

## ESTIMATIONS DE PARTS MODALES DES DÉPLACEMENTS À L'ÉCHELLE TERRITORIALE

David LEMARCHAND

Insee, Direction régionale Centre-Val de Loire, Services Études et Diffusion

david.lemarchand@insee.fr

**Mots-clés** : modèle prédictif, estimations petits domaines, régression pénalisée, *elastic net*, *bootstrap*.

**Domaines** : Statistique spatiale – Estimation sur petits domaines.

---

### Résumé

Cet article propose une méthode afin d'estimer des distances de déplacement parcourues par des individus à une échelle locale à partir de données d'enquête nationale.

A partir de l'enquête mobilité des personnes (EMP) réalisée par le Service des Données et Études Statistiques (SDES), un modèle d'apprentissage de type régression pénalisée (*Elastic Net*) est construit. Grâce à une implémentation dans le package *glmnet* (Friedman 2023), il est possible de générer plusieurs prédictions pour un même individu, correspondant à des distances tous motifs confondus et distances pour des motifs non professionnels pour plusieurs modes de transport (voiture, train...). L'article explicite une méthode afin de déterminer des valeurs d'hyperparamètres  $\alpha$  et  $\lambda$  du modèle.

Afin de mesurer la précision de ces prédictions, une méthode de rééchantillonnage intitulée *Bootstrap Bias Corrected Cross-Validation* (Tsamardinos 2018) a été appliquée. Elle permet de mesurer l'écart entre les prédictions du modèle et les distances mesurées dans l'enquête EMP. Cette méthode a l'avantage d'être parcimonieuse en temps de calcul grâce à la création d'une matrice contenant les prédictions individuelles pour différents jeux d'hyperparamètres du modèle. Cela évite de réestimer les prédictions à chaque rééchantillonnage.

Les déplacements et la part des distances parcourues selon différents mode de transport diffèrent selon les territoires de résidence des individus. On propose de vérifier l'adaptabilité du modèle en comparant les parts modales mesurées sur des territoires infra-départementales. Dans un premier temps les prédictions des distances tous motifs de déplacement sont comparés à des parts de déplacement pour différents modes de transport sur des territoires qui détiennent des compétences en matière de transport : les bassins de mobilité (Ménard, Movellan, Tillard, Olivier, Leconte 2024). Néanmoins, cette comparaison souffre d'un problème d'endogénéité : les parts de déplacements domicile-travail par modes de déplacement calculées à partir du recensement de la population sont corrélées à un des prédicteurs du modèle (le mode de déplacement principal pour se rendre sur le lieu de travail).

Dans un second temps, toujours dans l'objectif de vérifier l'adaptabilité du modèle, les prédictions du

modèle sont comparées aux parts modales calculés à partir de l'Enquête Mobilité Certifié Cerema (EMC<sup>2</sup>) réalisé dans l'agglomération tourangelle. Bien que le champ des déplacements considérés et les méthodes d'estimation des distances diffèrent selon les deux enquêtes, les résultats sont relativement proches.

## Abstract

This article suggests a method for estimating the distances travelled by individuals on a local scale using national survey data.

From the 'enquête mobilité des personnes' (EMP) conducted by the 'Service des Données et Études Statistiques' (SDES), a statistical learning model of the penalized regression family (Elastic Net) is carried out. Thanks to an implementation of the method in *glmnet* package (Friedman 2023), multiple outcomes (predictions) are generated for each observation. Travel distances for any purpose or for non professional reasons were hence estimated for several modes of transport (car, train...). The article highlights a way to calculate the  $\alpha$  and  $\lambda$  hyperparameters of the model.

A resampling method called Bootstrap Bias Corrected Cross-Validation (Tsamardinos 2018) is introduced in order to estimate the prediction precision. This method estimates the gap between model predictions and travel distances in the EMP survey. Saving time computation is a benefit of this process thanks to a matrix which summarized the individual predictions for different sets of model hyperparameters. This avoids having to re-estimate predictions at each resampling.

Travel patterns and the proportion of distances travelled by different modes of transport vary according to where individuals live. We propose to verify the adaptability of the model by comparing the modal shares estimated on local level. First, predictions of distances for all reasons of travel are compared with travel shares of travel for different modes of transport in territories with transport responsibilities: the mobility areas (Ménard, Movellan, Tillard, Olivier, Leconte 2024). However this comparison suffers from an endogeneous issue. The shares of commuting by mode of transport calculated from the population census are correlated with one of the input variable of the predictive model (the main mode of transport to get to work).

Secondly, still with the aim of verifying the adaptability of the model, the model's predictions are compared to the modal shares calculated from the Cerema Certified Mobility Survey (EMC<sup>2</sup>) carried out in the Tours metropolitan area. Although the scope of trips considered and the methods of estimating distances differ between the two surveys, the results are relatively close.

## 1 Introduction

Les données du recensement de la population (RP) permettent d'étudier jusqu'à un niveau territorial fin les déplacements des actifs pour se rendre sur leur lieu de travail. Cependant, la majorité des déplacements des Français en semaine est motivée par des raisons non liées au travail. Les trajets en semaine pour motifs liés au travail représentent 26 % des déplacements et 42 % des kilomètres parcourus en France métropolitaine en 2019 d'après l'enquête mobilité des personnes (EMP) produite par le Service des données et études statistiques (SDES).

L'EMP permet d'apprécier les motifs et les modes de déplacement des Français. Néanmoins l'échantillon des répondants ne permet pas de garantir une représentativité régionale. L'objectif de la démarche méthodologique présentée dans cet article est de pallier cette limite. Ces travaux s'inscrivent dans la continuité de ceux entrepris en 2012 (Le Guennec 2012 [1]) à partir de la précédente enquête nationale sur les Transports et Déplacements (SDES 2019 [2]).

Estimer les distances parcourues tous motifs de déplacements confondus par les individus résidant dans un territoire pour différents modes de transport permet d'établir un diagnostic aidant à guider

le développement de projets de services de mobilité, notamment pour renforcer les transports en commun et développer l'intermodalité. Ces résultats sont attendus par les acteurs publics régionaux compétents en matière de transports et d'aménagement du territoire.

La démarche proposée dans cet article est de construire un modèle prédictif grâce aux réponses de l'EMP afin d'estimer les distances et les nombres de déplacements à partir des enquêtes de recensement de la population.

## **2 L'enquête mobilité des personnes 2019**

### **2.1 Présentation de l'enquête mobilité des personnes**

L'enquête mobilité personnes (EMP) est produite par le Service des données et études statistiques (SDES). L'EMP permet d'apprécier les motifs et les modes de déplacement des Français.

Cette enquête a été conduite en face-à-face par des enquêteurs de l'Insee entre mai 2018 et avril 2019 auprès de 20 000 ménages de France métropolitaine. Un individu Kish par ménage est sélectionné afin de répondre à des questions pour un jour de la semaine sélectionnée. L'enquête permet d'obtenir des informations sur les déplacements, en particulier, leur nombre et le motif de chaque déplacement.

L'enquête a été conçue afin d'être représentative des individus de six ans ou plus au niveau national. Néanmoins l'échantillon des répondants ne permet pas de garantir une représentativité régionale. Par exemple, seulement 576 réponses ont été collectées en Centre-Val de Loire et ne peuvent pas représenter fidèlement les déplacements de plus de 2,5 millions d'habitants de la région.

### **2.2 Les variables à prédire**

L'objectif de la démarche est de produire des statistiques agrégées concernant les distances parcourues et le nombre de déplacements réalisés pour différents modes (train ; bus/car ; voiture ; moto ; vélo ; pied) et pour deux motifs de déplacement (tous motifs ; motifs non professionnels). L'indicateur principalement mis en avant dans cet article est la part modale en distance. Une part modale désigne la fraction de la distance parcourue pour un mode de transport donné divisée par la distance totale parcourue. Le nombre de déplacements des individus selon le mode de transport et le motif est également une information qui peut être déduite des résultats de l'EMP et donc estimé par le modèle construit.

Les variables prédictives ont été sélectionnées selon un critère de disponibilité simultanée dans l'EMP ainsi que dans la source du recensement de la population. On peut classer les variables prédictives selon quatre thèmes (Individu, Ménage, Logement, Géographie) (cf. annexe A).

Parmi les variables prédictives à disposition, la distance parcourue entre le domicile et le lieu de travail et le mode principal de transport pour aller au travail font partie de ceux dont le pouvoir de prédiction est vraisemblablement important, hormis pour les prédictions de variables dont le motif est non professionnel. C'est pourquoi, il a été décidé de prédire des distances et nombre de déplacements tous motifs et motifs professionnels et d'en déduire les distances parcourues pour des motifs non professionnels afin de bénéficier du pouvoir prédictif de variables fortement liées aux déplacements professionnels. Ainsi, 24 variables sont prédites simultanément pour chaque observation (cf. annexe A).

## 3 Préparation des données

### 3.1 Champ

Afin de réduire la variabilité des données et d'améliorer les performances en prédiction du modèle, on propose de restreindre le champ d'étude. Seuls les déplacements réalisés hors week-end et en Province sont considérés. Certains répondants à l'enquête EMP sont ainsi écartés :

- les répondants résidant en Île-de-France (2 142 observations) ;
- les répondants résidant en Corse (34 observations) ;
- les répondants qui ont été interrogés sur leur déplacement le samedi ou le dimanche (1 633 + 1 381 répondants).

Le choix d'écarter les répondants résidant en Île-de-France est cohérent avec le document de travail sur les estimations régionales dans l'enquête nationale transport et déplacements 2007-2008 (Le Guennec 2012 [1]). Dans ce document, les modèles avaient été réalisés sur deux champs géographiques distincts : Île-de-France et France de Province. De plus une enquête sur la sujet de la mobilité et dédiée à l'Île-de-France a été réalisée en 2022-2023 (Nguyen-Luong 2024 [3]). On conserve ainsi un échantillon de 8 708 répondants parmi les 13 825 répondants initiaux.

### 3.2 Retrait des valeurs atypiques

Certains répondants dont les distances de déplacement par mode sont éloignés de ce qui est observé (outliers) sont retirés du modèle. Deux méthodes ont été testées. La première basée sur l'écart interquartile : définition des valeurs atypiques si valeur supérieure à  $Q3 + 1,5 * IQR$  (écart interquartile). La seconde est une méthode de type z-score, avec un z égal à 3 comme seuil de détection.

La première méthode conduit à l'exclusion de 831 réponses rien que pour les distances tous motifs confondus, soit 9,5 % de l'échantillon. Compte tenu du nombre d'observations concernées, les outliers ont été sélectionnés à partir de la méthode du z-score. Cette méthode conduit à évincer moins de répondants (178 au regard uniquement des valeurs des distances tous motifs de déplacement). Après traitement des outliers pour chacune des 24 variables réponses, l'échantillon des répondants totalise 8 349 répondants au lieu de 8 708 (-4,1 %).

## 4 Modélisation Elastic Net

Afin de limiter le surapprentissage d'un modèle dans le cadre de prédictions, la famille des modèles de régression pénalisée ("regularized regression") est envisagée.

Les techniques de régression Lasso et Ridge sont fréquemment employées dans la cadre des régressions pénalisées. La régression Ridge est efficace dans les situations où les prédicteurs sont corrélés (multicolinéarité). La régression Lasso a tendance à effectuer une sélection automatique des variables, simplifiant le modèle. Pour combiner les avantages des deux modèles, un modèle de type Elastic Net est envisagé.

### 4.1 Généralités sur les modèles Elastic Net

Les éléments de cette partie sont issus de la documentation du package *glmnet* (Friedman 2023 [4]). Dans le modèle de régression avec plusieurs variables réponses ("multivariate regression" ou "multivariate multiple regression"), la variable réponse  $y$  n'est plus un vecteur mais une matrice de

valeur numérique (nombres de déplacements ; distances). Dans ce modèle, on suppose avoir  $N$  observations de  $K$  variables à prédire  $y_i \in \mathbb{R}^K$  et  $p$  variables prédictives  $x_i \in \mathbb{R}^p$ .

L'équation du modèle est :

$$\mathbf{Y} = \beta_0 + \mathbf{X}\beta + \epsilon$$

avec

- $\beta$  : la matrice des coefficients ( $p; K$ )
- $\epsilon$  : la matrice des termes d'erreur ( $N; K$ )

La fonction objectif à minimiser afin d'estimer les coefficients  $\beta$  est la suivante :

$$\min_{(\beta_0, \beta) \in \mathbb{R}^{(p+1) \times K}} \frac{1}{2N} \sum_{i=1}^N \|y_i - \beta_0 - \beta^T x_i\|_F^2 + \lambda \left[ \frac{(1-\alpha)}{2} \|\beta\|_F^2 + \alpha \sum_{j=1}^p \|\beta_j\|_2 \right]$$

Où :

- $\lambda \geq 0$  est le coefficient de pénalité
- $0 \leq \alpha < 1$  est un paramètre de mélange. Il s'agit de réaliser un compromis entre une régression Lasso ( $\alpha = 1$ ) et une régression Ridge ( $\alpha = 0$ )
- $\|\beta\|_F$  est la norme de Frobenius de la matrice  $\beta$  ( $\|\beta\|_F = \sqrt{\sum_{i,j} \beta_{ij}^2}$ )
- $\|\beta_j\|_2$  est la norme  $L_2$  de la  $j$ -ème ligne de la matrice des coefficients  $\|\beta_j\|_2 = \sqrt{\sum_{i=1}^K \beta_{ji}^2}$

Les variables explicatives (prédicteurs) sont centrées et réduites pour éviter les inégalités de traitement entre coefficient (Rakotomalala 2020 [5]). Le package *glmnet* "standardise" automatiquement les prédicteurs avant de résoudre le problème d'optimisation du modèle.

## 4.2 Choix de la valeur de pénalisation $\lambda$

Par défaut le package *glmnet* résout le problème d'optimisation en considérant jusqu'à 100 valeurs de  $\lambda$ . Si la part de la déviance expliquée par le modèle évolue peu entre une valeur de  $\lambda$  est la suivante, le processus est arrêté prématurément.

Dans cette partie, on fixe l'hyperparamètre  $\alpha$ . On cherchera à optimiser ce paramètre dans un deuxième temps.

### 4.2.1 Estimation de l'erreur de prédiction

Afin d'estimer l'erreur en prédiction, l'erreur quadratique moyenne entre les valeurs prédites et les valeurs mesurées dans l'enquête est calculée. La capacité prédictive du modèle est estimée à partir de la moyenne des erreurs quadratiques moyennes calculées pour chaque variable prédite. Les valeurs des variables prédites sont standardisées afin de ne pas accorder plus d'importance aux distances prédites par rapport aux nombres de déplacements prédits.

### 4.2.2 Estimation par validation croisée

Pour chaque valeur de  $\lambda$ , l'échantillon d'apprentissage est scindée en deux. Les coefficients  $\beta$  sont estimés à partir du premier échantillon puis l'erreur de prédiction en validation croisée (par défaut l'erreur quadratique moyenne) est estimée à partir du second. Par défaut, dans le package *glmnet*, une 10-folds validation croisée est réalisée.

La valeur de  $\lambda$  qui minimise la moyenne des erreurs de prédiction en validation croisée est notée  $\lambda.min$ . Du fait de la variabilité des erreurs en validation croisée qui diffèrent selon les échantillons, on peut vouloir être plus parcimonieux et choisir un modèle plus pénalisé donc plus simple et *a priori* avec moins de surapprentissage que le modèle associé à la valeur  $\lambda.min$ . Ainsi, la littérature (Tuffery 2024 [6]) suggère la prise en compte d'une autre valeur de  $\lambda$  (notée  $\lambda.1se$ ) qui correspond à la plus grande valeur de  $\lambda$  pour laquelle l'erreur de prédiction moyenne se situe à moins d'un écart type de l'erreur associée à  $\lambda.min$  (cf. figure 1). Cependant, dans le modèle étudié, cette valeur de  $\lambda.1se$  est trop importante si bien que peu de prédicteurs auraient été conservés si cette valeur avait été choisie pour réaliser les prédictions.

Dans le cadre du problème étudié, il a été choisi de considérer les valeurs de  $\lambda.min$  et  $\lambda.1se$  comme des bornes et de sélectionner une valeur de  $\lambda$  compris dans l'intervalle délimité par ces deux valeurs caractéristiques. Le choix de la valeur  $\lambda$  utilisée pour réaliser les prédictions sera discuté plus loin.

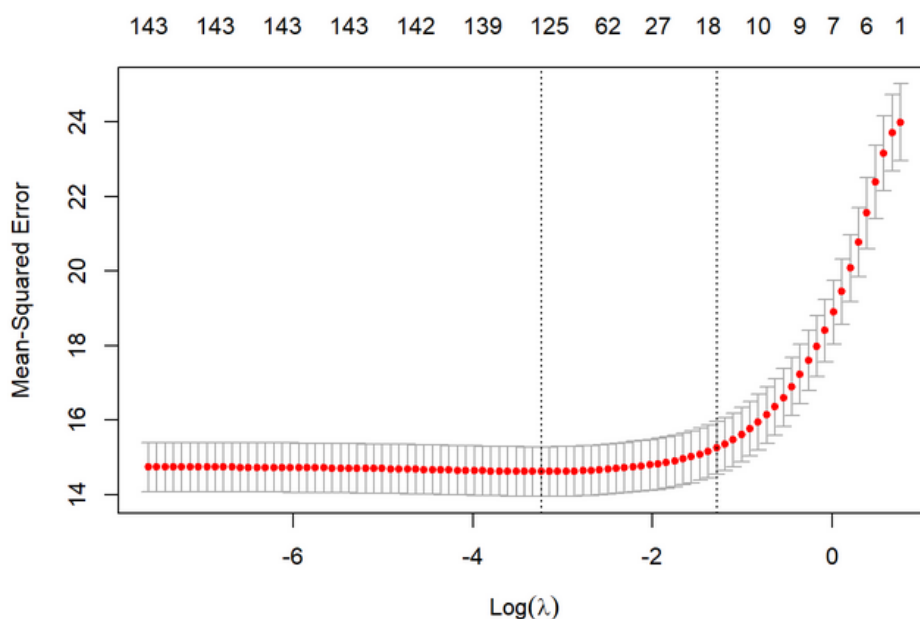


FIG. 1 – Évolution de l'erreur quadratique moyenne en fonction de  $\lambda$   
 Note : La ligne en pointillé à gauche matérialise la valeur de  $\lambda$  qui minimise l'erreur quadratique moyenne, celle de droite correspond à  $\lambda.1se$

### 4.3 Choix du paramètre de mélange $\alpha$

Nous venons d'étudier comment procéder pour identifier des valeurs caractéristiques de  $\lambda$ . La valeur de  $\alpha$  est déterminée de la manière suivante. Pour différentes valeurs de  $\alpha$  fixées, les valeurs minimales de l'erreur quadratique moyenne obtenu à partir de  $\lambda.min$  sont comparées. La valeur de  $\alpha$  qui minimise le critère du MSE est retenue.

Plus précisément, le choix de  $\alpha$  est effectué pour des valeurs comprises entre 0 et 1 avec un pas de 0,1. Pour chacune des valeurs de  $\alpha$ , une valeur de  $\lambda$  qui minimise l'erreur de prédiction sur l'échantillon de validation est déterminée. Au final, le paramètre  $\alpha$  qui minimisera l'erreur quadratique moyenne sur l'échantillon de prédiction (cf. figure 2) est choisie.

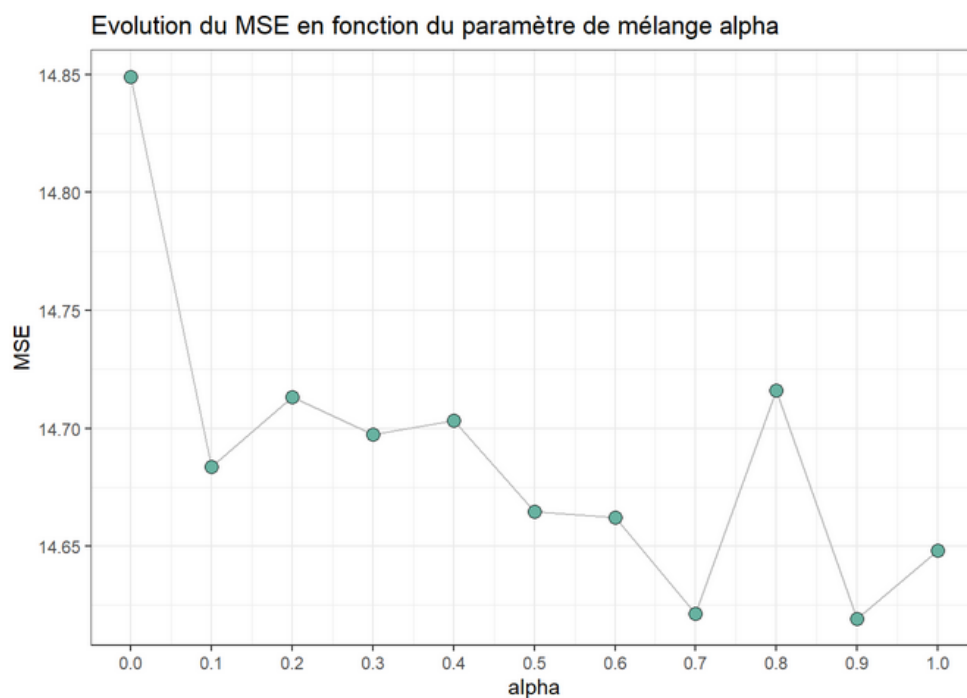


FIG. 2 – Évolution de l'erreur quadratique moyenne en fonction du paramètre de mélange  $\alpha$

Le résultat obtenu est soumis à des aléas. Si aucune graine n'est spécifiée, les erreurs quadratiques minimales à  $\alpha$  fixé diffèrent selon les individus tirés dans les échantillons. Afin d'améliorer la robustesse, le processus d'estimation de l'erreur quadratique moyenne est répétée arbitrairement 50 fois. L'évolution de l'erreur quadratique en fonction de la valeur de  $\alpha$  est représentée figure 3. Contrairement à la représentation précédente, on constate que la valeur minimale de prédiction est obtenue pour  $\alpha = 1$ .

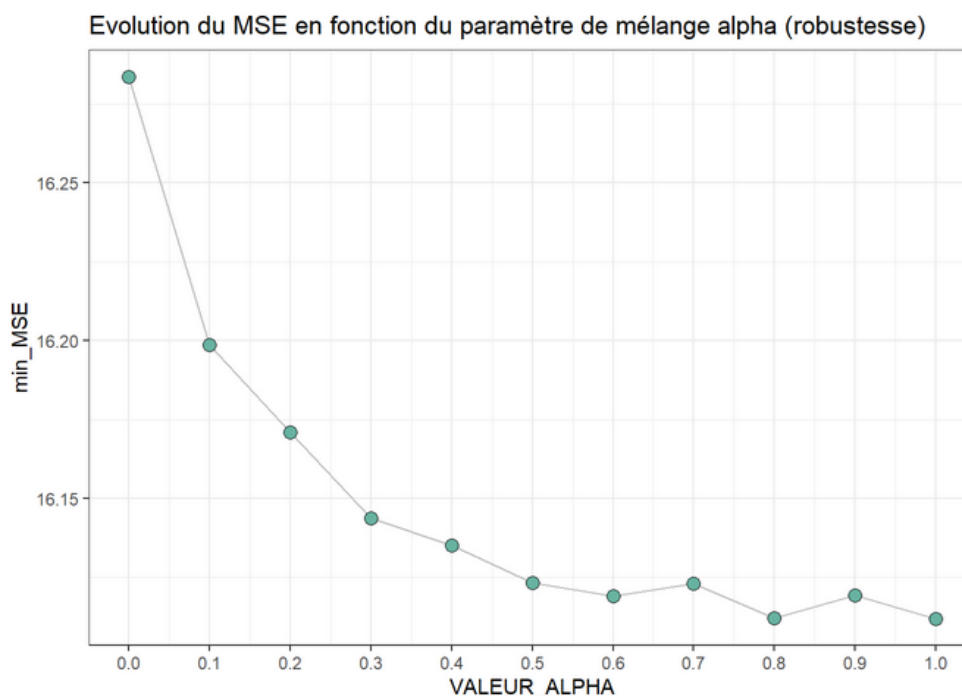


FIG. 3 – Évolution de l’erreur quadratique moyenne en fonction du paramètre de mélange  $\alpha$  - Valeur moyenne après 50 estimations

## 5 Plage des estimations

Dans cette partie, une méthode d’évaluation de la performance du modèle est présentée.

### 5.1 Principe

Afin de vérifier la capacité du modèle à estimer des parts modales, une méthode de rééchantillonnage (bootstrap) a été implémentée. Elle calcule l’écart entre la part modale estimée et la part modale observée sur un sous-échantillon. Ce calcul répliqué 1000 fois permet d’estimer les bornes d’une plage de 95 % des estimations pour chacune des parts modales. La méthode présentée est basée sur l’algorithme Bootstrap Bias Corrected Cross-Validation (BBC-CV) (Tsamardinos 2019 [7]).

#### Étape 1 : Construction d’un maillage de l’espace des hyperparamètres

La première étape consiste à construire un ensemble de couples  $\alpha$  et  $\lambda$  différents qui “quadrille” l’espace des hyperparamètres. Un modèle sera construit pour chacun des couples. Les valeurs de  $\alpha$  sont comprises entre 0 et 1 avec un pas de 0,1. Les valeurs de  $\lambda$  sont déterminés comme suit : pour chaque valeur de  $\alpha$  fixé, le modèle estime l’erreur de prédiction en validation croisée pour différentes valeurs de  $\lambda$ . Les valeurs de  $\lambda$  considérées sont comprises entre  $\lambda.min$  (valeur qui minimise l’erreur de prédiction) et  $\lambda.1se$  (le plus grande valeur de  $\lambda$  pour laquelle l’erreur de prédiction moyenne se situe à moins d’un écart type de l’erreur associée à  $\lambda.min$ ). Neuf valeurs de  $\lambda$  réparties uniformément entre  $\lambda.min$  et  $\lambda.1se$  sont sélectionnées. Nous avons donc onze valeurs de  $\lambda$  considérées pour chacune des onze valeurs de  $\alpha$  choisies, ce qui représente 121 configurations d’hyperparamètres.

#### Étape 2 : Prédictions individuelles pour chaque configuration d’hyperparamètres

Dès lors que les différentes configurations d’hyperparamètres sont déterminées, une prédiction individuelle de chaque variable à prédire est réalisée pour chaque couple d’hyperparamètres. Les

prédictions sont construites à l'aide d'une validation croisée (10-folds). Ainsi, à couple d'hyperparamètres  $(\alpha, \lambda)$  fixé, dix modèles sont estimés à partir de 90 % de l'échantillon et les prédictions sont réalisées pour le restant de l'échantillon. On obtient ainsi une matrice contenant 8348 lignes (taille de l'échantillon) et 2904 colonnes (nombre de configurations d'hyperparamètres (121) x nombre de variables à prédire (24)).

### Etape 3 : Construction des échantillons bootstrap et estimation du meilleur couple d'hyperparamètres pour chaque échantillon

Soit B le nombre d'échantillons bootstrap considérés. A partir de la matrice calculée à l'étape 2, B échantillons de 8348 individus sont générés par tirage aléatoire avec remise. Pour chaque échantillon bootstrap, un jeu d'hyperparamètres optimal est déterminé : il s'agit du couple  $\alpha$  et  $\lambda$  qui minimise l'erreur quadratique moyenne des prédictions de l'échantillon bootstrap. L'erreur quadratique moyenne est calculée simplement à partir des valeurs sauvegardées dans la matrice. Une valeur de 1 000 a été choisie pour B.

### Etape 4 : Performance des prédictions sur les échantillons out-of-sample

Pour chaque échantillon bootstrap considéré, un échantillon out-of-sample est associé. Il est défini comme l'ensemble des observations non tirées au sort dans l'échantillon bootstrap. La performance de la prédiction est mesurée à partir des prédictions sur l'échantillon out-of-sample. Pour chaque échantillon bootstrap, les parts modales sont calculées à partir des prédictions des observations de l'échantillon out-of-sample. L'écart entre les parts modales calculées à partir des prédictions et les parts modales déterminées à partir des réponses des enquêtés est mesuré pour chaque échantillon out-of-sample. On obtient ainsi B mesures de l'écart entre les parts modales prédites et calculées à partir de l'enquête EMP. Afin d'estimer une plage de ces écarts, les 2,5 % des écarts les plus faibles et les 2,5 % les plus élevés sont exclus et on détermine un intervalle de l'erreur absolue des prédictions à partir des valeurs extrêmes de 95 % des écarts pour les parts modales de chaque mode de déplacement.

## 5.2 Discussion

La méthode BBC-CV présente l'avantage d'être parcimonieux sur certains calculs. Le calcul des prédictions est réalisé une seule fois et est stocké dans une matrice. Cette matrice contient les informations nécessaires afin de déterminer l'erreur quadratique moyenne selon l'échantillon bootstrap tiré.

| Valeurs hyperparamètres                       | Nombre de configurations optimales parmi 1000 rééchantillonnages |
|---|--|
| alpha = 0,4   lambda min                      | 3  |
| alpha = 0,5   lambda min                      | 18   |
| alpha = 0,5   0,9*lambda min + 0,1*lambda 1se | 8  |
| alpha = 0,6   lambda min                      | 8  |
| alpha = 0,7   lambda min                      | 37   |
| alpha = 0,8   lambda min                      | 67   |
| alpha = 0,9   lambda min                      | 31   |
| alpha = 1,0   lambda min                      | 828  |

FIG. 4 – Plage des estimations pour les parts modales motifs non liés au travail

L'utilisation de cette méthode de rééchantillonnage conduit à une estimation de la précision des prédictions. Elle offre également une indication sur le choix du couple d'hyperparamètres  $(\alpha, \lambda)$  à privilégier pour effectuer les prédictions. En effet, l'évaluation 1 000 fois de l'erreur quadratique moyenne des prédictions sur différents échantillons permet d'exhiber les valeurs d'hyperparamètres optimaux choisis pour chaque rééchantillonnage (cf. figure 4).

On remarque que la valeur de  $\lambda = \lambda.min$  minimise l'erreur 992 fois sur 1000, renforçant l'intuition que la valeur de  $\lambda.1se$  pénalise trop le modèle pour obtenir des prédictions satisfaisantes. La valeur  $\alpha = 1$  est privilégiée plus de huit fois sur dix, confirmant qu'un modèle de type Lasso est plus adapté au problème. Cette vérification à partir des résultats du rééchantillonnage consolide également une démarche présentée précédemment : plusieurs réestimations du MSE en fonction de  $\alpha$  est plus prudente (cf. figure 3) qu'une estimation unique (cf. figure 2).

### 5.3 Résultats

On remarque que les plages des estimations sont les plus importantes pour les déplacements en voiture et en train, quelles que soient les motivations des déplacements (cf. figures 5, 6, 7). Cependant, les ordres de grandeur des parts modales ne sont pas les mêmes pour ces deux modes de transport. La part des distances parcourues en voiture étant de l'ordre de 80 %, l'erreur relative des estimations est de l'ordre de 2-3 %. En revanche, l'erreur relative pour les estimations des déplacements en train peut atteindre 50 %, voire plus pour les déplacements non liés au travail.

Par ailleurs, la plage des estimations des parts modales est plus importante pour les déplacements liés au travail par rapport aux déplacements tous motifs et ceux non liés au travail (environ +3 points de pourcentage). Cela peut paraître étonnant compte tenu de la présence de certains prédicteurs dans le modèle qui semblent fortement corrélés aux déplacements liés au travail : distance domicile-travail et mode de transport principal pour aller au travail. Le pouvoir prédictif de cette dernière variable est néanmoins restreint par l'agrégation des modes de transport 'train' et 'bus' dans le questionnaire du recensement de la population (présence de la modalité 'transport en commun' dans le questionnaire 2019).

| Mode de transport | Quantile 2.5% | Quantile 97.5% |
|-------------------|---------------|----------------|
| Voiture           | -2.2          | 2.3            |
| Train             | -2.5          | 2.2            |
| Bus/Car           | -0.6          | 0.5            |
| Moto              | -0.3          | 0.2            |
| Vélo              | -0.2          | 0.1            |
| Marche            | -0.2          | 0.1            |

**Lecture :** Parmi 1 000 rééchantillonnages, 95 % des parts modales en voiture tous motifs de déplacement prédites ont un écart compris entre -2,2 et 2,3 points de pourcentage par rapport aux parts modales calculées à partir des réponses à l'enquête EMP.

**Champ :** Individus âgés de 6 ans ou plus résidant en France métropolitaine hors Île-de-France et Corse. Déplacements réalisés hors week-end.

**Sources :** SDES, enquête mobilité des personnes 2018-2019.

FIG. 5 – *Plage des estimations pour les parts modales tous motifs*

| Mode de transport | Quantile 2.5% | Quantile 97.5% |
|-------------------|---------------|----------------|
| Voiture           | -3.8          | 3.6            |
| Train             | -4.0          | 4.1            |
| Bus/Car           | -0.5          | 0.5            |
| Moto              | -0.4          | 0.3            |
| Vélo              | -0.2          | 0.1            |
| Marche            | -0.1          | 0.1            |

**Lecture :** Parmi 1 000 rééchantillonnages, 95 % des parts modales en voiture tous motifs de déplacement prédites ont un écart compris entre -3,8 et 3,6 points de pourcentage par rapport aux parts modales calculées à partir des réponses à l'enquête EMP.

**Champ :** Individus âgés de 6 ans ou plus résidant en France métropolitaine hors Île-de-France et Corse. Déplacements réalisés hors week-end.

**Sources :** SDES, enquête mobilité des personnes 2018-2019.

FIG. 6 – *Plage des estimations pour les parts modales motifs liés au travail*

| Mode de transport | Quantile 2.5% | Quantile 97.5% |
|-------------------|---------------|----------------|
| Voiture           | -2.4          | 2.8            |
| Train             | -3.1          | 2.6            |
| Bus/Car           | -0.8          | 0.8            |
| Moto              | -0.3          | 0.3            |
| Vélo              | -0.3          | 0.2            |
| Marche            | -0.3          | 0.2            |

**Lecture** : Parmi 1 000 rééchantillonnages, 95 % des parts modales en voiture pour les déplacements non liés au travail prédites ont un écart compris entre -2,4 et 2,8 points de pourcentage par rapport aux parts modales calculées à partir des réponses à l'enquête EMP.

**Champ** : Individus âgés de 6 ans ou plus résidant en France métropolitaine hors Île-de-France et Corse. Déplacements réalisés hors week-end.

**Sources** : SDES, enquête mobilité des personnes 2018-2019.

FIG. 7 – Plage des estimations pour les parts modales motifs non liés au travail

## 6 Adaptabilité du modèle

Dans la partie précédente, nous avons cherché à estimer l'écart entre les parts modales prédites et celles mesurées sur différents échantillons. Cependant, ces erreurs ont été mesurées sur des populations tirées de manière aléatoire à partir d'un échantillon de répondants sur l'ensemble du territoire. Les comportements des Français en matière de déplacements ne sont pas uniformes selon leur lieu de résidence. Par exemple, la proximité ou non de services ou encore la présence d'équipements (gares, pistes cyclables) influent sur les distances parcourues et les recours aux divers modes de transport.

Afin de vérifier si le modèle bâti s'adapte à ces différences de comportement, les résultats des prédictions sont comparés à des statistiques estimées sur des populations localisées. Ces données sont tirées du recensement de la population ou encore d'enquêtes menées localement, par exemple les enquêtes Enquêtes Mobilité Certifiées Cerema (EMC<sup>2</sup>) labélisées par le Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement (Cerema)

### 6.1 Les autorités organisatrices de la mobilité (AOM) et les bassins de mobilité

Les autorités organisatrices de la mobilité (AOM) assurent l'organisation du réseau de transport urbain sur son territoire. Ces territoires correspondent à des établissements publics de coopération intercommunale (EPCI) lorsque celui-ci décide de ne pas déléguer la compétence à la région. "L'AOM dispose de la compétence d'organisation de la mobilité et l'exerce en choisissant de mettre en place les services adaptés aux besoins des habitants. Elle peut par exemple décider, au regard des



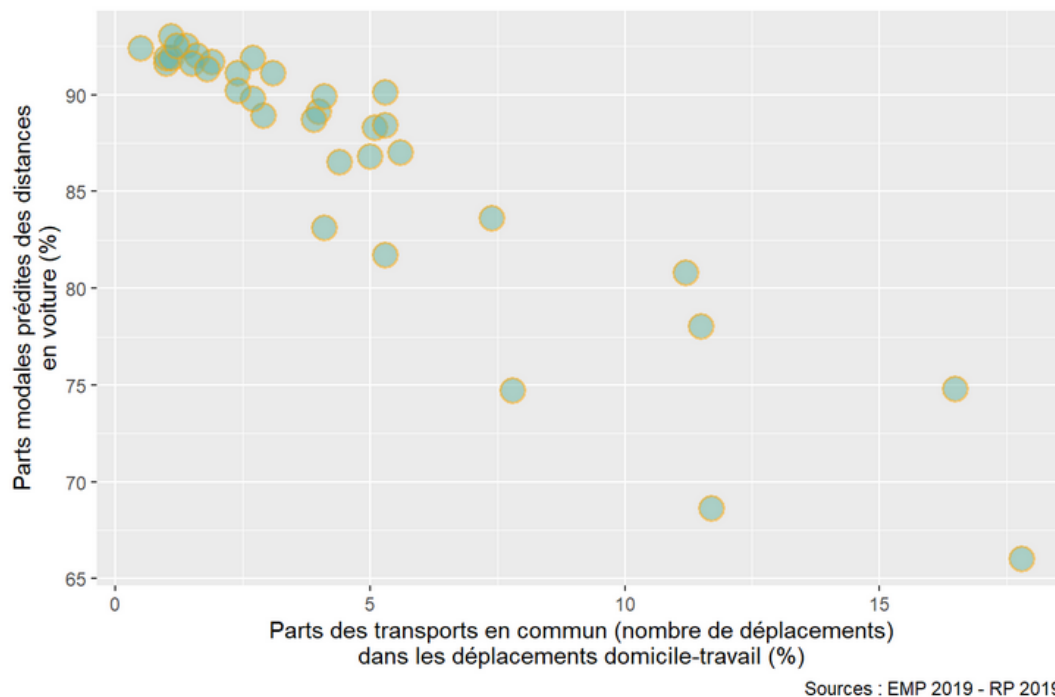


FIG. 9 – Représentation de la part modale prédite des distances en voiture en fonction de la part des déplacements réalisés en transport en commun pour chaque bassin de mobilité de la région Centre-Val de Loire

travailler, moins la part modale de la voiture prédite est importante.

Néanmoins, ce résultat souffre d'une limitation méthodologique. En effet, le mode de transport principal utilisé pour aller au travail est une information issue du recensement de la population qui sert à calculer les parts modales pour les déplacements domicile-travail (axe des abscisses) mais qui est aussi un des prédicteurs du modèle permettant d'estimer les parts modales tous motifs (axe des ordonnées). On peut donc légitimement mettre en doute le bien-fondé du résultat, bien que le modèle comporte bien d'autres prédicteurs et que la relation observée figure 9 ne pouvait pas être considérée comme totalement prévisible.

## 6.2 L'enquête mobilité certifiée Cerema (EMC<sup>2</sup>)

Afin de vérifier l'adaptabilité du modèle à partir de données exogènes à la modélisation, les parts modales estimées ont été comparées aux résultats obtenus par une enquête locale dédiée à la mobilité dans l'agglomération tourangelle réalisée la même année que l'enquête EMP, le dispositif EMC<sup>2</sup> – enquête mobilité certifiée Cerema. Les résultats de cette enquête ne sont pas prises en compte dans la construction du modèle prédictif.

L'enquête EMC<sup>2</sup> a été commanditée par le Syndicat des Mobilités de Touraine et le SMAT (Syndicat Mixte de l'Agglomération Tourangelle) en 2019. Cette enquête mesure la mobilité des habitants d'Indre-et-Loire (périmètre d'enquête) ainsi que sur des périmètres plus restreints au sein de l'agglomération tourangelle grâce aux réponses de 8 700 personnes de 5 ans ou plus. Afin de mesurer la capacité d'adaptation du modèle, les prédictions sont comparées aux résultats de l'enquête EMC<sup>2</sup>.

## 6.2.1 Résultats EMC<sup>2</sup> sur le département d'Indre-et-Loire et l'agglomération tourangelle

| Mode de transport   | Parts déplacements tous motifs prédites | Parts déplacements tous motifs EMC <sup>2</sup> |
|---------------------|---|---|
| Voiture             | 83.7                                    | 81.5  |
| Transport en commun | 12.9                                    | 13.7  |
| Marche              | 1.8                                     | 2.8   |
| Moto                | 0.9                                     | 0.8   |
| Vélo                | 0.7                                     | 1.0   |

**Lecture:** En 2019, 83,7 % de la distance parcourue lors de déplacements est réalisée en voiture. D'après l'enquête mobilité EMC<sup>2</sup> de Touraine, cette même part est estimée à 81,5 %.  
**Champ :** Individus âgés de 6 ans ou plus résidant en Indre-et-Loire.  
**Sources :** SDES, enquête mobilité des personnes 2018-2019 ; EMC<sup>2</sup> de Touraine 2019 ; Insee, Recensement de la population 2019, exploitation complémentaire.

FIG. 10 – Comparaison des parts modales tous motifs en Indre-et-Loire prédites et estimées par l'enquête EMC<sup>2</sup>

Concernant les distances parcourues en transport en commun, il n'est pas possible de distinguer la part parcourue en bus/car d'un côté et en train de l'autre. Dans l'enquête EMC<sup>2</sup>, les déplacements en transport en commun sont regroupés selon trois modalités : "Transport collectifs urbains" ; "Transports collectifs interurbains" ; "Autres transports collectifs" (cf. annexe B).

Les écarts absolus entre les parts modales prédites et celles diffusées par l'enquête EMC<sup>2</sup> Touraine diffèrent au maximum de 2,2 points pour le département de l'Indre-et-Loire (cf. figure 10) et de 1,7 points pour l'agglomération tourangelle (cf. figure 11).

| Mode de transport   | Parts déplacements tous motifs prédites | Parts déplacements tous motifs EMC <sup>2</sup> |
|---------------------|---|---|
| Voiture             | 78.3                                    | 78.6  |
| Transport en commun | 17.3                                    | 15.6  |
| Marche              | 2.4                                     | 3.8   |
| Moto                | 1.0                                     | 1.4   |
| Vélo                | 1.0                                     | 0.6   |

**Lecture:** En 2019, 78,3 % de la distance parcourue lors de déplacements est réalisée en voiture. D'après l'enquête mobilité EMC<sup>2</sup> de Touraine, cette même part est estimée à 78,6 %.  
**Champ :** Individus âgés de 6 ans ou plus résidant dans l'agglomération tourangelle (SMAT).  
**Sources :** SDES, enquête mobilité des personnes 2018-2019 ; EMC<sup>2</sup> de Touraine 2019 ; Insee, Recensement de la population 2019, exploitation complémentaire.

FIG. 11 – Comparaison des parts modales tous motifs dans l'agglomération tourangelle prédites et estimées par l'enquête EMC<sup>2</sup>

## 6.2.2 Discussion sur la comparaison avec les résultats de l'EMC<sup>2</sup>

Plusieurs éléments peuvent expliquer partiellement les écarts entre les estimations issues du modèle et celles de l'enquête EMC<sup>2</sup> :

- Le mode de calcul des distances est différente entre l'enquête EMC<sup>2</sup> et l'EMP. Pour l'EMC<sup>2</sup>, le calcul des distances est effectué a posteriori par le Cerema. Pour chaque trajet, une distance à vol d'oiseau est calculée, ainsi qu'une distance parcourue – totale et interne au périmètre d'enquête. Le calcul est différent selon le mode de transport :
  1. pour la marche, une vitesse constante de 4 km/h est supposée ;

2. pour les transports collectifs, la distance à vol d'oiseau est multipliée par un facteur d'1,5;
  3. pour la voiture et le vélo, les trajets sont projetés sur un réseau routier adapté au zonage fin.
- Dans le cas des déplacements multimodaux (plusieurs modes de transport utilisés pour un même déplacement), les enquêtes affectent un unique mode de transport au déplacement selon une hiérarchie des modes de transport. Cette hiérarchie est différente entre les enquêtes EMP et EMC<sup>2</sup>.
  - Pour certains résultats publiés à partir de l'enquête EMC<sup>2</sup>, seuls les kilomètres parcourus au sein du périmètre d'enquête sont comptabilisés (cf. figure 12).
  - Les modes de collecte des informations des deux enquêtes sont différents : EMP conduite en face-à-face alors que EMC<sup>2</sup> conduite en face-à-face et téléphone ; énoncés des questions différents...

Néanmoins, cette comparaison permet de vérifier que le modèle s'adapte relativement bien aux disparités locales mesurées par l'enquête EMC<sup>2</sup> dans l'agglomération tourangelle.

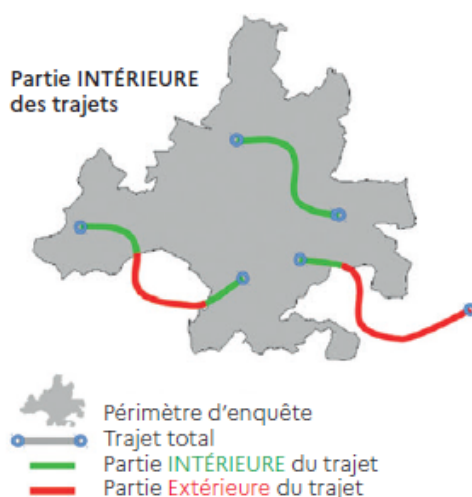


FIG. 12 – Distinction trajet à l'intérieur et à l'extérieur de la zone d'enquête (source : CEREMA)

## Références

- [1] J. L. Guennec, “Estimations régionales dans l’enquête nationale transport et déplacements 2007-2008,” Document de travail H2012-03, INSEE, 2012. Consulté le 5 août 2024.
- [2] SDES, “Résultats détaillés de l’enquête mobilité des personnes de 2019.” <https://www.statistiques.developpement-durable.gouv.fr/resultats-detaillés-de-lenquete-mobilite-des-personnes-de-2019>, 2021. Accédé en août 2025.
- [3] D. Nguyen-Luong, “Enquête régionale sur la mobilité des franciliens.” <https://www.institutparisregion.fr/mobilite-et-transport/deplacements/enquete-regionale-sur-la-mobilite-des-franciliens/>, 2024. Accédé en août 2025.
- [4] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent.” <https://glmnet.stanford.edu/articles/glmnet.html#linear-regression-family-gaussian-default>, 2023. Accédé en août 2024.
- [5] R. Rakotomalala, “Régression régularisée,” 2020. Slides de cours, Université Lumière Lyon 2.
- [6] S. Tufféry, *Modélisation prédictive et apprentissage statistique avec R*, vol. 3ème édition. 2024.
- [7] G. E. . B. G. Tsamardinos, I., “Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation,” *Mach Learn*, vol. 107, pp. 1895–1922, 2018.
- [8] collectivites locales.gouv.fr, “La mobilité.” <https://www.collectivites-locales.gouv.fr/competences/la-mobilite>, 2025. Accédé en octobre 2025.
- [9] T. T. J. O. L. E. Ménard Boris, Movellan Jean-Baptiste, “En Centre-Val de Loire, près de 70 % des trajets courts pour aller au travail sont effectués en voiture,” *Insee Centre-Val de Loire*, vol. 116, 2024.

## **A Liste des variables**

### **A.1 Variables à prédire**

#### **A.1.1 Distances**

- Distance parcourue en train tous motifs
- Distance parcourue en bus/car tous motifs
- Distance parcourue en voiture tous motifs
- Distance parcourue en moto tous motifs
- Distance parcourue en vélo tous motifs
- Distance parcourue à pied tous motifs
- Distance parcourue en train motifs professionnels
- Distance parcourue en bus/car motifs professionnels
- Distance parcourue en voiture motifs professionnels
- Distance parcourue en moto motifs professionnels
- Distance parcourue en vélo motifs professionnels
- Distance parcourue à pied motifs professionnels

#### **A.1.2 Nombre de déplacements**

- Nombre de déplacements en train tous motifs
- Nombre de déplacements en bus/car tous motifs
- Nombre de déplacements en voiture tous motifs
- Nombre de déplacements en moto tous motifs
- Nombre de déplacements en vélo tous motifs
- Nombre de déplacements à pied tous motifs
- Nombre de déplacements en train motifs professionnels
- Nombre de déplacements en bus/car motifs professionnels
- Nombre de déplacements en voiture motifs professionnels
- Nombre de déplacements en moto motifs professionnels
- Nombre de déplacements en vélo motifs professionnels
- Nombre de déplacements à pied motifs professionnels

### **A.2 Variables prédictives**

#### **A.2.1 Personnes enquêtées**

- Sexe
- Âge
- Âge au carré
- Indicateur de nationalité
- Diplôme le plus élevé
- Catégorie socio-professionnelle
- Situation principale

- Existence d'une activité antérieure
- Recherche d'un emploi
- Temps de travail (complet/partiel)
- Type emploi (cdd/cdd/interim/stage)
- CS de la mère
- CS du père
- CS du conjoint

### **A.2.2 Ménage**

- Nombre de personnes du ménage
- Nombre d'enfants dans le ménage
- Nombre d'actifs dans le ménage
- Type de ménage
- Vie en couple au sens du RP
- État matrimonial

### **A.2.3 Logement**

- Statut d'occupation (propriétaire/locataire. . .)
- Appartenance à un organisme HLM
- Nombre de voitures particulières, de véhicules utilitaires légers ou de campingcars à disposition du ménage

### **A.2.4 Géographie**

- Distance routière domicile – travail
- Distance routière domicile – études
- Mode de transport principal pour se rendre au travail
- Typologie rural/urbain de la commune de résidence
- Taille de la tranche d'unité urbaine de la commune de résidence
- Typologie rural/urbain de la commune de lieu de travail
- Taille de la tranche d'unité urbaine de la commune de lieu de travail

## B EMC<sup>2</sup> - Agrégation modes transport

Les parts modales calculées à partir des résultats de l'enquête EMC<sup>2</sup> ne sont pas diffusées selon les mêmes modes de transport que ceux proposés dans ce document. En particulier, la part des distances parcourues en train et en bus n'est pas diffusées dans les résultats de l'EMC<sup>2</sup>. Cependant, l'incertitude entre les modes de transport 'train' et 'bus/car' n'existe que pour les déplacements étiquetés 'transport collectifs interurbains'. En effet, ce regroupement agrège des transports en TER et également des transports interurbains routiers (TER routiers, lignes scolaires) ainsi que certains transports en autocar.

| libelle   | modCat1   | modCat2               | modCat3                 | modCEREMA          |
|---|-----------|-----------------------|-------------------------|--------------------|
| Marche à pied UNIQUEMENT  | actif     | marche                | marche                  | marche             |
| Conducteur Vélo Libre Service   | actif     | velo                  | velo                    | velo               |
| Conducteur de vélo classique  | actif     | velo                  | velo                    | velo               |
| Passager de vélo  | actif     | velo                  | velo                    | velo               |
| Conducteur de deux ou trois roues motorisés < 50 cm3  | motorise  | deuxroues             | deuxroues               | deuxroues          |
| Passager de deux ou trois roues motorisés < 50 cm3  | motorise  | deuxroues             | deuxroues               | deuxroues          |
| Conducteur de deux ou trois roues motorisés >= 50 cm3   | motorise  | deuxroues             | deuxroues               | deuxroues          |
| Passager de deux ou trois roues motorisés >= 50 cm3   | motorise  | deuxroues             | deuxroues               | deuxroues          |
| Conducteur Vélo Assistance électrique   | actif     | velo                  | velo                    | velo               |
| Conducteur Vélo Assistance électrique en Libre Service  | actif     | velo                  | velo                    | velo               |
| Conducteur de deux ou trois roues motorisés en libre-service  | motorise  | deuxroues             | deuxroues               | deuxroues          |
| Conducteur de véhicule particulier (VP)   | motorise  | voiture               | voiture conducteur      | voiture conducteur |
| Passager de véhicule particulier (VP)   | motorise  | voiture               | voiture passagers       | voiture passagers  |
| Passager bus urbain (réseau ville centre)   | collectif | collectif urbain      | filbleu                 | reseau urbain      |
| Passager tramway (réseau ville centre)  | collectif | collectif urbain      | filbleu                 | reseau urbain      |
| Passager bus à Haut Niveau de Service   | collectif | collectif urbain      | filbleu                 | reseau urbain      |
| TAD   | collectif | collectif urbain      | filbleu                 | reseau urbain      |
| Passager bus urbain autres villes du département  | collectif | collectif urbain      | collectif urbain autres | reseau urbain      |
| Passager autre réseau urbain  | collectif | collectif urbain      | collectif urbain autres | autres             |
| Passager transports interurbains routiers (TER routiers, lignes régulières départementales, scolaires)  | collectif | collectif interurbain | collectif interurbain   | autres tc          |
| Passager autres autocars (périscolaire, occasionnel, lignes internationales en cabotage , cars longues distance (Eurolines/Isilines, Ouibus, Flixbus...). | collectif | collectif interurbain | collectif interurbain   | autres tc          |
| Passager TGV  | collectif | collectif autres      | collectif autres        | autres tc          |
| Passager train TER  | collectif | collectif interurbain | collectif interurbain   | autres tc          |
| Passager Autres trains (Intercité, TET...)  | collectif | collectif autres      | collectif autres        | autres tc          |
| Passager taxi   | motorise  | voiture               | voiture passagers       | autres             |
| Passager VTC  | motorise  | voiture               | voiture passagers       | autres             |
| Transport employeur (exclusivement)   | collectif | collectif autres      | collectif autres        | autres tc          |
| Conducteur de fourgon, camionnette, camion (pour tournées professionnelles ou déplacements privés)  | motorise  | voiture               | voiture conducteur      | autres             |
| Passager de fourgon, camionnette, camion (pour tournées professionnelles ou déplacements privées)   | motorise  | voiture               | voiture passagers       | autres             |
| Transport Fluvial ou maritime (hors transrade militaire).   | collectif | collectif autres      | collectif autres        | autres             |
| Avion   | collectif | collectif autres      | collectif autres        | autres             |
| Roller, skate, trottinette non électrique.  | actif     | velo                  | velo                    | autres             |
| Fauteuil roulant  | actif     | marche                | marche                  | autres             |
| Autres modes (tracteur, engin agricole, quad, etc.)   | motorise  | voiture               | voiture conducteur      | autres             |
| Petits engins électriques (trottinette, segway, solowheel, etc.).   | motorise  | deuxroues             | deuxroues               | autres             |

FIG. 13 – EMC<sup>2</sup> - Agrégation modes transport