

Impact de l'aménagement du territoire sur la sécurité routière durable

Analyse de la situation belge.
Tâche A : Concentration spatiale des accidents de la route

Projet *SSTC MD/10/042*



**CENTRE D'ANALYSE SPATIALE ET URBAINE
UNIVERSITE CATHOLIQUE DE LOUVAIN**

Département de Géologie et Géographie
Unité de Géographie

Place L. Pasteur 3 B-1348 LOUVAIN-LA-NEUVE
Tel +32-(0)10-47.28.73 Fax 32-(0)10-47.28.77
<http://www.geo.ucl.ac.be>

Impact de l'aménagement du territoire sur la sécurité routière durable

Benoît FLAHAUT, sous la direction du Prof. I. THOMAS

Novembre 2000

Centre d'Analyse Spatiale et Urbaine, Unité de Géographie, U.C.L.

Place L. Pasteur 3 B-1348 LOUVAIN-LA-NEUVE

Tel +32-(0)10-47.28.73 Fax 32-(0)10-47.28.77

<http://www.geo.ucl.ac.be>

TABLE DES MATIERES

Table des matières	3
Introduction générale.....	5
Première partie : Délimitation des zones noires	6
1. Introduction	6
2. Données	7
3. Méthode d'analyse : l'autocorrélation spatiale.....	8
3.1 Pourquoi l'autocorrélation spatiale ?	8
3.2 L'autocorrélation spatiale globale.....	9
3.3 L'autocorrélation spatiale locale.....	10
3.4 Les pondérations w_{ij}	12
3.4.1 Le nombre de voisins.....	12
3.4.2 La valeur des pondérations.....	12
4. Application, analyse des résultats et étude de sensibilité pour la N29.....	13
4.1 Remarques préliminaires.....	13
4.2 L'indice de dangerosité.....	14
4.3 Étude de sensibilité.....	16
4.4 Conclusions	20
5. Comparaison avec la méthode par noyau : complémentarité et correspondances	20
5.1 Complémentarité	20
5.2 Correspondance des indices	21
5.3 Correspondance des classements des indices.....	22
5.4 Dissemblances entre les deux méthodes	22
6. Comparaison avec les zones rouges du MET.....	23
7. Conclusions	25

Deuxième partie : construction d'un modèle explicatif	27
1. Introduction	27
2. Méthode d'analyse : la régression logistique.....	28
2.1 La régression logistique	28
2.2 L'ajustement du modèle de régression	30
2.3 Construction du modèle de régression	31
2.4 Interprétation des coefficients de la régression	32
2.4.1 Une variable explicative dichotomique	32
2.4.2 Une variable explicative polychotomique	33
2.4.3 Plusieurs variables explicatives.....	34
2.5 Evaluation de la qualité de l'ajustement (goodness-of-fit) du modèle de régression	34
3. Zone d'étude et variables considérées	35
3.1 Le Brabant wallon	35
3.2 La variable dépendante (à expliquer).....	36
3.3 Les variables indépendantes (explicatives).....	38
4. Tableaux de contingence catégoriels : résumé.....	39
5. Construction d'un modèle multivarié	39
5.1 Tous types de routes confondus	39
5.2 Stratification suivant le type de route : les routes de type 1.....	46
5.3 Stratification suivant le type de route : les routes de type 2.....	47
5.4 Stratification suivant le type de route : les routes de type 3.....	48
5.5 Stratification suivant le type de route : les routes de type 4.....	49
6. Conclusions et perspectives	50
Bibliographie	53

INTRODUCTION GÉNÉRALE

Ce rapport s'inscrit dans le cadre du contrat de recherche SSTC N° MD/10/042 intitulé "L'impact de l'aménagement du territoire sur la sécurité routière durable : analyse de la situation belge". Dans ce projet, la sécurité routière belge est analysée en termes d'associations causales, avec une attention toute particulière sur l'impact de l'aménagement du territoire sur l'occurrence des accidents. L'aménagement, le trafic et l'infrastructure nous intéressent particulièrement.

Le projet est divisé en 4 tâches spécifiques, chacune étant elle-même subdivisée en plusieurs parties. Le présent rapport se rapporte à la tâche A. Celle-ci est divisée en 2 parties et est consacrée à l'étude de la structure spatiale des accidents de la route en Belgique. L'objectif est de modéliser la relation entre les *lieux de concentration d'accidents* et leur environnement, et d'identifier un lien éventuel entre lieux d'accidents et aménagement du territoire. Une meilleure compréhension des facteurs influençant la sécurité routière devrait permettre d'un point de vue pratique et appliqué, d'exercer une action concrète en aménagement du territoire en identifiant les structures environnementales et spatiales liées à la sécurité et à l'insécurité routières. La construction d'un tel modèle nécessite au préalable l'identification des zones d'accidents, et la première partie du projet a comme objectif de localiser les lieux de concentration spatiale d'accidents en les identifiant sous la forme de *zones noires*. La deuxième partie est consacrée à la construction du modèle explicatif de la co-occurrence spatiale des accidents de la route, autrement dit de la localisation des zones noires. Des rapports intermédiaires détaillent ces deux parties (Flahaut, 1999 et 2000).

Le travail portant sur la localisation des zones noires a donné lieu à deux publications dont voici les références :

FLAHAUT B., L'autocorrélation spatiale comme outil géostatistique d'identification des concentrations spatiales des accidents de la route, *Cybergéo*, accepté, à paraître début 2001.

FLAHAUT B. , MOUCHART M., SAN MARTIN E., THOMAS I., Local spatial autocorrelation and the kernel method for identifying black zones. A comparative approach, *Accident Analysis and Prevention*, en préparation, 2001.

PREMIÈRE PARTIE : DÉLIMITATION DES ZONES NOIRES

1. Introduction

En Belgique, les endroits dangereux du réseau routier sont classiquement représentés par des *points noirs*, soit des segments de route d'1 hectomètre qui comptabilisent au moins 3 accidents avec lésions corporelles pendant une année. Cette définition présente cependant les inconvénients (i) de ne reprendre que 10% du total des accidents de la route, (ii) de ne pas tenir compte de la migration des accidents dans l'espace au cours du temps, et (iii) de ne permettre l'identification que de problèmes très ponctuels. Cette situation est à l'origine d'une volonté d'identifier de façon plus satisfaisante les lieux d'accidents du réseau routier, qui seraient alors représentés sous la forme de *zones noires*.

Nous voulons proposer ici une définition des zones dangereuses (zones noires) qui soit basée sur des éléments statistiques : où et quand une zone est-elle *statistiquement* plus dangereuse qu'une autre ? Quelle longueur attribuer à chaque zone ? Longueur unique ou variable d'une zone à l'autre ?

Afin de répondre à ces questions, nous utiliserons des mesures de dangerosité basées sur l'*autocorrélation spatiale*. Les indices d'autocorrélation spatiale sont des indices qui permettent de mesurer la dépendance/association spatiale entre les valeurs x_i prises par une même variable X en des lieux caractérisés par une certaine proximité spatiale. Cette méthodologie permet de tenir compte de manière explicite de la distance dans sa formulation (nombre de voisins considérés comme proches) et de choisir la longueur de zone la plus appropriée à la réalité. Les résultats seront ensuite comparés à ceux obtenus par une méthode statistique d'estimation par noyau ; ils seront en outre utilisés par l'*IBSR* pour une étude en profondeur d'un échantillon de zones noires

2. Données

Nous nous limiterons ici aux données concernant les seuls *accidents avec lésions corporelles* qui ont eu lieu sur le réseau de *routes numérotées* gérées par les régions, car ce type d'accident est en principe toujours répertorié puisqu'il nécessite l'intervention des services de gendarmerie ou de police. Contrairement aux accidents avec dégâts matériels seulement, ils sont *obligatoirement* soumis à la rédaction d'un procès-verbal ainsi que d'un rapport pour l'*INS* (Institut National de Statistiques) sous la forme d'un formulaire à compléter. A ce type d'accident correspond en Wallonie un chiffre d'environ 8400 accidents par année en moyenne avec une décroissance observée dans le temps, soit au total 42116 accidents sur 5 ans durant la période de 1992 à 1996. En Flandre, les données font état de 95095 accidents avec lésions corporelles pendant la même période, soit plus du double du nombre d'accidents en Wallonie alors que le réseau routier considéré est plus petit.

Le réseau routier concerné par cette étude consiste en l'ensemble des routes numérotées régionales, plus les rings et les autoroutes. Cela représente au total pour la Wallonie plus de 8400 km de routes, décomposés en 840 kilomètres d'autoroutes (plus les rings avec numéro strictement inférieur à 9) et 7600 km de routes numérotées (plus les rings avec numéro supérieur à 9). Pour la Flandre, ce réseau représente un peu moins de 7000 km au total dont 825 km d'autoroutes et 6100 km de routes numérotées.

Un accident est localisé suivant sa position par rapport à la borne kilométrique la plus proche, soit avec une précision de 100 mètres; il est donc possible d'affecter un nombre d'accidents k à chaque hectomètre i de route.

Parmi les problèmes liés aux données, relevons le peu de confiance à accorder aux données d'accidents concernant les routes numérotées qui passent dans les grandes agglomérations notamment Charleroi, Liège et Namur pour la Wallonie. Dans les villes en effet, les accidents sont souvent localisés par le nom de la rue (et pas le numéro) et par le numéro du bâtiment devant lequel a lieu l'accident (et pas la *BK*). Ces accidents sont dès lors perdus pour les statistiques relatives aux routes numérotées à moins d'entreprendre leur relocalisation, et la perte de ces accidents a pour conséquence un manque de confiance des données urbaines. La délimitation de tronçons dangereux en zone urbaine est donc rendue moins sûre et c'est d'ailleurs pour cette raison que l'agglomération bruxelloise n'est pas reprise dans les analyses.

Tableau 2.1 : tableau récapitulatif des données

	Wallonie	Flandre	Bruxelles (non considéré dans cette étude)
Longueur du réseau de routes numérotées (1996)	7597km 54%	6097km 44%	232km 2%
Longueur du réseau autoroutier (1996)	840km 50%	825km 49%	11km 1%
Longueur du réseau total (revêtu) (estimation 96)	+ 47100 = 55537km 48 %	+ 52100 = 59020km 51 %	+ 1400 = 1643km 1 %
Accidents LC sur routes numérotées (1992-96)	36360	81779	-
Accidents LC sur autoroutes (1992-96)	5756	13318	-
Total accidents LC (1992-96)	42116	95097	-

Source : D. Antoine, MET

3. Méthode d'analyse : l'autocorrélation spatiale

Cette section 3 se présente de la manière suivante. La section 3.1. permettra de comprendre de manière générale le choix de la méthode de l'autocorrélation spatiale pour la construction d'un indice de dangerosité. L'autocorrélation spatiale globale est expliquée dans la section 3.2. tandis que la section 3.3. est consacrée à l'autocorrélation spatiale locale. Dans la section 3.4. sont expliquées les possibilités qui existent quant au choix des paramètres (pondérations w_{ij}) à introduire dans le calcul de l'indice.

3.1 Pourquoi l'autocorrélation spatiale ?

Une propriété souvent constatée lors de l'observation de données à caractère spatial est que les valeurs $\{x_i\}$ prises par une même variable X en différents lieux i présentent des relations entre observations proches dans l'espace. Cette idée est largement considérée depuis longtemps dans la littérature spatiale, par exemple pour Tobler (1970) : "everything is related to everything else, but near things are more related than distant things", ou Gould (1970) : "Why we should expect independence in spatial observations (...). All our efforts to

understand spatial patterns, structure, and process have indicated the lack of independence (...) of things in time and space".

Si les $\{x_i\}$ présentent de l'interdépendance dans l'espace, ce qui est fréquemment le cas, on dit alors que les données sont **autocorrélées spatialement** (Cliff et Ord, 1981). Des mesures d'autocorrélation spatiale permettent dès lors d'estimer l'association/dépendance/corrélation spatiale entre les valeurs d'une même variable en différents endroits de l'espace, plus ou moins proches les uns des autres. La relation entre l'interaction spatiale et l'autocorrélation spatiale a été démontrée par Getis (1991).

Les indices d'autocorrélation spatiale permettent de mettre en évidence des lieux qui présentent des relations par rapport à 2 critères simultanés :

- La proximité spatiale
- La ressemblance ou l'opposition entre les valeurs d'une même variable en différents endroits de la région d'étude

En appliquant ces considérations au problème de l'identification des zones noires du réseau routier, nous pouvons réécrire l'objectif de l'étude de la manière suivante, sachant que l'hectomètre constitue la plus petite unité spatiale à laquelle les accidents sont localisés : il s'agit de regrouper des hectomètres voisins en zones de façon à déterminer des zones dangereuses et des zones qui ne le sont pas, ou encore il s'agit d'agréger en tronçons de route suffisamment homogènes des hectomètres spatialement proches qui se ressemblent quant au nombre d'accidents qui les caractérise, de façon à identifier les zones dangereuses. Une association de valeurs élevées d'accidents entre des hectomètres voisins sera considérée comme formant une zone noire, et cette association est déterminée par les mesures d'autocorrélation spatiale.

Il s'agit donc de proposer la mesure d'un **indice de dangerosité de tronçons routiers qui soit basé sur une mesure d'autocorrélation spatiale**. Grâce à la méthode de l'autocorrélation spatiale, la longueur des zones peut être indiquée (elle dépend du nombre de voisins pris en compte dans le calcul de l'indice, voir section 3.4.1), de même que l'intensité du caractère dangereux de la zone (il dépend de la valeur de l'indice).

3.2 L'autocorrélation spatiale globale

Les mesures usuelles d'autocorrélation spatiale s'appliquent de façon globale à l'ensemble des lieux considérés. Elles permettent d'estimer si, globalement, les lieux appartenant à la région d'étude présentent le phénomène d'autocorrélation spatiale.

Pour des variables quantitatives, le coefficient de Moran est souvent utilisé dans la littérature pour tester la présence d'autocorrélation spatiale (voir notamment Upton et Fingleton, 1985). Il utilise au numérateur un terme de covariance pondérée entre observations contiguës, covariance nulle en absence d'autocorrélation spatiale, positive dans le cas d'autocorrélation positive, et négative en cas d'autocorrélation négative. Le dénominateur est constitué d'une mesure de la variance des observations.

$$\text{Moran's } I = (n/S_0) \sum_i \sum_j w_{ij} z_i z_j / \sum z_i^2 \quad [1]$$

Où w_{ij} sont les pondérations reflétant les relations de proximité

$$S_0 = \sum_i \sum_j w_{ij}$$

$$z_i = x_i - \bar{x}$$

$$z_j = x_j - \bar{x}$$

x_i = valeur de la variable X au lieu i

\bar{x} = valeur moyenne des x_i

n = nombre de lieux

Dans la construction de l'indice, c'est la valeur moyenne \bar{x} des observations qui constitue la valeur de référence, valeur qui permet de déterminer des valeurs x_i élevées ou faibles. A un z_i positif correspond à une valeur élevée de x_i , tandis qu'à un z_j négatif correspond une valeur faible de x_i .

3.3 L'autocorrélation spatiale locale

Des travaux plus récents ont permis de développer des *mesures locales* d'autocorrélation spatiale (Getis et Ord, 1992, Ord et Getis, 1995, Anselin, 1995). Ces mesures permettent de décomposer l'indice global de façon à identifier la contribution individuelle de chaque lieu de la région d'étude; ils mesurent successivement pour chacun la dépendance/association spatiale entre la valeur de la variable prise en ce lieu et l'ensemble de celles prises dans son voisinage. Ces indices locaux permettent donc de détecter les poches locales d'autocorrélation spatiale

L'indice local de Moran mesuré au lieu i est défini comme un *LISA* (Local Indice of Spatial Autocorrelation, voir Anselin, 1995) de la façon suivante :

$$I_i = z_i \sum_j w_{ij} z_j, \quad [2]$$

et

$$\sum_i I_i = \sum_i z_i \sum_j w_{ij} z_j \quad [3]$$

La somme des indices locaux est proportionnelle à l'indice global par le facteur $g = S_0 m_2$, avec $m_2 = \sum_i z_i^2 / n$. Il est également possible de standardiser l'indice local en le divisant par m_2 .

Puisqu'un indice local d'autocorrélation spatiale peut être mesuré pour chaque hectomètre i de route, et puisque cet indice local peut mettre en évidence une association entre cet hectomètre i et des hectomètres voisins j qui lui ressemblent quant à leur nombre d'accidents, l'indice local d'autocorrélation spatiale peut être utilisé comme **indice de dangerosité**. Par conséquent, pour l'identification des zones noires du réseau routier, chaque hectomètre de route est successivement considéré comme une observation individuelle, afin de découvrir s'il existe dans le voisinage de chacun d'eux, une tendance à présenter des valeurs similaires ou opposées (autocorrélation spatiale positive ou négative). Un **indice local** d'autocorrélation spatiale est calculé pour chaque hectomètre, et cet indice local peut être considéré comme une mesure locale de dangerosité. La mesure de la tendance du voisinage est réalisée par rapport à la moyenne ($\sum_j w_{ij} z_j$ dans la formule [2]), éventuellement pondérée selon la distance à l'hectomètre i considéré, du nombre d'accidents des hectomètres constitutifs du voisinage. Une zone, noire ou non, est constituée d'un hectomètre central i et d'hectomètres voisins j , et la mise en évidence d'une tendance vers des valeurs similaires dans le voisinage d'un hectomètre i (autocorrélation spatiale positive) peut permettre de les considérer comme un tronçon de route **homogène**, formant une **zone noire** si la tendance observée est un nombre élevé d'accidents à la fois pour l'hectomètre i et pour son voisinage.

Les indices locaux d'autocorrélation spatiale, considérés comme des indices de dangerosité, permettent donc d'identifier les zones noires du réseau routier et d'évaluer l'intensité de leur caractère dangereux.

Un paramètre important intervient dans la formulation de l'indice de Moran, et pour lequel les valeurs qui lui sont attribuées influencent fortement la valeur de l'indice : les **pondérations** w_{ij} qui représentent les relations de voisinage entre les hectomètres de route. Nous présentons dans la section 3.4 un résumé de cet aspect de la méthode, qui est largement développé dans le rapport intermédiaire correspondant (Flahaut, 1999).

3.4 Les pondérations w_{ij}

L'idée d'une mesure unique de l'autocorrélation spatiale n'est malheureusement pas réaliste en raison de l'inexistence d'une définition univoque et définitive des pondérations optimales devant représenter les relations de contiguïté entre les différents lieux. Les mesures d'autocorrélation spatiales sont toujours à mettre en relation avec la structure spatiale correspondante qui a été choisie. Des mesures différentes peuvent être obtenues en considérant des matrices de pondérations différentes, et le fait d'arriver à une conclusion de présence ou d'absence d'autocorrélation spatiale pour une structure particulière ne signifie pas qu'il en sera de même pour d'autres.

Deux considérations interviennent dans l'évaluation des coefficients de pondération : le **nombre de voisins** de chaque lieu (c'est-à-dire la distance de voisinage, ou encore le niveau de contiguïté) et la **valeur des poids** attribués à chacun d'eux.

3.4.1 Le nombre de voisins

Comme nous ne pouvons pas choisir *a priori* un nombre optimal de voisins, nous avons effectué **pour chaque hectomètre i** de chaque route une analyse systématique et exploratoire pour mettre en évidence la structure spatiale des données, c'est-à-dire pour évaluer le nombre de voisins le plus approprié. Dans ce but, des longueurs de zones variant entre 300 et 2100 mètres de 200 mètres en 200 mètres (soit un nombre de voisins variant de 2 à 20, répartis de façon symétrique de part et d'autre de l'hectomètre central i), sont successivement utilisées pour calculer un **indice local** d'autocorrélation spatiale I_i . Les valeurs de ces 10 indices sont comparées, et le nombre optimal de voisins est celui pour lequel la valeur de l'indice local d'autocorrélation spatiale est maximale, traduisant ainsi la meilleure association des valeurs élevées de la variable entre un hectomètre et ses voisins (c'est-à-dire le produit de deux valeurs positives). Cette démarche est répétée pour chaque hectomètre de chaque route, de façon à déterminer pour chacun une valeur de l'indice liée à un nombre de voisins adapté à la *structure spatiale locale*.

3.4.2 La valeur des pondérations

En ce qui concerne la **valeur des pondérations**, de nombreuses possibilités existent aussi (voir notamment Cliff et Ord, 1973 et 1981, Haining, 1990). Nous avons considéré des poids qui sont fonction de la distance au lieu considéré. Parmi 4 fonctions de décroissance des pondérations avec la distance (fonction de d_{ij}^0 , de d_{ij}^{-1} , de $d_{ij}^{-1.5}$, ou de d_{ij}^{-2}), nous avons choisi celle qui maximise l'indice moyen d'autocorrélation spatiale locale : d_{ij}^{-2} . Elle traduit de la sorte la plus forte association entre les nombres d'accidents observés dans les zones noires (centre et voisinage). Cette constatation est en cohérence avec la littérature relative à l'interaction spatiale qui considère généralement que les relations spatiales entre lieux sont

décroissantes avec l'inverse du carré de la distance. Ajoutons que la somme des poids par ligne de la matrice de proximité (c'est-à-dire les poids attribués aux voisins de chaque lieu de façon individuelle) est égale à l'unité ($\sum_j w_{ij} = 1, \forall i$). Cette modification présente l'intérêt de permettre une comparaison des mesures d'autocorrélation spatiale entre différentes distances de voisinage, puisque pour chacune d'elles une importance identique est accordée à l'ensemble des voisins indépendamment de leur nombre..

Après cette section sur l'autocorrélation spatiale, la section 4 est consacrée à l'application de la méthode. Cette application a été réalisée route par route pour l'ensemble du réseau numéroté de Belgique, et nous en présentons ici une illustration pour la N29.

4. Application, analyse des résultats et étude de sensibilité pour la N29

4.1 Remarques préliminaires

Afin de réduire quelque peu l'hétérogénéité des données à analyser, l'ensemble du réseau routier numéroté belge a été considéré en quatre parties : les autoroutes de Wallonie, les autres routes numérotées de Wallonie, les autoroutes de Flandre, et les autres routes numérotées de Flandre. L'ensemble des résultats pour ces quatre parties a été communiqué à l'*IBSR* pour servir de base à un échantillonnage pour une analyse en profondeur de plusieurs zones noires. Quant à la *N29* appartient au réseau des routes numérotées de Wallonie (hors autoroutes), et pendant la période d'étude (5 ans de 1992 à 96), 36360 accidents avec lésions corporelles ont eu lieu sur ce réseau wallon. Sur ces 36360 accidents, tous ne peuvent être considérés dans les analyses, certains d'entre eux n'étant pas localisables puisqu'ils présentent une borne kilométrique nulle (localisation inconnue). Leur nombre est ainsi ramené à 30729 accidents.

L'application de l'indice local d'autocorrélation spatiale de Moran à un ensemble donné de lieux nécessite le calcul du nombre moyen d'accidents par hectomètre de cet ensemble (\bar{x}). Pour cette application, l'ensemble de lieux est constitué des hectomètres auxquels au moins un accident a eu lieu. Sur les 17997 hectomètres du réseau wallon hors autoroutes caractérisés par au moins un accident, seuls 17792 sont retenus (élimination des hectomètres de début de route). Le nombre moyen d'accidents par hectomètre (\bar{x}) est donc de **1,73 accidents**. Quant aux pondérations w_{ij} , elles sont calculées avec un nombre de voisins variant entre 2 et 20 de façon à pouvoir s'adapter à la structure spatiale locale de chaque zone (en maximisant l'indice local, voir section 3.4.1), et leur décroissance avec d_{ij} (distance entre un hectomètre j

considéré comme voisin et l'hectomètre i auquel l'indice local d'autocorrélation spatiale est calculé) est fonction de l'inverse du carré de cette distance (d_{ij}^{-2}).

Il faut insister ici sur le fait que cette analyse par zones ne doit absolument pas se substituer à l'approche par points noirs qui permet quant à elle d'identifier des problèmes plus ponctuels. L'identification des endroits à risque du réseau routier par zones noires et par points noirs constituent des approches complémentaires indissociables.

Dans la section suivante (4.2) est étudiée la relation entre l'indice local d'autocorrélation spatiale et le nombre moyen d'accident par hectomètre de chaque zone auquel l'indice se rapporte; une carte des zones dangereuses de la N29 y est présentée, basée sur l'utilisation d'un nombre de voisins propre à chaque hectomètre de façon à maximiser chaque indice local, et sur la décroissance des pondérations w_{ij} avec d_{ij}^{-2} . Dans la section 4.3. est présentée une étude de sensibilité des résultats selon plusieurs niveaux de contiguïté et différentes fonctions de décroissance des pondérations.

4.2 L'indice de dangerosité

Les tableaux 4.1 et 4.2 présentent des informations relatives à la distribution du nombre d'accidents par hectomètre (variable Y) sur cette route longue de 527 hectomètres.

Tableau 4.1

Y	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
0	268	50.9	268	50.9
1	145	27.5	413	78.4
2	49	9.3	462	87.7
3	29	5.5	491	93.2
4	18	3.4	509	96.6
5	8	1.5	517	98.1
6	4	0.8	521	98.9
7	4	0.8	525	99.6
9	1	0.2	526	99.8
13	1	0.2	527	100.0

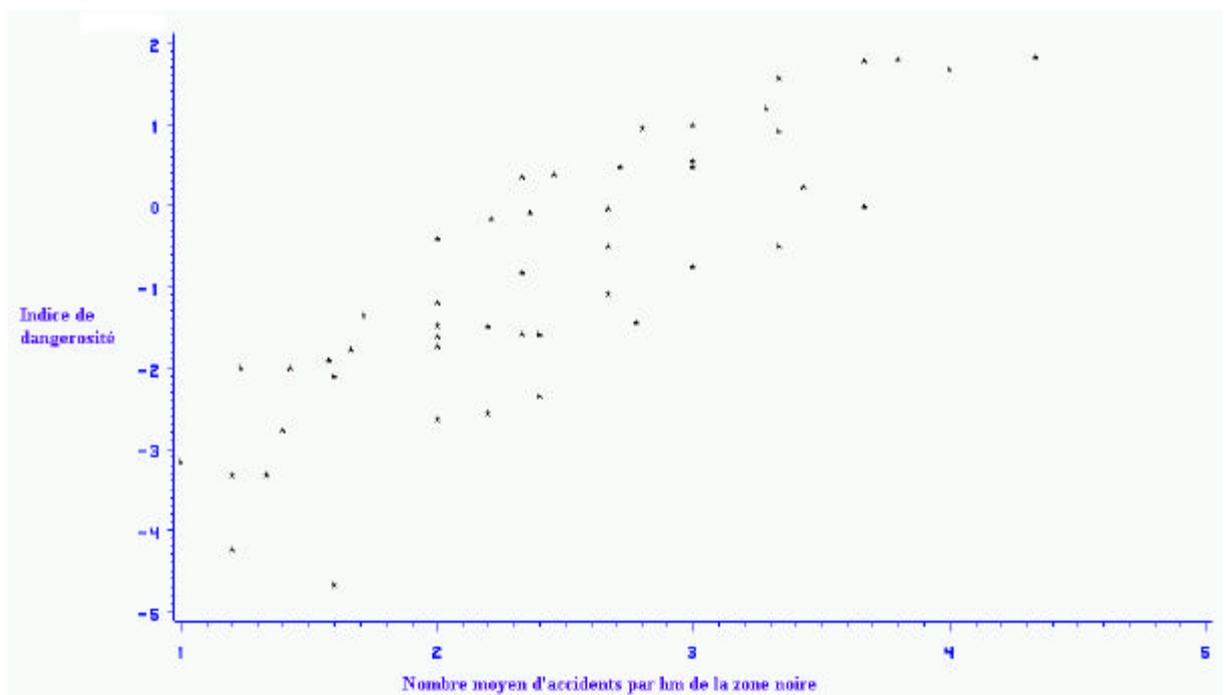
Tableau 4.2

N	Min	Max	Range	Mean	Variance	Std Dev	Std Error	CV
527	0	13	13	0.978	2.23	1.49	0.065	152.63

Un indice de dangerosité devrait présenter une corrélation positive élevée avec le nombre moyen d'accidents par hectomètre de la zone à laquelle il se rapporte. L'observation du graphique illustrant la relation entre ces deux variables montre qu'elle n'est pas linéaire mais plutôt exponentielle. Or, pour évaluer le degré d'association entre deux variables, le coefficient de corrélation de Pearson mesure la corrélation *linéaire* entre ces variables. Une transformation de la variable "indice d'autocorrélation" a donc été effectuée en prenant son logarithme (voir figure 4.1), et à ce moment, la corrélation est meilleure qu'auparavant et l'hétéroscédasticité est également réduite. Pour la N29, la corrélation avec l'indice local d'autocorrélation spatiale est de 0.74 (en enlevant les valeurs extrêmes perturbatrices) tandis qu'elle est de 0.87 avec le logarithme de l'indice (avec d_{ij}^{-2}).

De plus, le logarithme des indices de dangerosité présente aussi une meilleure correspondance avec les indices obtenus avec la méthode par noyau (voir section 5) que simplement les indices locaux d'autocorrélation spatiale.

Figure 4.1 : Relation entre l'indice de dangerosité ($\log(\text{indice local d'AC})$) et le nombre moyen d'accidents par hectomètre de la zone noire.



A titre d'exemple, la cartographie des zones noires de la N29 est présentée à la page suivante sur la figure 4.2. Les zones dangereuses sont classées selon les quantiles, en 5 catégories. Seules les zones appartenant aux 3 quantiles les plus élevés sont cartographiées. Il faut souligner ici que toutes les zones noires identifiées ne nécessitent pas forcément un aménagement, l'intensité du caractère dangereux de certaines d'entre elles n'étant en effet pas suffisamment élevée pour cela. Pour rappel, cette carte est construite avec des pondérations proportionnelles à d_{ij}^{-2} . L'observation de la carte montre qu'il peut exister des zones noires superposées, dont l'intensité du caractère dangereux peut alors varier : un hectomètre peut appartenir à plusieurs zones noires différentes, mais qui se recouvrent par endroits.

4.3 Étude de sensibilité

La sensibilité des résultats à la valeur prise par les paramètres intervenant dans la construction de l'indice d'autocorrélation a été étudiée, de façon à pouvoir prendre conscience des conséquences qu'implique le choix de certaines valeurs des paramètres plutôt que d'autres. L'ensemble des analyses de sensibilité est détaillé dans le rapport intermédiaire correspondant. Nous en présentons ici un résumé.

1) Le **coefficient global d'autocorrélation spatiale** a été successivement calculé avec un *même nombre de voisins fixe*, égal à 2 (soit seulement des zones de 300 mètres), 4 (zones de 500m), 6 (700m), ..., ou 20 (2100m). Pour chaque nombre de voisins, chacune des 3 fonctions de décroissance des pondérations w_{ij} (proportionnelles à d_{ij}^{-1} , $d_{ij}^{-1.5}$, d_{ij}^{-2}) a été considérée, ainsi que des pondérations indépendantes de d_{ij} (autrement dit proportionnelles à d_{ij}^0). Une variabilité des mesures de l'autocorrélation spatiale globale selon la longueur des zones est observée, ce qui montre l'intérêt de ne pas fixer à priori la longueur des zones pour le calcul des indices locaux (leur somme est proportionnelle à l'indice global), et de choisir pour chaque hectomètre la longueur la mieux adaptée à la structure spatiale locale. La figure 4.3 illustre ces variations pour tous les exposants et tous les niveaux de contiguïté, tandis que la figure 4.4 isole la relation pour un exposant de $d_{ij} = -2$.

Figure 4.2. Zones noires de la N29

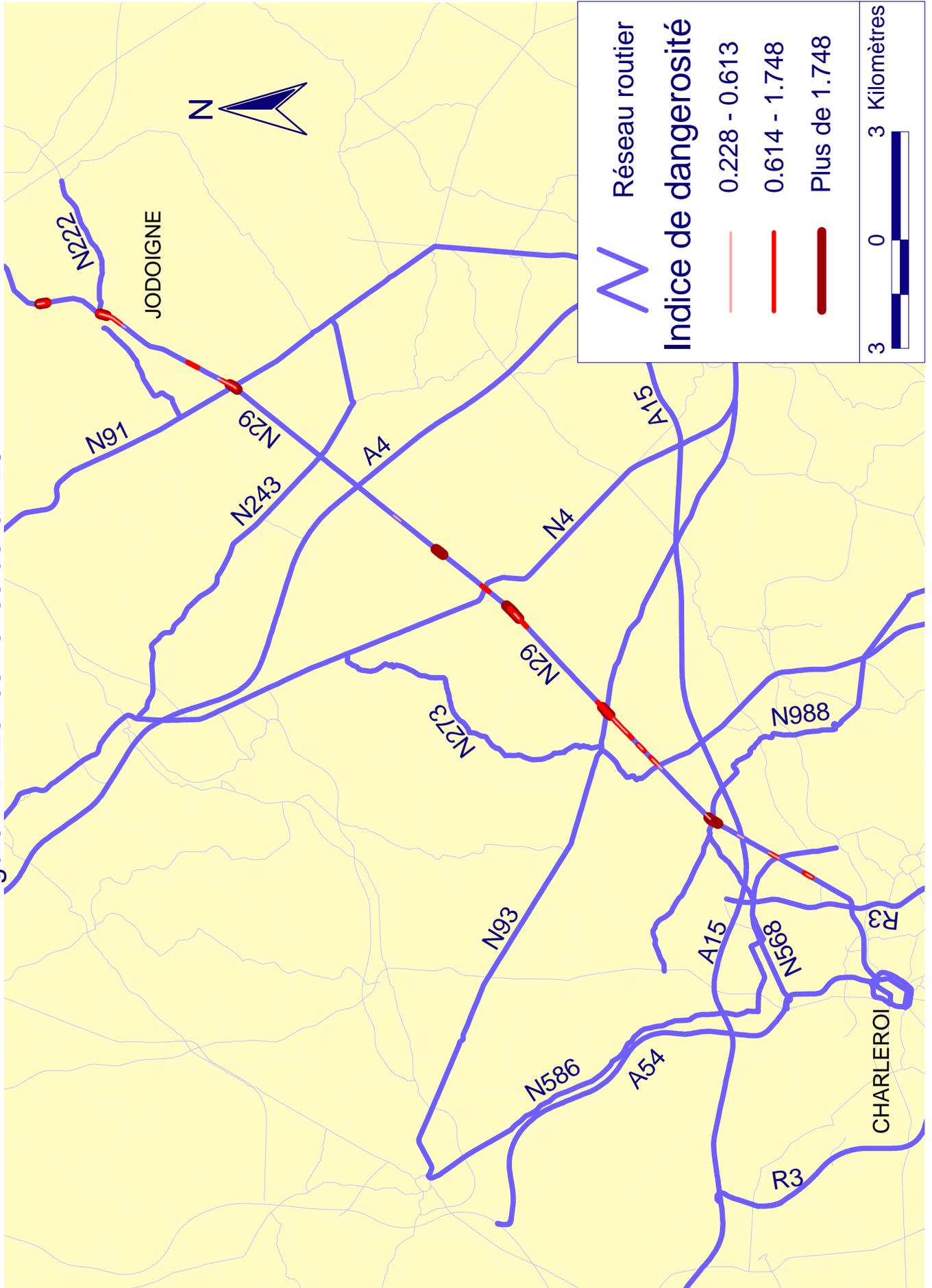


Figure 4.3 Variation de l'indice global avec le niveau de contiguïté, pour différentes fonctions de décroissance des pondérations avec dij (pour des zones de 300 à 2100 mètres)

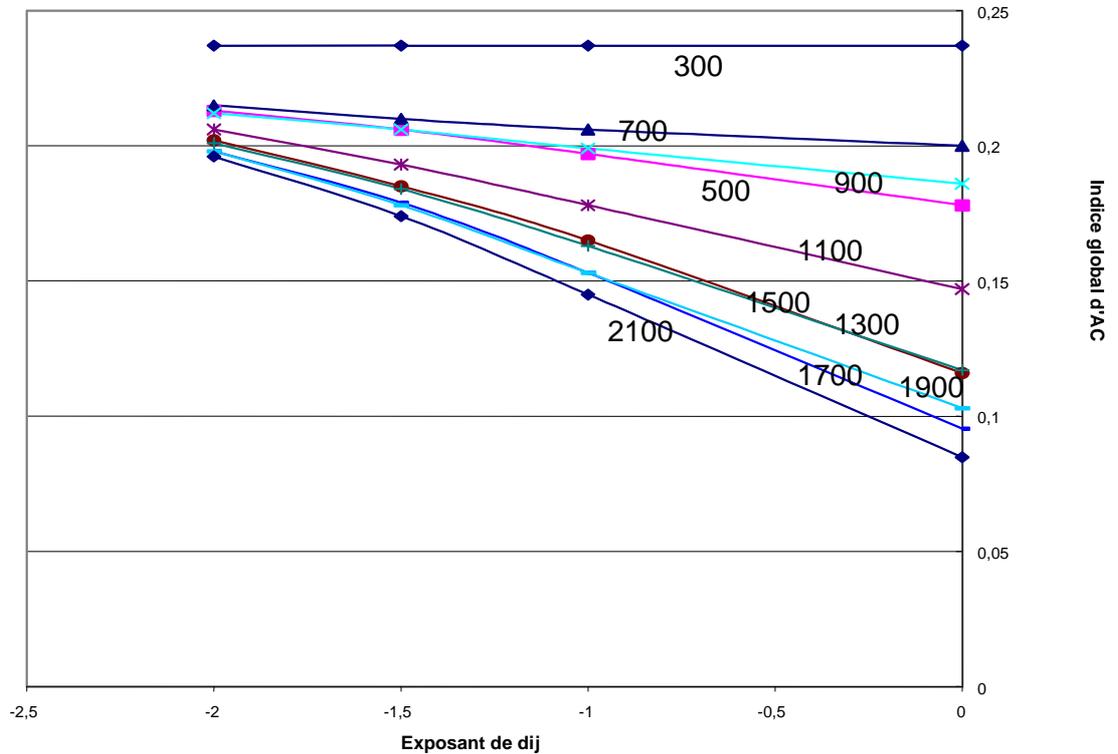
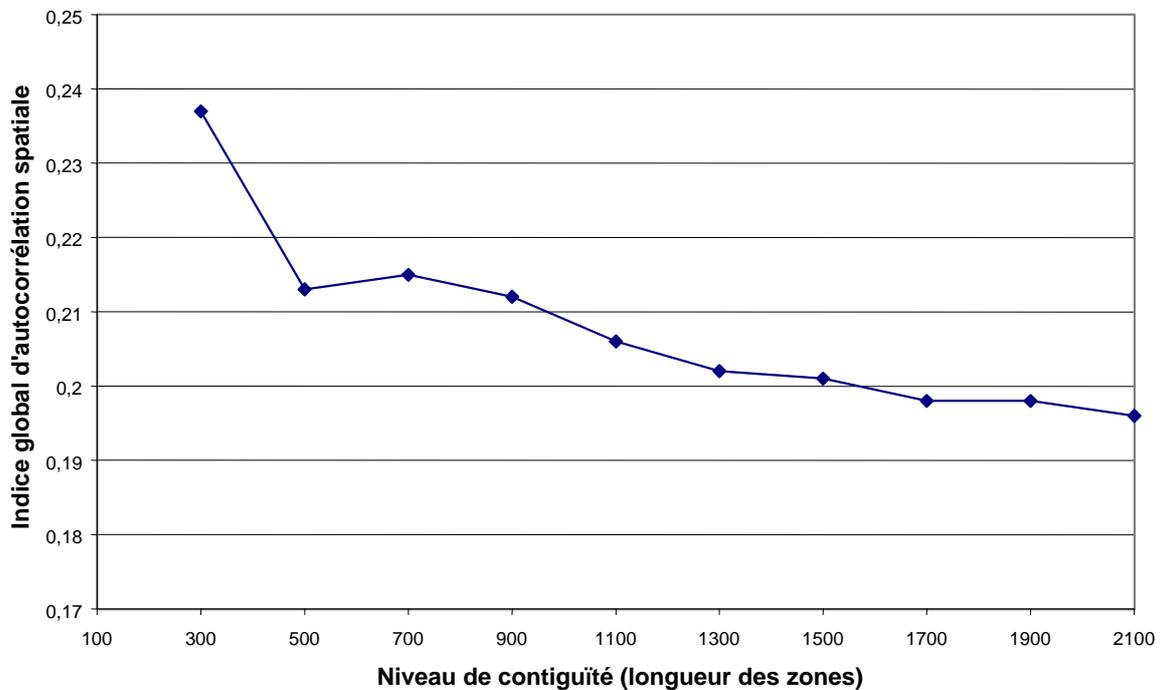


Figure 4.4 Variation de l'indice global d'autocorrélation spatiale selon le niveau de contiguïté (exposant de dij = -2)



2) En choisissant non plus une longueur fixe pour chaque zone noire, mais cette fois la *longueur adaptée* à la structure spatiale locale, l'influence de la fonction de décroissance des pondérations avec d_{ij} sur la détermination des zones noires a été étudiée. En attribuant une importance décroissante aux observations les plus éloignées (exposant de d_{ij} de plus en plus petit), on peut observer une diminution du nombre de zones noires identifiées, ainsi qu'une diminution de leur longueur moyenne, et une augmentation de la valeur moyenne de l'indice local d'autocorrélation spatiale. Puisque nous utilisons un exposant égal à - 2, les zones noires identifiées sont donc moins nombreuses qu'avec un exposant plus grand (ce sont les plus longues qui ne sont plus identifiées), leur longueur est en moyenne plus petite, et elles sont aussi en moyenne d'une intensité dangereuse plus élevée.

Il faut insister ici sur le fait que parmi le nombre relativement élevé de zones noires identifiées, toutes ne doivent pas être considérées comme un endroit nécessitant un aménagement visant à diminuer leur caractère dangereux : parmi ces zones, certaines sont dangereuses puisqu'elles ont été identifiées comme telles, mais elles le sont très peu, et il importe donc de classer les zones noires selon l'intensité de leur caractère dangereux afin de s'intéresser plus particulièrement à celles qui le sont vraiment.

3) La *correspondance de l'intensité du caractère dangereux* entre les zones noires identifiées avec plusieurs fonctions des pondérations a été analysée par des mesures de corrélation de Pearson. Celles-ci montrent l'existence d'une corrélation positive élevée (variant entre 0.87 et 0.98) permettant d'affirmer de façon significative que ces indices présentent une corrélation linéaire différente de 0. La variation de la fonction de décroissance des pondérations avec la distance influence donc les résultats au niveau du nombre de zones noires identifiées plutôt qu'au niveau de l'intensité de leur caractère dangereux. Ce sont donc les zones noires les moins significatives qui ne sont plus identifiées en diminuant l'exposant de d_{ij} .

Ces indices de dangerosité ont ensuite été *classés en 5 catégories, suivant la méthode des quantiles*, et l'appartenance des zones noires à la même classe quel que soit l'exposant a été étudiée. Les corrélations calculées (coefficient de Spearman et taux de Kendall) montrent à nouveau que la variation de la fonction de décroissance des pondérations avec la distance influence plutôt les résultats au niveau du nombre de zones noires identifiées (qui diminue avec un exposant de d_{ij} plus petit), tandis que le classement des zones noires est quant à lui significativement très semblable pour des exposants différents (Spearman varie entre 0.83 et 0.96, Kendall entre 0.74 et 0.94).

4.4 Conclusions

Le calcul d'indices de dangerosité basés sur l'autocorrélation spatiale apparaît être une méthode satisfaisante d'identification des zones dangereuses d'une route. Les indices de dangerosité présentent une relation positive élevée avec le nombre moyen d'accidents par hectomètre de chaque zone noire, et le choix d'une longueur de zone propre à chacune permet une adaptation à la structure spatiale locale observée. Quant à la fonction de décroissance des pondérations avec la distance, le choix de l'exposant (égal à -2) permet de limiter le nombre et la longueur des zones identifiées comme étant dangereuses et dès lors de gagner en précision. De plus, les zones noires déterminées de cette façon présentent en moyenne une dangerosité plus marquée qu'avec un exposant plus petit, liée à la non-identification des zones les moins dangereuses.

Afin d'évaluer la stabilité des résultats, ils ont été comparés avec ceux obtenus par 2 méthodes alternatives : la méthode statistique par noyau et la méthode des zones rouges développée par le *MET*.

5. Comparaison avec la méthode par noyau : complémentarité et correspondances

5.1 Complémentarité

Brièvement, il s'agit d'une méthode non-paramétrique utilisant une fonction d'estimation de densité de probabilité pour évaluer la probabilité d'occurrence d'un accident à un endroit donné, et donc la dangerosité de celui-ci. Cette approche a été développée en collaboration avec le département de statistique de l'UCL (M. Mouchart et E. San Martin). Les deux approches réalisées afin d'identifier les zones noires du réseau routier (la méthode basée sur l'autocorrélation spatiale et celle par noyau) sont conceptuellement différentes, ce qui rend leur comparaison intéressante. Elles ont comme points communs d'attribuer une pondération décroissante au nombre d'accidents des hectomètres les plus éloignés, mais les valeurs de ces pondérations sont estimées de façon différente, de même que le nombre de voisins qu'il convient de considérer. La méthode du noyau est détaillée en annexe du rapport intermédiaire consacrée à cette première partie du projet (Flahaut, 1999).

Pour l'autocorrélation spatiale, le nombre de voisins pris en compte est en relation avec la structure spatiale observée autour de chaque hectomètre de chaque route (puisqu'on prend l'indice maximal qui correspond à la meilleure association entre un hectomètre et ses voisins), et il est donc variable d'un hectomètre à un autre (entre 2 et 20 voisins, soit une longueur de zone entre 300m et 2100m). Pour le noyau par contre, ce nombre est déterminé

par la fenêtre $2h$ qui permet de lisser les données, et la valeur de cette fenêtre est constante pour tous les hectomètres d'une même route. La valeur optimale de cette fenêtre est déterminée statistiquement par la minimisation l'erreur quadratique moyenne intégrée (*mean integrated square error*, ou *MISE*). La fenêtre optimale $2h$ est fixe pour chaque route, et est en outre d'une taille relativement grande (5300m pour la N29, et fréquemment plus) par rapport aux zones déterminées avec la méthode de l'autocorrélation spatiale locale (au maximum 2100m). Ces deux approches peuvent donc être caractérisées comme **globale** pour la méthode par noyau (approche macro ou mésospatiale) et plus **locale** pour la méthode d'autocorrélation spatiale (approche microspatiale), ce qui permet de les considérer comme étant complémentaires.

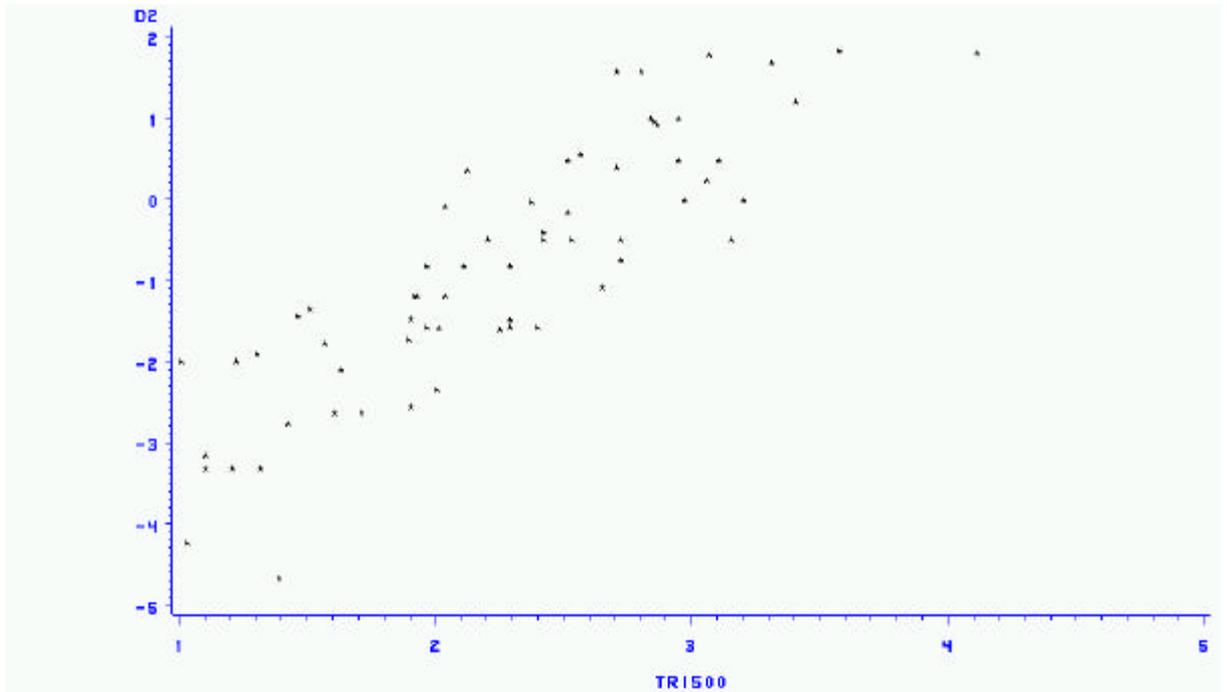
La comparaison des résultats des deux méthodes consiste en mesures de corrélation entre d'une part les *valeurs* des indices (section 5.2), et d'autre part entre les *classements* de ces indices (section 5.3). Quant à la section 5.4, elle est consacrée à l'examen des hectomètres dangereux non communs aux deux méthodes.

5.2 Correspondance des indices

Des corrélations ont été calculées entre les valeurs des indices obtenus par les deux méthodes pour la N29 avec différentes valeurs des paramètres, et il faut pour obtenir de bonnes corrélations, utiliser pour la méthode par noyau une fenêtre plus petite que la fenêtre statistiquement optimale. La taille est alors plus semblable à celles utilisées pour l'autocorrélation spatiale : des fenêtres de 2000m, de 1000m et de 500m sont utilisées. Les corrélations de Pearson calculées montrent qu'il existe une correspondance significative entre les résultats obtenus par ces 2 méthodes. La meilleure corrélation entre les indices des deux méthodes est obtenue avec une fenêtre de 500m pour la méthode par noyau et les pondérations proportionnelles à d_{ij}^{-2} pour la méthode de l'autocorrélation ($r = 0.86$). La figure 5.1 illustre cette relation.

Cette bonne correspondance est très intéressante puisque les résultats sont donc stables en appliquant deux méthodes différentes mais rendues comparables par une similarité des paramètres intervenant dans les deux méthodes à la fois.

Figure 5.1 Relation entre l'indice de dangerosité basé sur l'autocorrélation spatiale avec les pondérations proportionnelles à d_{ij}^{-2} (D2) et l'indice de dangerosité basé sur le noyau avec une fenêtre $2h$ de 500 mètres (TRI500).



5.3 Correspondance des classements des indices

La correspondance entre les classements obtenus par chacune des deux méthodes (en 5 catégories suivant les quantiles) a été évaluée par le calcul du coefficient de Spearman et du taux de Kendall. Ces mesures montrent à nouveau l'existence de relations significatives, la plus élevée étant obtenue de nouveau avec une fenêtre de 500m pour la méthode par noyau et les pondérations proportionnelles à d_{ij}^{-2} pour la méthode de l'autocorrélation spatiale (voir le rapport intermédiaire pour plus de détails, Flahaut 1999).

5.4 Dissemblances entre les deux méthodes

Nous avons montré l'existence de correspondances significatives entre les résultats des deux méthodes. Cette stabilité observée permet d'étayer la méthode de l'autocorrélation spatiale par des justifications fondées sur des éléments statistiques, et sa crédibilité en est par là renforcée. Cependant, il existe des zones noires non communes aux deux méthodes. En effet, la méthode du noyau attribue un indice de dangerosité à chaque hectomètre, ce qui contribue à considérer la moitié des observations comme plus ou moins sûres, et l'autre moitié comme plus ou moins dangereuses. Selon la méthode du noyau, 263 hectomètres

appartiennent donc à une zone noire. Par contre, la méthode basée sur l'autocorrélation spatiale ne relève que 75 centres de zones noires, dont 74 sont communs aux hectomètres dangereux selon la méthode par noyau. Comment expliquer ces différences ?

Tout d'abord, si la méthode basée sur l'autocorrélation spatiale ne détermine en effet que 75 centres de zones, ces zones noires sont composées non seulement d'un centre mais aussi des hectomètres voisins du centre, dont le nombre dépend de la longueur optimale de chaque zone selon la structure spatiale locale observée. Cela concerne 78 hectomètres, compte tenu du fait que certaines zones noires se recouvrent, en tout ou en partie.

Ensuite, comme la méthode par noyau est basée sur la construction d'un indice à partir des accidents dans une fenêtre fixe de 500 mètres, la détermination de ces indices reste tributaire de cette distance fixe même si celle-ci, n'est pas optimalement adaptée à la structure spatiale locale. Lorsque la structure spatiale locale est plus fine que 500 mètres, l'autocorrélation spatiale permet d'affiner l'identification des zones noires, de mieux les délimiter. Lorsqu'ils ne sont pas compris dans les zones noires, les hectomètres non communs aux deux méthodes peuvent souvent être des *imprécisions* de la méthode du noyau dues à sa rigidité au niveau de la taille de la fenêtre.

Enfin, une particularité supplémentaire intéressante propre à la méthode basée sur l'autocorrélation spatiale est qu'un même hectomètre peut appartenir à plusieurs zones noires, dont la longueur et/ou l'intensité du caractère dangereux diffère.

6. Comparaison avec les zones rouges du *MET*

Le Ministère Wallon de l'Équipement et des Transports (le *MET*) propose aussi une méthode d'identification des zones dangereuses (les zones rouges), qui apparaît peut-être davantage intuitive (absence de justification statistique de la valeur des différents paramètres intervenant dans le calcul de l'indice de dangerosité) (Romano et Heuchenne, 1996). Pour calculer l'indice de dangerosité, cette méthode tient compte de l'occurrence des accidents sur une période de 11 mois et sur une longueur fixe de 500 mètres en accordant une importance décroissante aux accidents les plus éloignés dans le temps et dans l'espace (par des courbes paraboliques). Voici une comparaison des résultats obtenus par la méthode de l'autocorrélation spatiale avec les résultats des zones rouges du *MET*, toujours pour la *N29*. Elle consiste aussi en mesures de corrélation entre les valeurs des indices de dangerosité.

En ce qui concerne les résultats des zones rouges communiqués par le *MET*, un indice de dangerosité a été calculé pour la période de 1992 à 1996, soit la même période que pour l'étude basée sur l'autocorrélation spatiale. Comme la méthode par noyau, la méthode du *MET* attribue un indice de dangerosité à chaque hectomètre de route, ce qui n'est pas le cas de la méthode par autocorrélation spatiale. De même, la longueur des zones est fixe, mais elle

est plus courte et est donc adaptée à une étude locale de dangerosité ; à ce point de vue, elle est donc comparable avec la méthode de l'autocorrélation spatiale.

Que ce soit en adaptant la longueur des zones noires selon la structure spatiale locale ou en considérant une longueur fixe de 500 mètres pour la méthode de l'autocorrélation spatiale, on peut remarquer une bonne correspondance des résultats en ce qui concerne les zones dangereuses identifiées par les deux méthodes (Pearson respectivement égal à 0.85 et 0.91). Il est intéressant de constater qu'une méthode basée sur des éléments plus intuitifs comme celle du *MET* donne des résultats comparables à ceux obtenus par la méthode basée sur l'autocorrélation spatiale. Grâce à ces éléments de comparaison, la méthode des zones rouges du *MET* en sort également renforcée.

Figure 6.1 Relation entre l'indice de dangerosité basé sur l'autocorrélation spatiale avec une longueur de zone variable (IDCOUNT) et l'indice de dangerosité du *MET*.

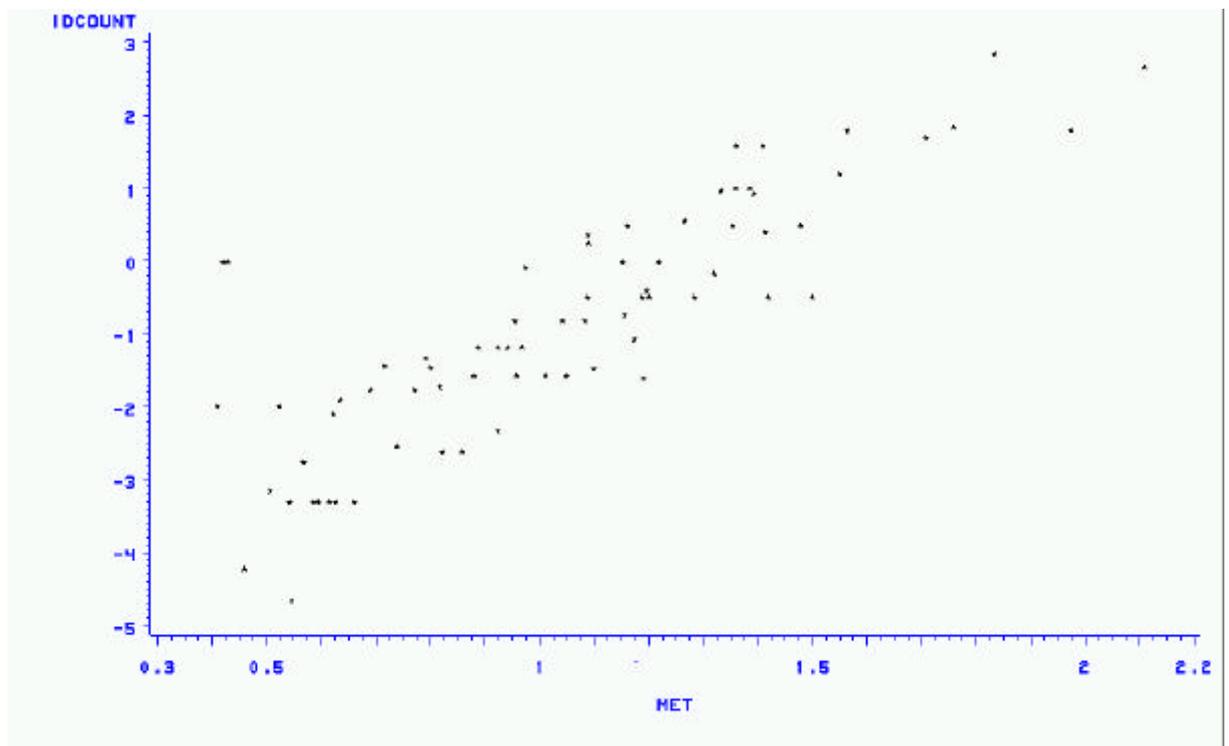
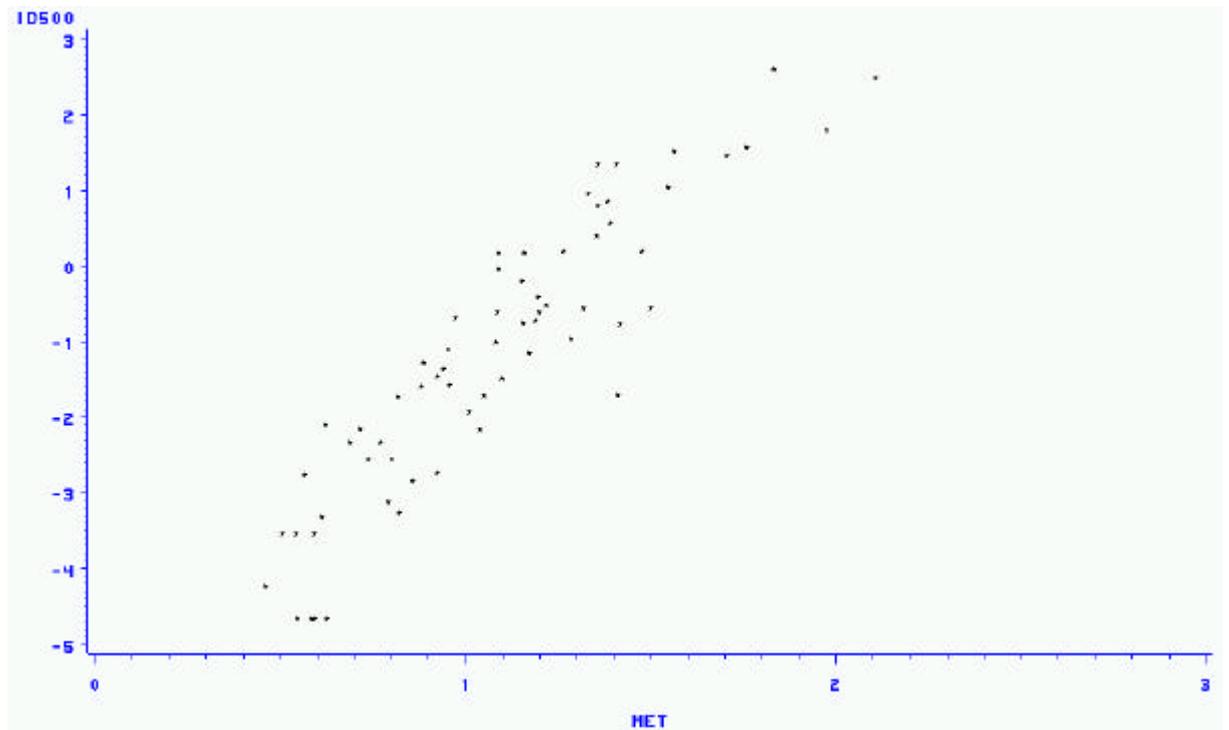


Figure 6.2 Relation entre l'indice de dangerosité basé sur l'autocorrélation spatiale avec une longueur de zone fixe de 500 mètres (IDCOUNT) et l'indice de dangerosité du MET.



7. Conclusions

Comme nous l'avons déjà souligné, la comparaison des résultats obtenus avec les deux méthodes utilisées pour identifier les zones dangereuses du réseau routier (par noyau et par autocorrélation spatiale locale) a permis de montrer l'existence d'une correspondance significative entre leurs résultats lorsque l'on attribue aux différents paramètres des valeurs permettant cette comparaison, que ce soit en considérant la valeur des indices ou le classement de ces indices en différentes catégories suivant l'intensité du caractère dangereux. Cette stabilité observée permet d'étayer la méthode de l'autocorrélation spatiale par des justifications fondées sur des éléments statistiques, et sa crédibilité en est par là renforcée.

En outre, les observations relatives aux zones noires non communes aux 2 méthodes permettent de considérer que la méthode d'identification des zones noires basée sur l'autocorrélation spatiale locale est plus fine que celle par noyau, car elle permet de s'adapter à la structure spatiale locale des accidents lorsque cela est nécessaire. Les avantages de la méthode basée sur des mesures locales de l'autocorrélation spatiale sont d'identifier les zones dangereuses en se basant sur la structure spatiale locale de chaque zone. En effet, outre les relations de voisinage et de proximité, ce sont des mesures d'association/dépendance spatiale entre le nombre d'accidents aux différents hectomètres de route qui permettent de localiser les

zones noires. Du fait de cette adaptation aux particularités locales, il n'est pas possible de déterminer LA longueur optimale d'une zone noire. S'il fallait cependant parler en ces termes, on peut observer que ce sont les zones noires de 300 mètres les plus nombreuses et que la longueur moyenne des zones noires en Belgique varie entre 545 et 648 mètres, selon qu'on se situe en Flandre ou en Wallonie et sur routes nationales ou sur autoroutes.

Sur les hectomètres appartenant à une zone noire selon l'une ou l'autre méthode, il faut encore insister sur le fait que tous ne doivent pas être considérés comme des endroits nécessitant une intervention visant à diminuer leur caractère dangereux : parmi eux, certains sont dangereux puisqu'ils ont été identifiés comme tels, mais ils le sont peu. Dès lors, il importe de classer les zones noires selon l'intensité de leur caractère dangereux afin de s'intéresser plus particulièrement à ceux qui le sont vraiment.

Un approfondissement qu'il serait intéressant de développer à propos de l'utilisation d'indices d'autocorrélation spatiale comme indicateurs de dangerosité concerne la complexification de la fonction de décroissance des pondérations (utilisation d'une fonction gaussienne par exemple). De plus, l'utilisation d'autres mesures d'association serait extrêmement intéressante en vue d'une évaluation de la stabilité des résultats. Il serait en outre intéressant de tenter d'étendre la méthode à un réseau dans son ensemble et sa continuité et non plus en considérant chaque route de façon séparée.

Après cette première partie dans laquelle nous avons développé une méthode d'identification des lieux dangereux du réseau routier, la deuxième partie du projet consiste à construire un modèle explicatif de la localisation de ces endroits, autrement dit de la co-occurrence spatiale des accidents de la route.

DEUXIEME PARTIE :

CONSTRUCTION D'UN MODELE EXPLICATIF

1. Introduction

Dans cette deuxième partie, il s'agit de se baser sur les résultats obtenus précédemment (identification des zones noires, première partie) de façon à tenter de construire un modèle explicatif de la co-occurrence spatiale des accidents de la route et de la localisation des zones noires.

Une part importante de la recherche internationale en accidentologie est consacrée à la description et à des tentatives d'explication de l'occurrence des accidents de la route. A cette problématique générale correspond un champ assez vaste et diversifié de centres d'intérêts, selon l'approche principale privilégiée par les chercheurs. Cela va des aspects juridiques et légaux aux aspects techniques des véhicules et des infrastructures, en passant par les aspects psychologiques, comportementaux et socio-économiques des usagers de la route. Certaines recherches sont davantage descriptives, tandis que d'autres sont axées sur une méthodologie statistique plus forte, avec un objectif explicatif et/ou prévisionniste.

L'objectif de cette deuxième partie est la construction d'un modèle explicatif de la co-occurrence spatiale des accidents de la route et de la localisation des zones noires. Il s'agit d'essayer de répondre aux questions suivantes :

- La co-occurrence des accidents de la route est-elle explicable *spatialement* ?
- Y a-t-il un lien entre concentrations d'accidents et environnement spatial ?
- Quels seraient les facteurs spatiaux explicatifs ?

L'approche choisie est donc une approche à caractère spatial, nous cherchons à expliquer pourquoi les accidents se concentrent à certains endroits du réseau et pas ailleurs. Les variables explicatives devront forcément être reliées à ces lieux de concentration, ce qui nécessite qu'elles présentent une dimension spatiale. Dès lors, d'autres types de variables à caractère non spatial (socio-démo-économique, comportemental, technique ou autre) ne seront

pas introduits dans l'explication. En outre, il faut aussi être conscient que nous ne pourrons prendre en compte qu'une partie de l'ensemble potentiel des variables à caractère spatial, ce qui rend aussi possible une éventuelle spécification incomplète de la modélisation.

Le modèle construit devra être capable de mettre en évidence l'impact des variables d'aménagement et de développement spatial sur la sécurité et l'insécurité routière, de façon à essayer de comprendre pourquoi et comment l'environnement local influence la co-occurrence spatiale d'accidents de la route. Afin de répondre à ces objectifs, des modèles régressifs de type logistique sont utilisés ; ils seront appliqués au réseau routier de la province du Brabant Wallon.

2. Méthode d'analyse : la régression logistique

2.1 La régression logistique

L'objectif d'une régression est d'expliquer la distribution (spatiale dans ce cas-ci) d'un phénomène par celle d'un ou de plusieurs autres. La régression implique une relation de dépendance entre les variables : il y a une variable à expliquer (Y , variable dépendante) et une ou plusieurs variables explicatives (X , variables indépendantes). Dans ce travail, il s'agit de tenter de mettre en évidence l'impact de variables d'environnement local (= les variables explicatives) sur la localisation des zones noires du réseau routier (= la variable à expliquer). Une analyse de régression permet de rencontrer cet objectif.

La régression linéaire (simple ou multiple) est souvent utilisée lorsque les variables intervenant dans la construction du modèle sont toutes de type cardinal. La formulation mathématique d'un modèle de régression linéaire est le suivant (en ne considérant, pour simplifier, qu'une seule variable explicative) :

$$Y = \hat{a} + \hat{a}X + \hat{a} \quad [1]$$

Y	est la variable dépendante
X	est la variable indépendante
\hat{a}	est le terme d'erreur
\hat{a} et \hat{a}	sont les paramètres du modèle, à estimer.

Lorsque toutes les conditions de la régression linéaire sont remplies, on peut aussi écrire que :

$$E(Y) = E(\hat{a} + \hat{a}X + \hat{a}) = \hat{a} + \hat{a}X \quad [2]$$

Et pour une valeur x_i donnée de la variable indépendante X (i est le lieu où cette valeur x_i est observée) :

$$E(Y/x_i) = \hat{a} + \hat{a}x_i \quad [3]$$

où l'expression $E(Y/x_i)$ doit être lue comme étant la valeur attendue de Y étant donnée la valeur x_i .

L'expression [3] implique que $E(Y/x_i)$ peut prendre n'importe quelle valeur comprise entre - et + , dès lors que x_i peut varier dans ce même intervalle.

Lorsque la variable dépendante Y est *discrète*, les conditions de la régression linéaire ne sont plus respectées et il faut modifier le modèle. C'est là qu'interviennent les modèles régressifs de type *logistique*. Dans ce travail, la variable à expliquer est effectivement une variable discrète : dans la première partie, nous avons identifié pour chaque hectomètre de route s'il se situe ou non dans une zone de concentration spatiale d'accidents (autrement dit, dans une zone noire). Dès lors, la variable dépendante est dichotomique : elle prend la valeur 1 si l'hectomètre appartient à une zone noire, la valeur 0 autrement. Cela correspond au cas le plus simple de variable discrète, lorsqu'elle ne peut prendre que 2 valeurs. C'est cette situation qui est développée dans cette section 2, le cas d'une variable polychotomique étant une extension de ce cas dichotomique.

Le fait que la variable dépendante soit discrète et ne puisse prendre qu'un nombre limité de valeurs est à l'origine de violations des hypothèses sous-jacentes à l'utilisation d'une régression linéaire classique. Une autre forme de modèle régressif doit être utilisée pour ce type de variable. Parmi les fonctions de distribution qui ont été proposées pour l'analyse de variables dichotomiques, c'est la distribution logistique qui est le plus souvent retenue, principalement pour deux raisons (Cox, 1970) : (1) d'un point de vue mathématique c'est une fonction très flexible et relativement facile à l'emploi, et (2) elle se prête facilement à une interprétation qui a du sens scientifiquement. La fonction $E(Y/x)$ est donc soumise à une transformation, la transformation LOGIT.

La forme d'un modèle régressif de type logistique est la suivante :

$$E(Y/x) = \frac{e^{a+bx}}{1 + e^{a+bx}} \quad [4]$$

La transformation logit est appliquée à $E(Y/x)$:

$$t(x) = \ln \frac{E(Y/x)}{1-E(Y/x)} = \hat{a} + \hat{a}x \quad [5]$$

Cette transformation logit $t(x)$ possède une bonne partie des propriétés intéressantes d'un modèle régressif linéaire : elle est linéaire en ses paramètres, peut être continue, et peut varier entre - et + suivant les valeurs que peut prendre la variable X .

Pour davantage de détails relatifs à cette transformation et à la régression logistique de façon générale, nous renvoyons le lecteur à Wrigley (1976 et 1985), et à Hosmer et Lemeshow (1989).

2.2 L'ajustement du modèle de régression

L'ajustement d'un modèle à l'ensemble de données observées consiste à estimer la valeur des paramètres \hat{a} et \hat{a} dans l'équation [5], qui sont inconnus au départ. C'est la méthode du *maximum de vraisemblance* qui est utilisée pour les régressions logistiques. Sans entrer dans les détails, cette méthode permet d'estimer la valeur des paramètres inconnus de façon à maximiser la probabilité d'obtenir l'ensemble de données observées. Il s'agit donc d'abord de construire une fonction de vraisemblance, exprimant la probabilité d'observer les données comme une fonction des paramètres inconnus. Les procédures de résolution de ces fonctions sont programmées dans les logiciels qui proposent des régressions logistiques. C'est le logiciel de traitements statistiques *SAS* que nous avons utilisé dans ce travail. Pour plus de détails et la formulation mathématique de ces fonctions, nous renvoyons à Hosmer et Lemeshow (1989).

Une fois que les coefficients ont été estimés, il importe aussi de tester leur degré global de significativité, de façon à évaluer si l'ensemble des variables indépendantes introduites dans le modèle a une incidence significative sur la variable dépendante. On compare donc l'apport du modèle qui intègre la (les) variable(s) concernée(s) avec un modèle qui ne l'(les) intègre pas. Le test est basé sur la fonction de maximum de vraisemblance utilisée pour estimer les paramètres du modèle. Dans le cas multivarié (plusieurs variables indépendantes), c'est la significativité du modèle dans sa globalité qui est testée. Il est possible de dire si le

coefficient *d'au moins une* des variables introduites est différent de zéro, et peut-être tous. Le test est basé sur une distribution de Chi-carré. Dans le cas univarié, cela correspond aussi à la significativité de la seule variable introduite. De nouveau, voir Hosmer et Lemeshow (1989) pour davantage d'explications.

Dans le cas d'un modèle multivarié, il est aussi possible d'évaluer le degré de significativité de chacune des variables introduites dans le modèle, par exemple par le test de Wald qui suit aussi une distribution de Chi-carré (voir section 2.5). Ce test permet de préciser quelle(s) variable(s) introduite(s) dans le modèle n'est (ne sont) éventuellement pas significative(s). Ajoutons tout de même que ces tests et statistiques ne doivent pas être les seuls critères à prendre en compte pour faire entrer ou sortir une variable dans le modèle, des interprétations des variables d'un point de vue scientifique doivent aussi être prises en compte.

2.3 Construction du modèle de régression

L'adéquation du modèle et des variables est évaluée d'un point de vue statistique, mais doit aussi s'intégrer dans le contexte scientifique du problème étudié. De façon traditionnelle, l'objectif est de proposer un modèle qui soit le meilleur possible avec le moins de variables possible, de façon à ce qu'il soit aisément généralisable. Lorsqu'un grand nombre de variables sont introduites dans un modèle, celui-ci sera probablement très bien adapté, mais il sera également très dépendant des données observées. Il existe en effet un risque réel lié à ce type d'approche visant à introduire le plus de variables possible, qui est d'obtenir un modèle trop précisément ajusté qui produirait alors des coefficients estimés instables lorsque le modèle est appliqué à un autre ensemble de données.

La construction d'un modèle multivarié pourrait être abordée en y introduisant les variables suffisamment significatives (suivant un seuil fixé) dans des analyses univariées préalables, tout en considérant aussi la pertinence des variables disponibles dans le contexte scientifique de l'étude.

Si l'on se base sur des justifications purement statistiques, les méthodes pas à pas (stepwise) sont utilisées. La première démarche (forward) consiste à introduire la variable qui améliore le modèle de façon la plus significative (selon un seuil de 0.25 par exemple), et ainsi de suite jusqu'à ce qu'aucune variable ne soit suffisamment significative par rapport au seuil de probabilité fixé. Suivant le même principe, la démarche inverse (backward) consiste à construire d'abord un modèle avec toutes les variables, et ensuite enlever progressivement celles qui n'apportent pas suffisamment au modèle. C'est cette démarche que nous avons appliquée, et dont les résultats sont présentés dans la section 6. D'autres méthodes moins utilisées consistent à intégrer un nombre donné de variables dans le modèle, et de choisir la configuration optimale suivant un (des) critère(s) spécifié(s). C'est la méthode de sélection des meilleurs sous-ensembles.

Que l'on opte pour un processus de construction ou un autre, il importe à chaque étape de vérifier la contribution de chaque variable introduite, de la comparer à sa contribution dans un modèle univarié, et également d'examiner l'ajustement du modèle dans sa globalité.

2.4 Interprétation des coefficients de la régression

On trouve dans les résultats de la régression logistique la valeur des paramètres \hat{a} de cette régression. A partir de ces paramètres (un pour chaque variable indépendante), on peut calculer ce qu'on appelle les "*odds ratios*", qui sont aussi donnés dans les résultats de la régression. Le odds ratio peut être interprété comme une mesure d'association entre la variable dépendante et la variable indépendante considérée. Il s'agit en fait d'un rapport de proportions. Mais avant d'aller plus loin, il faut introduire une distinction entre les variables explicatives dichotomiques et les variables explicatives polychotomiques. Ces deux types de variables sont présentes parmi les variables environnementales que nous considérons dans l'application présentée dans les sections 4 et 5.

2.4.1 Une variable explicative dichotomique

Prenons l'exemple simple d'une variable dépendante Y et d'une seule variable indépendante X , toutes deux dichotomiques. Dans ce cas, il est possible de construire un tableau de contingence permettant de calculer les diverses proportions pour les 4 cas de figures possibles, suivant la valeur prise par chacune des 2 variables.

	$X = 1$	$X = 0$
$Y = 1$	$P_{X=1, Y=1}$	$P_{X=0, Y=1}$
$Y = 0$	$P_{X=1, Y=0}$	$P_{X=0, Y=0}$

Ce qu'on appelle le "*odds*" pour la valeur 1 de la variable dichotomique explicative X est la probabilité pour un individu d'obtenir la valeur 1 pour Y lorsqu'il présente la valeur 1 pour X . De façon mathématique, cette probabilité s'écrit comme $P_{X=1, Y=1} / P_{X=1, Y=0}$. De la même façon, la probabilité (*odds*) d'obtenir $Y = 1$ parmi les individus qui présentent $X = 0$ s'écrit $P_{X=0, Y=1} / P_{X=0, Y=0}$. Le ***odds ratio*** est le rapport entre ces 2 probabilités, la probabilité pour $X = 1$ par rapport à la probabilité pour $X = 0$. De façon mathématique, il s'écrit :

$$OR = \frac{P_{X=1, Y=1} / P_{X=1, Y=0}}{P_{X=0, Y=1} / P_{X=0, Y=0}} \quad [6]$$

L'interprétation est la suivante : le odds ratio permet d'estimer de combien il est plus probable (ou improbable) d'obtenir $Y = 1$, si $X = 1$ et non 0. Il s'agit donc d'une MESURE COMPARATIVE entre les 2 probabilités.

Le odds ratio peut aussi être estimé de la façon suivante : si \hat{a} est l'estimation du coefficient de la variable X dans l'équation de la régression, et que la variable X est codée 0 ou 1, alors :

$$OR = e^{\hat{a}} \quad [7]$$

C'est le odds ratio qui est le paramètre habituellement considéré comme le plus intéressant dans les résultats d'une régression logistique. On peut aussi trouver dans les résultats de la régression un intervalle de confiance pour ce paramètre.

2.4.2 Une variable explicative polychotomique

Cette fois, la variable indépendante X se présente sous la forme d'un nombre de catégories strictement supérieur à 2, dont l'appartenance à l'une d'entre elles est mutuellement exclusive. Dans le cas d'une variable nominale, il ne convient pas de l'introduire telle quelle dans le modèle comme on pourrait le faire pour une variable ordinale. Il faut la transformer en plusieurs "dummy variables" représentant chacune une catégorie de la variable de départ. Prenons l'exemple d'une variable X qui se décomposerait en 4 catégories et qui pourrait alors prendre 4 valeurs : 1, 2, 3 ou 4. L'idée est de créer 3 nouvelles variables (dummy variables) à partir de celle d'origine (soit 1 de moins que le nombre de classes de départ) : X_1 , X_2 et X_3 . La variable X_1 prendra la valeur 1 si l'observation appartient à la 1^{ère} catégorie, et 0 autrement ; la variable X_2 prendra la valeur 1 si l'observation appartient à la 2^{ème} catégorie, et 0 autrement ; la variable X_3 prendra la valeur 1 si l'observation appartient à la 3^{ème} catégorie de route, et 0 autrement. Il n'est pas utile de créer une 4^{ème} variable qui serait une combinaison linéaire des 3 premières et perturberait alors le modèle.

Du fait de l'introduction dans le modèle régressif de 3 variables et non 4 (de façon générale $N-1$ si N représente le nombre de catégories de X), l'interprétation des odds ratios se fait par rapport à la catégorie non introduite dans le modèle. Si l'on souhaite garder toute l'information de la variable de départ X , il importe donc de ne pas enlever du modèle l'une des $N-1$ dummy variables même si elle n'est pas jugée significative lors de son ajustement.

L'estimation des odds ratios se fait de la même manière que pour une variable dichotomique (pour X_I , $OR = e^{b_I}$), de même que l'estimation de l'intervalle de confiance.

En ce qui concerne une variable explicative qui serait continue, cela ne pose pas de problème de l'intégrer telle quelle à condition que la fonction logit soit linéaire pour cette variable. Si ce n'est pas le cas (et c'est ce qui se passe la plupart du temps en pratique), il faut aussi la transformer en dummy variables. C'est ce qui a été fait pour la variable d'intensité de trafic, introduite dans le modèle explicatif développé à la section 4.

2.4.3 Plusieurs variables explicatives

La plupart du temps, plusieurs variables indépendantes sont introduites dans le modèle de régression, de façon à tenter de mieux cerner la variable à expliquer. Cela implique *d'ajuster statistiquement* les effets estimés de chaque variable (les odds ratios), en tenant compte de la présence et de l'effet des autres variables incluses dans le modèle. Le odds ratio pour une variable discrète X_I s'interprète alors comme celui que l'on s'attendrait à obtenir avec un modèle univarié si, pour chaque valeur de cette variable X_I , la distribution de la (des) autre(s) variable(s) indépendante(s) est identique. Pratiquement, les odds ratios ajustés sont estimés en comparant les individus pour lesquels seule la variable considérée X_I diffère, mais pas les autres variables pour lesquelles ces individus présentent la même valeur. Les odds ratios donnent donc l'effet attendu de chaque variable toutes choses égales par ailleurs. On considère donc qu'il n'y a pas d'interaction entre les variables.

2.5 Evaluation de la qualité de l'ajustement (goodness-of-fit) du modèle de régression

Après avoir introduit les variables pertinentes dans le modèle, il s'agit d'évaluer l'efficacité de ce modèle à expliquer la variable dépendante. L'idée de cette évaluation est de se baser sur des mesures globales et individuelles de l'écart entre la (les) valeur(s) prédite(s) de Y par le modèle et sa (ses) valeur(s) observée(s). Ces écarts sont appelés les résidus d'une régression.

Parmi les mesures globales, retenons-en principalement trois : la déviance, le Chi-carré de Pearson, et les tests de Hosmer et Lemeshow, qui peuvent tous les trois être approximés par une distribution de Chi-carré. Pour chaque mesure, l'hypothèse nulle est que le modèle est bien ajusté. Si ces mesures ont une valeur plus grande que celle liée au seuil de probabilité choisi pour accepter ou rejeter l'hypothèse nulle, c'est-à-dire qu'il n'est pas possible de rejeter cette hypothèse nulle, c'est qu'il est probable que le modèle n'intègre pas tous les facteurs explicatifs de la variation des données. D'un autre côté, si trop de variables sont introduites dans le modèle, la qualité de l'ajustement peut être très élevée, mais le modèle risque alors d'être trop dépendant des données et difficilement généralisable par extension à d'autres

ensembles de données. Nous avons donc choisi pour cette application, dans le but d'éviter une telle instabilité des coefficients des variables constitutives du modèle, de se limiter aux variables qui apportent de l'information significative à 95% (à 90% dans l'un ou l'autre cas) même si, comme nous l'avons observé, l'introduction de davantage de variables (moins significatives) permettrait d'obtenir un meilleur ajustement du modèle de régression.

En ce qui concerne les mesures individuelles, la démarche consiste à représenter graphiquement la relation entre la probabilité estimée par le modèle pour chaque individu, et une mesure basée sur le résidu de chacun de ces individus. On décompose donc ici une mesure globale en ses composants individuels. Une mesure de la déviance, de même qu'un résidu de Pearson peuvent être calculés pour chaque observation. Ces mesures sont utiles pour identifier les individus qui sont mal ajustés par le modèle.

Pour les formules de tous ces indices, nous conseillons de lire Hosmer et Lemeshow (1989), Stokes *et al.* (1996), et le manuel de SAS (1989).

3. Zone d'étude et variables considérées

3.1 Le Brabant wallon

L'échelle d'analyse adoptée pour tenter de mettre en évidence l'impact des variables d'aménagement et de développement spatial sur la sécurité et l'insécurité routière, est une échelle de niveau "macro" : on s'intéresse ici aux routes de la province du Brabant Wallon. La surface d'étude est réduite par rapport à l'ensemble de la Belgique ou de la Wallonie, de façon à faciliter les traitements des données ainsi que dans le but de pouvoir mieux identifier les facteurs explicatifs. En effet, cette province a l'avantage de présenter une large diversité parmi les variables considérées dans cette étude. Le réseau routier du Brabant Wallon compte 460km de routes numérotées, parmi lesquels 1202hm appartiennent à une zone noire (soit 26% du réseau numéroté de la province). Ces zones noires regroupent 2600 accidents qui ont eu lieu durant la période de 1922 à 1996 (soit 5 ans), sur 4078 accidents pour tous les hectomètres (soit 64% des accidents sur les routes numérotées de la province). Chaque hectomètre appartenant à une zone noire présente donc en moyenne 2,08 accidents. Les hectomètres n'appartenant pas à une zone noire (au nombre de 3401) présentent en moyenne 0,43 accident par hectomètre.

Tableau 3.1. Répartition des accidents pour les hectomètres routiers numérotés du Brabant Wallon.

	Zones noires	Zones non noires	Total
Routes numérotées du Brabant Wallon	1202hm (26%)	3401hm (74%)	4603hm
Nombre d'accidents (1992-1996)	2600 (64%)	1478 (36%)	4078
Nbre moyen d'acc/hm	2,08	0,43	0,88

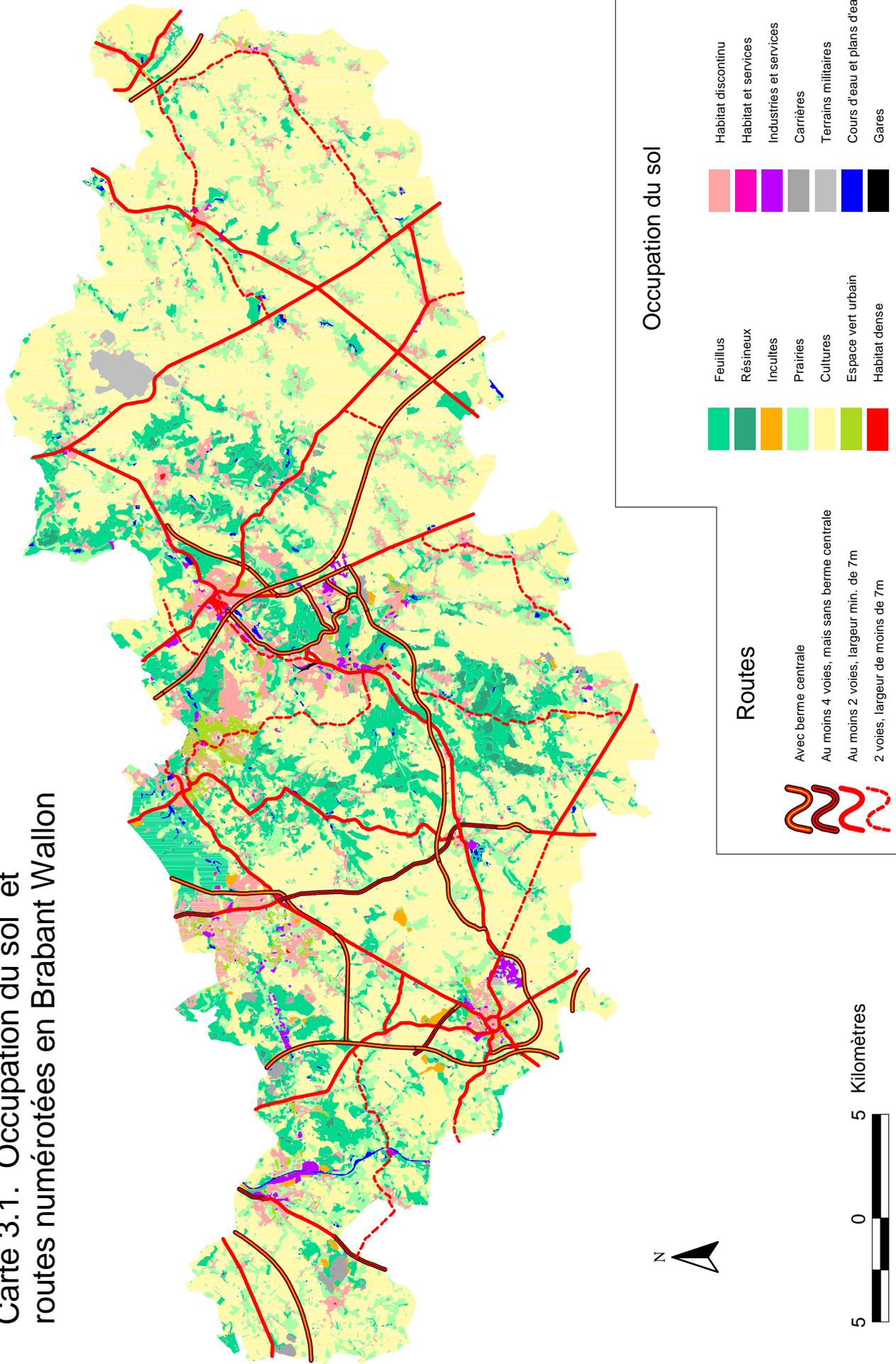
Le choix d'un niveau spatial d'analyse particulier (ici le niveau d'une province) conditionne l'ensemble des résultats obtenus. Il sera intéressant par la suite de pouvoir comparer les résultats obtenus à partir de cette approche à petite échelle avec ceux obtenus à une échelle plus grande, comme ont été abordées les analyses menées par l'*IBSR* (voir tâche B de ce projet : diagnostic en profondeur des zones noires).

Le Brabant wallon peut être considéré comme une province périurbaine et rurale ; même s'il n'y a pas vraiment de grande ville, il existe tout de même plusieurs centres importants à l'échelle de la région : Wavre, Ottignies-Louvain-la-Neuve, Nivelles, Tubize, Waterloo... Si l'ouest de la province présente un visage plus industrialisé, le nord et le centre sont plus orientés, d'un point de vue fonctionnel, vers du périurbain, tandis que l'est de la province peut être qualifié de milieu plus ouvert principalement consacré à l'agriculture. La carte de la page suivante (carte 3.1) illustre cette occupation du sol en Brabant wallon, et reprend également le réseau routier numéroté de la province.

3.2 La variable dépendante (à expliquer)

Nous prenons comme données de départ pour cette étude les résultats obtenus dans la première partie : la localisation des zones noires. A partir de ces résultats, nous pouvons donc connaître l'appartenance ou la non-appartenance d'un hectomètre à une zone noire. Dès lors, la variable dépendante est une variable dichotomique, prenant la valeur 1 si l'hectomètre appartient à une zone noire, la valeur 0 autrement. Puisque cette variable à expliquer est discrète, c'est donc une régression logistique qui va être utilisée (voir section 2.1).

Carte 3.1. Occupation du sol et routes numérotées en Brabant Wallon



Source : Région Wallonne

3.3 Les variables indépendantes (explicatives)

Nous disposons pour cette étude de plusieurs variables à caractère local, environnemental et spatial, qui pourraient avoir une influence sur la répartition spatiale des accidents de la route. Une meilleure compréhension des facteurs *géographiques et spatiaux* influençant la sécurité routière devrait permettre d'un point de vue pratique et appliqué, d'exercer une action concrète en aménagement du territoire en identifiant les structures environnementales et spatiales liées à la sécurité et à l'insécurité routières. Ces variables d'environnement sont de plusieurs types. Tout d'abord, des variables que l'on appellerait des variables d'environnement immédiat sont disponibles : elles concernent le design et la géométrie de la route, ainsi que sa fonctionnalité. Concrètement, il s'agit des variables suivantes : le **type de route** d'un point de vue "physique", divisée en 4 classes ; la **vitesse maximale autorisée**, divisée en 4 classes, et **l'intensité du trafic**, divisé en 5 classes. Un deuxième groupe de variables représentant l'environnement local, immédiat de la route, est également disponible. Concrètement, il s'agit des variables suivantes : le type **d'occupation du sol**, divisée à l'origine en 14 classes, et une variable **d'appartenance ou de non-appartenance à une agglomération**. L'ensemble de ces variables représente les caractéristiques de l'infrastructure routière, de la hiérarchie des routes et de leur environnement local. Ces variables sont les variables de base, à partir desquelles d'autres variables ont été construites. Il s'agit en fait de nouvelles variables indiquant la proximité d'une **discontinuité spatiale** des variables initiales.

Les variables de vitesse maximale autorisée et d'appartenance ou non à une agglomération (déterminée par les panneaux F1 et F3) proviennent de la base de données "accidents" du MET, et des interpolations ont dû être effectuées pour caractériser les hectomètres ne présentant pas d'accidents et donc non repris dans cette base de données. Les données de trafic ont également été fournies par le MET et ne sont pas tout à fait complètes, ce qui a aussi nécessité quelques interpolations. Quant aux données d'occupation du sol, elles proviennent de la Région wallonne : la carte d'occupation du sol de la Région wallonne a été réalisée par traitement numérique de données issues de la télédétection spatiale (satellites LANDSAT TM et SPOT XS). Elle est une représentation de la situation existante de l'occupation du sol de la Wallonie (en 1989), et rend compte de l'emprise spatiale des diverses activités de l'aspect paysager de la région. La distinction des types de routes provient d'une base de données belnet.

Après cette section introductive à l'application, les sections suivantes sont consacrées à l'application proprement dite. Nous proposons d'abord un résumé des analyses catégorielles qui sont développées dans le rapport intermédiaire complet (Flahaut, 2000 ; dans ce rapport, des analyses univariées sont aussi présentées et sont complétées d'une interprétation des résultats obtenus). La section 4 qui suit porte donc sur l'influence de chaque catégorie de

chaque variable sur la probabilité pour un hectomètre d'appartenir à une zone noire. Dans la section 5 ensuite, les variables indépendantes sont intégrées ensemble dans des modèles régressifs multivariés, avant de conclure dans la section 6.

4. Tableaux de contingence catégoriels : résumé

Certaines variables se retrouvent systématiquement (ou presque) en tête du classement, indépendamment de la stratification, en gardant cependant à l'esprit que certaines variables ont moins de sens pour certains types de route que pour d'autres.

Un *trafic élevé* augmente la probabilité de zone noire, de même qu'un environnement *urbain* (qui correspond en fait davantage à du bâti au sens large, puisqu'il intègre aussi l'habitat discontinu, voir Flahaut 2000). C'est aussi le cas, dans une moindre mesure, d'un environnement *boisé*. Les variables de *transition de limitation de vitesse* vont aussi dans le même sens, de même que *l'appartenance à une agglomération* (à relier probablement, au moins en partie, à l'occupation du sol de type bâti/urbain).

Quant aux variables qui agissent largement dans le sens d'une diminution du risque de zone noire, relevons d'abord un *faible trafic*. Un environnement de type *rural* (champs et prairies principalement) est également souvent important.

5. Construction d'un modèle multivarié

Suite à ces analyses uni-catégorielles et à des analyses univariées préalables (ces dernières sont aussi développées dans Flahaut, 2000), il s'agit maintenant de construire un modèle complet qui va intégrer *ensemble* les effets de ces variables. Il sera aussi intéressant d'examiner ici si l'influence des variables dans un contexte multivarié va dans le même sens que celle qui a été identifiée dans les régressions univariées.

La méthode choisie est une méthode de construction pas à pas, en partant d'un modèle qui intègre toutes les variables disponibles duquel on retire celles qui n'apportent pas d'information significative dans l'explication de la variable dépendante. Cette procédure est souvent appelée "backward".

5.1 Tous types de routes confondus

Un modèle a été construit qui intègre au départ toutes les variables disponibles. Après avoir enlevé successivement les variables non significatives, nous arrivons à un modèle multivarié composé de 7 variables : l'occupation du sol, le type de route, l'intensité du trafic, la vitesse maximale autorisée, la proximité d'une transition à un type de route différent (à

moins de 750 mètres), l'appartenance à une agglomération, et la proximité d'une transition à une limitation de vitesse différente (à moins de 250m). Ce modèle retenu ne permet pas d'obtenir un ajustement d'une qualité significative, si l'on se réfère à la valeur des statistiques de "goodness-of-fit" (la déviance et le chi-carré de Pearson). Pour obtenir un modèle mieux ajusté (c'est-à-dire une meilleure cohérence entre les estimateurs du modèle et les observations, voir section 2.5), il aurait fallu garder dans le modèle des variables non significatives individuellement. Nous avons fait le choix ici de présenter un modèle significatif dans son ensemble, et composé exclusivement de variables significatives individuellement non seulement parce que ce sont bien entendu les plus intéressantes, mais aussi afin d'éviter une trop grande dépendance des paramètres estimés à cet ensemble spécifique de données.

Comme certaines de ces variables sont polychotomiques, elles sont en fait intégrées dans le modèle sous la forme de plusieurs variables dichotomiques représentant les catégories de la variable de départ. Toutes ces dummy variables doivent être conservées pour pouvoir interpréter les odds ratios correctement (c'est-à-dire par rapport à l'une des catégories de chaque variable), mais elles ne sont pas nécessairement toutes significatives par rapport à la catégorie de référence (qui elle n'est pas introduite dans le modèle pour éviter de la redondance d'information).

L'influence de chaque variable, représentée par le odds ratio correspondant, doit s'interpréter toutes choses égales par ailleurs en ce qui concerne les autres variables introduites dans le modèle. Cela sous-entend donc l'hypothèse que le rôle d'une variable est identique quelle que soit la configuration présentée par les autres variables, en d'autres termes qu'il n'y a pas d'interaction entre les variables constitutives du modèle.

Voyons maintenant quels sont concrètement les résultats donnés par ce modèle multivarié. L'ensemble des variables introduites dans le modèle apporte de l'information significative, puisque le chi-carré associé à l'indice permettant de tester cette significativité est égal à 1037 (14 DF) avec une probabilité de 0.0001. Toutes les variables, considérées individuellement, sont significatives même si ce n'est pas nécessairement le cas de toutes les catégories de chacune. Voici, variable par variable, l'influence de chacune (et de chaque catégorie pour les variables polychotomiques) sur le risque pour hectomètre de route d'appartenir à une zone noire. Rappelons, et c'est important, que le odds ratio s'interprète de la façon suivante pour les variables dichotomiques : ***il permet d'estimer de combien il est plus probable (ou improbable) d'obtenir $Y = 1$, si $X = 1$ et non 0. Il s'agit donc d'une MESURE COMPARATIVE entre les 2 probabilités.*** Pour les variables polychotomiques, la comparaison se fait par rapport à la probabilité de zone noire de la catégorie non introduite dans le modèle.

1. *Le type de route*

Toutes choses égales par ailleurs, le type de route lié à la probabilité la plus élevée de zone noire est le type 3 (routes à 2(-3) voies d'une largeur supérieure à 7 mètres). Par rapport à ce type de route, la probabilité pour un hectomètre d'appartenir à une zone noire est diminuée de moitié s'il appartient à une route de type 1 (routes avec berme centrale), le facteur de risque étant de 0.5 avec un intervalle de confiance à 95% entre [0.337, 0.703]. S'il appartient à une route de type 4 (routes à 2 voies d'une largeur inférieure à 7 mètres), le facteur de risque est aussi égal à 0.5 dans l'intervalle [0.383, 0.716]. Le type de route 2 (routes à 4 voies mais sans berme centrale) est aussi associé à une probabilité plus faible de zone noire que le type 3, le facteur de risque relatif est de 0.7 (significatif à 90%).

Remarquons que dans ce modèle multivarié, l'influence du type de route diffère de celui qui avait été perçu lors de l'analyse univariée où c'est le type 2 qui était le plus dangereux, suivi par le type 1, le type 3 et le type 4. Cela montre bien l'intérêt de considérer aussi l'influence d'une variable dans un contexte multivarié plus large d'une simple régression univariée.

2. *La proximité d'une transition d'un type de route à un autre*

C'est la proximité à moins de 750m qui apporte le plus d'information significative au modèle. Toutes choses égales par ailleurs, cette proximité augmente le risque pour un hectomètre d'appartenir à une zone noire d'un facteur 2.3 ([1.695, 3.059]). Cette évaluation du risque va dans le même sens que celle qui avait été faite pour les analyses univariées, sauf qu'alors c'était la transition à moins de 1000m qui était la plus précise.

3. *L'intensité de trafic (5 classes, suivant les quantiles)*

Toutes choses égales par ailleurs, l'intensité de trafic liée à la probabilité la plus faible de zone noire est la classe de trafic la plus faible. Par rapport à cette classe de trafic, la probabilité pour un hectomètre d'appartenir à une zone noire est augmentée s'il appartient à une classe plus élevée, et ce d'autant plus que l'intensité de trafic est plus grande. Le facteur multiplicatif de risque est de 4.3 pour la 2^{ème} classe de trafic ([3.057, 6.176]), de 5.2 pour la 3^{ème} classe ([3.665, 7.318]), de 7.7 pour la 4^{ème} classe ([5.408, 11.078]) et de 69 pour la 5^{ème} classe ([45.063, 106.246]). Ce dernier chiffre peut sembler énorme, d'autant plus qu'il est associé à un intervalle de confiance relativement large, mais il illustre le fait que cette classe de trafic est nettement plus que les autres associée à une probabilité plus grande de zone noire.

Ces résultats sont semblables à ceux obtenus lors de l'analyse univariée.

4. *La vitesse maximale autorisée (3 classes)*

Toutes choses égales par ailleurs, la limitation de vitesse liée à la probabilité la plus faible d'appartenance à une zone noire est une limitation supérieure à 90km/h (appelée Vmax120 puisque la représentation principale de cette classe est une vitesse de 120km/h). Par rapport à cette limitation de vitesse, une limitation comprise entre 70 et 90km/h (Vmax90, avec principalement 90km/h) augmente la probabilité pour un hectomètre d'appartenir à une zone noire. Le facteur multiplicatif de risque est de 1.6 ([1.099, 2.448]). Une vitesse maximale autorisée de 50-60km/h (Vmax60) ne possède pas d'influence significative à 95%.

De la même façon que pour le type de route, l'influence de cette variable diffère de celle identifiée initialement avec une régression univariée, où la classe Vmax120 présentait un risque de zone noire plus élevé que Vmax90.

5. *La proximité d'une transition de limitation de vitesse à une autre (à moins de 250 mètres)*

Toutes choses égales par ailleurs, cette proximité augmente le risque pour un hectomètre d'appartenir à une zone noire d'un facteur 1.8 ([1.507, 2.175]). Cela va dans le même sens que l'analyse univariée.

6. *L'appartenance à une agglomération (définie par les panneaux F1 et F3)*

Toutes choses égales par ailleurs, cette variable augmente le risque pour un hectomètre d'appartenir à une zone noire d'un facteur 1.6 ([1.160, 2.172]). Cela est aussi conforme à l'analyse univariée.

7. *La proximité d'une entrée/sortie d'agglomération*

Toutes choses égales par ailleurs, cette variable n'apporte pas d'information significative (à 95%) au modèle, quelle que soit la distance de proximité considérée (à moins de 250m, de 500m, de 750m ou de 1000m). Cela va aussi dans le sens de l'analyse univariée, pour laquelle seule l'influence d'une proximité à moins de 250m était significative (donc influence très locale), mais relativement faible.

8. *Le type d'environnement local (occupation du sol).*

Les données relatives à l'occupation du sol sont initialement réparties en 13 catégories (outre les cours et plans d'eau qui ne sont pas pris en compte ici). Nous avons d'abord effectué un regroupement des 13 catégories initiales de façon à n'avoir plus que 4 classes principales :

- ✓ Urbain : reprend les classes de départ "habitat dense", "habitat discontinu", "espace vert urbain", "habitat et services", "industrie et services".
- ✓ Rural : "prairies", "cultures", "incultes".
- ✓ Bois : "feuillus", "résineux".
- ✓ Autres : "carrières", "terrains militaires", "gares".

Une petite précision concernant la variable "urbain" : celle-ci pourrait plutôt être interprétée comme "agglomération", comme "espace bâti", puisqu'elle reprend aussi l'affectation "habitat discontinu". Les villages (qui ne sont pas de l'urbain à proprement parler) sont donc inclus dans cette classe "urbain". Nous avons regroupé les affectations "habitat dense" et "habitat discontinu" parce que des surfaces très limitées sont considérées comme de l'habitat dense : ce type d'habitat correspond seulement aux centres-villes, voire même à une partie de ceux-ci. Cette catégorie est donc faiblement présente, et est loin de représenter l'habitat de type plus urbanisé comme on pourrait le croire. L'habitat discontinu correspond non seulement à une affectation de village, mais aussi à une affectation de milieux plus urbanisés (voir carte 3.1).

Toutes choses égales par ailleurs, c'est un environnement de type rural qui est lié à la probabilité la plus faible de zone noire. Par rapport à ce type d'occupation du sol, un environnement de type boisé ou urbain (en agglomération) augmente la probabilité pour un hectomètre d'appartenir à une zone noire. Le facteur multiplicatif de risque pour le type boisé est de 1.7 ([1.353, 2.126]), tandis que le facteur multiplicatif de risque pour le type urbain (agglomération) est de 3.1 ([2.504, 3.784]). De nouveau, ces résultats vont dans le sens de ce qui a été observé lors de l'analyse univariée.

Le tableau 5.1 illustre les résultats du modèle. Il est extrêmement important de garder à l'esprit que chaque facteur de risque d'une catégorie d'une variable doit s'interpréter par rapport à la catégorie de référence de cette variable, et que chaque facteur de risque doit aussi s'interpréter toutes choses égales par ailleurs en ce qui concerne les autres variables constitutives du modèle.

Tableau 5.1. Résultats du modèle : odds ratios et intervalles de confiance

	Odds ratio	Limite inf. (95%)	Limite sup. (95%)
Rural	1.000	-	-
Bois	1.696	1.353	2.126
Urbain	3.078	2.504	3.784
Fys3	1.000	-	-
Fys1	0.487	0.337	0.703
Fys2	0.700	0.465	1.053
Fys4	0.524	0.383	0.716
Traf1	1.000	-	-
Traf2	4.345	3.057	6.176
Traf3	5.179	3.665	7.318
Traf4	7.741	5.408	11.078
Traf5	69.192	45.063	106.241
Vmax60	1.276	0.795	2.047
Vmax90	1.640	1.099	2.448
Vmax120	1.000	-	-
Fys750	2.277	1.695	3.059
Agglo	1.588	1.160	2.172
Vmax250	1.811	1.507	2.175

En conclusion de cette section, voici un classement des catégories significatives, pour chaque variable, suivant leur facteur de risque sur la probabilité de zone noire (du plus élevé au plus faible) :

a) *Pour l'environnement en terme d'occupation du sol :*

1. Urbain
2. Boisé
3. Rural.

b) *Pour le type de route :*

1. Type 1 (avec berme centrale)
2. Type 4 (à 2 voies d'une largeur inférieure à 7m)
3. Type 3 (2-3 voies et une largeur d'au moins 7m).

c) *Pour l'intensité de trafic :* d'autant plus forte que le trafic est élevé.

d) *Pour la limitation de vitesse :*

1. Vmax90
2. Vmax120.

e) *Pour la proximité d'un changement du type de route (750m) :* augmente le risque.

f) *Pour l'appartenance à une agglomération :* augmente le risque.

g) *Pour la proximité d'un changement de limitation de vitesse (250m) :* augmente le risque.

Par contre, rappelons que cela n'aurait pas de sens de classifier *ensemble* toutes les catégories de toutes les variables polychotomiques selon l'importance du risque qui leur est associée, puisque ce risque est à chaque fois relatif à celui d'une des catégories, pour chacune des variables.

Les sections suivantes (5.2 à 5.5) sont consacrées à la construction d'un modèle complet, mais en effectuant une stratification suivant le type de route. De cette façon, nous voulons essayer d'identifier d'éventuelles différences de comportement entre ces différents sous-ensembles. D'un point de vue de la sécurité routière en effet, il est probable que la problématique diffère entre, pour prendre des extrêmes, des routes à 4 voies avec berme centrale et des petites routes étroites à 2 voies.

5.2 Stratification suivant le type de route : les routes de type 1

Pour rappel, les routes de type 1 sont les routes dont les directions de circulation sont séparées par une berme centrale. Elles représentent 26% des hectomètres du réseau numéroté du Brabant Wallon, soit environ 1200 hectomètres. Pour ces régressions stratifiées suivant le type de route, les variables de type de route et de transition de type de route à un autre ne sont bien entendu pas prises en compte.

Dans un modèle régressif complet, voici un classement des facteurs significatifs de risque suivant l'importance de leur influence sur la probabilité de zone noire (du plus élevé au plus faible), pour chaque variable :

- a) *Pour l'environnement en terme d'occupation du sol :*
 - 1. Urbain : 2.603 [1.578, 4.294]
 - 2. Boisé : 1.630 [1.146, 2.320]
 - 3. Rural : 1.000

- b) *Pour l'intensité de trafic : d'autant plus forte que le trafic est élevé :*
 - 1. Traf5 : 11.830 [5.735, 24.399]
 - 2. Traf4 : 1.652 (significatif à 80%)
 - 3. Traf3 : 1.000

- c) *Pour la proximité d'un changement du type d'occupation du sol (1000m) :*
augmente le risque : 1.896 [1.294, 2.778]

- d) *Pour la limitation de vitesse :*
 - 1. Vmax90 : 1.000
 - 2. Vmax120 : 0.396 [0.241, 0.650]
 - 3. Vmax60 : 0.350 [0.161, 0.763]

- e) *Pour la proximité d'un changement de limitation de vitesse (250m) :*
augmente le risque : 2.213 [1.267, 3.866].

Pour les hectomètres appartenant à une route de type 1, l'influence des variables est semblable à celles qui ont été identifiées pour l'ensemble des hectomètres, hormis la variable d'appartenance à une agglomération qui n'est pas significative ici, ce qui est somme toute assez logique sur ce type de route. Par contre, une variable représentant la proximité d'une

transition d'un type d'occupation du sol à un autre est significative pour ce sous-ensemble d'hectomètres.

5.3 Stratification suivant le type de route : les routes de type 2

Pour rappel, les routes de type 2 sont les routes avec au moins 4 voies dont les directions de circulation ne sont pas séparées par une berme centrale. Ces routes de type 2 représentent une petite partie de l'ensemble des hectomètres du Brabant wallon (5% seulement, soit un peu plus de 200 hectomètres). Il est par conséquent difficile d'obtenir des évaluations de l'influence des variables indépendantes qui soient statistiquement et interprétativement significatives, et les indications données pour ce sous-ensemble d'hectomètres doivent être considérées avec toutes les précautions nécessaires.

Dans le modèle complet, le classement des facteurs significatifs de risque suivant l'importance de leur influence sur la probabilité de zone noire (du plus élevé au plus faible), pour chaque variable, est le suivant :

a) *Pour l'environnement en terme d'occupation du sol :*

1. Urbain : 1.000
2. Rural. : 0.126 [0.017, 0.964]
3. Boisé : 0.032 [0.003, 0.355]

b) *Pour l'intensité de trafic : d'autant plus forte que le trafic est élevé :*

4. Traf5 : 1.000
5. Traf4 : 0.099 [0.035, 0.281]
6. Traf3 : 0.026 [0.006, 0.118]

c) *Pour la proximité d'une entrée/sortie d'agglomération (250m) : augmente le risque :*
20.812 [4.692, 92.321]

d) *Pour la proximité d'un changement de limitation de vitesse (1000m) : diminue le risque :*
0.129 [0.044, 0.377]

Un exemple de résultat difficilement crédible est le suivant : le odds ratio d'un environnement rural, significatif à 95%, est de 0.126 fois celui d'un environnement urbain. Cependant, l'intervalle de confiance est très large, pratiquement de 0 à 1 : [0.017, 0.964]. Sa valeur n'est donc pas très crédible. De la même façon, le risque associé à la proximité d'une

entrée/sortie d'agglomération es de 21, dans un intervalle de confiance très large : [4.692, 92.321].

5.4 Stratification suivant le type de route : les routes de type 3

Pour rappel, les routes de type 3 sont les routes avec au moins 2 voies et d'une largeur d'au moins 7 mètres. Elles représentent 47% des hectomètres du réseau numéroté du Brabant Wallon, soit environ 2100 hectomètres.

Dans le modèle complet, le classement des facteurs significatifs de risque suivant l'importance de leur influence sur la probabilité de zone noire (du plus élevé au plus faible), pour chaque variable, est le suivant :

a) *Pour l'environnement en terme d'occupation du sol :*

1. Urbain : 1.000
2. Boisé : 0.401 [0.277, 0.582]
3. Rural : 0.305 [0.238, 0.390]

b) *Pour l'intensité de trafic : d'autant plus forte que le trafic est élevé :*

7. Traf5 : 17.822 [10.070, 31.542]
8. Traf4 : 1.393 [1.058, 1.838]
9. Traf3 : 1.000
10. Traf2 : 0.662 [0.477, 0.920]
11. Traf1 : 0.309 [0.218, 0.438]

c) *Pour la limitation de vitesse :*

1. Vmax120 : 1.000
2. Vmax90 : 0.258 [0.074, 0.902]
3. Vmax60 : 0.234 [0.064, 0.850]

d) *Pour la proximité d'un changement de limitation de vitesse (250m) : augmente le risque : 1.531 [1.211, 1.937]*

e) *Pour la proximité d'une entrée/sortie d'agglomération (750m) : augmente le risque : 1.361 [1.044, 1.774]*

- f) *Pour l'appartenance à une agglomération* : augmente le risque: 1.361 (significatif à 90%).

Les variables explicatives pour le sous-ensemble des hectomètres de type 3 présentent aussi une tendance similaire à celle observée pour l'ensemble des hectomètres. Une variable supplémentaire est significative à 95%, celle représentant la proximité d'une entrée/sortie d'agglomération. Sinon, la différence principale observée ici est que c'est une limitation de vitesse de 120km/h qui est associée au risque de zone noire le plus élevé, au lieu de 90km/h pour l'ensemble des hectomètres.

5.5 Stratification suivant le type de route : les routes de type 4

Pour rappel, les routes de type 4 sont les routes avec 2 voies de circulation et d'une largeur inférieure à 7 mètres. Elles représentent 22% des hectomètres du réseau numéroté du Brabant Wallon, soit environ 1000 hectomètres.

Dans le modèle complet, le classement des facteurs significatifs de risque suivant l'importance de leur influence sur la probabilité de zone noire (du plus élevé au plus faible), pour chaque variable, est le suivant :

- a) *Pour l'environnement en terme d'occupation du sol* :

1. Urbain : 5.843 [3.512, 9.721]
2. Boisé : 3.387 [1.625, 7.061]
3. Rural : 1.000

- b) *Pour la proximité d'un changement du type d'occupation du sol (750m)* : diminue le risque : 0.265 [0.171, 0.412]

- c) *Pour l'appartenance à une agglomération* : augmente le risque : 1.812 [1.126, 2.917]

- d) *Pour la proximité d'un changement de limitation de vitesse (750m)* : augmente le risque : 19.597 [8.041, 47.758]

- e) *Pour la proximité d'une entrée/sortie d'agglomération (750m)* : diminue le risque : 0.314 [0.177, 0.556]

Les résultats obtenus pour ce sous-ensemble d'hectomètres de type 4 sont plus difficilement comparables aux précédents. En effet, si l'environnement en terme d'occupation du sol, l'appartenance à une agglomération, ainsi qu'un changement de la limitation de vitesse agissent dans le même sens que précédemment, l'orientation des autres variables significatives est moins évidente à comprendre.

Premièrement, les différences d'intensité de trafic ne sont plus ici un facteur significatif alors qu'elles l'ont toujours été jusqu'à présent, de même que les différences de vitesse maximale autorisée. Deuxièmement, la proximité d'un changement d'occupation du sol diminue ici le risque de zone noire, alors qu'il l'augmentait pour les hectomètres de type 1. Et troisièmement, la proximité d'une entrée/sortie d'agglomération diminue aussi le risque de zone noire, alors qu'elle l'augmentait pour les hectomètres de type 3.

Ces variations montrent l'intérêt de considérer de sous-ensembles comme nous l'avons fait, mais les tendances identifiées ci-dessus par le modèle sont cependant plus difficilement interprétables d'un point de vue explicatif. La logique voudrait en effet que les endroits de changement de la valeur d'une variable soient peut-être plus facilement liés à une probabilité de zone noire plus élevée.

6. Conclusions et perspectives

L'objectif de cette étude était de tenter d'identifier des variables à caractère environnemental et spatial pouvant avoir une influence sur la localisation des zones dangereuses du réseau routier numéroté en Belgique. En d'autres termes, nous voulions essayer d'expliquer spatialement les co-occurrences d'accidents de la route (correspondant à des agrégations spatiales d'accidents) par des variables explicatives liées à l'environnement local de ces zones d'accidents. Nous avons donc tenté de construire un modèle statistique explicatif afin de répondre à ces interrogations, qui a été appliqué à la province du Brabant wallon. Et de fait, nous avons effectivement pu constater que l'environnement spatial pouvait statistiquement expliquer les concentrations locales des accidents de la route (les zones noires). Cependant, il faut aussi rappeler que nous n'avons pris en compte qu'une partie de l'ensemble des variables à caractère spatial potentielles, ce qui rend possible une éventuelle spécification incomplète de la modélisation.

Des analyses univariées (voir Flahaut, 2000) ont permis de montrer quelle était l'influence individuelle de chaque variable sur la probabilité de zone noire, lorsque *seule cette variable* était prise en considération. De la même façon, les analyses catégorielles (section 4) ont permis de montrer l'influence individuelle de *chaque catégorie de variable*, non plus par rapport à une référence interne à la variable (une autre catégorie), mais par rapport à la non-réalisation de chaque catégorie.

Cependant, il importe d'être conscient que la prise en compte simultanée de *plusieurs* de ces variables dans un modèle multivarié (section 5) peut modifier la relation identifiée, puisque l'influence de chaque variable évaluée par le modèle explicatif dépend alors de celles des autres variables présentes dans le modèle. Une partie de la relation entre la variable à expliquer et une des variables explicatives peut être exprimée par une autre variable explicative introduite dans le modèle, ce qui peut entraîner une modification des facteurs de risque univariés. C'est le cas par exemple pour les variables "type de route" et "limitation de vitesse". Toujours dans le cadre de l'interprétation des résultats, rappelons que, de la même façon que pour un modèle univarié, l'interprétation du facteur de risque pour les variables dont le nombre de classes est supérieur à deux se fait aussi par rapport à l'une de ses catégories.

Dans un modèle de régression appliqué à l'ensemble du réseau routier numéroté du Brabant Wallon, plusieurs variables à caractère spatial apparaissent d'un point de vue statistique comme étant significativement explicatives des variations spatiales de la probabilité pour un hectomètre d'appartenir à une zone noire. L'influence de chaque variable doit s'interpréter *toutes choses égales par ailleurs* en ce qui concerne les autres variables introduites dans le modèle. Cela sous-entend donc l'hypothèse que le rôle d'une variable est identique quelle que soit la configuration présentée par les autres variables, en d'autres termes qu'il n'y a pas d'interaction entre les variables constitutives du modèle. Les variables significatives constitutives de ce modèle régressif multivarié sont les suivantes :

- ✓ *Le type de route*
- ✓ *La proximité d'une transition d'un type de route à un autre*
- ✓ *L'intensité de trafic*
- ✓ *La vitesse maximale autorisée*
- ✓ *La proximité d'une transition de limitation de vitesse à une autre*
- ✓ *L'appartenance à une agglomération*
- ✓ *Le type d'environnement local (occupation du sol)*

Un modèle de régression multivarié a également été appliqué à 4 sous-ensembles du réseau routier numéroté du Brabant wallon. La stratification de l'ensemble de départ est effectuée suivant le type de route. Dans ces régressions, les variables "type de route" et "proximité d'une transition d'un type de route à un autre" ne sont bien entendu pas introduites. Cette stratification a permis de montrer que globalement, les tendances principales observées sont similaires à celles constatées pour l'ensemble du réseau, même si des différences peuvent tout de même être relevées pour certaines variables. Ces variations laissent penser qu'il peut donc être intéressant d'identifier les variables significatives en effectuant une telle

stratification, lorsque l'on souhaite soustraire de l'explication l'influence du type de route sur la probabilité de zone noire. Cependant, une stratification différente pourrait aussi être utilement envisagée en se basant sur l'intensité du trafic. Nous avons en effet pu observer qu'au plus cette intensité de trafic est élevée, au plus le risque qu'un hectomètre appartienne à une zone noire est élevé, ce qui peut sembler logique. Dans la littérature cependant, même si ce type de relation est souvent observé, celle-ci tend aussi souvent à diminuer voire à s'inverser pour les intensités les plus élevées provoquant la saturation du réseau. Il pourrait aussi être intéressant de soustraire de l'explication l'influence de l'intensité du trafic sur la probabilité de zone noire.

Nous avons donc montré qu'il était possible de relier les lieux de concentration, de co-occurrence d'accidents de la route, à un ensemble de variables décrivant l'environnement local des zones noires. Dans la littérature, ce type de démarche est peu abordé, les auteurs préférant habituellement se diriger vers des variables caractérisant le véhicule et/ou ses occupants. Ce qui différencie clairement cette étude de la plupart de celles relevées dans la littérature, c'est qu'ici l'individu de base est *l'hectomètre et non l'accident* en tant que tel, c'est-à-dire un élément de l'espace, cet espace étant lui-même particulier puisqu'il s'agit d'un *réseau* routier. Les variables introduites dans le modèle doivent donc toutes pouvoir être reliées aux hectomètres par une composante spatiale.

Nous avons utilisé des modèles régressifs de type logistique pour établir ces relations. Au-delà de l'influence de ces variables sur la probabilité de zone noire, une extension intéressante de ce travail pourrait être de considérer *l'intensité* du caractère dangereux des hectomètres appartenant à une zone noire. L'objectif serait alors d'évaluer si certaines variables à caractère environnemental et spatial sont éventuellement liées plus que d'autres à des zones noires plus dangereuses. Dans le même ordre d'idée, il serait aussi intéressant de voir si la *longueur* des zones noires peut être ou non reliée à leur environnement, autrement dit si l'un ou l'autre type d'environnement local peut-être relié à des zones noires plus ou moins longues. Enfin, il existe encore d'autres variables à caractère environnemental et spatial qui n'ont pas été considérées dans cette étude ; nous pensons par exemple à des variables relevant d'une approche plus microspatiale comme l'état de la chaussée ou la géométrie de la route (virage, pente...). De la même façon, l'élargissement de la zone d'étude à d'autres provinces/régions pourrait apporter un éclairage différent et plus complet à notre problème : cela permettrait de mettre en évidence d'éventuelles disparités régionales, ou au contraire de confirmer les résultats identifiées pour le Brabant Wallon.

Pour terminer, ajoutons encore que d'un point de vue méthodologique, des extensions et améliorations du modèle explicatif sont possibles et souhaitables, comme par exemple l'utilisation de modèles autorégressifs et/ou autocorrélés qui tiennent compte de la présence éventuelle d'autocorrélation spatiale dans les données.

BIBLIOGRAPHIE

- ANSELIN L., Local indicators of spatial association-LISA, *Geographical Analysis*, 27, 2, 1995, pp. 93-115.
- CLIFF A.D. et ORD J.K., *Spatial autocorrelation*, London, Pion, 1973, 178p.
- CLIFF A.D. et ORD J.K., *Spatial processes. Models and applications*, London, Pion, 1981, 266 p.
- COX D.R., *The analysis of binary data*, London, Methuen, 1970.
- DEPARTEMENT LEEFMILIEU EN INFRASTRUCTUUR (*LIN*), Données réseau routier et accidents de Flandre 1992-96.
- FLAHAUT B., *Influence de l'aménagement du territoire sur la sécurité routière durable. Analyse de la situation belge. Concentration spatiale des accidents de la route: délimitation des zones noires*, Rapport de recherche inédit, 54p. + annexes, Louvain-la-Neuve, 1999.
- FLAHAUT B., *Influence de l'aménagement du territoire sur la sécurité routière durable. Analyse de la situation belge. Concentration spatiale des accidents de la route: construction d'un modèle explicatif*, Rapport de recherche inédit, 66p., Louvain-la-Neuve, 2000.
- GETIS A. et ORD J.K., The analysis of spatial association by use of distance statistics, *Geographical Analysis*, 24, 3, 1992, pp. 189-206.
- GETIS A., Spatial interaction and spatial autocorrelation: a cross-product approach, *Environment and Planning A*, 23, p. 1269-1277, 1991.
- GOULD P., Is statistix inferens the geographical name for a wild goose ?, *Economic geography*, 46, 1970, pp. 439-48.
- HAINING R., *Spatial data analysis in the social and environmental sciences*, Cambridge, University press, 1990, 409p.
- HOSMER D.W. et LEMESHOW S., *Applied logistic regression*, New York, Wiley, 1989.
- INSTITUT NATIONAL DE STATISTIQUES (*INS*), Données d'accidents 1992-1996 (non publiées).
- MINISTÈRE WALLON DE L'EQUIPEMENT ET DES TRANSPORTS (*MET*), Données réseau routier et accidents de Wallonie 1992-96.

- ORD J.K. et GETIS A., Local spatial autocorrelation statistics : distributional issues and applications, *Geographical analysis*, 27, 1995, pp. 286-306.
- ROMANO U. et HEUCHENNE D., Note d'information. Modèle mathématique d'évaluation de l'insécurité routière, *Collection Sécurité Routière*, 1, 7p., Ministère de l'Équipement et des Transports (MET), Namur, 1996.
- SAS INSTITUTE INC., *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2*, Cary, NC: SAS Institute Inc., 1989.
- STOKES M.E., DAVIS C.S. et KOCH G.G., *Categorical data analysis using the SAS System*, Heidelberg, SAS Institute, 1996.
- TOBLER W.R., A computer movie simulating urban growth in the Detroit region, *Economic Geography*, Supplement 46, 1970, pp. 234-40.
- UPTON G. et FINGLETON B., *Spatial data analysis by example*, New York, Wiley, 1985.
- WRIGLEY N., Introduction to the use of logit models in geography, *Catmog*, 10, 33 p., 1976.
- WRIGLEY N., *Categorical data analysis for geographers and environmental scientists*, London, Longman, 1985.