Accident Analysis and Prevention 103 (2017) 112–122

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Full length article

# Analysis of factors affecting the severity of crashes in urban road intersections

L. Mussone[a,*], M. Bassani[b], P. Masci[b]

[a] Politecnico di Milano, Department of Architecture, Built Environment and Construction Engineering (DABC), 9, Via Bonardi, Milano 20133, Italy
[b] Politecnico di Torino, Department of Environmental, Land and Infrastructures Engineering (DIATI), 24, Corso Duca Degli Abruzzi, Torino 10129, Italy

ARTICLE INFO

*Keywords:*
Urban roads
Road intersection
Crash severity level
5-min flow
Short-term data
Back-propagation neural network
Generalized linear mixed model

ABSTRACT

Road crashes are events which depend on a variety of factors and which exhibit different magnitudes of outputs when evaluated with respect to the effects on road users. Despite a lot of research into the evaluation of crash likelihood and frequency, only a few works have focused exclusively on crash severity with these limited to sections of freeways and multilane highways. Hence, at present there is a large gap in knowledge on factors affecting the severity of crashes for other road categories, facilities, and scenarios.

The paper deals with the identification of factors affecting crash severity level at urban road intersections. Two official crash records together with a weather database, a traffic data source with data aggregated into 5 min intervals, and further information characterising the investigated urban intersections were used. Analyses were performed by using a back propagation neural network model and a generalized linear mixed model that enable the impact assessment of flow and other variables. Both methods demonstrate that flows play a role in the prediction of severity levels

## 1. Introduction

In road safety research, analysts are interested in predicting road crashes in terms of location, frequency, pattern, and severity in order to ensure better traffic operations and save lives. A knowledge of the relationship between crash severity and the environmental and traffic conditions is fundamental to achieve this goal.

Regarding weather, the most important factors influencing crashes are those that affect the available friction between wheel and pavement, and/or driver visibility, resulting in crashes when the driver is unable to avoid collisions with moving or fixed obstacles. The conclusions of Caliendo et al. (2007) and Theofilatos et al. (2012) confirm that in the case of rainfall, the frequency of severe crashes increases. Rainfall may change its intensity rapidly, hence average annual rainfall precipitation, and even hourly rainfall may not be sufficient to capture the real-time rainy weather conditions prior to or during crash occurrence. But weather data cannot always be collected in very short intervals and close to each crash location, so Jung et al. (2010) suggested taking into consideration interpolation techniques to derive unmeasured (or even unmeasurable) data.

It is normally difficult to interpret and compare crash data in adverse weather conditions with those in good weather conditions. This is due to factors related to the ability of drivers to adapt their behaviour to weather conditions (Theofilatos and Yannis, 2014), and to possible changes in the composition of the driver population under adverse weather conditions where aggressive drivers (usually young males) tend to persist with their behaviour and speed, while the rest of the population tends to assume a more prudent attitude or avoid driving altogether (Hill and Boyle, 2007).

On the other hand, traffic flow can explain the number of conflicts between vehicles, hence if flow increases then the interferences between vehicles should increase, and their crash risk exposure as well. Theofilatos and Yannis (2014) stated that the influence of traffic flow has been considered more than other traffic related parameters such as speed, density and occupancy, mainly because it is simpler to measure. However, different circulation regimes are possible for the same vehicular volume (daily or even hourly), and this may lead to contrasting or difficult interpretations of crash outcomes. In the case of urban areas, Noland and Quddus (2005) observed that congestion does not significantly affect crash severity in the greater London metropolitan area.

One way to overcome the effect of traffic variables is to avoid the use of aggregate traffic data (i.e., AADT, hourly flow) which are not consistent with the traffic flow levels at the time of crashes. In fact, they hide and smooth out volume peaks, and provide a brief reference to the volume of traffic characterizing a road section or intersection.

* Corresponding author.
*E-mail addresses:* mussone@polimi.it, lorenzo.mussone@polimi.it (L. Mussone), marco.bassani@polito.it (M. Bassani), pietro.masci@gmail.com (P. Masci).

http://dx.doi.org/10.1016/j.aap.2017.04.007
Received 15 December 2016; Received in revised form 6 April 2017; Accepted 8 April 2017
Available online 19 April 2017

Furthermore, neither hourly traffic volumes, volume per lane, nor V/C (volume/capacity) ratio necessarily link the crash events to the explanatory traffic variables. This conclusion was also drawn for weather, therefore traffic and weather data available in an aggregated format cannot always depict the real event conditions.

Xu et al. (2013) carried out a crash severity analysis along a 29-mile of a freeway segment using real-time traffic data on flow, occupancy, and speed. They used a sampling frequency of 30 s to calibrate and validate a sequential logit model, linking the likelihood of crash occurrences at different severity levels (SL) to the previously mentioned traffic flow characteristics. Results showed that property-damage only (PDO) crashes tend to take place in congested conditions with a highly variable speed and frequent lane changes. Injury and fatal crashes occur more often in less congested flow conditions, while fatal crashes occur under uncongested conditions as well as when there are large differences in speed between adjacent lanes. Similar conclusions were drawn by Christoforou et al. (2010) with traffic data records of 6 min each. Yu and Abdel-Aty (2014) conducted a SL analysis on a mountainous freeway on the basis of 6-min traffic and weather data, with the SL classified into two levels (severe injury and PDO). Steep grades, standard deviation of speed, temperature, and snow were found to be the most influential variables for this type of facility. Other authors used real-time traffic data for crash and safety analysis in urban arterials and urban expressways (Shi et al., 2016a,b; Theofilatos and Yannis, 2016; Hossain and Muromachi, 2013a,b).

Shankar et al. (1996) observed that road safety studies were historically limited to the localization of fatalities, even though the estimation of consequences in terms of SL (from PDO to fatalities) could help in understanding the benefits accruing from countermeasures. Furthermore, at present there are no studies in literature that focus on the contributing causes to a specific SL in the event of a crash on urban road networks.

To bridge this gap, the paper aims at providing knowledge on factors contributing to crash severity in urban road intersections. Crash, traffic and the weather databases of the Turin road network in the North-West of Italy were collected and used to calibrate and validate predictive models for crash SL. The Artificial Neural Network (ANN), a robust tool used to investigate complex phenomena without assuming any preliminary hypotheses on the model, was used. Since the ANN cannot provide an analytical formulation, a Generalized Linear Mixed Model (GLMM) was also applied. Both models were subjected to sensitivity analysis to comprehend the effects of each variable.

The ANN method is well-known and there are many papers on its use for safety analysis in different scenarios (Karlaftis and Vlahogianni, 2011). There are fewer works using GLMM on the same subject. An example is the paper by Bailey and Hewson (2004) which analyses the incidence of fatal and serious crashes for different types of road users by a multivariate GLMM (Bailey and Hewson, 2004). Gargoum et al. (2016) explored the relationships between some features of road surroundings (geometric, temporal factors, and weather conditions), and driver compliance (a categorical variable) with speed limits by a GLMM (data are modelled using a cumulative logit model with random intercepts). A mixed logit model was also used by Milton et al. (2008) in an exploratory analysis of crash severity in highways. Similarly, Chen and Chen (2011) applied a mixed logit model to the study of the severity of traffic accidents involving trucks in single or multi-vehicle crashes. A comparison of three crash severity models, multinomial logit, ordered probit, and mixed logit with regard to crash data underreporting effects was proposed in Ye and Lord (2011). Qin et al. (2013) also compared the results obtained by three logistic regression models (multinomial logit, partial proportional odds and mixed logit) used to investigate the effects on crash severity of large trucks. With the aim of employing a multivariate approach to the investigation of crash severity, Ma and Kockelman (2006) proposed a Bayesian Poisson regression. Ma et al. (2008) introduced a multivariate Poisson approach to model injury counts by severity whereas Wang et al. (2017) identified the effects of a number of factors on different crashes by using a multivariate Poisson log-normal regression

## 2. Database formation

### 2.1. Crash and traffic data

The crash data used in this research were obtained from the *Istituto Nazionale di Statistica* (ISTAT). The database contains details on crash dynamics and location, on vehicles and on the individuals involved, but it does not include PDO events in accordance with current Italian legislation, specifically articles number 582, 583 and 590 of the Italian Penal Code 2015 (Repubblica Italiana, 2015). In fact, Italian law defines road accidents as crashes only when they result in at least one injury, and crash consequences are classified into five severity levels (SL) all of which refer to the most seriously injured road user in any particular crash:

- very slight injuries (VSI), when the most seriously injured person has a prognosis of fewer than 20 days;
- slight injuries (SLI), when the prognosis is between 21 and 40 days;
- severe injuries (SEI), if the event causes an illness that endangers the life of the injured party, and/or if the event results in the permanent weakening of brain function or a body organ;
- guarded prognosis (GPR), if the doctor cannot determine the disability, and he issues a report of "guarded prognosis" (until his reservations can be resolved, the road crash must be considered and treated as a determining factor); and
- fatalities (FAT), which include any injured persons who die within 30 days of the crash.

The ISTAT database was matched up with information from Turin's Municipal Police, to include: (a) historical data, in particular the time, to the nearest minute, day, month and year of the crash event; (b) locality data with the name of the street and house number where the crash took place along a road segment, or the denomination of two streets when it occurred at an intersection; and (c) generic information concerning crash SL.

Traffic data were provided by the 5T Company which uses induction-loop traffic sensors, located along the exiting lanes of the monitored intersections, with flow data collected every 5 min. Table 1 reports the crash data counts that were associated with 5-min flow data, while Fig. 1 shows the portion of the road network monitored by 5T in 2006, and includes the time scale used to estimate the seven 5-min flows of the 35 min before, during and after the crash event.

They are divided into function of SL, road typology, pavement conditions, vehicle type, gender and age of drivers (no more than two, and indicated as A and B). Variables relating to driver B have a relatively high percentage of unknowns since some crashes are single-vehicle collisions (e.g. rollover, roadway departure, collision with animals) thus involving only vehicle A. Driver B may be a pedestrian or even a child riding a bicycle.

Table 2 reports the traffic flow (TF) as the number of vehicles counted in 5-min intervals over a period of 35 min. Only crashes occurring along intersections yielding valid and reliable traffic data were extracted from the main database for further use. As a result, the database used for model calibration was obtained as a subset of the total number of crashes recorded in the official database, since only crashes that occurred on monitored roads were included (therefore it can be considered a random sample of crashes on monitored roads).

### 2.2. Weather data

The Environmental Protection Agency of the Piedmont Region (ARPA Piedmont) provided data on weather conditions from the Turin weather station. It is located in the city centre at 238 m a.s.l.,