INCIDENT PREDICTION: A STATISTICAL APPROACH TO DYNAMIC PROBABILITY ESTIMATION. APPLICATION TO A TEST SITE IN BARCELONA.

L. MONTERO[1], J. BARCELÓ[1,2], J. PERARNAU[1]

[1] Dept. Estadística i Investigació Operativa. Universitat Politècnica de Catalunya,

C/ Pau Gargallo 5, 08028 Barcelona (SPAIN). Tel. 34.93.401.10.38. Fax. 34.93.401.58.55. E-mail: LIDIA.MONTERO@UPC.ES

[2] TSS (Transport Simulation Systems). C/ Paris, 101 08029 Barcelona (Spain)

ABSTRACT - Real-time models for estimating incident probabilities (EIP models) are innovative methods for predicting the potential occurrence of incidents and improving the effectiveness of incident management policies devoted to increasing road safety. EIP models imbedded in traffic management systems can lead to the development of control strategies for reducing the likelihood of incidents before they occur. This paper presents and discusses the design, implementation and off-line testing of an EIP model in the PRIME (Prediction of Congestion and Incidents in Real Time for Intelligent Incident Management and Emergency Traffic Management) Project of the "Information Societies Technology Programme" of the EU. A statistically-oriented approach based on Generalized Linear Regression models with polytomous responses is developed: geometry, traffic and weather conditions are taken as explanatory variables at a road section level and a binary variable related to incident occurrence or otherwise for the prevailing conditions is taken as a response variable on the first level of decision. Once the probability of a generic incident has been predicted, the lower level models in the selected hierarchical approach will predict the probabilities of

incidents in a set of categories defined at a design level.

The EIP model has been incorporated in the AIMSUN microscopic simulation environment (developed by TSS[1]). AIMSUN is able to emulate a traffic management system, since it simulates traffic evolution including the replication of observed incidents and incorporates different modules of incident and traffic management in such a way that the impact of traffic management strategies can be evaluated by simulation.

A test site in Barcelona, located in a 15-km portion of the *Ronda de Dalt* ring road provided the data for calibrating and testing the EIP module. The selected site is equipped with 12 CCTV cameras for traffic monitoring, 18 local controllers, 12 detection stations, 10 variable message panels and 13 variable speed signals. Detection stations provide measures of different traffic variables in lane detail every minute.

Keywords - Transportation management, incident prediction, incident management, generalized linear regression models, microscopic simulation.

[1] TSS (TRANSPORT SIMULATION SYSTEMS). C/ Paris, 101   08029 Barcelona (Spain).

INCIDENT PREDICTION: A STATISTICAL APPROACH TO DYNAMIC PROBABILITY ESTIMATION. APPLICATION TO A TEST SITE IN BARCELONA.

L. MONTERO[1], J. BARCELÓ[1,2], J. PERARNAU[1]

[1] Dept. Estadística i Investigació Operativa. Universitat Politècnica de Catalunya,

C/ Pau Gargallo 5, 08028 Barcelona (SPAIN). Tel. 34.93.401.10.38. Fax. 34.93.401.58.55. E-mail: lidia.montero@upc.es

[2] TSS (TRANSPORT SIMULATION SYSTEMS). C/ Paris, 101 08029 Barcelona (Spain)

INTRODUCTION

Real-time models for automatic incident detection and estimation of incident probabilities help to improve the effectiveness of incident management policies devoted to increasing road safety. New, advanced-technology hardware and software and new models make it possible to improve monitoring, surveillance and management of high-risk road locations in urban and rural areas of the European Union (IN-RESPONSE, 1997). Fast and reliable detection and prediction models for incidents, imbedded in traffic management systems, are instrumental in the development of control strategies to reduce traffic delay and the likelihood of new incidents before they occur.

As part of the IN-RESPONSE Project, a procedure for predicting the probability of incidents based on the previous work of Hamerslag (Hamerslag et al., 1982) was developed. The procedure was 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 Hamerlag's approach. The idea of a logit statistical model as a way of improving Hamerlag's approach was suggested in IN-RESPONSE, but due to the lack of available data it was neither implemented nor tested. This was the reason for proposing a follow-up in the PRIME (**Prediction of Congestion and Incidents in Real Time for Intelligent Incident Management and Emergency Traffic Management**) Project in the "Information Societies Technology Programme" of the European Union, with the purpose of completing the research: the design, implementation and testing of the new approach. As a consequence of the ongoing research a more refined model became available for the dynamic estimation of incident probabilities (EIP). The model that was developed and implemented is based on Generalized Linear Regression models with polytomous responses (McCullagh et al., 1989, Dobson, 1990): geometry, traffic and weather conditions are taken as explanatory variables at a road section level and a polytomous variable related to incident type occurrence for the prevailing conditions is taken as a response variable. On the first level of decision, a binary variable related to incident occurrence or otherwise is taken as the response variable, since a hierarchical approach was chosen for dealing with the polytomous response variable.

The EIP model was included in the microscopic simulation environment GETRAM-AIMSUN (developed by TSS[2]) (Barceló et al., 1999, 1998, 1991). AIMSUN can emulate a traffic management system by simulating the traffic evolution, replicating observed incidents and interfacing modules of incident and traffic management. It is in this way that the impact of traffic management strategies can be evaluated by simulation (see Figure 1).

A test site in Barcelona, located in a 15-km portion of the *Ronda de Dalt* Ring Road, supplied data for building and calibrating the microscopic simulation model of the site and for calibrating, testing and validating the EIP module. The selected site is equipped with 12 CCTV cameras for traffic monitoring, 18 local controllers, 12 detection stations, 10 variable message panels and 13

[2] TSS (TRANSPORT SIMULATION SYSTEMS). C/ Paris, 101 08029 Barcelona (Spain)

variable speed signals. Detection stations provide measures of different traffic parameters in lane detail every minute.

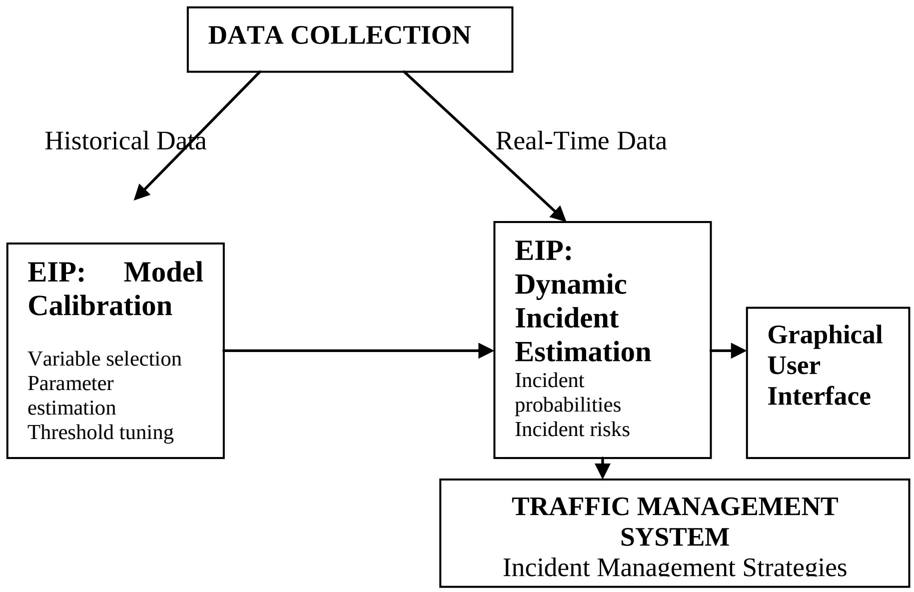


Figure 1. Estimation of Incident Probability Module in a Traffic Management System

APPROACH TO THE ESTIMATION OF INCIDENT PROBABILITIES

The aim of the dynamic incident probability estimation module (EIP module) 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 (Barceló et al., 2001). An alternative approach to the estimation of incident probabilities based on fuzzy logic and expert systems has also been proposed in the PRIME Project (Wilmink et al. 2001), but it is not discussed in this paper.

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

$$p_{ij} = \frac{\mathbf{exp}(\boldsymbol{\eta}_{ij})}{1 + \mathbf{exp}(\boldsymbol{\eta}_{ij})} \quad (1)$$

where $\boldsymbol{\eta}_{ij} = \mathbf{x}_{\mathbf{ij}}^{\mathbf{T}} \boldsymbol{\beta}_j$ 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* ($\mathbf{x}_{\mathbf{ij}}^{\mathbf{T}}$'s). Equation (1) is equivalent to

$$\mathbf{logit}\, p_{ij} = \mathbf{log}\frac{p_{ij}}{1 - p_{ij}} = \boldsymbol{\eta}_{ij} \quad (2)$$

which shows a clearer relation to generalized linear regression models for binary response for each hierarchical level and logit link function (McCullagh and Nelder, 1989).

An alternative approach that has not yet been considered relies on the proposal of a generalized linear regression model that deals directly with a polytomous response (as many possible responses as incident types or incident severity classes). This proposal can be studied only when a large amount of historical data is available for statistical analysis.

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., 1989, King et al., 2000), 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., 1996). 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 C++ procedure was developed during the project, and a particular case of the method of scoring for the estimation of generalized linear models in statistics (McCullagh et al., 1989, Dobson, 1990) was implemented.

DATA COLLECTION AT THE BARCELONA TEST SITE

In 1992, as part of the completion and improvement of the city road network infrastructure for the Olympic Games, a 43-km ring road was completed. The ring road is functionally split into two parts, the northern *Ronda de Dalt*, which crosses the upper part of the city at the feet of the Tibidabo hills, and the southern *Ronda Litoral* running parallel to the Mediterranean coast for most of its length. They have three lanes in both directions for most of their length, with the exception of the stretch of the *Ronda Litoral* along the harbour, which has only two lanes.

Traffic monitoring and management on the ring road is carried out from the Traffic Management Centre at the Collserola Node on the *Ronda de Dalt*, which also serves as a Traffic Information Centre. In the event of incidents requiring the intervention of emergency units (urban police, fire brigade, ambulances, etc.) the Traffic Management Centre activates the alarms and sends the available information to the Incident Management Centre, which is responsible for the intervention.

Figure 2 depicts the digital map of the whole ring road system (as a DXF background imported into the GETRAM environment, designed for building the AIMSUN micro simulation model). The highlighted area shows the site selected for the PRIME Project: the part of the *Ronda de Dalt* between the Trinitat Node roundabout in the top right-hand corner and the Diagonal Node in the bottom left-hand corner of the figure.

The ring road 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 the 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.

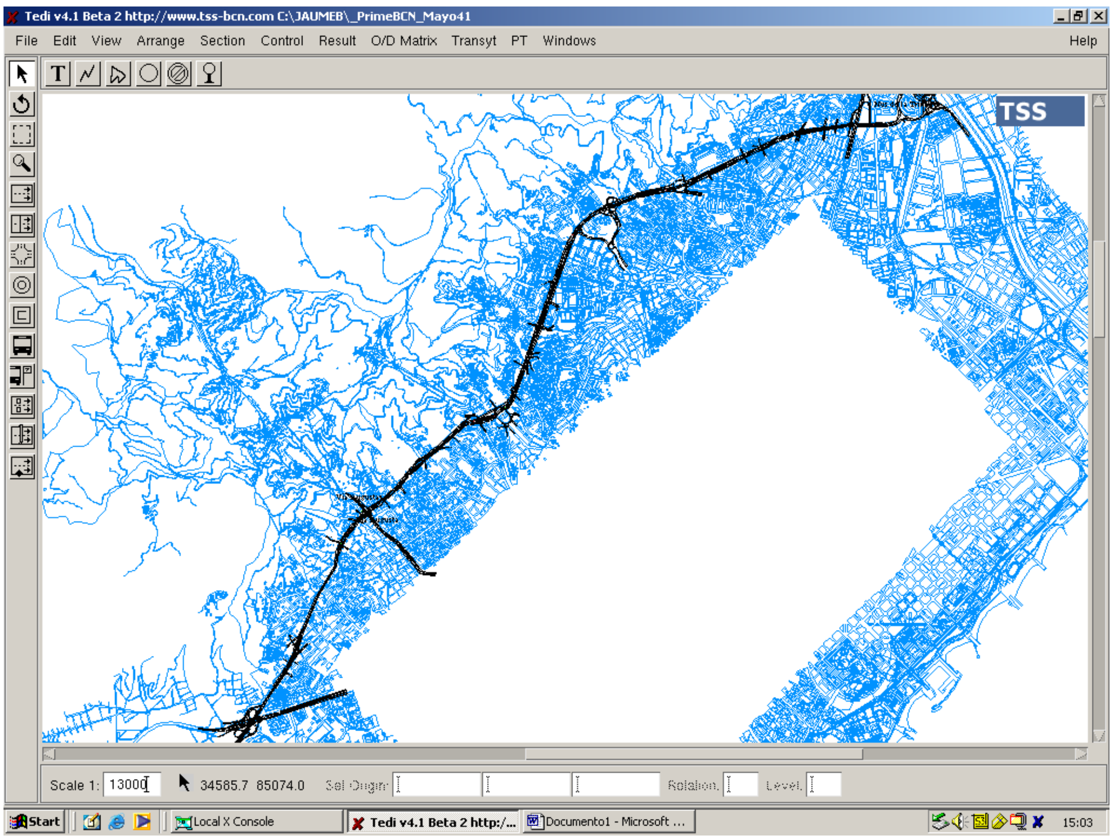


Figure 2. The Barcelona ring road and the PRIME site

The 15 km of urban motorway between the Trinitat and Diagonal Nodes interact closely with densely populated urban areas, causing traffic problems in the ring road to be easily overflow into the neighbouring 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. Table 1 provides an example of the collected data: detector identifier, date and time, aggregated link flow, speed and occupancy, and flow, speed and occupancy per lane every minute.

Detector ID	Date	Time stamp	Agg. flow	Agg. Occ.	Agg. Spd.	Flw. Lane 1	Occ. Lane 1	Spd. Lane 1	Flw. Lane 2	Occ. Lane 2	Spd. Lane 2	Flw. Lane 3	Occ. Lane 3	Spd. Lane 3
08 ESP-1	08-01-2001	12:05:00	2520	13	91	840	10	104	1440	24	87	240	6	78
08 ESP-1	08-01-2001	12:06:00	2820	15	91	1080	14	104	1380	25	83	360	7	87
08 ESP-1	08-01-2001	12:07:00	3360	19	90	1380	19	102	1440	26	84	540	12	79
08 ESP-1	08-01-2001	12:08:00	2340	12	92	720	9	103	1200	20	89	420	8	86

Table 1. Barcelona site: file format for traffic data

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 as Excel files, with the incident data recorded during the week from Monday 00:00 hours until Monday 00:00 hours. A sample of the contents of each record is shown in Table 2: 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.

Severity	Incident ID	Officer	Start date	Start time	End time	Duration	Direction	Km point	Zone	Type	Queue length	Delay	Road conditions

Table 2. Barcelona site: file format for incident data

The data sets employed for EIP-HLOGIT calibration and testing purposes are field data sets used in off-line analysis; no simulated data were included. The testing methodology was designed using independent data sets as much as possible at each step. The statistical process of model selection and calibration was undertaken using an external statistical package (MINITAB), for software validation purposes.

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.

The filtering results for each data set are described below as part of the description of the test results.

TESTING METHODOLOGY

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

- Data set 1: field data from 1.01.01 until 30.07.01 (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 1.01.01 until 31.03.01, to be used for the first testing on model selection, calibration and threshold tuning. It contains 295 incidents.

- Data set 3: field data from 1.03.01 until 30.06.01, to be used for the second testing. Recalibration and threshold tuning. It contains 335 incidents.

- Data set 4: field data from 1.07.01 until 31.08.01, 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 (King et al., 2000), 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 (Harrell et al., 1996). 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 results) 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.

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

Data used for model selection (study of statistically significant variables to be included in the model), calibration and threshold tuning. Time period: from 2001-01-03 18:40:00.00 to 2001-07-30 18:48:00.00. The data set consists of a total number of 283 incidents over seven months, related to consistent traffic data. The process of filtering inconsistent data and randomly selecting non-incident situations data yielded 6389 registers, each register belonging to a section and timestamp (one minute time period), some of them reflecting ‘incident conditions’ and the rest ‘incident free’. For each incident available, 6 minutes of traffic data from the same section were included in the current data set.

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. Table 3 details the results: for an incident situation (1536 registers in a total of 6389), 982 alarms were raised, 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 3).

Tabulated statistics: incident; indicator of results

Rows: incident   Columns: indicator

	fail	pass	all
0	1422	3431	4853
1	554	982	1536
All	1976	4413	6389

Cell contents -- Count

Table 3. Barcelona test site: results for Data Set 1

Threshold tuning for high-risk setting was established according to statistical analysis of the estimated probabilities in incident and non-incident situations: for non-incident situations, the estimated probability has a Q3 quartile of 0.21467 (Q3 is the probability value that is exceeded in only 25% of the situations in which incidents are not present); probabilities are extremely asymmetrical, and comparing to quartiles in incident situations, estimated probability intervals overlap. A good statistical criterion for setting the high-risk threshold seems to be the Q1 quartile, the 25% quartile of the estimated incident probability in an incident situation. The trade-off was chosen in such a way that probabilities higher than the Q1 quartile of probabilities in incident situations were considered to identify a ‘high-risk situation’, i.e., the threshold value for ‘high risk’ was established at 0.2 or 20%. Thresholds for ‘low’ and ‘medium’ risk situations were not used to validate and test the EIP-HLOGIT model. Taking into account the threshold setting procedure, fixed values are detailed in Table 4.

Estimated incident probability	***RISK***
$0.0 < p_i \leq 0.01$	*None-Low*
$0.01 < p_i \leq 0.2$	*Medium*
$0.2 < p_i \leq 1$	*High-Extreme*

Table 4. Barcelona test site: risk thresholds for Data Set 1

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 by means of the statistical package MINITAB 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 centred 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 included in the statistical package MINITAB (13.10) for Windows.

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 behaviour, 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.

Figure 3 illustrates the empirical probabilities of incident, in the logit scale (this means a logarithmic transformation of the empirical odds), depending on speed (X-axis), classified according to the level of occupancy (occupancy factor).

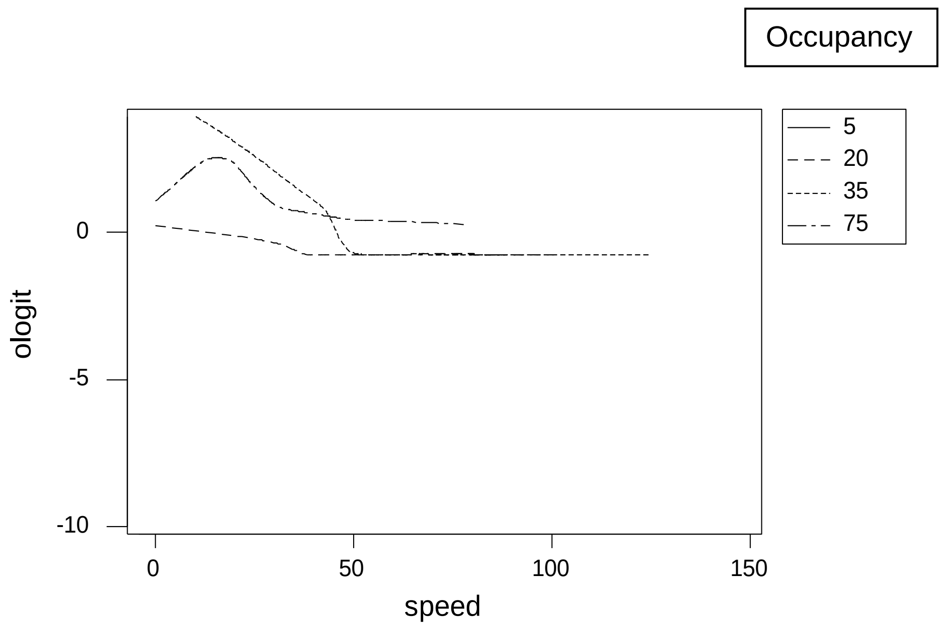


Figure 3. Observed incident probabilities (logit scale) vs. speed according to occupancy levels

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 behaviour is very difficult to understand and reproduce for high-speed situations. The estimated model (see Figure 4) is very *soft* (a third order curve is estimated for each occupancy level), compared to observed probabilities in the logit scale (Figure 3), 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.

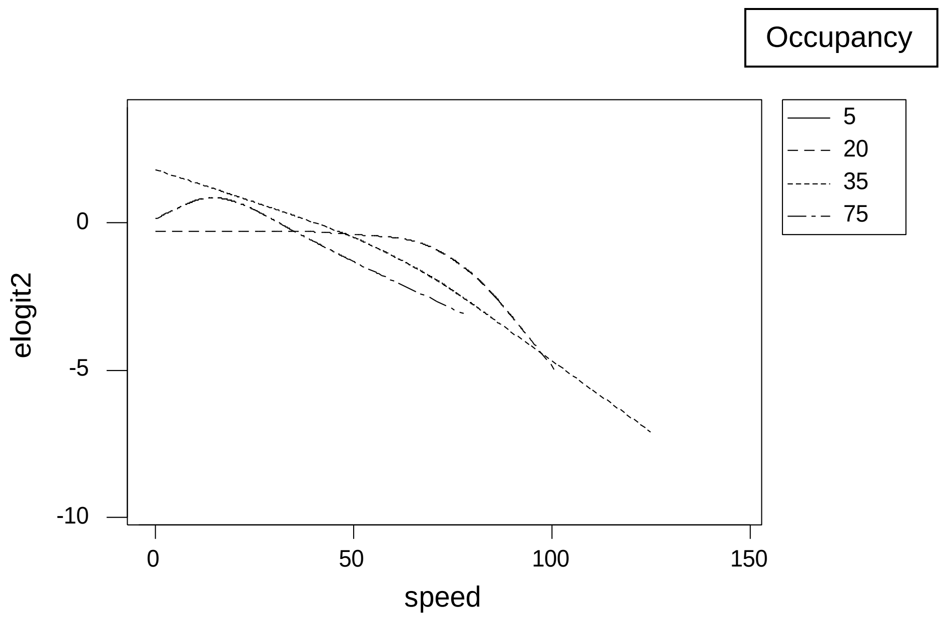


Figure 4. Estimated incident probabilities (logit scale) vs. speed according to occupancy levels

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%:

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

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

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

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

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

All coefficients are statistically significant.

Results for Data Set 2: Model Selection, Calibration and Threshold Tuning

It consists of a total number of 22 incidents over two weeks, related to consistent traffic data. Data Set 2 includes a test period from 2001-01-03 18:40:00.000 to 2001-01-17 18:15:00.000.

The process of filtering inconsistent data and randomly selecting non-incident situations data led to 2741 registers, each one belonging to a section and timestamp, some of them reflecting 'incident conditions' and the rest 'incident free'. For each incident available, 5 minutes of traffic data from the same section were selected to define Data Set 2.

The model selection phase was repeated in order to test whether the current model selected for Data Set 2 is the same as the one selected for Data Set 1 in terms of selected explanatory variables. The estimated parameters (coefficients in the linear predictor of the logit scale) varied, and their validity will be studied in Data Set 3. The discretization of occupancy studied in the model definition is the same as that proposed for Data Set 1.

Again a high order model was found to be the most significant, and the same model was selected, confirming the validity of the proposal found in Data Set 1. In this case the statistical model uses the same 16 degrees of freedom as before out of a total of 2470, and conforms to the Pearson and Hosmer-Lemeshow statistical tests of goodness of fit included in the statistical package MINITAB (13.10) for Windows. The model selection results provided by Data Set 2 are fully consistent with the model selection results found using Data Set 1.

In order to describe the estimated third order curve model in the logit scale of incident probability, let x be the covariate speed, and y be the covariate volume, then the estimated probability 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%:

$$\eta_i = -8.52 + 0.062x_i - 0.000259y_i - 0.000158(x_i - \bar{x})^2 - 0.0000415(x_i - \bar{x})^3 = \eta_i^{5\%}$$

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

$$\eta_i = \eta_i^{5\%} + 16.23 - 0.2095x_i + 0.000914(x_i - \bar{x})^2 + 0.0001832(x_i - \bar{x})^3$$

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

$$\eta_i = \eta_i^{5\%} + 11.434 - 0.1329x_i - 0.003229(x_i - \bar{x})^2 - 0.0000408(x_i - \bar{x})^3$$

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

$$\eta_i = \eta_i^{5\%} - 123.66 + 1.9955x_i + 0.05463(x_i - \bar{x})^2 + 0.000464(x_i - \bar{x})^3$$

The coefficients of the explanatory variables are shown to be quite different from the coefficients estimated with Data Set 1. This data set presents few incidents and the coefficient estimates are very imprecise. Even though selected model variables are consistent between Data Set 1 and Data Set 2, special care has to be taken in order to recalibrate the model parameters: recalibration on small data sets leads to high standard errors of the estimated coefficients and hence to low precision in the computation of the estimated incident probabilities.

The statistical results after model selection, calibration of model parameters and threshold tuning show: 988 alarms raised over 2741 time periods, 88 of them raised in an incident period situation, identifying 19 of the 22 incidents in Data Set 2, which means an Estimation Rate of 86% and a False Alarm Rate of 32%.

Thresholds for risk definition were established according to a statistical analysis similar to the one detailed for Data Set 1. For Data Set 2, Q1, the 25% quartile of the estimated incident probability in an incident situation, was considered as the high-risk threshold value (0.033).

Results for Data Set 3: Model Calibrationand Threshold Tuning

It resulted a test period from 2001-03-01 18:40:00.000 to 2001-06-18 0:0:00.000. The data set consists of a total of 254 incidents, related to consistent traffic data. The process of filtering inconsistent data and randomly selecting non-incident situations data yielded 4005 registers, each one belonging to a section and timestamp, some of them reflecting 'incident conditions' and the rest 'incident free'. For each incident available, 6 minutes of traffic data from the same section were selected for calibration and testing purposes. No model selection process was included in this test, only recalibration of model parameters and threshold tuning.

After threshold tuning, 1211 false alarms raised in the 4005 registers and 1000 alarms were raised in an incident situation, identifying 210 of the 254 incidents; this is an Estimation Rate of 83% and a False Alarm Rate of 30%.

Thresholds for high-risk definition were established according to statistical analysis: high-risk threshold probability was fixed to the Q1 quartile of the expected probability of incident, predicted in a real incident situation.

Model selection was not undertaken with this data set: speed is included in the model as a covariate, and terms of order 1, 2 and 3 are necessary (terms of order 2 and 3 are included, centred with respect to mean speed). Occupancy is included in the model as a factor and volume is included as a covariate. All main effects and interaction are statistically significant.

The statistical model uses 16 degrees of freedom out of a total of 2243 (number of covariate classes), and conforms to Pearson's statistical test of goodness of fit included in the statistical package MINITAB (13.10) for Windows.

Results for Data Set 4: Threshold Tuning

It consists of a total of 23 incidents over the whole month of August; the consistency of the traffic data was found to get worse. The process of filtering inconsistent data and randomly selecting non-incident situations data yielded 482 registers; each one belongs to a section and timestamp, some of them reflecting 'incident conditions' and the rest 'incident free'. For each incident available, 6 minutes of traffic data from the same section were selected for testing purposes. Neither model selection process nor recalibration of model parameters were included in this test; only the effect of threshold tuning was considered.

The statistical results **after threshold tuning**, but using model parameters estimated in Data Set 1, show: 255 alarms raised over 482 time periods, 73 of them raised in an incident period situation and identifying 18 of 23 incidents, which means an Estimation Rate of 78%. False alarms raised during the test period were 182 out of a total number of 482 registers, leading to a False Alarm Rate of 38%.

As August is the holiday month for people working in Barcelona, traffic density decreases noticeably, and so the border between indicators leading to ‘high-risk’ situations and ‘non-high-risk’ situations are much more difficult to establish. The current results indicate that model calibration and selection should take seasonal variation of traffic conditions into consideration; however, for this to be done, data from a longer period of time would have to be considered in order to have good quality model parameter estimates, that is, data from August in the last 3-5 years would be required.

The threshold for the ‘high-risk’ definition was established according to statistical analysis: as the Q1 or 25% quartile of the expected probability of incident in a real incident situation (the same criteria as the one selected previously).

Summary of Results for the Barcelona Test Site

The results of the tests for each data set yielded the following measures of effectiveness (see Table 5):

EIP Validation Measure of Effectiveness	***Data Set 1***	***Data Set 2***	***Data Set 3***	***Data Set 4***	***Target***
Estimation Rate	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%	**80%**
False Alarm Rate	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%	**10%**

Table 5. Barcelona test site: EIP-HLOGIT performance

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. 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, as reported in Deliverable 4.1 of the PRIME Project (PRIME, 2001). 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 shown numerically in Table 6 and graphically in Figure 5.

Incident situation	**High-risk threshold for incident probability**	**FAR (%)**	**ER (%)**
No	0.23	24	92
Yes	0.31	14	75
No	0.4	7.3	63
Yes	0.75	0.8	26

Table 6. EIP-HLOGIT: FAR-ER trade-off according to high-risk threshold tuning

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.

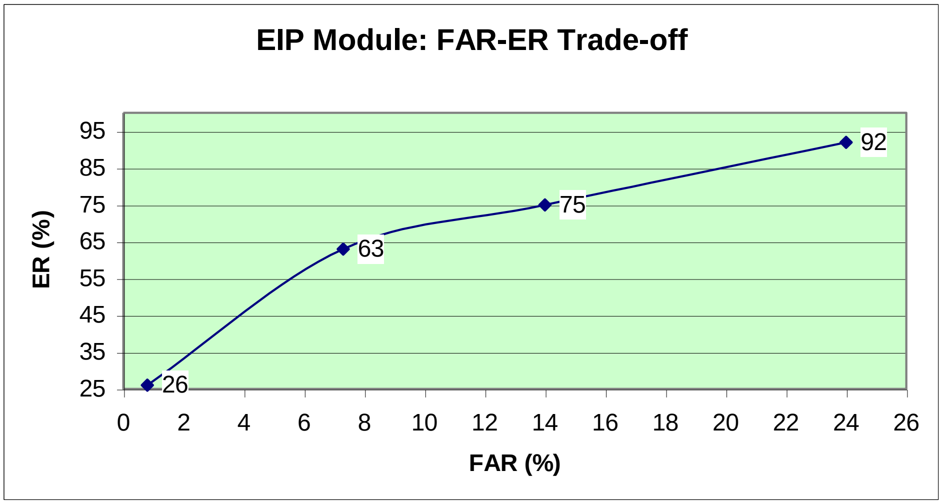


Figure 5. EIP-HLOGIT: FAR-ER trade-off for Data Set 3

IMPLEMENTATION AND GETRAM-AIMSUN INTEGRATION

The authors implemented and integrated this statistical model approach in the GETRAM-AIMSUN microscopic simulation environment, as a tool for the assessment of incident management strategies.

The GUI for EIP-HLOGIT manipulation and the EIP-HLOGIT module integrated into the GETRAM-AIMSUN environment with a GETRAM extension (a software application working within the GETRAM environment) is called AIMSUN2-KIT, and provides the user with the following functionalities:

- **Model definition and manipulation** (Load, Save, Change Model definition, etc.)
- **Calibration**. Computation of the values of model parameters from historical data, once an EIP-

HLOGIT model has been selected.

- **Estimation of dynamic probabilities of incidents** given dynamic weather and traffic conditions, road geometry and the statistical model of association (EIP-HLOGIT model).

- **Estimation of dynamic risks of incidents** given dynamic weather and traffic conditions, road geometry and the statistical model of association (EIP-HLOGIT model) and high-risk threshold definition.

The **Dynamic incident probability/risk estimation functionality** requires dynamic off-line traffic (speed, flow, occupancy, etc.) and weather data (rain, wind, fog and snow, etc.) and static geometric data (lanes, ramps, capacity, road/surface conditions, etc.). Off-line data could also be AIMSUN-simulated data loaded in what is called in the project the real-time database (RTDB). This possibility allows the testing by simulation of incident management strategies.

The AIMSUN2-KIT provides a graphical, user-friendly interface for operation with the Incident Detection module and EIP-HLOGIT module developed as part of the PRIME Project. From the point of view of network representation, the only specific requirement of the AIMSUN2-KIT is that the network geometry is represented using the TEDI set of graphic editors in the GETRAM environment (UPC, 1997). Below we present a brief description of how this KIT works.

Once a network has been selected in the AIMSUN2-KIT, and the EIP module starts its execution by activation in the PRIME menu, a default EIP-HLOGIT model is selected by loading a predefined model file; from here, the main PRIME Window allows the operator to zoom in and out and visualize incident risks/probabilities according to EIP-HLOGIT estimates, for each lane if input data is lane based (see Figure 6).

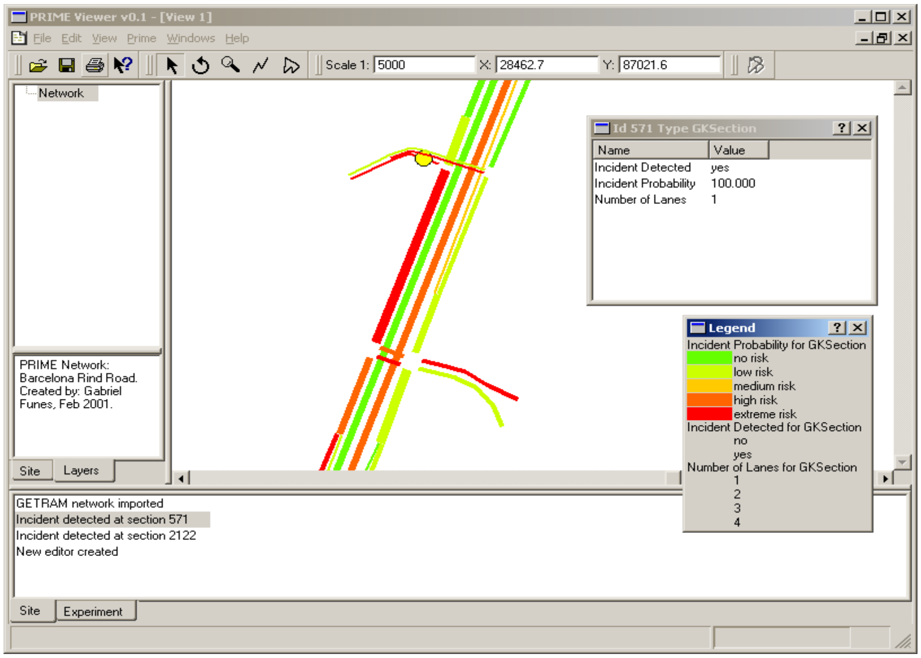


Figure 6. AIMSUN2-KIT: dynamic incident probability and risk estimation

Under the EIP/PRIME menu, there are asynchronous functions of Model Definition and Calibration, either variables or parameters or both (Calibration mode), Start-up and Stop module functions and Loading/Saving EIP-HLOGIT model variables and parameters (see Figure 7).

At any time, the operator might choose the option Definition and Calibration, which leads to the EIP Parameters window (see

Figure 8). The operator might change the description of the model, either variables or parameters. Once a variable has been included/eliminated from an EIP-HLOGIT model, the operator has to redefine the model parameters or select the calibration button in order for the EIP calibration process to compute the parameters according to currently available historical data.

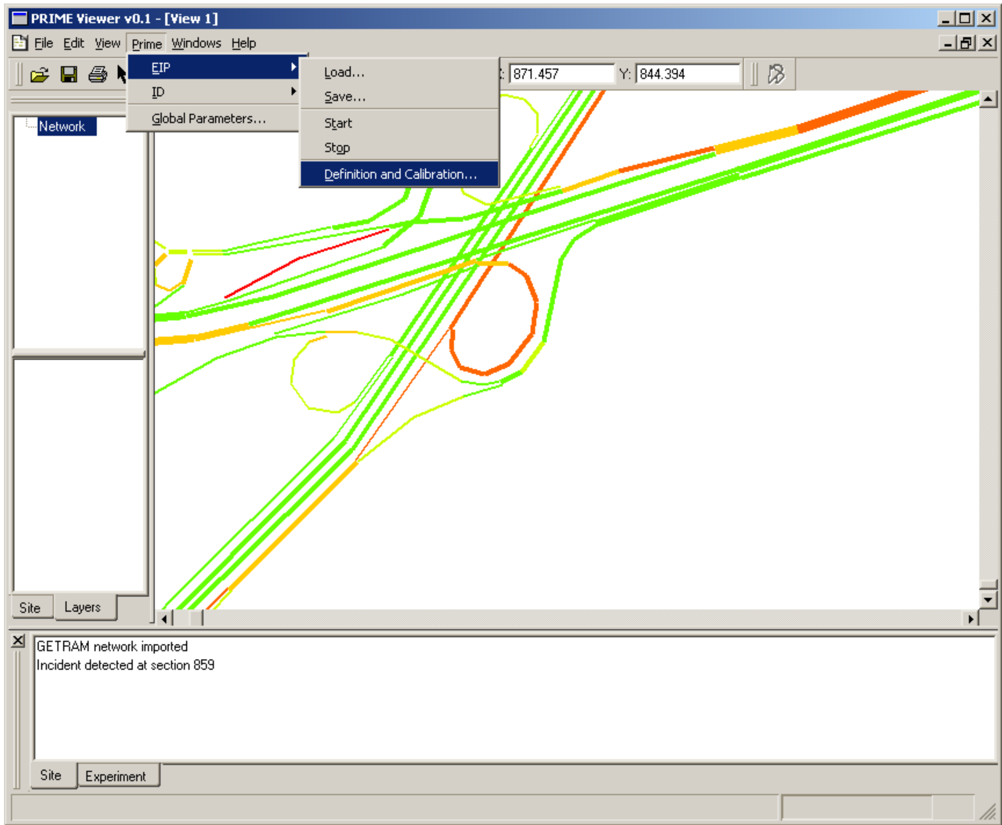
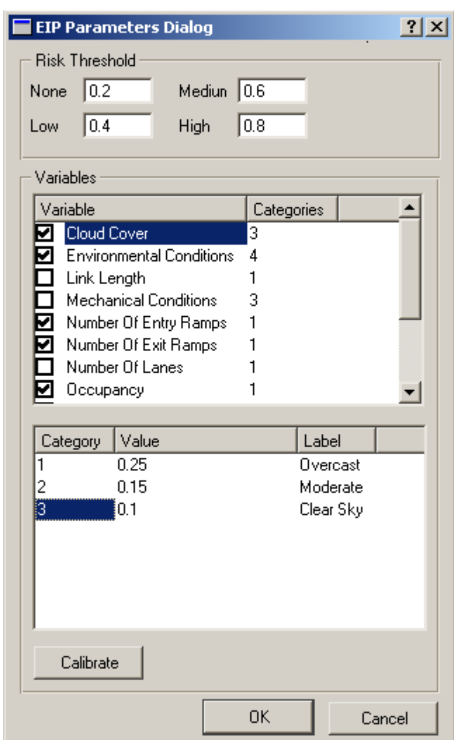


Figure 7. AIMSUN2-KIT: EIP-HLOGIT menu options



An EIP model is a set of variables plus a set of parameters for each variable in the model: for each continuous variable only one parameter has to be defined, but for each categorical variable in the model, a number of parameters (floating point values) equal to the number of categories have to be defined.

Figure 8. AIMSUN2-KIT: model definition and calibration for EIP-HLOGIT

Model variables and parameters can be loaded from a previous execution of the EIP module (Load option in the EIP-HLOGIT menu), calibrated for the EIP module (Calibrate option in

Figure 8) or input by the Operator (see

Figure 8, a situation in which the EIP model has been calibrated externally with a statistical package).The user must be aware of the fact that the underlying calibration process assumes a baseline type of re-parameterization and automatically assigns the last category as the baseline category (value of the parameter equals 0), otherwise an error indicating linear dependency in the columns of the design matrix will be given to the operator.

CONCLUSIONS

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 (Hamerslag et al., 1982, Miaou, 1994, TRAVELAID, 2001).

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, which we hope could be improved if better data, reliable data for longer time periods, and data for additional variables were available.

Reliable data collection for a longer period of time (around 3-5 years) would allow an incident probability/risk estimation **for incident type.** For example, we have been considering for tentative testing three types of incidents (tests not included in this paper): low serious damage (no injured persons), medium serious damage (with injured persons) and serious damage (with seriously injured persons or even deaths) situations.

One important problem to be considered is the huge amount of data that one year generates; statistical data mining techniques are suitable for processing such information. Seasonal variability of traffic behaviour directly affects seasonal variability in EIP-HLOGIT models. The authors believe that seasonal models should be developed to achieve the highest performance results in automatic real-time estimation of incident risks.

REFERENCES

Barceló, J., Montero, L., Perarnau, J., 2001. Automatic Detection and Estimation of Incident

Probabilities for Incident Management Purposes. A Case Study in Barcelona. The city at your fingertips: Proceedings of the 2nd ITS European Congress in Bilbao (Spain), 20-23 June 2001.

Barceló, J., Casas, J., Ferrer, J.L., García, D., 1999. Modeling Advanced Transport Telematic Applications with Microscopic Simulators: The Case of AIMSUN2, in Traffic and Mobility: Simulation, Economics, Environment. W. Brilon, F. Huber, M. Schreckenberg and H. Wallentowitz Editors, Springer Verlag.

Barceló, J., Ferrer, J.L., García, D., Florian, M., Le Saux, E., 1998. Parallelization of Microscopic Traffic Simulation for ATT Systems Analysis, in Equilibrium and Advanced Transportation Modeling, P. Marcotte and S. Nguyen Editors, Kluwer Academic Publishers pp. 1-26.

Barceló, J., 1991. Software Environments for Integrated RTI Simulation Systems. Advanced Telematics in Road Transport pp. 1095-1115, Proceedings of the DRIVE Conference, Brussels, Elsevier.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete choice analysis: theory and application to travel demand. MIT Press Cambridge.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. Chapman and Hall.

Dobson, A., 1990. An Introduction to Generalized Linear Models. Chapman and Hall.

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.

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.

IN-RESPONSE Project, 1997. Module Design: Task reports on Incident Detection and Incident Prediction. Deliverable 5.1, TR1030, EU-DG XIII.

King, G., Zeng, L. 2000. Logistic Regression in Rare Events Data. Harvard Burden of Disease Unit. Research Paper no. 2.

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.

PRIME Project (Prediction of Congestion and Incidents in Real Time for Intelligent Incident Management and Emergency Traffic Management), 2001. Deliverable 4.1. Project belonging to "*Information Societies Technology Program*" of the EU-DG XIII.

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.

UPC, 1997. Department of Statistics and Operational Research. LIOS. GETRAM Environment. Tedi: The GETRAM Editor User's Guide. AIMSUN2 User's Manual.

Wilmink, I., Tampère, C.M.J., Immers, L.H., Leeuwerke, A., Klunder, G., 2001. EIP Models. TNO-Description and Testing, Delft TNO Inro, November 2001.

Wilmink, I., Immers, B., Zwaneveld, P., Barceló, J. Montero, L., 1998. Incident Management combined with Traffic Management". ITS-1998 Seoul, 12-16 October 1998.

List of Figures:

FIGURE 1. ESTIMATION OF INCIDENT PROBABILITY MODULE IN A TRAFFIC MANAGEMENT SYSTEM ____ 5

FIGURE 2. THE BARCELONA RING ROAD AND THE PRIME SITE ____ 10

FIGURE 3. OBSERVED INCIDENT PROBABILITIES (LOGIT SCALE) VS. SPEED ACCORDING TO OCCUPANCY LEVELS ____ 18

FIGURE 4. ESTIMATED INCIDENT PROBABILITIES (LOGIT SCALE) VS. SPEED ACCORDING TO OCCUPANCY LEVELS ____ 19

FIGURE 5. EIP-HLOGIT: FAR-ER TRADE-OFF FOR DATA SET 3 ____ 26

FIGURE 6. AIMSUN2-KIT: DYNAMIC INCIDENT PROBABILITY AND RISK ESTIMATION ____ 28

FIGURE 7. AIMSUN2-KIT: EIP-HLOGIT MENU OPTIONS ____ 29

FIGURE 8. AIMSUN2-KIT: MODEL DEFINITION AND CALIBRATION FOR EIP-HLOGIT ____ 29

List of Tables:

TABLE 1. BARCELONA SITE: FILE FORMAT FOR TRAFFIC DATA ____ 11

TABLE 2. BARCELONA SITE: FILE FORMAT FOR INCIDENT DATA ____ 11

TABLE 3. BARCELONA TEST SITE: RESULTS FOR DATA SET 1 ____ 15

TABLE 4. BARCELONA TEST SITE: RISK THRESHOLDS FOR DATA SET 1 ____ 16

TABLE 5. BARCELONA TEST SITE: EIP-HLOGIT PERFORMANCE ____ 24

TABLE 6. EIP-HLOGIT: FAR-ER TRADE-OFF ACCORDING TO HIGH-RISK THRESHOLD TUNING ____ 25