# INCIDENT DURATION PREDICTION FOR IN-VEHICLE NAVIGATION SYSTEM

**Jun Hu**
Centre for Transport Studies, Department of Civil and Environmental Engineering, Skempton Building, South Kensington Campus, Imperial College London, London SW7 2BU, UK
Email: jh205@imperial.ac.uk

**Rajesh Krishnan**
Centre for Transport Studies, Department of Civil and Environmental Engineering, Skempton Building, South Kensington Campus, Imperial College London, London SW7 2BU, UK
Email: rajesh.k@imperial.ac.uk

**Michael G. H. Bell**
Centre for Transport Studies, Department of Civil and Environmental Engineering, Skempton Building, South Kensington Campus, Imperial College London, London SW7 2BU, UK
Email: mghbell@imperial.ac.uk

(5096 + 3· Figures (250) + 6· Tables (250) = 7346 words)

This paper was submitted to the Transportation Research Board.
1 August 2010

# INCIDENT DURATION PREDICTION FOR IN-VEHICLE NAGIVATION SYSTEM

**J. Hu, R. Krishnan, M.G.H. Bell**

Centre for Transport Studies, Imperial College London, UK

## ABSTRACT

Current vehicle navigation systems receive congestion information via RDS-TMC, to the extent that congestion is detected. A better quality of guidance would be possible, however, if short-term prediction of future congestion were available beside information about current congestion. This study used incident duration modelling techniques for the congestion prediction problem for use in-vehicle navigation systems. A unique database of TMC messages collected for the London road network is used in this study. A model that predicts incident duration is fitted with London TMC data using factors that affect the incident duration as explanatory variables are explored. The prediction accuracy of the model compared with other naïve predictors. It is shown that the proposed model is more accurate than the naïve predictors.

## 1. INTRODUCTION

The field of in-vehicle navigation system has quickly advanced in the last decade. The systems have evolved from the simple routing function of computing the fastest route from the current location to the destination, to the more sophisticated applications that take into account of the current and predicted traffic conditions. This is so-called dynamic routing function. This dynamic routing function is gaining importance and gradually being incorporated into newer navigation system.

One of the major benefits of route guidance system is better informed individual behaviour, in terms of departure time and route choice. The underlying assumption behind the usefulness of route guidance information is that the individual driver is a rational decision maker who makes his/her choice (e.g. departure time, route, etc) to minimize his/her expected travel costs (i.e. travel time or travel cost), subject to the constraints based on the requirements of each driver and their familiarity with the road network and their own preferences. Thus, reliability of route guidance information is vital to bringing out the full benefits of the system to the drivers. Reliability in this context means that the guidance provided by the navigation system is consistent with the traffic conditions experienced by drivers.

Congestion data are used in many new in-vehicle navigation systems. In current in-vehicle navigation systems, real-time congestion information can be accessed from a traffic information centre (TIC) to the vehicle via the Traffic Message Channel (TMC). If congestion has been detected on the recommended route, it will be displayed to the driver as a triangle on the map. These systems don't incorporate the anticipated traffic situations in route guidance. Hence, there is no guarantee that the recommended route will turn out to be congestion-free or the least congested one. Additionally, due to the latency between the time when the congestion occurs and drivers receiving the information, it may be too late for him/her to respond, and he/she will most likely be already held in congestion. Hence the guidance based on past or real-time traffic situation alone is not enough to provide reliable route guidance, the short-term congestion prediction information is required to give better quality of guidance to the driver.

Majority of congestions episodes on the road network are caused by incidents, such as accidents, vehicle breakdowns, roadworks etc. The prediction of congestion on a certain route is dependent upon the prediction of the incident duration on that route. The road incidents can be

classified into two categories: planned and unplanned incidents. For planned incidents, such as roadworks, the duration is generally provided to the driver in advance. Therefore the driver can make appropriate route choice at the start of the trip. One cannot predict the occurrence of unplanned incidents, such as accidents and vehicle breakdowns. However, once it is known that an unplanned incident has occurred, one can attempt to predict its duration. Then the prediction can be incorporated into the navigation system for selecting the route for the driver.

The motivation of this paper is to enhance the reliability of in-vehicle navigation system by predicting the duration of incidents that cause congestion. Concept from incident duration modelling is applied in the context of prediction in this paper. The rest of this paper is structured as follows: the next section provides a background on congestion prediction and duration modelling. The proposed modelling methodology is presented in the subsequent section and one year of traffic incident data from the London road network is used to calibrate and validate the proposed model. This dataset is described in the fourth section, followed by the results from model estimation in subsequent section. Finally, the findings are summarised and the paper is concluded in the last section.

## 2. BACKGROUND

A short background on congestion prediction methods and duration modelling is given in this section.

### 2.1 Congestion prediction for route guidance

The field of forecasting in the traffic domain is vast, with a large number of prediction studies appearing in the academic literature in the last decade. Most of the prediction studies focus on the short-term prediction of traffic parameters such as flow, occupancy, and speed.

Two different types of models are used in the literature for short-term prediction of traffic variables:simulation based models and data driven models.

The simulation-based approach makes use of traffic flow simulation models to predict the traffic conditions on the route of interest. Many simulation tools have been developed in this context, the most notable ones being DynaMIT (1) and SAVaNT (2). The advantages of simulation-based approaches are that the models used are largely based on traffic flow theory, they incorporate traffic control measures in the prediction, and they provide a detailed insight into the problem. However the main disadvantage is that these models are computationally demanding and hence not suitable for use in in-vehicle applications.

The data-driven approach, on the other hand, belongs to a broader category of prediction models, using machine learning models. The term machine learning is used to represent any method that can learn the relationship between two sets of related variables. These models have been widely documented in the past in transport literature. Examples includes statistical learning models (3),time series models (4), ARIMA models (5-7),ANN models (8) with combination of Bayesian method (9, 10) and also the KNN algorithm(11). The main advantage of data-driven approaches is that they do not require extensive expertise on traffic flow modelling, and as many ready-to-use software packages are available for model design and calibration, they are also fast and easy to implement (8). However, the drawback is all data-driven approaches is poor transferability since they need to be re-calibrated before being implemented on other networks.

Other hybrid approaches may also be found in literature, for example, (12) combines NN and ARIMA models, (13) uses a genetic algorithms with NN, (14) uses fuzzy sets and NN. Most of these models are not directly predicting the traffic congestion. They are predicting the traffic parameters, such as flow, occupancy etc.

Literature dealing directly with congestion prediction is from Horvitz's team in Microsoft research (10). In Horvitz's study, the congestion prediction problem is treated as the probability of congestion ("surprising event" in Horvitz's terminology) occurring on a given section of road during a

future period of time (30 minutes in Horvitz's case) and the likelihood of congestion occurring is predicted based on the current and historical traffic flow data using Bayesian statistical techniques. A threshold is set on the likelihood value to define the significance of the congestion occurrence. The model's prediction accuracy is also validated against real data from Greater Seattle area and is shown to be accurate for the test period. This model is the basis of a traffic service called INRIX available in the U.S.(15). Again, the traffic flow data is the basic input parameter for the model. The model is too computational intensive to be implemented on in-vehicle navigation units.

It must also be noted that many of the literature on predictive route guidance mainly concerns about the predictive models and algorithms, without much regard to the route guidance problem (16).

Input traffic data of the prediction model is also an issue for an application running in a vehicle on the road, which has access only to a limited data source. Generally, this is historical link travel time data stored in the navigation unit and current traffic condition from the TMC/TPEG data (event-based data). So the modelling approach developed for navigation systems should be tailored to these constraints.

### 2.2 Duration modelling

In the context of route guidance, it is crucial to consider what the most useful predictive information for the driver is. Nowadays, most of the drivers are able to access real-time traffic information, which provide information about current incidents in his route of travel. The next natural question is how long are the current incident is going to last, and will the incident end by the time the driver reaches the location of the incident. Duration modelling technique is a suitable candidate to provide this information.

The term 'duration modelling' often refers to as 'survival analysis'. It has been widely used in many fields, such as medical research, biometrics and mechanical engineering. The main aim of survival analysis is to represent the observed distribution of durations by the means of hazard models. In the transportation field, the applications of survival analysis and especially hazard models are employed in the accident analysis to explore the factors affecting the duration of accidents (17, 18), in traveller's activity patterns to analyze traveller's behaviour (19), and in automobile ownership to estimate the distribution of car ownership and the effects of characteristic of car and household (20). A thorough overview of existing transportation applications of duration models can be found in (21).

Using the survival analysis approach to predict the incident duration is not new in the literature, (22) has used survival analysis models to predict the urban traffic congestions for the application of traffic management. The results indicate the log-logistic functional form for incidents duration fit the data well and meaningful results are obtained. Other study has explored the different incident duration prediction models and concludes the best performance can be achieved by applying the different models at the different incident duration range (23). This study is for the application of incident management process.

However, there isn't any literature found so far that using the prediction of incidents duration for the in-vehicle route guidance application. This study will use the incidents data collected for the London network and apply the duration modelling technique to predict the incident duration. The source of the data is the same as that delivered to the driver and stored in the navigation unit, so that the model developed here can be later used for the application on the in-vehicle route navigation system.

## 3. METHODOLOGY

In developing the duration analysis model for our study, the variable of interest is the length of time that elapsed from the beginning of the incident until its end. The likelihood that the incident will end in the next time period given that it has lasted for so long can be expressed as a probability function

*h(t)*. The probability of incident continues to last *f(t)* mainly dependent on the time elapsed since the incident starts and will presumably decrease over time. Apart from the observed incident duration, other factors, such as incident characteristics and weather conditions, will also have an impact on the incident duration.

Before describing the actual models developed it is useful to describe the mathematical functions representing the probability distribution of time *T*.

*T* has a probability distribution *f(t)* and a corresponding cumulative density function *F(t)*. The probability distribution of *T* can be represented in a number of ways, among which the survival and hazard functions are the most useful. The survival function of incident duration is given by:

$$\boldsymbol{S(t) = Pr[T \geq t] = 1 - F(t), 0 < t < \infty} \quad (1)$$

Where *S(t)* is the probability that a duration of incident will be greater than or equal to some specified time *t*. That is, the probability that a congestion will continues ('survive') until time t. The survival function is monotone decreasing over *[0, ∞)*, with *S(0) = 1* and *limit*→∞*S(t)=0*. The probability density function of *T* is:

$$\boldsymbol{f(t) = lim_{\Delta t \to 0} \frac{P(t \leq T < t+\Delta t)}{\Delta t} = \frac{dF(t)}{dt}} \quad (2)$$

*F(t)* is corresponding cumulative density function and with $\int_0^\infty f(t)dt = 1$.

The most common function representing the distribution of duration is the hazard function *h(t)* which is essential for the further modelling process. All the three functions described above can be associated with the hazard function. The hazard function defined as:

$$\boldsymbol{h(t) = f(t)/(1 - F(t)) = f(t)/S(t)} \quad (3)$$

Where *h(t)* specifies the instantaneous failure rate at *T=t* conditional upon survival time. Put another way, it represents the rate at which congestion ends after it has lasted up to time t. Note that, in contrast to the survivor function, which focuses on the congestion will continue, the hazard function focuses on the congestion will end given it has lasted for time t. This conditional probability is extremely relevant to our study, because it provides the information on duration dependence (i.e. the probability of a duration ending is dependent on the length of time the duration has lasted).

Relationship among hazard, survival, density and cumulative functions are illustrated in figure 1:

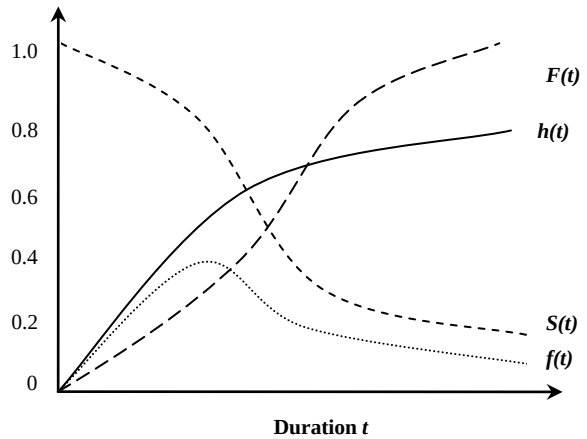



**Figure 1:** illustration of hazard (*h(t)*), survival (*S(t)*), density (*f(t)*) and cumulative distribution (*F(t)*) (21)

Using this hazard function, the effects of incident characteristics, also called covariates in the survival analysis terminology, can be introduced. The hazard function adopted in this study is shown below:

$$\boldsymbol{h(t|z) = h_0(t) exp\,(-\beta z)} \qquad \textit{(4)}$$

Where $\boldsymbol{z}$ is a vector of covariates and $\beta$ is the model parameter.

The function above can be viewed as the product of two functions. The function, $h_0(t)$, characterizes how the hazard function changes as a function of survival time. The other function, $exp(-\beta z)$, characterizes how the hazard function changes as a function of subject covariates. We note that $h_0(t)$ is the hazard function when $exp(-\beta z)=1$, therefore $h_0(t)$ is frequently referred to as baseline hazard function.

Baseline hazard function also can be shown by calculating the logarithm of both sides of the equation, and then one obtains:

$$\boldsymbol{log\, h(t|z) = \alpha(t) + \beta_1 x_1 + \cdots + \beta_k x_k} \quad \textit{(5)}$$

$\alpha\,(t)$ is the log of baseline hazard.

Parametric distributions are often used to implement the hazard model formulation introduced above. Typical duration distribution alternatives are: exponential, Weibull, lognormal and log-logistic. Choice of the distribution depends on the problem in question and replies on the knowledge of the data we have.

Different parametric distributions imply restrictive duration-dependence forms so it is important to select the correct distribution. Our approach here is to apply a generalized parametric form that includes specific distributions as a special case. The best-fitting model is chosen according to the Akaike Information Criterion (AIC) (24) and log likelihood value at the model convergence.

## 4. DATA DESCRIPTION

A new source of data called Transport Protocol Expert Group (TPEG) is used in this study. TPEG (Transport Protocol Expert Group) is a new standard format for delivering real-time traffic information to drivers over digital radio channels. The BBC has launched a pilot service that delivers a live feed of incident, congestion and roadwork information in tpegML format through their travel news website. TPEG is thought to be a future replacement for the TMC message broadcasting on FM (non-digital) radio. The content of the information will be exactly same as the one currently delivered into vehicle navigation system via the TMC. The quality and provenance of the TPEG data has been investigated and results can be found in (25). A computer program is set up to download the tpegML feed continuously from BBC travel news website and need for the offline analysis.

The data further filtered for the Greater London area, the region within the orbital motorway M25 outside the London, figure 2. They are referred to one year worth data, ranging from 01, May, 2007 to 30, April, 2008. The data later split into two parts for later analysis, the first 9 months for the model training and the rest for prediction. All the records in the database contain at least the following information: the starting and ending of the incident (duration of incident), the type of incident (accident, vehicle breakdowns, vehicle fires, and roadwork etc), and the location of the incident. One thing worth noting is that incident starting and end time continuously being updated as the time moving on, the closer to the finish of the incident the more accurate the end time of the incident, therefore, the true duration of the incident is calculated based on the last update of the incident. For the different type of incident, roadwork incidents are furthered filtered out from our database, they are not included in our further analysis, the reason has been explained earlier in the introduction, because the nature of these incidents are mostly planned and can be known by driver in advance.

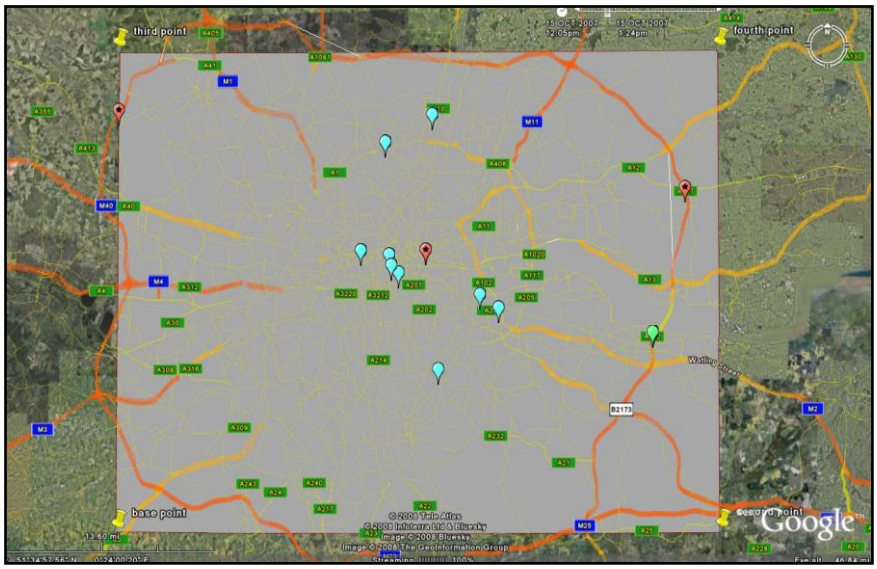



**Figure 3**: The coverage area of the data used for this study (Source: Google Earth)

Table 1 displays the summary statistics for different types of incidents in the training dataset with their duration. As it can be seen, there are 8 types of incidents are included in the database. The majority of the incidents are the accidents, which consist of more than half of the total incidents. The mean duration varies significantly among different incidents, from mean duration 108 minutes of broken-down vehicle to mean duration 8322 minutes for temporary traffic light. The first five types of incidents are having a mean duration around 180 minutes. The standard deviation also varies accordingly, generally the longer the mean duration the higher the standard deviation. From comparing the median and mean of the duration, it can be seen that the distribution of duration of the incidents are all right skewed.

**TABLE 1:** summary statistics for incidents

| Incident type | N | Mean(duration in mins) | Sd(duration in mins) | Med(duration in mins) |
|---|---|---|---|---|
| Accidents | 3,303 | 125.2579 | 99.43434 | 94 |
| Broken-down vehicles | 1,628 | 107.5412 | 66.82669 | 85 |
| Vehicle fire | 110 | 152.5 | 189.9609 | 103 |
| Flooding | 33 | 193.1515 | 109.0164 | 193 |
| Fuel Spillage | 26 | 152.8077 | 138.1092 | 96.5 |
| Gas leak | 33 | 8322.545 | 10423.83 | 4130 |
| Traffic light failure | 89 | 939.2697 | 1724.821 | 336 |
| Collapsed manhole | 28 | 443.2143 | 612.3703 | 263.5 |
| Total | 5250 | 187.919 | 1073.862 | 93 |

The characteristic of each incident are captured by following explanatory variables (also referred to as covariates in duration modelling): Severity factor, network performance, incident type. There are 4 categories for the severity factor: slight, medium, severe, very severe. Most of incidents have medium severity factor. For the network performance, there are 7 categories. Most of incidents have the network performance of ‘queuing traffic’. However, there are quite a lot of incidents not

specified with network performance. The types of incidents, as one of the major factors that affect the incident duration, are also used as explanatory variable in this study.

**TABLE 2:** Model covariates: severity factor (a), Network performance (b)

| (a) Severity factor | N | Mean(duration in mins) |
|---|---|---|
| Slight | 2 | 59 |
| Medium | 3,211 | 222.1529 |
| Severe | 1,909 | 128.7548 |
| Very severe | 128 | 213.5234 |

| (b) Network Performance | N | Mean(duration in mins) |
|---|---|---|
| Not specified | 2,291 | 263.7254 |
| Free flowing traffic | 11 | 158 |
| Slow traffic | 1,093 | 128.8829 |
| Heavy traffic | 143 | 141.5664 |
| Sheer weight of traffic | 2 | 106.5 |
| Queuing traffic | 1,705 | 128.3935 |
| Stationary traffic | 5 | 81 |

## 5. MODELLING RESULTS

Using the incident duration data and analysis methodology described in the previous sections, modelling results are presented here. The result of model functional form is presented first and then followed by choice of model explanatory variables, finally the model estimation results is presented.

### 5.1 Model functional form

Four different parametric distribution forms are fitted with the data. The criteria to select the best model form are according to the Akaike Information Criterion and log likelihood value. The preferred model is the one with lowest value of AIC and the largest log likelihood at model convergence.

According to table 3, the Log-logistic function provides the best fitting of the model. Therefore it is chosen as our model functional form for analyzing incidents duration data. Interesting about the log-logistic regression is it relaxes the monotonicity restriction of hazard function. The hazard function of the log-logistic specified as:

$$\boldsymbol{h(t)} = \frac{\boldsymbol{\lambda P(\lambda t)^{P-1}}}{\boldsymbol{1+(\lambda t)^{P}}} \qquad (6)$$

Where $\lambda$ and $P$ are distribution parameters, if P > 1, the hazard increases from zero to a maximum value and decreases toward zero thereafter, if P ≤1, the hazard is monotone decreasing.

**TABLE 3:** log likelihood and AIC values for different parameter distribution

| Model | obs | Log likelihood | AIC | df |
|---|---|---|---|---|
| Exponential | 5250 | -6274.206 | 12586.41 | 19 |
| Weibull | 5250 | -5500.433 | 11040.87 | 20 |
| Lognormal | 5250 | -4755.661 | 9551.321 | 20 |
| Loglogistic | 5250 | **-4706.9809*** | **9453.962*** | 20 |

## 5.2 Model explanatory variables

There are various factors affects the incident duration, such as incident severity, incident type, network performance and incident occurring during the time of the day, day of the week effects. The choice of variables for potential inclusion in the model is guided by the previous empirical work (26, 22) and checked by the coefficient's level of significance after the model being fitted. The procedure is as follows: The fully saturated model with all the explanatory variables included in the model is first fitted and identify those variables are not significant in the model according to its *P* value, and then fitting new model while excluding these insignificant variables, finally choose the variables that can best explain the data. The AIC criterion is used in selecting the model covariates. Table 4 listed all the covariates are selected in the final model. To be noted that the network performance variable is excluded during the selection process, because it produces a poor AIC value thus suggests including this variable doesn't explain the data well. In another word, the network performance doesn't affect the duration of the incident.

**TABLE 4:** Covariates included in the final model

| Temporal Characteristics | Incident severity | Incident type |
|---|---|---|
| Peak hour/non peak hour | Slight<br>Medium<br>Severe<br>Very severe | Accident<br>Broken-down vehicle<br>Fire<br>Flooding<br>Fuel spillage |

## 5.3 Model estimation results

The final model estimation results are displayed in following table, in interpreting the signs of the coefficients presented in this table, a negative sign increases the hazard (because of the use of $-\beta$ in equation) and thus the decreases duration and a positive sign decreases the hazard and increases duration.

During the model estimation, only five incident types are included in the final model. The other three types of incidents: gas leak, traffic light failure and collapsed manhole, are excluded from the model, because they are having significantly higher mean and standard errors comparing with the rest, according to Table 1. Inclusion of these variables will introduce a large variance and result in poor model fitting.

**TABLE 5:** model estimation results

| Variables | β | SEβ | Z | P | 95% confidence interval | |
|---|---|---|---|---|---|---|
| **Temporal characteristics** | | | | | | |
| Peak time (1=peak, 0=no-peak) | -0.04343 | 0.016673 | -2.61 | 0.009 | -0.07611 | -0.01075 |
| **Incident Severity** | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Slight (1=slight, 0=not) | -0.99927 | 0.348495 | -2.87 | 0.004 | -1.68231 | -0.31624 |
| Medium (1=medium, 0=not) | -0.55853 | 0.058407 | -9.56 | 0 | -0.67301 | -0.44406 |
| Severe(1=severe, 0=not) | -0.44024 | 0.058907 | -7.47 | 0 | -0.5557 | -0.32479 |
| Very severe | - | - | - | - | - | - |
| **Incident type** | | | | | | |
| Accident (1=accident, 0=others) | -1.64413 | 0.07425 | -22.14 | 0 | -1.78965 | -1.4986 |
| Broken-down vehicle(1=broken down vehicle, 0=not) | -1.72269 | 0.074887 | -23 | 0 | -1.86947 | -1.57592 |
| Vehicle fire(1= fire, 0=not) | -1.64211 | 0.094245 | -17.42 | 0 | -1.82682 | -1.45739 |
| Flooding(1=flooding, 0=not) | -1.09398 | 0.128265 | -8.53 | 0 | -1.34537 | -0.84258 |
| Fuel spillage | - | - | - | - | - | - |
| **Constant** | 6.76437 | 0.093703 | 72.19 | 0 | 6.580715 | 6.948025 |
| Number of observations | 5250 | | | | | |
| **Model structure parameter** | | | | | | |
| Gamma | 0.343624 | 0.003945 | | | 0.335979 | 0.351444 |
| **Overall model evaluation** | | | | | P | |
| Likelihood value | | | | -4875.9837 | 0.000 | |

From table 5, the coefficient on the temporal variable suggests that incident appears to last longer if it occurs during the peak hour. Incident types are the most influencing factors on the incident duration with largest absolute coefficient values. It indicates that if it's a broken down vehicle or accident or vehicle fire, the incidents appear to last longer.

### 5.4 Model prediction performance

After the model has been fitted with first 9 months data, the prediction performance of the model is checked using the later 3 months data. The results of the prediction of the model are compared with the naïve estimators' prediction results, mean and median listed in the Table 1 are chosen as naïve predictors corresponding for each different type of incidents. The first five types of incidents are calculated. The results are showing in table 6. The following statistics: Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are used.

**TABLE 6:** performance of the model

| **Model** | **N** | **MPE** | **MAPE** | **RMSE** |
|---|---|---|---|---|
| Naïve predictor (mean) | 1525 | 39% | 65% | 180 |
| Naïve predictor (median) | 1525 | 5% | 46% | 186 |
| Log-logistic hazard model | 1525 | -19% | 41% | 191 |

From table 6, for MPE, the median naïve predictor produces the lowest value of 5% comparing 39% from mean naïve predictor. It is not surprising the median give the best results for the MPE, because median is generally lesser biased estimator. The results from the MAPE are more robust. For MAPE, the best results are produced by log-logistic hazard model with a value of 41%, comparing with MAPE from mean Naïve predictor 65%. A large improvement is achieved by Log-logistic hazard model. However, for the RMSE value, the advantage of using Log-logistic hazard model isn't very obvious.

## 6. CONCLUSION

In this paper survival analysis technique is used to predict the duration of incidents. One-year worth of incidents data are used to fit the parametric duration model. Log-logistic model is finally chosen among the others as our candidate model, as it provides the best AIC. Factors affect the incident duration are explored, it appears the incident type is the most influencing factor on the incident duration. Other factors are the incident severity and time of the day effect. The prediction performance is examined and compared with that from naïve predictors. It suggests the proposed model gives more accurate prediction results.

In terms of future work, there is a need to further examine the appropriateness of using log-logistic for model form. The assumption of log-logistic distribution of incident duration should be further verified. Even the best model using all the available explanatory information has poor predictive performance. Hence, future research should focus on identifying other factors that has an effect on incident duration. The generators of TMC/TPEG data should incorporate these new elements of information in their data stream to enable more accurate incident duration prediction by in-vehicle navigation systems

## REFERENCES


(1) Ben-Akiva, M., M. Bierlaire, H. N. Koutsopoulos, and R. Mishalani. Dynamit: A Simulation-Based System for Traffic Prediction. Delft, The Netherlands, 1998.

(2) Wunderlich, K. E., D. E. Kaufman, and R. L. Smith. Link Travel Time Prediction for Decentralized Route Guidance Architectures. The University of Michigan ITS Research Centre, 1998.

(3) Hounsell, N. B., and S. Ishtiaq. Journey Time Forecasting for Dynamic Route Guidance Systems in Incident Conditions. *International Journal of Forecasting*, Vol. 13, 1997, pp. 33-42.

(4) Nihan, N. L., and K. O. Holmesland. Use of the Box and Jenkins Time Series Technique in Traffic Forecasting. *Tranportation*, Vol. 9, 1980, pp. 125-43.

(5) Kamarianakis, Y., and P. Prastacos. Space-Time Modeling of Traffic Flow. *Computer & Geosciences*, Vol. 31, 2005, pp. 119-33.

(6) Williams, Billy M., and Lester A. Hoel. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal Arima Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering*, Vol. 129, no. 6, 2003, pp. 664-72.

(7) Hamed, Mohammad M., Hashem R. Al-Masaeid, and Zahi M. Bani Said. Short-Term Prediction of Traffic Volume in Urban Arterials. *Journal of Transportation Engineering*, Vol. 121, no. 3, 1995, pp. 249-54.

(8) van Lint, H. *Reliable Travel Time Prediction for Freeways*. Delft University of Technology, Netherland, 2004.

(9) van Hinsbergen, C., and J. van Lint. Bayesian Combination of Travel Time Prediction Models. In *Transportation Research Record: Journal of the Transportation Research Board*, Washington, D.C., 2008, pp. 73-80.

(10) Horvitz, E., J. Apacible, R. Sarin, and Lin Liao. Prediction, Expectation, and Surprise: Methods, Designs, and Study of a Deployed Traffic Forecasting Service. In *Twenty-First Conference on Uncertainty in Artificial Intelligence*, Edinburgh, Scotland, 2005 2005.

(11) krishnan, R., and J.W. Polak. Short-Term Travel Time Prediction: An Overview of Methods and Recurring Themes. In *Proceedings of the Transportation Planning and Implementation Methodologies for Developing Countries Conference (TPMDC 2008)*, Mumbai, India, 2008.

(12) Van Der Voort, Mascha, Mark Dougherty, and Susan Watson. Combining Kohonen Maps with Arima Time Series Models to Forecast Traffic Flow. *Transportation Research Part C: Emerging Technologies*, Vol. 4, no. 5, 1996, pp. 307-18.

(13) Vlahogianni, Eleni I., Matthew G. Karlaftis, and John C. Golias. Spatio-Temporal Short-Term Urban Traffic Volume Forecasting Using Genetically Optimized Modular Networks. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 22, no. 5, 2007, pp. 317-25.

(14) Yin, Hongbin, S. C. Wong, Jianmin Xu, and C. K. Wong. Urban Traffic Flow Prediction Using a Fuzzy-Neural Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 10, no. 2, 2002, pp. 85-98.

(15) INRIX, inc. http://www.inrix.com/. Accessed 25, July, 2010.

(16) Schmitt, E. J., and H. Jula. Vehicle Route Guidance Systems: Classification and Comparison. In *Proceedings of the IEEE ITSC 2006*, Toronto, Canada, 2006 2006.

(17) Jones, Bryan, Lester Janssen, and Fred Mannering. Analysis of the Frequency and Duration of Freeway Accidents in Seattle. *Accident Analysis & Prevention*, Vol. 23, no. 4, 1991, pp. 239-55.

(18) Golob, Thomas F., Wilfred W. Recker, and John D. Leonard. An Analysis of the Severity and Incident Duration of Truck-Involved Freeway Accidents. *Accident Analysis & Prevention*, Vol. 19, no. 5, 1987, pp. 375-95.

(19) Schönfelder, S., and K. W. Axhausen. Analysing the Rhythms of Travel Using Survival Analysis. In *Annual Meeting of the Transportation Research Board* Washington, D.C., 2001.

(20) Gilbert, Carol C. S. A Duration Model of Automobile Ownership. *Transportation Research Part B: Methodological*, Vol. 26, no. 2, 1992, pp. 97-114.

(21) Hensher, David A., and Fred L. Mannering. Hazard-Based Duration Models and Their Application to Transport Analysis. *Transport Reviews: A Transnational Transdisciplinary Journal*, Vol. 14, no. 1, 1994, pp. 63 - 82.

(22) Stathopoulos, A., and M. G. Karlaftis. Modeling Duration of Urban Traffic Congestion. *Journal of Transportation Engineering-Asce*, Vol. 128, no. 6, 2002, pp. 587-90.

(23) Valenti, Gaetano, Maria Lelli, and Domenico Cucina. A Comparative Study of Models for the Incident Duration Prediction. *European Transport Research Review*, Vol. 2, no. 2 pp. 103-11.

(24) Akaike, H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, Vol. 19, no. 6, 1974, pp. 716-23.

(25) Hu, Jun, Rajesh Krishnan, and Michael G. H. Bell. Tpeg Feed from the Bbc: A Potential Source of Its Data? In *Road Transport Information and Control - RTIC 2008 and ITS United Kingdom Members' Conference, IET*, Manchester, UK, 2008, pp. 1-11.

(26) Nam, Doohee, and Fred Mannering. An Exploratory Hazard-Based Analysis of Highway Incident Duration. *Transportation Research Part A: Policy and Practice*, Vol. 34, no. 2, 2000, pp. 85-102.