



## Livrable L2

Analyse des comportements des 2RM et 4RM en lien avec la vitesse à partir des données issues du projet DYMOA

N° Livrable	L2	N° workpackage	WP2
<b>Statut</b>		Version remise à la DSR	
<b>Date</b>		22/07/2022	
<b>Responsable du document</b>		<b>Vincent Ledoux, Cerema Territoires et villes</b> Vincent.ledoux@cerema.fr	
<b>Auteur principal</b>		Vincent Ledoux (Cerema)	
<b>Contributeur(s)</b>		Henri Moulga, Guillaume Saint Pierre (Cerema). Lionnel Moussu (Logiroad)	
<b>Validation</b>			
<b>Enregistrement</b>		DYMOA+_WP2_L2.docx	



## Résumé

Le projet DYMOA a permis d'enregistrer en continu et en condition de conduite naturelle les vitesses de conducteurs qu'ils soient au volant de leur deux-roues motorisé (2RM) ou de leur véhicule léger (4RM). L'objectif du travail présenté dans ce livrable est, d'une part de déterminer s'il existe des différences entre les vitesses pratiquées par les 2RM et les 4RM et, d'autre part d'identifier les caractéristiques de l'infrastructure qui influent sur le non-respect des vitesses maximales autorisées (VMA).

### Méthode

Les vitesses ont été recueillies à l'aide d'enregistreurs de données embarqués sur une flotte de 2RM (26) et de 4RM (6) appartenant à des particuliers volontaires pour participer à l'expérimentation.

Les vitesses géolocalisées des conducteurs étaient relevées toutes les secondes. Afin de mieux les contextualiser, elles font l'objet d'un ensemble de traitements (mapmatching) et sont enrichies par croisement avec des bases de données routière : BD TOPO® de l'IGN (France entière) et BD Navstreet® de Here (limitées à la Normandie).

Les données sont ensuite analysées pour déterminer si les conducteurs pratiquent des vitesses différentes selon qu'ils circulent en 2RM ou en 4RM sur des tronçons routiers identiques. Des régressions logistiques à effets mixtes (GLMM) sont également mises en œuvre pour déterminer quels facteurs de l'infrastructure routière influent sur le dépassement de la vitesse maximale autorisée (VMA). Ces dernières analyses portent uniquement sur la Normandie et concernent 9 conducteurs 2RM dont 3 circulent également en 4RM.

### Résultats

L'analyse comparative des vitesses pratiquées, selon le type de véhicule, sur les sections routières circulées au moins une fois par un même conducteur indiquent que sur 82% des tronçons concernés les conducteurs se déplacent à des vitesses plus élevées en 2RM qu'en 4RM. Sur plus de 20% des tronçons la différence de vitesse entre 2RM et 4RM est supérieure à 20 km/h.

L'examen des données recueillies en Normandie indique également que les conducteurs de 2RM circulent plus longtemps en excès de vitesse de vitesse que ceux de 4RM (34% du temps vs. 20%) et avec des excès de vitesses plus importants.

Les résultats des GLMM confirment la plus grande propension des 2RM à être en excès de vitesse sur un tronçon routier donné par rapport aux 4RM (RR=1,6). Ces régressions indiquent également que les excès de vitesses sont d'autant plus fréquents que le réseau circulé est structurant et que les VMA sont moins élevées. Elles apportent également des connaissances sur certains éléments routiers qui influencent ou non, et parfois de manière différenciée entre 2RM et 4RM, les excès de vitesse.

### Conclusions

Le recueil de données embarquées mis en œuvre dans le projet DYMOA permet de mieux appréhender le comportement des conducteurs vis-à-vis de la vitesse au regard des caractéristiques des routes qu'ils empruntent. Cela nécessite toutefois de recourir à des traitements géomatiques et de disposer de bases de données routières pour enrichir les données brutes recueillies afin de les contextualiser.

Les résultats tendent à montrer que les conducteurs adoptent des vitesses et des comportements infractionnistes différents selon le type de véhicule qu'ils conduisent. Ces comportements sont toutefois modulés par de nombreux éléments de l'infrastructure.

## Table des matières

---

<b>1. INTRODUCTION .....</b>	<b>5</b>
<b>2. MÉTHODE .....</b>	<b>6</b>
2.1. RECUEIL DES DONNÉES.....	6
2.2. STATISTIQUES DESCRIPTIVES GLOBALES.....	6
2.3. MAP-MATCHING.....	7
2.3.1. <i>Principe général</i> .....	7
2.3.2. <i>Mise en pratique</i> .....	8
2.3.3. <i>Résultats</i> .....	10
<b>3. EXPLOITATIONS DES DONNÉES .....</b>	<b>12</b>
3.1. COMPARAISON 2RM / 4RM POUR UN MÊME CONDUCTEUR ET SUR UN MÊME ITINÉRAIRE .....	12
3.2. COMPARAISON 2RM / 4RM SUR LES TRONÇONS CIRCULÉS PAR UN MÊME CONDUCTEUR .....	13
3.3. DÉPASSEMENT DE LA VMA .....	15
3.4. FACTEURS FAVORISANT LE DÉPASSEMENT DE LA VMA.....	17
3.4.1. <i>Construction du jeu de données</i> .....	17
3.4.2. <i>Choix du modèle et méthodologie d'analyse</i> .....	18
3.4.3. <i>Influence du type de véhicule</i> .....	19
3.4.4. <i>Variables ayant une influence sur le dépassement de VMA</i> .....	19
<b>4. CONCLUSIONS.....</b>	<b>23</b>
<b>5. RÉFÉRENCES.....</b>	<b>24</b>
<b>6. ANNEXE .....</b>	<b>25</b>
6.1. RÉGRESSION LOGISTIQUE.....	25
6.2. MODÈLES MIXTES .....	26

## Financement

---

Les projets DYMOA et DYMOA+ ont bénéficié respectivement du soutien financier la Fondation Sécurité Routière et de la Délégation à la sécurité routière.

## 1. Introduction

---

Le projet DYMOA vise à développer de nouvelles méthodes de diagnostic de la sécurité des infrastructures routières et de leurs usages par les deux roues motorisés (2RM) à l'aide d'Enregistreurs de Données Routières (EDR).

La gestion et la maîtrise de la vitesse sont essentielles pour la sécurité routière. La vitesse constitue aussi un paramètre d'analyse de la sécurité routière en raison du lien statistique avéré entre vitesse pratiquée et accidentalité.

Jusqu'à ces dernières années, la détermination des vitesses pratiquées par les conducteurs provenait majoritairement de mesures effectuées avec des dispositifs bord de voie tel qu'une station de recueil de données de trafic ou un analyseur de trafic. Bien que présentant de nombreux avantages (connaissance précise et sans équivoque du positionnement du point de mesure, prise en compte quasi-exhaustive de l'ensemble des véhicules circulant en un point...), cette technique de recueil présente l'inconvénient majeur d'être spatialement très localisée.

Les évolutions technologiques récentes dans le domaine de l'électronique permettent de disposer désormais de dispositifs d'enregistrements embarqués rendant possible, à faible coût, le recueil en continu de la vitesse de déplacement des véhicules associée à leurs positions GPS.

Disposer, en continu, de la vitesse géoréférencée permet notamment de lier finement la vitesse pratiquée et les caractéristiques de l'infrastructure routière et laisse entrevoir la possibilité de mettre en œuvre de nouvelles analyses jusqu'alors difficilement réalisables.

Dans le cadre du projet DYMOA, l'intérêt de mettre en œuvre un observatoire des vitesses à partir de données embarquées avait été investigué en installant des Smartphones (EMMAPHONE) sur une flotte de 2RM et 4RM appartenant à des particuliers volontaires pour participer à l'expérimentation. Il s'agissait d'en étudier la faisabilité juridique et technique et de commencer à réfléchir à la définition d'indicateurs pertinents.

Cela avait conduit à la réalisation d'exploitations statistiques menées uniquement sur les premiers mois de recueil. Ce travail décrit dans le livrable L2.5 « Faisabilité et pertinence d'un observatoire des vitesses » du projet DYMOA (Ledoux et al., 2018) avait permis :

- d'appliquer des techniques de map-matching visant à raccrocher des points GPS à des bases de données géographiques ;
- de présenter, à titre illustratif, quelques exemples d'exploitation de ces données et produire quelques résultats qui restaient à consolider.

Compte-tenu de la faible proportion de données en provenance des 4RM disponibles au moment de son élaboration, le livrable traitait exclusivement des données vitesses des 2RM.

Le présent document s'inscrit pleinement dans la continuité de ce précédent travail :

- en se focalisant tout particulièrement sur les différences observées entre 2RM et 4RM en matière de vitesse pratiquées ;
- en étudiant les facteurs de contexte et d'infrastructure qui influent sur le dépassement de la Vitesse Maximale Autorisée (VMA).

En préambule sont rappelés quelques éléments relatifs aux données recueillies et aux traitements de raccordement géographique (map-matching) déployés pour les besoins de cette recherche.

## 2. Méthode

---

### 2.1. Recueil des données

Afin d'investiguer la faisabilité et la pertinence de construire un observatoire des vitesses basé sur la collecte des vitesses enregistrées par des EDR, les EMMAPHONE utilisés dans le projet ont été conçus de manière à enregistrer les vitesses pratiquées par les conducteurs de 2RM à une fréquence de 1Hz.

En pratique, la vitesse provient du GPS interne des EMMAPHONES (GPS 1Hz). Elle commence à être enregistrée une trentaine de secondes après la mise sous contact du véhicule (temps nécessaire à l'allumage de l'EMMAPHONE et au lancement du logiciel d'acquisition).

Ensuite, toutes les secondes, jusqu'à ce que le contact soit coupé, l'EMMAPHONE stocke dans un fichier spécifique les informations suivantes :

- la position du véhicule : latitude, longitude, altitude ;
- les vitesses du véhicule :
  - vitesse GPS : vitesse brute délivrée par le GPS ;
  - vitesse GPS corrigée : moyenne glissante sur 100 pts de la vitesse brute du GPS ;
  - vitesse bus CAN : vitesse lue sur le bus CAN du véhicule. Cette information n'étant pas disponibles sur les 2RM, elle n'a pas été utilisée.

Le fichier de stockage comprend également des informations permettant de connaître le type de jour (week-end ou semaine), la plage horaire d'enregistrement (ex de 17 à 20 heures), les conditions de luminosité (jour/nuit) et la catégorie de véhicule (2RM/4RM).

Les données sont alors transmises via GSM, de manière sécurisée et cryptée, vers un serveur hébergé par l'IFSTTAR à Salon-de-Provence puis viennent alimenter la base INCIDENTS hébergée au Cerema Sud-Ouest. Elles sont effacées du boîtier dès qu'elles ont été transmises sur le serveur de l'IFSTTAR.

Les données accessibles dans la base INCIDENTS ne sont malheureusement pas horodatées. L'analyse des données du fichier vitesse montrent que certains pas de temps d'acquisition sont manquants sans que cela soit mentionné. Deux lignes de mesure successives peuvent être, dans certains cas, espacées de plusieurs secondes ou minutes (voire plus de manière exceptionnelle).

Pour les besoins des exploitations mises en œuvre et afin de savoir quelles données vitesses étaient associées à un même conducteur (mais pas de qui il s'agissait), ces données sont associées à un identifiant conducteur par une lettre (propre à cette étude et différente de l'identifiant conducteur de la base INCIDENTS). Cette opération ne permet en aucun cas de remonter à l'identité des conducteurs : la table de correspondance entre cet identifiant et l'identité des conducteurs - ou toute information permettant de les identifier - ayant été d'une part définitivement supprimée, et d'autre part jamais en possession des auteurs de ce livrable.

### 2.2. Statistiques descriptives globales

Les données proviennent des 32 véhicules instrumentés dans le projet qu'il s'agisse de 2RM (26) ou 4RM (6). Ces véhicules ont été conduits par 26 conducteurs identifiés par une lettre comprise entre A à Z ; les conducteurs A à F ayant accepté d'instrumenter à la fois un 2RM et un 4RM de leur foyer.

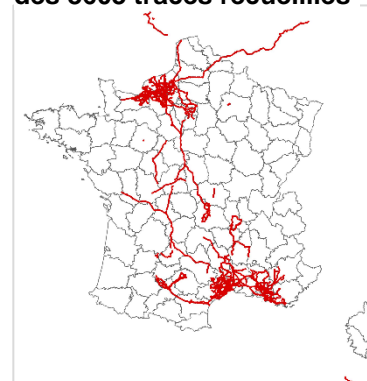
**NOTE :** L'identité du conducteur étant inconnue au moment de l'enregistrement, il est présumé dans le reste du rapport qu'un même et seul conducteur utilise le 2RM ou le 4RM d'un même foyer.

Après une étape de nettoyage de données, la base de données contient 8603 traces constituées de 8,15 millions de points de mesures soit l'équivalent de 94,3 jours d'enregistrements correspondant à une distance parcourue estimée <sup>1</sup> de plus 110 000 km.

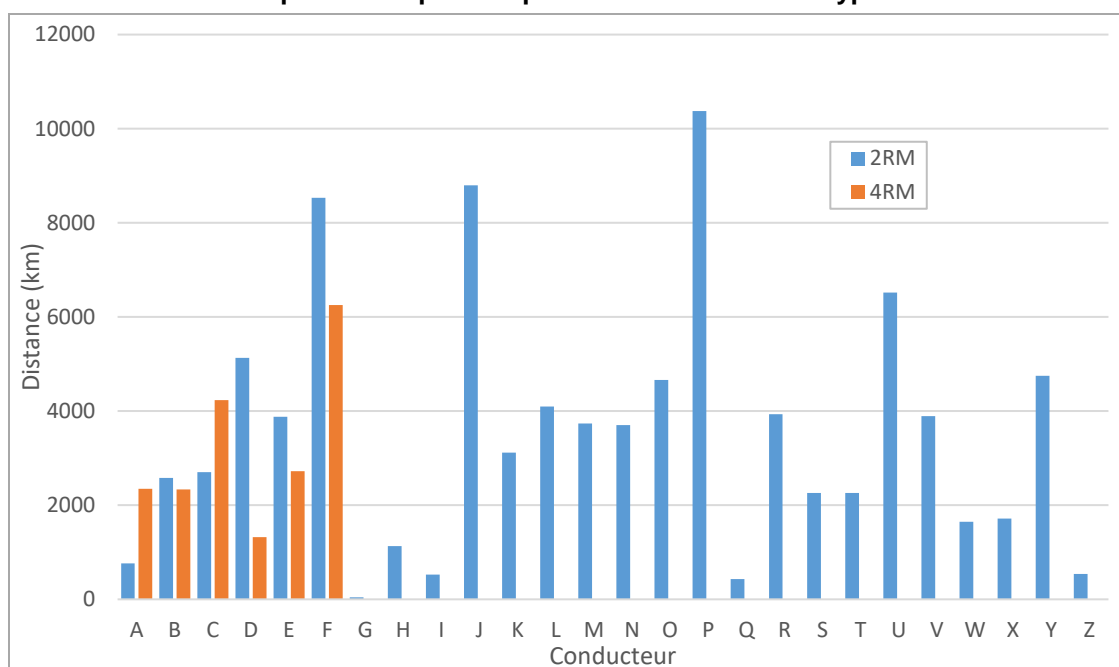
Les traces enregistrées (figure 1) sont majoritairement concentrées sur les régions d'habitation de conducteurs. Quelques traces sont enregistrées hors de France mais elles restent très minoritaires.

Ce kilométrage est imputable à près de 83% à des trajets réalisés en 2RM. Les distances parcourues par les conducteurs de la flotte sont par ailleurs assez inégales (figure 2).

**Figure 1 : Représentation des 8603 traces recueillies**



**Figure 2 : Distance estimée parcourue par chaque conducteur selon le type de véhicule**



## 2.3. Map-Matching

### 2.3.1. Principe général

L'objectif de l'exploitation géographique est de s'appuyer sur des informations externes à la chaîne de recueil des données pour enrichir les données recueillies avec des éléments de contexte, que ce soit sur leur localisation administrative, les réseaux empruntés, le type d'environnement (milieu urbain, espace rural, ...) voire la configuration routière immédiate (intersection, zone de virage, ...). Ces informations complémentaires sont obtenues en rapprochant les données des parcours effectués par les véhicules équipés avec des données géographiques externes, par le rattachement des traces GPS sur des éléments de réseau routier. Cette opération de raccordement est communément désignée par le terme « map-matching ».

Il existe désormais de nombreux algorithmes proposant de réaliser ce traitement des données. Toutefois, pour diverses raisons (écosystème informatique, flexibilité dans le choix du système

<sup>1</sup> Distance estimée en sommant les vitesses relevées par le GPS et multipliées par le temps.

d'information géographique (SIG) utilisé, absence d'horodatage des données...), le choix s'est porté sur développement d'un outil algorithmique spécifique au projet.

Son principe est basé sur l'utilisation d'un algorithme de routage. Ce type d'algorithme est généralement utilisé pour déterminer l'itinéraire routier le plus court (en temps, en distance...) entre 2 points situés sur un réseau routier en tenant compte des contraintes de circulation (sens uniques, double sens...).

Dans le cas présent, il ne s'agit pas de chercher le chemin le plus court mais l'itinéraire le plus vraisemblablement emprunté par le conducteur. Pour cela chaque tronçon routier à proximité de la trace GPS se voit attribuer un coût dépendant d'un d'ensemble de critères géométriques basés sur la proximité géométrique des points (moyenne géométrique) de la trace GPS avec les tronçons routiers du SIG utilisé et la similarité locale de forme (distance de Frechet, angle...) de la trace GPS avec ces tronçons.

La procédure de routage nécessite également de désigner un point de départ et d'arrivée sur le réseau. La détermination de ces points est toutefois loin d'être triviale : arrivée/départ situé sur un espace privé non cartographié, multiplicité des tronçons potentiels à égales distances de ces points...

La réponse apportée à ce problème a consisté à faire du multi-routage en considérant plusieurs points potentiels d'arrivée et de départ et de ne retenir que l'itinéraire se rapprochant le plus de la trace.

Une fois le chemin retenu, les points GPS sont projetés sur les tronçons qui le constituent ce qui permet d'associer à chaque point GPS toutes les informations qui lui sont afférentes.

### 2.3.2. Mise en pratique

Tous les traitements sont réalisés sous le logiciel Postgres (11.4) à l'aide des extensions Postgis (3.0) et pgRouting (3.0).

Deux SIG ont été utilisés selon le périmètre géographique. La BD TOPO® de l'IGN (3.0) pour les exploitations réalisées sur la France entière et la base de données Navtreets® de la société Here pour les exploitations spécifiques sur la région Normandie.

#### a. Préparation des traces

Afin de limiter les problèmes liés aux raccrochements des points d'arrivée et départ, les 50 premiers et derniers mètres de chaque trace sont supprimés.

Dans le cas où des points GPS sont manquants (voir figure 3) et que les distances entre deux points successifs sont supérieures à 300m, les traces sont découpées en sous-traces ; les sous-traces résultantes de longueur inférieure à 300 m étant supprimées.

**Figure 3 : Exemple de création de sous-traces**



Plus de 300 m séparent les points 1626 et 1627 de la trace 922 entraînant la création des sous-traces 922-2 et 922-3

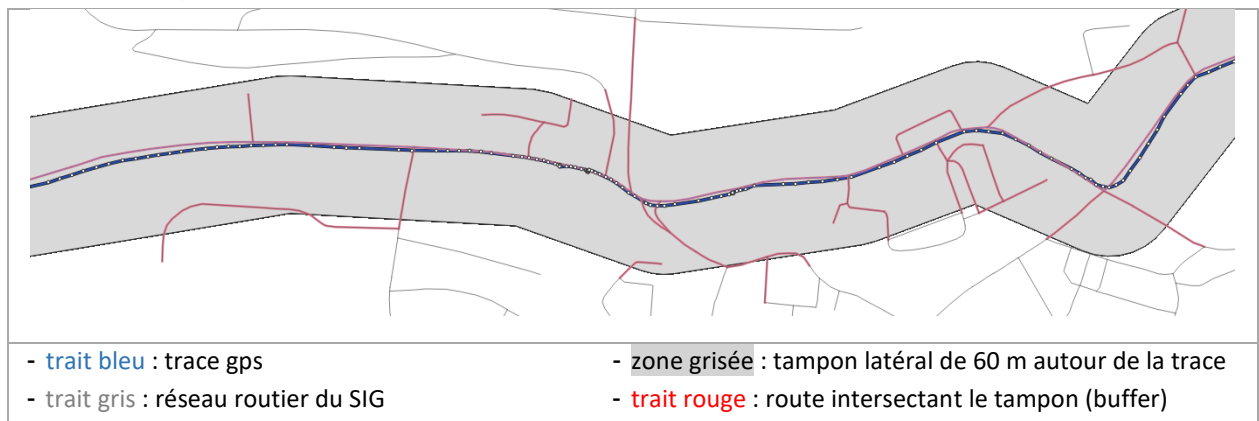
#### b. Sélection des tronçons potentiellement circulés et coûts associés

Dans un premier temps, une sélection est opérée sur le SIG pour ne conserver que les routes et rues sur lesquelles les conducteurs sont autorisés à circuler en fonction des attributs disponibles dans le SIG.



Puis pour chaque trace, ne sont sélectionnées que les routes qui intersectent un tampon latéral de 60m centré sur la trace GPS.

**Figure 4 : Tronçon potentiellement circulées**



À chaque tronçon retenu est associé un coût d'autant plus faible que sa représentation géométrique est proche des points GPS constituant la trace et que sa forme épouse au mieux la forme locale de la trace. Ce coût est calculé pour les deux sens de circulation. Un sens non circulaire (sens unique) se voit attribuer un coût négatif le rendant inempruntable.

**Figure 5 : Coût associé aux tronçons potentiellement circulés**



*Note : le coût de certains tronçons n'apparaît pas pour des questions de mises à l'échelle.*

### c. Réalisation du routage

Le routage est ensuite réalisé en utilisant plusieurs fonctions de pgRouting. Une topologie du réseau est tout d'abord créée. Elle associe chaque extrémité de tronçon à des nœuds *source* et *cible* et permet de déterminer tous les autres tronçons dont l'une des extrémités est reliée à l'un de ces nœuds. Cette étape préliminaire offre la possibilité de « naviguer » dans le graphe ainsi créé.

Divers algorithmes s'appuyant sur ce graphe sont alors opérables pour répondre à un problème de plus court chemin ou dans notre cas du chemin le plus probable.

Le choix s'est porté sur l'utilisation de l'algorithme de Dijkstra (du nom de son créateur) en raison de sa facilité d'utilisation. Cela s'est traduit par l'utilisation des fonctions `pgr_dijkstra` et `pgr_trsp` (Turn Restriction Shortest Path) proposé par pgRouting. La fonction `pgr_trsp` est uniquement employée avec la base Here qui, au contraire de la BDTopo, contient des informations sur un nombre plus important de règles de circulation telles que les restrictions de tourne à droite, à gauche...

Les fonctions de routage utilisées requièrent comme paramètres d'entrée la saisie des nœuds de départ et d'arrivée. La détermination de ces éléments à partir du début et de la fin de la trace est délicate en raison de l'imprécision des GPS (surtout dans les premières minutes qui suivent l'allumage du smartphone). La solution retenue consiste à effectuer plusieurs opérations de routage en sélectionnant les extrémités des tronçons les plus proches des points d'arrivée et de départ

comme nœuds d'origine et de destination et à ne retenir au final que le chemin le plus « ressemblant » à la trace GPS.

Afin de valider le routage effectué, la longueur trace GPS est comparée à celle du chemin proposé par le routage. Ne sont retenues de manière automatique que les situations pour lesquelles l'écart, en valeur absolue, est inférieur à 10%.

Dans le cas du map-matching réalisé à l'aide de la base Here sur la Normandie, des opérations de validation complémentaires ont été mises œuvre. Les routages donnant lieu à des écarts légèrement plus importants que cette limite de 10% ont été examinés manuellement et certains ont été « repêchés ».

Dans certains cas, l'opération de routage ne propose aucun chemin. Les principales raisons de ces échecs sont : des écarts entre les informations du SIG et la réalité terrain de la route au moment de la circulation des conducteurs (travaux, erreur sans les sens de circulation...) ou l'emprunt par les 2RM de voies non autorisées à leur circulation ou non représentées dans le SIG.

#### d. Rétroprojection des points

Une fois le chemin validé, les points GPS composants la trace GPS sont rétroprojetés orthogonalement sur le réseau et ainsi associés individuellement à un tronçon du SIG ; ils héritent par conséquent de ces caractéristiques afférentes. Le sens de circulation du tronçon est également déterminé.

**Figure 6 : Rétroprojection des points GPS de la trace (en noir) sur le chemin proposé par l'algorithme de map-matching (en rouge)**



#### e. Quelques limites sur l'algorithme de map-matching

Au-delà des imperfections imputables à la qualité des données utilisées déjà signalées précédemment (précision du GPS, fidélité du SIG utilisé, travaux de voirie...), l'algorithme développé présente plusieurs défauts parmi lesquels :

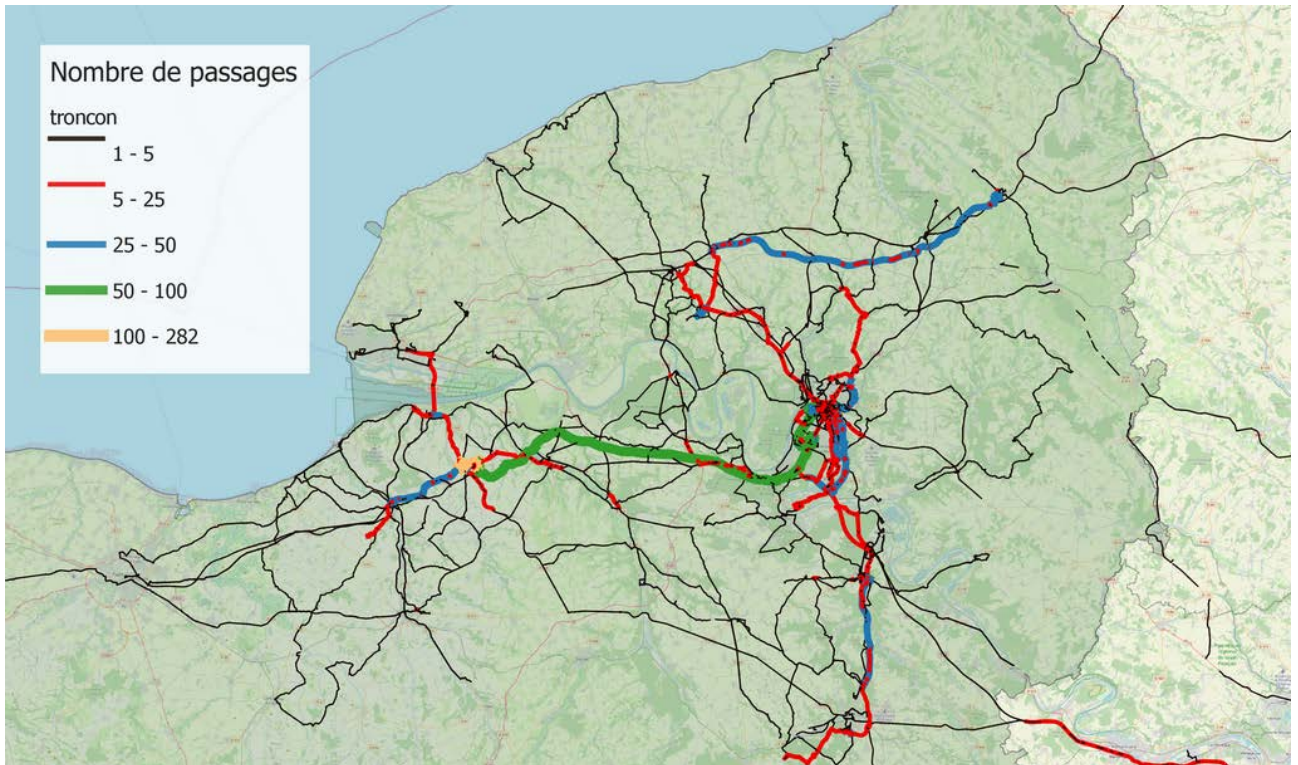
- son incapacité à gérer correctement les trajets comprenant des boucles (l'algorithme s'appuyant sur la minimisation des coûts ne rentre pas dans la boucle car elle allonge le coût) ;
- la possibilité de retenir sur certaines parties des traces des tronçons qui n'ont pas été empruntés ;
- de rétroprojeter de manière inappropriée, les points sur le réseau en choisissant, par exemple, un tronçon adjacent à celui auquel il appartient en réalité...

### 2.3.3. Résultats

Au final, France entière, le map-matching aboutit à un résultat jugé correct pour les données provenant de 7840 traces entraînant le rattachement de 6,4 millions de points de mesure aux tronçons de la BDTopo. Au total un peu plus de 100 000 tronçons de la BDTopo ont été circulés au moins une fois par l'un des véhicules de la flotte DYMOA.

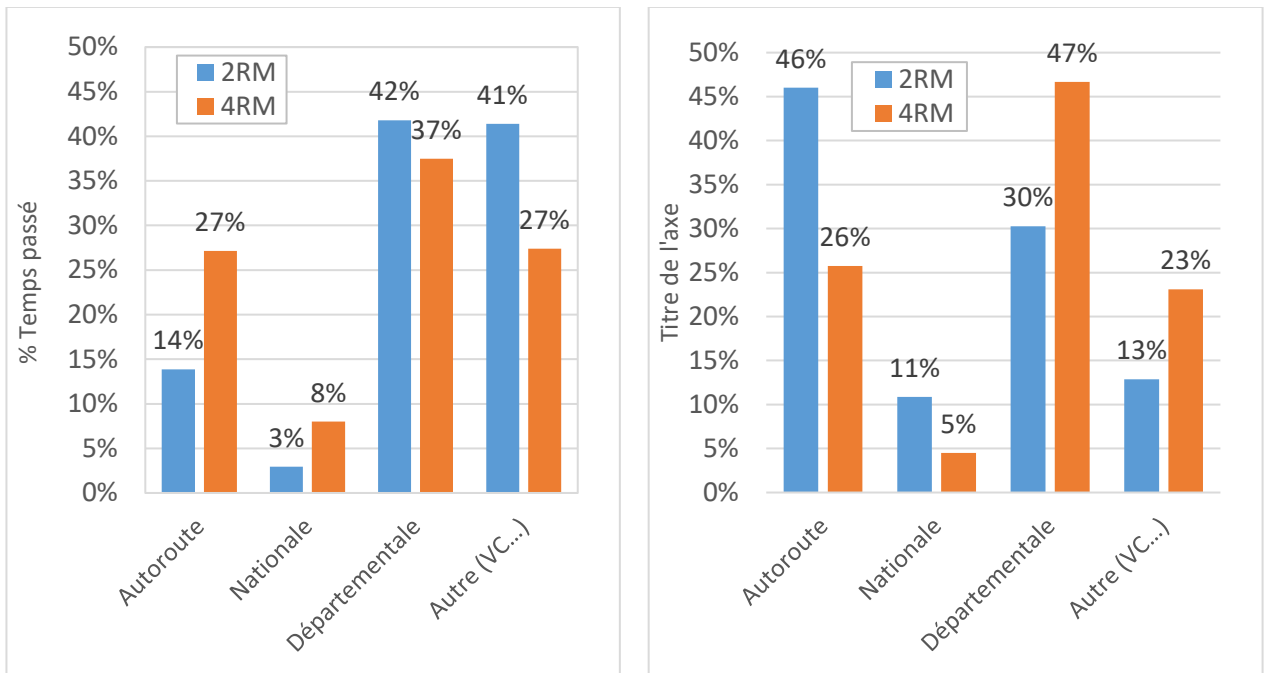
Ce travail permet d'établir une carte de « trafic » et de connaître le nombre de fois qu'un tronçon donné a été circulé, par quel conducteur (lettre) et avec quel type de véhicule (2RM / 4RM).

**Figure 7 : Carte du trafic de la flotte DYMOA sur la région Normandie**



La répartition des 102 000 km parcourus et 74,1 jours d'enregistrements de données associés selon la catégorie administrative des routes est donnée par la figure 8.

**Figure 8 : Distribution du temps passé (à gauche) et des distance parcourues (à droite) en 2RM et en 4RM selon la catégorie administrative de la route.**



### 3. Exploitations des données

Plusieurs niveaux d'analyse des données ont été conduits en essayant de tirer parti des informations disponibles dans les SIG utilisés.

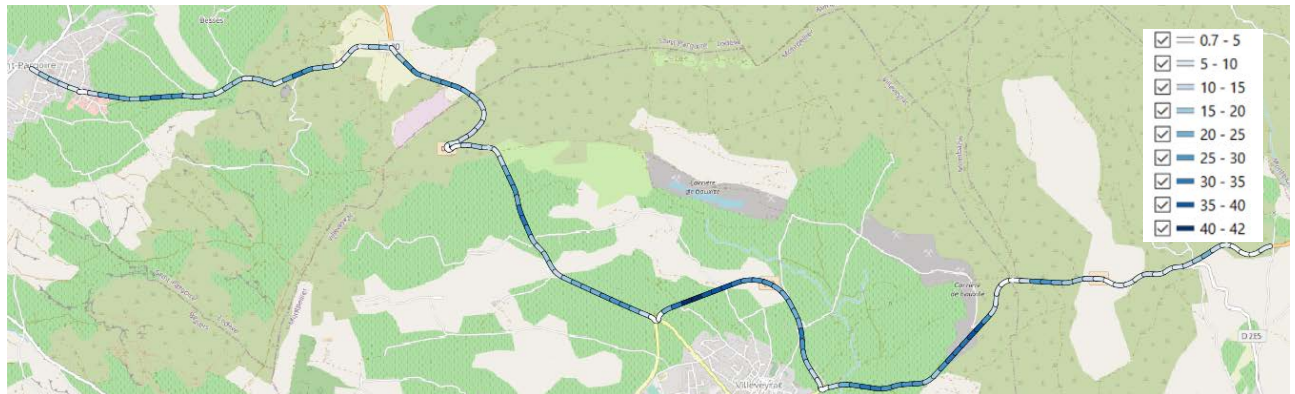
#### 3.1. Comparaison 2RM / 4RM pour un même conducteur et sur un même itinéraire

Afin de comparer de manière illustrative les vitesses adoptées par un conducteur selon qu'il conduise son 2RM ou 4RM, un itinéraire d'une quinzaine de kilomètres emprunté 26 fois en 2RM et 10 fois en 4RM est sélectionné.

L'itinéraire est découpé en sections de 100 m. La vitesse moyenne sur ces tronçons est calculée pour chaque passage et utilisée pour calculer les vitesses moyennes  $V_{2RM}$  et  $V_{4RM}$  de l'ensemble des passages sur chaque section respectivement en 2RM et 4RM.

Dans cet exemple, la vitesse moyenne sur chaque section en 2RM est systématiquement supérieure à celle des passages en 4RM (figure 9) ; les différences étant particulièrement importantes sur les parties rectilignes de l'itinéraire. En virage, et notamment en virage serré, les écarts de vitesse sont nettement moindres voir nuls.

**Figure 9 : Différence de vitesse moyenne  $V_{2RM} - V_{4RM}$  pour un même conducteur sur un même itinéraire par pas de 100 m (représentation en échelle de couleur).**



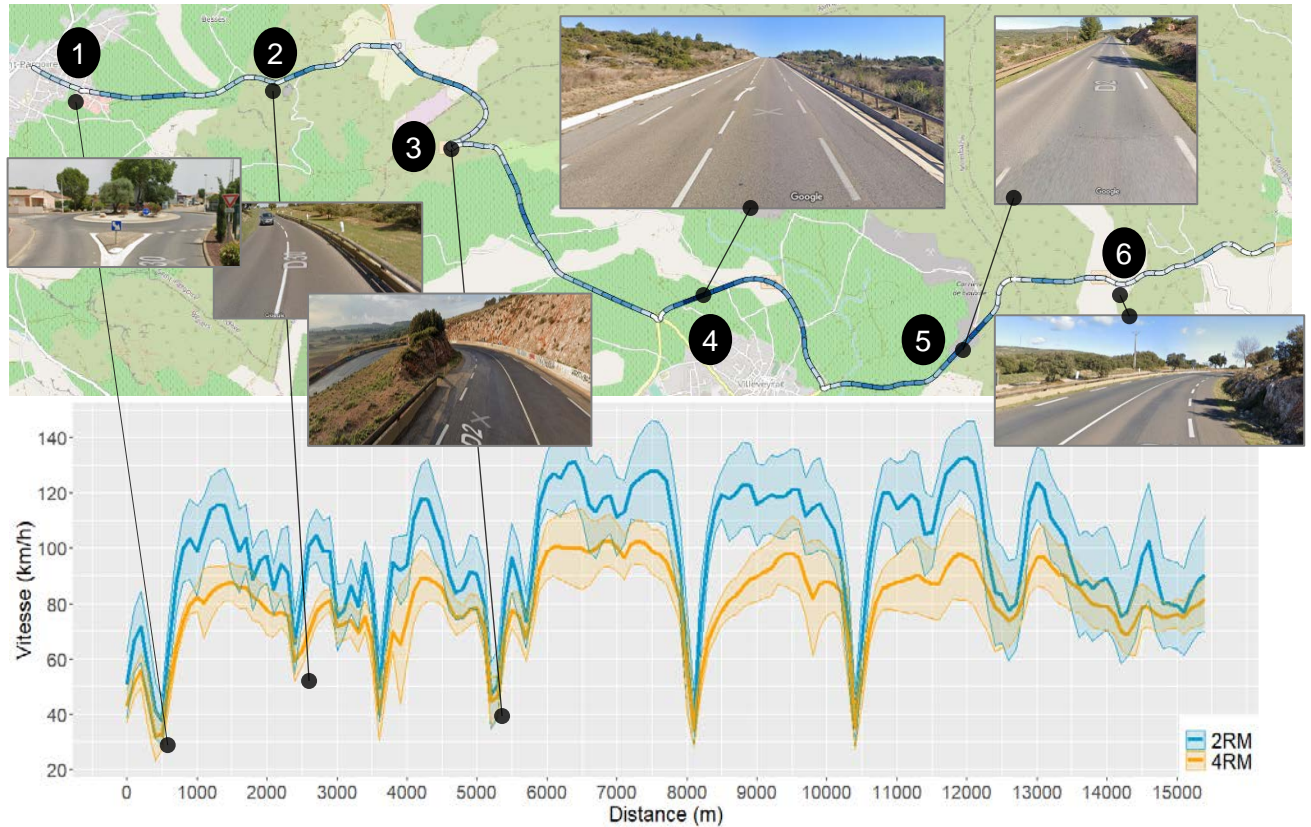
La mise en perspective des profils de vitesse des 2RM et 4RM avec l'itinéraire et ses aménagements corrobore ces premières impressions (figure 10).

Les écarts de vitesse les plus faibles sont observés sur les giratoires (1) et les courbes (2) pour lesquelles l'écart semble d'autant plus faible que le rayon courbure du virage est petit (3).

Les sections rectilignes (4,5 et 6), plutôt larges et dégagées sont quant à elles des portions de l'itinéraires sur lesquelles les vitesses en 2RM sont très largement supérieures à celles en 4RM.



Figure 10 : Profil des vitesses moyenne (et écart-type) des 2RM et 4RM en lien avec l'itinéraire et ses aménagements.



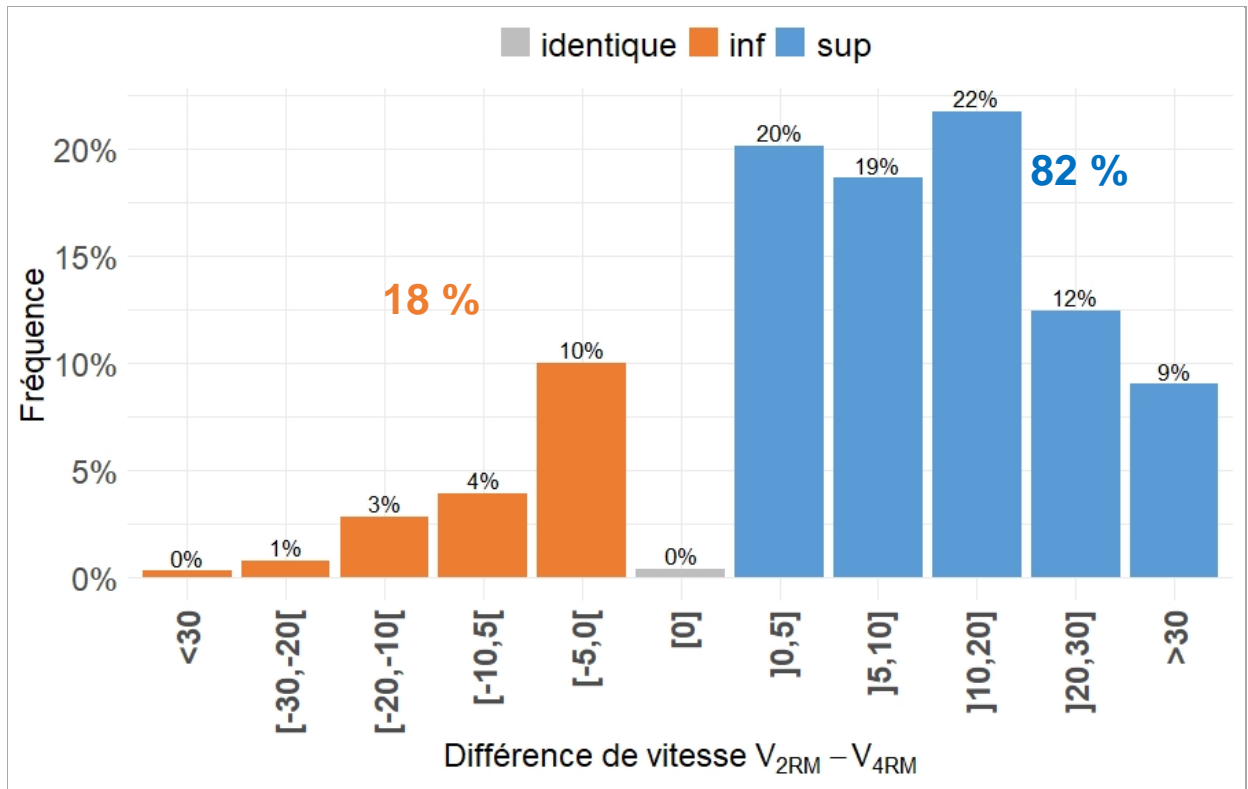
### 3.2. Comparaison 2RM / 4RM sur les tronçons circulés par un même conducteur

L'analyse des données porte sur les 9035 tronçons de la BD Topo qui ont été circulés au moins une fois dans le même sens de circulation par un même conducteur soit en 2RM soit en 4RM.

Les tronçons sont de dimensions variables. La vitesse  $V_{t,p}$  associée à un passage  $p$  sur un tronçon  $t$  est calculée en moyennant les vitesses des points de mesures appartenant à ce tronçon. La vitesse moyenne en 2RM  $V_{t,2RM}$  et en 4RM  $V_{t,4RM}$  sur chaque tronçon  $t$  est obtenue en calculant la moyenne des  $V_{t,p}$  propres à chaque catégorie de véhicule.

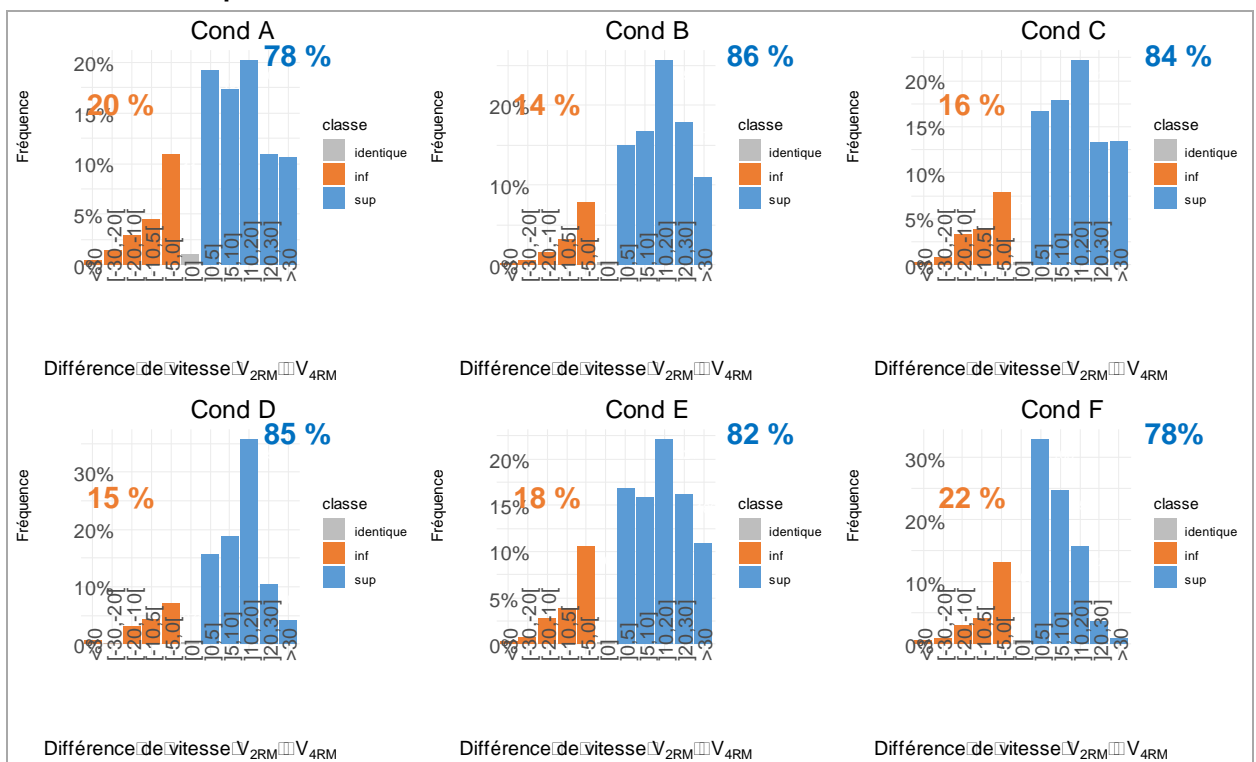
Sur 82% des tronçons la vitesse en 2RM est plus rapide qu'en 4RM (figure 11). Dans 21% des cas, l'écart est supérieur à 20 km/h.

Figure 11 : Répartition des différences de vitesse entre 2RM et 4RM sur les tronçons circulés au moins une fois par un même conducteur soit en 2RM soit en 4RM



La même analyse désagrégée au niveau chaque conducteur (figure 12) confirme cette tendance globale au niveau individuel avec toutefois quelques modulations pour les écarts supérieurs de 20 km/h et plus nettement moins fréquents pour certains des volontaires.

Figure 12 : Répartition des différences de vitesse entre 2RM et 4RM sur les tronçons circulés à la fois en 2RM et en 4RM par un même conducteur



### 3.3. Dépassement de la VMA

L'étude du dépassement de la Vitesse Maximale Autorisée requiert la connaissance de la limitation de vitesse qui s'applique sur chaque tronçon circulé par les conducteurs de la flotte. Cette information n'est malheureusement ni disponible dans la BD Topo, ni en open data à l'échelle de la France entière.

Cette information est par contre disponible dans la base de données de Here, pour laquelle le Cerema dispose d'un accès délimité uniquement à la Normandie. La version de la base de données utilisée est par ailleurs postérieure, d'une part à la période de l'expérimentation et, d'autre part à l'entrée en vigueur le 1<sup>er</sup> juillet 2018 du 80 km/h sur les routes bidirectionnelles hors agglomération. Ce dernier point a été traité en attribuant une VMA de 90 km/h aux tronçons dont la VMA de la base Here vaut 80 km/h.

Sur les près de 1 600 traces GPS dont au moins une partie s'effectue sur le réseau routier normand, 1 462 ont pu être, tout ou partie, raccrochées au réseau décrit dans la base Here. Cela correspond à 28 500 km parcourus sur 33 600 tronçons (3 600 km de réseau) et 43 700 tronçons en distinguant les sens de circulation et sur lesquelles 1,5 millions points de mesure vitesse sont rétroprojetés.

Ces trajets sont effectués par 9 conducteurs dont 3 ayant accepté d'équiper leur 2RM et 4RM.

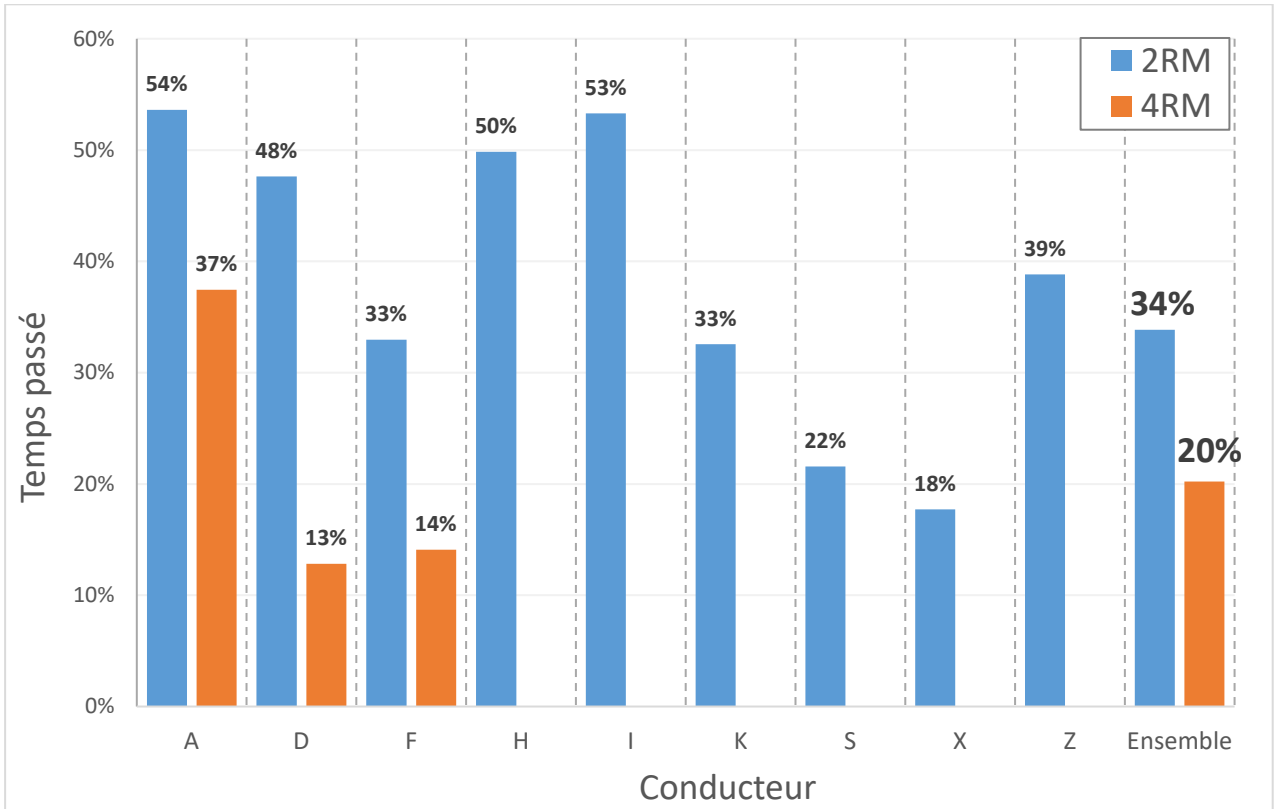
La comparaison des vitesses pratiquées avec la VMA se limite aux points de mesure associés aux tronçons de route dont la VMA est 30, 50, 70, 90, 110 ou 130 km/h. Elle est évaluée au regard de la proportion de temps passé ou de distance parcourue au-delà de la VMA mais également de la VMA+10 km/h.

Les résultats (tableau 1) indiquent que les 2RM circulent quasiment un tiers de leur temps et près de la moitié des distances qu'ils parcourent en excès de vitesse. La pratique de vitesses au-delà de la VMA+10 km/h est également assez fréquente : près 15% du temps de parcours et 21% des kilomètres parcourus. Le dépassement de la VMA pour les 4RM est moins fréquent mais reste néanmoins à des niveaux assez élevés. Ces constats globaux s'appliquent également aux 3 conducteurs de 2RM/4RM considérés individuellement (figures 13 et 14) même s'il existe une assez grande disparité interindividuelle dans les taux d'infraction.

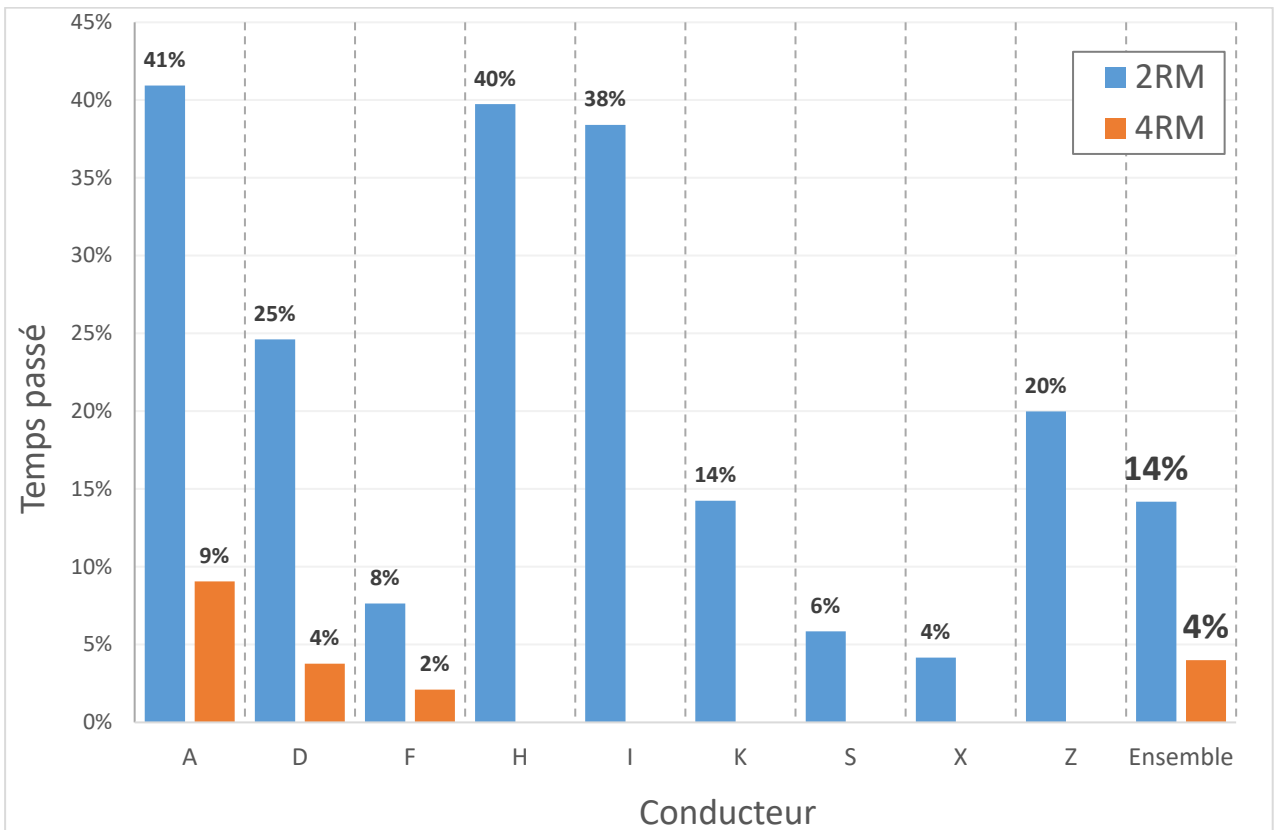
**Tableau 1 : Proportion de temps passé et de distance parcourue au-delà de la VMA et de VMA+10 km/h pour les 9 conducteurs lors de leurs trajets effectués en Normandie.**

Type de véhicule	> VMA		> VMA +10	
	% Temps passé	% Distance parcourue	% Temps passé	% Distance parcourue
2RM	34 %	47 %	14 %	21 %
4RM	20 %	28 %	4 %	5 %

**Figure 13 : Proportion de temps passé au-delà de la VMA selon le conducteur et le type de véhicule conduit lors de leurs trajets effectués en Normandie.**



**Figure 14 : Proportion de temps passé au-delà de la VMA+10 km/h selon le conducteur et le type de véhicule conduit lors de leurs trajets effectués en Normandie.**





### 3.4. Facteurs favorisant le dépassement de la VMA.

Après avoir constaté à l'aide d'une analyse univariée la forte propension des conducteurs de 2RM à circuler en excès de vitesse, ce chapitre vise à confirmer ce résultat par une analyse multivariée en y intégrant des variables descriptives des conditions de conduite et de l'infrastructure et à identifier celles qui favorisent ce comportement infractionniste.

#### 3.4.1. Construction du jeu de données

L'unité d'observation considérée est le tronçon routier T tel que défini dans la base Here associé aux sens de circulation S.

Une ligne de données correspond à la vitesse moyenne pratiquée  $V_{C,T,S}$  par un conducteur C pour chaque passage qu'il effectue sur un tronçon T, et pour chaque sens de circulation S.

Le tableau final qui en résulte est constitué de 234 607 observations dont la répartition par conducteur et type de véhicule est donnée dans le tableau 2.

Tableau 2 : Répartition du nombre d'observations par conducteur

Conducteur	Type de véhicule	Nbre de tronçon/sens	Type de véhicule	Nbre de tronçon/sens
A	2RM	7 003	4RM	16 645
D	2RM	20 534	4RM	1 500
F	2RM	49 598	4RM	45 941
H	2RM	11 270	-	-
I	2RM	3 094	-	-
K	2RM	2 9437	-	-
S	2RM	2 9079	-	-
X	2RM	13 602	-	-
Z	2RM	6 904	-	-

#### La variable à expliquer

La variable indépendante est la variable binaire excès qui vaut 1 si  $V_{C,T,S} > VMA_{T,S}$  et 0 dans le cas contraire.

#### Les variables explicatives

Les variables de contexte  $X_i$  sont uniquement celles provenant du recueil de données à savoir le type de jour de la semaine (Semaine/Weekend) et les conditions de luminosité (Jour/Nuit).

Les variables d'infrastructure proviennent de la base Here. Le choix des variables à prendre en compte dans le modèle s'est opéré en plusieurs étapes :

1. Un filtrage consistant à ne sélectionner que les variables disponibles dans la BD Here susceptibles d'avoir une incidence sur les vitesses pratiquées ;
2. Construction de certaines variables en s'appuyant sur les données primaires de la BD Here ;
3. Une opération de dénombrement visant à écarter les caractéristiques qui étaient trop peu présentes dans notre jeu de données ;
4. Une phase d'agglomération de certaines variables ou de modalités.

Au final 31 variables d'infrastructure ont été retenues (voir tableau 3). Sur le plan opérationnel, les caractéristiques sont attribuées à l'ensemble du tronçon. Les attributs ponctuels tels que les panneaux sont associés au tronçon sur lequel le message du signal s'applique. Elles sont susceptibles de différer selon le sens de circulation.

En complément de ces éléments est également associée la catégorie de véhicule (**2RM/4RM**).

**Tableau 3 : Variables d'infrastructure retenues pour l'analyse multivariée (en gras, la modalité retenue comme référence).**

Variable	Description et modalités
urban	Occupation des sols urbanisés ( <b>N/Y</b> )
adm_vma	Combinaison de la catégorie administrative de la route (Autoroute, <b>Nationale</b> , Départementale, Autre) et de la VMA (30, 50, 70, <b>90</b> , 110, 130).
ramp	Bretelle ( <b>N/Y</b> )
func_class	Classement hiérarchique de la route ( <b>1</b> à 5). Le niveau 1 étant le plus logique et le plus efficient pour l'usager
sens_unique	Un seul sens de circulation ( <b>N/Y</b> )
chauss_sep	chaussée séparée ( <b>N/Y</b> )
central_sep	Existence de séparation physique entre les voies (aucune ou non physique / physique < 3M). ( <b>N/Y</b> )
nb_voie_dir	Nombre de voies dans les sens de circulation ( <b>une</b> , deux, trois ou plus, indéterminé).
cat_lane_form	Changement du nombre de voies ( <b>profil constant</b> , une voie en plus, une voie en moins)
trans_area	Section de route avec changement du nombre de voies et absence de marquage ( <b>N/Y</b> )
acc_lane	Présence d'une voie d'accélération ( <b>N/Y</b> )
dec_lane	Présence d'une voie de décélération ( <b>N/Y</b> )
turn_lane	Présence d'une voie de tourne à droite ou tourne à gauche ( <b>N/Y</b> )
bike_lane	Présence d'une piste cyclable ( <b>N/Y</b> )
priorityrd	Route prioritaire ( <b>N/Y</b> )
roundabout	Circulation sur rond-point ( <b>N/Y</b> )
stop	Présence d'un stop ( <b>N/Y</b> )
cedez_pass	Présence d'un cédez-le-passage ( <b>N/Y</b> )
prio_droite	Présence d'une priorité à droite ( <b>N/Y</b> )
feux_tri	Présence d'un feu tricolore ( <b>N/Y</b> )
dep_interdit	Dépassement interdit ( <b>N/Y</b> )
dep_protege	Zone de dépassement spécifique ( <b>N/Y</b> )
vir_gauch	Présence d'un panneau annonçant à un virage à gauche ( <b>N/Y</b> )
vir_droit	Présence d'un panneau annonçant à un virage à droite ( <b>N/Y</b> )
vir_suc_gauch	Présence d'un panneau annonçant une succession de virage le premier étant à gauche ( <b>N/Y</b> )
vir_suc_droit	Présence d'un panneau annonçant une succession de virage le premier étant à droite ( <b>N/Y</b> )
pass_piet	Présence d'un passage piéton ( <b>N/Y</b> )
pan_enf	Présence d'un panneau annonçant la présence d'enfants ( <b>N/Y</b> )
trav_anim	Présence d'un panneau annonçant la traversée d'animaux ( <b>N/Y</b> )
retreci	Présence d'un panneau annonçant un rétrécissement de la chaussée ( <b>N/Y</b> )
ralentisseur	Présence d'un ralentisseur ( <b>N/Y</b> )
trav_anim	Présence d'un panneau annonçant la traversée d'animaux ( <b>N/Y</b> )

### 3.4.2. Choix du modèle et méthodologie d'analyse

#### Modèle retenu

S'agissant de modéliser une variable Y à valeurs binaires en fonction de plusieurs variables explicatives mixtes (c.a.d. quantitatives ou qualitatives), le choix s'est orienté vers l'utilisation d'un modèle linéaire généralisé, et plus particulièrement le modèle logistique. Ce modèle permet de se ramener à un modèle de régression linéaire classique en modélisant  $P(Y = 1 | X_1, \dots, X_p)$ .

En outre, le jeu de données est constitué d'observations appariées ne pouvant pas être considérées comme étant totalement indépendantes entre-elles. Les observations liées à un même conducteur

(en raison de facteurs propres à chaque individu) ayant vraisemblablement tendance à présenter plus de similitudes entre elles que les observations associées à des conducteurs différents.

Ce problème peut être adressé par des modèles mixtes qui ajoutent des effets aléatoires aux effets fixes des modèles de régression classiques afin de mieux prendre en compte les corrélations existantes entre observations issues d'un même individu ou groupe d'individu.

Au final les données sont ajustées par un modèle mixte de logistique binomial

$$P(\text{exces} = 1 | \text{cond}) = \frac{\exp(\beta_0 + \sum \beta_i \cdot X_i + \text{cond} \cdot u_{\text{cond}})}{1 + \exp(\beta_0 + \sum \beta_i \cdot X_i + \text{cond} \cdot u_{\text{cond}})}$$

Avec

- $\beta_0$ : La constante du modèle (dans notre cas  $\exp(\beta_0)$  est la probabilité d'un excès de vitesse pour les valeurs de références de toutes les variables du modèle, cf. tableau 3) ;
- $\beta_i$  : les coefficients d'ajustement des variables de contexte  $X_i$
- $u_{\text{cond}}$  : le coefficient de régression de la variable aléatoire pour le conducteur **cond**  $i$ . Le produit de ces deux termes correspondant à l'effet aléatoire.

Le lecteur pourra trouver en Annexe quelques détails techniques sur ces modèles. D'un point de vue pratique, les régressions sont réalisées à l'aide du logiciel R (R Core Team, 2021) et le recours au package lme4 (Bates et al., 2015).

#### Méthodologie d'analyse.

La démarche mise en œuvre poursuit deux objectifs, premièrement identifier les variables explicatives qui ont une influence statistiquement significative sur la probabilité d'être en excès de vitesse, et deuxièmement d'en quantifier les effets (taille de l'effet).

La stratégie retenue a consisté à réaliser la régression en utilisant l'ensemble des variables explicatives puis à relancer de manière itérative la régression en supprimant à chaque itération la variable dont la p-value était la plus élevée et supérieure à 0,05, soit un seuil de significativité de 5%.

Une fois cette procédure achevée, et dans l'optique de rendre plus facilement appréhendable et interprétable, les odds ratios (OR) résultants de ce traitement statistique sont transformés en risque relatif. Cette transformation est réalisée à l'aide du package effectsize (Makowski et al., 2019).

Les coefficients d'ajustement liés aux conducteurs ne sont pas étudiés, l'étude se focalisant sur les variables de contexte, et ne s'intéressant pas aux spécificités individuelles.

#### 3.4.3. Influence du type de véhicule

L'analyse multivariée conduite selon les modalités décrites précédemment et en écartant 6 variables non significative confirme (cf. 3.3) une plus forte occurrence des dépassements de la VMA par les 2RM avec un risque relatif d'être en excès sur un tronçon de route 1,61 [1,56 ; 1,64] (OR=2,77) fois plus élevé pour un 2RM que pour un 4RM. C'est-à-dire que cette augmentation est significative, et qu'elle est observée toutes choses égales par ailleurs puisque les facteurs de risques ont été conservés dans le modèle.

#### 3.4.4. Variables ayant une influence sur le dépassement de VMA

Les analyses sont conduites en séparant les données concernant les 2RM et 4RM. Les facteurs de risque relatifs résultant de la régression sont présentés dans le tableau 4. Les variables non significatives étaient (par ordre d'élimination du modèle) :

- Pour les 2RM : vir\_gauch, pan\_enf, vir\_droit, prio\_droite, vir\_suc\_droit, retreci, dec\_lane, turn\_lane ;
- Pour les 4RM retreci, trav\_anim, sens\_unique, dep\_protege , bike\_lane, vir\_suc\_droit, vir\_suc\_gauche, et prio\_droite.

**Tableau 4 : Facteur de risque relatif d'être en excès de vitesse sur un tronçon en fonction pour les différentes variables explicatives retenues dans les modèles (NS = Non significatif)**

Variable	Modalité vs Référence	2RM		4RM	
		RR	IC95%	RR	IC 95%
periode	Weekend vs. semaine	1,07	[1,06;1,08]	0,81	[0,77;0,85]
lumi	Nuit vs. jour	0,95	[0,94;0,97]	0,82	[0,79;0,85]
urbain	Urbain (oui vs. non)	0,64	[0,63;0,65]	0,41	[0,39;0,43]
adm_vma Catégorie administrative & VMA	Autoroute 130 vs. RN 90	0,83	[0,79;0,87]	0,23	[0,19;0,28]
	Autoroute 110 vs. RN 90	1,04 <sup>NS</sup>	[1,00;1,07]	1,13	[1,03;1,24]
	Autoroute 90 vs. RN 90	1,2	[1,16;1,23]	1,4	[1,25;1,55]
	Autoroute 70 vs. RN 90	1,36	[1,34;1,38]	1,57	[1,40;1,72]
	Autoroute 50 vs. RN 90	1,33	[1,28;1,37]	1,57	[1,35;1,75]
	RN 110 vs. RN 90	0,76	[0,73;0,79]	0,7	[0,64;0,75]
	RN 70 vs. RN 90	1,31	[1,26;1,35]	1,07 <sup>NS</sup>	[0,84;1,29]
	RN 50 vs. RN 90	1,2	[1,16;1,24]	1,22 <sup>NS</sup>	[0,98;1,45]
	RD 110 vs. RN 90	0,6	[0,52;0,68]	2,7x10 <sup>-5</sup> <sup>NS</sup>	[0,00;2,25]
	RD 90 vs. RN 90	0,6	[0,57;0,63]	0,23	[0,20;0,27]
	RD 70 vs. RN 90	1,06	[1,03;1,09]	1,3	[1,21;1,38]
	RD 50 vs. RN 90	1,23	[1,21;1,25]	1,62	[1,55;1,68]
	RD 30 vs. RN 90	1,46	[1,45;1,47]	1,97	[1,92;2,01]
	Autre 90 vs. RN 90	0,24	[0,19;0,29]	0,79	[0,65;0,95]
	Autre 70 vs. RN 90	0,72	[0,66;0,78]	1 <sup>NS</sup>	[0,83;1,18]
	Autre 50 vs. RN 90	1,13	[1,10;1,16]	1,76	[1,70;1,82]
Autre 30 vs. RN 90	1,43	[1,42;1,44]	1,98	[1,92;2,03]	
ramp	Bretelle (oui vs. non)	1,25	[1,21;1,28]	1,48	[1,37;1,58]
func_class Classe fonctionnelle	Niveau 2 vs. niveau 1	0,91	[0,88;0,93]	0,61	[0,53;0,69]
	Niveau 3 vs. niveau 1	0,79	[0,76;0,81]	0,47	[0,40;0,54]
	Niveau 4 vs. niveau 1	0,57	[0,54;0,60]	0,6	[0,51;0,69]
	Niveau 5 vs. niveau 1	0,5	[0,47;0,54]	0,11	[0,09;0,14]
sens_unique	Sens unique (oui vs. non)	0,9	[0,88;0,93]	NS	-
chauss_se	Chaussée séparée (oui vs. non)	1,06	[1,04;1,07]	1,3	[1,25;1,35]
central_sep	Séparateur central < 3m (oui vs. non)	0,56	[0,53;0,59]	0,24	[0,20;0,29]
nb_voie_dir (nb. voies ds le sens de circ.)	Deux vs. 1	1,04	[1,02;1,06]	1,17	[1,10;1,23]
	Trois ou plus vs. 1	0,99 <sup>NS</sup>	[0,96;1,01]	1,26	[1,18;1,34]
	Indéterminé vs. 1	1,01 <sup>NS</sup>	[0,99;1,03]	1,14	[1,08;1,20]
cat_lane_form	1 voie en + vs. profil constant	0,99	[0,97;1,02]	1,04 <sup>NS</sup>	[0,94;1,14]
	1 voie en - vs. profil constant	1,07	[1,03;1,10]	1,33	[1,23;1,42]
trans_area	Aire de transition (oui vs. non)	0,43	[0,34;0,53]	0,19	[0,11;0,34]
acc_lane	Voie d'accélération (oui vs. non)	0,95	[0,92;0,98]	0,68	[0,62;0,74]
dec_lane	Voie de décélération (oui vs. non)	NS	-	0,78	[0,71;0,85]
turn_lane	Voie de tad ou tag (oui vs. non)	NS	-	1,21	[1,10;1,32]
bike_lane	Piste cyclable (oui vs. non)	1,09	[1,06;1,12]	NS	-
priorityrd	Route prioritaire (oui vs. non)	1,15	[1,14;1,16]	1,33	[1,29;1,38]
roundabout	Rond-point (oui vs. non)	0,23	[0,21;0,25]	0,04	[0,03;0,05]
stop	Stop (oui vs. non)	0,21	[0,15;0,29]	0,14	[0,05;0,36]
cedez_pass	Cédez-le-passage (oui vs. non)	0,56	[0,50;0,62]	0,22	[0,15;0,31]
feux_tri	Feux tricolore (oui vs. non)	0,5	[0,47;0,52]	0,36	[0,29;0,45]
dep_interdit	Dépassement interdit (oui vs. non)	0,93	[0,91;0,96]	1,27	[1,17;1,36]
dep_protege	Dépassement protégé (oui vs. non)	1,28	[1,24;1,32]	NS	-
vir_gauch	Virage à gauche (oui vs. non)	NS	-	1,53	[1,34;1,69]
vir_droit	Virage à droite (oui vs. non)	NS	-	1,22	[1,03;1,41]
vir_suc_gauch	Success. Vira. 1 <sup>er</sup> à gauche (oui vs. non)	0,91	[0,83;0,99]	NS	-
pass_piet	Passage piéton (oui vs. non)	0,91	[0,90;0,93]	0,64	[0,59;0,69]
pan_enfant	Panneau enfant (oui vs. non)	NS	-	0,52	[0,40;0,66]
pan_trav	Panneau traversée d'animaux (oui vs. non)	1,07	[1,04;1,10]	NS	-
ralentisseur	Ralentisseur (oui vs. non)	0,69	[0,65;0,73]	0,46	[0,40;0,53]

**Note : Les RR des 2RM et 4RM ne sont pas directement comparables entre eux.**

L'exploitation met exergue plusieurs effets intéressants.

#### Le contexte

Les dépassements de VMA sont moins fréquents de nuit que de jour. Au contraire des 2RM, les jours de week-end sont associés à une plus forte probabilité de rouler au-delà de la VMA pour les 2RM.

#### La catégorie de route et la VMA (figure 15),

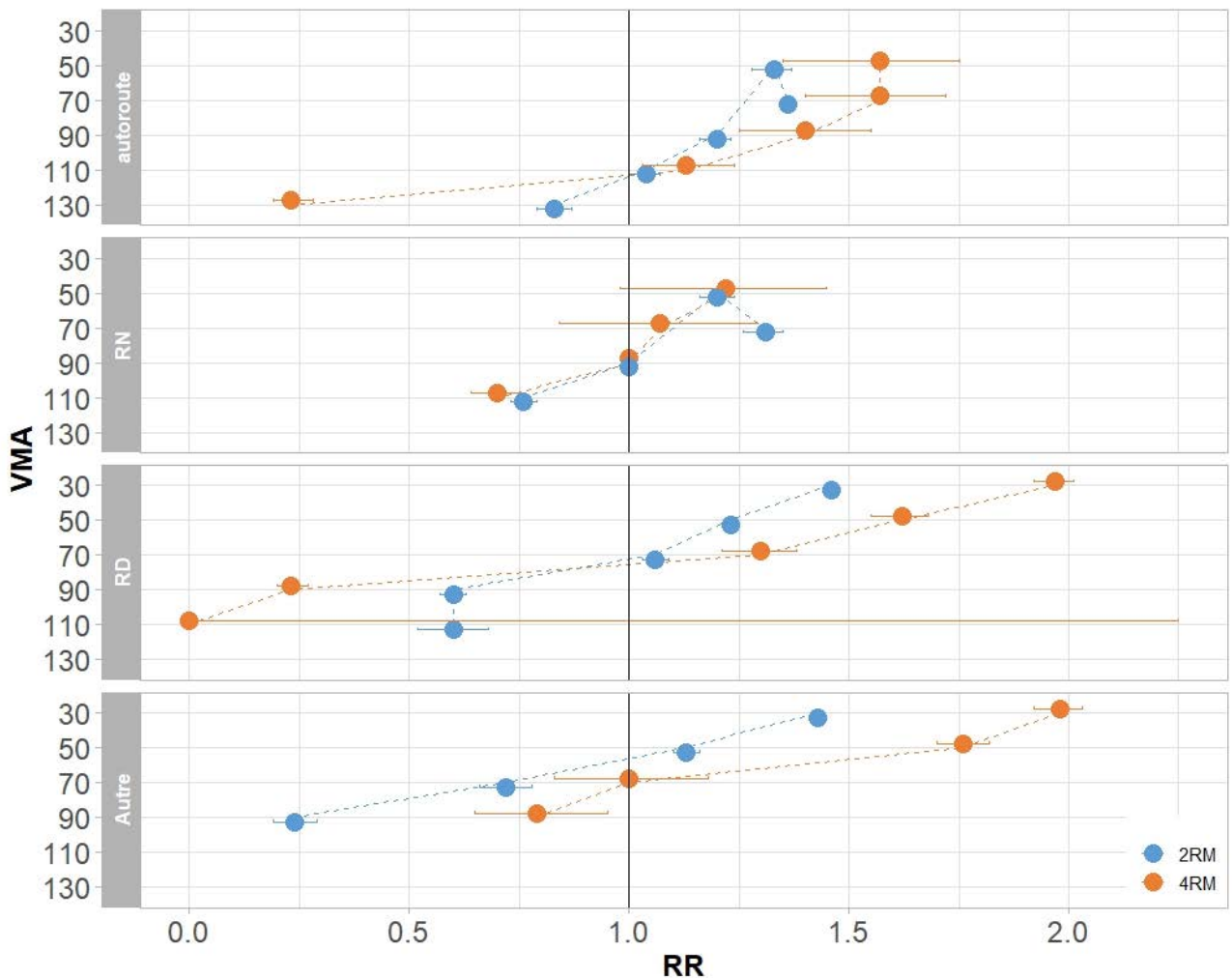
Les constats sont similaires pour les 2RM et les 4RM et une double tendance se dégage :

- Pour une même catégorie de réseau, le risque de dépassement est d'autant plus important que la VMA est basse ;
- Pour une même VMA, ce risque est d'autant plus grand que le réseau est le plus structurant. Ce point est par ailleurs confirmé par l'examen de la variable `func_class` ; le risque s'atténuant par ordre hiérarchique décroissant.

C'est par ailleurs sur les routes nationales (RN) que les comportements semblent les plus homogènes ; l'étendue des RR selon la VMA y étant la plus réduite.

Les éléments de réseau spécifiques que constituent les bretelles (ramp) (accès et sortie non différenciées ici) sont très propices au non-respect de la VMA tout spécialement pour les 4RM (RR=1,48).

**Figure 15 : Risque relatif d'être en excès de vitesse selon la catégorie de route et sa VMA pour les 2RM et 4RM (dans les deux cas le RR se rapporte à la situation d'une RN avec une VMA de 90 km/h).**



### Chaussée séparée / existence d'un séparateur central de moins de 3 m

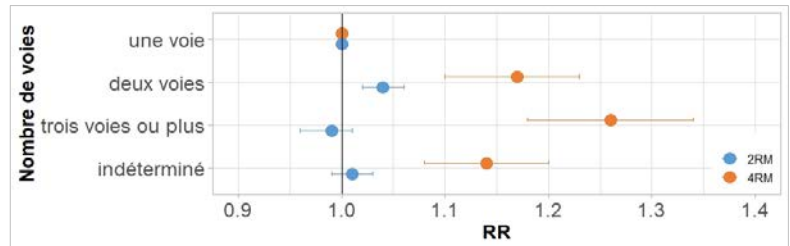
Les voies à chaussées séparées favorisent la pratique de vitesses au-delà des VMA notamment pour les 4RM (RR=1,3). La présence de délimiteur physique de largeur inférieure à 3 m est par contre associée à une forte réduction de ce type de comportement.

### Nombre de voies dans le sens de circulation et variation du profil en travers

Les tronçons comportant plusieurs voies de circulations (nb\_voie\_dir) accroissent assez fortement la probabilité de circuler en excès de vitesse pour les 4RM ; cet effet n'étant pas relevé pour les 2RM.

La modification du profil en travers de la route (cat\_lane\_form) avec réduction du nombre de voies s'accompagne d'une augmentation des excès vitesses tant pour les 2RM que les 4RM.

**Figure 16 : Risque relatif d'être en excès de vitesse selon le nombre de voies dans le sens de circulation**



Les aires de transitions (trans\_area) avec changement du nombre de voies et absence de marquage qui correspondent, en grande partie, aux sections routières situées en aval des barrières de péage sont propices à un meilleur respect des VMA.

### Les autorisations de dépassement

Les tronçons sur lesquels le dépassement est interdit sont associés à un risque plus élevé de dépassement de la VMA par les 4RM (RR=1,27) et un risque plus faible pour les 2RM (RR=0,93).

Les sections routières offrant une voie supplémentaire pour réaliser un dépassement dans de bonnes conditions incitent les 2RM à ne pas respecter la VMA (RR=1,28) ; l'effet étant non significatif pour les 4RM.

### Les régimes de priorité

Circuler sur une route prioritaire (priorityrd) favorise le dépassement de la VMA (RR<sub>2RM</sub>=1,15 ; RR<sub>4RM</sub>=1,33).

Les sections avec présence d'un stop, cédez-le-passage ou d'un feu tricolore, s'accompagnent, comme attendu, d'un meilleur respect de la VMA.

### Les virages

Les 4RM circulant sur les tronçons en aval d'un panneau de danger annonçant un virage (A1a ou A1b) affichent un fort risque de pas respecter la limitation de vitesse (aucun effet significatif pour les 2RM).

### Autres points

La présence de passages piétons et de ralentisseurs est également associée à un risque relatif plus faible d'excès de vitesse pour les deux populations étudiées.

La présence de certaines voies de circulation dédiées (accélération, décélération, tourne-à-gauche) semble également impacter le respect de la VMA. Il convient toutefois de signaler que l'analyse repose sur leur simple présence sur le tronçon circulé sans savoir si la dite voie est circulée ou non le conducteur considéré.



## 4. Conclusions

---

Le projet DYMOA a conduit à la constitution d'une base de données, relativement unique et à haute fréquence temporelle, des vitesses pratiquées par des conducteurs de 2RM et 4RM en situation de conduite naturelle.

Pour être pleinement exploitées, il convient toutefois de contextualiser les données recueillies aux caractéristiques de l'environnement auquel sont confrontés les conducteurs. Cela peut être mis en œuvre à grande échelle et de manière automatique à l'aide de traitements de raccordement géomatique (mapmatching), permettant ainsi de repérer des nuages de points issus de différents trajets sur un référentiel routier commun. Une fois associées à des objets géographiques bien définis et identifiés il devient alors possible de comparer les données provenant de plusieurs parcours, d'un ou de plusieurs conducteurs. La qualité et la fiabilité de ce type traitement dépend de nombreux paramètres qui lui sont intrinsèques (technique algorithmique retenue) mais surtout extrinsèques (précision du GPS, fidélité cartographique du SIG au monde réel...).

L'application de ce procédé aux données DYMOA a montré que sur des tronçons (au sens géomatique du terme) ou itinéraires identiques les conducteurs adoptent dans une très large majorité des cas (82 % des tronçons concernés) des vitesses plus élevées - voir bien plus élevées - au guidon de leur 2RM qu'au volant de leur 4RM.

Le second intérêt du raccordement géographique concerne l'enrichissement des données recueillies par l'adjonction des informations disponibles dans les SIG notamment en matière de connaissances des caractéristiques d'infrastructure et d'aménagements de la route. Les bases de données de données routières en open data (BD TOPO, OpenStreetMap) ne sont malheureusement pas très riches en informations routières. D'autres sources privées telle que la base Navstreets de la société Here utilisée pour ce travail, contiennent davantage d'attributs routiers. Conçues pour les navigateurs GPS et malgré les limitations afférentes à leur manque de complétude, plusieurs de leurs attributs, jugés pertinents, sont utilisés, d'une part, pour caractériser le comportement des conducteurs au regard du respect de VMA selon la catégorie de véhicule conduite et, d'autre part, étudier les caractéristiques de l'infrastructure, qui influent ou non sur ce respect.

Un premier niveau d'analyse univariée indique que les 2RM sont plus souvent en excès de vitesse que les 4RM aussi tant en temps passé (34 % vs. 20 %) qu'en distance parcourue (47% vs. 28%). Ce résultat est confirmé par une analyse multivariée recourant à une régression logistique binomiale à effet mixte (GLMM) pour mieux tenir compte des effets aléatoires propres à chaque conducteur. La probabilité qu'un 2RM soit en excès de vitesse sur un tronçon routier donné est, toute chose égale par ailleurs, 1,6 fois supérieure à celle d'un 4RM.

Divers facteurs de contexte et d'infrastructure modulent le risque d'être en excès de vitesse pour les 2RM et les 4RM. Ce risque est autant plus important que le réseau circulé est structurant et que les VMA sont moins élevées. Certaines configurations routières favorisent ce comportement infractionniste pour les deux catégories de véhicule : voies à chaussées séparées, bretelles, élargissement de profil en travers par ajout d'une voie, statut prioritaire de la route... D'autres semblent inciter au contraire à un meilleur respect de la VMA : existence d'un séparateur central de largeur inférieure à 3 m, intersections gérées par stop, cédez-le-passage ou feux tricolores...

Dans certains cas les situations sont différenciées pour les 2RM et les 4RM. Ainsi le taux de non-respect s'accroît pour les 4RM s'accroît, par exemple, avec le nombre de voies de circulation ou la présence de panneau signalisant alors qu'une telle tendance n'est pas observée pour les 2RM. À contrario, les sections routières disposant d'une voie supplémentaire pour réaliser un dépassement sont associés à une plus forte propension des 2RM à circuler au-delà de la VMA ; ce paramètre étant non significatif pour les 4RM.

Ces résultats sont à considérer avec précaution quant à leur généralisation en raison du nombre limité de conducteurs observés. Ils confirment néanmoins l'intérêt et la pertinence de mener ce type d'investigation et de disposer de bases de données routières fournies et fiables.

## 5. Références

---

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, **67**(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.

Ledoux V, Rongione B. (2018). "Faisabilité et pertinence d'un observatoire des vitesses" Livrable L2.5 du projet DYMOA.

Makowski D, Ben-Shachar M, Lüdtke D (2019). "Compute and interpret indices of effect size." *CRAN*. R package, <https://github.com/easystats/effectsize>.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.



## 6. Annexe

### 6.1. Régression logistique

Lorsque l'indépendance des données est supposée, la méthode usuelle pour étudier la relation entre une variable réponse binaire et plusieurs variables explicatives est la régression logistique. On introduit les notations suivantes :

Soit  $T_i$  le nombre d'observations réalisées sur le sujet  $i$  on note  $Y_i = (Y_{i1}, \dots, Y_{iT_i})$  le vecteur des observations, ces observations étant supposées indépendantes. Autrement dit, on note :

$$Y_{ij} = \begin{cases} 1 & i = 1, \dots, I; j = 1, \dots, T_i \\ 0 & \end{cases} \quad (1.1.)$$

La variable aléatoire  $Y_{ij}$  suit une loi de Bernoulli de paramètres  $p_{ij}$  où  $p_{ij} = P(Y_{ij} = 1)$ . Le modèle logistique ordinaire s'écrit :

$$\text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta' X_{ij} \quad (1.2.)$$

$$\text{Var}(Y_{ij}) = \mu_{ij}(1 - \mu_{ij}) \text{ (fonction de variance)} \quad (1.3.)$$

Où  $\mu_{ij} = p_{ij}$ ,  $X_{ij}$  est le vecteur des variables explicatives et  $\beta$  est le vecteur des paramètres. Dans ce cas, le vecteur  $\beta$  des paramètres est estimé par la méthode du maximum de vraisemblance. L'estimateur de  $\beta$  est solution des équations de vraisemblances (équations du score) obtenues en dérivant la fonction de log-vraisemblance (notée  $l$ ) par rapport à  $\beta$  :

$$U(\beta) = \frac{\partial l}{\partial \beta} = \sum_i D'_i \Sigma_i^{-1} (Y_i - \mu_i) = 0 \quad (1.4.)$$

Où  $D_i = \frac{\partial \mu_i}{\partial \beta'}$  (matrice diagonale),  $\Sigma_i$  est la matrice des variances-covariances (matrice diagonale  $\Sigma_i = \text{diag}(\text{Var}(Y_{ij}))$ ),  $Y_i$  est le vecteur des  $Y_{ij}$ , et  $\mu_i$  est le vecteur des  $\mu_{ij}$ . Les équations ci-dessus sont appelées "independence estimating equations"(IEE) et sont résolues par des méthodes itératives : algorithme du scoring de Fisher, méthode itérative des moindres carrés pondérés (IWLS) ou algorithme de Newton modifié (Quasi-Newton). L'estimateur obtenu  $\hat{\beta}$  est alors consistant (i.e. cv en probabilité vers  $\beta$ ) et distribué asymptotiquement selon une loi normale avec pour matrice de covariance  $V(\hat{\beta}) = (\sum_i D'_i \Sigma_i^{-1} D_i)^{-1}$ .

Cependant, dans notre cas, les trajets effectués par un même conducteur (ou les sous-sections de trajets) sont corrélés et ne peuvent donc pas être supposés indépendants. En effet, les trajets effectués par un même conducteur ont les mêmes caractéristiques car le comportement de conduite reste similaire. Ce problème de données corrélées apparaît dans de nombreuses situations : mesures répétées dans le temps appelées données longitudinales (principalement en épidémiologie), mesures issues d'une même famille ou cluster. Dans le domaine des transports, les observations issues d'expérience de type "naturalistic driving" appartiennent clairement à cette dernière catégorie : les observations issues d'un même conducteur sont, en général, fortement corrélées. L'hypothèse d'indépendance étant violée, on ne peut pas utiliser les modèles classiques comme les modèles linéaires généralisés dont la régression logistique est un cas particulier, et qui reposent sur l'indépendance des observations.

Il existe principalement deux méthodes permettant de prendre en compte la liaison entre observations réalisées sur un même sujet :

1. les modèles marginaux ou GEE (Generalized Estimating Equations)
2. les modèles mixtes.

Seuls les modèles mixtes seront décrits ici. Le lecteur pourra se référer à Liang et Zeger (1986) pour l'introduction à ces modèles, et à Diggle et al. [1994] pour une description détaillée des modèles GEE.

## 6.2. Modèles mixtes

Une autre approche pour modéliser des données corrélées est l'utilisation des modèles linéaires généralisés mixtes (GLMM). La méthode vue à la section précédente modélise les distributions marginales en traitant la corrélation comme un paramètre de nuisance. La méthode présentée ici introduit un effet aléatoire spécifique à chaque sujet. On modélise ainsi la variable réponse conditionnellement à ces effets aléatoires. L'inférence est donc individuelle ("subjects-specific approach"), contrairement aux modèles marginaux présentés à la section précédente où l'on modélisait la moyenne de la population ("population-averaged approach").

Le modèle logistique mixte s'écrit :

$$\text{logit}[P(y_{ij} = 1|u_i)] = x'_{ij}\beta + z'_{ij}u_i \quad (1.5.)$$

Où  $x_{ij}$  et  $\beta$  sont définis dans le modèle 1.2,  $u_i$  est le vecteur des effets aléatoires associés au sujet  $i$  et  $z_{ij}$  est la matrice associée.

Un cas particulier important est le cas où  $z_{ij} = 1$ . Ce modèle est appelé modèle logistique avec effet aléatoire sur la constante ("random-intercept logistic model" ou "logistic-normal model") et s'écrit :

$$\text{logit}[P(y_{ij} = 1|u_i)] = x'_{ij}\beta + u_i \quad (1.6.)$$

où les  $u_i$  sont indépendants et  $u_i \sim N(0, \sigma^2)$ . Soit  $y$  le vecteur des observations et  $u$  le vecteur des effets aléatoires. La vraisemblance marginale ("marginal likelihood") s'écrit :

$$l(\beta, \sigma^2; y) = f(y; \beta, \sigma^2) = \int f(y|u; \beta) f(u; \sigma^2) du \quad (1.7.)$$

où  $f(y|u; \beta)$  est la densité conditionnelle de  $y$  sachant  $u$  et  $f(u; \sigma^2)$  est la densité normale de  $u$ . En particulier, la vraisemblance marginale du modèle 1.11 s'écrit :

$$l(\beta, \sigma^2; y) = \int \prod_i \prod_j \left[ \frac{\exp(x'_{ij}\beta + u_i)}{1 + \exp(x'_{ij}\beta + u_i)} \right]^{y_{ij}} \left[ \frac{1}{1 + \exp(x'_{ij}\beta + u_i)} \right]^{1-y_{ij}} f(u_i; \sigma^2) \quad (1.8.)$$

L'estimation des paramètres se fait par approximation puis maximisation de cette vraisemblance. Il existe pour cela différentes méthodes :

### 1. Quadrature de Gauss-Hermite

Cette méthode consiste à approximer l'intégrale 1.13 par des méthodes numériques puis à maximiser cette approximation de la vraisemblance. L'approximation est réalisée par la formule de la quadrature de Gauss-Hermite :

$$\int_{-\infty}^{\infty} f(u) \exp(-u^2) du \approx \sum_{k=1}^q c_k f(s_k) \quad (1.9.)$$

où les  $c_k$  sont des poids et les  $s_k$  sont des points de quadrature. L'approximation est d'autant meilleure que le nombre de points de quadrature  $q$  est grand. Cette approximation de la vraisemblance est ensuite maximisée par des algorithmes classiques comme l'algorithme de Newton-Raphson.

### 2. Quasi-vraisemblance pénalisée ou "Penalized Quasi-Likelihood" (PQL)

Cette méthode consiste à approximer le modèle linéaire généralisé mixte par un modèle linéaire mixte ("linearization method"). La vraisemblance marginale 1.10 résulte de l'intégration de la

distribution jointe de  $y$  et  $u$  par rapport aux effets aléatoires  $u$ . En utilisant la représentation sous forme de famille exponentielle de chaque composant de cette loi jointe, l'intégrale 1.12 se ramène à une fonction exponentielle de  $u$ . Cette fonction peut alors être approximée par une série de Taylor du 2nd ordre ("Laplace approximation").

Cette approximation, appelé quasi-vraisemblance pénalisée, est maximisée en utilisant les méthodes de maximisation des modèles linéaires mixtes (REML : Restricted Maximum Likelihood).

### 3. Autres méthodes

Enfin, on peut citer aussi les méthodes de Monte Carlo ainsi que l'approche Bayésienne.

Pour un exposé plus détaillé des caractéristiques de ces modèles, on pourra se référer au livre de Stroup, (2012).

## REFERENCES

Diggle P. J., Liang K. Y. et Zeger S. L. (1994). « Analysis of longitudinal data ». Oxford Press.

Liang, K.Y. and Zeger, S. L. (1986) "Longitudinal Data Analysis Using Generalized Linear Models", *Biometrika*, Vol. 73, No. 1 (Apr., 1986), pp. 13-22.

Stroup, W.W. (2012), "Generalized Linear Mixed Models", CRC Press