Available online at www.sciencedirect.com

ScienceDirect

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

Original Research Paper

Modeling the effect of traffic regimes on safety of urban arterials: The case study of Athens

*Athanasios Theofilatos**, *George Yannis*, *Eleni I. Vlahogianni*, *John C. Golias*

School of Civil Engineering, National Technical University of Athens, Zografou, Athens, GR-15773, Greece

HIGHLIGHTS

- Traffic conditions are divided into meaningful regimes.
- The impacts of traffic regimes on accident likelihood and severity are investigated.
- Powered-two-wheelers are also examined.
- Potential hazardous traffic conditions are identified and discussed.

ARTICLE INFO

Article history:
Available online 24 May 2017

Keywords:
Real-time traffic
Regime
Urban arterials
Cluster analysis
Bayesian logistic regression
Accidents

ABSTRACT

This study aims to divide traffic into meaningful clusters (regimes) and to investigate their impact on accident likelihood and accident severity. Furthermore, the likelihood of powered-two-wheelers (PTWs) involvement in an accident is examined. To achieve the aims of the study, traffic and accident data during the period 2006–2011 from two major arterials in Athens were collected and processed. Firstly, a finite mixture cluster analysis was implemented to classify traffic into clusters. Afterwards, discriminant analysis was carried out in order to correctly assign new cases to the existing regimes by using a training and a testing set. Lastly, Bayesian logistic regression models were developed to investigate the impact of traffic regimes on accident likelihood and severity. The findings of this study suggest that urban traffic can be divided into different regimes by using average traffic occupancy and its standard deviation, measured by nearby upstream and downstream loop detectors. The results revealed potential hazardous traffic conditions, which are discussed in the paper. In general, high occupancy values increase accident likelihood, but tend to lead slight accidents, while PTWs are more likely to be involved in an accident, when traffic occupancy is high. Transitions from high to low occupancy also increase accident likelihood.

* *Corresponding author*. Tel.: +30 210 772 1575; fax: +30 210 772 1454.
E-mail addresses: atheofil@central.ntua.gr (A. Theofilatos), geyannis@central.ntua.gr (G. Yannis), elenivl@central.ntua.gr (E. I. Vlahogianni), igolias@central.ntua.gr (J. C. Golias).

Peer review under responsibility of Periodical Offices of Chang'an University.
http://dx.doi.org/10.1016/j.jtte.2017.05.003

1. Introduction

The effect of traffic characteristics on road safety has been investigated for many years. During the past decade, increased attention has been given to developing relationships between real-time traffic characteristics and road safety (Abdel-Aty et al., 2012; Christoforou et al., 2010; Lee et al., 2003; Oh et al., 2001, 2005, 2006; Xu et al., 2013; Yu and Abdel-Aty, 2014a; Zheng et al., 2010). The great majority of studies utilizes data from freeways, whilst there are some studies that investigate accident likelihood on urban expressways (Hossain and Muromachi, 2013).

Moreover, although literature has for long supported the regime-like evolution of traffic flow in both freeways and urban arterials (Hall et al., 1992; Kerner and Rehborn, 1996; Vlahogianni et al., 2007, 2008a, 2008b; Wu, 2002; Yildirimoglu and Geroliminis, 2013), limited knowledge is available on the manner the formation of these regimes and/or the transitions between them may affect traffic safety.

Towards this direction, Abdel-Aty et al. (2005) divided freeway traffic flow in high and low speed states and then examined severity and the mechanism of multi-vehicle accident occurrence under these two different states. Golob et al. (Golob and Recker, 2004; Golob et al., 2004) investigated the safety impact of traffic by dividing traffic flow into different traffic states (traffic regimes) by means of cluster analysis on the basis of traffic flow data collected from the nearest loop detector station from crash locations. The authors attempted to associate traffic regimes with accident type.

Xu et al. (2013) emphasized on the need to divide traffic in states and explored their effect on safety, due to the fact that different traffic states may have different influences on the risk of an accident. More specifically, the authors utilized traffic occupancy measured from nearby loop detectors and classified traffic flow into traffic states. Then, each traffic state was associated with a certain safety level. Moreover, it was found that the impact of traffic flow parameters on crash risk is not the same across different traffic flow state.

Yeo et al. (2013) defined the traffic states (free flow, back of queue, bottleneck front, and congestion) according to their distinctive patterns and attempted to model the crash involvement rate for each traffic state. It was concluded that crash involvement rate in free flow state is approximately 5 times lower than it in other traffic states.

The literature review reveals that a major limitation is data availability, due to the fact that real-time data is mainly regard freeways and not major urban arterials. Studies using real-time traffic data to investigate accident severity are relatively few (Jung et al., 2010; Yu and Abdel-Aty, 2014a, 2014b), while only a few studies investigate both accident likelihood and severity (Xu et al., 2013). Moreover, European countries are rarely considered, as only one study was found that explored safety of a motorway in Belgium (Pirdavani et al., 2015). Although there are many studies investigating powered-two-wheeler (PTW) accident risk (Montella et al., 2012; Maestracci et al., 2012; Harnen et al., 2003; Kasantikul et al., 2005), to the best of our knowledge no studies linking PTW accident risk with real-time traffic data are found.

Ensuring safety in major urban roads holds high priority. Consequently the primary objective of this study is to divide urban traffic flow into different regimes and to investigate their effect on accident likelihood and severity. Furthermore, PTW accident risk (involvement of a PTW in an accident) is also explored.

The remainder of the paper is organized as follows. Firstly, the proposed methodology is demonstrated (finite mixture cluster analysis, discriminant analysis, and Bayesian logistic regression). Then the data description and preparation are provided. Next, the application of the models is explained and the results are presented and discussed. The final section provides the conclusions.

2. Methodology

The proposed methodological approach involves a two-step methodology. First, expectation maximization clustering (EM) (or finite mixture) was used to classify traffic into different regimes. Second, Bayesian logistic regression models were applied in order to correlate traffic regimes with traffic safety.

2.1. *Finite mixture cluster analysis*

Cluster analysis is a widely used method for grouping observations on the basis of similar data structure. In this study, finite mixture cluster analysis is used to identify homogeneous groups of traffic conditions, which are called "regimes".

A Gaussian finite mixture model-based clustering (covariance parameterization and number of cluster selected via the Bayesian information criterion (BIC)) was followed. The models were fitted by EM algorithm. Fraley et al. (2012) and Fraley and Raftery (2002) provide a detailed description of normal mixture modeling. All following equations appear in Fraley et al. (2012).

A normal or Gaussian mixture model is assumed.

$$\Pi_{i=1}^{n}\Sigma_{k=1}^{G}\tau_k\varphi_k(x_i|\mu_k,\Sigma_k) \tag{1}$$

where x represents the data, $\sum_k$ is the covariance in this Gaussian mixture model, G is the number of components, τ_k is the probability that a case belongs to the kth component ($\tau_k \geq 0$ and $\sum_{k=1}^{G}\tau_k = 1$) and

$$\varphi_k(x|\mu_k,\Sigma_k) = (2\pi)^{-d/2}\left|\Sigma_k\right|^{-1/2}\exp\left\{-1/2(x_i-\mu_k)^{\mathrm{T}}\Sigma_{k=1}^{G}(x_i-\mu_k)\right\}\frac{n!}{r!(n-r)!} \tag{2}$$

where x_i is a d-dimensional vector that measures the components (clusters) φ_k $(x|\mu_k, \sum_k)$ in this ellipsoidal model, μ_k is the mean of component k in this Gaussian mixture model.

According to Fraley et al. (2012), the distribution for EM algorithm for multidimensional data, can be spherical, diagonal or ellipsoidal. The volumes and shapes of clusters can be equal or variable. The combination of these characteristics, defines each model (namely the covariance matrix $\sum_k$). For more information, the reader is encouraged to read the report from Fraley et al. (2012).