



Rapport de projet Stat'App

Elaboration de scénarios prospectifs sur la mobilité en France

Groupe 9

Louise LIGONNIERE, Amina MANSEUR, Lila MEKKI

Encadrant entreprise : M. Benjamin PETIAU

Encadrant école : M. Jaouad MOURTADA

Table des matières

1	Introduction	2
2	Problématique concrète pour EDF	2
3	Revue de littérature	4
4	Les données	4
4.1	L'Enquête sur la Mobilité des Personnes (EMP)	4
4.1.1	Présentation des données	4
4.1.2	Analyse descriptive	5
4.1.3	Variables cibles du projet	10
4.2	Les autres données utilisées	11
4.2.1	Projections démographiques	11
4.2.2	Projections météorologiques	11
5	Problématique technique : les pondérations	13
5.1	Qu'est-ce que les pondérations ?	13
5.1.1	Les pondérations individus	13
5.1.2	Les pondérations jours / déplacements	15
5.2	Démarche et difficultés rencontrées	15
5.2.1	Une première approche simple : la règle de trois	16
5.2.2	L'approche retenue de calage sur marges	17
5.2.3	Le calcul des pondérations jours / déplacements	18
6	Application aux scénarios et résultats	19
6.1	Scénario démographique	19
6.2	Scénario réchauffement climatique	21
7	Conclusion	21
8	Références	23
9	Annexes	24
9.A	Analyse descriptive	24
9.B	Méthode de repondération	26
9.C	Résultats	29

1 Introduction

Ce rapport s'inscrit dans le cadre du projet de Statistiques Appliquées du cursus de deuxième année du cycle ingénieur de l'ENSAE. L'objectif de ce projet est de mettre en œuvre une démarche scientifique et technique, basée sur les statistiques et des bases de données, afin de répondre à la problématique de l'entreprise encadrante. Nous synthétisons ici notre démarche et présentons les principaux résultats de notre étude.

Au vu de notre intérêt commun pour les sujets liés à l'environnement, l'énergie et la mobilité, nous avons choisi le sujet proposé par EDF visant à l'élaboration de scénarios prospectifs de mobilité. Dans le cadre de ce projet, notre groupe a eu la chance d'être encadré par Benjamin Petiau, ingénieur d'études en prospectives de la mobilité chez EDF R&D.

La prospective est un ensemble d'approches raisonnées du futur visant à éclairer les choix du présent. Elle ne consiste pas à prévoir l'avenir mais plutôt à élaborer des scénarios possibles pour le futur sur la base de l'analyse des données disponibles et la prise en compte de l'émergence de tendances. Elle est particulièrement utile aujourd'hui, alors que nous cherchons des solutions concrètes au changement climatique. En parallèle, la mobilité électrique se développe considérablement en France. D'après les chiffres de l'Observatoire Français de la Transition Écologique, plus d'un million de véhicules électriques circulent aujourd'hui en France. D'ici 2035, ce chiffre devrait atteindre les 17 millions. Pour accompagner ce changement, le nombre de bornes de recharge a augmenté de plus de 53% en l'espace d'un an seulement mais reste en deçà des besoins à venir. C'est donc dans cette optique de mitiger le risque climatique, en tenant compte du développement de la mobilité électrique, que s'inscrit notre projet.

Après une brève explication de la problématique concrète que représente ce projet pour EDF ainsi qu'un état de l'art sur la prospective en lien avec la mobilité, nous décrirons en détail notre démarche de recherche. Pour ce faire, nous observerons dans un premier temps les données à notre disposition ainsi que les données externes utiles aux scénarios de mobilité. Dans un second temps, nous développerons la problématique technique des pondérations qui a occupé une grande partie du projet et la façon dont nous l'avons abordée. Enfin, nous présenterons la démarche suivie pour mettre en œuvre les scénarios ainsi que nos conclusions et perspectives.

2 Problématique concrète pour EDF

Ce projet vise à relever un défi essentiel pour EDF, un acteur clé du domaine de l'énergie : anticiper la demande en électricité due au développement massif des véhicules électriques en France. En effet, avec l'émergence de cette technologie alternative, il est impératif de prévoir et de planifier les besoins énergétiques futurs, notamment en termes de recharge des véhicules électriques.

Pour cela, notre approche repose sur l'élaboration de scénarios de mobilité allant de l'horizon 2040 à 2070, en nous appuyant sur les données recueillies lors des deux dernières enquêtes nationales sur les déplacements des individus, réalisées par le Service Statistique Publique respectivement en 2008 et 2019. Ces enquêtes, riches en information sur les comportements et les habitudes de déplacement de la population, constituent une base solide pour établir des projections pertinentes quant à l'évolution de la mobilité électrique dans les années à venir.

Au cœur de notre démarche se trouve l'outil CharME (Charge des Mobilités Électriques), un instrument de simulation développé au sein de l'équipe de Benjamin Petiau à EDF R&D, et permettant d'estimer la consommation future liée aux recharges des véhicules électriques. CharME prend en compte comme données d'entrée, d'une part les paramètres décrivant les véhicules électriques à l'horizon souhaité (par exemple, l'énergie consommée par kilomètre par le véhicule), d'autre part les données de l'enquête de déplacement.

En sortie, l'outil génère la courbe de charge à une année d'horizon choisie, par exemple 2040. La courbe de charge représente la consommation heure par heure du 1er janvier 2040 au 31 décembre 2040, et fournit donc une vision détaillée des besoins énergétiques à anticiper par EDF.

La modélisation CharME suppose deux hypothèses importantes :

- a) Le type d'énergie de la voiture n'influe pas sur les déplacements des individus¹;
- b) La mobilité ne dépend pas de l'horizon de prévision².

Pour l'instant, les seules variables évoluant dans la prévision sont donc les caractéristiques des véhicules électriques, et le nombre de véhicules électriques en circulation. L'objectif de ce projet est de repenser l'hypothèse b) en introduisant des évolutions dans la modélisation des déplacements des individus.

Pour mener à bien cette mission, nous nous appuyerons sur les projections de population de l'Insee ainsi que sur une base de données essentielle, citée précédemment : l'EMP (Enquête Mobilité des Personnes) de 2019. Cette base regroupe une multitude de variables caractérisant les individus interrogés ainsi que leurs déplacements, fournissant ainsi une source riche et fiable d'informations pour notre étude. À chaque individu est associée une pondération de telle sorte que l'enquête représente les déplacements des Français en 2019.

Comme évoqué précédemment, notre démarche consiste en l'élaboration de scénarios de mobilité. Dans cette optique, nous avons réfléchi à plusieurs scénarios qui nous semblaient pertinents à mettre en place.

1. **Scénario démographique** : Ce scénario repose sur les projections démographiques de population élaborées par l'Insee aux horizons 2040 et 2070[1].
2. **Scénario climatique** : Ce scénario prend en compte les projections climatiques de plusieurs laboratoires français de modélisation du climat (Météo France et le Centre National de Recherche Météorologique[2]). Il matérialise le réchauffement climatique et montre l'impact de fortes chaleurs, ouragans, inondations sur les déplacements.
3. **Scénario sobriété** : Ce scénario implique une utilisation plus raisonnée des modes de transport énergivores via des changements structurels de modes de vie. Cela se traduit par une hausse de la marche à pied, du vélo et des transports en communs et une baisse de l'utilisation de la voiture couplé à de l'urbanisation.
4. **Scénario télétravail** : À la suite de la pandémie de COVID 19, on a observé une hausse de plus de 25% du nombre de personnes pratiquant le télétravail. Dans cette optique, les mobilités se voient être grandement modifiées. Ce scénario imagine donc une forte diminution des déplacements "domicile-travail" pour la partie active de la population.
5. **Scénario urbanisation/ruralisation** : Ce scénario façonne les mobilités selon que les populations se tournent vers un mode de vie plus urbain ou si les Français allaient s'installer en quantité à la campagne pour un retour à la nature.
6. **Scénario voitures autonomes** : Dans ce scénario, on observe un changement des usages. Cela permet par exemple aux personnes âgées de se déplacer lorsqu'elles ne sont plus capables de conduire. On peut imaginer une consommation moins élevée ainsi qu'un changement des heures de déplacements (notamment des déplacements tardifs ou très tôt le matin).
7. **Scénario transhumanisme** : Ce scénario s'appuie sur une hypothèse selon laquelle, grâce aux avancées technologiques, les populations vivraient beaucoup plus longtemps. Il se comprend dans un horizon plus lointain.

À partir de ces idées de scénarios, nous avons décidé de mettre en place deux scénarios principaux que nous estimons les plus pertinents, c'est-à-dire le scénario démographique ainsi que le scénario climatique.

L'approche centrale adoptée pour ce projet est de faire évoluer les pondérations de l'échantillon pour projeter l'enquête aux horizons temporels 2040 et 2070. L'idée est de définir un certain nombre de variables dont on prévoit une évolution d'ici 2040, et de modifier les pondérations de l'enquête pour refléter cette évolution. Par exemple, si on anticipe qu'il y aura plus de personnes de plus de 70 ans en 2040, on augmente le poids des répondants de plus de 70 ans dans l'enquête (sans modifier leurs réponses). Les variables seront choisies selon les scénarios considérés. Leurs évolutions dépendent des projections à l'horizon souhaité.

1. Si une personne parcourt 20km pour rentrer à son domicile le soir, on suppose qu'elle parcourt ces 20km aux mêmes horaires, qu'elle possède une voiture thermique ou électrique.

2. On utilise les données de l'enquête 2019, que ce soit à l'horizon 2019, 2040 ou autre : c'est-à-dire qu'on ne change pas la population ni la façon dont la population se déplace.

Ces différents scénarios servent de base crédible pour des projections prospectifs, mais au-delà des scénarios en eux-mêmes, c'est surtout la démarche technique qui est importante pour EDF.

3 Revue de littérature

Dans le cadre des objectifs de transition climatique, l'ADEME [3] propose quatre scénarios prospectifs menant à la neutralité carbone d'ici 2050. Chaque scénario repose sur des hypothèses distinctes en matière de modes de vie, de technologie et d'économie.

1. Le premier scénario décrit une **sobriété**, motivée à la fois par le choix et la contrainte, mettant l'accent sur une gouvernance locale, la **métropolisation**, ainsi qu'une réduction significative de la mobilité.
2. Le deuxième scénario présente des **coopérations territoriales** favorisant l'économie du partage et dynamisant les marchés économiques, avec une gestion maîtrisée de la mobilité.
3. Le troisième scénario privilégie le développement et l'utilisation de **technologies vertes** plutôt que la sobriété énergétique. Ce scénario insiste particulièrement sur l'utilisation de technologies en vue de décarboner l'énergie, développer des infrastructures et les pratiques de covoiturage, pour une mobilité accompagnée par l'Etat.
4. Enfin, le quatrième scénario se base sur le **déploiement de technologies de captage du CO₂** sans modifications majeures des modes de vie par ailleurs. Ces technologies sont loin d'être matures à l'heure actuelle.

Dans tous ces scénarios prospectifs, la mobilité des personnes est une variable centrale, de même que la consommation d'énergie. La question du développement des véhicules électriques notamment, est un enjeu important pour la neutralité carbone [4]. Il est donc crucial d'adapter les infrastructures aux besoins des populations futures et ce, dès les décennies à venir. Grâce aux données de l'Insee [1] ainsi que les projections météorologiques de Météo France [2], nous pouvons reproduire des scénarios prospectifs concentrés sur la mobilité en faisant évoluer les pondérations à différents horizons temporels à partir de l'Enquête de Mobilité des Personnes [5].

4 Les données

4.1 L'Enquête sur la Mobilité des Personnes (EMP)

4.1.1 Présentation des données

L'enquête sur la Mobilité des Personnes (EMP), menée entre mai 2018 et avril 2019, constitue une ressource essentielle pour comprendre et évaluer la mobilité des Français à l'échelle nationale et la comparer dans le temps. Cette base contient les réponses de **13 825 ménages**, ayant rapporté **45 169 déplacements**. Elle se place dans le cadre des enquêtes nationales de référence sur les déplacements des individus âgés de plus de 6 ans et fait suite à l'enquête nationale transport et déplacements (ENTD) conduite entre 2007 et 2008.

Son objectif est de fournir une description détaillée des pratiques de déplacement des individus, tant dans leur quotidien que pour des voyages de plus longue distance, et d'analyser les motifs et les modalités de ces déplacements. Elle offre également un aperçu de la composition du parc automobile des ménages, ainsi que des nouvelles tendances de mobilité telles que le covoiturage, l'utilisation de vélos en libre-service et l'équipement en recharge de véhicules électriques. Les données sont disponibles sous forme de bases de données individuelles anonymisées, que l'on utilise dans ce projet.

Au-delà de simplement capturer les déplacements à un instant donné, cette enquête, menée tous les dix ans environ, offre une perspective temporelle intéressante. En suivant l'évolution des habitudes de mobilité au fil du temps, elle permet aux chercheurs et aux décideurs d'identifier les tendances émergentes, de comprendre les changements sociétaux sous-jacents et d'anticiper les besoins futurs en matière d'infrastructures et de politiques publiques.

Dans le cadre de notre projet, nous nous intéresserons à l'analyse des déplacements locaux, qui englobent les trajets effectués pour des activités situées dans un rayon de 80 kilomètres à vol d'oiseau autour du domicile. Chaque individu a été interrogé sur tous ces déplacements **au cours d'une**

journée parmi les 7 jours précédents l'enquête. Rappelons également qu'à chaque individu est associé une pondération `pond_indC`.

4.1.2 Analyse descriptive

Nous avons entamé notre étude du projet en explorant les données disponibles dans l'EMP 2019. Cette phase initiale permet d'expliciter les variables **influant sur la mobilité des individus**, qui seront pertinentes à faire évoluer dans les scénarios prospectifs.

Nous nous sommes intéressées à quatre questions clés pour cette analyse :

- Qui se déplace ?
- Quand se déplace-t-on ?
- Pourquoi se déplace-t-on ?
- Comment se déplace-t-on ?

Nous présentons ci-dessous quelques résultats significatifs obtenus.

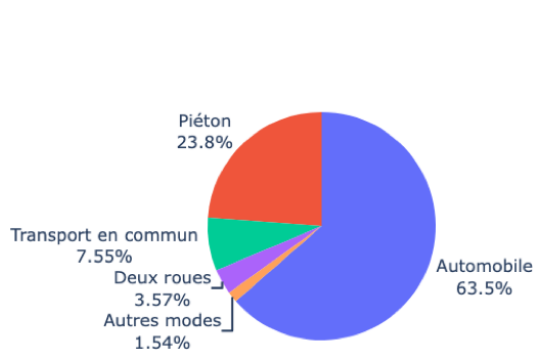


FIGURE 1 – Répartition des déplacements selon le mode

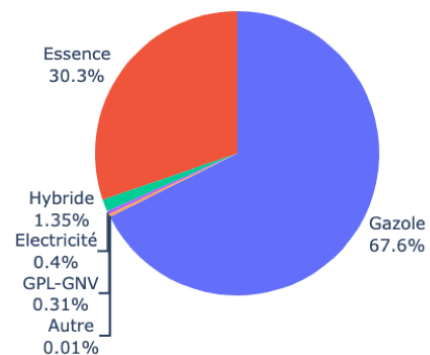


FIGURE 2 – Répartition des déplacements en voiture selon l'énergie du véhicule³

	Distance moyenne (km)	Ecart-type	Percentile 10%	Percentile 90%
Mode (libellé)				
Piéton	1.05	1.26	0.20	2.00
Vélo	4.20	5.60	0.36	10.00
Deux-roues motorisé	6.28	11.30	0.80	14.89
Automobile	15.03	34.73	1.33	30.99
Transport spécialisé, scolaire, taxi	16.11	22.63	3.27	40.00
Transport en commun	11.73	25.00	2.00	26.00
Train grande ligne	250.69	269.87	4.78	715.33
Avion	1740.91	1479.61	829.67	4714.26
Bateau	231.36	438.41	34.22	859.27
Autre	20.73	41.93	0.57	55.60

FIGURE 3 – Distance parcourue en un déplacement selon le mode de déplacement

L'essentiel des déplacements sont donc réalisés en voiture, dont une très faible part en voiture électrique en 2019. Depuis cette date, le nombre de véhicules électriques a considérablement augmenté. La distance parcourue dépend grandement du mode, un déplacement moyen en voiture représentant

3. À noter : l'EMP date de 2019, ce qui explique ces parts d'essence et gazole dans les déplacements.

15 km. Par ailleurs, on ne note pas d'effet majeur de l'énergie du véhicule sur la distance moyenne parcourue par déplacement. La distance semble simplement légèrement plus faible pour les voitures électriques, même si ce résultat est difficilement interprétable étant donné la faible représentation des déplacements en voiture électrique dans notre base.

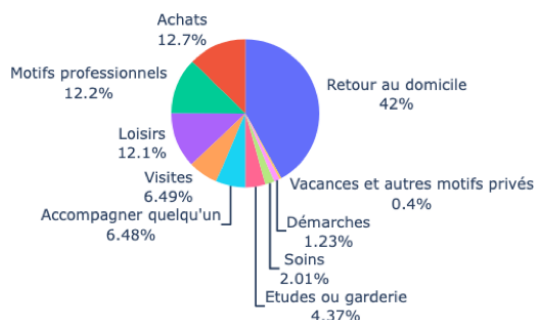


FIGURE 4 – Répartition des déplacements selon le motif⁴



FIGURE 5 – Répartition des déplacements selon l'horaire de départ

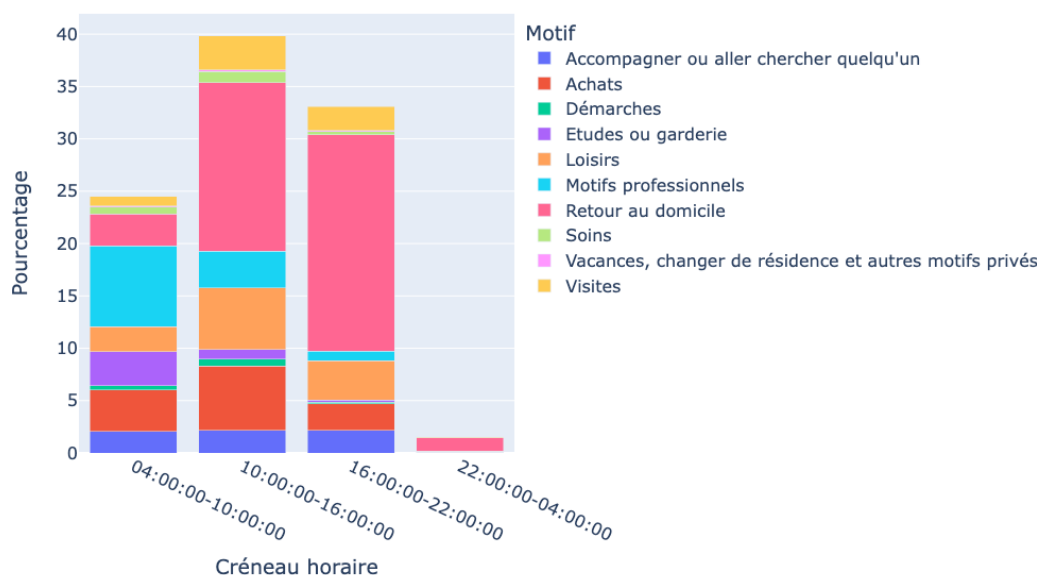


FIGURE 6 – Répartition des déplacements selon l'horaire de départ et le motif

S'agissant des motifs et horaires, au-delà de la prédominance du motif "Retour à domicile", les motifs les plus importants sont les achats, les motifs professionnels, ainsi que les loisirs. Les déplacements sont répartis au cours de la journée, réalisés pour la plupart entre 6h et 20h, avec un pic entre 16h et 18h. Les motifs sont par ailleurs très liés aux horaires de déplacement.

Au-delà de ces résultats généraux, on revient ci-dessous plus en détails sur quelques résultats liés aux scénarios envisagés.

Effets des variables évoluant dans le scénario démographie :

Pour rappel, dans ce scénario, nous utilisons les projections démographiques effectuées par l'Insee, qui incluent les variables suivantes : le genre, l'âge, le statut d'activité. Ces projections donnent lieu à horizons 2040 et 2070 à une population plus vieillissante. Afin d'étudier l'effet de ce vieillissement de la population, nous observons le lien entre les déplacements et l'âge. Les Figures 7 et 8 montrent que

4. Le graphe des motifs des déplacements, sans le motif Retour au domicile, est disponible en annexe 9.A

les seniors se déplacent en moyenne moins (2,24 déplacements par jour) et moins loin (7,56 km par déplacement).

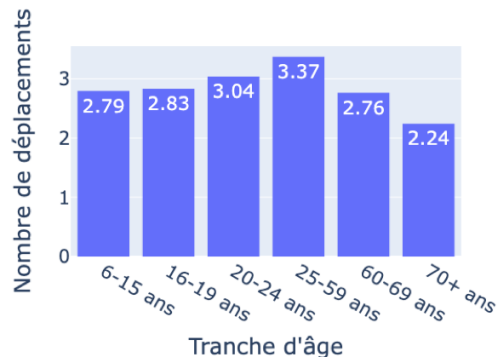


FIGURE 7 – Nombre journalier moyen de déplacements selon la tranche d'âge

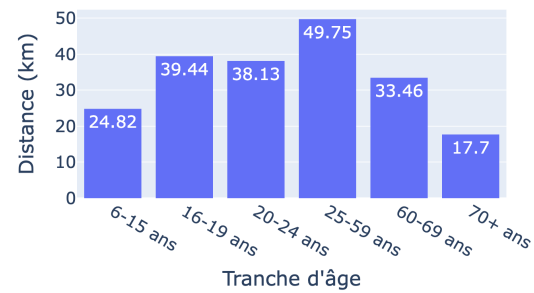


FIGURE 8 – Distance journalière moyenne selon la tranche d'âge

On regarde ensuite les motifs et modes de déplacement selon l'âge.

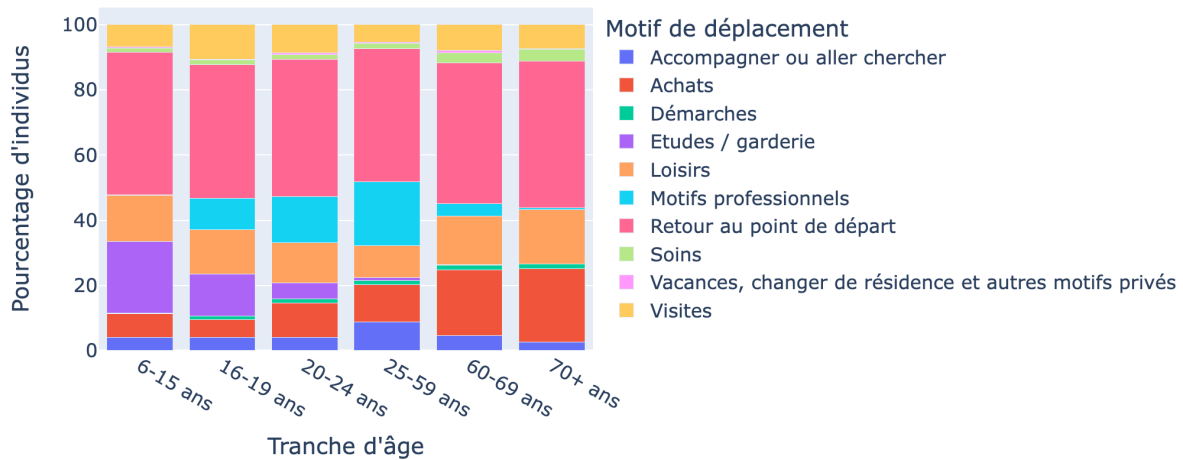


FIGURE 9 – Répartition des motifs de déplacement selon l'âge

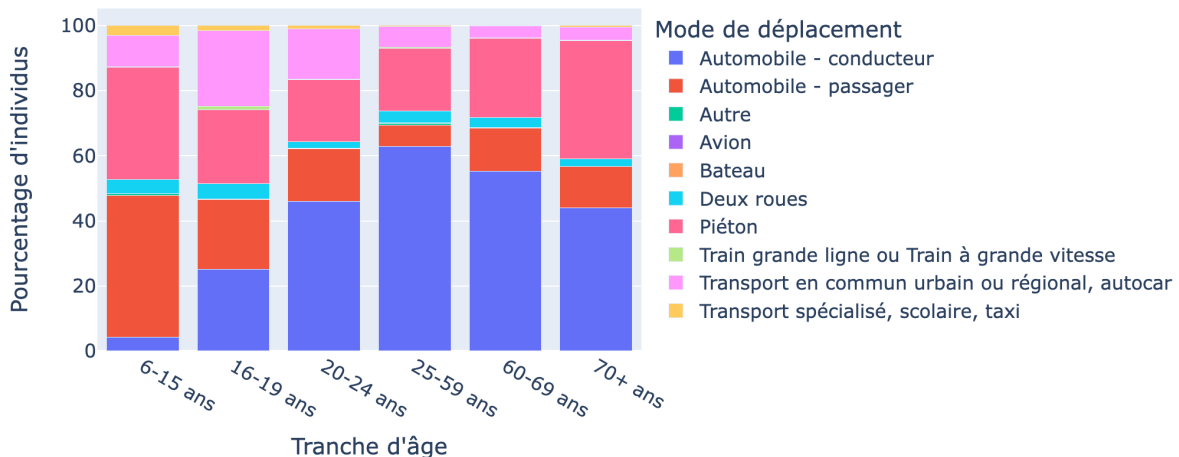


FIGURE 10 – Répartition des modes de déplacement selon l'âge

Les résultats obtenus coïncident avec l'intuition. Notamment :

- Les jeunes (jusqu'à 19 ans) se déplacent surtout pour se rendre dans leur lieu d'études (37% des déplacements des 6-15 ans et 26% pour les 16-19 ans). Ce motif reste le deuxième motif pour les 20-24 ans, puis diminue considérablement à partir de 25 ans au profit des déplacements pour se rendre au travail.
- Pour les plus de 60 ans, les achats sont le premier motif de déplacement (34% pour les 60-69 ans et 40% pour les 70 ans et plus), avec une prédominance des petits commerces par rapport aux grandes surfaces.
- Par ailleurs, les modes employés varient en fonction de l'âge. En voiture, les jeunes et les seniors sont beaucoup plus souvent passagers que conducteurs. Les moins de 24 ans ont tendance à plus emprunter les transports en commun, et les tranches de 16 à 59 ans se déplacent moins à pied.

Le statut d'activité a aussi un impact sur les déplacements : les actifs tendent à faire plus de déplacements, et parcourir plus de distance. Les actifs se déplacent aussi plus en automobile.

	Nombre de déplacements	Distance (km)
Actif	3.38	50.64
Non-actif	2.64	26.56

FIGURE 11 – Nombre de déplacements et distance journalière selon la situation d'activité

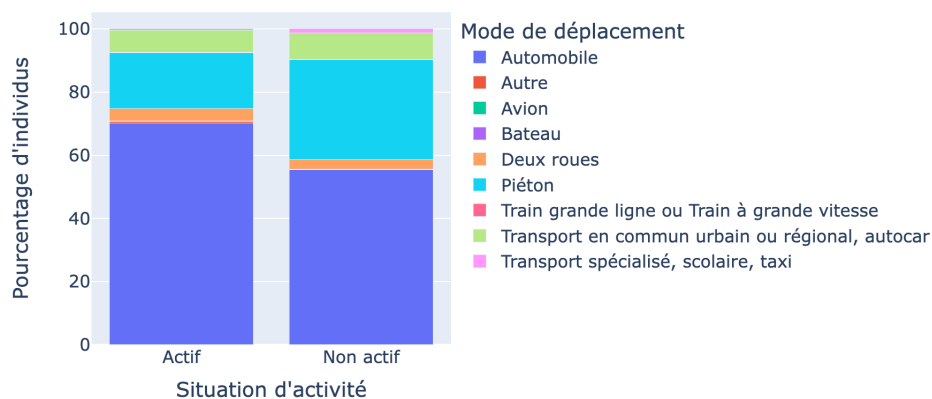


FIGURE 12 – Répartition des modes de déplacement selon la situation d'activité

Les graphes 13 et 14 permettent de préciser encore ces observations, en fonction de la situation vis-à-vis du travail. Néanmoins, cette variable n'est pas disponible dans les projections de l'Insee, elle ne pourra donc pas être utilisée dans les scénarios.



FIGURE 13 – Nombre journalier moyen de déplacements selon la situation vis-à-vis du travail

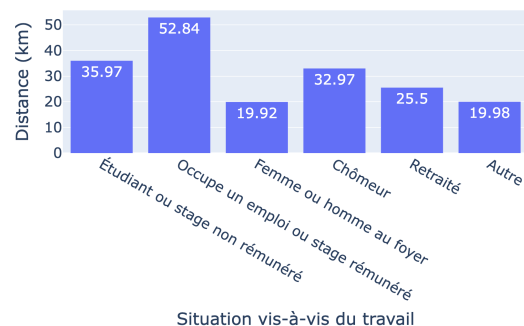


FIGURE 14 – Distance journalière moyenne selon la situation vis-à-vis du travail

Effets de la température :

La température est l'une des variables de l'EMP. Dans le cadre du scénario de changement climatique, l'influence de la température et des précipitations sur les habitudes de déplacement (distance moyenne d'un déplacement et nombre de déplacements par jour par individu) est un axe d'étude primordial.

Les Figures 15 et 16 témoignent d'un effet relativement faible de la température sur les déplacements, ce qui peut paraître étonnant de prime abord. On remarque seulement un peu moins de déplacements lorsqu'il fait moins de 2 degrés et plus de 20 degrés. Les distances sont par ailleurs un peu plus longues pour certaines températures élevées. Ces résultats sont néanmoins difficiles à interpréter, d'autant plus qu'ils peuvent recouper des effets saisonniers plutôt que de température.

C'est pourquoi, au lieu de considérer la température en valeurs absolues, nous avons cherché à construire une mesure relative de la température, en calculant l'écart de la température du jour de déplacement vis-à-vis d'une température moyenne définie. L'idée première était de comparer chaque température à la moyenne de la température calculée sur une fenêtre glissante de 9 jours autour du jour de déplacement. Cela aurait permis de tenir compte des comportements humains qui prévoient leurs déplacements en fonction de la météo des jours précédents ou suivants dans un horizon proche. Malheureusement, la base de données ne contient que le mois du déplacement ainsi que le type de jour (lundi, mardi,...), il est donc impossible d'identifier le jour exact de déplacement et donc de faire le lien avec la température des jours voisins. Nous n'avons donc pas pu calculer cette moyenne glissante. Pour pallier ce problème, l'approche retenue est de calculer la température moyenne sur le mois et de comparer la température du jour de déplacement à cette moyenne.

Les Figures 17 et 18 illustrent le nombre de déplacements et la distance moyenne de déplacement en fonction de ces écarts de température. Les variations demeurent faibles.

Pour expliquer cette absence de résultats significatifs concernant l'influence de la température sur les déplacements, on note d'abord sur les Figures 15 et 16 que les températures varient sur une plage assez restreinte (pour l'essentiel entre 0 et 25 degrés environ). Les écarts de température sont encore plus resserrés (entre -5 et 5 degrés). On dispose donc de peu d'observations pour des températures plus extrêmes, pour lesquels les résultats pourraient être plus forts. Ces observations ont un intérêt crucial car dans les décennies à venir, d'après les scénarios météorologiques, on s'attend à observer beaucoup plus d'événements extrêmes, ce qui devrait modifier considérablement les déplacements. La question reste donc ouverte quant à l'influence de la température sur les déplacements.

Une étude similaire a été faite dans le cas de l'influence des précipitations sur les déplacements. Les graphiques se trouvent en annexe 4.1.2 et montrent le même phénomène que décrit précédemment.

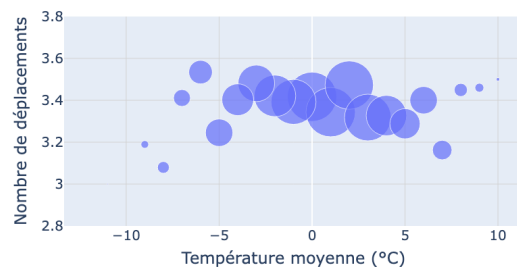


FIGURE 15 – Nombre de déplacements en fonction de la température⁵

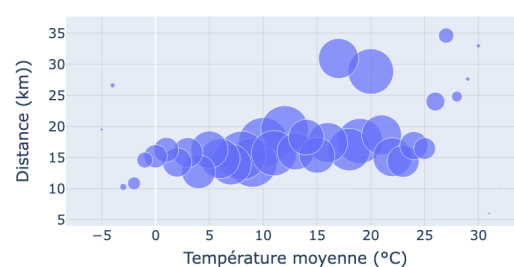


FIGURE 16 – Distance moyenne d'un déplacement en fonction de la température

5. NB : Sur ce graphique et les trois suivants, la taille des points indique le nombre d'observations sur lequel le calcul a été fait dans la base. Ainsi, les points les plus petits sont peu fiables. Par ailleurs, les températures ont été arrondies à l'unité la plus proche.

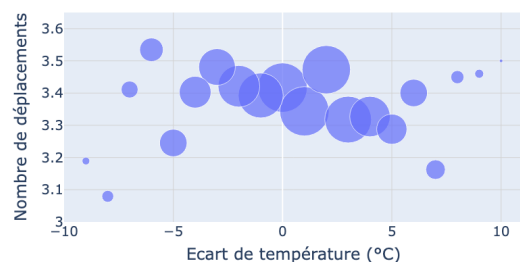


FIGURE 17 – Nombre de déplacements en fonction des écarts de température

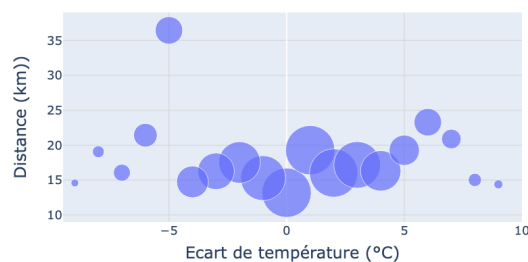


FIGURE 18 – Distance moyenne d'un déplacement en fonction des écarts de température

4.1.3 Variables cibles du projet

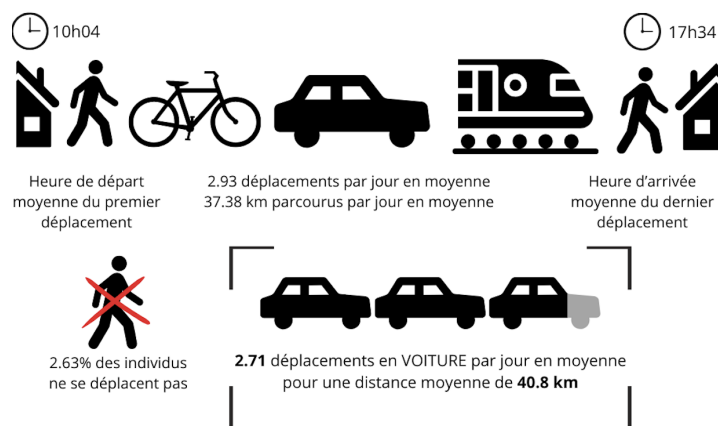
En parallèle de l'analyse descriptive, nous avons défini une première liste de variables cibles du projet. Il s'agit de variables qui peuvent être calculées à partir de l'enquête, et qui sont pertinentes vis-à-vis de l'objectif de CharME. Ce calcul permet de gagner du temps en obtenant des résultats similaires aux résultats de sortie de CharME, sans avoir à exécuter le programme CharME sur l'enquête (assez lourd en termes computationnels).

Il s'agit des variables suivantes :

- Nombre moyen de déplacements et distance moyenne parcourue par jour ;
- Nombre journalier moyen de déplacements et distance moyenne journalière en voiture, marche à pied, transports en commun ;
- Taux de journées sans déplacements ;
- Heure moyenne de départ du premier déplacement de la journée ;
- Heure moyenne d'arrivée du dernier déplacement de la journée ;

Ces variables peuvent être calculées sur l'ensemble de la population, ou sur des segments (en fonction de l'âge, du lieu de résidence, ou de l'activité professionnelle par exemple).

Dans un premier temps, nous avons calculé ces variables sur l'enquête 2019. Nous avons créé une fonction `var_cibles` qui, à partir de la base de réponses issue de l'enquête 2019 et d'un jeu de pondérations, produit les valeurs de ces variables. Cette fonction permettra donc d'obtenir les variables cibles dans les étapes suivantes, pour chaque scénario de projection de l'enquête.

FIGURE 19 – Quelques résultats illustrés de l'EMP 2019⁶

6. Par "individus qui ne se déplacent pas", on désigne les individus n'ayant rapporté aucun déplacement dans la semaine précédant l'enquête. Pour les déplacements en voiture, on ne regarde que les individus ayant effectué au moins un déplacement en voiture.

4.2 Les autres données utilisées

Pour élaborer les deux scénarios prospectifs, nous avons examiné l'incidence de deux facteurs clés : l'évolution démographique d'une part, et l'interaction entre le réchauffement climatique et l'évolution démographique d'autre part. Nous avons basé nos projections sur des données provenant de sources fiables telles que l'Insee[1] et Météo France[2].

4.2.1 Projections démographiques

Nous avons choisi de mettre en œuvre un scénario démographique appuyé sur les projections de l'Insee. Ces projections ont été actualisées en 2021 à partir des données du recensement 2018, et sont calculées jusqu'à 2070. Pour cette étude, on retient les données projetées à deux horizons temporels : 2040 et 2070. Ces projections donnent l'effectif total de population prévu, ainsi que les effectifs découpés selon trois variables : le genre, l'âge, le statut d'activité.

Les projections sont calculées selon différents scénarios, faisant varier plusieurs variables : la **fécondité**, **l'espérance de vie**, la **migration**.

Selon le scénario central, on considère :

- Une fécondité de 1,8 enfants par femme à partir de 2023
- Un âge moyen à la maternité en hausse jusqu'à 33,0 ans en 2052 et stable ensuite
- Une espérance de vie de 90,0 ans chez les femmes et 87,5 ans chez les hommes en 2070
- Un solde migratoire de +70 000 par an sur toute la période

Ainsi, selon le scénario central, la population active continue d'augmenter jusqu'en 2070 mais de manière nettement moins élevée que ces dernières décennies (car la fécondité est en baisse). Les variantes envisagées sur le solde migratoire ou la fécondité n'ont qu'un faible impact sur le rapport entre actifs et inactifs de 60 ans ou plus. Ces variantes, comme le scénario central, sont construites sous l'hypothèse que la législation des retraites reste inchangée d'ici 2070. Des réformes ayant eu lieu récemment à ce sujet, les projections de l'Insee devraient donc être modifiées en conséquence dans les années qui suivent.

Dans le scénario d'évolution démographique que l'on implémente, nous nous limitons à ce scénario central proposé par l'Insee. On pourrait dans un second temps élaborer des scénarios avec une population encore plus vieillissante, en considérant le scénario de l'Insee correspondant à une espérance de vie plus grande et une fécondité encore plus faible.

4.2.2 Projections météorologiques

Pour obtenir des projections météorologiques à l'horizon temporel 2040-2070, nous avons utilisé le portail DRIAS[2] mis à disposition sur le site du Ministère de la Transition Écologique, regroupant le travail de projections des principaux laboratoires français de modélisation du climat (IPSL, CERFACS, CNRM), en partenariat avec Météo France, l'Institut Pierre Simon Laplace et le CERFACS. Ce portail nous permet de sélectionner et télécharger gratuitement les variables et indicateurs climatiques saisonniers issus des simulations climatiques selon différents scénarios d'émission de CO_2 .

Les différents scénarios disponibles sur le portail sont :

- RCP2.6 : Scénario avec une politique climatique visant à faire **baisser les concentrations en CO_2** ,
- RCP4.5 : Scénario avec une politique climatique visant à **stabiliser les concentrations en CO_2** ,
- RCP8.5 : Scénario sans politique climatique.

Notre choix s'est porté sur le scénario RCP8.5, car il représente une trajectoire où les émissions de CO_2 continuent d'augmenter sans intervention significative pour les réduire. Ce scénario permet donc d'évaluer les implications les plus extrêmes du changement climatique, et est pertinent pour le scénario *Réchauffement climatique*.

Les différents horizons temporels disponibles sont :

- Période de référence : indicateurs calculés sur la période [1976-2005],

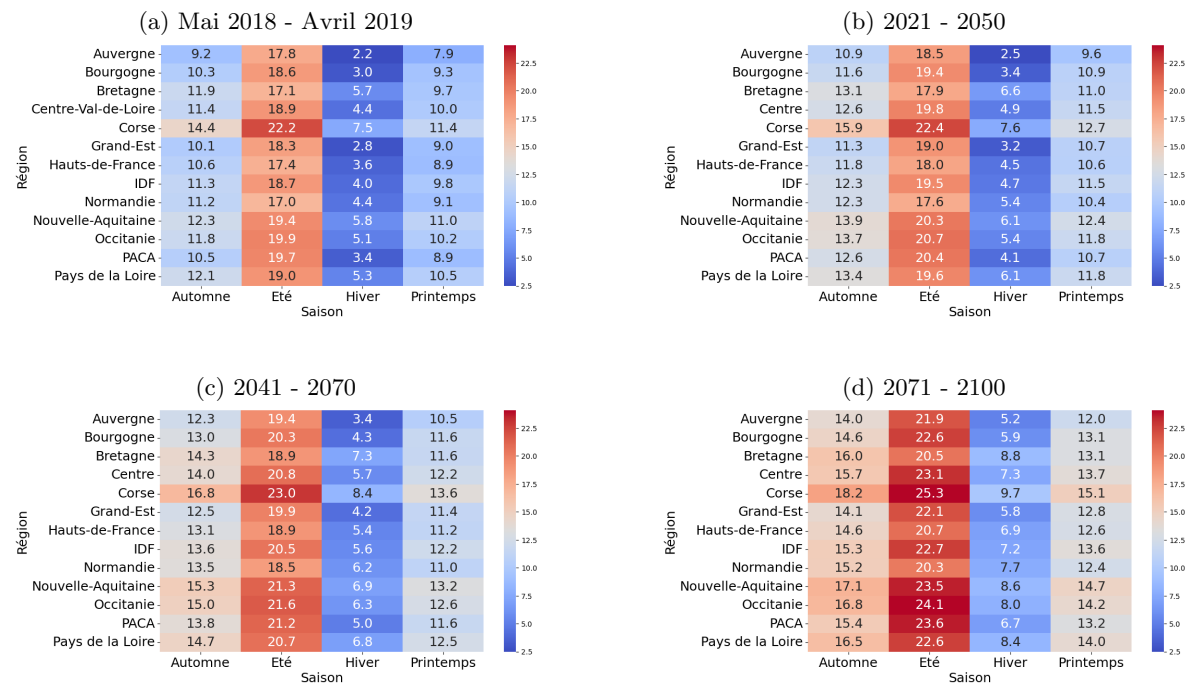


FIGURE 20 – Evolution de la température moyenne journalière par région par saison

- Horizon proche : indicateurs calculés sur la période [2021-2050],
- Horizon moyen : indicateurs calculés sur la période [2041-2070],
- Horizon lointain : indicateurs calculés sur la période [2071-2100].

Nous avons privilégié l'horizon moyen, couvrant la période de 2041 à 2070, pour anticiper les changements climatiques à moyen terme. Ce choix repose sur un compromis entre la fiabilité des résultats et la significativité des changements observés. En effet, alors que des projections sur des périodes plus éloignées pourraient révéler des évolutions plus marquées, elles sont généralement moins fiables en raison de l'incertitude croissante liée aux modèles climatiques. En se concentrant sur l'horizon moyen, nous devrions obtenir des résultats assez fiables tout en étant capables de discerner des tendances significatives.

Par ailleurs, différents modèles établis par des institutions françaises et européennes étaient disponibles. Notre choix s'est porté sur la modélisation proposée par Météo-France et le Centre National de Recherches Météorologiques.

D'autre part, le portail permet de sélectionner divers indicateurs météorologiques, comme la température, les précipitations, l'humidité ou encore le vent et d'en obtenir les projections. Les indicateurs que nous avons considérés sont : la température moyenne, les précipitations quotidiennes, le nombre de jours de vague de chaleur et de vague de froid, les périodes de sécheresse et le nombre de jours de précipitations intenses. Les indicateurs pouvant être obtenus par région ou à l'échelle de la France entière, nous avons utilisé les deux possibilités dans différents cas de notre analyse.⁷

Les indicateurs utilisés sont établis par région et par saison.

La comparaison de ces projections aux différents horizons temporels, dans le cas du scénario sans politique climatique, permet de remarquer une forte augmentation de la température moyenne journalière et la diminution des précipitations totales journalières.

Par ailleurs, les projections météorologiques révèlent une augmentation du nombre de jours de forte chaleur au fil du temps. De fait, une hausse constante du nombre de jours de vague de chaleur est observée pour toutes les saisons, qui devraient par exemple passer de 3.19 à 21.73 jours entre

⁷. Les données à l'échelle de la France ont été utilisées dans le cadre de la justification des scénarios établie précédemment dans la partie 4.1.2.

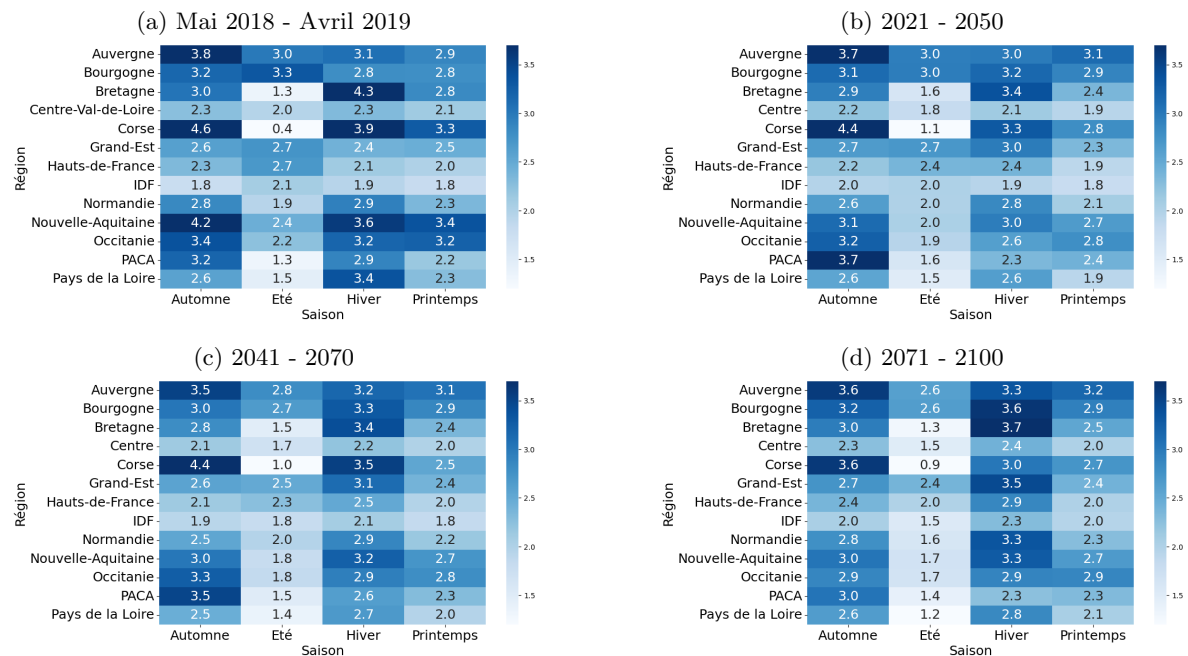


FIGURE 21 – Evolution des précipitations journalières par région par saison

les printemps 2021-2050 et 2071-2100. De plus, une nette diminution des jours de vague de froid est observée à travers les différents horizons temporels. Les périodes de sécheresse et les précipitations semblent maintenir une stabilité relative à travers les différentes périodes projetées, sans montrer de tendance significative à la hausse ou à la baisse.

5 Problématique technique : les pondérations

5.1 Qu'est-ce que les pondérations ?

L'EMP présente deux types de pondérations : les pondérations s'appliquant aux individus (`pond_indC`), et les pondérations s'appliquant aux déplacements (`pond_jour`).

5.1.1 Les pondérations individus

Les pondérations au niveau des individus (`pond_indC`) donnent un poids à chaque individu interrogé, et permettent d'obtenir des résultats représentatifs de la population française, même en n'ayant interrogé qu'un échantillon de la population. Elles ont été calculées par l'Insee en plusieurs étapes [5] :

- D'abord, la **non-réponse totale** a été corrigée, via la modélisation des probabilités individuelles de réponse et la constitution de groupes de réponses homogènes, permettant d'augmenter les poids des répondants pour compenser le biais introduit par les non-répondants.
- Les poids ont ensuite été calculés par la méthode de **calage sur marges**. Les marges à atteindre ont été fixées en simultanément au niveau ménage (7 marges : âge et catégorie sociale de la personne de référence, situation dans les Quartiers Prioritaires de la Ville, type de l'Unité Urbaine d'habitation, type de logement (individuel / collectif), type de ménage), et au niveau individu (4 marges : sexe, âge, diplôme, catégorie sociale).

Ce projet s'appuyant de façon centrale sur les pondérations, il est d'abord nécessaire de bien comprendre le principe des pondérations, et en particulier la méthode de calage sur marges. On revient ici rapidement sur la théorie de cette méthode [6].

Les méthodes de pondération visent à corriger un potentiel biais dans les enquêtes, lorsque l'échantillon n'est pas représentatif de la population sur certaines variables. Ceci arrive notamment en présence de

biais de sélection, lorsque certaines catégories de personnes sont plus faciles à atteindre que d'autres lors du sondage et de l'enquête. Si ces variables sont corrélées aux variables d'intérêt de l'enquête, l'estimation a tendance à donner des résultats biaisés. Ainsi, les techniques de pondérations permettent de redresser un échantillon pour le rendre représentatif de la population sur un certain nombre de variables, appelées variables de calage. On utilise une information auxiliaire disponible sur ces variables. Le calage sur marges en particulier permet de se caler sur des variables catégorielles, dont on doit connaître les effectifs pour chaque modalité sur la population.

Formellement, on considère une population U d'individus, dans laquelle on a sélectionné un échantillon s . Pour tout individu $k \in U$, sa probabilité d'inclusion dans s est notée π_k : c'est la probabilité pour chaque individu k d'être sélectionné dans l'échantillon. La probabilité d'inclusion découle directement du plan de sondage, c'est-à-dire la procédure selon laquelle on sélectionne l'échantillon dans la base de sondage.

Pour l'estimation d'une variable Y dans la population⁸, on utilise un estimateur pondéré : $\hat{Y} = \sum_{k \in s} d_k y_k$, où les d_k correspondent aux poids d'estimation associés aux observations de l'échantillon. Avant tout traitement et toute repondération, ces poids sont égaux à l'inverse des probabilités d'inclusion : $d_k = \frac{1}{\pi_k}$, et l'estimateur de Y est alors l'estimateur d'Horvitz-Thompson : $\hat{Y}_{HT} = \sum_{k \in s} \frac{1}{\pi_k} y_k$. Par exemple, dans le cas d'un plan de sondage uniforme, où l'on interroge par exemple 10 individus dans une population de 200 individus, les poids de sondage d_k sont égaux à 20 pour chaque individu. Intuitivement, on dit que chaque individu de l'échantillon en représente 20, et c'est donc le poids que l'on applique à son observation y_k .

Pour le calage, on suppose qu'on l'on connaît les totaux sur la population de J variables auxiliaires X_1, \dots, X_J . Par ailleurs, ces J variables auxiliaires sont disponibles pour toutes les observations de l'échantillon. Comme il s'agit ici de variables catégorielles, les variables X_j correspondent aux indicatrices associées aux modalités des différentes variables de calage. Pour le sexe par exemple, on aura besoin de connaître le nombre d'hommes et de femmes dans la population.

On cherche ainsi un nouveau jeu de pondérations $(w_k)_{k \in U}$, qui assure le calage sur les totaux des variables X_j , c'est-à-dire qu'elles permettent via l'estimation de retrouver les totaux exacts de ces variables sur la population. Cette contrainte s'écrit via les équations de calage :

$$\forall j = 1, \dots, J : \sum_{k \in s} w_k x_{j,k} = X_j \quad (1)$$

Dans le cas de l'EMP par exemple, le calage a été fait sur plusieurs variables dont le genre. Ainsi, si l'on estime le nombre de femmes dans la population à partir de l'échantillon EMP, en appliquant les pondérations individus (`pond_indC`), on retrouve bien les vrais effectifs dans la population française.

Ces poids de calage doivent par ailleurs être aussi proches que possible des pondérations initiales d_k , au sens d'une fonction de distance G fixée.

On résout donc le programme suivant :

$$\begin{aligned} \min_{(w_k)_k} \quad & \sum_{k \in s} d_k G\left(\frac{w_k}{d_k}\right) \\ \text{s.c.} \quad & \forall j = 1, \dots, J : \sum_{k \in s} w_k x_{j,k} = X_j \end{aligned} \quad (2)$$

La solution de 2 est obtenue via l'écriture d'un Lagrangien, puis la résolution du système de J équations à J inconnues par la méthode itérative de Newton.

Quatre fonctions de distance sont usuellement utilisées, donnant lieu à quatre méthodes de calage différentes :

— La méthode linéaire :

$$G(r) = \frac{1}{2}(r - 1)^2$$

Cette méthode est la plus rapide. Néanmoins, les poids ne sont pas bornés supérieurement, et elle peut conduire à des poids négatifs.

8. Si la variable d'intérêt est qualitative, on utilisera autant de variables indicatrices que de modalités.

— La méthode raking ratio :

$$G(r) = r \log r - r + 1$$

Cette méthode conduit à des poids toujours positifs, mais non bornés supérieurement. Les deux méthodes suivantes permettent de contrôler les rapports de poids, en assurant qu'ils soient compris entre deux bornes L et U fixées au préalable (telles que $L < 1 < U$).

— La méthode logit :

$$G(r) = \begin{cases} ((r-L) \log \frac{r-L}{1-L} + (U-r) \log \frac{U-r}{U-1}) \frac{(1-L)(U-1)}{U-L} & \text{si } L < r < U \\ +\infty & \text{sinon} \end{cases}$$

— La méthode linéaire tronquée :

$$G(r) = \begin{cases} \frac{1}{2}(r-1)^2 & \text{si } L \leq r \leq U \\ +\infty & \text{sinon} \end{cases}$$

Ces deux dernières méthodes sont les plus souvent utilisées, car elles permettent de contrôler les rapports des poids, et évitent ainsi des poids trop faibles ou trop élevés. Des poids trop élevés notamment entraînent une perte de précision dans les estimations, en utilisant trop fortement certaines observations par rapport aux autres.

La méthode de calage sur marges est mise en oeuvre dans le package R Icarus[7], que l'on utilisera dans ce projet. On utilise la fonction `calibration`, qui prend en entrée une base de données, une matrice de marges contenant les effectifs des variables de calage dans la population et un vecteur de poids initiaux (par exemple, les poids de sondage), et implémente la méthode itérative de Newton pour calculer le vecteur des poids de calage. Les quatre méthodes sont développées dans cette fonction.

5.1.2 Les pondérations jours / déplacements

Les pondérations au niveau des déplacements⁹ combinent la pondération relative à l'individu ayant effectué le déplacement, et donnent en plus un poids selon le jour de la semaine qui a été tiré dans l'enquête. Elles permettent d'obtenir des résultats représentatifs des jours de mobilités des Français au cours d'une semaine. Ainsi, si l'on somme les `pond_jour` sur la table déplacements, on obtient le nombre total de déplacements effectués sur une semaine pour l'ensemble des Français.

Néanmoins, la méthode de calcul des pondérations jour n'est pas du tout explicitée dans la documentation de l'enquête fournie par l'Insee. On reviendra donc sur notre méthode pour calculer les pondérations jours dans la section 5.2.3.

5.2 Démarche et difficultés rencontrées

L'approche centrale adoptée pour ce projet est de faire évoluer les pondérations de l'échantillon pour projeter l'enquête à deux horizons : un horizon proche en 2040, et un horizon plus lointain fixé en 2070. Dans les faits, les deux horizons sont intéressants pour EDF. Les projections à l'horizon proche sont plus fiables et moins disruptives par rapport à aujourd'hui, et les projections à l'horizon lointains offrent néanmoins de plus grandes évolutions dans la population.

L'idée est de définir un certain nombre de variables dont on prévoit une évolution d'ici l'horizon souhaité, et de modifier les pondérations (individus et jours) de l'enquête pour obtenir les bons effectifs projetés sur ces variables. Il s'agit donc bien d'une approche par les caractéristiques de population : on modifie les caractéristiques de la population aux horizons choisis. Par contre, on ne modifie pas la façon dont les individus, à caractéristiques contrôlées, se déplacent.

Notre travail méthodologique vise donc à définir une méthode permettant de calculer les pondérations selon les marges souhaitées pour un certain nombre de variables.

9. On les désigne par la suite par pondérations jours, celles-ci ayant été labellées de la sorte dans les documents de l'Insee. Il s'agit néanmoins bien de poids s'appliquant aux déplacements des individus.

5.2.1 Une première approche simple : la règle de trois

Une première approche simple, proposée initialement par notre tuteur Benjamin Petiau, consiste à recalculer les pondérations à l'horizon h souhaité à partir des pondérations de l'enquête 2019, en faisant évoluer la composition de la population sur une seule variable à la fois (par exemple, l'âge). Cette approche consiste en une règle de trois (avec les pondérations actuelles, le nombre d'individus actuel et le nombre d'individus projeté dans chaque catégorie d'âge) afin de trouver la pondération associée à 2040.

La limite principale de cette méthode est qu'en modifiant la pondération sur une seule variable (ici, l'âge), l'enquête n'est plus représentative sur les autres variables de calage. Cette méthode risque donc de biaiser les résultats obtenus, si la variable que l'on fait évoluer est corrélée à ces autres variables (ce qui est vraisemblablement le cas, il est par exemple raisonnable de faire l'hypothèse que l'âge est corrélé au statut d'activité).

Pour évaluer cette approche de re-pondération et en tester d'autres, nous avons développé trois bases de niveaux de complexité différents. L'idée est de pouvoir appliquer chaque méthode de re-pondération sur ces trois bases et d'évaluer ses résultats. Ces trois bases ont été créées selon les méthodes suivantes :

- Méthode 1 : Création (manuelle) d'une base très réduite et simplifiée, avec 10 lignes et 3 variables (âge, statut d'emploi, nombre de déplacements) ;
- Méthode 2 : Sélection d'un nombre réduit de variables parmi les variables de l'EMP (âge, genre, situation vis-à-vis du travail), pour lesquelles on conserve toutes les observations de l'EMP 2019 ;
- Méthode 3 : Essai d'une méthode de tirage dans la base de 2019. Cette méthode consiste à simuler une nouvelle base en tirant avec remise des déplacements parmi les lignes de l'EMP 2019. Les probabilités de tirage correspondent aux pondérations `pond_jour`. Ainsi, on obtient une enquête sans poids (ou de poids 1 pour chaque ligne).

Ces essais ont permis de confirmer notre intuition théorique. Pour exemple, on revient ici sur les détails de la méthode 2, à savoir l'application de la repondération par ratio à la variable d'âge, sur la base complète de l'EMP. On choisit ici un découpage en 9 tranches d'âge. On commence par calculer les effectifs de la population dans chaque tranche d'âge, en 2019, en appliquant les `pond_indC2019` (fournies par l'Insee). On obtient donc le nombre d'individus dans chaque catégorie d'âge $i \in 1, \dots, 9$ en 2019, notés $\text{tot}_{i,2019}$: ce sont nos effectifs de départ. En parallèle, on calcule les effectifs projetés de population pour chaque catégorie d'âge, en utilisant les projections de population de l'Insee à horizon 2040, notés $\text{tot}_{i,2040}$. Ce sont les effectifs objectifs à horizon 2040 : on souhaite que les poids 2040 permettent de retrouver ces effectifs par tranche d'âge¹⁰. On applique donc le ratio au sein de chaque catégorie d'âge. En notant $s_i = \{k \in s \mid \text{age}_k = i\}$, on a :

$$\forall k \in s_i : \text{pond_ind}_{k,2040} = \text{pond_ind}_{k,2019} \frac{\text{tot}_{i,2040}}{\text{tot}_{i,2019}}$$

On obtient bien :

$$\sum_{k \in s_i} \text{pond_ind}_{k,2040} = \text{tot}_{i,2040}$$

Pour vérification, on calcule ensuite les effectifs obtenus avec ces pondérations projetées en 2040, pour le statut d'activité. On obtient un total de 29 437 580 actifs en 2040. Or, d'après les projections de l'Insee sur le statut d'activité, le nombre d'actifs en 2040 sera de 30 460 662. La différence est donc de l'ordre du million d'actif. Ainsi, l'âge (que l'on modifie ici), est fortement corrélé au statut d'activité. La population vieillit entre 2019 et 2040, entraînant une diminution du nombre d'actifs. Néanmoins, le lien entre âge et activité évolue simultanément entre 2019 et 2040, avec une augmentation du taux d'activité des seniors. En prenant en compte seulement l'évolution de l'âge, et pas celle de son impact sur l'activité, on sous-estime le nombre d'actifs en 2040. Ce résultat correspond au principal biais de cette méthode de repondération sur une variable à la fois, que l'on cherche donc à améliorer par la suite.

10. Cela correspond aux équations de calage 1.

5.2.2 L'approche retenue de calage sur marges

Pour mettre en oeuvre une méthode permettant de contrôler plusieurs variables en parallèle, on s'appuie sur la méthode de calage sur marges décrite à la section 5.1.1.

Si l'on souhaite reproduire le calage sur marges à horizon h , avec le même niveau de précision que celui de l'Insee pour l'enquête 2019, les éléments suivants sont nécessaires :

- Les poids corrigés de la non-réponse totale, desquels on devrait repartir pour le calage
- Les marges projetées à horizon h , pour toutes les variables de calage définies par l'Insee¹¹ ;
- L'information de la méthode qui a été utilisée pour le calage (linéaire / raking ratio / logit / linéaire tronquée)

Néanmoins, ces informations ne nous sont pas toutes disponibles. Nous avons donc cherché à développer une méthode approchée. Ainsi :

- **Pour les poids initiaux**, on utilise des poids uniformes pour l'ensemble de l'échantillon. Ceux-ci correspondent aux poids de sondage si le sondage avait été fait selon un plan aléatoire simple sans remise, c'est-à-dire que chaque unité de la population U a la même probabilité d'être sélectionnée dans l'échantillon s . Chaque unité a donc une probabilité d'inclusion $\pi_k = \frac{\#s}{\#U}$, et les poids d'estimation sont tels que $d_k = \frac{1}{\pi_k} = \frac{\#U}{\#s}$ ¹². Cette méthode a l'avantage d'être simple et permet d'atteindre l'effectif de population désiré. Néanmoins, elle ne permet pas de prendre en compte le plan de sondage spécifique à l'enquête, à savoir un sondage aléatoire simple stratifié. Ainsi, la population a été découpée en strates selon 3 variables (le statut Quartiers Prioritaires de la Ville, la répartition entre communes urbaines et rurales, et la motorisation des ménages). L'échantillon a été tiré de façon non-uniforme entre ces strates. Ce plan de sondage permet d'augmenter la précision des estimations, en interrogeant plus d'individus dans les strates où l'on observe plus de variance, et moins dans les strates très homogènes en intra. Ici, par exemple, les logements multi-motorisés sont sur-représentés. Les probabilités π_k ne sont donc pas uniformes, et il faudrait normalement utiliser des poids de sondage d_k corrigeant cette sur-représentation de certains groupes d'individus. Par ailleurs, avant le calage sur marges, les poids de sondages sont aussi normalement traités pour corriger la non-réponse. Nous n'avons ici pas pu mettre en oeuvre de correction de la non-réponse.
- **S'agissant des marges**, on dispose seulement de projections pour certaines des variables de calage définies par l'Insee. On sera donc représentatif de ces variables à l'horizon h (par exemple, pour le scénario démographique, les variables de calage sont l'âge, le genre, et le statut d'activité). Pour les autres variables (par exemple, la catégorie sociale ou le type d'unité urbaine d'habitation), deux options sont possibles :
 - Utiliser les marges calculées en 2019. On fait ainsi l'hypothèse que la structure de la population du point de vue de ces variables n'évolue pas d'ici l'horizon h ;
 - Ne pas mettre de contrainte sur les effectifs de ces variables. On fait ainsi l'hypothèse qu'elles sont marginalement réparties de la même façon, relativement aux variables de calage que l'on contrôle. Par exemple, si les seniors habitent plus en milieu rural en 2019, on suppose que ce sera toujours le cas en 2040. S'il y a plus de seniors en 2040 (ce que l'on intègre via les marges au niveau de l'âge), il y aura aussi plus de personnes habitant en milieu rural. On ne contrôle donc pas le nombre de personnes habitant en milieu rural en 2040, qui mécaniquement augmentera via l'augmentation des poids des seniors. Cette deuxième hypothèse semble plus plausible et c'est donc celle que l'on implémentera.
- **S'agissant de la méthode de calage**, on opte pour une méthode linéaire tronquée, permettant une plus grande précision en contrôlant les rapports des poids.

11. Pour rappel, le calage a été réalisé sur 7 marges au niveau ménage (âge et catégorie sociale de la personne de référence, situation dans les Quartiers Prioritaires de la Ville, type de l'Unité Urbaine d'habitation, type de logement (individuel / collectif), type de ménage), et 4 marges au niveau individu (sexe, âge, diplôme, catégorie sociale).

12. Une façon simple de le comprendre est que chaque enquêté représente d_k personnes dans la population. Par exemple, si l'on interroge 10 personnes pour une population de 200 personnes, chaque enquêté représente 20 personnes.

5.2.3 Le calcul des pondérations jours / déplacements

Une fois les pondérations individus calculées, il est nécessaire de calculer les pondérations jours. Ce sont en effet les poids qui s'appliquent pour produire des estimations à partir des déplacements locaux déclarés par les enquêtés. Pour cette étape cruciale, ne disposant pas de détails sur la façon dont elles ont été estimées en 2019, nous avons dû inférer notre propre méthode de calcul. Cette étape a constitué une des difficultés principales du projet, et l'on présente ici l'approche adoptée pour la contourner.

Chaque enquête a d'abord indiqué tous les jours au cours desquels il a effectué au moins un déplacement parmi les 7 jours précédents l'enquête. Un jour a ensuite été tiré parmi les jours où il s'est déplacé. C'est ce jour que l'on appelle **jour d'interrogation** : c'est le jour pour lequel on dispose du détail des déplacements locaux de l'enquête. On le distingue du **jour d'enquête** (c'est-à-dire le jour où l'enquêteur a interrogé l'enquêté). Le tirage du jour d'interrogation a été fait de façon non uniforme, selon une loi choisie surtout pour des raisons méthodologiques, avec pour objectif d'atteindre un effectif suffisant de chaque type de jour¹³.

Le jour d'interrogation n'est donc pas réparti de façon uniforme dans la base : d'abord, car les enquêteurs se déplacent plus certains jours que d'autres (ainsi, 95% des enquêtés l'ont été en semaine) ; de plus, car le tirage du jour d'interrogation n'est pas uniforme dans les 7 jours précédents l'enquête. Ainsi, plus d'enquêtés ont été interrogés sur leurs déplacements un jour de semaine que sur un samedi ou un dimanche¹⁴.

Par ailleurs, les déplacements sont fortement corrélés au type de jour de la semaine. Ainsi, on a tendance à moins se déplacer le week-end, et les motifs et modes employés varient aussi grandement entre la semaine et le week-end. La représentation non-uniforme des différents types de jours doit donc être prise en compte pour pouvoir agréger les observations.

Pour éviter de perdre de l'information, il a été choisi par l'Insee de représenter la mobilité de la population au cours d'une semaine type (présentant donc chacun des cinq jours de semaine et un samedi et un dimanche), plutôt qu'au cours d'une journée moyenne. C'est là l'objectif des pondérations jour : de la même façon que la pondération individu permet de passer de l'échantillon à la population, la pondération jour permet de passer d'un jour d'observation à une semaine¹⁵.

Pour atteindre cette représentativité, nous avons envisagé deux idées, correspondant à deux niveaux de compréhension (ci-dessous, les pondérations jour sont notées $d_{2,k}$) :

- *Première idée* : On cherche à recréer une semaine représentative de la mobilité de la population. Il faut donc donner un certain poids $d_{2,k}$ à chaque ligne selon le type de jour d'interrogation pour obtenir une semaine complète. On note N la taille de la population. Pour reconstituer cette semaine, on doit additionner les déplacements des N individus pour chaque jour de la semaine. On a donc la contrainte suivante¹⁶ :

$$\sum_{k \in s} d_{2,k} = N * 7$$

On remarque néanmoins rapidement dans la base que ce résultat n'est pas vérifié.

- *Deuxième idée* : Le calcul doit prendre en compte les jours sans déplacements des individus. En effet, certains individus ne se déplacent pas certains jours. Ces jours sans déplacements sont potentiellement répartis non-uniformément dans la semaine (par exemple, plus souvent le dimanche). Or, aucun individu n'a été interrogé sur un jour où il ne s'est pas déplacé. La somme des poids doit donc bien correspondre au nombre d'individus-jours de déplacement sur une semaine, mais celui-ci est plus faible que $N * 7$. On doit avoir, en notant N_j^* le nombre

13. Les samedis et dimanches et la veille de l'enquête ont de plus grandes probabilités de tirage.

14. Le détail de la répartition des jours d'enquête et jours d'interrogation est disponible en annexe 9.A

15. Ainsi, si les déplacements étaient uniformes dans la semaine, il suffirait de multiplier le poids individu par 7 pour obtenir le poids jour : on a interrogé chaque individu sur une seule journée, et on souhaite des résultats sur une semaine, soit 7 fois cette journée. Comme ici, les jours d'interrogation ne sont pas uniformes, les poids jours ne le sont pas non plus : par exemple, moins d'enquêtés ont été interrogés sur leurs déplacements un samedi ou un dimanche qu'un jour de semaine, donc le poids donné à ces déplacements est plus important.

16. Les pondérations jour $d_{2,k}$ représentent en fait des individus-jours de déplacement. C'est pourquoi, en sommant ces individus-jours sur la base, on doit obtenir $N * 7$ individus-jours sur la population (puisque l'on souhaite représenter la mobilité de la population sur 7 jours).

d'individus de la population qui se sont déplacés le jour j :

$$\sum_{k \in s} d_{2,k} = \sum_{j=1 \dots 7} N_j^*$$

Marginalement, en notant jour_k le jour d'interrogation de l'individu k , on doit avoir les résultats suivants¹⁷ :

$$\forall j = 1, \dots, 7 : \sum_{k \in s, \text{jour}_k = j} d_{2,k} = N_j^*$$

Notre méthode consiste donc à calculer les N_j^* pour chaque type de jour $j = 1, \dots, 7$. Ceci est possible via la base individus, où l'on dispose d'une variable indiquant pour chaque jour de la semaine si l'individu s'est déplacé ou non. Ce calcul est fait en prenant en compte les pondérations individu $d_{1,k}$ déjà calculées à l'étape précédente. Les totaux obtenus prennent donc bien en compte la projection de la population à horizon h .

Les N_j^* correspondent aux nombre d'individus-jours de déplacement que l'on souhaite représenter pour chaque type de jour j . Ce sont les effectifs objectifs, que l'on doit atteindre en sommant les $d_{2,k}$ pour chaque jour j .

Pour chaque type de jour j , on estime ensuite le nombre d'individus dans la population dont on dispose des déplacements le jour j ¹⁸ :

$$n_j^* = \sum_{k \in s, \text{jour}_k = j} d_{1,k}$$

On applique ensuite le ratio suivant pour obtenir les pondérations jours :

$$d_{2,k} = d_{1,k} \sum_{j=1 \dots 7} \mathbb{1}(\text{jour}_k = j) \frac{N_j^*}{n_j^*}$$

Cette méthode permet bien d'obtenir une semaine de mobilité de la population française, en prenant en compte le taux de jours sans déplacement par type de jour.

Néanmoins, dans cette méthode, le ratio $d_{2,k}/d_{1,k}$ est le même pour tous les individus ayant été interrogés un même jour j . Dans l'EMP 2019, on constate que la relation entre la pondération individu et la pondération jour n'est pas proportionnelle, même lorsque l'on contrôle par type de jour¹⁹. Une solution pour aller plus loin pourrait être de construire des sous-groupes d'individus selon les catégories des variables de calage, et d'appliquer la méthode de ratio décrite ci-dessus au sein de chacun de ces sous-groupes. Cela permettrait de respecter la possible dépendance entre le fait de se déplacer ou non, et les caractéristiques de la population.

Une autre solution serait de mettre en oeuvre un calage simultané à plusieurs niveaux (individus et jours de déplacement). En effet, le calcul des poids jours dans notre méthode n'est fait que par type de jour, et comme une couche supplémentaire aux poids individus. Les pondérations jours risquent donc de biaiser légèrement l'échantillon par rapport aux variables de calage au niveau des individus. C'est aussi ce qui peut expliquer la non linéarité des pondérations jour 2019 dans l'EMP.

6 Application aux scénarios et résultats

6.1 Scénario démographique

Pour le premier scénario d'évolution démographique, on se cale sur une population projetée en 2040 et en 2070. La population totale est estimée à 64,9 millions d'individus de 6 ans et plus en 2040, et 64,2 millions en 2070. On choisit trois variables de calage :

17. Ceci explique pourquoi, dans la base, on observe toujours un peu moins de la population $N = 59.10^6$ lorsque l'on somme les pondérations jour pour chaque type de jour (et vraiment beaucoup moins pour les jours où l'on se déplace souvent pas : le dimanche, le lundi et le samedi surtout).

18. La pondération individu $d_{1,k}$ permet d'avoir un effectif au niveau de la population, pour le jour d'interrogation j .

19. Voir graphe en annexe 9.B

1. Le genre (en millions) :

	2019	2040	2070
Femmes	30,8	33,6	32,7
Hommes	28,7	31,3	31,5
Total	59,5	64,9	64,2

2. L'âge, selon un découpage en 9 catégories d'âge (en millions) :

	2019	2040	2070
6-14	7,7	6,6	6,1
15-29	10,2	11,7	11
30-39	8	8,2	7,5
40-49	8,3	7,9	7,5
50-59	8,2	8,3	8,3
60-69	8,4	8	8,1
70-79	5,4	7,4	7,1
80-89	2,9	5,1	6
90 et plus	0,5	1,6	2,6
Total	59,5	64,9	64,2

3. Le statut d'activité (en millions) :

	2019	2040	2070
Actif	28,7	30,5	29,2
Non-actif	30,8	34,4	35
Total	59,5	64,9	64,2

On implémente une méthode linéaire pour le calage, qui conduit ici à des rapports de poids dans un intervalle assez resserré²⁰. Il n'est donc pas nécessaire d'implémenter une méthode tronquée.

Les résultats sur les variables cibles les plus intéressantes sont les suivants²¹ :

Variable d'intérêt	2019	2040	2070
Pour rappel : population totale (6 ans et plus)	59,5 millions	64,9 millions	64,2 millions
Nombre total de déplacements en une journée	174 millions	178,1 millions	176,2 millions
Distance totale parcourue en une journée (km)	2,22 milliards	2,26 milliards	2,25 milliards
Nombre moyen de déplacements par jour par individu	2,93	2,74	2,75
Distance moyenne par jour par individu (km)	37,38	34,90	35,02
Nombre total de déplacements EN VOITURE en une journée	110,6 millions	116,2 millions	117,8 millions
Distance totale parcourue EN VOITURE en une journée (km)	1,66 milliards	1,72 milliards	1,72 milliards
Nombre moyen de déplacements EN VOITURE par jour par individu	1,86	1,79	1,84
Distance moyenne EN VOITURE par jour par individu (km)	27,94	26,58	26,87

FIGURE 22 – Résultats du scénario démographique à horizons 2040 et 2070

Ainsi, d'ici à 2040, le nombre de déplacements par individu diminue légèrement (sûrement lié au vieillissement de la population), mais cette diminution ne suffit pas à compenser l'augmentation de la population. Au total, le nombre de déplacements augmente donc. Le constat est le même en termes de distance, ainsi que quand on regarde les déplacements seulement en voiture.

Entre 2040 et 2070, la population diminue légèrement, entraînant une légère diminution du nombre total de déplacements ainsi que de la distance totale. Le nombre moyen de déplacements et la distance moyenne par individu réaugmentent très légèrement (alors qu'ils diminuaient entre 2040 et 2070). Par contre, les individus ont tendance à plus prendre la voiture. Les déplacements en voiture continuent donc d'augmenter malgré la baisse de la population.

20. Entre 0.7 et 2.5 pour 2040, et entre 0,7 et 2 pour 2070, avec seulement quelques valeurs extrêmes jusqu'à 4. Les graphes des rapports de poids sont disponibles en annexe 9.B

21. Les résultats détaillés de toutes les variables cibles sont disponibles en annexe 9.C

6.2 Scénario réchauffement climatique

En plus des évolutions démographiques, nous avons pour ambition d'intégrer les évolutions de température à notre travail. En effet, les températures en France sont amenées à évoluer grandement dans les décennies à venir, comme souligné dans la section 4.2.2. C'est donc un facteur qui pourrait avoir un impact important sur la mobilité. Nous avons néanmoins rencontré plusieurs difficultés, nous empêchant d'obtenir des résultats sur ce scénario.

D'une part, les seules projections que nous avons trouvées²² sont exprimées sous la forme d'une température journalière moyenne, par saison et par région. Ces températures sont disponibles pour des horizons assez lointains, mais sur des périodes très larges (par exemple, pour la période de 2041 à 2070). Globalement, on note une augmentation de la température moyenne d'ici à la période de 2041 à 2070. Intuitivement, il faudrait donc augmenter le poids des observations sur les jours les plus chauds dans l'EMP. La température étant une variable continue, il faudrait d'abord la discrétiser en construisant des petites tranches de température (par exemple, un découpage en tranches de 1 ou 2 degrés). Pour appliquer le calage sur marges, il serait nécessaire de disposer de la température journalière pour chaque jour d'une année (ou en considérant le découpage en tranches de température, le nombre de jours sur une année dans chaque tranche). L'échelle des projections dont on dispose n'est donc pas assez précise pour appliquer notre méthode de repondération.

Par ailleurs, même si l'on disposait des projections adéquates, nos analyses ont montré que, au sein de l'EMP, la température n'a pas un effet très fort sur les déplacements. Ainsi, on n'exclut pas une évolution de la mobilité en conséquence du réchauffement climatique, mais l'enquête a du mal à le capter.

Enfin, les observations présentes dans l'EMP sont resserrées dans un intervalle de températures assez restreint. Augmenter le poids des journées les plus chaudes, qui sont peu dans l'échantillon, aurait conduit à des rapports de poids assez élevés, et des résultats peu fiables. On disposait par exemple du nombre de jours de vagues de chaleur dans les projections à horizon 2040-2070. On aurait pu intégrer cette variable au calage, néanmoins, trop peu de jours de vague de chaleur étaient enregistrés dans l'EMP, empêchant d'obtenir des résultats.

7 Conclusion

Finalement, le scénario démographique nous donne des résultats satisfaisants : en 2040, une population vieillissante entraîne un nombre de déplacements par individu qui diminue légèrement, mais pas assez pour compenser l'augmentation de la population. On constate donc plus de déplacements. En voiture, la distance totale parcourue en une journée passe de 1,66 milliards de kilomètres à 1,72 milliards, soit une augmentation de 3,6%. À horizon plus lointain, en 2070, on constate une légère diminution du nombre total et de la distance totale de déplacements. Néanmoins, les individus tendent à plus prendre la voiture : ainsi, le nombre total de déplacements en voiture continue d'augmenter. En faisant en plus des hypothèses sur la composition du parc de véhicules, et notamment la part de véhicules électriques, ces résultats permettent de quantifier l'augmentation de la demande en électricité aux horizons considérés. En particulier, le logiciel CharME permet d'intégrer ces nouveaux jeux de pondérations pour prévoir la courbe de charge en 2040 et 2070, sous ce scénario d'évolution démographique.

La principale difficulté rencontrée durant cette étude a été le calcul des nouvelles pondérations. Nous avons vite réalisé qu'une méthode simple de type *"règle de trois"*, appliquée à une seule variable à la fois, risquait de biaiser les estimations. Nous avons donc construit une méthode permettant le contrôle de plusieurs variables, s'appuyant sur un algorithme de calage sur marges au niveau des individus, puis le calcul des pondérations jours pour obtenir des résultats représentatifs d'une semaine de mobilité.

Cette méthode présente plusieurs limites. Tout d'abord, elle est surtout adaptée pour des scénarios où l'on fait évoluer des variables socio-démographiques, comme par exemple notre scénario d'évolution démographique. On aurait aussi pu développer un scénario d'urbanisation ou de ruralisation, en faisant varier les variables décrivant le lieu d'habitation des ménages. En effet, il est clair que ces variables ont un impact significatif sur les déplacements²³. Néanmoins, des projections précises et fiables de ces

22. Le détail des projections est présenté en section 4.2.2.

23. Voir sur ce point l'annexe 9.A.

variables ne sont pas disponibles aux horizons considérés. A l'inverse, la méthode est peu adaptée pour des scénarios faisant évoluer des variables environnementales, ou des variables décrivant les pratiques individuelles²⁴. En effet, seules les variables présentes dans l'EMP et pour lesquelles on dispose de mesure des effectifs projetés peuvent être intégrées dans les scénarios. Les contraintes de calage sont telles que ces projections doivent être exprimées en effectifs (et pas en proportions d'un total non connu par exemple). C'est la raison pour laquelle nous n'avons pas pu obtenir de résultats pour le scénario réchauffement climatique.

S'agissant de la méthode même de repondération, nous avons opté pour l'utilisation de poids initiaux uniformes dans la fonction **calibrage**. Ces poids ne sont qu'une approximation et ne tiennent pas compte du plan de sondage. Ils ne corrigent pas non plus la non-réponse.

Enfin, il faut bien noter que la méthode repose surtout sur une hypothèse de départ : la non-évolution des pratiques de déplacement à variables de calage contrôlées. Ainsi, il s'agit d'une approche par les caractéristiques de population : on fait évoluer les caractéristiques de la population, mais on ne modifie pas la façon dont les individus, à caractéristiques contrôlées, se déplacent. Cette hypothèse peut paraître peu crédible. Il est raisonnable de penser que la mobilité des individus en 2040 ou 2070 ne sera pas la même que la mobilité en 2019, à variables modifiées égales par ailleurs. En effet, par exemple, des effets générationnels peuvent se superposer aux effets âge, les individus adaptent leurs pratiques aux facteurs environnementaux, les infrastructures évoluent, les politiques publiques ainsi que la conjoncture économique...

En tout état de cause, c'était avant tout la démarche technique et la méthodologie de recalcul des pondérations qui intéressait notre tuteur Benjamin Petiau et EDF. Malgré ses limites, la démarche mise en œuvre est donc un solide point de départ pour EDF dans la perspective d'élaborer d'autres scénarios prospectifs à l'avenir.

24. Par exemple, nous avons évoqué en introduction un scénario Télétravail, mais cette variable n'est pas disponible dans l'EMP.

8 Références

Références

- [1] INSEE. Projections de population par région et département entre 2024 et 2070.
- [2] Drias les futurs du climat.
- [3] ADEME. Transition(s) 2050.
- [4] SDES. Parc et circulation des véhicules routiers.
- [5] Ministère de la Transition écologique et de la Cohésion des territoires. Enquête de mobilité des personnes.
- [6] Olivier Sautory. Les méthodes de calage, 2018.
- [7] Antoine Rebecq. Icarus : un package r pour le calage sur marges et ses variantes, 2016.

9 Annexes

9.A Analyse descriptive

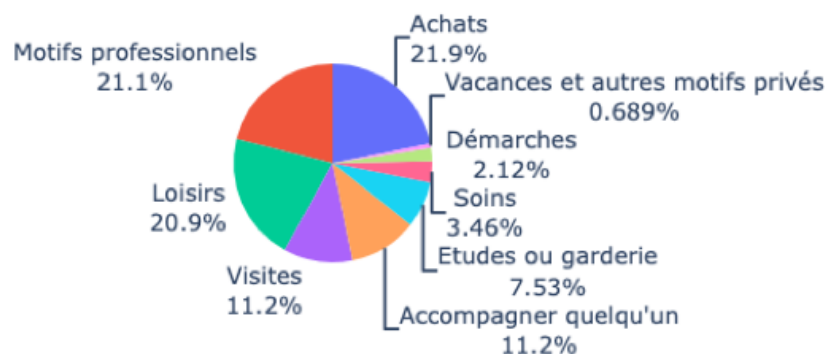


FIGURE 23 – Répartition des déplacements selon le motif (sans retour au domicile)

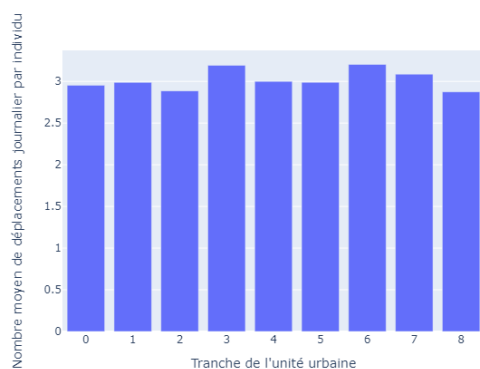


FIGURE 24 – Nombre de déplacements en fonction de la tranche d'unité urbaine

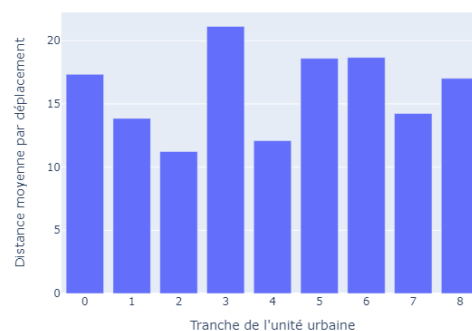


FIGURE 25 – Distance moyenne d'un déplacement en fonction de la tranche d'unité urbaine

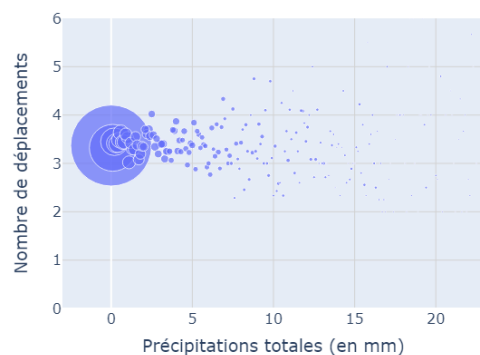


FIGURE 26 – Nombre de déplacements en fonction des précipitations ²⁵

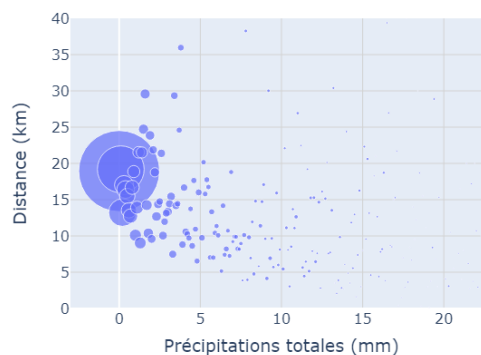


FIGURE 27 – Distance moyenne d'un déplacement en fonction des précipitations

25. NB : Sur ce graphique et le suivant, la taille des points indique le nombre d'observations sur lequel le calcul a

TABLE 1 – Description des tranches d'unité urbaine

Tranche	Description
0	Commune hors unité urbaine
1	Commune appartenant à une unité urbaine de 2 000 à 4 999 habitants
2	Commune appartenant à une unité urbaine de 5 000 à 9 999 habitants
3	Commune appartenant à une unité urbaine de 10 000 à 19 999 habitants
4	Commune appartenant à une unité urbaine de 20 000 à 49 999 habitants
5	Commune appartenant à une unité urbaine de 50 000 à 99 999 habitants
6	Commune appartenant à une unité urbaine de 100 000 à 199 999 habitants
7	Commune appartenant à une unité urbaine de 200 000 à 1 999 999 habitants
8	Commune appartenant à l'unité urbaine de Paris

	Nombre d'individus ayant été enquêtés	Nombre d'individus ayant rapporté leurs déplacements
Lundi	2897	1799
Mardi	2423	2238
Mercredi	2871	2068
Jeudi	2398	1956
Vendredi	2060	2295
Samedi	718	1633
Dimanche	3	1381

Champ : parmi les 13370 enquêtés ayant rapporté leurs déplacements locaux.

Lecture : 2897 individus ont été enquêtés un lundi. 1799 individus ont décrit leurs déplacements au cours du dernier lundi précédent le jour d'enquête.

FIGURE 28 – Nombre d'individus selon le jour d'enquête et le jour d'interrogation

été fait dans la base. Ainsi, les points les plus petits sont peu fiables. Contrairement aux analyses pour la température, nous n'avons pas représenté les déplacements en fonction de l'écart de précipitation avec les jours voisins. En effet, les observations sont déjà très condensées en valeurs absolues, les analyses seraient donc très limitées en valeurs relatives. Par ailleurs, les niveaux de précipitations ont été arrondis au dixième d'unité le plus proche.

9.B Méthode de repondération

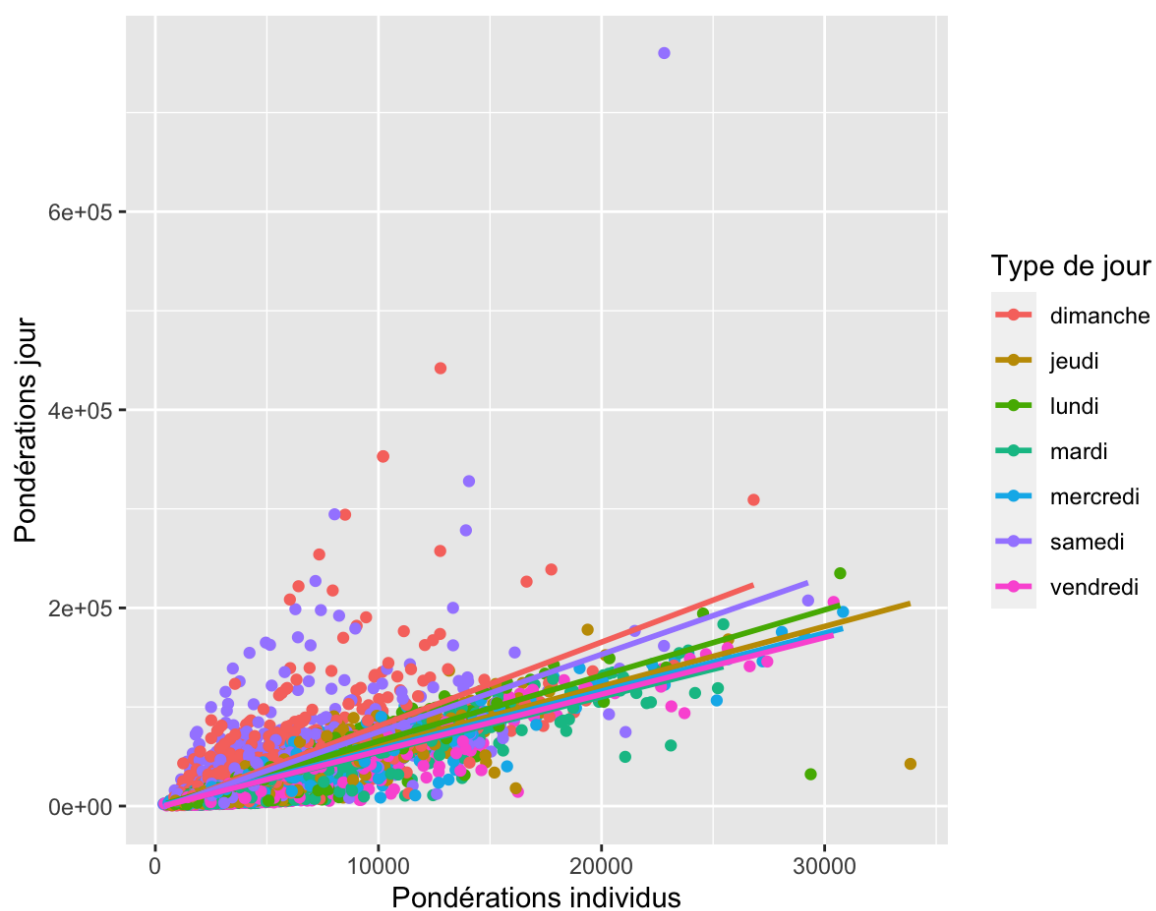


FIGURE 29 – Relation entre pondérations individus et pondérations jour

```
##### Summary of before/after weight ratios #####  
Calibration method : linear  
Mean : 1.0909  
0%    1%    10%   25%   50%   75%   90%   99%  100%  
0.6927 0.6927 0.7341 0.8079 1.0423 1.2889 1.5592 2.4292 2.5444
```

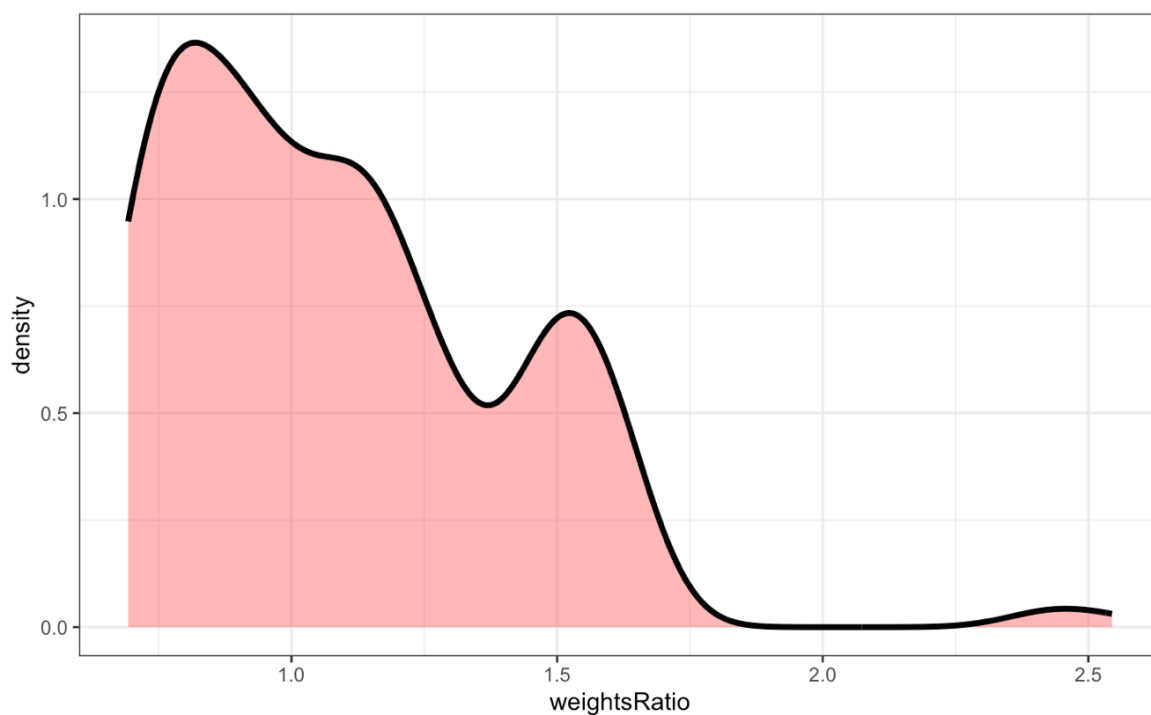


FIGURE 30 – Rapports des poids après calcul des pondérations individus pour le scénario démographie 2040

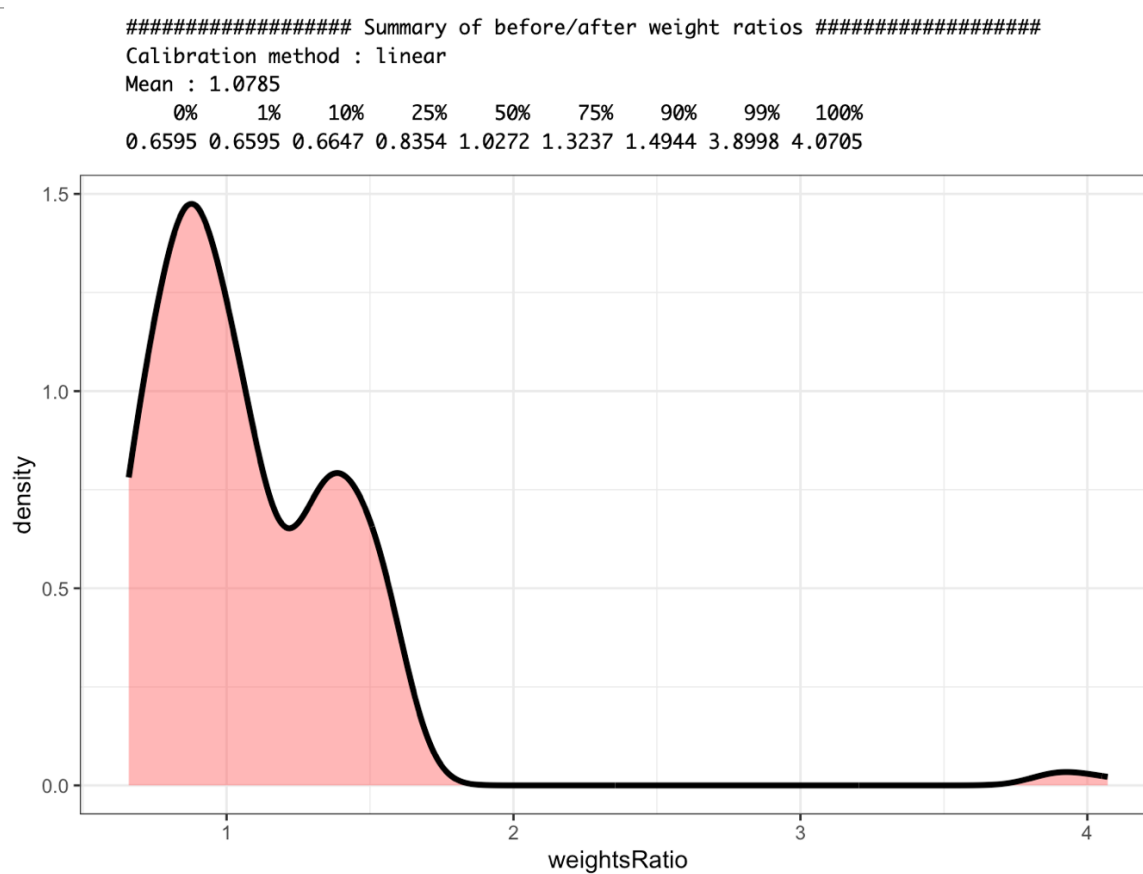


FIGURE 31 – Rapports des poids après calcul des pondérations individus pour le scénario démographie 2070

9.C Résultats

Mode de déplacement	Variable d'intérêt	2019	2040	2070
Résultats généraux	Pour rappel : population totale (6 ans et plus)	5,95E+07	6,49E+07	6,42E+07
	Nombre total de déplacements en une semaine	1,22E+09	1,25E+09	1,23E+09
	Nombre total de déplacements en une journée	1,74E+08	1,78E+08	1,76E+08
	Nombre moyen de déplacements par jour par individu	2.93	2.74	2.75
	Nombre moyen de déplacements par jour par individu EN SEMAINE	3.11	2.92	2.93
	Nombre moyen de déplacements par jour par individu LE SAMEDI	2.76	2.62	2.59
	Nombre moyen de déplacements par jour par individu LE DIMANCHE	2.15	1.99	2
	Distance totale parcourue en une semaine	1,56E+10	1,59E+10	1,57E+10
	Distance totale parcourue en une journée	2,22E+09	2,26E+09	2,25E+09
	Distance moyenne par jour par individu	37.38	34.9	35.02
	Distance moyenne par jour par individu EN SEMAINE	36.81	34.18	34.24
	Distance moyenne par jour par individu LE SAMEDI	39.13	37.19	36.79
	Distance moyenne par jour par individu LE DIMANCHE	38.48	36.22	37.16
	Part de la population n'ayant fait aucun déplacement dans les 7 jours précédents	2.63%	3.65%	3.2%
	Nombre de jours avec déplacements dans une semaine	5.89	5.74	5.75
	Heure de départ moyenne du premier déplacement de la journée	10:04	10:23	10:24
	Heure d'arrivée moyenne du dernier déplacement de la journée	17:34	17:19	17:12
En voiture	Nombre total de déplacements en une journée	1,11E+08	1,16E+08	1,18E+08
	Nombre moyen de déplacements par jour par individu	1.86	1.79	1.84
	Distance totale parcourue en une journée	1,66E+09	1,72E+09	1,72E+09
	Distance moyenne par jour par individu	27.94	26.58	26.87
	Part de la population ayant fait au moins un déplacement avec ce mode en une journée	68.5%	68.8%	70.7%
	Nombre moyen de déplacements par jour pour les individus qui utilisent le mode	2.71	2.6	2.6
A pied	Distance moyenne par jour par individu pour les individus qui utilisent le mode	40.8	38.62	38.01
	Nombre total de déplacements en une journée	4,14E+07	4,10E+07	3,97E+07
	Nombre moyen de déplacements par jour par individu	0.7	0.63	0.62
	Distance totale parcourue en une journée	4,33E+07	4,42E+07	4,38E+07
	Distance moyenne par jour par individu	0.73	0.68	0.68
	Part de la population ayant fait au moins un déplacement avec ce mode en une journée	29,6%	28,8%	28,4%
En transports en commun	Nombre moyen de déplacements par jour pour les individus qui utilisent le mode	2.35	2.19	2.18
	Distance moyenne par jour par individu pour les individus qui utilisent le mode	2.46	2.36	2.4
	Nombre total de déplacements en une journée	1,31E+07	1,25E+07	1,11E+07
	Nombre moyen de déplacements par jour par individu	0.22	0.19	0.17
	Distance totale parcourue en une journée	1,54E+08	1,44E+08	1,41E+08
	Distance moyenne par jour par individu	2.59	2.22	2.2
	Part de la population ayant fait au moins un déplacement avec ce mode en une journée	11,9%	11%	10%
	Nombre moyen de déplacements par jour pour les individus qui utilisent le mode	1.85	1.74	1.72
	Distance moyenne par jour par individu pour les individus qui utilisent le mode	21.68	20.11	21.98

Note : les distances sont exprimées en kilomètres.

FIGURE 32 – Résultats détaillés du scénario démographique à horizons 2040 et 2070