SCIENTIFIC SUPPORT PLAN FOR A SUSTAINABLE DEVELOPMENT POLICY

Intermediary report - January 2003

INNOVATIVE SPATIAL TECHNIQUES FOR THE ANALYSIS OF TRAFFIC SAFETY
CP-34

KUL – LUC - UCL

SPSD II

PART 1

SUSTAINABLE PRODUCTION AND CONSUMPTION PATTERNS

- GENERAL ISSUES
- AGRO-FOOD
- ENERGY
- TRANSPORT

This research project is realised within the framework of the Scientific support plan for a sustainable development policy (SPSD II)

Part I "Sustainable production and consumption patterns"

The appendixes to this report are available at :
http://www.belspo.be (FEDRA)

Published in 2003 by the
Belgian Public Planning Service Science Policy
Rue de la Science 8 - Wetenschapsstraat
B-1000 Brussels
Belgium
Tel : 32/2/238.34.11 – Fax 32/2/230.59.12
http://www.belspo.be (FEDRA)

Contact person :
Mrs Aurore Delis (deli@belspo.be)
Tel : 02/238.37.61

Innovative Spatial Analysis Techniques for Traffic Safety

Intermediary scientific report 01/2003

UCL Unité de Géographie	LIMBURGS UNIVERSITAIR CENTRUM PARTNER IN DE UNIVERSITEIT LIMBURG	SADL

Colofon

Title: Intermediary report - Innovative Spatial Analysis Techniques for Traffic Safety

Author(s): T. Steenberghen, G. Wets, I. Thomas, N. Eckhardt, F. Casaer, J. Wijnant

Participating research units:

SADL/K.U.Leuven

Faculty of Applied Economics/Limburg Univeristy Centre

Département de Géographie/Université Catholique de Louvain

Version: January 2003

Printed: 2003-10-22

File location: PODO2veilig/voortgangsverslagen/voortgangsverslag122002/scientific report 12003.doc

•

TABLE OF CONTENTS

1 INNOVATIVE SPATIAL ANALYSIS TECHNIQUES FOR TRAFFIC SAFETY

2 INTRODUCTION

2.1 Context and summary

During the last decades mobility of goods and persons –particularly the share of road transport– has increased significantly. This was accompanied by a dramatic number of traffic accident victims (table 1). Compared to the European average, Belgium is confronted with a poor record in terms of traffic safety (table 2).

Table 1. Evolution of road transport an of traffic (un)safety.

Year	Victims (deceased + wounded)	Vehicle km (billion km)
1970	107.777	Approx 2.500.000
1980	84.700	3.753.745
1990	88.160	4.594.058
2000	69.431	5.735.034

Source: BIVV 2001

Table 2. Number of fatalities per 100.000 inhabitants in Belgium and Europe.

	1997	**1998**	**1999**
Belgium	13,4	14,7	13,7
European Union (15 countries)	11,6	11,3	11,1

Source: IRTAD - International Road Traffic and Accident Database (OESO)

This (un)safety is frequently blamed on poor spatial planning. In a previous research project (PADDI, "Impact of spatial planning on traffic safety", 1998-2000), the impact of the spatial characteristics on traffic safety was examined. Much effort was spent to develop adequate tools for proper location of accidents on both numbered roads (highways and major roads), and secondary roads. This research resulted in new methods for the identification of statistically sound black zones (Flahaut 2001, Flahaut et. Al. 2002, Flahaut and Thomas, 2002), and in a number of case studies where relations were identified between type of urbanization and traffic safety (Dufays and Steenberghen, 1999, 2000, 2002, Steenberghen et al. 2003).

2.2 Objectives

The main objective of this research is to improve the explanatory model for traffic safety, in order to clarify the interactions between safety factors. The previous research was based on GIS and spatial statistics. In order to reach the objective, this project explores the potential of new data sources and analysis techniques.

Three innovative approaches are used:

High resolution satellite imagery: (KUL). This part of the project is focused on the identification of better land use, infrastructure and traffic characteristics by means of remote sensing techniques applied on high resolution satellite imagery.

Spatial Data Mining: determination of significant combinations of causal factors (LUC). The accident data used for this research consists of all the road accidents in Belgium for a period of 9 years (1991-1999), with approximately 100 attributes for each accident. Knowledge Discovery and Data Mining are explored here as tools to detect structures, patterns and relations in this large data set.

Multi Level Analysis: (UCL). Explanatory factors of traffic (un)safety appear to be very scale-dependent. Rather than developing an explanatory model for different scales, recent developments through multilevel frameworks provide the opportunity of examining interactions at different levels and of integrating interactions between scale levels.

2.3 Expected outcomes

The project is organised in two phases: a first analysis using each technique (first iteration) and a second iteration, where the results of the findings from the other techniques are included for new analyses.

The expected outcomes at the end of the first year are:

KUL:

- database of localized environmental elements, road infrastructure and accidents for the period until 2002;
- evaluation of the potential of IKONOS (QUICKBIRD if available) images for identification of land use, infrastructure and traffic characteristics, which may be useful for traffic safety;

LUC:

- set of explanatory variables and their importance for traffic safety;
- initial model based on a data mining technique

UCL:

- set of explanatory variables and their importance at different scale levels
- initial model based on multilevel analysis.

The expected outcomes of the second year are:

KUL:

- new indicators for traffic safety characteristics of roads and environments, with higher explanatory value in the explanatory model

LUC:

- Spatial Data Mining method for the clustering of accidents
- Data Mining method for the stressing of distinctive features of black zones

UCL:

- Set of explanatory variables and their importance at different scale levels.

3 DETAILED DESCRIPTION OF THE SCIENTIFIC METHODOLOGY

3.1 KUL

3.1.1 Accident location

Raw accident data are provided by the National Institute of Statistics (NIS) in ASCII format. The regional road administrations in Flanders and the Walloon Region check and correct the location attributes (since the mid 1990's). The locational information in the database is either a combination of a street name and house number, or a road number and hectometre number, or, in case of a crossroad accident, two roads (by number or by name). The location of accidents can be done rather easily in GIS with different techniques. We used *address matching* and *dynamic segmentation* techniques developed in the previous research project (Steenberghen et al. 2000).

Accident location based on road and hectometre number

Dynamic segmentation is a GIS technique which uses a digital road network where every road has its identification number (road number) and a measuring system (hectometres). Based on this route system accidents can be located in GIS as a route event.

Accident location based on street name and house number

Address matching is another GIS technique that locates accidents based on the street name and house number. This technique was only applied in the Brussels region because of the availability of an accurate base map containing streets and buildings. This information allowed us to locate many accidents with precision.

Finally, crossroad accidents were located according to the intersection of the two streets.

We finished this part of the research with some post processing steps and checks of the accident locations. Given an accident located in municipality X we can check if this is the same municipality X that is mentioned in the accident database. Another kind of post processing is the grouping of accidents near crossroads on the crossroads themselves and the alignment of accidents on the street axes.

For Brussels we then produced maps with black zones using standard GIS kernel density functions (cf. 4. Results).

3.1.2 Geometric evaluation and correction of Ikonos imagery

An Ikonos [1]image of the south eastern part of Brussels was provided by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (OSTC). For this region, a very accurate base map – street and parcels – exists (URBIS v2, C.I.B.G.[2]). This map is based on orthophotos and has an very high accuracy[3]. Therefore, this area is suited as a test area for evaluation and correction of the satellite image. At first glance it was obvious that these two geographic layers (map and Ikonos image) did not match quite well as the streets did not overlap in all areas. This is no surprise because the Ikonos image we disposed of is of the "Geo"

[1] Ikonos (image of Brussels, 2002), Space Imaging Inc.
[2] Brussels Urbis Adm v2.0 (Brussels Urban Information System), C.I.B.G. Kunstlaan 20, Brussel
[3] Distinct objects like poles etc. have a positional error of less than 20 cm.

type, the cheapest Ikonos product with an absolute planimetric accuracy of 24 metres, terrain displacements (due to terrain height differences) not included. More precise Ikonos products exist but they are 5-10 more expensive than Geo. We took several approaches to rectify the image:

Simple polynomial rectification

A first attempt to rectify the image was to apply a polynomial transformation on the image. In Erdas Imagine[4], this is a standard procedure: one collects several Ground Control Points (GCP's) from both the image and the reference map and then the software transforms the image. The resulting corrected image still showed a lot of deviations from the URBIS reference map. This simple method does not correct for terrain height differences.

Orthorectification

In the area covered by the image the difference in altitude is more than 100 meters which leads to a planimetric displacement of up to 44 metres given the elevation of the satellite at the time of image capture. These important terrain displacements should be corrected. A so called "relief corrected affine transformation" has been successfully applied on several Ikonos images to correct errors due to elevation differences (Baltsavias, Pateraki, Zhang, 2001). With this technique one can easily correct an image for terrain displacements: knowing the satellite's position in space and every pixel's height we can reproject each pixel to a reference plane (at a constant height) and then again apply a simple polynomial transformation in order to produce the final image. For the study we used a DTM (grid spacing of 30 m) and collected precise coordinates and elevations of ground control points located at the centroids of roundabouts, recognizable in both the reference map and the image.

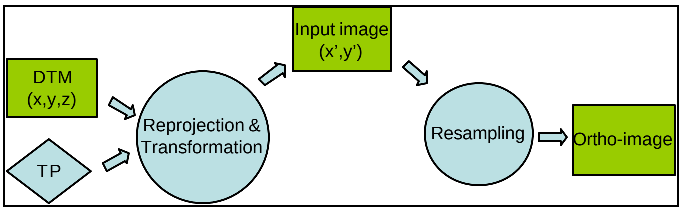

Figure 1: Orthorectification (TP=Transformation Parameters, base on reprojected GCPs)

Other methods

Unfortunately the geometric accuracy of the orthorectified image did not improve as much as expected (cfr. 4. Results). The error is probably due to the tilting of the sensor during image capture. These tilting movements along the three axes are responsible for complex anomalies. Other solutions will have to be searched for. We can try rubber sheeting with many GCP's or we can work with smaller subsets of the image.

If no geometric correction is applied then there won't be an exact and useful overlap with existing vector maps (like the Urbis street network). As a matter of fact this means an important loss of complementary information.

3.1.3 Fusion of image data

Image fusion is a way of integrating images of various spatial resolutions and it is mostly applied with a high resolution panchromatic band (in our case Ikonos Pan 1x1 metre) and several lower resolution multispectral bands (Ikonos MS 4x4 metre). The fused image is then a multispectral image with a higher spatial resolution (1x1 metre), called a pan-sharpened image.

[4] Erdas Imagine 8.5, Leica Geosystems

Adaptive Image Fusion or AIF (Steinocher, 1999) is specifically designed to handle high resolution images and to produce fused images with less unwanted detail and more homogeneous regions. This means that a subsequent multispectral classification will give better results (less “salt ‘n pepper” effect).

3.2 LUC

On the road to innovative spatial analyses techniques the LUC performed in the first place a research on the quality of the traffic accident data. A data dossier has been established containing descriptions concerning the database, its contents, the necessary structural adjustments and an investigation on the collection and registration quality.

Besides the attention for the data and the data quality, an exploratory data analysis was performed through the implementation of a data mining technique. The technique of association rules has been used to obtain a descriptive analysis of the accident data.

The methods and results are presented here as summaries of the produced articles.

3.2.1 Data Dossier & Data Quality

“A Research on the Quality of Belgian Traffic Accident Data” (Casaer et al.,2002)

The data dossier resulted predominantly in this article concerning a quality investigation. The article investigates the registration and the collection of Belgian accident data and aims at a contribution towards reliable applications on the basis of these data. It is addressed to all associated with this precious primary product (i.e. data analysts, police services, policy makers, …). It reports on the actual content, the accuracy, the completeness and the reliability of the data. They are examined through the use of cross tables displaying empty fields, missing values and inconsistencies.

The exploration regards the 1991-1999 data originating from the Flemish, Walloon and Brussels region and includes 505.880 registered traffic accidents with deceased or injured persons. The accidents are registered in the ‘Verkeersongeval Formulier voor ongevallen met doden en gewonden” form, which was brought into use on the 1st of July 1990 and which is considered to be a synthesis of the incident. It consists of 26 sections concerning the accident or the involved road user. Over the period 969.379 road users were registered.

The paper is organized in three sections. The first section aims at a verification of the different sections and fields of the accident form. It deals with the presence of empty fields, inaccuracies, missing values and under registration. The next section considers inconsistencies. Apart from numerical inconsistencies one is also confronted with inconsistencies within the various attribute values filled out. Section 3 concerns the reliability of the data input. The conclusions are presented in section 5.

3.2.2 Data mining technique : Association rules

A preliminary research (Geurts K., Wets G., Brijs T. and Vanhoof K. (2002), The Use of Rule Based Knowledge Discovery Techniques to Profile Black Spots. Paper presented at the the 6th Design and Decision Support Systems in Architecture and Urban Planning Conference, Ellecom, The Netherlands, July 7-10.) used the technique of association rules on the 1991-1996 data of Brussels. These data were already available at the beginning of the project. It was employed to recognize frequently occurring accident features. The features showed a variation which was correlated to the type of accident.

In the paper **"Profiling High Frequency Accident Locations Using Association Rules (Geurts et al. 2002)"**, association rules are used to identify accident circumstances that frequently occur together at high frequency accident locations. A comparative analysis between high frequency and low frequency accident locations is conducted to determine the discriminating character of the accident characteristics of black spots and black zones. In particular, the data mining technique of association rules is used to obtain a descriptive analysis of the accident data. In contrast with predictive models, the strength of this algorithm lies in the identification of relevant variables that make a strong contribution towards a better understanding of the circumstances in which the accidents have occurred. Hereby, the emphasis will lie on the interpretation of the results, which will be of high importance for improving traffic policies and ensuring traffic safety on the roads.

Technique : Association rules

Association rules is a data mining technique that can be used to efficiently search for interesting information in large amounts of data. More specifically, the association algorithm produces a set of rules describing underlying patterns in the data by means of the support parameter and the confidence parameter. Informally, the support of an association rule indicates how frequently that rule occurs in the data. The higher the support of the rule, the more prevalent the rule is. Confidence is a measure of the reliability of an association rule. The higher the confidence of the rule, the more confident we are that the rule really uncovers the underlying relationships in the data. It is obvious that we are especially interested in association rules that have a high support and a high confidence.

The concepts behind association rules and suggested algorithms for finding such rules were first introduced by Agrawal, Imielinski & Swami. The rule $X \Rightarrow Y$ holds in the transaction set *D* with confidence *c* if c% of transactions in *D* that contain *X* also contain *Y*. The rule $X \Rightarrow Y$ has support *s* in the transaction set *D* if s% of transactions in *D* contain $X \cup Y$. Generating association rules involves looking for so-called *frequent item sets* in the data. Indeed, the support of the rule $X \Rightarrow Y$ equals the frequency of the item set *{X, Y}*. Thus by looking for frequent item sets, we can determine the support of each rule.

Two properties of association rules can be used to distinguish trivial from non-trivial rules. A first, more formal method (9) to assess the dependence between the two item sets in the association rule is lift (L), a value that corrects the importance of each rule by taking the frequency of the attributes in the dataset into account. A second method to discern trivial from non-trivial rules is looking at the statistical rule significance. The statistical significance (T) of a rule is the validity of a rule, based on the influence of statistical dependency between the rule body and the rule head.

Data mining process

We distinguish three steps in the mining process: a preprocessing step in which the available data is prepared for the optimal use of the mining technique, a mining step for generating the association rules and a post-processing step for identifying the most interesting association rules.

To discern high frequency accident locations from low frequency accident locations, each accident needs to be linked with a location parameter that corresponds to a unique geographical location. Next, two different data sets were selected to explore association relationships between traffic accident attributes. Since our prime interest lies in the profiling and understanding of black spots and black zones, only the traffic accidents that occurred at a high frequency accident location were selected for the first analysis. To identify these locations, a criterion of minimum five accidents per location was used. This resulted in a total of 3.368 traffic accident records that were included in the first analysis. This number of accidents corresponds with the fact that in Flanders 15% of all the traffic accidents occur on so-called dangerous spots. The second association analysis is carried out on the remaining low frequency accident locations, including 30.985 accidents. By comparing the results from these two analyses, we can determine the discriminating character of the accident characteristics of high frequency accident locations. Furthermore, in the present data set, some attributes have a continuous character. Discretization of these continuous attributes is necessary, since generating association rules requires a data set for which all attributes are discrete. Finally, attributes with nominal values had to be transformed into attributes with binary attribute values.

A minimum support value of 5 percent was chosen for the analysis. This means that no item or set of items will be considered frequent for the first analysis if it does not appear in at least 165 traffic accidents. It could be argued that the choices for the values of these parameters are rather subjective. This is partially true, however a trial and error experiment indicated that setting the minimum support too low, leads to exponential growth of the number of items in the frequent item sets. Accordingly, the number of rules that will be generated will cause further research on these results to be impossible due to computer memory limitations. In contrast, by choosing a support parameter that is too high, the algorithm will only be capable of generating trivial rules. Analogously, the minimum confidence value was set at 30 percent. The algorithm obtained 187.829 frequent item sets of maximum size 4 for which 598.584 association rules could be generated. With the same parameters the second analysis resulted for the low frequency accident locations in 183.730 frequent item sets of maximum size 4 for which 575.974 association rules could be generated. These rules are further processed to select the most interesting rules.

The purpose of post-processing the association rules set is to identify the subset of interesting (i.e., non-trivial) rules in a generated set of association rules. Selecting the rules with a positive or a negative statistical significance from the association rules set narrowed down the results to 14.690 association rules for the high frequency accident locations and 77.282 association rules for the low frequency accident locations. We use the interestingness measure to limit the association rules to only the discriminating or useful ones. The interestingness measure is based on the deviation of the characteristic rules discovered for the accidents that occurred on high frequency accident locations from the accidents that occurred on low frequency accident locations.

3.3 UCL

The project "Innovative techniques of spatial analysis in road safety", brought us to realize a preliminary exploratory analysis in the development of an explanatory multilevel model. This analysis is submitted for publication in the journal " Recherche Transports Sécurité " and is entitled **Space-time variation of the road accidents in periurban environment. The example of Brussels. (ECKHART, FLAHAUT and THOMAS).**

The objective of this paper is to analyse the space-time variation in black zones within a given environment. The identification of the black zones for road accidents is based on local indices of spatial autocorrelation; the utility of this method was demonstrated in a previous SSTC contract and in the subsequent publications (Flahaut, 1999; Flahaut et al., 2002; Fhahaut and Thomas, 2002). This method presents the advantage of taking into account the nature of the space (measures of nearness), to be statistically based and to provide not only an indication of the dangerousness, but also the length of the dangerous segments. The black zone is less volatile and less mobile than the "black point" (see for example Silcock et Smyth, 1985; Nguyen, 1991; Joly *et al*., 1992; Hauer, 1996 et Vandersmissen *et al*., 1996).

If the operationality of the method has already been largely demonstrated, it is important to test if black zones have a strong and stable spatial structure in time. Indeed, before building an explanatory model of the concentrations of accidents (final objective of our project), we here propose a preliminary exploratory analysis which aims at testing the following hypothesis: the geographic distribution of black zones varies according to the chosen period of time. If this hypothesis is rejected, we can think that it is very likely that a strong environmental and spatial link exists, that the accidents often concentrate in the same places and that the results of the explanatory statistical model will be independent of the chosen period of time. If it is accepted, this means that space does not matter much in the explanation of road accidents; moreover, this also then means that the choice of the year for the data influence the explanatory results.

Our analysis is limited to one Belgian province: the Walloon Brabant, which is mainly periurban. The smallest spatial unity for which data on road accidents are available on numbered roads is the hectometer of road. Y_i is the studied variable; it takes the value 1 when hectometer i belongs to a black zone, and 0 otherwise. It is measured for several periods of time (1991-1993, 1994-1996, 1997-1999). The results obtained for each period of time are mapped and compared by means of several indices and statistical measures of similarity (see for example Gatrell, 1983; Kent and Coker, 1996; Legendre and Legendre, 1998). The chosen indices measure the statistical similarity between periods, by taking into account the disproportion inferred by the high number of hectometres not belonging to the black zones. Similarities are first computed for all black zones. In a second set of tests, black zones are divided into several subsets defined on their physical and environmental characteristics (e.g. type of road (highways and ring roads / national roads), volume of traffic, physical characteristics of the road and its environment (e.g. urban, wooded, opened rural environment).

The paper is organized in four sections. The data base is defined in section 2 and the hypotheses and selected criteria are justified in section 3. Section 3 also reminds the theoretical bases of analyses of accidents, notably the calculation of indices of spatial autocorrelation and the notion of black zone. The results are presented in section 4 and are followed by the conclusions in Section 5.

4 DETAILED DESCRIPTION OF THE INTERMEDIARY RESULTS, PRELIMINARY CONCLUSIONS AND RECOMMENDATIONS

4.1 KUL

4.1.1 Accident location

In the regions of Flanders and Wallony about 80% of all accidents on numbered roads are located. A major factor for non-located accidents is the fact that the hectometre value is missing. The graph in Figure 2 shows that the relative number of located accidents on crossroads is higher, probably because it is easier to locate a crossroad accident in the sense that there is no need for the hectometre value which is missing in many cases. Chronologically there is an upwards trend which can be attributed to the increased corrections and checks of the database by the regional administrations.

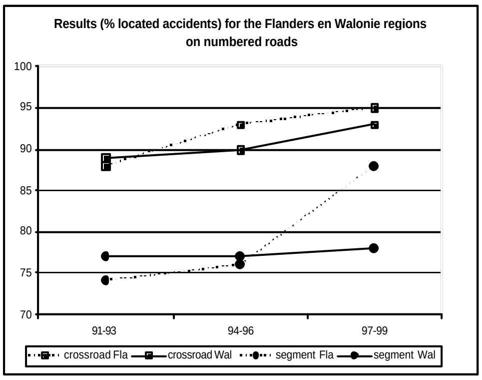

Figure 2: accident location results

Figure 3: black zones in the Brussels Region, period 1997-1999 (>45 accidents/km²)

In Brussels, 60% of the accidents were located. The lower percentage is explained by the location by means of address matching, while addresses tend not to be filled in as correctly as hectometre and road number (cf. 4.2.1). Figure 3 depicts a density map of the accidents in Brussels.

4.1.2 Geometric evaluation and correction

As mentioned before, the orthorectification method did not produce the expected results. Table 1 lists the differences between the predicted and true point locations for all GCPs and unfortunately the errors are quite high, up to 12,9 meters. From this information it is evident that there must be some other major source of error than the elevation differences, probably the satellite's tilting movements. These movements require specific software and knowledge to correct for. Figure 4 shows that the errors tend to cluster in the NW zone.

NR	Res.X	Res.Y	RMS
1	-0,4	-0,3	0,5
2	-2,0	-1,9	2,8
3	0,0	-0,5	0,5
4	**-2,1**	**-5,6**	**5,9**
5	1,0	0,6	1,1
6	**6,4**	**11,1**	**12,9**
7	-1,6	-1,7	2,4
8	0,0	-2,5	2,5
9	1,3	-0,9	1,5
10	**1,2**	**5,8**	**5,9**
11	**-2,4**	**-6,0**	**6,4**
12	**-1,5**	**-5,0**	**5,3**
13	-0,1	0,0	0,1
14	**0,3**	**5,5**	**5,5**
15	-1,6	-1,1	1,9
16	1,8	2,4	3,0

Table 1: residuals and RMS errors (meters)

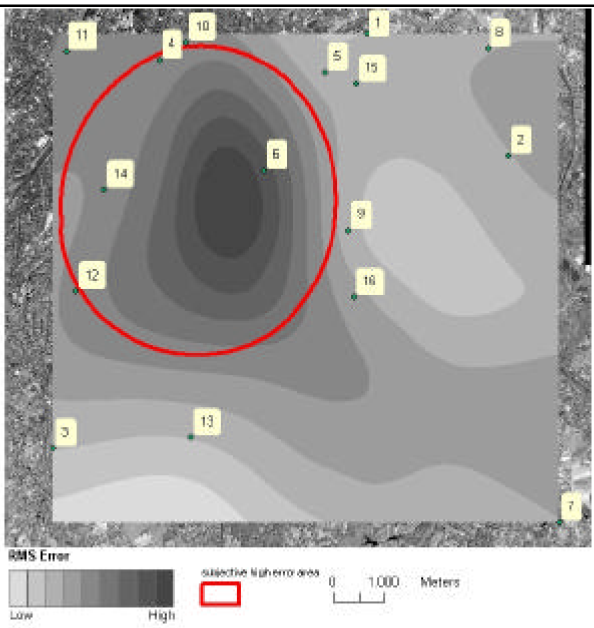

Figure 4: interpolated RMS errors (grey) and subjectively large error area (red ellipse)

4.1.3 Fusion of image data

Figure 5 is an example of the results from two different image fusion methods. The right image is fused in Erdas with a standard principal components algorithm. It is very suited for visual interpretation as it contains a lot of spatial detail. The left image, on the other hand, which is fused with AIF contains less detail but more homogenous regions with very distinct borders. Such an image is a much better input for multispectral classification, which may help identify zones with homogeneous characteristics important for traffic safety. On the other hand, small objects like individual cars have vanished into the street surface.

Figure 5: AIF fusion image (left) and Erdas (PCA) fused image (right

4.1.5 Ikonos' usefullnes for traffic information

There are difficulties in the use of Ikonos imagery for collecting traffic related information. The initial goals of the remote sensing research part were focussed on this dynamic traffic issue: identifying vehicles, calculating moving vehicle densities, parked vehicle densities etc. Although an interesting technique for the automatic detection of vehicles on Ikonos images was found in the literature (Dial et al. 2001), a major problem when working with Ikonos satellite images for traffic analysis, is the timing of the image-capturing in the study area (always around 11.00 am). Besides this important constraint, the technique needs a very precise geometric positioning which is very difficult to achieve with Ikonos as has been demonstrated. Multispectral classification of vehicles is not feasible because the spatial resolution of the spectral bands is too low (4 x 4 m) to recognize vehicles. Also fusion with the sharp PAN image does not yield better information because fast moving vehicles do not show up at exactly the same spot in the PAN and multispectral images probably due to a time delay between the capturing of the images. Moreover in a dense urban environment many vehicles are obscured by shadows or by buildings.

Our preliminary conclusions are that these techniques are not adequate to analyse traffic in urban environments.

4.2 LUC

4.2.1 Data Quality Investigation

Due to new legal and administrative procedures, the NIS could not provide accident data for the years after 1999. The update and corrections to the accident database were therefore conducted for the period 1991-1999. This not only consisted of adding the recent years. For the period 1991-1996 all the accident attribute data were added to the database, as opposed to the previous version, where only location and a limited number of driver and vehicle data were included.

This accident database holds millions of data fields containing accident factor registrations. However many fields remain empty, either because the specific section is not applicable or because certain relevant information - due to inaccuracy – was not properly stored. The present data collection does not make any distinction between both and fails to mention if a certain section or factor is not applicable. So it gets hard finding out the origin of the emptiness in a subsequent phase. This non-applicability occurs frequently for the sections which describe and explain the accident itself: course of the accident, accident factors.

Some of the attributes or sections which should be applicable to all accidents, have a striking lack of registrations. The next attributes or sections exhibit among others a significant percentage of missing values: cycle track milestone & house number, visibility of the pedestrian, dynamics, motion, passenger place.

The most important remark concerning accuracy concerns the location data: the sections which should contain either house number & street name, or milestone & road number. A digital road network is used to map the accidents. However, only Brussels is covered with a digital network able to deal with house numbers and street names. In Flanders and Wallonia a computerized localisation is solely feasible for the accidents on numbered roads (i.e. 57%). The location data of a significant part of these accidents is insufficient and makes the automated localisation impossible for 20% of the accidents on numbered road sections or cross roads. Both regions show better localisation percentages on the cross roads in comparison with the road segments between cross roads. Frequently missing data are the milestone data which are a necessary for the localisation of the accidents by means of dynamic segmentation.

The 1997-1999 period has a higher percentage of localised accidents on the Flemish road segments in comparison with the situation in Wallonia, nevertheless the Walloon region makes adequate corrections. Possibly some Walloon accidents have been wrongfully eliminated in the post processing checks.

In Brussels, the location technique by address is used. This type of localisation data appears to be more sensitive to erroneous registration. Only 60% of the 27.000 accidents were located.

The amount of registered accidents is smaller than the actual amount of traffic accidents. Accidents with exclusively material damage do not appear in the registrations. Although data of such accidents are very useful for analysis, they are not available. Some communities register those data, but they are not gathered at a national level. Furthermore not for every accident with deceased or injured a registration form is filled out, especially in case of light injuries. This under-registration can mount – up to 47% for the lightly injured – as a previous study (Beaucourt L.,1998) pointed out.

If inconsistencies show up in the registered amounts of the sections, one often has to conclude the presence of missing value. For about a thousand involved pedestrians there is no pedestrian section at our disposal and the bicycle section is missing for thousands of involved cyclists. Apart from these numerical inconsistencies one is also confronted with inconsistencies within the various attribute values filled in. Such inconsistencies appear for instance at the combination of sections motion and dynamics – where the vehicle can stand still and move at the same time – or between local characteristics and traffic regulation where roundabouts with traffic lights are registered. Although this last type of inconsistencies are rather limited, it is important to realize that they appear for they might be avoided in the future when using more automated registration. It is possible to be confronted with attribute values which were not postulated. This also results in missing values, e.g. section dynamics.

Last but not least the reliability of the data filled out in the various sections is evaluated. The remarks are based on our own data research and the inquiry of experts. It appears to be important to make a distinction between two kinds of sections: fixed element in the official report of the offence, and information collected from driver declaration or from conclusions at the moment of the ascertainment (the form often gets filled out at a later moment). The latter are obviously less reliable. Concerning specific reliabilities it is advisable to be cautious using the very subjectively measured value for the fields concerning 'personal consequences'. Other fields like crossing distance, speed limit, and alcohol test are also badly registered.

Conclusion

Traffic safety depends on a whole complex of interdependent features. A great deal of these features are registered and collected. In order to use these data it is crucial to examine the registration and collection quality. The actual content, the accuracy, the completeness and the reliability of the data have been evaluated. Different remarks have been made on these aspects, generally the traffic data of our country and her three regions can be considered interesting and useful to some extent. The relative high measure of emptiness within the explanatory sections, the under-registration, the inaccurate location data and the partly unreliable accident consequences take down the quality of the data. Apart from the higher percentage of localised accidents on the Flemish road segments, the data collection in the three regions is approximately of comparable quality.

The article concludes with hypothetical explanations for the sometimes reduced data quality. A first important hypothesis deals with a possible variation in local priorities concerning traffic safety which could influence the registration accuracy and completeness. Secondly, a reconstruction and localisation implemented in a later phase also can lead to inaccuracy. A more computerised and integrated data collection would not only decrease inconsistencies and missing values but would also make a higher form flexibility possible. Furthermore we point out the possible lack of road number maps. Often the accident localisation is still done with the available information on the spot. We finish by remarking the practice of data transformations (e.g. fom ASCII to Access) which also can result in a loss of quality since the NIS format is often difficult to interpret

4.2.2 High & Low Frequency Rules & Results

Selecting the association rules that appear in both the high frequency accident rules set and the low frequency accident rules set results in 3.670 statistically significant association rules. These can be further post-processed by means of the interestingness measure.

When ranking the association rules on their interest value we noticed for the 50 most discriminating rules, that a high interestingness value does not inseparably correlate with a strong lift value. The results show that accident characteristics that have the most discriminating power to identify black spots and black zones are not necessarily the most interesting rules according to their lift values.

Accordingly, when ranking the association rules on the lift value the results indicate that although the association rules identify frequently occurring patterns that are descriptive for the occurrence of accidents on high frequency accident locations, they are not necessarily discriminating between the profile of high frequency accident locations and low frequency accident locations.

When looking at the interpretation of the association rules, the rules with the 10 highest interestingness values were mostly related to geographical characteristics of the accidents. This means that the most discriminating characteristics between high frequency accident locations and low frequency accident locations are related to infrastructure or location features.

In total 46,4 percent of all accidents that occur at high frequency accident locations take place on a roadway with separated lanes outside the inner city. In comparison with the low frequency accident locations, where only 12,34 percent of the accidents can be attributed to this sort of location, these roadways with separated lanes outside the inner city are very characteristic for the occurrence of black spots and black zones. Although these results seem quite reasonable, they do indicate that roadways with separated lanes outside the inner city are an important problem for traffic safety. Therefore, further research on the cause of the unsafe character of these roads will be necessary. A high traffic intensity would seem the logical explanation for the high number of accidents that occur on these roads, but another possible explanation could also be the infrastructure of these roads. Depending on the results, government could consider restructuring these roads or changing their traffic regulation.

Furthermore, when an accident takes place on a roadway with separated lanes, the road user will more frequently than expected be of the age 30 to 45. Finally, a characteristic pattern for high frequency accident locations is the involvement of at least one passenger car among the road users when the accident occurs on a roadway with separated lanes. Accordingly to the previous results, this kind of accident accounts for 47,86 percent of all accidents on high frequency accident locations, whereas the same accident circumstances only occur in 15,02 percent of the low frequency accident locations.

When looking at the results of a final remark concerning the interest value of the rules should be made. The support values for the patterns in the low accident locations remain quite stable. However, the support values for the rules concerning the high frequency accident locations can be considerably high for some rules and considerably smaller for other. Therefore, the increase in the interest value of the rules and consequently the increase in the discriminating character of the rules are mainly related to the strong occurrence of the accident circumstances in the high frequency accident locations and not as much by the weak occurrence of these patterns in the low frequency accident locations.

It is possible, in the most pessimistic situation that these patterns do exist for the low frequency accident locations but they do not appear in the association rules set due to the value of the minimum support parameter (5 percent). Moreover, the choice of the minimum confidence parameter (30 percent) could inhibit the algorithm from generating a number of rules although the frequent item sets exceed the minimum support parameter.

Conclusion

The analysis showed that by generating association rules the identification of accident circumstances that frequently occur together is facilitated. This leads to a strong contribution towards a better understanding of the occurrence of traffic accidents. However, association rules do describe the co-occurrence of accident circumstances but they do not give any explanation about the causality of these accident patterns. Therefore, their role is to give direction to more profound research since the use of some additional techniques or expert knowledge will be required to identify the most important causes of these accident patterns. Furthermore, the results indicate that the use of the association algorithm not only allows to give a descriptive analysis of accident patterns on high frequency accident locations, it also creates the possibility to find the accident characteristics that are discriminating between high and low frequency accident locations. The most important results indicate that although human and behavioural characteristics play an important role in the occurrence of all traffic accidents, the main difference in accident patterns between high and low frequency accident locations can be found in infrastructure or location related circumstances. In conclusion, this analysis shows that a special traffic policy towards black spots and black zones should be considered, since these high frequency accident locations are characterized by specific accident circumstances, which require different measures to improve the traffic safety.

Several issues remain for future research. First, the skewed character of the accident data limits the amount of information contained in the dataset and will therefore restrict the number of circumstances that will appear in the results. Moreover, the choice for the minimum support and minimum confidence parameter can prevent the association algorithm from generating rules on the less frequent accident conditions. Secondly, the inclusion of domain knowledge (e.g. traffic intensities, a priori infrastructure distributions) in the association algorithm would improve the mining capability of this data mining technique and would facilitate the post-processing of the association rules set to discover the most interesting accident patterns. Finally, considering the large number of attributes in the traffic accident dataset, it seems interesting to explore the potential of techniques that generate rules with long patterns to uncover more complex associations in traffic accidents.

4.3 UCL

By observation of the maps realized for each period of time, we notice a rather high stability of the black zones on the motorways and more instability - explainable or not - on the main roads. The most stable concentrations of black zones are located on traffic axes going to Brussels (radial axes), whereas the transverse axes are less characterized by concentrations of accidents (less traffic). Environmental characteristics explain part of the stability.

Comparing maps is a clear objective on its own, but not that easy to perform statistically, especially in this case. The classic chi-square test is here indeed difficult to apply and interpret because of the high number of hectometres not belonging to any black zone. Methodological difficulties bound to the space-time comparison of maps brought us to choose **several indicators** of statistical similarity (see for example Gatrell, 1983; Kent and Coker, 1996; Legendre and Legendre, 1998).

Globally, the average rate of stability is 40 % for all black zones between two consecutive periods. If we subdivide the roads according to environmental factors, the highest stability (between 50 and 60 %) corresponds to highways and ring roads (94-96/97-99), to traffic measures greater than 17 300 vehicles (94-96/97-99), to roads with 2x2 separated lanes (94-96/97-99) and this especially if the traffic is greater than 24 000 vehicles, to roads with not separated lanes and with a traffic greater than 9400 vehicles (91-93/94-96), to dense urban environments (91-93/94-96), and to opened rural environments.

These indices support the tendency observed on maps: the important road axes going to Brussels constitute a strong spatial structure for the two last periods (94-96/97-99); the similarity - and thus the stability – scores higher in dense urban environments as well as on small roads for periods 91-93/94-96. The road safety measures realized during the last years seem to have improved the road safety for these conditions, but progress still remains to be done for the road infrastructures for freeways.

In fact,10 % of all the studied hectometres of the numbered roads of the Walloon Brabant belong to a black zone one period of time. They form a “hard core” that seems almost impossible to treat. Another part is more mobile, like black points (Silcock and Smyth, 1985; Nguyen, 1991; Joly *and al.*, 1992; Hauer, 1996 and Vandersmissen *and al.*, 1996). It is important to remember that the volume of traffic plays an important role: on the average, when traffic increases, the number of accidents also increases; this increase is however limited by the capacity of the road (Thomas, 1992).

This good stability of the black zones in time and space consolidates the idea that the choice of the year will not affect the results of the planned geographic analyses. We can now analyse space and explain the concentrations of road accidents without fear of a bad choice of the period of time. So, the results of this preliminary analysis allow us to continue our project and to build an explanatory model of the concentration of the accidents in Walloon Brabant. Hence, we reject the null hypothesis and state that the geographic distribution of the black zones does not vary much according to the chosen period of time. We can think that it is very likely that a strong environmental and spatial link exists and that the accidents tend to concentrate in the same places.

5 FUTURE PROSPECTS AND PLANNING

5.1 KUL

Two tracks will be further examined during the second year of the project. First, the urban structure of the black zones identified in Brussels will be further examined. The research question is: is it possible to find characteristics of these black zones which are different from the rest of the region? The physical characteristics determined from the satellite images will not be examined by themselves. LUC will examine if accidents in each black zone have specific characteristics that are different from the rest of the accidents. KUL will examine in how far the black zones have specific spatial characteristics. The identification of relevant scales of black zones will be based on the preliminary findings from the UCL.

The analysis of traffic characteristics based on RS will not be continued due to the incompatibility of the dynamic traffic and the static satellite imagery.

5.2 LUC

In the near future we still will be working with the current database and its accident attributes. It will be used for the clarification of accident concentrations. In Brabant Walloon and in the Region of Brussels zones with high accident concentrations have been established for the period 1997-1999. In the first region an autocorrelation technique was used to determine the different zones and the data mining technique ‘Association Rules’ will be used to profile these high concentration zones. In the second region a 2D-analyses was used to produce the zones and a decision tree algorithm will be utilized for the profiling. The algorithms will have to use the attributes that have been registered on the VOF-form. Profiling these zones will help in pointing out possible suggestions for the creation of new spatial attributes.

In addition to this profiling it’s necessary to develop another possible building stone. In order to apply a divide and conquer strategy we should try to perform a clustering of the 500.000 accidents and group them in several more similar partitions. So the clustering will be used as an explorative classification method. Due to the nature of the variables – mostly nominal or categorical – and due to the huge amount of accidents a specific clustering technique is needed.

5.3 UCL

This analysis is a first step before pursuing the explanatory analysis of the spatial structure of the road accidents with multilevel techniques.

Multilevel analysis enables one to investigate the effect of various spatial scales of aggregation of the data. Up to now, our effort is limited to a review of the literature (Goldstein 1995; Hox, 1995; Courgeau and Baccaïni, 1997; Snijders and Boskers 1999; Mathian and Piron, 2000). Applications to accidents are seldom found (Jones and JØrgensen, 2003). Applied to road accidents location, this method will help us to measure the various effect of space on accident location. We will analyse the characteristics of the accidents at a local level, while taking into account differentiations bound to the environments or the contexts in which they evolve. For road accidents, data are available at the level of the accidents and the implied users.

The various levels of analysis should be defined as well as variables attributed to every level to answer the fitting of scales of analysis. Multilevel modelling will be applied to significant variables defined by both other research teams (LUC according to Data Mining and SADL according to the techniques of remote sensing) by means of other techniques.

6 ANNEXES

6.1 References

Remote Sensing

BALTSAVIAS, PATERAKI, ZHANG (2001), "Radiometric and geometric evaluation of Ikonos Geo images and their use for 3D building modelling", *Proc. Joint ISPRS Workshop "High Resolution Mapping from Space 2001", Hannover*

BELGIAN INSTITUTE FOR TRAFFIC SAFETY (BIVV) and National Institute for Statistics (2001), Year Report on Traffic Safety 2001, BIVV v.z.w., Brussels. 115pp.

DIAL, GIBSON, POULSEN (2001) "Ikonos satellite imagery and its use in automated road extraction" *in: Baltsavias, Gruen, Van Gool (Eds.), Automated extraction of mand made objects from Aerial and space images. Balkema publishers, Lisse pp. 357-367*

FRASER, BALTSAVIAS, GRUEN (2002) "Processing Ikonos imagery for submetre 3D positioning and building extraction", *Journal of Photogrammetry & Remote Sensing pp.177-194*

MERRY,MCCORD, GOEL, COIFMAN (2001) "Annual Report - Project 00-01 – Satellite-based Data: Proof of Concepts and Operational Issues", National Consortium on Remote Sensing in Transportation, The Ohio State University

STEINNOCHER, K. (1999) "Adaptive image fusion of multisource raster data applying filter techniques", *International archives of Photogrammetry & Remote Sensing, Vol. 32*

Spatial Data Mining

AGRAWAL, R., IMIELINSKI, T. AND A. SWAMI (1993) Mining association rules between sets of items in large databases, *Proceedings of ACM SIGMOD Conference on Management of Data*, Washington D.C., USA, May 26-28, 1993, pp. 207-216.

AGRAWAL, R., MANNILA, H., SRIKANT, R. ET AL. (1996) Fast discovery of association rules, in Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. et al (eds.), *Advances in Knowledge Discovery and Data Mining.* AAAI Press, Menlo Park, California, USA, pp. 307-328.

BELGIAN INSTITUTE FOR TRAFFIC SAFETY (BIVV) and National Institute for Statistics (2000), Year Report on Traffic Safety 2000, BIVV v.z.w., Brussels.

BRIN, S., MOTWANI, R AND C. SILVERSTEIN (1997) Beyond market baskets: generalizing association rules to correlations, *Proceedings of the ACM SIGMOD Conference on Management of Data*, Tucson, Arizona, USA, May 13-15, 1997, pp. 265-276.

CHEN, W. AND P. JOVANIS (2002) Method for identifying factors contributing to driver-injury severity in traffic crashes, *Transportation Research Record* **1717**, pp. 1-9.

DAVID GIBSON, JON KLEINBERG AND PRABHAKAR RAGHAVAN (1998) Clustering Categorical Data: An Approach Based on Dynamical Systems in *VLDB Journal: Very Large Data Bases* (1988).

FRAWLEY, W., PIATETSKY-SHAPIRO, G., AND C. MATHEUS (1991) Knowledge discovery in databases: an overview, in Frawley, W. and G. Piatetsky-Shapiro (eds.), *Knowledge Discovery in Databases*. AAAI Press/ MIT Press, Menlo Park, California, USA, pp. 1-27.

Geurts K., Wets G., Brijs T. and Vanhoof K. (2002), The Use of Rule Based Knowledge Discovery Techniques to Profile Black Spots. Paper presented at the the 6th Design and Decision Support Systems in Architecture and Urban Planning Conference, Ellecom, The Netherlands, July 7-10

GEURTS, K. , WETS, G., BRIJS, T. AND VANHOOF, K. (2003). Profling High Frequency Accident Locations using Association Rules. Electronic Proceedings of the 82th Annual Meeting of the Transportation Research Board, Washington, January 12-16, USA, 18 pp.

GIRAUD-CARRIER, C. AND MARTINEZ, T.R. (1995) An efficient metric for heterogeneous inductive learning applications in the attribute-value language, . In *Intelligent Systems (Proceedings of GWIC'94)*, pages 341--350. Kluwer Academic Publishers, January 1995.

HOLTE, R.C. (1993) Very simple classification rules perform well on most commonly used datasets, *Machine Learning* **11**, pp. 63-90.

ICKJAI LEE (2002) Multi-Purpose Boundary-Based Clustering on Proximity Graphs for Geografical Data Mining" *The university of Newcastle Callaghan Australia* Februari 2002.

KONONOV, J. AND JANSON B. (2002) Diagnostic Methodology for the detection of safety problems at intersections, *Proceedings of the Transportation Research Board*, Washington D.C., USA, January 13-17, 2002, Paper No. 02-2148.

LEE, C., SACCOMANNO, F. AND B. HELLINGA (2002) Analysis of Crash Precursors on Instrumented Freeways, *Proceedings of the Transportation Research Board*, Washington D.C., USA, January 13-17, 2002, Paper No. 02-3790.

MANNILA, H (1997) Methods and problems in data mining, *Proceedings of the International Conference on Database Theory*, Delphi, Greece, January 8-10, 1997, pp. 41-45.

RAYMOND T. NG AND JAWEI HAN (2002) Clarans : a Method for Clustering Objects for Spatial Data Mining in *IEEE transactions on Knowledge and Data Engineering* 2002,Volume 14, nr 5.

SILVERSTEIN, C., BRIN, S. AND R. MOTWANI (1998) Beyond market baskets: generalizing association rules to dependence rules, *Data Mining and Knowledge Discovery* **2(1)**, pp. 39-68.

VENKATESH GANTI, JOHANNES GEHRKE AND RAGHU RAMAKRISHAN (1999)CACTUS-Clustering Categorical Data Using Summaries, in *Knowledge Discovery and Data Mining* (KDD 1999).

WILSON, D.R. AND MARTINEZ, T.R. (1997) Improved Heterogeneous Distance Functions, *Journal of Artificial Intelligence Research*, 6 (1997) 1-34. Submitted 5/96; published 1/97.

Multilevel analysis

COURGEAU D., BACCAÏNI B. (1997), Analyse multiniveau en sciences sociales. Population, n°4, 831-864.

FLAHAUT B., MOUCHART M., SAN MARTIN E., THOMAS I. (2002), Identifying black zones with a local spatial autocorrelation index and a kernel method. A comparative approach, *Accident Analysis and Prevention*, (sous presse).

FLAHAUT B., THOMAS I. (2002), Identifier les zones noires d'un réseau routier par l'autocorrélation spatiale locale. Analyses de sensibilité et aspects opérationnels, *Revue internationale de géomatique*, vol. 12, n°2, p. 245-261.

FLAHAUT B. (2001), L'autocorrélation spatiale comme outil géostatistique d'identification des concentrations spatiales des accidents de la route, Cybergéo, No. 185, 12 p. (http://www.cybergeo.presse.fr/revgeo/modelis/flahaut/flahaut1.htm).

FLEISS J.L. (1981), *Statistical methods for rates and proportions*. 2nd ed. USA : John Wiley & Sons, Inc.

GATRELL A.C. (1983), *Distance and space : a geographical perspective.* Oxford : Clarendon Press.

GOLDSTEIN H. (1995), Multilevel Statistical Models. London, Arnold, 178 p.

HAUER E. (1996), Identification of sites with promise, *Transportation Research Record*, vol.1542, 75th Annual Meeting, Washington D.C., p.54-60.

HOX J.J. (1995), Applied multilevel analysis. Amsterdam, TT-Publikaties, 119 p.

INSTITUT GÉOGRAPHIQUE NATIONAL (2002), Banques de données et cartes topographiques (Top50r), Bruxelles : IGN.

JOLY M-F., BOURBEAU R., BERGERON J., MESSIER S. (1992), *Analytical approach to the identification of hazardous road locations: a review of the literature.* Centre de recherche sur les transports, Université de Montréal.

JONES K. (1991), Concepts and techniques in modern geography. Multi-level models for geographical research. Norwich, Environmental Publications, University of East Anglia.

KENT M., COKER P. (1996), *Vegetation description and analysis. A practical approach.* Chichester : Wiley.

LEGENDRE P., LEGENDRE L. (1998), *Numerical ecology*. 2nd ed. Amsterdam : Elsevier Science B.V.

MATHIAN H., PIRON M. (2000), Echelles géographiques et méthodes statistiques multidimensionnelles in: SANDERS L. (éd.) Modèles en analyse spatiale, Paris, Hermès Science Publications, 62 -103.EVERITT B.S. (1977), *The analysis of contingency tables*. Great Britain : Chapman and Hall.

NGUYEN T.N. (1991), *Identification of Accident Blackspot Locations, an Overview.* VIC Roads/Safety Division. Research and Development Dept., Australia.

SILCOCK D.T., SMYTH A.W. (1985*), Methods of Identifying Accidents Blackspots.* Transport Operations Research Group. Dept. of Civil Engineering, University Of Newcastle Upon Tyne.

STOKES M.E., DAVIS C.S., KOCH G.G. (2000), *Categorical data analysis using the SAS system*. 2nd ed. NC, USA, SAS Institute Inc.

THOMAS I. (1992), La relation trafic-accidents sur autoroute. Approche statistique empirique. *Selected Proceedings of the Sixth World Conference on Transport Research, Politiques de transport*, Lyon.

THOMAS I. (1996), Spatial data aggregation : exploratory analysis of road accidents, *Accident Analysis and Prevention*, vol. 28, n°2, p.251-264.

THOMAS I., TULKENS H., BERQUIN P. (2000), Quelles frontières pour Bruxelles? La réponse d'un exercice statistique, géographique et économique. *In :* Combes P.-P.., Thomas I. (eds.), CIFOP, Rapport Commission 3, *XIVe Congrès des économistes Belges de Langue Française*, p. 73-88.

VANDERMISSEN M.H., POULIOT M., MORIN D.R. (1996), Comment estimer l'insécurité d'un site d'accident : état de la question, *Recherche Transports Sécurité*, vol.51, p.49-60.

VANDERMOTTEN C., MARISSAL P. (1993), Les structures démographiques et socio-économiques de quelques villes européennes : quelques réflexions comparatives. *Geographica Polonical*, n° 61, p. 103-120.

VANDERMOTTEN C., VERMOESEN F., DE LANNOY W., DE CORTE S. (1999), Villes d'Europe. Cartographie comparative, *Bulletin du Crédit Communal*, 53ème année, n° 207-208.

6.2 **Publications**

CASAER, F., WETS, G., THOMAS, I. , STEENBERGHEN, T. (2002) A Research on the Quality of Belgian Traffic Accident Data . (Draft)

DUFAYS T. AND T. STEENBERGHEN. (2002) "Cycle safe!" Geo:Connexion, Issue 3 Vol. 1, March 2002. pp.42-45.

GEURTS K., WETS G., BRIJS T. and VANHOOF K. (2002), The Use of Rule Based Knowledge Discovery Techniques to Profile Black Spots. Paper presented at the the 6th Design and Decision Support Systems in Architecture and Urban Planning Conference, Ellecom, The Netherlands, July 7-10

GEURTS, K. , WETS, G., BRIJS, T. AND VANHOOF, K. (2003). Profling High Frequency Accident Locations using Association Rules. Electronic Proceedings of the 82th Annual Meeting of the Transportation Research Board, Washington, January 12-16, USA, 18 pp.

STEENBERGHEN T., I. THOMAS, T. DUFAYS, EN B. FLAHAUT. (2003). Intra-urban location and clustering of road accidents using GIS: a Belgian example. Int. J. Geographical Information Science. 15 pp. (submitted)

ECKHARDT N., FLAHAUT B., THOMAS I. (2003), Spatio-temporalité des accidents de la route en périphérie urbaine. L'exemple de Bruxelles.» *Recherche Transports Sécurité* (soumis). This paper is joined by attachment.

Results of the precedent contract (MD/10/042 and MD/10/041):

DUFAYS T. (2000), Invloed van Ruimtelijke Ordening op Verkeersveiligheid; analyse van de Belgische situatie, Eindrapport van het contract MD/10/041.

FLAHAUT B. (2000), Influence de l'aménagement du territoire sur la sécurité routière durable. Analyse de la situation belge. SSTC, Rapport de recherche du contrat MD/10/042.

FLAHAUT B. (2001), L'autocorrélation spatiale comme outil géostatistique d'identification des concentrations spatiales des accidents de la route, *CYBERGEO*, n°185.

FLAHAUT B., MOUCHART M., SAN MARTIN E., THOMAS I. (2002), Identifying black zones with a local spatial autocorrelation index and a kernel method. A comparative approach, *Accident Analysis and Prevention*, (sous presse).

FLAHAUT B., THOMAS I. (2002), Identifier les zones noires d'un réseau routier par l'autocorrélation spatiale locale. Analyses de sensibilité et aspects opérationnels, *Revue internationale de Géomatique*, vol. 12, n°2, p. 245-261.

STEENBERGHEN, T. AND T. DUFAYS. (2000) "A GIS approach to the study of the impact of spatial planning on traffic safety; Case study of the historical centre of Mechelen" 4th GIS for European Historical Cities Euroconference, Seville, 10-12 May, 2000, 14pp.

STEENBERGHEN, T. AND T. DUFAYS. (2000) Impact of spatial planning on sustainable traffic safety; the mix of different transport modes as explanatory element, Velomondial2000 Conference, Amsterdam, 18-22 juni 2000, 14pp.

STEENBERGHEN, T. AND T. DUFAYS. (1999) Impact of spatial planning on sustainable traffic safety, Belgian situation analysis. Second Road Research Conference, Lausanne, Switserland 7-9/6/1999: 14 pp.