

Faculté des sciences

Impact de l'éclairage public sur la sécurité routière : une approche cas-contrôle par régression logistique et Random Forest

Au sein du Parc national de l'Entre-Sambre-et-Meuse

Auteur : **Romarc Pieroux**

Promoteur : **Patrick Meyfroidt**

Lecteur·rices : **Erasmus Zu Ermgassen, Sophie Vanwambeke**

Référent au Parc national de l'Entre-Sambre-et-Meuse : **Nicolas Goethals**

Année académique 2025-2026

Master [120] en sciences géographiques, orientation générale à finalité approfondie

Résumé

Les projets de Trame noire, bien développés en France, ont pour but de réduire la pollution lumineuse et donc l'éclairage artificiel avec l'objectif de valoriser des corridors écologiques et des réservoirs de biodiversité nocturnes. La pollution lumineuse, notamment créée par l'éclairage des routes, est responsable de la fragmentation des habitats et contribue donc à la crise de la biodiversité actuelle. Cependant, les projets de réduction de la pollution lumineuse s'opposent à trois principaux arguments : la sécurité routière, la criminalité et le sentiment d'insécurité. Ce mémoire vise donc à évaluer l'impact potentiel de l'éclairage public sur la sécurité routière dans le cadre de la mise en place d'une Trame noire au sein du Parc national de l'Entre-Sambre-et-Meuse, l'un des deux premiers Parcs nationaux de Wallonie. Pour cela, plusieurs variables, dont la présence d'un éclairage public, sont étudiées au moyen d'une approche cas-contrôle visant donc à expliquer la présence d'accidents à certaines localisations et l'absence à d'autres. Cette approche cas-contrôle est mise en œuvre via d'une part, un modèle de régression logistique, et d'autre part un modèle Random Forest, pour identifier l'effet des variables sur le risque d'accident. Le modèle Random Forest a l'avantage de prendre en compte l'aspect non-monotonique des variables. De ces résultats, il ressort notamment que l'éclairage public a un effet statistique positif significatif sur le risque d'accident, ce qui peut indiquer un biais d'endogénéité (les éclairages publics sont installés aux endroits accidentogènes). Il a également été possible, avec ces résultats, de prédire un risque d'accident pour l'ensemble du réseau routier étudié, mais également d'identifier les poteaux d'éclairage potentiellement inutiles, par rapport à la sécurité routière, car situés en des lieux ayant un risque d'accident faible, selon plusieurs seuils de risque d'accident.

Remerciements

Je tiens à remercier en premier lieu mon promoteur, Patrick Meyfroidt, de m'avoir accompagné dans ce mémoire. Merci pour les précieux conseils et pistes d'exploration liés à la méthodologie. Un tout grand merci aussi pour les nombreuses relectures qui ont donné une plus grande rigueur à ce mémoire.

Je remercie également Nicolas Goethals, chargé de mission au Parc national de l'Entre-Sambre-et-Meuse, de m'avoir proposé ce sujet de mémoire. Je le remercie également pour les nombreuses pistes données, les conseils et les discussions qui m'ont permis d'avancer dans mon mémoire. Sa relecture approfondie de mon mémoire m'a permis de m'assurer de la compréhension du sujet et d'améliorer la forme. Je le remercie également de m'avoir intéressé au sujet de la pollution lumineuse et de la Trame Noire. Je remercie aussi l'équipe du Parc national de m'avoir accueilli, d'abord pour un job étudiant, et ensuite pour mon mémoire.

Je remercie ma Maman qui m'a soutenu tout au long de mes études et de mon mémoire. Elle a su m'encourager et m'aider pour trouver de la motivation, merci ! J'ai également une pensée pour mon Papa, parti bien trop tôt, à qui je dois mon intérêt pour la géographie.

Un tout grand merci à mes amis avec qui j'ai partagé ces superbes années d'étude, Manu, Lucas, Antoine, Killian, Maxime et Nicolas. Les moments partagés en auditoire, en salle mem et surtout en dehors ont rendus ces années d'études mémorables. Merci à Victor, Tao et Olivier pour leur amitié et leur sympathie et avec qui j'ai partagé un bout de temps. Un merci à Simon Dek (« le propriétaire de la salle mem ») pour ces bons moments en salle mem. Je remercie également tous mes autres camarades de classe ainsi que les professeurs et assistants de l'école de géographie et de l'UCLouvain.

Merci à Laure pour les (très) nombreux chocolats chauds partagés ensemble, ainsi que pour tous ces moments ! Un merci aussi pour l'aide précieuse et la motivation pour travailler sur le cours de morpho.

Un merci à tous mes amis et mes proches !

Table des matières

1.	Introduction	1
2.	Revue de la littérature.....	3
2.1.	Pollution lumineuse	3
2.2.	Trame noire.....	4
2.3.	Accidents routiers	5
2.4.	Criminalité et sentiment d’insécurité la nuit	6
2.5.	Approche cas-contrôle	7
2.6.	Modèle de régression logistique et modèle Random Forest	8
3.	Matériel et méthode.....	9
3.1.	Zone d’étude.....	9
3.1.1.	Parc national de l’Entre-Sambre-et-Meuse	9
3.1.2.	Description du territoire.....	10
3.2.	Données	11
3.3.	Méthodes	14
3.3.1.	Impact potentiel de l’éclairage public sur l’environnement	15
3.3.2.	Approche cas-contrôle visant à identifier les facteurs influençant le risque d’accident : hypothèses, échantillonnage, et création de la base de données.....	15
3.3.2.1.	Hypothèses méthodologiques	15
3.3.2.2.	Hypothèses testées par les modèles de régression logistique.....	16
3.3.2.3.	Création des points de contrôle	16
3.3.2.4.	Présentation des variables étudiées.....	17
3.3.3.	Régression logistique (RL)	26
3.3.4.	Modèle Random Forest (RF).....	27
4.	Résultat.....	29
4.1.	Analyse exploratoire	29
4.1.1.	Caractéristiques des accidents	29
4.2.	Impact de l’éclairage public sur l’environnement	29
4.3.	Effet des variables sur le risque d’accident	30
4.3.1.	Analyse des modèles de régression logistique	30
4.3.2.	Analyse des modèles Random Forest	34
4.4.	Cartes du risque d’accident prédit.....	37
4.4.1.	Par régression logistique.....	37
4.4.2.	Par Random Forest	38
4.4.3.	Comparaison.....	39

4.5.	Poteaux d'éclairage public potentiellement inutiles (en ce qui concerne la sécurité routière)	40
5.	Discussion	43
5.1.	Impact de l'éclairage public sur le risque d'accident	43
5.2.	Impact des autres variables sur le risque d'accident.....	44
5.3.	Comparaison et évaluation des modèles	46
5.4.	Prédiction du risque d'accident et comparaison des deux modèles (RL et RF)	47
5.5.	Application des résultats pour des interventions de réduction de la pollution lumineuse	47
5.6.	Limites et pistes d'amélioration	48
6.	Conclusion.....	50
7.	Bibliographie.....	51
8.	Liste des figures	57
9.	Liste des tables	59
10.	Annexes	60
11.	Autobiographie	66

1. Introduction

Les effets néfastes de la pollution lumineuse sur l'environnement sont nombreux mais la fragmentation des habitats est sans doute l'effet le plus notable, favorisant la crise de la biodiversité actuelle (Bebronne *et al.*, 2024; Bennie *et al.*, 2014). La fragmentation physique du paysage et des habitats naturels par les zones artificialisées est accentuée par la pollution lumineuse, notamment par l'éclairage des routes. L'éclairage des routes s'est notamment intensifié avec l'urbanisation croissante et l'étalement urbain du siècle dernier lié, notamment, à la démocratisation de la voiture personnelle (Bebronne *et al.*, 2024). En plus des effets néfastes directs sur l'environnement, il est important de considérer le coût environnemental lié à la production d'électricité utilisée pour cet éclairage artificiel s'élevant à 182,6 GWh pour la région wallonne en 2022 (uvcw, 2022). De plus, cette production d'électricité a également un coût économique important s'élevant à 78,83 millions d'euros pour la Wallonie en 2022 (Bebronne *et al.*, 2024).

Plusieurs projets se sont créés en faveur d'une réduction de la pollution lumineuse comme les Trames noires en France visant la valorisation et la création de corridors écologiques et de réservoirs de biodiversité nocturnes (Sordello *et al.*, 2021). Le Parc national de l'Entre-Sambre-et-Meuse, l'un des deux premiers Parc nationaux de Wallonie, vise notamment la création d'un projet de Trame noire sur son territoire. Les projets de Trame noire passent notamment par la réduction et l'adaptation de l'éclairage public. Les principales actions pouvant être mises en œuvre sont le retrait de poteaux d'éclairage public, l'adaptation des horaires d'éclairage ou encore l'automatisation des ampoules basées sur la présence (Sordello *et al.*, 2021).

Cependant, ce type de projet en faveur de la réduction de la pollution lumineuse, se confronte à trois principaux arguments : la sécurité routière, la criminalité et le sentiment d'insécurité (Sordello *et al.*, 2021). Il est donc important de pouvoir évaluer ces éléments, de façon à pouvoir identifier des équilibres acceptables entre conservation de la biodiversité, sécurité routière, criminalité et sentiment d'insécurité. Néanmoins, si certaines études identifient un lien entre sécurité routière et éclairage public, il n'y a pas de consensus clair de l'effet de l'éclairage public sur le risque d'accident, des études récentes montrant une diminution de 3% du nombre d'accident pour des routes avec éclairage public (Donnell *et al.*, 2010a), là où des études anciennes, sujettes à des problèmes méthodologiques (qui sont développés dans la revue de la littérature), montrent une diminution jusqu'à 30% (Elvik, 1995). Les effets de l'éclairage public sur la criminalité et le sentiment d'insécurité sont également nuancés et seront développés dans la revue de la littérature.

Ce mémoire se concentre sur les facteurs de localisation influençant le risque d'accident puisque le Parc national est un territoire rural, et donc l'impact sur la biodiversité est principalement dû à l'éclairage public des routes. Dans ce territoire, les principaux risques sont donc plus liés au risque d'accident que liés à la criminalité ou au sentiment d'insécurité.

L'objectif de ce mémoire est donc d'étudier les facteurs de localisation influençant le risque d'accident et en particulier d'étudier l'effet ou non de l'éclairage public sur le risque d'accident. Le but sous-jacent est d'évaluer la pertinence d'utiliser autant d'éclairage public qu'actuellement malgré les effets néfastes liés à la pollution sur l'environnement. Les variables étudiées sont la classification du réseau routier, le volume de véhicules de nuit, la vitesse moyenne de nuit, la largeur des routes, la sinuosité des routes, la proximité à un éclairage public, la proximité à une intersection et enfin la présence ou non en zone bâtie. Ce mémoire prend place au sein des communes constituant le Parc national et des communes limitrophes. Il s'agit d'un territoire majoritairement rural et qui est traversé par la N5, une nationale faisant partie du réseau structurant de la Wallonie (SPW, 2025). Le but est donc

d'apporter une méthodologie permettant d'une part d'identifier l'impact potentiel de l'éclairage public sur le risque d'accident, en calibrant cette méthodologie sur le territoire étudié, et d'une autre part d'identifier les poteaux d'éclairage public potentiellement inutiles pour la sécurité routière, et ce dans le contexte de la mise en place d'une Trame noire au sein du Parc national de l'Entre-Sambre-et-Meuse.

Pour ce faire, ce mémoire utilise une approche cas-contrôle permettant d'étudier ce qui influence la présence d'un accident à certaines localisations et l'absence d'accident à d'autres. Les cas sont donc une collection d'accidents répertoriés entre 2017 et 2024, et les points de contrôle sont choisis aléatoirement sur le réseau routier hors des points d'accidents. Une classification du réseau routier, basée sur les volumes de véhicules par tronçon, est utilisée pour répartir proportionnellement ces points de contrôle à travers les différentes classes de routes ainsi créées. Cette approche cas-contrôle est mise en œuvre au moyen de deux types de modèles statistiques : la régression logistique et le modèle Random Forest. Le but du modèle de régression logistique, permettant d'étudier la variable dépendante binaire (présence ou absence d'accident), est surtout d'identifier les variables explicatives jouant un rôle dans le risque d'accident. Le modèle Random Forest, une méthode de machine learning basée sur les arbres de décision (Biau & Scornet, 2016), vise à identifier l'importance relative de chaque variable et à prédire également un risque d'accident pour l'ensemble du réseau routier. Cette seconde méthode a l'avantage de pouvoir prendre en considération l'effet non-linéaire et non-monotonique des variables. La classification du réseau routier permet également de créer un modèle de régression logistique et un modèle Random Forest par classe de route créée. Il est ainsi possible de décrire au mieux les variables pour chaque type de route (routes locales, nationales, autoroutes ou gabarit autoroutier). Sur base de ces deux approches, il est ensuite possible de créer une carte de risque d'accident sur le réseau routier étudié

Avec le risque d'accident prédit sur le réseau routier par les deux modèles, régression logistique et Random Forest, il est possible d'identifier, selon différents seuils de risque d'accident, les poteaux d'éclairage public qui seraient potentiellement inutiles pour la sécurité routière, car localisés à des endroits ayant un risque d'accident très faible. Le but serait alors de pouvoir prioriser les actions de réduction de l'éclairage artificiel, et donc de réduction de la pollution lumineuse, vers ces poteaux d'éclairage potentiellement inutiles.

Le deuxième chapitre de ce mémoire consiste en une revue de la littérature pour explorer les effets néfastes de la pollution lumineuse, la mise en place des projets de Trame noire, et ce que disent d'autres études sur l'impact potentiel de l'éclairage public sur la sécurité routière, sur la criminalité et le sentiment d'insécurité. Ce chapitre explore également les modèles utilisés dans ce mémoire, soit l'approche cas-contrôle, le modèle de régression logistique et le modèle Random Forest. Le troisième chapitre s'intéresse à expliciter les données utilisées ainsi que la méthode. Les résultats sont ensuite présentés, puis discutés dans les chapitres 4 et 5.

2. Revue de la littérature

2.1. Pollution lumineuse

La pollution lumineuse est engendrée par une utilisation excessive de lumière artificielle dans des zones d'obscurité qui entraîne des conséquences environnementales pour la faune et la flore (Bebronne *et al.*, 2024; Bennie *et al.*, 2014). La pollution lumineuse est notamment provoquée par l'éclairage public des voiries qui est vu comme une nécessité pour la sécurité routière. De ce fait, la Belgique est l'un des pays Européens qui éclaire le plus ses voiries et en particulier ses autoroutes et nationales en comparaison avec ses voisins.

L'éclairage public, en tant que service public, est réellement apparu à Paris au 17^{ème} siècle dans une volonté de mettre en valeur le règne du Roi Louis XIV (le Roi-Soleil) mais également d'apporter aux citoyens de Paris une hausse de la sécurité et de l'hygiène (McMahon, 2018). L'éclairage public, et l'éclairage des routes, s'est ensuite nettement intensifié au milieu du 20^{ème} siècle avec l'urbanisation croissante et l'étalement urbain vers les périphéries rendues possibles par la démocratisation de la voiture ainsi que par la forte croissance économique d'après-guerre. L'éclairage public est alors devenu un outil de l'aménagement du territoire en vue d'offrir du confort et de la sécurité dans les zones urbaines et périurbaines. L'éclairage public a également servi à soutenir l'activité économique nocturne et donc à améliorer les déplacements nocturnes. Mais cette augmentation de l'éclairage artificiel de nuit a également augmenté les nuisances liées à la pollution lumineuse (Meier *et al.*, 2015; Pérez Vega *et al.*, 2021).

La pollution lumineuse est une menace pour la biodiversité notamment par le fait que l'éclairage public des routes va agir comme une véritable barrière pour certaines espèces lucifuges. C'est d'autant plus inquiétant que 30% des mammifères et 60% des invertébrés sont nocturnes (Franchomme *et al.*, 2019a). Cela va amplifier d'autant plus la fragmentation des habitats par le réseau routier alors que la perte d'habitats est considérée comme la première cause de la crise de la biodiversité (Decocq, 2024; Jean, 2024). L'éclairage public va agir de deux manières sur le comportement des espèces : soit elle va piéger les individus en les attirant dans le halo lumineux, soit elle va repousser les espèces lucifuges. Il y a donc fragmentation par attraction et fragmentation par répulsion. Dans les deux cas, la pollution lumineuse va réduire la dispersion des espèces les empêchant d'atteindre des patchs d'habitats plus favorables. La flore subit aussi les effets néfastes de la pollution lumineuse, notamment par la raréfaction des pollinisateurs nocturnes (Longcore & Rich, 2004).

L'éclairage artificiel entraîne également des conséquences sur les rythmes biologiques (Tossa & Souques, 2019). Le rythme circadien est influencé par des facteurs environnementaux comme l'alternance chaud/froid, veille/sommeil et l'alternance lumière/obscurité. Ce dernier facteur, la luminosité, est d'ailleurs le plus important chez les mammifères. La synthèse de la mélatonine, une hormone nécessaire pour informer le corps de la photopériode, est impactée par l'éclairage artificiel nocturne puisque la production de mélatonine dépend du cycle jour/nuit. L'exposition à un éclairage artificiel nocturne tel un éclairage de rue provoque une inhibition de la production de cette hormone à un moment où l'organisme a besoin de se reposer et a donc besoin d'un taux de mélatonine plus élevé (Tossa & Souques, 2019). Cela provoque des conséquences sur le sommeil avec l'avance ou le retard des phases des sommeils ce qui provoque une baisse de la qualité du sommeil. La LED est de plus en plus choisie comme éclairage pour sa faible consommation d'électricité, mais la LED tend relativement plus vers des longueurs d'ondes bleues ce qui augmente l'impact sur l'environnement et la santé humaine par la diminution des taux de production de mélatonine (Tossa & Souques, 2019).

La pollution lumineuse a également un impact important sur le ciel étoilé rendant l'observation d'un ciel nocturne de plus en plus difficile. Il est notamment estimé que 83% de la population mondiale vit sous un ciel affecté par la pollution lumineuse (Falchi *et al.*, 2016).

L'éclairage public est un coût important pour les communes pouvant représenter jusqu'à 70% de leurs dépenses annuelles en électricité. En 2022, la Wallonie comptait 626813 points lumineux destinés à l'éclairage public pour une consommation totale d'électricité de 182,6 GWh représentant un coût de 78,83 millions d'euros (Bebronne *et al.*, 2024; uvcw, 2022). En plus du coût lié à la consommation d'électricité, il faut également ajouter le coût lié à l'entretien de ces points lumineux qui a représenté 14 millions d'euros pour la Wallonie en 2024 selon un rapport de la CWaPE (Commission wallonne pour l'Énergie) (CWAPE, 2025). Cette consommation a non seulement un coût économique mais également un coût environnemental lié à la production de l'énergie utilisée pour l'éclairage public. Ce coût environnemental, en émissions de gaz à effet de serre, dépend fortement du mix énergétique de chaque pays.

Comme vu dans cette section, il y a des avantages environnementaux et économiques, ainsi que pour la santé humaine, à repenser et réduire l'éclairage public pour limiter la pollution lumineuse. Des projets existent notamment en France avec les Trames noires en vue de réduire la pollution lumineuse pour valoriser la biodiversité nocturne (Sordello *et al.*, 2021). D'autres projets existent également comme les réserves internationales de ciel étoilé (RICE) dans le but de conserver et valoriser les ciels présentant une qualité exceptionnelle et donc les environnements avec peu de pollution lumineuse (Hunter *et al.*, 1991). Ces réserves internationales de ciel étoilé sont labellisées par l'Association Internationale Dark Sky (IDA). Pour autant, et comme montré par la suite, la volonté d'éteindre en partie l'éclairage public s'oppose à des arguments comme la sécurité routière et le sentiment d'insécurité la nuit (Sordello *et al.*, 2021).

2.2. Trame noire

Le concept de Trame noire rejoint les concepts de Trames bleue et verte qui consiste en la valorisation, la préservation et la création de corridors écologiques afin de réduire les effets négatifs de la fragmentation des habitats naturels (Sordello *et al.*, 2018). Les Trames bleues et vertes se concentrent donc principalement sur les espèces diurnes. La Trame noire rajoute la dimension nocturne en réduisant l'effet néfaste de l'éclairage artificiel. Les projets de Trame noire, en France, sont la plupart du temps portés par les Parcs naturels et les Parcs nationaux mais des communes, des communautés de communes et des métropoles ont déjà porté de tels projets. Il est notamment question du Parc national des Pyrénées, du Parc naturel régional des Causses du Quercy ou encore de la métropole de Lille (Sordello *et al.*, 2021). Ces projets visent donc la réduction de la pollution lumineuse et passent généralement par l'extinction locale ou le retrait de certains lampadaires d'éclairage public autant à l'intérieur des corridors écologiques et des réservoirs écologiques qu'en dehors puisque la pollution lumineuse a des impacts bien au-delà de sa source d'émission notamment via l'attraction et la répulsion de certaines espèces. D'autres mesures peuvent également être prises comme des horaires d'éclairage adaptés, le dimming des lampadaires (réduction l'intensité lumineuse) ou encore l'allumage des lampadaires avec capteur de présence. Le but est, pour les porteurs de projet de Trame noire, d'avoir une diminution globale de la pollution lumineuse sur le territoire et non exclusivement le long des corridors écologiques et des réservoirs écologiques d'espèces nocturnes pour diminuer les impacts de la pollution lumineuse sur la biodiversité (Franchomme *et al.*, 2019b; Sordello *et al.*, 2021).

La mise en place d'un projet de Trame noire repose sur plusieurs étapes dont l'essentielle consiste à identifier les points de conflits entre les zones à enjeux de biodiversité (corridors et réservoirs

écologiques) et la pollution lumineuse (Sordello *et al.*, 2021). Les zones à enjeux de biodiversité sont par exemple : des corridors écologiques, des zones Natura 2000, des réserves naturelles, des sites d'intérêt local (zones bocagères, forêts, grottes ou cavités, tout autre espace vert) et enfin, les cours d'eaux et zone humides. La pollution lumineuse est principalement identifiée par la localisation des points lumineux (lampadaires d'éclairage public) mais également à partir d'images satellites et données prises au sol. Vient ensuite l'étape de l'identification de la Trame noire en définissant les lieux et corridors étant encore peu affectés par la pollution lumineuse (Sordello *et al.*, 2021).

2.3. Accidents routiers

La sécurité routière est l'un des premiers arguments souvent cités en défaveur de la réduction de l'éclairage public des voiries, sur base de l'idée que l'éclairage public a un effet positif sur la sécurité routière et réduit donc le risque d'accident (Sordello *et al.*, 2021). L'éclairage des voiries a notamment pris de l'importance au cours des années 1950/1960 lorsque la voiture s'est démocratisée et a donc provoquée une augmentation considérable du nombre de véhicules sur les routes. Mais cette hausse du nombre de véhicules s'est également accompagnée par une hausse du nombre d'accidents qui sont non seulement un danger pour les personnes mais qui ont également un coût élevé pour la société. Il est notamment estimé que l'ensemble des coûts liés aux accidents routiers représentent une perte de 2% du PIB pour les pays de l'Union Européenne (Plainis *et al.*, 2006). L'éclairage public est donc vu comme une solution pour réduire les accidents de nuit puisque des études estiment que le coût des accidents de nuit est trois fois supérieur aux coûts liés à l'éclairage des voiries (Beyer & Ker, 2010). Mais il reste à évaluer si oui ou non l'éclairage public des voiries a une réelle incidence sur le risque d'accident et quels sont les facteurs impactant ce risque.

Plusieurs études se sont penchées sur l'analyse de l'impact de l'éclairage des routes sur le risque d'accident. La littérature montre que le risque d'accident est plus élevé la nuit jusqu'à 40% (Johansson *et al.*, 2009a). La question est de savoir si l'éclairage public améliore ou non la situation. Les études récentes montrent qu'éclairer des routes qui ne l'étaient pas auparavant, diminue le nombre d'accidents d'environ 3% et ce en particulier aux intersections (Donnell *et al.*, 2010b; Struyf *et al.*, 2019). Des études plus anciennes étaient sujettes à des problèmes méthodologiques mais montraient des diminutions jusqu'à 30% en moins d'accidents et jusqu'à 44% en moins aux intersections (Elvik, 1995; Morin, 2005; Struyf *et al.*, 2019). Les problèmes méthodologiques de ces anciennes études ont été analysés grâce à des méta-analyses et sont notamment liés au manque de contrôle des autres variables jouant sur la sécurité routière, à des biais statistiques dans le choix des localisations des échantillons (Johansson *et al.*, 2009b; Struyf *et al.*, 2019). D'autres études montrent que l'extinction de lampadaire le long de routes illuminées n'est pas associée à une hausse du nombre d'accidents et il en est de même pour les opérations de dimming ou d'horaire d'extinction des lampadaires (Green *et al.*, 2015; Meesmann, 2018; Struyf *et al.*, 2019). Cela est également montré par un rapport de la Police Fédérale Belge (Dienst Strategische Analyse Federale Politie West-Vlaanderen, 2017). Il est également montré que ces effets ne dépendent pas de la localisation des routes soit en milieu rural soit en milieu urbain (Green *et al.*, 2015; Meesmann, 2018; Struyf *et al.*, 2019). Cependant, les études soulignent qu'une diminution de l'éclairage public peut être associée à une hausse des collisions avec les piétons (principalement en milieu urbain), ceux-ci étant plus difficilement discernables dans l'obscurité (Green *et al.*, 2015; Retting *et al.*, 2003; Struyf *et al.*, 2019). Le risque de collision pour un piéton est d'autant plus important qu'il y a de voitures garées le long des rues, celles-ci obstruant encore plus la visibilité des automobilistes (Retting *et al.*, 2003; Struyf *et al.*, 2019). Concernant le risque d'accident, les études montrent que la vitesse est le facteur le plus important et que des interviews d'automobilistes ont montrés qu'entre 58 et 68% des répondants

affirment dépasser les limitations de vitesse durant la nuit sur autoroute et en dehors des agglomérations (zone bâtie) (Meesmann, 2018; Struyf *et al.*, 2019).

D'autres variables semblent également jouer un rôle dans le risque d'accident, qui peuvent être soit liées à l'environnement (comme la luminosité), soit au conducteur. Ce mémoire s'intéresse principalement aux facteurs environnementaux et non liés au conducteur mais il est tout de même à noter que l'âge, le sexe, l'expérience, le niveau de fatigue, le stress (ou tout autre état comme l'anxiété), la prise d'alcool ou de produits stupéfiants et la façon de conduire du conducteur jouent un rôle sur le risque d'accident (Peden, 2004). Les variables environnementales ayant un impact sur le risque d'accident sont, notamment, les conditions météorologiques, les conditions lumineuses (jour/nuit), la présence d'une intersection, le type de route (la classe de route : autoroute, nationale ou routes locales), la localisation en zone bâtie (et la densité de celle-ci), la sinuosité, la courbure des virages, la largeur de la route, la vitesse autorisée, le volume de véhicule et enfin l'état de la route (type de revêtement, dégradation de la route, présence d'obstacle, ...) (Flahaut, 2004; Khattak *et al.*, 2024; Papadimitriou *et al.*, 2019; Theofilatos & Yannis, 2014). Des composantes plus locales, mais plus difficiles à identifier, rentrent également en compte comme la visibilité aux intersections ou encore la signalisation.

2.4. Criminalité et sentiment d'insécurité la nuit

Le sentiment d'insécurité la nuit est également un argument important contre l'extinction de l'éclairage public en milieu urbain (Sordello *et al.*, 2021). Bien que ce mémoire se concentre principalement sur le lien entre éclairage public et sécurité routière, la criminalité et le sentiment d'insécurité sont des facteurs importants à présenter.

La littérature montre que l'éclairage public est effectivement une solution dans la réduction du nombre de crimes (Cadena-Urzúa *et al.*, 2025; Tealde, 2022). La littérature montre également que l'obscurité, en ville, ou une réduction de l'éclairage public est associée à une hausse de la criminalité (Chalfin *et al.*, 2022a; Fotios *et al.*, 2021). Une étude menée sur la ville de Chicago, a montré que l'amélioration de l'éclairage public dans une rue ne diminue que légèrement le nombre de crimes mais qu'il y a un important déplacement vers les rues adjacentes (Chalfin *et al.*, 2022b). D'autres études (Steinbach *et al.*, 2015) ont montré qu'il n'est pas évident de montrer un lien entre une augmentation et une amélioration de l'éclairage public dans certaines rues et une diminution de la criminalité, puisque certains criminels profitent de l'amélioration des conditions lumineuses pour mieux repérer leurs opportunités et leur moyen de fuite.

Concernant, le sentiment d'insécurité, celui-ci semble être plus important chez les femmes et les personnes âgées (Sutton, 2004). Une augmentation et une amélioration de l'éclairage public la nuit réduisent le sentiment d'insécurité chez les deux publics les plus touchés, les femmes et les personnes âgées, et une réduction de l'éclairage public semble changer les comportements des citoyens (Boyce, 2003; Steinbach *et al.*, 2015; Sutton, 2004). Les personnes vont avoir tendance à éviter certaines zones comme des rues ou des parcs la nuit, à limiter leurs déplacements seules (et donc privilégier les déplacements accompagnés), à utiliser une voiture ou un taxi plutôt que la marche et à être plus attentives à leur environnement. Les études expliquent ces changements de comportements, par un manque de visibilité réduisant les possibilités d'anticiper un danger, mais aussi par une présence plus faible dans l'espace public réduisant l'effet de surveillance informelle où finalement chaque personne dans l'espace public va exercer un contrôle social réduisant la peur du crime et donc le sentiment d'insécurité (Boyce, 2003; Haans & de Kort, 2012; Marchant, 2004; Pain *et al.*, 2006; Peña-García *et al.*, 2015; Perkins *et al.*, 2015; Steinbach *et al.*, 2015; Struyf *et al.*, 2019). Pour autant, les citoyens interrogés, dans certaines études, soulignent malgré tout les effets positifs sur l'environnement et les économies possibles d'une

réduction de l'éclairage public (Haans & de Kort, 2012; Struyf *et al.*, 2019). Mais l'effet de la réduction de l'éclairage public sur la diminution de la présence dans l'espace public ne dépend pas uniquement de l'éclairage public, puisque les études mentionnent que la configuration de l'espace public joue un rôle crucial (Innes, 2017; Struyf *et al.*, 2019). Une étude mentionne également que le sentiment d'insécurité et la peur du crime ne proviennent pas que de la configuration de l'espace public et de l'éclairage de celui-ci mais également de la médiatisation de plus en plus importante des crimes et faits divers (Innes, 2017). Il est donc possible, qu'en milieu urbain, la réduction de l'éclairage public ait des effets négatifs sur la présence dans l'espace public et donc sur le sentiment d'insécurité.

Pour autant les études soulignent l'importance de réduire l'éclairage public pour des raisons environnementale et certaines proposent donc un usage raisonné de l'éclairage public en proposant notamment de nouvelles méthodologies pour le placement des lampadaires (Murray & Feng, 2016; Truter, 2026), mais également avec des solutions de dimming et d'allumage automatique des lampadaires lors de la présence d'une personne (Sordello *et al.*, 2021; Steinbach *et al.*, 2015).

2.5. Approche cas-contrôle

Pour répondre à la question « quel est l'effet de l'éclairage public sur le risque d'accident » ou à la question plus globale de « quels sont les facteurs influençant le risque d'accident », il a été décidé de mettre en place une approche cas-contrôle constituant en l'étude de cas, les localisations d'accidents répertoriés, et de points de contrôle, à savoir les localisations d'endroits sans accident sur la période étudiée. Une approche cas-contrôle permet donc de comprendre et d'étudier les facteurs influençant la présence ou l'absence d'un accident à une localisation donnée. La localisation des points de contrôle est choisie de manière aléatoire ce qui permet d'éviter tout biais systématique. (Vandenbulcke *et al.*, 2014)

Les approches cas-contrôles ont d'abord été largement développées dans le monde de la recherche médicale et plus particulièrement en épidémiologie avant d'être transposées depuis quelques années à une méthodologie pour l'analyse des facteurs de risques d'un événement rare (tel qu'un accident ou une maladie) (Vandenbulcke *et al.*, 2014). En épidémiologie, les cas sont par exemple des individus atteints d'une maladie et les contrôles sont des individus sains. Là aussi les études portent sur la détermination des facteurs qui influencent le risque d'événement (être atteint de la maladie ou avoir un accident).

Plusieurs études ont déjà appliqué l'approche cas-contrôle à des études de facteurs de risque d'accident. Une étude s'est notamment intéressée aux accidents de vélos à Bruxelles en prenant comme cas une collection d'accident de vélos et comme points de contrôle des localisations aléatoires sur le réseau cyclable de Bruxelles (Vandenbulcke *et al.*, 2014). De plus, la localisation des points de contrôle est stratifiée selon le volume potentiel de vélo par secteur statistique et le nombre de points de contrôle choisi ici est le nombre de cas multiplié par quatre. Cette étude a ensuite étudié plusieurs variables au moyen d'un modèle de régression logistique dont les paramètres estimés ont permis de créer une carte du risque d'accident à vélo. Les variables, identifiées par cette étude, augmentant le risque d'accident à vélo sont notamment la présence de rails de tram, les ponts, les intersections complexes, la proximité à un centre commercial, le partage de la route avec des camions et la présence de voitures garées le long des rues. D'autres études se sont intéressées plus spécifiquement à l'impact de l'éclairage public sur le risque d'accident via une approche cas-contrôle en associant chaque localisation d'accident nocturne (cas) à une localisation sans accident (point de contrôle) mais ayant les mêmes caractéristiques (même type de route, conditions lumineuses similaires, même typologie du milieu, etc, ...) (Gross, 2013; Gross & Donnell, 2011; Li *et al.*, 2021; Xin *et al.*, 2019). L'impact de l'éclairage public est donc étudié en ayant des localisations d'accidents avec et sans lampadaire. Ces études ont identifié un impact positif de l'éclairage public sur la sécurité routière avec une baisse jusqu'à 22% du risque d'accident aux

intersections (Gross & Donnell, 2011). Mais il est à noter que cette baisse est obtenue en passant d'une absence d'éclairage faible (moins de 5 lux) à un niveau d'éclairage moyen (entre 5 et 10 lux correspondant à une rue faiblement éclairée) puisque les effets supplémentaires de l'éclairage public à des niveaux d'éclairement plus élevés sont réduits. Le risque de collision avec les piétons est également plus faible, selon ces études, puisque les piétons deviennent alors plus visibles.

2.6. Modèle de régression logistique et modèle Random Forest

Le modèle de régression logistique permet de réaliser une régression lorsque la variable dépendante est binaire (accident ou pas d'accident) (El Sanharawi & Naudet, 2013a; Rakotomalala, 2015). Ce modèle prédit alors une probabilité d'accident entre 0 et 1 pouvant être interprété comme le risque d'accident. Le modèle de régression logistique va estimer l'effet des variables (paramètres estimés, β) sur le risque d'accident. Les odds ratio (soit les ratios de chance suivant la valeur de la variable explicative) se calculent en prenant l'exponentiel de ces paramètres estimés. Plusieurs études ont déjà utilisé une régression logistique pour identifier les variables jouant un rôle dans le risque d'accident et ensuite pouvoir prédire le risque d'accident basé sur les paramètres estimés du modèle. Notamment, l'étude déjà citée concernant les accidents de vélos à Bruxelles a utilisé ce type de modèle (Vandenbulcke *et al.*, 2014).

Le modèle *Random Forest* fait partie de la famille des modèles de machine learning utilisant les arbres de décision aléatoire. C'est un modèle popularisé par Breiman au début des années 2000 (Biau & Scornet, 2016), faisant partie de la famille des modèles de machine learning dits à procédure supervisée, qui a la particularité de se montrer robuste tout en étant flexible. Un modèle Random Forest peut être utilisé autant pour réaliser de la classification supervisée que de la régression. Ce modèle a l'avantage de pouvoir prendre en considération l'aspect non-linéaire et non-monotonique des variables, contrairement à un modèle de régression linéaire ou logistique (Biau & Scornet, 2016). Le fonctionnement plus détaillé d'un modèle Random Forest est explicité dans la section correspondante dans la partie méthode.

Plusieurs études ont déjà utilisé un modèle Random Forest pour étudier le risque d'accident, notamment une étude sur 49 états des États-Unis (Yamarthi *et al.*, 2025). Le modèle Random Forest s'est alors montré particulièrement efficace pour déterminer les zones à risque d'accident élevé. La méthode reposait, comme pour une régression logistique, sur l'utilisation d'une collection d'accidents avec leur localisation. L'étude a donc prédit le risque d'accident basé sur des informations environnementales telles que l'heure de l'accident, les conditions météorologiques, la distance à une intersection et le type de celle-ci. Le but étant alors dans ce mémoire de prédire le nombre d'accidents probable, selon l'heure de la journée, les conditions météorologiques et la localisation (Yamarthi *et al.*, 2025).

3. Matériel et méthode

3.1. Zone d'étude

L'étude se focalise principalement sur les cinq communes constituant le Parc national de l'Entre-Sambre-et-Meuse soit Couvin et Viroinval dans la province de Namur ainsi que Chimay, Froidchapelle et Momignies dans la province du Hainaut (Figure 1). Cependant, pour calibrer au mieux le modèle, la zone d'étude est élargie en prenant les communes limitrophes : Sivry-Rance et Beaumont dans la province du Hainaut ainsi que Walcourt, Cerfontaine, Philippeville et Doische dans la province de Namur. Cet élargissement sert à avoir un plus grand échantillon d'accidents permettant une meilleure calibration du modèle tout en gardant une zone d'étude avec des communes et des territoires ruraux à semi-ruraux.

3.1.1. Parc national de l'Entre-Sambre-et-Meuse

Il existe plusieurs définitions pour définir ce qu'est un Parc national dont la plus mondialement acceptée est celle de l'UICN (Union Internationale pour la Conservation de la Nature) qui propose comme définition pour les parcs nationaux « Vastes aires naturelles ou quasi naturelles mises en réserve pour protéger des processus écologiques de grande échelle, ainsi que les espèces et les caractéristiques des écosystèmes de la région, qui fournissent aussi une base pour des opportunités de visites de nature spirituelle, scientifique, éducative et récréative, dans le respect de l'environnement et de la culture des communautés locales» (Leung *et al.*, 2018).

En 2021, la Région Wallonne a lancé un appel à projet visant à créer deux parcs nationaux sur son territoire. De cet appel à projet, une sélection s'est opérée après avoir initialement retenu 4 projets sur 7 proposés. Les deux lauréats de cet appel à projet ont été annoncés fin 2022 et sont le Parc national de l'Entre-Sambre-et-Meuse et le Parc national de la Vallée de la Semois qui ont pu démarrer leurs activités début 2023. Ces deux parcs nationaux ont pour but de préserver et conserver au maximum l'environnement de ces territoires reconnus pour leur biodiversité exceptionnelle. De plus les parcs nationaux de Wallonie répondent à un enjeu de tourisme durable en créant une offre touristique basée sur la nature et le respect de l'environnement. Cela participe au rayonnement international de la région Wallonne souhaitant s'inscrire dans une offre touristique globale s'articulant avec les parcs naturels et les points naturels touristiques (SPW, 2022).

Plus particulièrement le Parc national de l'Entre-Sambre-et-Meuse agit sur un territoire continu (périmètre du Parc national) de 22500 hectares au travers des 5 communes principales de cette étude (Couvin, Viroinval, Chimay, Froidchapelle et Momignies) (Figure 1). Ce périmètre regroupe des zones naturelles ayant une haute valeur biologique incluant des zones Natura 2000, des réserves naturelles agréées, des réserves forestières, des réserves naturelles domaniales et des zones humides d'intérêt biologique. Ce périmètre reprend, en plus de ces zones, d'autres zones naturelles pouvant être considérées comme à haute valeur biologique. L'un des projets du Parc national de l'Entre-Sambre-et-Meuse vise l'instauration d'une Trame noire sur son territoire passant donc par une réduction de la pollution lumineuse et une meilleure gestion de l'éclairage public.

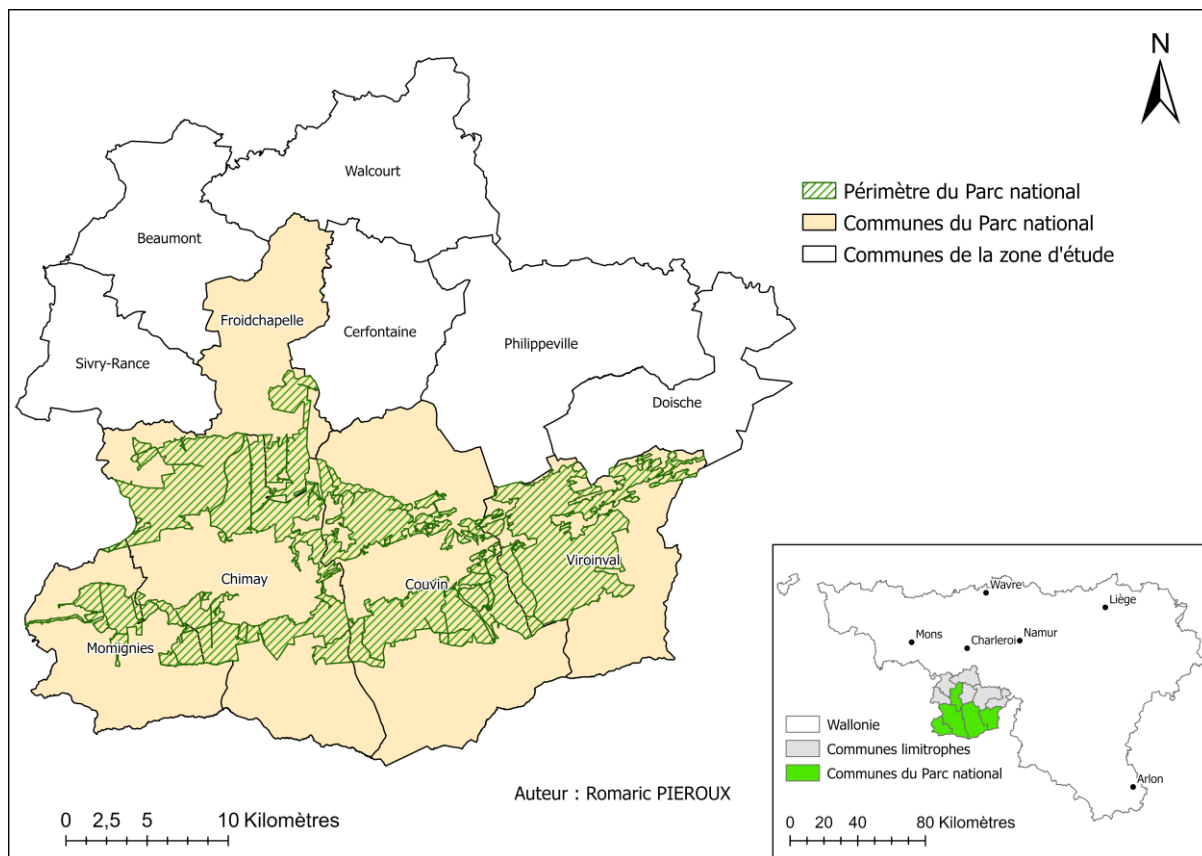


Figure 1 : Localisation du Parc national, des communes participantes et des communes limitrophes

3.1.2. Description du territoire

Le territoire étudié est principalement constitué d'espaces ruraux avec des densités de populations relativement faibles par rapport au reste du territoire wallon. La densité de population des communes du Parc national est de 53,6 hab/km² alors que la densité de la Wallonie est de 219,2 hab/km². L'ensemble des communes de la zone d'étude, à l'exception de la commune de Walcourt, a un degré de ruralité supérieur à 90% selon l'indicateur de ruralité du Service Public de Wallonie tel que visible en annexe (Figure A. 1) (SPW, 2021). La zone d'étude globale s'étend sur 1314,94 km² ou 699,37 km² en ne prenant que les communes du Parc national. Un détail important à prendre en compte est que la zone d'étude contient un site touristique majeur de la Wallonie, les Lacs de l'Eau d'Heure, situé sur les communes de Froidchapelle et de Cerfontaine.

En ce qui concerne le réseau routier, le territoire est relié du nord au sud par la N5 constituant l'axe majeur de la région. Cet axe relie Charleroi à Charleville-Mézières. Cet axe n'est pas considéré comme une autoroute en tant que telle mais comme une grande route. L'axe est tout de même considéré comme un axe structurant du réseau routier wallon. Le côté sud de la N5 (comprenant les communes de Couvin, Philippeville et Cerfontaine) est une route avec un gabarit autoroutier, avec peu d'intersection et où celles-ci sont construites avec des bretelles d'accès et non avec des intersections directes, mais le nord de cet axe suit une structure de nationale classique avec des intersections directes, mais l'axe reste un axe prioritaire. Le territoire est ensuite traversé d'ouest en est par plusieurs nationales comme la N99 entre Macquenoise (dans le sud de la commune de Momignies à la frontière avec la France), Chimay, Couvin et Viroinval (Nismes) pour rejoindre la N40 à Doische. La N40 passe entre Beaumont, Philippeville, Doische et Givet (F). La N97 passe entre Philippeville et Dinant (en dehors de la zone d'étude). Une nationale (la N53) rejoint Chimay à Beaumont selon un axe nord-sud. Ces nationales

représentent la structure locale du réseau routier et sont donc des routes assez empruntées par les véhicules. Le reste du réseau routier de la zone d'étude est constitué de petites nationales (NXXX) et de routes locales. Le réseau routier simplifié est visible sur la Figure 2.

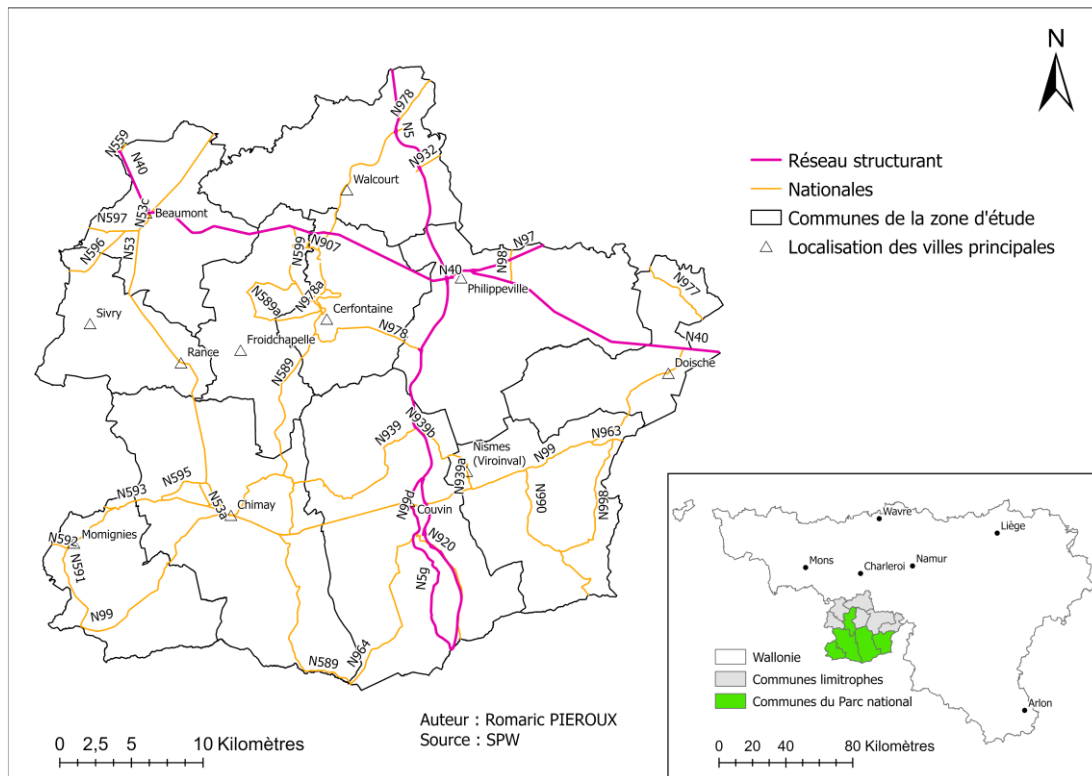


Figure 2 : Réseau routier de la zone d'étude avec le réseau structurant (N5 et N40) et le réseau de nationales selon le SPW

3.2. Données

Accidents routiers

Les données d'accident de la route proviennent de *Statbel*, l'institut national de statistique de Belgique (Statbel, 2025). Lorsque la police se rend sur les lieux d'un accident de la route, celle-ci doit écrire un procès-verbal reprenant diverses informations comme la présence ou non d'un carrefour, les conditions météorologiques, les conditions de la route (humide ou sec), les conditions lumineuses (de jour, de nuit avec éclairage public ou de nuit sans éclairage public à proximité), en agglomération ou hors agglomération, le type de route, la gravité de l'accident, le type de véhicule impliqué, le type de collision, la présence et le type d'obstacle. Ces accidents sont également géolocalisés permettant des analyses SIG. Ne sont donc pas repris dans cette base de données les accidents n'ayant occasionnés que des dégâts matériels et les accidents où la police n'est pas intervenue. Selon *Statbel*, les accidents n'impliquant que des blessés légers sont les plus sous-estimés et plusieurs études pointent le fait que seul 20% des accidents impliquant des personnes très légèrement blessées, soit sans besoin d'hospitalisation, sont répertoriés. Cela est notamment dû au fait que la police n'est pas toujours appelée pour des accidents mineurs et où un règlement à l'amiable est possible. Le taux d'enregistrement des accidents impliquant des hospitalisations est de plus ou moins 50%. Les accidents mortels sont répertoriés à plus ou moins 90% suite à des enquêtes judiciaires plus précises et plus intenses (Elvik & Mysen, 1999; Statbel, 2025).

Les données disponibles s'étendent de 2017 à 2024. Au total, 1968 accidents ont été répertoriés pour l'ensemble de la zone d'étude et 853 accidents sur les communes du Parc national. Cependant, suite à des problèmes de géolocalisation, connus de *Statbel*, seuls 1451 accidents (soit 73,73%) peuvent être

utilisés pour l'ensemble de la zone d'étude. En effet, certains accidents ont été mal géoréférencés rendant impossible les manipulations SIG et d'autres ne possèdent tout simplement pas de géoréférencement.

Étant donné que le mémoire porte uniquement sur la sécurité routière de nuit, et donc sur le risque d'accident de nuit, seuls les accidents de nuit sont utilisés dans ce mémoire, soit 420 accidents pour les communes du Parc national et les communes limitrophes. En réalité, 400 accidents seront utilisés dans ce mémoire, puisque 20 accidents ont eu lieu sur des routes non prises en compte, mais ce point est détaillé dans la section présentant la classification du réseau routier. Il est aussi à préciser que ce sont les rapports de police concernant les accidents qui définissent si un accident a eu lieu de nuit ou non. Sur l'ensemble des accidents étudiés de nuit, aucun accident n'a eu lieu à proximité d'un éclairage public présent mais éteint, chaque accident a soit eu lieu à proximité d'un éclairage public présent et allumé soit à un endroit sans éclairage public.

Données de trafic routier TomTom « floating car data »

TomTom, une entreprise privée, recueille chaque jour des millions de données provenant de véhicules via les GPS, les systèmes de navigation embarqués et certaines applications de navigation. Ces « floating car data » de *TomTom* concernent plus de 600 millions de véhicules dans le monde soit 1 véhicule sur 4 et analysent les déplacements en calculant à intervalles réguliers la position du véhicule ce qui permet de connaître la position du véhicule, sa vitesse, son sens de déplacement ainsi que des données temporelles. En regroupant les données de plusieurs véhicules et en les agrégeant par tronçon, il est notamment possible de connaître le nombre de véhicules par tronçons pour une période donnée ainsi que de connaître la vitesse moyenne (TomTom, 2024). TomTom permet un essai gratuit mais limité à ces données, en ne rendant disponible que les données récoltées durant le mois d'août 2024, qui sont donc utilisées dans ce mémoire. Malgré cette contrainte, ces statistiques restent intéressantes dans le but d'étudier l'impact du volume de véhicule sur le risque d'accident. Le volume de véhicules est également utilisé pour classer les routes. Les données issues de TomTom sont également utilisées afin de connaître la vitesse maximale pour chaque tronçon ainsi que la vitesse moyenne, médiane et les percentiles de 5 à 95 (par palier de 5) des véhicules empruntant ces tronçons. Les tronçons créés par TomTom sont de taille variable, dépendant de la distance entre deux intersections. Le réseau de tronçons ainsi obtenu depuis les données TomTom n'est pas précis et reprend, en plus, des routes privées, des chemins forestiers et des chemins agricoles ne faisant pas partie du réseau routier réellement utilisé par l'ensemble de la population.

Les données TomTom concernant le volume de véhicules et la vitesse moyenne par tronçon ont été utilisées selon deux plages horaires différentes. La première plage horaire concerne le mois d'août 2024 dans sa totalité et les journées entières. Cette première plage reprend donc la totalité du volume de véhicules sur le mois d'août 2024. La seconde plage horaire concerne également l'ensemble du mois d'août 2024 mais ne reprend que les données de nuit. La nuit a été ici définie entre 21H00 et 7H00 pour reprendre la nuit mais aussi le crépuscule et le lever du soleil pendant un mois d'août en Belgique.

Cartographie du réseau routier-PICC

Dans le but d'identifier la proximité des accidents de la route à différents types d'intersections, les données provenant du PICC (Projet Informatique de Cartographie Continue) sont utilisées. Le PICC est un projet du SPW visant à cartographier les éléments identifiables du territoire Wallon avec une haute précision de 25 cm. Le réseau routier est cartographié de manière assez complète en reprenant le réseau routier sous forme de polygones dont la largeur respecte la taille réelle des tronçons routiers. Cette première représentation du réseau routier est aussi accompagnée d'informations (toujours sous forme de polygones) comme la localisation des aires de repos, des carrefours et des ouvrages d'art mais seuls les

carrefours sont utilisés dans ce mémoire. Le réseau routier est également représenté sous forme de ligne reprenant les bords des voiries et d'autres informations non utilisées dans ce mémoire. Enfin, la représentation du réseau routier se fait également sous forme ponctuelle en reprenant également les intersections (mais en détaillant s'il s'agit d'un carrefour ou d'un rond-point) mais aussi en reprenant la localisation des poteaux le long des routes sans détailler s'il s'agit de poteau d'électricité ou de poteau d'éclairage. La réduction du nombre de voie est également considérée comme une intersection.

Localisation des lampadaires d'éclairage public

L'éclairage public est, comme la gestion du réseau routier, une compétence régionale. Cependant, les régions ne gèrent pas l'ensemble du réseau routier puisque celui-ci est partagé également avec les communes donnant lieu à des routes communales et des routes régionales. Les autoroutes et les principales routes nationales (type N4, N5, ...) forment le réseau structurant de la région Wallonne et sont gérées par la SOFICO (Société de financement complémentaire des infrastructures) qui est un organisme public de la région Wallonne. Le réseau secondaire est, par contre, géré directement par le SPW (Mobilités et infrastructures) et concerne les nationales plus petites (type NXX et NXXX) reliant les communes entre elles. Enfin, le réseau communal représente le maillage final du réseau routier et est géré par les communes et représente la plus grande partie en nombre de kilomètres du réseau routier Wallon. Ce réseau local relie les habitations et zones économiques aux autres axes du réseau routier. Le SPW et la SOFICO gèrent eux-mêmes l'éclairage public situé le long de leurs axes routiers (SPW, 2025). Le SPW met à disposition une couche cartographique reprenant l'ensemble des poteaux d'éclairage public régional, cette couche est donc censée reprendre les poteaux d'éclairages public pour l'ensemble du réseau routier à l'exception du réseau communal. Cependant, les poteaux d'éclairage public gérés par la SOFICO ne sont pas encore repris, ce qui a nécessité de cartographier manuellement chaque poteau d'éclairage manquant sur le réseau structurant à l'aide d'orthophotos et l'usage de Streetview de Googlemaps. Les communes délèguent la plupart du temps la gestion de l'éclairage public à des gestionnaires de réseau de distribution. Ces gestionnaires de réseau de distribution sont, pour la zone étudiée, ORES pour les communes de Doische, Philippeville et Walcourt, AIESH pour les communes de Beaumont, Chimay, Couvin, Froidchapelle, Momignies et Sivry-Rance ainsi que AIEG pour la commune de Viroinval. Les gestionnaires de réseau allument en temps normal l'éclairage public entre le coucher du soleil et le lever du soleil bien que des adaptations locales peuvent exister et des extinctions temporaires peuvent également avoir lieu comme en 2021-2022 lors de la crise énergétique.

Couverture du sol et zones naturelles à haute valeur biologique

La couverture du sol utilisée dans ce mémoire est celle provenant du Corine Land Cover 2018 qui est un projet pan-européen visant à catégoriser la couverture du sol de l'Europe par satellite. La couverture du sol est ici utilisée pour identifier les zones naturelles qui pourraient être impactées par la pollution lumineuse. La couche utilisée dans ce mémoire est celle de 2018, soit la plus récente actuellement disponible et est sous forme vectorielle. La classification de la couverture du sol est basée sur une méthode de classification semi-automatisée basée sur des images satellites de Sentinel-2 avec une résolution allant de 10 mètres à 60 mètres selon les bandes spectrales. Malgré une bonne précision, quelques limites apparaissent notamment à une échelle locale puisqu'il y a une unité minimale de cartographie (Minimal Mapping Unit) de 25 ha. Toute surface comme un bois ou une prairie de moins de 25 hectares n'est donc pas considérée comme une unité propre et est absente de la cartographie. Cela est justifié comme une forme de généralisation cartographique évitant de cartographier chaque détail. (Büttner *et al.*, 2017; Copernicus, 2021; Mingarro & Lobo, 2023).

Les zones naturelles à haute valeur biologique concernent notamment les zones Natura 2000, les réserves naturelles agréées, les réserves forestières, les réserves naturelles domaniales et les zones

humides d'intérêt biologique. Ces zones sont identifiées et localisées grâce au répertoire du Service Public de Wallonie. En plus de ces zones, le périmètre du Parc national, qui reprend en partie les zones précitées et des zones naturelles supplémentaires, est également considéré comme une zone naturelle à haute valeur biologique.

3.3. Méthodes

Étapes principales

La première partie de la méthode est une analyse descriptive des localisations des éclairages, de leur lien avec les zones naturelles, et des localisations des accidents. Cette première partie comprend la construction d'une carte présentant brièvement les impacts potentiels de l'éclairage public sur l'environnement. Il s'agit de décrire les endroits où un conflit existe entre les zones naturelles (dont principalement les zones naturelles à haute valeur biologique) et les routes éclairées. Cette première partie concerne également l'exploration descriptive des caractéristiques des accidents. Il s'agit donc d'une exploration au moyen de graphiques et cartes pour présenter les résultats de base avant de passer à la seconde partie avec l'identification des facteurs influençant le risque d'accident. Les variables construites dans ce modèle sont également brièvement décrites.

La deuxième partie s'intéresse donc à l'identification des variables jouant un rôle dans le risque d'accident et notamment l'éclairage public. Pour ce faire une approche cas-contrôle est créée avec les localisations d'accidents réels provenant de Statbel constituant donc les cas et les localisations d'endroits sans accidents sur la période d'étude (2017-2024) comme points de contrôle. Le but du modèle est de comprendre les facteurs expliquant la présence d'un accident à certains endroits et leur absence à d'autres. La localisation des points de contrôle est aléatoirement choisie sur le réseau routier mais est stratifiée par rapport au volume de véhicules dans trois classes de route. Ensuite, chaque cas et point de contrôle est associé à des variables telles que la localisation dans une zone bâtie ou non, la proximité à un éclairage public ou non, la proximité à une intersection ou non, la largeur du tronçon, la sinuosité du tronçon, le volume de véhicules de nuit du tronçon et la vitesse moyenne des véhicules de nuit du tronçon ainsi que la classe de la route (tel que construite dans ce mémoire). Ces variables ont été choisies, puisque comme montré dans la revue de la littérature, elles ont une influence ou peuvent avoir une influence sur la probabilité d'accident. Plusieurs modèles de régression logistique et Random Forest sont utilisés pour déterminer les variables significatives et leur impact sur le risque d'accident.

Ensuite, le modèle principal, reprenant l'ensemble des variables et des cas/points de contrôle, sert à construire une carte du risque d'accident basée sur les paramètres estimés des variables significatives.

Ensuite, un autre modèle de régression logistique est également créé en prenant la localisation d'un poteau d'éclairage public comme variable dépendante, et ce dans le but de vérifier si les localisations ayant répertoriés un accident ont plus de chance d'avoir la présence d'un poteau d'éclairage public, ce qui pouvant suggérer un risque d'endogénéité.

Les méthodes reposent sur l'utilisation de *RStudio* (version : 4.5.2) avec notamment les principaux packages « tidyverse », « randomForest » et « ggplot2 ». *RStudio* est utilisé uniquement pour les modèles de régression logistique et les modèles Random Forest ainsi que la création des graphiques. Le logiciel *ArcGIS* (version 3.1.0) est utilisé pour les manipulations SIG et la création des cartes.

La Figure 3 montre un résumé de la méthode, soit le « workflow » utilisé.

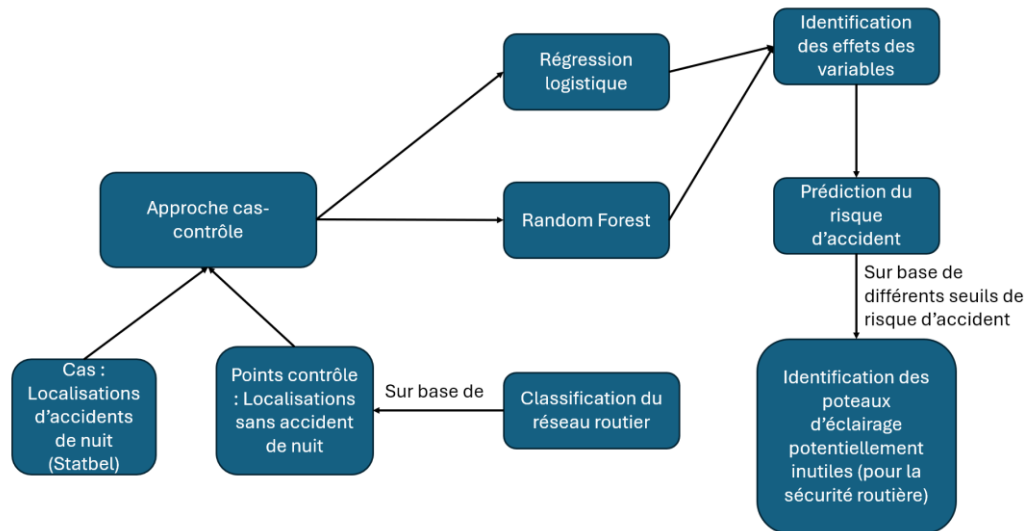


Figure 3 : Résumé de la méthode

3.3.1. Impact potentiel de l'éclairage public sur l'environnement

L'impact de l'éclairage public sur l'environnement se fait via une description d'une carte présentant les zones à haute valeur biologique qui pourraient être impactées par l'éclairage public et donc la pollution lumineuse. L'impact sur l'environnement n'est pas ici étudié plus en profondeur puisqu'il constitue l'enjeu qui sous-tend et justifie l'étude menée dans ce mémoire, mais celui-ci vise bien à étudier la contribution ou non de l'éclairage public à la sécurité routière.

Les zones à haute valeur biologique, constituant donc les zones les plus sensibles à la pollution lumineuse, sont notamment les zones Natura 2000, les réserves naturelles agréées, les réserves forestières, les réserves naturelles domaniales et les zones humides d'intérêt biologique. Le périmètre strict du Parc national, tel que visible sur la Figure 1, reprend en partie ces zones à haute valeur biologique mais reprend également d'autres zones n'étant pas déjà dans des pratiques de conservation mais qui sont ici considérées comme des zones importantes pour l'environnement. En plus de ces zones qui peuvent déjà être qualifiées de haute valeur biologique, une analyse descriptive des impacts de la pollution lumineuse est également réalisée sur d'autres milieux, non couverts par les mesures précitées, tels que les forêts et les prairies et ce via la couverture du sol du Corine Land Cover permettant d'identifier ces autres zones naturelles.

3.3.2. Approche cas-contrôle visant à identifier les facteurs influençant le risque d'accident : hypothèses, échantillonnage, et création de la base de données

3.3.2.1. Hypothèses méthodologiques

Plusieurs hypothèses méthodologiques ont été posées dans ce mémoire :

1) L'éclairage des routes

Une hypothèse importante a été formulée concernant l'éclairage des routes : l'ensemble des lampadaires d'éclairage public est allumé pendant la nuit excluant donc la possibilité d'extinction de certains

lampadaires et excluant aussi la possibilité d'un horaire d'extinction pendant la nuit par exemple entre minuit et 5H00. Cela est notamment dû à l'indisponibilité d'informations sur l'utilisation précise des lampadaires d'éclairage public.

2) La durée de la nuit

Comme déjà précisé, la nuit a ici été définie entre 21H00 et 7H00 pour reprendre la totalité d'une nuit avec le coucher du soleil (crêpuscule) et le lever du soleil (aube) d'un mois d'août en Belgique. Bien que cela puisse être une hypothèse importante, la définition de la nuit dans ce mémoire n'est pas aussi cruciale qu'il n'y paraît. En effet, comme les accidents ont déjà été catégorisés comme étant de nuit ou de jour par les services de police (et qui par conséquent prend en compte la saisonnalité dans la durée de la nuit), la définition de la nuit n'est déterminante dans ce mémoire que pour les variables concernant le volume de véhicules de nuit et la vitesse moyenne de nuit. Et comme la définition de la nuit dans ce mémoire exclue les heures de pointe, une différence d'une heure ou plusieurs dans la définition du début ou de la fin de la nuit ne va pas avoir une grande incidence sur ces variables puisque les volumes de véhicules sont déjà faibles de nuit. La définition de la nuit n'est pas non plus importante puisqu'il est fait comme hypothèse que l'ensemble de l'éclairage public est allumé toute la nuit, il n'y a donc pas de problèmes liés à une temporalité entre des accidents de nuit ou des points de contrôle étant à proximité d'éclairage public éteint ou allumés (d'autant plus qu'aucun accident n'a eu lieu à proximité d'un éclairage public éteint). Le fait de ne pas inclure la saisonnalité dans la durée de la nuit dans ce mémoire n'est donc pas un problème en soi.

3.3.2.2. Hypothèses testées par les modèles de régression logistique

Plusieurs hypothèses sont posées et sont testées grâce aux modèles de régression logistique. Ces hypothèses sont liées aux différentes variables étudiées en vue d'identifier les variables impactant le risque d'accident. **L'hypothèse principale à tester dans ce mémoire est « la proximité à de l'éclairage public a une influence sur le risque d'accident » et qui est ensuite dérivée avec « l'effet de l'éclairage public, s'il y a se concentre aux intersections ».**

Les hypothèses à tester pour les autres variables sont « être ou non en zone bâtie a une influence sur le risque d'accident », « La largeur des routes a une influence sur le risque d'accident », « La sinuosité des routes a une influence sur le risque d'accident », « la proximité à une intersection a une influence sur le risque d'accident », « la vitesse a une influence sur le risque d'accident », « le volume de véhicules a une influence sur le risque d'accident » et « la classe de route (autoroutes, nationales ou routes locales) a une influence sur le risque d'accident ». Une autre hypothèse est également testée pour vérifier s'il y a de l'endogénéité entre la présence d'un éclairage public et la présence d'un accident, l'hypothèse est alors « **les localisations ayant répertoriés un accident ont plus de chances d'avoir la présence d'un poteau d'éclairage public** ». Pour cette hypothèse, la variable dépendante change et devient donc la présence d'un éclairage public.

3.3.2.3. Création des points de contrôle

Les points de contrôle sont donc des points aléatoires où il n'y a pas eu d'accidents sur la période étudiée (2017-2024). Les points de contrôle sont choisis aléatoirement sur le réseau routier mais sont répartis entre les 3 classes de route selon leur volume de véhicules.

Le ratio cas/points de contrôle choisi ici est 1:2 (deux fois plus de points de contrôle que de cas), la classe 3 ayant une longueur restreinte (88,65 km) il n'est possible de créer plus de points de contrôle. Comme il y a 400 cas d'accidents étudiés il y a donc 800 points de contrôle créés. La répartition des points de contrôle est stratifiée sur le réseau routier selon le volume de véhicules par classe (Table 1).

De ce fait comme la classe de route 1 contient 17,86% du nombre de véhicules total, il y a donc 143 points de contrôle pour cette classe de route. La classe 2 contient 29,24% du nombre de véhicules total et donc 234 points de contrôle. La classe 3, qui contient la plus grande proportion du volume de véhicules (52,90%) contient également le plus de points de contrôle soit 423. Cette stratification permet d'étudier au mieux pourquoi la classe 3 qui, avec la plus grande proportion du volume de véhicules totale, a la plus faible proportion d'accidents. La localisation des points est ensuite choisie aléatoirement sur les tronçons de chaque classe en laissant un espace de minimum 100 mètres entre eux et de 100 mètres avec les accidents.

Table 1 : Répartition des 800 points de contrôle entre les 3 classes de route

Classe de route	% du nombre de véhicules total sur le mois d'août 2024	Nombre de points de contrôle
1 : Routes à faible trafic	17, 86%	143
2 : Routes à trafic moyen	29, 24%	234
3 : Routes à trafic élevé.	52, 90%	423

3.3.2.4. Présentation des variables étudiées

Classification du réseau routier

Le réseau routier a été classifié en 3 classes selon le volume de véhicules pour chaque tronçon grâce aux données GPS TomTom pour le mois d'août 2024. La classification s'est réalisée en prenant l'ensemble des données, de jour comme de nuit, malgré que l'étude porte uniquement sur les accidents de nuit et ce dans le but de classer le réseau routier de manière générale. Cependant, dans les variables utilisées dans les modèles statistiques, seules les données (volume de véhicules et vitesse moyenne) de nuit sont utilisées. Le choix des seuils s'est effectué en comparant les seuils obtenus selon plusieurs méthodes : méthode de Jenks, méthode des K-means et une méthode par clustering (avec la méthode « W.Ward »). Les seuils ont ensuite été légèrement adaptés pour correspondre au mieux à la réalité du réseau routier, c'est-à-dire, éviter les coupures entre des tronçons successifs proches. Le réseau routier est donc classifié dans le cadre de ce mémoire sans prendre en compte d'une classification qui aurait déjà été réalisée comme, par exemple, la distinction entre routes communales et régionales ou la distinction entre routes locales et autoroutes réalisées par le SPW, et ce pour avoir une méthode de classification homogène basée sur le volume de véhicules.

Les tronçons ayant moins de 62 véhicules pour le mois d'août 2024 (soit moins de 2 véhicules par jour) ont été retirés, d'une part pour avoir une significativité suffisante d'au minimum 30 observations par tronçons et d'une autre part car le seuil de 62 véhicules correspond au 40^{ème} percentile de la distribution des tronçons par nombre de véhicules. Donc, 40% des tronçons du réseau TomTom ont eu moins de 62 véhicules sur le mois d'août de 2024. Ces tronçons concernent notamment des chemins agricoles, des chemins forestiers (souvent interdits à la circulation en dehors des engins forestiers) ou encore des chemins privés qui ont été cartographiés uniquement sur base d'une cartographie réalisée en interne chez TomTom et non d'une cartographie officielle des voiries publiques (communales et régionales). Ces tronçons sont donc retirés pour garder une bonne significativité statistique permettant une meilleure calibration du modèle du risque d'accident. La distribution des tronçons par rapport à leur volume de véhicules est asymétrique avec un nombre important de tronçons ayant peu de volume de véhicules et peu de tronçons ayant un volume de véhicules relativement élevé (Figure 4).

La première classe de route choisie concerne les routes à faible trafic et comprend les tronçons ayant eu entre 62 et 5000 véhicules sur le mois d’août 2024. Cette classe contient donc le réseau local communal et quelques nationales de type NXXX. La deuxième classe reprend les tronçons ayant entre 5001 et 30000 véhicules, ce qui correspond à des routes avec un trafic moyen tel que des nationales de type NXX. La dernière classe de route contient les tronçons ayant plus de 30.001 véhicules jusqu’à 84.276 (qui constitue le tronçon avec le plus de véhicules sur le mois d’août 2024). La classe de route 3 correspond aux routes ayant un trafic élevé comme les nationales de type NX comme la N5.

Au total, le réseau routier étudié est constitué de 2700303 mètres de routes dont 2307631 mètres pour le réseau 1 (faible trafic routier), 304020 mètres pour le réseau 2 (trafic routier moyen) et 88276 mètres pour le réseau 3 (trafic routier élevé). Ces longueurs comprennent chaque sens de circulation. Bien que la classe 3 soit celle qui comporte le réseau routier le plus petit, cette classe concentre un volume de véhicules important avec 34606908 véhicules en août 2024. Le réseau 2 a eu volume de 19130245 véhicules tandis que le réseau 1, le réseau routier avec la longueur de réseau la plus élevée, a eu un volume 11679375 véhicules soit le volume le plus faible. En prenant l’indice véhicule-kilomètre qui consiste à diviser le nombre de véhicules par la longueur du réseau, la classe 3 a donc l’indice le plus élevé avec 390,97 véhicules/kilomètre, la classe 2 a un indice de 62, 92 véhicules/kilomètre tandis que la classe 1 a donc l’indice le plus faible avec 5,07 véhicules/kilomètre.

Le total du volume de véhicule est de 65416528 véhicules pour le mois d’août 2024 dont 52,9% pour le réseau 3, 29,24% pour le réseau 2 et 17,86% pour le réseau 1. Bien que le réseau 3 concentre la plus grande proportion de véhicules, ce réseau possède le nombre d’accidents le plus petit avec 46 accidents sur la période 2017-2024. Le réseau 2 arrive en deuxième position avec 152 accidents vient en dernier le réseau 1 avec 202 accidents. Il y a 20 accidents ayant eu lieu sur des tronçons avec moins de 62 véhicules ce qui porte à 400 le nombre d’accidents réellement utilisés dans ce mémoire. La Table 2 résume les caractéristiques des différentes classes de route.

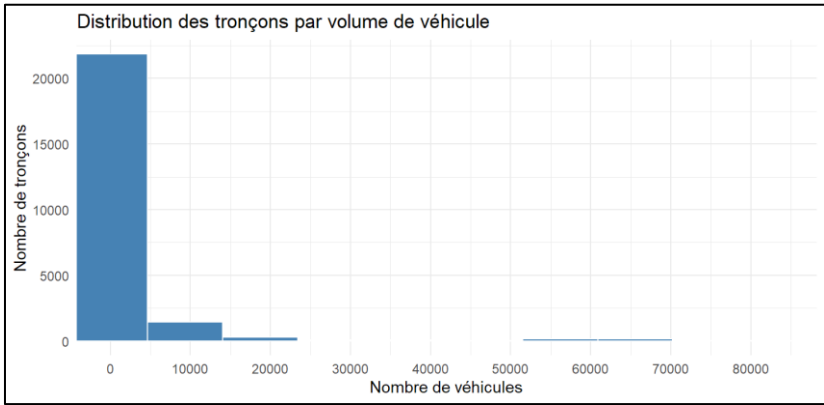


Figure 4 : Distribution des tronçons par volume de véhicules. Les données proviennent de TomTom et couvrent le mois d'août 2024.

Table 2 : Classification du réseau routier et caractéristique des classes de route ainsi créées

Classe de route	1 : Routes à faible trafic	2 : Routes à trafic moyen	3 : Routes à trafic élevé
Seuil (basé sur le volume de véhicules par tronçon)	62 – 5000	5001 – 30 000	30 001 – 84 276
Longueur du réseau (en mètres)	2307631	304020	88652

Total du nombre de véhicule sur le mois d'août 2024	11679375	19130245	34606908
%du nombre de véhicules total sur le mois d'août 2024	17, 86%	29, 24%	52, 90%
Véhicules/kilomètre	5, 07	62, 92	390, 37
Nombre d'accidents de nuit (2017-2024)	202 (50,5%)	152 (38%)	46 (11,5%)

Le réseau routier, tel que classifié dans ce mémoire, est visible sur la Figure 5. Où il est possible de voir que la classe 3 (en rouge) reprend bel et bien la N5 et un tronçon de la N40 à l'entrée de la ville de Philippeville.

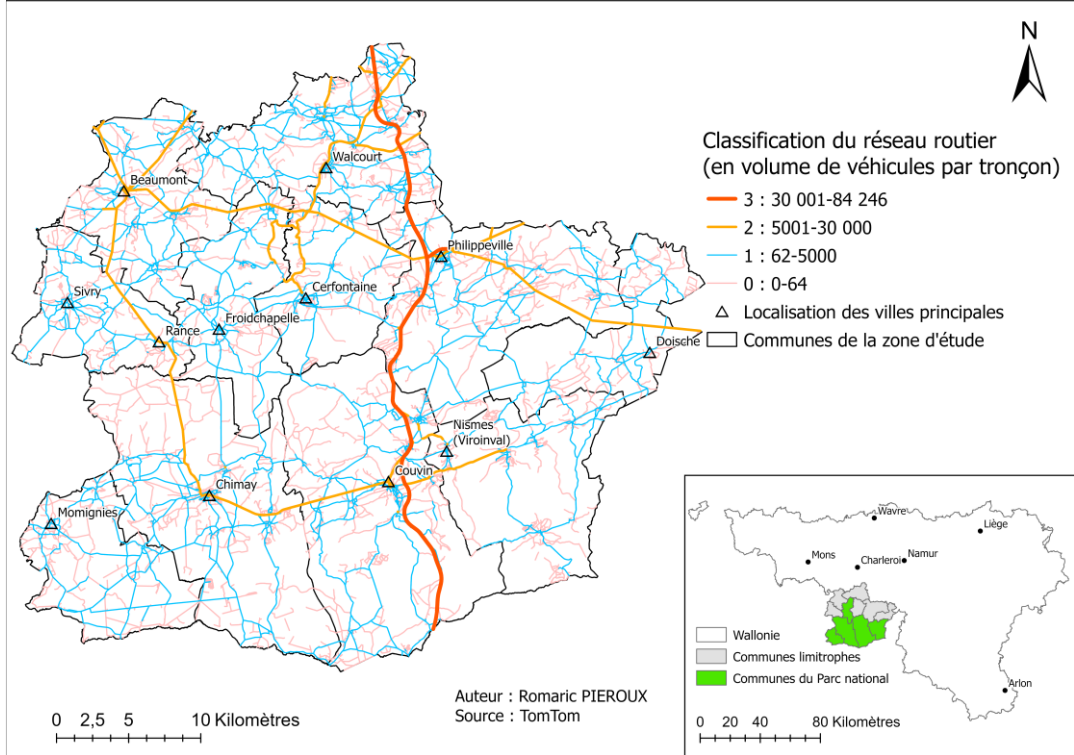


Figure 5 : Classification du réseau routier basée sur le volume de véhicules par tronçon. Source des données : TomTom

Volume de véhicules et vitesse moyenne de nuit

Chaque cas d'accident et point de contrôle est associé au tronçon routier correspondant. Les données TomTom concernant le volume de véhicules total passés de nuit sur le mois d'août 2024 sont utilisées. La vitesse moyenne de nuit de ces véhicules est également une variable utilisée. Pour rappel, dans ce mémoire, la nuit a été définie comme étant entre 21H00 et 7H00 et ce pour prendre en compte le coucher du soleil ainsi que le crépuscule et le lever du soleil du mois d'août 2024. La variable concernant le volume de véhicules de nuit est une variable discrète tandis que la variable concernant la vitesse moyenne de nuit est une variable continue.

Largeur des routes

La largeur d'une route influence le risque d'accident de diverses manières. Les routes plus larges vont notamment donner un sentiment de confiance supérieur amenant à une augmentation de la vitesse et à

accroître les comportements « dangereux » tels que des dépassements plus risqués. Mais les voies plus étroites peuvent également conduire à un risque d'accident plus élevé puisque ces voies vont avoir tendance à augmenter les sorties de routes laissant moins de place à des erreurs de conduite (Wu & Lin, 2022; Azin *et al.*, 2025). La largeur d'un tronçon est estimée en divisant l'aire du tronçon par sa longueur comme le montre l'Équation 1 (Brezina *et al.*, 2017). L'aire des différents tronçons est obtenue à partir de la cartographie du PICC du réseau routier. Il s'agit donc d'une largeur estimée et moyenne par tronçon. Les tronçons ont ici été définis comme une portion de route entre deux intersections. La largeur des routes constitue une variable continue.

Équation 1 : Calcul de la largeur moyenne (en mètre) des tronçons

$$\text{Largeur estimée moyenne} = \frac{\text{Aire du tronçon}}{\text{Longueur du tronçon}}$$

La sinuosité

La sinuosité des routes est calculée dans ce mémoire, avec l'Équation 2, comme le ratio entre la longueur réelle d'un tronçon et la longueur la plus courte (en ligne droite) entre son point de départ (A) et son point final (B) (Figure 6) (Hacar & Gökgöz, 2019; Ranapurwala *et al.*, 2016). Tout comme pour le calcul de la largeur, les tronçons sont ici définis comme une portion de route entre deux intersections. La sinuosité des routes varie entre 1 (une route parfaitement droite) à au maximum 11,1267 correspondant à une route avec un angle au centre de 330° soit une inversion totale du sens de circulation entre le début du tronçon et la fin. L'indice de la sinuosité peut tendre vers l'infini lorsque l'on s'approche d'une boucle fermée mais au-delà de 11,1267 il n'y a plus réellement de sens puisque non adapté à une étude sur un réseau routier étant donné qu'une telle configuration n'existe que rarement. En réalité, une route avec un indice entre 1,25 et 1,5 est déjà considérée comme une route sinueuse et une route avec un indice supérieur à 1,5 est considérée comme une route « à méandres ». Cette typologie pour la sinuosité provient principalement de l'étude de la sinuosité des cours d'eau mais qui a été adaptée pour les réseaux routiers (MUELLER, 1968; Wills, 2019).

Équation 2 : Calcul de la sinuosité moyenne des tronçons

$$\text{Sinuosité} = \frac{\text{Longueur réelle du tronçon } (l)}{\text{Longueur en ligne droite } (d, \text{ entre le point de départ A et final B})}$$

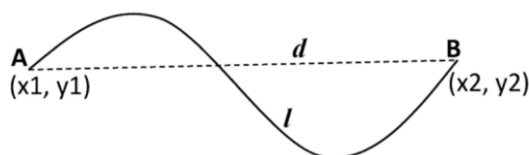


Figure 6 : Illustration du calcul de la sinuosité d'un tronçon (l). Image tirée de l'article Ranapurwala, S. I., Mello, E. R. & Ramirez, M. R., (2016), *A GIS-based Matched Case-control Study of Road Characteristics in Farm Vehicle Crashes*, *Epidemiology*, vol. 27, n°6, p. 827-834.

Proximité d'un éclairage public

Comme dit précédemment, dans le cas des accidents, les policiers se rendant sur place devaient noter une série d'informations sur les lieux de l'accident dont notamment la présence ou non d'un éclairage public et si celui-ci était allumé ou éteint. Cependant, pour les points de contrôle, cette variable doit être construite sur base des données de localisation des lampadaires d'éclairage public. Un seuil a donc été choisi pour déterminer la distance à partir de laquelle un point de contrôle est considéré comme étant à proximité d'un éclairage public. Ce seuil est basé sur la distribution des distances des accidents

catégorisés par les services de police comme ayant lieu à proximité d'un éclairage public au poteau d'éclairage le plus proche. Le but est donc de situer à partir de quelle distance les policiers jugeaient, en moyenne, que l'accident avaient eu lieu à proximité d'un éclairage public. La Figure 7 montre que 93,95% des accidents catégorisés comme à proximité d'un éclairage public se situent à une distance inférieure à 60 mètres et 87,90% se situent à une distance inférieure à 30 mètres. Le seuil choisi est 30 mètres ce qui est en lien avec plusieurs articles scientifiques utilisant ce seuil ou un seuil proche (Bullough *et al.*, 2013; Edwards, 2015; Ingi *et al.*, 2025; Keler *et al.*, 2025; Truter, 2026). Bien que les cas d'accidents possèdent déjà cette variable, ceux-ci ont malgré tout été recatégorisés comme ayant ou non la présence d'éclairage selon leur distance au poteau d'éclairage public le plus proche et ce pour plusieurs raisons. La première raison est d'abord une volonté d'homogénéiser la méthode pour l'ensemble des points (cas et contrôles) mais également car la catégorisation des accidents repose sur la subjectivité des policiers et le choix ne repose donc pas sur une méthode objective (comme, par exemple, moins de 50 mètres d'un poteau d'éclairage ou basé sur une mesure de l'éclairage). De plus des erreurs de catégorisation existent, puisque certains accidents sont catégorisés comme ayant eu lieu à proximité d'éclairage public alors que le premier poteau d'éclairage se situe à des distances relativement importante (plus de 100 mètres voir plus de 200 mètres). La variable est donc une variable binaire valant 1 si à moins de 30 mètres d'un éclairage public et 0 si non.

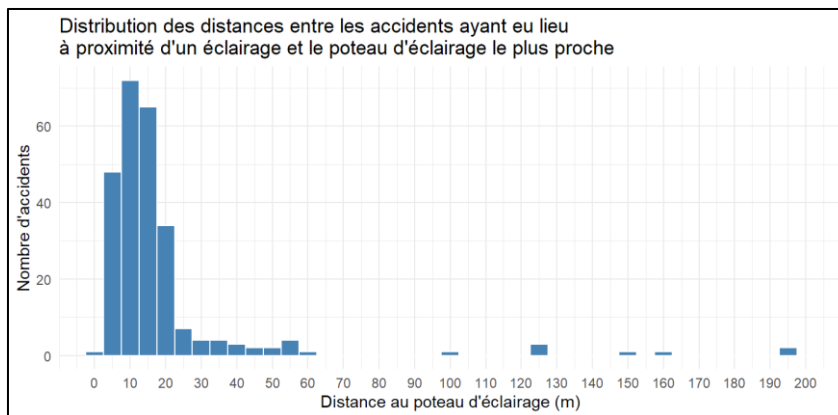


Figure 7 : Distribution des distances entre les accidents catégorisés par la police comme ayant eu lieu à proximité d'un éclairage et le poteau d'éclairage le plus proche

Proximité à une intersection

La proximité à une intersection est une variable construite à partir des données du PICC localisant les intersections. La proximité à une intersection est également une information présente dans la base des données des accidents routiers et la même méthodologie que pour la proximité à un éclairage public peut être utilisée. Sur la Figure 8, il est possible de voir que 86% des accidents catégorisés comme à proximité d'une intersection sont situés à moins de 40 mètres, ce qui correspond au seuil choisi pour définir si un cas ou un contrôle se situe à proximité d'une intersection. Ce seuil a été choisi sur base de la distribution mais également car ce seuil donnait les meilleurs résultats pour les modèles de régression logistique (les seuils de 30 et 50 mètres ont également été testés mais seul le seuil de 40 mètres a été testé puisque donnant les meilleurs résultats). Dans la littérature, il n'y a pas une convergence vers un seuil précis, mais plus vers une plage de valeur correspondant à être proche d'une intersection. La plupart des études tendent vers un buffer entre 30 et 50 mètres autour des intersections. Ce buffer peut même être appelé la phase d'approche ou d'entrée d'une intersection constituant en la décélération du véhicule. La grande majorité des études existantes trouvées prennent une distance de 30 mètres pour identifier les

accidents à proximité d'intersection mais comme expliqué le seuil de 30 mètres donnait de moins bons résultats pour le modèle de régression logistique concernant les variables significatives, ce qui explique que le seuil de 40 mètres ait été gardé. (Basu *et al.*, 2022; Kang *et al.*, 2019; Keler *et al.*, 2025; Ullah *et al.*, 2025)

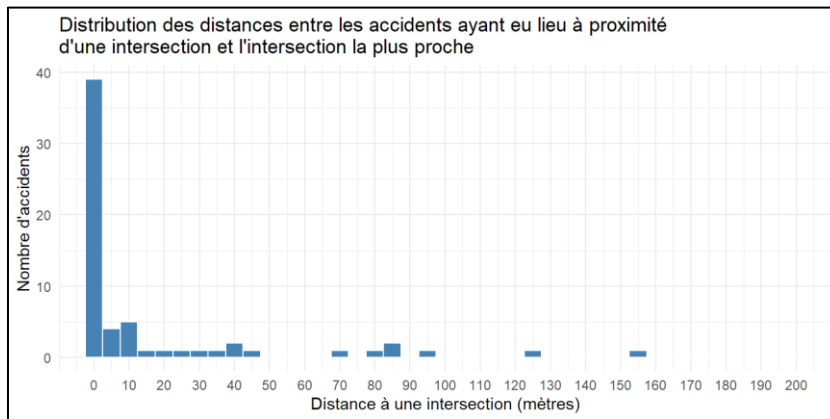


Figure 8 : Distribution des distances entre les accidents catégorisés par la police comme ayant eu lieu à proximité d'une intersection et l'intersection la plus proche

Les accidents et points de contrôle ont donc été catégorisés comme ayant lieu à proximité d'une intersection si l'intersection la plus proche se situe à moins de 40 mètres. C'est donc une variable binaire prenant la valeur de 1 si à moins de 40 mètres d'une intersection et 0 si non. Il s'agit donc bien ici de l'intersection accessible la plus proche depuis l'accident ou le point de contrôle. Cette précision est importante puisque les routes de classe 3 (gabarit autoroutier comme la N5) sont des routes possédant relativement moins d'intersections (et dont celles-ci sont plus espacées) et peuvent avoir des routes secondaires parallèles à leur tracé ce qui pourrait provoquer qu'un point situé sur une route de classe 3 puisse avoir comme intersection la plus proche une intersection qui ne lui est pas directement accessible. Un exemple d'une telle situation est visible en annexe (Figure A. 2) où l'accident (rond noir) pourrait capter l'intersection sur le pont alors que celle-ci n'est pas directement accessible puisque l'intersection la plus proche et qui est accessible est située plus loin avec la voie de rabattement de la bretelle d'entrée de l'autoroute.

Zone bâtie

Les zones bâties sont ici définies comme des zones où la majeure partie est couverte de constructions dont principalement des habitations telles que des villages ou des villes. Ces zones peuvent impacter le risque d'accident par une densité de bâtiments importantes ce qui réduit la visibilité aux intersections et diminue également le champ de vision. Être à l'intérieur d'un village ou d'une ville donne une perception différente que d'être à l'extérieur d'une zone bâtie et il est donc intéressant d'étudier l'impact de ces zones bâties sur le risque d'accident. Les zones bâties sont également un bon proxy concernant la densité d'intersections qui joue un rôle dans le risque d'accident en augmentant les phases d'accélération et de décélération ainsi que les phases où une grande attention et une grande vigilance sont demandées (Basu *et al.*, 2022).

Les services de police doivent noter si un accident a eu lieu au sein d'une zone bâtie ou non. Cependant, cette différenciation ne repose que sur le jugement de la personne au moment de la rédaction du rapport de l'accident (tout comme pour la proximité à un éclairage ou la proximité à une intersection). Dans les rapports des accidents, cette variable est exprimée comme le fait d'être ou non en agglomération sans rapport à la définition d'une agglomération au sens du code de la route. C'est dans ce but que les secteurs statistiques, provenant de Statbel, sont utilisés pour différencier les milieux bâtis des milieux non-bâtis.

En effet, les secteurs statistiques englobent les zones bâties comme les villages et les villes, créant ainsi des secteurs statistiques de taille relativement faible mais avec un nombre d'habitants relativement élevé. Au contraire, les zones non-bâties, sont de taille relativement élevée avec peu de populations. De ce fait, les secteurs statistiques des zones bâties ont des densités de population relativement élevées au contraire des secteurs statistiques des zones non bâties. Les secteurs statistiques sont utilisés et préférés dans ce mémoire par rapport au Land Cover CORINE 2018 ou à la couche de l'occupation du sol du SPW. En effet, pour le Land Cover CORINE 2018, des erreurs de catégorisation conduisent à avoir des zones bâties non catégorisées comme surfaces artificialisées. Et inversement, certaines zones bâties sont catégorisées comme terres agricoles, ce qui est notamment dû à la méthodologie du Land Cover liée à une classification automatique sur des images satellites. Là où les secteurs statistiques sont construits pour justement recouper assez précisément les zones habitées et donc bâties. Pour ce qui est de l'occupation du sol du SPW, la catégorisation est précise regroupant ainsi le réseau routier soit comme une classe à part entière soit dans la même classe que les surfaces artificialisées ce qui rend difficile la détermination d'une zone bâtie ou non bâtie puisque qu'un accident localisé sur une route traversant une surface agricole sera tout de même catégorisé comme ayant eu lieu dans une zone bâtie.

La Figure 9 montre le fait que les secteurs statistiques des zones bâties (densité supérieure à 150 hab/km²) comme les villages et villes sont de tailles relativement plus faibles. Il est également possible de voir sur cette figure, la catégorisation des accidents, effectuée par la police, localisés en agglomération ou non (sous-entendu en zone bâtie ou non). Même à cette échelle, il est déjà possible de voir que certains accidents ont été mal catégorisés vu que des accidents hors zone bâtie ont été catégorisés comme en agglomération et inversement.

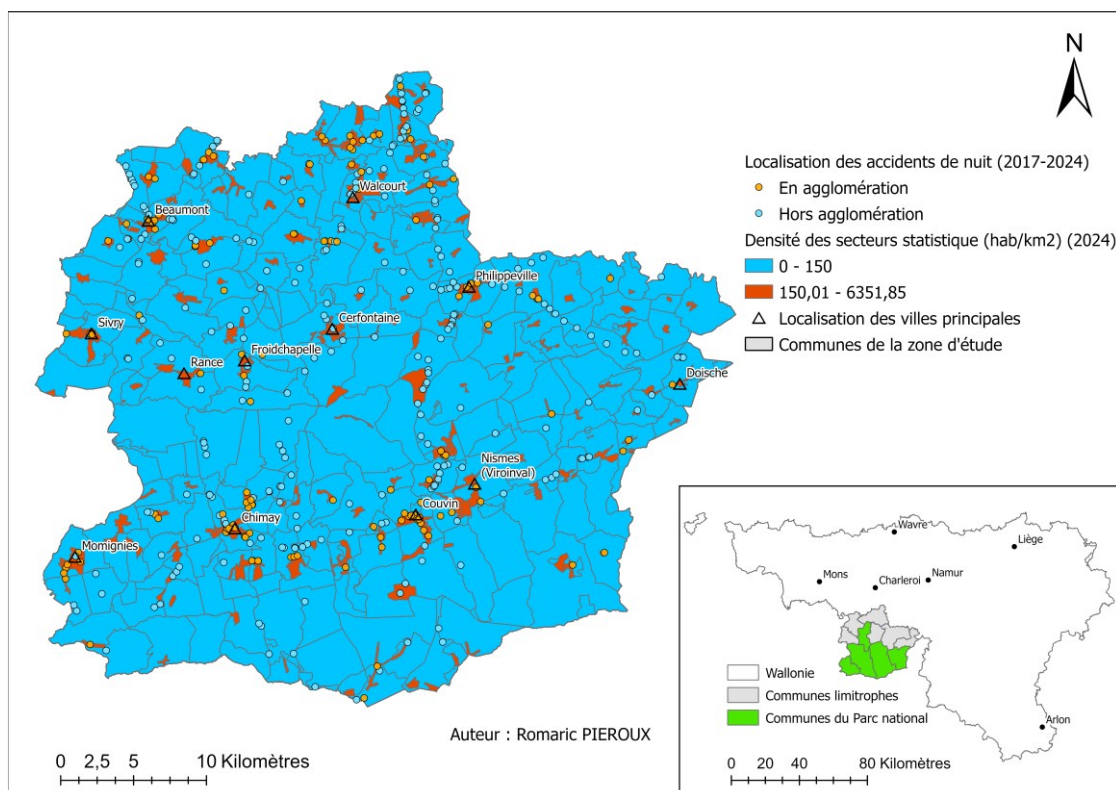


Figure 9 : Localisation des accidents de nuit et des secteurs statistiques de Statbel montrés avec leur densité de population (hab/km²). Fonds de carte : orthophotos 2024 (SPW).

Avec les densités par secteurs statistiques et la catégorisation des accidents opérée lors des rapports de police, il est possible de définir un seuil pour établir si les accidents et les points de contrôles sont

localisés dans une zone bâtie ou non en reprenant la même méthodologie que pour les variables sur la proximité à un éclairage public et de proximité à une intersection. La médiane des densités des secteurs statistiques avec des accidents ayant été catégorisés comme ayant eu lieu dans une zone bâtie est de 1106 hab/km² tandis qu'elle est de 25,5 hab/km² pour les secteurs statistiques avec des accidents ayant été catégorisés comme ayant eu lieu hors zone bâtie. La Figure 10 montre cette différence avec une distribution des densités des secteurs statistiques avec des accidents catégorisés comme ayant eu lieu dans une zone bâtie (en agglomération) plus élevée que pour les secteurs statistiques avec des accidents ayant été catégorisés comme ayant eu lieu hors zone bâtie (hors agglomération) où la distribution est entièrement en dessous des 150 hab/km² bien que des outlier existent avec des densités de populations supérieures à 150 hab/km². La Figure 11 montre également cette différence entre les deux catégories mais permet également de se rendre compte du nombre d'accident ayant été possiblement mal catégorisé. Cette figure montre également la concentration du nombre d'accident catégorisé en hors zone bâtie sous les 200 hab/km². Le seuil choisi dans ce mémoire pour déterminer les secteurs statistiques couvrant des zones bâtie est alors de 150 habitants par kilomètre carré ce qui permet également de reprendre le même seuil définissant un espace rural dans la classification (DGO3) des communes selon leur degré de ruralité du Service Public de Wallonie (SPW, 2013).

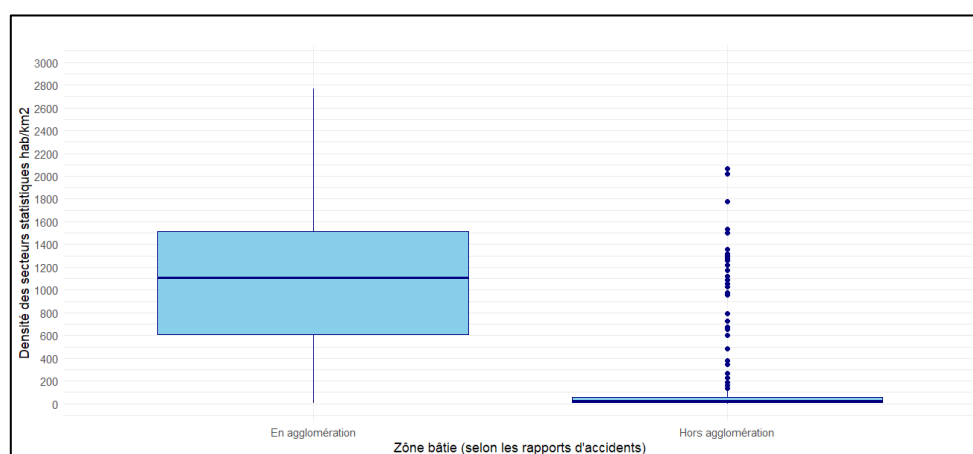


Figure 10 : Boxplot de la densité des secteurs statistiques où les accidents ont été catégorisés comme ayant eu lieu en agglomération (en zone bâtie) ou en hors agglomération (en zone non-bâtie) selon la police

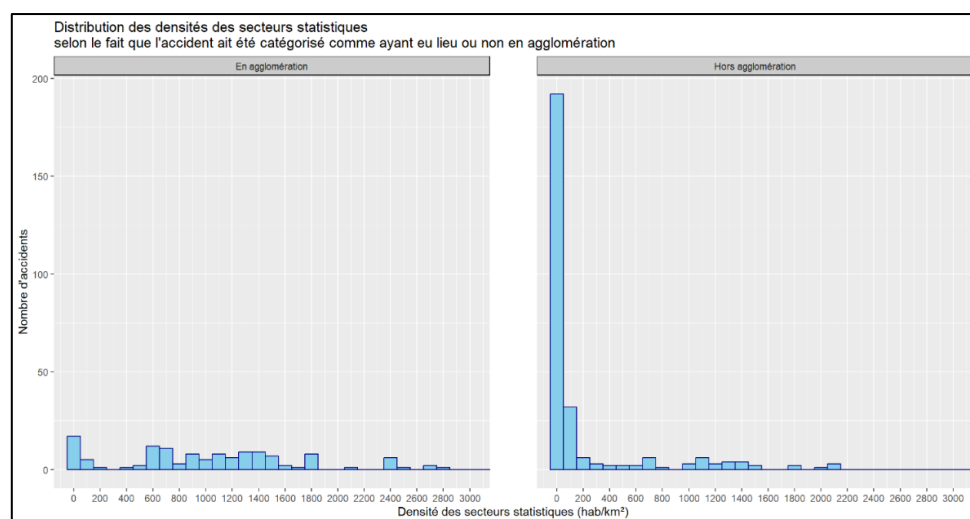


Figure 11 : Distribution des densités des secteurs statistiques selon le fait que l'accident ait été catégorisé comme ayant eu lieu ou non en agglomération

Les accidents et les points de contrôle sont donc catégorisés en zone bâtie si leur localisation se situe dans un secteur statistique avec une densité de population supérieure à 150 hab/km². Les accidents ont donc été recatégorisés afin d’avoir une méthode homogène pour les accidents et les points de contrôle. De plus, comme expliqué ci-dessus, la catégorisation opérée par les services de police ne repose que sur le jugement de la personne au moment de la rédaction du rapport d’accident. Un exemple d’une telle situation est disponible en annexe (Figure A. 3) montrant plusieurs accidents localisés dans un même secteur statistique d’une zone bâtie dont deux accidents qui ont été catégorisés en « hors agglomération ». D’autres erreurs à d’autres localisations sont également présentes.

Résumé des variables utilisées

Un résumé des variables étudiées est disponible dans la Table 3.

Table 3 : Résumé des variables utilisées dans ce mémoire

Variable	Description	Type	Source
Accident/point de contrôle	Variable à expliquer Décris si c’est un accident ou un point de contrôle	Binaire (1 si accident, 0 si point de contrôle)	Statbel pour les accidents. Les points de contrôle ont été créés dans ce mémoire avec une localisation aléatoire mais stratifiée par rapport au réseau routier.
Classification du réseau routier	Divise le réseau routier en 3 classes de routes (selon le volume de véhicules)	Variable catégorique mais est adaptée pour être une variable binaire en prenant une classe de référence (la classe 1) et en exprimant les deux autres par rapport à cette classe de référence	Selon le volume de véhicules par tronçon provenant des données TomTom
Volume de véhicules de nuit	Nombre de véhicules total sur le tronçon associé étant passés durant les nuits du mois d’août 2024	Variable discrète	Selon le volume de véhicules de nuit par tronçon provenant des données TomTom
Vitesse moyenne de nuit	Vitesse moyenne des véhicules sur le tronçon associé étant passés au total durant les nuits du mois d’août 2024	Variable continue	Selon le volume de véhicules de nuit et les vitesses associées par tronçon provenant des données TomTom
Largeur des routes	Largeur des routes estimées selon la longueur et l’aire des tronçons	Variable continue	Selon la cartographie du réseau routier par le PICC (SPW)
Sinuosité des routes	Sinuosité estimée selon la longueur réelle et la longueur euclidienne des tronçons	Variable continue mais comprise entre 1 et 11, 1267	Selon la cartographie du réseau routier par le PICC (SPW)
Proximité à un éclairage public	Si l’accident/point de contrôle est situé à moins de 30 mètres d’un éclairage public	Binaire (1 si à moins de 30 mètres, 0 si à plus de 30 mètres)	Selon la localisation à un lampadaire d’éclairage public régional ou communal.
Proximité à une intersection	Si l’accident/point de contrôle est situé à moins de 40 mètres d’une intersection	Binaire (1 si à moins de 40 mètres, 0 si à plus de 40 mètres)	Selon la localisation des intersections provenant du PICC (SPW)

Localisation dans une zone bâtie	Si l'accident est situé dans un secteur statistique avec une densité de population de plus de 150 hab/km ²	Binaire (1 si zone bâtie, 0 si hors zone bâtie)	Selon les densités de population des secteurs statistiques de Statbel (2024)
---	---	---	--

3.3.3. Régression logistique (RL)

Une fois les variables créées et associées à chaque cas et points de contrôle, il est possible de construire un premier modèle de régression logistique. Le modèle de régression logistique a donc deux buts : permettre d'identifier les variables qui ont un impact significatif sur le risque d'accident et, avec ces variables significatives, de modéliser le risque d'accident sur le réseau routier. Un modèle de régression logistique est un modèle de régression binomiale utilisé quand la variable à expliquer est binaire (0 ou 1). Le modèle de régression logistique est ici utilisé avec la fonction LOGIT et peut s'écrire comme l'Équation 3 :

Équation 3 : Équation générale d'un modèle de régression logistique utilisé avec la fonction LOGIT représentant le membre de gauche

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Dans cette formule, p représente la probabilité que l'évènement se produise, x représente une des variables étudiées et reprend donc la valeur pour un échantillon et β (paramètres estimés) représente l'effet de cette variable sur la probabilité. Les Odds, ou chances, sont $\left(\frac{p}{1-p}\right)$, soit le ratio entre la probabilité d'accident et la probabilité de non-accident (par exemple des odds de 2 signifient qu'il y a deux fois plus de probabilité d'accident que de non-accident, donc, 66% contre 33%). Les Odds-ratio sont le ratio des chances suivant la valeur de la variable, c'est à dire par exemple, pour une variable binaire prenant la valeur 0 ou 1, ce sont les $\text{Chances}_1 / \text{Chances}_0$, c'est à dire $\left(\left(\frac{p}{1-p}\right)_1 / \left(\frac{p}{1-p}\right)_0\right)$. Par exemple un Odd ratio de 1,5 en cas de présence d'une certaine variable signifie que les chances d'accident augmentent de 50% lorsque cette variable est présente. Les Odds ratio se calculent en prenant l'exponentiel de chaque paramètre estimé (β). Le membre de gauche représente le LOGIT. (Bouyer, 2024; El Sanharawi & Naudet, 2013b; Rakotomalala, 2015). La performance des modèles peut s'évaluer en regardant l'AIC (Akaike Information Criterion), permettant de comparer les modèles, et l'AUC (Area Under the Curve), mesurant la performance du modèle à détecter les cas et à éviter les faux positifs. Au plus l'AIC est faible, au plus le modèle peut être considéré comme meilleur et au plus l'AUC se rapproche de 1, au plus le modèle peut être évalué comme performant (Jr *et al.*, 2013).

Le premier modèle de régression logistique créé prend comme variable à expliquer la variable risque d'accident avec les accidents $Y=1$ et les points de contrôle $Y=0$ et est le modèle général de ce mémoire en reprenant l'ensemble des cas et points de contrôle ainsi que l'ensemble des variables étudiées. Les variables significatives seront identifiées par le modèle grâce à l'analyse des P-valeurs et le seuil de significativité ou de rejet de l'hypothèse nulle choisi est 0,05, soit le seuil classique. Les paramètres estimés du modèle de régression logistique permettent d'une part d'identifier si les variables significatives influencent positivement ou négativement le risque d'accident ainsi que l'importance de ces variables dans la régression et donc dans le risque d'accident. Si une variable impacte positivement le risque d'accident, cela signifie donc une augmentation de ce risque, à l'inverse si une variable impacte négativement le risque d'accident, cela signifie donc une diminution du risque d'accident.

D'une autre part, les paramètres estimés des variables significatives sont utilisées pour modéliser le risque d'accident sur le réseau routier. Étant donné que le modèle de régression logistique est construit

à partir d'une approche cas-contrôle, soit des points et non des tronçons agrégés, le risque d'accident est donc modélisé en générant des points tous les 50 mètres pour chaque tronçon routier. Les variables significatives sont alors calculées pour chacun de ces points et multipliées par les paramètres estimés du modèle de régression logistique. Ce résultat correspond ainsi au LOGIT et la probabilité p d'avoir un accident se calcule alors avec la formule suivante $p = \frac{1}{1+e^{LOGIT}}$ (Bouyer, 2024; El Sanharawi & Naudet, 2013b; Rakotomalala, 2015). La probabilité d'avoir un accident (ou le risque d'accident) est donc comprise entre 0 et 1. Cette carte du risque d'accident sert également à sélectionner les poteaux d'éclairage potentiellement inutiles, par rapport à la sécurité routière, selon plusieurs seuils de risque d'accident. Les poteaux d'éclairage étant reliés au tronçon routier le plus proche et donc au risque prédit le plus proche.

Un second modèle de régression logistique est également créé pour tester l'effet de l'intersection entre les variables « présence en zone bâtie » et « proximité à une intersection » dans le but de répondre à l'une des hypothèses de départ qui était que l'effet de l'éclairage sur le risque d'accident est concentré aux intersections et non sur l'ensemble du réseau routier. Une variable d'interaction est donc créée en multipliant les variables « proximité à un éclairage public » et « proximité à une intersection ». Ce modèle reprend également les autres variables tel qu'utilisés dans le premier modèle de régression logistique.

Ensuite, un troisième modèle de régression logistique est créé pour chaque classe de route dans le but de voir si des différences existent entre les classes de route concernant les variables significatives. Donc, pour les 3 classes de route, un modèle de régression logistique est créé en ne prenant que les accidents et les points de contrôle situés dans la classe de route correspondante. La classification du réseau routier n'est dès lors pas étudiée dans ces modèles spécifiques. Uniquement pour la classe de route 3 (gabarit autoroutier), un modèle de régression logistique supplémentaire a été testé en retirant la variable « zone bâtie » puisque les routes constituant cette classe ne subissent pas réellement l'impact des zones bâties. En effet, ces routes traversent à certains moments des zones bâties mais en sont isolées puisque non connectées avec des intersections au reste du réseau routier. Le reste du réseau routier est dévié ou aménagé en tunnels et ponts pour éviter les intersections avec les grandes routes ou autoroute et ainsi permettre une meilleure fluidité sur le réseau au gabarit autoroutier.

Enfin un quatrième modèle de régression logistique est créé en prenant cette fois-ci la variable « proximité à un éclairage public » comme variable dépendante pour tester l'hypothèse « les lampadaires d'éclairage public sont disposés à des localisations accidentogènes », avec donc comme variable explicative l'ensemble des variables du premier modèle de régression logistique et la présence ou absence d'accident en plus.

L'ensemble des modèles de régression logistique sont testés pour vérifier l'absence de multi colinéarité au moyen d'un test VIF (Variance Inflation Factor).

3.3.4. Modèle Random Forest (RF)

En plus d'un modèle de régression logistique, un modèle Random Forest est donc utilisé. Le modèle Random Forest est ici utilisé en reprenant les mêmes variables étudiées dans le premier modèle de régression logistique. Le but d'utiliser un modèle Random Forest dans ce mémoire est d'identifier les variables les plus importantes par rapport à l'augmentation du risque d'accident au moyen des importances relatives des variables. Un modèle Random Forest est un modèle de machine learning basé sur les arbres de décision. Plusieurs arbres de décision sont créés où pour chaque arbre de décision, un sous-ensemble du jeu de donnée va lui être attribué. Ce sous-ensemble est constitué d'échantillon (cas et points de contrôle) choisi aléatoirement parmi le jeu de données. Il s'agit d'un tirage avec remise, les

échantillons pouvant être choisis plusieurs fois pour être étudié. Cette étape s'appelle le rééchantillonnage ou « Bootstrap ». Chaque arbre est donc entraîné sur un échantillon aléatoire du jeu de données et chaque division d'un arbre va permettre d'étudier une variable. L'arbre va étudier l'ensemble des variables mais la variable étudiée à chaque division de l'arbre est choisie aléatoirement sans remise. Ensuite, chaque arbre va répartir le sous-ensemble d'échantillons entre les accidents et les points de contrôle en séparant les valeurs des variables. Prenons un exemple : un arbre reçoit 100 accidents et 100 points de contrôle avec comme variable la largeur de la route dont la moyenne pour les points de contrôle est de 1 mètre contre 5 mètres pour les cas, alors l'arbre va diviser d'un côté les échantillons avec une largeur de route de moins de 2 mètres et de l'autre les échantillons avec une largeur de route supérieure à 2 mètres et répéter ce mécanisme pour l'ensemble du reste des variables. Pour un autre arbre cette valeur sera de 3 mètres par exemple puisqu'il aura reçu d'autres échantillons à étudier. Cette méthode va être répétée pour chaque arbre (par défaut 500) qui vont ainsi étudier approximativement 70% du jeu de données de base, le reste (sous ensemble de validation) servant à valider le modèle. Chaque arbre étant entraîné et ayant ses propres paramètres estimés pour chaque variable est ensuite capable de faire des prédictions. Un échantillon du sous-ensemble de validation va être analysé par l'ensemble des 500 arbres qui vont donner une valeur prédite de la probabilité qu'il s'agisse ou non d'un accident. Une moyenne est réalisée sur les réponses des 500 arbres, il s'agit ici du Bagging ou Bootstrap Aggregating. En analysant l'ensemble du sous-ensemble de validation, la fiabilité du modèle à prédire des accidents ou des non-accidents va pouvoir être évaluée. Avec ce type de modèle il est donc possible de savoir l'importance relative des variables selon le fait qu'elles ont un poids ou non dans la « décision » d'un arbre. Ce type de modèle ne donne pas de p-valeur ni de paramètres estimés comme un modèle de régression logistique puisque ce modèle ne suppose pas la même forme mathématique mais les p-valeurs peuvent être approchées en mesurant la contribution locale des variables et les paramètres estimés en prenant l'effet moyen des variables notamment via les « Partial Dependence plot ». La performance des modèles est identifiée dans ce mémoire au moyen du pourcentage de variance expliquée et de la valeur du RMSE (Root Mean Square Error) (Biau & Scornet, 2016; Geurts *et al.*, 2006; Schlögl, 2020a).

Plusieurs modèles de Random Forest sont créés, et chacun utilise les mêmes paramètres initiaux dont le nombre d'arbres par modèle qui est de 500 dans ce mémoire et le nombre de divisions (split) par arbre, soit 2.

Le premier modèle Random Forest créé est semblable au premier modèle de régression logistique en étant un modèle général avec l'ensemble des cas et points de contrôle et des variables étudiées. Ce premier modèle est itéré 30 fois pour s'assurer de la constance des importances relatives de chaque variable et donc de la constance du modèle Random Forest. Ce premier modèle sert également à modéliser le risque d'accident à partir de point créé tous les 50 mètres. Avec le risque prédit, plusieurs seuils sont choisis pour identifier les poteaux d'éclairage public potentiellement inutiles pour la sécurité routière.

Ensuite, tout comme pour la régression logistique, un modèle Random Forest est créé par classe de route en retirant donc la variable classification du réseau routier puisque chaque modèle n'utilise que les cas d'accidents et les points de contrôle de la classe de route associée.

Une comparaison est réalisée entre le modèle de régression logistique général et le modèle Random Forest au moyen de t-test pour chaque classe de route. Le but est de voir s'il y a une différence significative dans le risque prédit entre les deux modèles.

4. Résultat

4.1. Analyse exploratoire

4.1.1. Caractéristiques des accidents

Répartition temporelle des accidents

La répartition temporelle des accidents répertoriés, entre 2017 et 2024, pour l'ensemble de la zone d'étude est montrée sur la Figure 12. Les accidents sont donc agrégés par heure et par mois. Il est possible de voir que les accidents se concentrent principalement entre 15H et 17H avec un pic à 22 accidents pour le mois de septembre à 16H. Comme montré dans la méthode, 1451 accidents correctement géolocalisés ont été répertoriés dont 420 accidents de nuit et donc 1031 accidents de jour.

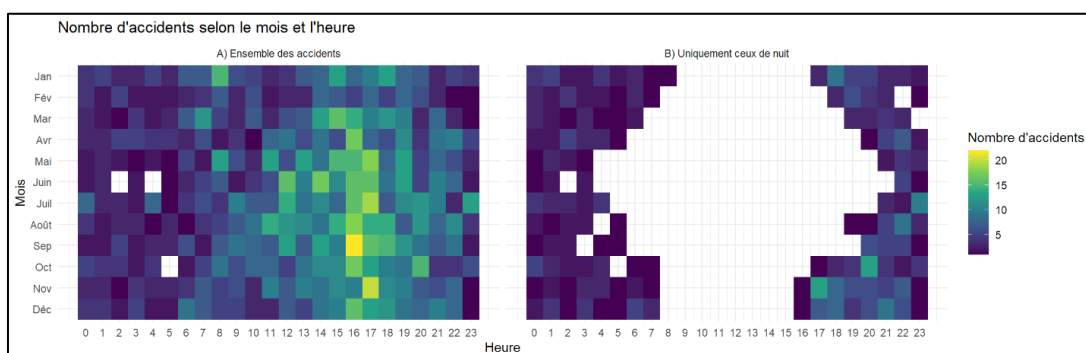


Figure 12 : Heatmap de la répartition temporelle des accidents de la zone d'étude entre 2017 et 2024. Le graphique de gauche (A) reprend l'ensemble des accidents tandis que le graphique de droite (B) ne reprend que les accidents catégorisés comme ayant eu lieu de nuit.

4.2. Impact de l'éclairage public sur l'environnement

La visualisation des localisations des lampadaires d'éclairage public permet de se rendre compte de la densité de points lumineux dans ce territoire ayant pourtant la présence de nombreuses zones à haute valeur de biodiversité (Figure 13). La densité de points lumineux pour l'ensemble de la zone d'étude est de 23647 lampadaires pour 1314,94 km² soit 17,98 points lumineux/km² ou 17,33 points lumineux/km² en ne prenant que les communes du Parc national (ayant alors 12119 lampadaires). La N5 est entièrement éclairée, à l'exception de la portion sud (entre la ville de Couvin et la frontière Française), ce qui fragmente les habitats et rend difficile les passages d'espèces d'est en ouest. Cela est encore plus visible au regard des zones Natura2000 et du périmètre du Parc national (Figure 13, Figure A. 4). De plus, la proximité d'éclairage public avec ces zones à haute valeur de biodiversité est également un risque pour la faune et la flore. La couverture du sol montre également l'impact de l'éclairage public sur des zones naturelles puisque de l'éclairage public est présent même en forêt et en prairie (Figure A. 5). Enfin, les réserves naturelles agréées, les réserves forestières, les zones humides à intérêt biologique et les réserves naturelles domaniales subissent des effets potentiellement néfastes de l'éclairage public par la proximité directe de poteaux d'éclairage public tel que visible sur la Figure A. 6 en annexe. L'analyse des impacts de l'éclairage public sur l'environnement n'est, comme expliqué dans la méthode, pas détaillé plus en profondeur puisqu'il ne s'agit pas du cœur du mémoire.

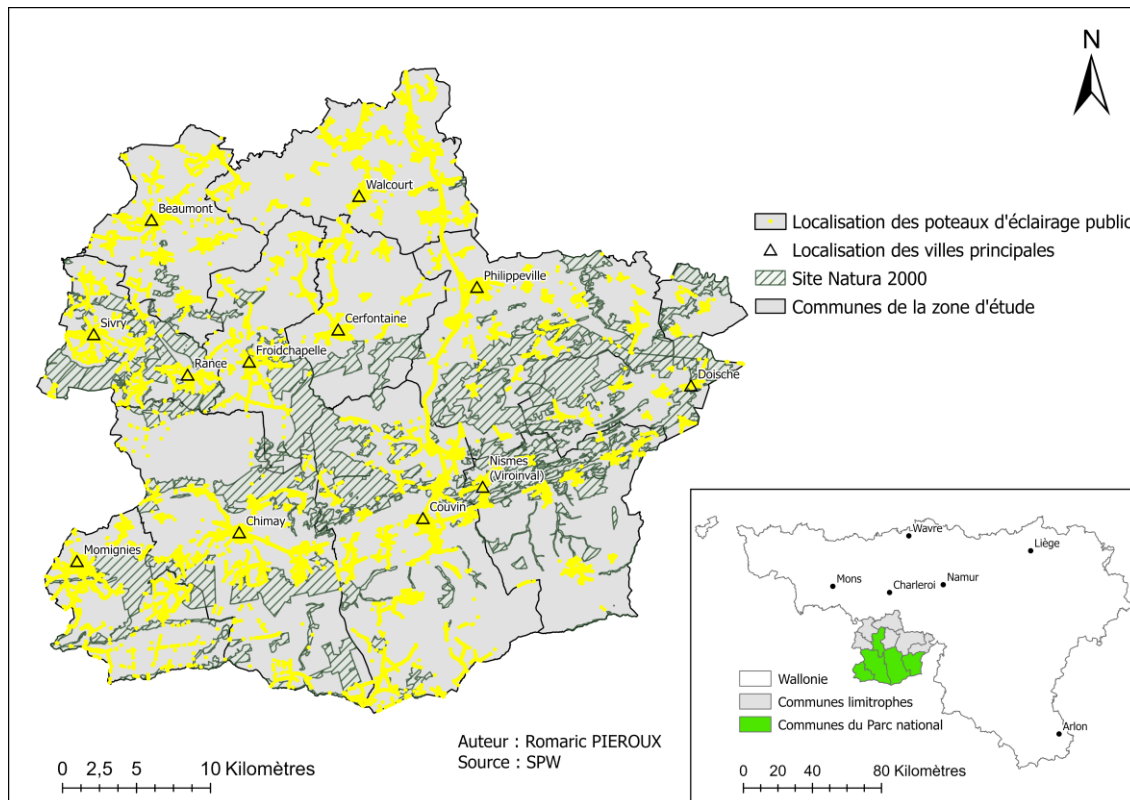


Figure 13 : Localisation des zones Natura2000 et des poteaux d'éclairage public

4.3. Effet des variables sur le risque d'accident

4.3.1. Analyse des modèles de régression logistique

Modèle de régression logistique général

Le premier modèle de régression logistique, qui est le modèle général de ce mémoire, reprend donc l'ensemble des variables étudiées et l'ensemble des cas et points de contrôle (soit 1200 cas et points de contrôle). Ce modèle montre que les variables ayant un impact significatif et positif sur le risque d'accident sont la présence en zone bâtie, la proximité à une intersection, la proximité à un éclairage public, la largeur de la route et la vitesse moyenne de nuit (Table 4). Ces variables augmentent donc le risque d'accident puisque leurs paramètres estimés sont positifs et ces variables sont significatives puisque leurs p-valeurs sont inférieures au seuil de confiance (0,05). Il est possible de dire, par l'analyse des paramètres estimés, et le calcul des Odds ratio des variables significatives, que le fait d'être situé en zone bâtie augmente de 73% les chances d'accident. La proximité à une intersection augmente les chances d'accident de 59% et la proximité à un éclairage public augmente les chances de 56%. Chaque km/h en plus de vitesse moyenne augmente les chances d'accident d'1,6% alors qu'augmenter la largeur d'une route de 1 mètre augmente les chances d'accident de 2,3%. La variable ayant un impact significatif mais négatif sur le risque d'accident est la classe de route signifiant qu'être sur une route de classe 3 ou 2 diminue le risque d'accident par rapport à être sur une route de classe 1. Circuler sur une route de classe 3 diminue les chances d'accident de 97% par rapport à une route de classe 1 (routes locales) et 66% s'il s'agit d'une route de classe 2. La sinuosité et le volume de véhicules de nuit sont des variables non significatives, n'influençant pas le risque d'accident selon ce modèle de régression logistique et ne sont donc pas utilisées pour construire la carte du risque d'accident prédit. L'AIC de ce modèle vaut 1254,7 et l'AUC vaut 0,8.

Ce modèle n'est pas impacté par un problème de multi colinéarité (l'ensemble des variables ayant un VIF inférieur à 3). Cela est également le cas pour l'ensemble des modèles de régression logistique suivants. Les valeurs de VIF pour chaque variable et pour chaque modèle de régression logistique sont disponibles dans la Table A. 1 (en annexe).

Table 4 : Résultats du modèle de régression logistique général réalisé sur les 1200 cas et points de contrôle. L'AIC de ce modèle est 1254,7. (« * » : <0,05, « ** » : <0,01, « *** » : 0)

Variable	Paramètres estimés	Odd- ratio	P-valeur
(Intercept)	-0,91	0,40	0,325
Zone bâtie	0,55	1,73	0,003**
Proximité à une intersection (0_1) (40 mètres)	0,46	1,59	0,021*
Proximité à un éclairage public (0_1) (30 mètres)	0,45	1,56	0,012*
Classe de route : 2	-1,09	0,34	6,57E-08***
Classe de route : 3	-3,53	0,029	2,24E-08***
Sinuosité	-0,50	0,61	0,515
Largeur	0,02	1,023	0,003**
Volume de véhicules de nuit	2,93E-05	1	0,599
Vitesse moyenne de nuit	0,02	1,016	0,004**

Modèle de régression logistique testant l'interaction entre la proximité à une intersection et la proximité à un éclairage public

L'interaction entre les variables proximité à une intersection et proximité à un éclairage public est testée en multipliant ces variables entre elles. L'interaction n'est pas significative, signifiant donc que l'effet de l'éclairage public sur les chances d'accident ne se concentre pas aux intersections mais est présent sur l'ensemble du réseau routier (Table A. 2). La proximité à une intersection et la proximité à un éclairage public augmentent toujours le risque d'accident, mais de façon légèrement différente que dans le premier modèle puisque la P-valeur de la proximité à un éclairage public est plus faible (P-valeur : 0,0044**) et donc plus significative. Les Odd-ratio de ces deux variables sont également légèrement plus élevés (2,37 pour la proximité à une intersection et 1,73 pour l'éclairage public). Les effets et significativités des autres variables n'ont pas sensiblement changé. L'AIC pour ce modèle vaut 1254,8 et l'AUC vaut 0,79.

Modèles de régression logistique pour la classe de route 3

Deux modèles de régression logistique ont donc été créés pour la classe de route 3 (soit 469 cas et points de contrôle), l'un avec la variable zone bâtie et l'autre sans cette variable puisque la classe de route 3 ne subit pas, à priori, les effets des zones bâties (Table 5). Les deux modèles ne sont pas significativement différents, ayant les mêmes valeurs pour les paramètres estimés et de P-valeur pour les variables, sauf pour la variable volume de véhicules de nuit puisque celle-ci n'est pas significative (P-valeur >0,05) dans le modèle sans la zone bâtie, mais l'est dans le modèle avec la zone bâtie (P-valeur : 0,021*). Concernant les autres variables, seule la sinuosité est une variable significative diminuant le risque d'accident. La proximité à un éclairage public n'influence donc pas le risque d'accident tout comme la proximité à une intersection, la présence en zone bâtie, la largeur de la route et la vitesse moyenne de

nuits. L'AIC pour le modèle avec zone bâtie vaut 282,83 (0,79 pour l'AUC) et vaut 284,02 pour le modèle sans la variable zone bâtie (0,79 pour l'AUC).

Table 5 : Résultats des deux modèles de régression logistique pour la classe de route 3 (soit 469 cas et points de contrôle). Le modèle avec la variable zone bâtie est présenté à gauche et le modèle sans la variable zone bâtie est présenté à droite. (« * » : <0,05, « ** » : <0,01, « *** » : 0)

Variable	Avec zone bâtie			Sans zone bâtie		
	Paramètres estimés	Odd-ratio	P-valeur	Paramètres estimés	Odd-ratio	P-valeur
(Intercept)	20,18	5,79 E08	0,98	20,31	6,60 E08	0,98
Zone bâtie	0,69	2	0,064	/	/	/
Proximité à une intersection (0_1) (40 mètres)	0,60	1,83	0,14	0,54	1,72	0,18
Proximité à un éclairage public (0_1) (30 mètres)	15,08	3,53 E06	0,98	15,11	3,64 E06	0,98
Sinuosité	-35,13	5,53 E-16	0,014*	-35,04	6,044 E-16	0,014*
Largeur	0,019	1,019	0,29	0,021	1,021	0,23
Volume de véhicules de nuit	-0,0001	0,99	0,021*	-0,00013	1	0,056
Vitesse moyenne de nuit	-0,0079	0,92	0,45	-0,012	0,88	0,22

Modèle de régression logistique pour la classe de route 2

Lorsque le modèle est effectué uniquement pour la classe de route 2 (soit 386 cas et points de contrôle) seule la variable volume de véhicules de nuit est significative augmentant alors le risque d'accident. Les autres variables ne sont donc pas significatives (Table 6). L'AIC pour ce modèle vaut 507,72 et l'AUC vaut 0,73.

Modèle de régression logistique pour la classe de route 1

Pour le modèle de régression de la classe de route 1 (soit 345 cas et points de contrôle), il est possible d'identifier que la présence en zone bâtie, la proximité à un éclairage public et la vitesse moyenne de nuit sont les trois seules variables significatives augmentant le risque d'accident, les autres variables n'étant pas significatives (Table 6). L'AIC pour ce modèle vaut et l'AUC vaut 0,73.

Modèle de régression logistique pour les classes de route 1 et 2 ensemble

Au vu des points communs entre les classes de route 1 et 2 et de leurs différences avec la classe de route 3, un modèle de régression logistique est créé en prenant ensemble les classes de routes 1 et 2 avec les cas et points de contrôles situés sur ces deux classes de routes (soit 731 cas et points de contrôle) (Table 6). Le modèle indique alors que les variables significatives augmentant le risque d'accident sont la présence en zone bâtie, la proximité à un éclairage public et la largeur de la route. Le volume de véhicules de nuit est une variable significative diminuant le risque d'accident. Les autres variables ne sont donc pas significatives et n'ont donc pas d'influence sur le risque d'accident selon ce modèle. L'AIC pour ce modèle vaut 978,36 et l'AUC vaut 0,66.

Table 6 : Résultats des modèles de régression logistique pour les classes de route 2 et 1 séparées et ensemble. (« * » : <0,05, « ** » : <0,01, « *** » : 0)

	Classe de route 2			Classe de route 1			Classes 1 et 2 ensemble		
	Paramètres estimés	Odd-ratio	P-valeur	Paramètres estimés	Odd-ratio	P-valeur	Paramètres estimés	Odd-ratio	P-valeur
(Intercept)	3,18	24,08	0,28	-4,042	0,018	0,0015**	-1,65	0,19	0,077
Zone bâtie	0,074	1,077	0,81	0,74	2,10	0,020*	0,61	1,84	0,0034**
Intersection	0,47	1,60	0,14	0,51	1,66	0,15	0,59	1,81	0,0077**
Éclairage public	-0,28	0,75	0,30	1,18	3,266	7,68E-05***	0,24	1,27	0,20
Sinuosité	-3,76	0,023	0,16	0,55	1,73	0,59	0,27	1,30	0,73
Largeur	0,018	1,018	0,070	0,023	1,024	0,14	0,016	1,016	0,043*
Volume de véhicules de nuit	0,00028	1,0003	0,041*	-0,0004	1	0,13	-0,0002	1	0,016*
Vitesse moyenne de nuit	-0,0065	0,99	0,51	0,045	1,046	3,75E-06***	0,0123	1,012	0,040*
Nombre d'échantillons	386			345			731		
AIC	507,72			427,25			978,36		
AUC	0,73			0,73			0,66		

Modèle avec la présence d'un éclairage public comme variable dépendante

Ce modèle, prenant la présence d'un éclairage public comme variable dépendante, sert à identifier si des variables favorisent l'installation ou non d'un éclairage public (Table 7). La présence en zone bâtie est une variable significative augmentant la probabilité d'avoir la présence d'un éclairage public tout comme la largeur de la route et le volume de véhicules de nuit. Donc, au plus une route est large et/ou reçoit un volume de véhicules de nuit plus élevé, au plus cette route a une probabilité d'avoir la présence d'un éclairage public. De même, la présence d'un accident répertorié augmente significativement la probabilité d'avoir un éclairage public. Il y a donc de l'endogénéité puisque les zones accidentogènes sont plus susceptibles d'avoir la présence d'éclairage public. Au plus un tronçon a une vitesse moyenne de nuit élevée, au moins il y a de probabilité d'avoir un éclairage public puisque cette variable est significative avec un paramètre estimé négatif. L'AIC pour ce modèle vaut 942,27 et l'AUC vaut 0,52.

Table 7 : Résultats du modèle de régression logistique présentant la variable proximité à un éclairage public comme variable dépendante. Modèle réalisé avec l'ensemble des données soit 1200 cas et points de contrôle. (« * » : <0,05, « ** » : <0,01, « *** » : 0)

Variable	Paramètres estimés	Odd-ratio	P-valeur
(Intercept)	3,35	28,41	0,0014**
Zone bâtie	0,91	2,48	0,00029***
Proximité à une intersection (0_1) (40 mètres)	0,25	1,29	0,37
Présence d'un accident	0,60	1,83	0,0024**
Classe de route : 2	0,81	2,26	0,0023**
Classe de route : 3	0,66	1,94	0,36
Sinuosité	0,34	1,41	0,69

Largeur	0,068	1,07	7,21E-05***
Volume de véhicules de nuit	0,00067	1,0006	1,11E-15***
Vitesse moyenne de nuit	-0,087	0,92	3,30E-28***

4.3.2. Analyse des modèles Random Forest

Modèle Random Forest général

Le modèle Random Forest général reprend, comme pour le premier modèle de régression logistique, l'ensemble des variables, des cas et des points de contrôle (soit 1200 échantillons). La première analyse de ce modèle concerne les importances relatives des variables (Table 8 et Figure 14). La variable ayant le plus d'importance relative est la vitesse moyenne de nuit (31,75%), suivie par le volume de véhicules par tronçon de nuit (31,26%) et par la classification du réseau routier (27,13%) qui sont les trois principales variables influençant le risque d'accident. La largeur de la route est également une variable avec une importance relative comparablement assez importante (22,21%) tout comme la proximité à une intersection (21,67%). Les variables ayant les importances relatives les plus faibles sont la présence en zone bâtie (13,02%), la proximité à un éclairage public (12,47%) et la sinuosité de la route (10,27%). Ce modèle est donc celui utilisé pour construire la carte du risque d'accident prédit par Random Forest. Il est à noter que ce modèle explique 24,95% de la variance avec un RMSE de 0,41. Comme expliqué dans la méthode, 30 itérations du modèle ont été réalisées pour vérifier la stabilité du modèle et donc vérifier que l'importance relative de chaque variable est stable. L'importance relative de chaque variable à chaque itération est visible sur la Figure 15 et montre qu'il n'y a pas de variations extrêmes des importances relatives puisqu'aucune variable importante (>25% d'importance relative) ne devient une variable faiblement importante (<15% d'importance relative) et inversement.

Table 8 : Importance relative des variables avec le modèle Random Forest général. Modèle réalisé avec l'ensemble des données soit 1200 cas et points de contrôle

Variable	%IncMSE (importance relative %)
Zone bâtie	13,02
Intersection	21,67
Eclairage	12,47
Classification du réseau routier	27,13
Sinuosité	10,27
Largeur	22,21
Volume de véhicules (de nuit)	31,26
Vitesse moyenne (de nuit)	31,75

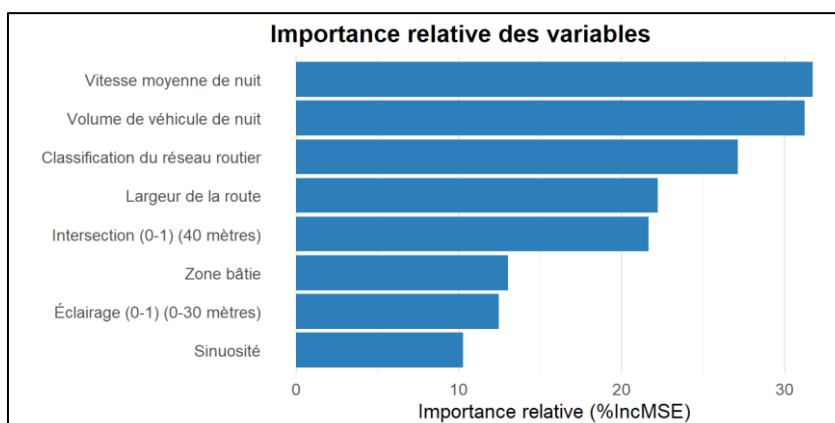


Figure 14 : Importance relative des variables avec le modèle Random Forest général.

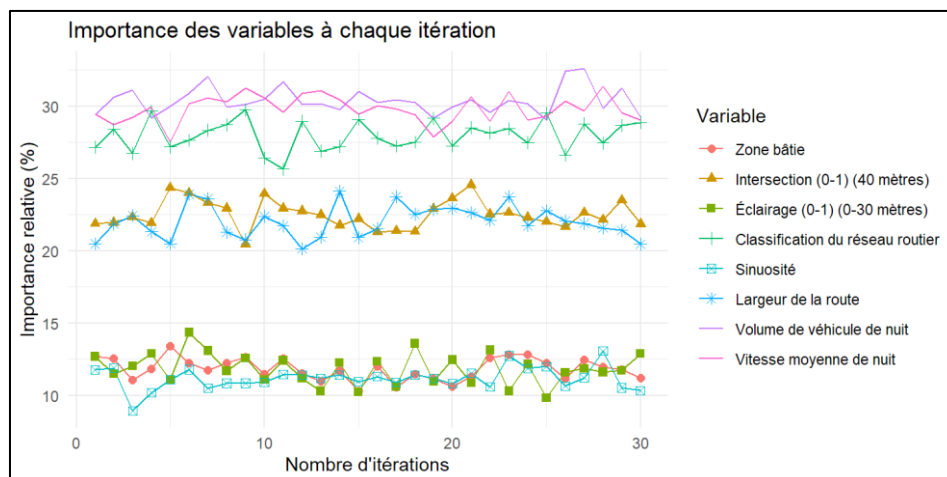


Figure 15 : Importance relative de chaque variable à chaque itération pour le modèle Random Forest général

Les Partial Dependence Plot (Figure 16) sont utilisés pour identifier si une variable (importante ou non) a un impact positif ou négatif sur le risque d'accident. Le modèle prédit que le fait d'être en zone bâtie augmente, légèrement, la probabilité d'accident de 5% puisque la probabilité d'accident hors zone bâtie est de 33% contre 38% en zone bâtie.

La proximité à une intersection augmente également la probabilité d'accident passant d'une probabilité de 31% en dehors d'une intersection, à 41% à proximité d'une intersection soit une augmentation de 10%.

Tout comme dans le modèle de régression logistique, le fait d'être sur une classe de route plus élevée diminue la probabilité d'accident. La probabilité d'accident est donc la plus faible sur la classe de route 3 (gabarit autoroutier) avec 25% soit une diminution de 18% par rapport à la classe de route 1 (ayant une probabilité d'accident de 43%) et une diminution de 7% par rapport à la classe de route 2 (ayant une probabilité d'accident de 25%).

L'effet de la proximité à un éclairage public est faible avec une augmentation de la probabilité d'accident de 3% soit une probabilité d'accident de 35% lorsqu'un éclairage public est présent et de 32% lorsqu'il n'y pas d'éclairage public.

La largeur de la route, une variable continue, montre une probabilité d'accident qui débute à 33% pour la largeur minimale des routes étudiées (2,97 mètres) pour augmenter jusqu'à une probabilité de 35% (correspondant à des routes de 7,69 mètres de largeur). Ensuite, alors que la largeur des routes continue d'augmenter, la probabilité d'accident va diminuer à 32% soit la largeur de route maximale de 12,40 mètres.

La probabilité d'accident liée à la sinuosité des routes est d'abord maximale (37%) pour les routes les moins sinueuses (valeur de 1 soit des routes rectilignes). Ensuite, lorsque la sinuosité augmente, la probabilité d'accident diminue pour atteindre un risque oscillant autour de 32% soit une probabilité 5% plus faible lorsque la sinuosité est plus élevée.

Au plus une route a un volume de véhicules de nuit élevé, au plus la probabilité d'accident diminue. La probabilité d'accident est donc plus élevée pour des tronçons ayant un volume de véhicules de nuit jusqu'à 3530 véhicules (soit une probabilité moyenne de 40%) pour ensuite atteindre une probabilité moyenne de 31% entre 3850 véhicules et 11220 véhicules de nuit. Finalement, la probabilité d'accident est relativement faible (25%) pour les tronçons ayant entre 11540 et 16030 véhicules de nuit.

La vitesse moyenne de nuit n'a d'abord pas d'effet sur la probabilité d'accident, avec une probabilité moyenne de 41% entre 0 et 92 km/h pour ensuite diminuer à une probabilité moyenne d'accident de 29% entre 94 et 120 km/h (correspondant à des routes principalement de classe 3) soit une baisse de 12% de la probabilité d'accident.

Il est possible de voir, contrairement au modèle de régression logistique général, que l'effet des variables ici est hétérogène selon les valeurs prises par les variables, puisqu'un modèle Random Forest est capable de prendre en compte l'effet non-linéaire et non-monotone des variables. Là où l'effet de la vitesse moyenne de nuit dans le modèle de régression logistique est constant et augmente la probabilité d'accident, ici dans le modèle Random Forest, l'effet est d'abord sensiblement constant avant de réduire la probabilité d'accident.

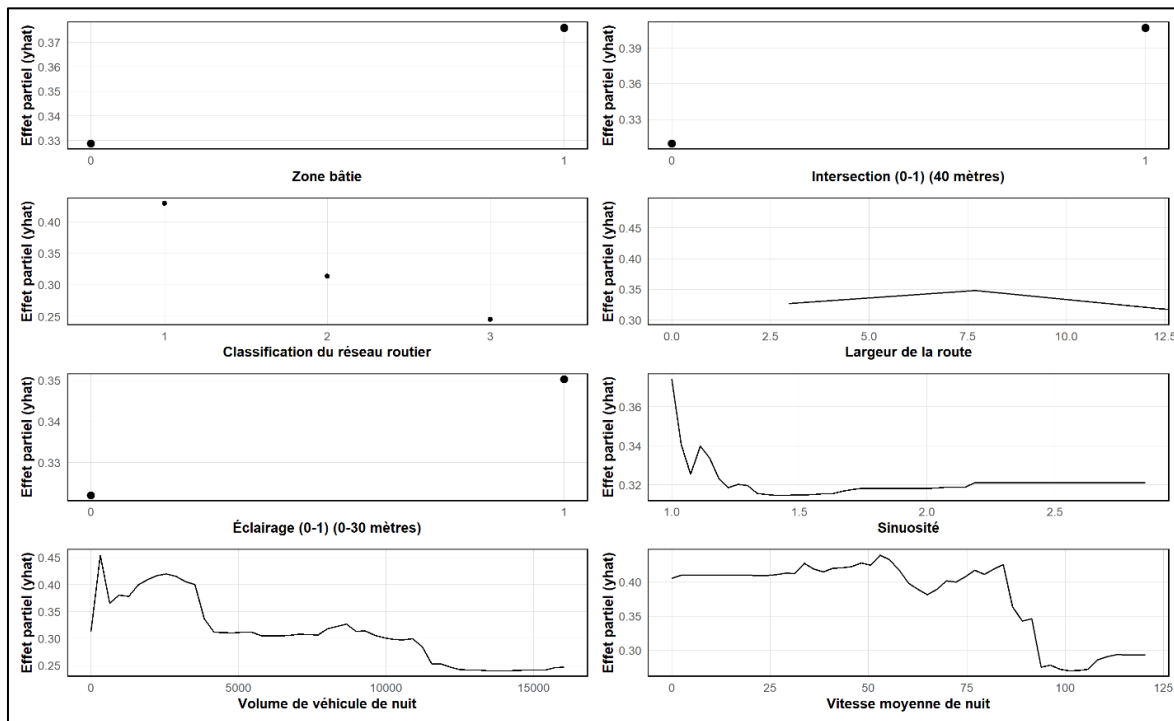


Figure 16 : Partial dependence plot : effet de chaque variable sur le risque d'accident ($\hat{y}$) selon le modèle Random Forest général

Modèle Random Forest pour la classe de route 3

Lorsque le modèle Random Forest est calibré uniquement pour la classe de route 3 (soit 469 échantillons), la vitesse moyenne de nuit reste la variable la plus importante dans le risque d'accident (Figure 17). Cependant, la largeur de la route devient la deuxième variable la plus importante suivie par le volume de véhicules de nuit et la sinuosité. Il est intéressant de noter que la proximité à un éclairage public est une variable avec une importance relative négative pouvant être considérée comme une variable avec une importance nulle. La variance expliquée par ce modèle est de 8,97% avec un RSME valant 0,29.

Modèle Random Forest pour la classe de route 2

Pour la classe de route 2 (386 échantillons), les variables ayant le plus d'importance sont la proximité à une intersection et le volume de véhicules de nuit (Figure 17). La proximité à un éclairage public, la largeur de la route et la vitesse moyenne sont des variables faiblement importantes (<7,5%) alors que la sinuosité et la présence en zone bâtie sont des variables avec une importance relative négative soit des

variables n'ayant pas d'importance. La variance expliquée par ce modèle est de -6,51% avec un RMSE de 0,50

Modèle Random Forest pour la classe de route 1

L'importance relative des variables du modèle pour la classe de route 1 (345 échantillons) (Figure 17) montre que la variable la plus importante, pour cette classe de route, est le volume de véhicules de nuit (34,31%). Les autres variables varient entre 7,65% (zone bâtie) et 14,63% (largeur de la route). La sinuosité présente l'importance la plus faible (<5%). La variance expliquée par ce modèle est de 21,92% avec un RMSE valant 0,43.

Modèle Random Forest pour les classes de route 1 et 2 ensemble

Enfin, lorsque les classes de routes 1 et 2 sont prises ensemble (735 échantillons), l'importance relative des variables montre que le volume de véhicules est, comme le modèle précédent, la variable la plus importante (Figure 17). Les autres variables moyennement importantes varient entre 9,83% (vitesse moyenne de nuit) et 19,75% (proximité à une intersection). La présence en zone bâtie et la sinuosité sont les deux variables les moins importantes (<5%) pour ce modèle. La variance expliquée par ce modèle est de 9,22% avec un RSME de 0,48.

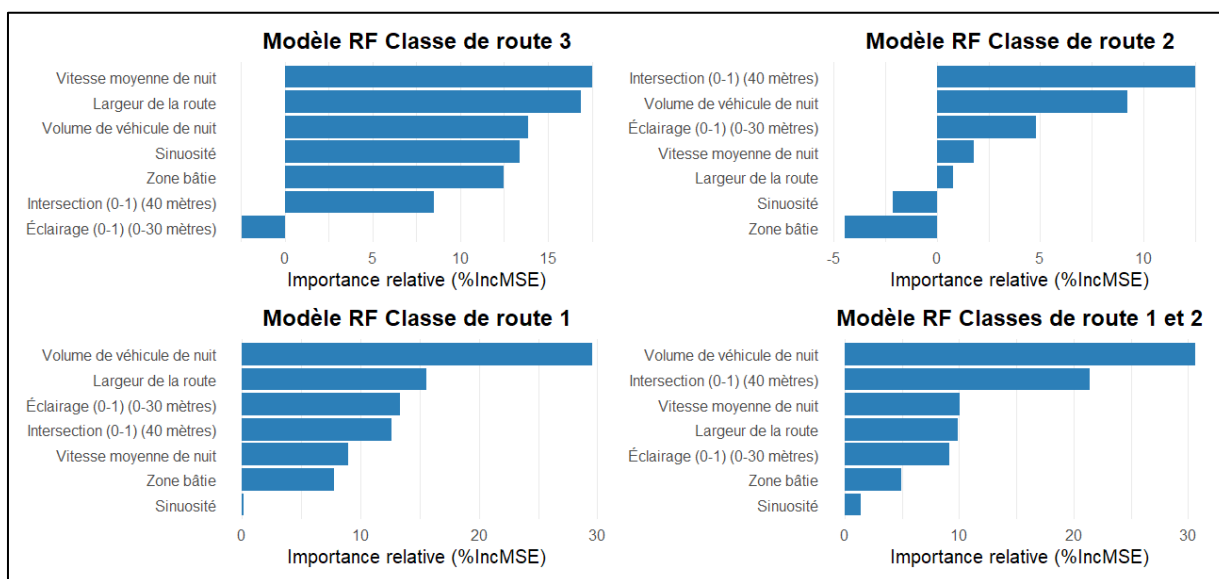


Figure 17 : Importance relative (%IncMSE) des variables pour les modèles Random Forest (RF) de classe de route 3, 2 et 1 séparés et pour le modèle Random Forest (RF) des classes de route 1 et 2 ensemble.

4.4. Cartes du risque d'accident prédit

4.4.1. Par régression logistique

La carte de prédiction du risque d'accident, basée sur les paramètres estimés des variables significatives du modèle de régression logistique général, est visible sur la Figure 18. La classification du réseau routier est représentée par la taille des points, avec la taille la plus élevée pour la classe de route 3 (gabarit autoroutier). Le risque d'accident prédit pour l'ensemble de cette classe est un risque faible inférieur à 0,4. Le risque pour cette classe varie de 0,017 à 0,39 avec un risque moyen de 0,12 et un écart type de 0,058 (Table 9). Les risques d'accident les plus élevés pour cette classe sont localisés à proximité d'intersections, ce qui est encore plus marqué au nord de la N5 (Walcourt), là où la N5 n'a pas un gabarit autoroutier et a donc un nombre plus élevé d'intersection traversant également plus de zone bâtie. La classe 2 (représentée avec la taille de point intermédiaire) présente un risque d'accident variant entre

0,089 et 0,996, le risque d'accident moyen étant de 0,42 et l'écart-type vaut 0,15. Enfin, pour la classe 1 (représentée par la taille de point la plus faible) le risque moyen d'accident est de 0,58, soit le risque moyen le plus élevé par rapport aux autres classes de route. Le risque d'accident varie entre 0,20 et 0,998 et l'écart-type vaut 0,16. Pour les classes de routes 1 et 2, le risque d'accident se concentre à proximité des intersections et à l'intérieur des zones bâties. C'est d'ailleurs avec la combinaison de ces deux variables que le risque d'accident devient élevé ($>0,81$). Les risques les plus élevés se localisent dans les zones bâties et à proximité d'une intersection, sur les nationales les plus importantes ayant alors une largeur de route augmentant encore le risque d'accident. Ces routes possèdent également des vitesses moyennes de nuit plus élevées et sont majoritairement éclairées.

Table 9 : Risque prédit moyen, minimum et maximal par le modèle de régression logistique (RL) pour chaque classe de route ainsi que l'écart-type

Classe de route	Risque moyen	Risque minimum	Risque maximum	Écart-type
3	0,12	0,017	0,39	0,058
2	0,42	0,089	0,996	0,15
1	0,58	0,20	0,998	0,16

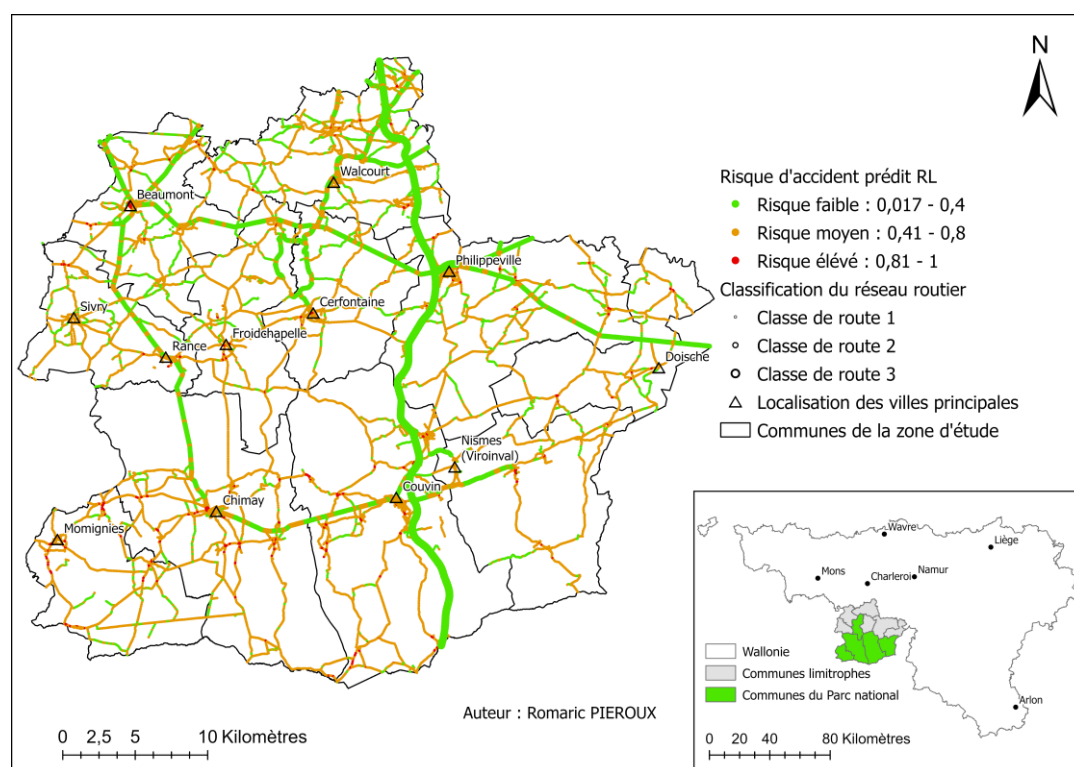


Figure 18 : Carte du risque prédit par le modèle de régression logistique pour les trois classes de route (RL). La classification du réseau routier est représentée par la taille des points.

4.4.2. Par Random Forest

Dans la globalité, le modèle Random Forest donne des résultats sensiblement semblables à ceux du modèle de régression logistique (Figure 19 et Table 10). Bien que le risque d'accident moyen prédit pour la classe 3 soit également de 0,12 comme pour le modèle de régression logistique, le risque d'accident varie entre 0,0024 et 0,79 ce qui est une variation bien plus importante du risque que pour le modèle de régression logistique. L'écart-type pour cette classe vaut 0,14, ce qui est également plus élevé que dans le modèle de régression logistique. Le phénomène observé avec le modèle de régression logistique,

c'est-à-dire, le fait que le sud de la N5 ait un risque d'accident plus faible que le nord se vérifie également avec le modèle Random Forest et est même amplifié. Pour la classe de route 2, le risque d'accident varie de 0,058 à 0,92 pour un risque moyen de 0,40 et un écart-type de 0,16 ce qui est sensiblement similaire aux prédictions du modèle de régression logistique. Enfin, le modèle Random Forest prédit, pour la classe de route 1, un risque moyen d'accident de 0,48, ce qui est un risque moyen plus faible par rapport au modèle de régression logistique. Le risque d'accident pour cette classe varie entre 0,025 et 0,94 pour un écart-type de 0,22 ce qui est plus élevé que dans le modèle de régression logistique. Le modèle Random Forest se montre plus précis grâce à la prise en compte de l'effet non linéaire et non monotonique des variables.

Table 10 : Risque prédit moyen, minimum et maximal par le modèle Random Forest (RF) pour chaque classe de route

Classe de route	Risque moyen	Risque minimum	Risque maximum	Écart-type
3	0,12	0,0024	0,79	0,14
2	0,40	0,058	0,92	0,16
1	0,48	0,025	0,94	0,22

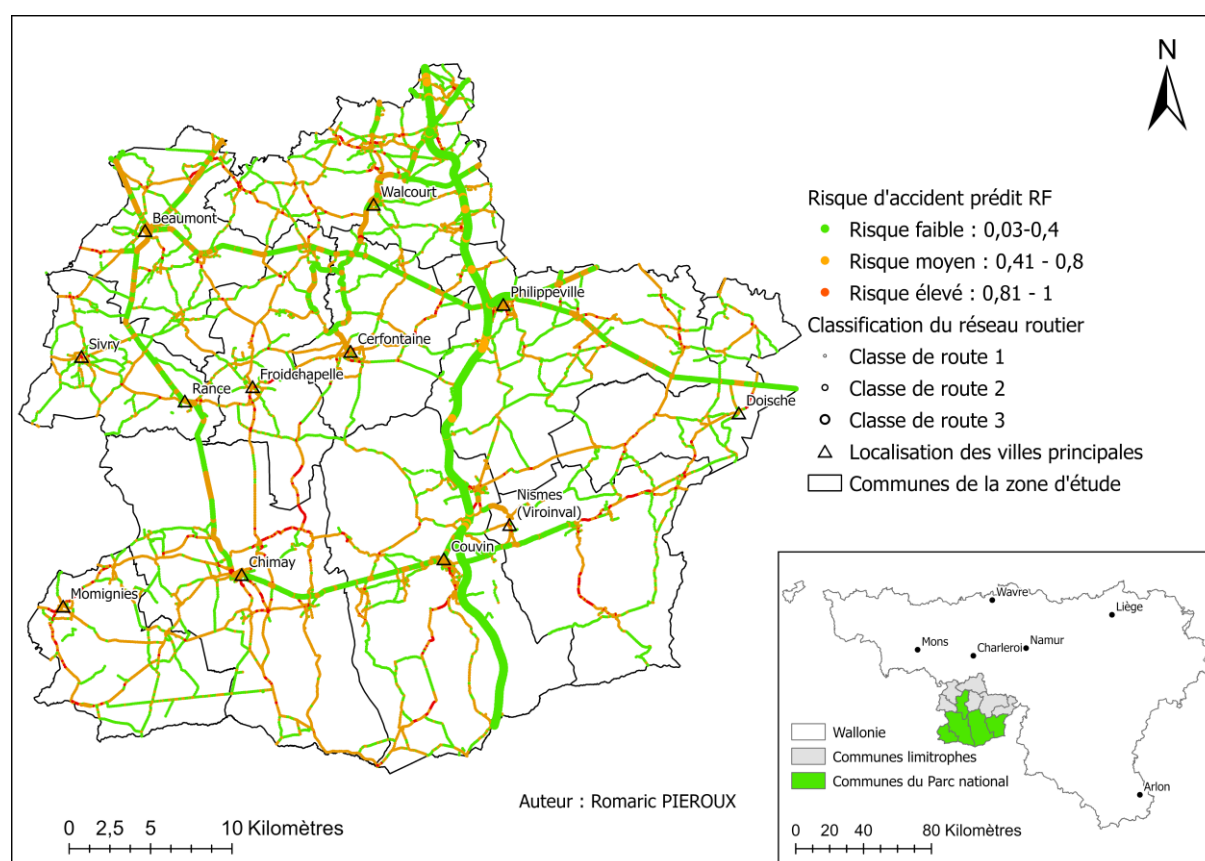


Figure 19 : Carte du risque prédit par le modèle Random Forest (RF). La classification du réseau routier est représentée par la taille des points

4.4.3. Comparaison

Comme déjà montré, les deux modèles montrent des différences de risque d'accident moyen pour certaines classes de routes et des variations du risque d'accident différentes. La différence entre les deux modèles est, en moyenne, de -0,0025 pour la classe 3, 0,021 pour la classe 2 et 0,093 pour la classe 3 (Table 11). Le modèle Random Forest prédit donc, en moyenne, un risque plus faible sauf pour la classe

de route 3. L'écart type de la différence de prédiction entre les deux modèles, montre un écart type relativement plus faible pour la classe de route 3 (avec un écart-type de 0,12). Les classes de route 1 et 2 ayant respectivement un écart-type de 0,17 et 0,15. Un t-test a été réalisé pour chaque classe de route afin de pouvoir dire s'il existe une différence significative entre les prédictions des deux modèles. Il s'avère que pour la classe de route 3, qu'il n'y a pas de différence significative entre les deux modèles, la P-valeur étant supérieure à 0,05. Pour les classes de routes 1 et 2, les deux modèles présentent une différence significative dans leur prédiction du risque d'accident.

Table 11 : Différences du risque d'accident prédit entre le modèle de régression logistique et le modèle Random Forest et résultats des t-tests testant la significativité de la différence entre les deux modèles.

Classe de route	Différence moyenne du risque prédit	Écart-type	P-valeur du t-test
3	-0,0025	0,12	0,39
2	0,021	0,15	2,2 E-16 ***
1	0,093	0,17	2,2 E-16 ***

4.5. Poteaux d'éclairage public potentiellement inutiles (en ce qui concerne la sécurité routière)

Cette section s'intéresse au nombre de lampadaires pouvant être identifiés comme potentiellement inutiles, par rapport à la sécurité routière, selon un seuil de risque d'accident choisi. Cette identification se fait principalement avec les prédictions de risque d'accident du modèle Random Forest général puisque ce modèle est mieux adapté aux données. Le risque d'accident peut être utilisé pour identifier les poteaux d'éclairage potentiellement inutiles puisque la présence d'un éclairage public n'a soit pas d'effet sur le risque d'accident, soit a un effet positif sur le risque d'accident. La répartition du nombre de lampadaires par classe de route est montrée dans la Table 12, permettant de visualiser la situation actuelle.

Il est déjà possible de classer l'ensemble des poteaux d'éclairage public placés le long des routes de la classe 0 (<62 véhicules sur le mois d'août 2024, ces tronçons n'ayant pas été étudiés dans ce mémoire) comme étant potentiellement inutiles en ce qui concerne la sécurité routière. Ces lampadaires sont placés sur des routes ayant moins de deux véhicules par jour et dont certaines routes n'ont aucun passage de nuit. Au total, cela concerne déjà 8707 lampadaires. Ensuite, la Table 13 et la Figure 20 montrent pour chaque classe de route, et chaque seuil de risque d'accident, le nombre de lampadaires potentiellement inutiles en ce qui concerne la sécurité routière. Il s'agit donc du nombre et du pourcentage de lampadaires ayant un risque d'accident inférieur au seuil correspondant. Ainsi, il est possible de choisir un seuil acceptable en faisant la balance entre le coût énergétique et environnemental des lampadaires d'éclairage public et le risque d'accident. La classe de route 3 est la classe ayant le plus de lampadaires potentiellement inutiles avec 78% des lampadaires identifiés comme tel pour un seuil de risque d'accident placé à 0,2 (78% des lampadaires de cette classe sont donc placés à des endroits ayant un risque d'accident inférieur à 0,2). À l'inverse, et pour le même seuil de risque d'accident, la classe de route 2 semble posséder le plus de poteaux d'éclairage potentiellement utiles (8,2% des poteaux d'éclairages pouvant être considérés comme potentiellement inutiles). Avec ce seuil de 0,2, 15945 lampadaires, soit 67,4% des 23647 lampadaires de la zone d'étude peuvent être considérés comme potentiellement inutiles (Table 14). Lorsque le seuil passe à 0,4, alors l'ensemble des classes ont plus de 59% de leurs lampadaires pouvant être identifiés comme potentiellement inutiles, et ce pourcentage monte à 92,4% pour la classe de route 3. Ces mêmes informations sont également disponibles pour le modèle de régression logistique en annexe (Figure A. 7).

La Figure 21 montre le risque d'accident prédit, par le modèle Random Forest, associé à chaque lampadaire (le risque associé est donc celui du tronçon routier le plus proche). Les lampadaires associés aux risques d'accident les plus élevés se situent dans les zones bâties et à proximité des intersections. Ceux-ci sont donc ceux qui seraient retirés en dernier en cas d'un retrait progressif, par seuil de risque, des poteaux d'éclairage. Les lampadaires ayant les risques d'accident associés les plus faibles sont situés en dehors des zones bâties et majoritairement sur des routes de classes 1 et 3. La Figure A. 8, en annexe, montre le risque d'accident prédit par le modèle de régression logistique, associé à chaque lampadaire.

Table 12 : Nombre de lampadaires actuels par classe de route

Classe de route	Nombre de lampadaires actuels
0	8707
1	11005
2	2349
3	1586
Total	23647

Table 13: Nombre et pourcentage de lampadaires potentiellement inutiles en ce qui concerne la sécurité routière pour chaque classe de route et selon différents seuils de risque d'accident prédit par le modèle Random Forest

Classe de route	Seuil de risque d'accident	Nombre de lampadaires potentiellement inutiles sous le seuil de risque	Pourcentage (%)
Classe de route 0	/	8707	100
Classe de route 1	0,2	5811	52,8
	0,4	6913	62,8
	0,6	8910	81
	0,8	10542	95,8
	0,9	10991	99,9
Classe de route 2	0,2	190	8,1
	0,4	1399	59,6
	0,6	2067	88
	0,8	2330	99,2
	0,9	2349	100
Classe de route 3	0,2	1237	78
	0,4	1466	92,4
	0,6	1572	99,1
	0,8	1586	100
	0,9	1586	100

Table 14 : Total du nombre de lampadaires potentiellement inutiles par seuil de risque d'accident prédit par le modèle Random Forest

Seuil de risque	Nombre total de lampadaires potentiellement inutiles par seuil de risque
0	8707
0,2	15945
0,4	18485
0,6	21256
0,8	23165
0,9	23633

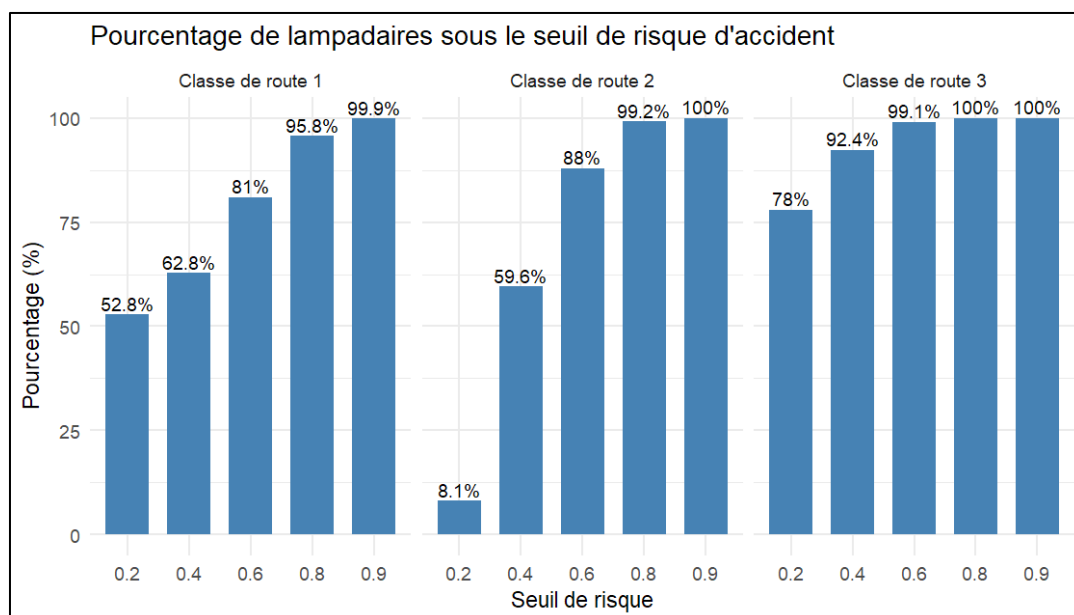


Figure 20 : Pourcentage de lampadaires potentiellement inutiles par classe de route et selon différents seuils de risque d'accident prédits par le modèle Random Forest

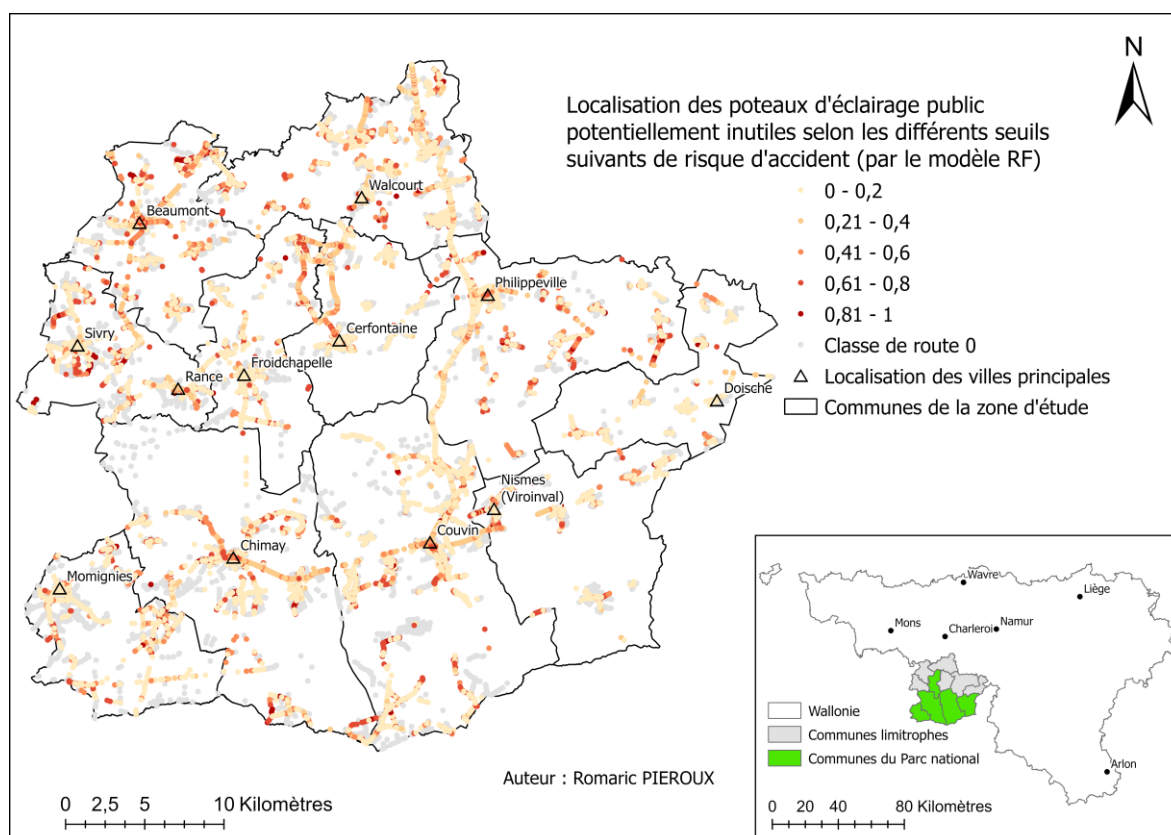


Figure 21 : Localisation des lampadaires potentiellement inutiles selon différents seuils de risque prédits par le modèle Random Forest. Les lampadaires situés sur une route de classe 0 (<2 véhicules par jour) sont représentés en gris

5. Discussion

Les résultats ont permis de mettre en lumière les variables influençant significativement le risque d'accident pour les deux modèles utilisés. Le modèle de régression logistique général a ainsi montré que la présence en zone bâtie, la proximité à une intersection, la proximité à un éclairage public, la largeur des routes sont des variables significatives augmentant le risque d'accident au contraire de la classification du réseau routier qui tend à diminuer le risque. La sinuosité et le volume de véhicules de nuit ne sont pas des variables significatives. Le modèle Random Forest a, quant à lui montré que la vitesse moyenne de nuit et le volume de véhicules de nuit sont les deux variables les plus importantes, la présence d'un éclairage public étant une variable moins importante et ayant un effet positif sur le risque d'accident. Chaque modèle associé à une unique classe de route a ensuite montré des résultats différents. Un modèle (de régression logistique) a permis de tester la présence d'éclairage public comme variable dépendante montrant ainsi que la présence en zone bâtie, la présence d'un accident répertorié, être sur une route classe 2, la largeur des routes et le volume de véhicules augmentent la probabilité d'avoir la présence d'un éclairage public au contraire de la vitesse moyenne de nuit qui diminue cette probabilité.

Avec les deux modèles généraux (régression logistique et Random Forest), il a été possible de prédire un risque d'accident pour le réseau routier étudié (excepté les tronçons avec moins de 62 véhicules sur le mois d'août 2024). Le modèle Random Forest s'est alors montré plus précis puisqu'il prend en compte l'effet non linéaire des variables. Avec la localisation des lampadaires et les prédictions de risque d'accident, il a été possible, pour chaque classe de route et pour différents seuils de risque d'accident, d'identifier les poteaux d'éclairage public potentiellement inutiles.

Ce mémoire a permis également de prendre en considération les problèmes méthodologiques d'anciennes études qui étaient notamment liés au manque de contrôle des autres variables jouant sur la sécurité routière et à des biais statistiques dans le choix des localisations des échantillons. Les biais statistiques dans le choix des localisations des échantillons ont été évités dans ce mémoire en prenant une approche cas-contrôle, dont la localisation des points de contrôle a été choisie aléatoirement évitant donc tout biais statistique de ce point de vue. Le manque de contrôle des autres variables a été évité en identifiant avec la littérature et en étudiant les principales variables jouant un rôle dans le risque d'accident. Bien entendu, ce mémoire ne permet pas de contrôler l'entière des variables jouant un rôle dans le risque d'accident tant le risque d'accident dépend d'un nombre élevé de facteurs dont certains, non pris en compte dans ce mémoire, sont explicités dans les pistes d'amélioration.

5.1. Impact de l'éclairage public sur le risque d'accident

Le modèle de régression logistique général et le modèle de régression logistique montrent un effet positif et significatif de l'éclairage public sur le risque d'accident. De même, le modèle Random Forest général montre un effet positif mais, la proximité à un éclairage est l'une des variables avec l'importance relative la plus faible. Il en est de même pour les autres modèles Random Forest par classe de route, cette variable ayant même une importance relative nulle dans le modèle pour la classe de route 3. Néanmoins, bien que ce mémoire montre que l'éclairage public a un effet statistique positif sur le risque d'accident, il est important de prendre en compte un important phénomène d'endogénéité. Le modèle exprimant l'éclairage public comme variable dépendante montre notamment que les localisations ayant répertorié un accident ont plus de probabilité d'avoir la présence d'un lampadaire d'éclairage public. Cela signifie donc que les zones accidentogènes ont plus de chances d'avoir un lampadaire et l'effet de l'éclairage public sur le risque d'accident. Il est donc difficile de définir réellement l'effet de l'éclairage public sur le risque d'accident. Ce problème d'endogénéité peut notamment être surmonté en étudiant des séries

temporelles et ayant des informations sur les poteaux d'éclairage réellement allumés. Il serait également possible, avec ces séries temporelles, d'étudier le retrait ou l'ajout de poteaux d'éclairage. Avec ces informations supplémentaires, il serait alors possible d'étudier s'il y a une augmentation ou une diminution significative du nombre d'accident lié à une hausse ou une diminution de l'éclairage public.

En théorie, ces résultats peuvent aussi s'expliquer par une prise de confiance des conducteurs la nuit lorsque les routes sont illuminées (Assum *et al.*, 1999). Cette prise de confiance mène à prendre plus de risques, comme par exemple, en dépassant les vitesses autorisées. Mais cela mène également à baisser sa concentration alors que la conduite de nuit est plus exigeante puisque la fatigue est plus présente. L'étude, (Assum *et al.*, 1999), montre alors que la diminution potentielle du risque provoquée par l'illumination des routes est compensée par le comportement des conducteurs prenant plus de risques.

L'effet positif de l'éclairage public sur le risque d'accident est à prendre avec précaution, notamment avec le problème d'endogénéité, mais également dû au fait que des études montrent l'importance d'éclairer les passages piétons, les piétons étant difficilement visibles la nuit par les automobilistes et ce par manque de contrastes (Retting *et al.*, 2003; Struyf *et al.*, 2019).

L'interaction entre l'éclairage public et la proximité à une intersection s'est révélée comme étant non-significative. D'ailleurs ce modèle n'a pas de réelle valeur ajoutée puisque qu'il s'agit sensiblement du même modèle que le modèle de régression logistique général et ne sera donc pas plus interprété dans la suite. Le fait que l'interaction ne soit pas significative peut notamment s'expliquer simplement, comme cela a été le cas dans la partie résultat, que l'effet de l'éclairage sur le risque d'accident ne se concentre pas uniquement aux intersections mais sur l'ensemble du réseau routier.

5.2. Impact des autres variables sur le risque d'accident

Zone bâtie

Comme montré en partie dans la méthode, être en zone bâtie augmente le risque d'accident selon la littérature. Les résultats obtenus dans ce mémoire vont également dans ce sens puisque le modèle de régression logistique général a montré un effet positif de cette variable sur le risque d'accident, l'effet étant significatif. L'augmentation du risque d'accident peut notamment s'expliquer par un nombre plus élevé d'intersection nécessitant une plus grande concentration sur une durée relativement importante, mais également par une diminution de la visibilité aux intersections (présence de bâtiments) (Basu *et al.*, 2022). Il est tout de même à noter que cette variable n'est pas significative dans les modèles de régression logistique pour chaque classe de route. En effet, cette variable est significative dans le modèle général, dans le modèle spécifique à la classe de route 1 et dans le modèle prenant les classes de routes 1 et 2 ensemble. Le fait que cette variable ne soit pas significative dans le modèle pour la classe de route 3 peut s'expliquer par le fait que les routes de cette classe (essentiellement la N5) ne rentrent pas dans les zones bâties étant donné le gabarit autoroutier. Ces routes évitent justement les zones bâties et il n'y a donc pas d'effet sur le risque d'accident. Ensuite, le fait que cette variable ne soit significative que pour la classe de route 1 et le modèle prenant les classes de routes 1 et 2 peut s'expliquer par le fait que ce sont les routes locales de classe 1 qui traversent réellement les zones bâties. En effet, lorsqu'une route de classe 2 (type nationale) arrive dans une ville, celle-ci se divise en plusieurs routes de classe 1 (le volume de véhicules se répartit sur plusieurs routes). Seules les plus grandes nationales restent de classe 2, mais celles-ci sont moins représentées par rapport aux routes de classe 1 en zone bâtie. Dans les villes, en milieu-rural, il n'y a souvent qu'une grande nationale traversante, le reste du réseau routier en ville étant constitué de routes locales de classe 1. Le même raisonnement s'applique aux modèles Random Forest où la variable est principalement importante dans le modèle pour la classe de route 1 mais aussi pour la classe de route 3 ce qui s'explique par le fait que le nord de la N5 (Walcourt) traverse tout de

même des zones bâties de façon directe expliquant également la hausse du risque prédit à cet endroit contrairement au modèle de régression logistique.

Proximité à une intersection

La proximité à une intersection augmente le risque de façon significative dans le modèle de régression logistique général et dans le modèle avec les classes de routes 1 et 2 ensemble. C'est une variable moyennement importante dans le modèle Random Forest général et les autres modèles sauf pour le modèle de la classe de route 2, où il s'agit de la variable la plus importante. Les intersections augmentent le risque d'accident, dans ce mémoire mais également selon la littérature (Basu *et al.*, 2022; Keler *et al.*, 2025; Ullah *et al.*, 2025), puisque qu'il s'agit d'endroits dangereux nécessitant une attention particulière des conducteurs et où les erreurs peuvent mener plus rapidement à un accident (traversée de routes, savoir qui a la priorité et la céder si nécessaire). Le fait que cette variable ne soit pas significative dans le modèle pour la classe de route 3 est un résultat attendu au vu de la construction des routes de cette classe. Celles-ci sont justement pensées pour éviter un maximum d'intersections, et les rares intersections sont adaptées avec des bretelles d'accès diminuant les difficultés d'insertion. Le fait que cette variable soit significative pour la classe de route 2 mais non pour la classe de route 1 peut s'expliquer par le fait que les intersections des routes de classe 2 sont plus importantes et plus dangereuses puisqu'il s'agit de routes avec des volumes de véhicules plus élevés mais également parce que ces routes ont des vitesses moyennes plus élevées laissant moins de place aux erreurs éventuelles. Là où les intersections sur routes de classe 1 sont relativement plus faciles avec moins de véhicules et des vitesses moyennes plus faibles permettant un freinage d'urgence. Enfin, les accidents sur routes de classe 1 sont plus sujets à des accidents légers (vitesses plus faibles, accidents seuls, ...) menant à une sous-estimation des accidents tandis que ceux sur routes de classe 2 sont plus visibles, ayant lieu sur des nationales plus importantes et menant donc plus souvent à l'appel des forces de police (Elvik & Mysen, 1999).

La visibilité aux intersections est un facteur encore plus précis du rôle joué dans le risque d'accident, mais n'a pas été étudiée dans ce mémoire en raison de la difficulté à construire cette variable (il aurait fallu des données précises de localisation des bâtiments, d'arbres, d'aides comme des miroirs en voirie, ...).

Classification du réseau routier

La classification du réseau routier s'est montrée comme étant l'une des variables les plus importantes ayant un effet sur le risque d'accident. Il est, cependant, important de noter que cette variable reprend en réalité d'autres variables bien qu'aucun problème de multi colinéarité ne se soit posé. Notamment, les routes de classes 3 sont celles-ayant les largeurs de route les plus élevées, les volumes et vitesses moyennes de véhicules de nuit les plus élevés, les valeurs de sinuosités les plus faibles, le moins de proximité à une intersection et le plus de proximité à un éclairage public (l'entière de la N5 étant éclairée sauf une portion au sud comme expliqué plus haut). De ce fait, cette classe de route est associée à des risques d'accident prédit les plus faibles, puisqu'assez peu d'accidents ont eu lieu, proportionnellement au volume de véhicules, sur les routes de classe de 3.

Vitesse moyenne de nuit et volume de véhicules de nuit

Ces deux variables sont étudiées ensemble tant leurs effets sont assez similaires dans le cadre de ce mémoire puisqu'en réalité, ces deux variables sont liées à la classification du réseau routier. Le volume de véhicules de nuit est proportionnel à la classification du réseau routier ce qui explique notamment que le risque d'accident diminue dans le modèle Random Forest quand le volume de véhicules de nuit augmente. De même, la vitesse moyenne de nuit est plus élevée sur les routes de classe 3 (entre 90 et

120 km/h), ce qui explique que le modèle Random Forest montre une diminution du risque d'accident après 90 km/h.

La vitesse étudiée dans ce mémoire est une vitesse moyenne. La littérature ainsi que les instituts de sécurité routière montrent que la vitesse des conducteurs est une variable importante augmentant le risque d'accident (OMS, 2023; Salathé *et al.*, 2024). Mais la vitesse utilisée est alors une vitesse à un instant t d'un conducteur qui conduit, selon les autres variables, à un risque d'accident plus élevé. La vitesse moyenne ne permet pas ici d'identifier les risques pris en dépassant les vitesses autorisées.

Largeur des routes

Tel que montré dans le modèle de régression logistique général et dans le modèle Random Forest général, la largeur des routes est donc une variable ayant un impact positif sur le risque d'accident et étant significative dans le modèle de régression logistique général. L'effet est toujours positif dans le modèle de régression logistique général tandis que dans le modèle Random Forest, l'effet varie selon la largeur des routes. Le fait que la largeur des routes augmente le risque d'accident s'explique notamment par un changement dans la perception du risque par les conducteurs (Azin *et al.*, 2025; Wu & Lin, 2022). Les routes de plus grande largeur donnent aux conducteurs un sentiment de sécurité plus important les poussant à conduire à des vitesses plus élevées et à dépasser les vitesses autorisées (et ce en particulier de nuit). La largeur augmente le risque d'accident dans le modèle de régression logistique général, dans le modèle pour les classes de route 1 et 2 ensemble. Par contre, l'effet de largeur des routes sur le risque d'accident est plus nuancé dans le modèle Random Forest, où l'augmentation de la largeur des routes augmente dans un premier temps le risque d'accident, comme expliqué, mais va ensuite diminuer le risque d'accident. Cette diminution du risque d'accident est visible pour des routes ayant une largeur entre 7,69 et 12,40 mètres, ce qui correspond à des routes de classe 3 (gabarit autoroutier), et ce en particulier au sud de la N5 qui devient alors une route à six voies et ayant un risque d'accident plus faible (aucun accident n'y a été répertorié).

Sinuosité

La sinuosité n'est pas une variable significative dans le modèle de régression logistique général. Cependant, dans les deux modèles de régression logistique liés à la classe de route 3, cette variable était significative avec un effet négatif sur le risque d'accident. De même, le modèle Random Forest montre que la sinuosité est une variable diminuant le risque d'accident même si cette variable est la moins importante. Bien que ce résultat puisse paraître dans un premier temps surprenant, celui-ci est en réalité assez semblable à d'autres études. Surprenant, puisqu'il est possible de penser que la sinuosité augmenterait le risque d'accident au vu des erreurs possibles de conduite (vitesse trop élevée faisant perdre le contrôle du véhicule, risque de glissement, ...). En réalité, et comme le montre la littérature, la sinuosité a tendance à rendre les conducteurs plus prudents par peur justement d'avoir un accident. Cette prudence plus élevée réduit le risque d'accident, d'autant plus que les conducteurs ont tendance à anticiper les routes sinueuses en adaptant leur vitesse, d'autant plus que les virages les plus dangereux sont la plupart du temps signalés à l'aide de panneaux routiers (Ranapurwala *et al.*, 2016). La sinuosité est d'ailleurs une variable significative, diminuant le risque d'accident, dans les modèles de régression logistique pour la classe de route 3. Le fait que la sinuosité ne soit pas une variable significative dans le modèle de régression logistique générale peut s'expliquer par le manque de sinuosité du réseau routier.

5.3. Comparaison et évaluation des modèles

Il est possible de comparer les modèles de régression logistique entre eux grâce aux valeurs d'AIC et d'AUC. L'AIC n'étant utile que dans la comparaison des modèles et n'ayant pas de réelle signification

seule. L'AUC du modèle de régression logistique général est de 0,8, signifiant une bonne performance du modèle à prédire les cas et donc à prédire le risque d'accident. De plus, cette valeur d'AUC (0,8), signifie que la plupart des prédictions sont proches de 1 ou de 0 mais non de 0,5 aidant sensiblement à interpréter les prédictions. La valeur de l'AUC est similaire pour les modèles de régression logistique de classe de route 1 et 2 séparées et vaut 0,73. L'AIC vaut respectivement 427,29 pour la classe de route 1 et 507,72 pour la classe de route 2, montrant une performance égale. Lorsque ces deux classes de routes sont étudiées ensemble alors l'AUC vaut 0,52 et l'AIC vaut 942,27. Le modèle avec les deux classes ensemble se montre donc moins performant. Les deux modèles de la classe de route 3 montrent une performance similaire, AUC de 0,79 et AIC de 282, 83 et 284,02. Ces deux modèles sont également ceux ayant l'AIC le plus faible signifiant une meilleure performance comparativement. Cela peut notamment s'expliquer par le grand nombre d'échantillons utilisé pour cette classe de route alors que cette classe possède la longueur de réseau la plus faible.

Le modèle Random Forest général explique 24,95% de la variance ce qui peut paraître faible mais est expliquée par le fait que l'étude du risque d'accident dépendant de tellement de variables qu'il n'est pas aisé d'avoir un meilleur score. Le modèle spécifique à la classe de route 2 montre une variance expliquée négative, signifiant qu'un modèle aléatoire aurait une meilleure prédiction du risque d'accident.

5.4. Prédiction du risque d'accident et comparaison des deux modèles (RL et RF)

Les deux modèles (RL et RF) montrent des différences dans leur prédiction du risque d'accident bien que les modèles prédisent les risques d'accidents les plus élevés à l'intérieur des zones bâties et à proximité des intersections. Les différences entre les deux modèles sont notamment dues au fait que le modèle de régression logistique fait l'hypothèse que l'effet estimé des paramètres est constant pour les variables tandis que le modèle Random Forest ne fait pas cette hypothèse et où donc l'effet estimé des variables sur le risque est non-monotonique. De ce fait, le modèle Random Forest semble plus adapté aux données (qui ont des effets variables) et donc plus adaptés à prédire le risque d'accident.

5.5. Application des résultats pour des interventions de réduction de la pollution lumineuse

Dans ce mémoire, les poteaux d'éclairage public potentiellement inutiles ont été identifiés sur base de deux hypothèses. La première est que les poteaux d'éclairage situés sur des routes ayant eu moins de 62 véhicules sur le mois d'août 2024 n'ont pas de risque d'accident associé puisque les tronçons routiers les plus proches n'ont pas été étudiés dans ce mémoire. De ce fait, ces poteaux sont considérés comme potentiellement inutiles au vu du faible volume de trafic. La seconde hypothèse est que le retrait ou l'adaptation des poteaux d'éclairage public, se faisant selon un seuil de risque d'accident acceptable, ne va pas augmenter sensiblement le risque et le nombre d'accident.

L'identification des poteaux d'éclairage public potentiellement inutiles permet notamment de prioriser les actions en faveur d'une réduction de la pollution lumineuse comme le retrait de certains poteaux d'éclairage public ou l'adaptation de ceux-ci via un horaire d'éclairage adapté, le dimming ou l'automatisation. Il est bien entendu que chaque adaptation devrait se faire au cas par cas suivant les conditions locales du poteau d'éclairage, ce mémoire ne remplaçant par une réelle identification des poteaux d'éclairage potentiellement inutiles par le terrain. Ces poteaux s'ajoutent ou se recoupent à ceux qui seraient déjà identifiés posant un conflit avec un projet de Trame noire (notamment à proximité de corridors écologiques ou de réservoirs de biodiversité).

5.6. Limites et pistes d'amélioration

Ce mémoire présente certaines limites dont certaines sont liées aux données mêmes de ce mémoire ou à la méthodologie utilisée dont la construction des variables. Des pistes d'amélioration sont proposées à la fin de cette section.

Limites liées aux données

Comme montré, le nombre d'accidents routiers réellement utilisable a dû être revu à la baisse suite à des problèmes de géolocalisation. Cette réduction réduit la taille de l'échantillon ce qui aurait permis une meilleure calibration des différents modèles. Pour autant, cette réduction du nombre d'accident n'introduit pas de biais statistique puisque les accidents non pris en compte ne sont pas sensiblement différents des accidents étudiés. Ensuite, et comme expliqué dans la section correspondante dans la partie donnée, il y a une sous-représentation des accidents légers n'ayant occasionné que des blessés légers (les accidents n'ayant occasionnés que des dégâts matériels n'étant pas répertoriés) ce qui réduit encore plus le volume d'échantillon réel. Cela mène aussi à prendre des localisations comme points de contrôle sans être réellement sûr qu'il n'y pas eu d'accident durant la période étudiée.

Les données TomTom, ne concernant que le mois d'août 2024, posent également quelques limites. Notamment, le fait que ces données ne couvrent qu'un seul mois réduit la significativité statistique des volumes de véhicules de nuit et vitesse moyenne de nuit par tronçon. Le mois d'août est, de plus, un mois particulier puisque c'est un mois de départ en vacances, où les trajets vers les écoles sont inexistantes (congés scolaires) et où les trajets vers les sites touristiques sont plus importants que dans le reste de l'année. À ce sujet, il est probable que les routes menant aux Lacs de l'Eau d'Heure et que la N5 (axe emprunté pour rejoindre la France et donc le sud de l'Europe) aient été surreprésentées par rapport au volume de véhicules et par rapport à un mois hors été. Ce manque de significativité a notamment rendu impossible l'utilisation de la V95, soit la vitesse sous laquelle circulent 95% des usagers. Cet indicateur est notamment intéressant pour étudier la vitesse excessive des conducteurs les plus imprudents, dépassant les vitesses autorisées.

Le manque d'informations concernant les réels horaires d'éclairage des lampadaires a rendu impossible l'utilisation de séries temporelles et a donc obligé de poser l'hypothèse que l'ensemble des poteaux d'éclairage sont allumés chaque nuit et avec un horaire similaire.

Limites liées à la méthodologie

L'approche par cas-contrôle a été utilisée pour ne pas avoir de biais statistiques liés au choix de la localisation des points de contrôle. Les points de contrôle ont été répartis aléatoirement sur le réseau routier et répartis entre les trois classes de route selon le volume de véhicules de chaque classe. Le choix du volume de véhicules pour stratifier les points de contrôle s'explique, d'abord, par une volonté de suivre ce qui a déjà été fait dans la littérature, notamment avec l'étude sur le risque d'accident à vélo à Bruxelles (Vandenbulcke *et al.*, 2014). Ce choix s'explique aussi par une volonté de contrôler l'exposition au risque d'accident, les risques d'accident trouvés dans ce mémoire sont donc relatifs au volume de véhicules. La classe de route 3 a un volume de véhicules élevé mais un nombre d'accident relativement faible par rapport aux autres classes de route, l'objectif est alors d'étudier pourquoi cette classe a moins d'accident en y mettant la plus grande proportion de points de contrôle. Il aurait été possible de stratifier les points de contrôle par rapport au nombre d'accident ce qui aurait totalement inversé la répartition des points de contrôle, la classe de route 3 aurait eu le moins de points de contrôle. Cependant, ce choix aurait pu ajouter des biais statistiques puisque la présence d'accidents routiers est la variable dépendante des modèles statiques de ce mémoire.

Le mémoire a ici uniquement utilisé la sinuosité des routes alors que la courbure des routes aurait pu être une autre variable intéressante à tester. La sinuosité des routes peut être comparée comme une valeur moyenne de la courbure des routes pour un tronçon, ayant été défini pour cette variable comme une portion de route entre deux intersections, alors que la courbure des routes sera une mesure locale de la courbure d'un virage.

La définition d'un tronçon routier peut également être une limite liée à la méthodologie. Dans ce mémoire, la définition choisie pour certaines variables a été qu'un tronçon routier est une portion de route entre deux intersections. Concernant les variables volume de véhicules de nuit et vitesse moyenne de nuit, les tronçons sont divisés en taille variant de 100 mètres à 500 mètres (selon la classification du réseau routier utilisée par TomTom).

Dans ce mémoire, les zones bâties ont été identifiées grâce aux secteurs statistiques. Ce qui dans l'ensemble donne d'assez bons résultats puisque, comme expliqué, les secteurs statistiques englobent assez bien les zones bâties. Cependant, et malgré ce bon proxy, il est important de noter que certaines zones bâties pourraient ne pas être prises en compte en ne prenant que la densité des secteurs statistiques. Par exemple, une zone commerciale ou une zone industrielle (qui constituent donc des zones bâties) constituent bien souvent un secteur statistique à part entière mais avec un nombre d'habitant faible voire nul et ne suffisant donc pas pour atteindre le seuil de densité d'habitants permettant de considérer la zone comme une zone bâtie. Ce phénomène est cependant rare sur la zone d'étude et ne constitue pas une limite majeure.

Pistes d'amélioration

Comme mentionné, l'ajout de la courbure des routes constitue une piste d'amélioration permettant de prendre en considération la courbure locale des tronçons. De même, l'utilisation de la V95, pourrait être un indicateur intéressant.

Ce mémoire n'a pas pu prendre en compte l'ensemble des facteurs jouant un rôle dans le risque d'accident. Notamment, les conditions météorologiques peuvent accentuer le risque d'accident comme par exemple, les distance de freinages qui sont plus élevées sur routes humides, la visibilité réduite en cas de brouillard ou de précipitations ou encore la présence de verglas. Les conditions météorologiques n'ont pas été prises en compte dans ce mémoire puisqu'il aurait fallu construire des séries temporelles nécessitant d'avoir des données avec une meilleure résolution temporelle.

Pour surmonter le problème d'endogénéité, il est possible de travailler avec des séries temporelles et en ne faisant plus l'hypothèse que l'ensemble des poteaux d'éclairage public sont actifs sur l'ensemble de la nuit et de l'année. Cela permet également d'étudier l'impact potentiel de l'ajout, du retrait ou de l'adaptation des horaires d'éclairages des lampadaires. Travailler avec des séries temporelles permettraient également d'étudier l'impact des conditions météorologiques.

6. Conclusion

La volonté de réduire la pollution lumineuse s'oppose à d'autres enjeux, comme la sécurité routière. L'objectif de ce mémoire était donc d'apporter une méthodologie pour identifier l'impact potentiel de l'éclairage public sur le risque d'accident. D'autres variables ont également été étudiées comme la classification du réseau routier, le volume de véhicules de nuit, la vitesse moyenne de nuit, la largeur des routes, la sinuosité des routes, la proximité à une intersection ainsi que la présence ou non en zone bâtie. Ces variables ont été étudiées au moyen d'une approche cas-contrôle, prenant les localisations d'accidents comme cas, et des localisations sans accident comme points de contrôle, ceux-ci étant répartis sur le réseau routier proportionnellement au volume de véhicules de chaque classe de route créée avec la classification du réseau routier. Ensuite, l'effet des variables a été étudié au moyen de deux modèles statistiques, le modèle de régression logistique et le modèle Random Forest.

Les résultats du modèle de régression logistique montrent ainsi que l'éclairage public est une variable qui augmente le risque d'accident tout comme la présence en zone bâtie, la proximité à une intersection, la largeur des routes et la vitesse moyenne de nuit. Le fait de circuler sur une route de classe 3 (gabarit autoroutier) réduit le risque d'accident par rapport aux deux autres classes de route. La sinuosité et le volume de véhicules de nuit n'étant pas des variables significatives. Le modèle Random Forest a permis d'être plus nuancé puisque ce modèle prend en compte l'effet non monotone des variables, contrairement au modèle de régression logistique. Notamment, le modèle Random Forest montre une diminution du risque d'accident avec les routes plus sinueuses, mais également avec les routes ayant une vitesse moyenne de nuit comprises entre 90 et 120 km/h (correspondants alors à des routes de gabarit autoroutier soit de classe 3). Malgré que ces deux modèles identifient un effet positif de l'éclairage public sur le risque d'accident, il est important de considérer un problème d'endogénéité. En effet, les zones accidentogènes ont plus de chances d'avoir la présence d'un lampadaire. Ce problème rend donc difficile l'interprétation de l'effet de l'éclairage public sur le risque d'accident. La hausse du risque d'accident dû à l'éclairage public pourrait également s'expliquer par une compensation dû au comportement des conducteurs ayant tendance à avoir des vitesses plus élevées et une baisse de la concentration lorsque les routes sont illuminées.

Avec ces résultats, provenant des deux modèles généraux, il a été possible de prédire un risque d'accident sur l'ensemble du réseau routier et d'associer chaque lampadaire au risque d'accident du tronçon routier le plus proche. Ensuite, les lampadaires potentiellement inutiles, par rapport à la sécurité routière, ont été identifiés selon plusieurs seuils de risque d'accident permettant donc de prioriser les actions de réduction de la pollution lumineuse vers ces points lumineux.

Ce mémoire n'est pas exempt de limites et des pistes d'améliorations ont pu être mises en lumière comme l'utilisation de séries temporelles permettant d'étudier l'installation, le retrait de poteaux d'éclairages public ou l'adaptation des horaires d'éclairages. L'ajout d'autres indicateurs paraît également être intéressant comme la courbure des routes, la V95, ou encore la prise en compte des conditions météorologiques. Les limites principales de ce mémoire concernent les données avec un nombre d'accidents étudiés qui a dû être revu à la baisse à la suite de problèmes de géolocalisations et de sous estimations des accidents routiers légers. Le fait que la classification du réseau routier repose sur l'utilisation de données de volume de véhicules ne provenant que du mois d'août peut poser des problèmes de représentativité des données.

7. Bibliographie

ASSUM, T., BJØRNSKAU, T., FOSSER, S. & SAGBERG, F., (1999), Risk compensation—the case of road lighting, *Accident Analysis & Prevention*, vol. 31, n°5, p. 545-553.

AZIN, B., EWING, R., YANG, W., PROMY, N. S., KALANTARI, H. A. & TABASSUM, N., (2025), Urban Arterial Lane Width Versus Speed and Crash Rates: A Comprehensive Study of Road Safety, *Sustainability*, vol. 17, n°2, p. 628.

BASU, S., FERRANTE, C. & DE BLASIIS, M. R., (2022), Influence of Intersection Density on Risk Perception of Drivers in Rural Roadways: A Driving Simulator Study, *Sustainability*, vol. 14, n°13, p. 7750.

BEBRONNE, E., HENG, S. & LIMBOURG, S., (2024), Towards sustainable nocturnal environment management: a social-ecological-technical system analysis in Wallonia (Belgium), *Discover Environment*, vol. 2, n°1, p. 98.

BENNIE, J., DAVIES, T. W., DUFFY, J. P., INGER, R. & GASTON, K. J., (2014), Contrasting trends in light pollution across Europe based on satellite observed night time lights, *Scientific Reports*, vol. 4, n°1, p. 3789.

BEYER, F. R. & KER, K., (2010), Street lighting for preventing road traffic injuries - Beyer, FR - 2009 | Cochrane Library.

BIAU, G. & SCORNET, E., (2016), A random forest guided tour, *TEST*, vol. 25, n°2, p. 197-227.

BOUYER, J., (2024), La régression logistique en épidémiologie.

BOYCE, P. R., (2003), *Human Factors in Lighting*, CRC Press, London, 602 p.

BULLOUGH, J. D., DONNELL, E. T. & REA, M. S., (2013), To illuminate or not to illuminate: Roadway lighting as it affects traffic safety at intersections, *Accident Analysis & Prevention*, vol. 53, p. 65-77.

BÜTTNER, G., KOSZTRA, B., SOUKUP, T., SOUSA, A. & LANGANKE, T., (2017), CLC2018 Technical Guidelines.

CADENA-URZÚA, P., GUARDIOLA, J. & MONTES, F., (2025), The impact of public lighting improvement on crimes and incivilities reduction in Chile, *Frontiers in Sociology*, vol. 10, .

CHALFIN, A., HANSEN, B., LERNER, J. & PARKER, L., (2022a), Reducing Crime Through Environmental Design: Evidence from a Randomized Experiment of Street Lighting in New York City, *Journal of Quantitative Criminology*, vol. 38, n°1, p. 127-157.

CHALFIN, A., KAPLAN, J. & LAFOREST, M., (2022b), Street Light Outages, Public Safety and Crime Attraction, *Journal of Quantitative Criminology*, vol. 38, n°4, p. 891-919.

COPERNICUS, (2021), Copernicus Land Monitoring Service-CORINE Land Cover.

CWAPE, (2025), Rapport coût de l'OSP éclairage public 2024.

DECOCQ, G., (2024), *Écologie historique*, ISTE Group.

DIENST STRATEGISCHE ANALYSE FEDERALE POLITIE WEST-VLAANDEREN, (2017), Analyse straatverlichting West-Vlaanderen. Wat is de impact van het doven van de straatverlichting op criminaliteit en verkeersveiligheid?

- DONNELL, E. T., PORTER, R. J. & SHANKAR, V. N., (2010a), A framework for estimating the safety effects of roadway lighting at intersections, *Safety Science*, vol. 48, n°10, p. 1436-1444.
- DONNELL, E. T., PORTER, R. J. & SHANKAR, V. N., (2010b), A framework for estimating the safety effects of roadway lighting at intersections, *Safety Science*, vol. 48, n°10, p. 1436-1444.
- EDWARDS, C. J., (2015), Lighting Levels for Isolated Intersections: Leading to Safety Improvements.
- EL SANHARAWI, M. & NAUDET, F., (2013a), Comprendre la régression logistique, *Journal Français d'Ophthalmologie*, vol. 36, n°8, p. 710-715.
- EL SANHARAWI, M. & NAUDET, F., (2013b), Comprendre la régression logistique, *Journal Français d'Ophthalmologie*, vol. 36, n°8, p. 710-715.
- ELVIK, R., (1995), Meta-Analysis of Evaluations of Public Lighting as Accident Countermeasure, *Transportation Research Record*, 1485(1), 12-24.
- ELVIK, R. & MYSEN, A., (1999), Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries, *Transportation Research Record*, vol. 1665, n°1, p. 133-140.
- FALCHI, F., CINZANO, P., DURISCOE, D., KYBA, C. C. M., ELVIDGE, C. D., BAUGH, K., ... FURGONI, R., (2016), The new world atlas of artificial night sky brightness, *Science Advances*, vol. 2, n°6, p. e1600377.
- FLAHAUT, B., (2004), Impact of infrastructure and local environment on road unsafety: Logistic modeling with spatial autocorrelation, *Accident Analysis & Prevention*, vol. 36, n°6, p. 1055-1066.
- FOTIOS, S. A., ROBBINS, C. J. & FARRALL, S., (2021), The Effect of Lighting on Crime Counts, *Energies*, vol. 14, n°14, p. 4099.
- FRANCHOMME, M., HINNEWINKEL, C. & CHALLÉAT, S., (2019a), La trame noire, un indicateur de la place de la nature dans l'aménagement du territoire, *Bulletin de l'association de géographes français. Géographies*, vol. 96, n°2, p. 161-180.
- FRANCHOMME, M., HINNEWINKEL, C. & CHALLÉAT, S., (2019b), La trame noire, un indicateur de la place de la nature dans l'aménagement du territoire, *Bulletin de l'association de géographes français. Géographies*, vol. 96, n°2, p. 161-180.
- GEURTS, P., ERNST, D. & WEHENKEL, L., (2006), Extremely randomized trees, *Machine Learning*, vol. 63, n°1, p. 3-42.
- GREEN, J., PERKINS, C., STEINBACH, R. & EDWARDS, P., (2015), Reduced street lighting at night and health: A rapid appraisal of public views in England and Wales, *Health & Place*, vol. 34, p. 171-180.
- GROSS, F., (2013), Case-control analysis in highway safety: Accounting for sites with multiple crashes, *Accident Analysis & Prevention*, vol. 61, p. 87-96.
- GROSS, F. & DONNELL, E. T., (2011), Case-control and cross-sectional methods for estimating crash modification factors: Comparisons from roadway lighting and lane and shoulder width safety effect studies, *Journal of Safety Research*, vol. 42, n°2, p. 117-129.
- HAANS, A. & DE KORT, Y. A. W., (2012), Light distribution in dynamic street lighting: Two experimental studies on its effects on perceived safety, prospect, concealment, and escape, *Journal of Environmental Psychology*, vol. 32, n°4, p. 342-352.

HACAR, M. & GÖKGÖZ, T., (2019), A New, Score-Based Multi-Stage Matching Approach for Road Network Conflation in Different Road Patterns, *ISPRS International Journal of Geo-Information*, vol. 8, n°2, p. 81.

HUNTER, T. B., CRAWFORD, D. L., HOWELL, L. & KNAUSS, D. G., (1991), The International Dark-Sky Association, Inc., *International Astronomical Union Colloquium*, vol. 112, p. 110-110.

INGI, D., BHUSAL, P., KYTTÄ, M. & PINHO, P., (2025), Public participation geographic information system as a tool to study the relationship between reassurance and outdoor lighting, *Lighting Research & Technology*, p. 14771535251324948.

INNES, M., (2017), From fear to understanding: 'Making' and managing public reactions to crime, disorder and policing, In : *Handbook of Crime Prevention and Community Safety*, Routledge2^e éd.

JEAN, S., (2024), *Modéliser la crise de la biodiversité : les rôles de l'espace et des comportements stratégiques dans la modélisation bioéconomique* (Theses, Université Paris-Saclay).

JOHANSSON, Ö., WANVIK, P. O. & ELVIK, R., (2009a), A new method for assessing the risk of accident associated with darkness, *Accident Analysis & Prevention*, vol. 41, n°4, p. 809-815.

JOHANSSON, Ö., WANVIK, P. O. & ELVIK, R., (2009b), A new method for assessing the risk of accident associated with darkness, *Accident Analysis & Prevention*, vol. 41, n°4, p. 809-815.

JR, D. W. H., LEMESHOW, S. & STURDIVANT, R. X., (2013), *Applied Logistic Regression*, John Wiley & Sons, 528 p.

KANG, M., MOUDON, A. V., KIM, H. & BOYLE, L. N., (2019), Intersections and Non-Intersections: A Protocol for Identifying Pedestrian Crash Risk Locations in GIS, *International Journal of Environmental Research and Public Health*, vol. 16, n°19, p. 3565.

KELER, A., MAEDA, D., NAKAO, S., YASUDA, K. & SCHMÖCKER, J.-D., (2025), Analysis of traffic crashes considering the field of view, *IATSS Research*, vol. 49, n°3, p. 305-313.

KHATTAK, M. W., DE BACKER, H., DE WINNE, P., BRIJS, T. & PIRDAVANI, A., (2024), Analysis of Road Infrastructure and Traffic Factors Influencing Crash Frequency: Insights from Generalised Poisson Models, *Infrastructures*, vol. 9, n°3, p. 47.

LEUNG, Y.-F., SPENCELEY, A., HVENEGAARD, G. & BUCKLEY, R. (ÉD.), (2018), *Tourism and visitor management in protected areas : guidelines for sustainability*, IUCN, International Union for Conservation of Nature.

LI, Q., WANG, Z., LI, M., YANG, R., LIN, P.-S. & LI, X., (2021), Development of crash modification factors for roadway illuminance: A matched case-control study, *Accident Analysis & Prevention*, vol. 159, p. 106279.

LONGCORE, T. & RICH, C., (2004), Ecological light pollution, *Frontiers in Ecology and the Environment*, vol. 2, n°4, p. 191-198.

MARCHANT, P. R., (2004), A Demonstration That the Claim That Brighter Lighting Reduces Crime Is Unfounded, *The British Journal of Criminology*, vol. 44, n°3, p. 441-447.

MCMAHON, D. M., (2018), Illuminating the Enlightenment: Public Lighting Practices in the Siècle Des Lumières*, *Past & Present*, vol. 240, n°1, p. 119-159.

- MEESMANN, U. ., TORFS, K. ., NGUYEN, H. ., & VAN DEN BERGHE, W., (2018), Do we care about road safety? Key findings from the ESRA1 project in 38 countries. ESRA project (E-Survey of Road users' Attitudes). Brussels, Belgium: Vias institute.
- MEIER, J., HASENÖHRL, U., KRAUSE, K., POTTHARST, M., & DEUTSCHLAND (ÉD.), (2015), *Urban lighting, light pollution and society*, Routledge, Taylor and Francis group, New York, NY London, 1 p.
- MINGARRO, M. & LOBO, J. M., (2023), European National Parks protect their surroundings but not everywhere: A study using land use/land cover dynamics derived from CORINE Land Cover data, *Land Use Policy*, vol. 124, p. 106434.
- MORIN, J.-F. B., DENIS, (2005), Standard and Nonstandard Roadway Lighting Compared with Darkness at Rural Intersections - Jean-François Bruneau, Denis Morin, 2005, *Transportation Research Record*.
- MUELLER, J. E., (1968), An Introduction to the Hydraulic and Topographic Sinuosity Indexes 1, *Annals of the Association of American Geographers*, vol. 58, n°2, p. 371-385.
- MURRAY, A. T. & FENG, X., (2016), Public street lighting service standard assessment and achievement, *Socio-Economic Planning Sciences*, vol. 53, p. 14-22.
- OMS, (2023), Rapport de situation sur la sécurité routière dans le monde - 2023 - Organisation mondiale de la Santé.
- PAIN, R., MACFARLANE, R., TURNER, K. & GILL, S., (2006), 'When, Where, if, and but': Qualifying GIS and the Effect of Streetlighting on Crime and Fear, *Environment and Planning A: Economy and Space*, vol. 38, n°11, p. 2055-2074.
- PAPADIMITRIOU, E., FILTNESS, A., THEOFILATOS, A., ZIAKOPOULOS, A., QUIGLEY, C. & YANNIS, G., (2019), Review and ranking of crash risk factors related to the road infrastructure, *Accident Analysis & Prevention*, vol. 125, p. 85-97.
- PEDEN, M. M., (2004), *World Report on Road Traffic Injury Prevention*, World Health Organization, 240 p.
- PEÑA-GARCÍA, A., HURTADO, A. & AGUILAR-LUZÓN, M. C., (2015), Impact of public lighting on pedestrians' perception of safety and well-being, *Safety Science*, vol. 78, p. 142-148.
- PÉREZ VEGA, C., ZIELINSKA-DABKOWSKA, K. M. & HÖLKER, F., (2021), Urban Lighting Research Transdisciplinary Framework—A Collaborative Process with Lighting Professionals, *International Journal of Environmental Research and Public Health*, vol. 18, n°2, p. 624.
- PERKINS, C., STEINBACH, R., TOMPSON, L. A., GREEN, J., JOHNSON, S., GRUNDY, C., ... EDWARDS, P., (2015), *What is the effect of reduced street lighting on crime and road traffic injuries at night? A mixed-methods study* (n° 3:11), Perth, UK, v-107 p.
- PLAINIS, S., MURRAY, I. J. & PALLIKARIS, I. G., (2006), Road traffic casualties: understanding the night-time death toll, *Injury Prevention*, vol. 12, n°2, p. 125-138.
- RAKOTOMALALA, R., (2015), *Pratique de la Régression Logistique*.
- RANAPURWALA, S. I., MELLO, E. R. & RAMIREZ, M. R., (2016), A GIS-based Matched Case-control Study of Road Characteristics in Farm Vehicle Crashes, *Epidemiology*, vol. 27, n°6, p. 827-834.
- RETTING, R. A., FERGUSON, S. A. & MCCARTT, A. T., (2003), A Review of Evidence-Based Traffic Engineering Measures Designed to Reduce Pedestrian-Motor Vehicle Crashes, *American Journal of Public Health*, vol. 93, n°9, p. 1456-1463.

SALATHÉ, M., MONDÉSIR, H. & EKER, F., (2024), Observatoire des vitesses (France métropolitaine) Résultats de l'année 2024.

SCHLÖGL, M., (2020a), A multivariate analysis of environmental effects on road accident occurrence using a balanced bagging approach, *Accident Analysis & Prevention*, vol. 136, p. 105398.

SCHLÖGL, M., (2020b), A multivariate analysis of environmental effects on road accident occurrence using a balanced bagging approach, *Accident Analysis & Prevention*, vol. 136, p. 105398.

SORDELLO, R., JUPILLE, O., DEUTSCH, É., VAUCLAIR, S., SALMON-LEGAGNEUR, L. & FAURE, J.-B., (2018), Trame noire : un sujet qui « monte » dans les territoires, *Sciences Eaux & Territoires*, vol. 25, n°1, p. 78-85.

SORDELLO, R., PAQUIER, F. & DALOZ, A., (2021), *Trame noire - Méthodes d'élaboration et outils pour sa mise en œuvre*, Office français de la biodiversité, 116 pages p.

SPW, (2013), Indicateur de ruralité Classification DGO3 des communes rurales, semi-rurales et non rurales.

SPW, (2021), Indicateur de ruralité - Degré de ruralité.

SPW, (2022), Les parcs nationaux.

SPW, (2025), Réseau routier.

STATBEL, (2025), Accidents de la circulation.

STEINBACH, R., PERKINS, C., TOMPSON, L., JOHNSON, S., ARMSTRONG, B., GREEN, J., ... EDWARDS, P., (2015), The effect of reduced street lighting on road casualties and crime in England and Wales: controlled interrupted time series analysis, *J Epidemiol Community Health*, vol. 69, n°11, p. 1118-1124.

STRUYF, P., ENHUS, D. E., BAUWENS, D. T. & MELGAÇO, D. L., (2019), Literature study: The effects of reduced public lighting on crime, fear of crime, and road safety.

SUTTON, R. M., (2004), Gender, Socially Desirable Responding and the Fear of Crime: Are Women Really More Anxious about Crime?, *British Journal of Criminology*, vol. 45, n°2, p. 212-224.

TEALDE, E., (2022), The unequal impact of natural light on crime, *Journal of Population Economics*, vol. 35, n°3, p. 893-934.

THEOFILATOS, A. & YANNIS, G., (2014), A review of the effect of traffic and weather characteristics on road safety, *Accident Analysis & Prevention*, vol. 72, p. 244-256.

TOMTOM, (2024), Inside TomTom's Traffic products: Traffic Stats.

TOSSA, P. & SOUQUES, M., (2019), Lumière artificielle nocturne et pollution lumineuse : une synthèse des effets sur les rythmes biologiques chez l'homme, *Environnement, Risques & Santé*, vol. 18, n°6, p. 477-487.

TRUTER, J. P., (2026), Research summary of a GIS-based analysis of the status of streetlighting on the WCG road network: towards a spatial asset repository to guide decision making and asset management, *South Africa*.

ULLAH, I., FARID, A., MUKNAHALLIPATNA, S. & KSAIBATI, K., (2025), Intersection sight distance evaluation through a comprehensive review and novel measurement framework, *Discover Civil Engineering*, vol. 2, n°1, p. 199.

- UVCW, (2022), L'éclairage public-Union des Villes et Communes de Wallonie.
- VANDENBULCKE, G., THOMAS, I. & INT PANIS, L., (2014), Predicting cycling accident risk in Brussels: A spatial case-control approach, *Accident Analysis & Prevention*, vol. 62, p. 341-357.
- WILLS, W., (2019), A Simple Sinuosity-Based Method using GPS data to Support Mitigation Policies for Public Buses GHG Emissions, *ArXiv*.
- WU, K.-F. (KEN) & LIN, T., (2022), Investigating the effects of travel lane configuration and lane width on traffic safety where powered-two-wheelers (PTWs) share with larger vehicles: A micro perspective, *Accident Analysis & Prevention*, vol. 172, p. 106682.
- XIN, C., WANG, Z., LEE, C., LIN, P.-S., CHEN, T., GUO, R. & LU, Q., (2019), Development of crash modification factors of horizontal curve design features for single-motorcycle crashes on rural two-lane highways: A matched case-control study, *Accident Analysis & Prevention*, vol. 123, p. 51-59.
- YAMARTHI, D. P., RAMAN, H. & PARVIN, S., (2025, avril 28), United States Road Accident Prediction using Random Forest Predictor, *arXiv*.

8. Liste des figures

Figure 1 : Localisation du Parc national, des communes participantes et des communes limitrophes -	10
Figure 2 : Réseau routier de la zone d'étude avec le réseau structurant (N5 et N40) et le réseau de nationales selon le SPW -----	11
Figure 3 : Résumé de la méthode -----	15
Figure 4 : Distribution des tronçons par volume de véhicules. Les données proviennent de TomTom et couvrent le mois d'août 2024. -----	18
Figure 5 : Classification du réseau routier basée sur le volume de véhicules par tronçon. Source des données : TomTom -----	19
Figure 6 : Illustration du calcul de la sinuosité d'un tronçon (l). Image tirée de l'article Ranapurwala, S. I., Mello, E. R. & Ramirez, M. R., (2016), A GIS-based Matched Case-control Study of Road Characteristics in Farm Vehicle Crashes, Epidemiology, vol. 27, n°6, p. 827-834. -----	20
Figure 7 : Distribution des distances entre les accidents catégorisés par la police comme ayant eu lieu à proximité d'un éclairage et le poteau d'éclairage le plus proche -----	21
Figure 8 : Distribution des distances entre les accidents catégorisés par la police comme ayant eu lieu à proximité d'une intersection et l'intersection la plus proche -----	22
Figure 9 : Localisation des accidents de nuit et des secteurs statistiques de Statbel montrés avec leur densité de population (hab/km ²). Fonds de carte : orthophotos 2024 (SPW). -----	23
Figure 10 : Boxplot de la densité des secteurs statistiques où les accidents ont été catégorisés comme ayant eu lieu en agglomération (en zone bâtie) ou en hors agglomération (en zone non-bâtie) selon la police -----	24
Figure 11 : Distribution des densités des secteurs statistiques selon le fait que l'accident ait été catégorisé comme ayant eu lieu ou non en agglomération -----	24
Figure 12 : Heatmap de la répartition temporelle des accidents de la zone d'étude entre 2017 et 2024. Le graphique de gauche (A) reprend l'ensemble des accidents tandis que le graphique de droite (B) ne reprend que les accidents catégorisés comme ayant eu lieu de nuit. -----	29
Figure 13 : Localisation des zones Natura2000 et des poteaux d'éclairage public -----	30
Figure 14 : Importance relative des variables avec le modèle Random Forest général. -----	34
Figure 15 : Importance relative de chaque variable à chaque itération pour le modèle Random Forest général -----	35
Figure 16 : Partial dependence plot : effet de chaque variable sur le risque d'accident (yhat) selon le modèle Random Forest général -----	36
Figure 17 : Importance relative (%IncMSE) des variables pour les modèles Random Forest (RF) de classe de route 3, 2 et 1 séparés et pour le modèle Random Forest (RF) des classes de route 1 et 2 ensemble. -----	37
Figure 18 : Carte du risque prédit par le modèle de régression logistique pour les trois classes de route (RL). La classification du réseau routier est représentée par la taille des points. -----	38
Figure 19 : Carte du risque prédit par le modèle Random Forest (RF). La classification du réseau routier est représentée par la taille des points -----	39
Figure 20 : Pourcentage de lampadaires potentiellement inutiles par classe de route et selon différents seuils de risque d'accident prédits par le modèle Random Forest -----	42
Figure 21 : Localisation des lampadaires potentiellement inutiles selon différents seuils de risque prédits par le modèle Random Forest. Les lampadaires situés sur une route de classe 0 (<2 véhicules par jour) sont représentés en gris -----	42
Figure A. 1 : Indicateur de ruralité de la Wallonie (SPW, 2021)	60

Figure A. 2 : Exemple d'un point où l'intersection la plus proche n'est pas celle qui est accessible	60
Figure A. 3 : Exemple de plusieurs accidents localisés dans une même zone bâtie mais dont deux accidents ont été mal catégorisés. Fonds de carte : orthophotos 2024 (SPW).....	61
Figure A. 4 : Périmètre du Parc national et localisation des poteaux d'éclairage public	63
Figure A. 5 : Couverture du sol (Corine Land Cover 2018) et localisation des poteaux d'éclairage	64
Figure A. 6 : Localisation des réserves naturelles agréées, des réserves forestières, des zones humides d'intérêt biologique, des réserves domaniales et des poteaux d'éclairage public.....	64
Figure A. 7 : Pourcentage de lampadaires potentiellement inutiles par classe de route et selon différents seuils de risque d'accident prédits par le modèle de régression logistique.	65
Figure A. 8 : Localisation des lampadaires potentiellement inutiles selon différents seuils de risque prédits par le modèle de régression logistique. Les lampadaires situés sur une route de classe 0 (<2 véhicules par jour) sont représentés en gris	65

9. Liste des tables

Table 1 : Répartition des 800 points de contrôle entre les 3 classes de route	17
Table 2 : Classification du réseau routier et caractéristique des classes de route ainsi créées	18
Table 3 : Résumé des variables utilisées dans ce mémoire	25
Table 4 : Résultats du modèle de régression logistique général réalisé sur les 1200 cas et points de contrôle. L'AIC de ce modèle est 1254,7. (« * » : <0,05, « ** » : <0,01, « *** » : 0)	31
Table 5 : Résultats des deux modèles de régression logistique pour la classe de route 3 (soit 469 cas et points de contrôle). Le modèle avec la variable zone bâtie est présenté à gauche et le modèle sans la variable zone bâtie est présenté à droite. (« * » : <0,05, « ** » : <0,01, « *** » : 0)	32
Table 6 : Résultats des modèles de régression logistique pour les classes de route 2 et 1 séparées et ensemble. (« * » : <0,05, « ** » : <0,01, « *** » : 0)	33
Table 7 : Résultats du modèle de régression logistique présentant la variable proximité à un éclairage public comme variable dépendante. Modèle réalisé avec l'ensemble des données soit 1200 cas et points de contrôle. (« * » : <0,05, « ** » : <0,01, « *** » : 0)	33
Table 8 : Importance relative des variables avec le modèle Random Forest général. Modèle réalisé avec l'ensemble des données soit 1200 cas et points de contrôle	34
Table 9 : Risque prédit moyen, minimum et maximal par le modèle de régression logistique (RL) pour chaque classe de route ainsi que l'écart-type	38
Table 10 : Risque prédit moyen, minimum et maximal par le modèle Random Forest (RF) pour chaque classe de route	39
Table 11 : Différences du risque d'accident prédit entre le modèle de régression logistique et le modèle Random Forest et résultats des t-tests testant la significativité de la différence entre les deux modèles.	40
Table 12 : Nombre de lampadaires actuels par classe de route	41
Table 13: Nombre et pourcentage de lampadaires potentiellement inutiles en ce qui concerne la sécurité routière pour chaque classe de route et selon différents seuils de risque d'accident prédit par le modèle Random Forest	41
Table 14 : Total du nombre de lampadaires potentiellement inutiles par seuil de risque d'accident prédit par le modèle Random Forest	41
 Table A. 1 : Valeur de VIF pour chaque variable étudiée et pour chaque modèle de régression logistique	62
Table A. 2 : Résultats du modèle de régression logistique testant l'interaction de la proximité à une intersection et de la proximité à un éclairage public réalisé sur les 1200 cas et points de contrôle. (« * » : <0,05, « ** » : <0,01, « *** » : 0)	63

10. Annexes

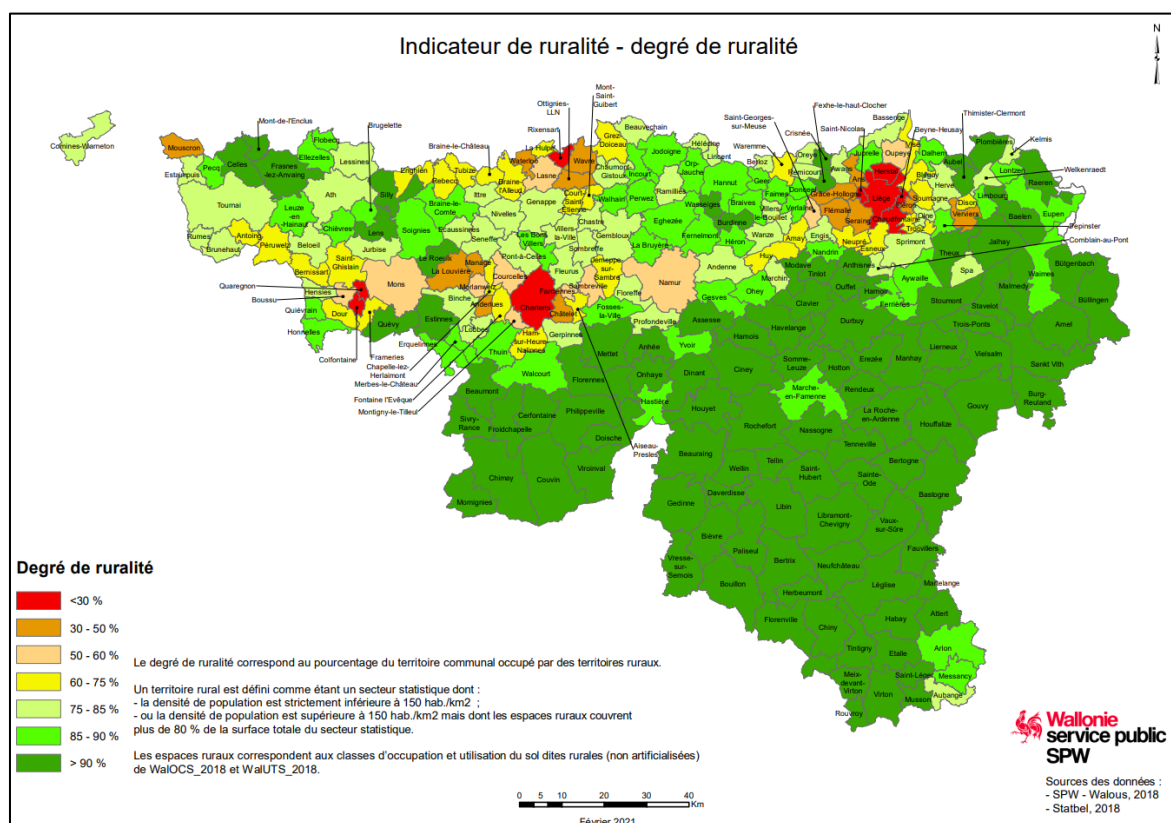


Figure A. 1 : Indicateur de ruralité de la Wallonie (SPW, 2021)



Figure A. 2 : Exemple d'un point où l'intersection la plus proche n'est pas celle qui est accessible

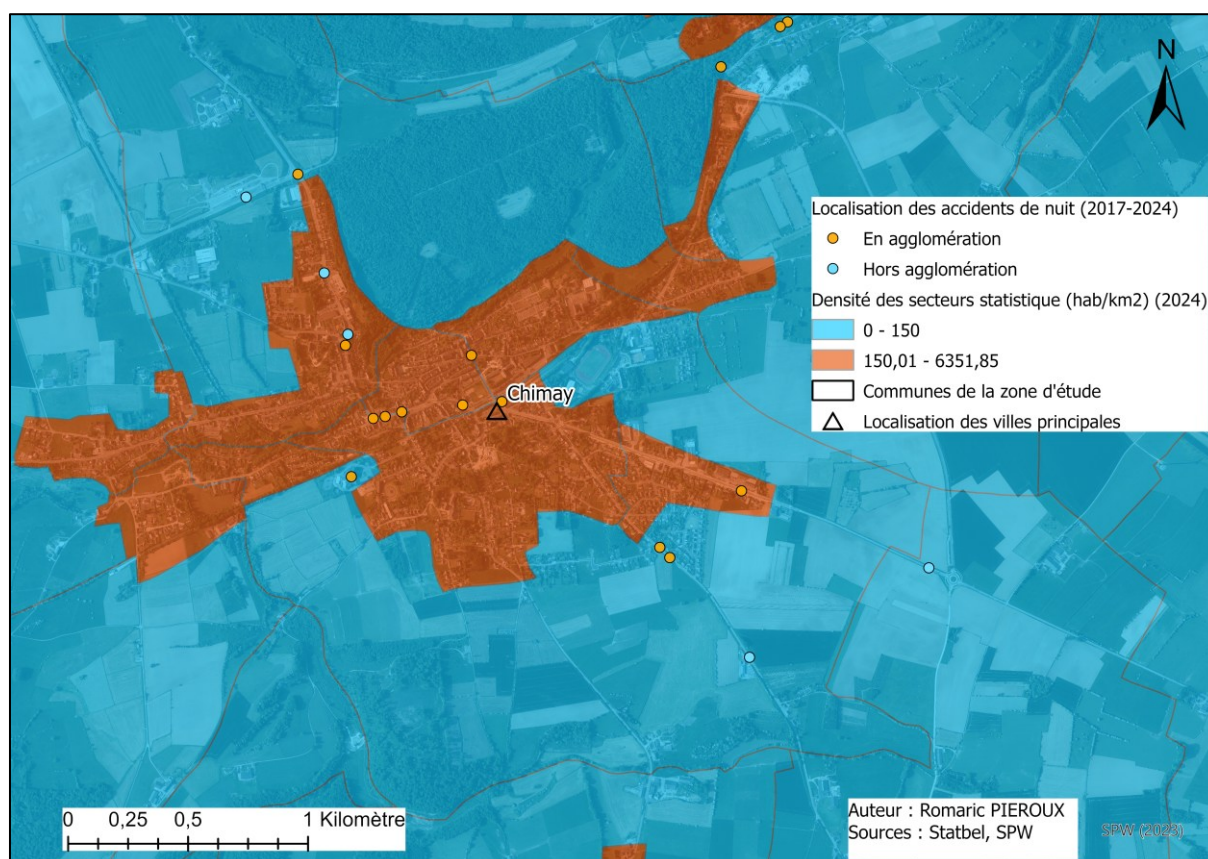


Figure A. 3 : Exemple de plusieurs accidents localisés dans une même zone bâtie mais dont deux accidents ont été mal catégorisés. Fonds de carte : orthophotos 2024 (SPW).

Table A. 1 : Valeur de VIF pour chaque variable étudiée et pour chaque modèle de régression logistique

VIF de chaque variable pour chaque modèle	Zone bâtie	Intersection	Eclairage	Classification du réseau routier	Sinuosité	Largeur	Volume de véhicules de nuit	Vitesse moyenne de nuit	Présence d'un accident
Modèle général	1,20	1,29	1,23	1,93	1,02	1,3	2,89	1,83	/
Modèle testant l'interactio n (Intersecti on : Eclairage)	1,21	2,31	1,33	1,93	1,016	1,32	2,91	1,84	/
Modèle classe de route 3 (avec zone bâtie)	1,21	1,25	1	/	1,038	1,58	1,27	1,26	/
Modèle classe de route 3 (sans zone bâtie)	/	1,23	1	/	1,030	1,55	1,15	1,18	/
Modèle classe de route 2	1,58	1,75	1,63	/	1,04	1,75	1,087	2,33	/
Modèle classe de route 1	1,72	1,90	1,56	/	1,025	1,72	1,24	2,01	/
Modèle classe de route 1 et 2	1,56	1,75	1,44	/	1,028	1,69	1,41	2,26	/
Modèle avec l'éclairage comme variable dépendant e	1,09	1,26	/	2,01	1,015	1,23	2,93	2,23	1,13

Table A. 2 : Résultats du modèle de régression logistique testant l'interaction de la proximité à une intersection et de la proximité à un éclairage public réalisé sur les 1200 cas et points contrôle. (« * » : <0,05, « ** » : <0,01, « *** » : 0)

Variable	Paramètres estimés	Odd-ratio	P-valeur
(Intercept)	-0,98	0,38	0,29
Zone bâtie	0,54	1,72	0,0034**
Proximité à une intersection (0_1) (40 mètres)	0,86	2,37	0,017*
Proximité à un éclairage public (0_1) (30 mètres)	0,55	1,73	0,0044**
Classe de route : 2	-1,081	0,34	8,41E-08***
Classe de route : 3	-3,49	0,031	3,51E-08***
Sinuosité	-0,49	0,61	0,52
Largeur	0,025	1,025	0,0019**
Volume de véhicule de nuit	2,31E-05	1	0,68
Vitesse moyenne de nuit	0,016	1,016	0,0046**
Proximité à une intersection : Proximité à un éclairage public	-0,55	0,58	0,18

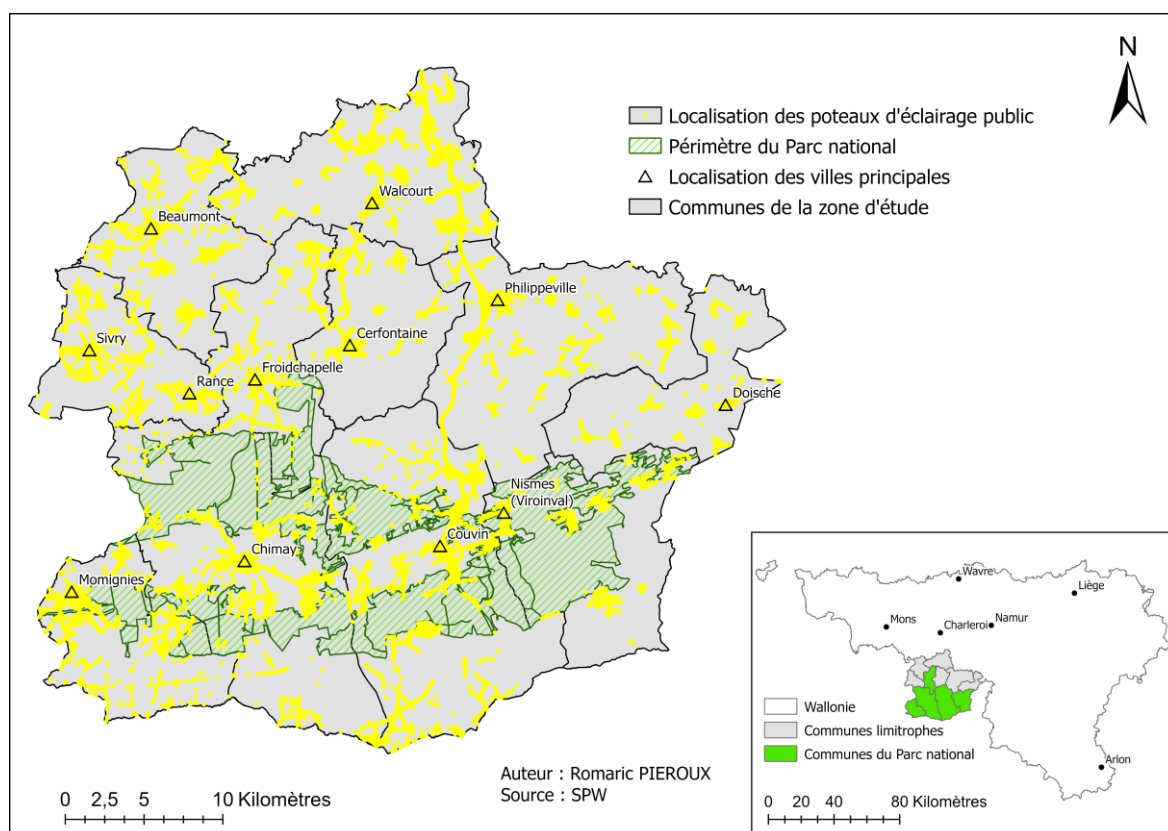


Figure A. 4 : Périmètre du Parc national et localisation des poteaux d'éclairage public

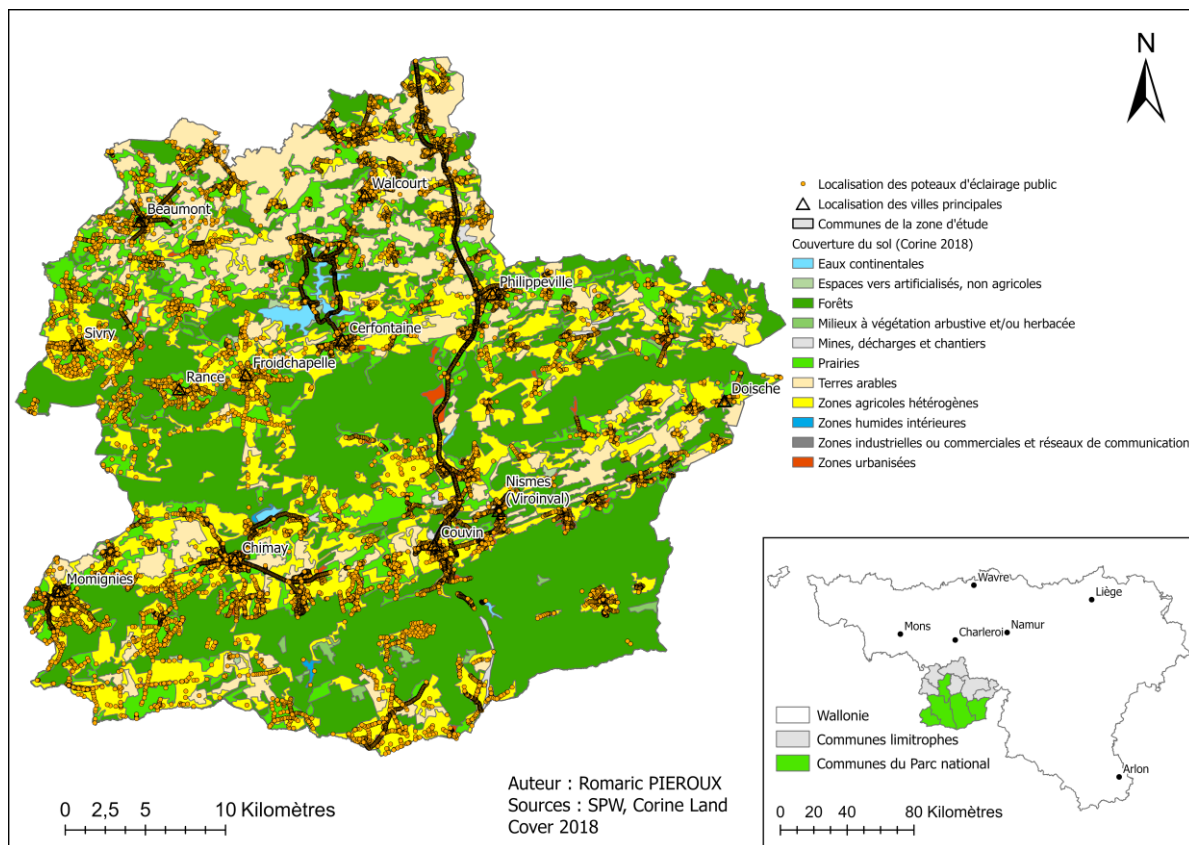


Figure A. 5 : Couverture du sol (Corine Land Cover 2018) et localisation des poteaux d'éclairage

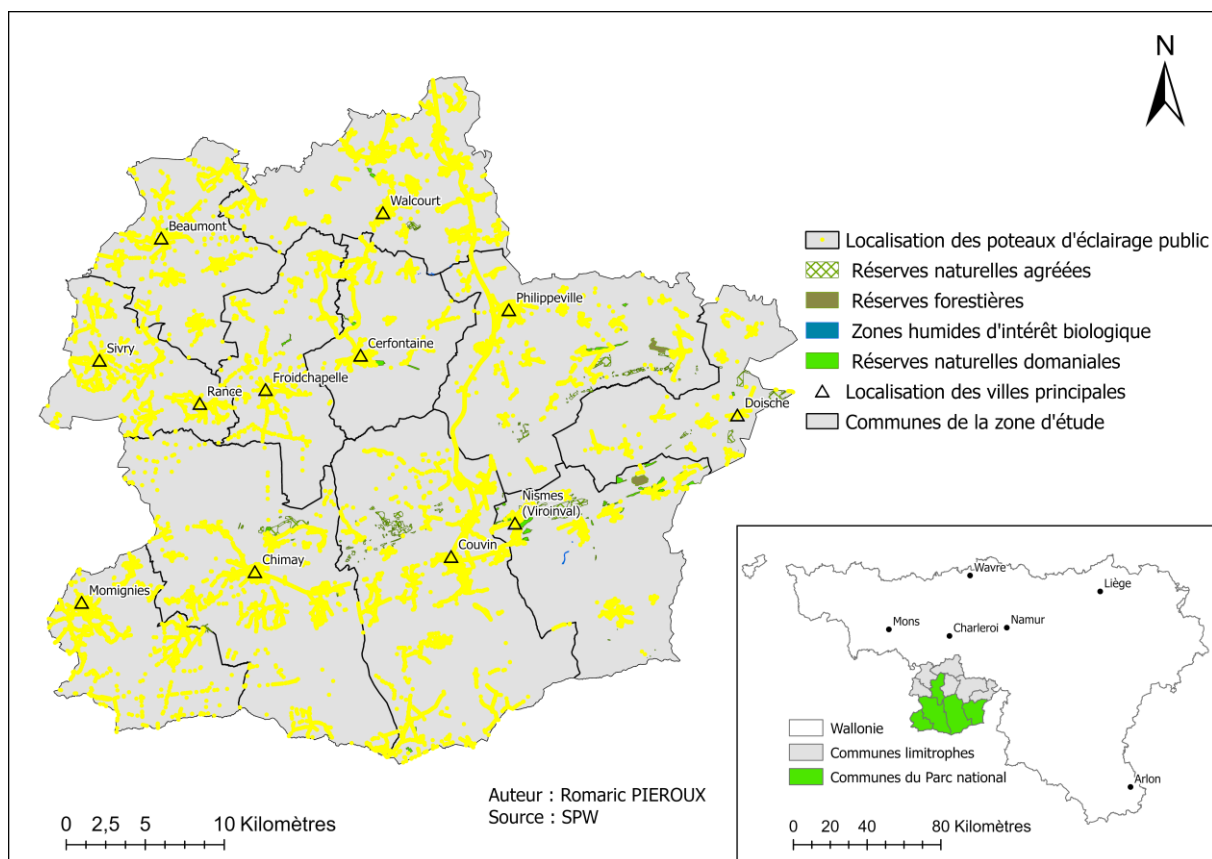


Figure A. 6 : Localisation des réserves naturelles agréées, des réserves forestières, des zones humides d'intérêt biologique, des réserves domaniales et des poteaux d'éclairage public

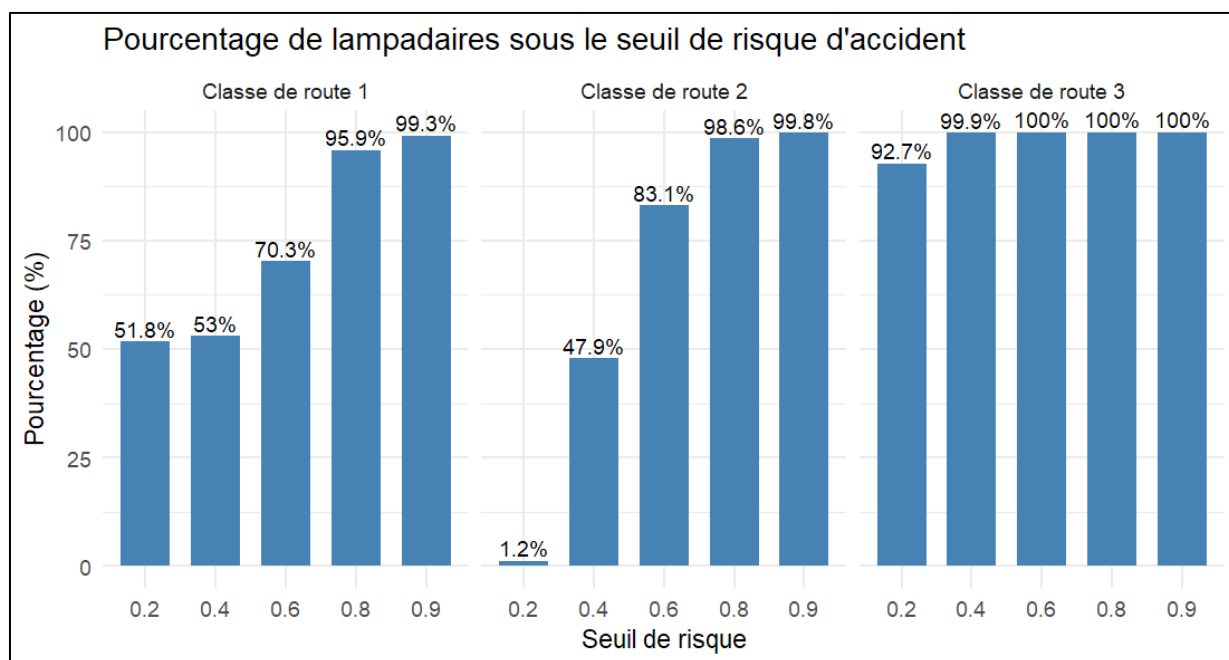


Figure A. 7 : Pourcentage de lampadaires potentiellement inutiles par classe de route et selon différents seuils de risque d'accident prédits par le modèle de régression logistique.

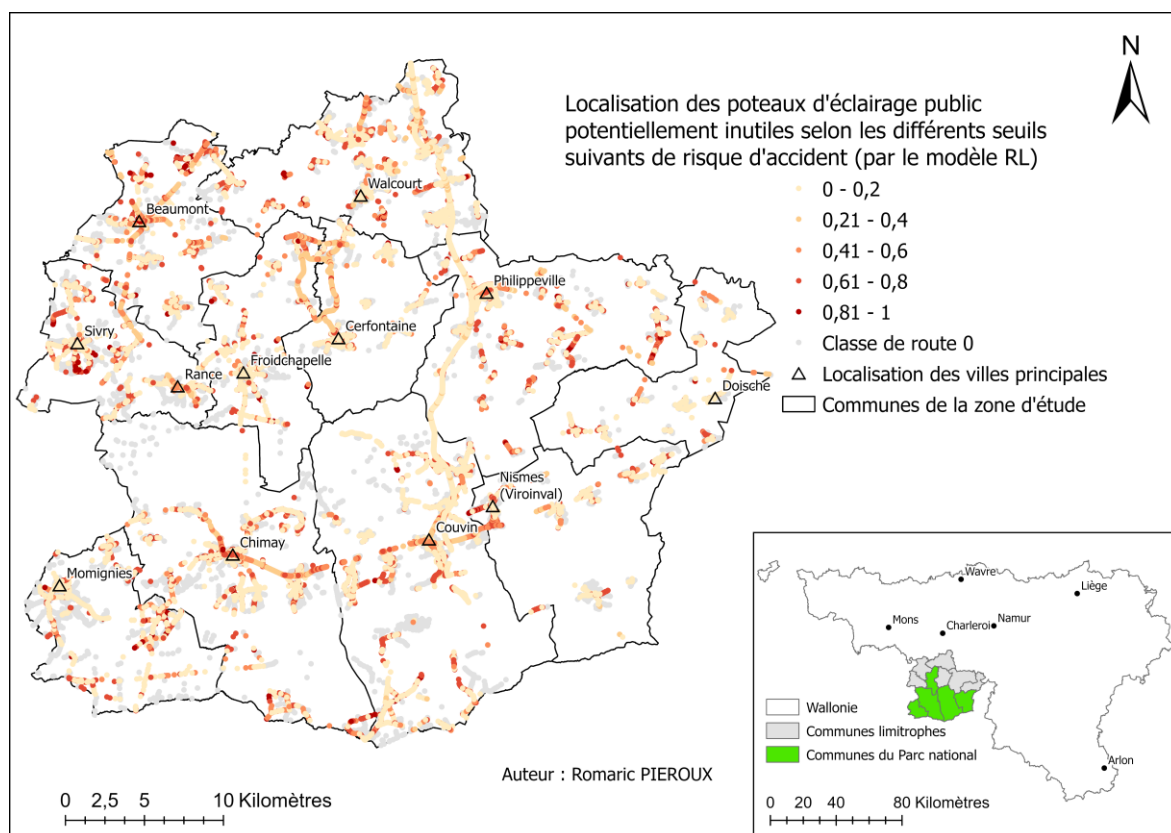


Figure A. 8 : Localisation des lampadaires potentiellement inutiles selon différents seuils de risque prédits par le modèle de régression logistique. Les lampadaires situés sur une route de classe 0 (<2 véhicules par jour) sont représentés en gris

11. Autobiographie

Coordonnées

Romaric Pieroux

03/08/2003

Chemin-Fosset, 2,

Boussu-en-Fagne, 5660 (Couvin)

E-mail : romaric.pieroux383@gmail.com

+32 470715227

Parcours académique

2024-2026 : Master [120] en sciences géographiques, orientation générale à finalité approfondie, à l'Université Catholique de Louvain (UCLouvain).

2021-2024 : Bachelier en sciences géographiques orientation générale, à l'Université Catholique de Louvain (UCLouvain).

2015-2021 : Études secondaires à l'Institut Sainte-Marie (ISM) de Pesche (Couvin)

2009 – 2015 : Primaires à l'école Saint-Joseph de Couvin.

