Power and Exponential Functions Relating Accidents to Traffic and Rain. Calibration on a French Network.

Maurice Aron, Romain Billot, Neila Bhouri, Nour Eddine El Faousi, Régine Seidowsky

Power and Exponential Functions Relating Accidents to Traffic and Rain. Calibration on a French Network.

Maurice Aron[1]*, Romain Billot[2], Neïla Bhouri[1], Nour-Eddin El Faouzi[3], Régine Seidowsky[1]

[1] Université Paris-Est, IFSTTAR, COSYS-GRETTIA Marne-la-Vallée, France
[2] Telecom-Bretagne, Plouzané, France
[3] Université Lyon, ENTPE, IFSTTAR, LICIT UMR_T9401, Lyon, France
mauraron@gmail.com, romain.billot@telecom-bretagne.eu, (neila.bhouri, nour-eddin.el-faouzi,regine.seidowsky)@ifsttar.fr

Abstract. Relations between the occurrence of road accidents, traffic and rainfall conditions are valuable in setting safety objectives for traffic management, and in assessing the safety impacts of new traffic management systems, prior to their implementation. Based on traffic, road accidents and rain data collected over one year, on a French urban motorway network, a set of safety performance functions were estimated; each of them provides the accident risk per vehicle- kilometer for a certain type of accident, according to the occurrence of rain, and to the level of a traffic variable (average speed, occupancy, percentage of tailgating...). Analyses were carried out separately by lane and for two types of accidents: single-vehicle accidents and multiple-vehicle accidents.The relationships, although statistically significant, have yet to be validated by the treatment of another set of accidents**.**

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

1 Introduction

In order to assess a new traffic management, before its implementation, it is necessary to assess the impact of the future values of the traffic variables on accidents. The aim of this paper is to establish the relationships which quantified this impact. For safety reasons, drivers adapt their 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. It must be remembered that the performance of vehicles and drivers decreases on slopes and bends, and some danger may come from close vehicles. Despite this adjustment, the accident rate remains related to infrastructure, [1], weather [2,3], and traffic conditions [4].

Relations between traffic conditions, infrastructure elements and accidents are discrete or continuous. In discrete relationships, the risk per vehicle-km is computed by traffic flow regime, a traffic flow regime being a homogeneous group of traffic flow

adfa, p. 1, 2011.

conditions, for different weather and infrastructure conditions. Golob *et al.* [4] 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.* [5] used traffic conditions and rain occurrence as accident precursors; [6] highlighted the impact of speed variation on accidents.

In continuous relationships, the rate of accident is a continuous function of the traffic variable. According to Nilsson or Elvik, the risk is a continuous power function of the speed [7,8]; Hauer and Elvik also proposed an exponential function [9, 10]. When appropriate, continuous relationships give a quick understanding of the risk and might be included in simulations or in traffic management algorithms.

The objective of this paper is to model the relationship between the risk of accidents and the traffic, according to the findings of Golob *et al.* [4] or Abdel-Aty *et al.* [5], who identified, among other things, the ways in which congestion affects road safety. Our approach is also in the continuity of Nilsonn, Elvik, Hauer [7,8,9,10], focusing on continuous relationships, and aims to demonstrate how speed, density and some other traffic variables are related to the occurrence of accidents.

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. Depending on what "speed" is, the analytical pattern, the numerical values, and the relevance of models vary. In this paper, we consider the 6-minute average speed observed. Nilsonn as well as Elvik and Hauer models have been tested on some French interurban motorways; they have been adapted to take into account the different motorway lanes and rain. Other models have also been considered.

In the following, we present the data in section 2, followed by some continuous models, which relate the risk to different traffic indicators, in section 3. Section 4 describes the relationships which have proved to be significant, while their limits are discussed in section 5. In section 6, the relationships between the obtained risk models are developed further. Section 7 contains the conclusion and perspectives; main numerical results are to be found in the Appendix.

2 Traffic, Accident and Meteorological Data

Meteorological Data. The occurrence of rain, at the time 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, less than 30 km from all points of the network. This station is managed by Météo-France, the French Agency in charge of weather forecasts. During rain, the percentage of injury accidents or fatalities (13%) is greater than the percentage of vehicle-kilometers travelled (5.3%); this confirms the danger due to rain.

Traffic Data. Between June 2009 and May 2010, the French centre for studies on risk, mobility and environment, CEREMA, collected traffic data (vehicle lengths, speeds) on the "Marius" network. This network is 150km long and is made up of the urban parts of motorways A7, A50, A51 and A55 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. Data are recorded every six minutes; the whole one-year database was used for calibrating the traffic-safety relationships. Given that there is not much missing data, the traffic pattern based on available data is assumed to be representative. A weather station is located in Marignane airport, in the North-West.

Fig. 1. The Marius urban motorway network, near Marseille (France)

The traffic counts and the distance between sensors lead to an estimated 1.5 billion vehicle-kilometers. 5.3% vehicles travel during rain, and 15% travel at night. Although defining night as being from sunset to sunrise would be a more accurate reflection of the contribution of darkness to the occurrence of accidents, here night is always defined as being from 8pm to 6am, which brings some homogeneity for the types of travel and driver, and for their tiredness.

Accident data. The French police collect the characteristics of all road fatalities and injury accidents. A few characteristics are used here: date, hour, minute, precise location, number of vehicles involved, rain information.

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. Table 1 gives the distributions of the remaining 274 accidents, according to rain and presence of a PTW.

Table 1. Distribution of accidents, according to the presence of rain and PTW

	Rain	No_Rain	Total	PTW	No PTW
Total	36	238	**274**	52	222
Nighttime	9	47	**56**	8	48
Daytime	27	191	**218**	44	174
Daytime, rain confirmed (*)	16	187	203	42	161
Daytime, rain not confimed(*)	**11**	**4**	15	2	13

(*) the rain information considered here comes from the police report; however, the meteorological station does not confirm rain information in 11 cases by rain, and in 4 cases by no-rain (daytime).

Relationships have been estimated by lane because averaging the traffic indicators between lanes may hide certain phenomena such as heterogeneous speeds or densities between lanes. Although the lane where an accident begins is generally unknown, the accidents occurring when inserting from or to a ramp are mentioned in the database. These accidents are not included in the estimations of the relationships for the middle and fast lanes. For every accident, traffic data from the first upstream sensor, when available, are 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 relationship between accidents and traffic conditions, to use traffic conditions before the accident. As accident times are only estimated by the police on their arrival on site afterwards, 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 single 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 (eighteen minutes was found to be the time-lag for which the sensitivity analysis conducted in [11] provides the highest correlation between speed and single vehicle accidents).

It is at night that 20% of all accidents occur, but for only 15% of the vehicle-kms traveled. This means an increase in risk at night

Types of accident. Relationships between traffic and accidents have been calibrated by type of accident. Two types of accident have been considered: accidents implying a single vehicle, and accidents implying multiple vehicles.

25 out of the 218 daytime accidents are linked either to a breakdown of a vehicle, or to the driver (drowsiness, alcohol...) or to the presence of a pedestrian; these accidents are probably not linked to the traffic conditions, so they have been discarded from the analysis. The distribution of the 193 remaining accidents according to the type of accident is given in Table 2.

Table 2. Number of accidents with available traffic data by lane and type of accident

	Single Vehicle	Multiple Vehicles	Total
Slow Lane	46	147	193
Slow Lane (*)	46	137	183
Middle Lane (**)	44	128	172
Fast Lane (***)	39	117	156

(*) Excluding accidents on ramps
(**)There are fewer accidents on the middle lane because of missing traffic data
(***)*There are fewer accidents counted on the fast lane because there is no "fast" lane for accidents occurring on a two-lane motorway section*.

3 Statistical Models and Traffic Indicators

Safety Performance Indicators are generally based on negative binomial models, or on distributions like the Poisson-Maxwell-Conway distribution which better fit the dispersion [12]. The risk "*R*" of accident by vehicle-kilometer is related in our approach to different variables within a logistic regression. Six types of relationships have been tested on thirteen variables (traffic indicators).

3.1 Six models

The name of a model is constituted by a part indicating its pattern ("*POW*" for the power function, "*EXP*" for the exponential function,..) and by subscripts indicating, if relevant, that the sets of accidents and traffic conditions have been restricted for the estimation of the model: The subscript "N" (for no-rain) indicates that accidents and traffic conditions during rain have been excluded. The subscript "R" (for "Ramp") indicates, when the model is estimated on the slow lane, that accidents occurring near a ramp have been included. The subscript "*" indicates that accidents implying a PTW have been excluded.

For a given weather condition (rain or no-rain), and a given type of accident, the six-minute periods are grouped by average speed intervals: the speed interval for a group "i" of periods is such that there is at least one accident of the given type occurring during the six-minute periods associated to this group.

1. In the power model, which is generally applied to speed only, risk is proportional to an exponent of the traffic indicator.

$$(POWER):\ R_i = \alpha_V V_i^{\beta+\gamma_{Rain}.Rain_i} + \varepsilon_i.\ \ \text{for } i = 1 \dots n \tag{1}$$

ε_i is the deviation, assumed to be Gaussian-distributed.

R_i is the risk by vehicle-km, V_i is the average speed for the group *i* of periods; $Rain_i$=1 if rain occurs during each period of this group (0 otherwise); the number of accidents (for a given type of accident) and the number of vehicle-kms are associated to this group.

α_V, β and γ_{Rain} are deterministic parameters to be estimated. *n* is the number of groups of periods. This type of model applies also to other relationships, where V_i is replaced by the traffic indicator of another variable, with an analogous process for forming the groups of periods.

2. 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 α as the logarithm of α_V:

$$(POW): \text{Log}[R_i /(1- R_i) = \alpha + \beta.\text{Log}(V_i) + \gamma_{Rain}.Rain_i + \varepsilon_{i.} \tag{2}$$

The number of accidents during rain being low, we have also proposed a simplified model, noted (POW_N), without the rain coefficient.

3. Hauer [9] and Elvik [10] proposed an exponential model; we add to this model a term which models the rain impact:

$$(EXPO): Ri = e^{\alpha+\beta V_i+\gamma_{Rain}.Rain_i} + \varepsilon_i \quad (3)$$

4. The associated logistic model reads:

$$(EXP):\text{Log}[R_i /(1- R_i)]= \alpha + \beta.\ V_i + \gamma_{\text{Rain}}\ .Rain_i + \varepsilon_{i.} \quad (4)$$

5. (V_i) being positive, its square is an increasing function, and replaces (V_i) without changing the sign of β in model (EXP^2):

$$(EXP^2): \text{Log}[R_i /(1- R_i)]= \alpha + \beta.\ V_i^2 + \gamma_{\text{Rain}}\ \text{Rain}_i + \varepsilon_{i.} \quad (5)$$

6. The function $\text{Log}^2(V_i)$ 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 β, unless considering ($-\text{Log}^2(V_i)$) as follows:

$$(LOG^2): \text{Log}\left[\frac{R_i}{1- R_i}\right] = \alpha - \beta.Log^2(V_i) + \gamma_{Rain}\ .Rain_i + \varepsilon_i \quad (6)$$

7. Parabolic models (excluding rainy conditions): When risk is not monotonous with the traffic indicator, modeling requires one more parameter. However, due to an insufficient number of accidents, it was not possible to estimate four parameters; models are therefore estimated here on datasets excluding rain, so the rain coefficient can be removed; "γ" 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 $-\beta /(2.\ \gamma)$.
   The (EXP_{NP}) model comes from the exponential model:

$$(EXP_{NP}):\text{Log}[R_i /(1- R_i)]= \alpha + \beta.\ V_i +\gamma\ V_i^2 + \varepsilon_i \quad (7)$$

   where the subscript "P" (for Parabolic) indicates a parabolic term.

8. The (MIX) model combines a part coming from a power model, and a parabolic part coming from an exponential model:

$$(MIX_{NP}):\text{Log}[R_i /(1- R_i)]= \alpha + \beta.\text{Log}(V_i) +\gamma\ V_i^2 + \varepsilon i. \quad (8)$$

3.2 Statistical analysis

The traffic and accident databases were separated into different cases, according to the time of day (night time/daytime), lane, weather (rain or not). Independent analyses were performed for every combination of cases.

Three accident datasets were considered: the whole dataset, or all accidents except those during rainy conditions, (disregarding the traffic data when raining), or all accidents except those involving Power-Two-Wheelers (PTW).
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 vehicle-kilometers (N_i) by group of periods.

3.3 Thirteen Traffic Indicators

Even with individual data on upstream sensors, it is impossible to identify, among others, the driver responsible for the accident. What we wanted to know was whether the parameters of the whole set of drivers are different just before an accident. The thirteen indicators presented here are computed from the aggregation of traffic data over 6-minute periods:

1. Average speed V_i, by 6-minute period in km/h. We use here the arithmetic speed average (time mean speed).
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 $3.6(L_j+ \lambda)/V^j$, λ being equal to 1 meter, the length of the inductive loop; the unit factor is "3.6".
3. Relative speed in km/h ("RelSpeed") is the difference between the speeds of two consecutive vehicles on the same lane. The sum of relative speeds on a period is of no interest, because the speed of a vehicle generally appears twice in the sum with opposite signs, and thus disappears. Since negative relative speeds have no safety impact, they were disregarded. The indicator proposed here is the sum of relative speeds, when positive, divided by the traffic count.
4. Indicators 4,5,6, and 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 6-minute average time headway ("Average TH"); indicators 5-7 are the percentages of tail-gating (less than 2 seconds, "%TH <2"), short (less than 1 second, "%TH <1") and very short headways (less than 0.5 seconds, "%TH <0.5").
5. Indicators 8,9,10, and 11: Uno defined the "*PICUD*" (Potential Index for Collision with Urgent Deceleration) [13]. It is the estimated difference (in meters) between the stopping locations of vehicles "j" and "j-1", assuming that the leader j-1 brakes at the very instant t^j when the follower "j" passes the sensor. The leader is then located at $V^{j-1}.(t^j - t^{j-1})$ meters downstream the sensor; its rear end is located at $V^{j-1}.(t^j - t^{j-1}) - L^{j-1}$ 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, which would have been avoided, if the follower had had a space gap greater by {-*PICUD*} before the braking of the leader. The follower brakes with the same deceleration (here the deceleration is Γ=6.25 m/s^2) at time $t^j + T$, after a reaction time T=1 second, when the vehicle is located at $V^j.T$ meters downstream the sensor:

$$PICUD^j = \frac{(V^{j-1})^2-(V^j)^2}{2\Gamma} - T.V^j + (t^j - t^{j-1}).V^{j-1} - L^{j-1} \quad (9)$$

The length L^j of the second vehicle replaces the length L^{j-1} 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 $PICUD^j$, over the 6-minute period, divided by the traffic count of the period.

Indicators 9, 10, and 11 are the percentages of drivers for which the *PICUD* is below a threshold (respectively 0 meters ("%*PIC*<0"), minus 10 meters ("%*PIC*<-10"), minus 20 meters("%*PIC*<-20").

6. Indicators 12 and13: *PICUD* becomes "*PICUDbis*", by removing the reaction time in equation (9).

   Indicator 12 ("*PICUDBis*") is the sum of negative $PICUDbis^j$, divided by the traffic count.

   Indicator 13 is the percentage of negative PICUDbis ("%*PICBis*<0").

4 Significant relationships

The risk of single-vehicle and multi-vehicle accidents is related to average speed in section 4.1, to occupancy in section 4.2, to time headway in section 4.3, to relative speed in section 4.4 and to PICUD indicators in section 4.5. The logistic form of the models has been 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 presented in the tables given in the Appendix. The lines of these tables are numbered for easy reference: the lines whose numbers end with the suffix "-1" represent the risk of single-vehicle accidents, while suffix "-2" is related to multiple-vehicle accidents. The relationships with a P-value lower than 5% are generally considered as significant. However, as the number of accidents in our database is not very large, and in the hope of identifying useful relationships, we also considered relationships with a P-value threshold of 10%; these relationships will have to be confirmed on a larger database. In the tables, for each relationship, the first line gives the number of the relationship, the type of model, the values of the two or three parameters, the Null deviance D_0, and the AIC criterion; the second line gives the name of the traffic indicator, the standard deviations "(σ)" of the coefficients, the P value (P%) rounded to zero when less than 0.1; the residual deviance D_R and the degree of freedom #Fre.

4.1 Risk and 6-minute average speed (daytime)

In normal weather conditions, there are relationships between single-vehicle accidents and speed (see the first subsection), between multiple-vehicle accidents and speed (the second subsection); for the whole set of weather conditions, there are relationships between single-vehicle accidents, speed and rain (the third subsection), We did not find any significant relationship between multiple-vehicle accidents, speed and rain, likely because, during rain, there is a concomitant (slight) decrease in speed and an increase in risk.

Models with a rain term, single-vehicle accidents. The risk of single-vehicle accidents (daytime) is significantly related to the 6-minute average speed of the fast lane by power and exponential relationships. Rain is a significant contributing factor. When Power Two Wheeler accidents are excluded (model POW*), the exponent β is very high (β =7.86, σ_β =2.2) with 52% of the deviance explained. The rain coefficient γ_{rain}=1.88 (its standard deviation σ_γ being 0.37); this corresponds to multiply the risk, in case of rain, by $e^{1.88}$ =6.5. When including PTW, the speed exponent decreases to β=2.66 (σ_β =1.3) with 38% of deviance explained, see Table 4 in the Appendix, lines 2-1 and 6-1; the decrease of β means that PTW accidents are not particularly correlated to a high observed average speed. We imagine that when the average speed is low or moderate, some PTW drive between the lanes with a higher speed, but this speed has no impact on the average speed because the magnetic loop of the traffic sensor is not implemented too close to the adjacent lane.

The exponential model (*EXP*) gives a speed exponent β=0.03 (σ_β =0.014) which passes to 0.073 (σ_β =0.020) when PTW accidents are excluded (model *EXP**), see Table 4 in the Appendix, lines 1-1 and 7-1.

On the middle lane, no logistic model appears significant. However, the second lane (from the right) of both two-lane sections and three-lane sections are grouped in this work as the "middle" lane, which brings some heterogeneity.

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.

Normal weather conditions, single-vehicle accidents. In normal weather conditions, i.e. when excluding traffic and accidents during rain, the percentage of explained deviance is only due to the traffic indicator, and is generally smaller than when the rain impact is added.

On the fast lane the model, the power model remains significant only when PTW are excluded (POW_{N*}), with a percentage of explained deviance of 26%, instead of 52% in (POW_*), and a slightly smaller exponent β, (Table 4, line 3_1).

Exponential models are also significant with smaller β, as well as with and without PTW (Models (EXP_N) and (EXP_{N*}) lines 5-1 & 8-1. Since many PTW postpone their trip when it rains, which decreases their risk exposure during rain, the comparison of risks between rainy and non-rainy conditions is fully justified only when excluding PTW.

Multi-vehicle accidents are related to the speed with negative β coefficients (Table 4, lines 9-2 to 14-2). This might mean that the presence of a high number of low-speed vehicles (trucks, entering vehicles) brings some danger. But it might just mean that, when speed decreases, accidents are more likely to result from crashes between vehicles than to be single-vehicle accidents.

Three tentative conclusions. First, the average speed is a better indicator when excluding PTW. Second, exponential models are robust. Third, speed must be estimated on the fast lane.

4.2 Risk and 6-minute occupancy, daytime, multiple-vehicle accidents.

Positive coefficient β are obtained, which means that the higher the occupancy, the higher the rate of multiple-vehicle accidents.

Including accidents related to a ramp, on the slow lane, model (EXP^2), with a rain coefficient γ_{rain}= 0.5 explains a limited percentage (20%) of the deviance of the logit of the risk of multiple-vehicle accidents (Table 5, line 26-2 in the Appendix).

Excluding rainy conditions, model (POWN) explains 36% of the deviance on the middle and slow lanes, and 10% only on the fast lane; on the fast lane, excluding rainy conditions and PTW, model (EXPN*) explains 23% of the deviance (Table 5, lines 25-2, 27-2, 28-2) - see also the comparison of the models, Figure 5.

However, the parabolic model EXPNP (with β>0 and γ<0) indicates that when occupancy is very high (greater than 32%), the risk decreases on the slow lane (relation 29-2, Table 5); it is the same for the middle lane (relation not included in Table 5).

4.3 Average time headway and percentages of short time headway.

Average time headway (TH). Average time headway is proportional to the inverse of the traffic count; thus a high TH (a low traffic flow) occurs either when there is a low traffic demand, or when there is congestion. The first case, which is the most frequent, implies that, for a high TH, when vehicles are far from each other, the higher the TH, the lower the rate of multi-vehicle accidents. Negative β and positive γ_{rain} have effectively been obtained on the middle lane, with a very limited percentage of explained deviance (Table 6, lines 31-2 and 32-2). At the same time, the rate of single-vehicle accidents is higher. A positive β explaining the single-vehicle accidents is thus expected and obtained on the three lanes for all models. Table 6, line 30-1 gives the results of the model (*POW*) for the slow lane; the coefficients are similar for other lanes and the percentage of explained deviance varies from 39% to 50% according to the lane.

Percentage of short time headway. The higher the percentage of short headway, the lower the single-vehicle accident rate. β exponents are negative (Table 6, line 33-1); when including rain, 38% to 68% of the deviance, according to the lane, are explained.

Relationships between risk of *multiple-vehicle* accidents 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 such as (EXP_{NP}) or (MIX_{NP}) (with β <0 and γ>0) indicate, for all lanes and for the three percentages, that the risk of *multiple-vehicle* accidents is high, both for low and high percentages of short time headway, (Table 6, lines 34-2 to 42-2).

4.4 Risk and sum of positive relative speeds (divided by the traffic count)

Abdel-Aty *et al* .found that the speed variance (obtained from a series of ten consecutive 30" average speeds) was a relevant accident precursor [14,15]. This means that changes in 30" traffic conditions are correlated to accident occurrence. This does not mean that heterogeneous speeds (between two consecutive vehicles on the same lane) are correlated to accidents and that more homogeneous speeds (as obtained with an adaptive speed control device) would be safer; although we tried to check this point with our relative speed indicator, which measures this speed heterogeneity, on our limited dataset we found little evidence on this point.

The risk of single-vehicle accidents increases with the sum of positive relative speeds (divided by the traffic count). In all lanes the β coefficients of the power model are positive: 27% to 64% of the deviance is explained, depending on the lane. The numerical values of the parameters are given here only for the fast lane (Table 6, line 43-1). This might come from correlations between relative speed and traffic density, and between traffic density and accident type. Symmetrically, with the model (*POW*), the risk of multiple-vehicle crashes seems to decrease (β negative) when the indicator increases; 33.6% and 45% of the deviance of the *multiple-vehicle* accident risk are explained on the middle and the slow lanes (Table 6, lines 44-2 and 45-2. However, with parabolic models MIX_{NP} or EXP_{NP}, the parabolic coefficients γ are positive, implying an increase in risk on the middle and slow lanes when the relative speed indicator is sufficiently high (above 1.5 km/h) (Table 6, lines 46-2 and 47-2).

4.5 PICUD-based indicators

Single-vehicle accident occurrence is related, with positive γ_{Rain} and β coefficients, to *PICUD*-based indicators:

- the absolute value of the average of negative *PICUDbis* on the fast lane (line 48-1),
- the percentage of *PICUD* less than minus 20 meters (fast lane, without Power–Two- Wheelers, line 49-1),
- the percentages of negative *PICUDbi*s on the fast and middle lane (lines 50-1 and 51-1).

Multi-vehicle accident occurrence is related to rain with a positive coefficient and, with a negative β, to the percentages of PICUD less than minus 10 or 20 meters (Table 6, lines 52-2 and 53-2); however these relationships are obtained only when including accidents related to the ramp entrance/exit (which was tested on the slow lane only); no significant relationship appears between multi-vehicle accident occurrence and the averages of negative PICUD or PICUDbis.

When rain is excluded, the rate of multi-vehicle accidents is related to the percentages of *PICUD* less than minus 10 meters or minus 20 meters. However, these relationships are obtained:

- on the fast lane when excluding PTW (model LOG^2_N*),with a small percentage of deviance explained (line 54-2); the risk increases with the percentage (β>0),

- on the slow and the middle lanes, when adding a parabolic term (EXP_{NP}, lines 55-2 and 56-2); the risk increases with high percentages ($\gamma > 0$).

The percentage of negative *PICUD*, when excluding rainy conditions, is related to the risk of multiple-vehicle accidents by the parabolic model EXP_{NP}, with a positive coefficient γ, which shows that risk increases with high percentages of negative *PICUD*. However, the percentage of explained deviance (26%) is limited (Table 6 line 57-2).

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 an aggregated 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 certain, that a high average speed results in a higher risk.
2. A good correlation does not mean causality. An accident due to a high density also appears with a low speed, because high density is correlated to a low speed, but this is misleading.
3. Misinterpretation. "A low speed" can be inappropriate if speed is not sufficiently low. During rain, the speed decreases, but too little, hence the risk increases. If the presence of rain is not identified, a misleading negative correlation between risk and speed would appear. All contributing factors have to be identified, and then introduced.
4. Accounting for the many contributing factors is difficult, with only a limited number of accidents. Accidents occur due to multiple factors: presence of rain, of a curve, of an access ramp, etc. Using the speed average may be inadequate, due to the sensitivity of the extremes ends of the distribution (the risky drivers). If all but the risky drivers respect the speed limit, this will imply a negative correlation between *average* speed and risk.
5. Classifying accidents as single or multiple-vehicle accidents: even if speed had no effect on risk, speed would have a positive correlation with single-vehicle accident risk and a negative one with multiple-vehicle accident risk: this is due to the negative correlation between speed and density, combined with the correlation between density and type of accident. When traffic density is low, few vehicles are close to another vehicle, and so more accidents are in fact "single-vehicle" accidents.
6. Non monotonous relationships. Multiple-vehicle accidents 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. Progress on that topic should be possible by refining the analysis on the basis of more accident data, thus allowing a more disaggregated level.

A lack of correlation may come from the absence of any relationship, from confounding various opposite effects, or from insufficient data; a lack of linear correlation may also come from a non linear relationship.

6 Deepening the relations between the relationships

In order better to understand the relevance of the models, the occurrence of accidents and the quality of data and data processing, three points need to be highlighted:

- Both the power and the exponential functions model the relationships between accident risk and speed or occupancy. How is this possible, given that the functions are different?
- What are the relationships between the total accident risk (sum of risks for single- and multiple-vehicle accidents) and speed?
- Combining the relationship between speed and risk with the relationship between risk and occupancy gives a speed-occupancy relationship. How credible is this new relationship, which comes under the category of speed-density relationships? The more incredible the new relationship, the more incredible the two risk models which have been combined.

6.1 Models relating the accident risk and speed

In the exponential model, the risk "R" is linked to the speed by a negative exponential function; this implies that the logarithm of the risk is a linear function of the speed.

Here, a linear relationship between the logit of the risk, defined as Log[R/(1-R)], and the speed has been calibrated within a logistic regression. As (1-R) is very close to 1, the logit of the risk is numerically very close to the logarithm of the risk. Also, in Figures 2, 3, 4 and 5, the Y-axis, instead of being entitled "logit of the risk" has been entitled "Logarithm of the risk" for the sake of readability. In these figures, the exponential model appears as a straight line, whereas the power model appears as a logarithm function.

Daytime accident risk and single-vehicle accidents. Figure 2 gives the modeled risks for single-vehicle accidents related to speed (daytime, fast lane, average 6-minute speeds between 60 and 130 km/h). The power model (represented by a diamond without any solid line, (parameters from Table 4, relation 2-1) is impressively close to the exponential model (the solid line) (parameters from Table 4, relation 1-1).

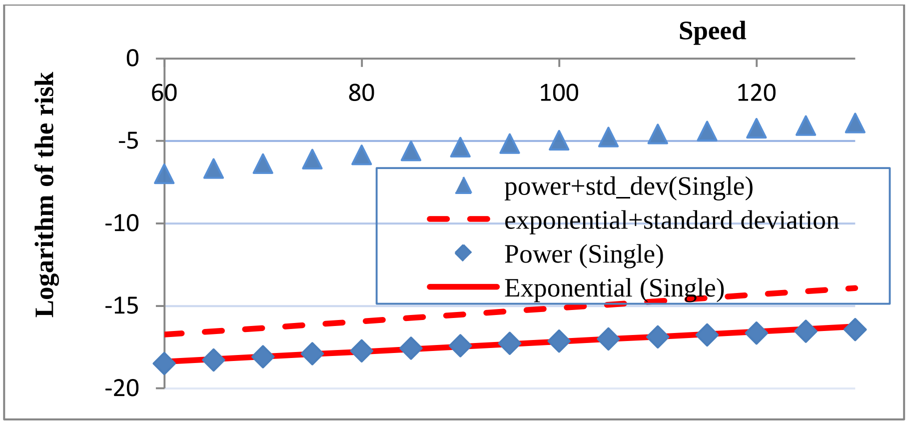

Fig. 2. Daytime risk of single-vehicle accident related to speed: Power & exponential models (fast lane)

In the case of rain - not displayed here - both models remain very close together: 1.6886 or 1.69669 need to be added to the logit of the risk for the power model and for the exponential model, respectively.

The curves "exponential +Standard Deviation (Single)" and "power+Std_Dev (Single)" have been built by adding one standard deviation to parameters α and β of the models; the first of these two curves is closer to the curve displaying the risk than the curve "power+ Std_Dev (Single)". This shows that the exponential model has a lower standard deviation than the power model and indicates that the exponential model should be chosen.

Accident risk and multiple-vehicle accidents. As already stated in section 4.1, the relationship between the average speed and risk is closer when excluding accidents implying PTW, perhaps because the speed of some PTW, which weave in and out between two lanes, is broadly independent of the recorded average speed; in addition, as the number of accidents in our database is limited, excluding rain decreases the number of parameters of the models and in turn makes the relationship easier to identify.

In the cases where there is no rain and no Power-Two-Wheelers are involved, Figure 3 gives the logit of the modeled risk for single-vehicle accidents and for multiple-vehicle accidents, assuming either a power model or an exponential model.

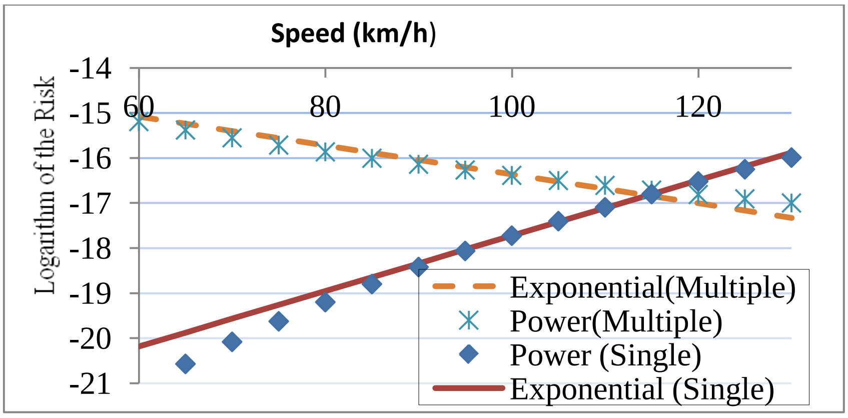

Fig. 3. Daytime risk related to the speed on the fast lane for single and multiple-vehicle accidents, power and exponential models; rain and PTW excluded.

These curves are based on relations 3-1, 8-1,13-2,14-2 of Table 4, Appendix.

The exponential model (dashed straight line) and the power model (stars) also coincide when considering the multiple-vehicle accidents. Both risks decrease with speed. Although we have no real proof, this decrease is probably not really due to speed, but to the combination of two factors:

1. Correlations between high speed and low density,
2. With low density traffic there are few vehicles on the road; accidents generally concern a single vehicle thus decreasing the occurrence of multiple-vehicle crashes.

Total accident risk and speed. The total daytime risk of accidents (single- or multiple-vehicle) is less sensitive to speed, because of a balance between the increase and decrease described previously.

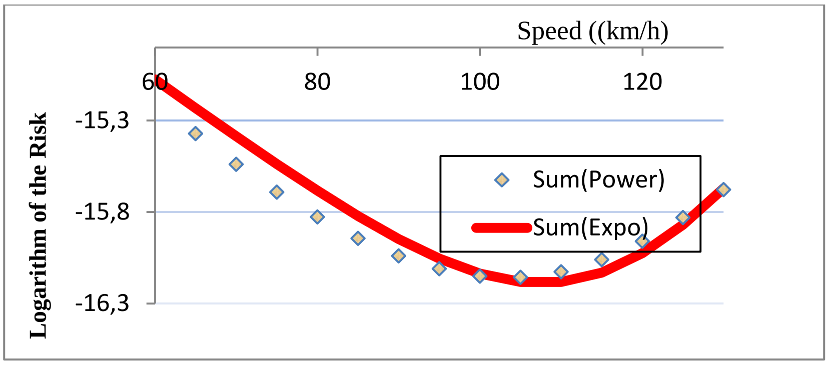

Fig. 4. Total daytime risk (single- and multiple-vehicle accident) related to the speed on the fast lane, power and exponential models; rain and PTW excluded.

The total risk at low speeds is high, probably due to congestion. The total risk decreases while speed increases up to 110 km/h and then increases for higher speeds. By definition, at least two vehicles are involved in a multiple-vehicle accident, whereas just one vehicle is involved in a single-vehicle accident[1]. It would be interesting to consider the risk by vehicle, which would be obtained by weighting each accident by the number of vehicles involved in the accident. Multiple-vehicle accidents would be weighted by a weight of at least two and would have a greater impact on the total risk.

6.2 Models relating the accident risk and occupancy

The power and exponential models have also been used to model the relationship between accident risk and occupancy.

In the case of no rain, both the power and exponential models relate significantly both risks (single- and multiple-vehicle accidents) and occupancy on every lane. Their numerical parameters are given in Table 5; these models are displayed for the fast lane on Figure 5, from the relations given Table 5 lines 17-2, 18-2, 21-1, 23-1) .

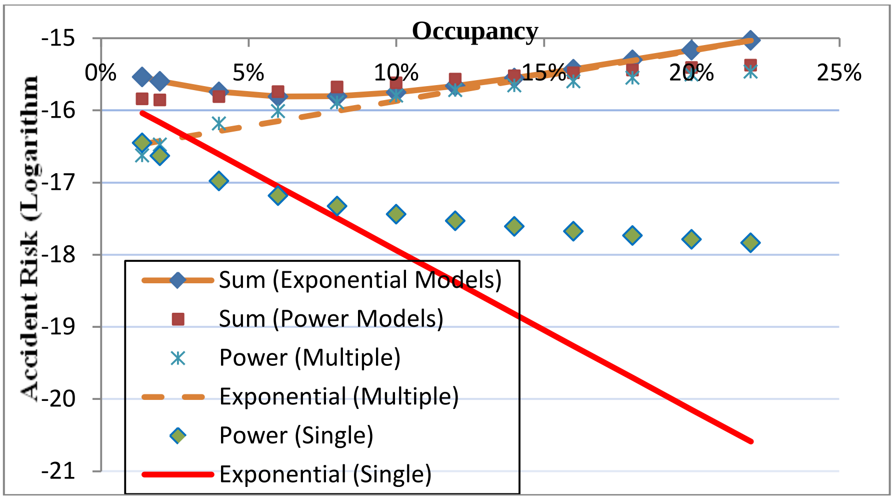

Fig. 5. Risks of single- and multiple-vehicle accidents, and total risk related to the occupancy of the fast lane, power and exponential models, daytime, no rain.

The risk of single-vehicle accidents *decreases* when occupancy increases, since with a high occupancy rate there are many vehicles on the road, and accidents generally concern multiple vehicles.

The power and the exponential models are close together in the case of multiple-vehicle accidents; however, they differ in the single-vehicle case. Remember that the number of single-vehicle accidents used to calibrate the models is rather low (23). For

[1] However, a number of drivers involved in a single-vehicle accident claim that the accident is due to a runaway vehicle.

single-vehicle accidents, the exponential model must be preferred to the power model because of the lower values of two criteria (Residual Deviance and AIC-Akaike Information).

The total risk is minimal (exponential model) when occupancy is 8%; for this occupancy, there are too few vehicles on the road to cause multiple-vehicle accidents, and average speed is generally not high enough to cause single-vehicle accidents. When occupancy is higher, there is a strong increase in the total risk.

6.3 Consistency analysis: the case of speed and occupancy risks.

There is a relationship between density ρ and occupancy *O*, assuming that all vehicles have the same constant speed during a period. Let L be the average length of the vehicles and ***l*** be the length of the traffic sensor (in meters); ρ =1000 .*O* /(L+ ***l***) or:

$$O = (L+ \boldsymbol{l}).\ \rho\ /1000 \tag{10}$$

In France, the values of *L* and ***l*** are 4.2 meters and 1 meter.

Let us now consider a model giving the risk related to speed, and a model giving the risk related to occupancy. If these models are consistent, their combination should be credible.

As the risk model of single-vehicle accidents related to speed is monotonous (increasing), it is possible to invert it and to produce a monotonous (increasing) "speed related to risk" model; combining this new model with the (decreasing) "risk related to occupancy" model results in a new "speed related to occupancy" model which is decreasing. Then, using the occupancy-density relationship, a speed-density relationship is obtained:

Another speed-occupancy or speed-density relationship is obtained when inverting the monotonous (decreasing), model giving the risk of multiple-vehicle accidents in function of speed, with the increasing model giving the risk of multiple-vehicle accidents in function of occupancy.

Our very aim is not to identify a speed-density relationship, but to analyze the relationship produced. The risk models will not be considered as credible if the speed-density relationship. produced by their combination is not credible.

Let us consider both exponential models for risk related to speed and risk related to occupancy; then the logarithm of the risk is linear related to speed and linear related to occupancy; it leads to a linear relation between speed and occupancy (or density) which is the relation proposed by Greenshields [17].

Considering both power models for risk related to speed and risk related to occupancy leads to a power speed-density relationship:

$$V = \mu.\ \rho^{\delta} \text{ with } \delta < 0 \tag{11}$$

Considering the power model (Eq.2) for risk related to speed and the exponential model (Eq.4) for risk related to occupancy leads to the Underwood speed-density relationship:

$$Log\ R/(1-R = \alpha_V + \beta_V.Log(V) = \alpha_O + \beta_O.O = \alpha_O + \beta O(L+ \boldsymbol{l}).\ \rho\ /1000 \tag{12}$$

where α_V and β_V are the parameters of the power model linking risk to speed; α_O and β_O are the parameters of the exponential model linking risk to occupancy.

Rewriting Eq.(12): Log(V) = (α_o - α_V)/β_V+ ρ. (β_o/β_V) (L+ $\boldsymbol{l}$). /1000 or:

$$V = e^{(\alpha_0 - \alpha_V)/\,\beta_V} \,.\, e^{\beta_0/\,\beta_V (L + \boldsymbol{l})\rho/1000} \qquad (13)$$

This is the Underwood [16] fundamental diagram:

$$V = V_f \,.\, e^{(-\rho/\rho_{cr})/} \text{ with: } V_f = e^{(\alpha_0 - \alpha_V)/\,\beta_V} \text{ and } 1/\rho_{cr} = -\,(\beta_o/\beta_V.\ (L + \mathbf{l})\ /1000) \qquad (14)$$

where V and ρ are the variables (average speed and density); V_f, is the free-flow speed and ρ_{cr} is the critical density at which the traffic flow is maximum.

When a credible speed-density relationship is, as inquired, obtained, this does not prove that the risk models are good, because the risk has disappeared from Eq.(13). Let us imagine that 50% of accidents are missing; the risk would be divided by two, but this would be unnoticeable in Eq.(13).

Fundamental diagrams which link speed to density are always using the space-mean speed (i.e. the harmonic mean). It is then necessary, for passing from risks to fundamental diagrams, to express risks in function of the space-mean speed rather than the time-mean speed (i.e. the arithmetic speed). The coefficients α, β, γ_{Rain} of the power model (Eq.(2)) relating risk and speed are replaced by the coefficients α_{Space}, β_{Space}, γ'_{Rain} of model (15) :

$$\text{Log}[R_i /(1- R_i) = \alpha_{Space} + \beta_{Space}.\text{Log}(V_{s,i,}) + \gamma'_{Rain}.Rain_{\,i} + \varepsilon_i \qquad (15)$$

where $V_{s,i}$ is the harmonic mean speed at period i; α_{Space} , β_{Space} and γ'_{Rain} are ob tained either by a direct calibration of model (15), or from the α, β and γ_{Rain} parameters of model (2) by minimizing the sum of the square of the differences between the risk related to the time-mean and the space-mean speeds:

$$\sum_i \left(\alpha_{Space} + \beta_{Space}\text{Log}\left(V_{s,i}\right) + \gamma'_{Rain}Rain_i - \alpha - \beta\text{Log}(V_i) - \gamma_{Rain}Rain_{,i}\right)^2 \quad (16)$$

This second approach has been done here. For the case of {fast lane, no rain, day-time}, three power models relating risk to space-mean speed have been calibrated:

- risk for single vehicle accidents: (PTW excluded): Relation 3-1 is modified in Relation 3-1$_{Space}$
- risk for multiple vehicle accidents: Relation 10-1 is modified in Relation 10-2$_{Space}$
- risk for multiple vehicle accidents (PTW excluded); Relation 14-2 is modified in Relation 14-2$_{Space}$

The values of the parameters of these relationships are in the Appendix, Table 4. The inverse of these three power models have been combined with exponential models relating risk to occupancy (relations 24-1, 17-2, 25-2); three Underwood speed-density relationships have been obtained; there parameters are in Table 3:

Table 3. Values of the parameters of the Underwood speed-density relationships obtained by power and exponential models relating risks of single- or multiple-vehicle accidents to space-mean speed and occupancy, daytime, no-rain, fast lane.

	V_f	ρ_{cr}	*Capacity*
Single-vehicle accident without PTW	132,6	53.1	2589
Multiple-vehicle accident	120.3	55.7	2465
Multiple-vehicle accident without PTW	119.3	58.1	2550

The free-flow speed is credible; the critical density and the capacity, although rather high, are not incredible, taking into account, first that it is a fast lane and second that a higher capacity appears when considering data aggregated on a 6-minute period rather than over one hour.

The three corresponding speed-density curves are given in Figure 6:

.

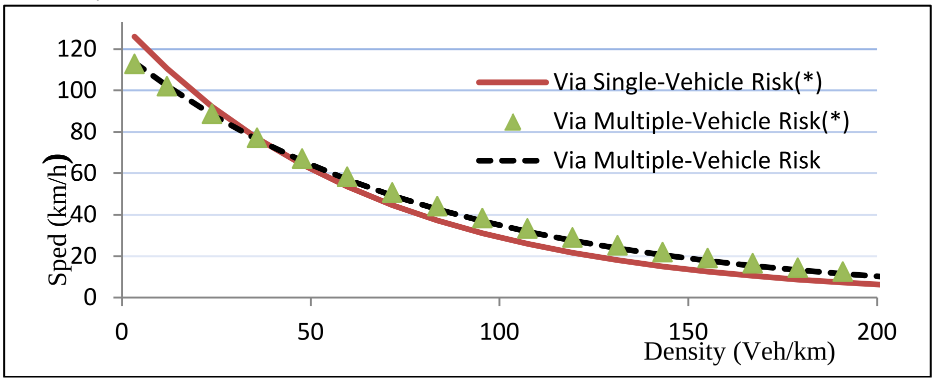

Fig. 6. Three Underwood Speed-Density relationships, no-rain, with and without (*) PTW, fast lane, daytime

The star(*) in the legend indicates than PTW have been excluded. The three speed-density relationships appear very close together.

7 Conclusion and perspectives

Using significant relationships, it has been shown that some variables are significantly linked to accidents:

- for single-vehicle accidents, the 6-minute average speed on the fast lane; and the average time headway (on every lane),
- for multiple-vehicle accidents, the percentages of short time headway (for every lane); excluding rain; the power and the exponential models relate significantly accidents and occupancy on every lane - but the correlation is low on the fast lane; moreover, for the slow and middle lanes, risk increases with occupancy until a threshold, then decreases.

It is more difficult to understand other relationships linking accidents to relative speed or to the "PICUD", a variable based on the collision computation. These relationships 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.

Moreover, the behavior of a single driver who is responsible for an accident does not systematically appear at an aggregated 6-minute level.

Such relationships, if correctly validated and integrated in traffic management tools, should be useful to anticipate the safety impact of a new traffic management scheme.

We also investigated the relationships *between* the risk relationships, and we think that this approach provides a better understanding of data, accidents and models.

The perspectives of this work are three-fold: improving the data processing, validating and assessing the predictive power of such relationships as accident precursors, and 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 relationships; the time of the accident is not perfectly known; the distance between two successive traffic stations is too high to capture very local problems; data are static, and do not describe the beginning or the ending of a bottleneck; Ppwer-Two-Wheels are not counted.

(2) Data processing. Some analyses must be disaggregated, according to the type of section and infrastructure. Various aggregation times should be tested.

(3) Traffic indicators. Other relevant indicators should be added (Time to Collision, Post-Encroachment Time, etc.). 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 that also take into account some interactions between the variables.

The perspective also is to go on to validation and to transferability. Validating the relationships on other periods is essential, 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 sensitive, the final step would consist in integrating such relationships as safety criteria in traffic management algorithms.

Acknowledgements. We are very grateful to the "COMET" project (https://sites.google.com/site/orsicomet/home), launched by IFSTTAR, which brings together different participants in order to share knowledge and to develop tools for traffic management during adverse meteorological conditions. We also thank the reviewers for their relevant and fruitful remarks.

8 Appendix. Significant relationships

Table 4. Parameters of significant relationships relating risk to average speed and rain (day-time)

Num-Ber	Model Lane	Constant (σ_α) (P%)	β (σ_β) (P%)	γ (σ_γ) (P%)	D_0 (D_R)	AIC #Fre
1-1	EXP	-20.205	0.030	1.697	59.9	131
Speed	Fast	(1.44)(0%)	(0.01)(2.5%)	(0.33) (0%)	(36.2)	#41
2-1	POW	-29.382	2.662	1.689	59.9	132
Speed	Fast	(6.13)(0%)	(1.31)(4.3%)	(0.33) (0%)	(37.0)	#41
3-1	POW$_N$*	-48.158	6.609		26.9	74
3-1$_{Space}$		-43.282	5.562			
Speed	Fast	(11.7)(0%)	(2.51)(0.8%)		(19.7)	#23
4-1	Pow$_N$	-27.399	2.235		28.9	98
Speed	Fast	(6.78)(0%)	(1.45)**(12%)**		(26.0)	#31
5-1	EXP$_N$	-19.584	0.025		28.9	97
Speed	Fast	(1.58)(0%)	(0.01)(10%)		(25.8)	#31
6-1	POW$_*$	-54.0	7.86	1.88	56.6	102
Speed	Fast	(10.3)(0%)	(2.21)(0%)	(0.37) (0%)	(27.2)	#33
7-1	EXP$_*$	-25.1	0.07	1.88	56.6	103
Speed	Fast	(2.2)(0%)	(0.02)(0%)	(0.37) (0%)	(27.3)	#33
8-1	EXP$_{N*}$	-23.9	0.06		26.9	75
Speed	Fast	(2.5)(0%)	(0.02)(0.8%)		(19.9)	#24
9-2	EXP$_N$	-12.828	-0.033		109.4	224
Speed	Fast	(0.35)(0%)	(0.003)(%)		(51.1)	#72
10-2	Pow$_N$	-5.149	-2.392		109.4	229
10-2$_{Space}$		-6.913	-2.015			
Speed	Fast	(1.26)(0%)	(0.28)(0%)		(56.0)	#72
11-2	Log^2	-10.173	-0.283		109.4	228
Speed	Fast	(0.66)(0%)	(0.03)(0%)		(54.6)	#72
12-2	EXP^2_N	-14.01	-2.109e-04		109.4	222
Speed	Fast	(0.22)(0%)	(2.5e-05)(0%)		(49.0)	#72
13-2	EXP$_{N*}$	-13.1543	-0.0321		89.4	197
Speed	Fast	(0.40)(0%)	(4.4e-3)(0%)		(46.9)	#65
14-2	Pow$_{N*}$	-5.620	-2.337		89.4	200
14-2$_{Space}$		-7.344	-1.969			
Speed	Fast	(1.43)(0%)	(0.3)(0%)		(49.5)	#65

Table 5. Parameters of significant relationships relating risk, occupancy and rain (daytime)

Num-ber	Model Lane	α (σ_α) (P%)	β (σ_α) (P%)	γ (σ_γ) (P%)	D_0 (D_{Res})	AIC (#Fr)
15-1 Occup.	POW Fast	-18.751 (0.31)(0%)	-0.545 (0.07)(0%)	1.796 (0.33) (0%)	94.8 (41.1)	126 #34
16-2 Occup.	POW_{N*} Fast	-15.238 (0.42)(0%)	0.352 (0.15)(2.4%)		75.0 (69.3)	224 #68
17-2 Occup.	EXP_N Fast	-16.565 (0.16)(0%)	6.969 (1.32)(0%)		97.2 (74.9)	256 #78
18-2 Occup.	POW_N Fast	-14.823 (0.38)(0%)	0.422 (0.14)(0.3%)		97.2 (87.2)	269 #78
19-1 Occup.	EXP Fast	-15.622 (0.26)(0%)	-24.474 (4.75)(0%)	1.993 (0.33) (0%)	94.8 (32.9)	117 #34
20-1 Occup.	POW_* Fast	-18.676 (0.39)(0%)	-0.466 (0.09)(0%)	2.033 (0.36) (0%)	64.2 (27.7)	99 #29
21-1 Occup.	POW_N Fast	-18.595 (0.36)(0%)	-0.503 (0.09)(0%)		47.4 (27.3)	87 #23
22-1 Occup.	EXP_* Fast	-15.762 (0.31)(0%)	-28.054 (6.1)(0%)	2.115 (0.37) (0%)	77.1 (59.9)	231 #73
23-1 Occup.	EXP_N Fast	-15.727 (0.30)(0%)	-22.108 (5.60)(0%)		47.4 (22.5)	82 #23
24-1 Occup	EXP_{N*} Fast	-16.1008 (0.37)(0%)	-20.1865 (7.11)(0.4%)		21.581 (13.2)	64 #20
25-2 Occup.	EXP_{N*} Fast	-16.7577 (0.17)(0%)	6.5275 (1.48)(0.4%)		75.0 59.4	213 #68
26-2 Occup.	EXP^2_R Slow	-16.117 (0.1)(0%)	14.423 (2.26)(0%)	0.482 (0.28)(9%)	144 (116)	351 #98
27-2 Occup.	POW_N Slow	-13.069 (0.4)(0%)	1.182 (0.16)(0%)		128 (81.6)	182 #82
28-2 Occup.	POW_N Middle	-12.787 (0.4)(0%)	1.321 (0.18)(0%)		142 (91)	178 #79
29-2 Occup.	EXP_{NP} Slow	-17.473 (0.3)(0%)	20.029 (4.17)(0%)	-31.340 (10.5)(0%)	127 (75)	283 #84

Table 6. Significant relationships. relating risk, various traffic indicators and rain (daytime)

Number Variable	Model Lane	α (σ_α)(P%)	β (σ_α) (P%)	γ (σ_γ) (P%)	D_0 (D_{Resi})	AIC #Fr
30-1	POW	-18.6	1.15	1.25	86	154
Average TH	Slow	(0.3)(0%)	(0.22)(0%)	(0.35)(0%)	(51)	#46
31-2	EXP*	-15.8	-0.12	0.52	63	246
Average TH	Middle	(0.2)(0%)	(0.07)(6.5%)	(0.3)(11%	(56)	#80
32-2	MIX_{NP}	-15.2	-1.08	0.014	61	237
Average TH	Middle	(0.3)(0%)	(0.41)(0.8%)	(0.007)6%	(54)	#75
33-1	EXP	-15.4	-3.58	1.67	69.3	133
%TH<2s	Middle	(0.4)(0%)	(0.85)(0%)	(0.34)(0%)	(30.5)	#46
34-2	EXP_{NP}	-14.0	-8.84	8.56	70.8	243
%TH<2s	Fast	(0.4)(0%)	(1.67)(0%)	(1.7)(0%)	(48.9)	#85
35-2	EXP_{NP}	-14.8	-7.44	8.60	70.7	253
%TH<2s	Middle	(0.5)(0%)	(2.0)(0%)	(2.2)(0%)	(57.5)	#82
36-2	EXP_{NP}	-15.4	-4.38	6.59	74.2	287
%TH<2s	Slow	(0.4)(0%)	(2.24)(5%)	(2.6)(1%)	(64.4)	#95
37-2	EXP_{NP}	-14.36	-14.47	26.21	94.8	235
%TH<1s	Fast	(0.2)(0%)	(2.5)(0%)	(5.5)(0.0)	(62.1)	#72
38-2	MIX_{NP}	-17.2	-0.47	6.62	68.4	242
%TH<1s	Middle	(0.4)(0%)	(0.14)(0.1%)	(4.1)(11%)	(59.7)	78
39-2	MIX_{NPR}	-16.9	-0.34	6.87	59.3	247
%TH<1s	Slow	(0.4)(0%)	(0.13)(0.7%)	(4.1)(9(%)	(52.8)	78
40-2	MIX_{NPR}*	-18.4	-0.47	168.5	56.9	145
%TH<0.5s	Fast	(0.5)(0%)	(0.11)(0%)	(51.4)0.1%	(40.8)	42
41-2	EXP_{NP}	-15.7	-63.1	1565	36.1	131
%TH<0.5s	Slow	(0.2)(0%)	(37.6)(9.3%)	(969)(10%	(33.3)	#37
42-2	EXP^2_N	-16.0	-308.5		54.6	183
%TH<0.5s	Middle	(0.1)(0%)	(136)(2.4%)		(47.7)	#56
43-1	POW	-17.0	1.95	1.59	57.2	111
RelSpeed	Fast	(0.16)0%	(0.45)(0%)	(0.33)0%)	(20.6)	#36
44-2	POW_R*	-16.2	-1.41	0.52	122	277
RelSpeed	Slow	(0.1)(0%)	(0.2)(0%)	(0.3)(9.8%	(67)	#85
45-2	POW_N	-16.1	-1.58		106	228
RelSpeed	Slow	(0.1)(0%)	(0.2)(0%)		(50)	#71
46-2	MIX_{NP}	-16.8	-2.00	0.42	100	225
RelSpeed	Middle	(0.3)(0%)	(0.4)(0%)	(0.2)(10%)	(60)	#66
47-2	EXP_{NP}	-13.0	-4.60	1.42	106	228
RelSpeed	Slow	(0.5)(0%)	(1.2)(0%)	(0.66)(3%)	(48)	#71
48-1	EXP	-18.2	1.78	1.6	28	63

PicudBis	Fast	(0.3)(0%)	(0.9)(3.9%)	(0.5)(0.2%	(18)	#18
49-1 %PIC<-20	EXP^2* Fast	-17.7 (0.3)(0%)	60.66 (36.8)(9.9%)	1.9 (0.4)(0%)	40 (20)	94 #33
50-1 %PiCbis<0	EXP^2 Middle	-17.3 (0.2)(0%)	246.5 (140)(7.8%)	1.6 (0.3)(0%)	52 (33)	113 #35
51-1 %PiCbis<0	EXP^2 Fast	-17.4 (0.2)(0%)	211.2 (110)(5.5%)	1.8 (0.3)(0%)	57 (33)	92 #24
52-2 %PIC <-10	POW_R Slow	-17.3 (0.3)(0%)	-0.48 (0.08)(0%)	0.51 (0.28)(7%)	97 (62)	288 #95
53-2 %PIC <-20	POW_R Slow	-18.4 (0.5)(0%)	-0.55 (0.1)(0%)	0.54 (0.28)(5%)	78 (43)	183 #54
54-2 %PIC <-20	LOG^2_N* Fast	-17.0 (0.2)(0%)	0.05 (.01)(0%)		63 (35)	142 #44
55-2 %PIC <-20	EXP_{NP} Middle	-15.2 (0.2)(0%)	-41.30 (8.0)(0%)	238.09 (64)(0%)	67 (31)	177 #59
56-2 %PIC <-20	EXP_{NP} Slow	-15.2 (0.2)(0%)	-73.26 (17.4)(0%)	725.46 (250)0%	61 (28)	148 #46
57-2 %PIC < 0	EXP_{NP} Fast	-14.2 (0.4)(0%)	-11.50 (2.6)(0%)	16.4 (4.4)(0%)	71 (63)	245 #84

References

1. Venkataraman, N. S., Ulfarsson, G. F., Shankar, V., Oh, J., Park, M.: Model of Relationship between Interstate Crash Occurrence and Geometrics: Exploratory Insights from a Random Parameters Negative Binomial Approach. Transportation Research Record, No 2236, 41-48, Washington D.C. (2011)
2. Brodsky H, Hakkert A.S.: Risk of a Road Accident in Rainy Weather. Accident Analysis and Prevention, 20(3), 161-176 (1988)
3. Bergel-Hayat,R., Debarrh, M., Antoniou, C., Yannis, G.: Explaining the Road Accident Risk: Weather Effects. Accident Analysis & Prevention, 60, 456-465 (2013)
4. Golob, T. F., Recker, W.W., Alvarez, V.M.: Freeway Safety as a Function of Traffic Flow. Accident Analysis& Prevention, 36(6), 933–946 (2004)
5. Abdel-Aty, M., Uddin N., Pande, A.: Split Models for Predicting Multivehicle Crashes during High-Speed and Low-Speed Operating Conditions on Freeways. In Transportation Research Record , No 1908, Washington, D.C., 51–58 (2005)
6. Park, P. Y-J, Saccomanno, F.: Reducing Treatment Selection Bias for Estimating Treatment Effects Using Propensity Score Method. Journal of Transportation Engineering. American Society of Civil Engineering, Manuscript No. TE/2005/023489, Feb. (2007)
7. Nilsson, G.: Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety. Bulletin 221. Lund Institute of Technology, Department of Technology and Society, Traffic Engineering, Lund (2004) http://lup.lub.lu.se/record/21612
8. Elvik, R.: A re-parameterisation 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 (2013)

9. Hauer, E.: Speed and Safety. Transportation Research Record, No 2103, pp. 10–17 Washington.,D.C.,(2009)
10. Elvik, R.: Speed and Road Safety – New Models. TØI Report 1296/2014 Oslo, 39 pages (2014) https://www.toi.no/getfile.php/Publikasjoner/T%C3%98I%20rapporter/2014/1296-2014/1296-2014-elektronisk.pdf
11. Aron, M., Seidowsky, R., Ditchi, N.: Traffic Indicators and Accidents: The Case of a Motorway Network in the South of France. Transportation Research Board, 92th Annual meeting, paper#13-4638 Washington D.C. (2013)
12. Lord, D., Guikema, S.D.: The Conway–Maxwell–Poisson Model for Analyzing Crash Data. Journal Applied Stochastic Models in Business and Industry 28 (2) 122-127 (2012)
13. Uno, N., Iida, Y., Itsubo, S., Yasuhara, S. : A Microscopic Analysis of Traffic Conflict Causes by Lane-Changing Vehicle at weaving section. In: Proc. of the 13th Mini-Euro Conf. " Handling Uncertainty in Transportation Analysis of Traffic", pp. 143-148 (2003)
14. Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F. M., & Hsia, L.: Predicting Freeway Crashes from Loop Detector Data by Matched Case-Control Logistic Regression. Transportation Research Record, No 1897(1), 88-95, Washington D.C. (2004)
15. Abdel-Aty, M. A., Pemmanaboina, R.: Calibrating a Real-Time Traffic Crash-Prediction Model Using Archived Weather and ITS Traffic Data. In: IEEE Transactions on Intelligent Transportation Systems, 7(2), pp. 167-174 (2006)
16. Underwood, R. T.: Speed, Volume, And Density Relationship: Quality and Theory of Traffic Flow Yale Bureau of Highway Traffic, pp. 141–188, New Haven (1961)
17. Greenshields, B D, Bibbins, J R, Channing, W S and Miller, H H. A Study of Traffic Capacity. In: Highway Research Board Proceedings 14, pp. 448–477 (1935)