Figure 7.

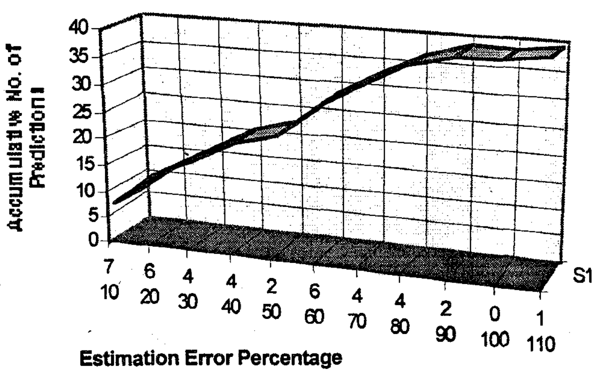

Figure 7. Distribution of Estimation Error Percentage

It is seen that near 50% of predictions have an *Estimation Error Percentage* less than 50%. A almost linear relationship is observed between the *Cumulative No. of Predictions* and *Estimation Error Percentage* before *Estimation Error Percentage* reaches about 80%. Number of predictions with *Estimation Error Percentage* over 80% increases remarkably slowly.

6.2. Incident Classification Problem

Use of WAIMSS model requires finding the most appropriate node in the estimation tree for a specific incident. The current tree representation has only less than ten first level nodes such as Road Hazard, Property Damage, Personal Injury, Disabled Truck, Vehicle Fire, Hazmat, Weather Related, and Car Breakdown. These are major incident characteristics currently used in WAIMSS to define estimation tree structure, and nodes defined by these major incident represents major incident types. If we call an incident with only one major incident characteristic a simple incident, and an incident with multiple major incident characteristics a complex incident, the current estimation tree structure does not suit complex incidents because a complex incident cannot be accurately represented.

A number of such situations were encountered in the validation. A natural choice is to choose the incident characteristic that has the most importance, however, by doing that, the effect of the minor incident characteristics are ignored. The choice is harder when an incident can belong to two major categories having comparable importance from the duration point of view. Suppose an incident involving both personal injury and property damage, as the former has a higher importance the incident duration should be dominated by the personal injury characteristic. But, property damage as a secondary incident characteristic should make the incident duration longer than a pure personal injury incident. This fact is also confirmed by V-data.

In V-data setwe have 7 cases of only property damage, 12 cases of both property damage and personal injury, and 14 cases of personal injury with or without other features. The duration mean of these three groups, as in Table 4, shows exactly the pattern discussed above.

Table 4. Comparison of the Mean of Incidents with or without Multiple Characteristics

	Only Property Damage	Property Damage & Personal Injury	Personal Injury (with or without other features)
Mean (Mins)	26.85	51.3	49.4
No. of Cases	7	12	14

For the group of incidents with both personal injury and property damage, we tried two different classification. Figure 8 shows the estimation errors resulted from two different classifications. It is apparent that classifying incident using its most significant incident characteristic results in a better prediction.

7. CONCLUSIONS

This study tested the statistical assertions on which WAIMSS incident duration estimation model is based. Statistical tests using the overall V-data showed that:

- The incident duration distribution of V-data is very similar to that of C-data. Both show a double peak pattern. The global maximum appears around 30 minutes, which corresponds to the largest cluster of minor incidents, and a secondary local maximum appears around 60 minutes, which corresponds to the cluster of moderate incidents.
- While most incidents last less than one hour, the spread of the incident duration is much wider than 1 hour.

- Non-normal duration distribution assumption for overall incidents is strongly supported by the V-data set.
- V-data sets mean matches extremely closely with that of C-data extremely (43, 45).

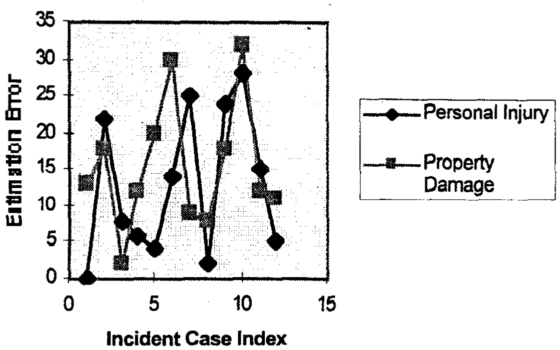

Figure 8. Estimation Errors Resulted from Different Incident Classifications

Tests using partitioned V-data subsets showed that:

- Most duration estimations of WAIMSS are accepted with large p-values (average 50%), except for two nodes. One of the two node data subsets shows apparent bias.
- In general, WAIMSS' model estimations is consistent with the observed, the average prediction error of V-data is 14.2%.
- In general, a higher consistency is observed when the subset data size is bigger.

Error analysis and additional studies show that:

- About 38% of all the estimationerrors are less than 10 minutes, and about 75% of all the estimation errors are less than 20 minutes, and almost all the prediction errors are less than half an hour.
- About 50% of all estimations have an estimation error percentage less than 50%.
- Estimation errors distribute uniformly within an error percentage ranging from zero to 80 percent. Estimation errors with error percentage higher than 80% are rare.
- For complex incidents, it is better to use the most significant incident characteristics to identify their most appropriate tree node.

7.2. Future Improvements

Tree Structure: In general, the current tree structured estimation model does not make full use of the available incident information, only the information used in locating the proper node contribute to the estimation. In case of a complex incident, only the most significant major incident characteristic is used while other secondary major incident characteristics are ignored. This is very likely to be the cause of under-estimation observed in complex incident cases that were in the V-data set. As a complex incident is a very frequent phenomenon, structural changes in the estimation tree are necessary.

Node Prediction Mechanism: Under the tree structure, the duration estimation at each node is simply the mean of the calibration data subset. Therefore, all the incidents belonging to a node have the same duration prediction regardless of their other features. This might be the reason why we see very close means of V-data and C-data, but still observe considerable errors for some individual cases. To avoid this limitation, more incident features need to be incorporated into the prediction mechanism at each node. A natural choice can be to replace the current mean estimation with a multivariate regression model at each node.

Effect of Congestion: While congestion is considered an adverse factor during incident clearance, congestion is not currently used as a factor affecting incident duration. Congestion lengthens the incident duration mainly by affecting the travel time of response vehicles, and in some cases, affecting operations at incident scene. Visual recordings of incidents show that congestion effects are magnified in certain roadway environments. It is seen that response times to incidents on bridges are substantially extended, and efficiency of incident removal operations on bridges and ramps are reduced by limited space. A special study to determine the effect of congestion on the incident clearance is necessary.

REFERENCES

1. Ozbay, K., Narayanan, A., Jonnalagadda, S., "Wide-Area Incident Management Support System (WAIMSS) Software," Paper presented at the Third Annual World Congress on ITS, Orlando, Florida, October 14-18, 1996.
2. Ozbay, K., Hobeika, A., and Zhang, Y.L., "Estimation of Duration of Incidents in Northern Virginia", Paper presented at the 75th TRB Annual Meeting, Washington D.C., January, 1997.
3. Jones, B., Janssen,L., and Mannering, F., "Analysis of the Frequency and Duration of Freeway Accidents in Seattle," *Accident Analysis and Prevention*, Vol. 23, No.4, pp. 239-255, 1991.
4. Golob T.F. ,Recker W.W. and Leonard J.D., "An Analysis of the severity and Incident Duration of Truck Involved Freeway Accident," *Accident Analysis and Prevention*, 19, pp. 375-395, 1987.
5. Giuliano, G., "Incident Characteristics, Frequency, and Duration on a High Volume Urban Freeway," *Transportation Research-A*, Vol. 23A, No. 5., pp 187-396, 1989.
6. Wang, M., *Modeling Freeway Incident Clearance Times*, Unpublished MS thesis, Northwestern University, Evanston, Illinois, 1991.
7. Sethi, V., Koppleman, F.S., Flannery, C.P., Bhandari, N., and Schofer, J., "Duration and Travel Time Impacts of Incidents," *ADVANCE Project Technical Report*, Northwestern University, 1994.
8. Frank, H., and Althoen, S.C., "Statistics: Concepts and Applications", Cambridge University Press, 1994.