

# SAFETY INDICATORS FOR MICROSIMULATION-BASED ASSESSMENTS

**Jaime Barceló<sup>1\*</sup>**

**André-Gilles Dumont<sup>2</sup>**

**Lidia Montero<sup>1</sup>**

**Josep Perarnau<sup>1</sup>**

**Alexandre Torday<sup>2</sup>**

1. Dept. of Statistics and Operations Research  
Technical University of Catalonia  
Pau Gargallo, 5  
08028 Barcelona  
Spain  
Phone +34.93 401 7033  
Fax +34.93 401 5881  
Email(s): [jaume.barcelo@upc.es](mailto:jaume.barcelo@upc.es)  
[Lidia.montero@upc.es](mailto:Lidia.montero@upc.es)  
[Josep.perarnau@upc.es](mailto:Josep.perarnau@upc.es)

2. Swiss Federal Institute of Technology  
Laboratory of Traffic Facilities  
LAVOC-EPFL  
CH-1015 Lausanne  
Switzerland  
Phone +41.21 693 24 19  
Fax +41.21 693 63 49  
Email: [andre-gilles.dumont@epfl.ch](mailto:andre-gilles.dumont@epfl.ch)  
[alexandre.torday@epfl.ch](mailto:alexandre.torday@epfl.ch)

\*All correspondence send to first author.

Submitted for presentation and publication  
Transportation Research Board 2003 Annual Meeting  
January 2003  
Washington, D.C.

# WORDS: 6886

November 2002

**ABSTRACT:**

In the field of ITS applications evaluation, micro-simulation is becoming more and more a useful and powerful tool. In the evaluation process, one of the most important steps is the safety analysis. For that purpose, classical micro-simulation outputs give some helpful information, but which aren't sufficient for an accurate analysis in many cases. Nevertheless, the microscopic level of traffic description offers the possibility of tracking the simulated vehicles getting at each time step their relative position, speed and deceleration. This paper explains how a safety indicator can be calculated with these different parameters. This safety indicator is used in a ramp metering case study to illustrate the utility of such output for a safety analysis. However, this indicator is limited to the linear collision probability and gives therefore no information on crossing trajectories conflicts like in junctions. On the other hand the likelihood of an incident to happen depends not only on traffic conditions but on the influence of many other factors as for example the geometry of the road, the visibility or the pavement conditions (wet, dry, etc.). When significant statistical information is available an estimation of the probability of an incident to happen can be computed, and used in micro-simulation analysis. The paper is completed with the development and testing of hierarchical logit based model to estimate this probability.

## 1. INTRODUCTION

There is a general agreement on the importance of safety analysis in the implementation and operation of traffic systems. This analysis can be conducted from a preventive as well as from a reactive point of view. A preventive safety analysis will have the objective of identifying the incident prone traffic conditions, determine the relevant factors and evaluate their importance. This type of preventive analysis can provide indices and indicators which can help the traffic operator to intervene, implementing management strategies which could hopefully help to reduce the incident likelihood. A reactive analysis concerns the intervention when an incident has occurred, an early intervention being critical to alleviate or, ideally avoid the traffic disturbance caused by the incident.

The typical methodology to conduct these types of safety analysis is usually based on a long term carefully designed incident data collection process to gather sufficient statistical data on which conduct the analysis. The object of the analysis is very often to derive indicators helping to understand or to evaluate the phenomena. These indicators usually depend on threshold values that have to be previously calibrated. An example of that, in the case of the reactive analysis would be the thresholds that activate most of the Automatic Incident Detection Algorithms. Thresholds for operational algorithms have been typically calibrated by trial and error, empirical experimentation on historical data, and performance curves obtained from multiple runs of the respective algorithm on the data with incrementally changing thresholds. On the other hand, to remain operational, these algorithms should be frequently recalibrated.

Improvements on these estimation processes could be expected from the possibility of reproducing the traffic conditions before and after the incident. Also it would be desirable to explore the sensitivity of the estimated thresholds by changing some of the incident related parameters, i.e. location, severity in terms of the length of the blocked lane by the incident, etc. Microscopic traffic simulation can provide the technological framework to achieve these objectives. Microscopic traffic simulation has proven to be a useful tool to capture the full dynamics of time dependent traffic phenomena, but also being capable of dealing with behavioral models accounting for drivers' reactions, it is not only a simulation tool that emulates realistically the flow of vehicles on a road network but it is also capable of reproducing accurately the situation generated by an incident in terms of length of the blocked lane and duration of the incident, allowing in this way the replication of the traffic congestion due to the incident. All these characteristics make microscopic simulation a tool to assist the analyst in the preventive as well as in the reactive safety analysis.

A typical use of microscopic traffic simulation is for the evaluation of traffic systems, namely those involving ITS applications. Given the relevance of safety in traffic systems it becomes obvious that safety analysis should be an important issue in an ITS application evaluation process. The implementation of an ITS application cannot be justified only by the increase in performance of a network if it implies a decrease in user safety. So, the knowledge of the safety level is crucial for taking good decisions. But safety analysis has also an indirect impact on the evaluation process. Indeed, if the safety level of a network decreases, the number of accident rises. The presence of new accidents will create congestions and thus a decrease in network performance. The safety and performance evaluations are directly linked, the first having an important influence on the second. In microscopic simulation based evaluation, this phenomenon becomes critical. However, most of the currently existing microscopic traffic simulators are based on the family of car-following, lane changing and

gap acceptance models to model the vehicle's behavior [1], and that makes of microscopic traffic simulation an ideal world where no incidents can occur, as far as the basic modeling hypothesis in the underlying car-following models is that vehicles should keep a "safety to stop distance" [2], [3]. Hence, a decrease in safety doesn't imply a decrease in network performance as it must do. This particular aspect of the microscopic simulation increases the need of a safety analysis tool providing useful micro-simulation safety indicators to be used in the evaluation processes. In the literature, few articles have dealt with this particular field of the micro-simulation. Among them are the research works of Archer [4], Minderhoud & Bovy [5] and Kosonen & Ree [6]. The first one has demonstrated the potential of micro-simulation for safety assessments. The second one has presented a first safety indicator but which was only based on the time to collision (TTC) parameter. The last one has introduced the SINDI project and proposed an indicator combining the TTC and the speed for a non-constant reaction time simulator (HUTSIM).

The situation described so far has provided the motivation for the research whose results are presented in this paper. The research has been carried out with the microscopic traffic simulator AIMSUN, [7], [8], [9], embedded in the GETRAM software environment for traffic modelling and analysis, [10]. The paper is structured as follows, section 2 presents the development of a different approach to get a micro-simulation safety indicator, and its testing on a site in the peripheral motorway of Lausanne in Switzerland. Section 3 addresses the development of an incident probability estimator, to be used in preventive safety analysis to help the traffic management in trying to prevent incident, it is part of the results of PRIME, a EU project of the 5<sup>th</sup> Framework Programme, for the test site of Barcelona, Spain. Section 4 summarizes the results achieved.

## **2. SAFETY INDICATOR FOR MICRO-SIMULATION BASED ASSESSMENTS**

### **2.1 The Approach**

As explained in the introduction, micro-simulation models prevent all type of collision between vehicles. In the particular case of linear conflicts, which is one of the topics in this paper, the car-following model is in charge of avoiding collision situations. Micro-simulation software has its own car-following model, an improved version evolved from the seminal Gipps model, [11], in the case of AIMSUN, but all models are generally based on an important behavioural parameter: the driver's reaction time. Depending on the software, the reaction time can be a global parameter for all the vehicles (including their driver) or differentiated for each class of vehicles and it can be a deterministic value or a stochastic one (following a distribution rule). But the reaction time of a particular vehicle remains constant during all the simulation. In every case, the car-following model controls the acceleration and deceleration and consequently the headway of the follower vehicle depending on its reaction time. Obviously, the less a vehicle's reaction time is, the less its minimum acceptable headway is.

If this approach offers an excellent approximation of the traffic flows and the relative position of the vehicles, it doesn't permit to extract potential collision situation from the simulation process as they exist in reality. The main reason is that, in the simulation, the headway between two vehicles is in accordance to the reaction time of the follower, but in real world this accordance is not always guaranteed. The big difference between the model and the real behaviour of drivers is that in reality the reaction time is always changing and is not constant during a travel. The reaction time and, consequently, the concentration of the

driver are permanently influenced by his state of tiredness, a phone call, a dialog with other passengers, a look in another direction, etc.

As the behaviour model approximates with a satisfactory accuracy the vehicle's movements, the reaction time, which is obtained after the calibration process, represents then the average of the reaction times of the vehicles in reality. Better said, it represents the average of the reaction times the drivers believe they have! A lot of standards fix limits or standard values for reaction time in the field of road transport. Usually, a standard reaction time represents a maximum limit that only few drivers exceed and only during some limited moments of their journey. This standard reaction time is generally used in road geometry studies and planning. For example, the Swiss standards have adopted 2 seconds as standard reaction time which is divided in a physiological reaction period and a mechanical one, [12].

The definition of the standard reaction time implies that the potential of collision becomes significant if the headway between two vehicles is below this value. In fact, this statement is only valid if both vehicles are driving with the same speed and have the same deceleration capacity, which is rarely the case. More precisely, the approach to determine the crash potential between two vehicles (a follower and a leader) is to respond to the following question:

*If the follower vehicle's reaction time is equal to the standard time reaction (2 seconds in the Swiss case) and the leader vehicle breaks with its maximum deceleration capacity, will a crash occur?*

Obviously, the importance of this "hypothetical" crash will be proportional to the difference of speed between both vehicles at collision time (first impact) but also to the speed of the follower vehicle at collision time (potential impacts with other cars or lateral obstacles after the first impact).

## 2.2 The "unsafe" density parameter

The application of this approach during a micro-simulation process is:

During each simulation step, the position, the speed and the maximum braking capacity of a particular vehicle is known and can be obtained. The same parameters can be obtained for its leader vehicle. The two following hypotheses are taking into account:

- Follower vehicle's reaction time = standard reaction time (2 seconds for the Swiss case)
- Leader vehicle breaks with its maximum braking capacity

With this information and by applying the basic dynamic rules, it's possible to determine if the "hypothetical" crash will occur or not. If it occurs, the speed of the follower vehicle  $S$  and the difference of speed  $\Delta S$  at collision time can be calculated.

As explained in the previous section, the importance of this "hypothetical" crash is proportional to  $S$  and  $\Delta S$ . An "unsafe" parameter could then be defined as the multiplication of both parameters. But this value represents the maximum importance possible. But the real deceleration of the leader vehicle can be obtained (if it is decelerating). So, this maximum value must be multiplied by the ratio  $R_d$  between the deceleration of the leader vehicle and its maximum deceleration capacity. The "unsafe" parameter can then be defined as:

$$\text{unsafety} = \Delta S \cdot S \cdot R_d$$

This parameter determines the level of "unsafe" in the relation between two consecutive vehicles on the road for a determined simulation step. If the "hypothetical" crash doesn't occur or the leader vehicle isn't breaking, the value of the "unsafe" parameter is zero.

But this parameter doesn't give a global situation of the safety in a network or part of it. For that purpose, an "unsafe" density parameter must be calculated. It will be done for each link of the microsimulation model network and for each aggregation period as follows:

$$\text{unsafety density} = \frac{\sum_{s=1}^{S_t} \sum_{v=1}^{V_t} \text{unsafety}_{v,s} \cdot d}{T \cdot L}$$

Where:

$V_t$  = nb of vehicles in the link

$S_t$  = nb of simulation steps within aggregation period

$d$  = simulation step duration [s]

$T$  = aggregation period duration [s]

$L$  = section length [m]

The "unsafe" density (UD) parameter allows to compare the safety level between different links of the network, and to observe its evolution from one time period to another. But the most significant is that it permits comparison between different simulation scenarios and can therefore be the principal indicator to use in a safety assessment process.

### 2.3 Case study

The UD parameter was used within the framework of an evaluation study of a ramp metering implementation on one of the Lausanne (Switzerland) by-pass junctions. In this section of the by-pass, frequent congestion problems are reported during peak hours. The traffic flow on this section increases dramatically in some few kilometers by an important traffic input coming from Morges-West (first entrance) and Morges-East (second entrance) junctions. Generally, congestions appear in the second entrance area.

A micro-simulation based performance evaluation using AIMSUN, shows that the implementation of a ramp metering on the second entrance permits to limit the length of the congestion queue and, consequently, the duration of the congestion. At the safety assessment level, clear conclusions are more difficult to get. Indeed, typical micro-simulation outputs give two contradictory results:

- Presence of the ramp metering decreases the duration of congestion  $\Rightarrow$  decrease in accidents potential  $\Rightarrow$  increase in user safety



- The ramp metering strategy implies important variations of speed in the junction area and on the on-ramp  $\Rightarrow$  increase in accidents potential  $\Rightarrow$  decrease in user safety

Without safety indicators, it's difficult to get a global balance between these two conclusions, but the application of the UD parameter allows it. But, before being applied to this ramp metering study, a validation process has been conducted. It demonstrated a good correlation between the UD evolution in the studied area and the accident reports provided by the police.

Figure 1 shows the one hour average UD for each section of the AIMSUN micro-simulation model. The case with ramp metering (RM) is compared to the one without. In both cases, the UD before the first entrance is close to zero because the normal flow allows regular headways and a fluid traffic without important breaking manoeuvres. Between both entrances, the UD is much more important in the case without RM, because the back of the queue moves back close to the first entrance and because the congestion duration is more important. In the case with RM, the safety problem due to congestion appears only after section ID 35. The phenomenon of speed variations in the second entrance area in the case with RM is confirmed by a UD level more important. Figures 2 and 3 illustrate more accurately both phenomena. However, from a global point of view, the network overall “unsafe” density calculation permits to say that the case with RM is a safer scenario.

This indicator could be also used for the evaluation of applications like Intelligent speed adaptation, variable speed limitation signs and so on.

### **3. A STATISTICAL APPROACH TO DYNAMIC INCIDENT PROBABILITY ESTIMATION**

A contribution to the preventive safety analysis is a model to estimate incident probabilities. The aim of the dynamic incident probability estimation model (EIP) is to establish the association between traffic conditions, weather conditions, road geometry and incident occurrence. The association is established according to statistical models that take dynamic traffic and weather conditions and static road geometry as explanatory variables and the presence of incidents as a response variable (possibly for each incident type), giving dynamically as a result the probability of incident occurrence for each segment of the road network. A previous research on this topic was done by Hamerslag, [13], who proposed a procedure based on a statistical approach following a Poisson regression model. The procedure was conceived and implemented as a static tool for incident occurrence analysis, but real-time traffic management systems require procedures that are able to work in real-time and cannot be derived from Hamerslag's approach. The idea of a logit statistical model as a way of improving Hamerslag's approach was the objective of the research in PRIME. That the topic is of interest is proven by other recent research as the reported in [14], namely the TRAVELAID project sponsored by FHWA, [15].

The model proposed by the authors is not considered previously in the literature, since the response variable in the Hamerlag's model is the number of expected incidents during a long-period (year) in a section given its average level of daily volume and length; for a shorter period of time, as for example 5 minutes, as incidents are rare events, it makes no sense to formulate an statistical model with response variable the number of expected incidents for 5 minutes !!! The response variable must be reformulated to be the presence or absence of incidents and thus the expected presence of incidents is the incident probability (Bernoulli response variable). This is the proposal of the authors and it is properly captured by a binary response logit model in particular or in general, a hierarchical binary response logit model to

predict the expected probability by each incident type (given a previous classification leading to the structure of the hierarchical model).

The statistical approach developed in this paper for the dynamic estimation of incident probabilities is called EIP-HLOGIT. The association between the probability of occurrence of an incident and the explanatory variables is established by means of a generalized linear regression model that for each incident type  $j$  (under a hierarchical underlying structure) establishes a logit link relationship for each section  $i$  (EIP-HLOGIT). The estimated probability for section  $i$  and incident type  $j$ ,  $p_{ij}$ , can be described as

$$p_{ij} = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \quad (1)$$

where  $\eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}_j$  is a linear predictor defined as a linear combination of model parameters ( $\boldsymbol{\beta}$ 's) and current values of explanatory variables defined for section  $i$  and incident type  $j$  ( $\mathbf{x}_{ij}^T$ 's). Equation (1) is equivalent to

$$\mathbf{logit} p_{ij} = \log \frac{p_{ij}}{1 - p_{ij}} = \eta_{ij} \quad (2)$$

which shows a clearer relation to generalized linear regression models for binary response for each hierarchical level and logit link function, see McCullagh and Nelder, [16].

The EIP-HLOGIT model is built after:

- Model variable definition: identification of variables that play a role as predictors of the incident occurrence. This set of variables is clearly site-dependent and must be defined/identified in the **model selection** stage. Variables can be continuous variables (covariates) or factors (discrete variables). Factor variables can be included in generalized linear regression models by means of dummy variables related to each category of factors. Interactions between factors and covariates are technically possible and have been considered in this case. Predictor variables are considered at section level and vary for each interval period (1 minute in this case): dynamic section predictor variables are used to predict incident probability in the current section and time period.
- Model parameter estimation: each variable selected in a EIP-HLOGIT model, either covariate or dummy variable related to a level factor, has an associated real number that plays the role of the coefficient in the linear combination defining the contribution to the prediction. The values of these parameters, once the model variables are selected, are estimated by the **calibration of model parameters**.
- Threshold tuning for risk level definition: thresholds are numerical values related to probabilities, and are used to classify a computed incident probability as a low, medium or high-risk probability situation. The stage of threshold setting is called **threshold tuning** or **calibration of high-risk threshold**.

The model selection and calibration stages require a significant amount of recorded historical data related to:

- Traffic data for incident and non-incident periods. In the latter case only a proportion of the total time is registered, otherwise the huge amount of data would be impossible to manipulate. The result is an increase in the incident probability values, given a random selection of the non-incident registers, Ben Akiva et al., [17], and this affects the risk



threshold tuning stage. Traffic data variables depend on the data available at the site. In our case volume, speed and occupancy per lane per minute were considered.

- Incident data: location of the incident on the road section, time stamp (time at which the incident occurred), severity, duration, etc.
- Weather data: sun, rain, wind, fog, snow, etc.
- Road surface conditions: dry, wet, etc.
- Geometric data: straight or curved road sections, presence of entrance or exit ramps, gantries, etc.

General multivariable regression models are powerful tools that can use a mixture of categorical and continuous variables; however, uncritical application of modelling techniques can result in models that fit the available data set poorly, or even more likely, predict incident risk on new situations inaccurately. To avoid these risks we measured model fits in order to avoid poorly fitted or overfitted models by assessing calibration quality and measuring the predictive accuracy, using the Somers D rank correlation index to quantify the predictive discrimination, Harrell et al., [18]. Discrimination measures the model's ability to separate situations with different responses (high-risk and non-high-risk in our case).

In order to estimate the values of the parameters by maximum likelihood in the generalized linear regression proposal for modelling the association between incident type probabilities and the explanatory variables (*Hierarchical Logit approach*), a particular case of the method of scoring for the estimation of generalized linear models in statistics, McCullagh et al.[16], Dobson [19], has been implemented.

### 3.1 Data collection at the Barcelona test site

The ring road of Barcelona is an urban freeway that articulates the main accesses/exits to and from the city, distributes the traffic around the city and channels the main traffic streams, between two main industrial areas north (from the Trinitat Node) and south (from the Diagonal Node) of the city respectively. The Trinitat Node is an urban/interurban interchange node acting as a collector/distributor of all traffic from/to the motorways A-18 (to/from the Vallès industrial and residential area, generating a large amount of business and commuter traffic every day), A-7 (to/from Girona and the French border) and A-19 (to/from the Maresme industrial area along the Mediterranean coast north of the city, which also generates a large amount of commuter traffic). The Diagonal Node is an urban/interurban interchange node distributing the flows from/to the A-2 motorway, which links the city to other major industrial and residential areas and to the cities of Madrid, Zaragoza and Valencia. The selected site has been the 15 km section of urban motorway between the Trinitat and Diagonal Nodes, a section that interacts closely with densely populated urban areas, causing traffic problems in the ring road to be easily overflow into the neighboring urban arterials and streets and vice versa. The whole test site was used to calibrate, test and evaluate the EIP-HLOGIT module using the field data collected from detectors at the site.

Traffic data (traffic volumes, occupancies and speeds, per lane and aggregated) at each detection station are recorded every minute for all detectors in the site, 24 hours a day, 7 days per week, from Monday 00:00 hours until Monday 00:00 hours. Incident data are recorded by the urban police; all the information gathered corresponds to incidents that have been verified by a police officer. Incident data are recorded during the week also from Monday 00:00 hours

until Monday 00:00 hours. The data recorded for each incident are: incident severity, incident identifier, police officer reporting the incident, incident starting date and time, ending time, duration, location (direction, km point and description of the zone), incident type, queue length and delays caused by the incident and description of road conditions.

These are the data sets employed for EIP-HLOGIT calibration and testing purposes. The testing methodology was designed using independent data sets at each step. The full set of explanatory variables for the Barcelona site is composed for each section and minute from September 2000 to September 2001, namely: length, volume, occupancy, speed, indicator of existence of VMS, indicator of existence of entry and exit ramps. Problems were found concerning the reliability of the collected data, therefore it was necessary to filter out the field data so that only time periods for which reliable traffic and incident data were available were used in the calibration and testing processes.

### 3.2 Testing methodology

The division of the collected data into test data sets was as follows:

- Data set 1: field data from 01.01.01 until 07.30.01 (the whole reliable time period, not including the summer seasonal effect), in order to have the maximum number of incidents, to be used for model selection, calibration and threshold tuning. It contains 646 incidents.
- Data set 2: field data from 01.01.01 until 03.31.01, to be used for the first testing on model selection, calibration and threshold tuning. It contains 295 incidents.
- Data set 3: field data from 03.01.01 until 06.30.01, to be used for the second testing. Recalibration and threshold tuning. It contains 335 incidents.
- Data set 4: field data from 07.01.01 until 08.31.01, to be used for the third testing. It contains 170 incidents.

The testing methodology was as follows:

- A data set (Data Set 1) was used to select the statistical model (specification of explanatory variables), calibrate the model (estimation of model parameters) and carry out threshold tuning. The false alarms, correct estimates and failures of this data set were also analysed. This data set will be referred to as the training data set, using a statistical terminology.
- Subsets of Data Set 1 (i.e., Data Sets 2 to 3) and Data Set 4, reflecting summer seasonal conditions, were used for model testing. They also estimate the incident probability for each of the recorded incident periods and reliable incident-free data during the time period under study, and determine false alarms, correct estimates and failures after partial model redefinition.

Due to the lack of data, since incidents are rare events, a single level (non-hierarchical) model was considered that takes as a response a dichotomic variable indicating the occurrence or otherwise of an incident: no specific incident type model was tested.

Model validation by examination of the apparent accuracy of a generalized linear regression model using the training data set is not very useful [18]. The most stringent test of a model is an external validation -- the application of the 'frozen' model to a new data set -- but data is very expensive to obtain. There are several methods for obtaining internal assessments of accuracy: internal model validation, data-splitting, cross-validation and

bootstrapping. An approach to data-splitting internal validation was selected, given the constraint that incidents are rare events and incident data are too precious to waste.

The Estimation Rate, ER, is defined as the number of pass results (successful prediction of incident) divided by the total number of incidents selected from the historical database. The target value defined in the PRIME Project was 80%.

The False Alarm Rate, FAR, is defined as the number of false alarms raised over the total test period divided by the total number of test records in the test period. The target value defined in the PRIME Project is 10%. A false alarm occurs when a high risk is predicted for a section in a given time period in a non-incident situation; this happens when the probability exceeds the high-risk threshold, but no incident occurred.

The purpose of the EIP-HLOGIT model is to identify dangerous situations that might lead to an incident occurrence. Thresholds are tuned in such a way that incidents are properly identified, all incidents (or at least as many as possible), and yielding an Estimation Rate, ER, that should be as high as possible. There is a trade-off between Estimation Rate and False Alarm Rate. Since incidents are rare events, risk thresholds must be tuned in such a way that the Estimation Rate is high and the False Alarm Rate is as low as possible. FAR proved to be very high and the target values in the PRIME Project were not met, since the threshold for high-risk situations must be tuned to a very low probability value.

### 3.3 Results for Data Set 1: Model Selection, Calibration and Threshold Tuning

The statistical results after model selection, calibration of model parameters and threshold tuning show: 2404 alarms raised over 6389 time periods, 982 of them raised in an incident period situation (1536 registers in a total of 6389), some of them corresponding to 219 incidents of a total of 283 included in Data Set 1, leading to an ER of 78%. In contrast, 1422 alarms raised in a non-incident situation and 554 incident situations were not detected as 'high-risk', leading to a failure result of 1976 of 6389 registers leading to a Failure Rate of 31%, and a FAR of 1422 to 6389 or 22% (see Table 1).

Threshold tuning for high-risk setting was established according to statistical analysis of the estimated probabilities in incident and non-incident situations

In the model selection phase, the model containing direct variables (from the Historical Datastore) was not explanatory. In order to improve the model quality, the statistical analysis conducted led to a discretization of some continuous variables available for the Barcelona site in order to smooth real data and clarify the model selection process. The discretization considered was the following:

- Volumes were classified into four groups: Low (0 to 1500, represented by 600), Medium (1500 to 3000, represented by 2100), High (3000 to 4500, represented by 3600) and Jam (more than 4500, represented by 5100).
- Speeds were classified into four groups: Low (0 to 50, represented by 30), Medium (50 to 75, represented by 60), High (75 to 100, represented by 90) and Extreme (more than 100, represented by 120).
- Sections were classified according to their length: Short, less than 300 m (represented by 200); Medium, 300 to 600 m (represented by 450); Long, more than 600 m (represented by 700).
- Occupancies were classified in four groups: Low, from 0 to 15% (represented by 5%), Medium, from 15 to 25% (represented by 20%), High, from 25 to 50% (represented by 35%) and Jam, over 50% (represented by 75%).

The most significant variables found were speed and occupancy. Speed is included in the statistical model as a covariate, and terms of order 1, 2 and 3 are necessary (terms of order 2 and 3 are included centered with respect to the mean speed). Occupancy is included in the model as a factor and volume is included as a covariate. From a statistical point of view, section length has no validity as an explanatory variable.

The statistical analysis showed that a simple additive model including speed and occupancy was not statistically significant. However, interactions between linear, quadratic and cubic terms of speed and discretized occupancy were found to be significant, and therefore it was necessary to include them in the model.

The resulting statistical model uses 16 degrees of freedom out of a total of 5697 (number of covariate classes in statistical terms), and it conforms to Pearson and Hosmer-Lemeshow statistical tests of goodness of fit.

The constant (intercept) coefficient is not meaningful for the proposed model. The coefficient in the additive scale represented by the logit transformation of estimated probabilities for each level of the factor occupancy shows non-linear behavior, since the likelihood (of incident occurrence) in the 20% level of occupancy increases by 6182% with respect to that of the reference occupancy (5%, the lowest), but the likelihood decreases for 75% occupancy. First order effects of the covariates speed and volume show a negative coefficient, indicating a reduction in the probability of incident occurrence as speed and volume increase. The key variables in the model are the second and third order interactions between the factor occupancy and speed, which lead to significant third order curves on the logit scale on speed, one for each level of the factor occupancy. These have the corrective effect of decreasing the probabilities of the linear effects with low occupancies and increasing the probabilities with high occupancy.

For low occupancy situations, lower speeds identify higher risk situations. For high occupancy (75%), the observed risk increases from 0 to 15 kph, but decreases as speed increases. Obviously, no simple model can be adjusted to observed probabilities in the transformed logit scale. For low speeds, the model is able to reproduce observed data, but the behavior is very difficult to understand and reproduce for high-speed situations. The estimated model is very *soft* (a third order curve is estimated for each occupancy level), compared to observed probabilities in the logit scale which are extremely non-linear (and show a very different pattern for each occupancy level), but from a statistical point of view, given the small number of incidents present in the reference data, more complex models could be too sparse, leading to non-valid inference results, *in other words, for the amount of available data, the model is complex enough*, and in fact, observed incident probabilities (in the logit scale) have been shown to be very difficult to model.

In order to mathematically describe the estimated third order curves in the logit scale of incident probability, let  $x$  be the covariate speed, and  $y$  be the covariate volume, then the estimated probability of incident occurrence (any incident type) for section  $i$  can be described as:

$$p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$$

Where for current section  $i$ , if its level of occupancy is 5%:

$$\eta_i = -0.3504 - 0.014x_i - 0.0000739y_i + 0.000083(x_i - \bar{x})^2 + 0.0000013(x_i - \bar{x})^3 = \eta_i^{5\%}$$

For current section  $i$ , if its level of occupancy is 20%:

$$\eta_i = \eta_i^{5\%} + 4.14 - 0.0489x_i - 0.002(x_i - \bar{x})^2 - 0.0000197(x_i - \bar{x})^3$$

For current section  $i$ , if its level of occupancy is 35%:

$$\eta_i = \eta_i^{5\%} + 4.524 - 0.0688x_i - 0.000566(x_i - \bar{x})^2 - 0.00000308(x_i - \bar{x})^3$$

For current section  $i$ , if its level of occupancy is 75%:

$$\eta_i = \eta_i^{5\%} - 14.603 + 0.2454x_i + 0.00873(x_i - \bar{x})^2 + 0.000079(x_i - \bar{x})^3$$

All coefficients are statistically significant.

A similar analysis was conducted for each data set. The results are summarized in Table 1. A detailed description can be found in [20].

The results of testing the performance of the EIP-HLOGIT can be summarized in terms of False Alarm Rate versus Estimation Rate, the most significant MOE's for the authors (see Table 1). The trade-off between these two performance indicators achieved by the tests conducted with the data sets collected during the PRIME Project's life shows that the EIP-HLOGIT can be interpreted as a promising method for early identification of incident conditions and represents an added value to the traditional analysis based only on aggregated traffic variables. The quality of the results, in spite of the reported inadequacy of a significant part of the collected data, is good enough to encourage the follow-up of the tests with new and more accurate data. In particular, from the available data sets and corresponding results it is clear that:

- The quality can be improved if data for longer periods are available.
- The observed discrepancies between the results for the various data sets indicate a clear seasonal impact. Data sets for longer periods will allow us to verify this perceived tendency and, if applicable, to adjust the models for different seasons.
- Unfortunately the current recording methods have not allowed us to collect data for some other variables initially suspected as being of potential interest. The achieved results suggest that if it were possible to collect data for additional variables, the quality of the estimation would be substantially improved.

According to high-risk threshold tuning criteria, Estimation Rate and False Alarm Rate can vary substantially for each data set. The results for Data Set 3 are depicted in Figure 4. The criteria for fixing the 'high-risk' threshold are the 25% and 75% quartile expected probability of incident (predicted for the model) under a real incident situation and under a real incident-free situation.

## 4. CONCLUSIONS

### 4.1 Safety indicators for micro-simulation assessment

The "unsafe" density parameter is an important indicator for safety assessment and gives more accurate information than typical micro-simulation outputs. It allows, among other things, to highlight the difference in safety level between a fluid and a "jerked" traffic flow situation, which cannot be shown by using traditional macroscopic outputs like speed, flow or occupancy. The "unsafe" density parameter is based on the direct interaction between pairs of vehicles, which is the most appropriate for treating safety problems.



However, some limitations must be taken into account. The value of this parameter doesn't really have a sense in itself and must be used only for comparison purposes. The "unsafe" density parameter takes into account only potential for rear-end collision and is therefore particularly planned for highways network assessments.

#### 4.2 Estimation of Incident Probability

The EIP-HLOGIT statistical approach for the estimation of dynamic incident probabilities has achieved promising results in terms of the Estimation Rates and False Alarm Rates reached, over 80% for ER and an average FAR in the range 20-40%, making the approach suitable for application as an early indicator of static and dynamic variables affecting incident conditions and for assisting traffic managers in applying preventive strategies. Furthermore, the innovative approach that the flexibility of a statistical general regression model applied to dynamic data represents provides information for real-time traffic management purposes that goes beyond what is achievable using traditional static analysis based on aggregated variables [13], [14], [15].

The strongest points are the relative simplicity of the statistical model and the systematic procedure provided for the model calibration process. The EIP-HLOGIT concept is applicable to any road network provided it is possible to collect the sensitive data. The integration of the EIP-HLOGIT in the GETRAM-AIMSUN micro-simulation environment allows the assessment of incident management strategies and improves the possibilities of AIMSUN as a component of a traffic management tool.

As reported in the results of EIP-HLOGIT for Barcelona's test site, data collection is still the bottleneck to applying this statistical approach. Nevertheless, despite the problems with field data in the Barcelona site (inconsistent field data, different departments of Barcelona City Council as sources for traffic and incident data), results in terms of statistical association between traffic data and incident occurrence are very promising, leading to meaningful explanatory models in all cases. The weak point in the approach identified during testing is the high False Alarm Rate (it does not meet the target value of 10%), which we hope could be improved if better data, reliable data for longer time periods, and data for additional variables were available.

#### 5. REFERENCES

- [1] J.F. Gabbard, Car-Following Models. In: M. Papageorgiou (Ed.), Concise Encyclopedia of Traffic and Transportation Systems, Pergamon Press, Oxford, (1991).
- [2] D.L. Gerlough and M.J. Huber, Traffic Flow Theory – A Monograph, Transportation Research Board, Special Report 165, Chapter 6, TRB, 1975.
- [3] M. Mahut, 2000, Behavioral Car Following Models, Ph. D. Thesis, Chapter 3, Centre de Recherche sur les Transports, CRT, Université de Montréal,.
- [4] J. Archer, Developing the potential of micro-simulation modelling for traffic safety assessment, 13th ICTCT workshop, Corfu, 2000.
- [5] M. Minderhoud, P. Bovy, Extended time-to-collision measures for road traffic safety assessment, Accident Analysis and Prevention 33, p. 89–97, 2001.
- [6] I. Kosonen, and S. Ree, The Potential of Microscopic Simulation in Traffic Safety and Conflict Studies, Traffic Safety on Three continents Conference, Pretoria, South-Africa, 2000.



- [7] J. Barceló, J.I. Ferrer, R. Grau, M. Florian, I. Chabini and E. Le Saux, 1995, A Route Based variant of the AIMSUN Microsimulation Model. Proceedings of the 2<sup>nd</sup> World Congress on Intelligent Transport Systems, Yokohama.
- [8] J. Barceló, J.L. Ferrer, D. García, M. Florian and E. Le Saux, 1998, Parallelization of Microscopic Traffic simulation for ATT Systems Analysis. In: P. Marcotte and S. Nguyen (Eds.), Equilibrium and Advanced Transportation Modeling, Kluwer Academic Publishers.
- [9] Barceló J., Casas J., Ferrer J.L. and García D (1999) Modeling Advanced Transport Telematic Applications with Microscopic Simulators: The case of AIMSUN, in: Traffic and Mobility, Simulation, Economics, Environment, W. Brilon, F. Huber, M. Schreckenberg and H. Wallentowitz (Eds.), Springer Verlag.
- [10] AIMSUN and TEDI, Version 4.1 User's Manuals, TSS-Transport Simulation Systems, 2002.
- [11] P.G. Gipps, A behavioral car-following model for computer simulation. *Transp. Res. B*, Vol. 15, pp 105-111, (1981).
- [12] F. Bühlmann, H.P. Lindenmann and Spacek P. Sichtweiten, Überprüfen der Grundlagen zur Norm SN 640 090, Projektierungsgrundlagen, Sichtweiten, mandat de recherche VSS 17/84, IVT ETH Zürich, 1991.
- [13] Hamerslag, R., Roos, J.P., Kwakernaak, M., 1982. Analysis of Accidents in Traffic Situations by Means of Multiproportional Weighted Poisson Model. *Transportation Research Record* 847, 29-36.
- [14] Miaou, S.P. 1994: The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis and Prevention*, 26:471-482.
- [15] TRAVELAID, 2001. Washington State Transportation Commission, Dep. of Transportation and U.S. Department of Transportation, Federal Highway Administration, Final Report Research Project T9903, December 2001.
- [16] McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*. Chapman and Hall.
- [17] Ben-Akiva, M., Lerman, S.R., 1985. *Discrete choice analysis: theory and application to travel demand*. MIT Press Cambridge.
- [18] Harrell, F.E., Kerry, L.L., Mark, D.B. (1996). *Multivariable Prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors*. *Statistics in Medicine* (John Wiley & Sons), Vol. 15, 361-387.
- [19] Dobson, A., 1990. *An Introduction to Generalized Linear Models*. Chapman and Hall.
- [20] L. Montero, J. Barcelo and J. Perarnau, A Statistical Approach to Dynamic Incident Probability Estimation. Application to a test Site in Barcelona, May 2002, submitted for publication in *Transportation Research Series C*.

## LIST OF TABLES

Table 1. Barcelona test site: EIP-HLOGIT performance

## LIST OF FIGURES

Figure 1: One hour average UD for each of the AIMSUN micro-simulation model's sections

Figure 2: Evolution of the UD on a section (ID 35) between both entrances during the simulation process

Figure 3: Evolution of the UD on a the entrance area section (ID 43) during the simulation process

Figure 4. EIP-HLOGIT: FAR-ER trade-off for Data Set 3

**TABLES**

<i><b>EIP Validation Measure of Effectiveness</b></i>	<i><b>Data Set 1</b></i>	<i><b>Data Set 2</b></i>	<i><b>Data Set 3</b></i>	<i><b>Data Set 4</b></i>
<b>Estimation Rate</b>	219 of 283 incidents; i.e., 78%	19 of 22 incidents; i.e., 86%	210 of 254 incidents; i.e., 83%	18 of 23 incidents; i.e., 78%
<b>False Alarm Rate</b>	1422 of 6389 registers; i.e., 22%	900 of 2741 registers; i.e., 32%	1211 of 4005 registers; i.e., 30%	182 of 482 registers; i.e., 37%

Table 1. Barcelona test site: EIP-HLOGIT performance

FIGURES

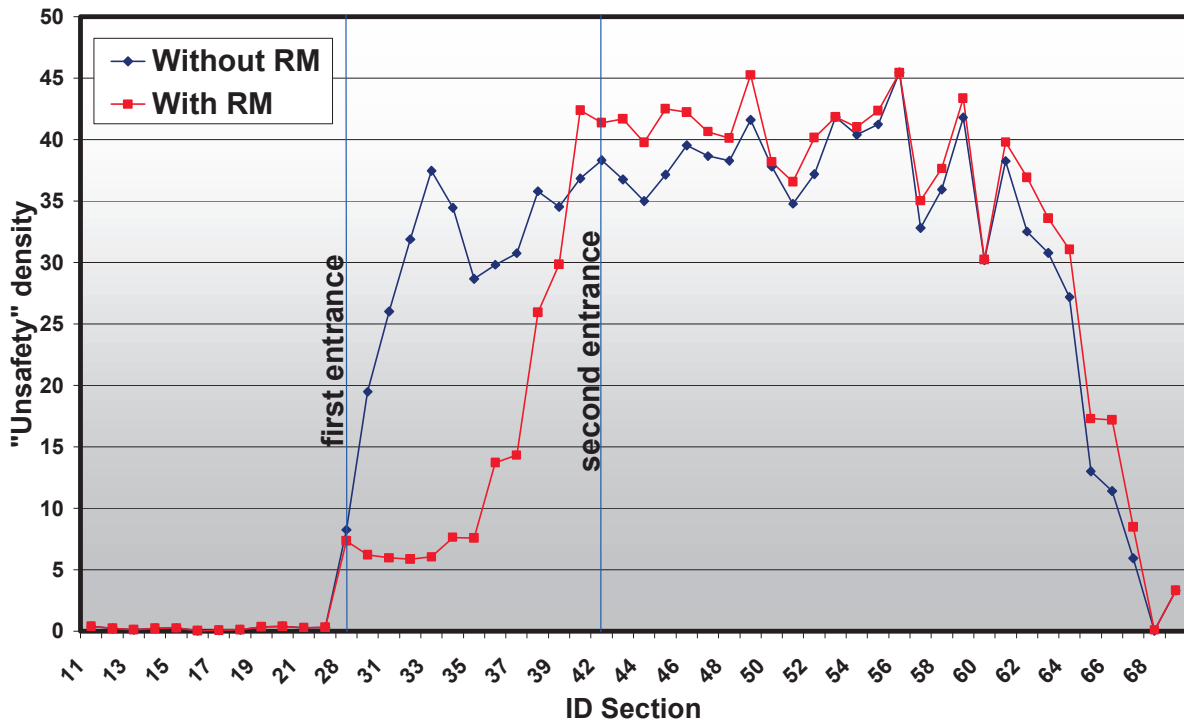


Figure 1: One hour average UD for each of the microsimulation model's sections

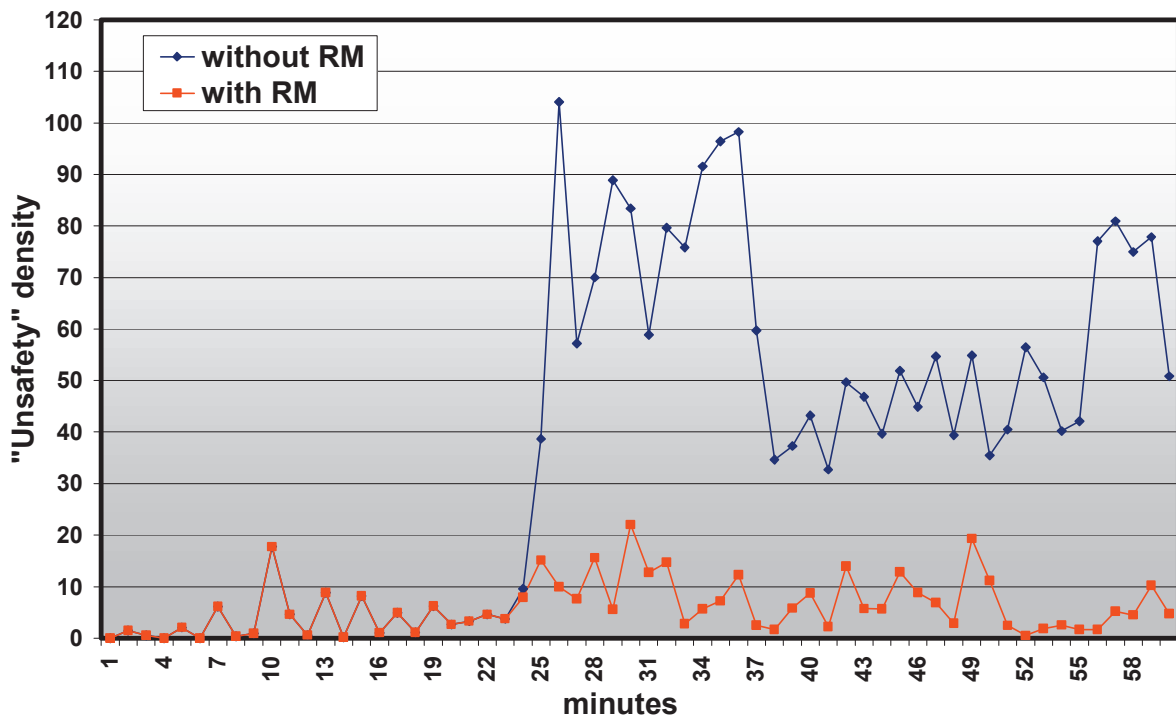


Figure 2: Evolution of the UD on a section (ID 35) between both entrances during the simulation process

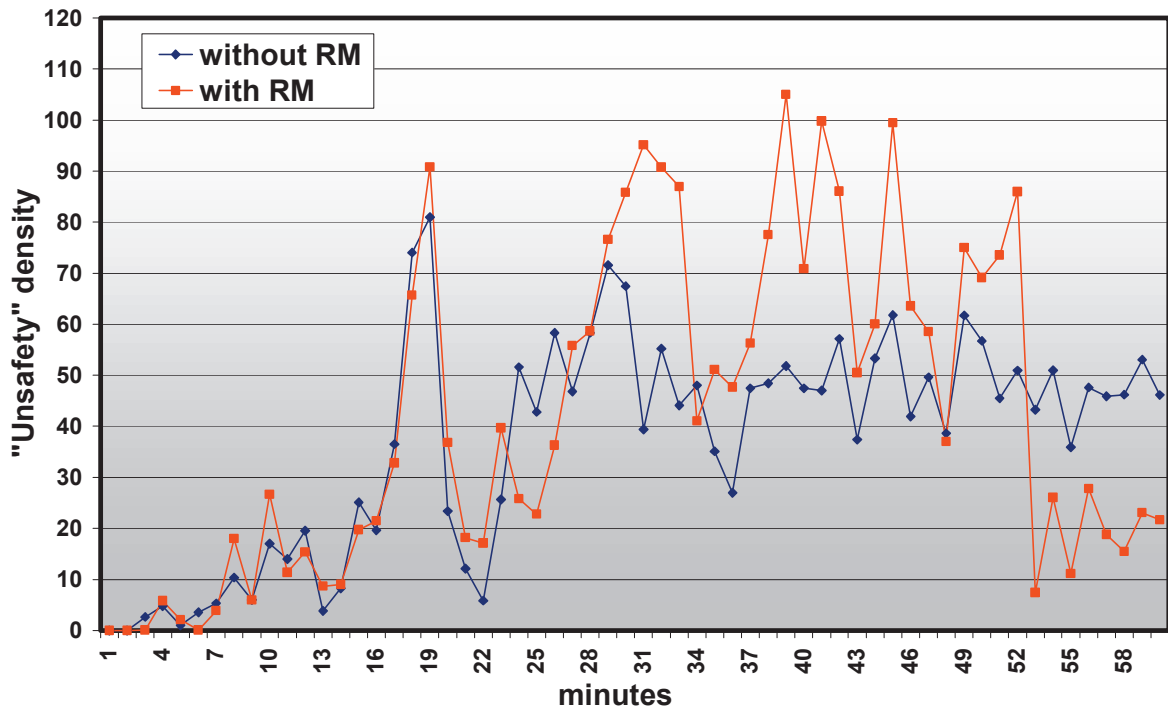


Figure 3: Evolution of the UD on a the entrance area section (ID 43) during the simulation process

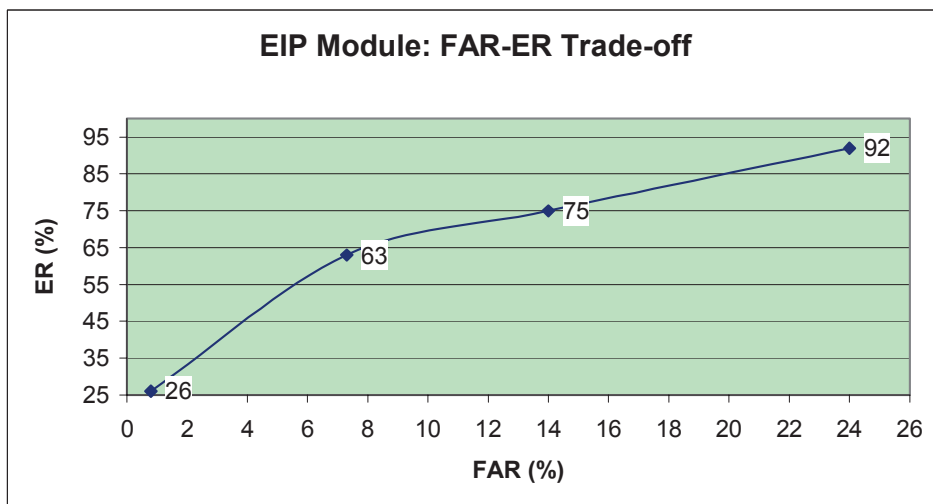


Figure 4. EIP-HLOGIT: FAR-ER trade-off for Data Set 3