


Modelling traffic accident duration on urban roads with high traffic variability using survival models: a case study on Fortaleza arterial roads

Modelagem da duração de acidentes de trânsito em vias urbanas com alta variabilidade de tráfego usando modelos de sobrevivência: estudo de caso das vias arteriais de Fortaleza

Vandeyberg Nogueira de Souza¹, Francisco Moraes de Oliveira Neto¹

¹Universidade Federal do Ceará, Fortaleza, Ceará, Brasil

Contat: vandeyberg.souza@gmail.com (VNS); moraes@det.ufc.br,  (FMON)

Submitted:

13 September, 2022

Accepted for publication:

25 May, 2023

Published:

11 July, 2023

Associate Editor:

Sara Ferreira
Universidade do Porto,
Portugal

Keywords:

Traffic accidents.
Non-recurring congestion.
Survival analysis.

Palavras-chave:

Acidentes de trânsito.
Congestionamento não-recorrente.
Análise de sobrevivência.

DOI: 10.58922/transportes.v31i2.2837



ABSTRACT

Unexpected congestions are a common problem in the lives of urban citizens who need to travel to carry out their activities. This type of congestion causes unexpected delays to drivers and has traffic accidents and their duration as the main factor for their formation. To contribute to this problem, this study aimed to analyze the duration of traffic accidents on arterial roads of Fortaleza, Brazil, and its potential explanatory factors. The duration of accidents was estimated based on traffic data obtained from electronic surveillance equipment, as the accident databases did not have this information. For this purpose, we generated profiles of speed and flow proportion per lane for days with accident and typical days to differentiate the impact on traffic caused by an accident from a typical traffic variability. The method detected the duration of 316 accidents with an average duration of 71 minutes and a standard deviation of 43 minutes. Next, a set of suggested hypotheses to explain the variability of accident duration was analyzed using survival models. The calibrated model showed that the severity of the accident, the traffic conditions at the accident location, the quantity and scheduling of the traffic agents, and the number of vehicles involved can have a significant impact on accident duration.

RESUMO

Congestionamentos inesperados são um problema frequente na vida de habitantes de grandes centros urbanos que precisam se deslocar para realizar suas atividades. Esse tipo de congestionamento causa atrasos imprevistos aos motoristas e tem os acidentes de trânsito e a sua duração como principal fator para sua formação. Visando contribuir com essa problemática, este estudo teve como objetivo analisar a duração de acidentes de trânsito ocorridos em vias arteriais de Fortaleza, Brasil, e a sua relação com seus fatores causais. Uma vez que não existem dados disponíveis com a duração desses acidentes, foram utilizados dados de tráfego, obtidos de equipamentos de fiscalização eletrônica, para extrair essa informação de forma indireta. Para isso, foram gerados

perfis da velocidade e proporção do fluxo veicular por faixa no dia do acidente e em dias típicos para diferenciar o impacto causado no tráfego por um acidente de trânsito de suas variações normais. Ao final, foi obtida uma amostra de 316 acidentes com uma duração média de 71 minutos e um desvio padrão de 43 minutos. Em seguida, foram levantadas hipóteses sobre potenciais fatores causais da variabilidade observada na duração dos acidentes, as quais foram testadas usando modelos de sobrevivência. O modelo calibrado mostrou que a severidade do acidente, as condições de tráfego no local do acidente, a quantidade e programação dos agentes de trânsito e o número de veículos envolvidos no acidente possuem uma influência significativa na duração dos acidentes.

1. INTRODUCTION

Unexpected congestion is often faced by urban people who need to travel to perform their activities, such as work and study. Known in the literature as non-recurrent, this type of congestion causes unexpected delays in travel times, increases fuel consumption, the level of air pollution and negatively impacts the economy and quality of life of the population. Non-recurrent congestions are triggered by random events that temporarily reduce the capacity of the road and can represent up to 30% of the total delay observed on highways (Skabardonis, Varaiya and Petty, 2003). Typically, these events involve incidents that include traffic accidents, broken vehicles on the road, adverse weather conditions, work zones and other random events.

To reduce non-recurrent congestion, several prior studies have sought to understand which factors contribute to the variability observed in the duration of traffic incidents or, more specifically, traffic accidents, which account for most traffic incidents. The biggest challenge in these studies is the lack of reliable and complete data. For example, in some locations, some incidents characteristics are not registered or even the duration data may not be recorded by the traffic agency responsible for managing incidents. To overcome this challenge, the main approaches used in previous studies is to obtain the duration of incidents by observing anomalies in traffic variables such as speed and flow (Haule et al., 2018; Hojati et al., 2014). However, in general, these methods were developed for freeways or highways and may not be suitable for arterial roads since speed and flow are recurrently affected by traffic lights, crosswalks, intersections and land use in this type of roads.

As highlighted by Li, Pereira and Ben-Akiva (2018), previous studies focused on modeling the duration of traffic incidents that occurred on some segments of freeways between cities or urbanized regions with few studies focused on analyzing the duration of incidents that occurred on arterial roads. In this regard, there is a lack of knowledge about incidents duration behavior in urban arterial roads and whether it is impacted by different factors than those studied in freeways such as zone of the city where the incident occurs.

To contribute to these issues, this study aimed to analyze the duration of traffic accidents on arterial roads and its potential explanatory factors. It is worth mentioning that while previous studies have primarily investigated traffic incidents, our study specially focused on traffic accidents due to data limitations. Nonetheless, the approach used in this study can be adapted for analyzing both incidents and accidents. A method to estimate accidents duration from traffic data was proposed based on methods presented in previous works for freeways such as in Hojati et al. (2014). Next, an initial exploratory analysis

was undertaken to understand accident duration behavior and to support the hypotheses raised about factors that contribute to the observed variability in accident duration. To test the hypotheses a survival model was developed. The proposed methodology was applied to a database of accidents that occurred between 2015 and 2017 in the city of Fortaleza, which is in the northeast of Brazil and had a population of approximately 2.7 million people and a fleet of 1.2 million vehicles in 2020 (IBGE, 2020).

2. LITERATURE REVIEW

2.1. Incident duration: definitions and methods of estimation

In general, traffic incident duration is defined as the elapsed time from the occurrence of an incident until the clearance of the road (Khattak et al., 2016). This period consists of four phases: incident detection/reporting time, preparation/dispatching time, travel time and clearance/treatment time (Li, Pereira and Ben-Akiva, 2018). However, some authors have analyzed only some of these phases (Hou et al., 2014; Li, 2015). While others considered a fifth phase - the time elapsed between incident clearance and non-recurrent congestion clearance - called the recovery phase (Haule et al., 2018; Hojati et al., 2014).

The different definitions for the duration of traffic incidents are, in part, due to the limitations in the data availability. In general, the source, size and quality of the available data vary from study to study (Li, Pereira and Ben-Akiva, 2018). To overcome the lack of information of incident duration, previous studies proposed to use data from other sources, such as traffic data recorded by electronic surveillance equipment. The duration of incidents is then estimated by detecting anomalies in traffic conditions close to the location and time reported in the incident database (Haule et al., 2018; Hojati et al., 2014). However, in general, methods for estimating the duration of incident from traffic data have been developed and applied strictly to incident occurring on uninterrupted traffic facilities - highways or freeways (Li, Pereira and Ben-Akiva, 2018). Since arterial roads have interrupted flow, these methods need to be adjusted to consider the extra variability in traffic conditions due to the existence of traffic lights and crosswalks, for example.

2.2. Factors affecting incident duration

The results presented in previous works showed that the duration of incidents can have a great variability due to the influence of several factors, such as: incident characteristics (severity, type, location, number of vehicles involved), weather conditions (rain, snow, fog), road characteristics (type of layout, existence of vertical and horizontal signs, shoulder), traffic conditions (volume, speed, density) and time factors (day, time, month) (Hou et al., 2014; Hojati et al., 2013).

According to Hojati et al. (2014), the set of factors that influence the duration of incident varies from place to place due to differences in some attributes, such as: traffic conditions, geographical and social characteristics. In addition, the availability and quality of data also vary greatly which makes it difficult to use studies developed in other locations to guide local interventions. This scenario is even worse for incident happened in arterial roads (or urban streets) and for developing countries due to the lack of studies in literature for these cases (Li, Pereira and Ben-Akiva, 2018).

2.3. Modelling of incident duration

Over the past few decades, several models have been developed to analyze the duration of traffic incidents and their relationship with their explanatory factors, especially in freeways (Hojati et al., 2014). Among these are: ANOVA tests (Hojati et al., 2012), Linear Regression Models (Valenti, Lelli and Cucina, 2010; Zhang and Khattak, 2010), Structural Equation Models (Lee, Chung and Son, 2010), Spatial Models (Xie et al., 2015) and Survival or Hazard-Based Models with parametric (Junhua, Haozhe and Shi, 2013; Shi, Zhang and Liu, 2015) or semi-parametric formulation (Hou et al., 2014; Shi, Zhang and Liu, 2015).

Although classical linear regression models provide results that are easier to interpret and understand when compared to survival models (Khattak, Schofer and Wang, 1995), the latter have been applied in several studies on the analysis of the duration of incidents in recent years (Li, Pereira and Ben-Akiva, 2018). This occurs due to the fact that survival models is suitable for data analysis for which the outcome variable of interest is time until an event occurs such as the time until a road clearance after an incident (Kleinbaum and Klein, 2012). This length of time is assumed to be a continuous random variable, T , with a cumulative distribution function, $F(t)$, and probability density function, $f(t)$. $F(t)$ is also known in literature as the failure function and gives the probability of having an event before some time t . Conversely, the survival function, $S(t)$, is the probability of the duration being greater than some specific time t (Hojati et al., 2013). The instantaneous potential of an incident end, per unit of time, given that the incident lasted until time t is given by the hazard function $h(t)$. Equation 1 shows the mathematical relationship between $F(t)$, $f(t)$, $S(t)$ e $h(t)$.

$$h(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{S(t)} \quad (1)$$

Where $h(t)$: hazard function;

$S(t)$: survival function;

$f(t)$: probability density function;

$F(t)$: Failure function.

In addition, survival models allow to study the effect of explanatory variables on the probability of an incident lasting more than a specific time t (Nam and Mannering, 2000). As can be seen in the literature, the most used survival models are: 1) proportional hazards, which assume that the factors act multiplicatively on some underlying hazard function; 2) Accelerated Failure Time models (AFT), which assume that the covariates rescale (accelerate) time directly in a baseline survivor function (Washington, Karlaftis and Mannering, 2003). The choice among proportional hazard and AFT models relies on knowledge about underlying distribution: proportional hazard is a preferred choice when underlying distribution is unknown, whereas AFT models are more suitable when underlying distribution is known or are theoretically justified (Lee, Chung and Son, 2010). The AFT models have been chosen in most previous works to model the duration of incidents due to their accelerate assumption and the availability of large databases that allow the estimation of underlying distribution.

According to Washington et al. (2003), the efficiency of AFT models in estimating parameters depends on the proper choice of a known probability distribution for the survival time. In this regard, the results obtained in previous studies show that the main distributions to

model the survival time variable are: Weibull which allows for positive duration dependence (the probability of the incident ending increase over time), negative duration dependence and no duration dependence (Hojati et al., 2013; Washington et al., 2003)); Log-normal and Log-logistic which allows for nonmonotonic hazard function where the hazard increases from zero to a maximum point and then decreases toward zero (Xie, Ozbay and Yang, 2015; Junhua, Haozhe and Shi, 2013, 2013; Nam and Mannering, 2000; Washington et al., 2003). These different distributions can result from different factors, including drivers' cultures and traffic laws as well as the size and quality of the available data (Li et al., 2017).

3. DATA DESCRIPTION

The accident data were obtained from the Fortaleza Traffic Accident Information System for January 2015 to August 2017. The database contains 52,503 records of traffic accidents which does not include vehicles broken down and other random incidents. However, only accidents that occurred on workdays, between 06:00 and 19:00, were considered, which corresponds to 30,014 accidents. The database includes information on the characteristics of accidents, such as: time reported by the traffic officer, geographic location, type and severity of the accident and number of vehicles involved. However, the database does not provide accurate information about accidents duration. The traffic agent only records the time in which he/she became aware of the accident, missing the information about each accident duration phase.

Since the accident database does not provide information about the duration of the accidents, a traffic database was used to estimate the accident duration from disturbances observed in traffic conditions. The database was provided by the Municipal Traffic and Citizenship Authority of Fortaleza and contains data collected by 251 loop-detectors. Each equipment records the date and time information for the passage of each vehicle, the lane used, the speed and the type of vehicle detected.

The traffic speed and flow have been used in several previous studies to estimate the duration of accidents that occurred on freeways or highways. However, on arterial roads these variables have greater variability due to the existence of intersections, traffic lights and crosswalks, which may limit the capacity of the proposed methods to detect disturbances caused on traffic by accidents. To deal with this variability, we proposed to analyze the distribution of traffic flow among lanes (fp_i) - which measures the proportion of the flow on a road, at a given section, that uses a lane i . We argue that fp_i is a more suitable variable to detect disruptions in arterial roads, since it is not so much affected by the traffic signal operations, crosswalks and entry or exit of vehicles on the road, while it is strongly impacted by events which close partially and temporarily the road.

4. METHODOLOGY TO ESTIMATE ACCIDENT DURATION

The accident duration was estimated by applying the methodology presented by Souza and Oliveira Neto (2020). Historical traffic data and accident details (including the location and time) were used to compare the traffic at the day of the accident with its expected pattern. The historical profiles of speed and flow proportion per lane were built using data from one month before and one month after the day of the accident, without considering atypical days

(i.e., Saturdays, Sundays or holidays) and days with the occurrence of non-recurrent events. This period was defined in such way that reduces the effect of traffic seasonality, but without compromising the sample size and, consequently, the statistical analyzes performed. The expected traffic profiles were defined based on the calculation of the 100 (1- α)% prediction interval (PI_t), at a moving time window t , for the speed (s) and the flow proportion (fp). The PI_t predicts, with a confidence level of 100 (1- α)%, the range in which a future individual observation (speed or flow proportion caused by a single accident) is likely to fall. For each variable, the PI_t was calculated around the average profile (\bar{x}_t), 10% trimmed to exclude days with unexpected events, such as traffic incidents, adverse weather and so on. The moving average technique was applied to smooth the generated profiles of the variables s and fp , since these variables have a high variability even in normal traffic conditions without the occurrence of accidents. The different phases of the accident were identified using the methodology proposed by Souza and Oliveira Neto (2020). The first step was to identify a minimum sequence of (m) aggregation intervals (t) in which the values of both variables (s) and (fp_i) on the day of the accident were outside the limits of the prediction interval defined by the historical profile. After identifying this period, the moment when the accident occurred was determined by analyzing the 15 minutes prior to the first aggregation interval (t_1) of the sequence of m intervals in which traffic conditions were outside the prediction interval. It was assumed that the accident occurred at the interval before t_1 where it was observed an abrupt change (d) in the distance between the daily and historical profiles. Finally, the moment of road clearance was identified by analyzing the aggregation intervals between t_1 and t_m to search for the moment when the traffic variables on the day of the accident start to return to the expected conditions.

Figure 1 shows the daily and historical profiles of variables s and fp for one day in which one of the accidents analyzed occurred on a road with two lanes. The profiles show that at 16:15 the traffic speed abruptly reduced at both two lanes and the majority of vehicles start using lane 1. This fact indicates that the accident occurred at this moment and obstructed lane 2. At 16:30, the traffic speed started to increase and the vehicles returned to use the lane 2, which indicates the moment that road was cleared. Lastly, the effect of the accident on the traffic conditions lasted until 16:50, when the daily profiles returned back to the expected conditions.

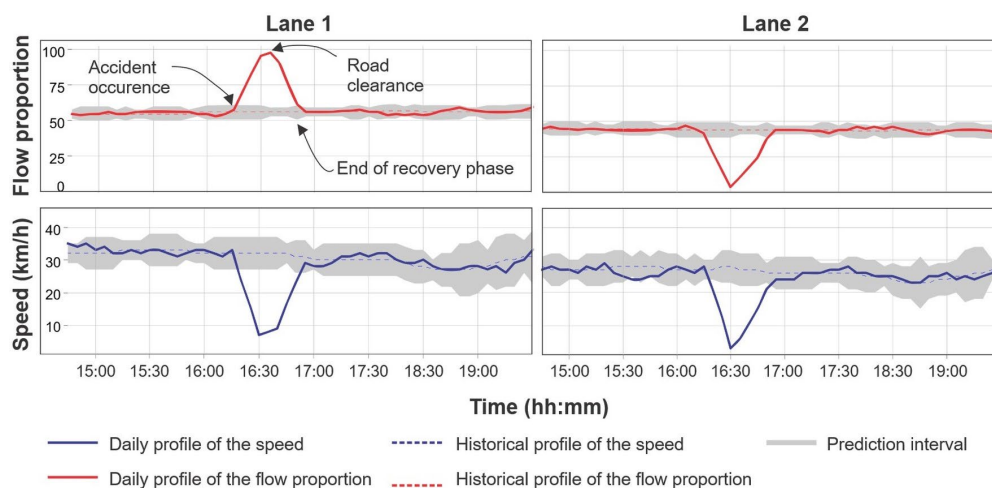


Figure 1. Profiles of speed and flow proportion by lane for estimation of accident duration.

5. EXPLORATORY ANALYSIS

Not all accidents in the database were suitable for estimation of accident duration. The number and location of the loop-detectors restricted the sample of accidents to be analyzed. Therefore, it was assumed that only accidents within 50 meters of a loop-detector were suitable for duration estimation, what reduced the sample to 1,785 accidents. Furthermore, in some cases the accident occurred in the opposite direction to that inspected by the loop-detector, reducing even more the sample size. The estimation of accident's duration was successful for 316 accidents, having a broad spatial coverage on the traffic network system of the city, as shown in Figure 2.

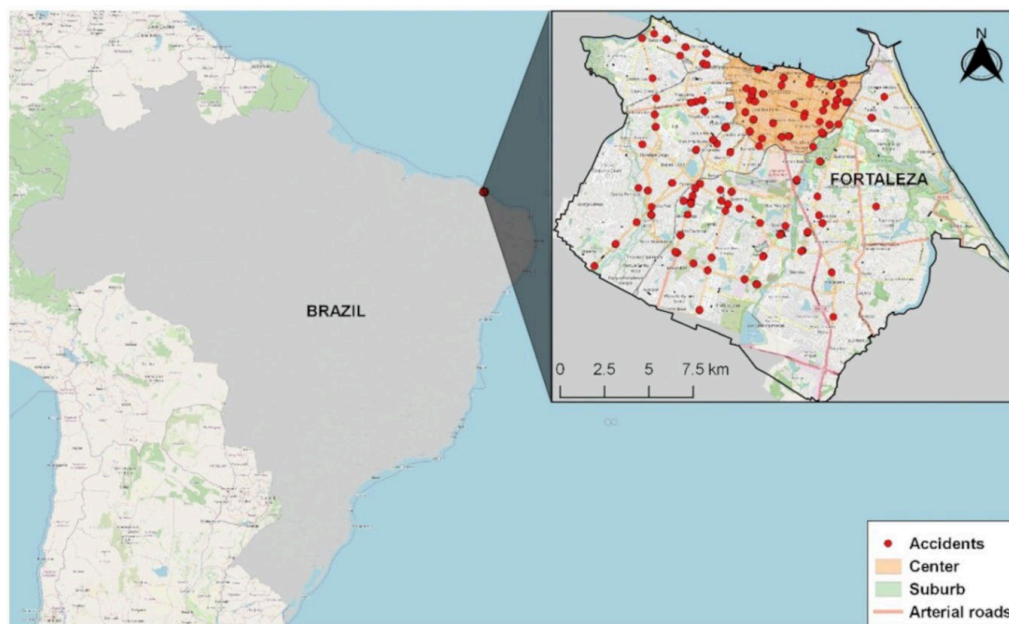


Figure 2. Spatial distribution of detected accidents.

The accidents had a minimum duration of 15 minutes and a maximum of 210 minutes. The average duration was 71 minutes with a standard deviation of 43 minutes, a skewness of 0.9 and a kurtosis of 0.32. This high variability makes it difficult to understand and predict the duration of accidents and their impact on traffic congestion without knowing how this variability can be influenced by factors that act to “stretching out” (or contracting) the accident duration. To analyze this influence, we used the AFT model to test the factors.

5.1. Factors affecting accident duration

The potential explanatory factors for the variability in the duration of accidents on arterial roads were defined based on factors studied in previous work for accidents that occurred on freeways, such as severity, accident type, period (peak or off-peak), number of vehicles involved, and involvement of large vehicles. In addition to these factors, the following hypotheses were taking into account, related to the specificities of arterial roads and the region of study, as suggested by Souza and Oliveira Neto (2020):

- I. Drivers involved in traffic accidents without injuries comply with the Brazilian Traffic Code and remove vehicles from the road immediately.

- II. The level of congestion on the roads close to the accident site is taken into account by traffic agents to define the priority of service.
- III. Accidents in the central region, the business center of the city where more traffic agents are assigned, have a shorter duration because it is required less travel time for traffic and rescue agents to arrive at accident site.
- IV. The duration of accidents is affected by the reason for drivers to travel, since drivers traveling for work reasons are less tolerant of unexpected delays and tend to solve the problem and clear the road more quickly.
- V. More experienced drivers tend to have a greater knowledge of traffic laws and the steps to be followed after an accident, solving the problem more quickly.

The hypothesis I could not be tested directly since the method to estimate accident duration from traffic profiles detected only accidents with at least 15 minutes. However, this hypothesis will be discussed further in the next section. Hypothesis II was tested by the V/C factor which measures the volume/capacity ratio of the road at the time of the accident. The hypothesis III was tested by the zone factor (dummy variable that indicate whether the accident occurred in the center or on the periphery of the city). The hypothesis IV could not be tested directly but it is correlated with the period factor, as drivers who travel for work/study reasons tend to travel at peak hours. Lastly, the hypothesis V could not be tested because information about drivers age was incomplete in the database.

The analysis of the existence of a statistically significant influence of the factors was carried out using the Kruskal-Wallis non-parametric test. The test has as its null hypothesis that the means of accident duration among different factor levels are the same. Table 1 presents the factors analyzed, their levels and the p-value for the Kruskal-Wallis test. Three factors showed statistical significance, at a 10% confidence level, as follows: Period, Zone and Number of vehicles involved. The other 4 factors tested did not show a significant impact on accident duration which means that some hypotheses raised were not supported by the data.

The results showed that accident type does not seem to impact its duration, what contradicts the expected behavior that drivers involved in minor rear-end collisions would quickly come to an agreement, removing the vehicles from the road, since the Brazilian traffic laws state that a driver of the car that hit a leading car is usually considered the faulty driver. Most surprised was the result that the severity of accident may not affect the duration of accidents. Two reasons may explain this last result: 1) the classification of accidents into only two levels of severity (i.e., accidents with or without injuries) in the dataset is not enough to represent all possible types of injuries in accidents, as for example, minor injuries may not affect accident duration, since the victims usually do not need any medical attention; 2) drivers usually do not follow the guidelines of the Brazilian Traffic Code and do not clear the road after an accident with non-injuries, what makes this type of accidents last longer than expected.

To further verify whether the type of classification on accident severity had impact on the analysis results, the difference in accident duration due to accident severity was tested controlling by the zone of the city (center or periphery). The main hypothesis of this analysis was that accidents on roads in the peripheral region tend to cause more serious

injuries due to the higher speeds developed on those roads that usually have less traffic. Therefore, by separating the data by the type of zone may reveal the effect of traffic severity. The Kruskal-Wallis test performed to the two data sets showed that severity had a significant influence on accidents that occurred in the periphery (p-value of 0.031) and there was no significant difference for accidents that occurred in the city center (p-value of 0.78). Thus, the severity factor was considered together with the factors zone, period, and number of vehicles to calibrate the AFT model in the next section and measure the impact of these factors on accident duration.

Table 1: Potential explanatory factors and their description

Factor	Levels	Value	Kruskal-Wallis test (p-value)
Period	2	0 – Peak period; 1 – Off-peak period	0.026
Zone	2	0 – Center; 1 - Periphery	0.005
V/C	3 (2 dummies)	00 - $V/C < 0.50$; 01 - $0.50 \leq V/C < 0.75$ 10 - $V/C \geq 0.75$	0.2
Severity	2	0 – No injuries; 1 – injuries	0.332
Accident type	2	0 – Rear-end collision 1 – Side-impact or sideswipe collision	0.769
Number of vehicles	2	0 – Up to two; 1 – Three or more	0.1
Large vehicles	2	0 – No; 1 - Yes	0.354

6. AFT MODEL

6.1. Accident duration adherence test for a known distribution

The AFT parametric models assume that the response variable (duration of accidents in this study) follows a known probability distribution (Lee, Chung and Son, 2010). As seen in section 2.3, in previous studies accident duration generally has better fit to Weibull, Lognormal or Loglogistic distributions (Hojati et al., 2013). The Weibull distribution has a positive/negative monotonic duration dependence (i.e., the probability of the duration ending increase/decreases over time). On the other hand, Loglogistic and Lognormal distributions allow for a nonmonotonic hazard function where probability of the duration ending increases from zero to an inflection point and decreases toward zero thereafter (Washington et al., 2003).

The fit of accident duration to Weibull, Lognormal and Loglogistic probabilities distributions was tested using the Anderson-Darling statistical test (A-D). We did not test the Gamma distribution, as suggested by Souza and Oliveira Neto (2020), due to its similarity to the Weibull distribution and because it is not well established in the literature for modelling accident duration (Valenti, Lelli and Cucina, 2010). The null hypothesis of the A-D test, that the variable accident duration follows a known distribution, was rejected at a 5% confidence level for the Lognormal and Loglogistic distributions. On the other hand, the null hypothesis was not rejected for the Weibull distribution, what is an indication that hypothesis I is not valid. Figure 3 shows the histogram of accidents duration and the adjusted distributions.

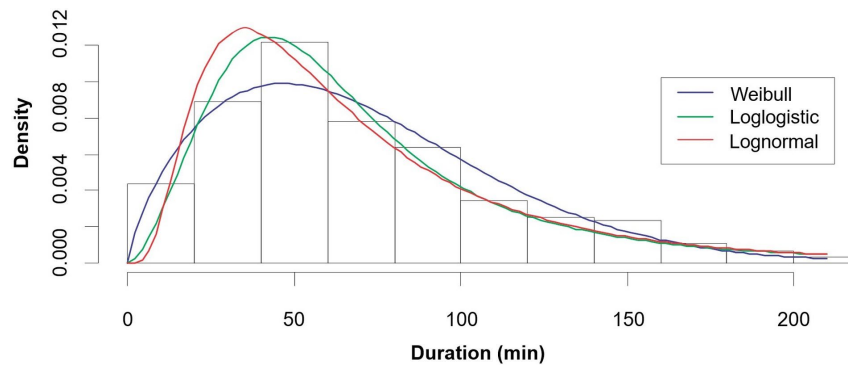


Figure 3. Histogram and fitting distributions for accident duration.

6.2. Model formulation

The survival and hazard function for the Weibull AFT model are showed in Equation 2 and Equation 3 respectively, where p and λ are the shape and scale parameters of Weibull distribution respectively. The parameter p defines whether the Weibull distribution has a positive/negative monotonic function: for $p > 1$ the function is monotonically positive, for $p < 1$ it is monotonically negative and for $p = 1$ it is constant (Washington et al., 2003). The effect of explanatory variables on the duration of accidents is usually added to Weibull AFT model by reparametrizing (λ) in terms of explanatory variables (Kleinbaum and Klein, 2012). Equation 4 shows the reparametrized λ as a function of the explanatory variables defined in the exploratory analysis, whereas Equation 5 shows the relationship between the duration of the accident and the explanatory variables.

$$S(t) = 1 - F(t) = e^{(-\lambda t^p)} \quad (2)$$

$$h(t) = \lambda p t^{p-1} \quad (3)$$

$$\frac{1}{\lambda^{\frac{1}{p}}} = e^{(\beta_0 + \beta_1 ZO + \beta_2 PE + \beta_3 SE + \beta_4 NV)} \quad (4)$$

$$t = \left\{ [-\ln [S(t)]]^{1/p} \right\} * \exp (\beta_0 + \beta_1 ZO + \beta_2 PE + \beta_3 SE + \beta_4 NV) \quad (5)$$

where $S(t)$: survival function;

$F(t)$: failure function;

$h(t)$: hazard function;

t : duration of the accident, in min.;

p : shape parameter of Weibull distribution;

λ : scale parameter of Weibull distribution;

ZO : zone;

PE : period;

SE : severity;

NV : number of vehicles.

The use of the AFT model has as implicit assumption that the survival function is constant for all observations, in other words, that any variation observed in the response

variable can be associated only to the vector of explanatory variables added to the model. However, other factors not added to the model may cause the observed variability for these accidents to be greater than expected. This problem is known in literature as unobserved heterogeneity and can result in inconsistent estimates of the model parameters (Washington et al., 2003).

A common approach to solve the problem of unobserved heterogeneity is to introduce a random component α (frailty) to capture variability due to unobservable effects, as showed in Equation 6 (Hojati et al., 2014). Accidents with $\alpha > 1$ have a decreased survival function and tend to be solved quickly. Similarly, accidents with $\alpha < 1$ have an increased survival function and tend to have longer duration. Since it is not possible to estimate α for each accident, it is assumed that the term α follows a distribution $g(\alpha)$ with a mean of α equal to 1 and variance θ . The unconditional survival function, which represents a population average, is found by integrating over the conditional survival function times $g(\alpha)$ with respect to α , as showed in Equation 7. The distribution $g(\alpha)$ is parameterized in term of the variance θ which is estimated from the data. If the value of θ is close to zero, it is assumed that there is no unobserved heterogeneity in the data. For more details see (Kleinbaum and Klein, 2012).

$$S(t|\alpha) = S(t)^\alpha \quad (6)$$

$$S_u(t) = \int_0^\infty S(t|\alpha) g(\alpha) d\alpha = \{1 - \theta \ln[S(t)]\}^{-\frac{1}{\theta}} \quad (7)$$

where α : random component (frailty);

t : accident duration, in min;

$S(t)$: survival function;

$S(t|\alpha)$: conditional survival function;

$S_u(t)$: unconditional survival function;

$g(\alpha)$: assumed distribution for α with mean 1 and variation θ ;

To analyze the effect of explanatory variables on the duration of accidents and the existence of unobserved heterogeneity in the data, two Weibull AFT models (with and without the frailty component) were calibrated. The calibration of the model parameters was performed using the maximum likelihood estimation technique.

7. Study results and discussions

As seen in section 6.1, the duration of accidents that occurred on urban arterial roads of Fortaleza-CE better fits to the Weibull distribution. This result indicates that the hypothesis I raised in section 5.1 may not be valid, possibly reflecting the driver's culture of leaving their vehicles on the road after an accident. Besides, this result also suggests that traffic agents do not effectively respond to clear the road right after an accident in an urban environment. This could be due to either an insufficient number of traffic agents to cover the entire traffic network or the difficulty of solving the problem, given that traffic accidents are typically more severe and challenging to manage than other types of traffic incidents examined in previous studies. Hojati et al. (2014), for example, found that the type of incident may be a determinant

for its duration distribution, showing that the duration of traffic accidents had a better fit to the Weibull distribution while the duration of incidents caused by hazards and stationary vehicles to the Loglogistic distribution, indicating that the latter are more likely to have shorter durations, possibly ending within the first few minutes.

In order to analyze the influence of other factors on the duration of accidents on urban arterial roads based on the hypotheses raised in Section 5.1, the Weibull AFT model was calibrated as shown in the previous section. To test the significance of the parameter θ , the likelihood ratio test was performed. The test had a p-value of 0.225, indicating that there is not enough evidence to believe that there is unobserved heterogeneity in the model for the dataset used. These results indicate that the factors added to the model capture all the variability of the duration of accidents. Therefore, to analyze the effect of the explanatory variables on the duration of accidents, the Weibull model without the frailty component was chosen. However, as the frailty component captures an average effect of the variables that have not been added to the model, it is possible that the effects of these variables are canceled out by each other, which can make it difficult to detect these effects in the dataset used in this study.

The estimated p parameter has a value of 1.819 what indicates that the probability of the accident ending increases monotonically as its duration increases ($p > 1$). This happens due to events that can occur throughout the duration of the accident to solve the problem and clear the road. However, there is no specific period when these events are most likely to occur as found in some previous works as discussed at the beginning of this section.

Table 2 presents the estimated values for the parameters of the calibrated AFT models. The effects of factors on accident duration are analyzed by calculating the acceleration factor, which for binary variables (such as the ones used in our model) can be obtained by calculating the exponential of its coefficients as showed in Equation 8.

$$\gamma = \frac{t_{x_i=1}}{t_{x_i=0}} = e^{\beta_i} \quad (8)$$

where γ : acceleration factor;

β_i : coefficient of explanatory variable i ;

t_{xi} : average duration of accidents for a specific value (0 or 1) of variable i .

Table 2: Summary of AFT models estimations

	Weibull		Weibull with Gamma frailty	
Variable	Parameter estimate	Accelerated factor (%)	Parameter estimate	Accelerated factor (%)
Zone	0.138 (0.038)	14.80	0.143 (0.037)	15.36
Period	-0.175 (0.010)	-16.05	-0.178 (0.011)	-16.32
Nº of vehicles	0.183 (0.047)	20.08	0.194 (0.043)	21.40
Severity	0.128 (0.054)	13.66	0.135 (0.050)	14.41
Constant	4.343 (0.057)	-	4.296 (0.053)	-
p	1.819 (0.042)	-	1.929 (0.036)	-
Theta (ϑ)	-	-	0.102	-
LL (0) ¹	-229.45	-	-229.18	-
LL (β) ²	-221.13	-	-220.85	-
LR ³	16.64	-	16.66	-

Notes: The parameter estimation is followed by p-value in parentheses ().

Likelihood ratio test of the null hypothesis that $\vartheta = 0$: χ^2 (01) = 0.57; p-value = 0.225.

1 - Log-likelihood for constant-only model

2 - Log-likelihood for full model

3 - Likelihood ratio statistics

All the estimated parameters are statistically significant, at about 5% level. Based on the model results, accidents that occur during off-peak hours (period = 1) last, on average, 16% less than accidents that occur during peak hours (period = 0). This result supports the hypothesis III raised in section 5.1, suggesting that drivers are more willing to remove their vehicles from the roads because of traffic-related concerns. In addition, this result suggests that the number of traffic agents in the city is not enough to cover all the accidents during peak hours.

Regarding the location, the results show that the accidents in the periphery (zone = 1) last, on average, about 15% more than accidents in the center (zone = 0). This result may reflect the difficulty of the traffic agents in covering the entire road network of a large city, possibly leading to a prioritization strategy for accident response based on severity and location, which in turn may increase the travel time phase, as discussed in Section 2. Additionally, accidents that occur in the periphery tend to be more severe than those in the city center, likely due to higher speeds and less traffic in the periphery. In this regard, the location seems to play a more critical role in accidents that occur in an urban arterial road network compared to those on a freeway. This is because the area covered by the urban arterial road network is larger, and the speed, traffic, and response team strategy/priority may vary more throughout the road.

Other factor that impacts accident duration based on the model results is the number of vehicles involved. The results show that accidents involving more than two vehicles last, on average, 20% longer than accidents involving up to two vehicles. This finding corroborates the hypothesis that the accident complexity is directly related to the number of vehicles involved and is consistent with the outcomes reported in prior studies conducted on freeways and arterial roads.

Finally, accidents with injuries last, on average, 14% longer than accidents without injuries. However, it was expected an even greater impact of the severity on accident duration based on results found in previous works for freeways. According to Hojati et al. (2014), accidents that require medical attention increase accident duration by 33%. In addition, the Brazilian Traffic Code states that in case of accidents with only property damage, all vehicles must be removed from the road immediately. However, one possible explanation for this outcome could be attributed to the dataset's limitation in categorizing the accident severity, which only differentiates accidents into those with injuries and those without injuries. This classification does not provide enough information to determine the actual severity of an accident and whether it requires medical attention. Another explanation for this result could be attributed to the culture of drivers in Brazil of not following the recommendation of the Brazilian Traffic Code for accidents without victims, and instead waiting for the traffic agents to arrive at the accident site to remove the vehicles from the road.

8. CONCLUSIONS

This work presented an analysis of the duration of traffic accidents in arterial roads in the city of Fortaleza, Brazil. First, we proposed a method to estimate the duration of accidents by combining accidents and traffic databases. The method relies on historical traffic data to detect the different phases of accident duration by comparing the traffic condition in the day of the accident to a typical traffic pattern. The daily traffic conditions are defined by the daily profiles of the flow speed and the flow proportion per lane, measured at a given time window. The flow proportion per lane was chosen, instead of the absolute flow (largely used in studies for freeways), because in arterial roads this variable is less sensitive to the effect of traffic lights, land use, intersections and so on, that usually cause a high variability on speed and flow. Moreover, since flow proportion is highly sensitive to events, such as accidents, that partially obstruct urban roads, it can quickly differentiate a traffic state impacted by an accident from a typical variation of traffic.

The exploratory analysis of the duration of the 316 accidents showed that the average duration of accidents on the arterial roads of Fortaleza was about 71 minutes, with a standard deviation of 43 minutes. This high variability in the accident duration can be a result of several factors, that, for the case of an urban region, can be related to drivers' behavior and experience, the severity of the accidents, the traffic conditions at accident location, the scheduling of the traffic agents, the type of collision, the type and number of vehicles involved, and the trip purposes of the drivers. It is worth noting that the data does not permit to test all hypotheses, since it had no information about drivers' experience and behavior. The Kruskal-Wallis non-parametric test was used to test the factors' significance and the result indicated that the duration of the accidents may be influenced by the zone of the city where the accident occurred (center or suburb), the period of the day (peak or off-peak), the number of vehicles involved and the severity of the accident (with injured or non-injured people). The traffic condition at the time of accident did not significantly impact the accident duration, maybe due to the existence of alternative routes for the drivers on the network, which is not typically the case on most freeways. Moreover, there was no significant difference in the duration of rear-end and side collision accidents, which could be indicative of a cultural tendency among drivers to wait for traffic authorities to arrive and resolve the issue before clearing the road.

The Weibull AFT model was applied to understand how these factors impact the duration of the accidents. An analysis of the data distribution by an Anderson-Darling test revealed that the Weibull was a better probability distribution to represent the variation of accident duration. This result shows that duration of accidents in Fortaleza usually behaves in such way that the hazard function is a monotonic function. In other words, drivers usually do not follow the traffic rules saying that the vehicles must be removed from the road immediately after the occurrence of an accident without injuries. Moreover, apparently there is no strategy for the traffic agents team to respond quickly to accidents on the urban environment. Then, in general, all accident events tend to last more than expected and finish only after the arrival of traffic agents, which possibly makes the termination probability increases over time.

Considering that some factors related to the drivers' behavior could not be observed, Weibull AFT models with and without Gamma heterogeneity were calibrated. The results indicated that the unobserved factors, such as driver experience and trip purposes, did not cause unobserved heterogeneity in the model. Therefore, we assumed that the AFT model without unobserved heterogeneity component is more adequate for analyzing the accident data. However, as the frailty component captures the average effect of variables that were not added to the model, it is possible that the effects of these variables are canceled out by each other, thereby making it difficult to detect these effects in the dataset used in this study.

The estimated coefficients of the explanatory variables showed the following main effects: accidents involving more than two vehicles tend to last 20% longer; accidents in the periphery of the city tend to last 15% longer; accidents during off-peak hours tend to last 16% less; and accidents with injuries tend to last 14% longer. The great impact of the number of vehicles is related to complexity of resolving an accident involving more vehicles and people. The effect of the period of the day can be attributed to drivers being more willing to remove their vehicles from the road due to traffic-related concerns. In addition, this result also suggests that there may not be enough traffic agents in the city to adequately cover accidents during peak hours, resulting in slower response times during these periods. The different effects of geographical locations on accident duration may reflect the challenges faced by the traffic agents in covering the extensive road network of a large city. As a result, there may be a prioritization strategy for accident response based on the severity and location of the accident. Moreover, accidents in the center tend to be less serious than those that occur in the periphery. Finally, the effect of the accident severity was not well measured because the accident database did not allow to categorize the different degrees of severity.

The methodology and analysis in this paper can guide and inform city planners and traffic managers on making decisions to reduce the impact of accidents on traffic in urban regions. It seems that the quantity and scheduling of the traffic agents, as well as the rescue teams, are important issues that when well-planned can reduce the response time of accidents. Further studies in other cities should be undertaken since the realities in other urban regions can be quite different. Furthermore, future works with larger databases may allow a deeper analysis of the explanatory factors not considered in this study, such as the experience of the drivers and the existence of a park to remove the vehicles from road.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERENCES

- Haule, H.J.; T. Sando; R. Lentz et al. (2018) Evaluating the impact and clearance duration of freeway incidents. *International Journal of Transportation Science and Technology*, v. 8, n. 1, p. 13-24. DOI: 10.1016/j.ijtst.2018.06.005.
- Hojati, A. T.; Ferreira, L.; Charles, P. et al. (2012) Analysing freeway traffic-incident duration using an Australian data set, *Road and Transport Research*, v. 21, n. 2, p. 19-31.

- Hojati, A. T.; L. Ferreira; S. Washington et al. (2013) Hazard based models for freeway traffic incident duration, *Accident Analysis and Prevention*, v. 52, p. 171-81. DOI: 10.1016/j.aap.2012.12.037. PMid:23333698.
- Hojati, A. T.; L. Ferreira; S. Washington et al. (2014) Modelling total duration of traffic incidents including incident detection and recovery time, *Accident Analysis and Prevention*, v. 71, p. 296-305. DOI: 10.1016/j.aap.2014.06.006.
- Hou, L.; Y. Lao; Y. Wang et al. (2014) Time-varying effects of influential factors on incident clearance time using a non-proportional hazard-based model. *Transportation Research Part A, Policy and Practice*, v. 63, p. 12-24. DOI: 10.1016/j.tra.2014.02.014.
- IBGE (2020) Brazilian Census of 2020. Available at: <<https://cidades.ibge.gov.br/brasil/ce/fortaleza/panorama>> (accessed 05/25/2023).
- Junhua, W.; C. Haozhe and Q. Shi (2013) Estimating freeway incident duration using accelerated failure time modeling, *Safety Science*, v. 54, p. 43-50. DOI: 10.1016/j.ssci.2012.11.009.
- Khattak, A.J.; J. Liu; B. Wali et al. (2016) Modeling traffic incident duration using quantile regression, *Transportation Research Record: Journal of the Transportation Research Board*, v. 2554, n. 1, p. 139-48. DOI: 10.3141/2554-15.
- Khattak, A.J.; J.L. Schofer and M. Wang (1995) A simple time sequential procedure for predicting freeway incident duration, *IVHS J.*, v. 2, n. 2, p. 113-38. DOI: 10.1080/10248079508903820.
- Kleinbaum, D.G. and M. Klein (2012) *Survival Analysis: A Self-Learning Text* (3th ed.). New York: Springer. DOI: 10.1007/978-1-4419-6646-9.
- Lee, J.; J. Chung and B. Son (2010) Incident clearance time analysis for Korean freeways using structural equation model, *Journal of the Eastern Asia Society for Transportation Studies*, v. 8, n. 1, p. 1878-91. DOI: 10.11175/eastpro.2009.0.360.0.
- Li, R. (2015) Traffic incident duration analysis and prediction models based on the survival analysis approach, *IET Intelligent Transport Systems*, v. 9, n. 4, p. 351-8. DOI: 10.1049/iet-its.2014.0036.
- Li, R.; F.C. Pereira and M.E. Ben-Akiva (2018) Overview of traffic incident duration analysis and prediction, *European Transport Research Review*, v. 10, n. 22, p. 22. DOI: 10.1186/s12544-018-0300-1.
- Li, R.; M. Guo; H. Lu et al. (2017) Analysis of the different duration stages of accidents with hazard-based model. *International Journal of Intelligent Transportation Systems Research*, v. 7, n. 16, p. 7-16. DOI: 10.1007/s13177-015-0115-6.
- Nam, D. and F. Mannering (2000) An exploratory hazard-based analysis of highway incident duration, *Transportation Research Part A, Policy and Practice*, v. 34, n. 2, p. 85-102. DOI: 10.1016/S0965-8564(98)00065-2.
- Shi, Y.; L. Zhang and P. Liu (2015) Survival analysis of urban traffic incident duration: a case study at shanghai expressways, *Journal of Computing*, v. 26, n. 1, p. 29-39. Available at: <<http://www.csroc.org.tw/journal/JOC26-1/JOC26-1-3.pdf>> (accessed 05/25/2023)
- Skabardonis, A.; P. Varaiya and K. Petty (2003) Measuring recurrent and nonrecurrent traffic congestion. *Transportation Research Record: Journal of the Transportation Research Board*, v. 1856, n. 1, p. 118-24. DOI: 10.3141/1856-12.
- Souza, V.N. and F.M. Oliveira Neto (2020) Analysis of accident duration on urban arterials roads, *TRANSPORTES*, v. 28, n. 3, p. 266-80. DOI: 10.14295/transportes.v28i3.2142.
- Valenti, G.; M. Lelli and D. Cucina (2010) A comparative study of models for the incident duration prediction, *European Transport Research Review*, v. 2, n. 2, p. 103-11. DOI: 10.1007/s12544-010-0031-4.
- Washington, S.P.; M.G. Karlaftis and F.L. Mannering (2003) *Statistical and Econometric Methods for Transportation Data Analysis*. New York: Chapman & Hall/CRC. DOI: 10.1201/9780203497111.
- Xie, K.; K. Ozbay and H. Yang (2015) Spatial analysis of highway incident durations in the context of Hurricane Sandy, *Accident Analysis and Prevention*, v. 74, p. 77-86. DOI: 10.1016/j.aap.2014.10.015. PMid:25463947.
- Zhang, H. and A.J. Khattak (2010) Analysis of cascading incident event durations on urban freeways, *Transportation Research Record: Journal of the Transportation Research Board*, v. 2178, n. 1, p. 30-9. DOI: 10.3141/2178-04.