PolyPublie
Polytechnique Montréal

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

Titre: Title:	Étude comparative des accidents impliquant des véhicules lourds avec ou sans matière dangereuse au Québec.
Auteur: Author:	Benoît Prufer
Date:	2010
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Prufer, B. (2010). Étude comparative des accidents impliquant des véhicules lourds avec ou sans matière dangereuse au Québec. [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/432/

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/432/
Directeurs de recherche: Advisors:	Martin Trépanier
Programme: Program:	Génie industriel

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal

https://publications.polymtl.ca

UNIVERSITÉ DE MONTRÉAL

ÉTUDE COMPARATIVE DES ACCIDENTS IMPLIQUANT DES VÉHICULES LOURDS AVEC OU SANS MATIÈRES DANGEREUSES AU QUÉBEC

PRUFER BENOIT

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION

DU DIPLÔME DE MAITRISE ES SCIENCES APPLIQUÉES

(GÉNIE INDUSTRIEL)

NOVEMBRE 2010

© Benoit PRUFER, 2010.

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

ÉTUDE COMPARATIVE DES ACCIDENTS IMPLIQUANT DES VÉHICULES LOURDS AVEC OU SANS MATIÈRES DANGEREUSES AU QUÉBEC

Présenté par : PRUFER Benoît

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

Mme DE MARCELLIS-WARIN Nathalie, Ph.D, Présidente

M. TRÉPANIER Martin, Ph.D, Membre et directeur de recherche

M. ADJENGUE Luc, Ph.D, Membre

REMERCIEMENTS

En tout premier lieu, je tiens tout particulièrement à remercier mon directeur de recherche, M. Martin Trépanier, qui m'a fait confiance pour mener à bien ce projet. Sa disponibilité, ses conseils précieux, ses connaissances et son soutien m'ont permis d'avancer rapidement et efficacement tout au long de ma maitrise.

Je tiens aussi à remercier M Raynald Boies du Ministère des Transports du Québec pour avoir si gentiment répondu à mes très nombreuses questions ainsi que M. Carl St-Pierre qui m'a grandement aidé, par ses connaissances et ses conseils, dans la réalisation de mon étude statistique.

Enfin j'aimerais remercier mes proches de m'avoir soutenu dans cette aventure, et particulièrement mon amie, qui a supporté mes explications et m'a très souvent éclairé d'un point de vue original et pertinent.

RÉSUMÉ

Les accidents routiers impliquant des matières dangereuses présentent, du fait de la présence de ces matières, de sérieux risques d'aggravations. Aussi il est essentiel d'étudier leur transport afin de permettre aux autorités de réguler de façon efficace et sécuritaire ce secteur du transport. Cette présente étude vise à déterminer l'influence de la présence de matières dangereuses sur l'accidentologie des véhicules lourds au Québec, à l'aide d'une analyse statistique des données puis d'une analyse spatiale.

On trouve dans la littérature de nombreuses études descriptives et spatiales du transport des matières dangereuses (TMD) mais peu traitent spécifiquement du Québec. Les études descriptives du TMD étudient de nombreux aspects du domaine comme l'évolution du TMD, la répartition entre les types de transport, les matières dangereuses impliquées, les conséquences humaines et financières des accidents ou encore les circonstances des accidents. Les études spatiales s'intéressent plus à la détermination des zones à risque (« points chauds ») ainsi qu'à l'estimation de l'impact des accidents sur la population et à la minimisation des risques par l'optimisation des itinéraires des camions de TMD.

Cette étude s'appuie sur deux bases de données québécoises recensant d'une part les accidents impliquant les véhicules lourds (« base globale ») et d'autre part les accidents impliquant des véhicules de transport de matière dangereuse (« base TMD »). Dans un premier temps un travail de mise en forme et de préparation des données a été réalisé. Les accidents de la base TMD ont été identifiés dans la base globale grâce à un programme Visual Basic, et la base TMD a été consolidée. De plus les accidents ont été localisés, et un outil de visualisation des rapports d'accidents intégrant le positionnement de l'accident sur une carte a été mis en place dans les deux bases de données.

Une analyse comparative des champs communs a alors été menée en utilisant des outils statistiques comme la statistique descriptive, des graphiques et des tests d'hypothèses. Tous d'abord le champ « date » a été étudié. L'analyse de l'évolution annuelle du nombre d'accidents n'a pas mis en évidence de tendance claire. Il semble que les années à forte proportion d'accidents de véhicules lourds soient aussi des années à fort taux d'accidents TMD, mais le lien n'est pas très marqué. On peut cependant noter qu'à proportions égales les variations annuelles des accidents TMD sont plus de 2 fois supérieures aux variations annuelles des accidents de la

base globale. D'autre part la saison hivernale est particulièrement propice aux accidents TMD avec des taux d'accidents particulièrement élevés. La répartition horaire obtenue est similaire entre les deux bases de données, les accidents ayant majoritairement lieu la journée avec deux pics vers 10h et 14h et une « pause » à midi.

Ensuite les champs quantitatifs ont été étudiés. Le nombre de véhicules impliqués apparait significativement plus faible dans le cas des accidents TMD avec une différence de 0,2 véhicule par accident, ce qui traduit une proportion plus élevée d'accidents impliquant 1 seul véhicule. L'analyse du champ indiquant la limitation de vitesse sur la route où l'accident a eu lieu semble montrer que les accidents TMD ont lieu sur des routes moins rapides. Cependant ce résultat est à nuancer puisque dans près de 50% des rapports d'accidents cette information est manquante. Les résultats de l'analyse sur les conséquences humaines sont par contre plus fiables. On obtient que les accidents TMD sont en proportion plus mortels et font plus de blessés qu'un simple accident de la route et que la probabilité d'avoir un grand nombre de personnes impliquées est plus élevée pour un accident TMD.

Cette analyse statistique s'est terminée par l'étude des champs qualitatifs. Ils permettent de souligner certaines différences dans les conditions ou les circonstances favorisant des accidents. Ainsi, on constate que les accidents de la base TMD sont plus souvent dus à une sortie de route que ceux de la base globale, même si le cas le plus fréquent est, dans les deux cas, une collision entre deux véhicules. De plus, plusieurs champs soulignent la plus grande dangerosité du transport MD dans des conditions hivernales, avec par exemple plus d'accidents par temps de neige ou sur des surfaces glacées ou enneigées que pour la base globale. Enfin, les accidents TMD ont plus souvent lieu en milieu rural ou forestier que les accidents globaux, qui ont majoritairement lieu en zone commerciale.

Dans le second volet de cette étude, une analyse spatiale des répartitions des accidents a été menée. Les répartitions obtenues montrent que les accidents se concentrent dans les zones habitées du Québec particulièrement le long du Saint-Laurent. La région regroupant le plus grand nombre d'accident TMD est la Montérégie avec 20% des accidents TMD. Pour la base globale c'est Montréal qui concentre une grande majorité d'accidents avec 26% des accidents. L'analyse centrographique réalisée confirme cette observation, le centre moyen de la base globale se trouvant plus proche de Montréal que celui de la base TMD. Ce point moyen correspond

d'ailleurs presque à celui de la population québécoise. Ainsi les accidents TMD semblent mieux répartis sur le territoire québécois et moins liés à la répartition de population, ce que confirment les ellipses d'écart-type. L'analyse de l'évolution annuelle de la position des centres moyens montre que globalement la tendance est à un rapprochement vers Montréal notamment pour les accidents de la base globale, les centres moyens des accidents TMD semblent eux plus aléatoires et varient sur une plus grande surface d'une année sur l'autre. Une analyse des plus proches voisins vient confirmer l'éparpillement plus important des accidents de la base TMD sur le territoire avec une valeur du NNI plus de 10 fois supérieure à celui de la base globale. L'analyse des points chauds met en évidences que si l'on étudie avec de « gros » points chauds les répartitions, elles se comportent de façon similaire. La concentration principale des accidents se trouvant comme le confirme l'analyse des densités autour de Montréal. Par contre lorsque l'on fait des regroupements plus fins des différences apparaissent. La concentration plus forte des accidents de la base globale est alors mise en évidence, avec notamment des ellipses NnH de niveau 2 couvrant une faible partie du territoire comparativement aux ellipses de la base globale.

Enfin une analyse spatiale pondérée des accidents a été réalisée, selon le nombre de véhicules impliqués, de décès et de blessés. Les densités des véhicules impliqués et des blessés sont similaires, les zones de fortes densités correspondants aux endroits de forte concentration d'accidents. En revanche la répartition des décès souligne, en plus des zones de forte concentrations d'accidents, des sites particulièrement dangereux, pour la plupart communs aux deux distributions.

ABSTRACT

Road accidents involving hazardous materials present, due to such materials, serious risk of exacerbations. Thus it is essential to study their transport, to enable the authorities to regulate effectively and safely this transport sector. The present study aims to determine the influence of hazardous materials presence on the heavy vehicle accident in Quebec, using a statistical analysis of data and spatial analysis.

Many descriptive and spatial studies of hazardous materials transport (hazmat) exist in the literature but few deal specifically with the Quebec. Descriptive studies of hazmat study many aspects of the field as the evolution of hazmat, the distribution between types of transport, hazardous materials involved, the human and financial consequences of accidents or the circumstances of accidents. Space studies are more interested in identifying risk areas (hot spots) and the estimation of the accidents impact on the population, and minimizing risks by optimizing hazmat truck routes.

This study uses two databases of the Quebec identifying, on one hand accidents involving heavy vehicles (global basis) and on the other hand other accidents involving vehicles transporting dangerous materials (hazmat basis). Initially a work of editing and data preparation has been done. Hazmat accidents have been identified in the global base with a VB program, and the hazmat base has been completed and corrected. Moreover, accidents have been located, and a visualization tool of accident reports, integrating the accident's position on a map, has been realized in both databases.

A comparative analysis of common fields was then conducted using statistical tools such as descriptive statistics, charts and hypothesis testing. First the field "date" was studied. The analysis of the number of accidents per year did not reveal any clear trend. It seems that years with high proportion of heavy vehicle accidents are also years of high hazmat vehicle accident rates, but the link is not very marked. However, in equal proportions, hazmat accidents annual variations are more than 2 times superior to the annual variations of global base. Moreover, the winter is especially conducive to hazmat accidents with particularly high accident rate. The hourly distribution obtained is similar between the two databases, with a majority of accidents during the day, with two peaks around 20h and 14h and a pause at noon.

Then the quantitative fields have been studied. The number of vehicles involved appears significantly lower in the case of hazmat accidents with a difference of 0.2 vehicles per accident, reflecting a higher proportion of accidents involving a single vehicle. Analysis of the “speed limit” field seems to show that hazmat accidents occur on slower roads. However, this result must be qualified since in nearly 50% of accident reports this information is missing. The results of human consequences analysis are more reliable. We get that hazmat accidents are proportionally more fatal and cause more injuries than global accidents and that the probability of having a large number of people involved is higher for a hazmat accident.

This statistical analysis was completed by the study of qualitative fields, which can point out some differences in conditions or circumstances favoring accidents. Thus, hazmat accidents are more often due to an exit route than the global ones, although the most frequent is, in both cases, a collision between two vehicles. In addition, several fields emphasize the greater danger of DM transportation in winter conditions, with for example more accidents in the snow or on icy or snowy surfaces for the hazmat base. Finally, hazmat accidents occur most often in rural or forest areas than general accidents, which occur mainly in commercial areas.

In the second part of this study, a spatial analysis of accident distributions was conducted. The distributions obtained show that accidents are concentrated in the populated areas of Quebec, especially along the Saint-Laurent. The region with the largest number of hazmat accidents is the Montérégie with 20%. For the global base it is Montreal, which focuses the vast majority of accidents with 26% of accidents. The centrographic analysis confirms this observation, since the mean center of the global base is located closer to Montreal than hazmat base’s one. This point also corresponds to that of the Quebec population. Hazmat accidents appear to be better distributed in Quebec and less related to population distribution, as confirmed by the standard deviation ellipses. The analysis of annual changes of the centers means position shows that the overall trend is to get closer to Montreal, particularly for global accidents base, the hazmat accidents centers means seem more random and vary over a larger surface from one year to another. An analysis of nearest neighbors confirms the larger scatter of hazmat accidents base with a value of the NNI over 10 times higher than the global base. The analysis of hotspots shows that with "large" hot spot distributions studies, the two distributions behave similarly. The main concentration of accidents is found around Montreal, as confirmed by the analysis of densities. On the contrary, when making groupings finer, differences appear. The highest concentration of

global base's accidents is then highlighted, with ellipses NNH Level 2 covering a small part of the territory compared to the ellipses of the hazmat base.

Finally a weighted spatial analysis of accidents has been achieved, depending on the number of vehicles involved, deaths and injuries. The densities of the vehicles involved and casualties are similar with high densities areas corresponding to locations of high concentration of accidents. In contrast, the distribution of deaths stresses, in addition to areas of high concentrations of accidents, particularly dangerous sites mostly common to both distributions.

TABLE DES MATIÈRES

REMERCIEMENTS III

RÉSUMÉ IV

ABSTRACT VII

TABLE DES MATIÈRES X

LISTE DES TABLEAUX XIV

LISTE DES FIGURES XVI

LISTE DES SIGLES ET ABRÉVIATIONS XXII

LISTE DES ANNEXES XXIII

INTRODUCTION 1

CHAPITRE 1 PROBLÉMATIQUE ET MISE EN CONTEXTE 3

1.1 Problématique 4

1.2 Buts et objectifs 5

1.3 Définitions 6

1.3.1 Les matières dangereuses 6

1.3.2 Transport de Matières Dangereuses (TMD) 6

1.3.3 Accidents et accidents MD 7

1.4 Cadre législatif en vigueur au Québec 8

1.4.1 Rôle des différentes institutions 8

1.4.2 Les contenants 9

1.4.3 Les indications de danger 10

CHAPITRE 2 REVUE DE LITTERATURE 11

2.1 Études descriptives du TMD 12

2.1.1 Distribution dans le temps des accidents 12

2.1.2 Type de transport....17
2.1.3 Matières dangereuses impliquées....17
2.1.4 Conséquences des accidents....19
2.1.5 Type d'accident MD....23
2.1.6 Circonstances des accidents....24
2.2 Études spatiales du TMD....27
2.2.1 Analyse centrographique....27
2.2.2 Analyse des points chauds....29
2.2.3 Ajout de couche d'informations spatiales....31
2.3 Utilisation d'une base de données d'accidents....36
2.3.1 Sélection des données....36
2.3.2 Le biais dans les bases de données....37
2.3.3 Cas des bases de données incomplètes....38
CHAPITRE 3 MÉTHODOLOGIE....40
3.1 Démarche suivie....40
3.2 Appropriation des bases de données....41
3.2.1 Présentation des bases de données....41
3.2.2 Identification des accidents TMD dans la base globale....47
3.2.3 Corrections et enrichissement de la base TMD....52
3.3 Analyse comparative des données....53
3.3.1 Problème de la dépendance des données....53
3.3.2 Données quantitatives....54
3.3.3 Données qualitatives....61
3.3.4 Cas particulier du champ « Date »....62

3.4 Analyse de la répartition spatiale des accidents 62
3.4.1 Localisation des accidents 63
3.4.2 Montage du système d'information géographique 67
3.4.3 Analyse spatiale comparative 70
CHAPITRE 4 ANALYSE COMPARATIVE DES DONNÉES 76
4.1 Étude du champ « Date » 77
4.1.1 Distribution annuelle 77
4.1.2 Distribution mensuelle 81
4.1.3 Distribution saisonnière 83
4.1.4 Distribution selon les jours de l'année 84
4.1.5 Distribution selon les jours de la semaine 87
4.2 Études des champs quantitatifs 89
4.2.1 Champ « Heure » 89
4.2.2 Champ « Nombre de véhicules impliqués » 92
4.2.3 Champ « Limitation de Vitesse » 96
4.2.4 Champs concernant les impacts sur la santé humaine 98
4.3 Étude des champs qualitatifs 105
4.3.1 Champ « État de la météo » 105
4.3.2 Champ « Genre d'accident » 106
4.3.3 Champ « Code d'impact » 107
4.3.4 Champ « Aspect de la chaussée » 108
4.3.5 Champ « Éclairement » 109
4.3.6 Champ « Environnement » 110
4.3.7 Champ « État de la surface » 111

4.3.8 Champ « Nature de la Chaussée » 112
CHAPITRE 5 ETUDE SPATIALE DES ACCIDENTS 113
5.1 Analyse des répartitions d'accidents 113
5.1.1 Aperçu global des répartitions d'accidents 113
5.1.2 Répartition par région économique (RE) 115
5.1.3 Répartition des accidents sur le réseau routier 117
5.2 Analyse statistique des répartitions d'accidents 119
5.2.1 Analyse Centrographique 119
5.2.2 Évolution des points moyens selon les années 121
5.2.3 Analyse des plus proches voisins 124
5.2.4 Analyse des points chauds 125
5.2.5 Analyse des densités 133
5.3 Analyse pondérée des répartitions d'accidents 136
5.3.1 Analyse spatiale du nombre de véhicules impliqués 136
5.3.2 Analyse spatiale de la répartition des décès 137
5.3.3 Analyse spatiale de la répartition des blessés 139
CONCLUSION 141
REFERENCES 145
ANNEXES 151

LISTE DES TABLEAUX

Tableau 1.1 : Principales conséquences d'un accident MD, adapté de (Brun-Picard, Gominet, & Louis, 2008) 7

Tableau 1.2 : Principaux contenants de matières dangereuses (Direction de la sécurité en transport et du camionnage, 2006) 9

Tableau 2.1 : Comparaison des distributions temporelles d'accidents dans plusieurs études 12

Tableau 2.2 : Compilation des résultats des courbes p-N cumulative concernant le nombre de décès dans différentes études 21

Tableau 3.1 : Étapes de génération de la carte de visualisation 46

Tableau 3.2 : Organisation des données utilisées.....55

Tableau 3.3 : statistique descriptive à deux variables 57

Tableau 3.4 : Paramètres d'analyse centrographique (Ned Levine & Associates, 2010a) 71

Tableau 3.5 : Description des paramètres calculés par CrimeStat pour la NNA (Ned Levine & Associates, 2010b) 72

Tableau 4.1 : Les 18 champs communs 76

Tableau 4.2 : Distribution annuelle des observations d'accidents 77

Tableau 4.3 : Statistique descriptive des distributions annuelles 79

Tableau 4.4 : Distribution mensuelle des observations d'accidents.....81

Tableau 4.5 : Distribution saisonnière des accidents dans les deux bases de données 83

Tableau 4.6 : Statistique descriptive des distributions d'accidents selon les 366 jours de l'année (∑ des accidents entre 1995 et 2007).....84

Tableau 4.7 : Jours où il y a eu le plus d'accidents.....85

Tableau 4.8 : Jours où il y a eu le moins d'accidents.....85

Tableau 4.9 : Distribution journalière des observations d'accidents (1995-2007).....88

Tableau 4.10 : Distribution selon les heures des observations d'accidents (1995-2007) 89

Tableau 4.11 : Valeurs des paramètres de statistique descriptive pour les champs « Nombre de véhicules impliqués » des deux bases de données (1995-2007) 92

Tableau 4.12 : Paramètre de forme des variables normalisés 93

Tableau 4.13 : Fréquence d'apparition des valeurs des « limitation de vitesse » pour la période 1995-2007.....96

Tableau 4.14 : Quelques paramètres de statistique descriptive pour les répartitions des valeurs des limitations de vitesse (1995-2007) 97

Tableau 4.15 : Fréquence d'apparition des nombres de décès pour la période 1995-2007 99

Tableau 4.16 : Valeurs des paramètres de statistique descriptive pour les décès (1995-2007) ...100

Tableau 4.17 : Proportion d'accident mortel.....101

Tableau 4.18 : Fréquence d'apparition des nombres de blessés pour la période 1995-2007.....102

Tableau 4.19 : Valeurs des paramètres de statistique descriptive pour les blessés (1995-2007).103

Tableau 4.17 : Proportion d'accident mortel.....104

Tableau 5.1 : Résultats de l'analyse des plus proches voisins (ordre 1) 124

Tableau 5.2 : Quelques résultats numériques de l'analyse NnH.....126

Tableau 5.3 : Quelques résultats numériques du K-Means clustering 130

LISTE DES FIGURES

Figure 0.1 : Renversement d'un camion-citerne à Brossard le 23 octobre 2008 2

Figure 0.2 : Semi-remorque chargé d'hydrosulfate de sodium en feu[2].....2

Figure 1.1 : Organigramme du TMD (de Marcellis-Warin, Favre, et al., 2006) 9

Figure 1.2 : Exemple de plaques d'indication de danger (Direction de la sécurité en transport et du camionnage, 2006) 10

Figure 2.1 : Nombre de publications par année selon la base de données Compendex, concernant le transport des matières dangereuses 11

Figure 2.2 : Évolution des accidents routier qui ont mis en cause des marchandises dangereuses au Canada (Transports Canada, 2009a).....14

Figure 2.3 : Répartition horaire des accidents TMD dans le comté de Chester en Pennsylvanie (Shorten et al, 2002).....16

Figure 2.4 : Répartition horaire des accidents TMD au Québec (Boies & Beaulieu, 2009).....16

Figure 2.5 : Nombre d'accidents en fonction de la matière dangereuse (Shorten, et al, 2002) 18

Figure 2.6 : Répartition des classes de MD impliquées dans les accidents aux États-Unis en 2003 (Erhan. Erkut, et al., 2007).....19

Figure 2.7 : Évolution des nombres de décès et de blessés dans les accidents impliquant des matières dangereuses au cours du transport et lors de la préparation du transport au Canada (Transports Canada, 2009a) 20

Figure 2.8 : Courbe f-N pour le transport routier, ferroviaire et aérien (1967-2001) (Evans, 2003) 21

Figure 2.9 : Coûts comparés en US$ de différents types d'accident routier aux États-Unis (Erhan. Erkut, et al., 2007).....22

Figure 2.10 : Statistique québécoises sur les indemnités liées aux accidents du travail (en $CAD) (de Marcellis-Warin, Leroux, et al., 2006).....23

Figure 2.11 : Arbre des probabilités de chaque type d'accident (Oggero, et al., 2006).....24

Figure 2.12 : Nombre d'accidents en fonction de la longueur du réseau dans chaque ville (Shorten, et al., 2002) 25

Figure 2.13 : Causes des accidents TMD selon le mode de transport (USA-2003) (Erhan. Erkut, et al., 2007) 26

Figure 2.14 : Centre de gravité de la population du Québec en 1951, 1961 et 1971 (Dean, et al., 1974) 28

Figure 2.15 : Déplacement de la population francophone au Nouveau-Brunswick de 1961 à 2001 (Cao, et al, 2005) 29

Figure 2.16 : Distribution spatiale des accidents piétonniers de Las Vegas (en haut) et leurs zones de forte concentration (Pulugurtha, Krishnakumar, & Nambisan, 2007) 30

Figure 2.17 : « Hot spots » des rejets de MD et des usines et installations utilisant des MD au sud de la Californie (Schweitzer, 2006) 31

Figure 2.18 : Caractérisation d'un nuage de dispersion (Luketa-Hanlin, 2006) 32

Figure 2.19 : Estimation des conséquences sur la santé humaine d'un déversement à Edmonton, Canada (Zhang, et al., 2000) 32

Figure 2.20 : Déversement d'ammoniac à Châteauguay, Canada (Drouin & Leroux, 2004) 33

Figure 2.21 : Optimisation selon 3 critères différents du routage entre Chicago et Atlanta (E Erkut & Verter, 1998) 34

Figure 2.22 : Critères de sélection d'un accident de MHIDAS (Ronza et al, 2003) 36

Figure 2.23 : Interface de l'outil HECOT (Trépanier, et al., 2009) permettant de matcher des accidents de base de données différentes 38

Figure 3.1 : Aperçu des premiers champs de la BDD globale 42

Figure 3.2 : Aperçu du fichier Excel de la base TMD 43

Figure 3.3 : Exemple de champ corrigé dans la base TMD 44

Figure 3.4 : Visualisation des accidents de la base TMD 45

Figure 3.5 : Visualisation des accidents de la base globale 45

Figure 3.6 : Méthode de simplification des recherches....48
Figure 3.7 : Méthodologie d'identification et résultats obtenus....49
Figure 3.8 : Comparateur....50
Figure 3.9 : Bilan des associations....51
Figure 3.10 : Comparaison des résultats de chacune des étapes....51
Figure 3.11 : Exemple de cellules complétées....52
Figure 3.12 : Diagramme de Venn (période 1995-2007)....53
Figure 3.13 : Réduction numérique de données (Carol, et al., 2002)....56
Figure 3.14 : Exemple de test de Student avec Excel....59
Figure 3.15 : Probabilité cumulée d'un accident avec N décès (courbe f-N) (Darbra & Casal, 2004)....61
Figure 3.16 : Les champs qui permettent de localiser un accident de la base globale....63
Figure 3.17 : Localisation de la base TMD via MapPoint....64
Figure 3.18 : Exemple de localisation Google Maps des accidents initialement localisés à Lévis (centre-ville) par MapPoint....65
Figure 3.19 : Localisation des accidents TMD de Montréal et de ces environs (en vert = localisation MapPoint, en jaune = localisation manuelle)....66
Figure 3.20 : Origine de la localisation de la base TMD....66
Figure 3.21 : Interface de MapWindow GIS....68
Figure 3.22 : Fond de carte générique du Québec....69
Figure 3.23 : Population en fonction des aires de diffusion (Ile de Montréal)....70
Figure 3.24 : Paramétrage des données dans CrimeStat III....71
Figure 3.25 : Exemple Hot Spot obtenu par la méthode NnH (N. Levine, 2006)....73
Figure 3.26 : Principe de sommation des fonctions kernel (Ned Levine & Associates, 2010d)....74
Figure 3.27 : Grille utilisée pour l'estimation des densités et ses paramètres dans CrimeStat....75

Figure 3.28 : Paramétrage de CrimeStat pour l'estimation de la densité 75

Figure 4.1 : Évolution annuelle comparée du nombre d'accidents (% du total) 78

Figure 4.2 : Nuage de points des effectifs annuel d'accident 80

Figure 4.3 : Évolution du nombre d'accidents TMD pour 1000 accidents de véhicule lourd 80

Figure 4.4 : Distribution mensuelle comparée du nombre d'accidents par jour par an selon les mois (en % du total) 82

Figure 4.5 : Distribution saisonnière des accidents (% du total) 83

Figure 4.6 : Histogramme des observations d'accidents selon les jours de l'année (% du total) ..86

Figure 4.7 : Distribution des accidents selon les jours de la semaine (1995-2007) (% du total) ...88

Figure 4.8 : Distribution horaire des accidents sur la période 1995-2007 (% du total) 90

Figure 4.9 : Nuage de point des fréquences horaire des accidents 91

Figure 4.10 : Fréquences des nombres de véhicules impliqués (% du total) 94

Figure 4.11 : Évolution annuelle du nombre moyen de véhicules impliqués par accident 95

Figure 4.12 : Pourcentage cumulé des fréquences d'apparition des limitations de vitesse sur la période 1995-2007 97

Figure 4.13 : Probabilité cumulée d'un accident TMD avec N décès 100

Figure 4.14 : Probabilité cumulée d'un accident TMD avec N blessés (grave ou léger) 103

Figure 4.15 : Fréquence comparée des valeurs du champ « météo » (% du total) 105

Figure 4.16 : Fréquence comparée des genres d'accidents (% du total) 106

Figure 4.17 : Animaux impliqués dans des collisions TMD 107

Figure 4.18 : Fréquence comparée des valeurs du champ « code d'impact » (% du total) 107

Figure 4.19 : Fréquence comparée des valeurs du champ « aspect de la chaussée» (% du total) 108

Figure 4.20 : Fréquence comparée des valeurs du champ « éclairement» (% du total) 109

Figure 4.21 : Fréquence comparée des valeurs du champ « environnement » (% du total) 110

Figure 4.22 : Fréquence comparée des valeurs du champ « état de la surface » (% du total) 111

Figure 4.23 : Fréquence comparée des valeurs du champ « nature de la chaussée » (% du total) 112

Figure 5.1 : Répartition des accidents TMD (rouge) et des accidents de la base globale (bleu) dans le Québec « habité » et principales routes 114

Figure 5.2 : Répartition par RE des accidents des deux bases (% du total) 116

Figure 5.3 : Répartition des accidents obtenues à l'échelle de Montréal (bleu = base globale, rouge = base TMD) 118

Figure 5.4 : Point Moyen, Distance type XY (rectangle), Ellipse de l'écart-type (rouge=TMD, bleu= base globale, jaune=population).....119

Figure 5.5 : Point moyen (carré), Centre des distance minimum (losange) et distances-types elliptiques (rouge=TMD, bleu= base globale, jaune=population) 121

Figure 5.6 : Évolution des points moyens entre 1995 et 2007 (bleu=base globale, rouge=TMD) 122

Figure 5.7 : Évolution des points moyens de la base globale entre 1995 et 2007.....122

Figure 5.8 : Évolution des points moyens de la base TMD entre 1995 et 2007 123

Figure 5.9 : Indice des plus proches voisins (NNI) en fonction de l'ordre K 125

Figure 5.10 : Nearest Neighbor Hierarchical Clustering pour un nombre minimal d'accidents « grand » (bleu = base globale, rouge = TMD).....127

Figure 5.11 : Nearest Neighbor Hierarchical Clustering pour un nombre minimal d'accidents « petit » (bleu = base globale, rouge = TMD).....128

Figure 5.12 : Zoom sur la région de Montréal du NNH de la base globale pour un nombre minimum d'accidents par cluster « petit » 129

Figure 5.13 : K-Means Clustering pour K=5 (bleu=base globale, jaune=TMD).....131

Figure 5.14 : K-Means Clustering pour K=30 (à gauche résultat pour la base Globale, à droite pour la base TMD) 132

Figure 5.15 : Estimation de la densité par la méthode du noyau, échelle comparable (bleu=base globale, rouge=base TMD) 134

Figure 5.16 : Estimation de la densité par la méthode du noyau pour la région entre Montréal et Québec (bleu=base globale, rouge=base TMD).....135

Figure 5.17 : Densité des accidents, point moyen et ellipse d'écart-type pondéré par le nombre de véhicules impliqués (haut=base globale, bas=base TMD).....137

Figure 5.18 : Estimation de la densité par la méthode du noyau pour le nombre de décès (haut=base globale, bas=base TMD).....138

Figure 5.19 : Estimation de la densité par la méthode du noyau pour le nombre de blessés (haut=base globale, bas=base TMD).....140

LISTE DES SIGLES ET ABRÉVIATIONS

AD	Aire de diffusion
BDD	Base de données
CIRANO	Centre Interuniversitaire de Recherche en Analyse des Organisations
CRPQ	Centre de Renseignements Policiers du Québec
CSST	Commission de la Santé et de la Sécurité du Travail (au Québec)
MD	Matière Dangereuse
MHIDAS	Major Hazardous Incident Data Service
MTQ	Ministère des Transports du Québec
NNI	Nearest Neighbor Indices (Indice du plus Proche Voisin)
NNH	Nearest Neighbor Hierarchical clustering (regroupement hiérarchique par plus proche voisin)
RE	Région Économique
SAAQ	Société de l'Assurance-Automobile du Québec
SIG	Système d'Information Géographique (GIS en anglais)
TMD	Transport des Matières Dangereuses
VB	Visual Basic

LISTE DES ANNEXES

ANNEXE 1 – Les neuf classes de matières dangereuses ……..……..…151

ANNEXE 2 – Structure du fichier de la base globale .…….……...152

ANNEXE 3 – Structure du fichier de la base TMD ……...…….....154

ANNEXE 4 – Signification du code d'impact ……..……..…….....156

ANNEXE 5 – Procédures VB permettant d'identifier les accidents de la base TMD dans la base globale ...…….…….158

ANNEXE 6 – Description des paramètres de statistique descriptive utilisés dans l'analyse comparative des données ……...…….....160

ANNEXE 7 – Définition des principales subdivisions du territoire Québécois...……....161

ANNEXE 8 – Distribution mensuelle des accidents ……...……162

ANNEXE 9 – Répartition spatiale des accidents ……..…….....165

ANNEXE 10 – Diagramme de Gantt de la maitrise ……167

INTRODUCTION

Le développement de la société et de l'industrie nécessite de plus en plus de matières dites « dangereuses », c'est-à-dire de substances qui par leurs propriétés physiques ou chimiques peuvent présenter un danger pour l'homme et pour l'environnement. Ces substances, d'une grande variété qui peuvent être inflammables, toxiques, explosives, corrosives ou encore radioactives, sont ainsi quotidiennement manipulées, consommées, chargées, déplacées ou déchargées.

Chacune de ces opérations présente un risque supplémentaire et en particulier le transport de ces matières par voie routière qui n'est pas exempt de collision, dérapage, capotage ou toute autre forme d'accident routier. Du fait de la présence de matières dangereuses, chacun de ces accidents comporte un sérieux risque d'aggravation. Par exemple un simple accrochage peut rapidement conduire, en fonction des circonstances, à une situation particulièrement dangereuse tant pour les personnes impliquées dans l'accrochage que pour les autres usagers, voire pour les résidents à proximité du lieu de l'accrochage. Car en plus du choc, ce type d'accident comporte un risque d'aggravation que l'on peut qualifier de « primaire », immédiatement ressentie, comme une explosion ou un déversement ainsi qu'une combinaison d'effets « secondaires » comme la propagation d'un nuage toxique ou la pollution des sols.

Ces situations sont particulièrement délicates à gérer, avec des conséquences humaines et environnementales potentiellement graves. D'autant que de mauvaises décisions peuvent encore venir aggraver la situation, comme, par exemple, lorsque le comportement de la matière dangereuse est mal apprécié par les équipes d'intervention d'urgence.

Plusieurs accidents récents ont impliqué ces matières au Québec, on peut citer par exemple l'accident qui a eu lieu à Brossard, le 23 octobre 2008, dans la bretelle d'autoroute qui permet de passer de la 10 à la 30. Un camion-citerne rempli de propane a violemment percuté un lampadaire avant de se renverser en travers de la route. Fort heureusement il n'y a eu aucun déversement de son chargement, hautement inflammable à température ambiante. En raison des risques d'explosion, un large périmètre de sécurité a été mis en place. L'autoroute 30 a dû être complètement fermée à la circulation entre le boulevard Matte et la Grande-Allée pendant 18 heures pour laisser le temps au personnel d'urgence de vider le camion de son contenu puis de le remettre sur ses roues.

Figure 0.1 : Renversement d'un camion-citerne à Brossard le 23 octobre 2008[1]

Fort heureusement le pire a pu être évité lors de cet accident mais la présence de matières dangereuses a tout de même sérieusement compliqué et ralenti l'intervention des équipes d'urgence. On peut aussi rappeler un autre accident avec cette fois « participation » de la matière dangereuse, survenu le 12 octobre 2005 sur l'autoroute 40 aux environs du viaduc Alfred-Nobel sur l'île de Montréal. Un camion semi-remorque chargé de 2 700 kg d'hydrosulfate de sodium (UN 1384), s'est renversé sur le côté avant de prendre feu sur la voie de desserte de l'autoroute 40. L'autoroute a là aussi dû être fermée dans les deux sens et les pompiers ont eu beaucoup de mal à éteindre le feu car ce produit dégage un gaz hautement toxique au contact de l'eau. Ces deux événements nous montrent bien que l'utilisation de ces substances n'est pas sans risque et que l'on ne peut pas se permettre de faire n'importe quoi avec ces matières. Ainsi afin d'éviter des accidents aux conséquences dramatiques tant pour l'homme que pour l'environnement il est essentiel que les pouvoirs publics encadrent de façon pertinente l'utilisation, le stockage et le transport des matières dangereuses.

Figure 0.2 : Semi-remorque chargé d'hydrosulfate de sodium en feu[2]

Ce mémoire est structuré en cinq chapitres. Le premier chapitre dévoile la problématique et les objectifs de cette étude. Ensuite le second chapitre passe en revue les principaux résultats des publications du domaine et le troisième chapitre présente la méthodologie qui a été utilisée. Enfin les deux derniers chapitres exposent les résultats de l'étude, selon les deux axes de recherche qui ont été choisis.

---

[1] http://espace.canoe.ca/mediathunder2005/album/view/364412, consulté le 19 octobre 2010

[2] http://lcn.canoe.ca/lcn/infos/regional/archives/2005/10/20051012-130446.html, consulté le 19 octobre 2010

CHAPITRE 1 PROBLÉMATIQUE ET MISE EN CONTEXTE

Au Québec, depuis quelques années, le groupe Risque du CIRANO (Centre Interuniversitaire de Recherche en Analyse des Organisations) mène plusieurs projets de recherche sur différents aspects de la gestion des risques liés au stockage et au transport de matières dangereuses. Ces travaux impliquent de nombreux partenaires, tant gouvernementaux (fédéral et provincial), que municipaux et industriels.

L'ensemble des discussions, échanges et travaux qui ont été réalisés ont mené au lancement du projet de recherche intitulé GESRISQ pour « gestion des risques liés au stockage et au transport de matières dangereuses ». L'objectif global du projet est de trouver un cadre durable de développement des transports et de l'industrie des matières dangereuses, en produisant des recommandations pour les évolutions réglementaires futures. En ce sens un guide de bonnes pratiques pour le transport des matières dangereuses à l'intention des transporteurs et des industriels doit être réalisé ainsi qu'un outil d'aide à la décision concernant la sélection des transporteurs de matières dangereuses et la gestion des contrats de sous-traitance des entreprises.

Ce projet GESRISQ a été structuré en cinq volets et le présent travail s'insère dans le second volet du projet. Le premier volet concerne l'analyse des stratégies logistiques pour le transport de matières dangereuse au Québec, le second, qui nous intéresse particulièrement ici; a pour but l'analyse des accidents impliquant des matières dangereuses. Un troisième volet s'intéresse à l'analyse et à la modélisation des réglementations liées au stockage et au transport de matières dangereuses au Québec. Le quatrième volet vise à dresser le portrait socio-économique des secteurs industriels employant des matières dangereuses au Québec et enfin le cinquième sera l'aboutissement des quatre volets précédents en permettant de traduire leurs conclusions en outils et guides de bonnes pratiques pour améliorer la gestion des risques liés au stockage et au transport de MD.

Le contexte de l'étude étant maintenant précisé, la problématique puis les buts et objectifs vont être décrits dans les deux sections suivantes. On trouvera ensuite les définitions des principaux termes du domaine utilisés puis une description rapide du cadre législatif au Québec.

1.1 Problématique

Comme nous l'avons évoqué en introduction, les conséquences potentielles d'un accident impliquant des matières dangereuses sont telles que les autorités ne peuvent se permettre de ne pas encadrer spécifiquement ce secteur du transport. Il est cependant clair que pour mettre en place des règles à la fois efficaces et adaptées aux réalités économiques, il est indispensable d'étudier le transport des matières dangereuses (TMD).

On trouve dans la littérature de nombreuses études concernant le TMD, mais rare sont les études s'intéressant au TMD québécois, alors qu'il est probable que des spécificités régionales comme le contexte économique particulier du Québec ou les conditions hivernales rigoureuses ont des impacts majeurs sur le TMD. Les travaux récents du CIRANO dans ce domaine, réalisés dans le cadre du projet GLOBAL, avec le soutien du ministère des Transports du Québec, vise justement à combler ce manque, en étudiant de nombreux aspects de ce secteur au Québec, comme les règlementations, l'analyse des bases de données, les coûts du TMD, la santé et la sécurité du travail, et le stockage et la logistique des MD. Ce projet « global » a mis en évidence certaines lacunes et souligné la nécessité d'approfondir les travaux réalisés, ce qui a donné naissance au projet GESRISQ dans lequel s'inscrit ce travail. En particulier l'analyse comparée de bases de données d'accidents concernant le Québec, le Canada et les États-Unis a mis en évidence que les données québécoises n'étaient pas adéquates, trop disparates et imprécises, pour étudier de façon précise ce phénomène (de Marcellis-Warin, Leroux, Peignier, & Trépanier, 2006).

On trouve par ailleurs peu de statistiques, au Québec, sur les accidents impliquant les véhicules de transport des matières dangereuses. Il existe pourtant de nombreuses réglementations et mesures spécifiques concernant le transport des matières dangereuses (de Marcellis-Warin, Leroux, et al., 2006), mais leur impact, positif ou négatif, est difficile à évaluer.

Une bonne façon d'estimer cet impact serait d'étudier comparativement des bases de données d'accidents impliquant des véhicules transportant des matières dangereuses et des véhicules lourds conventionnels afin d'en savoir un peu plus sur l'impact de la présence des matières dangereuses. Cependant dans le monde peu d'études ont mené ce genre d'analyse comparative, et au Québec, à notre connaissance, aucune étude ne s'est intéressée à cet aspect du TMD. Nous allons nous pencher sur cet aspect du problème.

1.2 Buts et objectifs

Tous d'abord l'objectif général de cette étude est de faire avancer les connaissances sur les accidents impliquant des matières dangereuses afin d'aider les autorités à mieux encadrer cette activité et à se rapprocher d'un compromis « optimal » entre risque et coût.

L'objectif plus précis de cette étude est, comme évoqué dans la problématique, de déterminer l'influence de la présence de matières dangereuses sur l'accidentologie des véhicules lourds au Québec. Cette étude s'appuiera sur des données fournies par le ministère des Transports du Québec et consistera à comparer une base d'accidents impliquant tous types de véhicules lourds à une base d'accidents TMD.

On étudiera dans un premier temps de façon statistique les informations contenues dans les champs de ces deux bases de données. En examinant l'ensemble des champs « comparables » on espère identifier des éléments caractéristiques de chacun de ces événements, en utilisant des méthodes empruntées aux statistiques. On s'intéressera en particulier à l'évolution dans le temps du phénomène, ainsi qu'aux conséquences humaines des accidents :

- Quelle est la tendance en termes de nombre d'accidents par an?
- Y a-t-il des mois, des heures ou des jours plus favorables aux accidents TMD qu'aux accidents de véhicules lourds?
- Y a-t-il plus de blessés ou de morts pour les accidents TMD?

On n'oubliera pas de s'intéresser aussi aux autres attributs des bases de données qui concernent les circonstances, les conditions ou encore les causes des accidents.

Dans un second temps on étudiera les répartitions spatiales des accidents. Il faudra pour cela localiser les accidents, puis identifier dans la littérature des méthodes de comparaison spatiale pertinentes afin de caractériser les répartitions obtenues et de déterminer les similarités et les différences d'un point de vue spatial.

1.3 Définitions

Dans cette section nous allons définir ce que l'on entend par « matières dangereuses », par « transport des matières dangereuse » et enfin par ce que l'on définit comme un accident impliquant ou n'impliquant pas des matières dangereuses.

1.3.1 Les matières dangereuses

Les matières dangereuses sont « des substances qui, soit par leurs propriétés chimiques ou celles de leurs composants, soit par la nature des réactions physiques qu'elles sont susceptibles de mettre en œuvre, présentent un risque pour l'homme ou son environnement » (Saint-Éloi, 2004).

L'ONU (Organisation des Nations Unies) en tant qu'institution qui établit les règles du transport international de ces matières, a établi une classification en neuf classes (elles-mêmes subdivisées en divisions) des matières dangereuses selon le risque principal qu'elles présentent. L'ordre des classes et divisions est sans rapport avec le degré de dangerosité des matières. D'autre part les marchandises dangereuses sont identifiées par l'ONU grâce à un numéro ONU (ou UN) à quatre chiffres et par une désignation officielle de transport. La classification ONU ainsi que la liste officielle des numéros et des désignations se trouve dans le volume I des « recommandations relatives au Transport des Marchandises Dangereuses » (seizième édition révisée) (Nations Unies, 2009). L'annexe 1 présente les neuf classes de cette classification.

1.3.2 Transport de Matières Dangereuses (TMD)

Le transport de matières dangereuses (TMD) désigne le déplacement d'un endroit à un autre d'une matière répertoriée par l'ONU comme matière dangereuse. Le TMD peut s'effectuer par voie routière, par chemin de fer, par voie maritime (ou fluviale), par canalisation ou encore par avion. Les règles du TMD sont fixées par des accords internationaux ainsi que par des spécificités nationales. Transport Canada estime à environ 30 millions le nombre de chargement de matières dangereuses en transit par route, rail, air et voie maritime par an au Canada (Transports Canada, 2009b). Erkut et al. estiment qu'en 2002, 99% de ces transferts de matières dangereuses se sont passés sans problème au Canada, ce qui fait du TMD un des secteurs du transport les « plus sûr » (Erhan. Erkut, Tjandra, & Verter, 2007).

1.3.3 Accidents et accidents MD

Un accident est un « évènement fortuit qui implique un ou plusieurs véhicules » causant des préjudices à la santé humaine et/ou à l'environnement et/ou aux véhicules impliqués (de Marcellis-Warin, Leroux, et al., 2006).

Nous considérerons, dans cette étude, qu'un accident implique des matières dangereuses si au moins un véhicule impliqué est un véhicule enregistré par la SAAQ comme un véhicule de transport de MD. Un accident considéré comme impliquant des matières dangereuses (accident TMD) peut ainsi être avec ou sans participation de la matière dangereuse. Sans participation de la MD, l'accident a alors les mêmes conséquences qu'un accident routier « normal ». Mais lorsqu'il y a participation de la matière dangereuse les conséquences prennent généralement les formes suivantes (potentiellement combinées) :

Tableau 1.1 : Principales conséquences d'un accident MD, adapté de (Brun-Picard, Gominet, & Louis, 2008)

	Causes	**Conséquences**
Explosion	Choc, étincelles, échauffement d'une cuve, mélange de MD, allumage d'explosifs ou de munitions	Effet thermique et mécanique à cause de l'onde choc
Nuage Toxique	Déversement, résultat d'une combustion	Intoxication
Dispersion dans le sol ou dans l'eau	Déversement	Pollution du sol, de l'eau, de la nappe phréatique, intoxication par ingestion directe ou indirecte
Incendie	Choc, étincelles, échauffement, inflammation d'une fuite, explosion,…	Effets thermiques, émission de fumées toxiques
Radioactivité	Déversement de matières radioactives	Émission de rayonnement

Les conséquences sur la santé humaine vont d'une simple gêne à la mort en passant par des blessures graves ou légères.

1.4 Cadre législatif en vigueur au Québec

Nous allons maintenant décrire brièvement le cadre législatif en vigueur au Québec, ce qui nous permettra de mieux appréhender les informations contenues dans les bases de données.

Chaque mode de transport possède une réglementation propre, mais dans le cadre du TMD les législations sont très abondantes, ce qui fait toute la complexité du secteur. L'objectif de cette réglementation est d'assurer un transport sécuritaire des matières dangereuses et de régir l'ensemble des opérations de transport, de l'expédition à la réception. Ces nombreuses contraintes font que les barrières d'entrée dans le secteur sont assez importantes et que la plupart des opérations de transport sont réalisées par des entreprises spécialisées. On trouve au Québec une centaine d'entreprises opérant dans ce secteur.

1.4.1 Rôle des différentes institutions

Le rapport du projet global intitulé « Revue des réglementations applicables au stockage et au transport des matières dangereuses au Québec » réalisé par l'équipe du CIRANO (de Marcellis-Warin, Favre, Peignier, & Trépanier, 2006) détaille en profondeur les réglementations en vigueur au Québec. Selon ce rapport le domaine routier relève de la compétence du gouvernement québécois, le Ministère des Transports du Québec (MTQ) édite ainsi un « Règlement sur le transport des matières dangereuses » (Ministère des Transports du Québec, 2010) qui fait partie du Code de la sécurité routière (article 622) et qui s'applique sur tous les chemins publics. Cependant en vertu des lois fédérales ce règlement reprend les normes du « Règlement sur le Transport des Marchandises Dangereuses » (RTMD) édité par le Ministère des Transports du Canada (Transports Canada, 2010) qui lui s'applique aux entreprises. Cela permet une harmonisation des réglementations dans l'ensemble du Canada mais également dans les échanges avec les États-Unis. Pour simplifier voici un organigramme du TMD provenant du rapport du projet Global :

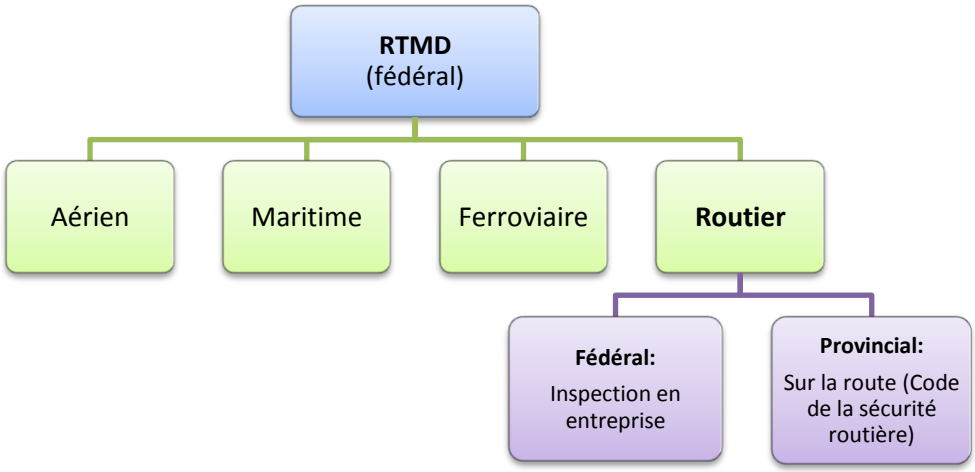


Figure 1.1 : Organigramme du TMD (de Marcellis-Warin, Favre, et al., 2006)

Le « Guide sur le transport des matières dangereuses » (Direction de la sécurité en transport et du camionnage, 2006), permet d'avoir un rapide aperçu de la réglementation. Le guide décrit la documentation obligatoire, les contenants et véhicules autorisés pour le transport, les indications de dangers à apposer sur les contenants, le franchissement des tunnels et passage à niveau, la procédure obligatoire en cas de rejet, la liste des exemptions,…

1.4.2 Les contenants

Les matières dangereuses doivent être transportées dans des contenants normalisés, c'est-à-dire qui ont été fabriqués selon des normes qui garantissent la qualité de la conception et de la fabrication afin d'éviter tout rejet. Les contenants doivent aussi être inspectés et testés à des intervalles précis. On distingue trois catégories de contenants selon leurs tailles :

Tableau 1.2 : Principaux contenants de matières dangereuses (Direction de la sécurité en transport et du camionnage, 2006)

	Forme	Capacité
Petits contenants	Cylindres, jerricans, seaux, tonneaux, futs ou bouteilles	450 litres d'eau
Grands contenants	Citernes routières, citernes amovibles, grands récipients	Entre 450 et 3 000 litres d'eau
Camion-citerne	Citerne	16000 litres max pour une citerne

1.4.3 Les indications de danger

Des indications de danger doivent être apposées sur les contenants servant au TMD. Ces indications prennent la forme d'étiquettes, de plaques (pour des contenants d'une capacité de plus de 450 litres) ou de marques et signes. Elles doivent indiquer la classe primaire et subsidiaire de la matière dangereuse ainsi que son numéro ONU. En annexe 1 se trouve un tableau indiquant les neuf classes de matières dangereuses ainsi que le pictogramme correspondant, en voici deux exemples à apposer sur le contenant d'une matière comburante (classe 5.1) et d'une matière corrosive (classe 8) :

Figure 1.2 : Exemple de plaques d'indication de danger (Direction de la sécurité en transport et du camionnage, 2006)

Maintenant que nous avons introduit et présenté l'objet de cette étude, nous allons passer en revue les principaux résultats des études publiées qui concernent notre problématique.

CHAPITRE 2 REVUE DE LITTERATURE

Le transport des matières dangereuses (TMD) se distingue fortement des autres secteurs du transport, en raison du caractère dangereux de l'objet transporté, et donc des conséquences potentiellement graves d'un accident, et d'une réglementation complexe. Pour ces raisons le TMD fait l'objet de nombreuses recherches à travers le monde. L'objectif commun de ces études est d'améliorer la connaissance du TMD, en vue d'adapter les législations et de réduire les risques. A l'aide des outils de recherche de la base de données Compendex, on peut se faire une idée du nombre de publications par année. En définissant une recherche expert, avec des mots clé désignant les accidents TMD, environ 4200 résultats ont été obtenus dont voici la décomposition par année :

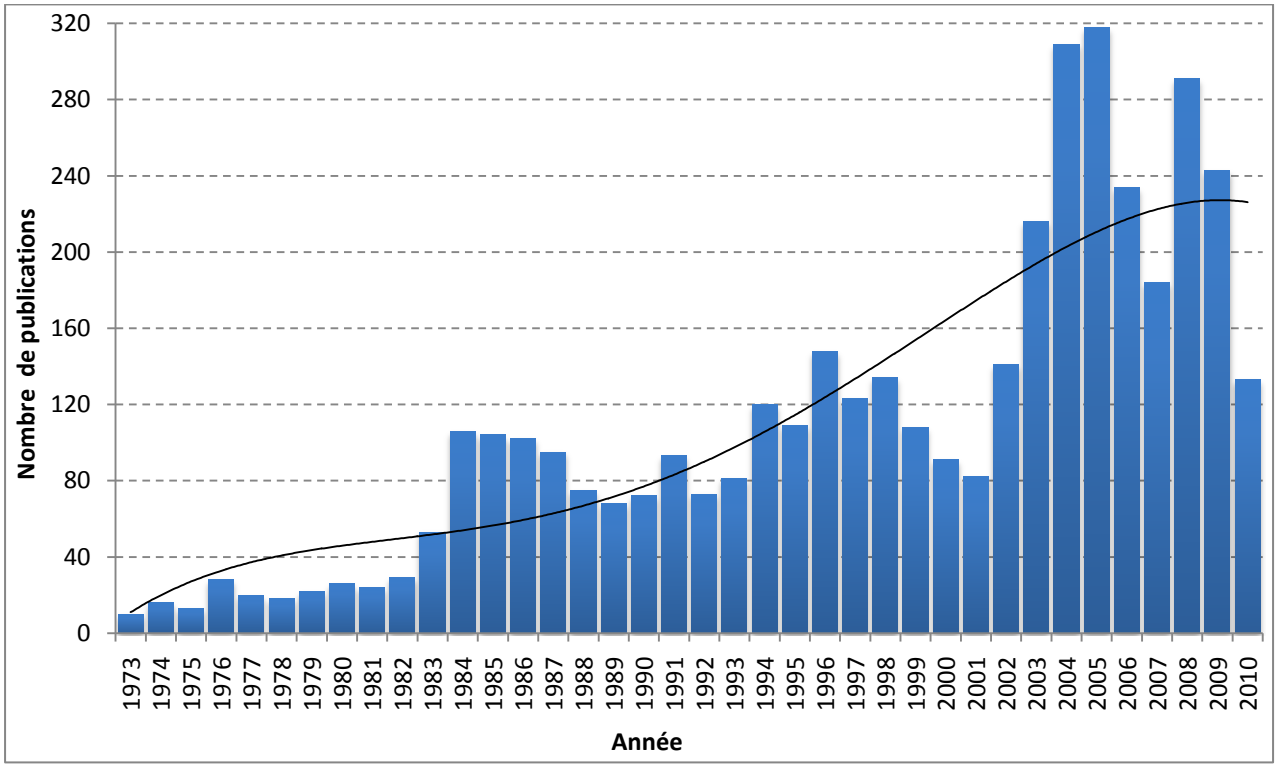


Figure 2.1 : Nombre de publications par année selon la base de données Compendex, concernant le transport des matières dangereuses

Cette courbe montre que l'intérêt pour le TMD a réellement commencé dans les années 80. La tendance est depuis à la hausse avec un pic de publications important en 2005. Le pays le plus

producteur est, selon Compendex, de très loin les États-Unis. Les principaux thèmes abordés sont l'évaluation des risques, le routage des chargements, l'ordonnancement et la conception du réseau. Tous ces travaux ne concernent pas directement notre étude. Dans cet état des lieux des recherches effectuées, nous allons nous intéresser dans un premier temps aux études descriptives du TMD (§ 2.1), qui visent à découvrir les causes et les conséquences des accidents, puis aux travaux d'analyses spatiales des répartitions d'accidents (§ 2.2). Le troisième paragraphe s'intéressera aux différentes façons de travailler avec les bases de données et aux problèmes rencontrés par les auteurs dans ces opérations (§ 2.3).

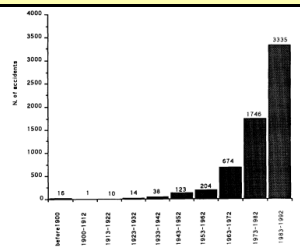
2.1 Études descriptives du TMD

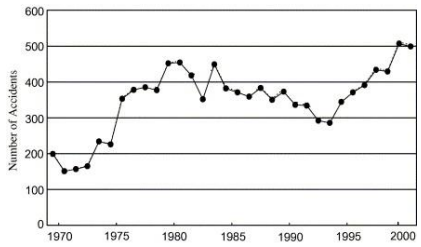
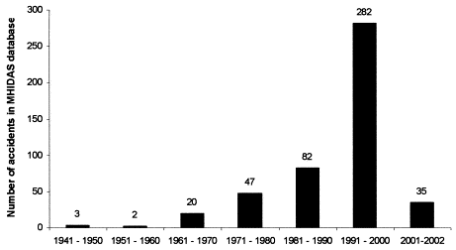
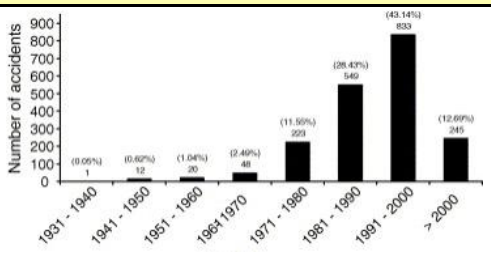
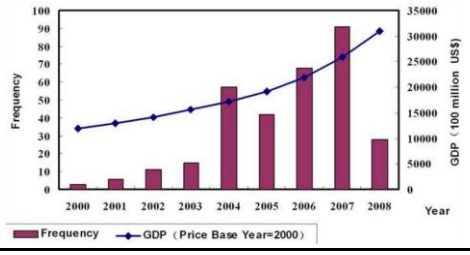
L'étude descriptive du transport des matières dangereuses est une préoccupation majeure des recherches dans l'étude du TMD. Elle consiste à examiner des bases de données d'accidents afin de déterminer les causes possibles et les conséquences des accidents. Afin de favoriser une approche comparative les résultats des travaux ont été regroupés en six thèmes : la distribution dans le temps des accidents, le type de transport, les matières dangereuses impliquées, les conséquences des accidents, les types d'accidents et les circonstances des accidents. Les résultats d'une même étude peuvent donc se retrouver dans plusieurs sections. Par ailleurs, pour chacune de ces sections des exemples concernant le Québec ont été favorisés.

2.1.1 Distribution dans le temps des accidents

De nombreuses études s'intéressent à l'évolution dans le temps du nombre d'accidents, le tableau suivant présente les courbes d'évolution annuelles de quelques études du TMD :

Tableau 2.1 : Comparaison des distributions temporelles d'accidents dans plusieurs études

(Vílchez, Sevilla, Montiel, & Casal, 1995)		
Nombre d'accidents	5325	
Période :	[1900 - 1992]	
Zone géographique	Monde	
BDD :	MHIDAS	

(Ohtani & Kobayashi, 2004)		
Nombre d'accidents	>10 000	
Période :	[1970 - 2001]	
Zone géographique	Japon	
BDD :	From Fire and Disaster Management Agency	
(Darbra & Casal, 2004)		
Nombre d'accidents	471	
Période :	[1941 – 2002]	
Zone géographique	Monde (zone portuaire)	
BDD :	MHIDAS	
(Oggero, Darbra, Muñoz, Planas, & Casal, 2006)		
Nombre d'accidents	1932	
Période :	[1931 – 2004]	
Zone géographique	Monde	
BDD :	MHIDAS	
(Yang et al., 2010)		
Nombre d'accidents	322	
Période :	[2000 – 2008]	
Zone géographique	Chine	
BDD :	Construite par les auteurs	

Comme les courbes précédentes le montrent, de nombreuses études constatent une augmentation des fréquences d'accidents, ce qui peut s'expliquer par la croissance importante des besoins en matières dangereuses dans le monde, et donc de l'augmentation des opérations de transport de matières dangereuses (Oggero, et al., 2006).

Au Canada il y a eu ces dernières années, selon les chiffres de Transport Canada, un peu plus d'une centaine d'accidents routiers avec participation des matières dangereuses. La courbe d'évolution suivante construite à partir des données de Transport Canada montre une tendance

plutôt à la baisse pour ces trois dernières années avec moins de 90 accidents en 2009 (les chiffres de 2009 sont cependant encore préliminaires) :

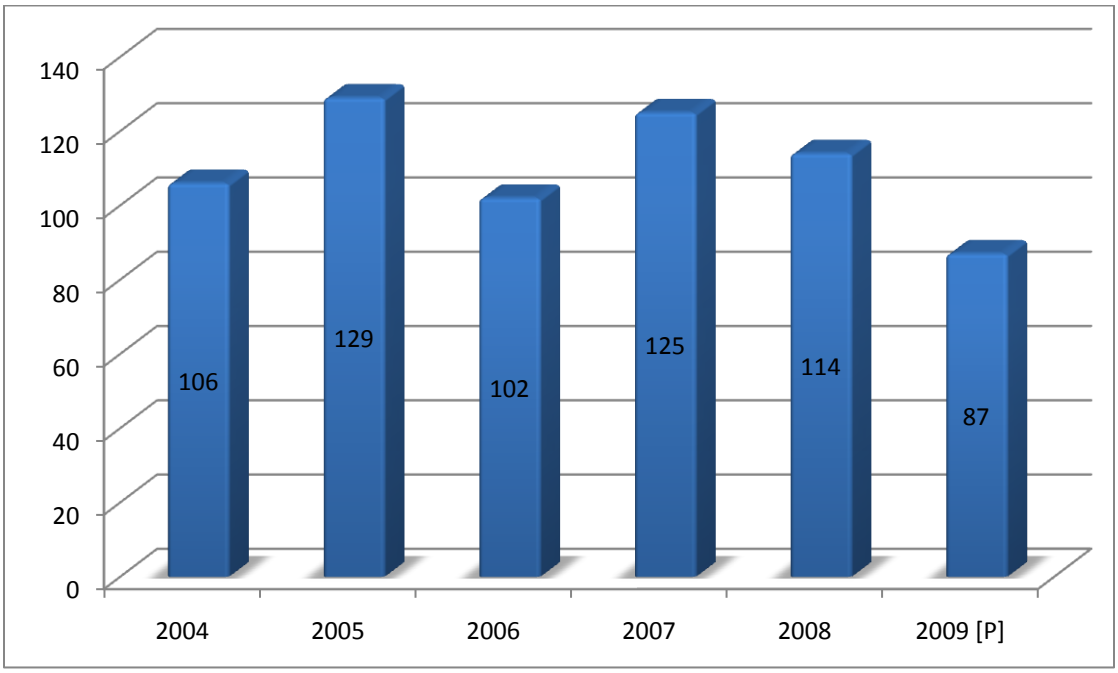


Figure 2.2 : Évolution des accidents routier qui ont mis en cause des marchandises dangereuses au Canada (Transports Canada, 2009a)

En France, Saint-Eloi estime qu'il y a environ 150 à 200 accidents par an impliquant des véhicules transportant des matières dangereuses, soit 1,5% des accidents de poids lourds alors que ce trafic représente plus de 15% de la circulation des poids lourds. Il obtient donc un taux d'accident 10 fois faible pour les véhicules TMD que pour les autres véhicules (Saint-Éloi, 2004). Le BARPI (Bureau d'Analyse des Risques et Pollutions Industriels) considère toutefois dans son « Inventaire 2010 des accidents technologiques » qu'en France, le taux d'accidents des véhicules transportant des matières dangereuses est seulement trois fois faible que pour l'ensemble des poids lourds (Bureau d'Analyse des Risques et Pollutions Industriels (BARPI), 2010).

Enfin au Québec, selon les chiffres de l'étude menée par Boies et Beaulieu, 2 317 accidents impliquant des véhicules de transport de matières dangereuses ont été recensés sur la période 1995-2007 avec une moyenne de 178 accidents TMD par an (Boies & Beaulieu, 2009). Le maximum a été observé en 1995 avec 208 accidents et le minimum en 1998 avec 152 accidents. Ces chiffres peuvent sembler important comparativement aux chiffres français notamment, mais cela peut s'expliquer par une définition plus large des accidents TMD au Québec.

- Conditions hivernales

Boies et Beaulieu notent qu'au Québec les conditions hivernales semblent être plus propices aux accidents TMD, avec près de 46% des accidents pour les seuls mois de décembre, janvier, février et mars (Boies & Beaulieu, 2009). Les victimes sont aussi plus nombreuses dans cette période, le mois le plus dommageable étant mars. L'étude comparative menée par l'équipe du CIRANO sur des bases de données canadiennes et américaines montrent par ailleurs que cette caractéristique se vérifie pour la totalité du Canada mais ne se retrouve aux États-Unis où les mois les plus touchés sont au contraire ceux de l'été (de Marcellis-Warin, Leroux, et al., 2006). Il semble donc que l'hiver rigoureux joue un rôle important en favorisant les accidents. Ainsi Boies et Beaulieu conseillent aux conducteurs d'être particulièrement prudent dans ces conditions, et invitent les responsables à tout mettre en œuvre pour que les chauffeurs partent « l'esprit tranquille », consultent les prévisions météo, adaptent leur conduite, vérifient l'état des pneus,…

- Répartition horaire et journalière

Certains travaux s'intéressent plus particulièrement à la répartition de ces accidents à l'échelle hebdomadaire ou quotidienne. Selon les travaux menés par Shorten et al. en 2002 sur les accidents TMD survenus dans le comté de Chester en Pennsylvanie, vendredi est le jour où il y a le plus d'accidents et dimanche le jour où il y en a le moins (Shorten, Galloway, Krebs, & Fleming, 2002). Ce résultat est confirmé par les travaux de (de Marcellis-Warin, Leroux, et al., 2006) sur des bases d'accidents américaines, canadiennes et québécoises. Selon Shorten et al., cela reflète que le vendredi la prudence des conducteurs est plus faible, peut-être à cause de la fin de semaine toute proche.

En ce qui concerne la répartition selon les heures Shorten et al. (2002) ont montré que les accidents TMD ont le plus souvent lieu au milieu de la journée, lorsque le trafic est le plus intense selon une répartition presque parabolique comme le montre la courbe ci-dessous :

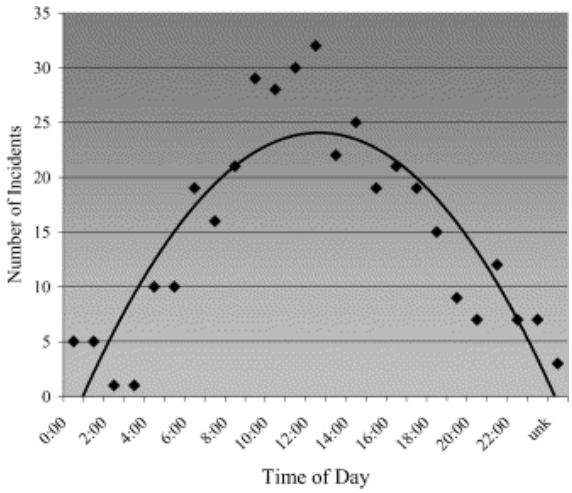


Figure 2.3 : Répartition horaire des accidents TMD dans le comté de Chester en Pennsylvanie (Shorten et al, 2002)

Au Québec, la courbe obtenue par Boies et Beaulieu présente des différences étonnantes. En effet si les accidents ont bien majoritairement lieu en journée, ils constatent deux pics d'accident dans la journée (qu'ils qualifient « d'heures critiques »), le premier vers 10h et le second à 15h avec entre ces heures un creux correspondant à 12h qui est totalement absent de la courbe précédente :

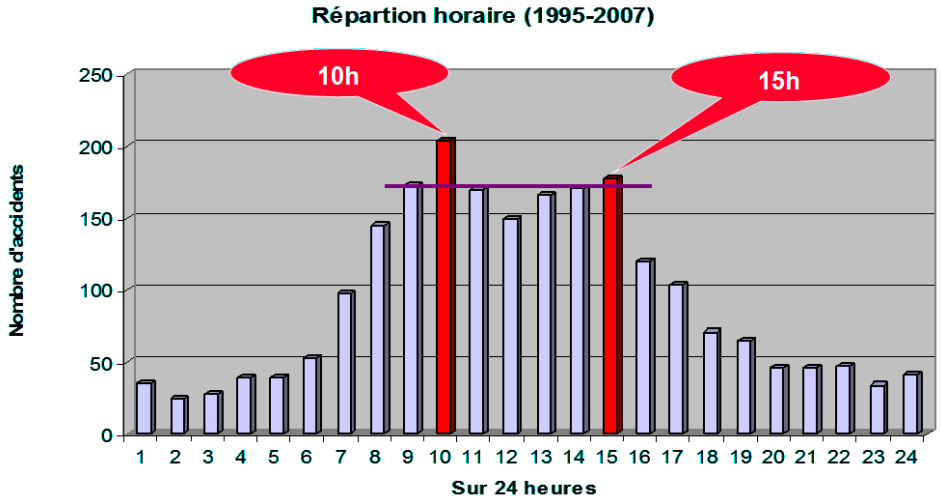


Figure 2.4 : Répartition horaire des accidents TMD au Québec (Boies & Beaulieu, 2009)

Cela montre certainement des habitudes de vie différentes, avec une pause de midi plus marquée au Québec qu'en Pennsylvanie.

Nous allons maintenant nous intéresser aux distributions des accidents TMD selon les types de transports.

2.1.2 Type de transport

L'influence du type de transport sur les accidents TMD est souvent abordée par les auteurs. Ainsi de nombreuses études comparent la dangerosité des différents modes de transport. Selon Erkut et al, aux USA près de 94% des transferts de MD se font par voie routière, mais ces transferts ne concernent que 43% du tonnage total de MD en transit. Les 57% restant sont essentiellement transportés par rail, bateaux et pipelines, qui sont des modes de transport permettant de transporter de grosses quantités de marchandises en une fois, contrairement aux camions (Erhan. Erkut, et al., 2007).

Au Canada, selon Transport Canada, plus de la moitié des chargements de matières dangereuses sont déplacés par voie routière (Transports Canada, 2009b). La part du transport routier dans les accidents mettant en cause des matières dangereuses est par contre prépondérante, avec environ 90% des accidents contre 6% pour le transport ferroviaire et 4% pour le transport aérien (Transports Canada, 2009a). Ces proportions ne semblent cependant pas être les mêmes selon les régions du monde. En effet une étude comparant le TMD routier et ferroviaire à partir d'une base d'accidents recensant des événements du monde entier a permis d'évaluer à 63% la proportion d'accidents routiers et 37% celle d'accidents ferroviaires (Oggero, et al., 2006). Dans des travaux antérieurs qui comprenaient en plus le transport maritime et par canalisation, sur 691 accidents impliquant des matières dangereuses, 30% des accidents sont survenus sur la route et 37% sur le rail (Haastrup & Brockhoff, 1990). En ne considérant que les accidents routiers et ferroviaires cela nous donne des proportions de 45% pour les accidents routiers et de 55% pour les accidents ferroviaires, ce qui est totalement différent. Cependant cette étude date de 1990 et ces différences peuvent s'expliquer par le fait que le transport routier semble de plus en plus utilisé par rapport aux autres modes de transport, ce qui a entrainé l'inversion du rapport entre les accidents routiers et par rail.

2.1.3 Matières dangereuses impliquées

Dans cette section on aborde la question des matières dangereuses qui sont le plus souvent impliquées dans les accidents, avec ou sans participation effective. Cette démarche est cependant souvent gênée par un manque de précision dans les données, la ou les matières dangereuses transportées n'étant pas toujours identifiées et les quantités étant encore plus rarement indiquées comme le constate Ronza et al. (2003). Ainsi dans le travail de Shorten et al., dont la courbe est

présenté ci-dessous, pour près de 35% des accidents le chargement est inconnu (Shorten, et al., 2002).

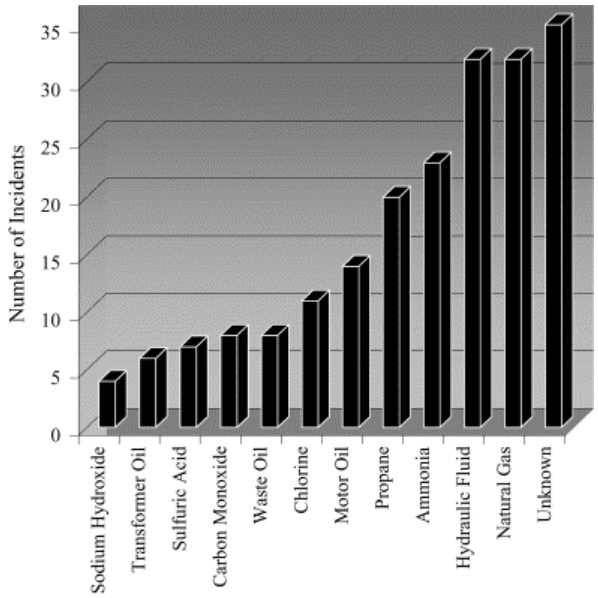


Figure 2.5 : Nombre d'accidents en fonction de la matière dangereuse (Shorten, et al, 2002)

Ils ont cependant montré que dans le comté de Chester en Pennsylvanie, les hydrocarbures sont les matières dangereuses les plus couramment impliquées avec une présence dans 74% dans les accidents TMD.

Les chiffres obtenus en Australie par Winder, et al. sont similaires avec les 2/3 des accidents impliquant des hydrocarbures (Winder, Tottszer, Navratil, & Tandon, 1992). On peut aussi noter qu'en zone portuaire, les proportions sont équivalentes avec 62% des accidents concernant le pétrole brut ou des produits pétroliers (Ronza et al., 2003).

Erkut, et al. se sont intéressés aux classes de matières dangereuses les plus fréquentes. Selon eux les classes 3 (liquides inflammables) et 8 (matières corrosives) de matières dangereuses (cf. annexe 1) sont les plus souvent impliquées avec environ 40% des accidents chacune dans les accidents TMD survenu aux États-Unis en 2003 (Erhan. Erkut, et al., 2007) :

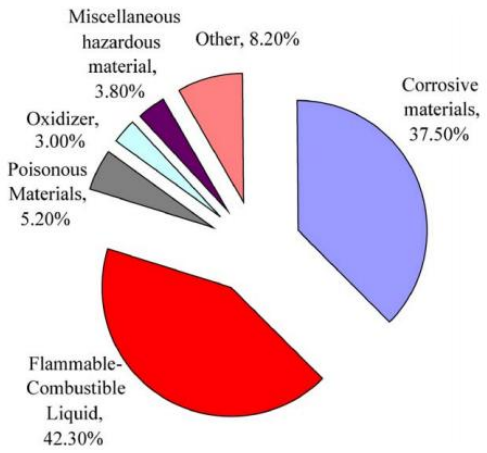


Figure 2.6 : Répartition des classes de MD impliquées dans les accidents aux États-Unis en 2003 (Erhan. Erkut, et al., 2007)

Au Québec, de Marcellis-Warin et al. ont également obtenu que les classes 3 et 8 sont les plus impliquées dans les incidents TMD avec respectivement environ 35% et 30% des accidents. Elles sont suivies par la classe 2 (gaz comprimé) avec 16% des accidents, et par les autres classes qui tournent autour de 5% des accidents (de Marcellis-Warin, Leroux, et al., 2006).

Du point de vue des conséquences, les accidents impliquant des matières explosives (classe 1) sont pour Carol et al. nettement plus sévères que ceux impliquant d'autres classes de matières dangereuses (Carol, Vilchez, & Casal, 2002). Forts de cette constatation, Guo et Verma ont étudié une façon originale de diminuer les risques lors du transport en jouant sur le choix de la capacité des véhicules transportant de l'essence ou des explosifs (Guo & Verma, 2010).

Nous allons maintenant nous intéresser aux conséquences des accidents TMD.

2.1.4 Conséquences des accidents

Cette section s'intéresse aux conséquences des accidents. Il est tout d'abord intéressant de noter que plusieurs études ont constaté que les accidents survenant dans les pays en voie de développement sont plus susceptibles d'avoir des conséquences graves qu'en Europe ou en Amérique du Nord (Haastrup & Brockhoff, 1990) (Oggero, et al., 2006) (Carol, et al., 2002). Cela prouve que le développement d'une société est un facteur important dans la gestion d'accidents TMD.

- Conséquences humaines

En France selon l'« Inventaire 2010 des accidents technologiques » du BARPI, le TMD routier et ferroviaire est à l'origine de 271 accidents mortels en France sur la période 1992 – 2009. 90% de ces accidents sont d'origine routière et la matière dangereuse n'intervient pas directement dans la cause des décès dans plus de 80% des cas. Les matières dangereuses en soi ne sont responsables, selon ce document, « que » d'une cinquantaine de morts en France depuis 1992 (Bureau d'Analyse des Risques et Pollutions Industriels (BARPI), 2010).

Selon Transport Canada, il y a une cinquantaine de victimes dans les accidents impliquant des matières dangereuses au Canada par an comme le montre la courbe ci-dessous :

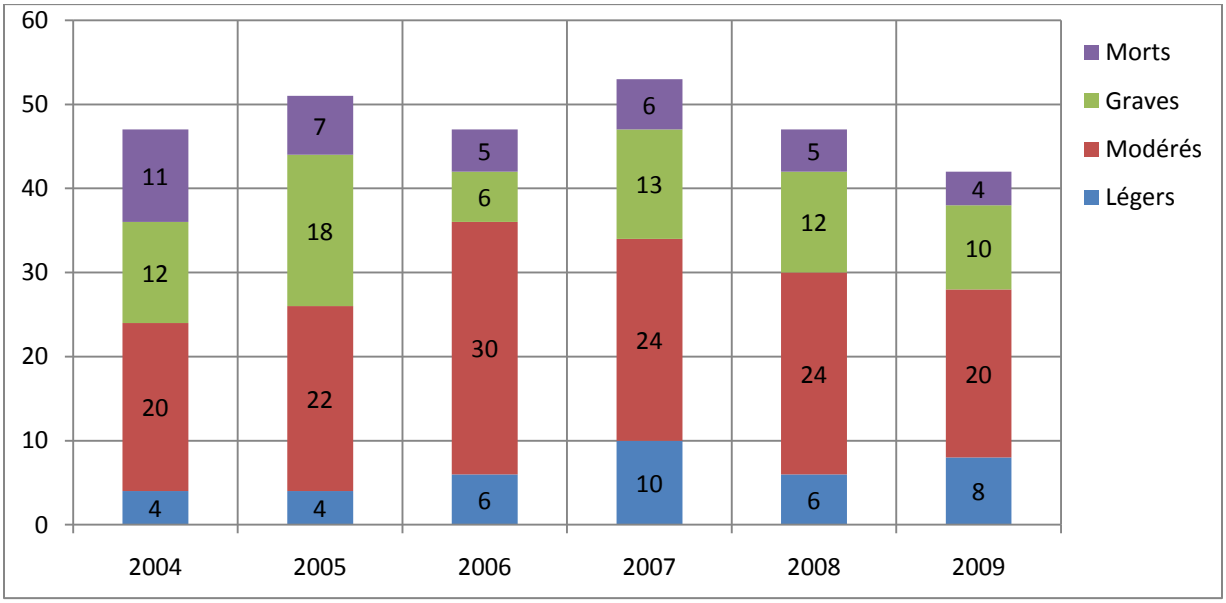


Figure 2.7 : Évolution des nombres de décès et de blessés dans les accidents impliquant des matières dangereuses au cours du transport et lors de la préparation du transport au Canada (Transports Canada, 2009a)

La tendance semble être depuis 2007 à la baisse, comme on l'avait constaté pour le nombre d'accidents par an (cf. § 2.1.1). Il faut cependant noter qu'un accident en 2007 en Ontario (lors du transbordement d'acide chlorhydrique d'un wagon-citerne vers un camion-citerne) est responsable à lui tout seul de 15 blessés modérés. Parmi ces victimes seules une douzaine par an sont directement imputables aux matières dangereuses, (Transports Canada, 2009a).

Pour étudier les conséquences humaines des accidents, les auteurs utilisent fréquemment les courbes f-N, dont la réalisation est détaillée dans la partie méthodologie au § 3.3.2.6. Ces courbes

sont très utiles pour comparer des résultats. Par exemple Evans se sert d'une telle courbe pour comparer le nombre de décès en fonction du mode de transport dans les accidents anglais entre 1967 et 2001 (Evans, 2003) :

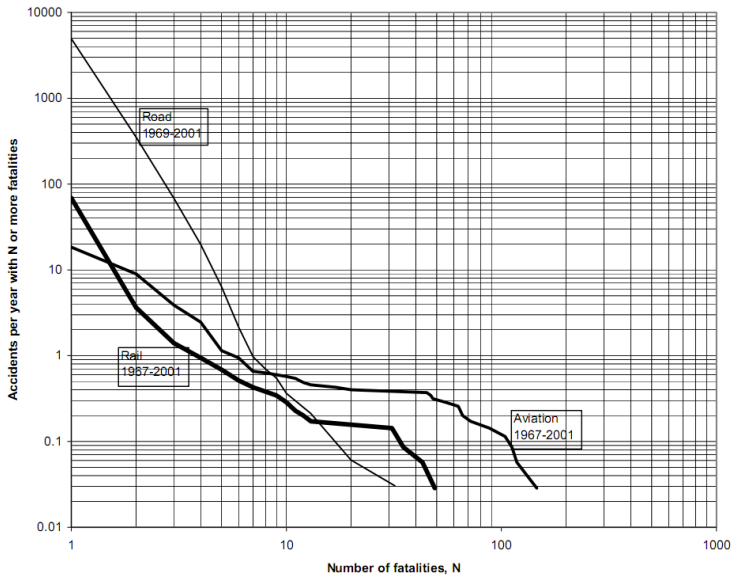


Figure 2.8 : Courbe f-N pour le transport routier, ferroviaire et aérien (1967-2001) (Evans, 2003)

Cette courbe montre que le transport routier provoque avec une plus forte probabilité des accidents entrainant moins de 5 décès, tandis que les transports aérien et ferroviaire ont une probabilité plus importante de provoquer des accidents avec un grand nombre de morts. Généralement les chercheurs obtiennent une courbe décroissante proche d'une droite, ce qui montre une relation de proportionnalité entre le nombre d'accidents et le nombre de morts. La pente de cette droite est alors calculée ce qui permet de comparer facilement les courbes entre elles. Le tableau suivant présente quelques valeurs des pentes qui ont été obtenues :

Tableau 2.2 : Compilation des résultats des courbes p-N cumulative concernant le nombre de décès dans différentes études

Référence	Type d'étude	Pente obtenue
(Haastrup & Brockhoff, 1990)	1793 accidents TMD entre 1960 et 1988	≈ -1
(Vílchez, et al., 1995)	5325 accidents MD (transport + stockage) issus de MHIDAS entre le début du siècle et1992	$-0{,}84$
(Planas-Cuchi, Montiel, & Casal, 1997)	Accident impliquant le feu issu de MHIDAS entre 1900 et 1993	$-0{,}909$

(Darbra & Casal, 2004)	471 accidents MHIDAS en milieu portuaire entre 1900 et 2002	$-0,7$
(Oggero, et al., 2006)	1932 accidents routiers et ferroviaires entre 1931 et 2004	Route : $-0,81$ Rail : $-0,75$
(Yang, et al., 2010)	322 accidents TMD en Chine entre 2000 et 2008	$-0,988,$

Ces pentes tournent toute autour de -1 et il sera intéressant de les comparer aux résultats que l'on obtiendra.

Ces courbes peuvent aussi servir à déterminer si le risque est acceptable en définissant des critères de tolérance. Ces critères définissent ainsi une zone dite ALARP (As Low As Reasonably Practicable), souvent une droite déterminée par 2 points, donnant la probabilité « acceptable » d'un accident entrainant un certain nombre de morts. A l'inverse, certains ont utilisé les critères ALARP Britannique et Néerlandais dans leur étude afin de savoir où se situait la courbe f-N qu'ils ont obtenue par rapport aux tolérances de ces pays (Carol, et al., 2002).

- Conséquences financières

Des estimations monétaires des conséquences des accidents sont très difficiles à obtenir et assez peu d'auteurs s'y risquent. Erkut, et al. ont réalisé une estimation comparée très intéressante des coûts estimés d'un accident en dollars américains en fonction du type d'accident, en prenant en compte les coûts directement liés à l'accident, mais aussi les conséquences plus indirectes comme les délais de livraison ou l'immobilisation du véhicule accidenté (Erhan. Erkut, et al., 2007) :

Type of accident/incident event	Average cost (in US$)	Average traffic delay (in hours)
Nonhazmat events	340,000	2
All hazmat events	414,000	–
Hazmat events with spill/release	536,000	5
Hazmat events with fire	1,200,000	8
Hazmat events with explosion	2,100,000	12

Figure 2.9 : Coûts comparés en US$ de différents types d'accident routier aux États-Unis (Erhan. Erkut, et al., 2007)

On constate ainsi que les accidents impliquant des matières dangereuses ont des coûts généralement plus élevés que les autres accidents, notamment si l'accident provoque un rejet de ces matières comme un incendie ou une explosion.

Au Québec, à partir des données de la CSST, de Marcellis-Warin et al. ont déterminer quelques caractéristique des indemnités versé en fonction du type d'industrie :

Industrie	Nb accidents	Indemnité moyenne[48]	Perte de salaire/ accident[49]	Jours perdus/accident
Industrie pétrolière	196	$16,928	$8,337	136
Industrie chimique	23	$18,746	$10,214	127
Industrie gazière	13	$15,437	$10,765	132
Autres	26	$14,758	$7,114	107

Figure 2.10 : Statistique québécoises sur les indemnités liées aux accidents du travail (en $CAD) (de Marcellis-Warin, Leroux, et al., 2006)

2.1.5 Type d'accident MD

Fort heureusement la matière dangereuse n'intervient pas toujours dans les accidents impliquant des véhicules transportant des MD. Saint-Eloi estime qu'un tiers des accidents TMD en France sont avec participation effective de la matière dangereuse, les 2/3 des accidents étant de simples accidents de la route (Saint-Éloi, 2004). Certaines études s'intéressent cependant au type d'accident survenu lorsque la matière dangereuse est amenée à participer à l'incident.

En Chine 84,5% des accidents sont des rejets suivis par des nuages de gaz (13,0%), des incendies (10,2%), et des explosions (5,9%) (Yang, et al., 2010). De plus toujours selon cette étude, dans 9,9 % des cas, un rejet a pu être évité en raison d'une prise en charge suffisamment rapide. Au niveau mondial, Oggero et al. obtiennent une répartition similaire. Selon eux 78% des accidents sont des rejets, 28% des incendies, 14% des explosions et 6% des nuages de gaz (Oggero, et al., 2006). Le total est ici supérieur à 100% car un accident peut combiner plusieurs types d'évènements. Pour résumer ces chiffres, Oggero et al. ont réalisé un arbre de probabilités qui indique l'ensemble des cas possibles ainsi que leur probabilité :

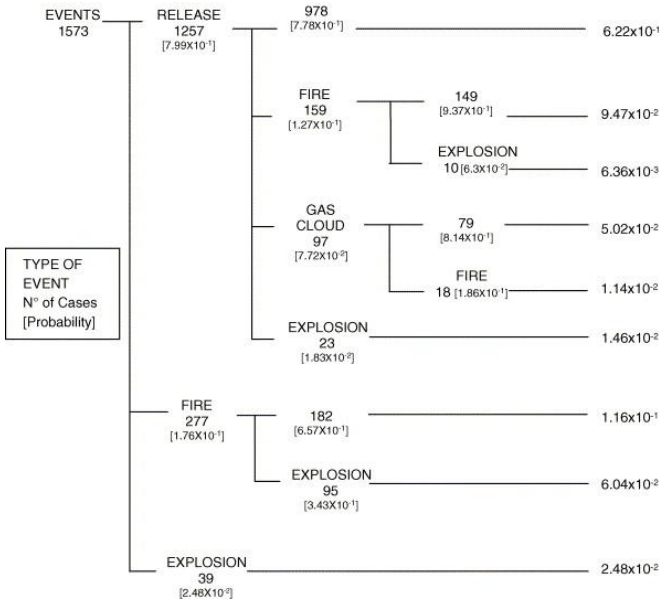


Figure 2.11 : Arbre des probabilités de chaque type d'accident (Oggero, et al., 2006)

En zone portuaire, les proportions sont selon Darbra et Casal légèrement différentes. Les rejets sont toujours le type le plus commun d'accident avec seulement 51% des cas devant les incendies avec un taux de 29%, suivis des explosions (17%) puis des nuages de gaz (3%) (Darbra & Casal, 2004).

Au Canada, les types d'accidents les plus fréquents impliquant des camions transportant de l'essence et du GPL observés en 1986 et 1987 (ce qui correspond aux deux premières années où la loi a rendu obligatoire la déclaration des accidents) sont des renversements, des collisions ainsi que des collisions suivi d'un déversement (Stewart & Van Aerde, 1990). Au Pays-Bas, Tiemessen et Van Zweeden ont constaté la même chose; selon eux le renversement d'un camion-citerne suivi d'une libération de la matière est le type d'accident TMD de transport le plus fréquent (Tiemessen & Van Zweeden, 1998).

2.1.6 Circonstances des accidents

Enfin selon les informations dont ils disposent les chercheurs peuvent aussi décrire les diverses circonstances des accidents TMD, comme le type de route et l'origine de l'accident.

- Type de route

Au Québec, globalement 40% des accidents surviennent dans des conditions « normales » c'est-à-dire le jour sur chaussée sèche (de Marcellis-Warin, Leroux, et al., 2006). De plus, selon Boies et Beaulieu, 55% des accidents ont lieu en ligne droite, 23% à une intersection et 13% dans une courbe. Le mouvement du véhicule TMD avant l'impact est dans 50% des cas « tout droit », dans 14% des accidents, il reculait et dans 13% des accidents il tournait (Boies & Beaulieu, 2009).

Selon Oggero, et al. environ 80% des accidents routiers ont lieu sur l'autoroute (Oggero, et al., 2006). Aux États-Unis ce chiffre est de 90% (Erhan. Erkut, et al., 2007). Par ailleurs, la circulation dans les tunnels des véhicules TMD entrainent selon Bubbico et al. une augmentation non négligeable des risques (Bubbico, Di Cave, Mazzarotta, & Silvetti, 2009). Les auteurs recommandent d'ailleurs de limiter leur circulation ou de l'encadrer lorsque cela est nécessaire en prenant en compte le type de matières dangereuses.

Shorten et al. ont montré que chaque secteur de la route 30 qui traverse le comté Chester en Pennsylvanie possède la même probabilité d'accident (Shorten, et al., 2002). Ils ont pour cela réalisé une régression linéaire astucieuse entre le nombre de miles de cette route dans les villes traversées en fonction du nombre d'accidents ayant eu lieu sur cette route dans ces villes :

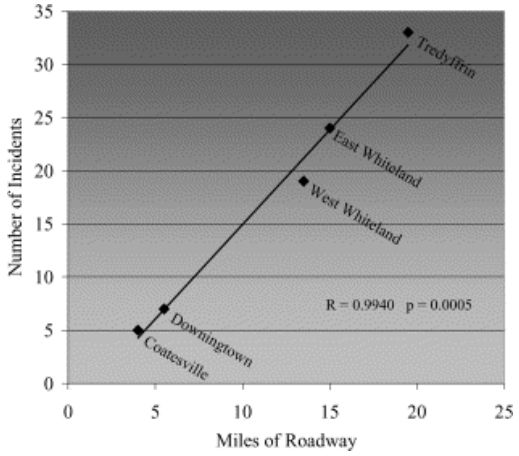


Figure 2.12 : Nombre d'accidents en fonction de la longueur du réseau dans chaque ville (Shorten, et al., 2002)

Ils obtiennent une droite de pente 1 ce qui leur permet de conclure qu'il n'y pas de « miles plus dangereuse qu'un autre » sur cette route, le nombre d'accidents dans chaque ville étant proportionnel au nombre de miles de cette route dans la ville.

- Origine des accidents

L'erreur humaine semble être la principale origine des accidents. Aux États-Unis selon Erkut et al. 40% des accidents routier sont imputables à des erreurs humaines. En France cette proportion est en 2010 légèrement supérieure avec 48% des accidents dus à une erreur humaine; 28% des accidents provenant de causes externes et 15% de défaillances mécaniques selon les chiffres de « l'Inventaire 2010 des accidents technologiques » (Bureau d'Analyse des Risques et Pollutions Industriels (BARPI), 2010). Au Québec la proportion d'accident dus à une erreur humaine grimpe à 70% des accidents TMD (de Marcellis-Warin, Leroux, et al., 2006). Cela rejoint les propos de Dionne et al., qui pensent qu'un facteur significatif expliquant les taux d'accidents routier au Québec impliquant des chauffeurs professionnels est le nombre de points d'inaptitude accumulé par ces conducteurs (Dionne, Laberge-Nadeau, Desjardins, Messier, & Vanasse, 1995).

Voici selon Erkut et al. les répartitions des principales causes d'accidents aux USA en fonction du type de transport :

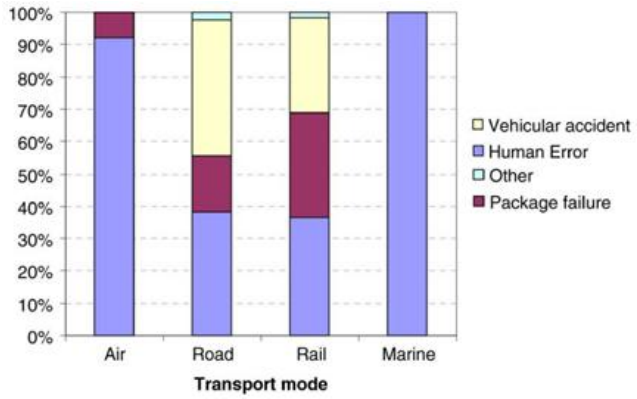


Figure 2.13 : Causes des accidents TMD selon le mode de transport (USA-2003) (Erhan. Erkut, et al., 2007)

Ainsi selon ces chiffres l'erreur humaine est la cause la plus fréquente des accidents, notamment en ce qui concerne le transport aérien et par bateau. Notons par ailleurs que dans le cas du transport par voie maritime, ce chiffre peut s'expliquer par la difficulté de mouvement des bateaux dans les ports et par l'influence importante du facteur humain dans leurs manœuvres ainsi que dans les opérations de chargement et de déchargement (Ronza, et al., 2003). D'ailleurs selon cette étude 40% des accidents se produisent en mer lors de l'approche et des manœuvres des navires, 34% lors du chargement ou du déchargement et 12% au cours du stockage. Ces chiffres sont similaires à ceux obtenus par Dabra et Casal (Darbra & Casal, 2004).

Ohtani et Kobayashi ont trouvé une forte corrélation entre le nombre d'accidents et le nombre de travailleurs. Ainsi selon ces travaux, plus le nombre de salariés est faible, plus les risques d'accidents sont importants. Ils remarquent tout de même qu'en pratique, le nombre d'employés ne peut être modifié uniquement à des fins de sécurité, et ils pensent que des progrès importants peuvent être fait en matière de sécurité en mettant l'accent sur la formation des salariés particulièrement lorsqu'ils sont peu nombreux (Ohtani & Kobayashi, 2004).

2.2 Études spatiales du TMD

Après avoir passé en revue un certain nombre de résultats d'études descriptives du TMD, nous allons nous intéresser aux études spatiales du TMD. Les études spatiales constituent un aspect de plus en plus important de l'étude du TMD depuis l'apparition de logiciels d'analyse spatiale toujours plus performants. Ces outils permettent aux chercheurs de localiser les accidents sur une carte, et ouvrent le champ à de multiples analyses utilisant des données géographiques (répartition de population, réseau routier,…).

La première section aborde les analyses centrographiques, une méthode d'analyse spatiale ancienne rarement utilisée dans l'étude du TMD, que nous utiliserons dans notre étude. La seconde section abordera les analyses par la méthode des points chauds (« Hot Spots »). Enfin le dernier point traitera de l'utilisation des deux types de données spatiales les plus couramment utilisées pour étudier le TMD : les données concernant la population et le réseau routier.

2.2.1 Analyse centrographique

L'analyse centrographique constitue un ensemble d'indices qui permettent de décrire et de mesurer les caractéristiques globales de la distribution spatiale d'un phénomène. Cela en fournissant un équivalent des mesures statistiques des tendances de centralisation et de dispersion, adaptées à un espace géographique à deux dimensions. Ce type d'analyse est né à la fin du XIXème siècle lorsque J.E. Hilgard essayait de tracer la marche vers l'Ouest du peuplement américain. Il a alors eu l'idée de mesurer la position centrale d'une distribution de population, pour suivre « le centre de population de la nation américaine » (Hilgard, 1872), ce qui

a donné naissance aux études centrographiques. Ce type d'étude est principalement utilisé pour étudier des répartitions de population, comme les deux exemples suivant le montrent.

En 1974, Dean et al. ont étudié la population du Québec selon l'idée de Hilgard. Ils ont calculé pour les 1596 municipalités du Québec trois mesures centrographiques, le centre de gravité (ou point moyen), la distance-type et un indice de dispersion relative pour trois dates (1951, 1961, 1971) afin de faire ressortir les tendances évolutives de la distribution spatiale de la population (Dean, Bisson, & La Rochelle, 1974). D'après leurs résultats, la tendance est à la concentration de la population et on peut observer un déplacement du centre de gravité le long d'un axe parallèle au Saint-Laurent vers Montréal qui se trouve en 1971 au milieu du lac Saint-Pierre. Voici les positions qu'ils ont obtenues :

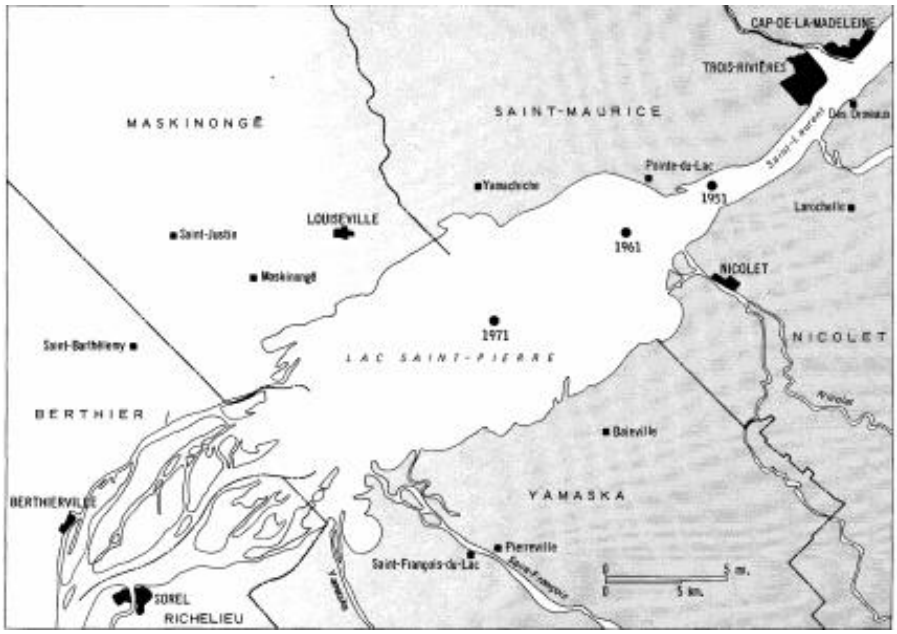


Figure 2.14 : Centre de gravité de la population du Québec en 1951, 1961 et 1971 (Dean, et al., 1974)

Une analyse centrographique a également été utilisée par Cao et al. pour comparer les phénomènes migratoires de francophones urbains du Nouveau-Brunswick en 1961, 1981 et 2001 (Cao, Chouinard, & Dehoorne, 2005). A partir des données de recensement de ces trois années, les auteurs ont tracé les centres de gravité ainsi que les ellipses centrographiques de la répartition de la population francophone de cette province. Ils constatent que, par rapport à sa position de 1961, le centre de gravité pondéré de la répartition spatiale des francophones s'est déplacé en 1981 de 9 kilomètres vers le Nord-Est où se trouve la région qui compte la plus grande concentration de francophones. À l'époque, cet attrait s'expliquait surtout par la prospérité de l'industrie de la pêche. En 2001, ce centre de gravité s'est dirigé de manière significative vers le Sud-Est. Ce déplacement s'explique selon les auteurs par l'expansion considérable de la région urbaine de Moncton. D'autre part, la superficie de l'ellipse qui représente la répartition de la population francophone urbaine s'est considérablement réduite depuis 1961. Ce phénomène indique la tendance à la concentration de plus en plus marquée de cette population.

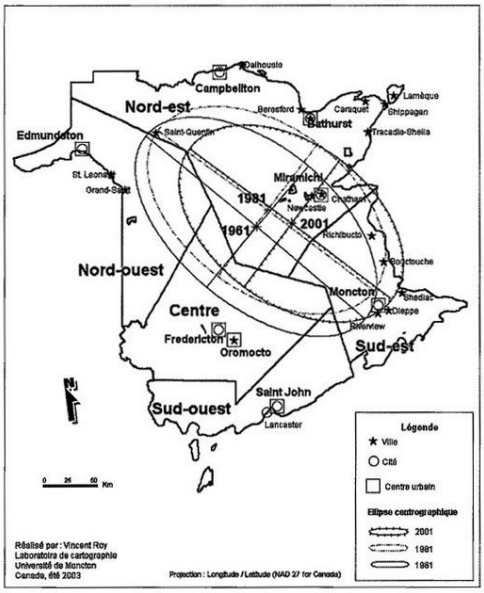


Figure 2.15 : Déplacement de la population francophone au Nouveau-Brunswick de 1961 à 2001 (Cao, et al, 2005)

2.2.2 Analyse des points chauds

Les analyses centrographiques permettent d'étudier la centralité et la dispersion d'une répartition. Cependant depuis les années 1990, les technologies SIG sont devenues des outils de plus en plus populaires et permettent grâce à leur puissance de calcul de mener des études plus poussées. Un de ces outils est la détermination et l'affichage des « points chauds », ainsi que des densités d'un phénomène grâce aux fonctions « kernel ». Une description de ces méthodes est faite dans la section méthodologie (§ 3.4.3).

Ce type d'étude peut permettre de mettre en évidence les zones les plus dangereuses et ainsi aider les autorités à améliorer la sécurité en agissant sur ces zones. Par exemple Erdogen et al. ont géo-référencé les accidents survenus sur les autoroutes de la région d'Afyonkarahisar en Turquie, et

ont ensuite estimé la densité d'accidents par la méthode du noyau (kernel) (Erdogan, Yilmaz, Baybura, & Gullu, 2008). Ils se sont aperçus que les zones à forte densité d'accidents correspondaient à des intersections menant à des villages, des virages serrés ou encore des aires glissantes durant les saisons humides. Les auteurs ont ainsi pu suggérer aux autorités d'augmenter la signalisation dans les zones sensibles relevées.

Selon le même principe, Pulugurtha et al. ont repéré et classé les zones où se sont produits le plus d'accidents piétonniers dans la région métropolitaine de Las Vegas sur la période 1998-2002, voici la carte qu'ils ont produit :

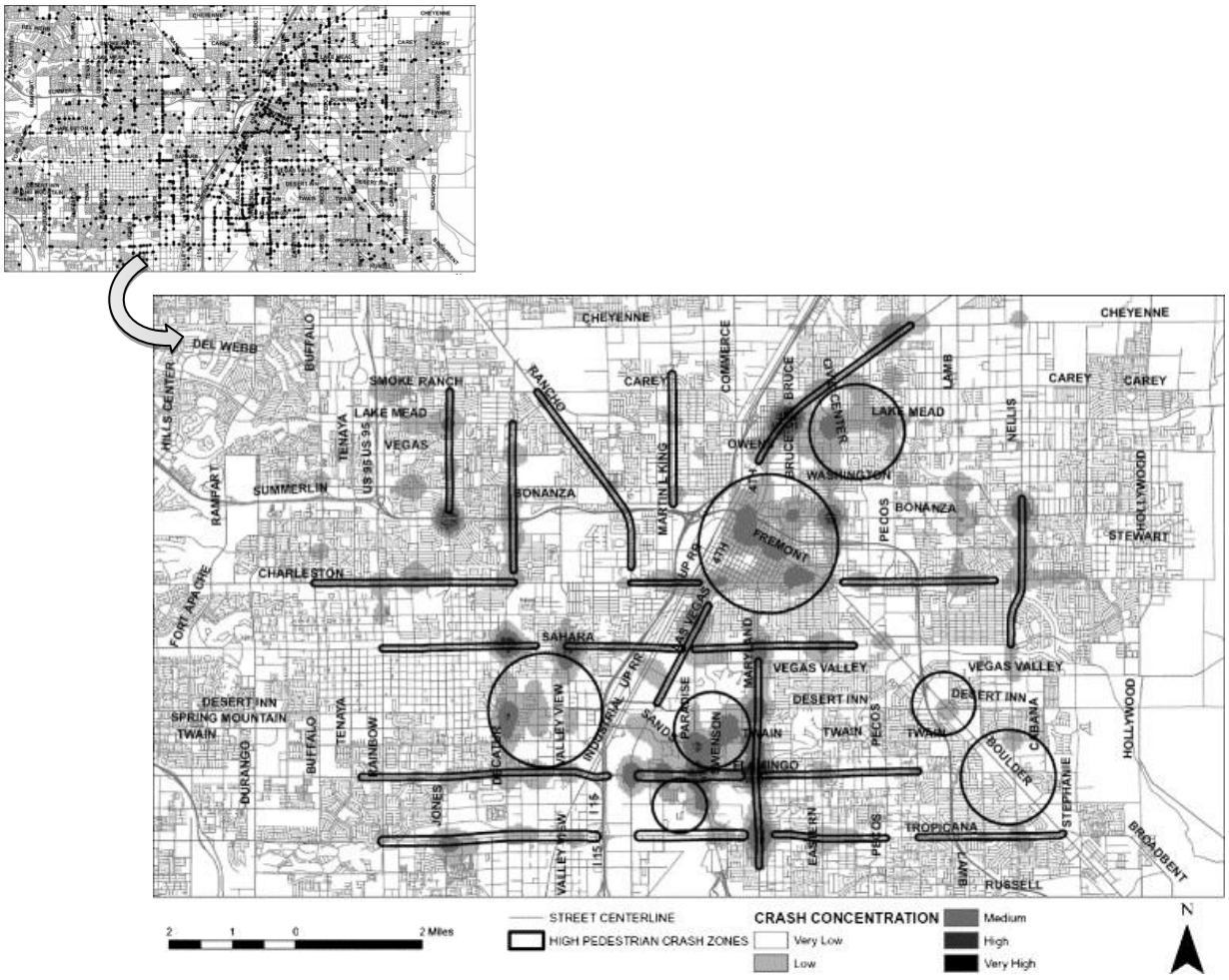


Figure 2.16 : Distribution spatiale des accidents piétonniers de Las Vegas (en haut) et leurs zones de forte concentration (Pulugurtha, Krishnakumar, & Nambisan, 2007)

A partir de cette carte, 29 zones dangereuses ont été identifiées et classées selon leur dangerosité, dans le but de permettre aux autorités de prioriser leurs actions (Pulugurtha, et al., 2007).

Enfin une analyse des densités peut aussi permettre de comparer deux répartitions. Schweitzer utilise cette méthode pour prouver qu'il y a un lien entre les lieux où des accidents MD de type rejet ont été signalés, et les lieux où les MD sont utilisés :

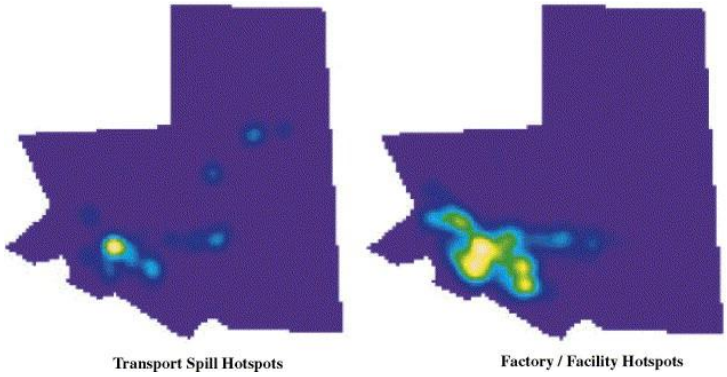


Figure 2.17 : « Hot spots » des rejets de MD et des usines et installations utilisant des MD au sud de la Californie (Schweitzer, 2006)

Schweitzer remarque par ailleurs que les zones à forte concentration d'accidents correspondent à des zones où habitent majoritairement des populations latinos. C'est un exemple des informations que l'on peut obtenir en ajoutant les données liées à la population à l'étude des points chauds.

2.2.3 Ajout de couche d'informations spatiales

Les SIG permettent, de par leur fonctionnement, de superposer des « couches » d'informations spatiales d'une grande variété. Dans l'étude du TMD les chercheurs utilisent principalement des informations sur les répartitions de population, ainsi que sur le réseau routier. Dans cette section nous allons voir de quelle façon ils utilisent ces données.

2.2.3.1 Utilisation de données sur la population

Une préoccupation majeure des chercheurs dans le domaine du TMD, est d'évaluer l'impact d'un accident MD sur les populations. Il faut pour cela utiliser des méthodes d'estimation de la zone d'impact des accidents impliquant des matières dangereuses. Dans le cas d'un nuage de dispersion, Luketa-Hanlin propose de modéliser son comportement en cas de déversement en fonction de la quantité de produit, afin en de prévoir les impacts potentiels d'un accident (Luketa-Hanlin, 2006).

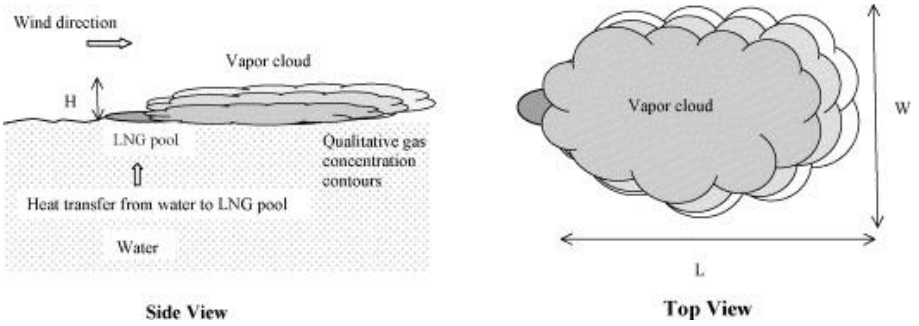


Figure 2.18 : Caractérisation d'un nuage de dispersion (Luketa-Hanlin, 2006)

La méthode généralement utilisée est un modèle dit « Gaussien Plume » qui permet de déterminer les concentrations dans l'atmosphère lors de la dispersion de contaminant comme l'ammoniac et le chlore. Dès lors en estimant les conséquences sur la santé humaine comme une fonction de cette concentration et en utilisant les fonctions de recouvrement des GIS, on est capable d'estimer l'impact d'un nuage de dispersion sur une population. C'est exactement ce qu'on fait Zhang et al. avec une précision remarquable (Zhang, Hodgson, & Erkut, 2000) :

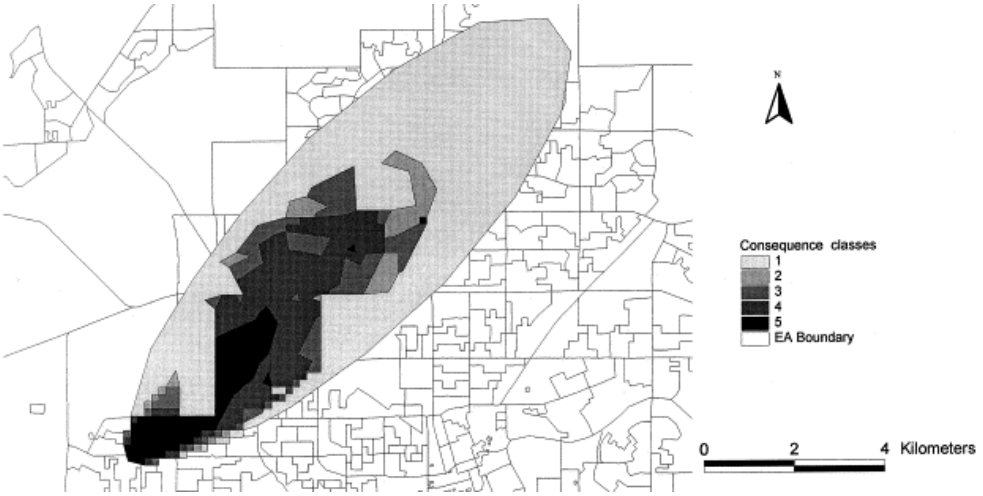


Figure 2.19 : Estimation des conséquences sur la santé humaine d'un déversement à Edmonton, Canada (Zhang, et al., 2000)

Au Québec, Drouin et Leroux ont mené une étude similaire. Ils ont simulé des déversements d'acide chlorhydrique, d'ammoniac et de formaldéhyde sur une route empruntée par les camions qui en transportent régulierement à St-Hyacinthe et à Châteauguay. A l'aide d'un logiciel développé par l'US Environmental Protection Agency (EPA), ils ont estimé les panaches de dispersion d'un déversement. En superposant ces panaches aux données de population (Aires de diffusion, recensement de 2001) il ont alors été capable d'estimer la population impactée. Pour

résoudre le problème des aires de diffusion partiellement impactées par un panache, ils ont astucieusement utilisé une formule liant le nombre de personnes impactées de l'AD au nombre de mètres de rue inclus dans l'aire de diffusion, en remarquant qu'au Québec la population vie le plus souvent au bord des routes. Voici par exemple la carte qu'ils ont obtenue pour un déversement d'ammoniac à Châteauguay :

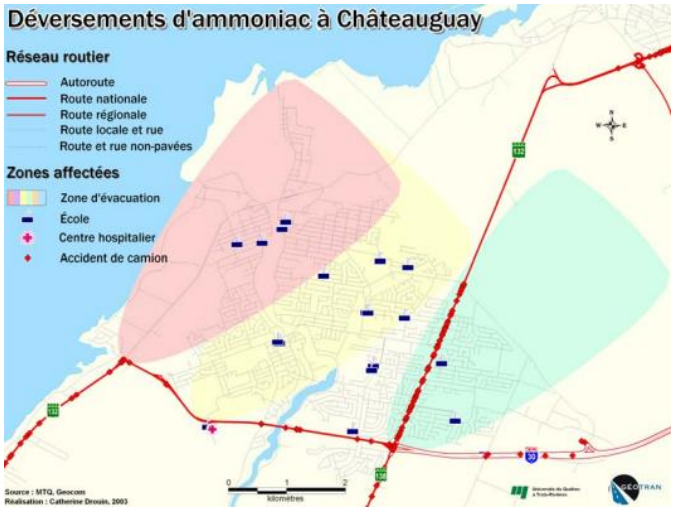


Figure 2.20 : Déversement d'ammoniac à Châteauguay, Canada (Drouin & Leroux, 2004)

Sur les trois panaches visibles sur cette carte, c'est le jaune qui concerne le plus de population avec plus 19 500 personnes touchées (Drouin & Leroux, 2004).

2.2.3.2 Utilisation de données sur le réseau routier

Le second type de couche très souvent utilisé dans les analyses spatiales du TMD concerne le réseau routier.

Un nombre très important de travaux concerne la conception et l'optimisation du trajet des véhicules transportant des matières dangereuses. Beaucoup d'auteurs y voient une façon efficace de diminuer les risques d'impact sur la population. Verter et al. soutiennent que le réseau routier accessible aux transporteurs de matières dangereuses se doit d'être limité par le législateur afin de limiter les risques encourus par la population (Vedat Verter & Kara, 2008). Les auteurs sont cependant bien conscients que chaque solution de routage doit être un compromis entre risque minimum et viabilité économique. Ils présentent d'ailleurs une méthodologie permettant de générer un routage satisfaisant selon plusieurs critères, qu'ils appliquent pour un itinéraire local (Ouest de l'Ontario) et plus vaste (Toronto – Québec).

Il y a cependant des divergences entre les auteurs concernant la définition du risque sur un trajet et donc sur les itinéraires minimisant ce risque. Ainsi, Erkut et Verter soulignent qu'un chemin optimal selon un modèle peut être très mauvais selon un autre modèle. Voici par exemple l'optimisation selon trois critères du trajet entre Chicago vers Atlanta qu'ils présentent (E Erkut & Verter, 1998) :

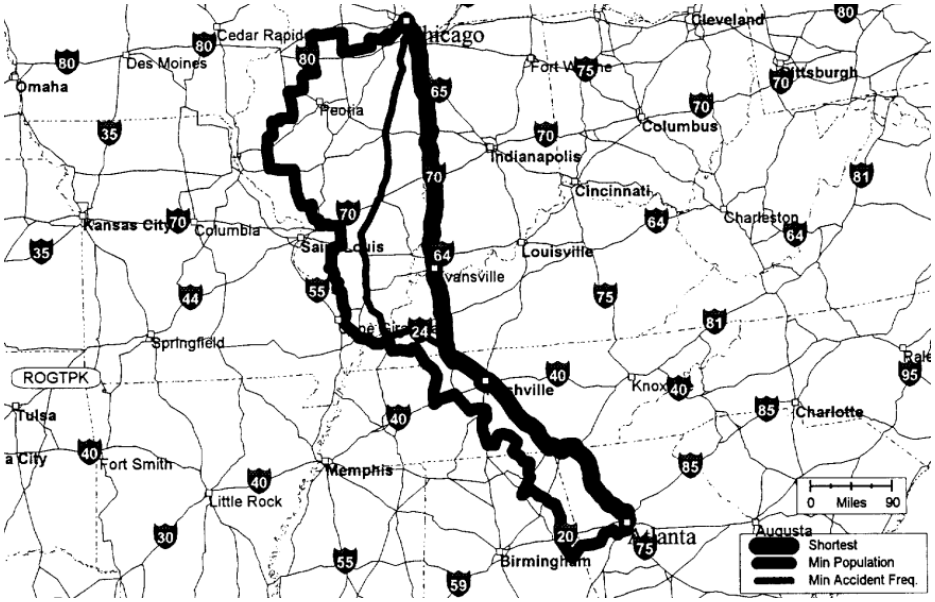


Figure 2.21 : Optimisation selon 3 critères différents du routage entre Chicago et Atlanta (E Erkut & Verter, 1998)

Une mesure souvent retenue, que Batta et Chiu ont été parmi les premiers à proposer, est de prendre en compte le nombre de personnes vivant à proximité des routes (Batta & Chiu, 1988). Une autre étude, ayant pour objectif de déterminer l'itinéraire de chargement de matières radioactives, utilise une fonction pondérée prenant en compte l'exposition de la population et les coûts de transport (ReVelle, Cohon, & Shobrys, 1991). Ak et Bozkaya proposent quant à eux une méthode prenant en compte la minimisation du temps de parcours plutôt que de la distance. Ils jugent cette approche plus efficace en milieu urbain où des problèmes de congestion peuvent allonger considérablement les trajets et faire ainsi augmenter les risques (Ak & Bozkaya, 2008). Verter et Kara proposent pour leur part trois optimisations des chemins minimisant l'exposition de la population, le nombre estimé de personnes à évacuer en cas d'accident et la probabilité qu'un accident ait lieu, afin de permettre aux organismes gouvernementaux d'estimer l'impact des modifications des routes (V. Verter & Kara, 2001).

La tendance actuelle est à la prise en compte d'un grand nombre de facteurs, ce qui est possible grâce à l'évolution de l'outil informatique et à l'utilisation d'heuristiques. Zographos et Androutsopoulos ont ainsi mis point une méthode de calcul d'un itinéraire dit « de moindre risque » à partir d'un algorithme heuristique, prenant en compte la vitesse du vent, les caractéristiques de la matière dangereuse et la quantité transportée (Zografos & Androutsopoulos, 2004).

D'autres études utilisent le réseau routier dans d'autre optique que l'optimisation du parcours. Ainsi, pour Liu et al. une autre façon d'améliorer la sécurité du TMD est d'optimiser l'emplacement des postes de secours en fonction du risque, ce qui est particulièrement pertinent dans des régions fortement urbanisées, comme à Singapour, où une étude a été menée. L'objectif est pour eux de réduire à 5 minutes (contre 8 initialement) le temps d'intervention sur les routes empruntées par des chargements de matières dangereuses. Pour cela un algorithme ANT couplé à un GIS a été utilisé afin de déterminer l'emplacement de nouvelles stations de lutte anti-incendie dans la ville de Singapour (Liu, Huang, & Chandramouli, 2006).

Enfin on peut citer une étude originale, qui a utilisé des données sur le réseau routier et sur la population afin de déterminer s'il était possible d'évacuer de façon efficace les personnes potentiellement touchées par un accident impliquant des matières dangereuses (Lozano, Muñoz, Antún, Granados, & Guarneros, 2010). Pour cela les auteurs ont déterminé les endroits et les types d'accidents qui, selon les dernières données sur le TMD de Mexico, sont le plus susceptibles de se produire dans la ville. Ils ont ensuite estimé le nombre de personnes à évacuer et, en comparant avec la capacité des rues, discuté de la réalisation de l'évacuation dans un délais acceptable suivant le type de matière dangereuse impliquée. La situation est selon eux problèmatique car des congestions risquent d'apparaitre, empéchant les personnes de partir et ralentissant l'intervention des secours.

Après ce tour d'horizon des principales études spatiales que l'on trouve dans la littérature, nous allons pour finir évoquer les méthodes misent en oeuvre par les chercheurs pour travailler avec des bases de données d'accidents.

2.3 Utilisation d'une base de données d'accidents

L'étude du TMD passe très souvent par l'exploitation de bases de données, et pour cela il est intéressant de voir comment les auteurs les ont utilisées et les problèmes qu'ils ont rencontrés. On s'intéresse ici plus à la méthodologie qu'aux résultats obtenus, les travaux présentés dépassent donc parfois le cadre du transport des matières dangereuses, même si les exemples provenant d'études du TMD ont été préférés.

2.3.1 Sélection des données

La première étape d'une analyse de base de données est de choisir les données à analyser. Le plus souvent les chercheurs ont accès à une base de données existante. Mais parfois il n'y a pas de base de données publique et les auteurs doivent se constituer leur propre base de données à partir de sources d'informations diverses comme Yang et al. (2010) ont dû le faire pour faire face à l'absence de BDD chinoise. Cette situation est cependant loin d'être idéale car il est difficile d'être exhaustif dans ce cas, et les accidents avec peu d'impact sont difficiles à recenser.

Dans le cas où les chercheurs ont accès à une base de données, ils doivent le plus souvent en extraire les informations qui les intéressent. Ils doivent pour cela mettre en place une méthode de sélection d'un type d'événement particulier dans la base globale.

Cette sélection peut être manuelle, c'est-à-dire que le chercheur choisit lui-même si un accident doit ou ne doit pas être intégré dans son étude. Cela peut être toutefois un travail fastidieux, puisque tous les accidents de la base globale doivent être passés en revue, et un biais lié à l'utilisateur est difficile à éviter.

Les auteurs choisissent généralement d'automatiser cette tâche en codant des critères de sélection basés sur les champs de la base de données. De cette façon plusieurs groupes de chercheurs ont extrait d'une même base de données (MHIDAS) des informations différentes. Voici par exemple ci-contre les critères

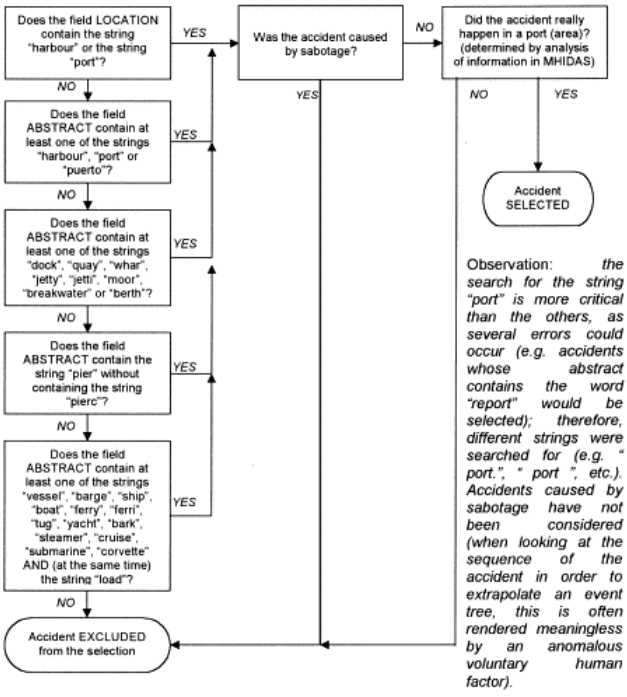


Figure 2.22 : Critères de sélection d'un accident de MHIDAS (Ronza et al, 2003)

utilisés par Ronza et al. pour sélectionner les accidents MD ayant eu lieu en zone portuaire (Ronza, et al., 2003). Ces critères ont permis de sélectionner environ 830 accidents de la base MHIDAS. Les auteurs de l'étude ont par contre dû vérifier « manuellement » cette sélection car ils se sont aperçus que des accidents pertinents avaient été omis et que certains avaient été sélectionnés à tort. Avec une méthode similaire 1 900 accidents MD routiers et ferroviaires ont été identifiés (Oggero, et al., 2006). Une autre étude a extrait 131 accidents de cette même base MHIDAS afin d'étudier les accidents de transport concernant le gaz naturel (Montiel, Vílchez, Arnaldos, & Casal, 1996).

2.3.2 Le biais dans les bases de données

Il faut cependant se méfier des bases de données utilisées et toujours se renseigner sur leurs origines. Pour en revenir à notre exemple précédent, la base de données anglaise MHIDAS (maintenue par la HSE, pour Health and Safety Executive), recense depuis 1980 des accidents du monde entier qui ont eu lieu lors du transport, du stockage ou de la fabrication de matières dangereuses excluant les accidents concernant des matières radioactives. L'origine anglaise de la base de données n'est pas sans effet sur la répartition géographique des données. En effet, 47% des accidents qu'ils ont sélectionnés sont survenus aux États-Unis et 30% en Angleterre (Oggero, et al., 2006). Ces résultats sont assez étonnants car pour des pays comme la France et l'Allemagne ayant pourtant une économie de taille similaire à celle de l'Angleterre la proportion est faible (de l'ordre de 4% pour les deux pays seulement!). Toute analyse sur la répartition spatiale de ces accidents semble donc biaisée.

On peut considérer qu'il y a deux principales sources de biais. La première est celle venant des personnes qui réalisent les bases de données, qui ont un accès « sélectif » à l'information. Ainsi, par exemple, les événements causant peu de dommages sont sous-représentés (comme un « petit » déversement) alors que les événements causant des conséquences plus importantes (explosion, décès, blessés,…) sont plus susceptibles d'être reportés car plus marquant. La seconde source de biais est d'ordre géographique, les gestionnaires de bases de données ayant tendance à favoriser des événements se produisant dans leur propre pays ou à proximité (Ronza, Vílchez, & Casal, 2007). L'exemple précédent de la base MHIDAS en est la parfaite illustration.

Pour éviter ces deux principaux biais il est préférable de travailler avec des bases de données concernant une zone géographique clairement définie (limite le biais géographique) et tirant leur

origine d'une obligation légale de déclaration (limite la sous-représentation de certains événements) (Ronza, et al., 2007).

2.3.3 Cas des bases de données incomplètes

D'autre part de nombreuses études ont souffert du manque de précision de certaines bases de données. Par exemple, les chercheurs qui ont utilisé la base MHIDAS ont estimé que le manque d'informations précises dans la base de données était un obstacle important à l'utilisation de celle-ci. Au Japon, les rapports d'accidents réalisés sont, malgré un format standard bien défini, imprécis et incomplets, ce qui a causé l'échec de nombreuses analyses, qui visaient notamment à expliquer l'augmentation du nombre d'accidents (Ohtani & Kobayashi, 2004).

Pour pallier au manque de données d'une base on peut, si cela est possible, utiliser les informations d'autres bases. Ainsi, trois bases de données concernant la même zone géographique ont été utilisées dans une étude (Trépanier, Leroux, & de Marcellis-Warin, 2009). Pour parvenir à croiser ces trois différentes sources d'informations les auteurs ont développé un outil Access, appelé HECOT pour Hazmat Event Cross-Observer Tool, qui, après qu'ils aient formaté les données, permet de chercher des données liées à un événement préalablement sélectionné selon un ou plusieurs critères comme la date et le lieu. L'utilisateur peut ainsi identifier les enregistrements d'un même accident dans les trois bases de données.

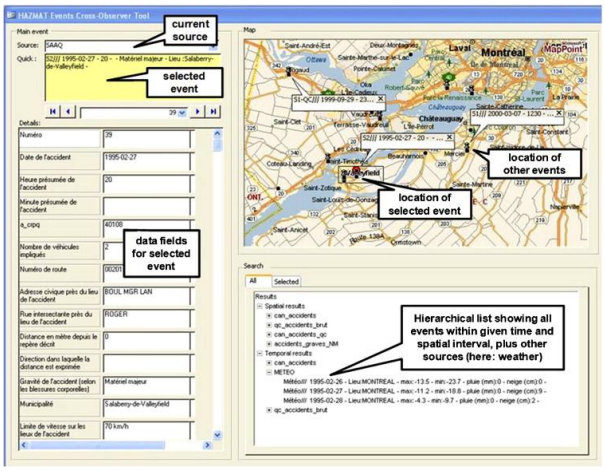


Figure 2.23 : Interface de l'outil HECOT (Trépanier, et al., 2009) permettant de matcher des accidents de base de données différentes

Dans une autre étude, les auteurs sont parvenus à utiliser six bases de données d'accidents différentes afin de développer une base de données centralisée d'accidents MD maritimes. Cela dans le but d'aider les secours à prendre les bonnes décisions en cas d'urgence (Dobbins & Abkowitz, 2003).

Cependant, travailler avec plusieurs sources de données n'est pas toujours sans poser de problèmes. Certaines études sont en effet confrontées à un faible niveau de correspondance dans des bases de données censées se superposer. Cela a été le cas dans celle d'Hobeika et Kim, qui ont constaté le faible niveau de correspondance de deux bases de données d'accidents américaines concernant la Pennsylvanie (Hobeika & Kim, 1993).

Enfin, un même accident peut être décrit plusieurs fois dans la littérature avec des divergences notables dans les informations disponibles. Une analyse des différences a montré que les informations concernant la date et le lieu de l'accident sont de bonne qualité, tandis que la description de l'événement et les données sur le nombre de victimes sont de qualité qualifiée de « moyenne ». Les informations les moins concordantes sont les données sur la matière dangereuse (nom et quantité) (Haastrup & Brockhoff, 1991).

Dans l'étude que nous avons menée, les données de deux bases ont été utilisées et cela a permis de mettre en évidence que certains champs ont été mal renseignés dans l'une d'entre elles, ce qui a incité à apporter des modifications et donc engendré des résultats plus proches de la réalité.

Ainsi les observations relevées dans cette revue de littérature ont permis d'éviter certains écueils et ont donc été largement pris en compte dans la méthodologie.

CHAPITRE 3 MÉTHODOLOGIE

Ce chapitre présente la démarche puis les méthodes utilisées pour produire les résultats présentés dans les deux chapitres suivants. Dans une première partie la démarche générale suivie est présentée, puis nous détaillerons les outils et méthodes utilisés pour l'analyse comparative des bases de données et l'analyse spatiale.

3.1 Démarche suivie

Après avoir parcouru la littérature afin de voir ce qui avait été fait dans le domaine, le travail a été réalisée en trois grandes étapes. Il a tout d'abord fallu faire un travail d'appropriation des données avant de pouvoir commencer l'analyse comparative de ces données pour terminer avec leur analyse spatiale. Le diagramme de Gantt disponible en annexe indique son déroulement temporel.

Le travail d'appropriation des bases de données a consisté, une fois que les données ont été collectées selon les modalités définies avec le MTQ, à comprendre leur fonctionnement et à les préparer pour l'analyse. Pour comprendre le fonctionnement des deux bases de données, leur origine respective a été étudiée ainsi que la composition des différents champs présents. Cette étude a mis en évidence la nécessité de réaliser certaines tâches afin de travailler efficacement avec les données, comme la réalisation d'un outil de visualisation des données, la normalisation et la correction de certaines données et l'identification des accidents TMD dans la base globale. Le paragraphe 3.2 détaille ces travaux.

Une fois les bases de données assimilées et préparées, l'analyse comparative a pu commencer. L'ensemble des champs comparables dans les deux bases de données a été parcouru afin de mettre en évidence d'éventuelles différences ou similitudes. Les outils de la statistique, comme des graphiques, la statistique descriptive ou des tests statistiques, ont été principalement utilisés comme détaillé au paragraphe 3.3.

Enfin, les deux bases de données comportant des informations sur la localisation des accidents, il a été possible de les géo-localiser (cf. §3.4.1). Pour exploiter ces données géographiques, un système d'information géographique a été mis en place, ce qui a permis de générer des cartes d'analyse et de calculer des données géographiques.

3.2 Appropriation des bases de données

Cette section détaille la première partie de la démarche qui a été suivie. Nous présentons tout d'abord les deux bases de données utilisées, puis l'outil de visualisation mis en place avant d'expliquer comment les données ont été liées entre elles.

3.2.1 Présentation des bases de données

Notre étude utilise deux bases de données d'accidents distinctes concernant le Québec. La première base concerne l'ensemble des accidents impliquant des véhicules lourds et la seconde concerne uniquement les accidents impliquant des matières dangereuses.

3.2.1.1 Description de la BDD des accidents impliquant des véhicules lourds

La base de données concernant les accidents impliquant des véhicules lourds est basée sur les registres des rapports d'accident de la SAAQ (Société de l'Assurance-Automobile du Québec). Ce registre est tenu depuis 1978, dans le cadre du mandat de la SAAQ qui est «de remplir, de conserver et de tenir à jour tous les registres officiels en sécurité routière» (de Marcellis-Warin, Leroux, et al., 2006). Il est constitué à partir des formulaires de rapports d'accidents remplis par les policiers du Québec lorsqu'ils sont appelés sur les lieux. Cela n'est pas toujours le cas, ce qui limite l'exhaustivité du registre, notamment pour les accidents de faibles conséquences. Malgré tout, selon le Code de la Sécurité routière du Québec, les rapports sont obligatoires dans plusieurs situations. Notamment s'il y a des dommages corporels ou un délit de fuite, ou encore si l'accident a eu lieu sur le réseau du Ministère des Transports du Québec (MTQ) et que les dégâts causés excèdent 1000$. En ce qui concerne les accidents de poids lourds, ils doivent être répertoriés dès que les dommages matériels sont supérieurs à 1000$, quelle que soit la route.

Ces données ont été dénominalisées et transmises par la SAAQ au MTQ, avec qui nous avons une entente pour l'accès à ces données. Par la suite cette BDD sera souvent désignée par « base globale » car elle concerne tous les accidents impliquant des véhicules lourds, qu'ils transportent ou non des matières dangereuses.

La base de données couvre la période entre 1995 et 2009, elle contient 33 champs, et 262 287 enregistrements. La structure détaillée du fichier, indiquant pour tous les champs le type de

donnée, le nombre de valeurs disponibles, le pourcentage de valeurs non vide ainsi qu'une description est disponible en annexe 2. Les champs permettent de connaitre :

- La date et l'heure de l'accident
- La localisation de l'accident par une description de l'intersection
- Le nombre de véhicules impliqués
- Le nombre de décès et de blessés
- La gravité et le type des accidents
- Les circonstances de l'accident (météo, état de la chaussée, de la surface, limitation de vitesse…)

ID	Jour de la semaine	Date complète	Heure	Minute	Numéro CRPQ	Nb de véhicules impliqués selon le rapport SUM	Route	Adresse près de l'accident	Intersection près de l'accident	Distance du repère en mètre	Direction	Gravité	Municipalité RA act.	Limite de vitesse GLV
1	lundi	02/01/95	18	0	31405	2	-	3320 BL ST F	-	0	-	Matériel majeur	Saguenay	-
2	mardi	03/01/95	17	0	40170	2	40	-	CHAMP D EAU	0	-	Matériel majeur	Montréal	70 km/h
3	mercredi	04/01/95	14	0	40170	2	40	EST	QUERBES	0	-	Matériel majeur	Montréal	100 km/h
4	jeudi	05/01/95	16	0	30234	2	-	GILFORD	DE LORIMIER	0	-	Matériel majeur	Montréal	-
5	jeudi	05/01/95	0	0	30712	1	-	ST JOSEPH	PARIS	0	-	Matériel majeur	Drummondville	-
6	jeudi	05/01/95	19	0	40104	3	20	KM 118	SORTIE 120	0	-	Mortel	La Présentation	100 km/h
7	vendredi	06/01/95	2	0	31417	1	167	966 3E RUE	-	0	-	Matériel majeur	Chibougamau	50 km/h
8	samedi	07/01/95	11	0	40170	1	25	TUNNEL L H	-	0	-	Matériel majeur	Montréal	70 km/h

Figure 3.1 : Aperçu des premiers champs de la BDD globale

Les informations des champs, malgré le grand nombre d'enregistrements, sont étonnamment « stables » (peu de fautes d'orthographe, de valeurs aberrantes,…) ce qui prouve que les données ont été consolidées et peuvent être utilisées « tel quel ». Cependant il n'y a pas d'informations particulières relatives aux transports des matières dangereuses. Le champ « camion lourd » indique toutefois le type de camion lourd impliqué, ce qui signifie que si c'est un véhicule de transport de matières dangereuses, la mention « matières dangereuses » apparaitra.

Le champ « Code d'impact », qui renseigne sur le mouvement des véhicules lors de l'impact, nécessite l'utilisation de la table d'interprétation disponible en annexe 4 pour être compréhensible.

3.2.1.2 Description de la base TMD

Cette base de données ne comprend que des accidents qui ont impliqué des matières dangereuses. Elle a été compilée par le service de Mr Raynald Boies du MTQ à partir des données de la SAAQ, et complétée en accédant au microfilm des rapports des policiers et à diverse sources d'information comme des coupures de journaux. Pour qu'un accident soit ajouté dans cette base, il faut qu'au moins un des véhicules impliqués dans l'accident soit un véhicule de transport de matières dangereuses.

Le fichier original se présente sous la forme d'un fichier Access qui comprend toutes les informations des rapports d'accident, ainsi que des informations sur les véhicules impliqués. Les rapports d'accident ont été exportés dans un fichier Excel dont la structure détaillée est disponible en annexe 3. Les informations décrivant les véhicules impliqués n'ont pas été utilisées dans cette étude. Enfin le fichier Excel a été mis en forme afin de correspondre à celui de la base de données globale et ainsi faciliter la comparaison des deux BDD, voici un aperçu du fichier Excel obtenu :

N° dossier	ID	Date	Heure	Minute	Route	Adresse près de l'accident	Intersection près de l'accident	Municipalité	Info Localisation	Nombre de véhicule	Dommages matériels	Code l'impact	Vitesse autorisée	Morts (code 1)	Blessés graves (code 2)	Blessés mineurs (code 3)	Victime sans blessure (code 4
212	490	15/12/95	19	15	109	KM8 Route 109	Rang 1 et 2 Camotte	La Motte	MapPoint	1	501$ et plus	99	90 km/h	0	0	0	0
10	560	06/03/95	7	58	746	Chemin Aylmer	CASTELBEAU	Aylmer	Google Maps	2	501$ et plus	1	70 km/h	0	0	1	0
137	650	04/07/95	13	55	372	rue Bagot	Boul. Baie Nord	La Baie	Google Maps	2	501$ et plus	1	50 km/h	0	0	0	0
185	786	11/11/95	12	22	811	4 Avenue	entré du stationnement	Saint-Romuald	Google Maps	2	501$ et plus	1		0	0	0	0
68	864	21/01/95	22	15		Alphonse de Roy	Notre-Dame	Montréal	Google Maps	2	501$ et plus	2	70 km/h	0	0	0	0
78	902	11/02/95	22	5		Boul. Saguenay	Lalonde	Jonquière	Google Maps	2	501$ et plus	2	50 km/h	0	0	0	0
108	969	28/03/95	18	20	148	Boul. Sauvé	Rond-Point 640	Saint-Eustache	MapPoint	2	501$ et plus	2	50 km/h	0	0	0	0
157	1127	12/08/95	21	50	73	Entrée du pont Pierre Laporte	Bretelle Accès BL Champlain	Sainte-Foy	Google Maps	2	501$ et plus	2	90 km/h	0	0	0	0

Figure 3.2 : Aperçu du fichier Excel de la base TMD

Le fichier couvre la période 1995-2007 et comprend 43 champs pour 2 379 enregistrements. Globalement les informations répertoriées permettent :

- D'identifier l'accident
- De localiser l'accident
- De décrire le lieu de l'accident
- De décrire les circonstances de l'accident
- De faire état des victimes et des dégâts

Contrairement à la base globale, les champs contiennent des données avec des syntaxes différentes, et des fautes de frappes. Afin de pouvoir travailler avec ces données il a donc fallu consolider les données. Par exemple les municipalités en « Saint » étaient parfois écrites avec l'abréviation « st », parfois sans. La même syntaxe que dans la base globale a été utilisée, c'est-à-dire dans ce cas en remplaçant « St- » par « Saint » pour les municipalités concernées.

Par ailleurs, l'ensemble des fautes d'orthographes, valeurs aberrantes ou manifestement fausses détectées ont été corrigées au fur et à mesure de l'avancement du travail (ex : Montéal pour Montréal). Raynald Boies s'étant montré intéressé par ces corrections, à chaque fois qu'une modification a été faite, un « style » de correction (texte en gras et fond vert) a été appliqué. Il est ainsi très facile en triant par couleur de remplissage de voir les champs qui ont été modifiés :

N° dossier	Date	Heure	Minute	Route	Adresse près de l'accident	Intersection près de l'accident	Municipalité	Nombre de véhicule	Code l'impact	Vitesse autorisée
324	13/12/96	6	30				**Obedjiwan**	1	16	
261	27/03/96	14	37		Face 2189 StMaurice		**Jonquière**	2	1	
424	**19/06/1996**	15	30	**169**		**Rue de l'Hopital**	**Métabetchouan**	2	3	90 km/h
398	01/08/96	14	0	169	Sud	Côte Sables	**Chambord**	3	3	90 km/h
294	09/08/96	10	46		**128 PRINCIPA**		**Saint-Basile-le-Grand**	2	3	
419	27/12/96	3	5	132	est	sortie Rolland-Therrien	**Longueuil**	2	3	100 km/h
322	06/12/96	17	32	397	Œufs Beauvais		**Val-d'Or**	2	4	
244	16/02/96	15	10		Bouchette	Benni	**Huntingdon**	2	6	
385	12/03/96	7	55	20	Km 12	Autoroute 55	**Drummondville**	2	15	100 km/h

Figure 3.3 : Exemple de champ corrigé dans la base TMD

Les champs concernant la localisation des accidents (adresse, intersection, municipalité et route) sont particulièrement concernés par ces corrections car la base a été par la suite localisée en partie « manuellement ».

3.2.1.3 Visualisation des données

Afin de pouvoir visualiser les données contenues dans les nombreux champs des deux BDD en un « coup d'œil » nous avons réalisé sous Excel, pour les deux bases de données, un outil permettant de visualiser les données concernant un accident sur une interface plus pratique que l'imposant tableau. Cet outil comprend une partie affichant les informations des champs et une carte interactive indiquant la localisation de l'accident (les accidents ont été localisés selon la méthode indiquée dans le paragraphe 3.4.1). Son utilisation est très simple, il suffit d'inscrire le

numéro d'identification (clé primaire) d'un accident, et toutes les informations concernant l'accident contenues dans la base de données s'affichent. Les fonctions Excel utilisées sont les fonctions INDEX et RECHERCHEV. Voici comment se présente cet outil dans les deux bases de données :

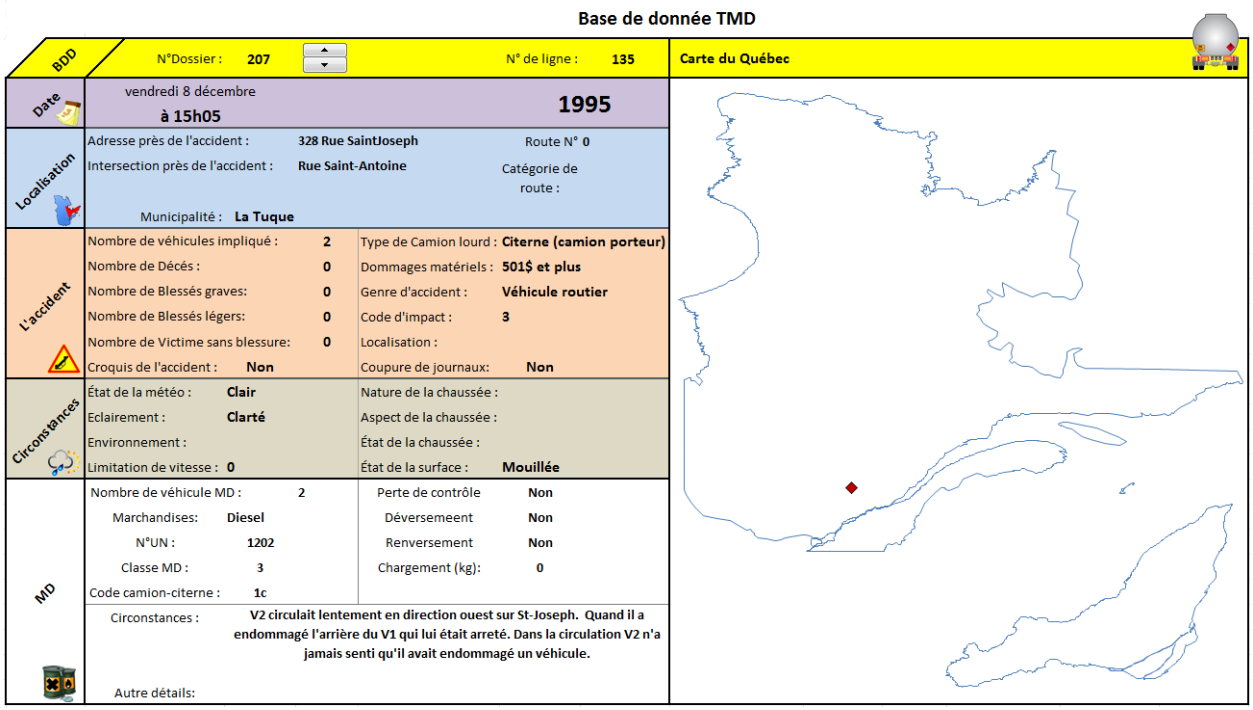


Figure 3.4 : Visualisation des accidents de la base TMD

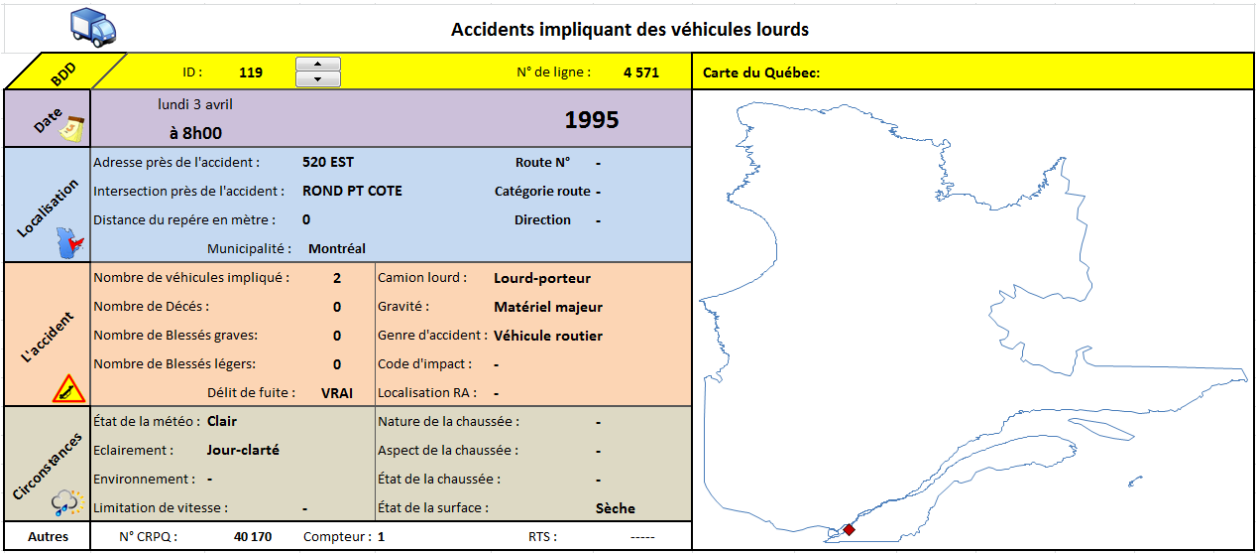
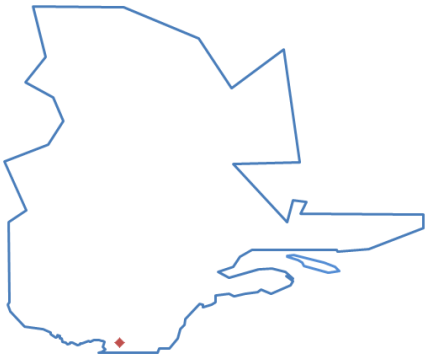
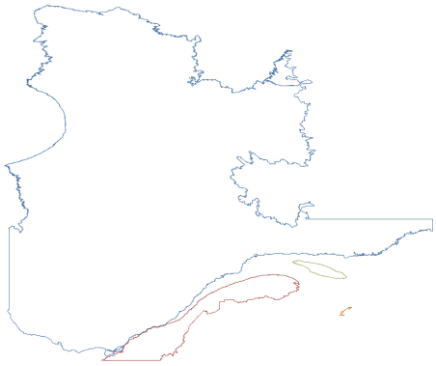
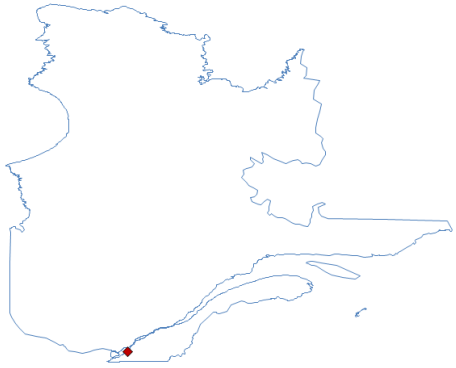


Figure 3.5 : Visualisation des accidents de la base globale

Les coordonnées des accidents ont été obtenues selon la méthode décrite dans le paragraphe 3.4.1 et la carte de visualisation est obtenue à partir d'un nuage de points ayant pour coordonnées la latitude et la longitude des limites du Québec. Dans un premier temps, la forme du Québec a été grossièrement dessinée avec une cinquantaine de points (latitude/longitude) entrées manuellement. Afin d'améliorer le résultat, les données issues du fichier des limites des provinces et territoire du Québec (cf. § 3.4.2) ont ensuite été utilisées. A l'aide du logiciel SIG (MapWindows) les données géographiques (contenues dans le shape file) ont été exportées dans Excel et on a obtenu 255 000 couples de points qui définissent « officiellement » les 1 233 lignes fermées formant les frontières géographiques du Québec.

Le résultat obtenu est d'une précision remarquable, mais pour donner une idée grossière de la situation d'un accident, c'est bien entendu inutile, d'autant que le temps d'affichage est assez long. Pour réduire le nombre de points chaque « paquet » de 50 points a été remplacé par sa moyenne et seules les quatre lignes fermées les plus pertinentes, c'est-à-dire la rive Nord, la rive sud, et les îles Anticosti et celles de la Madeleines ont été conservés. Ainsi les îles de Montréal et de Laval sont absentes de la carte car leurs contours ne se voient quasiment pas. De cette façon une carte suffisamment précise pour l'utilisation qui en est faite a été obtenue avec « seulement » 3 575 points. Voici en image le résultat des différentes étapes :

Tableau 3.1 : Étapes de génération de la carte de visualisation

1er Étape	2nd Étape	3ème Étape
Points entrées manuellement	Points issues du fichier des limites	Réduction du nombre de point
		

3.2.2 Identification des accidents TMD dans la base globale

Afin de pouvoir faire une analyse comparative des accidents en général et des accidents TMD, il est nécessaire d'identifier dans la base globale les accidents de la base TMD, notamment pour des raisons de dépendance de données pour la comparaison statistique, mais cela permettra également de vérifier la compatibilité des données dans les deux bases.

Pour identifier un accident de la base TMD dans la base globale, il a fallu comparer les informations des champs afin de choisir la bonne combinaison permettant de conclure à une correspondance. Si on l'avait fait manuellement, on aurait pu trier par la date puis par la municipalité pour enfin choisir le bon accident, parmi les accidents restants, grâce au code d'impact par exemple. Il n'est cependant pas envisageable de faire ce travail « à la main » pour les quelques 2 400 accidents de la base TMD. En s'inspirant de travaux précédents (Trépanier, et al., 2009) ce travail a été automatisé au maximum.

3.2.2.1 « Simplification » du problème

Pour identifier un accident il faut comparer à un moment ou un autre un accident de la base TMD à tous les accidents de la base globale via la recherche d'une valeur par exemple. Cette opération avec les quelques 250 000 accidents constituant la base globale est particulièrement longue surtout si la recherche porte sur plusieurs champs. Dans un premier temps, nous avons donc cherché à « simplifier » le problème en éliminant les accidents ayant peu de chance de se trouver dans la base TMD. Pour cela les informations des champs date et heure, qui sont toujours remplies, ont été utilisée et que l'on peut supposer être la plupart du temps exactes, selon la logique suivante :

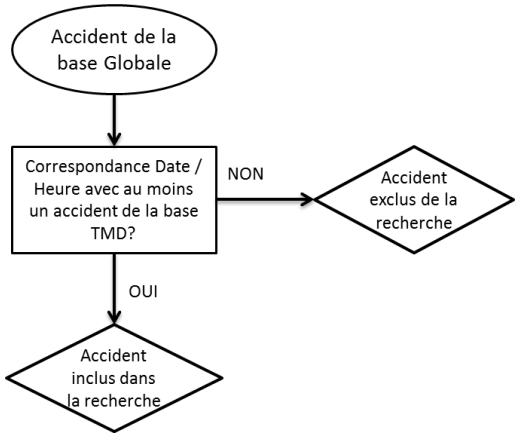


Figure 3.6 : Méthode de simplification des recherches

Le programme VB est disponible dans l'annexe 5. Ainsi tous les accidents de la base globale ayant une date et une heure différentes des accidents TMD ont été marqués de façon à éviter de rechercher dans ces valeurs par la suite.

Ce programme a tout de même été long à exécuter (environ 6 h) mais a permis d'éliminer 236 355 accidents (environ 95% des accidents de la base globale) pour les recherches futures. Le problème simplifié est donc maintenant d'associer les 2 379 accidents de la base TMD parmi les 10 461 accidents restant dans la base globale.

3.2.2.2 Méthode d'identification

Pour associer le plus d'accidents possible, il faut déterminer la combinaison de champ donnant les résultats les plus pertinents et les plus nombreux possible. Pour cela il faut partir d'un accident de la base TMD en essayant de lui trouver une correspondance unique dans la base globale. Mais plus on essaie d'être précis (et donc d'avoir beaucoup de champs communs) moins les résultats sont nombreux et inversement si on n'est pas assez précis les résultats peuvent se révéler non pertinents. Le fait d'avoir simplifié le problème rend les recherches rapides et plusieurs combinaisons de champs ont pu être testées en un temps raisonnable afin de déterminer une méthode de recherche efficace. Voici la méthode d'identification retenue et réalisée sous VB avec les nombres d'accidents TMD correspondant à chaque étape :

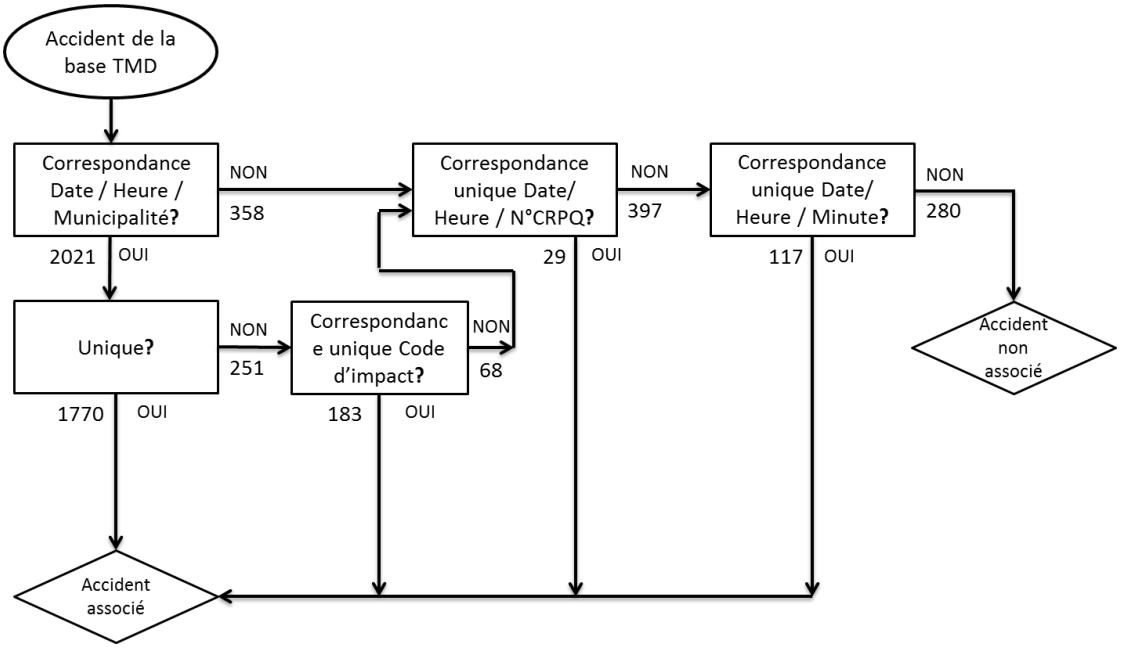


Figure 3.7 : Méthodologie d'identification et résultats obtenus

Environ 80% des accidents ont ainsi pu être identifié. La programmation de la première recherche par Date / Heure / Municipalité est fournie dans l'annexe 5.

3.2.2.3 Vérification des associations trouvées

Il n'est cependant pas sûr que les associations trouvées soit toutes pertinentes. Afin de ne vérifier que les accidents « douteux » on peut calculer un indicateur de qualité de l'association. Le nombre de champs communs entre deux accidents associés a été choisi comme indicateur, et a été calculé via un programme VB disponible dans l'annexe 5. 21 champs potentiellement identiques ont été identifié (Date, Heure, Minute, municipalité, N°CRPQ, météo….) et les résultats obtenus varient entre 3 et 19 champs identiques. Les accidents avec une les plus faible valeurs ont donc été vérifié en premier jusqu'à atteindre une valeur suffisamment élevée pour qu'il n'y ait manifestement plus d'erreur.

Afin de vérifier en un coup d'œil une association, un « comparateur » a été réalisé. Basé sur l'outil de visualisation des deux bases de données (cf. §3.2.1.3), ce comparateur permet d'afficher sur une même page toutes les informations concernant les deux accidents associés. Voici comment il se présente :

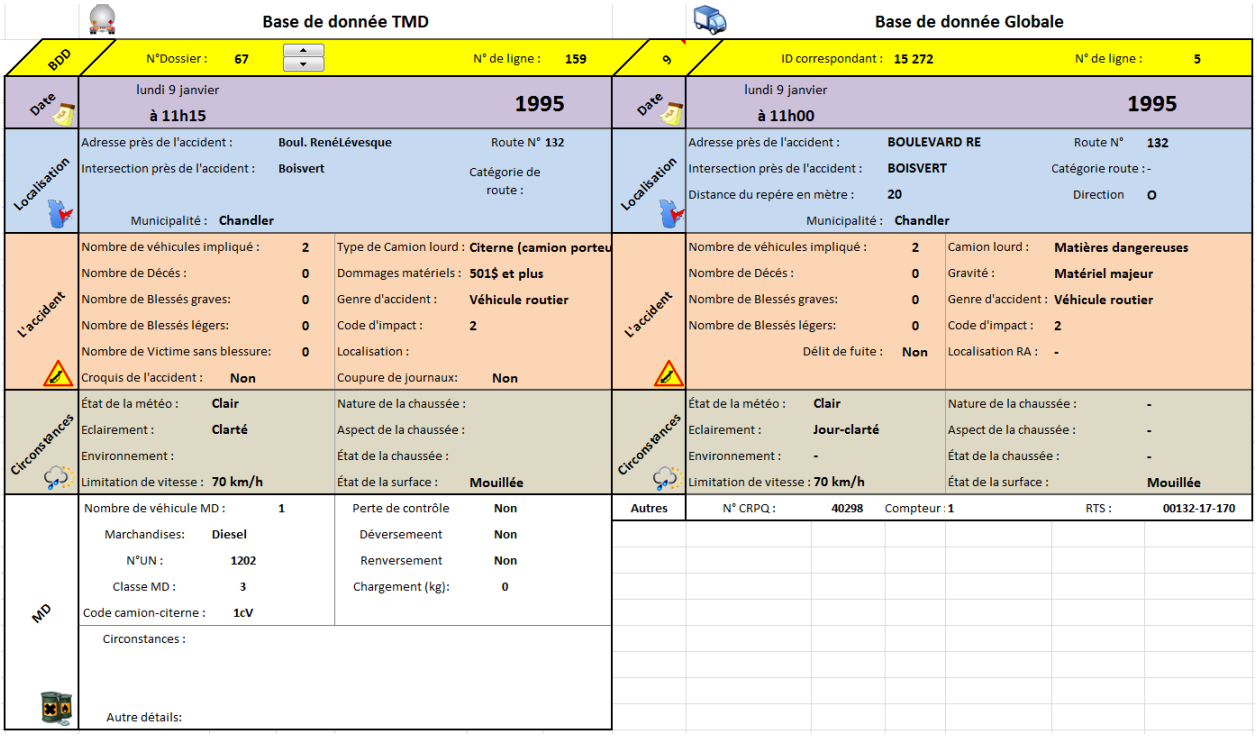


Figure 3.8 : Comparateur

Une vingtaine d'association manifestement fausses ont de cette façon été identifiées.

3.2.2.4 Recherche « manuelle »

Au final 2086 accidents ont été associés, il reste donc 293 accidents non associés. Nous avons essayé d'élargir la recherche en utilisant d'autres champs mais les résultats n'ont pas été convaincant notamment car la majorité des accidents concernés ont un grand nombre de champs vides. Il a donc été choisi à ce stade de rechercher manuellement les accidents restant. Pour cela la fonction « filtrer » d'Excel ainsi que le comparateur ont été utilisé et nous sommes parvenu à associer des accidents ayant des erreurs de date, d'heure, ou de municipalité en utilisant des tris sur des champs non utilisés dans la recherche VB (ex : la météo, l'état de surface,…). Cependant tous les accidents n'ont pas été identifiés. Par exemple certains accidents de la base MD comportent la mention « Pas accident de la route » (16 accidents), c'est-à-dire qu'après avoir saisi l'accident dans la base, le rapporteur a jugé qu'il ne s'agissait pas d'un accident routier. Par exemple un camion ayant eu une fuite à l'arrêt. Ces accidents étant introuvables dans la base globale, nous en avons déduit qu'ils ne figuraient pas dans la base de façon certaine. Une croix a donc été mise dans la colonne ID afin de bien indiquer qu'aucune association n'est possible.

Cette recherche manuelle à tout de même permis de trouver 194 accidents sur les 293 restants. 99 accidents de la base TMD (≈ 4%) n'ont pas été trouvés dans la base globale.

3.2.2.5 Conclusion

2280 accidents ont au final été identifiés, voici le bilan des associations trouvées :

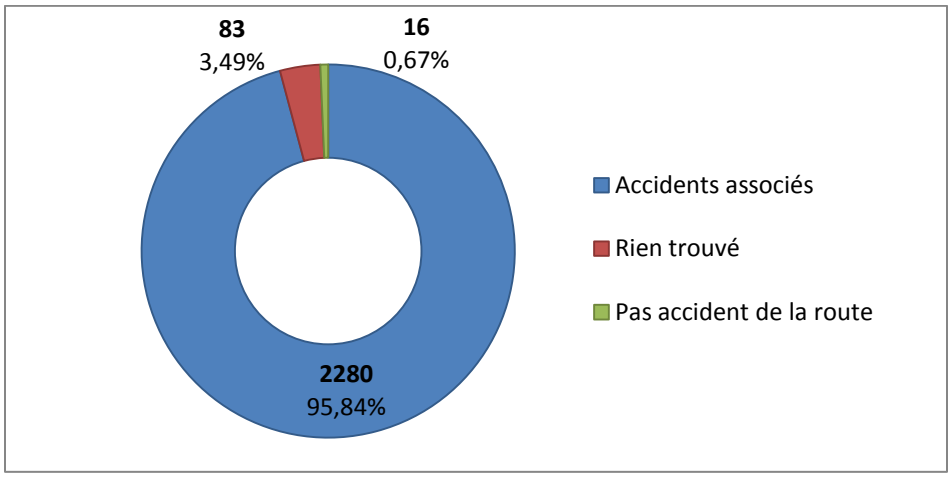


Figure 3.9 : Bilan des associations

Ces résultats sont tout à fait satisfaisants. Voici pour finir un comparatif de l'efficacité des cinq étapes de la recherche d'association :

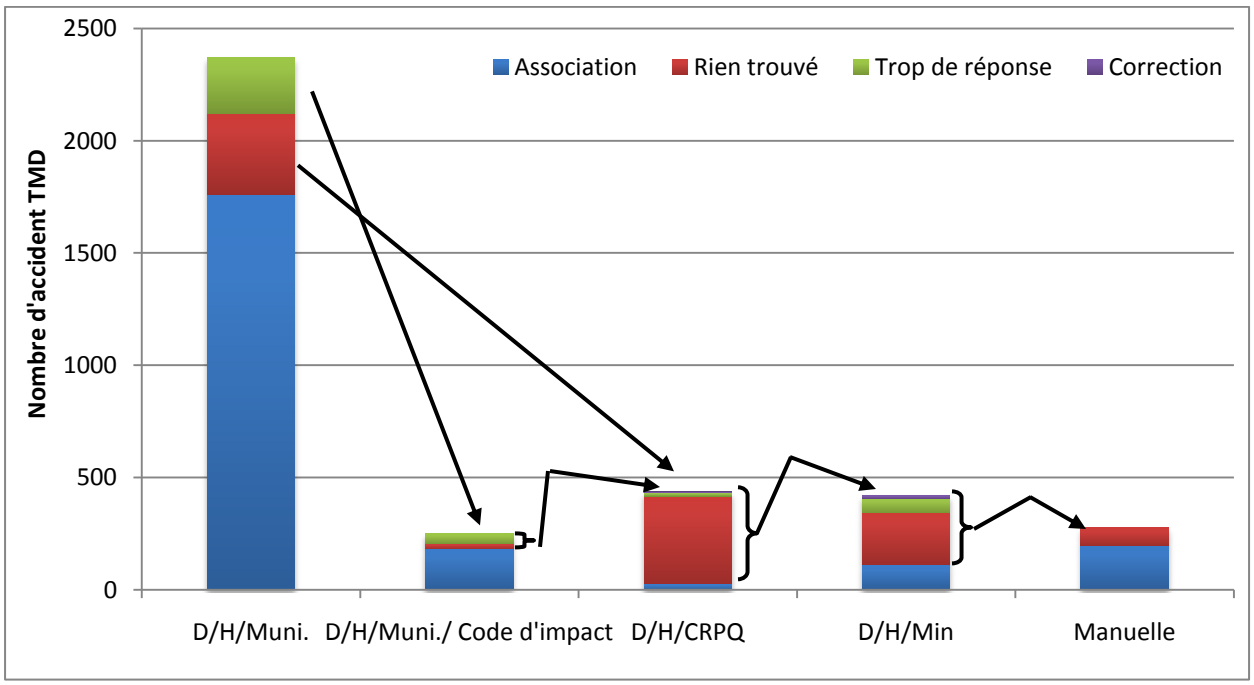


Figure 3.10 : Comparaison des résultats de chacune des étapes

On constate que l'étape 1 a été de loin la plus importante source d'associations pertinentes. Les recherches suivantes ont donné des résultats plus mitigés avec pour les étapes 3 et 4 beaucoup d'échecs dans la recherche, et des valeurs qui ont nécessité des corrections manuelles.

Un champ « matières dangereuses » de type Oui/Non a ainsi pu être ajouté à la base globale, permettant d'indiquer si l'accident se trouve également dans la base TMD, et dans celle-ci un nouveau champ indique l'ID de l'accident dans la base globale.

3.2.3 Corrections et enrichissement de la base TMD

Les accidents TMD ayant été majoritairement identifiés dans la base globale il est possible afin d'améliorer la finesse des résultats de corriger et de compléter les informations sur un accident TMD en combinant ces deux sources d'informations. Ainsi à l'aide d'une procédure VB, les champs vides de la base TMD ont été complétés par les informations de la base globale (lorsqu'elles étaient disponibles). Afin d'identifier ces modifications un fond rouge a été appliqué à toutes les cellules ayant été complétées :

Adresse près de l'accident	Intersection près de l'accident	Municipalité	Nombre de véhicule	Dommages matériels	Code l'impact	Vitesse autorisée	Morts (code 1)	Blessés graves (code 2)	Blessés mineurs (code 3)	Victime sans blessure (code 4	5 (Genre d'accident)	6 (Environnement)	7 (Catégorie de route)	8 (État de la surface)
180 RG BUREAU		Saint-Léonard d'Aston	1				0	0	0	0	Quitter chaussée			
Notre Dame Est	Sainte-Catherine Est	Montréal	1		16		0	0	0	0	Capotage	Industriel/Manu fac.	Rue	Mouillée
KM 184,7	Boul. Gagne	Sorel	1		15	70 km/h	0	0	0	0	Garde-fou	Résidentiel	Route numérotée	Sèche
2016 Saint Joseph		Nominingue	1		15		0	0	0	0	Lampadaire/Pot eau			Glacée
Ile aux chats	Voie Ferré subdivison Valleyfield	Salaberry-de-Valleyfield	2		9		0	0	1	0	Train	Rural	Rue	Sèche
LAC DES SITELLES	DES HETRES	Austin	2		10	50 km/h	0	0	1	0	Véhicule routier	Résidentiel	Rue	Enneigée
219 LAURIER	CHAUREST	Ile Bizard	2		10		0	0	1	2	Véhicule routier	Résidentiel	Rue	Mouillée

Figure 3.11 : Exemple de cellules complétées

Ensuite, avant chaque analyse d'un nouveau champ, nous avons essayé de corriger (ou d'éliminer) les valeurs aberrantes en s'aidant éventuellement des informations de la base globale. Pour des données quantitatives l'utilisation des extremums permet souvent de détecter de tel valeurs (Trépanier, 2010). Ainsi par exemple un accident avait un nombre de véhicule impliqué de 368, la seconde plus grande valeur observé étant de 16 véhicules. En comparant avec les données de la base globale il s'est avéré que ce nombre correspondait en réalité au numéro de la route où l'accident a eu lieu, le nombre de véhicules impliqués étant de deux. Cette valeur extrême avait une influence importante sur les valeurs des paramètres de statistique descriptive avec une moyenne de près de deux véhicules par accident au lieu de 1,8 et surtout un écart type

de 7,6 au lieu 0,7 (l'écart-type est très sensible aux valeurs extrêmes). Cette correction s'est donc avérée particulièrement importante.

3.3 Analyse comparative des données

Maintenant que nous avons deux bases de données aussi complètes et fiables que possible, l'exploration des données peut commencer.

L'ensemble des champs communs (une vingtaine) entre les deux bases ont été étudiés. Les champs des bases de données peuvent être quantitatifs (numériques), qualitatifs (variable nominale) ou binaires (de deux type). Il n'y a pas de champ commun binaire donc seul les deux autres types de champs ont été étudiés. Les deux sections suivantes présentent comment ces champs ont été étudiés.

3.3.1 Problème de la dépendance des données

Afin de faire des comparaisons pertinentes il convient de faire le point sur les liens entre nos variables. Nous avons d'une part une base de données d'accidents impliquant des véhicules lourds, c'est une image de l'ensemble des accidents impliquant des véhicules lourds qui se veut aussi complète que possible. Cette base de données contient donc aussi les accidents impliquant des véhicules lourds transportant des matières dangereuses, accidents qui constituent notre seconde base de données. Comme le montre l'étape d'identification des accidents TMD dans la base globale il y a des accidents TMD a priori non présent dans la base globale (99 accidents) et l'inverse peut également être vrai (difficile à quantifier). Pour résumer la situation voici un diagramme de Venn :

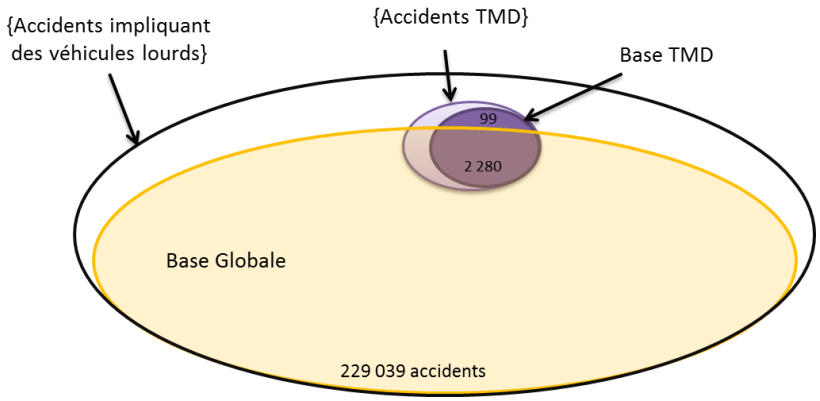


Figure 3.12 : Diagramme de Venn (période 1995-2007)

Ce diagramme met bien en évidence, parmi les accidents renseignés dans nos données, l'existence d'accidents MD hors de la base globale et d'accidents MD dans la base globale hors de la base TMD.

Dans notre analyse on cherche à étudier l'influence du caractère « présence de matières dangereuses lors du transport » par rapport aux autres types de marchandises. Si l'on conserve les accidents TMD de la base globale dans la comparaison on vient biaiser celle-ci, car on compare deux variables dépendantes. Il faudrait donc, lorsque l'on compare deux champs, éliminer de la base de données globale les accidents TMD qui ont été identifiés. Cependant cela ne permet pas d'éliminer totalement le biais car il y toujours des accidents TMD qui n'ont pas été identifiés. On peut cependant considérer que les échantillons sont alors « quasi-indépendants » notamment à la vue des ordres de grandeur des nombres d'échantillons.

Si cette opération n'est pas réalisée, l'erreur est au pire d'environ 3000 accidents sur 229039 accidents, soit 1% d'erreur. Toutefois, pour la précision et la cohérence des résultats, cette opération a été faite pour chaque champ.

3.3.2 Données quantitatives

Pour ces champs nous pouvons utiliser les outils de la statistique descriptive. La statistique descriptive (ou exploratoire) permet de résumer un échantillon de données quantitatives par des caractéristiques numériques et à l'aide de graphiques. Cela permet de mettre en évidence des différences ou des similitudes dans les données des deux bases.

Les paragraphes suivants passent en revue les méthodes utilisées pour l'analyse. Les informations proviennent essentiellement des ressources éducatives de statistique Canada (Statistique Canada, 2010) ainsi que du cours de statistiques suivi à l'ENSAM (Bastet, 2008).

3.3.2.1 Organisation des données

Il existe trois façons d'organiser les données d'un champ. La plus simple est de conserver la suite de valeurs du champ, on parle alors de série. Suivant le type de données on peut aussi utiliser un tableau d'effectif, indiquant la fréquence d'apparition des différentes valeurs du champ, ou une série classée, indiquant la fréquence d'apparition d'une classe de valeurs. Les fréquences peuvent être suivant les besoins estimées en %, on peut aussi utiliser les fréquences cumulées.

Pour déterminer les effectifs des valeurs ou des classes définies, on peut utiliser soit la fonction tri d'Excel (lorsque le nombre de valeurs est faible), ou l'utilitaire d'analyse « histogramme » d'Excel. Cet utilitaire d'analyse calcule les fréquences d'apparition de classes préalablement définies sur la feuille de calcul. C'est particulièrement utile lorsque le nombre de valeurs ou de classes est grand (comme par exemple pour les dates de l'année).

Pour étudier un même champ, plusieurs organisations peuvent être utilisées, lorsque cela est pertinent, pour faire ressortir des caractéristiques différentes. Le tableau suivant présente les organisations des données qui ont été utilisées pour étudier les champs :

Tableau 3.2 : Organisation des données utilisées

Série	Tableau d'effectifs	Série classée
		Date (classée selon année, mois, saison, jour de l'année et de la semaine)
	Heure et minute	
Nombre de véhicules impliqués	Nombre de véhicules impliqués	
Limitation de vitesse	Limitation de vitesse	
	Champ sur les impacts humains	

3.3.2.2 Réduction numérique des données

Pour résumer des données on peut utiliser tout un lot de paramètres représentant les caractéristiques des données. On parle de statistique descriptive. Ces calculs peuvent s'appliquer à la série de données « brutes » si la variable est quantitative, mais aussi à un tableau d'effectif ou une série classée. On peut regrouper ces différentes valeurs en trois catégories selon ce qu'elles représentent : les paramètres de position, de variabilité et de forme. Pour calculer ces paramètres on utilise les fonctions de calcul d'Excel qui sont décrites dans l'aide du logiciel (Microsoft, 2010) ainsi que l'utilitaire d'analyse d'Excel « statistiques descriptives ». Une description des paramètres utilisés, ainsi que leur fonction Excel est fournie en annexe 6. La comparaison de ces valeurs fournie un nombre important d'indications sur les points communs et les différences des deux échantillons.

Voici un exemple d'utilisation d'une réduction numérique portant sur la gravité des accidents du travail impliquant des matières dangereuses (Carol, et al., 2002) :

	No. of dead	No. of injuries	No. of evacuated	Material damage[a]
No. of accidents	1694	1482	554	280
Min	0	0	0	0
Max	2001	10,000	200,001	428.46
Average	5.090	26.20	2157.7	23.98
Mode	0	0	0	5.469
Median	1	3	2	3.845
Standard deviation	51.91	280.6	13,932	60.208

Figure 3.13 : Réduction numérique de données (Carol, et al., 2002)

Ce tableau permet, par exemple, de détecter que dans la répartition du nombre de morts il y a certains évènements aux conséquences hors de l'ordinaire (2001 morts au maximum alors que la moyenne est 5, le mode de 0 et la médiane de 1).

3.3.2.3 Réduction graphique des données

Les méthodes de réduction graphique sont très utiles car elles résument de façon visuelle les données. On peut ainsi comparer facilement des données en comparant deux graphiques ou même, lorsque c'est possible, en superposant leurs représentations graphiques. Dans notre cas il faut cependant faire attention aux problèmes de la différence de « taille » entre les deux bases de données, avec une base de données avec 222 000 lignes contre seulement 2 400 pour la seconde. Pour s'affranchir de ce problème une bonne façon est de déterminer la fréquence d'apparition d'une valeur (que l'on exprime souvent en pourcentage), puis de tracer une courbe comparative avec les données obtenues. La fréquence s'obtient en divisant l'effectif d'une valeur par le nombre total d'échantillon, ce qui élimine l'effet de taille.

$$X' = \frac{X}{\sum x_i} \times 100$$

On peut aussi standardiser les variables ce qui consiste à transformer cette variable en une nouvelle variable de moyenne nulle et de variance 1. Ainsi en appliquant cette transformation aux deux distributions de moyenne et d'écart-type différents on obtient deux distributions de même moyenne et de même écart-type, donc deux distributions comparables. Voici la formule pour transformer la variable aléatoire X :

$$X' = \frac{X - Moyenne(X)}{Ecart_type(X)}$$

3.3.2.4 Statistique descriptive à deux variables

Les paramètres décrits précédemment permettent de résumer une série de données. Cependant pour déterminer s'il existe ou pas une corrélation entre deux séries de taille identique il faut utiliser les outils de la statistique descriptive à deux variables.

- Nuage de point

Un nuage de point est la représentation des valeurs d'un échantillon en fonction des valeurs de l'autre échantillon. Ce graphique permet de déterminer qu'elle est la nature de la relation qui lie les deux variables. Si les points semblent suivre une courbe on peut penser que les variable sont dépendantes.

- Caractéristiques numériques

Tableau 3.3 : statistique descriptive à deux variables

	Définition	Fonction Excel	Remarques
Covariance	Moyenne de la somme du produit des écarts des valeurs des deux variables par rapport à leur moyenne arithmétique	=COVARIANCE.STANDARD()	Si nulle, pas de relation Si grande, relation forte entre les deux variables Si négative, varient en sens inverse
Coefficient de corrélation	Covariance divisé par les écart-types de chacune des variables	=COEFFICIENT.CORRELATION()	Plus il est proche de ±1 plus la relation linéaire est forte. Le signe indique le sens de la relation
Coefficient de détermination	Carré du coefficient de corrélation	=COEFFICIENT.DETERMINATION()	S'il est proche de 1, l'ajustement linéaire est légitime

- Ajustement

Le problème de l'ajustement est de déterminer la fonction f, qui, d'un point de vue graphique, s'ajuste le mieux au nuage de point. L'allure du nuage de point permet bien souvent de déterminer la famille de fonctions pertinentes (linéaire, exponentielle, polynôme,…). Les valeurs des coefficients de corrélation ou de détermination permettent de juger d'une relation linéaire (ajustement affine). Les travaux d'Ohtani et Kobayashi (2004) utilisent par exemple des régressions linéaires afin de chercher des liens entre le nombre d'accidents et certains facteurs.

3.3.2.5 Test d'hypothèse

Un test d'hypothèse est une démarche statistique permettant d'accepter ou de rejeter une hypothèse dite hypothèse nulle (notée Ho). Il existe de nombreux tests, cependant dans notre étude, seul des tests sur la moyenne (test d'homogénéité) ont été réalisés. Ces tests permettent de tester l'hypothèse nulle « les deux variables aléatoires ont la même moyenne ». Ils peuvent être facilement réalisés avec la formule =TEST.STUDENT() d'Excel. Pour appliquer ce test il faut suivre les étapes suivantes :

1. Variable indépendante

Les variables étudiées doivent être indépendantes. C'est-à-dire dans notre cas qu'il faut éliminer de la base globale les accidents identifiés comme impliquant des matières dangereuses. Dans le cas d'un test sur une série de données issues de nos deux bases de données cela n'aura presque aucune influence sur le résultat final car l'échantillon de la base globale est beaucoup plus grand que celui de la base TMD. Il faut cependant le faire pour vérifier cette hypothèse.

2. Hypothèse de normalité

Le test de Student suppose que les données suivent une distribution selon la loi normale. Afin de faire le test il faut donc vérifier que la distribution est normale. Pour cela on s'aide des valeurs des paramètres de forme (cf. 3.3.1.2), c'est-à-dire du coefficient d'asymétrie et d'aplatissement. S'ils sont nuls la distribution est normale. On rejettera l'hypothèse de normalité lorsque leurs valeurs dépassent l'intervalle [-2,2].

Lorsqu'une distribution n'est pas normale il faut la normaliser avant d'appliquer un test de Student. Pour cela on utilise souvent les fonctions logarithmiques, racine carrée, inverse…(Daumas, 1982) Il faut alors déterminer en comparant les valeurs des paramètres de forme quelle est la transformation la plus efficace.

3. Test bilatéral ou unilatéral?

Un test peut être unilatéral ou bilatéral. Dans notre cas on cherche toujours une différence entre deux estimations sans se préoccuper du signe ou du sens de la différence. On choisira donc un test bilatéral.

4. Choisir le type de test

Le test peut être « sur des observations pairées » lorsque les deux échantillons ont des valeurs pairées (ce qui implique forcément deux échantillons de même taille). Sinon il faut déterminer si les échantillons sont de variances égales ou différentes. Pour cela on utilise un test de Ficher, que l'on réalise sous Excel avec la fonction =TEST.F(matrice1; matrice2).

- Réalisation et interprétation du test

Nous avons maintenant tous les éléments pour réaliser le test. La syntaxe sous Excel est la suivante : TEST.STUDENT(matrice1; matrice2; uni/bilatéral; type). Le type vaut 1 pour un test sur des observations pairées, 2 sur deux échantillons de variances égales et 3 sur deux échantillons de variances différentes. La valeur calculée est la probabilité de se tromper en acceptant H_0. On considère généralement comme acceptable un seuil de 5%. Donc si la valeur-p est inférieure à 0,05, H_0 sera rejetée.

Les tests de Student et le test de Fischer peuvent aussi être réalisés avec les utilitaires d'analyses correspondant dans Excel (« Test d'égalité des espérances : … » et « Test d'égalité des variances (F-Test) » (Microsoft)). Excel affiche alors plus d'informations sur le test.

Test de Student:		
Test d'égalité des espérances: deux observations de variances différentes		
	Variable 1	*Variable 2*
Moyenne	1,860552133	1,796973518
Variance	0,505257532	0,53613886
Observations	229039	2379
Différence hypothétiq	0	
Degré de liberté	2425	
Statistique t	4,214581026	
P(T<=t) unilatéral	1,29701E-05	
Valeur critique de t (u	1,645482227	
P(T<=t) bilatéral	2,59402E-05	
Valeur critique de t (bi	1,960942721	

Figure 3.14 : Exemple de test de Student avec Excel

Dans l'exemple ci-dessus en trouve en jaune la valeur calculée par la formule TEST.STUDENT. Celle-ci est inférieure à 0,05, dans ce cas on rejetterait donc H_0, c'est-à-dire que les moyennes sont différentes.

3.3.2.6 Courbes f-N

Les courbes f-N (ou p-N) sont souvent utilisées dans les études concernant les matières dangereuses pour étudier les conséquences humaines des accidents (cf. § 2.1.4). Ces courbes permettent de présenter la fréquence f(N) ou la probabilité p(N) d'un type d'événement (ex : accidents impliquant des matières dangereuses) en fonction d'une variable N d'étude (souvent le nombre de décès ou de blessés). Les valeurs des fréquences ou des probabilités et de N étant généralement sur plusieurs ordres de grandeur une échelle logarithmique est souvent utilisée. Ce type de graphique est utile pour déterminer si la probabilité qu'un événement avec une variable N se produise est proportionnelle à la probabilité qu'un autre accident avec une variable N' se produise.

Pour estimer les fréquences ou les probabilités on peut utiliser les données d'une base de données ou un modèle de probabilité (Bubbico, Ferrari, & Mazzarotta, 2000). Dans notre cas, nous utiliserons les données de nos deux bases de données d'accidents, en calculant la probabilité cumulative $P(x \geq N)$ comme on le trouve dans plusieurs études ((Planas-Cuchi, et al., 1997; Vílchez, et al., 1995; Yang, et al., 2010) par exemple) :

$$P_{(j \geq N)} = F_j = \frac{\sum_{i=j}^{n} N_i}{\sum_{i=1}^{n} N_i}$$

Ainsi, si par exemple la variable d'étude N (axe x) est le nombre de décès, $P(j \geq N)$ est la probabilité qu'un accident avec j décès ou plus se produise, n est le nombre maximum de décès observé.

Avec cette formule la courbe passera par le point (1;1), ce qui permettra de comparer par superposition les courbes obtenues pour nos deux bases de données. On peut alors calculer la pente (slope) qui correspond au coefficient de puissance d'une régression polynomiale, ce qui nous permettra de quantifier l'écart entre nos deux courbes et de comparer nos résultats à ceux obtenus dans d'autres études. Voici par exemple la courbe obtenue par Darbra et Casal sur le nombre de décès dans les accidents ayant eu lieu dans un environnement portuaire:

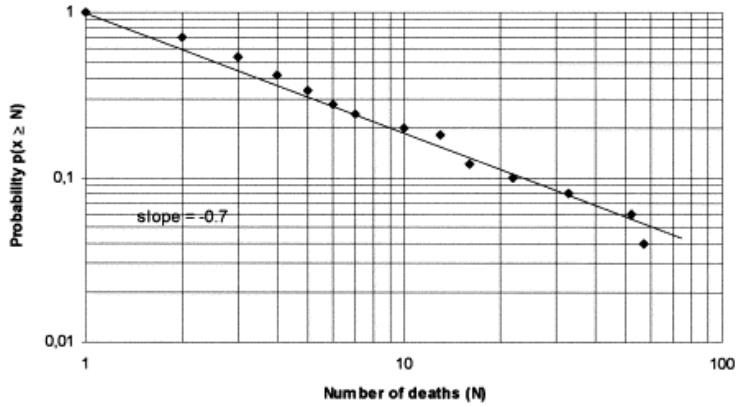


Figure 3.15 : Probabilité cumulée d'un accident avec N décès (courbe f-N) (Darbra & Casal, 2004)

La pente de $-0{,}7$ obtenue par exemple peut s'interpréter comme suit : la probabilité d'accident avec 1 décès ou plus est 7 fois supérieure à la probabilité d'un accident avec 10 décès ou plus.

3.3.3 Données qualitatives

Les données qualitatives concernent une dizaine de champs communs aux deux bases de données. Elles indiquent par exemple les conditions météorologiques ou les conditions de visibilité lors de l'accident. Pour traiter ces données les outils de la statique descriptive sont inutilisables.

- Détermination des fréquences

Pour étudier un champ qualitatif on détermine tout d'abord les fréquences d'apparition des différentes valeurs. Pour se faciliter la tâche on peut affecter une valeur numérique à chaque information. Par exemple dans une étude (Qiao, Keren, & Mannan, 2009), la variable « état de surface » a été codifiée de la façon suivante : 1=sec, 2=humide, 3=boueux, et 4=neige / glace. Cela permet de diminuer la taille de la base de données mais facilite aussi le traitement informatique du champ. Les fréquences peuvent alors être déterminées comme pour un champ quantitatif, avec l'utilitaire d'analyse « histogramme » d'Excel.

- Détermination du mode et diagramme en tuyaux d'orgue

Une fois les fréquences d'apparition connues, on peut déterminer le mode (valeur la plus fréquente) et tracer un diagramme en tuyaux d'orgue ou un diagramme circulaire. Pour contrer l'effet de taille, on détermine les pourcentages des fréquences d'apparition tout comme pour les

variables quantitatives. Cependant si le champ est codé numériquement tout autre paramètre numérique (moyenne, médiane,…) n'a aucun sens.

- Test de proportion

Afin de comparer les distributions dans nos deux bases de données on peut appliquer un test de proportion. Il faut pour cela préparer les données en codant le paramètre que l'on teste. Pour cela on attribue une valeur de 1 lorsque le paramètre est vérifié et 0 sinon. Par exemple si on compare les proportions d'accidents ayant lieu sur une surface humide, on attribue un 1 à chaque accident dont la surface est indiquée comme humide et un 0 pour les autres enregistrements. On peut alors appliquer un test de Student (cf. § 3.3.2.5) aux deux champs obtenus en déterminant au préalable par un test de Fisher si les variances sont égales ou différentes. L'hypothèse de normalité n'est pas à vérifier dans ce cas.

3.3.4 Cas particulier du champ « Date »

Le champ date contient des données particulières au format « date ». Ce n'est pas vraiment une variable quantitative puisqu'elle combine en réalité trois données quantitatives qui sont le jour, le mois et l'année. Une bonne façon d'étudier ce champ est donc d'extraire une variable quantitative du format date d'Excel (par exemple l'année) et de l'étudier séparément. Pour extraire ces données d'une cellule au format « date » on utilise principalement la fonction =TEXTE(cellule;"Format"). On peut ainsi étudier la répartition annuelle, mensuelle et journalière des accidents.

Enfin le champ date permet d'étudier la variation dans le temps des valeurs d'autres champs. Par exemple on peut tracer une courbe d'évolution comparée du nombre de véhicules impliqués par accident sur la période 1995-2007.

3.4 Analyse de la répartition spatiale des accidents

Une fois l'analyse statistique terminée, une analyse spatiale comparative a été menée. Les accidents n'étant pas localisés pour les deux bases de données il a fallu commencer par cela (§ 3.4.1). Un système d'information géographique permettant de générer des cartes d'analyse a ensuite été mis en place (§ 3.4.2) et enfin une analyse spatiale comparative a été réalisée (§

3.4.3). Les trois paragraphes suivant décrivent les méthodes utilisées pour mener à bien cette étude.

3.4.1 Localisation des accidents

On trouve dans les deux bases de données des informations permettant de localiser les accidents. Ces champs donnent l'adresse de l'accident, l'intersection, le N° de la route, la distance à un repère (borne kilométrique par exemple), la direction et la municipalité. Nous avons cependant besoin des coordonnées en terme de longitude et latitude pour pouvoir placer les accidents sur une carte. Toute la difficulté réside donc en la conversion des diverses informations de localisation contenues dans ces champ en coordonnées. Il est assez difficile d'automatiser le processus en utilisant toutes ces données car les champs sont souvent incomplets, pas toujours remplis selon la même syntaxe, utilisent des abréviations,… Dans l'extrait suivant on peut voir par exemple que l'adresse n'est pas toujours remplie et que les noms des rues sont souvent « coupés » comme pour la rue Papineau qui devient PAPINEA :

Route	Adresse près de l'accident	Intersection près de l'accident	Distance du repère en mètre	Direction	Municipalité RA act.
-	455 MC CONNE	VANIER	0	-	Gatineau
540	-	KM 1	0	-	Vaudreuil-Dorion
-	620 ST CHARL	VICTORIA	10	E	Longueuil
-	2599 PAPINEA	-	0	-	Longueuil
20	-	SORTIE 12	0	-	Les Coteaux

Figure 3.16 : Les champs qui permettent de localiser un accident de la base globale

Le champ « Municipalité » est par contre beaucoup plus fiable et une solution simple, mais peu précise, aurait été de localiser les accidents en utilisant uniquement ce champ. On peut cependant utiliser un peu plus de champ grâce aux fonctionnalités offert par le logiciel MapPoint (Microsoft, 2009). Ce logiciel est capable de localiser et de placer des points en fonction de l'adresse, en plus du champ municipalité. Lorsque qu'il parvient à comprendre l'adresse, la localisation est alors précise. Cependant lorsque ce n'est pas le cas, l'accident sera localisé seulement avec la municipalité, c'est-à-dire que par exemple pour Montréal ce sera au point « Montréal » de MapPoint, c'est-à-dire en centre-ville. En première approximation cette méthode

de localisation a cependant été choisie. Voici par exemple la répartition obtenue pour la base TMD :

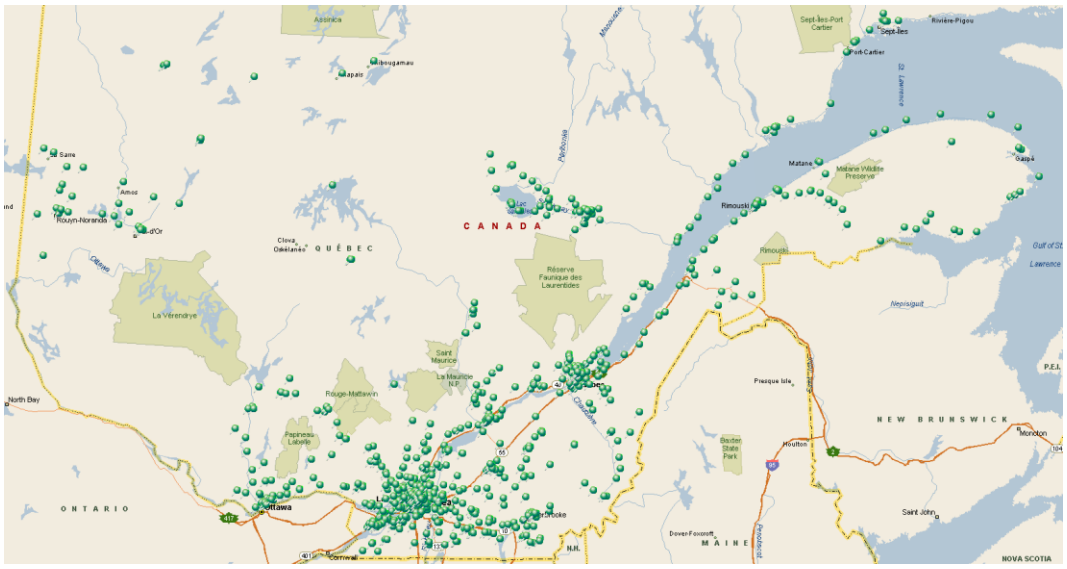


Figure 3.17 : Localisation de la base TMD via MapPoint

On constate cependant que dans les grandes villes, il y a beaucoup d'accidents localisés au même point, c'est-à-dire pour lesquels le champ adresse n'a pas permis à MapPoint de situer avec plus de précision l'accident.

Afin d'augmenter la qualité des résultats, les accidents TMD mal placé ont été relocalisés « manuellement » à l'aide Google Maps, ce qui a permis de prendre en compte l'ensemble des informations localisant les accidents (intersection, distance du repère, direction,…). Pour la base globale il n'est cependant pas envisageable, vu le nombre d'accidents à relocaliser (de l'ordre de 130 000 accidents) de faire la même chose. Pour se faire une idée, le rythme maximal de localisation manuelle, une fois notre méthode rodée, nous permettait de traiter environ 200 accidents par jour. A cette vitesse il aurait fallu près de 2 ans pour relocaliser les accidents de la base globale...

Pour positionner manuellement un accident, on a dans un premier temps identifié les accidents à relocaliser en détectant des concentrations de point. En utilisant les fonctions de Google Maps nous avons ensuite positionné ces accidents en utilisant toute les informations disponibles sur la localisation (adresse, intersection, N° route, direction, distance au repère kilométrique). Enfin les nouvelles coordonnées ont été exportée en utilisant les fichiers au format kml que Google Maps

est capable de générer. La localisation des accidents de l'ensemble des grandes villes québécoises comme Montréal, Longueuil, Gatineau, Laval, Québec, Sherbrooke, Lévis, Saguenay, Trois-Rivières, Terrebonne ont ainsi été précisé. L'amas le plus grand était situé au centre-ville de Montréal avec plus de 200 accidents. Voici par exemple la relocalisation de « l'amas » de Lévis.

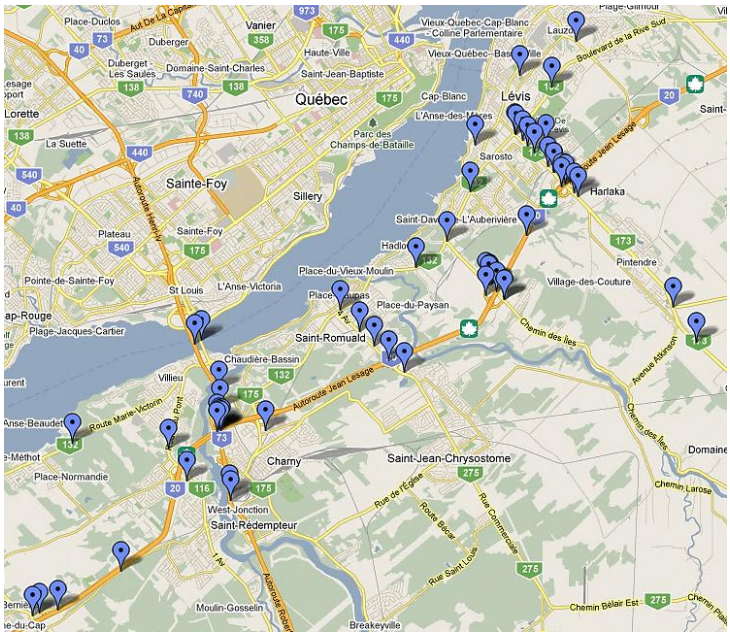


Figure 3.18 : Exemple de localisation Google Maps des accidents initialement localisés à Lévis (centre-ville) par MapPoint

Tous les accidents localisés sur cette carte avaient été positionnés par MapPoint au même endroit, on peut ainsi constater le gain de précision apportée par cette relocalisation. En cherchant des « amas », un amas imprévu de taille conséquente (250 accidents!) a été trouvé en plein milieu du Québec, dans la région du Nord du Québec. Tous ces points correspondent en fait au cas où MapPoint n'a pas été capable de reconnaitre la municipalité. Ne connaissant alors que la province il a placé les accidents au « Québec », ce qui correspond dans les données du logiciel à un point au milieu du Québec. Une partie de ces municipalités « inconnues » sont en fait de petites localités comme par exemple Lac-Ashuapmushuan, Saint-Maxime-du-Mont-Louis, Saint-Pierre-les-Becquets,… Pour les localiser le répertoire des municipalités (MAMROT, 2010) a été utilisé. D'autres cas « d'échec » étaient dus à une ou des fautes d'orthographes sur la municipalité (ex : Montéal au lieu de Montréal), où à une absence d'information (champ vide). En comparant avec la base globale ces champs ont été corrigés. L'ensemble des erreurs qui ont été repérées en parcourant tous ces accidents ont d'ailleurs été corrigées (adresse coupée, incomplète, fautes

d'orthographe,…). Comme mentionné dans le paragraphe 3.2.3, elles ont été identifiées dans le fichier par un fond de cellule de couleur verte. Voici le résultat obtenu pour l'île de Montréal :

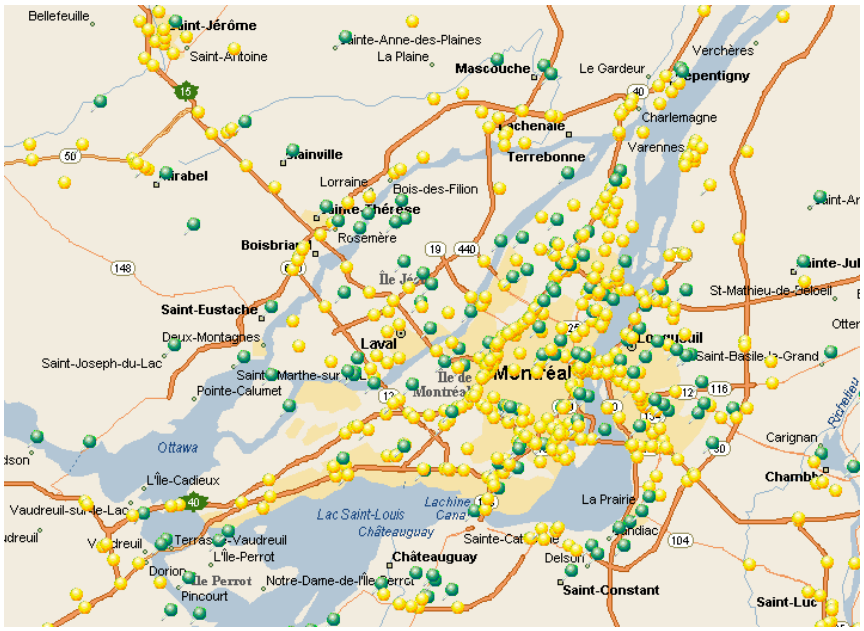


Figure 3.19 : Localisation des accidents TMD de Montréal et de ces environs (en vert = localisation MapPoint, en jaune = localisation manuelle)

En tout 1175 accidents de la base TMD ont été placés « à la main ». Une trentaine d'accidents n'ont pas pu être localisés en raison du manque d'information sur leur localisation (municipalité vide ou introuvable). Voici le bilan de la localisation des accidents TMD :

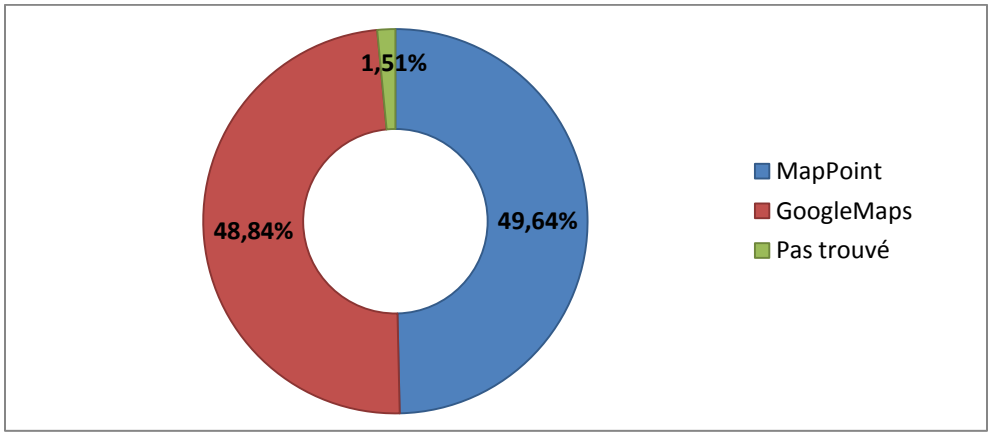


Figure 3.20 : Origine de la localisation de la base TMD

La moitié de la base TMD a donc été localisée par MapPoint, l'autre moitié ayant été localisée manuellement. Le résultat nous semble tout à fait acceptable en termes de précision. On peut

raisonnablement estimer, en se basant sur ce ratio, que la moitié des accidents de la base globale ne sont pas placés de façon très précise, c'est à dire au niveau d'une municipalité.

3.4.2 Montage du système d'information géographique

Une fois les accidents localisés, nous pouvons mettre en place le système d'information géographique. Un système d'information géographique (SIG en français, GIS en anglais) est un outil qui permet le stockage, le traitement et l'affichage de données à caractère géographique (Trépanier, 2010). Les données sont regroupées en objets de même nature pour constituer des couches (layers). Le montage du système d'information géographique est l'étape qui vise à déterminer puis à assembler dans un même projet les données spatiales nécessaires à l'étude. La première étape consiste à choisir les données dont nous avons besoin. Dans notre cas nous souhaitons étudier comparativement deux répartitions d'accidents de véhicules lourds aux Québec. Nous avons donc besoin d'informations concernant la géographie du Québec, et comme les accidents concernent le réseau routier, des données sur celui-ci sont aussi nécessaires. Enfin, afin de discuter de la répartition des accidents par rapport à la répartition de la population nous utiliserons aussi des données sur la répartition de la population québecoise. L'ensemble de ces données sont disponibles sur le site internet de la CREPUQ (Bibliothèques universitaires québécoises, 2007) dont l'accés est réservé aux universités.

3.4.2.1 Présentation de l'application SIG utilisée

Nous avons choisi d'utiliser une application SIG gratuite. Parmi les nombreux programmes existant nous avons retenu le logiciel MapWindow GIS (Ames & MapWindows GIS Team, 2010), une application open source. Ce programme a été développé à partir de 1998 dans un laboratoire de recherche sur les eaux à Logan dans l'Utah, aux Etat-Unis. Le logiciel possède une interface simple partiellement traduite en français et il est compatible avec de nombreux formats de fichier. Il permet, entre autre d'importer des données depuis Excel. Un guide de prise en main rapide (Croft, Watry, Ames, & Michaelis, 2007) est disponible ainsi qu'une documentation en anglais complète et gratuite (MapWindows GIS Team, 2010). Voici un aperçu de son interface graphique :

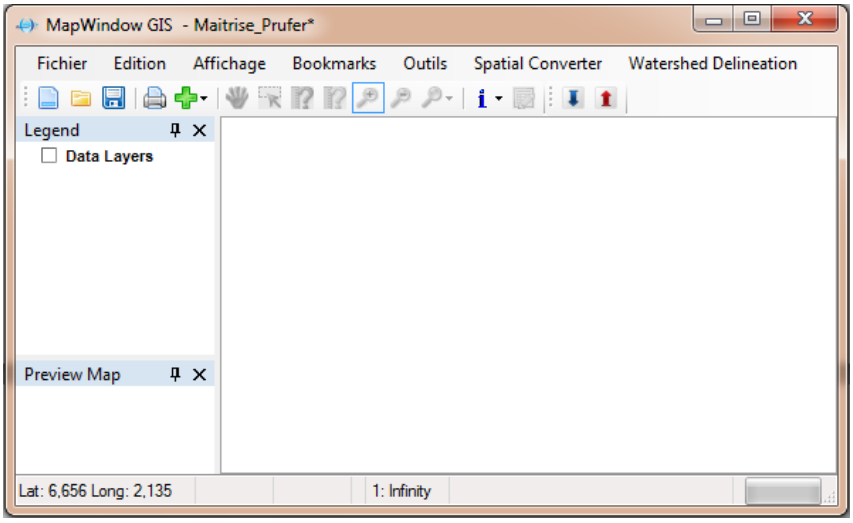


Figure 3.21 : Interface de MapWindow GIS

3.4.2.2 Système de projection utilisé

Lors du montage d'un système d'information géographique il est important de définir et d'utiliser le même système de projection. En effet, les valeurs des coordonnées doivent être projetées dans un système de projection afin de représenter la surface sphérique du globe terrestre en deux dimensions. Cette projection produit toujours une distorsion concernant la surface, l'échelle, la forme ou la direction. En fonction des applications, de très nombreux systèmes de projection ont ainsi été créés. Le système de projection que nous utiliserons est le « North American Datum 1983 » (NAD83), qui est utilisé par Statistique Canada. Cette projection est particulièrement bien adaptée pour les latitudes moyennes et pour des territoires orientés est-ouest comme le Canada.

3.4.2.3 Fichier des limites

Les premières couches que l'on va insérer dans notre projet forment ce que l'on appelle généralement le « fond de carte ». Ce sont les données qui concernent les caractéristiques géographiques du Québec et que l'on trouve dans les données publiées par Statistique Canada sous le nom de « fichier des limites ». Ces fichiers contiennent les données géographiques sur les limites des provinces et territoires, sur l'hydrographie (côtes, lacs et rivières) ainsi que sur les 12 subdivisions du territoire du Canada (recensement, région, circonspection électorale,…). On trouve dans le guide de référence du fichier des limites (Statistique Canada, 2007b) une description précise de ces données. Leur date de référence est le 1er janvier 2006. L'annexe 7

définie et présente les différentes subdivisions qui ont utilisées dans l'étude (Région économique, Aire de diffusion,…).

En combinant ces couches géographiques et en jouant sur les couleurs et la transparence on peut ainsi générer un fond de carte générique du Québec avec des degrés de précision adaptables à l'échelle. En voici un exemple :

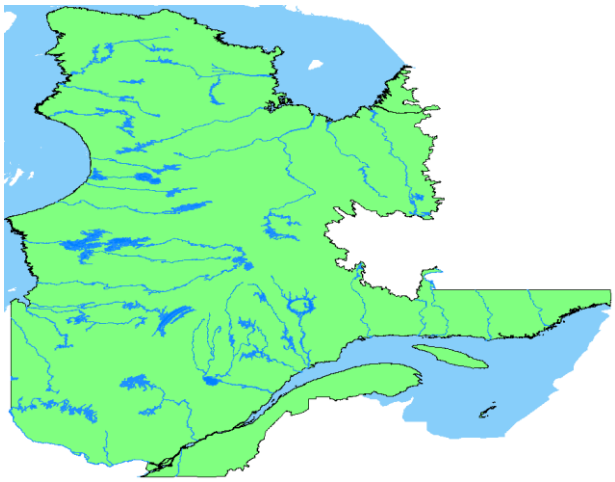


Figure 3.22 : Fond de carte générique du Québec

3.4.2.4 Fichier du réseau routier

Pour étudier un phénomène se manifestant sur les routes il est utile d'ajouter une couche concernant le réseau routier. Le fichier disponible est assez lourd avec plus de 400 000 enregistrements et de nombreux attributs (Statistique Canada, 2007a). Pour limiter sa taille et les délais d'affichage, les attributs inutiles pour notre étude ont été supprimés, et les routes majeures ont été extraites pour en faire une couche à part entière beaucoup moins lourde. Pour une échelle grande on peut ainsi se contenter de n'afficher que les route principales.

3.4.2.5 Ajout des données de recensement

Enfin les dernières informations à ajouter concernent la répartition de la population. Les résultats des recensements de 2006, 2001 et 1996 sont accessibles depuis sur le site CREPUQ (Bibliothèques universitaires québécoises, 2007). Ce sont les données de 2006 qui ont été utilisées. Les aires de diffusion sont la plus petite unité géographique pour laquelle toutes les données de recensement sont diffusées, c'est donc cette subdivision qui a été utilisée. A l'aide du visualisateur Beyond 20/20 il faut alors sélectionner, parmi l'ensemble des données collectées

lors du recensement, les champs concernant les chiffre de la population, puis importer ces données par aire de diffusion selon leur numéro d'identification. On a ainsi ajouté des attributs portant sur la répartition des populations sur la couche des Aires de diffusion. Voici par exemple le type de carte que l'on peut générer :

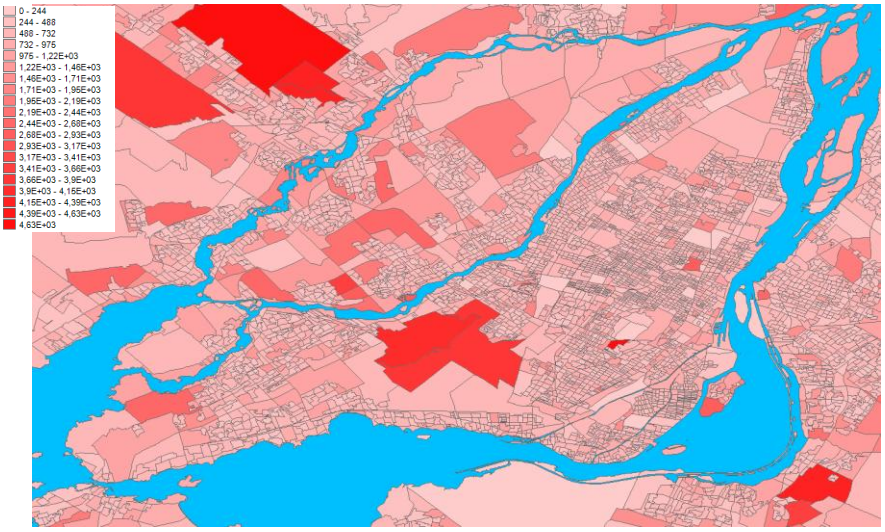


Figure 3.23 : Population en fonction des aires de diffusion (Ile de Montréal)

Pour comparer la répartition de la population et celle des d'accidents, il a fallu créer une couche de points avec des informations sur la population. Pour cela le centre de chaque aire de diffusion a été calculé et la population de l'AD a été appliquée à ce point.

3.4.3 Analyse spatiale comparative

Nous avons maintenant mis tous les éléments en place pour commencer l'analyse spatiale des répartitions d'accidents obtenues. Nous allons dans cette analyse principalement utiliser les outils du logiciel d'analyse spatiale CrimeStat. Commençons donc par présenter ce logiciel.

3.4.3.1 Présentation de CrimeStat

CrimeStat (Ned Levine & Associates, 2010e) est un logiciel américain, disponible gratuitement, qui permet de réaliser des analyses spatiales. Ce logiciel a été développé il y a une dizaine d'années pour aider les départements de police américains à analyser les répartitions d'infractions et les aider ainsi à agir de façon plus efficace. Il peut tout à fait être utilisé pour étudier les répartitions d'accidents. Son interface n'est pas des plus intuitives mais il a l'avantage d'être léger et s'il est bien paramétré il est très efficace, même avec des fichiers importants.

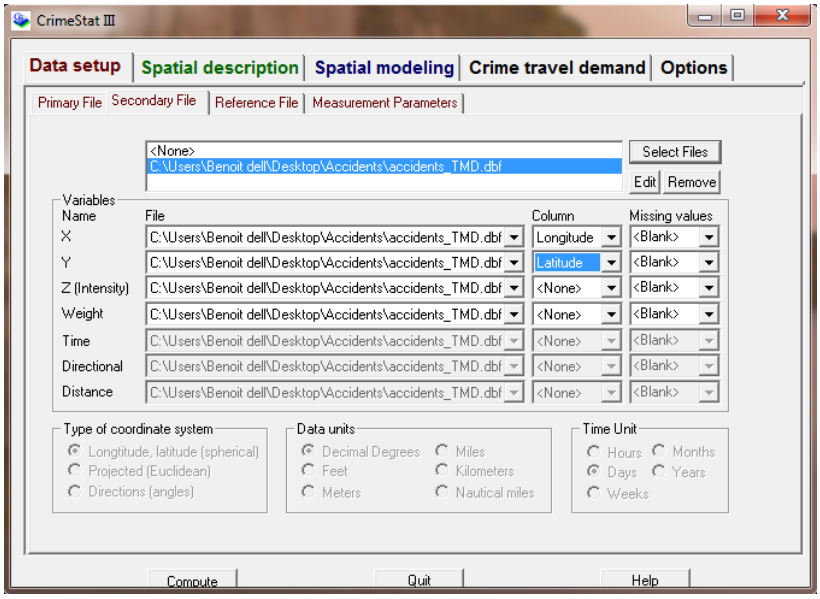


Figure 3.24 : Paramétrage des données dans CrimeStat III

Le logiciel est capable de réaliser de nombreuses études, qui sont regroupées par thème dans les onglets en haut de la fenêtre. Voici la description des outils qui ont été utilisés dans cette étude.

3.4.3.2 Analyse centrographique

Une analyse centrographique permet de mesurer la position centrale d'une distribution. CrimeStat calcule tout un ensemble de paramètres dont voici la description (Ned Levine & Associates, 2010a) :

Tableau 3.4 : Paramètres d'analyse centrographique (Ned Levine & Associates, 2010a)

	Traduction	Type de couche	Description
Mean Center	Point Moyen	1 point	Moyenne (pondérée ou non) respective des coordonnées X et Y
Median center	Point Médian	1 point	Médiane (pondérée ou non) respective des coordonnées X et Y
Center of Minimum distance	Centre des distances Minimum	1 point	Point pour lequel la somme des distances aux autres points est minimale. CrimeStat utilise un algorithme itératif pour obtenir les coordonnées de ce point. La distance utilisée peut-être directe, indirect ou suivre un réseau routier.
Standard deviation of X and Y coordinates	Distance type	1 carré	Écart-type respectif des coordonnées X et Y

Standard distance deviation	Distance-type elliptique	1 ellipse	Écart-type de la distance de chaque point au centre moyen. C'est l'équivalent en 2D de l'écart-type.
Standard deviational ellipse	Ellipse de l'écart-type	1 ellipse	Ellipse construite à partir des Écart-type respectif des coordonnées X et Y
Geometric Mean	Moyenne géométrique	1 point	Moyenne minimisant l'effet des valeurs extrêmes par l'emploi de logarithme
Harmonic Mean	Moyenne harmonique	1 point	Inverse de la moyenne de l'inverse respective des coordonnées X et Y S'il est différent du Centre moyen, cela signifie que la distribution est asymétrique
Convex Hull	Enveloppe convexe	1 polygone	Frontière tracée autour de la répartition de point

Nous allons dans notre étude comparer les valeurs ces paramètres entre nos deux distributions ainsi que vis-à-vis de la répartition de population. Pour les appliquer à la population, il a fallu créer une nouvelle couche de points correspondants au centre des Aires de diffusion et ayant dans leurs attributs le nombre de personnes.

3.4.3.3 Analyse du plus proche voisin

Une analyse du plus proche voisin (Nearest Neighbor Analysis, NNA) permet de quantifier les écarts entre les points des distributions. Voici la description des paramètres calculés pour CrimeStat dans le cadre de cette analyse :

Tableau 3.5 : Description des paramètres calculés par CrimeStat pour la NNA (Ned Levine & Associates, 2010b)

Nom	Notation	Description
Nearest Neighbor distance	$d(NN)$	Moyenne de la distance entre un point et son plus proche voisin
Mean Random Distance	$d(ran)$	$d(NN)$ que l'on obtiendrait si la distribution était complétement aléatoire
Nearest Neighbor index	NNI	$NNI = \frac{d(NN)}{d(ran)}$ Si $NNI \approx 1$ la répartition est aléatoire Si $NNI < 1$ les points sont plus proches qu'attendu Si $NNI > 1$ les points sont plus dispersés qu'attendu
Test Statistic (Z)		CrimeStat réalise un Z test afin de déterminer si la différence par rapport au $d(NN)$ est significative

Pour comparer deux répartitions on calcule le rapport $\frac{NNI(1)}{NNI(2)}$ afin de déterminer quelle est la distribution la plus dispersée. Par ailleurs CrimeStat permet aussi d'analyser la distance au K^{ième} plus proche voisin et trace la courbe du NNI obtenu en fonction de K. Nous pourrons ainsi tracer la superposition de ces deux courbes pour nos deux distributions afin de les comparer.

3.4.3.4 Analyse de type « Hot Spot »

On appelle « Hot Spots », ou « points chauds » les zones où il y a une concentration d'accidents. Il existe de nombreuses méthodes permettant de déterminer ces zones, CrimeStat en propose sept qui sont décrites dans la documentation du logiciel (Ned Levine & Associates, 2010c). Parmi ces méthodes certaines ne sont pas ou peu pertinentes pour notre étude et nous avons utilisé que les deux méthodes décrites ci-dessous. Dans les deux cas le résultat est un groupement d'accidents que CrimeStat propose de représenter soit avec une ellipse soit par leur enveloppe convexe. L'une ou l'autre de ces présentations sera utilisée en fonction des cas.

- Nearest Neighbor Hierarchical Clustering (NnH)

Cette approche sur plusieurs niveaux (hiérarchique) consiste à regrouper les accidents qui sont spatialement proches. Pour être conservé au premier niveau, un groupe doit contenir un nombre minimum de points préalablement défini par l'utilisateur, et ces points doivent être suffisamment proches les uns des autres (CrimeStat calcule une distance seuil). Les clusters de premier niveau sont ensuite utilisés pour produire des clusters de second niveau et ainsi de suite jusqu'à ce qu'il n'y ait plus de possibilité.

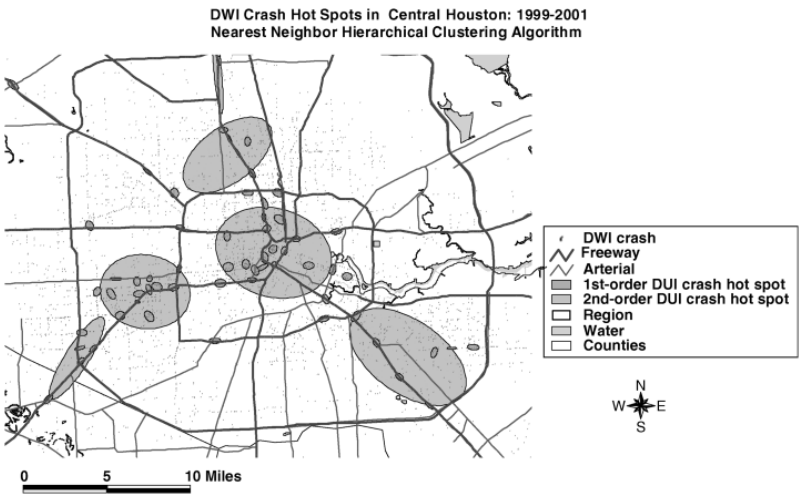


Figure 3.25 : Exemple Hot Spot obtenu par la méthode NnH (N. Levine, 2006)

Pour la base TMD (≈2500 points) le calcul n'a posé aucun problème, par contre il a été impossible de le calculer avec les accidents de la base globale, CrimeStat arrivant systématiquement à saturation de la mémoire (« Out of memory »). Pour pallier à ce problème un échantillon aléatoire de 35 000 accidents issus de la base globale a été généré, ce qui a permis à CrimeStat de s'en sortir (avant d'arriver à ce chiffre 150 000 et 60 000 accidents ont été testé sans succès…). Afin d'avoir des clusters de tailles comparables malgré la différence d'échelle entre le nombre d'accidents des deux bases les nombres minimums d'accidents pour former un cluster ont été choisi de façon à ce qu'ils soient proportionnels.

- K-Means Partitioning Clustering

Dans cette approche, l'utilisateur choisi le nombre de cluster à générer. Le programme cherche alors la meilleure position des K clusters puis assigne chaque point à un cluster. Il n'y a ici qu'un seul niveau hiérarchique. Contrairement à l'approche précédente (NnH) nous n'avons pas eu de problème pour appliquer ce calcul à la base globale, ce sont bien les 250 000 accidents qui ont été pris en compte dans le calcul.

3.4.3.5 Estimation de la densité par la méthode du noyau

Une estimation de la densité par la méthode du noyau, en anglais, Kernel Density Interpolation, est une méthode qui permet d'estimer une variable d'intensité sur une région en appliquant une fonction noyau (kernel) symétrique sur chaque point. Il faut au préalable définir une grille de découpage de la zone géographique d'étude. CrimeStat calcule alors pour chaque élément de la grille une variable d'intensité, résultant de la somme des contributions de chacune des fonctions kernel symétriques (Ned Levine & Associates, 2010d).

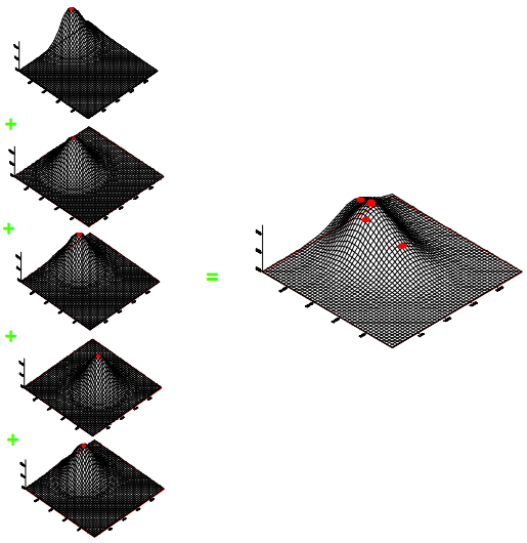


Figure 3.26 : Principe de sommation des fonctions kernel (Ned Levine & Associates, 2010d)

Pour appliquer ce principe il y a de nombreux paramètres à choisir. En premier lieu il faut sélectionner la région d'étude et la grille de découpage associée. Il a été choisi d'étudier uniquement le sud du Québec et de « couper » le nord du

Québec où il y a très peu d'accidents. Cela permet de diminuer le nombre de calculs à faire et ainsi de choisir une grille plus fine. Voici la grille qui a été utilisée :

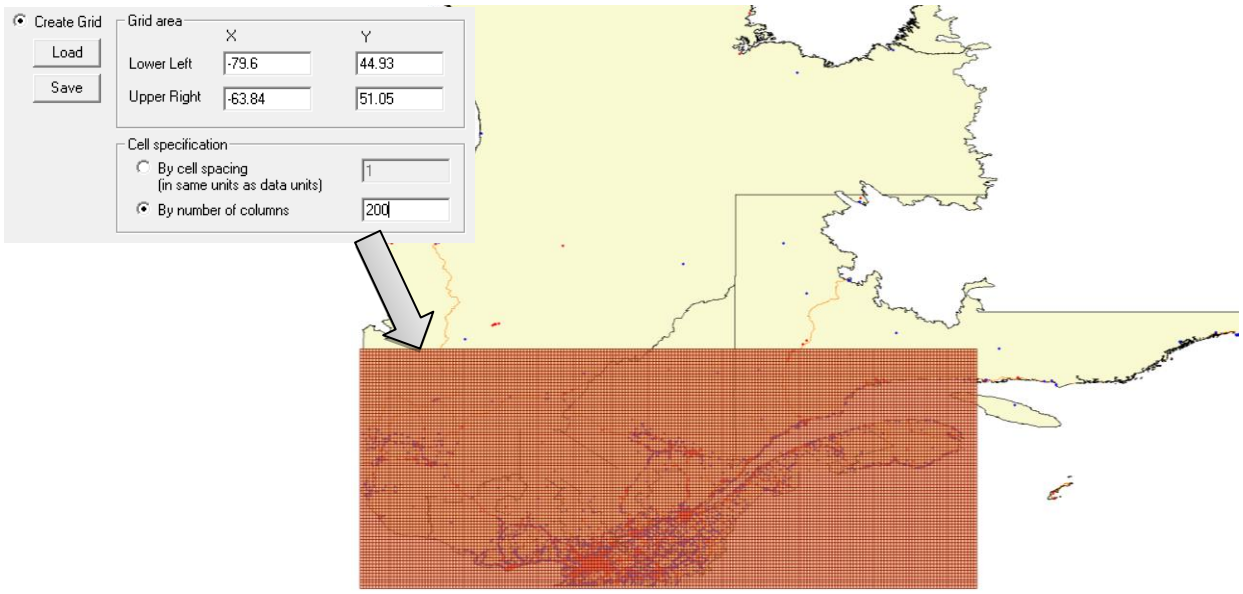


Figure 3.27 : Grille utilisée pour l'estimation des densités et ses paramètres dans CrimeStat

Voici les valeurs des autres paramètres qui ont été choisis :

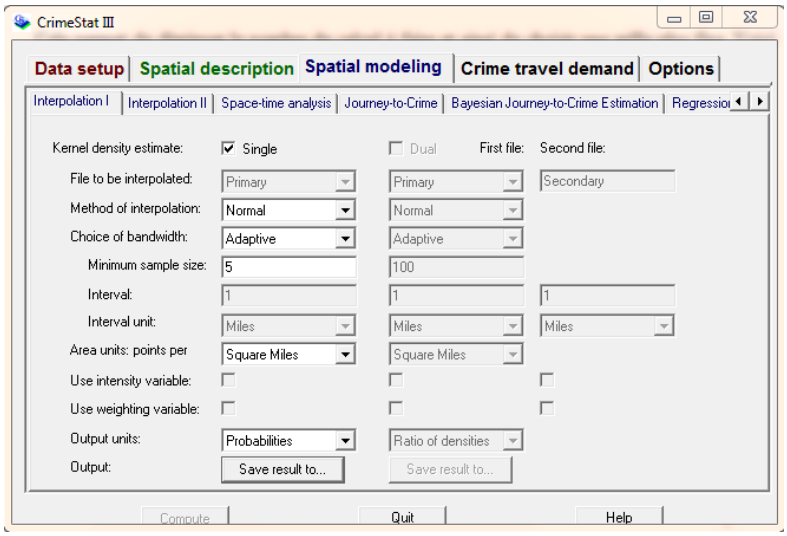


Figure 3.28 : Paramétrage de CrimeStat pour l'estimation de la densité

L'unité de sortie a été choisie en termes de probabilité car alors la densité obtenue est divisée par le nombre d'échantillons, ce qui annule l'effet de taille et fournit des valeurs comparables pour nos deux répartitions. A noter que des variables pondérées peuvent être utilisées, ce que l'on fera pour identifier les endroits les plus critiques en termes de décès par exemple.

CHAPITRE 4 ANALYSE COMPARATIVE DES DONNÉES

Les deux bases de données ayant été présentées et préparées, nous pouvons maintenant procéder à l'analyse descriptive comparative de la base de données globale et de la base TMD. Pour cela nous allons parcourir les champs communs entre ces bases en essayant de faire apparaitre d'éventuelles différences ou similitudes et de les expliquer. Le tableau suivant présente les 18 champs communs qui ont été identifiés :

Tableau 4.1 : Les 18 champs communs

Champ commun	Format des données	Taux de remplissage		Étudié?	Commentaires
		Base Globale	Base TMD		
Date	Format Date	100%	100%	✅	Nécessite d'être transformé en données quantitatives
Heure	Quantitatif discret	97,72%	97,90%	✅	
Minute	Quantitatif discret	97,72%	97,90%	❌	Données « arrondies », interprétation difficile
Nombre de véhicules impliqués	Quantitatif discret	100%	100%	✅	
Limitation de vitesse	Quantitatif discret	39,87%	55,65%	☑	Peu de données, difficile de compléter car les limitations de vitesse peuvent varier (travaux,…)
Code d'impact	Qualitatif ordinal	97,08%	95,96%	✅	
Décès	Quantitatif discret	100%	100%	✅	
Blessés	Quantitatif discret	100%	100%	✅	
Genre d'accident	Qualitatif nominal	98,70%	97,06%	✅	
Environnement	Qualitatif nominal	28,23%	31,78%	☑	Peu de données
État de la surface	Qualitatif nominal	97,20%	97,35%	✅	
Catégorie de route	Qualitatif nominal	27,32%	32,37%	❌	Interprétation difficile, peu de données
Conditions météo	Qualitatif nominal	99,25%	97,31%	✅	
Éclairement	Qualitatif nominal	99,26%	97,23%	✅	
Localisation	Qualitatif nominal	15,37%	35,86%	❌	Peu de données, interprétation difficile
Aspect de la chaussée	Qualitatif nominal	15,12%	29,34%	☑	Peu de données
Nature de la chaussée	Qualitatif nominal	15,14%	27,87%	☑	Peu de données
État de la chaussée	Qualitatif nominal	15,12%	27,57%	❌	Peu de données

Le tableau précédent montre que la principale difficulté de cette analyse est le taux de remplissage particulièrement faible de certains champs. Dans ce contexte, l'étude de 3 champs n'est pas présentée car les résultats sont difficilement interprétables. L'analyse des 4 autres champs ayant un faible taux de remplissage sont présentés, mais les résultats sont à relativiser car la majorité des rapports d'accidents ne renseignent par leurs valeurs. On peut noter que les champs de la base globale sont plus fournis, ce qui prouve que plusieurs sources d'informations ont été utilisées pour sa construction. Une petite part de cette différence est aussi due au fait que les champs vides de la base TMD ont été complétés par les informations de la base globale lorsque cela été possible. Le champ minute n'a pas non plus été utilisé, car bien qu'il soit très souvent renseigné les données sont trop approximatives (cf. §4.2.1). Les 10 champs restants n'ont pas posé de problème particulier.

Dans un premier temps le champ « date » sera analysé selon plusieurs distributions temporelles (§ 4.1), puis les champs quantitatifs (§4.2) et les qualitatifs (§4.3) seront étudiés.

4.1 Étude du champ « Date »

Cette section présente les analyses réalisées à partir des informations contenues dans le champ « date » des deux bases de données. Ce champ, toujours renseigné, permet de multiples analyses sur les répartitions dans le temps. Il a été étudié selon cinq répartitions temporelles, qui sont dans l'ordre : les années, les mois, les saisons, les jours de l'année et les jours de la semaine.

4.1.1 Distribution annuelle

Nous allons tous d'abord étudier la distribution du nombre d'accidents selon les années.

- Tableau d'effectifs

Tableau 4.2 : Distribution annuelle des observations d'accidents

Année	Base Globale	Base TMD
1995	18 439	215
1996	16 557	197
1997	18 332	176
1998	17 208	155
1999	17 342	193

2000	18 423	190
2001	16 816	154
2002	17 493	194
2003	18 067	191
2004	17 829	202
2005	18 000	174
2006	16 624	162
2007	17 909	176
2008	17 777	×
2009	15 471	×

Ce tableau montre qu'il y a environ 18 000 accidents impliquant des poids lourds par année et environ 200 accidents TMD par an.

- Évolution comparée

À partir du tableau précédent on peut tracer l'évolution annuelle comparée des rapports d'accidents. Comme l'ordre de grandeur du nombre d'accidents des deux bases est très différent (18 000 contre 200 accidents), le pourcentage du total a été utilisé :

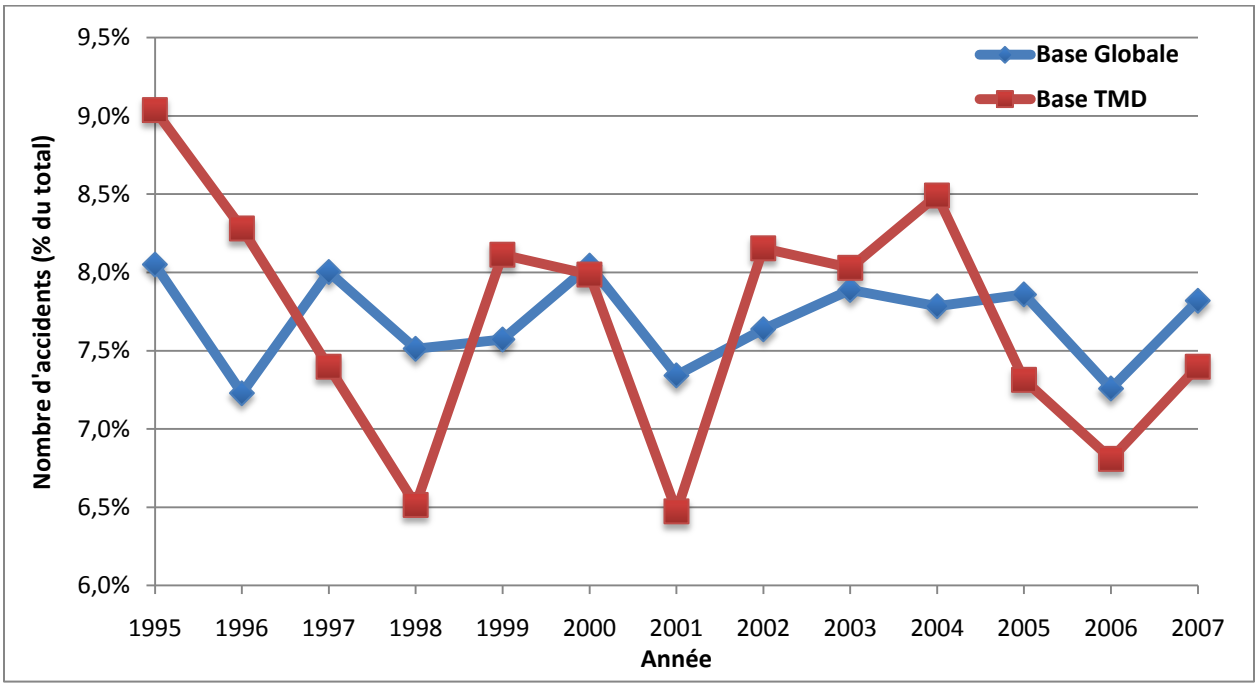


Figure 4.1 : Évolution annuelle comparée du nombre d'accidents (% du total)

Aucune tendance claire ne dégage de ce graphique, les deux courbes semblant osciller autour de 7,5% de leur total respectif d'accidents. On peut noter que le nombre d'accidents TMD a considérablement baissé de façon linéaire de 1995 à 1998 (-30% par rapport à la valeur de 1995),

mais est depuis remonté. Les variations des deux courbes ne semblent pas liées par une relation marquée. On peut cependant remarquer que sur la période allant de 1997 à 2007 il y a une ressemblance dans les variations, ce qui peut indiquer qu'une année où il y beaucoup d'accidents de véhicules lourds est une année où il y aussi beaucoup d'accidents TMD.

- Statistique descriptive de la distribution annuelle

Afin de nous éclairer un peu plus sur les caractéristiques de ces deux distributions annuelles nous pouvons calculer quelques paramètres de statistique descriptive :

Tableau 4.3 : Statistique descriptive des distributions annuelles

Statistiques descriptives		Base Globale (1995-2007)	Base TMD
Paramètres de position	Minimum	16 557	154
	Maximum	18 439	215
	Moyenne	17 618,38	183,00
	Erreur Type	184,76	5,16
	Médiane	17 829	190
Paramètres de variabilité	Étendue	1 882	61
	Écart moyen	564,66	15,54
	Variance	443 787,4	346,66
	Écart-type	666,17	18,62
	Coefficient de variation	3,78%	10,17%

Le nombre moyen d'accidents TMD sur la période 1995 – 2007 est de 183 accidents par an avec un écart-type d'environ 18 accidents. Pour la base globale, sur la même période le nombre moyen d'accidents est de 17 618 par an, avec un écart-type de 666 accidents. On constate un écart conséquent entre les deux coefficients de variation. Celui de la base globale est 2,7 fois plus petit, ce qui signifie que d'une année sur l'autre les variations des accidents TMD sont plus importantes en proportion équivalente, que celle des accidents de véhicule lourd.

- Nuage de points et tentative d'ajustement des distributions annuelles

Les variations des deux courbes précédentes ne semblent pas liées, les maxima et les minima ne correspondant pas toujours. On a cependant constaté une ressemblance de l'allure sur la période 1997-2007, ce qui peut laisser penser à un lien affine entre les distributions. Voici le nuage de points de la répartition des accidents :

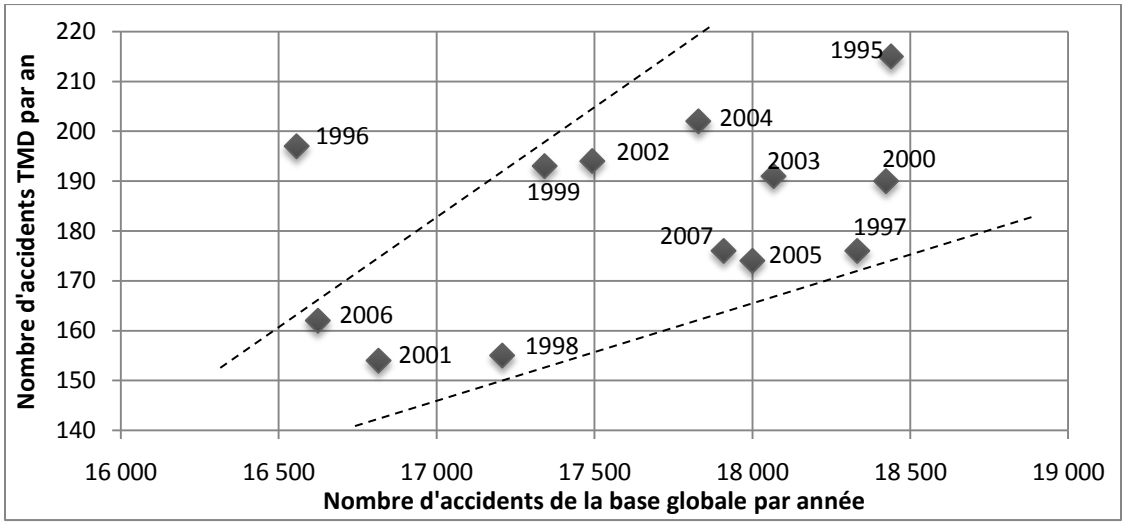


Figure 4.2 : Nuage de points des effectifs annuel d'accident

Le résultat n'est pas très probant et aucun ajustement ne se dégage. On peut cependant remarquer que si l'on excepte le point qui correspond à l'année 1996, il semble que plus le nombre d'accident global est grand, plus le nombre d'accidents TMD l'est aussi.

- Évolution du ratio nombre d'accidents globaux/Nombre d'accidents TMD

Pour terminer l'étude de ce champ voici tracé l'évolution du nombre d'accidents TMD pour 1000 accidents de véhicules lourds :

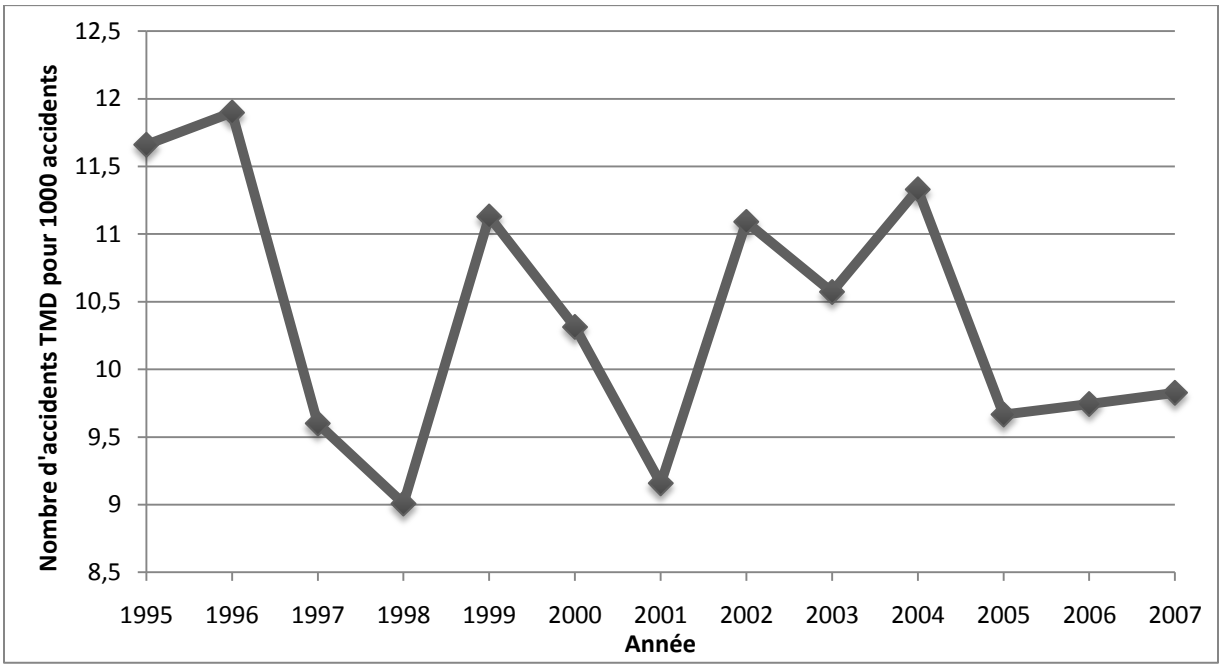


Figure 4.3 : Évolution du nombre d'accidents TMD pour 1000 accidents de véhicule lourd

Il y a en moyenne environ 10 accidents TMD pour 1000 accidents de véhicules lourds. Entre 1995 et 2005 cette valeur a oscillé entre 12 et 9 accidents TMD pour 1000. Elle semble se stabiliser depuis cette 2005 légèrement en dessous de 10. On constate que comme pour la courbe d'évolution annuelle aucune tendance claire ne se dégage.

4.1.2 Distribution mensuelle

Après l'analyse de la distribution annuelle des accidents nous allons étudier leur distribution selon les 12 mois de l'année.

- Tableau d'effectifs

Le tableau suivant indique le nombre d'accidents selon les mois sur la période 1995-2007 ainsi que le nombre moyen d'accidents par jour par an selon le mois, afin de gommer l'influence du nombre de jour de chaque mois :

Tableau 4.4 : Distribution mensuelle des observations d'accidents

Mois	Nombre de jours	Base Globale		Base TMD	
		Nombre total d'accidents	Nb moyen d'accidents/jour/an	Nombre total d'accidents	Nb moyen d'accidents/jour/an
janvier	31	22 541	60,59	278	0,75
février	28,25	19 943	58,83	259	0,76
mars	31	18 149	48,79	255	0,69
avril	30	13 914	38,65	138	0,38
mai	31	17 904	48,13	164	0,44
juin	30	20 255	56,26	174	0,48
juillet	31	18 177	48,86	143	0,38
août	31	19 044	51,19	163	0,44
septembre	30	17 882	49,67	157	0,44
octobre	31	19 544	52,54	173	0,47
novembre	30	20 240	56,22	186	0,52
décembre	31	21 446	57,65	289	0,78
Moyenne		**19 086,58**	**52,28**	**198,25**	**0,54**

On obtient une moyenne de 52 accidents de véhicule lourds par jour/an pour la base globale et une moyenne d'un accident TMD tous les deux jours. En comparant les valeurs minimales et maximales selon les mois, on constate qu'il y des variations notables du nombre d'accidents notamment pour les accidents TMD. Un diagramme nous permettra de mieux visualiser ces différences.

- Diagramme en radar

Un diagramme « en radar » a été choisi parce qu'il permet de représenter le cycle annuel des mois. Sur ce diagramme, en % du total, le nombre d'accidents par jour par année en fonction du mois, a été tracé afin de comparer des mois « équivalents » en nombre de jour. En pratique la différence est notable uniquement pour le mois de février.

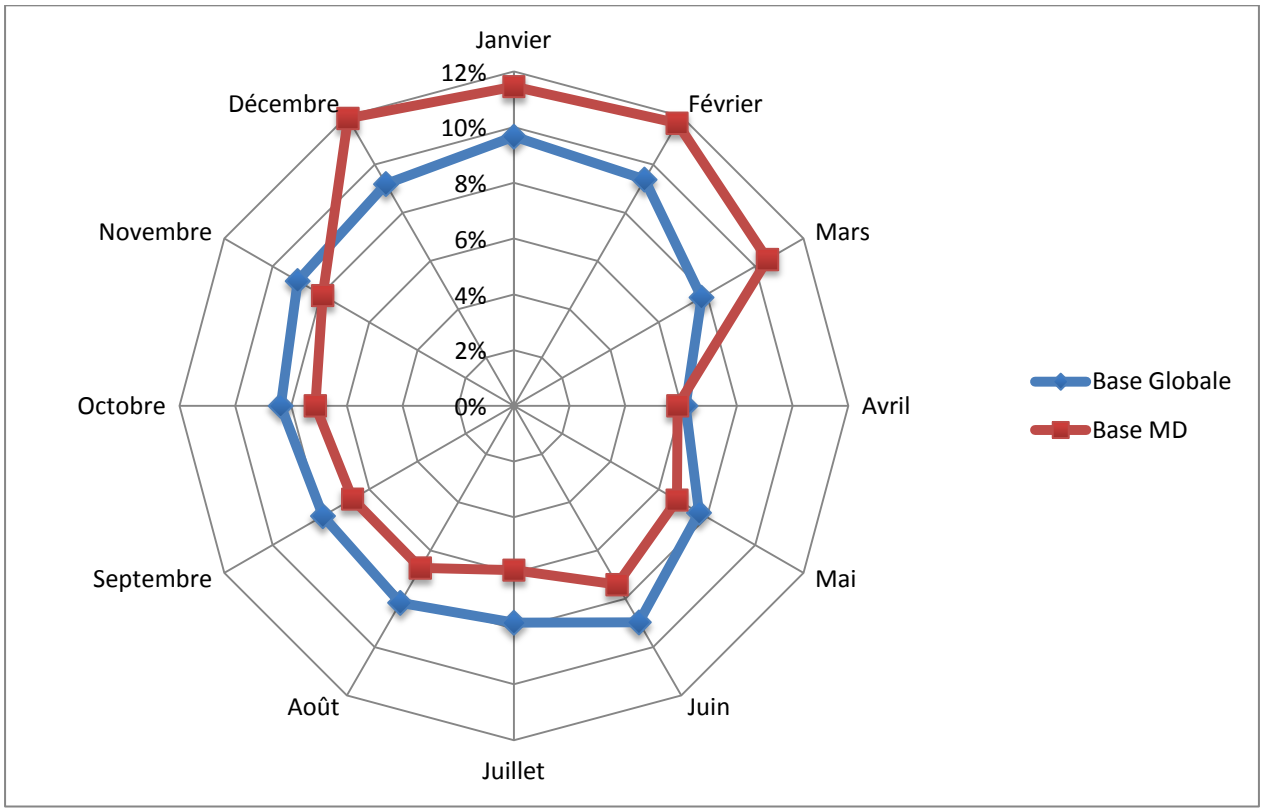


Figure 4.4 : Distribution mensuelle comparée du nombre d'accidents par jour par an selon les mois (en % du total)

On remarque immédiatement que les mois d'hiver présentent des taux d'accident important, notamment pour les accidents TMD, avec des valeurs autour des 12% en décembre, janvier, février et mars (10% pour la base globale). En avril la proportion d'accidents chute pour les deux bases de données jusqu'à 6% avant de remonter en mai et en juin. Le reste de l'année, le pourcentage demeure stable autour de 6% pour la base MD et 8% pour la base globale. Les variations mensuelles sont au cours de l'année plus importantes pour les accidents TMD (12% - 6%), que pour les accidents de la base globale (9% - 6%).

4.1.3 Distribution saisonnière

D'après la distribution mensuelle des accidents, les mois d'hiver semblent avoir une influence significative sur la répartition des accidents, il serait intéressant de voir si ce phénomène se retrouve à l'échelle des saisons. Pour définir les saisons une définition astronomique basée sur les dates des solstices et des équinoxes a été utilisée.

- Tableau d'effectifs

Tableau 4.5 : Distribution saisonnière des accidents dans les deux bases de données

Saison	Période retenue	Nombre de jours	Nombre d'accidents Base Globale (95-07)	Nombre d'accidents TMD
Hiver	22 déc. - 20 mars	89,25	59 373	807
Printemps	21 mars - 20 juin	92	50 944	491
Été	21 juin - 22 sept	94	56 847	469
Automne	23 sept - 21 déc.	90	61 875	612
Somme		365,25	229 039	2 379
Moyenne		91,3125	57 259,75	594,75

- Diagramme à barre verticale

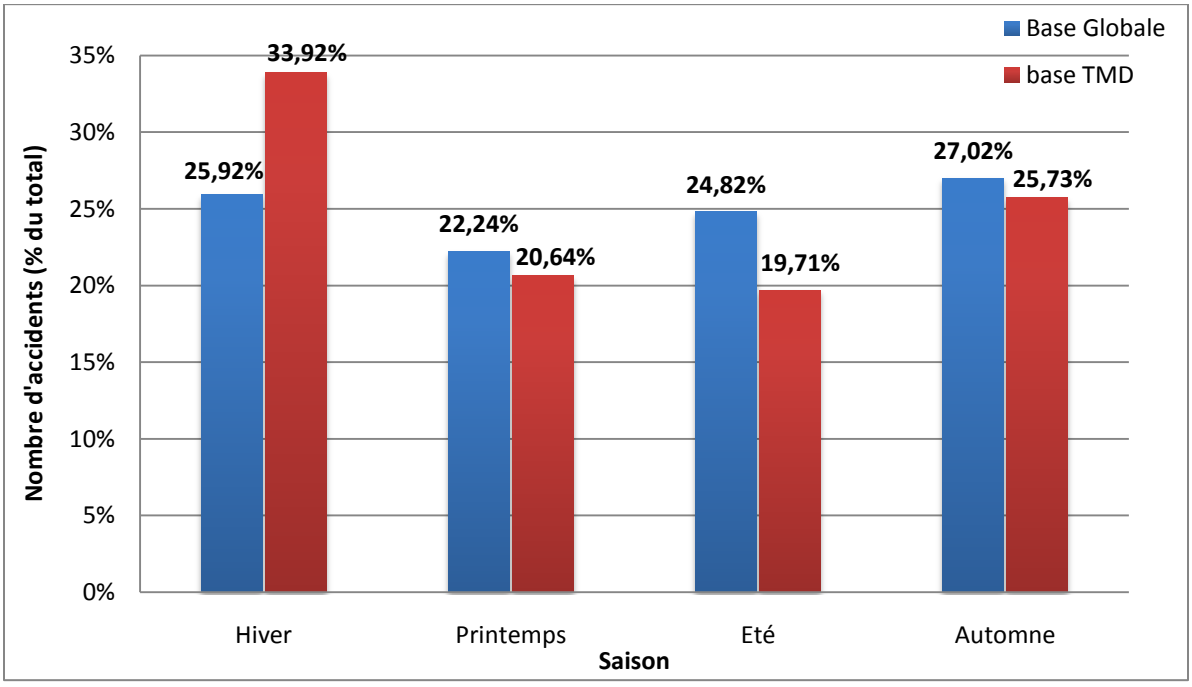


Figure 4.5 : Distribution saisonnière des accidents (% du total)

Ce diagramme met bien en évidence l'influence importante de l'hiver sur le nombre d'accident TMD avec près de 34% des accidents durant cette saison. Les accidents impliquant des véhicules lourds sont par contre moins touchés par ce phénomène avec 26% des accidents sur cette saison. Leur pire saison n'est d'ailleurs pas l'hiver mais l'automne avec 27% des accidents. Le printemps est pour les deux bases de données une saison relativement « calme ».

La forte proportion d'accidents TMD en hiver est une caractéristique spécifique au TMD. On peut supposer que le transport de matières dangereuses est plus fréquent en hiver avec par exemple de nombreuses livraisons de combustible pour le chauffage, ce qui augmente la probabilité d'un accident. Les mauvaises conditions climatiques jouent très probablement aussi.

4.1.4 Distribution selon les jours de l'année

Nous allons maintenant analyser la répartition des accidents selon les jours de l'année. Le tableau des fréquences d'apparition pour la période 1995-2007 est beaucoup trop long pour figurer ici, puisqu'il comporte 366 lignes pour les 366 jours possibles dans une année. Pour résumer ces données nous utiliserons les outils de la statistique descriptive, puis nous les visualiserons à l'aide d'un histogramme.

- Statistique descriptive de la distribution selon les jours

Tableau 4.6 : Statistique descriptive des distributions d'accidents selon les 366 jours de l'année (∑ des accidents entre 1995 et 2007)

Statistiques descriptives		Base Globale	Base TMD
Paramètres de position	Minimum	96	0
	Maximum	1 076	23
	Moyenne	625,79	6,50
	Médiane	629	6
	P_{25}	554	4
	P_{75}	698	8
Paramètres de variabilité	Étendue	980	23
	Intervalle interquartiles	143,75	4
	Écart moyen	97,07	2,57
	Variance	17 134,6	11,20
	Écart-type	130,9	3,35
	Coefficient de variation	20,92%	51,56%
Paramètres de forme	Coefficient d'asymétrie	-0,3469	0,9318
	Coefficient d'aplatissement	2,0492	1,8498

On constate qu'il y a pour les deux bases de données un écart important entre l'étendue et l'intervalle interquartiles (de l'ordre de 60%), ce qui signifie qu'il y a des valeurs extrêmes importantes dans les deux bases de données et que 50% des valeurs des fréquences sont assez proches, l'intervalle interquartile étant faible. D'autres parts les variations selon les jours de l'année des accidents TMD sont plus de 2 fois supérieures à celles des accidents de la base globale, comme l'indique leur valeur respective du coefficient de variation.

- Comparaison des dates les plus et les moins touchées par les accidents

Avant de présenter les deux histogrammes obtenus voici une petite comparaison des valeurs extrêmes qui ont été obtenue. Tous d'abord les valeurs maximales :

Tableau 4.7 : Jours où il y a eu le plus d'accidents

Base globale		Base TMD	
∑ Nombre d'accidents	Date	∑ Nombre d'accidents	Date
1076	20 décembre	23	20 décembre
976	18 décembre	19	6 décembre
975	13 janvier	18	10 mars

Les trois valeurs extrêmes observées se trouvent toute en période hivernale, essentiellement en décembre. Pour les deux bases de données le jour le plus « noire » en termes d'accident est le 20 décembre. Les conditions hivernales ainsi que l'effervescence d'avant les fêtes de fin d'année peuvent expliquer cette date. Voici maintenant les trois jours où il y a eu le moins d'accidents :

Tableau 4.8 : Jours où il y a eu le moins d'accidents

Base globale		Base TMD	
Nombre d'accidents	Date	Nombre d'accidents	Date
96	25 décembre	0	1 janvier / 29 février / 25 avril/ 29 avril/ 6 & 8 septembre
97	29 février		
121	1er janvier		

Il est remarquable que les valeurs extrêmes se concentrent essentiellement en fin et en début d'année. En seulement 5 jours, du 20 décembre au 25 décembre on passe même d'une valeur maximale à une valeur minimum.

Afin d'avoir une vision d'ensemble des distributions des accidents selon les jours de l'année voici les deux histogrammes qui ont été obtenu.

- Histogrammes des distributions des accidents selon les jours de l'année

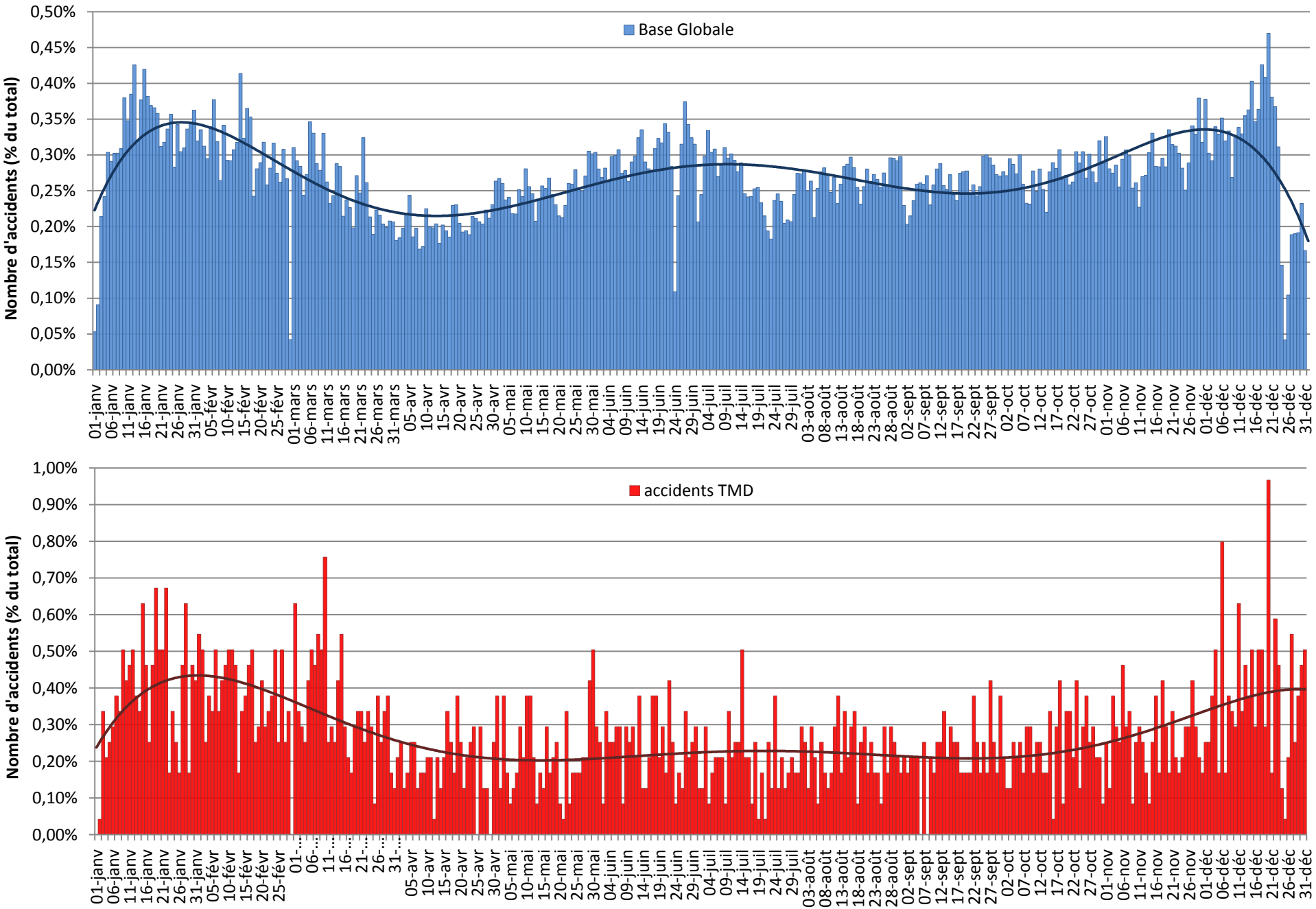


Figure 4.6 : Histogramme des observations d'accidents selon les jours de l'année (% du total)

Ces deux graphiques permettent de se faire une idée assez précise de la répartition des accidents dans l'année. Comme nous l'avons constaté précédemment on voit clairement que les mois hivernaux sont les plus favorables aux accidents notamment pour la base TMD. On remarque cependant que la période d'avant les fêtes est particulièrement touchée par les accidents et que leur nombre chute pendant les fêtes. Ensuite les taux d'accidents remontent jusqu'à la fin du mois de janvier à des valeurs comparables aux taux d'avant les fête, puis baissent tranquillement jusqu'à la fin de l'hiver. Pour la base TMD le reste de l'année le nombre d'accident reste alors équivalent jusqu'à l'hiver suivant. On remarque pour la base globale une embellie des accidents pour le mois de juin, qui ne se retrouve pas pour la base globale.

Comme les valeurs des coefficients de variation le laissaient présager, on constate que les variations des accidents TMD sont plus de 2 fois importantes que celle de la base globale. La courbe de la base globale est aussi plus régulière, avec moins de « pics » et de « creux ». On peut d'ailleurs y voir clairement deux creux, en plus de celui de noël et du jour de l'an. Le premier est le 29 février, qui n'a lieu qu'une fois tous les quatre ans, et le second le 24 juin qui est la Fête Nationale du Québec. Il y aussi un « creux » pour le 1er juillet, qui correspond à la Fête du Canada mais celui-ci est 2 fois moins important que celui de la fête du Québec… Les autres jours fériés sont moins marqués, surement parce qu'il n'a pas toujours lieu le même jour. On peut d'ailleurs noter, pour la base globale, une sorte de cycle sur semaine avec des petits pics et creux. Ce cycle est dû au fait que sur les 12 ans de données qui sont ici compilées, certain jour ont été plus souvent des jours ouvrés que d'autre. On peut en effet penser que les accidents ont plus lieu en semaine. Pour le prouver nous allons d'ailleurs maintenant nous intéresser à la distribution des accidents selon les jours de la semaine.

4.1.5 Distribution selon les jours de la semaine

Pour conclure l'étude du champ, nous allons étudier la distribution des accidents suivant les jours de la semaine. Voici le tableau d'effectif, suivi du diagramme qui a été obtenue :

- Tableau d'effectifs

Tableau 4.9 : Distribution journalière des observations d'accidents (1995-2007)

Jour	Base Globale	Base TMD
lundi	37 331	384
mardi	40 880	426
mercredi	40 795	434
jeudi	42 554	434
vendredi	42 940	422
samedi	15 261	187
dimanche	9 278	92

- Diagramme linéaire double

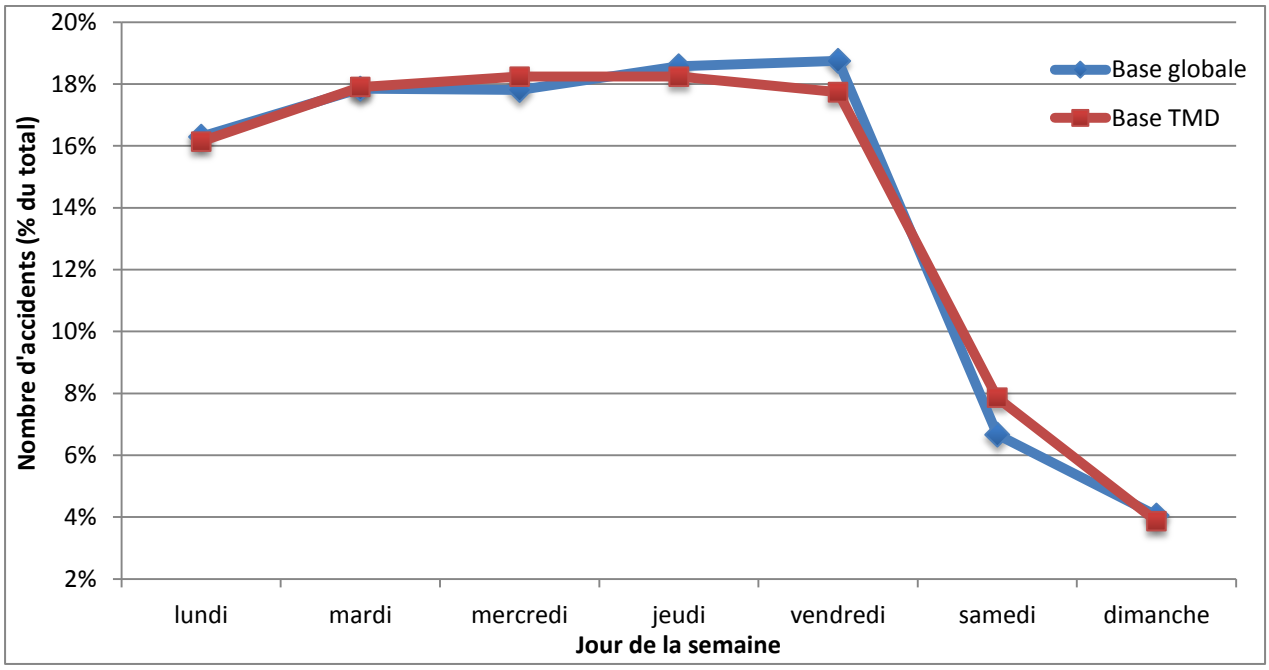


Figure 4.7 : Distribution des accidents selon les jours de la semaine (1995-2007) (% du total)

Les jours ouvrés sont plus beaucoup plus touchés que les jours de fin de semaine. Le lundi est le jour ouvré le moins touché par les accidents, les autres jours le sont sensiblement de la même façon (autour de 18% pour les 2 bases). On constate un petit « pic » pour la base globale le vendredi, alors que le % d'accident baisse légèrement pour la base TMD. Le dimanche est le jour où il y a le moins d'accidents de transport avec 4% des accidents devant le samedi.

Les deux répartitions sont très similaires, on peut donc affirmer que la présence de matières dangereuses n'influe pas, ou très peu, sur la distribution des accidents selon les jours de la semaine.

4.2 Études des champs quantitatifs

Après l'étude du champ date, nous allons maintenant étudier les champs quantitatif commun entre les deux bases, à savoir, les champs « heure », « nombre de véhicules impliqués », « limitation de vitesse » et les champs concernant les impacts sur la santé humaine.

4.2.1 Champ « Heure »

Cette section présente l'analyse réalisée à partir des informations contenues dans les champs « Heure ». Initialement il a été essayé de traiter les informations des champs « Heures » et « Minutes » ensemble, mais les résultats n'ont pas été très pertinents. En effet on obtenait pour les deux bases de données des pics importants pour les heures « rondes », puis des pics de moindre importance pour les demi-heures et les $15^{ème}$ et $45^{ème}$ minutes, les heures « intermédiaires » étant très peu fréquentes. Cette constatation peut facilement s'expliquer par le fait que les rédacteurs des rapports d'accidents, ne connaissant pas toujours l'heure exacte d'un événement, et arrondissent l'heure qu'ils saisissent dans le rapport. En conséquence seul l'étude du champ « heure » est ici présentée.

- Tableau d'effectifs

Tableau 4.10 : Distribution selon les heures des observations d'accidents (1995-2007)

Heure	Base Globale	Base TMD
0:00	2 542	43
1:00	2 017	36
2:00	1 860	25
3:00	2 039	31
4:00	2 026	41
5:00	2 580	40
6:00	5 245	54
7:00	10 338	105
8:00	15 177	149
9:00	15 782	177
10:00	17 413	211
11:00	17 972	172
12:00	15 303	153
13:00	17 014	177
14:00	17 742	179
15:00	17 262	180
16:00	16 936	123

17:00	13 338	106
18:00	8 339	75
19:00	6 021	65
20:00	4 835	50
21:00	4 531	49
22:00	3 709	50
23:00	3 026	37
Moyenne	**9 294**	**97**

- Diagrammes à barres verticales

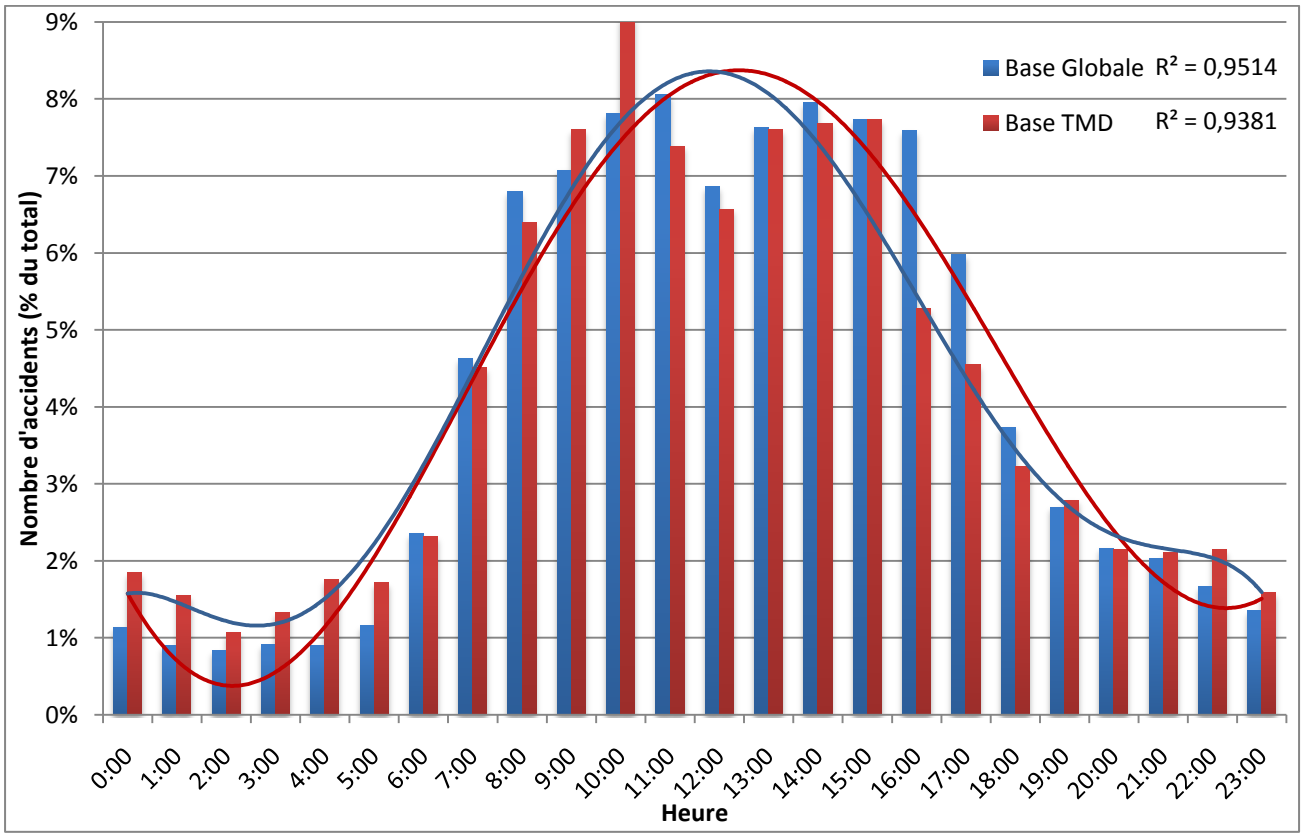


Figure 4.8 : Distribution horaire des accidents sur la période 1995-2007 (% du total)

On constate que les accidents ont essentiellement lieu pendant la journée, avec 75% des accidents entre 7h et 17h. Pour les deux bases de données on constate un creux à 12h qui correspoind certainement à une pause pour diner. Cette baisse des accidents à 12h semble d'ailleurs être une caractéristique québécoise, car cela ne se retrouve pas dans les bases d'autre base de données. En Pennsylvanie par exemple 12h correspond au contraire au maximum de la courbe (cf. § 2.1.1) (Shorten, et al., 2002).

Les répartitions entre les deux bases de données sont très proches, la principale différence étant la présence d'un maximum important pour les accidents TMD entre 10h et 11h comme l'ont remarqué Boies et Beaulieu et un plus faible taux d'accidents TMD à 16h et 17h. La régression polynomiale réalisée, d'ordre 6 (qui est l'ordre maximum proposé par Excel), donne des courbes très ressemblantes avec deux bons coefficients de détermination. Ces courbes ne prennent cependant pas en considération le creux du midi, où se trouve leur maximum. Les principaux écarts entre ces deux paraboles se trouvent dans les heures nocturnes où le pourcentage d'accidents TMD est toujours légérement supérieur à celui de la base globale. Ils semblent donc qu'il y est plus d'accidents TMD la nuit.

- Ajustement linéaire

Afin de prouver que les répartitions sont similaires voici une analyse de régression linéaire :

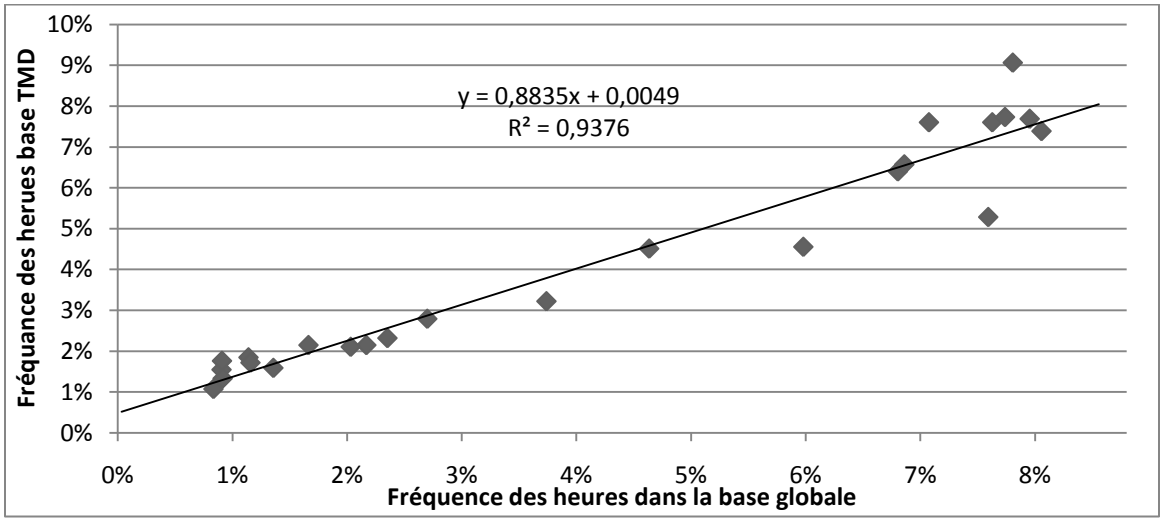


Figure 4.9 : Nuage de point des fréquences horaire des accidents

La régression linéaire permet d'obtenir une droite avec un bon coefficient de corrélation et une pente proche de 1, on peut donc considérer que la présence des matières dangereuses n'affecte globalement pas la répartition des accidents dans les heures de la journée.

4.2.2 Champ « Nombre de véhicules impliqués »

Le champ « nombre de véhicules impliqués » indique le nombre de véhicules qui ont été impliqués dans l'accident. La valeur de ce champ ne peut être nulle, un accident impliquant forcément au moins une voiture. Il bénéficie d'autre part d'un très bon taux de remplissage puisqu'il est toujours renseigné. Le tableau des fréquences étant trop long, voici une réduction numérique des données grâce aux outils de la statistique descriptive :

- Réduction numérique des deux séries de données

Tableau 4.11 : Valeurs des paramètres de statistique descriptive pour les champs « Nombre de véhicules impliqués » des deux bases de données (1995-2007)

Statistique descriptive		Base Globale	Base TMD
Paramètres des données	Nb d'échantillons total	226 759	2 379
	% vide	0,00%	0,00%
	Somme des échantillons	421 988	4 275
Paramètre de position	Minimum	1	1
	Maximum	98	16
	Mode	2	2
	Moyenne	1,8610	1,7970
	Médiane	2	2
	P_{25}	2	1
	P_{75}	2	2
Paramètre de variabilité	Étendue	97	15
	intervalle interquartiles	0	1
	Écart moyen	0,3876	0,4670
	variance	0,5054	0,5359
	Écart-type	0,7109	0,7322
	coefficient de variation	38,20%	40,75%
Paramètre de forme	Coefficient d'asymétrie de Fisher	24,97	5,50
	Coefficient d'aplatissement	2 492,58	84,34

Le nombre moyen de véhicules impliqués par accident est assez proche pour les deux bases de données. On remarque néanmoins que la moyenne de la base de matières dangereuses est légèrement plus faible de 0,06 véhicules par accident. Cependant dans les deux cas le nombre de véhicules impliqués le plus fréquemment est de 2. Toutefois $P_{25} = 1$ pour la base de matières dangereuses alors qu'il vaut 2 pour la base globale nous met sur la piste qu'il y a

vraisemblablement plus souvent des accidents impliquant un seul véhicule pour le transport des matières dangereuses. Pour le prouver il faut faire un test statistique sur la moyenne.

- Test sur la moyenne

Nous souhaitons tester l'hypothèse nulle selon laquelle les moyennes des nombre de véhicule impliqués sont égales, avec un T-test (test de Student). On constate que les deux paramètres de forme sont grand, et donc que les distributions ne sont pas normales. Avant d'appliquer le test il va donc falloir normaliser les distributions. C'est la fonction logarithme népérien qui a donné les meilleurs résultats :

Tableau 4.12 : Paramètre de forme des variables normalisés

Paramètre de forme	Base Globale	Base TMD
Coefficient d'asymétrie de Fisher	-0,4701	-0,1356
Coefficient d'aplatissement	1,0676	0,6082

Les paramètres de forme étant maintenant dans l'intervalle [-2;2] on peut considérer les distributions transformées comme normale. Par ailleurs le test de Fisher nous permet de déterminer que les variances sont différentes (=> type 3) :

$$TEST.F(serie\ BG; Série\ TMD) = 0{,}0399$$

Nous avons maintenant tous les éléments et hypothèse pour réaliser le T-test. On obtient :

$$TEST.STUDENT(serie\ normalisé\ BG; Série\ normalisé\ TMD; bilatéral; type\ 3) = 3{,}39 \times 10^{-10}$$

La probabilité obtenue est extrêmement faible, on peut donc rejeter notre hypothèse nulle. Les moyennes sont significativement différentes.

- Diagramme linéaire double

Les paramètres de forme et les valeurs des modes, moyenne et médiane nous indiquent que la courbe sera très « pointue » et dissymétrique. Les tableaux d'effectifs de chacune des distributions des nombres de véhicules impliqués ont été déterminés à l'aide de l'utilitaire d'analyse « histogramme » d'Excel. Exprimée en pourcentage du total, voici la courbe obtenue (afin de la visualiser une abscisse en échelle logarithmique a été choisie) :

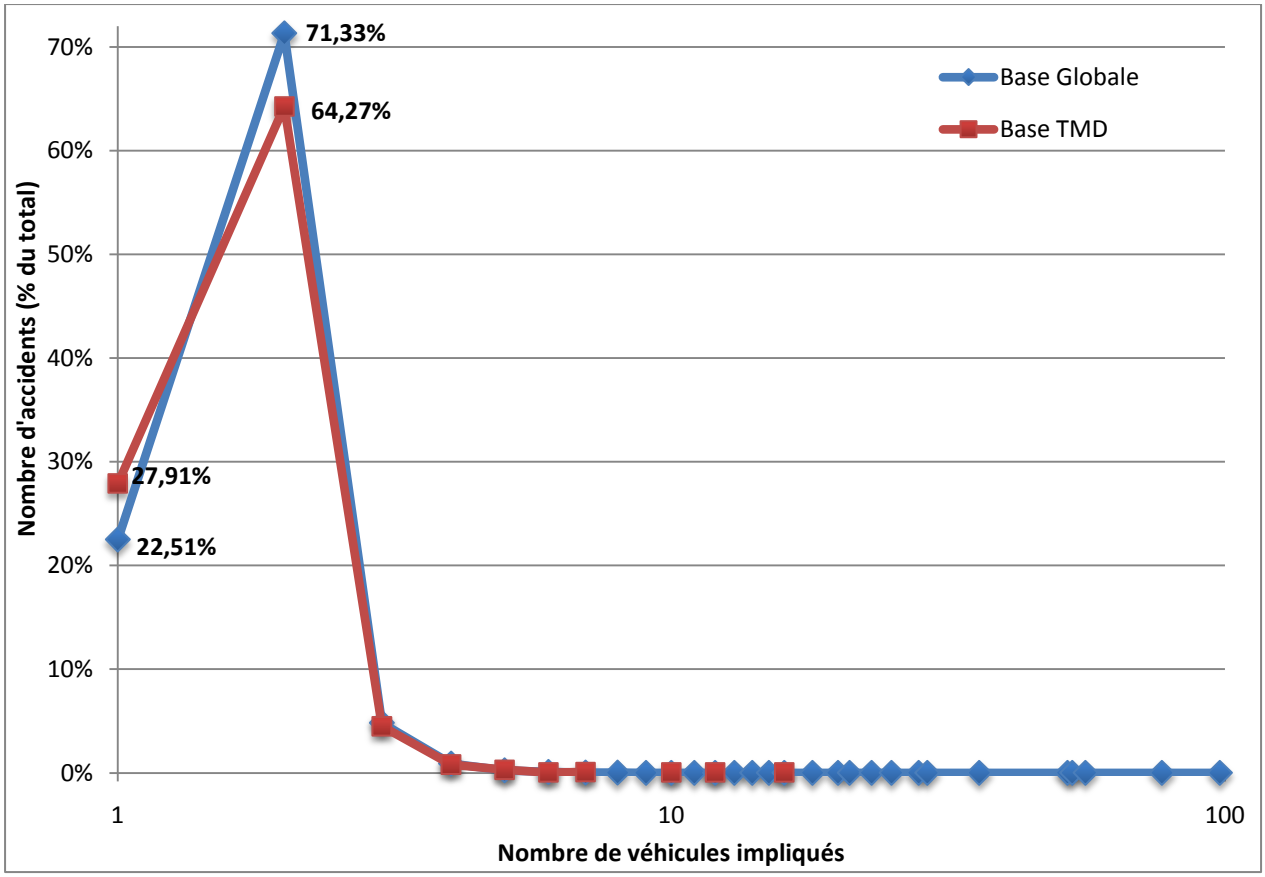


Figure 4.10 : Fréquences des nombres de véhicules impliqués (% du total)

Les incidents impliquent le plus fréquemment, dans les deux cas, deux véhicules, suivis par les accidents impliquant un seul véhicule. Les accidents impliquant plus de trois véhicules sont assez rares. Par contre si les deux courbes sont similaires pour les accidents avec plus de trois véhicules on remarque une différence notable pour les accidents à 1 et 2 véhicules, avec un croisement de courbes particulièrement visible. On voit ici pourquoi les deux moyennes sont significativement différentes : les accidents TMD impliquent plus fréquemment un seul véhicule.

Cela peut s'expliquer par le fait que certains types d'accidents sont propres au transport de matières dangereuses, comme un déversement de matières, et n'impliquent la plupart du temps qu'un seul véhicule.

- Évolution annuelle comparé du nombre moyen de véhicules impliqués

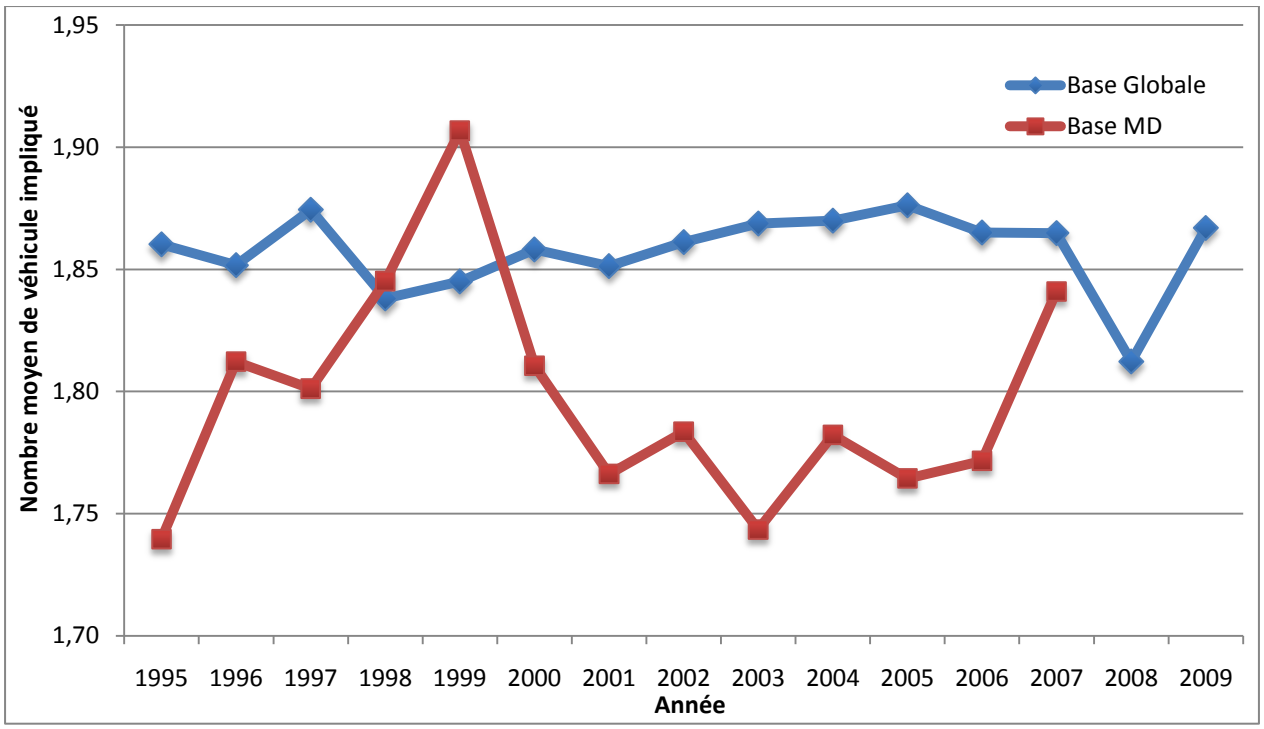


Figure 4.11 : Évolution annuelle du nombre moyen de véhicules impliqués par accident

Le nombre moyen de véhicules impliqués varie assez peu selon les années pour la base globale, valant entre 1,85 et 1,86 véhicule par accidents. Une baisse importante de 0,04 véhicule a cependant été enregistrée en 2008, mais en 2009 a marqué un retour à une valeur « normale ». On a montré que les moyennes des deux distributions étaient significativement différentes avec une différente d'environ 0,06 véhicule entre les deux distributions. On constate cependant sur cette courbe que la moyenne du nombre de véhicule impliqué par accident TMD n'est pas toujours en-dessous de celle de la base globale. En effet en 1998 et surtout en 1999 cette valeur est supérieure. Le « pic » de 1999 résulte en réalité d'un accident TMD particulier qui a impliqué 16 véhicules (c'est le maximum sur la base de données). Mises à part ces deux années le nombre moyen de véhicules impliqués est toujours en-dessous de la valeur observée dans la base globale, ce qui explique la différence des moyennes significatives.

On peut en conclure que les accidents de transport de des matières dangereuses impliquent moins de véhicules que les autres accidents de camion.

4.2.3 Champ « Limitation de Vitesse »

Dans cette section l'étude du champ indiquant la limitation de vitesse sur la route où l'accident a eu lieu est présentée. Ce champ, présent dans les deux bases de données, est renseigné pour environ un accident sur deux.

- Tableau d'effectifs

Tableau 4.13 : Fréquence d'apparition des valeurs des « limitation de vitesse » pour la période 1995-2007

Limitation de vitesse	Fréquence dans la Base Globale	Fréquence dans la Base TMD
10 km/h	0	3
15 km/h	0	5
20 km/h	0	1
30 km/h	49	10
40 km/h	0	11
45 km/h	0	4
50 km/h	18 760	355
60 km/h	913	11
65 km/h	0	3
70 km/h	18 695	234
80 km/h	3 349	58
90 km/h	23 957	387
100 km/h	27 7992	242

On remarque que la base TMD comporte une plus grande gamme de limitations de vitesse renseignée. Il n'y a que sept limitations de vitesse pour la base globale contre 13 pour la base TMD. Cela peut s'expliquer par le fait que la base globale ait été normalisée et que les valeurs de limitation « anormale » aient été arrondies. Les principales limitations de vitesse en vigueur au Québec hors autoroute sont de 90, 80, 70, 60, 50, 40,et 30 km/h (Ministère des Transports du Québec, 2008). D'autre part la vitesse maximale sur autoroute est de 100 km/h. Ces valeurs peuvent en revanche changer et prendre des valeurs différentes de façon temporaire en raison essentiellement de travaux (Direction de la sécurité en Transport, 2008) ou de circonstances particulières. Des limitations de 15 ou 45 km/h sont donc assez rares et cela indique que l'accident a probablement eu lieu sur une portion en travaux.

- Diagramme de fréquences cumulées

Dans ces conditions la meilleure façon de comparer les deux répartitions est l'évolution des pourcentages cumulés des fréquences de chacune des limitations de vitesse :

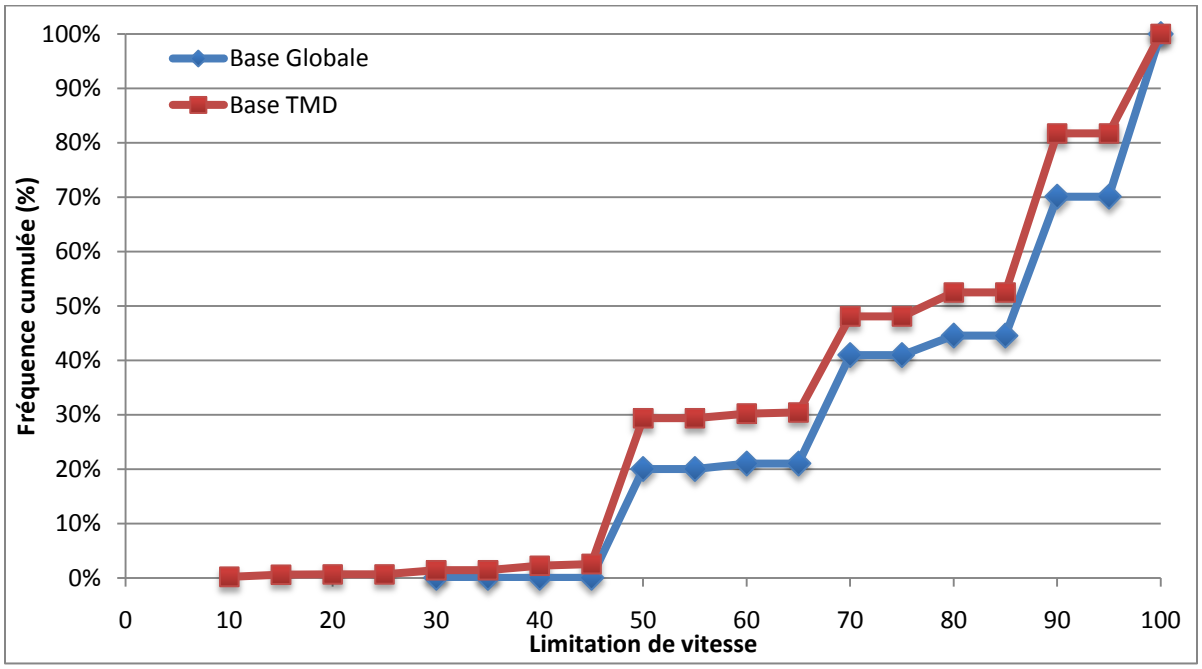


Figure 4.12 : Pourcentage cumulé des fréquences d'apparition des limitations de vitesse sur la période 1995-2007

Les deux courbes sont assez similaires. La principale différence est le fait que 30% des accidents TMD ont eu lieu sur des routes limitées à 50 km/h et moins, contre 20% des accidents de la base globale. Entre 60 et 90 km/h les proportions sont tout à fait similaires. Il y a par contre 10% d'accident TMD en moins sur les autoroutes (où la limitation est de 100 km/h) par rapport aux accidents de la base globale. La vitesse semble donc jouer un rôle moins important.

- Statistique descriptive

Tableau 4.14 : Quelques paramètres de statistique descriptive pour les répartitions des valeurs des limitations de vitesse (1995-2007)

	Base Globale	Base TMD
Nombre d'échantillons	93 717	1324
Minimum	30	10
Maximum	100	100
Plage	70	90
Moyenne	80,309	75,302
Médiane	90	80

Mode	100	90
Écart-type	18,71	20,23
Variance	350,87	409,48
Coefficient de variation	23,30%	26,86%
Coefficient d'aplatissement	-1,1795	-0,9814
Coefficient d'asymétrie	-0,5228	-0,4155

On constate que les deux moyennes présentent une différence d'environ 5 km/h. Afin de voir si cette différence est significative nous allons faire un test statistique. Par ailleurs on peut noter que les valeurs des coefficients de variation sont assez proches.

- Test sur la moyenne

Les deux distributions sont suffisamment normales, au vu des valeurs des leurs coefficients de forme, nous n'avons donc pas besoin de les normaliser. Un test F permet de montrer que les variances sont différentes :

$$TEST.F(serie\ BG; Série\ TMD) = 4{,}57 \times 10^{-5}$$

On peut donc appliquer le test de Student suivant :

$$TEST.STUDENT(serie\ normalisé\ BG; Série\ normalisé\ TMD; bilatéral; type\ 3) = 1{,}11 \times 10^{-18}$$

La probabilité obtenue est extrêmement faible, on rejette l'hypothèse nulle. Les moyennes sont significativement différentes.

- Conclusion

On peut en conclure que les accidents TMD ont plus souvent lieu sur des routes moins rapides que les autres accidents. On peut donc penser que la vitesse à moins d'influence sur les accidents impliquant des matières dangereuses. Ce résultat est cependant à utiliser avec précaution car pour seulement 40% des accidents une vitesse est renseignée.

4.2.4 Champs concernant les impacts sur la santé humaine

Cette section présente une étude des informations concernant les impacts sur la santé humaine qui ont été notés dans les rapports d'accidents. Pour les deux bases de données on retrouve le décompte du nombre de morts, de blessés graves et de blessés légers. Dans une première partie nous allons comparer les répartitions des décès, puis dans une seconde partie les répartitions de la somme blessés graves + blessés légers. On trouve par ailleurs dans la base de données concernant

le TMD un champ supplémentaire indiquant le nombre de « victime sans blessure », comme il n'y a pas d'information équivalente dans la base globale, ces informations n'ont pas été exploitées.

4.2.4.1 Nombre de décès

On compare les données sur les nombres de décès observés dans nos deux bases de données.

- Tableau d'effectifs

Tableau 4.15 : Fréquence d'apparition des nombres de décès pour la période 1995-2007

Nombre de décès	Nombre d'accidents Base globale	Nombre d'accidents TMD
0	225 323	2 309
1	1 287	58
2	114	8
3	26	2
4	3	1
5	6	1

Il est étonnant de noter que dans les deux bases de données le nombre maximum de décès pour un accident est de 5 décès. On constate, fort heureusement que la majorité des accidents ne font aucun décès, et que les accidents mortels sont assez rares. Afin de préciser cela, nous allons calculer les paramètres de statistique descriptive des deux distributions, mais auparavant traçons les courbes p-N à obtenues à partir de ces chiffres.

- Courbes p-N

Pour étudier les conséquences sur la santé humaine, les chercheurs utilisent souvent des courbes p-N cumulative (cf. § 2.1.4). La méthode de tracée de ce type de courbe est expliquée dans la section 3.3.2.6. Voici les courbes obtenues pour le nombre de décès :

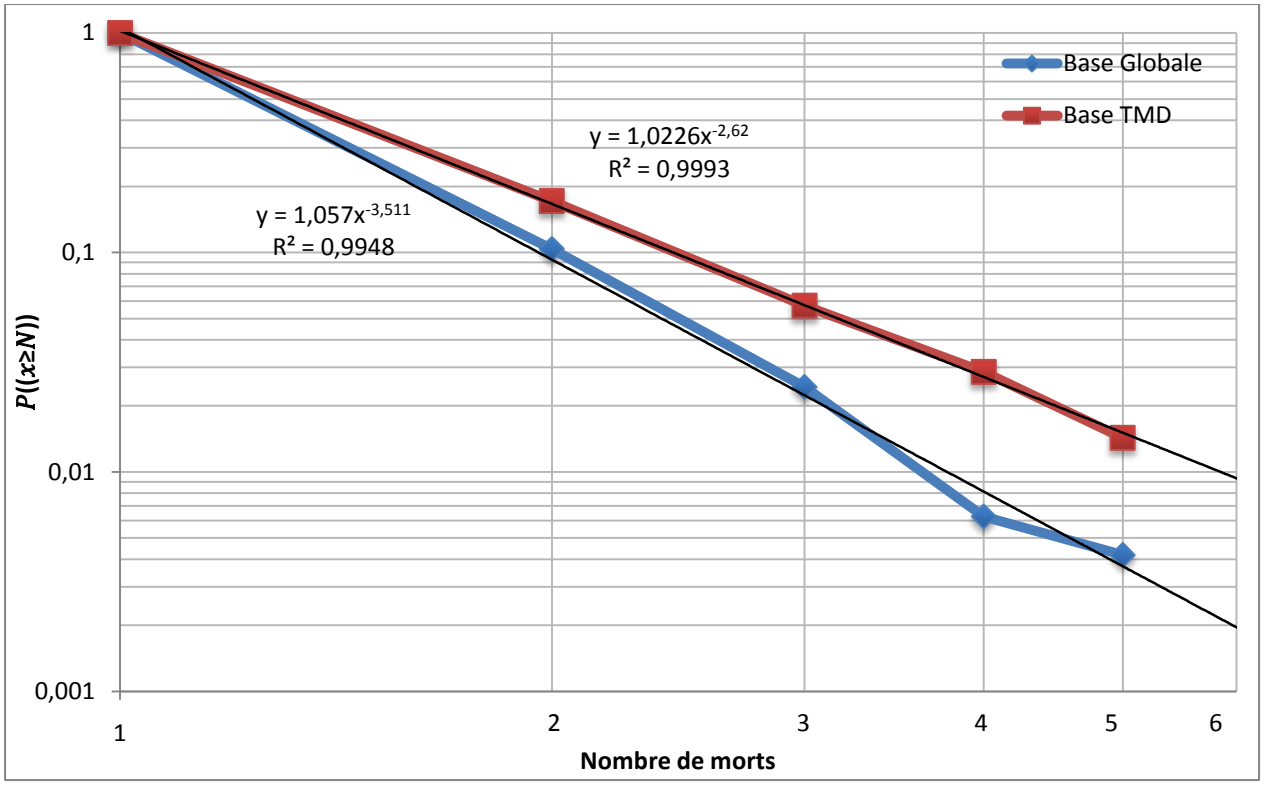


Figure 4.13 : Probabilité cumulée d'un accident TMD avec N décès

Les pentes obtenues sont de $-2{,}62$ pour la base TMD et de $-3{,}51$ pour la base globale, avec dans les deux cas un coefficient de détermination très bon ($\approx 0{,}99$). Les accidents TMD ont donc plus une probabilité plus forte de donner lieu à des accidents avec un nombre de décès élevé. Les pentes obtenues sont beaucoup plus forte que celles obtenue dans les études mentionnées dans la revue de littérature qui tourne autour de -1. Cela peut s'expliquer par le fait que ces études ont étudiés des bases de données avec une forte représentation des accidents avec un fort impact, comme la base de données MHIDAS par exemple, alors que nos données ne favorisent à priori pas ce type d'évènement.

- Statistique descriptive

Tableau 4.16 : Valeurs des paramètres de statistique descriptive pour les décès (1995-2007)

Statistique descriptive		Base Globale	Base TMD
Paramètres des données	Nb d'échantillons total % vide Somme des échantillons	226 759 0,00% 1635	2 379 0,00% 89
Paramètre de	Minimum	0	0

position	Maximum	5	5
	Mode	0	0
	Moyenne	0,0072	0,0374
	Médiane	0	0
Paramètre de variabilité	Étendue	5	5
	variance	0,0095	0,0613
	Écart-type	0,0977	0,2475
	coefficient de variation	1 354,61%	661,58%
Paramètre de forme	Coefficient d'asymétrie de Fisher	17,97	9,67
	Coefficient d'aplatissement	463,41	129,07

Les valeurs la plus fréquente est dans les deux cas, de 0 morts par accidents comme l'indique la valeur du mode. Il y a eu sur la période étudiée de 12 ans, 1 635 décès dans les accidents de véhicules lourds et 89 décès dans des accidents impliquant un transport de matières dangereuses. On constate de gros écart entre les moyennes et les écart-types dans les deux distributions, ce qui se traduit par des coefficients de variation important. La moyenne du nombre de mort par accident près de 5 fois supérieur à celle des accidents de la base globale. Les courbe p-N ont cependant montré que les accidents TMD ont un nombre de mort par accident plus important, il serait donc intéressant de déterminer quelles sont les proportions d'accident mortel dans les deux bases de données. Voici les proportions observées :

Tableau 4.17 : Proportion d'accident mortel

	Base Globale	Base TMD
Nombre d'accidents mortels pour 100 accidents	0,63	2,94

Il y 4,6 fois plus d'accident mortel dans la base TMD que dans la base globale. Nous allons réaliser un test de proportion afin de déterminer si cette différence est significative.

- Test de proportion

Pour réaliser ce test on laisse les enregistrements sans mort à zéro, et on attribue un 1 à celles ayant un mort ou plus. On peut alors appliquer un test de Student pour comparer les moyennes de ces nouvelles distributions :

$$TEST.STUDENT(serie\ binaire\ BG; Série\ binaire\ TMD; bilatéral; type\ 3) = 3{,}49 \times 10^{-11}$$

La probabilité obtenue est extrêmement faible. Les proportions sont donc significativement différentes.

- Conclusion

Les accidents de transport de matières dangereuses sont à proportion égale responsable de plus d'accidents mortels que les autres accidents. De plus les accidents TMD mortel font plus de victimes par accident que les accidents de véhicules lourds.

4.2.4.2 Nombre de blessés

On compare maintenant les données concernant le nombre de blessés dans les rapports d'accidents à notre disposition selon la même méthode que pour les nombre de décès. Dans les deux bases de données les nombre de blessés graves et légers ont été sommés pour obtenir le nombre total de blessés.

- Tableaux de fréquence

Tableau 4.18 : Fréquence d'apparition des nombres de blessés pour la période 1995-2007

Classe	Base Globale	Base TMD
0	199 197	1 878
1	21 271	380
2	4 568	76
3	1 070	30
4	390	7
5	132	2
6	62	2
7	20	1
8	15	0
9	5	2
10	3	0
11	0	0
12	1	0
13	5	0
14	2	0
15	3	0
16	3	0
17	5	0
18	0	1
19	0	0
20 et plus	7	0

Tout comme pour les accidents mortels, on constate que la majorité des accidents ne font pas de blessés. La proportion de blessé est cependant bien plus importante que celle des décès, et l'étendue des valeurs reportées est plus grande.

- Courbes p-N

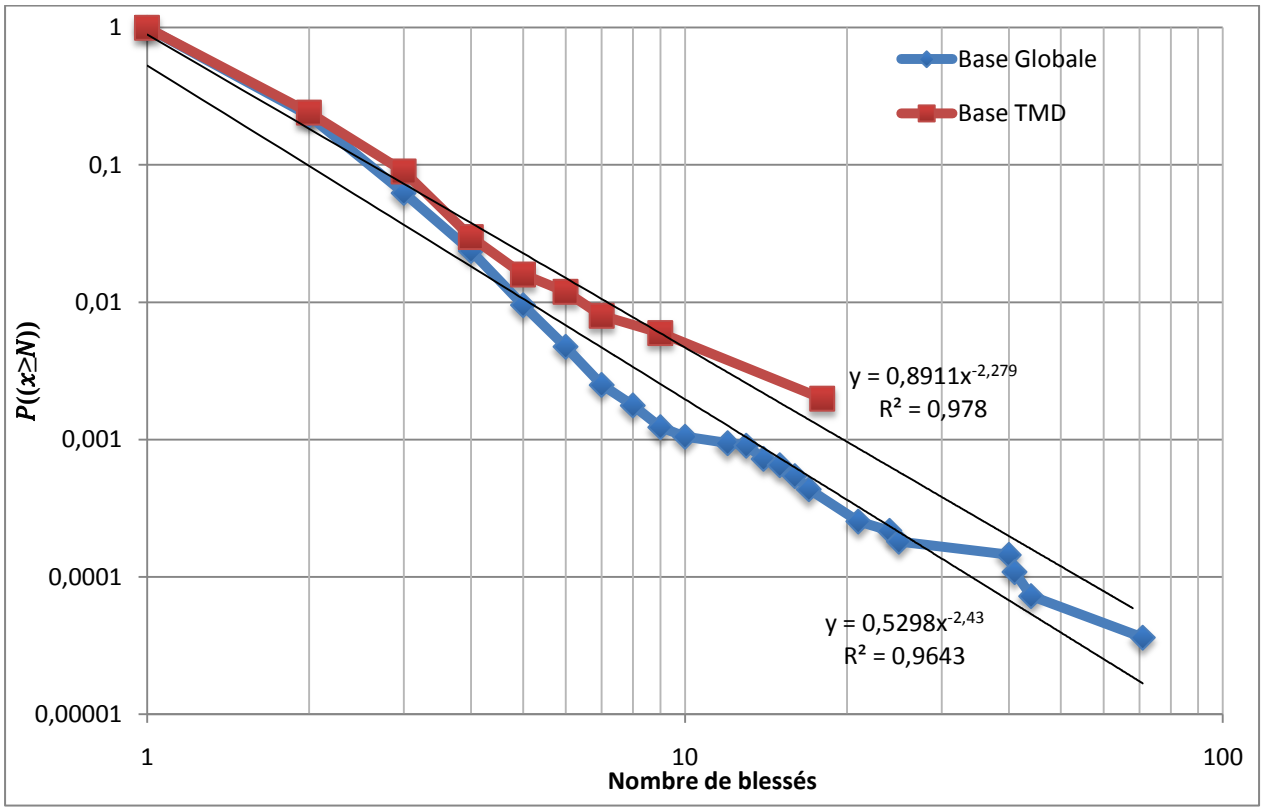


Figure 4.14 : Probabilité cumulée d'un accident TMD avec N blessés (grave ou léger)

Les pentes obtenues sont $-2{,}28$ pour la base TMD et de $-2{,}43$ pour la base globale, avec dans les deux cas un bon coefficient de détermination ($\approx 0{,}97$). Tout comme pour la courbe concernant les décès, la courbe de la base TMD est au-dessus de la courbe de la base globale ce qui signifie que les accidents TMD ont plus forte probabilité d'avoir des accidents avec beaucoup de blessés. Cependant l'écart est faible avec des courbes qui se chevauchent au début.

- Statistique descriptive

Tableau 4.19 : Valeurs des paramètres de statistique descriptive pour les blessés (1995-2007)

Statistique descriptive		Base Globale	Base MD
Paramètres des données	Nb d'échantillons total	226 759	2 379
	% vide	0,00%	0,00%
	Somme des échantillons	37 093	626
Paramètre de position	Minimum	0	0
	Maximum	71	14
	Mode	0	0

	Moyenne Médiane	0,1636 0	0,2631 0
Paramètre de variabilité	Étendue variance Écart-type coefficient de variation	71 0,3250 0,5700 348%	14 0,4680 0,6842 260%
Paramètre de forme	Coefficient d'asymétrie de Fisher Coefficient d'aplatissement	19,17 1523,08	6,15 82,22

On constate qu'entre 1995 et 2007, il y a eu 37 093 blessés dans les accidents de véhicules lourds dont 626 dans des accidents impliquant un transport de matières dangereuses. Le nombre moyen de blessé par accident n'est « que » 1,6 fois plus important pour les accidents matières dangereuses, ce qui fait une différence plus faible que pour le nombre de morts. Toutefois, en parcourant les données on constate que le nombre de blessés dans la base globale peut aller jusqu'à 71 et il est plusieurs fois supérieur à 20. Cela peut faire « gonfler » la moyenne par rapport aux accidents matières dangereuses qui impliquent au maximum 14 blessés. Les deux coefficients de variation sont particulièrement grands, ce qui s'explique par le fait que l'écart-type est supérieur à la moyenne. Voici les proportions observées :

Tableau 4.20 : Proportion d'accident mortel

	Base Globale	Base TMD
Nombre d'accidents avec au moins un blessé pour 100 accidents	12,15	21,06

Tout comme pour les nombre de décès, nous allons essayer de déterminer si la proportion d'accident impliquant des blessés est plus importante pour la base TMD.

- Test de proportion

Pour réaliser ce test on laisse les enregistrements sans blessé à zéro, et on attribue un 1 à ceux ayant un blessé ou plus. Pour comparer les moyennes on applique alors le test de Student suivant :

$$TEST.STUDENT(serie\ binaire\ BG; Série\ binaire\ TMD; bilatéral; type\ 3) = 1{,}28 \times 10^{-18}$$

La probabilité obtenue est extrêmement faible, les proportions sont significativement différentes. Les accidents de véhicules lourds transportant des matières dangereuses impliquent significativement plus souvent des blessés que les autres accidents de camions.

4.3 Étude des champs qualitatifs

Après les champs quantitatifs nous allons maintenant aborder l'étude de huit champs qualitatifs communs aux deux bases de données, qui précisent les circonstances des accidents (météo, impact, environnement,…). Pour chaque champ un diagramme en tuyaux d'orgue présente les fréquences observées en pourcentage des valeurs non nulles.

4.3.1 Champ « État de la météo »

Ce champ indique les conditions météorologiques au moment de l'accident. Voici les fréquences observées dans les deux bases de données exprimées en pourcentage des champs remplis :

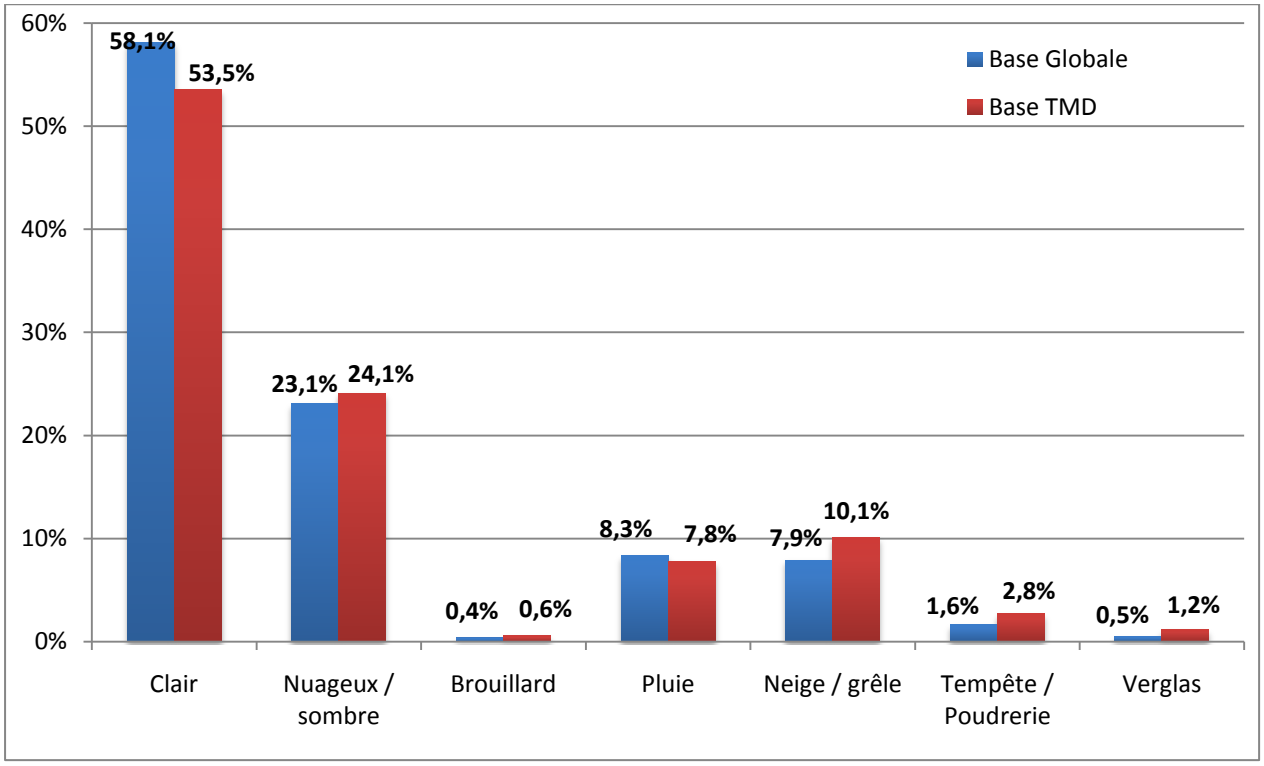


Figure 4.15 : Fréquence comparée des valeurs du champ « météo » (% du total)

Les proportions d'accidents par condition météo sont sensiblement les mêmes. Cependant les accidents TMD semblent, comme l'étude saisonnière et mensuelle l'ont suggérées, plus sensibles aux conditions hivernales. Cela se traduit ici pour un pourcentage plus élevé d'accident par temps de neige, de poudrerie et de verglas.

4.3.2 Champ « Genre d'accident »

Le champ « genre d'accident » indique le type d'accident. Voici la répartition comparée des valeurs du champ :

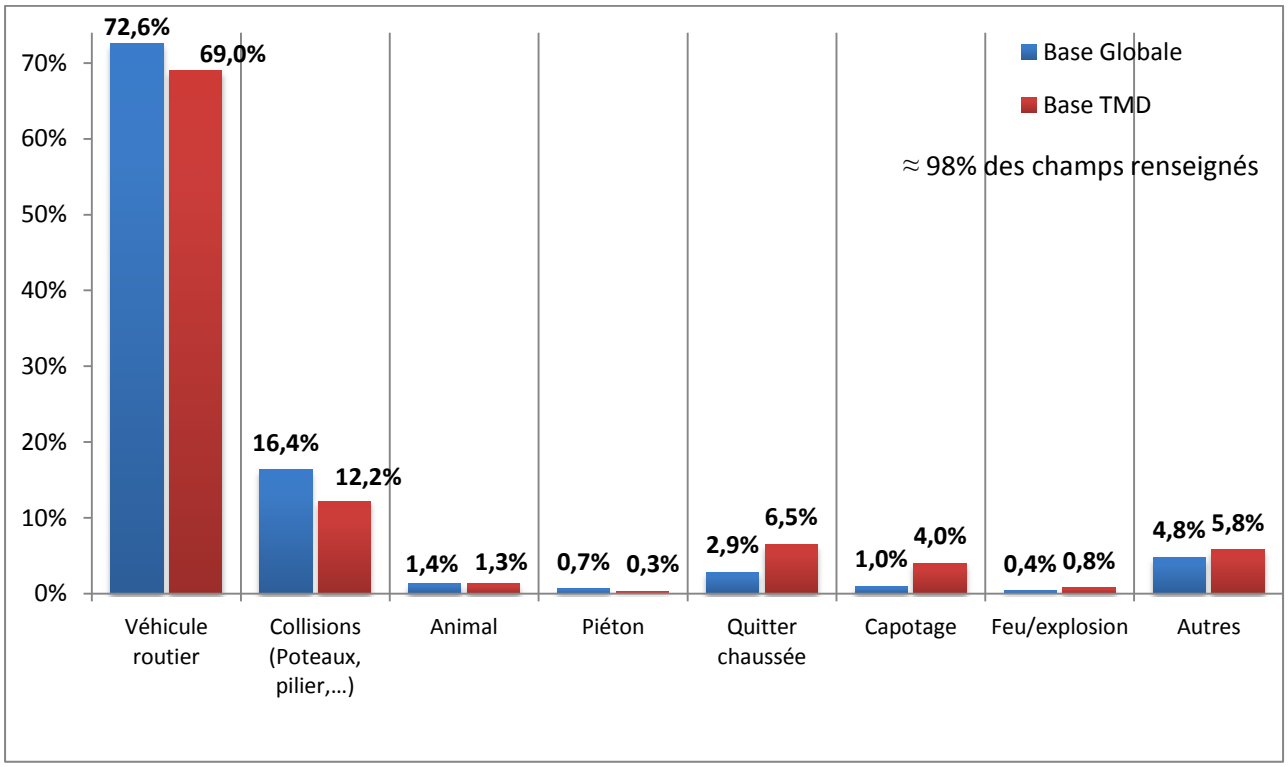


Figure 4.16 : Fréquence comparée des genres d'accidents (% du total)

On constate que la majorité des accidents sont des collisions entre véhicules (70 % des accidents) largement devant les collisions avec des objets fixes (15% environ). Les différences entre les deux bases sont assez faibles. On peut noter que les accidents TMD sont un peu plus fréquemment des capotages et des sorties de route et un peu moins des collisions entre véhicules et objets fixes (environ 3% d'écart). On peut aussi noter que les collisions avec des animaux se font dans les mêmes proportions quelques soit le type de marchandises. Cela montre que les animaux ne « choisissent » pas leurs camions. Le champ description de la base TMD apporte des précisions sur les animaux impliqués, voici la répartition constatée :

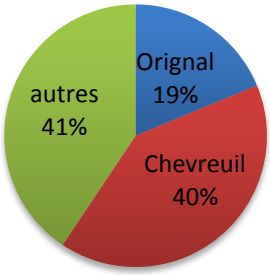


Figure 4.17 : Animaux impliqués dans des collisions TMD

4.3.3 Champ « Code d'impact »

Le code d'impact est un code décrivant le mouvement des véhicules impliqués dans l'accident. La signification précise de ce code est disponible en annexe. Voici le diagramme de répartition obtenu :

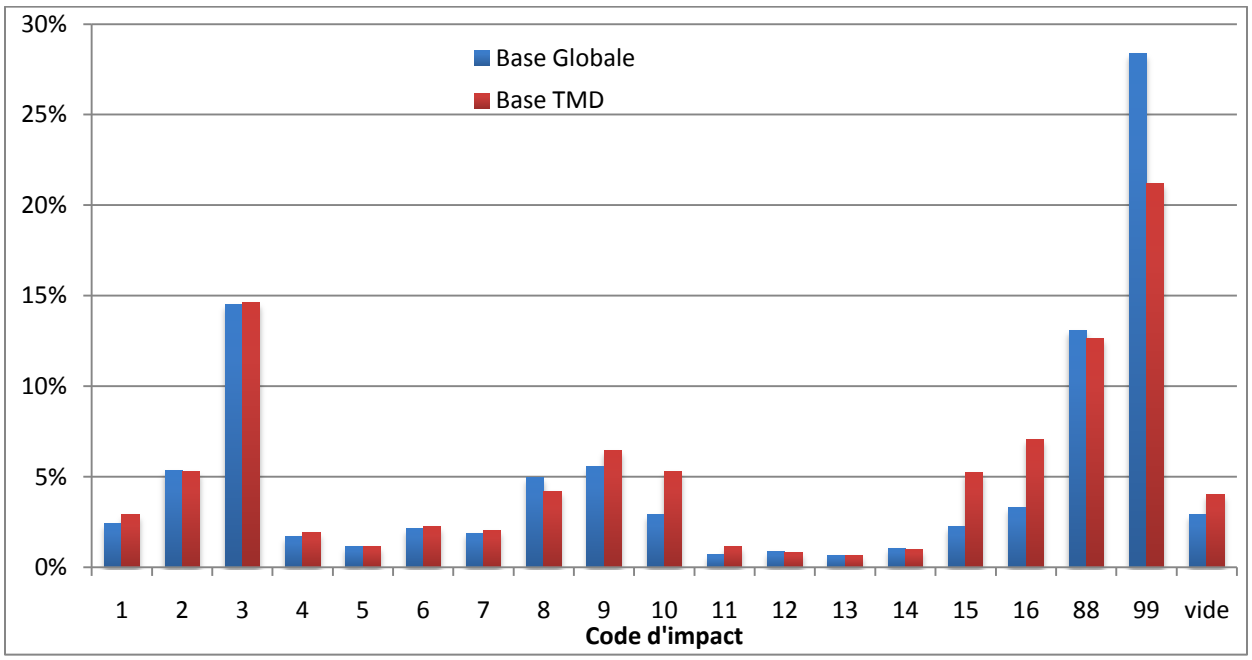


Figure 4.18 : Fréquence comparée des valeurs du champ « code d'impact » (% du total)

Le mode pour les deux bases de données est le code 99 qui concerne les accidents impliquant plusieurs véhicules dont le code n'est pas les codes précédents ou lorsqu'un un des deux premiers véhicules impliqués dans l'accident est un véhicules stationné. Le second code le plus fréquent est le code 3, qui correspond à un impact entre deux véhicules allant dans la même direction et dont le deuxième percute le premier par l'arrière.

Les deux répartitions sont assez similaires. On constate néanmoins que les accidents TMD sont plus souvent concernés par un code d'impact de 10, 15 et 16. Un code d'impact de 10 indique une collision frontale tandis que les codes 15 et 16 concernent une sortie de route d'un véhicule seul sur la droite ou la gauche. Ce n'est pas surprenant de retrouver les codes 15 et 16, la section 4.3 ayant mis en évidence que le nombre de véhicules impliqués dans des accidents TMD est plus faible. Le code 10 est plus surprenant.

4.3.4 Champ « Aspect de la chaussée »

Le champ aspect de la chaussé » décrit le type d'alignement de la chaussée (plane, courbe,…). Ce champ est cependant peu remplis (15% pour la base globale et 39% pour la base TMD) et donc difficile à interpréter. Voici les fréquences observées dans les deux bases de données exprimées en pourcentage des champs remplis :

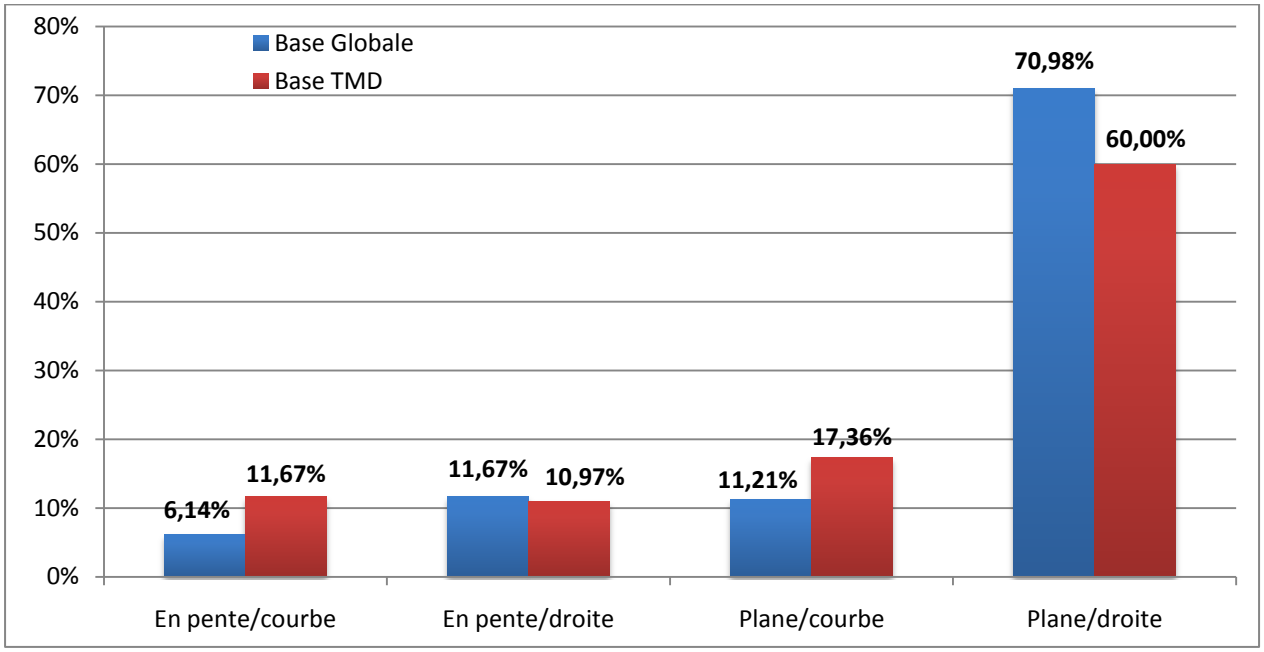


Figure 4.19 : Fréquence comparée des valeurs du champ « aspect de la chaussée» (% du total)

Les accidents se produisent sans surprise le plus souvent sur des routes droites. On constate cependant que les accidents TMD sont plus sensibles à la courbure de la route avec des fréquences d'accidents supérieures en proportion. Cela est peut-être dû au fait que les véhicules de transport des matières dangereuses sont en général plus difficile à manier dans une courbe.

4.3.5 Champ « Éclairement »

Le champ éclairement indique les conditions d'éclairage et de visibilité au moment de l'accident. Voici les fréquences observées dans les deux bases de données exprimées en pourcentage des champs remplis :

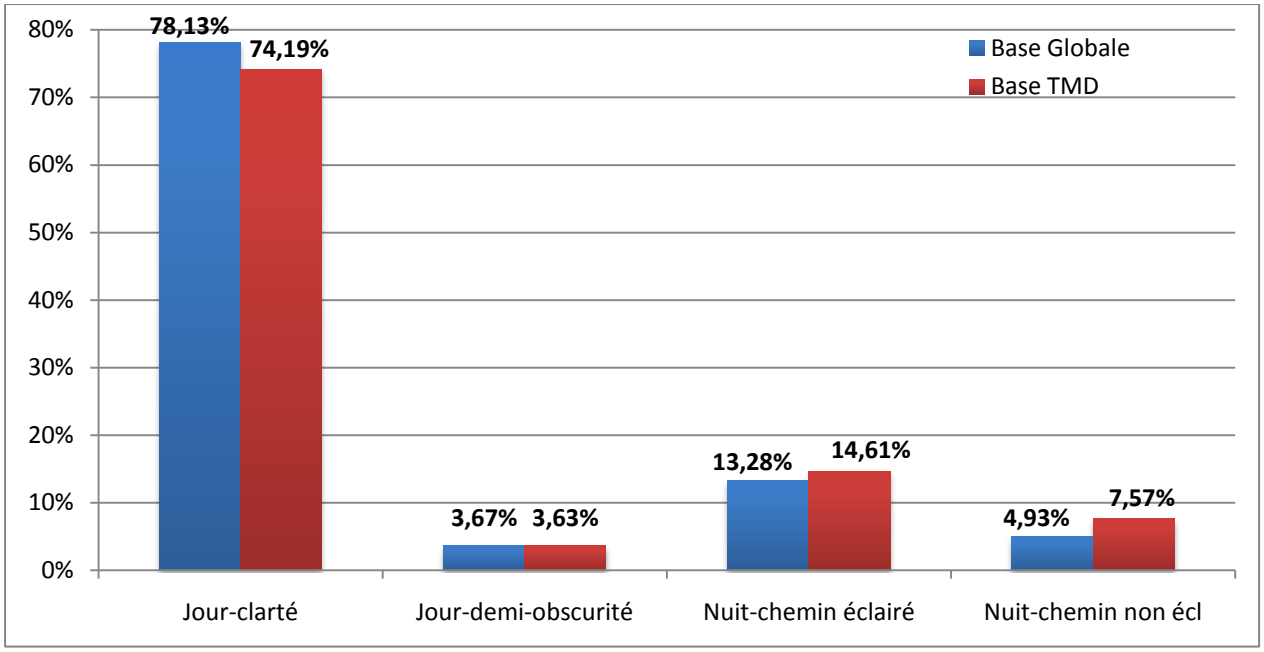


Figure 4.20 : Fréquence comparée des valeurs du champ « éclairement» (% du total)

Les accidents ont majoritairement lieu le jour, ce qui est normal car c'est à ce moment que l'essentiel des déplacements ont lieu, comme le montre la répartition horaire obtenue. La répartition est assez similaire dans les deux bases de données mais les accidents TMD se produisent un peu plus souvent de nuit. Cela se remarque aussi dans la répartition des heures (cf. section 4.2.2).

4.3.6 Champ « Environnement »

Le champ « environnement » indique l'environnement dans lequel l'accident a eu lieu. Dans les deux bases de données environ 70% des enregistrements ont ce champ vide. Voici les fréquences observées dans les deux bases de données exprimées en pourcentage des champs remplis :

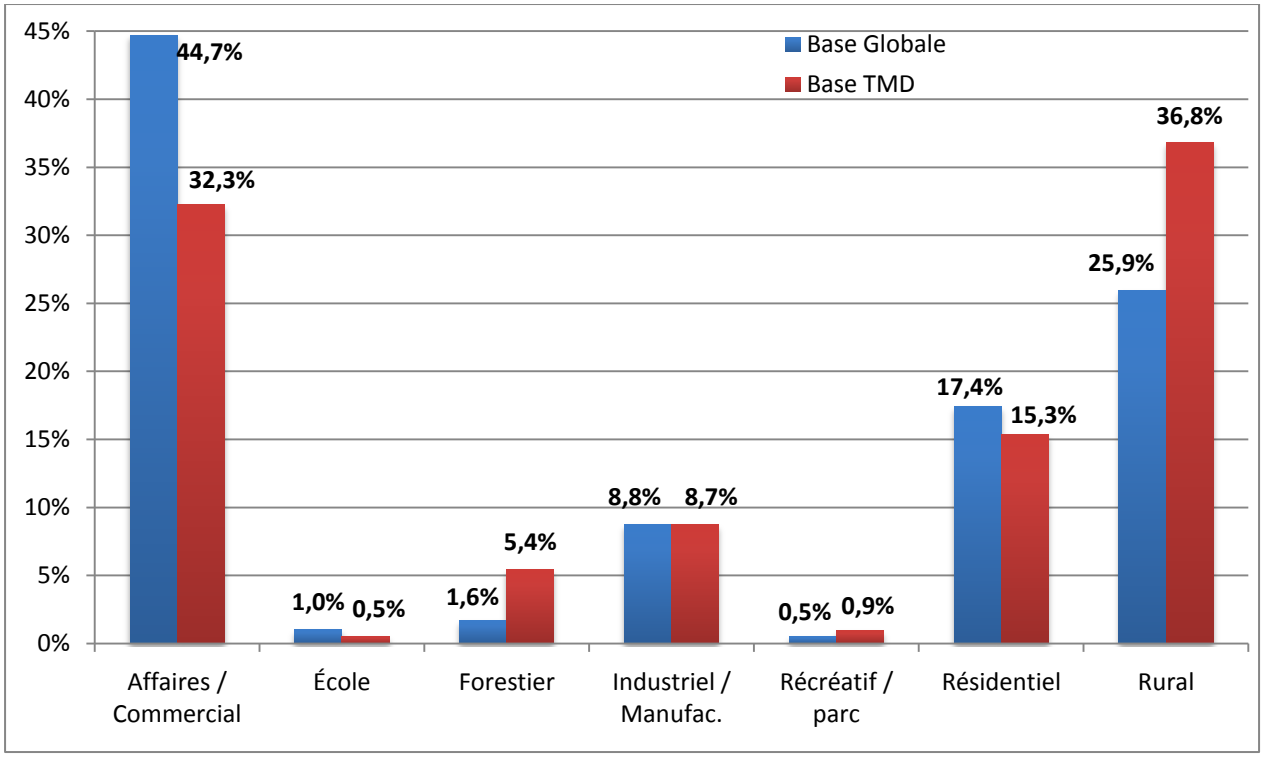


Figure 4.21 : Fréquence comparée des valeurs du champ « environnement » (% du total)

On constate que, par rapport aux accidents globaux, les accidents TMD ont plus souvent lieu dans un milieu rural (30 % de plus) et forestier (4% de plus) et moins souvent dans un environnement résidentiel et commercial. Cela est logique, des matières dangereuses étant souvent utilisées pour l'agriculture ou pour l'industrie forestière. Pour le milieu industriel la proportion d'accident est exactement la même.

4.3.7 Champ « État de la surface »

Le champ état de la surface décrit l'état de la surface lors de l'accident (mouillée, sèche,…). Voici les fréquences observées dans les deux bases de données exprimées en pourcentage des champs remplis :

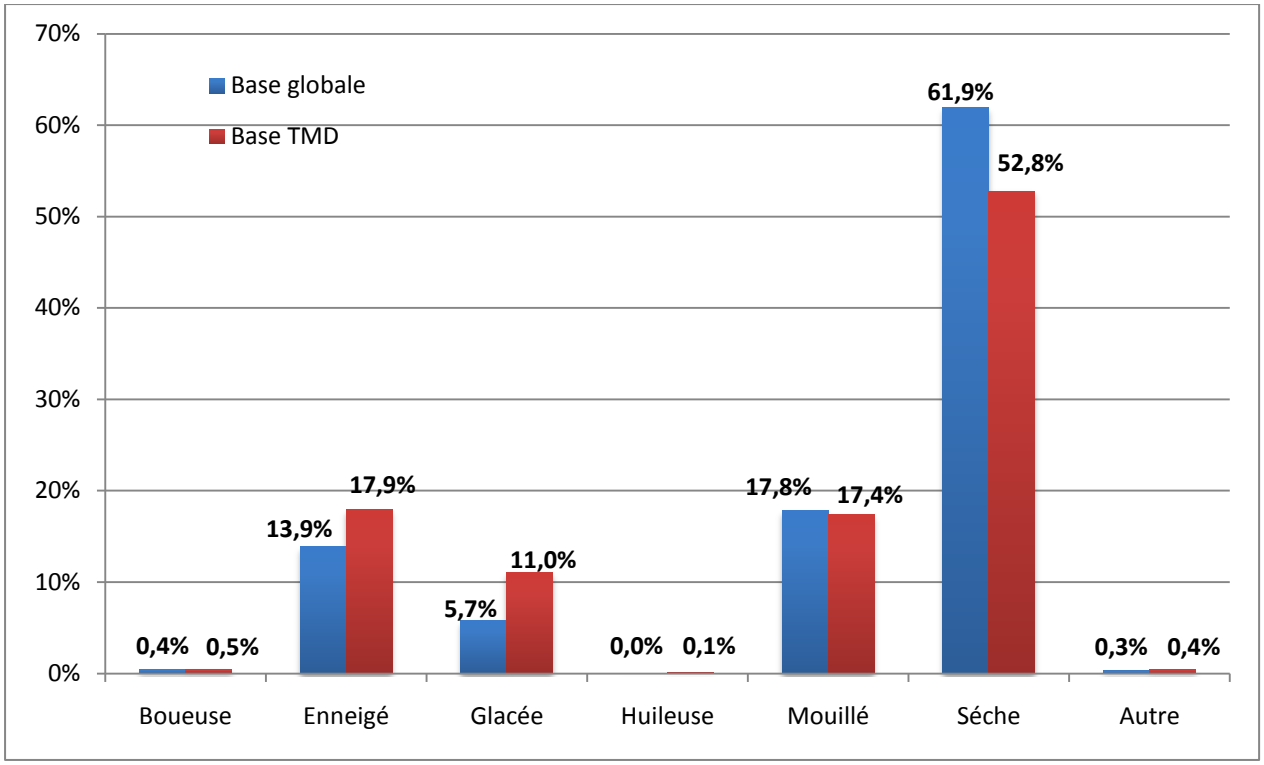


Figure 4.22 : Fréquence comparée des valeurs du champ « état de la surface » (% du total)

Les accidents ont le plus souvent lieu dans des conditions sèches dans les deux bases de données. On retrouve ici la conséquence du fait que les accidents TMD semblent plus affectés par les conditions hivernales (cf. § 4.1.2 et 4.1.3) en remarquant que ces accidents ont plus souvent lieu sur des surfaces enneigées et glacées et moins souvent sur une surface sèche.

4.3.8 Champ « Nature de la Chaussée »

Le champ « nature de la chaussée » décrit le type de chaussée sur lequel l'accident a eu lieu. Voici les fréquences observées dans les deux bases de données exprimées en pourcentage des champs remplis :

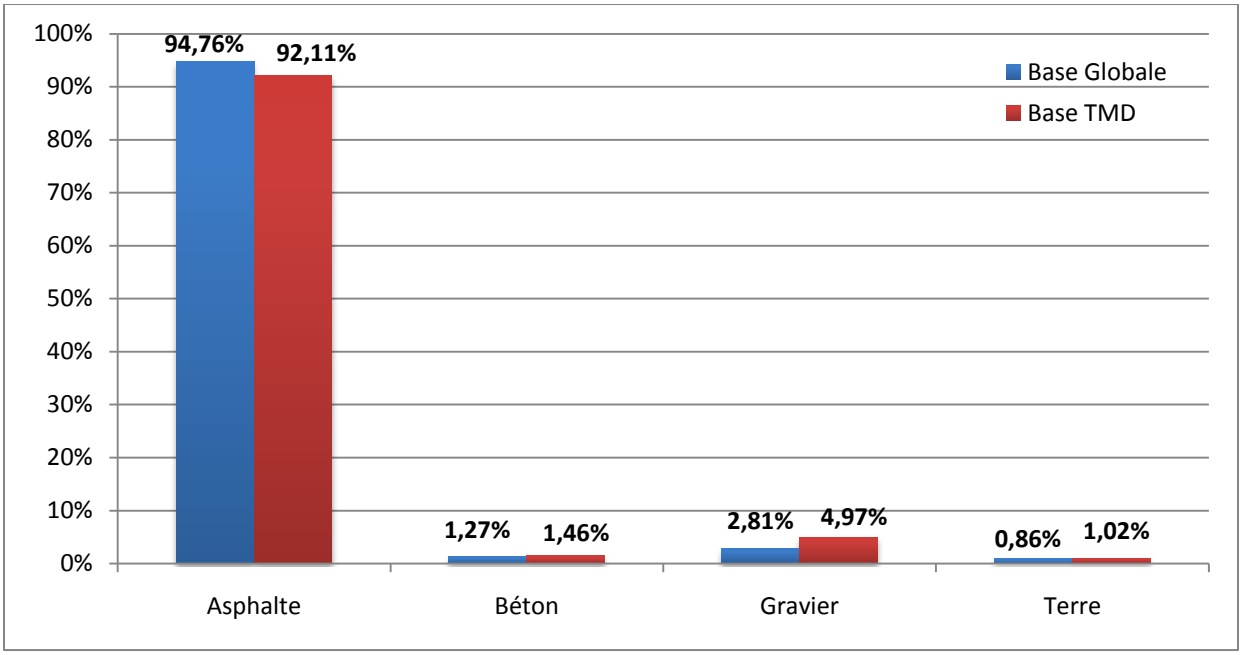


Figure 4.23 : Fréquence comparée des valeurs du champ « nature de la chaussée » (% du total)

L'asphalte est le type de chaussée le plus fréquent et logiquement c'est le type de chaussée que l'on retrouve largement en tête des fréquences d'accidents. Les accidents ont un petit peu plus souvent lieu sur les autres surfaces mais c'est minime.

CHAPITRE 5 ETUDE SPATIALE DES ACCIDENTS

Dans cette section l'analyse spatiale des données est présentée. Comme expliqué dans la section 3.4, le logiciel SIG utilisé pour afficher les cartes est MapWindows GIS, tandis que le logiciel CrimeStat est utilisé pour calculer les données statistiques spatiales présentées. On verra dans un premier temps la répartition spatiale des accidents qui a été obtenue (§ 5.1) avant d'en faire une analyse statistique (§5.2). Enfin la dernière partie présente une analyse spatiale pondérée, par le nombre de véhicules impliqués, de décès et de blessés (§5.3).

5.1 Analyse des répartitions d'accidents

Dans un premier temps nous allons présenter les deux répartitions que l'on a obtenues pour les accidents de la base globale et pour les accident impliquant des matières dangereuses. Il faut bien noter, comme indiqué dans le paragraphe 3.4.1, que les deux bases de données ont été localisées à l'aide de MapPoint, mais que seule la base TMD à fait l'objet d'une correction manuelle approfondie.

5.1.1 Aperçu global des répartitions d'accidents

La figure ci-dessous présente les deux répartitions obtenues à l'échelle du Québec habité. En annexe 9 on trouve les répartitions d'accidents à différentes échelles du Québec. On constate logiquement pour les deux bases que les accidents se concentrent au sud du Québec, le long du Saint-Laurent, là où réside la majorité de la population.

On distingue plus précisemment une forte densité d'accidents dans toutes la région sud du Québec notamment entre Montréal et Québec, qui sont les deux principaux pôles. Le fjord du Saguenay et le lac Saint-Jean, et la région de Gatineau sont également, dans une moindre mesure, le théâtre de nombreux accidents pour les deux bases. Enfin, la dernière zone de forte concentration se distingue aux alentours des villes de Rouyn-Noranda et Amos. Les routes qui relient ces différentes zones sont également bien visibles sur les cartes, comme par exemple la 155 qui relie Trois-Rivières au Lac saint-Jean et la 117 qui se rend en Abitibi. Enfin, les deux routes longeant de part et d'autre le Saint-Laurent regroupent la plupart des accidents survenus à l'Est du Québec, notamment la route 132, qui borde aussi la Gaspésie.

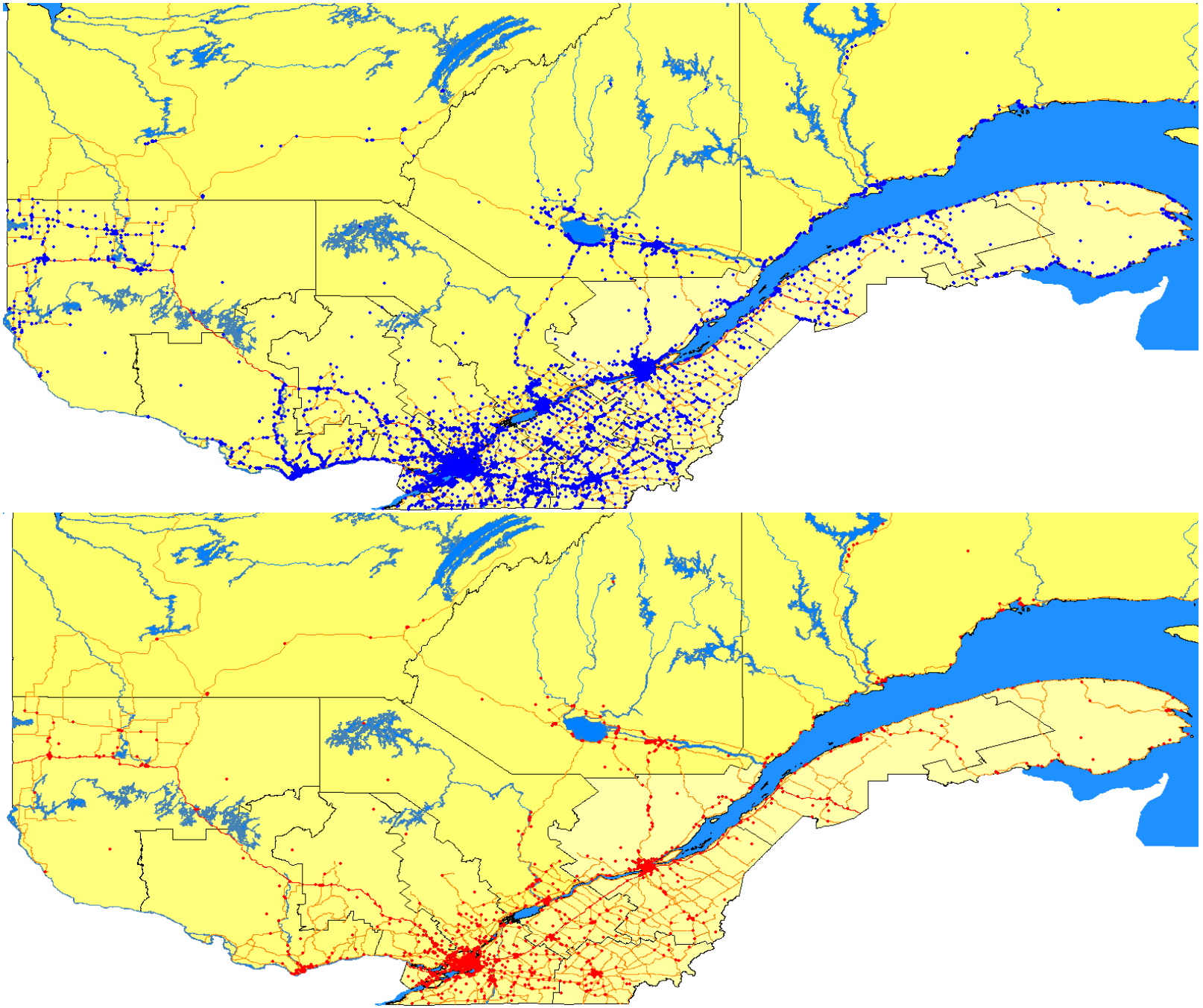


Figure 5.1 : Répartition des accidents TMD (rouge) et des accidents de la base globale (bleu) dans le Québec « habité » et principales routes

En ce rapprochant de la zone de forte densité, dans le Sud du Québec (voir la figure 2 de l'annexe 9), on distingue clairement sur les 2 répartitions les villes de Montréal, Trois-Rivières, Québec et Sherbrooke, qui constituent des agglomérats importants d'accidents. On peut aussi identifier, notamment sur la carte de la base globale, des villes intermédiaires, qui sont pour la plupart des nœuds routiers au croisement de plusieurs grosses routes, comme Drummondville, Victoriaville ou Granby. En dehors de ces amas, les accidents semblent répartis de façon assez homogène.

Il est toutefois difficile d'analyser précisemment la répartition avec ce type de carte, il faut pour cela pouvoir quantifier le nombre d'accident par zone déterminée.

5.1.2 Répartition par région économique (RE)

Voici, sur la page suivante, les distributions des accidents selon les régions économiques obtenues pour les deux bases de données. Une analyse selon une division géographique plus fine serait plus précise mais les accidents de la base globale ne sont pas situés avec sufisament de précision pour que les résultats soient interprétés de façon pertinente. Une analyse selon les RE permet tout de même d'avoir un bon aperçu des répartitions d'accident sur le territoire du Québec.

On constate pour les accidents TMD que la RE la plus touchée est la Montérégie avec près de 20% des accidents devant Montréal (15%) et la Capitale-Nationale avec environ 8% des accidents TMD sur la période 1995-2007. Pour les accidents de la base globale, près de 26% des accident ont eu lieu pour la RE de Montréal, et 17 % en Montérégie. Ainsi par rapport aux accident TMD, on constate que les accidents de véhicules lourds ont plus souvent lieu à Montréal. La répartition des accidents TMD est plus étalée le long du Saint-Laurent, moins centrée dans la région de Montréal.

On remarque ainsi que pour la base TMD 3,8% des accidents ont lieu sur la Côte-Nord contre seuleument 1,8% pour la base globale. On peut avancer comme explication la réalisation, dans cette région, de nombreux grands chantiers, notamment la construction d'ouvrage hydro-éléctriques, qui ont nécessité l'utilisation de matières dangereuses (explosifs par exemple). Ces accidents, anecdotiques pour la base globale, ont une influence relativement importante dans la base TMD.

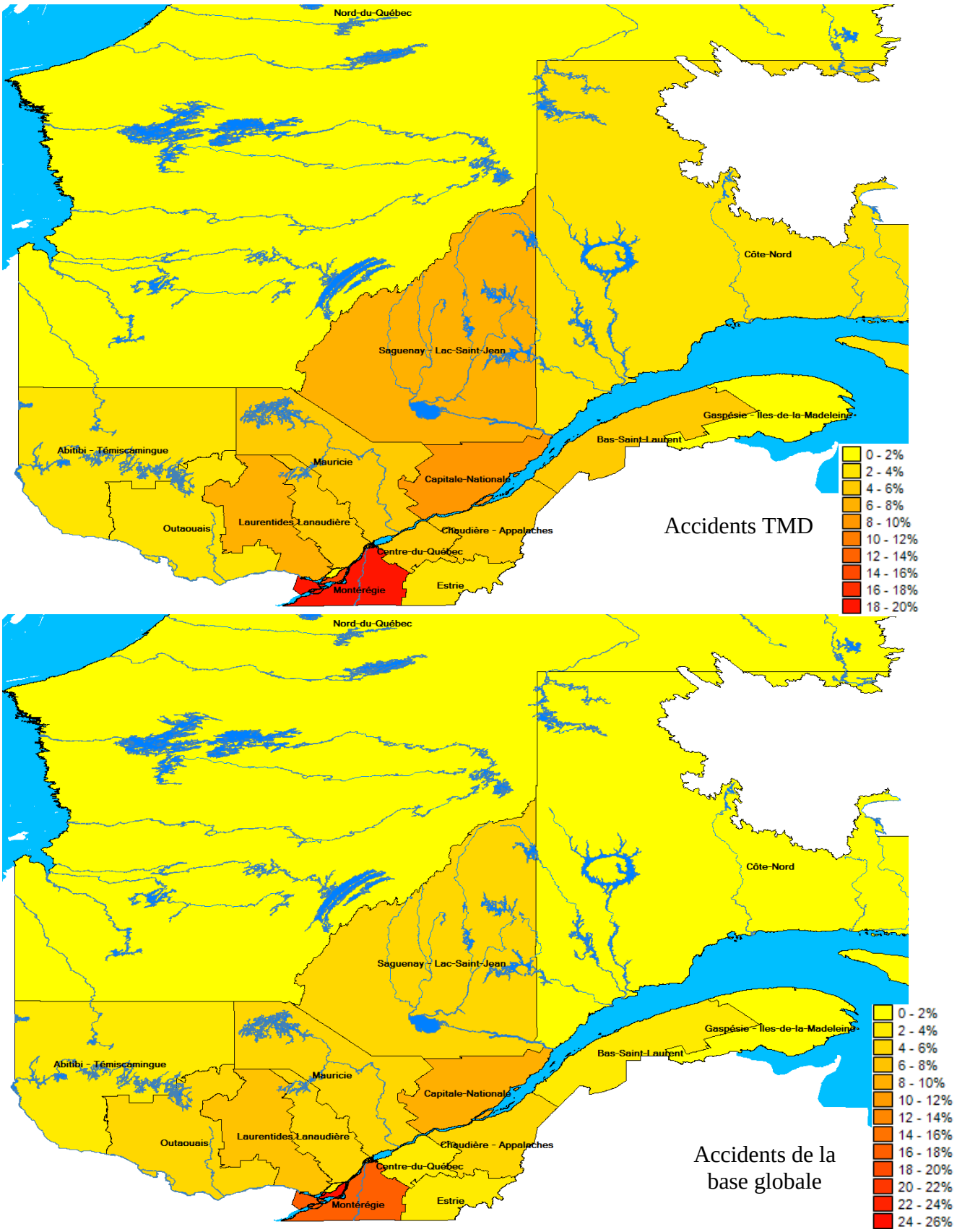


Figure 5.2 : Répartition par RE des accidents des deux bases (% du total)

Il est toutefois évident que les accidents ne sont pas répartis de façon homogène dans une région. Ces cartes donnent donc une idée du poids de chaque région, mais pour une analyse plus fine, nous déterminerons plus loin les densités d'accidents.

5.1.3 Répartition des accidents sur le réseau routier

Il est interressant d'observer comment les accidents se situent sur le réseau routier. Sur les deux cartes de la figure 5.1, présentant les accidents dans la partie habitée du Québec, ainsi que sur celles en annexe 9, le réseau routier principal est présent. Il apparait clairement que dans certaines régions il n'y a que très peu d'accidents en-dehors des routes principales. C'est le cas par exemple de la 155 qui relie Trois-Rivières au Lac saint-Jean, de la 175 entre Québec et Chicoutimi ou encore de la route 132, qui borde la Gaspésie.

Il est plus difficile de distinguer les routes dans les zones de très forte densité, mais en utilisant la figure 2 de l'annexe 9, il apparait également que les accidents sont en majorité sur des routes importantes, notamment dans le cas de la base TMD, pour laquelle peu de points sont situés dans les zones vides correspondant aux routes secondaires. Ainsi, sur l'autoroute 20, reliant Montréal et Québec, on compte environ 80 accidents, alors que la portion de la 132 entre ces deux villes en compte moitié moins. Toutefois ces chiffres restent assez imprécis. Concernant le base globale, il est encore plus difficile de tirer des conclusions à cette échelle en raison de la localisation imparfaite des accidents. En effets, les accidents ayant eu lieu sur une autoroute ou une route principale sont souvent uniquement localisés par leur numéro ainsi que par le kilomètre où l'accident a eu lieu, ce que MapPoint n'est pas capable de comprendre. L'accident est alors localisé dans la municipalité associée. MapPoint aura donc tendance à mieux placer les accidents ayant lieu sur route qui à un nom, c'est-à-dire les chemins, les rues plutôt que sur les principales artéres reliant les villes entre-elles.

Descendons donc à l'échelle d'une ville, où, en-dehors des sites non localisés assignés au même point, on peut espérer que les accidents sont plus précisemment placés. Les deux cartes ci-dessous montrent la répartition des accidents au centre ville de Montréal.

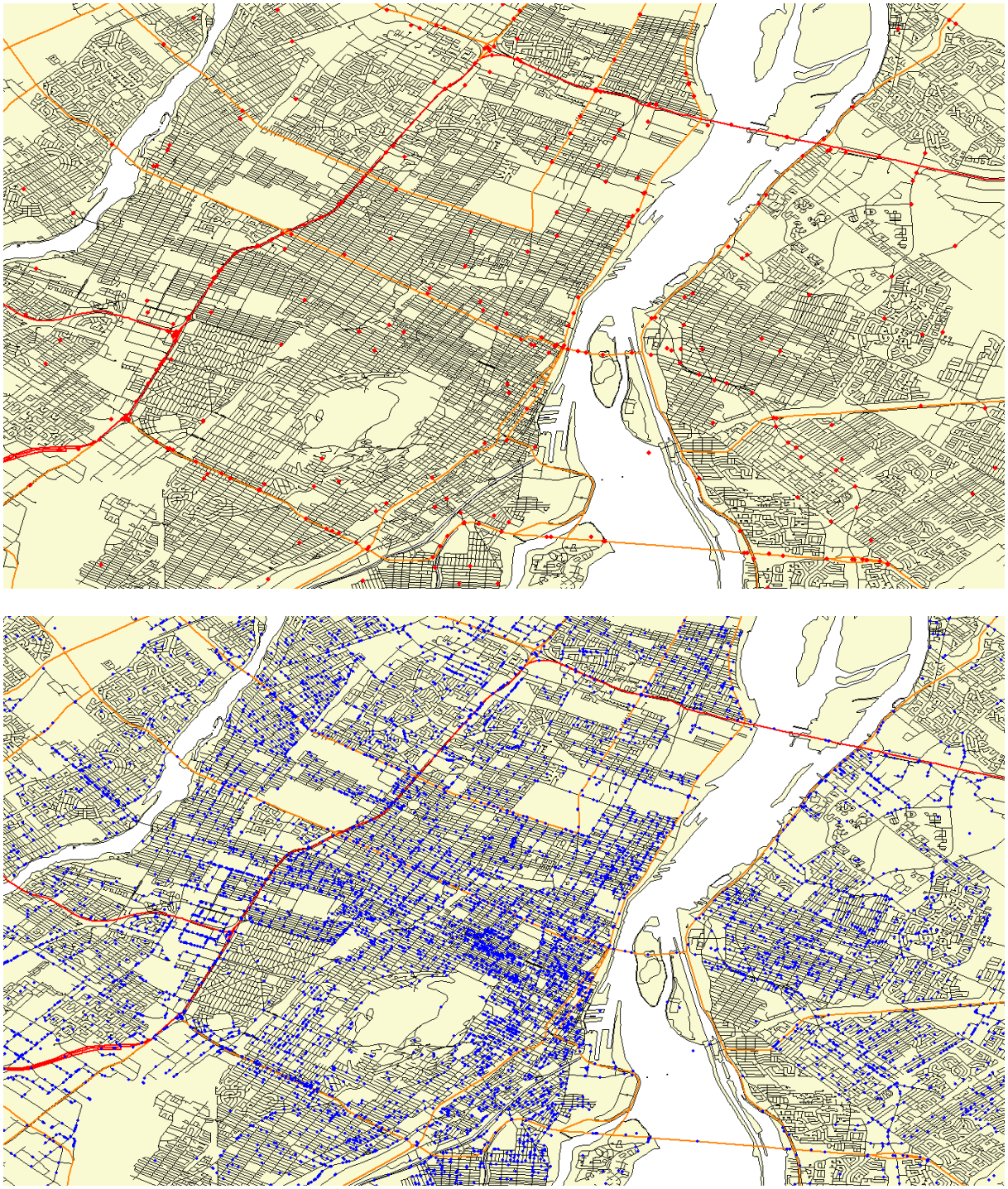


Figure 5.3 : Répartition des accidents obtenues à l'échelle de Montréal (bleu = base globale, rouge = base TMD)

Il apparait clairement que les accidents de matières dangereuses sont majoritairement situés sur les routes importantes, notamment les autoroutes, comme par exemple sur la 40 qui traverse l'île

dans sa longueur. Les accidents de la base globale semblent dispersées de façon plus homogène sur toutes les routes de l'île, même si l'on observe une forte concentration au centre ville. Cependant on ne pas se fier à ces apparences pour conclure quoi que ce soit car un nombre très important d'accidents se trouve superposés en un point au centre-ville, MapPoint ne les ayant pas localisé avec plus de précision que la municipalité. Pour étudier comparativement les répartitions d'accident sur le réseau routier il faudrait relocaliser de façon plus précise la base globale, une tâche qui longue et délicate, comme expliqué dans le chapitre 3.

5.2 Analyse statistique des répartitions d'accidents

Nous allons maintenant mener une étude statistique des répartitions à l'aide du logiciel CrimeStat. Cette analyse comprend une analyse centrographique, une analyse des plus proches voisins, une analyse Hot Spot et une évaluation des densités à l'aide d'une fonction kernel.

5.2.1 Analyse Centrographique

Une analyse centrographique va nous permettre de comparer la position centrale et la dispersion de nos distributions. Cette analyse a aussi été menée pour la répartition de population afin de la comparer aux répartitions d'accidents. Voici tout d'abord la position des centres moyens obtenus ainsi que les caractéristiques de dispersion :

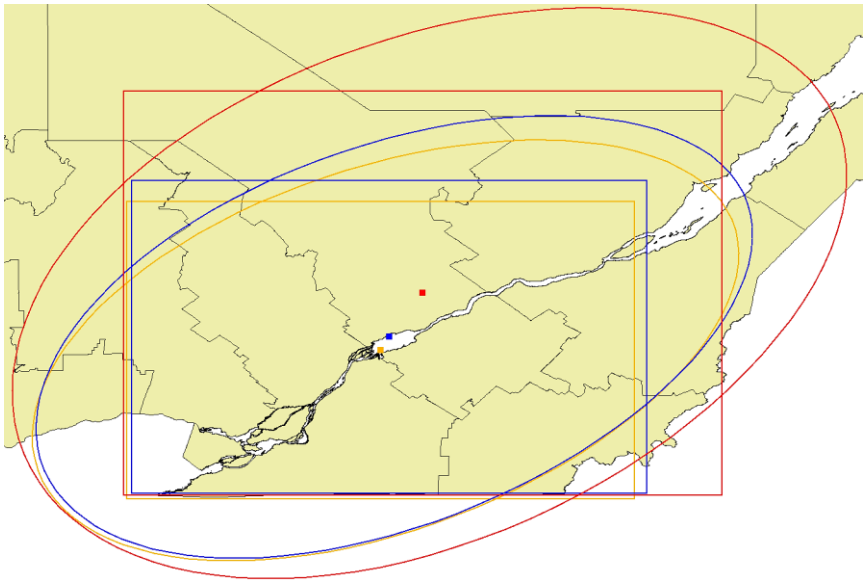


Figure 5.4 : Point Moyen, Distance type XY (rectangle), Ellipse de l'écart-type (rouge=TMD, bleu= base globale, jaune=population)

On constate immédiatement que le centre de gravité de la population du Québec et celui de la base globale sont très proches, ce qui confirme les observations précédentes, à savoir que les accidents de la base globale sont plus concentrés vers Montréal. On peut noter que l'évolution du centre de gravité de la population du Québec vers Montréal se poursuit, comme Dean et al. l'avaient remarqué sur les années 1951-1961-1971 (Dean, et al., 1974).

Le point moyen de la base TMD est décalé vers le nord-est, presque à mi-distance entre Montréal et Québec ce qui indique que les accidents TMD sont répartis de façon plus homogène entre les deux villes majeures du Québec. Ce point moyen est aussi légèrement au-dessus du Saint-Laurent, et donc que les régions du sud Québec sont un peu moins prépondérantes que pour les deux autres distributions.

Afin de mesurer la dispersion des points on a représenté la distance type, c'est-à-dire l'écart-type respectif des coordonnées X et Y qui peut se représenter par un rectangle, ou par une ellipse, qui donne des précisions sur l'orientation des distributions. On constate que la distribution la moins dispersée est la population, concentrée très majoritairement entre Montréal et Québec, la distance type cernant environ 68% du phénomène qui la constitue (Dean, et al., 1974). Les accidents de la base globale ont une distance-type très proche de celle la population, alors que ceux de la base TMD sont clairement plus dispersés. Enfin on peut noter que l'orientation des ellipses sont assez similaires et qu'elles semblent dirigées selon l'axe du Saint-Laurent.

Le point moyen, obtenu en calculant la moyenne respective des coordonnées, n'est cependant pas une mesure parfaite de la centralité, et on lui préfère généralement le centre des distances minimums qui prend mieux en compte les deux dimensions d'une carte. Ce point est par contre beaucoup plus difficile à calculer; CrimeStat utilise un algorithme itératif. Voici le résultat obtenu :

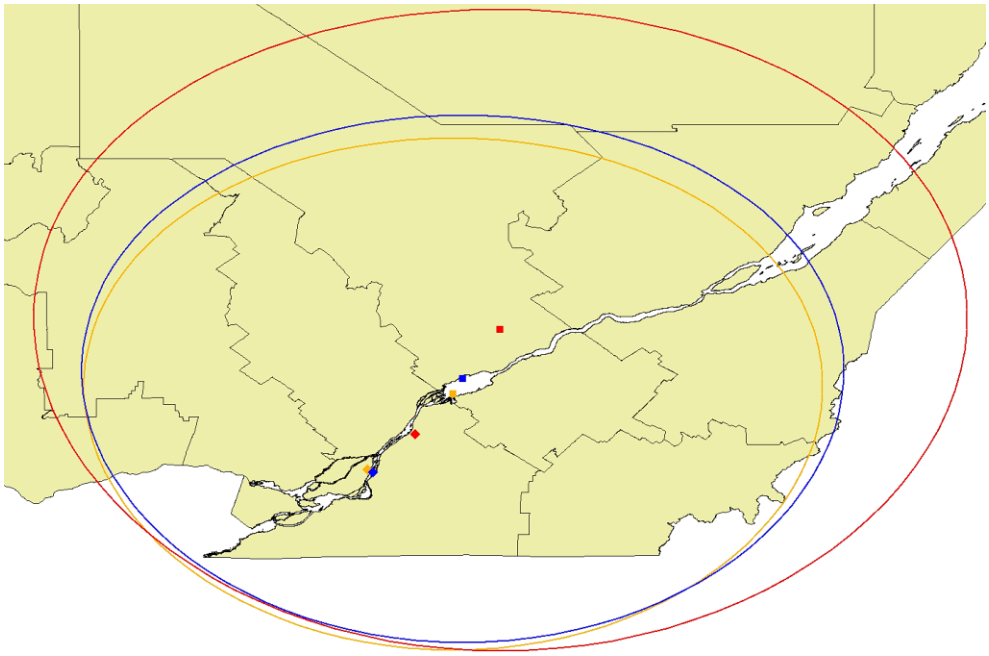


Figure 5.5 : Point moyen (carré), Centre des distance minimum (losange) et distances-types elliptiques (rouge=TMD, bleu= base globale, jaune=population)

On remarque que les centres des distances minimum de nos trois répartitions se sont déplacés vers Montréal en suivant le Saint-Laurent, ce qui prouve que la majorité des accidents se concentrent dans la région de Montréal tout comme la population. Le centre des distances minimum de la base globale et celui de la population sont maintenant presque superposés. Le centre de la base TMD s'est quant à lui déplacé de façon identique vers le sud-ouest en gardant un écart similaire, ce qui confirme nos observations précédentes.

De la même façon, les distances-types elliptiques sont cette fois calculées à partir d'écart-type de chaque point par rapport au point moyen, plutôt qu'à partir de l'écart-type des coordonnées X et Y comme précédemment. On observe toujours une dispersion plus importante des accidents de la base TMD et une dispersion similaire pour la population et la base globale, mais l'orientation des ellipses suit moins nettement le Saint-Laurent. Elles sont plutôt horizontales, ce qui indique une répartition des distributions selon l'axe est-ouest.

5.2.2 Évolution des points moyens selon les années

Historiquement les analyses centrographiques ont été inventées pour suivre le déplacement des populations notamment dans la cadre de la conquête de l'ouest américain (voir revue de

littérature). Dans cet esprit une analyse de l'évolution de la position des points moyens a été menée sur la période 1995 - 2007. Voici l'évolution obtenue.

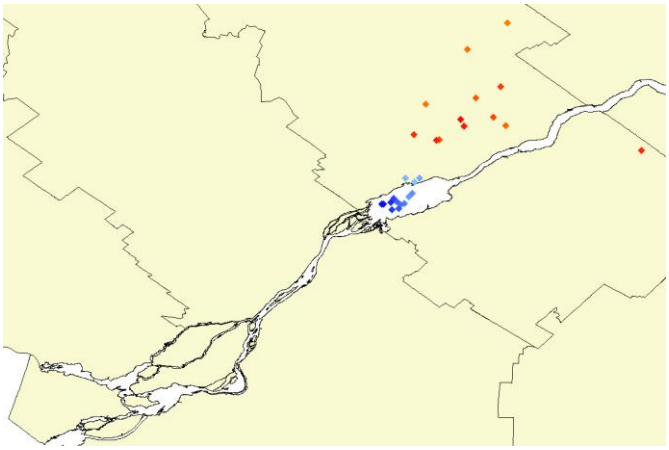


Figure 5.6 : Évolution des points moyens entre 1995 et 2007 (bleu=base globale, rouge=TMD)

Les points moyens de la base globale sont beaucoup plus « stables » que ceux de la base TMD qui « s'éparpillent » plus. Les centres moyens TMD ont toutefois toujours été plus au nord-est que ceux de la base globale. Afin de voir s'il y a une tendance dans l'évolution de ces points voici un zoom sur leur domaine :

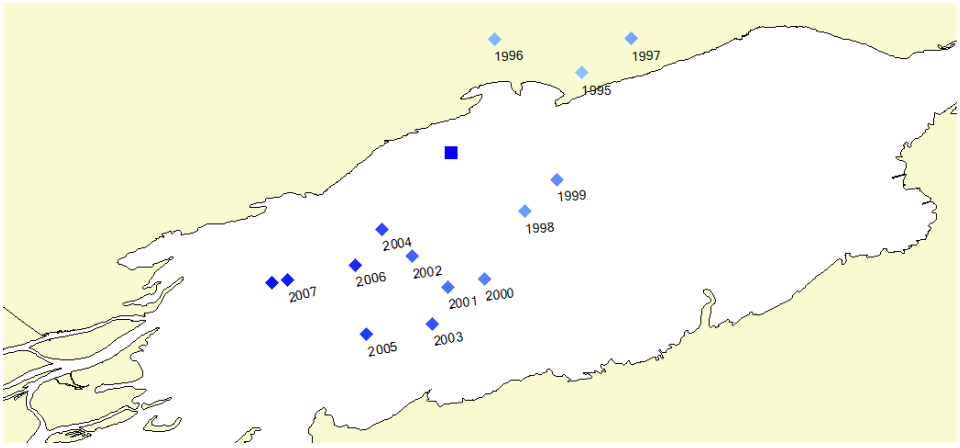


Figure 5.7 : Évolution des points moyens de la base globale entre 1995 et 2007

On constate que les centres de gravité de la base globale se déplacent vers le Sud-Ouest, vers Montréal. Ces variations ne sont cependant pas très importantes, presque tous les points étant contenus dans le lac Saint-Pierre.

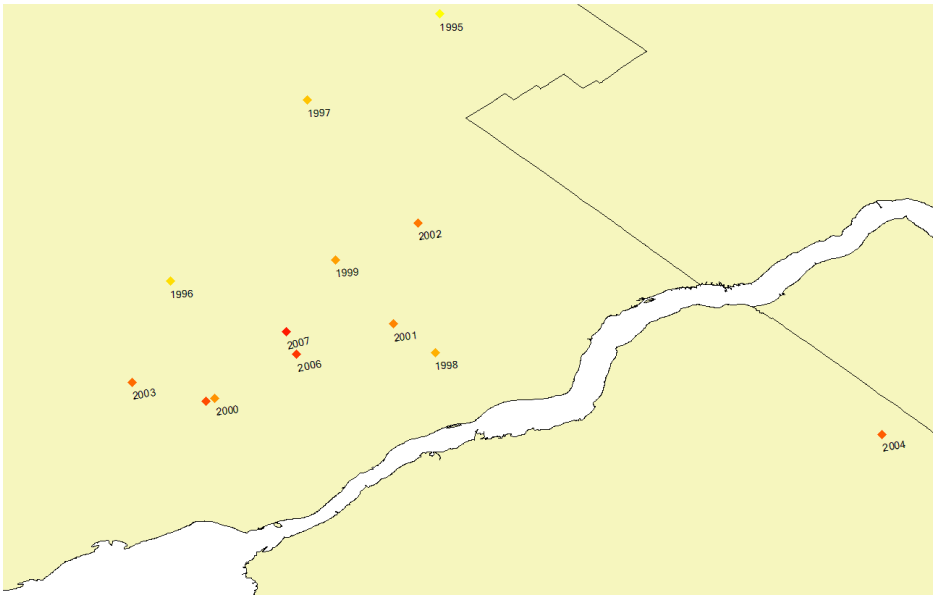


Figure 5.8 : Évolution des points moyens de la base TMD entre 1995 et 2007

Les points moyens de la base TMD sont répartis sur une plus grande surface, et ne semblent pas forcément se rapprocher de Montréal comme en témoigne la position des points des années 1996 et 2007. Cela semble beaucoup plus aléatoire que pour la base globale. On peut aussi remarquer que tous les points se trouvent au nord du Saint-Laurent, à l'exception du point de l'année 2004 qui est au sud du fleuve.

Cela nous a permis de visualiser l'écart entre les points-moyens de chaque distribution pour chaque année. Il peut être intéressant de visualiser l'écart entre les points eux-mêmes.

5.2.3 Analyse des plus proches voisins

Une analyse des plus proches voisins va nous permettre de quantifier les écarts entre les points de nos deux distributions. Voici les résultats de l'analyse des plus proches voisins réalisée à l'aide de CrimeStat :

Tableau 5.1 : Résultats de l'analyse des plus proches voisins (ordre 1)

	Véhicules lourds	TMD
Distance min	0 m	0 m
Distance max	1 928 338,93 m	1 693 064,52 m
$d(NN)$	24,498 m	2248,7045 m
$d(ran)$	1770,86 m	15440,0199 m
NNI	0,01383	0,14564 m
Standard Error	1,81 m	166,70 m
Test Statistique (Z)	-965,99	-79,13
p-value (one tail)	0,0001	0,0001
p-value (two tail)	0,0001	0,0001

La distance minimale moyenne entre deux accidents de véhicules lourds est de 24,5 m, elle est de 2 250 m pour les accidents TMD. Il est cependant difficile de comparer ces deux valeurs au vu de la différence d'ordre de grandeurs des nombres d'accidents. Le NNI permet de comparer ces deux répartitions. On obtient pour les deux distributions un NNI largement inférieur à 1, ce qui indique que les distribution sont, à l'échelle du Québec, beaucoup plus rapprochées que ne le serait une distribution aléatoire sur ce territoire. Les valeurs des deux « p-value » montrent par ailleurs, que cette différence par rapport à une distribution parfaitement aléatoire est significative pour les deux bases de données. Ce résultat, tout à fait prévisible, traduit le fait que les accidents se concentrent, comme on l'a déjà montré, dans la partie habitée du Québec. Il est par contre intéressant de comparer les deux valeurs du NNI obtenues :

$$\frac{NNI(TMD)}{NNI(Acc_véhi_lourd)} = \frac{0{,}14564}{0{,}01383} \approx 10{,}5$$

Les accidents TMD apparaisent donc près de 10,5 fois plus « éparpillés » que les accidents de la base globale qui sont eux très proches les uns des autres. Enfin les courbes d'évolution pour des ordres supérieurs confirment cette tendance pour les Kième voisins, avec même une augmentation de l'écart jusqu'au cinquième voisin (le temps de calcul très long pour la base globale ne permet pas d'aller en un temps raisonnable plus loin que le $5^{ème}$ voisin) :

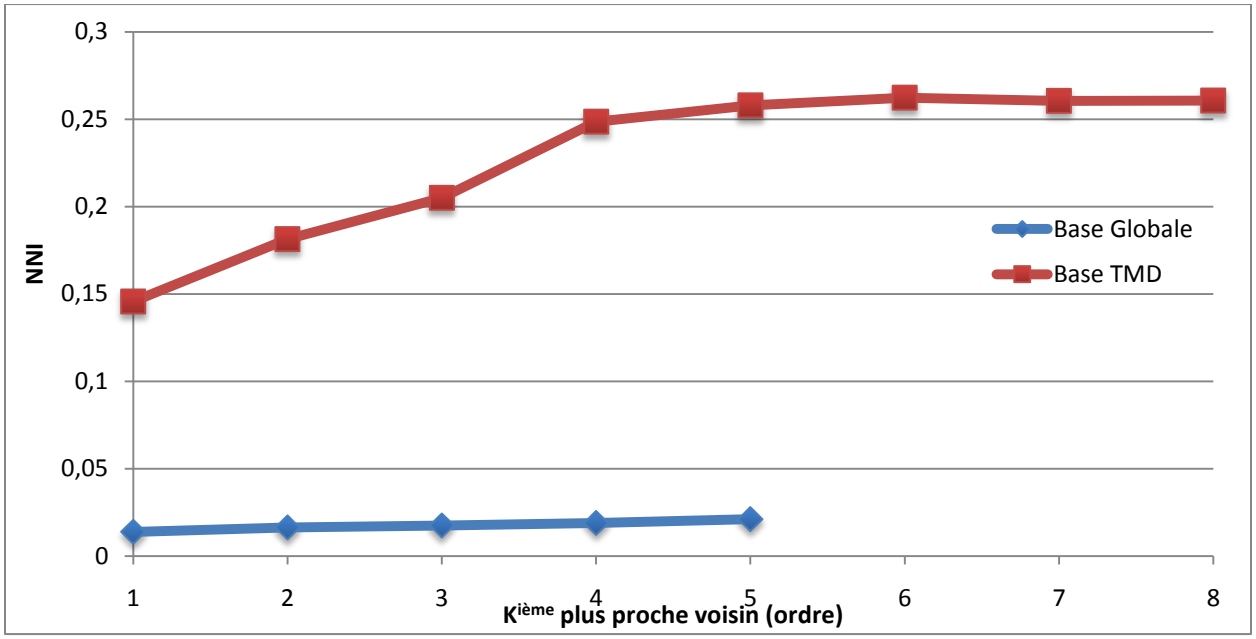


Figure 5.9 : Indice des plus proches voisins (NNI) en fonction de l'ordre K

Cette étude montre que les accidents de la base globale sont plus proches les uns des autres que les accidents de la base TMD. Voyons maintenant si des regroupements remarquables existent.

5.2.4 Analyse des points chauds

Cette section présente les résultats de deux méthodes permettant de déterminer les points chauds (Hot Spot) des deux répartitions d'accidents.

5.2.4.1 Nearest Neighbor Hierarchical Clustering (NnH)

Dans cette analyse il faut définir le nombre minimal d'accidents pouvant former un cluster. Afin d'obtenir des résultats comparables, le nombre minimum de points par cluster a été défini dans un premier temps pour la base TMD puis adapté à la base globale de façon proportionelle à la taille de l'échantillon. Toutefois le calcul n'a pas pu être mené à son terme par Crimestat avec les 250 000 accidents de la base globale. Pour faire malgré tout ce calcul la taille de cet échantillon a été réduite en choisissant de façon aléatoire 35 000 accidents de la base globale. Les valeur des paramètres d'analyse centrographique obtenus sont assez proches, ce qui permet de supposer que cet échantillon est bien représentatif de l'ensemble. Enfin le calcul a été mené avec un nombre minimum de points « petit » et « grand ». Voici quelques résultats numériques qui ont été obtenus :

Tableau 5.2 : Quelques résultats numériques de l'analyse NnH

	Base globale		TMD	
Taille de l'échantillon	35 000		2 344	
Nombre minimum de points par cluster	76	303	5	20
Nombre de cluster de niveau 1	57	16	81	22
Nombre de cluster de niveau 2	3	1	8	1
Nombre de cluster de niveau 3	0	0	1	0
Pop max dans cluster de niveau 1	9 117	9 117	458	458
Pop min dans cluster de niveau 1	77	330	5	21
∑ des accidents « clusterisé »	24 966 (71%)	18 573 (53%)	1 964 (84%)	1 417 (60%)
Nb moyen d'accidents par cluster	438	1 160,8	24,25	64,4

On constate que pour les deux nombres minimums de points, le nombre de clusters de niveau 1 est plus important pour la base TMD, ce qui se retrouve dans la proportion d'accidents regroupés dans ces clusters, puisqu'il y a environ 10% d'accidents de plus dans les clusters de la base TMD. Ce résultat est assez étonnant puisque les analyses précédentes avaient montré que les accidents de la base globale semblaient plus proches et donc plus facilement « clusterisables », ce qui n'est pas le cas.

On constate par ailleurs que plus le nombre mininum de points par cluster est faible, plus le pourcentage d'accidents associés à un cluster est important. La différence entre notre « petit » et notre « grand » nombre minimum de points est de l'ordre de 20% d'accidents clusterisés en moins pour les deux bases de données.

Enfin on peut noter que seul le calcul pour un nombre petit de point minimum par cluster de la base TMD a donné lieu à un cluster de niveau 3. Voyons maintenant où sont situés sur une carte ces clusters :

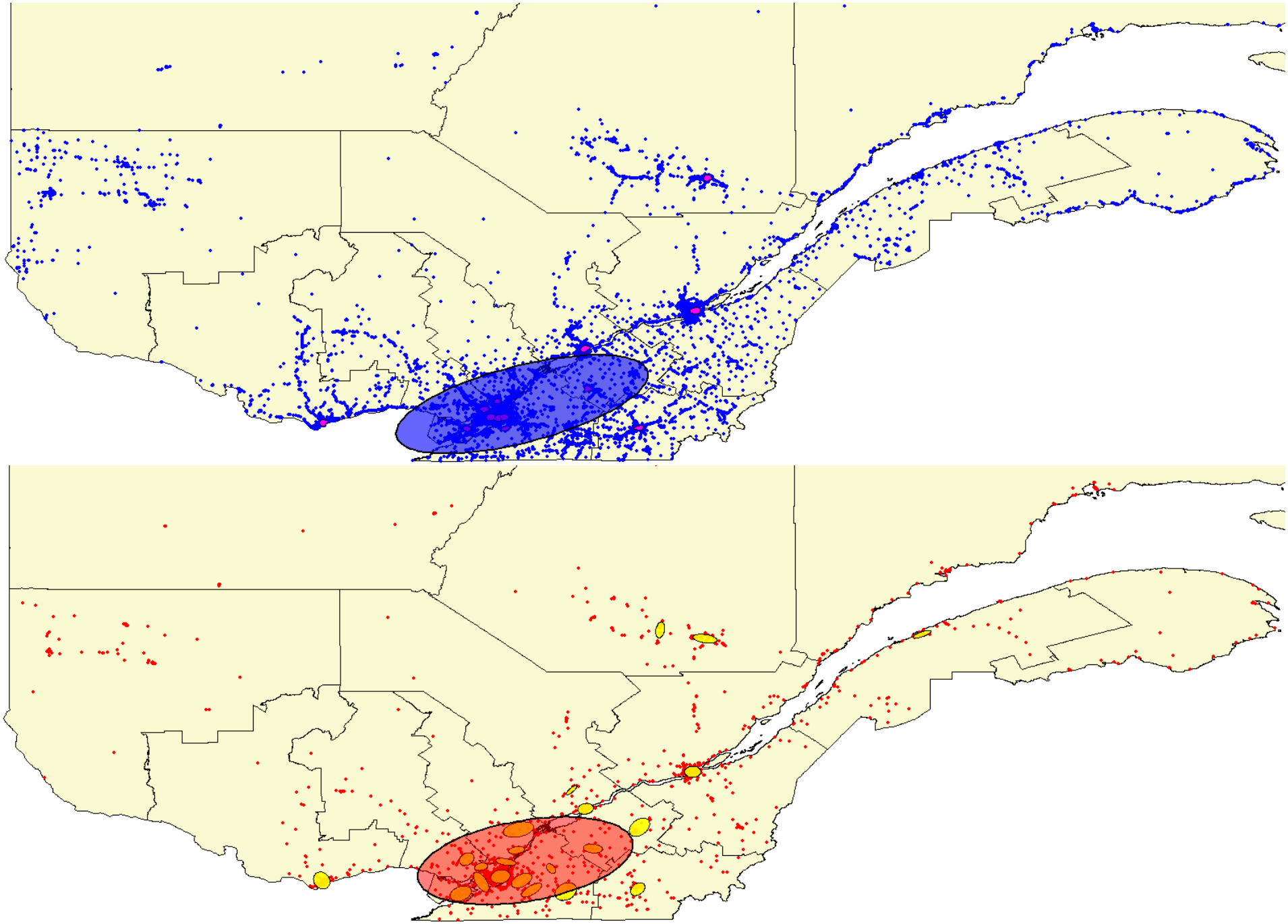


Figure 5.10 : Nearest Neighbor Hierarchical Clustering pour un nombre minimal d'accidents « grand » (bleu = base globale, rouge = TMD)

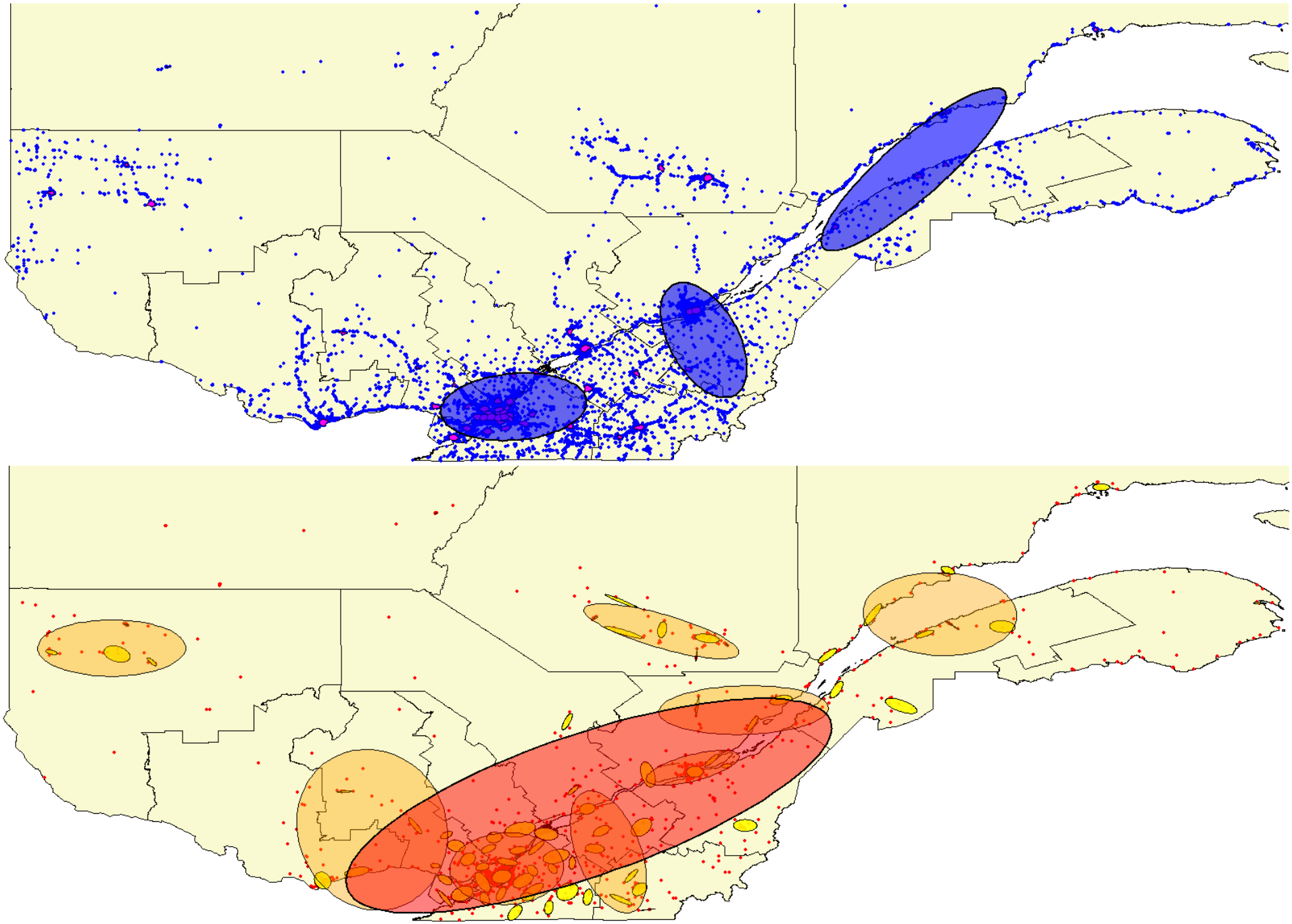


Figure 5.11 : Nearest Neighbor Hierarchical Clustering pour un nombre minimal d'accidents « petit » (bleu = base globale, rouge = TMD)

Pour un nombre minimum de points par cluster plutôt « grand » (premier couple de carte), on obtient un cluster de niveau 2 pour les deux répartitions d'accidents. Ces deux clusters sont de taille et d'orientation similaires, couvrant la région de Montréal, une bonne partie de la Montérégie et la moitié de la région du Centre-du-Québec. De plus les clusters de niveau un semblent aussi correspondre globalement. Cela montre qui si on regarde « en gros » ces deux répartitions, elles ont de nombreux points communs, le plus important étant que les accidents de véhicules lourds, qu'ils transportent ou non des matières dangereuses, ont majoritairement lieu dans la région de Montréal.

Le second couple de carte permet d'observer les deux répartitions de façon plus fine avec un nombre minimum d'accidents par cluster plus petit, et donc plus de clusters. Il y a maintenant trois clusters de niveau 2 pour la base globale contre 8 pour la base TMD. Cela montre bien que les accidents impliquant des matières dangereuses sont mieux répartis sur le territoire, les clusters de niveau 2 couvrant l'ensemble des grands centres d'habitation du Québec, contrairement aux clusters de la base globale. Il est par ailleurs assez étonnant de constater que la plupart des clusters de niveau 2 se trouvent à cheval sur les deux rives du Saint-Laurent, et donc qu'une zone accidentogène sur une rive l'est aussi au même niveau de l'autre côté. De plus le cluster de niveau 3 obtenu par la base TMD couvre l'ensemble de la vallée du Saint-Laurent quasiment jusqu'au fjord du Saguenay.

Les cluster de niveau 1 correspondent la plupart du temps aux plus grosses municipalités québecquoises. Il n'y a que Montréal qui regroupe plusieurs clusters de niveau 1 sur son territoire, notamment pour la base globale, comme le montre le zoom suivant :

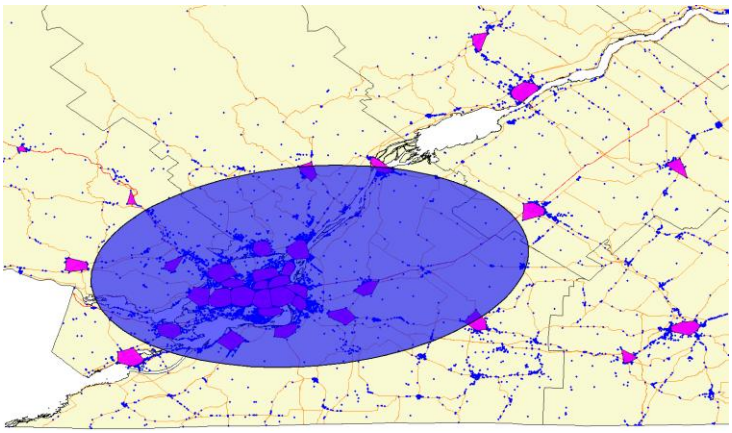


Figure 5.12 : Zoom sur la région de Montréal du NNH de la base globale pour un nombre minimum d'accidents par cluster « petit »

Enfin on constate que les groupements de niveau 1 de la base globale sont beaucoup plus petits que leurs homologues de la base TMD. Cela est dû, d'une part à une distance seuil calculée par CrimeStat plus petite en raison du grand nombre d'accidents, mais aussi à la méthode moins précise de localisation de ces accidents qui favorise des concentrations importantes d'accidents en un point.

5.2.4.2 K-Means Partitioning Clustering

Contrairement à l'approche précédente, il faut dans cette méthode choisir le nombre K de clusters à placer, que le logiciel localisera au mieux en leur assignant TOUS les points. Le nombre minimal de cluster est de 2 et il s'est avéré que le nombre maximum possible de cluster pour la base TMD est de 45. Comme précédemment il a été choisi de comparer les résultats pour un nombre de clusters petit (5) et grand (30). Voici quelques résultats numériques obtenus pour ces deux calculs :

Tableau 5.3 : Quelques résultats numériques du K-Means clustering

	Base globale		TMD	
Taille de l'échantillon	262 172		2 344	
Valeur de K	5	30	5	30
Pop max dans un cluster	141552 (53%)	95124 (36%)	988(42,1%)	985 (42%)
Pop min dans un cluster	19778 (7,5%)	1138 (0,4%)	204 (8,7%)	2 (0,08%)
Nb moyen d'accidents par cluster	52 434,4	8 739,07	468,8	78,13

On constate que pour la base globale la valeur de K a une grande influence sur le nombre d'accidents maximum dans un cluster. Elle est de 53% lorsque K=5, ce qui énorme, pour fondre à 36% lorsque K=30. Dans le même temps ce chiffre est resté constant à 42% dans la base TMD. Cela prouve une nouvelle fois que la répartition des accidents TMD est plus étalée sur le territoire. On peut aussi remarquer que pour k=30, CrimeStat a généré de tout petits groupes avec seulement 2 accidents, ce qui montre que 30 est vraiment un « grand » nombre de clusters pour une répartition de cette taille. Il aurait été possible d'utiliser un K beaucoup plus grand pour la base globale, mais il n'aurait pas été possible de comparer les résultats avec la répartition TMD et cela n'a donc pas été fait.

Voici maintenant les clusters obtenus pour K=5 (La représentation par leurs enveloppes convexes a été préferée aux ellipses, afin de bien montrer que tous les accidents sont clusterisés) :

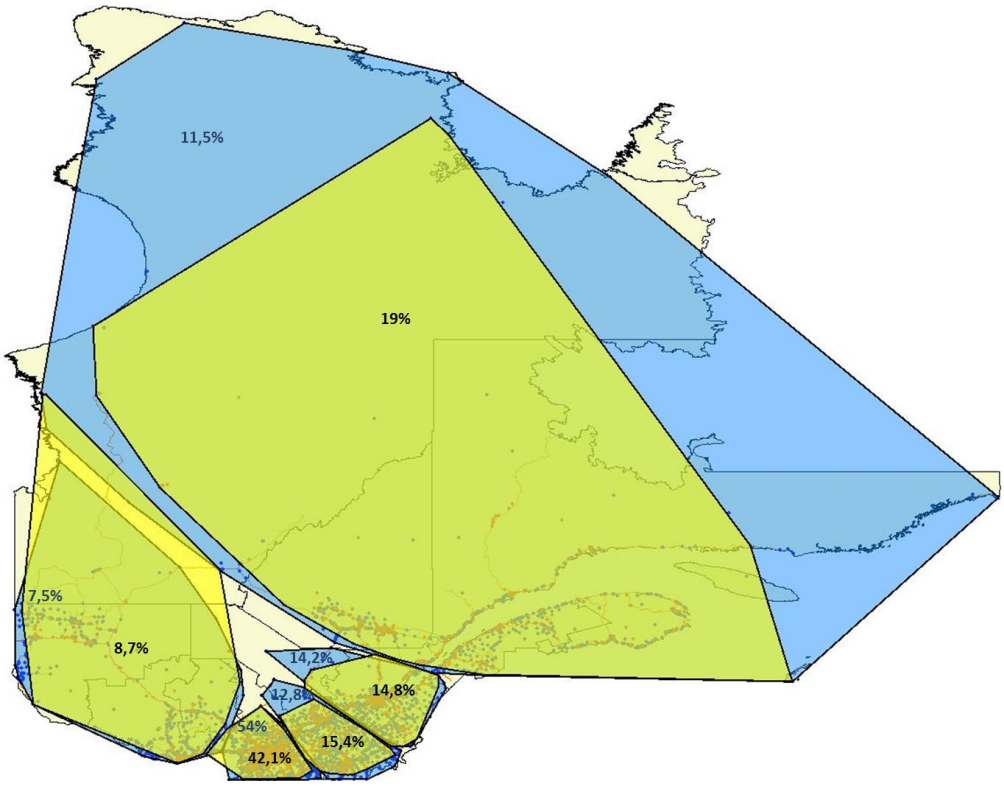


Figure 5.13 : K-Means Clustering pour K=5 (bleu=base globale, jaune=TMD)

On constate immédiatement les similitudes entre la forme et la position des cinq clusters. Dans les deux cas un premier cluster couvre le Nord-du-Québec, la Côte-Nord, le fjord du Saguenay et la Gaspésie, trois autres plus petits semblent descendre le Saint-Laurent de Québec à Montréal et un dernière couvre la région de Gatineau et l'Abitibi.

Pour chaque cluster le pourcentage d'accident qu'il contient est indiqué. On constate que globalement les pourcentages sont proches, ce qui confirme l'observation précédente de l'analyse NnH pour un nombre minimal d'accidents par cluster « grand » : de loin, les deux répartitions semblent similaires. On peut toutefois noter que les accidents de la base globale sont plus concentrés aux environs de Montréal avec 54% des accident contre 42% pour la base TMD et moins dans le Nord Est du Québec avec 11,5% des accidents contre 19% pour la base TMD. Cette forte concentration dans ce grand cluster vient, comme on l'a montré dans la réparition par région économique, des régions de la Côte-Nord et du Saguenay essentiellement.

Voici maintenant les enveloppes convexes des clusters obtenues pour k=30 :

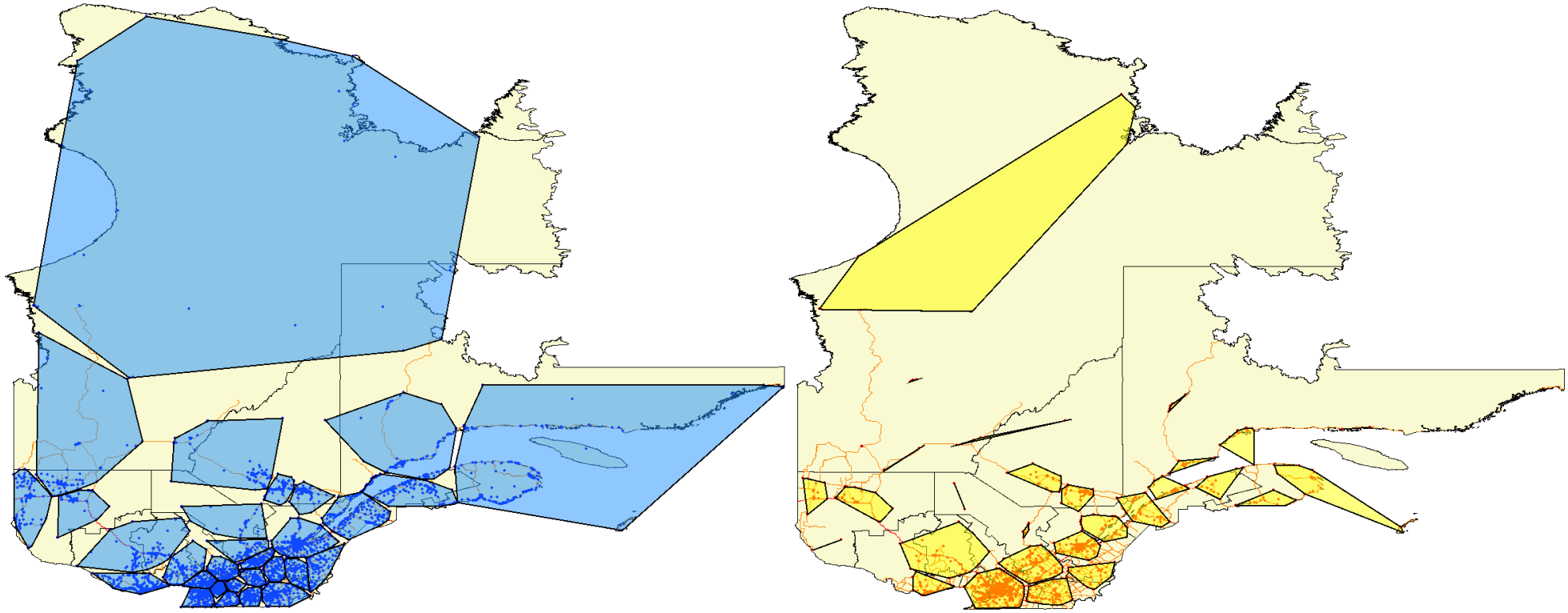


Figure 5.14 : K-Means Clustering pour K=30 (à gauche résultat pour la base Globale, à droite pour la base TMD)

On constate maintenant avec 30 clusters des différences notables entre les groupements réalisés par CrimeStat. Tous d'abord les clusters de la base globale couvrent de façon plus large l'ensemble du territoire québecois que ceux de la base TMD. On voit très bien que pour trouver 30 clusters dans la base TMD CrimeStat a été obligé de générer des groupements « tout petits » avec seulement quelques accidents, et au final des clusters couvrant une surface toute petite. Le découpage obtenu dans la région de Montréal est beaucoup plus fin dans la base globale et même sur l'ensemble de la vallée du Saint-Laurent avec très peu de « trous ». Ces cartes viennent nuancer ce que l'on a dit précédement concernant la répartition des accidents. En effet si les accidents de la base globale semblent se concentrer majoritairement dans la région de Montréal, il n'en demeure pas moins qu'ils concernent une plus grande partie du Québec que les accidents TMD.

5.2.5 Analyse des densités

Les analyses précédentes ont montré qu'il semble que les accidents de véhicules lourds se concentrent dans la région de Montréal notamment pour la base globale. Afin d'estimer de déterminer précisement les zones de concentration des accidents, nous allons mener une analyse des densités selon la méthode des noyaux tel que proposé par CrimeStat.

Les deux cartes ci-dessous (figure 5.15) présentent les densités obtenues, selon des échelles comparables, ainsi que les clusters de type NnH qui permettent de souligner les zones à forte densité d'accidents. On constate à nouveau que globalement, à l'échelle du Québec, les deux distributions sont assez similaires, bien qu'encore une fois il apparait clairement que les accidents de la base globale sont beaucoup plus concentrés aux alentours de Montréal. D'ailleurs, trois ellipses encadrent le pic de densité, quand une seule suffit pour la base TMD. De plus, il y a de plus nombreuses zones jaunes sur la carte TMD, ce qui traduit des densités comparables d'accidents en de nombreux endroits, quand la densité aux environs de Montréal est, pour la base globale, inégalable si ce n'est pour les principaux pôles déjà cités.

Sur les deux cartes suivantes (figure 5.16) présentent un zoom sur la région entre Montréal et Québec. Pour un résultat plus précis le maillage de calcul a été réduit à cette zone et découpé de façon beaucoup plus fine. On distingue sur les deux cartes les mêmes zones principales, Montréal, Québec, Trois-Rivières, Sherbrooke et Drummondville. Mais Victoriaville apparait également en jaune sur la carte TMD. De plus, en regardant plus précisemment la région de Montréal, on constate que les densités sont légérement différentes. Dans le cas de la base globale on distingue des cercles concentriques tant la densité augmente uniformément en s'approchant de la métropole. Les couleurs sont réparties de façon bien moins homogène pour la base TMD, et le jaune couvre une plus grande superficie. Enfin la zone de Québec est d'un orange soutenu, alors que pour la base globale, on note à peine une différence, la densité pour la base globale par rapport à Montréal étant plus faible que la base TMD.

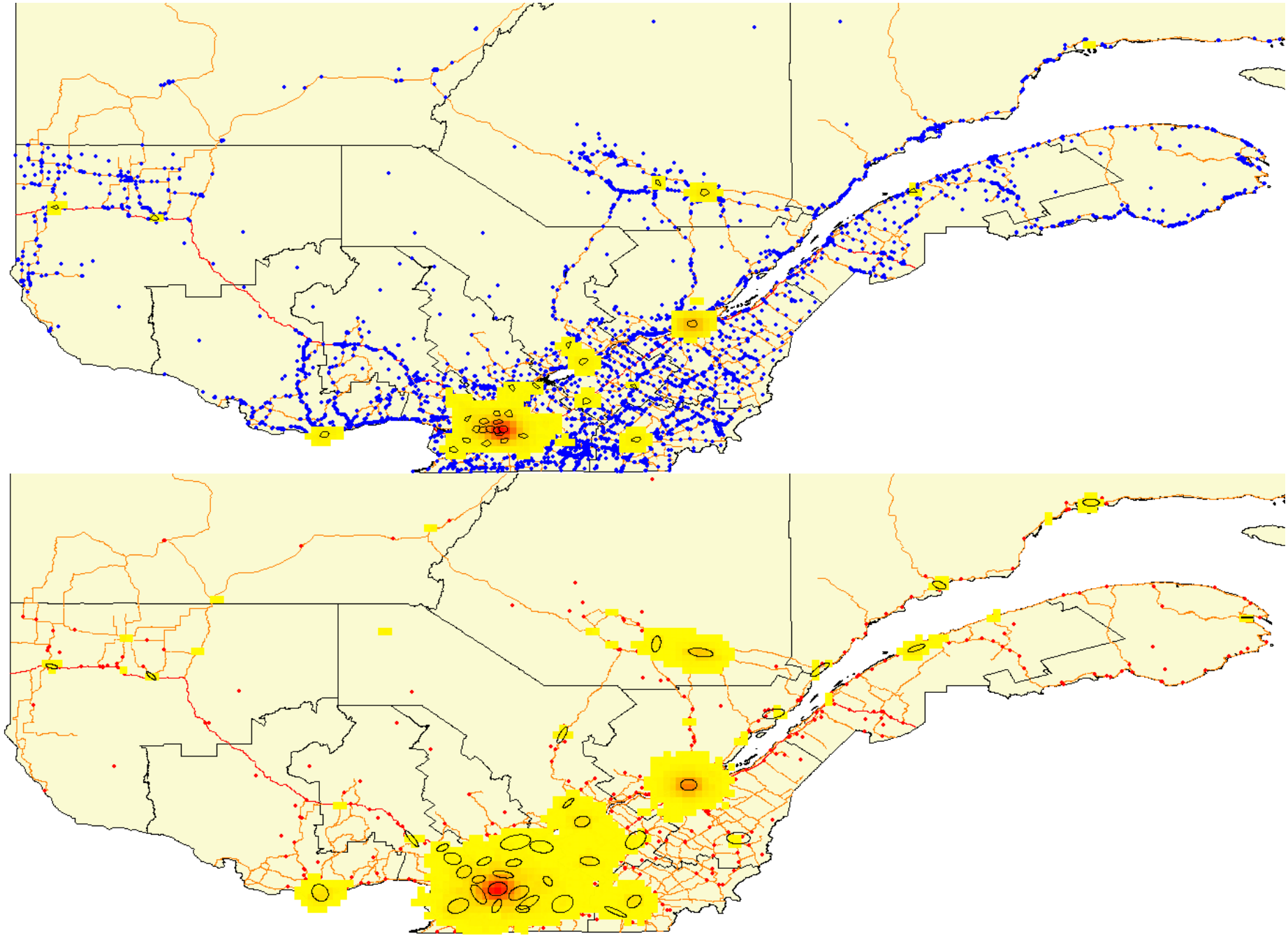


Figure 5.15 : Estimation de la densité par la méthode du noyau, échelle comparable (bleu=base globale, rouge=base TMD)

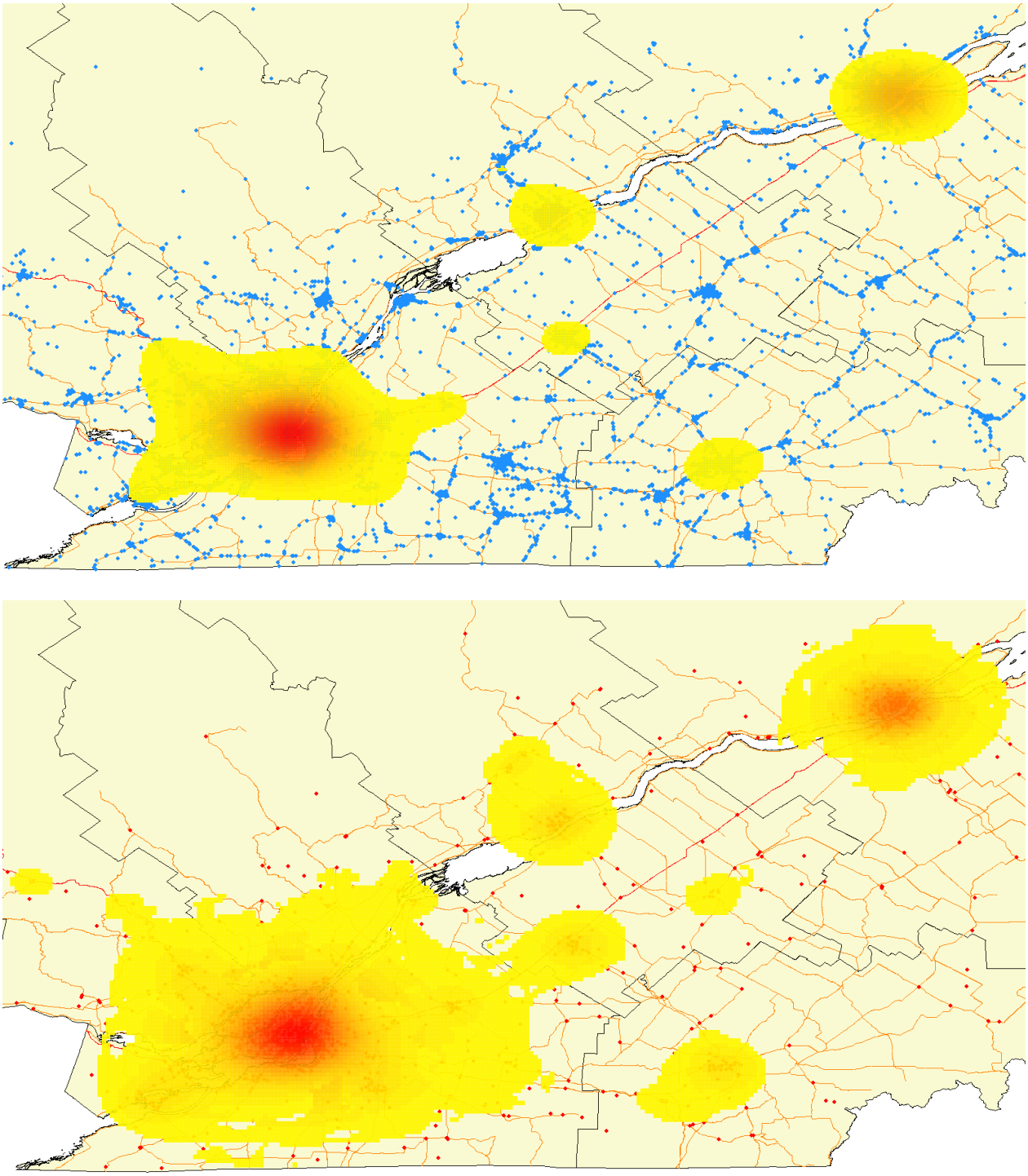


Figure 5.16 : Estimation de la densité par la méthode du noyau pour la région entre Montréal et Québec (bleu=base globale, rouge=base TMD)

5.3 Analyse pondérée des répartitions d'accidents

Après avoir analysé la simple répartition des accidents, nous allons nous intéresser à la répartition spatiale de certaines caractéristiques quantitatives des accidents. Pour cela des calculs équivalents à ce qui été fait précédement seront menés en incorporant une pondération par la caractéristique étudiée. Les caractéristiques étudiées sont le nombre de véhicules impliqués et le nombre de décés et de blessés.

5.3.1 Analyse spatiale du nombre de véhicules impliqués

Les deux cartes de la page suivante, présentent les densités pondérées ainsi que le centre moyen et l'ellipse d'écart type pondéré obtenue. Les accidents sont aussi présentés, avec un code de couleur fonction du nombre de véhicules impliqués. On distingue nettement, comme cela a été prouvé dans l'analyse statistique, que la majorité des accidents ont impliqué 1 ou 2 véhicules.

Il ne semble pas y avoir de tendance particulière sur cette répartition. On retrouve les mêmes zones à forte densité que précédemment, c'est-à-dire essentiellement Montréal et Québec. Pour la base TMD il y a là aussi plus de zones de densité moyenne que pour la base globale.

On constate par ailleurs que le point moyen se trouve quasiment au même endroit que pour le calcul non pondéré. Pour la base TMD en revanche, le point moyen est situé plus haut, dans les environs de Shawinigan, alors qu'il était précédement à ce niveau dans le Saint-Laurent. Cela indique que les accidents TMD situés au plus au nord, impliquent en moyenne un peu plus de véhicules.

En ce qui concerne les deux ellipses d'acrt-type elles sont similaire en taille et sont toujours orientées dans l'axe du Saint-Laurent.

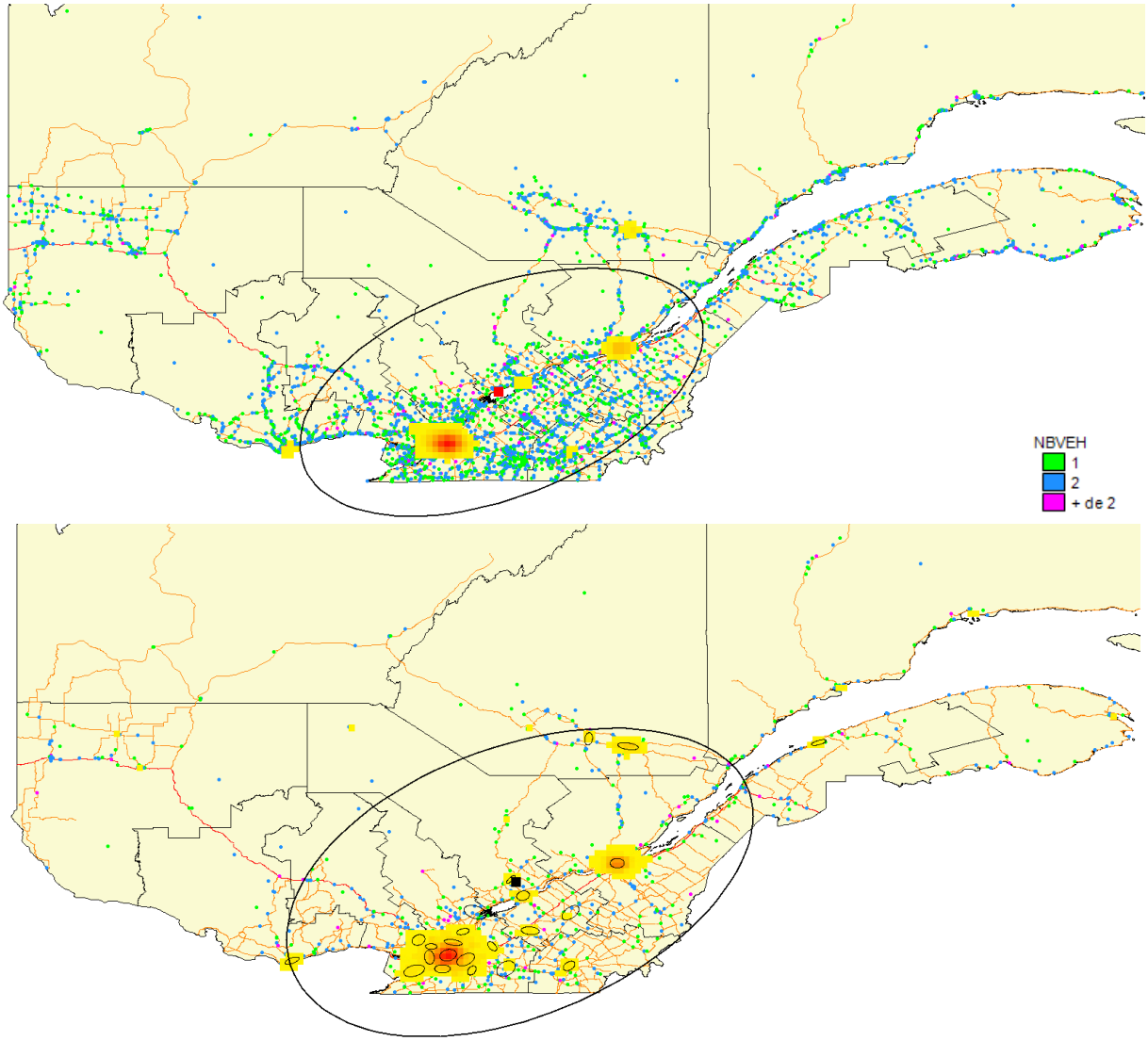


Figure 5.17 : Densité des accidents, point moyen et ellipse d'écart-type pondéré par le nombre de véhicules impliqués (haut=base globale, bas=base TMD)

5.3.2 Analyse spatiale de la répartition des décès

Voici maintenant les deux cartes similaires obtenues en pondérant les accidents par le nombre de décès :

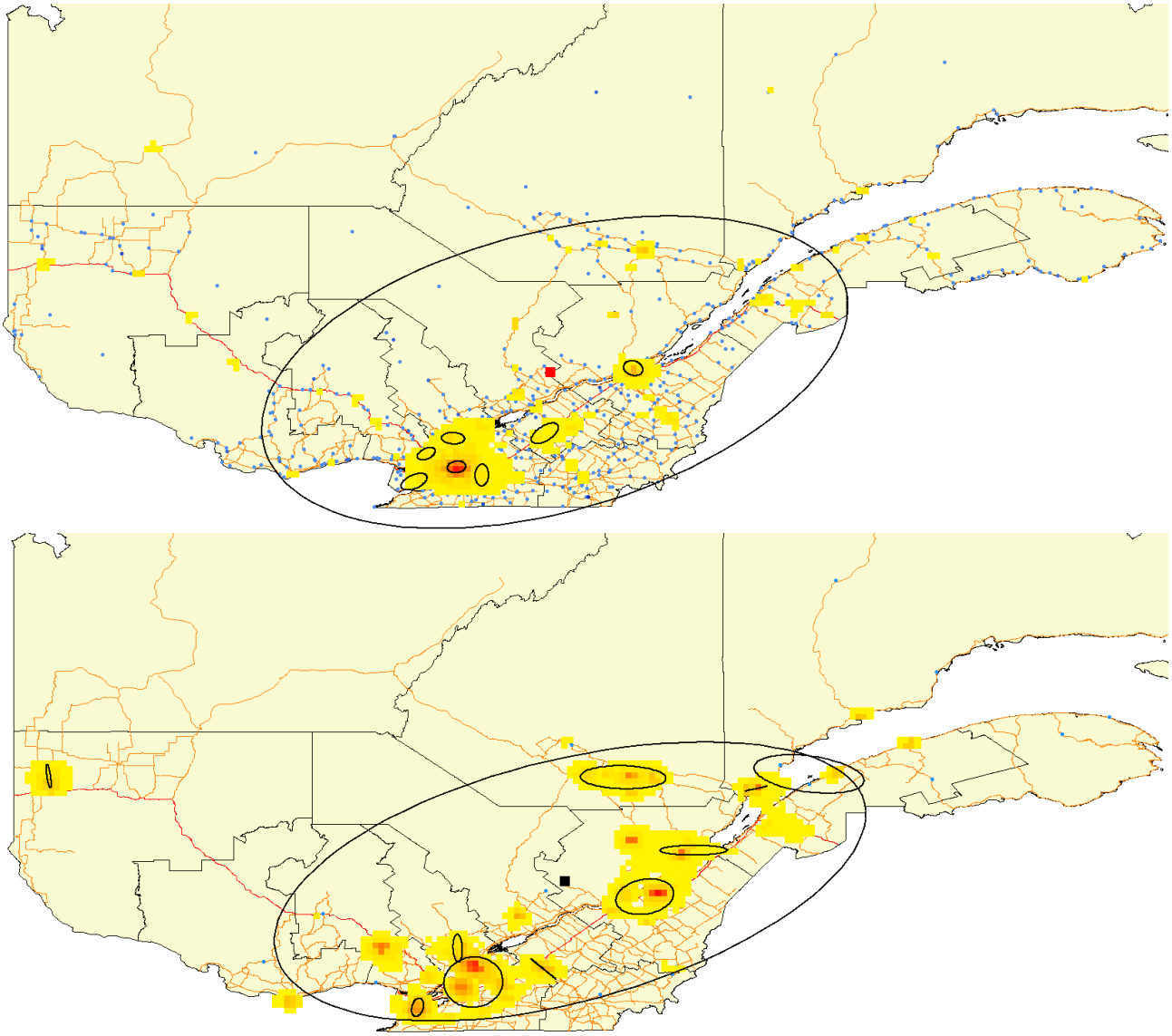


Figure 5.18 : Estimation de la densité par la méthode du noyau pour le nombre de décès (haut=base globale, bas=base TMD)

On constate de nombreuses différences avec les cartes précédentes non pondérées. On retrouve une zone de forte densité dans la région de Montréal sur les deux cartes, mais le pic est légèrement en-dessous de la ville pour la base globale et largement au-dessus pour la base TMD, même si l'on remarque une zone orange au même endroit que le maximum de la base globale. De plus dans le cas de la base TMD, cette répartition est très hétérogène, quand elle reste concentrée sur la première carte. Par ailleurs, il y a pour les deux répartitions de nombreuses petites zones jaunes un peu partout, ce qui montrent que les morts peuvent avoir lieu n'importe où, y compris dans des endroits peu accidentogènes, ce qui se tradauit par des ellipses d'écart-type couvrant une

plus grande surface. Les points moyens sont aussi tout les deux décalés vers le nord, et sont beaucoup plus proche que pour les répartitions non pondérée.

Pour la base globale ils semblent qu'il y ait beaucoup de morts dans la région de Québec avec des densité aussi étendue que pour Montréal, alors qu'on le sait il y a moins d'accidents dans cette zone. D'autre zones, assez peu peuplées et souvent absentes des cartes précédentes, qui en général correspondant entre les deux bases de données, semblent aussi plus dangereuses. On peut citer par exemple la route 175 au-dessus du parc de la Jacques-Cartier, la 381 dans les environs du parc des Grands-Jardins (il s'agit aux abords de ces parcs de routes plus sinueuses, avec des pentes plus importantes, et elles sont peu fréquentées, ce qui peut en partie expliquer ce fait), l'embouchure du fjord du Saguenay, Rivière-du-Loup, la 138 au niveau de Baie-Comeau et la 132 à Matane.

5.3.3 Analyse spatiale de la répartition des blessés

Enfin pour terminer cette étude voici, ci-dessous, les deux cartes de la répartition des blessés. Elles se rapprochent des cartes vues précédemment, alors qu'on aurait pu penser qu'elles seraient similaires à celles des décès. Comme pour la carte des véhicules impliqués, les densités observées correspondent aux zones de forte concentration d'accidents. Donc, plus il y a d'accidents à un endroit, plus le nombre de blessés est important. On retrouve ainsi la zone de forte concentration de Montréal, plus étendue pour la base TMD, avec une ellipse entourant le pic alors que la présence de plusieurs petites montrent la densité proportionnellement plus importantes du pic de la base globale. On remarque que la base globale a peu d'autres zones de densité comparable aux environs de Montréal, alors que de nombreuses tâches jaunes indiquent une répartition plus homogène des blessés dans les sites accidentogènes de la base TMD. Les centres moyens sont différents, pour la base globale dans l'axe Montréal-Québec, mais plus proche de la métropole et pour la base TMD plus proche de Québec et au-dessus du Saint-Laurent, l'ellipse associée couvre donc une surface plus importante.

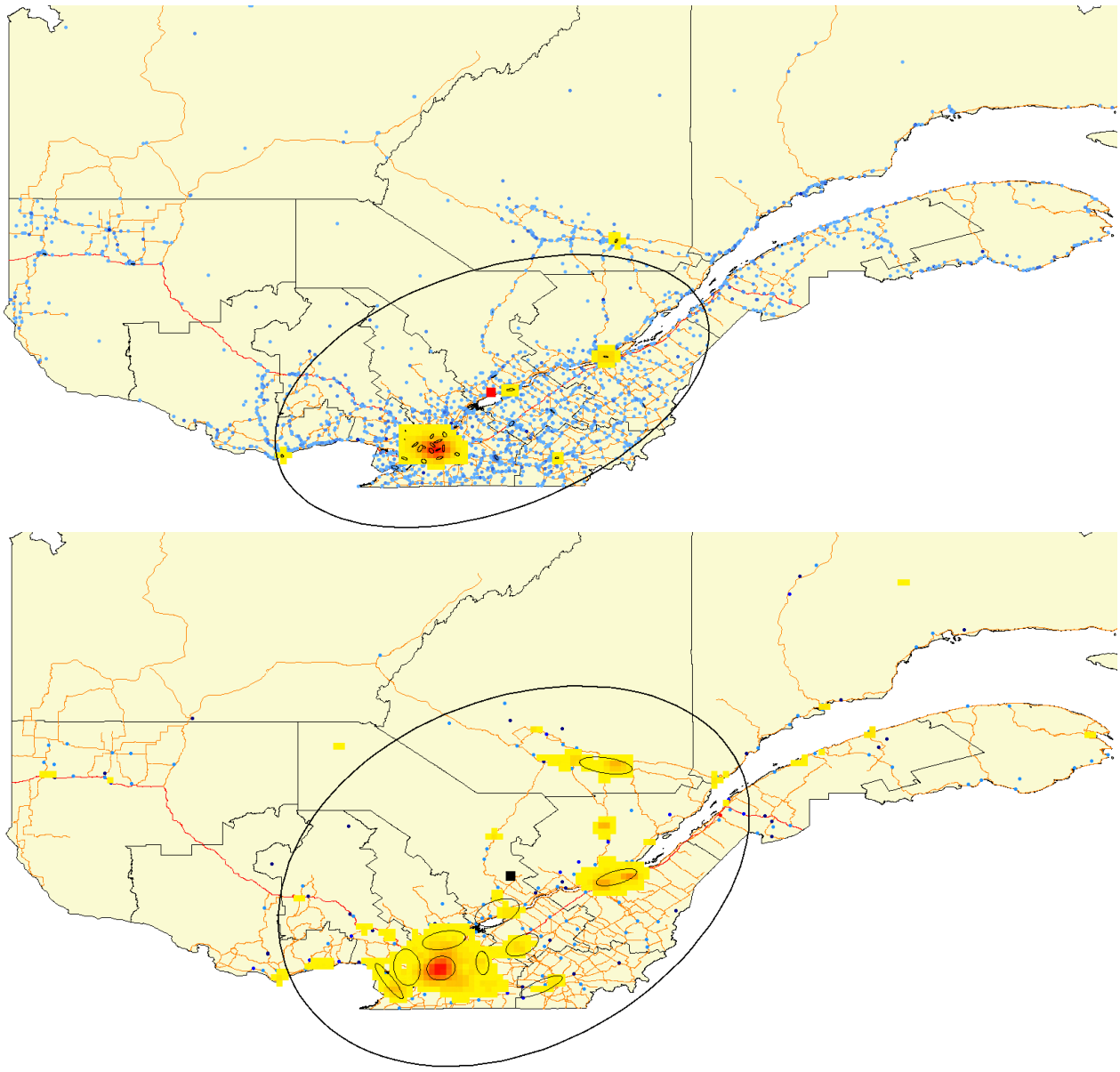


Figure 5.19 : Estimation de la densité par la méthode du noyau pour le nombre de blessés (haut=base globale, bas=base TMD)

On constate que ces analyses pondérées ne présentent pas tout à fait les mêmes résultats. Les cartes des véhicules impliqués et des blessés sont similaires et montrent qu'il n'y a pas de tendance particulière sur la répartition, les zones de fortes densités correspondants aux endroits de forte concentration d'accidents. En revanche la répartition des décès souligne, en plus des zones prévisibles, des sites particulièrement dangereux, pour la plupart communs aux deux distributions. Cela clôt l'étude, nous allons maintenant récapituler l'ensemble des résultats obtenus.

CONCLUSION

Cette analyse comparative des accidents impliquant des véhicules lourds avec ou sans transport de matières dangereuses, nous a permis de décrire de nombreux aspects sur l'accidentologie des véhicules lourds au Québec avec ou sans matières dangereuses. Cette étude s'appuie sur deux bases de données recensant d'une part les accidents impliquant les véhicules lourds et d'autre part les accidents impliquant des véhicules de transport de matières dangereuses.

Tous d'abord tout un travail de mise en forme et de préparation des données a été réalisé. Les accidents de la base TMD ont été identifiés dans la base globale grâce à un programme VB, et la base TMD a été complétée et consolidée. Les accidents ont été localisés, et un outil de visualisation des rapports d'accidents intégrant le positionnement de l'accident sur une carte a été mis en place dans les deux bases de données. Ces opérations contribuent à l'amélioration de la qualité de l'information et faciliteront certainement les études futures.

Le premier volet de ce travail s'est intéressé à la comparaison des différents champs communs dans ces deux bases de données en utilisant des méthodes statistiques. Tous d'abord le champ date a été étudié. L'analyse de l'évolution annuelle du nombre d'accidents n'a pas mis en évidence de tendance claire. Il semble que les années à forte proportion d'accidents de véhicules lourds soient aussi des années à fort taux d'accidents TMD, mais le lien n'est très marqué. On peut cependant noter qu'à proportion égales les variations annuelles des accidents TMD sont plus de 2 fois supérieures aux variations annuelles des accidents de la base globale. D'autre part la saison hivernale est particulièrement propice aux accidents TMD. C'est en effet durant cette saison que l'on observe les plus forts taux d'accidents TMD avec près de 34% des accidents contre seulement 26% pour la base globale. La répartition horaire obtenue est similaire entre les deux bases de données, les accidents ayant majoritairement lieu la journée avec deux pics vers 210h et 14h et une « pause » à midi.

Ensuite les champs quantitatifs ont été étudiés. Le nombre de véhicules impliqués apparait significativement plus faible dans le cas des accidents TMD avec une différence de 0,2 véhicule par accident, ce qui traduit une proportion plus élevée d'accidents impliquant 1 seul véhicule. L'analyse du champ limitation de vitesse semble montrer que les accidents TMD ont lieu sur des routes moins rapides. Cependant ce résultat est à nuancer puisque dans près de 50% des rapports d'accidents cette information est manquante. Les résultats de l'analyse sur les conséquences

humaines sont par contre plus fiables. On obtient que les accidents TMD sont en proportion plus mortels et font plus de blessés qu'un simple accident de la route. Il y en effet en moyenne environ 3 accidents TMD mortels sur 100 accident TMD contre seulement 0,6 accident mortel pour 100 accidents dans de la base globale. Les tests de proportion réalisés montrent que cette différence est significative. Enfin les courbes p-N montrent que la probabilité d'avoir un grand nombre de personnes impliquées est plus élevée pour un accident TMD, que pour un accident « normal ».

Cette analyse statistique s'est terminée par l'étude de 8 champs qualitatifs. Ils permettent de souligner certaines différences dans les conditions ou les circonstances favorisant des accidents. Ainsi, on constate que les accidents de la base TMD sont plus souvent dus à une sortie de route que ceux de la base globale, même si le cas le plus fréquent est, dans les deux cas, une collision entre deux véhicules. De plus, plusieurs champs soulignent la plus grande dangerosité du transport MD dans des conditions hivernales déjà constaté, avec par exemple plus d'accidents par temps de neige ou sur des surfaces glacées ou enneigées que pour la base globale. Enfin, les accidents TMD ont plus souvent lieu en milieu rural ou forestier (42% des accidents) que les accidents globaux, qui ont majoritairement lieu en zone commerciale (45% des accidents).

Dans le second volet de cette étude, une analyse spatiale des répartitions a été menée. Les accidents ont été localisés de façon automatisée en utilisant les champs adresse et municipalité présents dans les rapports d'accidents. Les répartitions obtenues montrent que les accidents se concentrent dans les zones habitées du Québec particulièrement le long du Saint-Laurent. La région regroupant le plus grand nombre d'accident TMD est la Montérégie avec 20% des accidents TMD; pour la base globale c'est Montréal qui concentrent une grande majorité d'accidents (26% !) sur une des plus petites RE du Québec. L'analyse centrographique réalisée confirme cette observation; Le centre moyen de la base globale se trouve en effet plus proche de Montréal que celui de la base TMD. Ce point moyen correspond d'ailleurs presque à celui de la population québécoise. Ainsi les accidents TMD semblent mieux répartis sur le territoire québécois et moins liés à la répartition de population, ce que confirment les ellipses d'écart-type. L'analyse de l'évolution annuelle de la position des centres moyens montre que globalement la tendance est à un rapprochement vers Montréal notamment pour les accidents de la base globale, les centres moyens des accidents TMD semblent eux plus aléatoires et varient sur une plus grande surface d'une année sur l'autre. Une analyse des plus proches voisins vient confirmer l'éparpillement plus important des accidents de la base TMD sur le territoire avec une valeur du

NNI plus de 10 fois supérieure à celui de la base globale. L'analyse des points chauds met en évidences que si l'on étudie « en gros » (ie. avec de « gros » points chauds) les répartitions se comportent de façon similaire. La concentration principale des accidents se trouvant comme le confirme l'analyse des densités autour de Montréal. Par contre lorsque l'on fait des regroupements plus fins des différences apparaissent. La concentration plus forte des accidents de la base globale est alors mise en évidence, avec notamment des ellipses NnH de niveau 2 couvrant une faible partie du territoire comparativement aux ellipses de la base TMD.

Enfin une analyse spatiale pondérée des accidents a été réalisée, selon le nombre de véhicules impliqués, de décès et de blessés. Les densités des véhicules impliqués et des blessés sont similaires, les zones de fortes densités correspondants aux endroits de forte concentration d'accidents. En revanche la répartition des décès souligne, en plus des zones de forte concentrations d'accidents, des sites particulièrement dangereux, pour la plupart communs aux deux distributions.

La principale limitation de cette étude concerne la validité des données qui nous ont été fournies. Il est difficile de savoir à quels points des données sont représentatives de la réalité. Il est certain qu'il manque dans les deux bases de données des événements qui devraient y figurer. Pour la base globale on peut espérer que ce phénomène soit assez limité par l'obligation légale qu'ont les policiers de rapporter les accidents impliquant des véhicules lourds. Par contre pour les événements impliquant les matières dangereuses les entreprises peuvent être tentées de ne pas rapporter des accidents mineurs comme un « petit » déversement. Il faut donc être bien conscient, et les auteurs l'ont souvent noté dans leurs études, que ce sont les événements aux impacts les plus majeurs (matériel ou humain) qui sont le plus susceptibles d'être rapportés. Dans notre cas, au vu du nombre d'événements répertoriés, on peut cependant estimer que ce phénomène est assez limité.

Par ailleurs les rapports d'accidents n'ont pas tous le même degré de précision et certains champs souffrent d'un manque important d'information. Certains champs sont remplis pour à peine 15% des accidents ce qui est très faible et vient fausser leurs études. Les champs plus remplis ne sont aussi exempt d'erreur et de fautes qui viennent gêner leurs études. Cela est particulièrement vrai pour le champ adresse qui est trop souvent incompréhensible et la localisation qui en résulte est alors approximative (municipalité uniquement). Si le problème a été résolu pour la base TMD par

une relocalisation manuelle, cela n'a pas été possible pour la base globale en raison du trop grand nombre d'accidents. Ce problème influence certainement notre étude spatiale et limite la force et la finesse de notre étude spatiale. Il pourrait être pertinent, à l'avenir, de mettre en place une méthode permettant de mieux localiser la base globale. On peut ainsi recommander à la SAAQ de revoir sa méthode de localisation, en indiquant par exemple directement les coordonnées longitude et latitude des accidents. Avec la démocratisation d'outil comme les GPS cela ne semble pas très couteux et difficile à mettre en œuvre.

Ainsi les perspectives de recherche, à partir d'une localisation précise sont importantes et variées. Un premier travail pourrait s'intéresser de façon détaillée à l'impact des accidents TMD sur la population. En estimant les rayons d'impacts, où même plus précisément les panaches d'impact des accidents en fonction de la matière dangereuses et de la masse du chargement, il serait alors possible, à l'aide des données de recensement, d'estimer le nombre de personnes touchées, voire les conséquences possibles sur la vie humaine, en termes de décès et de blessés. Un autre axe de recherche spatial serait d'étudier finement la répartition des accidents sur le réseau routier. Grace à un travail précis de localisation il serait possible d'identifier les points ou zones « noires » du réseau routier, un peu comme Pulugurtha et al (2007) l'ont fait pour les accidents piétonniers de Las Vegas. Enfin il serait aussi intéressant d'analyser la répartition des accidents TMD par rapport au tissu industriel, particulièrement par rapport aux lieux où les matières dangereuses sont stockées et/ou utilisées. Il faudra pour cela ajouter une couche de données présentant la répartition de ces installations, ce qui demanderait certainement un travail de collecte de données important.

L'analyse statistique des champs réservent aussi un grand nombre d'analyses possibles. Un arbre des probabilités tel que réalisé par Oggero et al. (2006) (cf. 2.1.5) pourrait être construit. On peut aussi chercher à combiner plusieurs champs ensemble pour étudier l'influence d'un paramètre sur un autre. Par exemple on peut essayer de déterminer s'il y a plus de décès ou de blessés la nuit que le jour, les possibilités sont presque infinies. Enfin les informations sur les véhicules impliqués dans les accidents n'ont pas été utilisées dans notre étude. Une étude future pourrait utiliser ces données, afin d'étudier le profil des véhicules qui ont des accidents (âge, marque, type, provenance,…).

REFERENCES

Ak, R., & Bozkaya, B. (2008). A proposed risk model and a GIS framework for hazardous materials transportation. *Engineering Management Conference, 2008. IEMC Europe 2008. IEEE International* (pp. 1-5). 10.1109/IEMCE.2008.4617959

Ames, D., & MapWindows GIS Team. (2010). *MapWindows GIS* (Version 4.8.1) [Open Source GIS]. http://www.mapwindow.org/

Bastet, P. (2008). Petit formulaire de statistique, à l'usage des élèves 1ère année ENSAM. Cours SIM (Systèmes Informatiques et outils Mathématiques pour l'ingénieur). 30.

Batta, R., & Chiu, S. S. (1988). Optimal obnoxious paths on a network: Transportation of hazardous materials. *Operations Research, 36*(Compendex), 84-92.

Bibliothèques universitaires québécoises. (2007). *Données statistiques et géographiques du Canada et du Québec*. Consulté le 06/09/2010, tiré de http://ivt.crepuq.qc.ca/.

Boies, R., & Beaulieu, D. (2009 07/10). Les accidents TMD 1995 à 2007. *Colloque Transport des Matières Dangereuses, Montréal*.

Brun-Picard, C., Gominet, S., & Louis, S. (2008). *Risques liés au transport de marchandises dangereuses (TMD), Fiche RT3*. Consulté le 25 novembre 2010, tiré de http://www.mementodumaire.net/02risques_technos/RT3.htm.

Bubbico, R., Di Cave, S., Mazzarotta, B., & Silvetti, B. (2009). Preliminary study on the transport of hazardous materials through tunnels. *Accident Analysis & Prevention, 41*(6), 1199-1205.

Bubbico, R., Ferrari, C., & Mazzarotta, B. (2000). Risk analysis of LPG transport by road and rail. *Journal of Loss Prevention in the Process Industries, 13*(1), 27-31.

Bureau d'Analyse des Risques et Pollutions Industriels (BARPI). (2010). Inventaire 2010 des accidents technologiques d. l. E. Ministère de l'Ecologie, du Développement Durable et de l'Aménagement du Territoire,, (Éd.). www.aria.developpement-durable.gouv.fr

Cao, H., Chouinard, O., & Dehoorne, O. (2005). « De la périphérie vers le centre : l'évolution de l'espace francophone du Nouveau-Brunswick au Canada ». *Annales de géographie, 2*, 115-140.

Carol, S., Vilchez, J. A., & Casal, J. (2002). Study of the severity of industrial accidents with hazardous substances by historical analysis. *Journal of Loss Prevention in the Process Industries, 15*(6), 517-524.

Croft, T., Watry, G., Ames, D. P., & Michaelis, C. (2007). Quick Guide to MapWindow GIS Desktop Application, (pp. 24). Consulté le 06/09/2010, tiré de: http://www.mapwindow.org/download/Quick_Guide_to_MapWindow_GIS.pdf

Darbra, R.-M., & Casal, J. (2004). Historical analysis of accidents in seaports. *Safety Science, 42*(2), 85-98.

Daumas, F. (1982). Méthodes de normalisation des données. *Revue de Statistique Appliquée, 30*(4), 23-38.

de Marcellis-Warin, N., Favre, S., Peignier, I., & Trépanier, M. (2006). *Revue des réglementations applicables au stockage et au transport des matières dangereuses au Québec*. 2006RP-11, Projet "GLOBAL". Ministère du Transport du Québec, Chaire CN en Intermodalité des Transports, IRSST, Ville de Montréal – CIRANO. Montréal.

de Marcellis-Warin, N., Leroux, M., Peignier, I., & Trépanier, M. (2006). *Revue et analyse des bases de données canadiennes et américaines touchant les accidents durant le transport et le stockage des matières dangereuses*. 2006RP-12, Projet "GLOBAL". Ministère du Transport du Québec, Chaire CN en Intermodalité des Transports, IRSST, Ville de Montréal – CIRANO. Montréal.

Dean, R., Bisson, M., & La Rochelle, P. (1974). Analyse centrographique de la population du Québec de 1951 à 1971. *Cahiers de géographie du Québec, 18*(45), 421-444.

Dionne, G., Laberge-Nadeau, C., Desjardins, D., Messier, S., & Vanasse, C. (1995). *Analyse des facteurs qui expliquent les taux et les gravités des accidents routiers impliquant des chauffeurs professionnels au Québec* (Rapport R-111). Montréal.

Direction de la sécurité en Transport. (2008). Limite de vitesse légale temporaire dans les zones de travaux routiers [Note d'information technique], (mai 2008), Consulté le 10/09/2010, tiré de http://www.mtq.gouv.qc.ca/portal/page/portal/6A5E4A3921506A9EE04400144F0104BD

Direction de la sécurité en transport et du camionnage. (2006). Guide sur le transport des matières dangereuses Direction des communications, (Éd.). Consulté le 24/08/2010, tiré de: http://www.mtq.gouv.qc.ca/portal/page/portal/1399408A0A973ECEE04400144F0104BD

Dobbins, J. P., & Abkowitz, M. D. (2003). Development of a centralized inland marine hazardous materials response database. *Journal of hazardous materials, 102*(2-3), 201-216.

Drouin, C., & Leroux, D. (2004). Transport et Environnement: Analyse des risques associés au transport des matières dangereuses en milieu urbain [En ligne]. *Congrés et exposition annuels de l'Association des Tranports du Canada, Québec (Canada)* (pp. 16). Consulté le 25/08/2010, tiré de TRIS Online. http://www.tac-atc.ca/english/pdf/conf2004/DROUIN.pdf

Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. *Accident Analysis & Prevention, 40*(1), 174-181.

Erkut, E., Tjandra, S. A., & Verter, V. (2007). Hazardous Materials Transportation. *Handbooks in Operations Research and Management Science, 14*(C), 539-621.

Erkut, E., & Verter, V. (1998). Modeling of transport risk for hazardous materials. *Operations Research*, 625-642.

Evans, A. (2003). *Transport fatal accidents and FN-curves 1967-2001* (Research Report). Londres: University College London.

Guo, X., & Verma, M. (2010). Choosing vehicle capacity to minimize risk for transporting flammable materials. *Journal of Loss Prevention in the Process Industries, 23*(2), 220-225.

Haastrup, P., & Brockhoff, L. (1990). Severity of accidents with hazardous materials : A comparison between transportation and fixed installations. *Journal of Loss Prevention in the Process Industries, 3*(4), 395-405.

Haastrup, P., & Brockhoff, L. H. (1991). Reliability of accident case histories concerning hazardous chemicals : An analysis of uncertainty and quality aspects. *Journal of hazardous materials, 27*(3), 339-350.

Hilgard, J. (1872). The advance of population in the United States. *Scribner's Monthly, 4*(2), 214-218.

Hobeika, A., & Kim, S. (1993). Databases and needs for risk assessment of hazardous materials shipments by trucks. *Transportation of hazardous materials: issues in law, social science, and engineering*, 135.

Levine, N. (2006). Crime Mapping and the Crimestat Program. *Geographical Analysis, 38*(1), 41-56.

Levine, N., & Associates. (2010a). Chapter 4 - Spatial Distribution *CrimeStat Documentation* (pp. 73).

Levine, N., & Associates. (2010b). Chapter 5 - Distance Analysis I and II *CrimeStat Documentation* (pp. 42).

Levine, N., & Associates. (2010c). Chapter 6 & 7 - "Hot Spot" Analysis I & II *CrimeStat Documentation* (pp. 98).

Levine, N., & Associates. (2010d). Chapter 8 - Kernel Density Interpolation *CrimeStat Documentation* (pp. 43).

Levine, N., & Associates. (2010e). *CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations* (Version 3.3). Houston, TX: National Institute of Justice, Washington, DC. Consulté le Octobre 2010, tiré de http://www.icpsr.umich.edu/CrimeStat/

Liu, N., Huang, B., & Chandramouli, M. (2006). Optimal Siting of Fire Stations Using GIS and ANT Algorithm. *Journal of Computing in Civil Engineering, 20*(5), 361-369.

Lozano, A., Muñoz, Á., Antún, J. P., Granados, F., & Guarneros, L. (2010). Analysis of hazmat transportation accidents in congested urban areas, based on actual accidents in Mexico. *Procedia - Social and Behavioral Sciences, 2*(3), 6053-6064.

Luketa-Hanlin, A. (2006). A review of large-scale LNG spills: Experiments and modeling. *Journal of hazardous materials, 132*(2-3), 119-140.

MAMROT. (2010). Répertoire des Municipalités. from Ministère des Affaires municipales, des Régions et de l'Occupation du territoire du Québec http://www.mamrot.gouv.qc.ca/repertoire_mun/repertoire/repertoi.asp

MapWindows GIS Team. (2010, 7 Juillet 2010). *MapWindow GIS Documentation*. Consulté le 06 septembre 2010, tiré de http://www.mapwindow.org/wiki/index.php/Main_Page.

Microsoft. *À propos des outils d'analyse statistique, Support Excel*. Consulté le 18 septembre, tiré de http://office.microsoft.com/fr-fr/excel-help/a-propos-des-outils-d-analyse-statistique-HP005203873.aspx.

Microsoft. (2009). *Microsoft MapPoint 2010 North America* (Version 17.00.19.2900) [GIS software]. http://www.microsoft.com/mappoint/en-us/default.aspx

Microsoft. (2010). *Aides et procédures Excel*. Consulté le septembre, tiré de http://office.microsoft.com/fr-be/support/aide-et-procedures-excel-FX010064695.aspx.

Ministère des Transports du Québec. (2008). Aide à la détermination des limites de vitesse sur le réseau routier municipal [Guide], 8. Consulté le 11/09/2010, tiré de http://www.securite-routiere.qc.ca/2e_rapport_Tqsr_Web.pdf

Ministère des Transports du Québec. (2010). *Règlement sur le transport des matières dangereuses*. Consulté le 24/08/2010, tiré de http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=3&file=/C_24_2/C24_2R4_2_1.HTM.

Montiel, H., Vílchez, J. A., Arnaldos, J., & Casal, J. (1996). Historical analysis of accidents in the transportation of natural gas. *Journal of hazardous materials, 51*(1-3), 77-92.

Nations Unies. (2009). *Recommandations Relatives au Transport des Marchandises Dangereuses* (Seizième édition révisée[e] éd. Vol. I). New York et Genève: Publication des Nations Unies.

Oggero, A., Darbra, R. M., Muñoz, M., Planas, E., & Casal, J. (2006). A survey of accidents occurring during the transport of hazardous substances by road and rail. *Journal of hazardous materials, 133*(1-3), 1-7.

Ohtani, H., & Kobayashi, M. (2004). Statistical analysis of dangerous goods accidents in Japan. *Safety Science, 43*(5-6), 287-297.

Planas-Cuchi, E., Montiel, H., & Casal, J. (1997). A Survey of the Origin, Type and Consequences of Fire Accidents in Process Plants and in the Transportation of Hazardous Materials. *Process Safety and Environmental Protection, 75*(1), 3-8.

Pulugurtha, S., Krishnakumar, V., & Nambisan, S. (2007). New methods to identify and rank high pedestrian crash zones: An illustration. *Accident Analysis & Prevention, 39*(4), 800-811.

Qiao, Y., Keren, N., & Mannan, M. S. (2009). Utilization of accident databases and fuzzy sets to estimate frequency of HazMat transport accidents. *Journal of hazardous materials, 167*(1-3), 374-382.

ReVelle, C., Cohon, J., & Shobrys, D. (1991). Simultaneous Siting and Routing in the Disposal of Hazardous Wastes. *Transportation Science, 25*(2), 138.

Ronza, A., Félez, S., Darbra, R. M., Carol, S., Vílchez, J. A., & Casal, J. (2003). Predicting the frequency of accidents in port areas by developing event trees from historical analysis. *Journal of Loss Prevention in the Process Industries, 16*(6), 551-560.

Ronza, A., Vílchez, J. A., & Casal, J. (2007). Using transportation accident databases to investigate ignition and explosion probabilities of flammable spills. *Journal of hazardous materials, 146*(1-2), 106-123.

Saint-Éloi, J. P. (2004). Transport de matières dangereuses. In (Vol. AGL2). Paris, FRANCE: Techniques de l'ingénieur. Consulté le 25/08/2010, tiré de http://www.techniques-ingenieur.fr/book/ag8170/transport-de-matieres-dangereuses.html

Schweitzer, L. (2006). Environmental justice and hazmat transport: A spatial analysis in southern California. *Transportation Research Part D: Transport and Environment, 11*(6), 408-421.

Shorten, C. V., Galloway, J., Krebs, J. G., & Fleming, R. S. (2002). A 12-year history of hazardous materials incidents in Chester County, Pennsylvania. *Journal of hazardous materials, 89*(1), 29-40.

Statistique Canada. (2007a). Fichier du réseau routier et des attributs géographiques, guide de référence, Année de recensement 2006, (Statistique Canada ed., pp. 37). Ottawa. Consulté le 07/09/2010, tiré de: http://geodepot.statcan.gc.ca/2006/18050605180514030 5/06180505162102/92-155-G2006001-fra.pdf

Statistique Canada. (2007b). Fichiers des limites, Guide de référence, Recensement de 2006 (Troisième édition), (Statistique Canada, ed., pp. 60). Ottawa. Consulté le 06/09/2010, tiré de: http://geodepot.statcan.gc.ca/2006/18050605180514030 5/06180505162102/92-160-G2006003-fra.pdf

Statistique Canada. (2007c, 23/04/2007). *Glossaire Illustré* Consulté le 06/09/2010, tiré de http://geodepot.statcan.ca/Diss2006/Reference/COGG/Index_f.jsp#cont.

Statistique Canada. (2010). Les statistiques : le pouvoir des données! [Ressources éducatives], Consulté le 24 Octobre 2010, tiré de http://www.statcan.gc.ca/edu/power-pouvoir/toc-tdm/5214718-fra.htm

Stewart, A. M., & Van Aerde, M. (1990). An empirical analysis of Canadian gasoline and LPG truck releases. *Journal of hazardous materials, 25*(1-2), 205-217.

Tiemessen, G., & Van Zweeden, J. (1998). Risk assessment of the transport of hazardous materials. (pp. 299–307).

Transports Canada. (2009a). *Les Transports au Canada 2009 - Tableaux de l'addenda*. Ottawa: Ministère des travaux publics et des services gouvernementaux. Consulté le 30/08/2010, tiré de http://www.tc.gc.ca/media/documents/politique/addenda2009.pdf

Transports Canada. (2009b). *Les Transports au Canada 2009 - Un survol*. Ottawa: Ministère des travaux publics et des services gouvernementaux. Consulté le 30/08/2010, tiré de http://www.tc.gc.ca/media/documents/politique/survol2009.pdf

Transports Canada. (2010). *Règlement canadien sur le transport des marchandises dangereuses (RTMD)*. Consulté le 25/08/2010, tiré de http://www.tc.gc.ca/fra/tmd/securite-menu.htm.

Trépanier, M. (2010). Système de transport intelligents: notes de cours: cours IND6214. *Ecole Polytechnique de Montréal*.

Trépanier, M., Leroux, M.-H., & de Marcellis-Warin, N. (2009). Cross-analysis of hazmat road accidents using multiple databases. *Accident Analysis & Prevention, 41*(6), 1192-1198.

Verter, V., & Kara, B. Y. (2001). A GIS-Based Framework for Hazardous Materials Transport Risk Assessment. *Risk Analysis, 21*(6), 1109-1120.

Verter, V., & Kara, B. Y. (2008). A path-based approach for hazmat transport network design. *Management Science, 54*(Compendex), 29-40.

Vílchez, J. A., Sevilla, S., Montiel, H., & Casal, J. (1995). Historical analysis of accidents in chemical plants and in the transportation of hazardous materials. *Journal of Loss Prevention in the Process Industries, 8*(2), 87-96.

Winder, C., Tottszer, A., Navratil, J., & Tandon, R. (1992). Hazardous materials incidents reporting: Results of a nationwide trial. *Journal of hazardous materials, 31*(2), 119-134.

Yang, J., Li, F., Zhou, J., Huang, L., Zhang, L., & Bi, J. (2010). A survey on hazmat accidents during road transport in China from 2000 to 2008. *Journal of hazardous materials, In Press, Accepted Manuscript*.

Zhang, J., Hodgson, J., & Erkut, E. (2000). Using GIS to assess the risks of hazardous materials transport in networks. *European Journal of Operational Research, 121*(2), 316-329.

Zografos, K. G., & Androutsopoulos, K. N. (2004). A heuristic algorithm for solving hazardous materials distribution problems. *European Journal of Operational Research, 152*(2), 507-519.

ANNEXE 1 – Les neufs classes de matières dangereuses

Classe	Risque principal présenté	Pictogrammes
Classe 1	Matières et objets explosibles	
Classe 2	Gaz comprimés, liquéfiés ou dissous sous pression	
Classe 3	Matières liquides inflammables	
Classe 4	4.1 : Matières solides inflammables	
	4.2 : Matières sujettes spontanément à l'inflammation	
	4.3 : Matières dégageant au contact de l'eau des gaz inflammables.	
Classe 5	5.1 : Matières comburantes	
	5.2 : Peroxydes organiques	
Classe 6	6.1 : Matières toxiques	
	6.2 : Matières infectieuses	
Classe 7	Matières radioactives	
Classe 8	Matières corrosives	
Classe 9	Matières et objets dangereux divers	

ANNEXE 2 – Structure du fichier de la base globale

N° de Champ	Titre du Champ	Nombre de valeur non nulles	% non vide	Description
1	ID	262 287	100,00%	N° d'identification unique
2	Jour de la semaine	262 287	100,00%	Jour de l'accident
3	Date complète	262 287	100,00%	Date complète de l'accident
4	Heure	262 287	100,00%	Heure présumée de l'accident
5	Minute	262 287	100,00%	Minute présumée de l'accident
6	Numéro CRPQ	262 287	100,00%	N° du Centre de Renseignements Policiers du Québec
7	Nombre de véhicules impliqués selon le rapport SUM	262 287	100,00%	Nb de véhicule impliqués dans l'accident
8	Route	105 602	40,26%	N° de la route sur laquelle l'accident a eu lieu
9	Adresse près de l'accident	245 772	93,70%	Adresse civique près du lieu de l'accident
10	Intersection près de l'accident	209 433	79,85%	Rue intersectante près du lieu de l'accident
11	Distance du repère en mètre	262 287	100,00%	Distance en mètre depuis le repère décrit
12	Direction	91 254	34,79%	Direction dans laquelle la distance est exprimée
13	Gravité	262 287	100,00%	Gravité de l'accident (Léger, matériel, grave ou mortel)
14	Municipalité RA act.	262 287	100,00%	Municipalité où l'accident a eu lieu
15	Limite de vitesse GLV	104 572	39,87%	Limite de vitesse sur les lieux de l'accident
16	Décès SUM	262 287	100,00%	Nombre de décès
17	Blessés graves SUM	262 287	100,00%	Nombre de blessés graves
18	Blessés légers SUM	262 287	100,00%	Nombre de blessés légers
19	Victimes SUM	262 287	100,00%	Nombre total de victimes
20	Genre d'accident	258 876	98,70%	Type d'accident (capotage, pilier,…)
21	Localisation RA	40 313	15,37%	Localisation de l'accident sur la chaussé (trottoir,…)
22	Aspect de la chaussée	39 663	15,12%	Alignement de la chaussée (plane, courbe,…)
23	Nature de la chaussée	39 722	15,14%	Type de chaussée (Asphalte, terre,…)

24	État de la chaussée	39 646	15,12%	État de structure de la chaussée (bon état, en construction,…)
25	État de la surface	255 981	97,60%	État de la surface (Sec, mouillée,…)
26	État de la météo	260 329	99,25%	Conditions météorologique
27	Éclairement	260 334	99,26%	État de l'éclairage sur les lieux de l'accident
28	Code d'impact	254 628	97,08%	Code décrivant l'impact sur le ou les véhicules
29	Environnement	74 041	28,23%	Milieu environnant l'accident (urbain, rural,…)
30	Catégorie de route	71 662	27,32%	Catégorie de la route (classe)
31	Délit de fuite	260 374	99,27%	Indique si l'accident a donné lieu à un délit de fuite (O ou N)
32	Camion lourd	262 287	100,00%	Type de camion lourd
33	Compteur	251 664	95,95%	?
34	RTS	100 754	38,41%	?

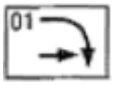
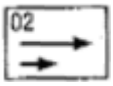
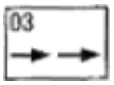
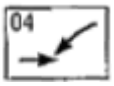
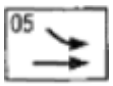
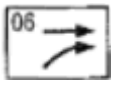
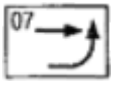
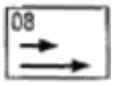
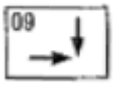
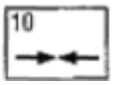
ANNEXE 3 – Structure du fichier de la base TMD

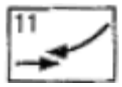
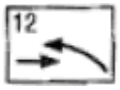
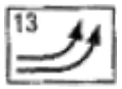
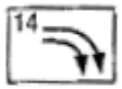
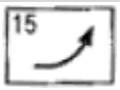
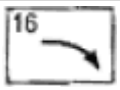
N° Champ	Champ	Type de données	Nb de valeurs non nulles	% non vide	Description
1	N° Dossier	Numéro	2 379	100,00%	N° d'identification unique
2	N° du microfilm	Texte	2 379	100,00%	N° de microfilm
3	N° CRPQ	Texte	962	40,44%	?
4	Code Camion-Citerne	Texte	1 486	62,46%	Code précisant le rôle du camion-citerne dans l'accident et la courbure de la route
5	Date	Date	2 379	100,00%	Date complète de l'accident
6	Heure	Heure	2 325	97,73%	Heure complète présumé de l'accident
7	Nombre de véhicule	Numérique	2 287	96,13%	Nombre de véhicule impliqué
8	DT	Texte	263	11,06%	Localisation de l'accident
9	CS	Texte	265	11,14%	?
10	Municipalité	Texte	2 368	99,54%	Municipalité où l'accident a eu lieu
11	Route	Numérique	1 112	46,74%	N° de la route sur laquelle l'accident a eu lieu
12	Nom du chemin	Texte	1 913	80,41%	Adresse civique près du lieu de l'accident
13	Route intersection	Texte	1 750	73,56%	Rue intersectante près du lieu de l'accident
14	Dommages matériels	Texte	2 041	85,79%	Estimation de l'ordre de grandeur des dommages matériels (généralement en Dollars)
15	Le véhicule lourd (Partie?)	Texte	2 288	96,17%	N° de parties du véhicule lourd
16	Type de véhicule lourd	Texte	2 246	94,41%	Type de véhicule lourd (semi-remorque, camion,…)
17	Coupures de journaux	Oui/Non	2 379	100,00%	Champ indiquant si l'accident a donné lieu à des coupures de journaux
18	Perte de contrôle	Oui/Non	2 379	100,00%	Champ indiquant si lors de l'accident il y a eu perte de contrôle
19	Déversement	Oui/Non	2 379	100,00%	Champ indiquant si l'accident a donné lieu à un déversement de chargement de matières dangereuses
20	Renversement	Oui/Non	2 379	100,00%	Champ indiquant si l'accident a

					donné lieu à un renversement
21	Croquis de l'accident	Oui/Non	2 379	100,00%	Champ indiquant si un croquis de l'accident a été réalisé
22	Type marchandises	Texte	1 277	53,68%	Type de marchandise transporté
23	Masse chargement (kg)	Numérique	25	1,05%	Masse en kilogramme du chargement en matière dangereuse
24	Matière dangereuse (NIP)	Texte	916	38,50%	N°UN de la matière dangereuse
25	Classe	Texte	1 176	49,43%	Classe si matière dangereuse
26	Autre détail	Mémo	64	2,69%	Détails sur la cargaison
27	Circonstances	Mémo	1 458	61,29%	Petit texte précisant les circonstances de l'accident
28	Vitesse autorisée	Numérique	1 039	43,67%	Limite de vitesse sur les lieux de l'accident
29	Morts (code 1)	Numérique	1 286	54,06%	Nombre de décès
30	Blessés graves (code 2)	Numérique	1 289	54,18%	Nombre de blessés graves
31	Blessés mineurs (code 3)	Numérique	1 448	60,87%	Nombre de blessés légers
32	Victime sans blessure (code 4)	Numérique	1 026	43,13%	Nombre de victime sans blessure
33	Code l'impact	Numérique	2 243	94,28%	Code décrivant l'impact sur le ou les véhicules
34	5 (Genre d'accident)	Texte	2 303	96,81%	Type d'accident (capotage, pilier,…)
35	6 (Environnement)	Texte	961	40,40%	Milieu environnant l'accident (urbain, rural,…)
36	7 (Catégorie de route)	Texte	974	40,94%	Catégorie de la route (classe)
37	8 (État de la surface)	Texte	2 282	95,92%	État de la surface (Sec, mouillée,…)
38	9 (Temps)	Texte	2 279	95,80%	Conditions météorologique
39	10 (Éclairement)	Texte	2 271	95,46%	État de l'éclairage sur les lieux de l’accident
40	13 (Localisation)	Texte	1 027	43,17%	Localisation de l'accident sur la chaussé (trottoir,…)
41	14 (Aspect de la chaussée)	Texte	920	38,67%	Alignement de la chaussée (plane, courbe,…)
42	15 (Nature de la chaussée)	Texte	887	37,28%	Type de chaussée (Asphalte, terre,…)
43	16 (État de la chaussée)	Texte	878	36,91%	État de structure de la chaussée (bon état, en construction,…)

ANNEXE 4 – Signification du code d'impact

Voici la signification des code d'impact indiqué dans les rapports d'accidents. Les codes d'impacts (01 à 16) déterminent le mouvement du ou des véhicules. Le code 15, 16 ou 88 doit être inscrit lorsqu'un **seul** véhicule est impliqué.

Symbole	Code	Description
	01	Les deux véhicules se dirigent dans le même sens : l'impact se produit lorsque celui circulant du côté gauche effectue un virage vers sa droite pour s'engager dans une direction différente.
	02	Les deux véhicules se dirigent dans le même sens : l'impact se produit lorsque l'un frôle l'autre en le dépassant par la gauche ou en changeant de voie de gauche à droite.
	03	Les deux véhicules se dirigent dans le même sens : l'impact se produit lorsque l'un s'approche de l'autre en venant par l'arrière.
	04	Les deux véhicules se dirigent dans les sens inverses : l'impact se produit lorsque l'un des deux effectue un virage vers sa gauche pour s'engager dans une direction différente.
	05	Les deux véhicules se dirigent l'un vers l'autre dans des directions qui se croisent : l'impact se produit lorsque celui venant du côté gauche effectue un virage vers sa gauche pour s'engager sur la même chaussée et dans le même sens que l'autre véhicule.
	06	Les deux véhicules se dirigent l'un vers l'autre dans des directions qui se croisent : l'impact se produit lorsque celui venant du côté droit effectue un virage vers sa droite pour s'engager sur la même chaussée et dans le même sens que l'autre véhicule. Cette case inclut également les accidents survenant sur les voies d'accès de boulevards et d'autoroutes.
	07	Les deux véhicules se dirigent dans le même sens : l'impact se produit lorsque celui circulant à droite effectue un virage vers sa gauche.
	08	Les deux véhicules se dirigent dans le même sens : l'impact se produit lorsque l'un frôle l'autre en le dépassant par la droite ou en changeant de voie de droite à gauche.
	09	Les deux véhicules se dirigent l'un vers l'autre dans des directions qui se croisent ; l'impact se produit lorsque les deux véhicules s'engagent dans l'intersection sans changer de direction.
	10	Les deux véhicules se dirigent dans des sens inverses : l'impact se produit lorsque l'un s'approche de l'autre en venant de l'avant (collision frontale).

	11	Les deux véhicules se dirigent l'un vers l'autre dans des directions qui se croisent : l'impact se produit lorsque celui venant du côté gauche effectue un virage vers sa droite pour s'engager sur la même caussée et en sens inverse de l'autre véhicule.
	12	Les deux véhicules se dirigent l'un vers l'autre dans des directions qui se croisent : l'impact se produit lorsque celui venant du côté droit effectue un virage vers sa gauche pour s'engager sur la même chaussée et en sens inverse de l'autre véhicule.
	13	Les deux véhicules circulent côte à côte et effectuent simultanément un virage à gauche ; l'impact se produit lorsque les deux véhicules se touchent.
	14	Les deux véhicules circulent côte à côte et effectuent simultanément un virage à droite ; l'impact se produit lorsque les deux véhicules se touchent.
	15	Véhicule seul quittant la chaussée vers la gauche.
	16	Véhicule seul quittant la chaussée vers la droite.
	88	Pour tous les accidents impliquant un seul véhicule dont le code d'impact est différent des codes 15 ou 16.
	99	Pour tous les accidents impliquant plus d'un véhicule dont le code d'impact est différent des codes 01 à 14.
		* Lorsqu'un des deux premiers véhicules impliqués dans l'accident est un véhicule stationné, inscrire 99.

ANNEXE 5 – Procédures VB permettant d'identifier les accidents de la base TMD dans la base globale

```
Sub Elimination()
'Elimination des dates non correspondantes (par année)
'Tri par date de la BDD indispensable!
Dim k As Long


'Année 1995
For i = 2 To 18440
k = 0
    For j = 2 To 217
        If Worksheets("BDD_Globale").Cells(i, 3) = Worksheets("BDD_MD").Cells(j, 5) Then
            If Worksheets("BDD_Globale").Cells(i, 4) = Worksheets("BDD_MD").Cells(j, 46) Then
            k = k + 1
            End If
        End If
    Next j

    If k = 0 Then
    Worksheets("BDD_Globale").Cells(i, 43) = "non"
    End If
Worksheets("BDD_Globale").Cells(2, 44) = i
Next i
```

Figure 1 : Extrait du programme de simplification de la recherche (année 1995)

```
Sub matchBDD_Vérification()

For j = 2 To 2380
Worksheets("Suivi").Cells(2, 8) = j

If Worksheets("BDD_MD").Cells(j, 48) <> "" Then
    If Worksheets("BDD_MD").Cells(j, 48) <> "x" Then

 'Champ date
 If Worksheets("BDD_Globale").Cells(Worksheets("BDD_MD").Cells(j, 49), 3) = Worksheets("BDD_MD").Cells(j, 5) Then
 Worksheets("BDD_MD").Cells(j, 51) = Worksheets("BDD_MD").Cells(j, 51) + 1
 End If

 'Champ Heure
 If Worksheets("BDD_Globale").Cells(Worksheets("BDD_MD").Cells(j, 49), 4) = Worksheets("BDD_MD").Cells(j, 6) Then
 Worksheets("BDD_MD").Cells(j, 51) = Worksheets("BDD_MD").Cells(j, 51) + 1
 End If

 'Champ Minute
 If Worksheets("BDD_Globale").Cells(Worksheets("BDD_MD").Cells(j, 49), 5) = Worksheets("BDD_MD").Cells(j, 7) Then
 Worksheets("BDD_MD").Cells(j, 51) = Worksheets("BDD_MD").Cells(j, 51) + 1
 End If

 'Champ N°CRPQ
 If Worksheets("BDD_Globale").Cells(Worksheets("BDD_MD").Cells(j, 49), 6) = Worksheets("BDD_MD").Cells(j, 3) Then
 Worksheets("BDD_MD").Cells(j, 51) = Worksheets("BDD_MD").Cells(j, 51) + 1
 End If
```

Figure 2 : Extrait du programme VB calculant le nombre de champ commun sur 21 possible

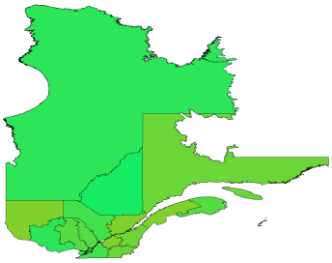
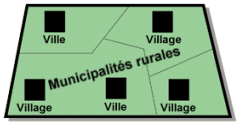
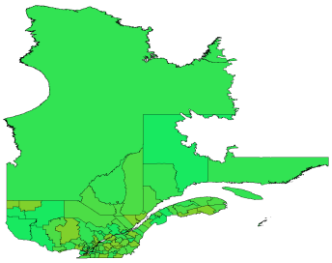
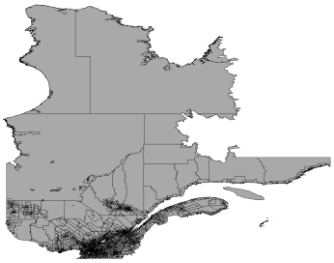
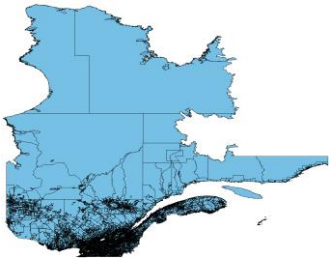
```
Sub matchBDD_MD1()
a = 43 'N° colonne matières dangereuse dans BDD Globale
b = 48 'N° de l'ID de l'accident MD trouvé dans BDD_Globale


For j = 2 To 2380
Worksheets("Suivi").Cells(3, 4) = j


    'Recherche Date/Heure/Municipalité
    k1 = 0: k2 = 0
    For i = 236357 To 246817
    If Worksheets("BDD_Globale").Cells(i, 3) = Worksheets("BDD_MD").Cells(j, 5) Then
        If Worksheets("BDD_Globale").Cells(i, 4) = Worksheets("BDD_MD").Cells(j, 7) Then
            If Worksheets("BDD_Globale").Cells(i, 14) = Worksheets("BDD_MD").Cells(j, 13) Then
            k1 = k1 + 1
            k2 = i
            End If
        End If
    End If
    Next i

    If k1 = 0 Then
    Worksheets("BDD_MD").Cells(j, b + 3) = "rien trouvé"
    Worksheets("Suivi").Cells(6, 4) = Worksheets("Suivi").Cells(6, 4) + 1
    End If

    If k1 = 1 Then
    Worksheets("BDD_Globale").Cells(k2, a) = "oui"
    Worksheets("BDD_Globale").Cells(k2, a + 1) = Worksheets("BDD_MD").Cells(j, 1)
    Worksheets("BDD_Globale").Cells(k2, a + 2) = j
    Worksheets("BDD_MD").Cells(j, b) = Worksheets("BDD_Globale").Cells(k2, 1)
    Worksheets("BDD_MD").Cells(j, b + 3) = "Date/Heure/Municipalité"
    Worksheets("BDD_MD").Cells(j, b + 4) = k1
    Worksheets("Suivi").Cells(7, 4) = Worksheets("Suivi").Cells(7, 4) + 1
    End If



    If k1 > 1 Then
    Worksheets("BDD_MD").Cells(j, b + 3) = "k1>1"
    Worksheets("Suivi").Cells(8, 4) = Worksheets("Suivi").Cells(8, 4) + 1
    End If


Next j

End Sub
```

Figure 3 : Programme de recherche d'associations uniques sur les champs Date / Heure / Municipalité

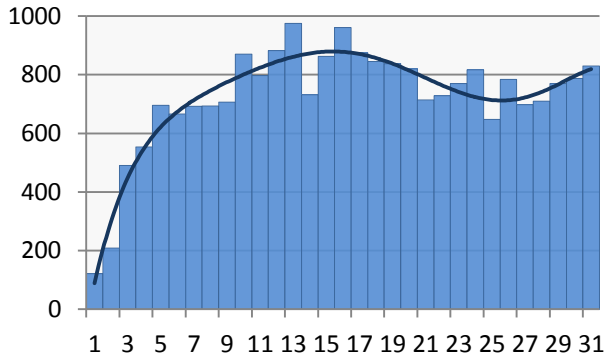
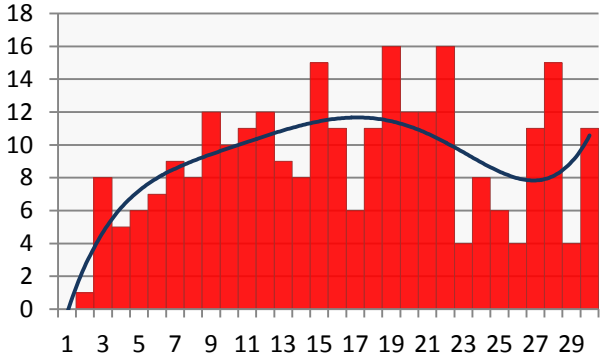
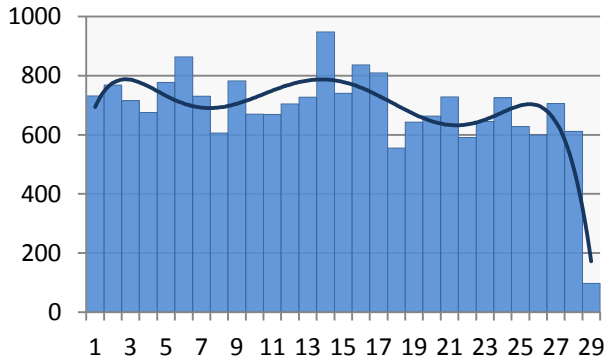
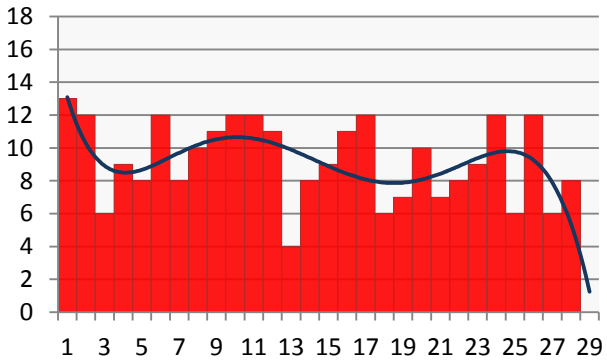
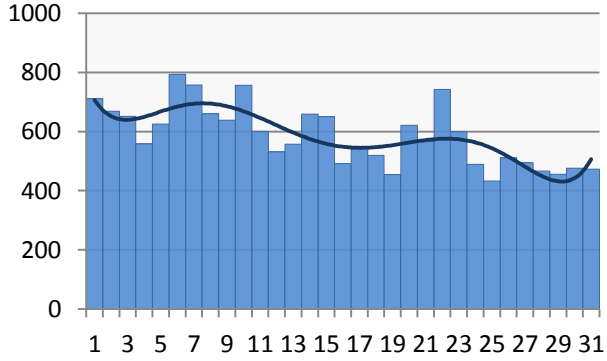
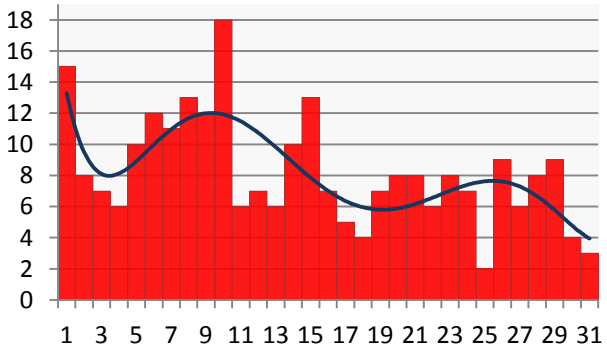
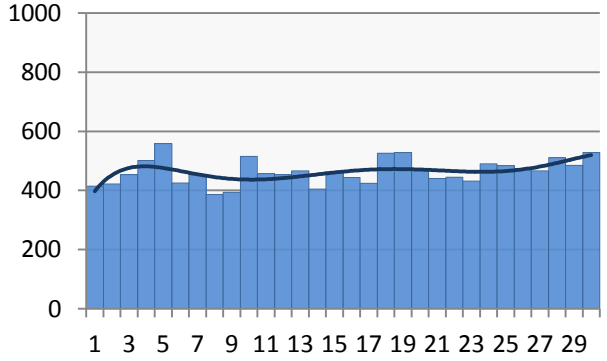
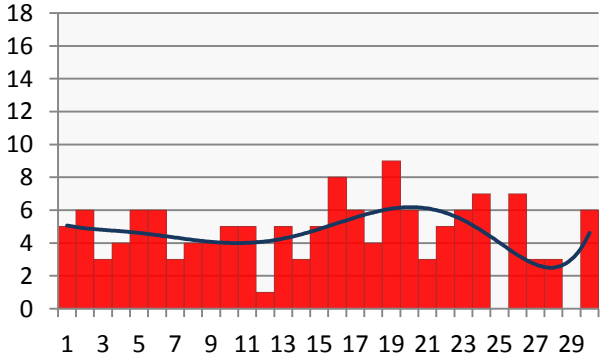
ANNEXE 6 – Description des paramètres de statistique descriptive utilisés dans l'analyse comparative des données

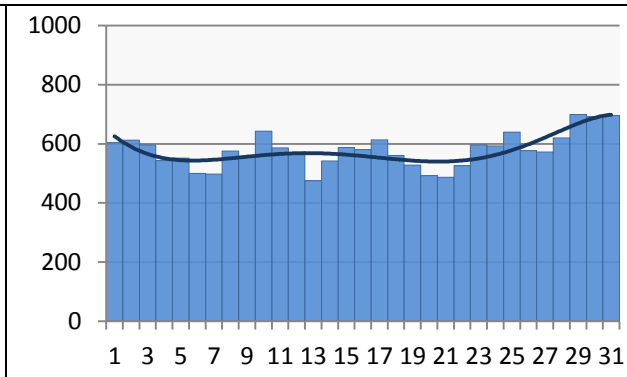
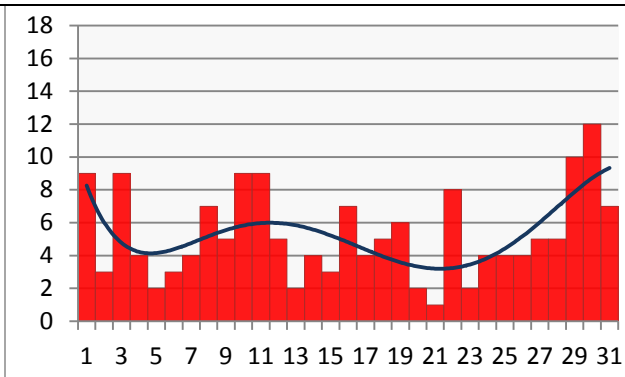
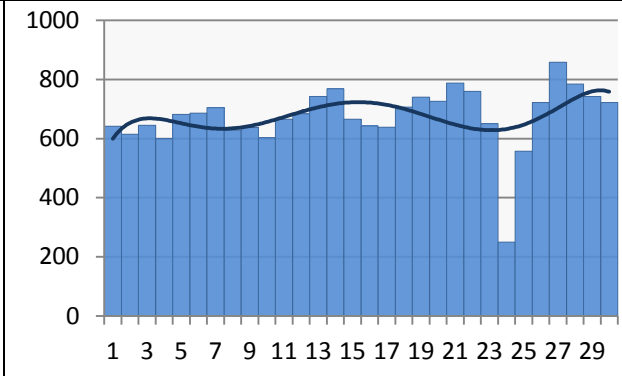
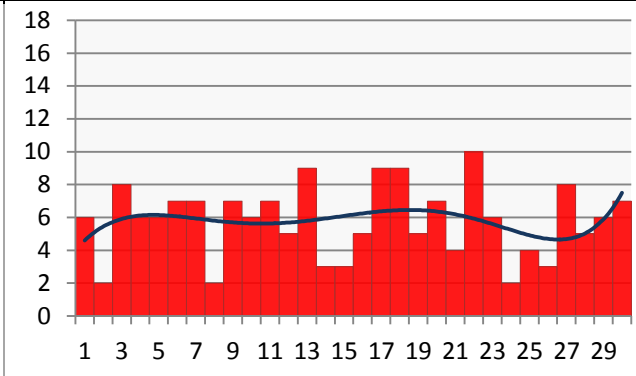
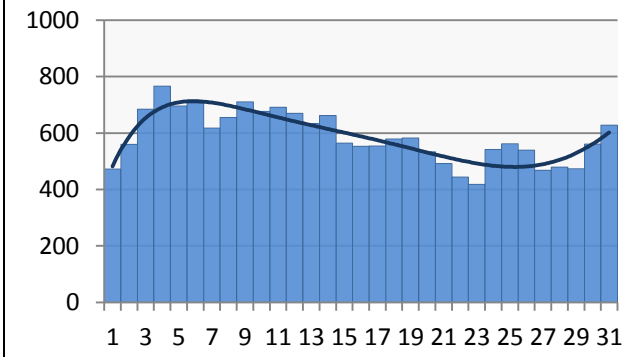
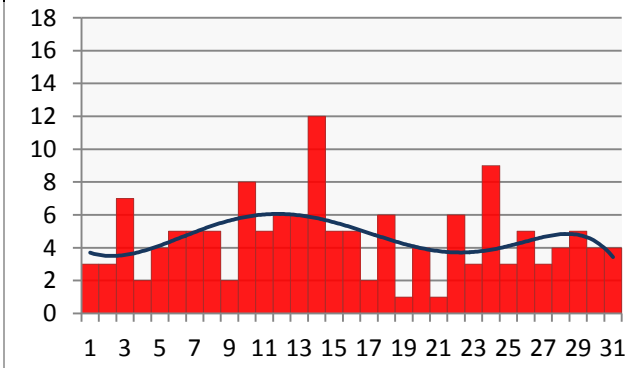
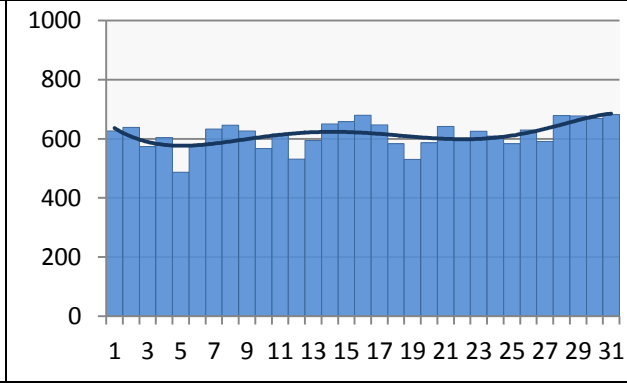
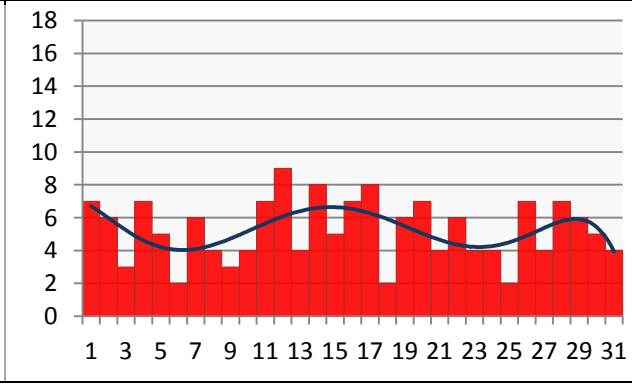
	Définition	Fonction Excel	Remarques
Paramètres de position			
Mode	Valeur la plus fréquente d'un échantillon	=MODE.SIMPLE()	
Moyenne	Grandeur de chacun des membres de l'échantillon s'ils étaient tous identique sans changer la dimension globale de l'ensemble	=MOYENNE()	
Médiane	Valeur qui laisse 50% des observations en dessous et 50% au-dessus	=MEDIANE()	
Quantile	Le quantile α est la valeur P_α qui laisse α% des observations en-dessous.	=QUARTILE.INCLURE(…;quart)	En général on calcule P_{25} et P_{75}
Paramètres de variabilité (ou critère de dispersion)			
Étendue (ou plage)	Écart entre la plus grande valeur et la plus petite valeur		Très sensible aux valeurs extrêmes, donc aux éventuelles valeurs aberrantes
Intervalle interquartiles	Écart $P_{75} - P_{25}$		Peu sensible aux valeurs extrêmes
Écart Moyen	Moyenne de la valeur absolue de l'écart avec la valeur moyenne	=ECART.MOYEN()	On préfère utiliser l'écart-type en général
Variance	Moyenne des carrés des écarts à la moyenne	=VAR.P()	Mesure la dispersion des valeurs (variance élevé => dispersion importante)
Écart-Type	Racine carré de la variance	=ECARTYPE.STANDARD()	Même unité que le caractère
Coefficient de variation (CV)	Quantifie en % ce que représente l'écart-type par rapport à la moyenne	=Écart-type/Moyenne	Évalue la dispersion relative (très utiles pour comparer deux séries qui n'ont pas le même ordre de grandeur)
Paramètres de forme			
Coefficient d'asymétrie (Fisher)	Indicateur de symétrie si proche de 0	=COEFFICIENT.ASYMETRIE(	Dissymétrie à droite si >>0, à gauche si <<0
Coefficient d'aplatissement (Kurtosis)	Indique si le sommet de la courbe est « pointu » ou « plat »	=KURTOSIS(	indique si le sommet est « pointu » (>>0) ou « plat » (<<0)

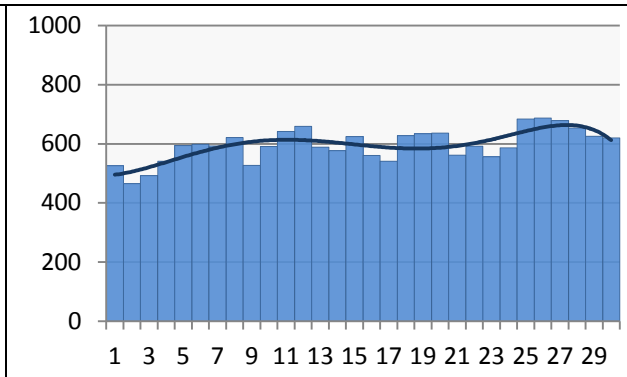
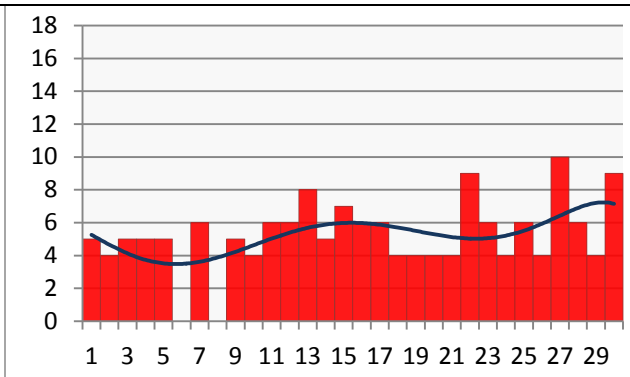
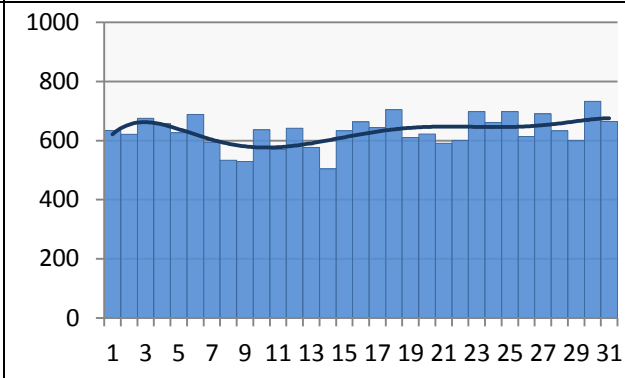
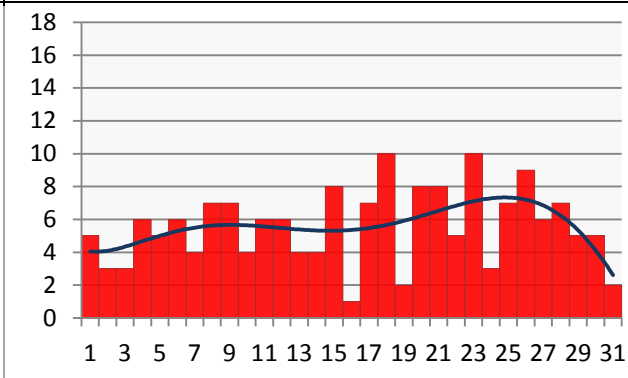
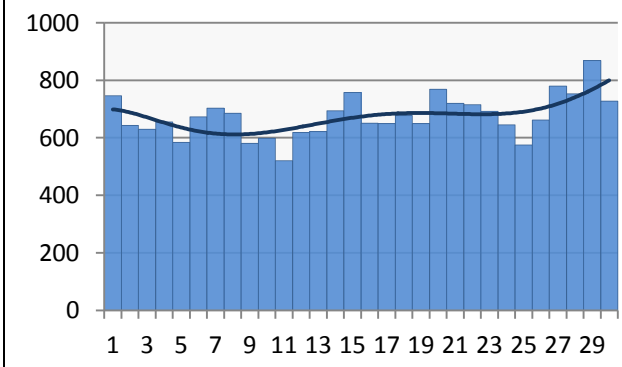
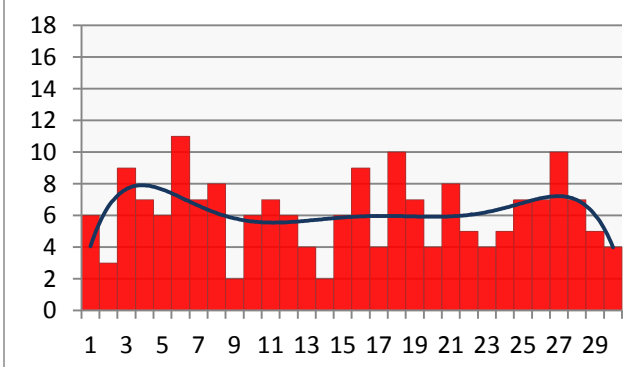
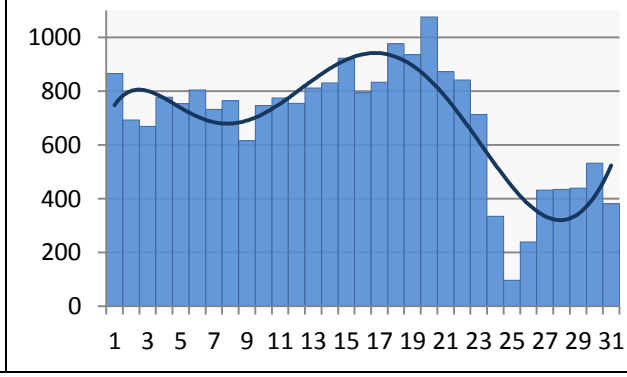
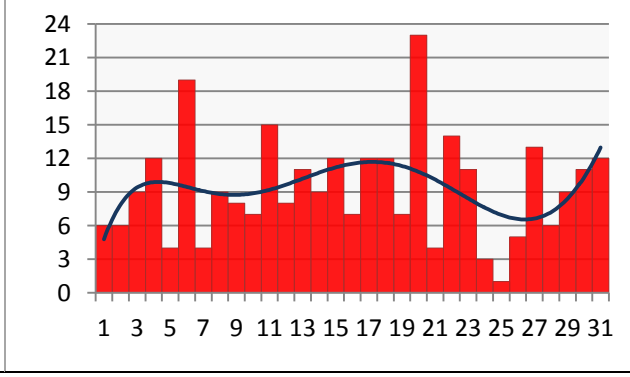
ANNEXE 7 – Définition des principales subdivisions du territoire Québécois

	Définition (Statistique Canada, 2007c)	Aperçu de la couche
Provinces et territoires	Délimitation des 10 provinces et des trois territoires constituant le Canada :	
Régions économiques (RE)	Groupes de divisions de recensement (17 au Québec)	
Divisions de recensement (DR)	« Groupes de municipalités voisines les unes des autres qui sont réunies pour des besoins de planification régionale et de gestion de services communs » (98 au Québec) 	
Aire de diffusion (AD)	« Petite région composée de un ou de plusieurs îlots de diffusion avoisinants et regroupant de 400 à 700 habitants ».	
Ilot de diffusion (ID)	« Territoire équivalant à un pâté de maisons dont les côtés sont délimités par des rues formant des intersections »	

ANNEXE 8 – Distribution mensuelle des accidents

Mois	Histogramme Base globale	Histogramme accidents TMD
Janvier		
Février		
Mars		
Avril		

Mai		
Juin		
Juillet		
Août		

Septembre		
Octobre		
Novembre		
Décembre		

ANNEXE 9 – Répartition spatiale des accidents

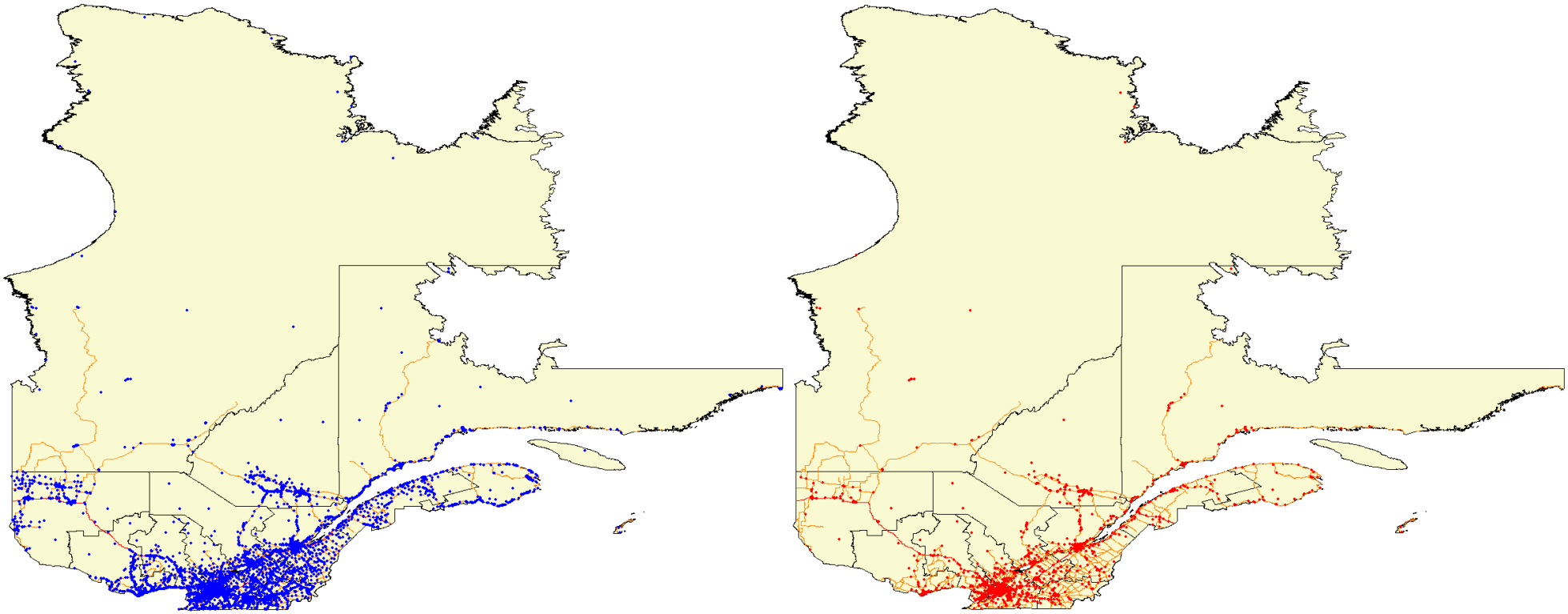


Figure 1 : Répartition des accidents obtenus à l'échelle du Québec (en bleu = base globale, en rouge = base TMD)

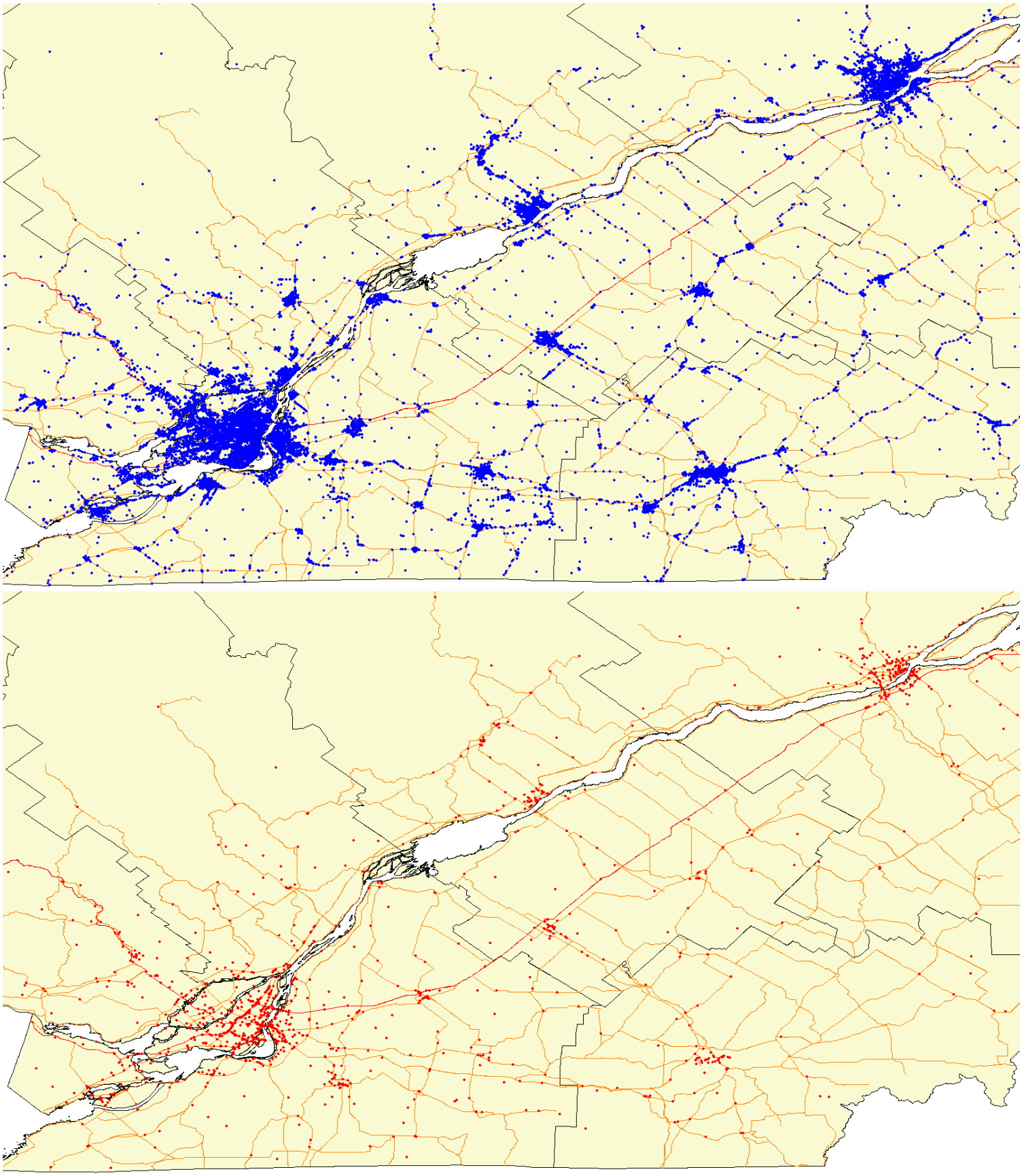
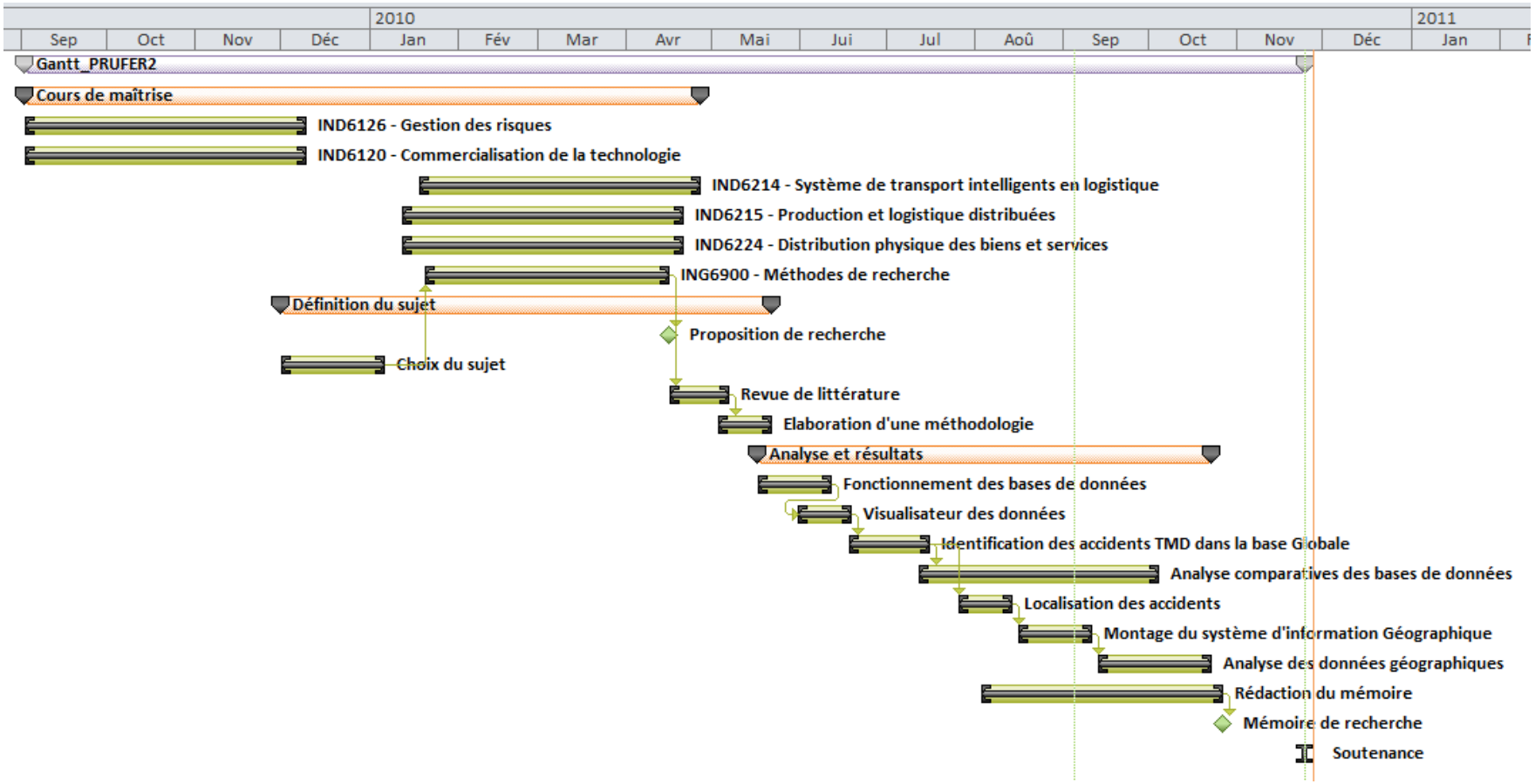


Figure 2 : Répartition des accidents entre Montréal et Québec (en bleu = base globale, en rouge = base TMD)

ANNEXE 10 – Diagramme de Gantt de la maîtrise