Traffic indicators, accidents and rain: some relationships calibrated on a French urban motorway network.

Maurice Aron*, Romain Billotb, Nour-Eddin EL Faouzib, Régine Seidowskya

a Université Paris-Est, IFSTTAR, COSYS-GRETTIA Marne-la-Vallée 77447, France
b Université de Lyon, F-69000, Lyon, France
IFSTTAR, LICIT, F-69500, Bron
ENTPE, LICIT, F-69518, Vaulx en Velin

Abstract

The purpose of this paper is to study the link between the occurrence of injury road accidents, the prevailing traffic conditions, and the occurrence of rain. This is useful for assessing, before its implementation, the safety impact of a new traffic management. Traffic conditions were extracted from a one year traffic database which covers 150 kilometres of two or three lanes urban motorways near the city of Marseille, in the south of France. 208 loop detectors provide the individual speeds, headways, arrival times and lengths of vehicles. Based on this information, thirteen aggregated traffic variables were constituted every six minutes, such as the average speed, occupancy, short time headways and a few combinations of speed, relative speed and time gaps. 292 injuries or fatal accidents occurred on the network during the same year. The French accident database provides their characteristics - location, time and type of accidents, meteorological conditions and other parameters addressing the infrastructure, the driver and the vehicle. The rain occurrence is provided, every six minutes, from a meteorological station. A set of safety performance functions were estimated, each one giving the risk of injury accident by vehicle-kilometre according to the level of one traffic variable and according to the occurrence of rain. Generally based on logistic regression models, analyses were carried out separately by lane and for two types of accidents - single vehicle accidents and crashes between vehicles. Some relations linking accidents with traffic variables are significant: the occurrence of single vehicle accidents is related to the speed on the fast lane; the occurrence of multiple vehicle accidents is related to occupancy.

Keywords: Traffic data; accident; rain; surrogate data; traffic indicators; urban motorway; risk, safety performance function, logistic regression

* Corresponding author. Tel.: 0033181668687.
E-mail address: maurice.aron@ifsttar.fr

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of Delft University of Technology
1. Introduction

The assessment of a new traffic management, before its implementation, requires to assess the impact of the future values of the traffic variables on accidents. The aim of this paper is to establish the relations which quantified this impact. For safety reasons, a driver adapts his speed, relative speed, time gap and lane, according to the infrastructure (bends, slopes, intersections), traffic conditions (speed of the vehicle ahead or on the adjacent lane, density) and weather conditions. Indeed the performances of vehicles and drivers decrease on slopes and bends, and some danger may come from close vehicles. Despite this adjustment, the accident rate remains related to infrastructure, (Venkataraman et al., 2011), weather (Brodsky and Hakkert, 1988), (Bergel-Hayat et al., 2013), and traffic conditions: Golob et al. (2004) found different accident rates according to the type of crash and traffic conditions, to the temporal variations in volume and speed; Abdel-Aty et al. (2005) used traffic conditions as accident precursors; (Park and Saccomanno, 2007) highlighted speed variation. According to Nilsson (1984) or Elvik (2013), the risk is a continuous power function of the speed; it is an exponential function according to Hauer (2009) and Elvik (2014).

Relations between traffic conditions, infrastructure elements and accidents are discrete or continuous. Discrete relations provide the risk by vehicle-kilometer and by class. According to the authors, a class may combine traffic, weather and infrastructure conditions. Making a discrete relation "continuous" is possible by varying the thresholds defining the classes- see (Aron et al., 2013). This paper focuses on continuous relations. When appropriate, they give a quick understanding of the risk and might be included in simulations or in traffic management algorithms.

The role of speed in road safety has been demonstrated. "Speed" refers to different quantities: the speed limit at a national level, on a network, on a particular section; speeds of individual drivers recorded at particular points, or their distribution on a route; average speed on a spatial range; temporal average at a given point, by lane, or for all lanes. According to what "speed" is, the analytical pattern, the numerical values, the relevance of models vary.

In the following, we present the data in section 2 and then some continuous models, which relate the risk to different traffic indicators (section 3). We give in section 4 the relationships which proved to be significant; we discuss their limits in section 5. Section 6 is the conclusion and perspectives. Numerical results are in the Appendix.

2. Traffic, Accident and Meteorological Data

2.1. Traffic Data

CEREMA, the French centre for studies on risk, mobility and environment, collected, from June 2009 to May 2010, time-stamped data in terms of lengths, speeds of vehicles on the “Marius” network.

This network is 150 km-long. It consists in the urban parts of motorways A7, A50, A51 and A55 (in red on the Figure) around Marseille. The sections have either two lanes per direction (here called middle and slow lanes) or three lanes (fast, middle and slow lanes); 104 available traffic stations by direction, (one station every 750 meters) are available on the main carriageway and on the ramps. Missing data being not too numerous, the traffic pattern based on available data is assumed to be representative. The traffic counts and the distance between sensors lead to the estimation of 1.5 billions of vehicle-kilometres. 5.3% vehicles travel during rain, and 15% during night time (defined here between 8PM to 6AM).

Fig. 1. The Marius urban motorway network, near Marseille (France); A weather station is located in Marignane airport, at North-West.

2.2. Accident Data

The French Police collects the characteristics of all road fatalities and injury accidents. A few characteristics are used here - date, hour, minute, accurate location, number of implied vehicles, atmospheric conditions.
Over one year, 292 injury accidents or fatalities were recorded on the Marius network. Missing data affects 18 accidents for which the direction of the accident is missing, or its location, or the traffic data shortly before.

Relationships have been estimated by lane because averaging the traffic indicators between lanes may hide certain phenomena as heterogeneous speeds or densities between lanes. However, the lane where an accident begins is generally unknown, except for those occurring when inserting from or to a ramp, which is mentioned in the database (15 accidents, including 10 daytime accidents). These accidents are not included in the estimations of the relations for the middle and the fast lane. For every accident, traffic data from the first upstream sensor, when available, is considered; when unavailable, traffic data at the first downstream sensor, or at the second upstream or downstream sensors, are considered. Traffic conditions may change at the moment of the accident; also it is mandatory, when estimating a relation between accidents and traffic conditions, to use traffic conditions before the accident. However, accident times are only estimated by the police on their arrival on site afterward. Also, for every accident, we examined the series of speeds recorded at the upstream sensor for forty minutes until the time estimated by the police. When traffic conditions did not change, we considered the 6-minute period ending before the accident time (taking into account a time offset equal to the average travel time between the sensor and the accident location); when one unique drop in speed occurred, we considered the 6-minute period ending before the drop; when several drops were recorded, we selected the period ending eighteen minutes before the accident.

During night time, there are 20% of accidents for only 15% of the vehicle-kilometres travelled. This means a risk increase at night which must be modelled or, as it is done here, a separate analysis for daytime and night time must be achieved. 218 daytime accidents remain, including 44 accidents (20%) where a Power-Two-Wheeler (PTW) is involved. Accidents are split in three classes; the third one consists of the 25 daily accidents where a contributing factor appears, which is linked to the vehicle (breakdowns), to the driver (drowsiness, alcohol), or to the presence of a pedestrian; these accidents have been discarded from the analysis. The first class addresses the 46 daytime accidents where only one vehicle is involved; the second class addresses the 147 daytime accidents where two or more vehicles are involved. These numbers apply for the slow lane. The numbers for the first and the second class are respectively 44 and 128 on the middle lane because accidents occurring near a ramp are no more considered, and because of missing data. There are only 39 and 117 for the fast lane (defined in France as the third lane from the right), because there is no “fast” lane hence no data related to this lane for accidents occurring on a two-lane motorway section.

2.3. Meteorological Data

The occurrence of rain, at the moment and place of the accident, is recorded in the accident database; in the case of no-accident, the rain occurrence is provided, every six minutes, at the meteorological station of Marignane, Marseille airport, at less than 30 km of all points of the network. This station is managed by Météo-France, the French Agency in charge of meteorological forecasts. During rain, the percentage of injury accidents or fatalities (13%) is greater than the percentage of vehicles-kilometers travelled (5.3%); this confirms the danger due to rain.

3. Statistical Models and Traffic Indicators

3.1. Eight Models

Safety Performance Indicators are generally based on negative binomial models, or on distributions like the Poisson-Maxwell-Conway distribution which better fit the dispersion (Lord and Guikema, 2012); here we regress the risk "R" of accident by vehicle kilometer by 2 methods (linear regression on the logarithm, logistic regression). Three types of relations have been tested - power, exponential, and exponential with a parabolic term.

3.2. Power model

In the power model, generally applied to speed only, risk is proportional to an exponent of the traffic indicator. The name of the model is constituted by a part indicating its pattern (here “POWER”), by a number designating its equation number (here 1), and by the subscript “Rain” or “NoRain whether it includes or not a term modeling rain.
\( R_i \) is the risk by vehicle-kilometer, \( V_i \) is the speed for observation \( i \), \( \text{Rain}_i = 1 \) if rain occurs during observation \( i \) (0 otherwise); \( \alpha, \beta \) and \( \gamma_{\text{Rain}} \) are deterministic parameters to be estimated. \( n \) is the number of observations, defined as follows: records are grouped in the same “observation” when they have the same weather condition (rain or not), the same speed group named by its center \( V_i \). Speed groups are determined in two steps. First, from 100 pre-defined speed thresholds, 100 speed classes are formed. Second, groups of consecutive classes are formed in order to have at least one accident in the group. This constitutes an observation. The number of accidents (for a given type of accidents) and the number of vehicle-kilometers are associated to this observation.

This type of model applies also for other relationships, where \( V_i \) is replaced by the traffic indicator of another variable, with an analogous process for forming the observations.

We introduce “logistic” power models, linking the logit of the risk \( \log[R_i/(1-R_i)] \) to the indicator (\( V \)) and to the occurrence of rain. Computing confidence intervals on parameters requires an assumption on the distribution of deviations - here assumed to be Gaussian. The second model is written as follows, using \( \alpha \) as the logarithm of \( \alpha_V \):

\[
(\text{POWER2}_{\text{Rain}}): \log[R_i/(1-R_i)] = \log[R_i] - \log(1-R_i) = \alpha + \beta \log(V_i) + \gamma_{\text{Rain}} + \varepsilon_i. \tag{2}
\]

Dropping the term \( \log(1-R_i) \) of a logit model gives a logarithm form which is estimated using a linear regression; the name given to such models is the name of the logit model, plus the subscript "Bis": “\( \text{POWER2}_{\text{Rain Bis}} \).”

Hauer (2009) and Elvik (2014) proposed an exponential model; we propose a model including the rain impact:

\[
(\text{EXPO3}_{\text{Rain}}): R_i = e^{\alpha + \beta V_i + \gamma_{\text{Rain}} \cdot \text{Rain}_i} + \varepsilon_i. \tag{3}
\]

The associated logistic model reads \( \text{EXPO4}_{\text{Rain}}: \log \left[ R_i / (1 - R_i) \right] = \alpha + \beta V_i + \gamma_{\text{Rain}} \cdot \text{Rain}_i + \varepsilon_i \) \( (V_i) \) being positive, its square is an increasing function, and replaces \( (V_i) \) without changing the sign of \( \beta \) in:

\[
(\text{EXPO5}_{\text{Rain}}): \log \left( R_i / (1 - R_i) \right) = \alpha + \beta V_i^2 + \gamma_{\text{Rain}} \cdot \text{Rain}_i + \varepsilon_i \tag{5}
\]

\( \log^2(V_i) \) has been successfully tested for some indicators (the percentages); it is a decreasing function when \( V_i \) is less than 1; this would imply a change of interpretation of the sign of \( \beta \), unless considering \( - \log^2(V_i) \) as follows:

\[
(\text{MODEL6}_{\text{Rain}}): \log \left( R_i / (1 - R_i) \right) = \alpha - \beta \log^2(V_i) + \gamma_{\text{Rain}} \cdot \text{Rain}_i + \varepsilon_i \tag{6}
\]

### 3.3. Parabolic models (excluding rainy conditions)

When risk is not monotonous with the traffic indicator, modelling requires one more parameter. However, due to an insufficient number of accidents, it was not possible to estimate four parameters; models are then estimated here on datasets excluding rain, allowing removing the rain coefficient; “\( \gamma \)” designs the new coefficient – the coefficient of the square of the indicator in parabolic models, which takes into account the traffic indicator and its square; the direction of variation of the risk depends on whether the traffic indicator \( V \) is below/above the value \(-\beta/(2 \cdot \gamma)\).

\[
(\text{POWER7}_{\text{NoRainPara}}): \log \left( R_i / (1 - R_i) \right) = \alpha + \beta \cdot \log(V_i) + \gamma \cdot V_i^2 + \varepsilon_i \tag{7}
\]

\[
(\text{EXPO8}_{\text{NoRainPara}}): \log \left( R_i / (1 - R_i) \right) = \alpha + \beta V_i^2 + \gamma \cdot V_i^2 + \varepsilon_i \tag{8}
\]
The name of the model includes the subscript "Para" for parabolic models, or/and "Ramp" when the model is estimated on the slow lane, including accidents occurring near a ramp.

3.4. Statistical processing

The traffic and accident databases were separated in different cases, according to the time of the day (night time/daytime), lane, weather (rain or not). Independent analyses are performed for every combination of cases. Three accident datasets were considered: the whole dataset, or excluding accidents during rainy conditions, (disregarding the traffic data when raining), or accidents excluding those implying Power Two Wheelers (PWT).

The logistic regression Generalized linear Model (GLM) was used with the software R®; it processes the vectors of number of accidents \((Acc_i,)\) and number of vehicles-kilometers \((N_i)\) by observation. The estimations of linear regressions (models "bis") were achieved without weighting the observations, the endogenous variable being the logarithm of the risk.

3.5. Thirteen Traffic Indicators

Even with individual data on upstream sensors, it is impossible to identify, among others, the driver responsible for the accident. We seek to know if the parameters of the whole set of drivers are peculiar just before an accident. The thirteen indicators presented here are computed from the aggregation of traffic data over six-minute periods:

1. Average speed \(V_i\), by six-minute period in kilometer/hour,
2. Occupancy – it is the percentage obtained by summing the “occupancy times” (in seconds) of vehicles passing in a 360-second period, and then by dividing the sum by 360; the occupancy time of vehicle \(j\) of length \(L_j\) and speed \(V_j\) is equal to \(3.6(L_j+\lambda)/V_j\), \(\lambda\) being equal to 1 meter, the length of the inductive loop; the unit factor is "3.6"
3. Relative speed in kilometers/hour is the difference between the speeds of two consecutive vehicles on the same lane. The sum of relative speeds on a period is not of interest, because the speed of a vehicle generally appears twice in the sum with opposite signs, thus disappears. Negative relative speeds, having no safety impact, were disregarded. The indicator proposed here consists in the sum of relative speeds, when positive, divided by the traffic count.

Indicators 4,5,6,7: Time headway is here the difference between the arrival times at the sensor of the fronts of two successive vehicles. Indicator 4 is the six-minute average time headway; indicators 5-7 are the percents of tailgating (less than 2 seconds), short (less than 1 second) and very short headways (less than 0.5 second).

Indicators 8,9,10,11: Uno et al., 2003) defined the “PICUD” (Potential Index for Collision with Urgent Deceleration). It is the estimated difference (in meters) between the stopping locations of vehicle “j” and “j-1”, assuming that the leader \(j-1\) brakes at the very instant \(t^i\) when the follower “j” passes the sensor - the leader is then located at \(V_{j-1}^i\cdot(t^i-t_{j-1}^i)\) meters downstream the sensor; its rear end is \(L_{j-1}^i\) meters before, assuming that the traffic sensor records times of passage of the front of vehicles. When the PICUD is negative, there is a collision danger, that would have been avoided had the follower a space gap greater by \{-PICUD\} before the braking of the leader. The follower brakes with the same deceleration (here the deceleration is \(\Gamma=6.25\ m/s^2\)) at time \(t^{j+T}\), after a reaction time \(T=1\) second, when the vehicle is located at \(V^{j+T}\). T meters downstream the sensor:

\[
\text{PICUD}^j = \left[\left(V_{j-1}^i \cdot V_j^i\right)^2 / (2 \cdot \Gamma) - \left(t^i - t_{j-1}^i\right) - V_j^i \cdot L_{j-1}^i\right] \cdot \left(1 - V_j^i / V_{j-1}^i\right) + \left(t^i - t_{j-1}^i\right) - V_j^i \cdot L_{j-1}^i \right] \ (9)
\]

The length \(L_j^i\) of the second vehicle replaces the length \(L_{j-1}^i\) of the leader in equation (9) when the sensor records times of passage of the rear-end of vehicles.

Indicator 8 is the absolute value of the sum of negative PICUD, on the six-minute period, divided by the traffic count of the period. Indicators 9, 10, 11 are the percentages of drivers for which the PICUD is below a threshold (respectively 0 meter, minus 10 meters, minus 20 meters).
Indicators 12-13: \( \text{PICUD} \) becomes “\( \text{PICUDbis} \)”, by removing the reaction time in equation (9). Indicator 12 is the sum of negative \( \text{PICUDbis} \), divided by the traffic count. Indicator 13 is the percentage of negative \( \text{PICUDbis} \).

### 4. Significant relationships

Risks of single-vehicle accidents and of crashes were related with the average speed in section 4.1, occupancy in section 4.2, time headway in section 4.3, relative speed in section 4.4 and \( \text{PICUD} \) indicators in section 4.5. The logistic form of the models was generally preferred. Results are presented by lane, with or without the impact of rain, only for daytime accidents and traffic. Numerical values of the parameters of logistic regressions are in the Appendix, Table A.1. The lines of this table are numbered for easy reference - the lines whose numbers end with the suffix “ _1” stands for the risk of single-vehicle accidents, while suffix “ _2” is related to crashes between vehicles. The relationships the P-value of which is less than 5%, are considered as significant. However, as the number of accidents in our database is not large, hoping to reveal interesting relationships, we considered also relationships until a P-value threshold of 10%; these relationships will have to be confirmed on a larger database.

#### 4.1. Risk and 6-minute average speed (daytime)

(I) Models with a rain term

The risk of single-vehicle accidents (daytime) is significantly related to the six-minute average speed of the fast lane by power and exponential relationships. Rain is a significant contributing factor. Model (\( \text{POWER2_Rain} \)) provides a rain exponent \( \gamma_{\text{rain}}=1.69 \) (its standard deviation \( \sigma_r \) being 0.33); the speed exponent is \( \beta=2.66 \) (\( \sigma_{\beta}=1.3 \)) with 38% of deviance explained; when Power Two Wheelers accidents are excluded, \( \beta \) jumps to 7.86 (\( \sigma_{\beta}=2.2 \)) with 52% of the deviance explained. The exponential model (\( \text{EXPO4_Rain} \)) gives a speed exponent \( \beta=0.03 \) (\( \sigma_{\beta}=0.014 \)) which jumps to 0.073 (\( \sigma_{\beta}=0.020 \)) when PWT accidents are excluded, see Table A.1 in Appendix, lines 1_1 to 4_1.

These exponents are high compared to value two, given by Nilsson (2004); recall that "speed" is not the same, it is related to the fast lane only; the relation applies only to single-vehicle accidents, which reduces its extent.

On the middle lane, no logistic model appears significant. However the second lane (from the right) of two-lane sections and of three-lane sections, are grouped in this work as the “middle” lane, which brings some heterogeneity. With the metrics induced by an ordinary non-weighted linear regression, when excluding PTW, significant exponents \( \beta=3.34 \) (\( \sigma_{\beta}=1.4 \)) and \( \beta=0.034(\sigma_{\beta}=0.015 \)) were obtained respectively with model (\( \text{POWER2_RainBis} \)) and (\( \text{EXPO4_RainBis} \)). However we did not check the normality of the errors or their heteroscedasticity.

On the slow lane, no significant relationship was found. Speed is less homogeneous on this lane, due to the presence of trucks and ramps. The speed average, an average of inhomogeneous quantities, is not a good indicator.

Crashes between vehicles are not significantly related to speed, likely because, during rain, there is concomitantly a decrease in speed and an increase in risk.

(II) Normal weather conditions, excluding traffic and accidents during rain

By normal weather, the part of deviance explained is only due to the traffic indicator, and is generally smaller than when the rain impact is added.

On the fast lane Model (\( \text{POWER2_NoRain} \)) remains significant only when PTW are excluded, with a percentage of explained deviance of 26%, instead of 52% in (\( \text{POWER2_Rain} \)), and a slightly smaller exponent \( \beta \), (Table 1, line 5_1). Model (\( \text{EXPO4_NoRain} \)) is significant, also with smaller \( \beta \), as well as with and without PTW, (lines 6_1 &7_1). Many PTW postpone their trip when it rains, which decreases their risk exposure during rain, the comparison of risks between rainy and no-rainy conditions is fully justified only when excluding PTW.

On the middle lane, the negative \( \beta \) (Table 1, line 8_1, model \( \text{EXPO4_NoRain} \)) is surprising, see the discussion in section 5 on the limits of average indicators.

On the slow lane, excluding accidents involving a PTW, (\( \text{EXPO8_NoRain_ParBis} \)), a non-weighted parabolic linear model, implies an increase of risk with speed only until speed reaches 88 km/h (\( \beta=0.467(0.27) \) \( \gamma=-0.0026 (0.0016) \)).

Crashes between vehicles are related to the speed with negative exponents \( \beta \). This might mean that the presence of numerous low-speed vehicles (trucks, entering vehicles) brings some danger. It might just mean that, when speed decreases, accidents are rather crashes between vehicles than single-vehicle accidents.
(III) Three tentative conclusions.
First, the average speed is a better indicator when excluding PWT. Second, exponential models are robust. Third, speed must be estimated on the fast lane.

4.2. Crashes between several vehicles and occupancy, daytime

Positive coefficient $\beta$ are obtained, which means that the more occupancy, the more the crash rate. Including accident related to a ramp, on the slow lane, model (EXPO5_Rain), with a rain coefficient $\gamma_{\text{Rain}}= 0.5$, explains a limited part (20%) of the deviance of the logit of the crash risk - Table 1, line 9_2.
Excluding rainy conditions, model (EXPO4_NoRain) explains 23% of the deviance on the fast lane; this part is higher (36%) on the middle and slow lanes with Model (POWER2_NoRain) - Table 1, lines 10_2 to 13_2.

4.3. Average time headway and percentages of time headway less than 2, 1 or 0.5 seconds

(I) Average time headway (TH)
The average time headway (TH) is proportional to the inverse of the traffic count; a high TH (thus a low traffic flow) occurs either by low traffic demand, or by congestion. The first case, being the most frequent, implies that, for high TH, vehicles are far from each other, the higher the TH, the lower the rate of crashes. Negative $\beta$ and positive $\gamma_{\text{Rain}}$ are actually obtained on the middle lane, with a very limited part of explained deviance (Table 1, lines 15_2&16_2). The lower the flow, the higher the TH, the speed and the rate of single-vehicle accidents. A positive $\beta$ explaining the single vehicles accidents is thus expected and obtained on the three lanes for all models. Table 1, line 14_1 gives the results of Model (POWER2_Rain) for the slow lane; the coefficients are similar for other lanes; the percentage of explained deviance varies from 39% to 50% according to the lane.

(II) Percentage of time headway less than two seconds, one second or half a second

The higher the percentage of short headway, the lower the rate of single-vehicle accidents. $\beta$ exponents are negative (Table 1, line 17_1); when including rain, 38% to 68% of the deviance, according to the lane, are explained.
Relationships between crash risk and the three percentages of time headway are either insignificant, or explain a very small part of the deviance, except when a parabolic term is added. Excluding rainy conditions, parabolic models as (EXPO8_NoRain) or (POWER7_NoRain) (with $\beta <0$ and $\gamma>0$), indicate for all lanes and for the three percentages, that the crash risk is high, both for low and high percentages of short time headway, see (Table1, line 18_2 to 26_2).

4.4. Risk and average of positive relative speed

(Abdel-Aty et al., 2004) and (Abdel-Aty and Pemmanaboina, 2006) found that the speed variance (obtained from a series of ten consecutive 30" average speed) was a relevant accident precursor. This means that changes in 30" traffic conditions are correlated to accident occurrence. This does not mean that heterogeneous speeds (between a vehicle to the next one on the same lane) are correlated to accidents and that more homogeneous speeds (as obtained with an adaptive speed control device) would be safer; also we tried to check this point with our relative speed indicator, which measures this speed heterogeneity, but, on our limited dataset, we found little evidence on this point.

The average of positive relative speeds explains, according to the lane, with model (POWER2_Rain), from 27% to 64% of the deviance of the single-vehicle accidents; $\beta$ are positive; it explains 33.6% and 45% of the deviance of the crash risk on the middle and the slow lanes; then $\beta$ are negative. This might come from correlations between relative speed and traffic density, and between traffic density and accident type. Rain coefficients are positive (Table 1, lines 27_2 to 29_2). Parabolic models POWER7NoRainPara or EXPO8NoRainPara, estimated excluding rainy conditions, have a positive $\gamma$ coefficient, implying an increase of risk on the middle and slow lanes when the relative speed indicator is greater than 1.5 km/h - Table 1, lines 30_2 and 31_2.

4.5. PICUD-based indicators

Single-vehicle accident occurrence is related, with positive $\gamma_{\text{Rain}}$ and $\beta$ coefficients, to PICUD-based indicators:
• the absolute value of the average of negative \( \text{PICUDbis} \) on the fast lane (line 32_1),
• the percentage of \( \text{PICUD} \) less than -20 meters (fast lane, without Power Two Wheelers, line 33_1),
• the percentages of negative \( \text{PICUDbis} \) on the fast and middle lane (lines 34_1 and 35_1).

Crashes between vehicles occurrence is related to rain with a positive coefficient and, with a negative \( \beta \), to the percentages of \( \text{PICUD} \) less than -10 meters or -20 meters; however these relations are obtained only when including accidents related to the ramp entrance/exit (which was tested on the slow lane only, Table 1, lines 36_2&37_2); no significant relation appears with the averages of negative \( \text{PICUD} \) or \( \text{PICUDbis} \).

Crashes between vehicles occurrence, when rain is excluded, is related with a negative \( \beta \) to the percentages of \( \text{PICUD} \) less than -10 meters or -20 meters; however these relations are obtained:
- on the fast lane when excluding PTW (MODEL6NoRain,line 38_2), with a small part of deviance explained,
- on the slow and the middle lane, when adding a parabolic term (EXPO8NoRainPara, , lines 39_2 and 40_2).

This last model also links the crash risk and the percentage of negative \( \text{PICUD} \), only for the fast lane, when excluding rainy conditions; however the part of explained deviance (26\%) is limited, see Table 1 line 41_2. lane.

5. Discussion

Accidents are due to an inappropriate speed, relative speed or lane change at an individual level; however, at an aggregated level, a correlation between risk and a traffic variable might come via another correlated variable. Different interpretations/misinterpretations have to be considered. They are discussed here only in the case of speed:
(1) Continuity between individual and aggregated traffic values: a high average speed results from many risky drivers with high individual speeds: it is likely, but not more, that a high average speed results into a higher risk.
(2) Misinterpretation of "a low speed" which can be inappropriate if not sufficiently low. During rain, the speed decreases, but too slightly, hence the risk increases. If the presence of rain was not identified, a misleading negative correlation between risk and speed would appear. All contributing factors have to be identified, then introduced.
(3) Inadequacy of the arithmetic mean due to the sensitivity to the extreme of the distribution (the risky drivers). Accidents occur due to multiple factors: presence of rain, of a curve, of an access ramp... If most drivers, except the risky driver, reduce their speed, this will imply a negative correlation between average speed and risk.
(4) The classification in single/multiple vehicle accidents: even if speed had no action on risk, speed would have a positive correlation with single-vehicle accident risk and a negative one with crashes: this is due to the negative correlation between speed and density, combined with the correlation between density and type of accidents: indeed, by low density, few vehicles are close to a vehicle; thus accidents are rather "single-vehicle" accidents.
(5) Non monotonous relations. Crashes increase with traffic density, but for very high density, some crashes only lead to material damage (excluded here). The pattern of the relationship linking the risk to the traffic indicator is not monotonous when several basic phenomena occur. Progresses on that topic should come by refining the analysis, with more accident data, allowing a more disaggregated level.
A lack of correlation may come from the absence of any relation, from confounding various opposite effects, from insufficient data or insufficient analysis framework.

6. Conclusion and perspectives

Some variables are significantly linked to accidents, thanks to interpretable relationships:
- for single vehicle accidents, the 6-minute average speed on the fast lane; and the time headway (on every lane),
- for multiple vehicle accidents, the occupancy, and the time headway (for the middle lane).

It is more difficult to understand other relations linking accidents to relative speed or to the "PICUD", a variable based on the collision computation. These relations might come from a correlation between relative speed or the PICUD with the traffic density, followed by a mechanical effect of the level of density on the type of accident. Besides, the behavior of a single driver, responsible of an accident, does not systematically appear at an aggregated six-minute level.
Such relations, correctly validated and integrated in traffic management tools, should be useful to anticipate the safety impact of a new traffic management scheme.

The perspectives of this work are three-fold: improving the data processing, validating and assessing the predictive power of such relationships as accidents precursors, integrating them into traffic management systems.

Improving data processing is at three levels:

1. Improving data. The number of accidents considered is not sufficient to validate the relations; the time of the accident is not accurate; the distance between two successive traffic stations is too high to catch very local problems; data are static, and do not describe the beginning or the ending of a bottleneck; PTW are not counted.

2. Data processing. Some analyses must be disaggregated, according to the type of section, to the infrastructure.

3. Traffic indicators. Other relevant indicators should be added (Time to Collision, Post-Encroachment Time...). Selected percentiles might replace the 6-minute average in the indicators. Inter-lane indicators, based on relative speed between lanes and on gap availability, and platoon indicators, should be introduced. Future research should also include the development of models that take into account the various contributing variables in the same model, and examining interactions between the variables.

Validating these relationships on other periods is mandatory, as well as studying their transferability onto other sites. The power of such relationships as "accident precursors" should be assessed first by analyzing the rates of false alarm and of "no detection" they imply (Abdel-Aty et al., 2005) and second by checking whether, beyond correlation, the traffic indicators really contribute to the risk. In the case where this power is sensible, the final step would consist in integrating such relationships as safety criteria in traffic management algorithms.

Acknowledgements

We thank the reviewers for their relevant remarks. We acknowledge the "COMET" project (http://orsi-comet.org/index.php/), launched by IFSTTAR, which brings together different participants in order to share knowledge and to develop tools for traffic management during adverse meteorological conditions.

References


Elvik, R. 2013. A re-parametrisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims. Accident Analysis & Prevention,50 854-860.


Appendix. Numerical results associated to significant relationships

Table A.1. Numerical values associated to significant relationships obtained with a logistic regression

<table>
<thead>
<tr>
<th>Line</th>
<th>Variable</th>
<th>Lane</th>
<th>Model</th>
<th>α (σ(α))</th>
<th>z(α)</th>
<th>β (σ(β))</th>
<th>z(β)</th>
<th>( Y_{\text{Rain}} ) or ( Y_{\text{NoRain}} )</th>
<th>Deviance</th>
<th>Null Residual</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_1</td>
<td>Speed</td>
<td>Fast</td>
<td>POWER2Rain</td>
<td>-54.0 (10.3)</td>
<td>0.0%</td>
<td>7.86 (2.21)</td>
<td>0.0%</td>
<td>1.88 (0.00)</td>
<td>33</td>
<td>56.6</td>
<td>102.4</td>
</tr>
<tr>
<td>2_1</td>
<td>Speed</td>
<td>Fast</td>
<td>POWER2Rain*</td>
<td>-20.2 (1.4)</td>
<td>&lt; 0.03 (0.01)</td>
<td>2.5%</td>
<td>1.70 (0.00)</td>
<td>42</td>
<td>59.9</td>
<td>36.3</td>
<td>130.7</td>
</tr>
<tr>
<td>3_1</td>
<td>Speed</td>
<td>Fast</td>
<td>EXPO4Rain</td>
<td>-25.1 (2.2)</td>
<td>&lt; 0.07 (0.02)</td>
<td>0.0%</td>
<td>1.88 (0.00)</td>
<td>33</td>
<td>56.6</td>
<td>27.3</td>
<td>102.5</td>
</tr>
<tr>
<td>4_1</td>
<td>Speed</td>
<td>Fast</td>
<td>EXPO4Rain*</td>
<td>-48.2 (11.8)</td>
<td>0.0%</td>
<td>6.61 (2.51)</td>
<td>0.9%</td>
<td>-</td>
<td>56</td>
<td>28.9</td>
<td>58.7</td>
</tr>
</tbody>
</table>

(*) excluding PTW accidents

Lines i_1 refer to single-vehicle accidents; Lines i-2 refer to crashes between vehicles; \( n \) is the number of "observations"