# Risk Index Model Building: Application for Field Ramp Metering Safety Evaluation on Urban Motorway Traffic in Paris

*Habib Haj Salem, Nadir Farhi, Jean Patrick Lebacque*

*Institut Français des Sciences et Technologies des Transport de l'aménagement et des réseaux (IFSTTAR) COSYS/GRETTIA, Cité Descartes, 14-20 Boulevard Newton, 77447 Marne–La-Vallée; France*
*Emails: habib.haj-salem@ifstttar.fr   nadir.farhi@ifsttar.fr   jean-patrick.lebacque@ifsttar.fr*

***Abstract:*** *This paper aims at developing a risk index based on real-data measurements, which can be used either off-line as an evaluation index during the evaluation process which leads to the dramatical reduction of the field test periods, or in real-time like: a safety monitoring tool (e.g. safety user warning system), or a multi-criterion function to be optimized in real time (safety index combined with a traffic index) within several control strategies, such as coordinated ramp metering, speed limit control, route guidance, etc.*

***Keywords:*** *Risk modeling, safety, clustering ramp metering, field evaluation.*

## 1. Introduction

Control measures introduced to ameliorate traffic performance in motorway traffic include speed limit control, ramp metering, user information aiming at homogenizing the practical speed along the motorway sections and at minimizing the number and the severity of accidents and consequently increasing safety [7]. On the other hand, the introduction of electronics and computerization systems in vehicle technologies has significantly contributed to safety and comfort. However, the prediction of a crash in real time is still in an investigation phase and some research efforts are dedicated in this area. During the last decade, there is an

increasing focus on the development of real time ("potential crash") prediction algorithm on urban motor way traffic [2, 3, 5, 6].

In the field of safety analysis, the classical traffic evaluation approaches consist in collecting incident/accidents traffic data during the experimented scenarios (traffic control strategies, modification of the infrastructure, etc.), and in proceeding to traffic impact and statistical safety analysis of the number of accidents before and after the implementation of these scenarios. Generally, the collection of the accident numbers must get statistical significance before undertaking an evaluation process. This remark imposes a long time of field data collection (5-10 years), which is the "price to pay" for having a correct safety evaluation.

This paper aims at developing a risk index based on real-data measurements, which can be used either off-line, as an evaluation index during the evaluation process which leads to the dramatical reduction of the field test periods, or in real-time, like: a safety monitoring tool (e.g. safety user warning system), or a multi-criterion function to be optimized in real time (safety index combined with a traffic index) within several control strategies, such as coordinated ramp metering, speed limit control, route guidance, etc.

The developed risk index is based on the collection of measured traffic data synchronized with incidents/accidents data on the ring way of Paris.

The paper is organized as follows: Section 2 is dedicated to the description of the collected data base, Sections 3 and 4 are focused on the development of the used methodologies. Section 5 includes the description of the best scenarios for the final risk model building. Section 6 is dedicated to the application of the risk index model for the evaluation of the safety impact on the implemented ramp metering strategies.

## 2. Data base characteristics

The traffic dataset and accident characteristics are collected from historical database stored in the Ville de Paris operating system. The considered sites are fully equipped with real traffic measuring detectors located at around every 500 m apart. The incidents/accidents data characteristics include: time of day, location of the accident, involved vehicle categories, weather conditions and severity (number of lanes blocked).
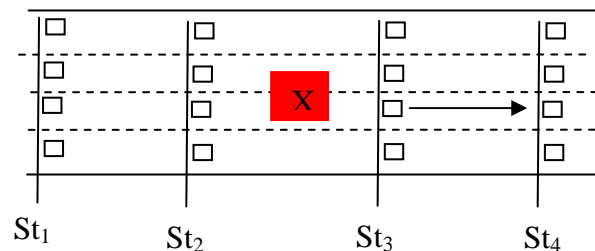


Fig. 1. Topology of the considered stretch measurements for each crash

The collected traffic data covers two hours (one before and one after the crash) at two upstream and two downstream measurement stations (Fig. 1). The time intervals of the traffic measurements are equal to one minute.

The final constituted database includes the overall accidents that occurred and traffic data during 4 years (2003-2005). The total number of the accidents collected is around 900 on the ring way of Paris. After traffic data cleaning, 300 sets of accidents are retained. During the selection of the accidents, the following criteria are considered: same weather condition (sunny), same topology (number of lanes). Among the 300 sets of accidents data, the remaining sets were equal to only 90 accidents sets which are used for statistical analysis.

## 3. Methodology

The applied methodologies are mainly based on statistical analysis of the collected traffic measurements around the accident (see Fig. 1). A series of multivariate statistical methods are used, with the aim to find the relationship between the occurrence time of the accident and the traffic conditions. Two well-known statistical methods are applied: cluster analysis and the most common form of factors analysis. In particular, the principal components analysis is applied to find the non-correlated variables to be used for building the risk model. In our case, the total number of variables characterizing the dataset is equal to 4(stations)×2(volume, occupancy rate)×4 (number of lanes) = 32 variables.

For the clustering analysis, several possibilities are investigated:
- Clustering by upstream occupancy rates/lane
- Clustering by downstream occupancy rates/lane
- Clustering by all occupancy rates/lane

The same clustering method is applied for the measurement stations including four lanes. Lastly, based on the clustering output results, linear regression and nonlinear logistic modelling approaches are applied for computing the risk index.

The hierarchical ascending clustering via SAS is performed, using a Ward's criterion [2], in order to exhibit the particular class of traffic conditions which prevail at the time just before the accident.

## 4. Clustering by occupancy rate/lane results

In this case, the application of SAS clustering method leads to finding five main representative clusters. The first cluster is characterized by a homogeneous average occupancy (MOcc) on the 16 measurement stations. The Occupancy rates (Occ) are comprised between 9, and 15%, and characterize a low occupancy value and consequently light traffic conditions. This cluster contains 2191 observations and is representing 42.96% of all measurements.

Cluster 2 gathers the observations with higher and inhomogeneous average occupancy. Indeed, the MOcc are lower on the fast lane; their values vary around the critical occupancy (from 18 up to 23%). The two central lanes have higher

occupancy rates and correspond to unstable traffic states. All lanes of the last station ($St_4$) are congested. The occupancy rates range from 24 up to 27.5%. This cluster represents 27.37% of the samples.

Regarding cluster 3 (representing 8.8% of the data), a clear transition is observed between the MOcc of the upstream stations, which are very high (from 36 up to 52%), and the low MOcc of the downstream stations (from 6 up to 11 %).

The MOcc of cluster 4 are homogeneous on the 16 measurement points, with high values, ranging from 30 up to 40 %. This cluster represents 19.39 % of the population. These states correspond to a high level of congestion.

Lastly, cluster 5 is characterized by an average upstream MOcc (16 up to 21 %), particularly on the first two lanes, and very congested downstream (52 up to 68 %). Moreover, we observe that station 4 is more fluid than station 3. This cluster is less representative (1.15 % of the data).

Screening the time evolution of the clusters (one hour before the crash) of all records (85 in total), 41 accidents indicate a change of the cluster during the last six minutes, i.e., in 48% of the cases. If only the last observation before the accident is observed, among the total number of accidents, 39 (46%) are moved to cluster 3. Cluster 3 represents upstream congestion and downstream fluid conditions. The risk modelling is based on the traffic state of this cluster.


## 5. Logistic regression

The constituted accident database is split into two parts. The first half is dedicated to the calibration of the linear regression using SAS tool. The second half is used for validation of the found risk model.

During the calibration process, the results given by the clustering are used. The logistic regression is performed by considering that cluster 3 presents the highest level of a crash risk. In this case, the risk model value is set to 1, otherwise – to 0. Hence, the risk model parameters and variables given by SAS are the following:

$$
\begin{aligned}
\text{Risk} = 1/[1 + \exp{} & -(-5.7335 + 0.01107\,\text{st1\_to}(1) - 0.0827\,\text{st1\_to}(3) + \\
& + 0.02601\,\text{st2\_to}(1) + 0.1102\,\text{t2\_to}(3) + 0.1886\,\text{st2\_to}(4) - \\
& - 0.5798\,\text{st3\_to}(2) - 0.3851\,\text{st3\_to}(4) - 0.4483\,\text{st4\_to}(2) - \\
& - 0.5809\,\text{st4\_to}(4) - 0.00407\,\text{st2\_q}(2) + 0.00663\,\text{st3\_q}(2) + \\
& + 0.00449\,\text{st4\_q}(1) )].
\end{aligned}
$$

(1)

As indicated, the obtained risk model includes 13 parameters and 12 variables. According to these numbers of parameters and variables, the use of this model seems to be very complicated. On the other hand, the results obtained during the validation process are not satisfactory. In fact, applying the Risk model on the second half of the database generates large oscillation of the Risk value between 0 and 1. In order to reduce the number of parameters and variables, the same approach is applied on the aggregate variables by the measurement stations. In this case the number of variables is limited to 8 instead of 32.

## 6. Model based on the clustering by station measurements

In order to minimise the model parameters and its interpretation, we aggregate our variables by averaging the occupancy rates on the lanes of the same station and by summing the flows at each station. Our variables then reduce to one flow and one occupancy rate at each station.

The SAS clustering procedure output gives five clusters. Cluster 1 is the densest (more than 36%). It is characterized by quite homogeneous Occ and an average flow over the 4 stations, (an Occ of 11 up to 12% and a flow of 1450 up to 1500 vehicles per hour and per lane) characterizing the fluid traffic conditions. Cluster 2 presents a very high Occ on the (upstream) stations 1 and 2 and rather average downstream (from 14 up to 18%). As for the flow, it is rather stable and low compared to other clusters. This cluster contains 20% of the time steps. Cluster 3 presents high occupancy rates over all stations. The flow is a higher upstream. Cluster 4 has an average Occ close to the usual 20% critical value, increasing from upstream to downstream (26.7% at station 4). The flows are higher than the other clusters, up to 1774 veh/h/lane at station 2. Cluster 5 has a high average Occ (around 37%) at all stations and a lower flow (around 1230 veh/h per lane).

When we consider the accidents and attribute to each time step of the cluster number to which it belongs, we observe that 43 accidents out of 85 studied (50.58 %) present a cluster change during the last six minutes. For 60 accidents (i.e., more than 70 % of them), the last time step belongs to cluster 2, characterized by a rare faction shock wave (congested upstream and fluid downstream).

The same approach as the one previously described is applied. However, the Risk model is set to 1 for the observations belonging to cluster 2 and 0 elsewhere. The calibration of the Risk model is based on 80 % of the full observations. The logistic regression model output by SAS is given as:

$$\text{(2)} \quad \begin{aligned} \text{Risk} = 1/\{1 + \exp[-(-7.1677 + 0.2122 \text{ Oc\_st1} + 0.1383 \text{ Oc\_st2} - \\ - 0.1061 \text{ Oc\_st3} - 0.2052 \text{ Oc\_st4} + 0.000385 \text{ q\_st1})]\}. \end{aligned}$$

The dataset remaining (20% of the observations) is used for the model validation. The same model obtained is applied on 1000 observations which are not used for the calibration. The output results of this model are depicted in Figs 2 and 3.
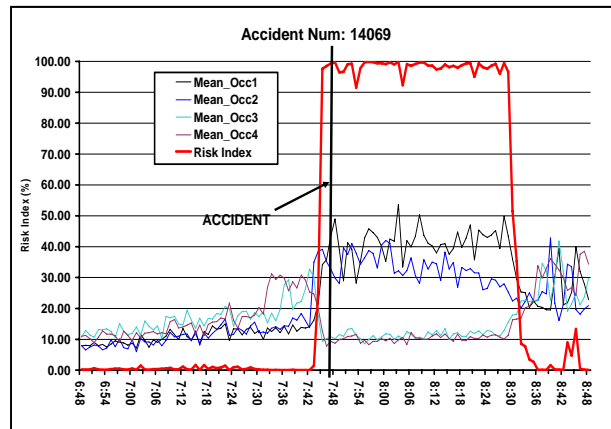
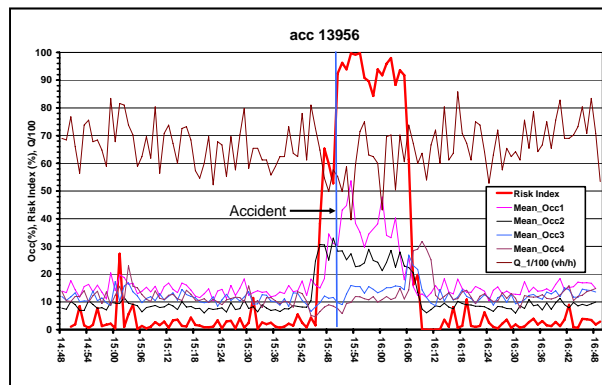Fig. 2. Evolution of the risk index for accident Num: 14069



Fig. 3. Evolution of the risk index for accident Num: 13956

Screening the time evolution of the clusters (one hour before the crash), the results obtained demonstrate that among all found clusters a critical cluster leads (48% in total) to the occurrence of the accident. Consequently, the risk modelling is based on the traffic state of this cluster.

## 7. Risk Model application for ramp metering safety evaluation

In frame of the European project "EURAMP", the field trials have been conducted aiming at traffic impact evaluation of several ramp metering strategies. Four control strategies have been tested:

1. No control : reference case
2. ALINEA: traffic responsive strategy
3. Variable cycle ALINEA
4. Coordinated strategy.

The test site is located in the south of the Ile de France Motorway network (Fig. 4). The total length of the experimental area is approximately 20 km (only the direction towards Paris is considered). This part of the motorway includes five

consecutive on-ramps, which are fully equipped with loop detectors and traffic signals. The carriage way is equipped with detector stations (each 500 m) for traffic volume, occupancy and speed measurements.
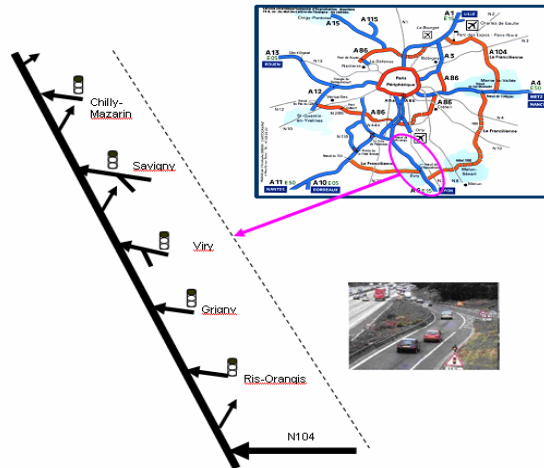


Fig. 4. Field trial test site

The overall period of these field trials is limited to around 16 months. During the evaluation process, the risk index was applied for safety evaluation. However, before using the safety index model, it is necessary to proceed to the risk model validation. The used data corresponds to the collected accidents and the collected measurement traffic data.

The accident data was collected from September 2006 to mid-January 2007 (the holiday periods are excluded). Table 1 shows the number of accidents per strategy that occurred between 5:00 and 12:00 period.

As expected, the number of accidents is not statistically significant and therefore it will not be possible to draw any conclusion for the safety assessment from these accidents data.

Table 1. Number of collected accidents per strategy

| NC | AL | VC | COORD |
|----|----|----|-------|
| 7  | 5  | 6  | 2     |

In order to extend the number of accidents, all accidents occurring during both years 2006 and 2007 are considered. The total number of collected accidents is equal to60. After data cleaning and accidents selection criteria (winter and night are excluded), only 20accidents are selected for the analysis and in particular the risk model validation. The same risk model (equation (2) with the same parameters) found on the ring way of Paris is applied on A6W motorway. Figs 5 and 6 depict the risk index time evolution of two selected accidents.
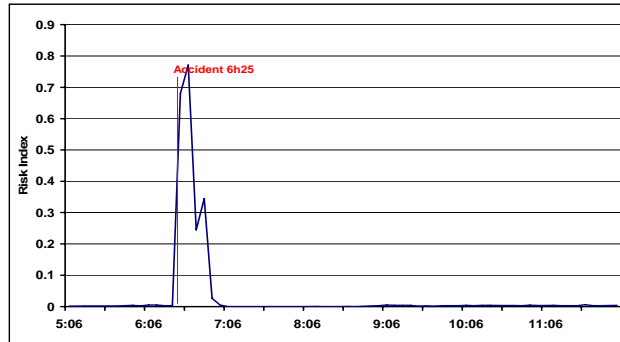
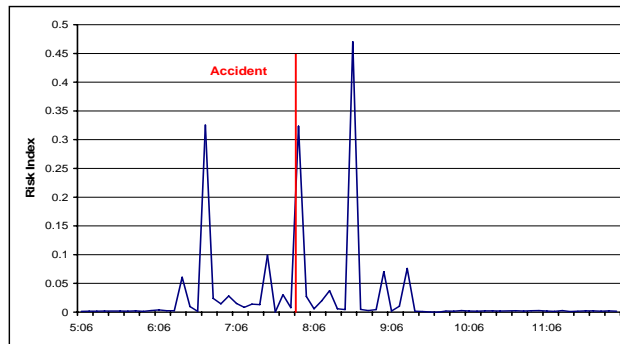Fig. 5. A6W Risk index time evolution: Accident 1


Fig. 6. Risk index time evolution: Accident 2

We can underline that, using the same parameters of the risk model by the station found on the ring way of Paris and applied to the A6W motorway, the obtained results of the time evolution of the risk index is very promising. Without any calibration, the risk index value is maximal before the accident occurs (see Figs. 5 and 6). Consequently, we can assume that the computation of the risk index can be considered as a safety index to be compared between the candidate ramp metering strategies.

The cumulative risk index by strategy is computed on the overall motorway sections (19 sections) and on the overall time period (6-12 h). The obtained results indicated that the implementation of the ramp metering strategies improves the safety aspect by 20%. In particular, the safety benefit is more important in case of the coordination. The obtained results are very similar to other safety evaluation impacts of the ramp metering. Extensive results can be found in [2].

Table 2. CRI: Cumulative Risk index assessment

| Strategies | CRI | Gain |
|------------|------|--------|
| NO Contr | 1.33 | --- |
| ALINEA | 1.05 | − 20.8% |
| VC | 1.09 | − 18.8% |
| Coord | 1.03 | − 23.2% |

40

## 8. Conclusion and next steps

The results obtained are very promising. The number of parameters was limited to 5, which can minimise the effort for the calibration process. The interesting point to underline is the limited effort of the risk model calibration. As a matter of fact, the same parameters found on the ring way of Paris are valid for the A6W motorway. However, more investigations are needed in order to take into account other parameter conditions, such as the weather, luminosities (night) and the modification of the geometric topology (different lanes number of the upstream and downstream measurement stations).

Therefore, investigations are on the way in order to combine safety and traffic criteria for the development of integrated control strategies including simultaneously e.g. ramp metering, route guidance, speed limits. The objective function to be optimized corresponds to a multi-objective function.

R e f e r e n c e s

1. E v e r i t t, B. Cluster Analysis Heinemann Educ. London, Books, 1974.
2. EURAMP Project. Deliverable D3.5. Safety Critical Issue for Ramp Metering. European Commission, Brussels 2006.
3. G l o b, F. T., W. W. R e c k e r. A Method for Relating Type of Crash to Traffic Characteristics on Urban Freeway. – Transportation Research. Part A. Vol. **38**, 55-80.
4. A b d e l - A t y, M., A. P a n d e. Identifying Crash Propensity Using Specific Traffic Speed Conditions. – Journal of Safety Resaerch, Vol. **36**, 2005, 97-108.
5. L e e, C., B. H e l l i n g a, F. S a c c o m a n n o. Evaluation of Variable Speed Limits to Improve Traffic Safety. – Transportation Research, Part C, Vol. **14**, 2006, 213-228.
6. S a l e m, H. H., C. R a m a n a n j o a n a, J. P. L e b a c q u e. Risk Index Model for Real-Time Prediction of Potential Traffic Crash on Urban Motorway. – In: Proccedings of EWGT'2006. Advances in Traffic and Transportation Systems Analysis, BARI, Italy, 2006.
7. D i l m o r e, J. Implementation Strategies for Real-Time Traffic Safety Improvement on Urban Freeways. MA Thesis, University of Central Florida, Orlando, 2005.