Investigation of Powered-Two-Wheeler accident involvement in urban arterials by considering real-time traffic and weather data

Theofilatos Athanasios[a, *], Yannis George[a]

[a]National Technical University of Athens, School of Civil Engineering, Dept. of Transportation Planning and Engineering, 5, Iroon Polytechniou Str., Zografou Campus, Zografou-Athens, GR-15773, Greece
Email: atheofil@central.ntua.gr, geyannis@central.ntua.gr
* Corresponding author.

Abstract

Objective: Understanding the various factors that affect accident risk is of particular concern to decision makers and researchers. Recently, the incorporation of real-time traffic and weather data constitutes a fruitful approach when analysing accident risk. However, the vast majority of relevant research has no specific focus on vulnerable road users such as Powered-Two-Wheelers (PTWs). Moreover, studies using data from urban roads and arterials are scarce. This study aims to add to the current knowledge by considering real-time traffic and weather data from two major urban arterials in the city of Athens, Greece, in order to estimate the effect of traffic, weather and other characteristics on PTW accident involvement.

Methods: Because of the high number of candidate variables, a Random Forest model was applied to reveal the most important variables. The significant variables according to the Random Forest model were used as input to a Bayesian logistic regression model in order to reveal the magnitude of their effect on PTW accident involvement.

Results: The results of the analysis suggest that PTWs are more likely to be involved in multi-vehicle accidents than in single-vehicle accidents. It was also indicated that increased traffic flow and variations in speed have a significant influence on PTW accident involvement. On the other hand weather characteristics were found to have no effect.

Conclusions: The findings of this study can ~~be helpful~~ contribute to the understanding of accident mechanism of PTWs ~~to deploy in real-time safety management strategies aiming~~ and ~~to~~ reduce accident risk of PTWs on urban arterials.

Keywords: Powered-Two-Wheelers, accidents, real-time data, urban arterials

INTRODUCTION

Mopeds and motorcycles (together as Powered-two-wheelers or PTWs) constitute a flexible and economic alternative for commuters and are widely used in dense urban areas where traffic density is high. However, PTW occupants face a much higher risk of being fatally injured than car occupants (Lin and Kraus, 2009), mainly due to the lack of protective equipment which is well-known to enhance passive safety (Viano, 1991). In 2013, circa 3.862 riders (drivers and passengers) of motorcycles were killed in the European Union (ERSO, 2015). Most of moped fatalities occur on urban areas while the majority of motorcycle fatalities occur in rural areas (ERSO, 2015). In Greece the majority of PTW fatalities are motorcyclists (ERSO, 2015).

In order to improve safety of PTWs, great efforts have been made explore PTW accident risk and a variety of accident-related factors have been identified. The road environment (road type, road geometry, roadside installations etc.) plays a very significant role in PTW accident risk (Haque et al., 2009; Haque et al., 2010; Harnen et al., 2003; Kasantikul et al., 2005; Wanvik, 2009; Daniello and Gabler, 2011). However, the effect of traffic and weather characteristics has not been extensively explored (Branas and Knudson, 2001; Houston and Richardson, 2008; Xuequn et al., 2011; Abdul Manan and Várhelyi, 2012). It is noted that behavioural risk factors are considered important and have deeply explored in literature but do not fall in the scope of this study and thus are not discussed here.

Recently, an increasing number of studies exploit real-time traffic and weather data in order to investigate accident risk on freeways. To the best of our knowledge there are no studies with real-time data that explore PTW accident risk, as the relevant literature has a more general scope, exploring accident risk (Abdel-Aty and Pande, 2005; Abdel-Aty et al., 2007; Ahmed and Abdel-Aty, 2012; Xu et al., 2013a and 2013b; Hassan and Abdel-Aty, 2013; Abdel-Aty et al. 2012; Yu and Abdel-Aty, 2013b), accident frequency (Yu and Abdel-Aty, 2013a; Yu et al., 2013) or accident severity (Christoforou et al., 2010; Golob et al., 2008; Yu and Abdel-Aty, 2014a and 2014b; Xu et al., 2013b; Jung et al., 2010). It is also noted, that the vast majority of studies exploit freeway data. Concerning accident likelihood in particular, previous research on this topic suggests that common risk factors are mainly the variations in traffic conditions (Ahmed et al., 2012a and 2012b; Ahmed and Abdel-Aty 2012; Xu et al, 2013a and 2013b; Zheng et al., 2010) and low visibility or adverse weather conditions (Xu et al., 2013a; Ahmed et al., 2012b; Abdel-Aty et al., 2012).

Because of the low weight, the high acceleration and manoeuvring capabilities of PTWs it is very important to understand especially the influence of existing traffic conditions on PTW risk. In addition, PTW movements with simulation and also the various interactions taking place on dense urban environments are of also of great interest and have been investigated by various researchers (Dey et al., 2008; Vlahogianni et al., 2012; Nikias et al., 2012; Vlahogianni, 2014; Barmpounakis et al., 2014, Barmpounakis et al., forthcoming)

Consequently urban PTW data need to be further explored. However, existing studies utilizing real-time traffic and weather data do not have particular focus on Powered-Two-Wheeler risk, as the literature review has showed. The present paper aims to add to the current knowledge by explaining Powered-Two-Wheeler (PTW) accident risk on urban arterials of Athens, Greece. Aside from the traditional accident attributes, real-time traffic and weather data are attempted to be correlated with PTW accident involvement. It is noted that the term “PTW accident risk”, refers to whether a PTW is involved or not in an accident that has already occurred.

METHODS

Data preparation

Data for this study have been collected for the period 2006-2011 to investigate the relationship between traffic, weather and other characteristics with PTW accident risk. More specifically, the area of interest is in the Greater Athens area; two central densely urban arterials with very similar geometrical and traffic characteristics.

All accident data were collected from the Greek accident database SANTRA, which is provided by the National Technical University of Athens. It provides access to road accidents in Greece since 1985 in high detail and includes all relevant information about each accident (persons injured, severity of injuries, location, weather, accident type and so on). The traffic data were extracted from the Traffic Management Centre (TMC) of Athens, which operates since 2004 and covers several major roads in Athens by having 550 loop detectors, 217 cameras and 24 variable message signs. It also controls more than 800 junctions.

The data includes traffic flow ~~(number of vehicles per 5 min)~~, traffic occupancy ~~(%)~~ and mean-time speed ~~(km/h)~~. Weather data were collected from the Hydrological Observatory of Athens (HOA, 2012), which provides an online open-access database. HOA covers more than 10 meteorological stations located in the greater Athens area, measuring rainfall ~~(mm)~~, temperature ~~(°C)~~, relative humidity ~~(%)~~, solar radiation ~~(W/m²)~~, wind direction ~~(degrees)~~, wind speed ~~(m/sec)~~ etc.

Then, each accident was assigned to the closest upstream loop detector and to the closest weather station. Note that, spacing between consecutive loop detectors is not consistently the same, since they were placed at specific places of interest (in terms of traffic interest) by the Traffic Management Centre. Some of the loop detectors were very closely spaced (the minimum distance between loop detectors was found to be about 68m), while a few detectors were very far from each other (the maximum distance between loop detectors was identified to be about 1,13km). Accident cases where the location from the closest upstream loop detector was more than 600m were generally excluded, as such measurements could not considered to be highly reliable due to the complex nature of the urban environment. However, in cases where there were only a few minor roads merging on the major arterial, higher distances were included as well, although in general distances higher 750m were not considered reliable.

Consequently, an observation is a record of each accident, the corresponding traffic and weather conditions and also various external factors. ~~Traffic data from the closest upstream loop detector were considered.~~ The 5-min raw traffic data were further aggregated to 1-hour level to obtain averages and standard deviations prior to an accident. Traffic flow was divided by the number of lanes for consistency reasons, as these arterials have segments with different number of lanes. It is noted that data from bus lanes were not considered in this study. When the time or location of an accident was not known, this case was deleted from the dataset.

The 10-min raw weather data were aggregated in order to obtain maxima, averages and standard deviations in the time-slice of 1-hour prior to the time of the accident occurrence. Regarding rainfall, the sum and standard deviation of rainfall has also been calculated for 1h, 2h, 6h and 12h prior to the time of the accident. Moreover, the qualitative variable *weather* (good/adverse weather) and the pavement condition (good/wet) as originally coded in SANTRA database system were also considered for further analysis.

In this study, a time lag of 20 minutes was used, meaning that the time of each accident case was recalculated by subtracting 20 minutes, in order to avoid the impact of the accident itself on the traffic variables and to compensate for any potential inaccuracies in the precise time of the accident. Other similar studies have used similar or even larger time lags (Christoforou et al., 2010; Quddus et al., 2009). The following example illustrates the approach. If an accident occurred on 10 February at 10:00, then the relevant data are extracted for the time period 8:40 to 9:40 from the closest upstream loop detector and from the closest meteorological station.

The final dataset included 527 accident cases. PTWs were involved in 326 of those accidents (61.9% of the accidents). For the needs of the study, a new binary variable is created (1 for PTW accident involvement, 0 when no PTWs are involved in an accident) and expresses whether a PTW is involved in an accident. Powered-Two wheelers did not have separate lanes, but they share lanes with other vehicles (but not permitted to enter bus lanes).Tables 1, 2 and 3 provide ~~a~~ data description as well as some basic descriptive statistics.

Please insert Table 1 here

Please insert Table 2 here

Please insert Table 3 here

Statistical Analysis

Random Forests ~~–~~A random forest is a classifier including a collection of tree-structured classifiers {h(x, Θ_k), k = 1,...}, where the {Θ_k} are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x (Breiman, 2001). Strobl and Zeileis, (2008) suggest that in order to construct a random forest, a number of bootstrap samples from the original sample have to be drawn and afterwards a classification tree to each bootstrap sample has to be fitted (number of trees).

Random Forests (RF) have been used in various similar traffic safety studies (Abdel-Aty and Haleem, 2011; Ahmed and Abdel-Aty, 2012; Yu and Abdel-Aty, 2014b) by ranking the explanatory variables according to their relative importance. Thus, this method is very useful when dealing with a high number of candidate explanatory variables, because it assists in selecting the most significant variables and then enter them in other statistical models. However, a limitation of this method is that the magnitude of the effect and the sign of each variable effect are not revealed. Therefore, it is used mainly as a preliminary analysis.

Bayesian logistic regression ~~—~~The Bayesian logistic regression approach is different than the traditional frequentist approach, in a sense that prior distributions for each parameter are defined and then the data are used to update beliefs about parameters. Therefore, the posterior distributions and the 95% credible intervals are produced. Any prior distribution can be potentially used, however, it is preferable to use "vague" or "non-informative" priors if little is known about the coefficient values (Lunn et al., 2012). The likelihood function for

Bayesian logistic regression is the same as in the frequentist inference. Therefore, the logistic regression equation is:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \qquad \text{(Eq. 1)}$$

where β_0 is the constant term, β_i is a beta coefficient of the explanatory variable x_i.

While the "frequentist" approach considers the model parameters as fixed unknown constants and uses the data solely to best estimate the unknown values of the parameters, the Bayesian approach treats parameters as random variables and the data are used to update beliefs about the behaviour of the parameters. Moreover, Bayesian inference can effectively avoid the problem of over fitting that occurs when there is a low number of observations but high number of variables.

A parameter is statistically significant if the 95% credible interval (2.5%-97.5%) of the beta coefficient does not ~~contain~~ include zero (Lunn et al., 2012). The DIC as the Bayesian generalization of Akaike information criterion (Akaike, 1974), is a measure of model fit and to compare models. More information about the Bayesian logistic regression technique can be found in Lunn et al. (2012).

RESULTS AND DISCUSSION

The preliminary analysis of the Random Forests (RF) served as a first screening to identify potential important variables that should be entered in the Bayesian logistic regression models. All traffic (flow, speed, occupancy) and weather (temperature, humidity, rainfall, wind speed, wind direction, solar radiation) related variables were considered as potential risk factors and were entered in the RF model. Traditional accident variables (such as accident type, location, pavement condition, etc.) were considered as well. All variables are listed in Tables 2, 3 and 4.

The variable importance ranking was explored by constructing 1000 trees and using mtry = 7. Figure 1 illustrates the variable importance rankings by ranking and sorting the ~~.~~variables according to their relative importance. The blue vertical line was added for illustration purposes. More specifically, the vertical blue line was placed where a "gap" appears between more important and less important variables in order to visually separate them. Similarly, the red threshold line was added to aid the interpretation of the results. More specifically, this line was added at the absolute value of the lowest ranking predictor. Strobl et al. (2009) suggest that variables can be considered informative and potentially important if their variable importance value is above the absolute value of the lowest negative-scoring variable. All variables right to the red vertical line were further considered in the Bayesian logistic models.
~~Explanatory variables to the right of dashed vertical line are identified to be potentially significant. However, another vertical line was set in order to highlight the very highly important predictors (to the right of the blue line).~~

The type of accident was identified as a considerably highly important predictor of PTW accident occurrence. Furthermore, the 1-h average speed upstream (V_avg_1h_up), the 1-h average occupancy upstream (Occ_avg_1h_up), 1-h average flow upstream (Q_avg_1h_up) and lastly the 1-h median flow upstream (Q_median_1h_up) were identified as highly significant as well. ~~As in previous analyses, the variables had to be checked once more for potential correlations.~~ However, as stated earlier, all variables right to the red vertical line are potentially significant. For example, standard deviation of occupancy (Occ_stdev_1h_up), coefficient of variation of speed (V_cv_1h_up), average temperature (T_1h_avg), maximum temperature (T_1h_max), maximum wind speed (W.Sp_1h_max) and average humidity (Hum_1h_avg) have also to be considered for the final models.

Please insert Figure 1 here

Then, a correlation matrix of the variables entered in the final models has been checked to avoid multicollinearity problems. For example, the median flow although indicated by the RF model as significant, it was found to be correlated with the average flow (r=0.62) and thus removed from the model. Moreover, only the coefficient of variation of speed was retained to the model since it was correlated with standard deviation of occupancy (r=0.91).

Afterwards, the Bayesian logistic regression was performed on the basis of the Random Forest model findings. The next table (Table 4) summarizes the findings of the Bayesian logistic regression model. This was the model with the lowest DIC value (640.05) and thus having the best fit. Furthermore, the area under the ROC curve (AUC) was used to assess the model fit. The area under the ROC curve shows how well the model discriminates between accidents with a PTW (y=1) and without a PTW (y=0) and it was found to be 0.724, indicating that the model can provide good discrimination.

All prior distributions of the parameters were chosen to be non-informative, following a normal distribution with zero mean and a very low precision of 0.0001, namely ~dnorm(0, 0.0001). The first 5,000 iterations were discarded and used as burn-in and 3 chains of 20,000 iterations were set up. Regarding the two traffic variables in the model, they were not highly correlated (r=0.16).

The beta coefficient of average flow has a positive sign, indicating that when flow increases, there is an increase in the probability of a PTW to have been involved in an accident. This means that in more congested traffic conditions, the probability that PTWs are ~~more likely to be~~ involved in accidents is high. One explanation could be the existence of increased interaction with other motorized traffic and the potential need for manoeuvres under these dense traffic conditions.

Another interesting finding was that the coefficient of variation of speed (basically expressing variations in mean time speed) increase the probability of an accident involving a PTW. In literature, speed variations have been linked to high risk of accidents on freeways (Zheng et al., 2010; Xu et al., 2013b; Ahmed et al., 2012b; Hassan and Abdel-Aty, 2013). The finding of the Bayesian model suggests that large variations in speed, have an influence on PTW accidents in urban roads as well. It is important to comment on the high odds ratio (3.487) of the coefficient of variation of speed variable, meaning that 1 unit increase results in 3.487 higher odds of an accident involving a PTW than that before the increase.

Lastly, it was found that accident type was strongly associated with PTW accidents. It was found that PTWs are more associated with head-on collisions (Acc.type1), side (Acc.type3) and sideswipe collisions (Acc.type4). The 95% credible intervals of the beta coefficient of Acc.type2 (rear-end collisions) include zero and therefore is non-significant. The reference category was set ~~us~~ as an accident with a fixed object or run-off road collisions (Acc.type0). What is interesting is that by interpreting the odds ratios, PTWs are more likely to be involved in head-on, side and sideswipe collisions rather than a collision with a fixed-object or to run-off-road (7.737, 7.546 and 2.652 times more likely respectively). This finding means that PTWs are more vulnerable and thus are more affected by interactions with other motorized traffic, as they are more likely to be involved in multi-vehicle accidents than in single vehicle accidents.

In order to further support this finding, a table of descriptive statistics was constructed. Table 5 shows that 107 PTW accidents are single-vehicle, while 219 PTW accidents are multi-vehicle. It is also shown that that the proportion of PTWs is significantly higher in head-on, side and sideswipe collisions, while in off-road and fixed object collisions the PTWs and other vehicles are almost equally involved.

Please insert Table 4 here

Please insert Figure 2 here

Please insert Table 5 here

CONCLUSIONS

The aim of the present study was to investigate Powered-Two-Wheelers (PTWs) accident risk and more specifically to understand the mechanism behind PTW involvement in accidents. For that reason, real-time traffic and weather data as well as other traditional accident characteristics from urban roads in Athens, Greece were exploited. Initially, a Random Forests (RF) model was utilized in order to rank the candidate variables according to their relative importance and provide a first insight on the potential significant variables affecting PTW accident risk. The RF technique revealed a significant effect of real-time traffic variables. However, no weather variables were indicated as significant.

Afterwards, a Bayesian logistic regression model was applied on the basis of the RF preliminary analysis and provided useful information regarding the likelihood of PTW accidents in urban roads of the city of Athens. PTW accident probability was found to be positively influenced by high values of traffic flow and by variations in speed. Moreover, PTWs are more likely to be involved in multi-vehicle accidents than in single-vehicle accidents. However, rear-end accidents might probably occur in high congestion or in near signalized intersections and seem to affect more other types of motorized traffic, such as private cars. The fact that PTWs seem to be prone to be involved in such types of accidents when fluctuations in traffic conditions and conditions of increased traffic flow exist, implies that in urban road roads, PTWs are particularly affected by the interaction with other motorized traffic. Therefore, PTW accident occurrence seems to be also a matter of behavioural interaction with other motorized traffic.

The authors recognize the limitations of the study. The distance between consecutive loop detectors was not optimal as the placement of loop detectors served traffic management and not safety purposes. The best would be to have very dense network of detectors or perhaps keep only accident cases very close to detectors. This was not feasible since it would have resulted in further reduction in our sample size, which was already not very large. For that reason, it was decided to include also real-time weather variables and traditional accident variables in the study and not rely solely on the traffic variables. It is noted however, that studies in literature have considered measurements even longer than 1 mile (although such measurements concern freeways); see for example Zheng et al. (2010). Moreover, data were not as microscopic as in other studies in the field and such time intervals used may be too large to capture short-term variations. In addition, traffic conditions and interactions at intersections need to be further examined in order to better understand the effect of real-time traffic at complex urban environments. Future studies on other cities' urban arterials could utilize more microscopic traffic and weather data and perhaps consider carrying out separate analyses on intersections by including approaching traffic flows from all merging roads.

However, to the best of our knowledge this study could be considered as one of the first attempts to utilize real-time traffic and weather data from urban arterials in order to model PTW accident risk. The results of the study can potentially be used as a direction towards further research and also further actions from policy makers for better traffic monitoring in major urban arterials. By understanding the accident mechanism behind PTW accidents, accident risk of PTWs on urban arterials could be reduced.
~~The authors are aware that data were not as microscopic as other studies in this field and that such time intervals used in the study may be too large to capture short-term variations. However, to the best of our knowledge this study could be considered as one of the first few attempts to utilize real-time traffic and weather data from urban arterials in order to model PTW accident risk. Future relevant studies on other cities' urban arterials could utilize more microscopic traffic and weather data.~~

~~The findings of this study can contribute to reducing accident risk of PTWs on urban arterials by deploying real-time safety management strategies. For example,~~ Lastly, after high risk conditions are identified (e.g. congested traffic, variations in mean speed), variable messages signs could be potentially used.

REFERENCES

Abdel-Aty M., Haleem K. Analyzing angle crashes at unsignalized intersections using machine learning techniques. Accident Anal Prev. 2011; 43: 461–470

Abdel-Aty M., Hassan H.M., Ahmed M., Al-Ghamdi A.S. Real-time prediction of visibility related crashes. Transportation Res C. 2012; 24: 288–298

Abdel-Aty M., Pande, A. Identifying crash propensity using specific traffic speed conditions. Journal of Saf Res. 2005; 36: 97– 108

Abdel-Aty, M., Pande A., Lee C., Gayah V., Dos Santos C. Crash risk assessment using intelligent transportation systems data and real-time intervention strategies to improve safety on freeways. Journal of Int Trans Syst. 2007; 11(3): 107-120

Abdul Manan M.M., Varhelyi A. Motorcycle fatalities in Malaysia. IATSS Res. 2012; 36(1): 30-39

Ahmed M., Abdel-Aty M. The viability of using Automatic Vehicle Identification data for real-time crash prediction. IEEE Transactions of Intel Transp Syst. 2012; 13(2): 459-468

Ahmed M., Abdel-Aty M., Yu R. A bayesian updating approach for real-time safety evaluation using AVI data. Transportation Res Rec. 2012a; 2280: 60–67

Ahmed M., Abdel-Aty M., Yu R. Assessment of the interaction between crash occurrence, mountainous freeway geometry, real-time weather and AVI traffic data. Transportation Res Board. 2012 January 22-26; Washington, D.C.

Akaike H. A new look at the statistical model identification. IEEE Transactions on Aut Contr. 1974; 19: 716–723

Barmpounakis E.N., Vlahogianni E.I., Golias J.C. Investigating Powered Two-Wheelers Overtaking Behavior in Urban Arterials. Transportation Res Board. 2014 January 12-16; Washington, DC.

Barmpounakis E.N., Vlahogianni E.I., Golias J.C. Vision-based multivariate statistical modeling for powered two-wheelers maneuverability during overtaking in urban arterials. Transportation Letters. In press; doi: http://dx.doi.org/10.1179/1942787515Y.0000000020

Branas C.C., Knudson M.M. Helmet laws and motorcycle rider death rates. Accident Anal Prev. 2001; 33: 641-648

Breiman, L. Random Forests. Machine Learning. 2001; 45: 5-32

Christoforou Z., Cohen S., Karlaftis M. Vehicle occupant injury severity on highways: An empirical investigation. Accident Anal Prev. 2010; 42: 1606-1620

Daniello A., Gabler H.C. Fatality risk in motorcycle collisions with roadside objects in the United States. Accident Anal Prev. 2011; 43: 1167-1170

Dey P.P., Chandra S., Gangopadhyay S. Simulation of Mixed Traffic Flow on Two-Lane Roads. Journal of Transp Eng. 2008; 134: 361-369

European Road Safety Observatory, ERSO. Traffic safety basic facts "Motorcycles and mopeds". European Commission, Traffic Safety Basic Facts on Motorways, European Commission, Directorate General for Transport. 2015

Golob T.F., Recker W.W., Pavlis Y. Probabilistic models of freeway safety performance using traffic flow data as predictors. Safety Science. 2008; 46: 1306-1333

Haque M.M., Chin H.C., Huang H. Modelling fault among motorcyclists involved in crashes. Accident Anal Prev. 2009; 41: 327-335

Haque M.M., Chin H.C., Huang H. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Anal Prev. 2010; 42: 203-212

Harnen S., Wong S.V., Radin Umar R.S., Wan Hashim W.I. Motorcycle crash prediction model for non-signalized intersections. IATSS Res. 2003; 27(2): 58-65

Hassan H.M., Abdel-Aty M.A. Predicting reduced visibility related crashes on freeways using real-time traffic flow data. Journal of Saf Res. 2013; 45: 29–36

Houston D.J., Richardson L.E. Motorcyclist fatality rates and mandatory helmet-use laws. Accident Anal Prev. 2008; 40: 200-208

Jung S., Qin X., Noyce D.A. Rainfall effect on single-vehicle crash severities using polychotomous response models. Accident Anal Prev. 2010; 42: 213-224

Kasantikul V., Ouellet J.V., Smith T., Sirathranont J., Panichabhongse V. The role of alcohol in Thailand motorcycle crashes. Accident Anal Prev. 2005; 37: 357-366

Lin M.R., Kraus J.F. A review of factors and patterns of motorcycle injuries. Accident Anal Prev. 2009; 41:710-722

Lunn D.J., Jackson C., Best N., Thomas A., Spiegelhalter D. The BUGS Book: A practical introduction to Bayesian Analysis. 2012; 1st ed., Boca Raton, FL., Chapman and Hall/CRC

Nikias, V., Vlahogianni E., Lee T.-C., Golias J. Determinants of Powered Two-Wheelers Virtual Lane Width in Urban Arterials. International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2012). 2012; 1205-1210: Anchorage, USA

Quddus M.A., Wang C., Ison S.G. Road traffic congestion and crash severity: econometric analysis using ordered response models. Journal of Transp Eng. 2010; 136(5): 424–435

Strobl C., Malley J., Tutz G. An Introduction to Recursive Partitioning: Rational, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. Psychol Meth. 2009; 14(4): 323-348

Strobl C., Zeileis A. Danger: High power! – exploring the statistical properties of a test for random forest variable importance. 2008; International Conference on Computational Statistics: Porto, Portugal

The Hydrological Observatory of Athens (HOA). 2012; Retrieved from http://hoa.ntua.gr/

Viano D. Effectiveness of Safety Belts and Airbags in Preventing Fatal Injury. SAE Technical Paper. 1991; 910901~~.~~

Vlahogianni E.I. Kinematic Characteristics and Interactions of Powered-Two- Wheelers Filtering and Overtaking in Urban Arterials. Transportation Res F. 2014; 24: 133-145

Vlahogianni E.I., Golias J.C. Bayesian modeling of the microscopic traffic characteristics of overtaking in two-lane highways. Transportation Res F. 2012; 15: 348-357

Wanvik P.O. Effects of road lighting: An analysis based on Dutch accident statistics 1987-2006. Accident Anal Prev. 2009; 41: 123-128

Xu C., Tarko A.P., Wang W., Liu P. Predicting crash likelihood and severity on freeways with real-time loop detector data. Accident Anal Prev. 2013b; 57: 30-39

Xu C., Wang W., Liu P. Identifying crash-prone traffic conditions under different weather on freeways. Journal of Saf Res. 2013a; 46: 135-144

Xuequn Y., Ke L., Ivers R., Du W., Senserrick T. Prevalence rates of helmet use among motorcycle riders in a developed region in China. Accident Anal Prev. 2011; 43: 214-219

Yu R., Abdel-Aty M. Investigating the different characteristics of weekday and weekend crashes. Journal of Saf Res. 2013a; 46: 91-97

Yu R., Abdel-Aty M. Utilizing support vector machine in real-time crash risk evaluation. Accident Anal Prev. 2013b; 51: 252-259

Yu R., Abdel-Aty M. Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. Accident Anal Prev. 2014a; 62: 161-167

Yu R., Abdel-Aty M. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Safety Sci. 2014b; 63: 50-56

Yu R., Abdel-Aty M., Ahmed M. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Anal Prev. 2013; 50: 371-376

Zheng Z., Ahn S., Monsere C.M. Impact of traffic oscillations on freeway crash occurrences. Accident Anal Prev. 2010; 42: 626– 636

List of Tables and Figures

Table 1: Summary of accident related variables.

Variable	Type	Abbreviation	Description	Descriptive Statistics	
PTW accident involvement	Dummy	MC.involvement.no	Yes=1	326	61.9%
			No=0	201	38.1%
Median	Dummy	Barrier	Yes=1	458	86.9%
			No=0	69	13.1%
Illumination	Dummy	Illumination	Day=1	332	63.0%
			Night/Dusk=0	195	37.0%
Accident Type (collision type)	Dummy	Acc.type	Off road/Fixed object/Other	226	42.9%
			Head-on	43	8.2%
			Rear-end	91	17.3%
			Side	67	12.7%
			Sideswipe	100	19.0%
Road curvature	Dummy	Euthigrammia	Straight line=1	497	94.3%
			Curve=0	30	5.7%
Intersection	Dummy	Intersection	Yes=1	174	33.0%
			No=0	353	67.0%
Traffic control	Dummy	Rythmisi.kukloforias	Traffic lights=1	169	32.1%
			No/Other=0	358	67.9%
Weather conditions	Dummy	Weather	Good=1	489	92.8%
			Adverse=0	38	7.2%
Pavement conditions	Dummy	Pavement.conditions	Good=1	486	92.2%
			Wet=0	41	7.8%

Table 2: Summary of traffic related variables.

				Descriptive Statistics			
Variable	**Type**	**Unit**	**Description**	**Min**	**Median**	**Mean**	**Max**
Q_avg_1h_up	Continuous	Vehicles/hour/lane	1h average flow per lane upstream	52.82	835.64	793.66	1848.91
Q_stdev_1h_up	Continuous	Vehicles/hour/lane	1h st.deviation of flow per lane upstream	10.58	81.81	270.17	1165.50
Q_median_1h_up	Continuous	Vehicles/hour/lane	1h median of flow per lane upstream	13.50	681.63	602.28	1889.13
Q_cv_1h_up	Continuous	unitless	1h coefficient of variation of flow per lane upstream	0.02	0.08	0.11	0.58
V_avg_1h_up	Continuous	Km/hour	1h average speed upstream	4.50	45.90	45.52	104.60
V_stdev_1h_up	Continuous	Km/hour	1h st.deviation of speed upstream	0.00	3.06	5.13	28.08
V_cv_1h_up	Continuous	unitless	1h coefficient of variation of speed upstream	0.00	0.08	0.15	0.89
Occ_avg_1h_up	Continuous	Percentage %	1h average occupancy upstream	0.15	12.94	15.74	57.52
Occ_stdev_1h_up	Continuous	Percentage %	1h st.deviation of occupancy upstream	0.00	2.06	3.96	30.84
Occ_cv_1h_up	Continuous	unitless	1h coefficient of variation of occupancy upstream	0.00	0.17	0.25	1.76

			Descriptive Statistics			
Variable	**Type**	**Description**	**Min**	**Median**	**Mean**	**Max**
Q_avg_1h_up	Continuous	1h average flow per lane upstream	52.82	835.64	793.66	1848.91
Q_stdev_1h_up	Continuous	1h st.deviation of flow per lane upstream	10.58	81.81	270.17	1165.50
Q_median_1h_up	Continuous	1h median of flow per lane upstream	13.50	681.63	602.28	1889.13
Q_cv_1h_up	Continuous	1h coefficient of variation of flow per lane upstream	0.02	0.08	0.11	0.58
V_avg_1h_up	Continuous	1h average speed upstream	4.50	45.90	45.52	104.60
V_stdev_1h_up	Continuous	1h st.deviation of speed upstream	0.00	3.06	5.13	28.08
V_cv_1h_up	Continuous	1h coefficient of variation of speed upstream	0.00	0.08	0.15	0.89
Occ_avg_1h_up	Continuous	1h average occupancy upstream	0.15	12.94	15.74	57.52
Occ_stdev_1h_up	Continuous	1h st.deviation of occupancy upstream	0.00	2.06	3.96	30.84
Occ_cv_1h_up	Continuous	1h coefficient of variation of occupancy upstream	0.00	0.17	0.25	1.76

Table 3: Summary of weather related variables.

Variable	Type	Unit	Description	Descriptive Statistics			
				Min	Median	Mean	Max
T_1h_max	continuous	^{o}C	1h maximum temperature	-1.70	17.85	18.72	42.17
T_1h_avg	continuous	^{o}C	1h average temperature	-2.14	17.23	18.20	41.99
T_1h_stdev	continuous	^{o}C	1h st.deviation of temperature	0.02	0.30	0.39	2.84
Hum_1h_max	continuous	%	1h maximum humidity	12.49	58.96	58.27	97.50
Hum_1h_avg	continuous	%	1h average humidity	12.20	55.37	55.84	97.20
Hum_1h_stdev	continuous	%	1h st.deviation of humidity	0.06	1.37	1.82	13.12
Rain_1h_sum	continuous	mm	1h sum of rainfall	0.00	0.00	0.05	6.60
Rain_1h_st.dev	continuous	mm	1h st.deviation of rainfall	0.00	0.00	0.01	1.20
Rain_2h_sum	continuous	mm	2h sum of rainfall	0.00	0.00	0.15	23.60
Rain_2h_st.dev	continuous	mm	2h sum of rainfall	0.00	0.00	0.02	1.84
Rain_6h_sum	continuous	mm	6h sum of rainfall	0.00	0.00	0.33	35.60
Rain_6h_st.dev	continuous	mm	6h st.deviation of rainfall	0.00	0.00	0.02	1.90
Rain_12h_sum	continuous	mm	12h sum of rainfall	0.00	0.00	0.53	52.80
Rain_12h_st.dev	continuous	mm	12h st.deviation of rainfall	0.00	0.00	0.02	1.67
W.Sp_1h_max	continuous	m/sec	1h maximum wind speed	0.00	2.32	2.72	9.67
W.Sp_1h_avg	continuous	m/sec	1h average wind speed	0.00	1.75	2.16	7.95
W.Sp_1h_stdev	continuous	m/sec	1h st.deviation of wind speed	0.00	0.34	0.39	1.39
W.Dir_1h_avg	continuous	degrees*	1h average wind direction	0.00	141.00	128.36	351.67
Sol_1h_max	continuous	W/m^2	1h maximum solar radiation	0.00	264.75	354.07	1100.00
Sol_1h_avg	continuous	W/m^2	1h average solar radiation	0.00	164.38	284.92	1007.29

*0 degrees is North wind, 90 degrees is East wind etc.

Variable	Type	Description	Descriptive Statistics			
			Min	Median	Mean	Max
T_1h_max	continuous	1h maximum temperature	-1.70	17.85	18.72	42.17
T_1h_avg	continuous	1h average temperature	-2.14	17.23	18.20	41.99
T_1h_stdev	continuous	1h st.deviation of temperature	0.02	0.30	0.39	2.84
Hum_1h_max	continuous	1h maximum humidity	12.49	58.96	58.27	97.50
Hum_1h_avg	continuous	1h average humidity	12.20	55.37	55.84	97.20
Hum_1h_stdev	continuous	1h st.deviation of humidity	0.06	1.37	1.82	13.12
Rain_1h_sum	continuous	1h sum of rainfall	0.00	0.00	0.05	6.60
Rain_1h_st.dev	continuous	1h st.deviation of rainfall	0.00	0.00	0.01	1.20
Rain_2h_sum	continuous	2h sum of rainfall	0.00	0.00	0.15	23.60
Rain_2h_st.dev	continuous	2h sum of rainfall	0.00	0.00	0.02	1.84
Rain_6h_sum	continuous	6h sum of rainfall	0.00	0.00	0.33	35.60
Rain_6h_st.dev	continuous	6h st.deviation of rainfall	0.00	0.00	0.02	1.90
Rain_12h_sum	continuous	12h sum of rainfall	0.00	0.00	0.53	52.80
Rain_12h_st.dev	continuous	12h st.deviation of rainfall	0.00	0.00	0.02	1.67
W.Sp_1h_max	continuous	1h maximum wind speed	0.00	2.32	2.72	9.67
W.Sp_1h_avg	continuous	1h average wind speed	0.00	1.75	2.16	7.95
W.Sp_1h_stdev	continuous	1h st.deviation of wind speed	0.00	0.34	0.39	1.39
W.Dir_1h_avg	continuous	1h average wind direction	0.00	141.00	128.36	351.67
Sol_1h_max	continuous	1h maximum solar radiation	0.00	264.75	354.07	1100.00
Sol_1h_avg	continuous	1h average solar radiation	0.00	164.38	284.92	1007.29

Table 4: Summary of the Bayesian logit model for PTW accident probability.

Variables	Parameters Estimates			Credible Intervals	
	Mean	St.Deviation	Odds Ratio	2.50%	97.50%
constant	-1.321	0.295	0.267	-1.905	-0.740
Q_avg_1h_up	0.001	0.000	1.001	0.001	0.002
V_cv_1h_up	1.249	0.603	3.487	0.077	2.441
Acc.type0 (ref)	-	-	-	-	-
Acc.type1	2.046	0.479	7.737	1.165	3.053
Acc.type2	0.435	0.260	-	-0.071	0.948
Acc.type3	2.021	0.417	7.546	1.244	2.888
Acc.type4	0.976	0.263	2.652	0.478	1.497
DIC	640.095				

Table 5: Descriptive statistics of accident type in relation to PTW accident involvement.

Accident	Accident type				
	Off-road/fixed object	Head-on	Rear-end	Side	Sideswipe
Without a PTW	119	6	37	8	31
With a PTW	107	37	54	59	69

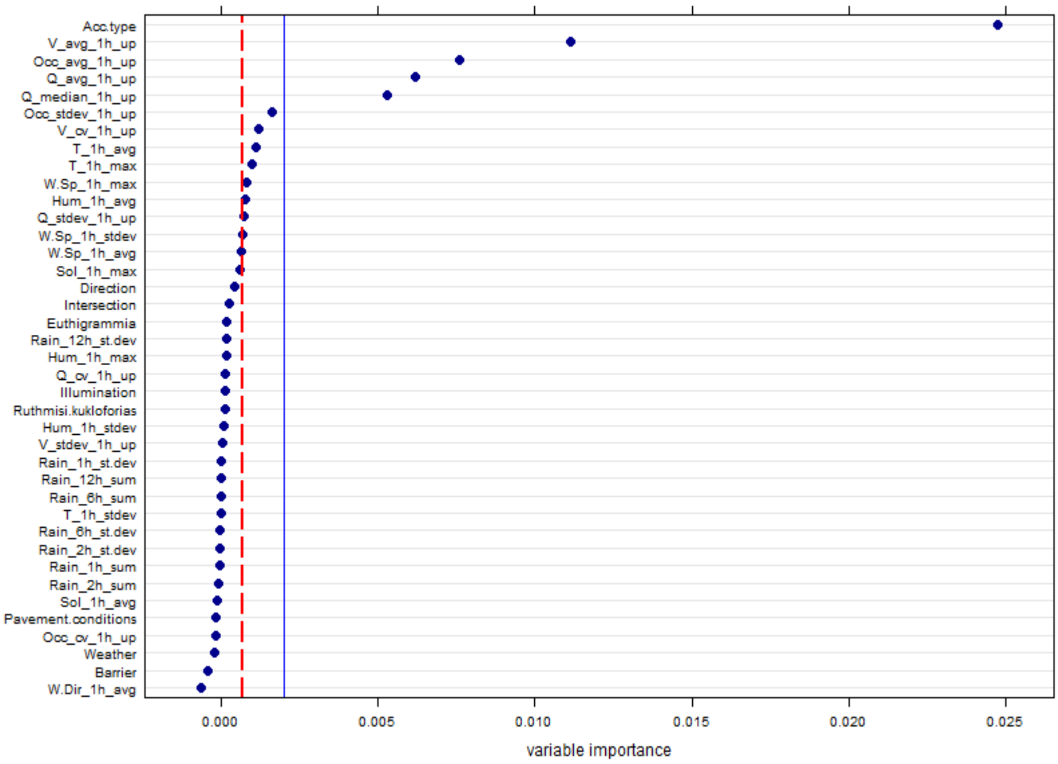

Figure 1: PTW accident probability variable importance provided by Random Forests.

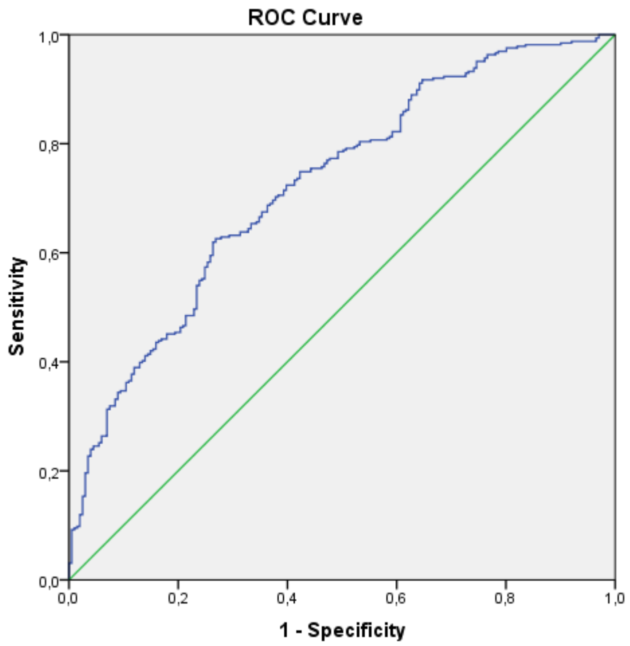

Figure 2: Receiver Operating Characteristic (ROC) curve.