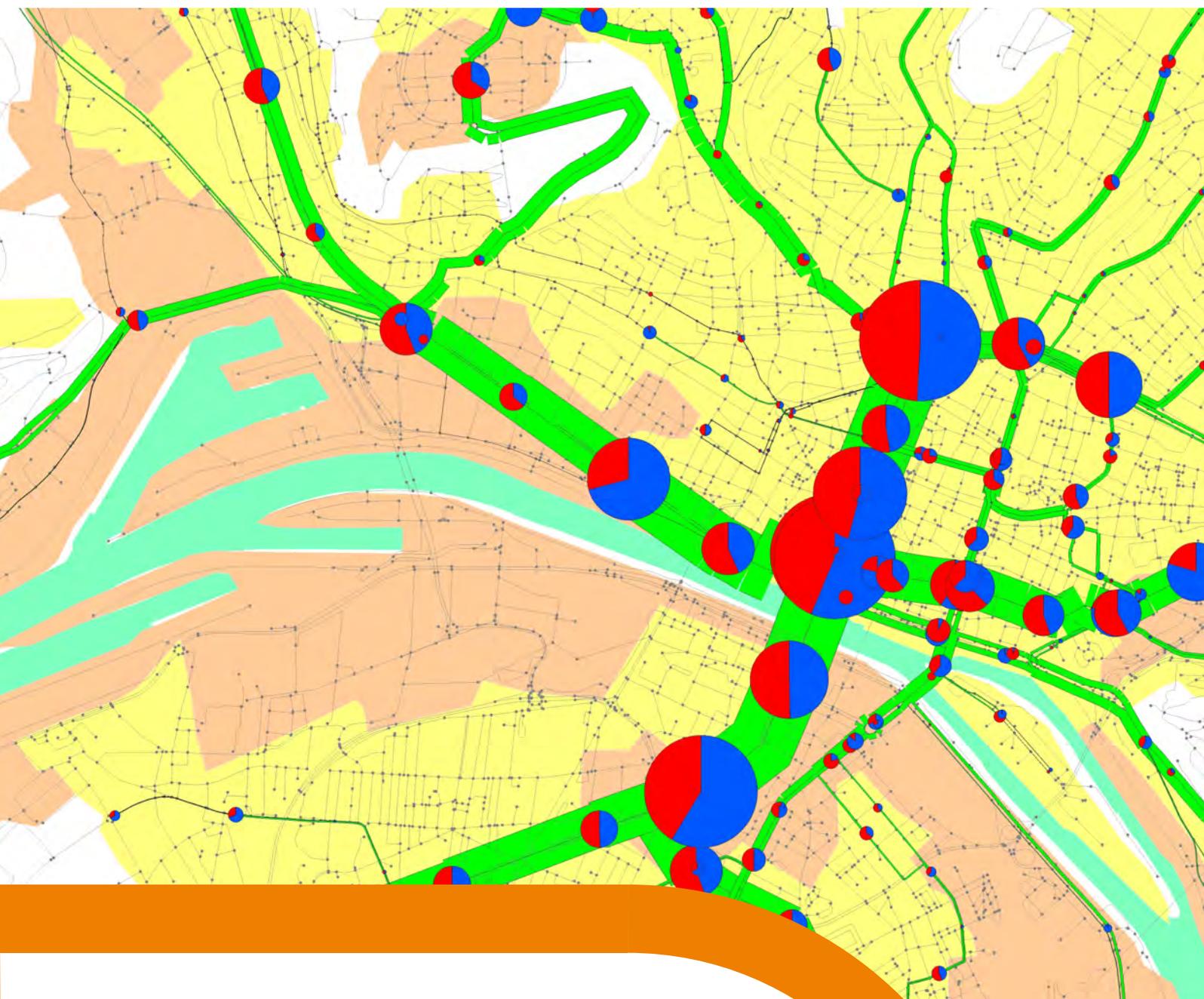


Modélisation multimodale des déplacements de voyageurs
Concevoir un modèle de choix modal



*Modélisation multimodale
des déplacements de voyageurs*

Concevoir un modèle de choix modal

Références

Cette collection regroupe l'ensemble des documents de référence portant sur l'état de l'art dans les domaines d'expertise du Cerema (recommandations méthodologiques, règles techniques, savoir-faire...), dans une version stabilisée et validée. Destinée à un public de généralistes et de spécialistes, sa rédaction pédagogique et concrète facilite l'appropriation et l'application des recommandations par le professionnel en situation opérationnelle.

Remerciements

La publication de cet ouvrage a été rendue possible grâce au soutien de la Direction générale des infrastructures des transports et de la mer (DGITM) du ministère de l'Écologie, du Développement durable et de l'Énergie. qui a confié au Cerema le soin d'éditer des guides méthodologiques de référence sur le domaine de la modélisation afin d'identifier les « bonnes pratiques » en la matière. Ces guides doivent servir aussi bien aux modélisateurs qu'aux maîtres d'ouvrage souhaitant s'assurer de la qualité des prévisions de trafic réalisées.

Le présent guide est le résultat d'un groupe de travail thématique animé par le Cerema et qui s'est déroulé de 2012 à 2014.

C'est un ouvrage collectif du Cerema : Thomas Anselme (Cerema-Normandie Centre) et Mathieu Jacquot (Cerema-Est) ont piloté la rédaction du document, en association avec :

Michael Savary (Cerema-Normandie Centre) ;

David Savalle (Cerema-Normandie Centre) ;

Vincent Perrier-David (Cerema-Est) ;

Nicolas Speisser (Cerema-Est).

Le travail de coordination de l'ouvrage a été assuré par Damien Verry (Cerema-Territoires et ville) et Charlotte Coupé (Cerema-Infrastructures, Transports et Matériaux)

Nous remercions pour leur relecture attentive de l'ouvrage :

Mathieu de Lapparent (EPFL) ;

Patrick Bonnel (LET) ;

Jacques Bougnol (Cerema-Ouest) ;

Patrick Palmier (Cerema-Nord Picardie).

Nous remercions pour leurs contributions et conseils l'ensemble des personnes ayant pu participer à une ou plusieurs journées de travail : Jean-Baptiste Ray (Arcadis), Samuel Mainguy (Arcadis), Cunxiao Zhang (Citec), Matthias Lenz (PTV), Éric Magdelenat (DREIA IF), Justine Song (DREIA IF), Isabelle Cabanne (CGDD), et pour le Cerema : Aurélie Bousquet, François Combes, Richard Grimal, Étienne Hombourger, Aurélie Herve, Maxime le Corre, Samuel Melennec et Olivier Richard.

Sommaire

1. Introduction	5
1.1 Contexte de création de ce guide	5
1.2 Champs d'application	5
1.3 Démarche adoptée	6
2. Principes de base	8
2.1 Choix des alternatives	8
2.2 Théorie des choix discrets	9
2.3 Quelle formulation de l'utilité ?	13
2.4 Disponibilité des alternatives	15
3. Bases de données	16
3.1 Bases de données disponibles	16
3.2 Avec quelles bases de données estimer le modèle ?	17
3.3 Combinaison d'enquêtes	18
3.4 Les bases de données sur l'offre de transports	20
4. Du bon calcul du coût généralisé	23
4.1 Périodes de modélisation	23
4.2 Coût généralisé TC	24
4.3 Coût généralisé VP	26
4.4 Coût généralisé MAP	29
4.5 Cas particulier des coûts généralisés intrazonaux	30
4.6 Incidence de la congestion sur le calibrage du modèle	31
5. Les outils statistiques disponibles pour estimer le modèle de choix modal	32
5.1 Les indicateurs statistiques de base	32
5.2 La propriété IIA	38
5.3 L'endogénéité	39
5.4 Le biais d'agrégation	40
5.5 La segmentation	40
6. Processus de construction d'un modèle	43
6.1 Problème préliminaire : le modèle dans son univers d'utilisation	43
6.2 Proposition d'un processus pratique de construction du modèle	45
6.3 Différents problèmes spécifiques	68
6.4 Valider le modèle	76
7. Faire des simulations avec un modèle de choix modal	78
7.1 Méthodes d'application du modèle calibré	78
7.2 Analyser les élasticités	82
7.3 Intégrer des variables supplémentaires	85
7.4 Réutiliser un autre modèle	85
8. Perspectives	89
9. Annexes	90

9.1 Annexe 1 - modèles agrégés	90
9.2 Annexe 2 - méthode de décomposition des flux VP en flux VPC/VPP	91
10.Bibliographie	94
11.Liste des sigles	98

1. Introduction

1.1 Contexte de création de ce guide

La volonté accrue de réduire l'usage de la voiture au profit des modes alternatifs s'est traduite notamment par un renforcement du rôle des plans de déplacements urbains et un engagement en faveur du mode ferroviaire, ainsi que par l'émergence de politiques de transports globales visant à optimiser l'ensemble du système de déplacements tous modes confondus, y compris leurs interfaces.

Les modèles de choix modal sont au centre de ce processus. Ils permettent en effet d'effectuer un diagnostic de la situation actuelle, notamment via la détermination des facteurs explicatifs des parts de marché des différents modes, facteurs qui constituent autant de leviers d'action potentiels pour la puissance publique. Les modèles permettent également de tester l'influence de la mise en place de diverses politiques, par exemple le renforcement des fréquences de transport collectif, la mise en place d'une nouvelle ligne ou d'un nouveau mode de transport, sur les parts de marché futures.

Cette évolution du contexte a imposé un renouvellement des méthodes de modélisation des déplacements dans le réseau scientifique et technique (RST) du ministère de l'Écologie, du Développement durable et de l'Énergie (MEDDE), et en particulier un regain d'intérêt pour la modélisation du choix modal.

Ce guide a pour vocation de synthétiser ces avancées en formulant un certain nombre de recommandations pratiques pour la constitution d'un modèle de choix modal, afin de garantir un meilleur usage des modèles comme outils d'aide à la décision des politiques publiques. Le guide s'applique à tous les modèles multimodaux, tout en s'appuyant fortement sur des exemples puisés dans le domaine de la modélisation multimodale plutôt urbaine.

En l'état des connaissances et des pratiques, l'analyse est restreinte aux choix de mode effectués par des voyageurs pour leurs déplacements du quotidien. Les déplacements interurbains, peu fréquents dans ce cas, ne sont donc pas traités dans ce document.

1.2 Champs d'application

Les modèles de choix de mode sont marqués par le débat entre deux types d'approche concernant le niveau d'agrégation de l'analyse. La plupart du temps, on oppose l'approche agrégée, où l'on raisonne à partir d'un individu « moyen », à l'approche désagrégée, où cette fois chaque individu est pris en considération séparément par l'intermédiaire de ses propres caractéristiques.

Parmi les **modèles agrégés** les plus utilisés, on peut citer par exemple les grilles et les courbes de partage modal, ainsi que les modèles prix-temps. L'avantage de ces modèles agrégés est leur simplicité d'estimation et d'utilisation.

On retrouve aussi, souvent classés dans cette famille des modèles agrégés, les modèles logit multinomiaux (parfois également classés dans la famille des **modèles désagrégés**). Dans les deux cas, il s'agit de modèles de choix discrets (la famille de modèles dont il est question dans ce guide). Mais dans un cas (modèles agrégés), les paramètres du modèle ne dépendent généralement que des caractéristiques de l'offre de transport, alors que dans l'autre (modèles désagrégés), ils dépendent aussi des caractéristiques de l'individu considéré.

La différence entre ces deux types d'approche porte également sur le type de données utilisées pour l'estimation, et sur la méthode d'estimation : les modèles désagrégés ne peuvent être estimés que sur des données désagrégées, alors que les modèles agrégés peuvent être estimés aussi bien sur des données agrégées que des données désagrégées. Enfin, la méthode d'estimation la plus souvent utilisée pour estimer des modèles agrégés est la régression linéaire, alors que les modèles désagrégés appellent couramment en pratique l'utilisation de la méthode du maximum de vraisemblance. Pour plus de détails sur les

modèles agrégés les plus fréquemment utilisés en pratique et sur leur méthode d'estimation, le lecteur pourra utilement se reporter à l'annexe 1.

Ainsi, ce guide sera consacré à l'étude des **modèles de choix discrets « estimés sur données désagrégées »**. Ces modèles s'intègrent rigoureusement dans le cadre de l'analyse des choix discrets (Cf. § 2.2) et de la théorie de la maximisation de l'utilité aléatoire. Le niveau de désagrégation du modèle lui-même (c'est-à-dire de ses variables) sera un des points abordés dans ce guide.

Les modes de transport en compétition sont les suivants :

- Véhicule particulier (VP), qui pourra éventuellement être décomposé en :
 - VP conducteur et VP passager
 - VP seul et VP à plusieurs
- Transports en commun (TC), urbains et interurbains (à l'exception des TGV, et GL¹)
- Marche à pied (MAP)
- Enfin, les pratiques émergentes et nouveaux services de mobilité (vélo, intermodalité VP/TC, covoiturage, autopartage, et transport à la demande) seront également abordés dans ce guide, même si pour certains les possibilités et l'intérêt de la modélisation sont en pratique limités voire nuls à l'heure actuelle

Les modèles dont il est question ici peuvent s'appliquer aussi bien sur un déplacement que sur une boucle de déplacements. L'estimation conjointe avec un modèle de choix de destination n'est pas entrée dans les pratiques : elle impose d'utiliser un système de tirage aléatoire d'un nombre réduit de destinations lors de l'estimation du modèle, qui sera plus ou moins difficile à mettre en œuvre en fonction de la spécification retenue pour le modèle (Cf. § 2.2). En revanche, on retrouve souvent dans la pratique des modèles de choix de destination et de mode estimés de façon séquentielle, et où un indicateur composite des fonctions d'utilité est utilisé dans la distribution. L'approche présentée dans ce guide pourra donc s'inscrire dans ce dernier type d'utilisation.

1.3 Démarche adoptée

Pour rédiger ce guide, l'ensemble de la communauté des modélisateurs du réseau scientifique et technique du ministère de l'Écologie, du Développement durable et de l'Énergie a été mobilisé dans le cadre d'un groupe de travail dédié, copiloté par les directions techniques Infrastructures de Transports et Matériaux et Territoires et Ville (ex-SETRA et CERTU), et le PCI MOD2* (Directions territoriales Normandie-Centre et Est, ex-CETE Normandie-Centre, et ex-CETE de l'Est) du Cerema.

La méthode de travail adoptée pour concevoir ce guide a reposé sur les trois piliers que sont la capitalisation des expériences de modélisations multimodales existantes, qu'elles soient réalisées par le Cerema ou les bureaux d'études privés, la réalisation de travaux exploratoires sur des modèles existants pour approfondir un certain nombre de points techniques, et l'appropriation des publications issues du monde de la recherche les plus récentes en la matière.

¹ GL : Grandes Lignes

* Pôle de Compétence et d'Innovation (PCI) Méthodes, outils et démarches pour la modélisation et l'organisation des déplacements (MOD2).

De manière à faciliter la lecture de ce guide, chaque point sera abordé selon la trame suivante :

- problématisation pour comprendre les enjeux,
- état des meilleures pratiques observées dans la littérature ou dans les modèles existants,
- identification des pratiques à éviter,
- état des pratiques futures (c'est-à-dire souhaitables, mais pas forcément réalisables compte tenu des connaissances actuelles ou des données disponibles).

2. Principes de base

2.1 Choix des alternatives²

Le choix des modes de transport qui composeront le modèle de choix modal est une étape déterminante. Le choix des alternatives proposées résulte d'un arbitrage entre :

- la recherche d'une qualité minimale : l'intégration des principaux modes semble indispensable, ne serait-ce que pour bien représenter les principales possibilités qu'ont les usagers au moment de leur choix de mode et donc éviter des comportements aberrants (ex : modèle MAP / TC seul, qui ne traiterai pas de la concurrence avec la VP, y compris pour les déplacements longs...) ;
- les effets de politiques que l'on cherche à mesurer ; c'est généralement le souhait de tester de nombreux effets différents qui pousse à chercher à intégrer le plus de modes possibles ;
- les données disponibles : à l'inverse, ce facteur oblige souvent à limiter le nombre d'alternatives du modèle, la taille des échantillons relatifs aux modes peu usités étant souvent trop faible ;
- la complexité du modèle de choix modal, lors de son estimation et pour son application : il faut être en mesure de décrire précisément les caractéristiques de chaque alternative que l'on cherche à intégrer au modèle, ce qui nécessite souvent des efforts assez importants (Cf. § 4) ; par ailleurs l'augmentation du nombre d'alternatives oblige souvent à complexifier le modèle de choix modal, notamment en termes de structure, ce qui génère des temps de calcul plus importants.

Pour répondre au premier point, la sélection des modes VP, TC, et MAP semble indispensable si l'on veut garantir la qualité minimale du modèle.

En ce qui concerne la problématique de la décomposition des modes VP selon différentes modalités, le choix à effectuer correspond, par exemple, à des logiques diverses :

- mesurer l'effet de voies de covoiturage, ce qui impose de disposer d'une matrice VP à plusieurs /VP solo pour l'affectation ;
- améliorer la pertinence du modèle en ne proposant pas à l'individu des modes non disponibles (Cf. § 2.4), comme le mode VP conducteur (modèle décomposé VP conducteur /VP passager) à une personne n'ayant pas le permis de conduire.

En tout état de cause, aucune pratique ne s'est imposée aujourd'hui. La méthode développée par la DTer Ouest du Cerema et présentée à l'annexe 2 est également une possibilité intéressante : elle consiste à retenir un mode VP unique pour le choix modal, puis à répartir en VP conducteur / VP passager selon une grille de répartition.

Par ailleurs, la différenciation des modes TC pourra être envisagée en vue d'évaluer des projets ou des politiques spécifiques à un sous-mode TC donné, et d'estimer si les comportements des usagers sont vraiment différenciés en fonction de ces sous-modes : la décomposition entre les transports urbains et interurbains (à condition de disposer de données sur ces derniers) et/ou la décomposition modes TC légers / modes TC lourds (Cf. § 6.3.1) pourront ainsi être utilisées.

Enfin, l'intégration des nouveaux services et nouvelles pratiques de mobilité représente encore un défi important (Cf. § 6.3.3), qui n'a de sens que si les données le permettent et si certaines politiques à tester portent spécifiquement sur ces modes.

² Dans les publications, le terme « univers de choix » est parfois préféré à la notion d'alternative plus couramment usitée dans le domaine des transports.

2.2 Théorie des choix discrets

2.2.1 Principe de l'utilité aléatoire

Les différentes alternatives qui se présentent à un usager dans le choix de son mode de transport sont limitées en nombre et facilement énumérables.

Aussi, la théorie de l'analyse des choix discrets, qui sert à modéliser les comportements de choix des individus parmi un ensemble fini d'alternatives, énumérables et exclusives, est utilisée dans les modèles de choix modal. De nombreuses règles de choix ont été proposées dans la littérature. [BEN-AKIVA et LERMAN, 1985] cite quatre règles : la dominance, la satisfaction, l'ordre lexicographique et l'utilité. Parmi ces exemples, le concept d'utilité a la particularité d'être fondé sur le postulat de commensurabilité des attributs. Ainsi, une valeur unique est attribuée à l'attractivité de chaque alternative : l'utilité, que l'individu cherche à maximiser. Cette théorie présente l'avantage considérable d'être particulièrement adaptée à la réalisation d'analyses mathématiques et de tests statistiques.

L'utilité pour l'individu i de l'alternative m sera notée $U_{i,m}$; elle sera fonction de ses caractéristiques (par ex. le temps de parcours pour le choix du mode de transport) et des préférences propres de l'individu (par ex. spécifiques à sa catégorie socioprofessionnelle) .

La probabilité que l'individu i choisisse l'alternative m est supposée être une fonction des utilités de chacune des alternatives disponibles, dont les paramètres sont estimés sur des recueils de données individuelles afin de reproduire les comportements observés. Le modélisateur considère par la suite que ce modèle est explicatif et qu'il fournit donc des enseignements utiles pour tester les choix effectués par les individus en modifiant différentes caractéristiques des alternatives et/ou en analysant les choix effectués par différents types de population.

Dans les modèles actuels, l'utilité d'un mode de transport m pour un individu i est couramment définie par la formule :

$$U_{i,m} = [\beta_{m,s}] \times f_X(X_{i,m}, \lambda_{m,s}) + ASC_{m,s} + \varepsilon_{i,m}$$

où :

- i : l'individu
- m : le mode de transport
- s : le segment de la population auquel appartient l'individu i
- $[\beta]$: les paramètres, qui peuvent être distribués au sein de la population (par ex. la valeur du temps)
- X : les variables explicatives du comportement de choix modal (par ex : le temps de parcours, le coût, le temps d'attente)
- f_X : une fonction généralement de forme linéaire, mais pouvant également être un logarithme népérien (modèle de Kirchhoff ou loi d'Abraham) ou une transformation puissance, la plus utilisée étant la transformation Box-Cox (voir encadré ci-après) – λ étant le vecteur de paramètre utilisé dans la fonction
- ASC : les constantes d'alternatives³ qui peuvent être spécifiques aux modes et aux segments de population. Ces constantes intègrent les termes que l'on ne sait pas expliquer à l'aide des seules variables explicatives X de la fonction d'utilité
- ε : la portion aléatoire de l'utilité, autrement appelée le terme d'erreur (non corrélé aux variables explicatives)

3 La distinction de vocabulaire entre les paramètres β et les constantes d'alternatives ASC , même si les deux sont en fait des paramètres du modèle, est maintenue dans l'intégralité de ce document. Le terme constantes d'alternatives a été étendu par abus de langage aux paramètres pondérant les variables socioéconomiques (par ex. une valeur binaire exprimant l'appartenance ou non de l'individu i au segment de population s). Il nous apparaît important de bien différencier les deux types de paramètres, notamment pour mettre en avant ceux qui sont associés aux variables sur lesquelles les politiques de transports ont un impact direct (temps de parcours par exemple).

Transformation Box-Cox

$$x_i(\lambda, \alpha) = \begin{cases} \frac{(x_i + \alpha)^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln(x_i + \alpha) & \text{si } \lambda = 0 \end{cases}$$

Remarque : l'application f_x peut également consister à transformer une variable continue en une variable discrète. Cette transformation ne s'applique qu'à des variables à valeurs réelles strictement positives.

La forme de la distribution du terme d'erreur joue un rôle central, car elle détermine la structure, les propriétés, ainsi que les limites du modèle.

Dans le cas où les termes d'erreur sont supposés suivre une loi normale, on définit le modèle probit, que l'on ne détaillera pas ici car il n'est quasiment jamais utilisé au niveau opérationnel. Lorsque les termes d'erreur sont indépendants et identiquement distribués selon une loi de Gumbel, on est dans le cas du modèle logit multinomial (Cf. § 2.2.2), qui est le modèle le plus couramment utilisé.

2.2.2 Cas du modèle logit multinomial (MNL)

Le modèle le plus utilisé en pratique, du fait de sa forme analytique extrêmement simple, est le modèle logit multinomial (MNL). Ce modèle, qui est principalement fondé sur trois hypothèses (Cf. § 2.2.2.1), implique en effet la formule suivante de détermination de la probabilité d'utilisation d'un mode :

$$P_m(i) = \frac{e^{\mu \times U_{i,m}}}{\sum_n e^{\mu \times U_{i,n}}}$$

où :

- μ est un paramètre appelé le facteur d'échelle du modèle.

Dans les modèles MNL, μ prend une valeur conventionnelle égale à 1. Il dépend de la variance des termes d'erreur au sein de la population étudiée : il en résulte que les paramètres et constantes d'alternatives de différents modèles de choix modal ne sont pas comparables directement mais uniquement au facteur d'échelle près.

2.2.2.1 Hypothèses du modèle logit multinomial (MNL)

Le modèle logit multinomial repose sur l'hypothèse que les parties aléatoires de l'utilité des différentes alternatives sont **indépendantes** entre elles et **identiquement distribuées** (IID) selon une **loi de Gumbel**. Cette formulation mathématique a deux conséquences importantes :

- La covariance entre les termes d'erreur de deux alternatives distinctes m et n est supposée nulle. Cela signifie qu'il n'y a pas de facteurs non observés, constituant le modèle, qui soient communs à ces deux alternatives (autrement dit, le modèle n'est valable que si la préférence pour l'alternative m est indépendante de celle pour l'alternative n , hors composante systématique de l'utilité). Cette propriété s'appelle la propriété d'indépendance vis-à-vis des autres alternatives (propriété IIA en anglais, fréquemment citée dans la littérature).
- La variance des termes d'erreur est identique pour toutes les alternatives et tous les individus. Il n'y a donc pas de variation de la perception des facteurs non observés en fonction des alternatives (autrement dit, il n'y a pas de différence entre la préférence pour une alternative m et celle pour tout autre alternative n , hors composante systématique de l'utilité).

La pratique courante des modèles logit multinomiaux implique généralement une troisième hypothèse, qui permet, par sa simplicité, de réaliser des gains importants en temps de calcul : il n'y a pas de variation de la perception au sein des groupes homogènes d'individus. Autrement dit les paramètres relatifs à chacune des variables doivent être identiques pour tous les individus d'un même groupe.

2.2.2.2 Propriété d'indépendance vis-à-vis des autres alternatives (propriété IIA)

Le non-respect de la **propriété IIA** est une cause importante de problèmes de spécification des modèles logit multinomiaux, conduisant à des résultats inutilisables en prévision. Cette propriété Celle-ci implique que le rapport de probabilité de choix entre deux alternatives **m** et **n** dépend seulement de la composante observée de l'utilité de ces deux alternatives, selon la formule :

$$\frac{P(m)}{P(n)} = \frac{e^{U_m}}{e^{U_n}}$$

Cette propriété est assez contraignante, il arrive fréquemment dans les modèles de choix modal qu'elle ne soit pas respectée. Prenons l'exemple du choix entre trois modes de transport : la voiture, une ligne de bus bleue et une ligne de bus rouge. Supposons en outre que les deux lignes de bus partagent exactement les mêmes caractéristiques, hormis leur couleur. Imaginons maintenant qu'un modélisateur cherche à estimer deux modèles différents :

- un modèle où le bus est en concurrence avec la voiture, indépendamment de sa couleur,
- un autre modèle où le bus bleu, le bus rouge et la voiture sont en concurrence.

L'utilisation des propriétés IIA de ces différents modèles conduit à l'égalité suivante :

$$U_{bus} = \ln(e^{U_{busbleu}} + e^{U_{busrouge}})$$

$$U_{busbleu} = U_{busrouge} = U_{bus}$$

Or Il en résulte :

$$U_{bus} = \ln(2) + U_{bus}$$

On aboutit à une contradiction, ce qui prouve que le modèle logit multinomial ne peut traiter convenablement des situations de concurrence où certaines périodes sont uniquement assurées par le bus bleu et d'autres périodes à égalité par le bus rouge et le bus bleu.

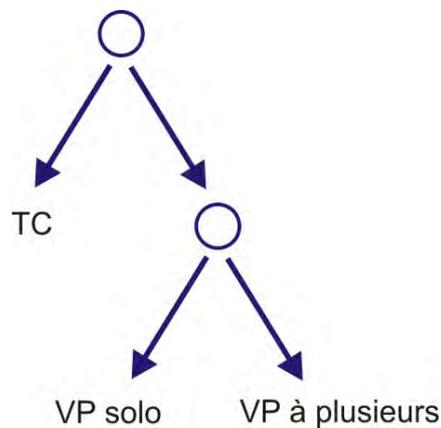
En conséquence, il faut recourir à une autre structure de modèle, dite nested logit (voir plus loin), qui est un des moyens de limiter les erreurs de spécification liées au non-respect de la propriété IIA.

Cette propriété IIA est donc très utile (car simplificatrice) mais elle constitue également la faiblesse du modèle logit multinomial. Dans le cas où les termes d'erreur de certaines alternatives sont corrélés, il est nécessaire de recourir à d'autres modèles permettant d'intégrer des valeurs non nulles des covariances entre différents termes d'erreur. Les alternatives dont les termes d'erreur sont corrélés sont alors regroupées en « nids ». C'est le cas des modèles nested logit (covariance non nulle et identique entre les termes d'erreur du nid) et cross-nested logit (les covariances entre termes d'erreur du nid sont supposées non nulles ; leur valeur dépend du facteur d'échelle du nid et du degré d'appartenance de l'alternative au nid).

2.2.3 Autres types de modèles

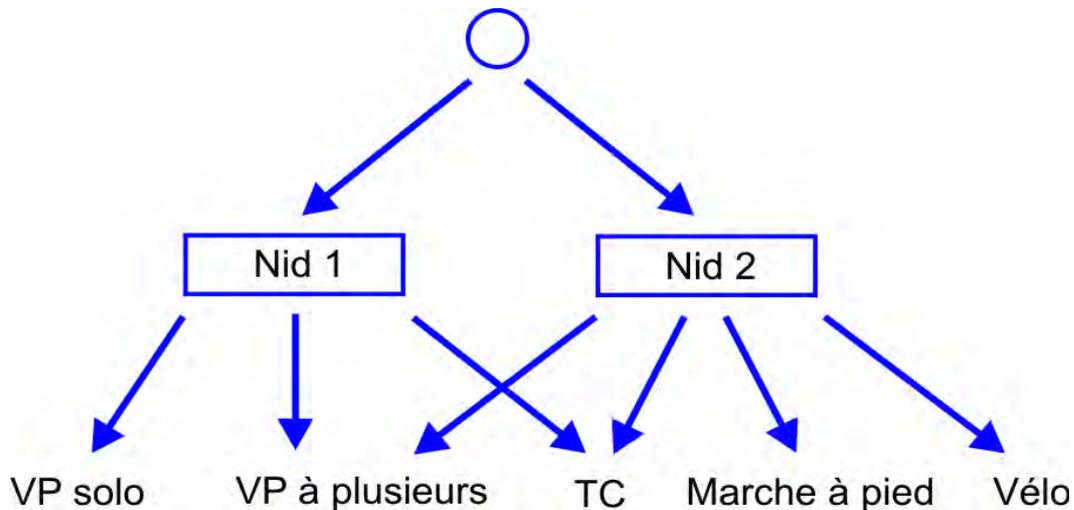
Le modèle **nested logit** (logit emboîté) permet d'introduire une covariance non nulle entre les termes d'erreur de deux ou plusieurs modes. Il en résulte que les violations les plus évidentes de la propriété IIA peuvent être évitées (comme entre les modes bus bleu et bus rouge de l'exemple du 2.2.2.2). Concrètement, on estime plusieurs modèles logit de façon hiérarchisée : le choix parmi les alternatives d'un même nid se fait à l'aide d'un modèle logit multinomial, et le nid constitue en lui-même une alternative dont l'utilité est égale au $\log\text{sum}^4$ des utilités des alternatives qui le composent (auxquelles on applique un facteur d'échelle qui est propre au nid). Ce résultat provient d'une autre propriété de la famille de modèles logit (voir par ex. [BEN-AKIVA et LERMAN, 1985]). Ainsi, si les termes d'erreur $(\varepsilon_1, \dots, \varepsilon_n)$ sont distribués selon une fonction Gumbel de paramètres $((\eta_1, \mu), \dots, (\eta_n, \mu))$, alors leur maximum est distribué selon une loi de Gumbel de paramètres $(\frac{\log\text{sum}(\eta)}{\mu}, \mu)$. D'après l'hypothèse de rationalité des individus, l'utilité de plusieurs modes est égale au maximum de l'utilité et a pour valeur moyenne $\frac{\log\text{sum}(\eta)}{\mu} - \frac{\gamma}{\mu}$ (γ est la constante d'Euler $\approx 0,57$). Il en résulte immédiatement que le $\log\text{sum}$ est la forme correcte pour évaluer l'utilité systématique de plusieurs modes.

Voici un exemple de modèle nested logit, dans le cas où les modes TC, VP solo et VP à plusieurs sont en concurrence et où l'on suppose que les termes d'erreur des modes TC et MAP sont corrélés :



4 Rappel : $\log\text{sum}(U_a, U_b) = \ln[\exp(U_a) + \exp(U_b)]$

Il est également possible de définir une appartenance d'un mode à plusieurs nids : on parle alors de modèle **cross-nested logit**. C'est le cas des modes TC et VP à plusieurs dans l'exemple suivant (ils appartiennent au nid 1 en tant que modes motorisés et au nid 2 en tant que modes alternatifs au mode VP conducteur) :



Enfin, tout en conservant la structure du modèle logit multinomial, il est possible de supposer que certains paramètres sont distribués au sein de la population (ex : la valeur du temps – Cf. § 6.3.2). L'estimation de ce modèle, qui est fondée sur la réalisation de tirages aléatoires pour le paramètre distribué, nécessite des temps de calcul plus importants.

2.3 Quelle formulation de l'utilité ?

La première composante de l'utilité est constituée du coût généralisé⁵ de transport, dont la formulation générale peut être résumée par la formule suivante :

$$CG_{i,m} = \beta_{m,s}^{temps} [revenus_i, distance_i] \times T_{i,m} + \beta_{m,s}^{coût} [revenus_i, distance_i] \times C_{i,m}$$

où :

- $CG_{i,m}$: coût généralisé pour l'individu i pour le mode m
- s : segment de la population auquel appartient l'individu i
- β^{temps} : paramètres pour la variable temps, qui peuvent dépendre des revenus et de la distance
- $T_{i,m}$: temps de parcours pour l'individu i pour le mode m
- $\beta^{coût}$: paramètres pour la variable coût, qui peuvent dépendre des revenus et de la distance
- $C_{i,m}$: coût payé par l'individu i pour le mode m

Les paramètres de pondération du temps devront être décomposés selon les différents éléments homogènes du trajet, de même que le coût de déplacement, si suffisamment d'éléments et de variabilité sont disponibles. Dans l'exemple du mode VP, les différents éléments du trajet seraient (sans exhaustivité) les temps d'accès à la voiture, de parcours en voiture, de recherche de stationnement et d'accès à la destination. Les différents éléments du coût pourraient être, dans ce cas, les coûts de stationnement, de carburant, de péage autoroutier, ainsi que le coût d'usure de la voiture.

⁵ Remarque : le guide propose une rédaction basée sur la notion de coût généralisé, mais le raisonnement peut porter, de façon équivalente, sur la notion de temps généralisé.

La définition correcte des éléments du coût généralisé fait l'objet d'une partie spécifique (Cf. § 4). Les éléments de coût de transport et de temps de transport sont reliés entre eux par la valeur du consentement à payer, communément appelée valeur du temps. Ce paramètre est une fonction du revenu de la personne, du mode de transport et de la distance de déplacement. Il peut également suivre une distribution au sein de la population de forme lognormale. Le calibrage de ce paramètre étant en général très délicat (insuffisance de variabilité dans les données, nécessité de recourir à des enquêtes de préférences déclarées...), des valeurs par défaut à retenir sont indiquées dans le paragraphe spécifique (Cf. § 6.3.2).

Pour toute variable continue, et en particulier pour les temps de parcours, les modèles tenant compte d'effets linéaires de la variable sur l'utilité sont des restrictions du cas général. Il est par exemple assez bien établi qu'une différence de temps de parcours de 5 minutes n'est pas perçue de manière identique pour un trajet de 5 km ou de 50 km. Un test de transformation Box-Cox des variables doit donc être envisagé de manière systématique dès lors que la quantité de données est suffisante au vu du nombre de paramètres et de constantes d'alternatives.

Enfin, outre les coûts généralisés de parcours des différents modes, l'utilité pourra être composée des éléments suivants :

- les contraintes de stationnement si les temps de recherche et coûts de stationnement ne sont pas explicitement pris en compte dans la formule de coût généralisé (Cf. [CERTU, 2009]) ;
- la proximité des arrêts au point de départ du trajet, qui est un élément incontournable pour diminuer l'endogénéité d'auto-sélection du mode TC (Cf. § 5.3) ;
- les paramètres de préférence d'alternatives, qui seront en lien binaire avec différents paramètres socioéconomiques (dont les principaux sont la motorisation, la possession du permis et la possession d'un abonnement TC).

A contrario, les éléments suivants, que l'on retrouve parfois dans certains modèles opérationnels, n'auront *a priori* pas besoin d'être intégrés dans le modèle, ou alors seulement en cas de nécessité impérative, car ils sont partiellement redondants avec les éléments qui viennent d'être cités :

- la densité de population et d'emploi sur la surface urbanisée d'une zone, qui permet de diminuer l'endogénéité d'auto-sélection (les usagers ayant une préférence pour les modes alternatifs à la voiture conducteur ont une plus grande propension à habiter en zones denses pour bénéficier d'un réseau de transports en commun de bonne qualité) ; cette information peut parfois servir à estimer les contraintes de stationnement ;
- la distance à vol d'oiseau au centre-ville, qui est un indicateur alternatif ou complémentaire au précédent ;
- la distance d'éloignement à la destination, qui permet de prendre en compte l'éventuelle croissance de la préférence pour un mode particulier avec la distance, et qui est calculée à partir de la distance à vol d'oiseau au sein de secteurs sans coupures physiques et de la distance à vol d'oiseau via un ou des points de passage obligés si des coupures physiques limitent les communications entre secteurs (ces points seront par exemple les ponts dans le cas où la coupure est une rivière, les tunnels ou cols s'il s'agit d'une barrière montagnaise...).

2.4 Disponibilité des alternatives

Dans la théorie fondatrice de ce type de modèle, la définition des alternatives constituant effectivement l'univers de choix des individus est un élément essentiel pour l'obtention d'un modèle correctement estimé, c'est-à-dire un modèle qui capte des sensibilités réalistes aux différentes variables. Il est ainsi évident que l'absence du permis de conduire et/ou de la disponibilité d'une voiture modifie complètement l'univers de choix des individus. Il convient également de citer :

- la proximité des arrêts TC, ce mode devenant indisponible si la distance à parcourir est trop longue ;
- les temps de parcours, qui, s'ils sont trop importants, peuvent rendre le mode en question indisponible dans l'esprit des individus ;
- la distance de déplacement, qui peut également présenter un seuil infranchissable pour les modes MAP et vélo ;
- la fréquence des TC, le nombre de correspondances, ou le caractère très indirect des trajets imposant de passer par le centre-ville pour des déplacements de rocade, qui peuvent rendre le mode TC indisponible.

Ainsi, le modélisateur devra définir différentes situations pour lesquelles certains modes deviennent indisponibles. Dans la mesure où il ne s'agit pas d'informations aberrantes ou très spécifiques (Cf. exemple § 6.2.2), il faudra veiller à ce que cela ne corresponde pas à des situations où le mode a été choisi d'après les enquêtes. Le modélisateur pourra tester la pertinence de sa formulation en comparant les valeurs des paramètres estimés avec et sans l'indisponibilité, ainsi que la qualité de reproduction des données.

3. Bases de données

Pour estimer le modèle de choix modal, il est nécessaire de disposer d'une ou de plusieurs bases de données sur le comportement des individus en situation de choix du mode de transport. Le choix de mode dépend à la fois des caractéristiques de l'individu (revenus, motorisation du ménage ou de l'individu lui-même, etc.) et de celles du déplacement (motif, niveau de service des différentes alternatives modales, etc.) : il faut donc utiliser une base de données qui croise les caractéristiques socioéconomiques des individus, les comportements de mobilité, ainsi que les caractéristiques de l'offre de transport (pour chacune des alternatives). La qualité du modèle ainsi estimé dépend fortement de celle des données utilisées.

Par ailleurs, puisqu'il s'agit de déterminer la sensibilité d'un certain nombre de variables explicatives du choix modal, la base de données doit permettre de caractériser les préférences modales des individus dans des situations de choix très variées.

Pour répondre à ce besoin de données, il existe deux types d'enquêtes : les **enquêtes de préférences révélées** (enquêtes RP), qui révèlent les choix de mode effectivement réalisés lors de situations de choix déjà vécues, et les **enquêtes de préférences déclarées** (enquêtes SP), qui permettent de caractériser les choix des individus dans des situations de choix hypothétiques. Habituellement, ce sont plutôt les données de type préférences révélées qui servent à estimer les modèles.

Dans l'hypothèse où le modélisateur ne dispose pas des données nécessaires (ou de qualité suffisante), le lecteur pourra utilement se référer au paragraphe 7.4 dédié à la problématique de transférabilité de modèles de choix de mode.

3.1 Bases de données disponibles

En France, plusieurs bases de données de préférences révélées sont disponibles.

3.1.1 Recensement de la population

Le recensement permet de connaître de façon très précise les caractéristiques socioéconomiques des individus ainsi que leur choix de mode « habituel » pour leurs mouvements pendulaires (domicile-travail et domicile-études). Cependant, il ne s'agit pas d'une base de données sur les déplacements à proprement parler (pas d'information sur la fréquence des déplacements, notion de choix « habituel » trop imprécise, une partie des déplacements non renseignée, etc.) et les caractéristiques de l'offre de transports n'y sont pas décrites, ce qui fait que le recensement est rarement utilisé comme base de données d'estimation en milieu urbain.

3.1.2 Enquêtes « voyageurs »

Ces enquêtes concernent le plus souvent les déplacements en voiture particulière et en transports collectifs, mais elles peuvent parfaitement s'adapter à n'importe quel mode de transport (ex : enquête vélo, enquête aux parcs relais, etc.). Elles nous renseignent sur le choix de mode, les caractéristiques du déplacement (généralement le motif, parfois la fréquence), une partie des caractéristiques de l'offre de transports (l'itinéraire emprunté) du mode effectivement choisi, et, plus rarement, quelques caractéristiques socioéconomiques des individus (informations rapides à recueillir comme l'âge, la motorisation ou la catégorie socioprofessionnelle (CSP)). Même si cette fois on intègre bien l'exhaustivité des différents types de déplacements, là encore les caractéristiques précises de l'offre de transports, en particulier pour les modes non utilisés, n'y sont pas décrites, et cette fois les informations sur les caractéristiques socioéconomiques des individus ne sont que très partielles. Ces bases de données peuvent néanmoins être utiles pour le calibrage des modèles, notamment pour connaître les pratiques relatives à des modes marginaux ou émergents (c'est-à-dire avec de

faibles parts modales) qu'il est difficile d'interroger en nombre suffisant avec des méthodes d'échantillonnage classique.

3.1.3 Enquêtes ménages / déplacements (EMD)

Ces enquêtes consistent à interviewer un échantillon de ménages de l'ordre de 1 à 2 % de la population d'un territoire donné, tirés aléatoirement, mais sous condition de représentativité. La plupart des caractéristiques socioéconomiques des individus du ménage utiles à la modélisation du choix modal (âge, CSP, possession du permis, motorisation du ménage, etc.) sont relevées, et l'ensemble des déplacements effectués la veille de l'enquête sont soigneusement décrits (zone origine-destination, heure de départ, mode utilisé pour chacun des trajets qui composent le déplacement, motif, temps estimé, etc.). Ces enquêtes permettent de disposer d'une solide base de données sur les comportements de choix dans des situations de concurrence modale très variées et pour des typologies d'usagers et de déplacements très diversifiées.

Leur principal défaut est qu'elles ne comportent pas la description complète des déplacements effectués (notamment, les itinéraires empruntés ne sont pas décrits), et surtout qu'elles n'intègrent pas de description précise de l'offre de transport des modes non utilisés. Par ailleurs, dans les villes moyennes ou dans certains secteurs de tirage éloignés du cœur urbain du périmètre de l'enquête, un questionnaire simplifié est utilisé : notamment, celui-ci ne comprend pas la fiche trajet, qui décompose de manière fine la succession des modes de transport utilisés. Néanmoins, l'information sur la combinaison de modes utilisés reste disponible, ainsi la perte d'information n'est pas très pénalisante.

3.2 Avec quelles bases de données estimer le modèle ?

Les EMD sont généralement considérées comme la seule source de données réellement utile pour estimer un modèle de choix modal sur les déplacements du quotidien⁶. Pour pouvoir les utiliser, il faut néanmoins compléter la base de données en y ajoutant la reconstitution de l'offre de transport des différentes alternatives entre lesquelles l'utilisateur a eu à faire un choix, au moment précis de son choix (Cf. § 4). A ce propos, les EMD fixent des limites pratiques quant au potentiel du modèle. Les origines, destinations et arrêts intermédiaires des déplacements sont connus à un niveau zonal, avec un degré de liberté important de la maîtrise d'ouvrage sur la finesse de définition des zones. Ceci a diverses conséquences : biais d'agrégation (Cf. § 7.1) de reconstitution du niveau de service d'une part, impossibilité d'identifier la proximité de la résidence (ou du lieu de travail) d'un arrêt TC d'autre part. Le passage à une identification en longitude/latitude apporterait un gain considérable pour les modèles de choix modal.

Les enquêtes « in situ » peuvent également s'avérer extrêmement utiles pour compléter les données d'EMD là où elles sont insuffisantes (d'un point de vue de la fiabilité statistique), puisqu'elles peuvent permettre de sur-échantillonner certains comportements et/ou certains modes de transport. Dans la mesure où elles n'interrogent qu'un seul mode de déplacement, elles ne peuvent se suffire à elles-mêmes et doivent donc être utilisées en combinaison avec les données d'EMD, ce qui nécessite quelques précautions statistiques lors du calibrage (Cf. § 3.3).

La taille de l'échantillon disponible dans l'enquête de préférences révélées constitue l'un des éléments les plus importants pour définir les grands principes de construction du modèle. La qualité de l'estimation des paramètres est ainsi directement reliée au nombre d'observations par paramètre. Les tests statistiques permettent d'effectuer des vérifications au cas par cas, et de rejeter les paramètres qui ne peuvent être estimés avec une marge d'erreur suffisamment faible. Cependant, il reste souhaitable de connaître *a priori*, avant de lancer les calculs d'estimation du modèle, un ordre de grandeur du nombre de paramètres potentiellement acceptables au vu de la taille de l'échantillon, une remise en cause trop

⁶ L'ENTD 2008 (enquête nationale transports et déplacements) permet également d'apporter des éléments complémentaires d'information.

systematique du modèle, notamment des alternatives retenues, étant à éviter lors de la phase d'estimation. Un seuil de 200 observations par paramètre constitue à ce titre un ordre de grandeur raisonnable.

Au-delà de ces problématiques liées aux échantillons, il est important de préciser que les modèles calibrés sur des enquêtes de préférences révélées ne sont valables que dans la plage de variabilité existante et observée dans l'enquête. Pour étudier les comportements des usagers au-delà de cette plage de variabilité, il est donc nécessaire de faire appel à d'autres types d'enquêtes portant sur les comportements des usagers dans des situations de choix hypothétiques : les enquêtes de préférences déclarées. Bien que relativement peu utilisées en France à l'heure actuelle, ces enquêtes sont pourtant indispensables pour modéliser certains contextes nouveaux et inexpérimentés par les usagers en situation actuelle (ex : introduction d'un nouveau mode de transport, comme le covoiturage dans l'étude [PREDIT, 2010]), ou pour quantifier la sensibilité de certains indicateurs spécifiques difficiles à estimer (ex : valeurs du temps, régularité ou cadencement comme dans la thèse [PONS, 2011]). Là encore, leur combinaison avec les enquêtes de préférences révélées nécessite des précautions statistiques spécifiques (Cf. § 3.3) et une rigueur importante dans la réalisation de ce type d'enquête. En effet, il est primordial que la méthodologie permette de positionner l'individu interrogé en réelle situation de choix sans prise en compte de son opinion (les individus sont souvent d'accord avec un choix hypothétique mais l'écartent lorsqu'il existe réellement).

3.3 Combinaison d'enquêtes

L'échantillon disponible dans l'enquête RP n'est pas toujours suffisant quantitativement ou qualitativement (insuffisance de variabilité dans les valeurs prises par les variables, par ex. un seul prix du billet TC) pour estimer correctement certains paramètres, voire l'intégralité de la fonction d'utilité pour certains modes peu utilisés. Dans ce cas, il est utile de réaliser un recueil complémentaire ou particulier, visant à combler ces lacunes. Il existe principalement trois types de pratiques :

- sur-échantillonner l'enquête RP pour un (ou des) mode(s) particulier(s),
- réaliser une enquête de préférences déclarées,
- utiliser plusieurs recueils RP différents (par ex. lorsqu'un recueil RP local de dimension modeste est réalisé pour transférer dans un contexte local un modèle de référence basé sur un recueil RP conséquent, par ex. l'ENTD (enquête nationale transports et déplacements)).

La combinaison de ces différentes enquêtes est à effectuer selon des procédures particulières, dont les principes sont décrits ci-dessous.

3.3.1 Sur-échantillonnage d'un mode de transport

Certaines alternatives (par ex. le vélo) ont des parts modales constatées très faibles (de l'ordre de quelques pour-cent). Par conséquent, un tirage aléatoire de l'échantillon conduit dans ces conditions à une quantité d'observations insuffisante pour estimer plusieurs paramètres spécifiques à cette alternative. Il faut donc effectuer un tirage fondé sur l'alternative choisie par l'individu : le plan d'échantillonnage est conçu pour obtenir une quantité minimale d'observations pour chaque alternative. Ce résultat peut être obtenu, dans le cas d'une EMD, par la réalisation d'une enquête supplémentaire dite « aux modes », par exemple en positionnant des enquêteurs sur des lieux d'usage quotidien du mode à sur-échantillonner (une piste cyclable, un parking de covoiturage, une gare routière). Il faut veiller à ce que ce complément d'enquête contienne toutes les informations nécessaires en fonction de la spécification du modèle : une planification minutieuse est donc requise. Un échantillon plus important pour certains modes peut aussi être obtenu lors de l'enquête RP en sélectionnant les répondants via une première question filtre sur leur choix de mode de transport.

Pour combiner l'EMD avec la (ou les) enquête(s) complémentaire(s) « aux modes », la procédure à suivre dépend du modèle qui est construit :

- Dans le cas des modèles logit multinomiaux, le modèle de choix modal peut, d'après [BEN-AKIVA et LERMAN, 1985], être estimé directement sur une base de données comportant l'ensemble des observations (c'est-à-dire la combinaison de l'EMD et de l'enquête complémentaire « aux modes »). Les paramètres obtenus seront consistants et seules les valeurs des constantes d'alternatives devront être ajustées pour reproduire les parts modales réelles. Cette technique simple à mettre en œuvre nécessite donc une connaissance préalable des parts modales (généralement, l'EMD possède une précision suffisante pour cela).
- Dans le cas de modèles logit plus complexes, l'estimation des modèles sur une base constituée de l'ensemble des observations selon la procédure « standard » n'est plus possible. Il convient de faire appel à des procédures spécifiques disponibles dans certains logiciels. L'utilisation de ces techniques présente toutefois quelques avantages : ainsi, dans le cas de la méthode décrite dans [BIERLAIRE et al., 2008], la connaissance préalable des parts de marché n'est, selon les auteurs, plus nécessaire.

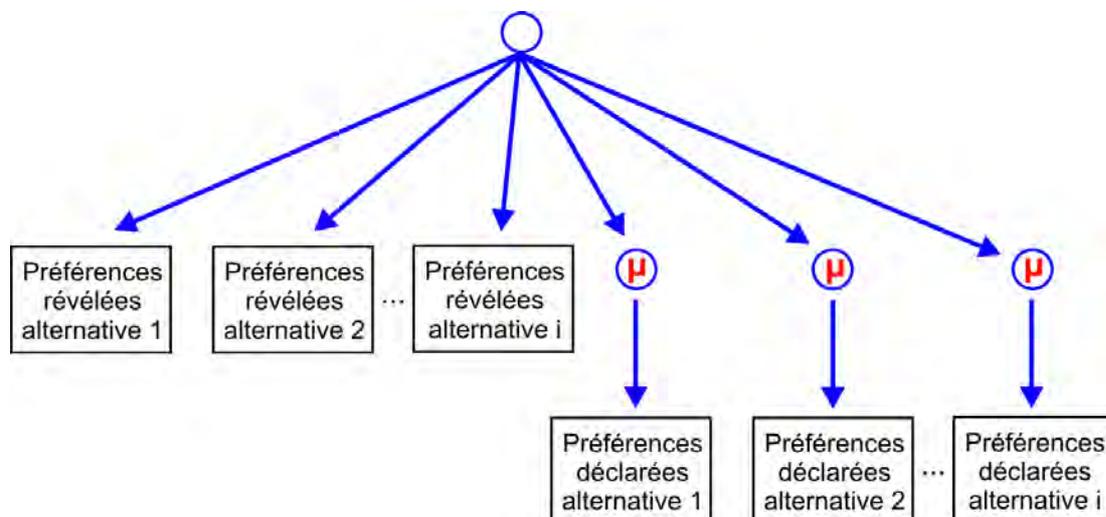
3.3.2 Enquête complémentaire de type préférences déclarées

La réalisation d'une enquête de préférences déclarées (SP) en complément d'une enquête de préférences révélées (RP) a différents usages :

- analyse de l'introduction d'un mode de transport radicalement nouveau,
- détermination de la sensibilité à certaines variables impossible à effectuer avec l'enquête RP de départ.

Pour enrichir les données de l'enquête RP à l'aide des données de l'enquête SP lors de l'estimation du modèle, la méthode suivante peut être utilisée :

1. Deux modèles de choix modal sont postulés (un pour l'enquête RP, un pour l'enquête SP), la contrainte étant que les paramètres des modes ou des variables qui sont communs aux deux enquêtes soient définis comme étant communs à ces deux modèles.
2. De façon arbitraire, on fixe le facteur d'échelle pour l'enquête RP à 1. L'enquête SP aura un facteur d'échelle μ qui sera estimé lors de la procédure de calibrage. Le modèle global RP+SP est alors estimé sur les deux bases de données en même temps, en reproduisant artificiellement la structure hiérarchique suivante (cela revient à multiplier chaque terme de l'utilité des alternatives de l'enquête SP par μ et à estimer ce paramètre μ comme n'importe quel autre paramètre du modèle) :



Le modèle final à retenir pour la prospective est le modèle global RP+SP comportant toutes les variables ainsi calibrées, mais sans tenir compte du facteur d'échelle de la SP.

3.3.3 Combinaison de plusieurs enquêtes de préférences révélées

Ce type de combinaison peut être utilisé par exemple lorsque l'on cherche à transférer un modèle « national », qui aurait été estimé à l'aide d'une combinaison d'EMD sur différentes agglomérations, sur un contexte local où un recueil RP plus modeste qu'une EMD aurait été effectué. La méthode à utiliser est exactement la même que celle utilisée pour combiner des enquêtes SP et des enquêtes RP, le recueil RP local jouant ici le rôle de l'enquête de préférences déclarées dans le schéma fourni au § 3.3.2. Une description plus détaillée (et accompagnée d'un exemple) de cette problématique du transfert de modèle sera réalisée au § 7.4.

3.4 Les bases de données sur l'offre de transports

La codification de l'offre de transports à l'aide d'une base de données géographique est un passage incontournable, indépendamment de la base de données disponibles sur les choix de mode de transport des usagers. En effet, dans les recueils assimilables à des enquêtes RP les plus couramment disponibles (EMD, Cf. § 3.1.3), aucune donnée n'est collectée sur les caractéristiques des alternatives non choisies et les données disponibles sur l'alternative choisie sont souvent de mauvaise qualité. *A contrario*, une enquête SP spécifique contient un recueil des caractéristiques de l'offre des alternatives non choisies. Cependant, pour obtenir des informations de bonne qualité, il convient d'utiliser « in situ » un logiciel de vérification de la pertinence des réponses proposées, ce qui impose la disponibilité d'un outil de représentation de l'offre de transports.

L'offre de transports est donc codifiée à l'aide d'un outil informatique comprenant :

- une représentation graphique géolocalisée des réseaux de transport (SIG, Système d'Information Géographique), incluant une description du tracé des réseaux et un choix du maillage (par ex. suppression des rues en impasse, des chemins ruraux ou forestiers) ;
- une description du niveau de service des tronçons (temps de parcours, péage éventuel...) et des nœuds (type de carrefour, arrêt d'une ligne TC...) de ces réseaux ;
- un ensemble d'informations particulières pour les transports en commun : identification des lignes, des itinéraires de ces lignes (succession de nœuds par lesquels passe la ligne), politique de desserte (nœuds qui correspondent à des arrêts), fréquence ou horaires de passage et type de matériel roulant utilisé pour cette ligne.

De nombreuses bases de données existantes peuvent servir de point de départ à la constitution de cet outil informatique, mais il n'existe pas de référence sur laquelle s'appuyer systématiquement. La liste étant trop importante pour être traitée ici, le lecteur pourra utilement consulter le guide CERTU des pratiques sur la modélisation des déplacements urbains de voyageurs [CERTU, 2003] ou le rapport SETRA sur l'accessibilité des territoires et services [SETRA, 2008].

Si le guide des pratiques sur la modélisation des déplacements urbains de voyageurs [CERTU, 2003] recommandait une certaine prudence vis-à-vis de l'usage des SIG décrivant les réseaux routiers (et par extension ferroviaires ou piétons), ils sont devenus incontournables depuis. Leur mise à jour régulière les rend particulièrement attractifs : l'offre de transports évoluant continuellement, maintenir une base de données à jour constituerait un écueil récurrent pour la pérennité des modèles multimodaux.

Néanmoins, aucun SIG ne contient l'intégralité des informations nécessaires ni n'est utilisable directement : un travail de compilation et de simplification est impératif, de même qu'un travail de vérification/correction « à la main ». Il rend ainsi la mise à jour des réseaux plus délicate : sa mise en œuvre doit principalement pouvoir être effectuée par des algorithmes et être utilisée avec parcimonie pour les cas particuliers. Il en résulte qu'il sera souvent préférable d'utiliser un SIG très détaillé (comme par ex. la BD TOPO® de l'IGN)

qui pourra être simplifié selon des règles plutôt qu'un SIG pour lequel il faudra ajouter quelques tronçons (tel que la BD CARTO® de l'IGN). La définition du niveau de service est exposée en détail dans la partie 4. Elle peut être en bonne partie réalisée à partir des informations qui sont renseignées dans ces SIG : nature de la voirie (réseau national, régional, local), nombre de chaussées, nombre de voies, et dans certains cas vitesse réglementaire ou même vitesse pratiquée permettent de construire une typologie de voirie appropriée. L'information sur le type des nœuds (feu de circulation, giratoire, carrefour à priorité...) est manquante dans la plupart des cas et doit faire l'objet d'un recueil spécifique. Celui-ci peut être largement réalisé à l'aide des images satellitaires (du type www.geoportail.gouv.fr) ou d'images capturées par des caméras embarquées à bord de véhicules (du type <http://sirnet.setra.i2/isimage/>⁷ pour le réseau routier national). Les vérifications et corrections du réseau, pour les informations les moins fiables telles que le nombre de voies arrivant sur des carrefours complexes, pourront également être effectuées par ce biais. La réalisation d'un relevé terrain sera réservée aux sites les plus importants (carrefours très complexes, forte congestion aux heures de pointe). Enfin, la réalisation de relevés de temps de parcours par la méthode du véhicule flottant (principalement en utilisant un GPS relevant la position à intervalle régulier, par ex. toutes les secondes) pourra utilement compléter ces informations.

En ce qui concerne les transports en commun, les informations complémentaires sont généralement issues de l'utilisation d'un plan (SIG ou image Raster) du réseau et des grilles horaires de l'exploitant, supposées construites à partir de données de temps de parcours des véhicules et représentatives du niveau de service réel. Ces éléments permettent de reconstruire les temps de correspondance aux arrêts et les vitesses commerciales par secteur selon les périodes horaires. Ces informations viennent alimenter l'estimation d'un modèle de temps de parcours en fonction de la vitesse de la circulation générale et du nombre d'arrêts. A noter que le développement des SIM (Systèmes d'Information Multimodaux) et le développement de la mise à disposition de l'offre théorique en open data devraient considérablement simplifier cette étape de codification des alternatives TC ainsi que leur maintenance.

Aujourd'hui, la réutilisation d'un outil de codification de l'offre de transport existant, mais ancien et non actualisé, est une pratique courante, afin de limiter le coût de développement des modèles multimodaux. Elle s'intègre dans une stratégie d'usage par « à-coups » du modèle, en fonction des besoins d'études ponctuels. Dans la mesure où la méthodologie de construction de l'offre de transports ne peut plus être reproduite, la construction d'une nouvelle base de données, avec un suivi continu de son actualisation, est donc la plupart du temps préférable. En outre, l'ancienne base de données peut contenir des informations qui ne s'intègrent plus dans le nouveau modèle (par ex. capacités bridées sur les arcs pour « représenter » les limites d'un carrefour alors qu'un modèle d'écoulement aux nœuds est mis en œuvre, utilisation de malus pour caler le modèle d'affectation alors que celui-ci peut faire l'objet d'améliorations méthodologiques rendant cette utilisation obsolète).

Enfin, l'usage massif d'applications de guidage satellitaire (GPS) pouvant être connectées en direct ou en différé à internet (données FCD⁸ - qui sont de plus en plus utilisées de manière opérationnelle pour alimenter les modèles), la possibilité de suivre la trace des téléphones portables (données FMD⁹), ainsi que la commercialisation de ces données par les opérateurs, ouvrent la possibilité d'une amélioration très importante de la modélisation des temps de parcours et notamment des phénomènes de ralentissement. La différenciation des types d'usagers (VL / PL / bus / 2 roues, véhicules en section courante de l'autoroute / véhicules entrant sur l'autoroute) est pour l'instant impossible et ces données sont surtout à utiliser de manière relative.

Les gestionnaires de réseaux de transports en commun collectent également des données de suivi de la flotte (notamment par GPS) qui restent, pour diverses raisons, insuffisamment

7 Outil accessible sur le réseau intranet du MEDDE

8 FCD = *Floating Car Data*

9 FMD = *Floating Mobile Data*

utilisées. Un accès plus courant à ces éléments permettrait une nette amélioration de la description du niveau de service TC : prise en compte de l'irrégularité (ouvrant ainsi la voie à une analyse de politiques d'amélioration de la régularité), du temps perdu aux arrêts (avec la possibilité de modéliser le temps d'arrêt en fonction de différents paramètres tels que le nombre de portes et la quantité de montants/descendants) ou encore du temps perdu dans la circulation générale.

4. Du bon calcul du coût généralisé

Le calcul du coût généralisé est en réalité un résultat du modèle d'affectation. Les différents itinéraires empruntés pour une origine-destination (OD) sont estimés, puis la part de marché de chacun de ces itinéraires est évaluée et enfin le coût généralisé est obtenu par produit de ces parts de marché et du coût généralisé de chaque itinéraire¹⁰. Cette situation a deux conséquences pratiques :

- dans un secteur fortement congestionné, le partage modal est nécessairement issu d'une procédure itérative ;
- les formules de coût généralisé entre les modèles d'affectation, de distribution et de choix modal doivent être les plus cohérentes possibles.

4.1 Périodes de modélisation

La définition des périodes de modélisation de l'offre de transports doit correspondre à deux grands critères :

- critère de **cohérence** : le coût généralisé utilisé pour estimer le modèle de choix modal doit être représentatif du niveau de service ressenti par l'utilisateur dont on cherche à reproduire le choix. Il faut ainsi éviter d'utiliser un niveau de service d'heure de pointe pour déterminer l'utilité d'un usager se déplaçant la nuit. En outre, le calcul doit être cohérent entre la calibration et l'application du modèle.
- critère d'**homogénéité** : le niveau de service est supposé homogène sur l'ensemble de la période modélisée. Cela peut impliquer, par exemple, l'utilisation d'une période de pré-pointe du matin, avec une montée en puissance des TC par rapport au service de nuit, alors que la circulation VP est encore fluide.

Afin de satisfaire à ces critères, l'option la plus souhaitable, mais aussi la plus coûteuse, consiste à décomposer un jour ouvrable en périodes homogènes du point de vue du niveau de service et couvrant les 24 heures (au minimum : une période de nuit, la pointe du matin, une période creuse et la pointe du soir, mais des périodes de pré-pointe du matin et de post-pointe du soir peuvent également être utiles) et à codifier l'offre de transports pour chacune de ces périodes. Le modèle de choix modal peut alors être calibré sur l'ensemble de la base de données de préférences révélées et/ou déclarées (typiquement l'EMD, Cf. § 3) à l'aide de niveaux de service correspondant au mieux à la situation de choix qui se présente aux individus. Cette option est également particulièrement intéressante dans l'optique de mieux reproduire les comportements de choix pour l'ensemble des déplacements d'une même boucle d'activités (Cf. § 6.1). Estimer un modèle d'affectation pour qu'il reproduise correctement les temps de parcours de chaque période implique toutefois l'introduction d'un facteur d'erreur important qu'il convient de ne pas perdre de vue pour identifier la meilleure option. Il sera en outre possible (mais non nécessaire) de tester des paramètres de pondération du temps distincts selon les périodes modélisées (autrement dit, cela revient à tester la construction d'un modèle différent pour chaque période horaire).

Il est également possible de codifier l'offre sur la journée. Ceci est largement envisageable sur des secteurs où la congestion VP est faible et où l'offre TC est relativement homogène selon les périodes (faible fréquence de desserte toute la journée, par ex. dans les villes moyennes), mais aussi lorsque l'utilisation du modèle n'est pas centrée exclusivement sur des questions d'amélioration du niveau de service (par ex. si l'on s'intéresse plutôt à l'évaluation des effets de la modification de la composition de la population).

Dans la deuxième option, si certains phénomènes localisés de congestion VP interviennent, il faudra recourir à des facteurs de concentration, permettant de calculer un

¹⁰ Ce calcul s'appuie sur une formulation déterministe du choix d'itinéraire. Dans le cas de modèles stochastiques, l'utilité de l'ensemble des itinéraires sera calculée différemment, en particulier via le logsum (Cf. 2.2.3) dans le cas de modèles logit.

débit représentatif des conditions moyennes de circulation (Cf. [SETRA, 2009]). Ce recours peut également s'avérer nécessaire dans la première option si les périodes de pointe sont très étalées (afin de bien représenter l'offre TC) et la congestion très variable au sein de cette pointe (par ex. une période de pointe de 6h30 à 9h00 avec des phénomènes de goulot d'étranglement seulement entre 7h30 et 8h30 connaît des temps de parcours moyens bien plus élevés que si les bouchons sont à peu près constants entre 6h30 et 9h00).

Au niveau des pratiques actuelles, on observe parfois que l'offre, notamment TC, n'est codifiée que pour une (ou des) période(s) de pointe, afin de diminuer les coûts de développement du modèle. Cette solution est notamment intéressante si le modèle est utilisé exclusivement à des fins de dimensionnement des infrastructures. La restriction du calibrage du modèle aux seules données de déplacements effectués dans la période en question conduirait à des échantillons trop faibles pour permettre d'estimer des modèles performants. La pratique courante est donc de conserver l'intégralité des observations, ce qui conduit à une perte du pouvoir comportemental du modèle.

Enfin, notons que si le choix modal peut en général être calibré de manière souple, selon les trois options précédentes, il n'en est pas forcément de même au stade de l'application du modèle. Ainsi, il n'est pas possible d'utiliser des niveaux de service distincts selon les périodes horaires dans le modèle de choix modal si la décomposition horaire de la matrice n'a pas été effectuée au préalable (cela soulève notamment la question du traitement du choix de l'heure de départ, au niveau de l'étape de génération, qui ne sera pas traitée ici). Le choix de l'option retenue est donc intrinsèquement dépendant de la spécification générale du modèle multimodal.

4.2 Coût généralisé TC

4.2.1 Traitement des trajets terminaux

Par sa nature même, le réseau TC inclut une très forte hétérogénéité d'accessibilité qu'il convient de prendre en compte dans le modèle de choix modal. La définition d'un zonage fin est certes un premier élément de solution de ce problème, mais cela ne saurait être suffisant :

- d'une part, parce que dans les secteurs plus éloignés du cœur des agglomérations, où l'urbanisation est très diffuse, les zonages sont généralement plus grossiers (du fait de la connaissance statistique sur ces secteurs et des contraintes de temps de calcul) ;
- d'autre part, parce que la finesse du zonage des EMD est très variable et peut même s'avérer insuffisante dans le cœur dense de l'agglomération.

La meilleure pratique est donc de conserver le zonage fin de l'EMD pour l'estimation du modèle. Les centroïdes sont connectés à dire d'expert sur le réseau routier et sur le réseau marche à pied pour le calcul des trajets terminaux. Si le zonage de l'EMD n'est pas suffisamment fin, plutôt que de positionner des connecteurs de centroïde à dire d'expert, censés représenter de manière homogène la distance d'accès aux stations TC, des connecteurs directs des zones vers les stations desservant la zone seront préférables. Ils porteront deux attributs : la médiane et la moyenne des coûts généralisés d'accès de la zone vers les stations. Ces éléments pourront être calculés par tout moyen approprié, notamment après utilisation d'un calcul lissé de densité de bâti sur la zone (Cf. [CERTU, 2005]). Ces coûts généralisés seront déterminés selon les méthodologies propres au mode considéré pour le trajet terminal (modes actifs, VP).

Enfin, la méthode probabiliste d'affectation avec les points d'ancrage, qui a été l'objet de travaux de recherche très récents (Cf. [LEURENT et al., 2011]), permet :

- une prise en compte plus systématique de la répartition spatiale de la demande de déplacements ;
- une moindre nécessité de la recherche de zonages toujours plus fins au détriment des temps de calcul et de la fiabilité statistique, en permettant par exemple une

répartition stochastique de la demande sur les différents arrêts TC desservant une zone ;

- et enfin d'éviter la solution des connecteurs « à dire d'expert ».

4.2.2 Calcul des itinéraires TC

Les services TC étant des constructions complexes, avec différents types de services possibles sur une même ligne (direct, semi-direct, omnibus, embranchements...), différents niveaux d'accessibilité du service selon les périodes horaires, différentes possibilités de combinaison des lignes pour effectuer le trajet complet et différentes options pour les trajets terminaux, déterminer le niveau de service passe nécessairement par un calcul d'itinéraire multi-chemins. Deux principales méthodes se sont imposées dans les pratiques : l'affectation selon les fréquences et l'affectation selon les horaires.

L'affectation selon les fréquences est à privilégier, notamment pour la description des TC urbains. Pour mettre en œuvre cette méthode, une ligne TC est définie par :

- une succession de nœuds définissant le parcours ;
- une liste d'arrêts desservis en montée et en descente¹¹ ;
- des fréquences de desserte selon différentes périodes de la journée / de l'année ;
- des temps de parcours inter-arrêts selon différentes périodes de la journée (basés sur le temps de parcours VP pour les TC intégrés à la circulation générale ou calculés à partir des fiches horaires) ;
- des temps d'arrêt en station¹² ;
- l'irrégularité des arrivées du bus. Elle sera généralement estimée par une variable alpha variant, lorsque la demande arrive à l'arrêt de manière uniforme, de 0,5 (arrivées uniformes des véhicules) à 1 (arrivées très irrégulières des véhicules). Elle est utilisée pour le calcul du temps d'attente initial, calculé comme le produit de alpha par le temps moyen entre les passages à l'arrêt de deux véhicules. Le temps d'attente initial ne peut cependant être réellement calibré que lorsque le modélisateur dispose de données de suivi des véhicules et donc de régularité des lignes. Il peut ensuite être transformé à l'aide d'une fonction, par exemple de type Box-Cox.

Le réseau TC est quant à lui défini par :

- la liste des lignes et leur description ;
- la définition des systèmes tarifaires ;
- l'organisation des correspondances selon les périodes : pour chaque station, des temps de correspondance entre toutes les lignes desservant l'arrêt doivent être définis ; ceux-ci seront calculés à partir d'une analyse de la grille horaire pour les arrêts où des optimisations des correspondances sont définies, voire d'une analyse des suivis de véhicules ; des pénalités spécifiques peuvent être intégrées pour « indiquer » que certaines stations ne sont pas utilisées pour effectuer des correspondances ;
- les interconnexions avec les autres modes (parcs relais notamment).

Ainsi, le coût généralisé du mode TC est composé (avec une pondération permettant la meilleure reproduction des fréquentations constatées) :

- du temps d'accès à l'arrêt,
- du temps d'attente initial,

11 Pour certaines lignes, ou lorsqu'une ligne traverse une zone de grande taille, une simplification de la desserte pourra être effectuée.

12 Si les temps de parcours et les temps d'arrêt sont souvent intégrés dans une variable globale de temps reliée à la vitesse commerciale, pour des commodités de calcul, cela peut engendrer des difficultés lors de la codification de lignes directes et omnibus empruntant le même réseau.

- du temps passé à bord du véhicule,
- du temps d'attente en correspondance,
- des éventuels temps de marche intermédiaires pour effectuer les correspondances,
- du temps d'accès à la destination finale,
- d'une pénalité de correspondance,
- du coût du trajet TC.

L'accès est généralement effectué en marche à pied. La question de la prise en compte de l'intermodalité VP/TC est traitée de manière particulière (cf § 6.3.3).

De très nombreuses méthodes de mise en œuvre de cette affectation selon la fréquence existent, dont l'usage est à examiner au cas par cas. Elles correspondent toutes à un principe déterministe de sélection des lignes attractives, aucun modèle stochastique n'étant réellement disponible.

De manière générale, et en particulier pour les TC urbains, l'affectation « selon les horaires » n'est pas à retenir pour un usage dans un modèle d'agglomération, pour différentes raisons : difficulté de constituer les grilles horaires futures de manière cohérente, choix délicats des algorithmes de sélection des itinéraires et de répartition sur ceux-ci et temps de calcul plus importants.

Néanmoins, pour les lignes de faible fréquence (moins de 3 services par heure), notamment interurbaines, l'affectation selon les fréquences pose un problème pour le calcul du temps d'attente initial (les usagers connaissent les horaires d'arrivée théoriques et se présentent aux arrêts en fonction de ceux-ci ; *a contrario*, ils ne peuvent pas choisir précisément leur horaire de départ : le bon indicateur d'utilité est donc pour eux la fréquence et non le temps d'attente) et du temps de correspondance. L'utilisation des modèles fondés sur les grilles horaires peut être une meilleure solution, en particulier lorsque les correspondances sont optimisées après mise en œuvre d'un cadencement. Une autre possibilité consiste à définir le temps d'attente initial comme une fonction du nombre de services dans la période modélisée, fonction qui pourra être définie lors du calibrage du modèle de choix modal.

Enfin, il faut noter que la prise en compte de la congestion n'est pas entrée dans les pratiques actuelles, pour différentes raisons (notamment du fait qu'il n'y a souvent qu'un itinéraire TC pour une OD donnée). Elle fait l'objet de nombreux développements au niveau de la recherche, du développement des logiciels, voire de la mise en œuvre de méthodes *ad hoc*. Au-delà de la sélection des itinéraires, elle peut s'avérer très utile pour la description du niveau de service réel offert par les TC. Cependant, cela nécessite de disposer de données relatives aux caractéristiques des véhicules et/ou des stations (nombre de places debout, assises, taille des portes, capacité des quais), et cela peut également influencer la codification de l'offre dans le modèle en augmentant le nombre de lignes à codifier (autant de lignes que de véhicules différents), ce qui n'est pas sans conséquences en termes de coûts de développement.

4.3 Coût généralisé VP

De très nombreux algorithmes d'affectation VP sont disponibles sur le marché. Effectuer un choix ici n'est donc ni souhaitable, ni possible. Pour effectuer un choix, trois éléments principaux sont à prendre en considération :

- Il s'agit naturellement d'une affectation à contrainte de capacité permettant d'atteindre un équilibre de Wardrop par une procédure itérative.
- L'intégration des mécanismes bi-critères prix-temps, que ce soit par une affectation multi-classes ou par une affectation tenant compte d'une distribution lognormale de la valeur du temps, est à évaluer au cas par cas (autoroute à péage actuellement ou dans le futur notamment).

- Les modèles sophistiqués de prise en compte des encombrements (conflits aux carrefours, remontées de file...) sont à utiliser avec prudence, leur solidité en pratique restant à démontrer par un retour d'expériences.

La pratique courante repose sur l'utilisation d'un modèle d'affectation statique, tenant compte de la congestion à l'aide de simples courbes débit-vitesse. Cette méthode peut néanmoins être enrichie par différentes analyses particulières.

Sur des secteurs urbanisés, les temps de parcours VP sont principalement déterminés par le mode de gestion des carrefours, selon deux modalités :

- ralentissement lié à un arrêt obligatoire (feu rouge, stop) ou à une cession de priorité, en situation fluide ;
- formation de files d'attente en situation congestionnée.

Le temps perdu lié à un arrêt obligatoire est composé de trois termes : décélération, temps d'arrêt minimal et accélération. Les valeurs de ces trois termes sont inscrites dans les tableaux suivants en fonction du type de carrefour :

Type de carrefour	Ralentissement + accélération	Temps d'arrêt minimal
Stop	2 s (30 km/h) à 7 s (50 km/h)	5 s
Feu	2 s (30 km/h) à 7 s (50 km/h)	$\frac{(1-X)^2 \times cycle}{2}$ où X : % de vert du mouvement considéré cycle : durée du cycle (s)
Cession de priorité	0 s (pas d'arrêt) à 7 s (arrêt et ré-accélération à une vitesse de 50 km/h)	0 s

Source : estimations issues des travaux du groupe de travail

Ces temps sont à intégrer en tant que pénalités de mouvements tournants aux nœuds. Les mouvements spécifiques (TAD tourne-à-droite, TAG tourne-à-gauche) peuvent faire l'objet des pondérations usuelles (par ex. 1,1 pour les TAD et 1,5 pour les TAG). En combinant cette méthodologie avec une codification des vitesses à vide correspondant aux limites de vitesse, des résultats satisfaisants seront obtenus en termes de reproduction des temps de parcours en heures creuses.

Concernant les temps perdus liés à la congestion, une analyse précise des conflits reste à l'heure actuelle un cas particulier. Les valeurs de capacité sont donc largement conventionnelles ; elles doivent par contre impérativement être issues d'une analyse du type d'intersection et des modalités de priorité.

Des analyses ponctuelles sur des carrefours précis connus pour leur situation de saturation seront néanmoins utiles :

- pour les feux, par relevé ou obtention du phasage et détermination des capacités afférentes pour chaque mouvement (ne nécessitant *a priori* que rarement un examen spécifique des TAG non protégés, qui ont soit une capacité de 2 à 3 véhicules/cycle s'ils sont contraints d'utiliser le temps de dégagement, soit une capacité égale à celle du TAD divisée par 1,5 si les flux avec lesquels ils entrent en conflit sont loin de la saturation) ;
- pour les giratoires, par application d'un calcul Girabase© ;
- pour les cessions de priorité ou les stops, par application de la méthode de calcul du HCM¹³.

13 HCM : Highway Capacity Manual

En secteur interurbain ou périurbain, ainsi que sur les axes autoroutiers, la congestion peut également provenir de goulots d'étranglement en section courante, générant la formation de files d'attente se propageant en amont. Le modèle d'affectation statique étant fondé sur la demande, cette répercussion de la file d'attente sur les temps de parcours de véhicules non directement concernés par le goulot d'étranglement est impossible à prendre en compte. Il conviendra donc de retenir des courbes débit/vitesse à deux portions (avant et après saturation). La première portion de la courbe débit/vitesse pourra être calibrée sur des données locales (recueils de données individuelles à partir de SIREDO¹⁴, ou de données FCD...) ou, par défaut, s'appuiera sur les éléments du tableau suivant (correspondant aux recommandations du SETRA) :

Type de route	γ	α
Autoroute à 2x2 voies	0,51	4,0
Autoroute à 2x3 voies	0,51	6,0
VRU à 2x2 voies, routes express, 2x2 voies avec carrefours plans	0,41	4,0
VRU à 2x3 voies	0,41	6,0
Routes avec créneaux de dépassement	0,5	3,0
Urbain : 2 voies normales	5,6	3,4
Urbain : 2 voies larges ou 3 voies normales	4,0	3,5
Urbain : profil en travers plus important	2,7	3,4

Source : extrait de [MEDDE, 2007]

Les paramètres du tableau ci-dessus correspondent à une courbe débit/vitesse de type BPR2, formulée de la manière suivante

$$T_{charge} = T_{vide} \times (1 + \gamma \times sat^\alpha)$$

où

- **T_{charge}** : temps de parcours en charge
- **T_{vide}** : temps de parcours à vide
- **sat** : pourcentage de saturation défini comme le ratio du trafic horaire sur la capacité (tous deux exprimés en UVP)

*Rq : pour les autoroutes, il est conseillé de calculer le paramètre γ en fonction de la vitesse à vide, de telle sorte que la vitesse critique (lorsque **sat** = 100 %) vaille 85 km/h.*

La deuxième portion devra être ajustée à dire d'expert de telle sorte que des mesures de temps de parcours sur l'axe congestionné soient correctement reproduites. Là encore, des données FCD pourront utilement être exploitées à cette fin.

Dans tous les cas, il est indispensable de s'assurer de la cohérence des différentes courbes débit/vitesse utilisées. Par ailleurs, il est nécessaire de rappeler que le principe qui prévaut sur tous les autres est le traitement homogène de l'ensemble du réseau : autrement dit, il est nécessaire d'éviter de modéliser les retards aux carrefours au niveau de certaines zones seulement du modèle, sous peine de reconstituer des temps de parcours avec des erreurs hétérogènes (selon les zones) et donc de fausser les estimations des paramètres associés aux temps de parcours dans le modèle de choix modal.

¹⁴ SIREDO : Système Informatisé de Recueil de Données, pour les données de trafic routier.

Les coûts de déplacement VP sont quant à eux constitués :

- des coûts « d'usage » (entretien et dépréciation évalués respectivement à 8,9 c€₂₀₁₀/km et 1,3 c€₂₀₁₀/km),
- des coûts de péage autoroutier (à reconstituer précisément),
- des coûts de carburant, pouvant être calculés en fonction de la vitesse,
- des coûts kilométriques par typologie de route (malus), reflétant la préférence des usagers pour certains itinéraires. Ces coûts ne sont utilisés que pour l'affectation.

Finalement, le coût généralisé du mode VP est donc composé (avec une pondération permettant la meilleure reproduction des comptages de trafic) :

- du temps d'accès à la voiture,
- du temps de parcours à vide,
- du temps supplémentaire perdu lié à la congestion,
- du temps d'accès à la destination finale,
- du coût du trajet.

Pour être complet, il faut également évoquer la recherche de stationnement, qui implique un temps supplémentaire, qui conditionne le temps d'accès final, et qui génère un coût supplémentaire. Les modèles traitant de cette question étant encore au stade du développement dans les laboratoires de recherche, d'autres approches, basées sur l'estimation du niveau de contrainte de stationnement (permettant ainsi la définition d'une variable discrète), restent aujourd'hui à privilégier (Cf. [CERTU, 2009]).

Ces éléments doivent être calculés pour chaque OD, et correspondent donc, à l'issue de la convergence du modèle d'affectation, à une moyenne des caractéristiques de chaque itinéraire pondérée par le trafic de l'OD supporté par chaque itinéraire.

La modélisation statique des phénomènes d'encombrement ne constitue plus le seul outil disponible pour le modélisateur. Différents outils se développent, permettant en théorie une meilleure reproduction des temps de parcours :

- Le plus prometteur à court terme concerne la modélisation macroscopique détaillée des conflits au droit des intersections.
- La détermination des remontées spatiales et temporelles de file d'attente à l'aide de simples modèles pseudo-dynamiques de stocks (nécessitant que l'utilisateur connaisse la densité maximale) commence également à être régulièrement proposée par les bureaux d'étude.
- L'utilisation de techniques plus élaborées de modélisation dynamique macro- ou microscopique à l'échelle d'une agglomération fera sans doute l'objet de premiers tests dans les années à venir.

Ces différents outils sont à manier avec précaution et dans tous les cas à comparer avec la méthode traditionnelle avant de décider d'y recourir.

4.4 Coût généralisé MAP

Les déplacements piétons ne font généralement pas l'objet de modèles d'affectation cherchant à reproduire les flux. Viser un tel objectif ne serait possible qu'au prix de très importants recueils de données complémentaires à la fois sur les comportements et la quantification des usages et probablement d'une finesse de zonage peu compatible avec les besoins de modélisation. Le surcoût serait incompatible avec l'utilité sociale de l'usage de telles données. Même si les comportements sont complexes et impliquent l'usage d'itinéraires divers, le calcul du plus court chemin sera retenu dans la quasi-intégralité des modèles.

Le calcul du coût généralisé du mode MAP devra intégrer la distance, éventuellement différenciée par type de voie, afin de ne pas émettre d'hypothèse intermédiaire inutile : la sensibilité au temps sera un résultat implicite de la calibration du paramètre lié aux variables de distance par type de voie.

Les itinéraires sont définis par un calcul du plus court chemin en distance, impliquant éventuellement des itérations entre le calibrage du modèle de choix de mode et celui du modèle d'affectation (les pondérations des distances étant idéalement identiques entre les deux modèles).

Des temps de pénalités de mouvements tournants aux nœuds sont à introduire au niveau des carrefours à feu (qui sont théoriquement répertoriés pour le modèle VP). Pour tous les mouvements, la pénalité à prendre en compte vaut :

$$\frac{(1-X)^2 \times cycle}{2}$$

où :

- **X** : temps de vert piéton / durée du cycle
- **cycle** : durée du cycle (s)

De même, les traversées d'axes prioritaires peuvent imposer des rallongements importants de temps de parcours qu'il est nécessaire de prendre en compte.

Ces pénalités de mouvements tournants sont transformées en distance initialement avec une vitesse conventionnelle puis en utilisant les vitesses implicites issues du paramétrage du choix modal (obtenues à partir du ratio entre le paramètre du temps perdu aux intersections et le paramètre de la distance parcourue).

La création d'une typologie de voirie caractérisant un niveau d'attractivité pour les piétons sera un complément utile. Cette typologie reposera sur un ensemble de critères prédéfinis plutôt que sur un dire d'expert, critères tels que la nature de la voie (plateau piétonnier, zone de rencontre...) et le relief.

4.5 Cas particulier des coûts généralisés intrazonaux

La prise en compte des déplacements intrazonaux n'est pas indispensable à la réalisation d'un modèle de choix modal, notamment si l'objectif de modélisation est principalement centré sur des questions de dimensionnement d'infrastructure. Le choix de les intégrer est à définir en fonction du compromis à trouver entre la perte de pertinence du modèle (qui sera inévitable puisque les indicateurs de niveaux de service sont nécessairement mal reproduits) et la nécessité « politique » d'évaluer des objectifs de parts modales (en quantité de déplacements). La fiabilisation des étapes de génération/distribution, que permet l'utilisation de l'intégralité de la demande, est également un facteur à prendre en compte.

Si les déplacements intrazonaux sont conservés, le calcul d'impédance s'appuiera sur la forme et la taille de la surface bâtie de chaque zone. Ainsi, pour l'estimation des impédances VP, piéton et vélo, les éléments suivants, basés sur la définition de trois types de configurations, pourront être utilisés :

- **Type 1** : surface bâtie de forme « ramassée », carrée, rectangulaire ou ronde.
Distance moyenne $\approx 0,5 \times \sqrt{surface}$
- **Type 2** : surface allongée, quasi linéaire, en croix ou en étoile. Distance moyenne $\approx 0,50 D$, où D est la plus grande distance entre les extrémités des branches.
- **Type 3** : surface multipolaire, modulation des deux formules.

Pour les TC, le calcul sera plus complexe, car il dépend des configurations spécifiques de chaque zone. Cependant, le mode TC sera certainement inaccessible pour la plupart des déplacements intrazonaux et ce calcul ne concernera qu'un nombre limité de zones.

4.6 Incidence de la congestion sur le calibrage du modèle

La congestion implique une dépendance entre le calcul de partage modal et l'estimation des temps en charge rendant nécessaire l'emploi d'une procédure itérative. Cette pratique est néanmoins surtout utile dans les agglomérations fortement congestionnées et ayant une forte proportion de population non captive empruntant les modes alternatifs à la voiture (Paris, Strasbourg, Lyon...) ou dans les agglomérations où les politiques publiques vont conduire à une forte augmentation de la congestion.

Ce point reste encore largement méconnu dans les pratiques et outils : il n'existe donc pas d'algorithme s'étant imposé et garantissant convergence et solution unique. Cependant, il apparaît que dans la plupart des cas pour lesquels une méthode itérative de prise en compte de la congestion sur le calcul du choix modal a été utilisée, le modèle a convergé en une dizaine d'itérations en utilisant un bouclage simple :

- A l'étape **n+1**, les coûts généralisés de l'étape précédente **n** sont utilisés comme données d'entrée.
- Les parts modales obtenues par application du modèle de choix modal à l'étape **n+1** sont retenues comme valeur de sortie.

Néanmoins, s'il s'avérait sur un cas particulier fortement congestionné ou nécessitant des résultats très précis sur certaines OD que la convergence n'est pas assurée, il est possible d'utiliser la méthode classique et robuste dite *Method of Successive Average*.

Idéalement, l'estimation du modèle devrait généralement être effectuée en deux temps. Dans un premier temps, les coûts généralisés sont approchés en utilisant une matrice VP « observée¹⁵ », ce qui permet de travailler finement sur la spécification du modèle. Dans un deuxième temps, le modèle est estimé sur les coûts généralisés obtenus après avoir fait tourner l'ensemble de la chaîne de modélisation. Une vérification de la cohérence de cette spécification est finalement effectuée en comparant les coûts généralisés obtenus avec le modèle final et avec le modèle initial. Dans la pratique, différents facteurs contribuent à l'absence de ce recours à une estimation en deux temps. D'une part, les modèles pour lesquels la matrice issue de l'étape de demande est directement affectée sont très rares ; un pivot additif ou multiplicatif est généralement utilisé, ce qui signifie que la matrice « observée » fournit directement les coûts généralisés en situation d'estimation du modèle. D'autre part, il pourrait s'avérer nécessaire d'effectuer également plusieurs itérations de cette méthode « idéale ». Ainsi, le recours à une estimation du modèle à l'aide des coûts généralisés issus de la matrice « observée » sera dans la plupart des cas satisfaisant ; la mise en oeuvre de la méthode « idéale » apportera des informations complémentaires et intéressantes dans certains cas particuliers (chaîne de modélisation sans pivot, matrice « observée » peu fiable, forte congestion).

La matrice « observée » peut être reconstruite selon différentes méthodes, en tenant compte des incertitudes sur les enquêtes utilisées. Différentes techniques spécifiques existent à cette fin (Cf. par ex. [FELDMAN et al., 2012]), dont les deux grandes familles suivantes, qui peuvent être combinées :

- recourir à un modèle de distribution gravitaire pour décomposer une matrice observée à un niveau d'agrégation pour lequel les OD sont fiables, en tenant compte des temps de parcours ;
- construire des matrices de manière automatique de façon à reproduire les comptages. La matrice observée est basée sur la déformation de matrices initiales, qui peuvent être directement les matrices de l'EMD ou de manière préférentielle une matrice obtenue selon une méthode similaire à celle décrite ci-dessus. Les logiciels de modélisation proposent généralement des procédures automatiques pour réaliser ces déformations ; il sera utile de contraindre les valeurs des déformations des OD dans leur intervalle de confiance, par exemple à 90 %.

15 La matrice est dite « observée », bien que les résultats d'enquête ne soient pas fiables aux niveaux de finesse du zonage nécessaires pour reproduire les choix d'itinéraire. Souvent, il s'agira d'une matrice observée sur un zonage plus agrégé, puis décomposée sur un zonage plus fin.

5. Les outils statistiques disponibles pour estimer le modèle de choix modal

5.1 Les indicateurs statistiques de base

Afin de juger de la pertinence de la formulation du modèle, aussi bien du point de vue de sa structure que de celui des variables utilisées et des paramètres qui leur sont associés, il est nécessaire de disposer d'une batterie d'indicateurs efficaces permettant de faire les choix les plus pertinents. Pour chaque formulation du modèle testée, il existe plusieurs types d'indicateurs, permettant de juger de la qualité générale du modèle mais aussi de la qualité spécifique des paramètres associés à chaque variable explicative.

5.1.1 Indicateurs généraux

De façon pratique, ces indicateurs doivent permettre d'une part de comparer deux modèles entre eux afin de retenir la meilleure formulation, et d'autre part de connaître la qualité intrinsèque du modèle afin de pouvoir décider le moment venu de l'achèvement du processus de calage.

Il existe principalement deux indicateurs (le ρ^2 et le ratio de vraisemblance) permettant de juger de la qualité générale du modèle. Tous deux sont déduits de la valeur de la fonction de log-vraisemblance définie dans l'encadré ci-dessous.

Définition de la log-vraisemblance (logV)

L'estimation d'un modèle désagrégé s'effectue par maximisation de la fonction de vraisemblance définie comme suit :

$$L(\beta) = \prod_{i,m} (P_i(m|\beta))^{y_{i,m}}$$

où :

- β est le vecteur de paramètres à estimer
- $P_i(m|\beta)$ = probabilité que l'individu i choisisse le mode m , sachant le vecteur de paramètres β
- $y_{i,m} = 1$ si l'individu i a effectivement choisi le mode m , 0 sinon.

En pratique, c'est le logarithme népérien de cette fonction que l'on cherche à maximiser (il est plus aisé de maximiser une somme qu'un produit), ce qui permet de définir la log-vraisemblance, notée ici **logV** :

$$\log V(\beta) = \sum_{i,m} \ln (P_i(m|\beta))^{y_{i,m}}$$

La log-vraisemblance est, par construction, toujours négative. Quel que soit le logiciel de calibration utilisé, le calibrage s'effectue toujours par maximisation de cet indicateur (soit une diminution de sa valeur absolue).

La valeur de log-vraisemblance obtenue permet de juger de la capacité d'un modèle à reproduire l'échantillon observé, et également de comparer deux modèles entre eux à l'aide de deux indicateurs qui en sont issus : le ρ^2 (ou $\bar{\rho}^2$ ou ρ^2 ajusté qui en sont des variantes) et le ratio de vraisemblance.

5.1.1.1 ρ^2 , $\bar{\rho}^2$, et $\bar{\rho}^2$ ajusté

Le $\bar{\rho}^2$ (rho-carré) est un indicateur qui se veut analogue au R^2 de la régression linéaire. Il est défini de la manière suivante :

$$\bar{\rho}^2 = 1 - \frac{\log V(\hat{\beta})}{\log V(c)}$$

où :

- $\hat{\beta}$: vecteur de paramètres obtenu à l'issue du calibrage
- 0 : vecteur de paramètres nuls (ce qui revient à considérer l'équiprobabilité des alternatives modales)
- c : le vecteur de paramètres constitué à l'aide de simples constantes d'alternatives ou de segments.

Cet indicateur est compris entre 0 et 1. Il permet de juger de l'amélioration apportée par le modèle par rapport à une simple reproduction des parts modales : elle sera d'autant plus importante que la valeur sera proche de 1. Il est difficile en pratique de préconiser une valeur minimale qui assurerait un « bon » calage du modèle, car celle-ci peut dépendre :

- de la part des facteurs non observés (ou non observables) dans les choix constatés des usagers ;
- de la nature du modèle, notamment de son niveau de segmentation et du nombre de modes pris en compte. Le $\bar{\rho}^2$ obtenu sera naturellement d'autant plus élevé que le nombre de segments sera important, et que le nombre de modes pris en compte sera faible ;
- des objectifs du modèle (précision recherchée) : les modèles de type stratégique n'ont pas forcément autant d'exigence de qualité que les modèles de projet pour lesquels un très bon calage est nécessaire, et surtout une bonne reproduction des parts modales sur les OD concernées par le projet.

En tout état de cause, la littérature montre que, quel que soit le modèle construit, une valeur de 0,3 est systématiquement atteinte. Cette valeur peut donc être considérée comme la valeur minimale requise, pour les modèles les plus simples et les moins exigeants.

En pratique, le ρ^2 est également utilisé. Il est calculé de la façon suivante :

$$\rho^2 = 1 - \frac{\log V(\hat{\beta})}{\log V(0)}$$

Toutefois, cet indicateur n'est pas recommandé. Ainsi, dans le cas où les parts modales des alternatives modales observées sont éloignées les unes des autres, la spécification d'un modèle à l'aide de simples constantes d'alternatives permet d'obtenir une très bonne valeur de ρ^2 , alors même que ce modèle n'a aucune valeur prédictive réelle.

Par ailleurs, comme pour le R^2 de la régression linéaire, cet indicateur croît avec le nombre de variables introduites dans la fonction d'utilité. Ainsi, afin de comparer deux modèles entre eux, un troisième indicateur est alors proposé afin de juger de l'opportunité de l'introduction de nouvelles variables dans la fonction d'utilité. L'indicateur ρ^2 ou ρ^2 ajusté est alors défini comme suit :

$$\bar{\rho}^2 \text{ ajusté} = 1 - \frac{\log V(\hat{\beta}) - K}{\log V(c)}$$

où K est le nombre de variables introduites dans la fonction d'utilité.

Si la valeur de ce $\bar{\rho}^2$ ajusté n'augmente pas après l'introduction de nouvelles variables dans la fonction d'utilité, alors ces nouvelles variables peuvent être rejetées. En revanche, si la valeur de $\bar{\rho}^2$ ajusté augmente, il est nécessaire du point de vue de la rigueur statistique de

recourir à un autre indicateur pour vérifier si le gain lié à l'introduction de ces variables est significatif ou non : le **test du ratio de vraisemblance**.

Enfin, précisons que la valeur des indicateurs ρ^2 , $\overline{\rho^2}$, et $\overline{\rho^2}$ ajusté a tendance à diminuer avec le nombre d'alternatives du modèle, et qu'il faudra prendre en compte cette caractéristique lors de la « comparaison » de deux modèles entre eux.

5.1.1.2 Ratio de vraisemblance (LR)

Le ratio de vraisemblance **LR (pour *likelihood ratio* en anglais)**, qui sert à comparer deux modèles entre eux, est un indicateur donné par la formule :

$$LR = -2 \times (\log V(\beta_r) - \log V(\beta_{nr}))$$

où :

- β_r : vecteur de paramètres du modèle restreint
- β_{nr} : vecteur de paramètres du modèle non restreint

Ce paramètre suit asymptotiquement une loi du χ^2 avec r degrés de liberté, où r est le nombre de restrictions du modèle restreint (autrement dit, le nombre de variables en moins dans le modèle restreint par rapport au modèle non restreint). Le ratio de vraisemblance est donc analogue au **F** de Fischer de la régression linéaire : si sa valeur est supérieure à la valeur du χ^2 avec r degrés de liberté au seuil de confiance choisi, alors cela signifie que le modèle non restreint est significativement meilleur que le modèle restreint.

En pratique, s'il doit être effectué rigoureusement, ce test permet surtout d'écarter les modèles les moins pertinents (il suffit que la log-vraisemblance augmente de quelques dizaines de points pour que l'amélioration soit jugée significative du point de vue de ce test statistique). Pour les autres modèles, le bon sens doit jouer un rôle aussi important que ce test. A cet égard, le groupe de travail a relevé des valeurs de log-vraisemblance typiquement de l'ordre de -10 000 sur les modèles pour lesquels il a disposé d'informations.

Par ailleurs, le test du ratio de vraisemblance étant trop peu contraignant, il est plus efficace en pratique de se contenter de regarder l'augmentation du ρ^2 ajusté pour comparer deux modèles entre eux, en choisissant un seuil minimum relativement élevé en dessous duquel l'amélioration peut être jugée comme insuffisante. La pratique montre qu'une augmentation de l'ordre de 0,05 peut être considérée comme significative. En deçà, la décision de l'abandon de telle ou telle formulation doit être prise précautionneusement, notamment à l'aide des indicateurs spécifiques construits au niveau des variables.

Enfin, il sera utile de calculer $\overline{\rho^2}$, et $\overline{\rho^2}$ ajusté par mode de transport, afin de diagnostiquer si un mode particulier présente des difficultés en termes de reproduction des comportements des individus.

Par ailleurs, pour améliorer la valeur du $\overline{\rho^2}$ et donc la qualité « affichée » du modèle, l'utilisation d'un nombre important de constantes de segments ou de zones, appelées *dummies*, s'avère assez efficace et est donc assez tentante, alors qu'elle n'améliore pas nécessairement la qualité prédictive du modèle (elle peut même la dégrader dans certains cas). Ainsi, il convient de ne pas perdre de vue l'objectif du modèle lors de son calibrage, qui est de bien identifier et de bien reproduire les mécanismes microéconomiques à l'œuvre lors du choix de mode de transport, et non de reproduire de manière artificielle la vraisemblance du modèle en utilisant ce type de constantes de façon excessive.

5.1.2 Indicateurs spécifiques au niveau des variables

Les indicateurs présentés ci-dessus sont utiles pour comparer deux modèles entre eux et pour juger de la capacité du modèle à reproduire les comportements observés, mais ils ne permettent pas à eux seuls de juger de la qualité prédictive du modèle. En effet, des investigations spécifiques au niveau de chaque variable doivent également être menées, afin de vérifier que le comportement de chaque variable du modèle est conforme à la logique économique, et que les paramètres qui leur sont associés sont statistiquement significatifs (autrement dit que la sensibilité du modèle par rapport à chaque variable est fiable). Les tests à disposition portent sur le signe des paramètres, sur leur fiabilité statistique (quasi-t tests), et sur leurs corrélations.

5.1.2.1 Signe (et valeur) des paramètres

Il s'agit du premier examen à mener lors du calibrage du modèle : le signe des paramètres estimés doit être cohérent avec la logique économique (exemple : le tarif des TC ne doit pas être positif dans l'utilité TC, car cela signifie que l'augmentation des tarifs TC engendrerait une augmentation de la part modale TC).

Le fait d'obtenir un paramètre du mauvais signe peut être causé par plusieurs sources d'erreurs :

- Variable non significative : si la valeur est du mauvais signe mais proche de 0, il se peut que le paramètre ne soit tout simplement pas significatif (Cf. § 5.1.2.2).
- Corrélations entre paramètres (Cf. § 5.1.2.3) : pour reprendre l'exemple précédent, si les temps et les coûts sont corrélés, il est possible que le coût ressorte avec une valeur positive, l'effet coût pouvant être englobé dans le paramètre de temps (qui dans ce cas sera surestimé en valeur absolue).
- Problème dans la structure du modèle (Cf. § 5.2) : si la structure du modèle est inappropriée, par exemple avec des modes qui sont corrélés entre eux, le signe de certains paramètres peut être également faussé.
- Endogénéité (Cf. § 5.3).

Le cas où le signe n'est pas le bon est le plus simple à détecter. Dans la plupart des cas, les sources d'erreurs mentionnées ci-dessus ne vont influencer que la valeur des paramètres obtenus, à la hausse ou à la baisse, sans aller jusqu'à en changer le signe. Ces erreurs sont alors beaucoup plus difficiles à détecter. Les parties suivantes ont pour objectif de traiter ces différents problèmes afin d'éviter ce type d'erreurs d'estimation.

5.1.2.2 t tests asymptotiques

L'estimateur du maximum de vraisemblance ayant une distribution asymptotiquement normale, il est possible, comme pour la régression linéaire, de faire des tests sur la représentativité de chaque paramètre. Le test en question s'appelle le **t test de Student**. Il permet de tester l'hypothèse H_0 : « le paramètre est nul ». La valeur de test (« t »), fournie par les logiciels de simulation, doit ensuite être comparée à la valeur de la table de Student pour un nombre infini de degrés de liberté (car l'effectif sur lequel le modèle est estimé est grand) au seuil de confiance désiré. Si l'on se place à un seuil de confiance de 95 % (valeur classique utilisée), on obtient 1,96. Autrement dit, le paramètre dont le quasi-t test est inférieur à 1,96 pourra être rejeté, car il n'y a pas de certitude, au seuil de confiance de 95 %, qu'il soit significativement différent de 0 (l'hypothèse H_0 n'est pas rejetée).

En fonction des résultats de ce quasi-t test, il est ensuite nécessaire de prendre une décision quant au choix de conserver ou de rejeter la variable. Celui-ci, en lien avec la nature de la politique à tester, peut être résumé par le tableau suivant :

Signe du coefficient	t-test	Nature de la variable	
		Liée à la politique à tester	Autre
Conforme	significatif	conserver	conserver
	non significatif	problème → respécifier	problème → respécifier ou rejeter
Non conforme	significatif	problème important → respécifier	problème important → respécifier
	non significatif	problème → respécifier	problème → rejeter

Source : tableau issu des travaux du groupe de travail

Le quasi-t test permet également de connaître l'intervalle de confiance du paramètre obtenu. Ce test est toutefois peu utilisé dans la pratique. Pourtant, cette information permet de juger de l'incertitude sur l'estimation du coefficient et donc de la fiabilité du modèle.

A la place de ce t de Student, il arrive également que les logiciels d'estimation fournissent la probabilité critique de vérification de l'hypothèse H0. Cette probabilité est calculée en inversant le raisonnement précédent. Elle doit rester inférieure au seuil désiré (classiquement la valeur de 5 % est retenue), sinon le paramètre peut (ou doit, selon les cas) être rejeté.

5.1.2.3 Corrélations entre paramètres

De la même façon que le quasi-t test permet de tester l'hypothèse H0, il est possible de tester l'hypothèse « les deux paramètres sont égaux », à l'aide d'un autre quasi-t test (que nous noterons t_{correl} dans la suite de ce guide), défini de la manière suivante :

$$t_{\text{correl}} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\text{var}(\hat{\beta}_1 - \hat{\beta}_2)}}$$

où :

- $\hat{\beta}_1$ et $\hat{\beta}_2$: valeurs des deux paramètres estimés
- $\text{var}(\hat{\beta}_1 - \hat{\beta}_2)$: variance¹⁶ de $(\hat{\beta}_1 - \hat{\beta}_2)$

Cet indicateur est généralement calculé par le logiciel d'estimation, pour chaque couple de paramètres. Si la statistique t_{correl} est inférieure à 1,96, alors deux explications sont possibles :

- Soit la valeur des deux paramètres estimés est très proche ($\hat{\beta}_1 = \hat{\beta}_2$). Ce cas n'est pas forcément problématique : si leur covariance (généralement calculée par le logiciel d'estimation également) est positive ou proche de 0, il n'y a pas de raison (statistiquement parlant) pour que les valeurs de ces deux paramètres aient été mal estimées. Il peut donc être envisagé de les regrouper dans le modèle, si ce regroupement a du sens. Par exemple, si le modèle est segmenté par motif de déplacement, regrouper certaines constantes de motif peut s'avérer nécessaire. En revanche, si leur covariance est fortement négative, cela peut s'avérer plus problématique (voir point suivant).

16 Pour rappel : $\text{var}(\hat{\beta}_1 - \hat{\beta}_2) = \text{var}(\hat{\beta}_1) + \text{var}(\hat{\beta}_2) - 2 \times \text{cov}(\hat{\beta}_1, \hat{\beta}_2)$

- Soit la variance de $(\hat{\beta}_1 - \hat{\beta}_2)$ est importante, ce qui signifie que β_1 et β_2 sont corrélés et diffèrent de leur moyenne dans le sens opposé (autrement dit leur covariance est fortement négative). Ce cas est plus problématique puisque cela engendre une erreur dans l'estimation de ces deux paramètres, la sensibilité propre de ces deux paramètres ne pouvant être estimée correctement (les variations de l'un sont en quelque sorte compensées par les variations de l'autre). Dans ce cas, il est nécessaire, soit d'abandonner l'une des deux variables, soit de rechercher les motifs de cette corrélation et de modifier ces deux variables en conséquence de façon à supprimer ce problème. **Les cas où la covariance de deux paramètres est négative sont donc à surveiller de près.**

Remarque: le cas où la covariance de $(\hat{\beta}_1 - \hat{\beta}_2)$ est fortement positive (ce qui signifie que β_1 et β_2 sont corrélés et diffèrent de leur moyenne dans le même sens) ne pose pas de problème dans l'estimation des deux paramètres, car il n'y a pas le même phénomène de compensation des erreurs. Cependant, il est probable que dans ce cas le test du ratio de vraisemblance ou du ρ^2 ajusté (entre d'une part, le modèle intégrant seulement l'un des deux coefficients, et d'autre part, le modèle intégrant simultanément les deux coefficients) conduise à rejeter l'un ou l'autre de ces deux paramètres.

Il est souvent constaté dans la pratique que ces tests statistiques sont trop peu utilisés, hormis peut-être le test portant sur le signe du coefficient. L'utilisation du quasi-t test est pourtant indispensable afin de s'assurer de la fiabilité statistique de chaque coefficient et donc de la pertinence du modèle qui en découle.

De même, les corrélations entre paramètres doivent être étudiées scrupuleusement car elles peuvent occasionner des erreurs d'estimation tout à fait préjudiciables à la qualité du modèle. En pratique, il ne s'agit pas de chercher à supprimer toute corrélation entre paramètres, car cela est presque impossible à réaliser, surtout lorsque les échantillons sont faibles : il s'agit plutôt de savoir maîtriser ces erreurs potentielles afin d'être en mesure d'identifier les vrais dysfonctionnements qui sont de nature à fausser l'estimation des paramètres importants du modèle (autrement dit, ceux auxquels il est le plus sensible). Des corrélations entre constantes pourront par exemple être tolérées, alors qu'inversement des corrélations entre paramètres de temps et de coûts seront à éviter à tout prix car elles pourront avoir des conséquences importantes sur les résultats du modèle en prospective.

5.1.3 Récapitulatif

Le tableau ci-après permet de synthétiser l'ensemble des tests disponibles :

	Valeur minimale suggérée	Principe de l'indicateur	Objectif
$\bar{\rho}^2$ et $\bar{\rho}^2$ ajusté	0,3 mais dépend de la nature et des objectifs du modèle	Comparer le modèle avec le modèle constitué de « constantes modales »	Juger de la qualité du modèle
LR	χ^2 avec r degrés de liberté	Comparer le modèle avec un modèle restreint (comportant r variables en moins)	Décider de rejeter ou non des variables supplémentaires
$\bar{\rho}^2$ et $\bar{\rho}^2$ ajusté par mode	0,3 mais dépend de la nature et des objectifs du modèle	Comparer le modèle avec le modèle « constantes modales »	Juger de la qualité de reproduction par mode du modèle
Signe des paramètres	Cohérent avec la logique économique	Valider le comportement du modèle	Décider de rejeter ou non des variables
Quasi-t tests (sur chaque paramètre)	1,96	Vérifier que chaque paramètre est significativement non nul	Décider de rejeter ou non des variables, et/ou de respécifier le modèle
Corrélations entre paramètres (t_{correl})	1,96 ou $<1,96$ mais covariance >0	Tester la corrélation entre paramètres	Décider de rejeter ou de regrouper certaines variables

5.2 La propriété IIA

La propriété IIA (indépendance des alternatives non utilisées) est fondamentale pour comprendre le fonctionnement du modèle logit. Quelle que soit sa qualité statistique en termes de reproduction de l'échantillon, un modèle logit ne pourra être validé si sa structure n'est pas vérifiée en testant la validité de la propriété IIA. Ce test est très important car une structure de modèle inadaptée engendre un mauvais comportement du modèle en prévision.

Il existe principalement deux méthodes pour tester cette hypothèse :

- le test d'Hausman-McFadden : il s'agit d'estimer le modèle sur l'ensemble des alternatives, puis sur un nombre restreint d'alternatives, et de vérifier la stabilité des paramètres (sinon la propriété n'est pas respectée). En pratique, il n'y a jamais complète stabilité des paramètres, donc l'idée est simplement de vérifier que les signes et les ordres de grandeur des paramètres sont respectés, ce qui comporte une part de subjectivité, c'est pourquoi le second test ci-après peut lui être préféré ;
- le *McFadden's Omitted Variables Test* : il s'agit de vérifier que le ratio des probabilités de choix entre deux alternatives ne dépend pas des attributs d'une troisième alternative (autrement dit que les attributs de cette troisième alternative n'entrent pas significativement dans l'utilité d'une des deux premières). Concrètement, il faut procéder pas à pas de la manière suivante :
 - estimation du modèle complet (noté modèle 1) ;
 - création, pour les 2 modes soupçonnés d'être corrélés (notés 1 et 2), de variables auxiliaires (ou variables croisées) qui dépendent des utilités de ces deux modes ;
 - estimation d'un second modèle (noté modèle 2) en ayant ajouté ces variables auxiliaires aux utilités des modes 1 et 2 ;

- réalisation d'un test de ratio de vraisemblance : si le modèle 2 améliore significativement le modèle 1, et que les paramètres relatifs à ces variables auxiliaires sont significatifs, alors l'hypothèse IIA peut être rejetée.

Les logiciels de calibration des modèles de choix discrets permettent souvent de réaliser automatiquement ces tests. Si ce n'est pas le cas, le second test pourra être mis en œuvre, en se référant à la littérature pour la création de ces variables auxiliaires.

Si ce test conduit au rejet de l'hypothèse IIA, alors la structure du modèle logit doit être abandonnée au profit d'une structure plus complexe du type nested ou cross-nested. Il est également important de préciser que le choix des nids à constituer dépend du résultat de ce test IIA : les modes à inclure dans le même nid sont tout simplement les modes qui ressortent comme étant corrélés dans le test IIA, ce qui souligne encore plus l'importance de ce test.

La vérification de l'hypothèse IIA doit théoriquement être réalisée **à chaque nouvelle formulation du modèle logit**. Cependant, en pratique, le test étant relativement lourd, il est conseillé de l'effectuer seulement lorsque l'on teste une nouvelle formulation très différente de la précédente (nouvelle segmentation, introduction de transformations non linéaires, etc.), et bien évidemment, avant de valider le modèle qui sera retenu à l'issue de la procédure de calibrage. L'expérience montre qu'une segmentation importante du modèle tend à retenir des structures de modèles assez simples (logit multinomial), et qu'inversement les transformations non linéaires des variables clés (ex. : transformation Box-Cox sur le temps) peuvent entraîner le rejet de l'hypothèse IIA, ce qui oblige alors à aller vers une structure de modèle plus complexe de type nested logit.

5.3 L'endogénéité

L'endogénéité apparaît lorsque, dans la fonction d'utilité, le terme d'erreur ε est corrélé avec une ou plusieurs variables \mathbf{X} . Les problèmes d'endogénéité engendrent une sur- ou sous-évaluation des paramètres β et donc une mauvaise reproduction de l'élasticité des parts modales à ces variables \mathbf{X} .

Dans les modèles de choix discrets de la famille logit, trois sources d'endogénéité existent :

- L'erreur d'estimation : les variables \mathbf{X} estimées avec une erreur importante peuvent générer un problème d'endogénéité. Par exemple, l'utilisation des temps de parcours déclarés dans l'EMD est potentiellement endogène.
- L'auto-sélection : ce problème intervient lorsque le choix effectué par l'individu détermine la valeur d'une des variables. Par exemple, les usagers qui ont une préférence pour les TC ($\varepsilon > 0$) sont susceptibles d'habiter près des stations et donc d'avoir un temps de parcours faible.
- L'omission de variables : l'omission d'une variable lors de la spécification du modèle est également source d'endogénéité, car le terme d'erreur inclut alors l'effet de cette variable sur le choix de l'individu. Par exemple, la non prise en compte de la fréquence TC peut conduire à obtenir un paramètre de temps de parcours VP positif (si le temps de parcours TC est inférieur au temps de parcours VP).

Des tests existent pour évaluer l'endogénéité (Cf. par ex. [GUEVARA et BEN-AKIVA, 2006]), mais pas de méthode standard similaire au test IIA. Une première solution peut également consister à effectuer un examen critique du modèle au regard des trois critères ci-dessus.

Une autre solution consiste, lorsque le problème d'endogénéité concerne un segment de population, à utiliser des constantes d'alternatives. Ainsi, l'endogénéité d'auto-sélection pour les usagers habitant près des stations TC peut être traitée en introduisant une constante d'alternative pour des segments de population séparés selon ce critère.

De même il conviendra d'effectuer des calibrations du modèle avec et sans les variables suspectées de générer de l'endogénéité lorsqu'elles sont omises. Si les valeurs des autres

paramètres ne sont pas stables, alors cela signifie qu'il s'agit effectivement de variables à ne pas omettre car leur omission génère de l'endogénéité.

Concernant le biais d'auto-sélection, des méthodes correctives sont disponibles qui font l'objet de travaux spécifiques (Cf. [YÁÑEZ et al., 2009]). Elles consistent à introduire dans le modèle des variables (appelées variables latentes) représentatives des attitudes et perceptions propres aux individus. L'introduction de ces variables latentes permet alors de ré-estimer les paramètres du modèle de façon plus fiable, c'est-à-dire sans biais d'endogénéité. L'emploi de cette méthode est coûteux et cantonné à des modèles à portée de recherche à l'heure actuelle, et donc à réserver à des modèles complexes.

5.4 Le biais d'agrégation

La théorie des choix discrets s'appuie sur la connaissance des valeurs individuelles de chaque variable : le modèle est estimé sur des données désagrégées. Ce choix s'explique par l'apparition d'un biais lors de l'estimation sur données agrégées : le biais d'agrégation. En effet, l'erreur d'agrégation, commise en donnant à la valeur d'une variable la moyenne de cette variable pour un groupe de population, entre dans le terme d'erreur de la fonction d'utilité. Il en résulte, ainsi qu'exposé au § 5.3, un biais dans l'estimation des paramètres. Ainsi, [ORTUZAR et WILLUMSEN, 2011] conseille d'éviter au maximum l'emploi de variables agrégées dans les fonctions d'utilité.

Toutefois, dans les pratiques de modélisation du choix modal en France, cette méthode d'estimation sur données agrégées est la seule disponible pour estimer les paramètres des variables d'offre de transports. Ainsi qu'exposé au § 3., les EMD (ou leur équivalent sur d'autres territoires que les grandes agglomérations de province), ne fournissent que peu d'informations individuelles sur les niveaux de service et elles ne concernent que l'alternative choisie. Qui plus est, l'erreur commise sur les niveaux de service des variables d'offre de transport de l'alternative choisie est suspectée d'être importante, ce qui introduit également un biais d'endogénéité ! Ainsi, les niveaux de service sont en fait reconstitués par un modèle d'affectation agrégé. L'erreur pourra toutefois être limitée en définissant des segments de population pour lesquels la variance de la composante systématique de l'utilité est faible.

5.5 La segmentation

Les modèles logit, nested logit ou cross nested logit supposent que les paramètres relatifs à chacune des variables présentes dans la fonction d'utilité sont identiques entre tous les individus (il ne doit donc pas y avoir de variation de perception). Si cette hypothèse n'est pas vérifiée, alors le comportement du modèle ne sera pas conforme à la réalité. Pour résoudre ce problème, la solution proposée dans la littérature est de segmenter le modèle en groupes de comportement (c'est-à-dire de perception) homogène de manière à ce que cette hypothèse soit cette fois vérifiée au sein de chaque segment.

Pour tester la validité de cette hypothèse, il existe un test statistique appelé test de la variation de perception (*test of taste variation*).

Pour ce faire, il faut avoir un *a priori* sur les caractéristiques socioéconomiques qui pourraient influencer cette variation de la perception : on segmente alors l'échantillon selon ces caractéristiques (ex. : ménages motorisés et ménages non motorisés), et on estime séparément un modèle pour chaque segment. Deux tests statistiques successifs doivent alors être menés :

- test du ratio de vraisemblance (Cf. § 5.1.1) ; l'indicateur est ici donné par la formule suivante :

$$LR = -2 \times (\log V(\beta) - \sum_s (\log V(\beta_s)))$$

où

β : vecteur de paramètres obtenu pour le modèle non segmenté

s : ensemble des segments formés

β_s : vecteur de paramètres obtenu pour le modèle spécifique au segment s

Cet indicateur doit être comparé avec le résultat de la table du χ^2 à k degrés de liberté, où k est le nombre de nouvelles variables créées par la segmentation :

$$k = (s - 1) \times \dim(\beta)$$

Si le test est négatif ($LR < \chi^2(k)$) alors cela signifie que le modèle segmenté n'est pas significativement différent du modèle non segmenté. Il n'y a donc pas lieu de segmenter le modèle, et ainsi de remettre en cause l'hypothèse d'homogénéité de la perception au sein de la classification socioéconomique testée ici. Si en revanche le test est positif, alors il est nécessaire d'en comprendre les raisons en comparant les valeurs des coefficients obtenus pour chaque segment. Pour ce faire :

- un second test permet d'évaluer l'hypothèse d'égalité, pour chaque couple de segments (ici notés 1 et 2), des paramètres associés à une même variable (ici notée p). Il s'agit du quasi-t test (que nous noterons t_{seg} dans la suite de ce guide) défini par la formule suivante :

$$t_{seg} = \frac{\hat{\beta}_1^p - \hat{\beta}_2^p}{\sqrt{\text{var}(\hat{\beta}_1^p) + \text{var}(\hat{\beta}_2^p)}}$$

où :

$\hat{\beta}_1^p$: paramètre associé à la variable p du segment 1

$\hat{\beta}_2^p$: paramètre associé à la variable p du segment 2

Cet indicateur est à comparer avec la table de Student avec un nombre infini de degrés de liberté, au seuil de confiance choisi. Ainsi, si la statistique t_{seg} dépasse en valeur absolue le seuil de 1,96 (seuil de confiance 95 %), alors les paramètres peuvent être considérés comme statistiquement différents : il y a donc variation de la perception de cette variable p au sein de ces segments, ce qui implique que le modèle nécessite d'être segmenté pour cette variable. Dans le cas contraire, les paramètres ne peuvent pas être considérés comme statistiquement différents, donc il est préférable de ne pas créer de segments pour cette variable.

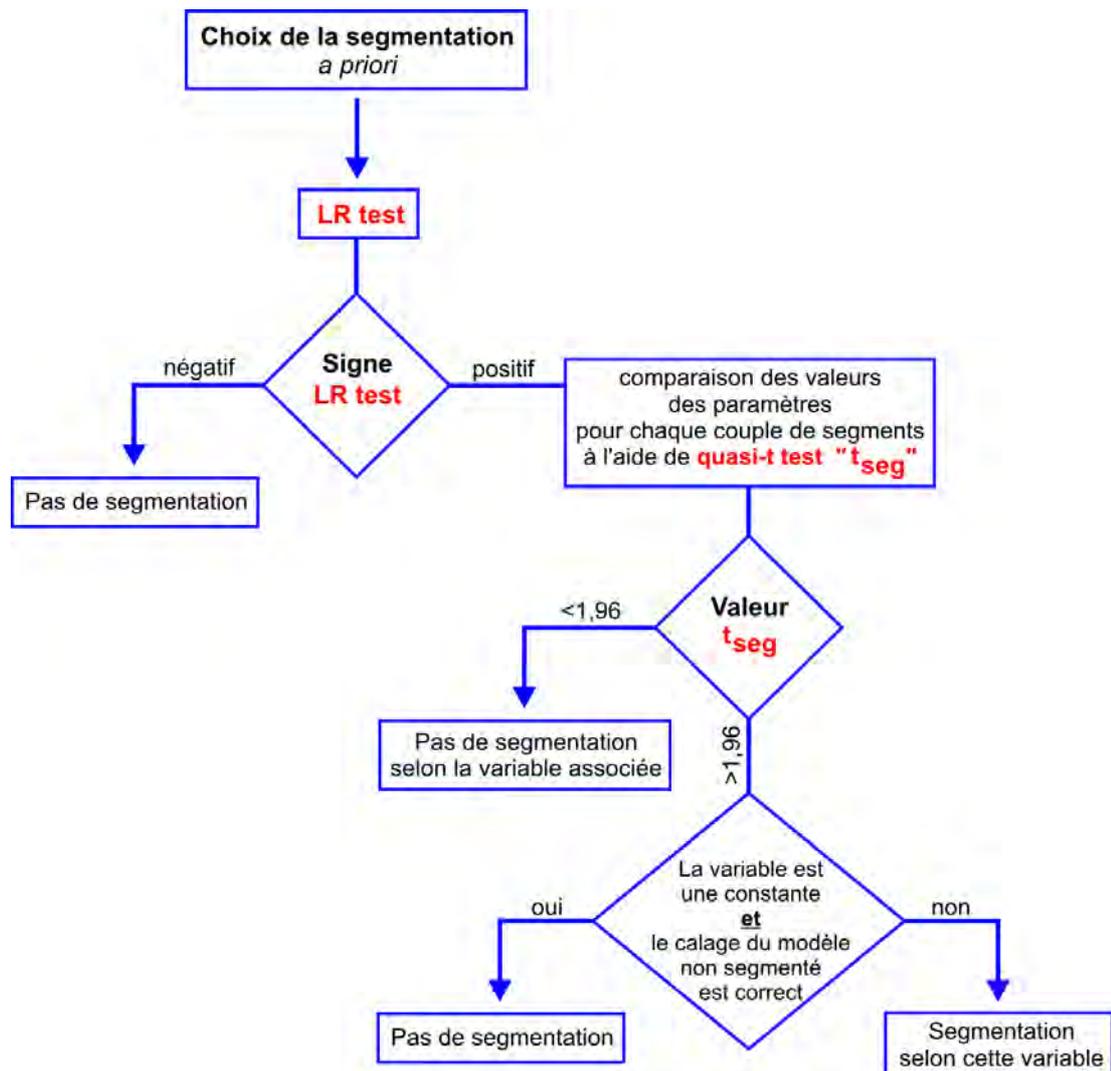
Cette méthode permet donc d'aboutir à un modèle segmenté selon certaines variables et non segmenté pour d'autres. Il est généralement peu probable d'aboutir à un modèle non segmenté, de même qu'il est peu probable d'aboutir à un modèle complètement segmenté.

Si la variable qui ressort comme étant à segmenter à l'issue des tests statistiques n'est pas une constante d'alternative (ex. : le temps), il est nécessaire de segmenter pour éviter les erreurs de comportement du modèle (sinon la sensibilité au temps en sera effectivement faussée).

A contrario, si la variable qui ressort comme étant à segmenter à l'issue des tests statistiques est une constante d'alternative (ex. : constante liée à la catégorie socioprofessionnelle), le choix de segmenter dépend également du contexte de modélisation (objectifs, organisation globale de la chaîne de modélisation). En effet, le fait de segmenter un modèle oblige, pour pouvoir l'appliquer, à reconstituer la demande de transports avec le même niveau de segmentation (concrètement, une matrice OD « tous modes » pour chaque segment) ce qui peut s'avérer assez lourd en pratique selon le type de segmentation obtenu. Cela oblige également à prévoir l'évolution de ces segments de demande en prospective (ce qui nécessite des hypothèses supplémentaires, nécessairement subjectives), ou à construire

des modèles de prévision spécifiques à ces segments de demande (ex. : modèles de motorisation) car ceux-ci peuvent être influencés par les politiques de transports à tester. Une segmentation obtenue à l'issue des tests statistiques peut donc être abandonnée, si elle engendre plus d'imprécision sur la construction de la matrice de déplacements tous modes confondus (en situation actuelle comme en situation de prévision). En effet, puisqu'il ne s'agit que d'une constante d'alternative, la sensibilité du modèle vis-à-vis de la politique à tester ne sera pas modifiée, donc il ne s'agit que d'un raffinement permettant d'améliorer le calage.

En conclusion, le processus à mettre en œuvre, pour chaque variable socioéconomique soupçonnée d'influencer les comportements de choix de mode, est le suivant :



6. Processus de construction d'un modèle

6.1 Problème préliminaire : le modèle dans son univers d'utilisation

Les outils statistiques développés dans la partie 5 délimitent assez précisément les structures possibles du modèle, même si certaines nuances ont été définies. En particulier, la réalisation de segmentations du modèle est reliée à la question de l'existence de différences de perception d'une variable au sein de la population. L'expérience (en accord avec la théorie) montre clairement que cette logique va notamment conduire à segmenter le modèle selon le niveau de disponibilité d'un véhicule particulier. Cependant, les sensibilités du modèle aux niveaux de service vont en être fortement affectées et, en l'absence d'un modèle d'équipement automobile des ménages, les capacités en prospective seront réduites. De plus, comme ce dernier doit intégrer, pour être opérationnel, les niveaux de service des différents modes de transport pour l'ensemble des déplacements effectués par un individu, cette problématique de la segmentation se retrouve directement reliée au choix d'un principe de modélisation par les schémas d'activité (Cf. encadré ci après).

Dès que le modèle de choix modal entre dans une dimension prospective, les options disponibles pour définir sa structure sont limitées par la combinaison des données d'entrée des étapes précédentes et des hypothèses d'évolution des compositions de population des différentes zones. Tant que la segmentation n'est pas reliée à une stratégie d'utilisation des modes de transport (c'est-à-dire qu'elle ne concerne ni la motorisation du ménage ni la possession d'un abonnement TC, ni les habitudes...), il reste possible d'émettre des hypothèses, dont la plus courante sera de supposer le maintien de la composition de la population. Dans le cas contraire, il convient de bien identifier dans quelle mesure la segmentation risque d'influencer la sensibilité du modèle au niveau de service : si celle-ci diminue trop fortement, le recours à un modèle non segmenté pourra alors être envisagé. Pour différentes raisons (hors contraintes liées aux vérifications statistiques), certains segments de population pourront ainsi être abandonnés.

Il existe des pratiques consistant à estimer tout d'abord un modèle partiellement segmenté en ne retenant que les constantes d'alternatives distinctes, puis à estimer à nouveau les constantes d'alternatives du modèle en ne conservant que les segments de population utilisables, tout en maintenant les valeurs des autres paramètres. Cette pratique n'est pas recommandée : elle n'est utile que si les valeurs des paramètres évoluent fortement par rapport au modèle restreint (segmenté uniquement selon les segments de population utilisables). Or cette modification est le signe d'un problème important de spécification du modèle. Ainsi, il semble largement préférable de garantir une cohérence entre l'estimation et l'utilisation du modèle, et donc de retenir l'une des deux options suivantes :

- estimation et utilisation du modèle restreint,
- estimation et utilisation du modèle segmenté, en émettant des hypothèses sur la composition future de la population ou en utilisant un modèle de prévision de cette composition (ex : modèle de motorisation, d'abonnement, etc.)

Prise en compte des boucles de déplacements

Dans un modèle à 4 étapes, les choix de déplacements des individus se font successivement de manière indépendante. *A contrario*, dans une modélisation par boucle, et pour notre contexte de modélisation du choix modal sur une agglomération, les différents choix effectués par les individus tiennent compte de l'ensemble des déplacements effectués du départ du domicile au retour vers le domicile.

D'un point de vue théorique, cette méthode doit mieux capter l'influence de l'amélioration d'un mode de transport, puisqu'elle tient notamment compte de la cohérence des modes de transport successivement choisis (dans le cas général, il n'est pas possible de rentrer en voiture si l'aller s'est fait en transports en commun), ou du caractère déterminant qu'une éventuelle indisponibilité du mode représente pour l'ensemble des choix faits pour la boucle. De telles interactions ne peuvent pas être modélisées selon l'approche par déplacements.

D'après le guide américain NCHRP 406 [NCHRP, 2010], la prise en compte des boucles de déplacements tire son plein potentiel dans le cadre de la modélisation selon les schémas d'activité. D'autres avantages de cette révolution méthodologique apparaissent alors : la capacité pour le modélisateur d'évaluer plus directement différentes politiques que dans le cadre d'un modèle à 4 étapes, qui reste très centré sur les problématiques de dimensionnement. Cette capacité s'appuie principalement sur la technique de génération de populations synthétiques, ce qui permet d'obtenir des résultats similaires dans leur forme à ceux d'une enquête. Pour ces raisons, l'étude des besoins des différentes agences métropolitaines américaines montre un fort intérêt pour la modélisation basée sur les activités, plus d'une dizaine étant impliquées dans l'utilisation ou le développement de tels modèles. Il apparaît cependant également que ces modèles s'appuient sur des programmations *ad hoc* et ne peuvent simplement utiliser des logiciels existants. Le succès d'une telle démarche est donc lié à une stratégie à long terme et à un besoin impératif de dépasser les limitations du modèle à 4 étapes.

Il n'en reste pas moins que la modélisation par boucles peut également être intégrée au sein d'un modèle « plus classique ». Les étapes amont (génération, distribution...) devront également être formulées selon une approche par boucles. Une simplification des comportements sera nécessaire, car les échantillons ne seront pas suffisants pour traiter l'ensemble des alternatives possibles (par ex. pour un modèle traitant jusqu'à 4 chaînes dans les boucles et avec 3 modes disponibles, il y a jusqu'à 81 combinaisons de modes théoriquement possibles).

Comment cette évolution possible des pratiques modifiera-t-elle les recommandations de ce guide ?

Le problème du choix de mode au sein d'un modèle basé sur les activités est le suivant : trouver pour la boucle de déplacements y de l'individu i la succession de modes $M = \{m_1, m_2, \dots, m_n\}$ utilisés pour se rendre successivement aux centroïdes $C = \{c_1, c_2, \dots, c_n\}$. L'utilité de la chaîne modale M peut s'écrire :

$$U(M) = \sum_k U_{m_k}(c_k, c_{k+1})$$

Les données disponibles sont évidemment insuffisantes pour couvrir l'exhaustivité des situations (par ex. les boucles de plus de 4 déplacements auront un échantillon très faible dans une EMD). Cette formulation générale implique donc un travail de simplification des comportements : nombre de déplacements au sein de la boucle, combinaison de modes disponibles... Elle est par exemple compatible avec une pratique courante consistant à simplifier les boucles en allers/retours, avec le même mode pour l'aller et le retour. Surtout, elle permet d'intégrer les deux mécanismes centraux du choix modal dans le cadre d'une modélisation des boucles de déplacements :

- l'utilité de la boucle dépend des utilités de chacun des éléments de la boucle,
- certains modes peuvent être non interchangeables : il suffit de déclarer certaines successions de modes de transport comme non disponibles (par ex. : pour une boucle en 3 chaînes, la succession VP conducteur - piéton - VP conducteur est non disponible, tandis qu'est tout à fait possible pour une boucle en 4 chaînes la succession VP conducteur - piéton - TC - VP conducteur).

En particulier, cette formulation permet d'éviter que la probabilité d'une chaîne modale soit indépendante de l'utilité d'un des maillons de la chaîne (par ex. : le dernier déplacement).

Cette formulation est tout à fait cohérente avec le cadre conceptuel de la modélisation du choix modal à 4 étapes, à la fois pour les dimensions de calcul de l'utilité et de calibration du choix modal. Seuls les calculs d'application seront nettement plus compliqués et gourmands en temps. Il est donc possible de développer une approche de modélisation par boucle sur la base des conseils pratiques de ce guide.

Ceci implique néanmoins que les autres éléments de choix (décision de faire la boucle, choix de la chaîne de destination, choix des horaires pour chaque élément de la chaîne de destination) puissent être disponibles en tant que données d'entrée du modèle de choix modal. A ce titre, il faudra observer avec prudence les pratiques de modélisation conjointe choix modal / choix de destination, qui soulèvent des défis complexes de calibration du modèle et de respect de l'hypothèse IIA, et qui sortent complètement du cadre de ce guide.

6.2 Proposition d'un processus pratique de construction du modèle

6.2.1 Principes de construction du modèle

Un modèle est par nécessité une simplification de la réalité. Il n'y a donc pas de modèle qui permette d'évaluer exactement l'ensemble des projets ou politiques souhaités. Il s'agit donc de construire un modèle qui reproduise suffisamment bien les données observées tout en étant opérationnel. Tout l'art de la modélisation consiste donc à trouver le bon niveau de compromis entre la complexité du modèle et son utilisation. Ainsi, le rasoir d'Ockham sera un critère utile au modélisateur : « **Les hypothèses suffisantes les plus simples sont les plus vraisemblables¹⁷.** »

La présentation des paramètres et indicateurs statistiques d'un unique modèle après estimation dans une boîte noire n'apporte rien à la compréhension des limites du modèle, ce qui est pourtant l'objet de cette étape de construction. Avec de telles pratiques, ce n'est qu'au moment de l'utilisation du modèle que ces limites apparaîtraient, conduisant le maître d'ouvrage à une forte défiance vis-à-vis de l'outil voire à son abandon pur et simple.

Pour éviter cet écueil, la meilleure pratique consiste à mettre au point plusieurs modèles « tests » de spécifications différentes, qui auront été calibrés de manière rigoureuse selon les principes du § 5. En particulier, la nécessité ou non d'emboîter le modèle, de même que la nécessité de segmenter ou non certains paramètres / constantes d'alternatives seront des résultats de ce processus de vérification statistique. La comparaison de ces modèles permettra de tirer des conclusions quant à l'outil satisfaisant le mieux au principe du rasoir d'Ockham. Ainsi, il sera d'usage d'examiner :

- quels paramètres / constantes d'alternatives permettent les améliorations les plus significatives de la log-vraisemblance,

¹⁷ Formulation moderne de « *Pluralitas non est ponenda sine necessitate* » (Une pluralité ne doit pas être posée sans nécessité). Guillaume d'Ockham (1319).

- si l'introduction de certains paramètres / constantes d'alternatives rendent significatives (ou non significatives) les estimations d'autres paramètres / constantes d'alternatives,
- si les valeurs des paramètres / constantes d'alternatives sont fortement dépendantes de la spécification.

Outre la reproduction des données à l'aide de la log-vraisemblance et d'indicateurs sur données agrégées, l'élasticité agrégée du modèle à certaines variables sera un critère décisif pour choisir quel modèle doit être conservé (Cf. § 7.2) parmi l'ensemble des modèles « tests » qui auront été estimés.

La modélisation du choix modal réalisée sur l'aire urbaine de Rouen, et décrite dans [ANSELME et COMBES, 2012], servira d'exemple pour illustrer ce processus. Trois modes de transport sont retenus : marche à pied (MAP), transports en commun (TC) et voiture particulière (VP).

Dans le § 6.2.3, 5 modèles de natures différentes vont ainsi être estimés :

1. modèle incluant des paramètres de niveau de service pour chaque mode (incluant les contraintes de stationnement), et des constantes d'alternatives distinctes selon les motifs ;
2. modèle incluant des paramètres de niveau de service pour chaque mode (incluant les contraintes de stationnement), et complètement segmenté selon les motifs ;
3. modèle incluant des paramètres de niveau de service pour chaque mode (incluant les contraintes de stationnement), et complètement segmenté selon les motifs x le niveau de captivité à la voiture (basé sur la définition de la captivité proposée par [HASIAK et al., 2011]) ;
4. modèle incluant des paramètres de niveau de service pour chaque mode (incluant les contraintes de stationnement), avec transformation Box-Cox de la variable temps, et complètement segmenté selon les motifs ;
5. modèle incluant des paramètres de niveau de service pour chaque mode (incluant les contraintes de stationnement), et des constantes d'alternatives distinctes selon les motifs, la captivité, la CSP, le niveau d'études, les habitudes d'utilisation des différents modes de transport, le sexe, et la possession d'un abonnement de transport.

6.2.2 Détection des valeurs aberrantes dans les bases de données

L'échantillon servant à estimer le modèle peut contenir des situations de choix observées conduisant à une estimation faussée des paramètres :

- soit à cause d'erreurs importantes sur les valeurs des variables entrant dans la fonction d'utilité (mauvais codage de l'origine ou de la destination, problème de calcul d'itinéraire...) ;
- soit à cause de la capture aléatoire par l'enquête d'un comportement très marginal (probabilité $< 10^{-40}$) ou relevant de considérations autres que celles entrant dans le champ du modèle (par ex. une randonnée).

Chaque observation entrant dans l'estimateur de la log-vraisemblance comptant pour le logarithme népérien de sa probabilité, ces données aberrantes ont un poids disproportionné lors du processus de maximisation de la log-vraisemblance (ainsi qu'indiqué sur le tableau ci-dessous). En outre, elles ralentissent très fortement la procédure d'estimation.

Probabilité	Contribution à la log-vraisemblance
50 %	-0,69
10 %	-2,30
1 %	-4,60
0,1 %	-6,90
10⁻³⁸ %	-92,10

Lorsqu'un modèle a été estimé, il est donc impératif de vérifier en tout premier lieu l'existence de valeurs de probabilité aberrantes pour certains choix.

Dans un premier temps, un examen de la distribution de la contribution de chaque observation à la log-vraisemblance permettra de déterminer la présence de données aberrantes et la valeur du seuil en deçà duquel un examen particulier des données doit être conduit.

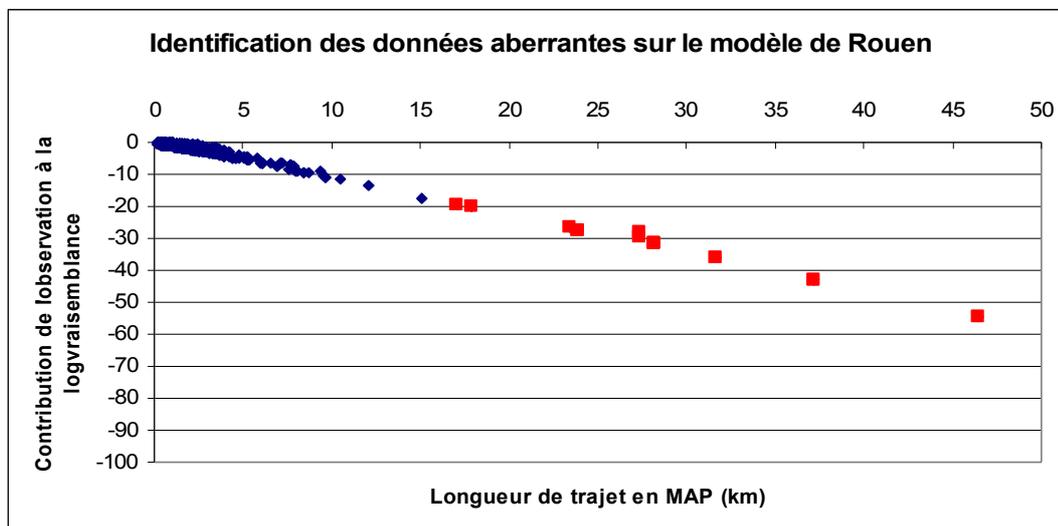
Dans notre exemple du modèle de choix modal de l'aire urbaine de Rouen, nous obtenons (à partir du modèle de base proposé au § 6.2.3.1) le tableau ci-dessous et choisissons ainsi la valeur de 0,000001 %, comme seuil : 15 observations sur un total de 25 615 apportent une contribution de -545 à la log-vraisemblance (sur -10 395), soit 5,3 % !

Probabilité du choix observé	Contribution d'une observation à la log-vraisemblance	Nombre d'observations	Part des observations	Part de la log-vraisemblance
> à 50 %	> à -0,69	21 114	82 %	29,7 %
de 10 % à 50 %	De -2,3 à -0,69	3 902	15 %	45,7 %
De 1 % à 10 %	De -4,61 à -2,3	527	2 %	15,3 %
De 0,1 % à 1 %	De -6,91 à -4,61	27	0,11 %	1,5 %
De 0,01 % à 0,1 %	De -9,21 à -6,91	18	0,07 %	1,4 %
De 0,001 % à 0,01 %	De -11,51 à -9,21	9	0,04 %	0,9 %
De 0,0001 % à 0,001 %	De -13,82 à -11,51	2	0,01 %	0,2 %
De 0,00001 % à 0,0001 %	De -16,12 à -13,82	1	0,00 %	0,2 %
De 0,000001 % à 0,00001 %	De -18,42 à -16,12	0	0,00 %	0,0 %
< à 0,000001 %	> à -18,42	15	0,06 %	5,3 %

Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

La deuxième étape consiste à examiner les raisons de ces valeurs aberrantes de la probabilité modélisée du choix observé. Si la quantité d'observations est restreinte, un examen observation par observation devrait apporter très rapidement des enseignements très précieux : s'agit-il de comportements non modélisables ? Y-a-t-il une erreur de codage sur une zone ? Dans le cas contraire, il faudra rechercher des invariants au sein de ce groupe, par exemple en traçant une courbe des valeurs prises par chaque variable en fonction de la contribution de l'observation à la log-vraisemblance.

L'une ou l'autre méthode (voir en particulier le graphique ci-dessous) conduisent ainsi dans l'exemple à identifier des distances en marche à pied très importantes comme la cause unique des 15 valeurs aberrantes sélectionnées (en rouge dans le graphique ci-dessous). Il s'agit probablement soit d'erreurs de codage des OD, soit de déplacements de type randonnée que le modèle ne saurait reproduire. Nous avons donc éliminé ces observations pour la suite des calculs.



Le diagnostic effectué sur les valeurs aberrantes permet finalement d'écarter certaines observations (après justifications) ou de modifier la spécification du modèle.

6.2.3 Mise en œuvre du processus

S'il y a des « recettes » pour les vérifications statistiques, aucune n'existe pour comparer des modèles. Pour obtenir un outil qui soit réellement utile, la seule méthode consiste à présenter les résultats obtenus de la manière la plus transparente et ouverte possible, afin qu'un consensus puisse s'établir. Des indications pratiques sont exposées dans la suite de ce paragraphe pour montrer comment parvenir à ce résultat.

Cinq modèles de spécifications différentes ont été construits pour le modèle de l'aire urbaine de Rouen. Sans être exhaustive (dans la plupart des cas, plus de cinq modèles devront être testés avant d'aboutir), cette démarche donne un aperçu suffisamment complet des analyses à conduire.

6.2.3.1 Modèle 1 : modèle simple

Le premier modèle testé est très simple dans sa construction et servira de référence pour évaluer l'intérêt de modèles plus complexes. Il est ici explicitement choisi de ne retenir que des constantes d'alternatives par motif et non des paramètres. Ceci permettra, via le second modèle (Cf. 6.2.3.2 Modèle 2 : modèle segmenté selon les motifs de déplacement), de mesurer l'intérêt d'une segmentation complète vis à vis d'une segmentation partielle. Il est à noter que l'offre TC a été codifiée pour reproduire correctement l'offre scolaire.

Les fonctions d'utilité utilisées sont les suivantes :

$$U_{MAP} = \beta_{T_{MAP}} \times T_{MAP} \quad 18$$

$$U_{VP} = \beta_{T_{VP}} \times T_{VP} + \beta_{STAT_{moyen}} \times STAT_{moyen} + \beta_{STAT_{fort}} \times STAT_{fort} + ASC_{motif_{VP}}$$

$$U_{TC} = \beta_{T_{TC}} \times T_{TC} + \beta_{IWT_{TC}} \times IWT_{TC} + \beta_{ACC_{TC}} \times ACC_{TC} + ASC_{motif_{TC}}$$

où :

- T_{MAP} : temps de marche à pied
- T_{VP} : temps de parcours en voiture
- $STAT_{moyen}$: variable binaire valant 1 si la valeur la plus élevée des contraintes de stationnement des zones origine et destination est moyenne
- $STAT_{fort}$: variable binaire valant 1 si la valeur la plus élevée des contraintes de stationnement des zones origine et destination est forte
- T_{TC} : temps de parcours en TC de l'arrêt de départ à l'arrêt final, incluant tous les temps de parcours intermédiaires liés aux correspondances
- IWT_{TC} : temps d'attente à l'arrêt de départ, calculé comme la moitié de l'inverse du nombre de services par heure
- ACC_{TC} : somme des temps de marche à pied de l'origine vers l'arrêt de départ et de l'arrêt final vers la destination
- **motif** : motif du déplacement, selon 6 niveaux : domicile-études dans l'enseignement supérieur (**DE_SUP**), domicile-études dans l'enseignement secondaire (**DE_SECONDAIRE**), domicile-études dans l'enseignement primaire (**DE_PRIMAIRE**), domicile-travail (**DT**), domicile-autres (**DA**)

En premier lieu, un modèle logit multinomial est estimé, avec une vérification de l'hypothèse IIA, en vue de déterminer la structure de modélisation adaptée. Trois nids sont envisageables dans ce modèle :

1. modes individuels : VP et MAP,
2. modes motorisés : VP et TC,
3. modes alternatifs à la voiture : MAP et TC.

Selon la démarche exposée au § 5.2, une variable auxiliaire est calculée pour chaque mode de chaque nid, captant les corrélations entre les modes. Un modèle supplémentaire est alors testé pour chaque nid, en introduisant les deux variables auxiliaires dans la formule d'utilité. Les résultats sont exposés dans le tableau suivant :

18 Pour des raisons de lisibilité, la notation choisie pour l'exemple diffère de celle présentée dans le reste du document. Le mode m n'est pas indiqué en indice des variables ou des paramètres lorsque l'intitulé de la variable est suffisamment clair (par ex. la contrainte de stationnement). L'indice i de l'individu a également été supprimé.

Modèle	Base		Nid indiv		Nid motor		Nid alterVP	
	Valeur	t-tests	Valeur	t-tests	Valeur	t-tests	Valeur	t-tests
ASC_TC_DA	-2.92	-34.85	-3.62	-33.44	-2.79	-28.95	-2.40	-23.06
ASC_TC_DE_SUP	-2.48	-14.95	-3.71	-17.34	-2.47	-14.53	-1.97	-11.01
ASC_TC_DE_PRIMAIRE	-2.96	-14.81	-3.72	-17.38	-2.84	-13.90	-2.17	-9.81
ASC_TC_DE_SECONDAIRE	-1.99	-19.12	-3.23	-20.49	-1.98	-17.42	-1.59	-13.76
ASC_TC_DT	-2.72	-23.10	-3.44	-25.12	-2.60	-21.17	-2.31	-17.58
ASC_TC_SEC	-3.01	-31.70	-3.64	-31.94	-2.87	-26.44	-2.45	-21.37
ASC_VP_DA	-1.07	-20.26	-1.57	-22.13	-0.894	-15.85	-1.09	-16.22
ASC_VP_DE_SUP	-2.62	-15.64	-3.64	-17.41	-2.82	-16.84	-2.61	-15.20
ASC_VP_DE_PRIMAIRE	-1.54	-13.27	-2.19	-16.32	-1.33	-11.21	-1.56	-12.86
ASC_VP_DE_SECONDAIRE	-3.45	-32.65	-4.58	-28.46	-4.07	-31.55	-3.39	-30.63
ASC_VP_DT	-0.959	-10.69	-1.44	-14.46	-0.785	-8.45	-0.974	-9.88
ASC_VP_SEC	-0.703	-11.48	-1.08	-15.14	-0.542	-8.16	-0.740	-10.14
BETA_ACC_TIME_TC	-0.0522	-12.63	-0.0526	-12.38	-0.0530	-11.01	-0.0520	-12.49
BETA_IN_TIME_TC	-0.0149	-8.80	-0.0159	-9.10	-0.0153	-8.26	-0.0187	-10.32
BETA_IWT	-0.0243	-5.51	-0.0209	-4.96	-0.0258	-5.26	-0.0311	-6.33
BETA_STATIO_FORTE	-1.50	-32.62	-1.80	-30.36	-1.46	-26.75	-1.43	-31.46
BETA_STATIO_MOY	-0.505	-9.19	-0.642	-11.18	-0.494	-8.95	-0.494	-9.15
BETA_TIME_MAP	-0.103	-49.87	-0.150	-32.15	-0.102	-48.95	-0.0997	-37.91
BETA_TIME_VP	-0.0300	-6.52	-0.0347	-7.42	-0.0302	-6.38	-0.0145	-3.03
alterVP_MAP							-0.214	-3.52
alterVP_TC							0.658	14.18
indiv_MAP			-0.651	-12.65				
indiv_VP			0.0584	0.97				
motor_TC					0.0654	1.41		
motor_VP					-0.916	-9.36		
LogV	-11323.701		-11236.677		-11279.554		-11201.610	

Résultats de l'estimation du modèle 1, avec et sans tests de la propriété IIA. Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Ici, le nid « modes motorisés » ne doit pas être retenu : l'amélioration de la log-vraisemblance (+44 points) est certes suffisante pour passer le test du ratio de vraisemblance, mais elle reste néanmoins très mesurée, et par ailleurs la valeur du paramètre **motor_TC** (pondérant la variable auxiliaire captant les corrélations VP - TC pour le mode TC) n'est pas significativement non nulle (valeur absolue du t-test inférieure à 1,96).

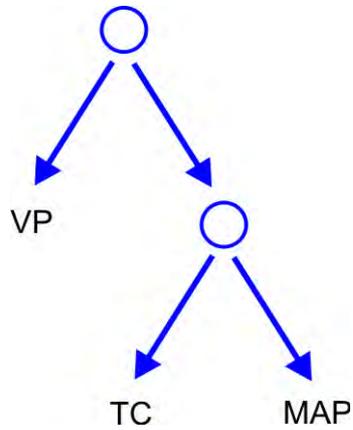
De même, pour le nid « modes individuels », la valeur du paramètre relatif à l'une des deux variables auxiliaires n'est pas significativement non nulle. Toutefois, l'amélioration de la log-vraisemblance est ici bien supérieure (+87 points), et impose de se questionner sur l'éventualité de garder ce nid.

Enfin, le nid « modes alternatifs à la VP » engendre une amélioration de la log-vraisemblance qui est encore supérieure (+122 points), et les deux paramètres des variables auxiliaires sont significativement non nuls. De plus, les valeurs des paramètres et constantes d'alternatives sont très fortement modifiées par l'introduction des variables auxiliaires : par

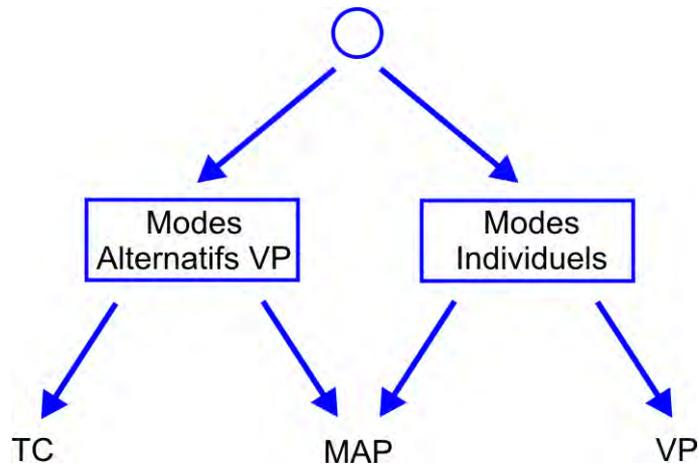
exemple le paramètre **BETA_TIME_VP** passe de -0,03 (dans le modèle de base) à -0,015. En conclusion, ce nid doit être conservé.

Ces premiers résultats conduisent donc à poursuivre l'élaboration du premier modèle en testant :

- un modèle nested logit avec un nid « alternatives à la voiture » :



- un modèle cross nested logit avec deux nids « alternatives à la voiture » et « modes individuels ».



Les résultats pour ces deux modèles sont présentés dans le tableau suivant :

Modèle	Nested Logit		Cross Nested Logit	
	Valeur	t-tests	Valeur	t-tests
ASC_TC_DA	-2.04	-21.23	-1.96	-38.32
ASC_TC_DE_SUP	-1.69	-12.63	-1.68	-15.75
ASC_TC_DE_PRIMAIRE	-2.09	-12.57	-1.89	-13.47
ASC_TC_DE_SECONDAIRE	-1.36	-14.94	-1.43	-54.47
ASC_TC_DT	-1.95	-17.73	-1.91	-17.73
ASC_TC_SEC	-2.08	-20.23	-1.96	-27.11
ASC_VP_DA	-0.806	-15.01	-0.795	-6.44
ASC_VP_DE_SUP	-2.42	-16.47	-2.46	-13.99
ASC_VP_DE_PRIMAIRE	-1.34	-12.06	-1.32	-8.18
ASC_VP_DE_SECONDAIRE	-3.27	-33.99	-3.33	-25.69
ASC_VP_DT	-0.713	-8.60	-0.709	-5.81
ASC_VP_SEC	-0.457	-7.50	-0.44	-2.25
BETA_ACC_TIME_TC	-0.0514	-14.97	-0.0513	-15.88
BETA_IN_TIME_TC	-0.0205	-11.20	-0.0209	-11.29
BETA_IWT	-0.032	-7.44	-0.0302	-7.44
BETA_STATIO_FORTE	-1.46	-32.39	-1.45	-32.50
BETA_STATIO_MOY	-0.517	-9.60	-0.519	-9.75
BETA_TIME_MAP	-0.0838	-35.79	-0.0819	-37.62
BETA_TIME_VP	-0.0163	-3.43	-0.0126	-3.11
MU_ALTERVP	1.63	24.32	2.74	18.25
MU_INDIV			1.00	
AlphaMAP_ALTERVP			0.536	49.56
AlphaMAP_INDIV			0.464	5.15
LogV	-11246.187		-11166.977	

Résultats de l'estimation du modèle 1, selon une spécification nested logit ou cross nested logit ; Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Le gain de log-vraisemblance apporté par le modèle cross nested logit reste modéré. En outre, les estimations sont quasiment identiques, en dehors du paramètre de temps de parcours VP. La valeur de ce dernier est plus faible pour le modèle cross nested logit, ce qui témoigne d'une élasticité agrégée moindre. On conclut dans le cas présent à l'abandon du modèle cross-nested logit, pour ne conserver finalement que la structure nested logit, avec un nid « modes alternatifs à la VP ».

L'étape suivante consiste à regrouper les segments de population, qui sont établis dans l'exemple sur la base des motifs de déplacement. Les valeurs des quasi t-tests de segmentation t_{seg} (test d'égalité pour chaque couple de segment, exposé au § 5.5), sont présentées dans les tableaux suivants.

Pour le mode TC, les valeurs des t_{seg} entre les motifs **DT**, **DA**, **DE_PRIMAIRE** et **SEC** (cases jaunes dans le tableau ci-dessous) sont très faibles. Ainsi, il apparaît que les motifs **DT**, **DA**, **DE_PRIMAIRE** et **SEC** ne font pas l'objet de préférences distinctes. D'autres valeurs de t_{seg} sont légèrement inférieures au seuil théorique (1,96) mais elles restent néanmoins nettement supérieures aux valeurs des motifs précédents.

Pour le mode VP, la valeur du t_{seg} entre les motifs **DT** et **DA** (case orange dans le tableau ci-dessous) est très faible. Ainsi, théoriquement, seuls les motifs **DT** et **DA** devraient être regroupés selon ce test. Il apparaît toutefois également que les valeurs des t_{seg} entre le motif **SEC** et respectivement **DA** et **DT** (cases roses dans le tableau ci-dessous) sont nettement plus faibles que les valeurs des autres motifs.

tseg		ASC_TC _DA	ASC_TC_ DE_SUP	ASC_TC_D E_PRI- MAIRE	ASC_T C_DE_ SECON DAIRE	ASC_T C_DT	ASC_TC_SE C
Mode TC	ASC_TC_DA		-1,65	0,21	-3,57	-0,44	0,20
	ASC_TC_DE _SUP			1,62	-1,63	1,22	1,83
	ASC_TC_DE _PRIMAIRE				-3,18	-0,58	-0,04
	ASC_TC_DE _SECON- DAIRE					3,06	3,75
	ASC_TC_DT						0,64
	ASC_TC_SE C						
tseg		ASC_TC _DA	ASC_VP_D E_ SUP	ASC_VP_D E_PRI- MAIRE	ASC_V P_DE_ SECON DAIRE	ASC_VP _DT	ASC_VP_SE C
Mode VP	ASC_VP_DA		10,11	3,97	19,71	-0,82	-3,71
	ASC_VP_DE _SUP			-5,75	4,70	-9,86	-12,17
	ASC_VP_DE _PRIMAIRE				12,12	-4,17	-6,47
	ASC_VP_DE _SECON- DAIRE					-18,03	-22,14
	ASC_VP_DT						-2,21
	ASC_VP_SE C						

Tests de segmentation du modèle 1, selon une spécification nested logit avec un nid « alternative à la VP », sans regroupements de motifs. Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Au vu de ces résultats, afin de garantir un traitement homogène des modes VP et TC, il est proposé de regrouper les motifs **DT**, **DA** et **SEC** pour les constantes d'alternatives de ces deux modes. Le nouveau modèle estimé conduit aux valeurs de t_{seg} présentées dans le tableau ci-dessous :

tseg		ASC_TC_DE_ SUP	ASC_TC_DE_ PRIMAIRE	ASC_TC_DE_ SECONDAIRE	ASC_TC_DT_ DA_SEC
Mode TC	ASC_TC_DE_ SUP		1,61	-1,68	1,67
	ASC_TC_DE_ PRIMAIRE			-3,20	-0,22
	ASC_TC_DE_ SECONDAIRE				3,67
	ASC_TC_DT_ DA_SEC				

tseg		ASC_TC_DE_ SUP	ASC_VP_DE_ PRIMAIRE	ASC_VP_DE_ SECONDAIRE	ASC_VP_DT_ DA_SEC
Mode VP	ASC_VP_DE_ SUP		-5,82	4,75	-10,95
	ASC_VP_DE_ PRIMAIRE			12,22	-4,80
	ASC_VP_DE_ SECONDAIRE				-20,92
	ASC_VP_DT_ DA_SEC				

Tests de segmentation du modèle 1, selon une spécification nested logit avec un nid « alternative à la VP », avec regroupements de motifs. Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Ces résultats mettent en évidence que la segmentation choisie traduit des différences de perception effectives, sauf pour le motif **DE_PRIMAIRE** pour le mode TC dont le t_{seg} avec le motif regroupé **DT_DA_SEC** reste faible (0,22). Cependant, les estimations des constantes d'alternatives sont respectivement de -2,05 et de -2,00, donc suffisamment proches pour qu'une nouvelle estimation du modèle ne soit pas nécessaire.

Le résultat final pour le premier modèle est exposé dans le tableau suivant :

Modèle 1		
Paramètres / constantes	Valeur	t-tests
ASC_TC_DE_SUP	-1.66	-12.59
ASC_TC_DE_PRIMAIRE	-2.05	-12.52
ASC_TC_DE_SECONDAIRE	-1.33	-14.89
ASC_TC_DT_DA_SEC	-2.00	-21.85
ASC_VP_DE_SUP	-2.40	-16.56
ASC_VP_DE_PRIMAIRE	-1.32	-11.97
ASC_VP_DE_SECONDAIRE	-3.25	-34.09
ASC_VP_DT_DA_SEC	-0.686	-13.54
BETA_ACC_TIME_TC	-0.0513	-15.13
BETA_IN_TIME_TC	-0.0207	-11.31
BETA_IWT	-0.0317	-7.50
BETA_STATIO_FORTE	-1.41	-31.74
BETA_STATIO_MOY	-0.503	-9.39
BETA_TIME_MAP	-0.0827	-35.77
BETA_TIME_VP	-0.0166	-3.52
MU_ALTERVP	1.67	24.35
LogV	-11279.544	

Résultat final de l'estimation du modèle 1 ; Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Ce premier modèle est d'ores et déjà efficace pour reproduire les choix modaux observés. Un modèle 0 fondé uniquement sur la reproduction des parts modales globales à l'aide de constantes d'alternatives a une log-vraisemblance de -17998. Le $\bar{\rho}^2$ ajusté vaut donc 0,37.

Ce premier modèle n'était que partiellement segmenté selon les motifs : les paramètres étaient supposés identiques pour tous les motifs (autrement dit seules les constantes d'alternatives étaient segmentées). Cette hypothèse est donc ensuite testée en développant un modèle totalement segmenté.

6.2.3.2 Modèle 2 : modèle segmenté selon les motifs de déplacement

Une démarche de segmentation complète (puis éventuellement de regroupement en fonction des résultats des tests statistiques) selon les motifs de déplacement est alors effectuée et permet d'obtenir les résultats présentés dans le tableau suivant :

Modèle 1			Modèle 2		
Paramètres / constantes	Valeur	t-tests	Paramètres / constantes	Valeur	t-tests
ASC_TC_DT_DA_SEC	-2.00	-21.85	ASC_TC_DA	-1.55	-16.38
ASC_TC_DE_SUP	-1.66	-12.59	ASC_TC_SUP_DT_SEC_ SECONDAIRE	-1.96	-23.21
ASC_TC_DE_SECONDAIRE	-1.33	-14.89			
ASC_TC_DE_PRIMAIRE	-2.05	-12.52	ASC_TC_PRIMAIRE	0.00	
ASC_VP_DE_SUP	-2.40	-16.56	ASC_VP_SUP	-2.68	-16.84
ASC_VP_DE_PRIMAIRE	-1.32	-11.97	ASC_VP_PRIMAIRE_DT_DA	-0.835	-15.73
ASC_VP_DT_DA_SEC	-0.686	-13.54			
			ASC_VP_SEC	-0.389	-5.94
ASC_VP_DE_SECONDAIRE	-3.25	-34.09	ASC_VP_SECONDAIRE	-3.39	-34.97
BETA_ACC_TIME_TC	-0.0513	-15.13	BETA_ACC_TIME_TC_DA_DT	-0.0671	-15.64
			BETA_ACC_TIME_TC_PRIMAIRE	-0.204	-10.63
			BETA_ACC_TIME_TC_SEC_SUP_ SECONDAIRE	-0.0375	-8.62
BETA_IN_TIME_TC	-0.0207	-11.31	BETA_IN_TIME_TC_DA_SUP	-0.0245	-9.30
			BETA_IN_TIME_TC_DT_SEC	-0.0317	-10.50
			BETA_IN_TIME_TC_PRIMAIRE_ SECONDAIRE	0.00	
BETA_IWT	-0.0317	-7.5	BETA_IWT_DA_SEC_PRIMAIRE	-0.0999	-11.90
			BETA_IWT_SUP_DT_SECONDAIRE	0.00	
BETA_STATIO_FORTE	-1.41	-31.74	BETA_STATIO_FORTE_DA_SEC_ PRIMAIRE	-1.60	-31.01
			BETA_STATIO_FORTE_DT	-1.14	-11.63
			BETA_STATIO_FORTE_SUP	-0.768	-3.48
			BETA_STATIO_FORTE_ SECONDAIRE	0.00	
BETA_STATIO_MOY	-0.503	-9.39	BETA_STATIO_MOY_DT_DA_ PRIMAIRE_SEC	-0.552	-9.64
			BETA_STATIO_MOY_SUP_ SECONDAIRE	-0.241	-1.54
BETA_TIME_MAP	-0.0827	-35.77	BETA_TIME_MAP_PRIMAIRE	-0.0630	-12.28
			BETA_TIME_MAP_autres	-0.0870	-38.34
BETA_TIME_VP	-0.0166	-3.52	BETA_TIME_VP_DA_SUP	-0.0192	-3.15
			BETA_TIME_VP_SECONDAIRE	0.00	
			BETA_TIME_VP_SEC_DT	-0.0341	-5.24
			BETA_TIME_VP_PRIMAIRE	0.00	
MU_ALTERVP	1.67	24.35	MU_ALTERVP	1.61	26.34
LogV	-11279.544			-11105.603	

Résultat final de l'estimation du modèle 2 ; Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

D'après le test du ratio de log-vraisemblance (LR=347 à comparer à une valeur de 1,64), le modèle 2 améliore suffisamment la reproduction des observations pour ne pas être un simple *artefact* de l'augmentation du nombre de paramètres et constantes d'alternatives. Néanmoins, cette amélioration de la log-vraisemblance reste faible. En outre, le modèle 2 n'ajoute que 6 nouveaux paramètres significativement non nuls et traduisant des différences de perception par rapport au modèle 1 à 16 paramètres. De nombreux résultats doivent être interprétés correctement au regard de l'ensemble des modèles qui seront construits :

- les déplacements **DE_PRIMAIRE** ne sont sensibles qu'au temps d'accès, au temps d'attente initial et aux contraintes de stationnement ;
- les déplacements **DE_SECONDAIRE** sont insensibles aux niveaux de service des modes motorisés, à l'exception du temps d'accès aux TC et de la contrainte de stationnement moyenne ;
- le temps d'attente initial (soit le nombre de services par heure) ne joue un rôle que pour trois motifs (**DA**, **SEC** et **DE_PRIMAIRE**), mais la pondération est très élevée (elle vaut 6 si elle est comparée à la pondération du temps VP du modèle 1) ;
- le temps d'accès est pondéré très fortement pour les déplacements **DE_PRIMAIRE** et plus fortement d'un ratio de 1 à 2 pour les motifs **DA** et **DT** que pour les autres motifs ;
- les contraintes de stationnement influencent assez faiblement les choix modaux pour les motifs **DE_SUP** et **DE_SECONDAIRE**.

Ainsi, les parts modales de différents motifs se retrouvent absolument insensibles à certains facteurs du niveau de service selon ce modèle, en diminuant d'autant son champ d'application.

Ce deuxième modèle met ainsi en évidence que les différences de perception des niveaux de service selon les motifs ne joue pas un rôle clé sur la répartition modale dans l'aire urbaine de Rouen.

6.2.3.3 Modèle 3 : modèle segmenté selon la captivité

Le troisième modèle est construit sur une variante complètement segmentée du premier modèle. Chaque paramètre / constante d'alternative du premier modèle est décomposé en trois éléments correspondant au niveau de captivité de l'individu à la voiture (niveau 1 : pas de voiture disponible / niveau 2 : voiture disponible mais pas de permis / niveau 3 : voiture disponible).

Les résultats obtenus sont les suivants :

Modèle 1			Modèle 3		
Paramètres / constantes	Valeur	t-tests	Paramètres / constantes	Valeur	t-tests
ASC_TC_DE_SUP	-1.66	-12.59	ASC_TC_DE_SUP_12	-1.54	-9.92
			ASC_TC_DE_SUP_3	-2.98	-16.31
ASC_TC_DE_PRIMAIRE	-2.05	-12.52	ASC_TC_DE_PRIMAIRE_123	-2.64	-15.92
ASC_TC_DE_SECONDAIRE	-1.33	-14.89	ASC_TC_DE_SECONDAIRE_123	-1.52	-23.30
ASC_TC_DT_DA_SEC	-2.00	-21.85	ASC_TC_DT_DA_SEC_12	-2.10	-53.70
			ASC_TC_DT_DA_SEC_3	-3.61	-0.00
ASC_VP_DE_SUP	-2.40	-16.56	ASC_VP_DE_SUP_12	-3.03	-10.84
			ASC_VP_DE_SUP_3	-1.56	-7.91
ASC_VP_DE_PRIMAIRE	-1.32	-11.97	ASC_VP_DE_PRIMAIRE_123	-1.27	-11.10
ASC_VP_DE_SECONDAIRE	-3.25	-34.09	ASC_VP_DE_SECONDAIRE_1	-5.36	-7.79
			ASC_VP_DE_SECONDAIRE_2	-2.97	-28.04
			ASC_VP_DE_SECONDAIRE_3	-0.632	-2.29
ASC_VP_DT_DA_SEC	-0.686	-13.54	ASC_VP_DT_DA_SEC_12	-1.83	-25.28
			ASC_VP_DT_DA_SEC_3	0.00	
BETA_ACC_TIME_TC	-0.0513	-15.13	BETA_ACC_TIME_TC_123	-0.0382	-12.19
BETA_IN_TIME_TC	-0.0207	-11.31	BETA_IN_TIME_TC_13	-0.0241	-10.30
			BETA_IN_TIME_TC_2	0.00	
BETA_IWT	-0.0317	-7.5	BETA_IWT_23	-0.0374	-8.95
			BETA_IWT_1	0.00	
BETA_STATIO_FORTE	-1.41	-31.74	BETA_STATIO_FORTE_123	-1.53	-33.87
BETA_STATIO_MOY	-0.503	-9.39	BETA_STATIO_MOY_123	-0.487	-8.52
BETA_TIME_MAP	-0.0827	-35.77	BETA_TIME_MAP_12	-0.0822	-34.55
			BETA_TIME_MAP_3	-0.101	-69.30
BETA_TIME_VP	-0.0166	-3.52	BETA_TIME_VP_13	-0.0858	-17.59
			BETA_TIME_VP_2	0.00	
MU_ALTERVP	1.67	24.35	MU_ALTERVP	1.36	105.61
LogV	-11279.544			-9735.680	

Résultat final de l'estimation du modèle 3 ; Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Ce troisième modèle permet une amélioration considérable de la log-vraisemblance : la valeur de \bar{p}^2 ajusté passe de 0,37 (pour le modèle 1) à 0,46. Ce modèle comporte des résultats très logiques : les signes et valeurs des constantes d'alternatives et des paramètres conduisent à ce que l'utilité des TC diminue lorsque la disponibilité d'une voiture augmente tandis que l'utilité de la voiture augmente. Seuls trois résultats sont moins intuitifs :

- les paramètres des temps TC à bord du véhicule et du temps de parcours en voiture ne sont pas significativement non nuls pour la valeur intermédiaire de disponibilité de la voiture ;
- le paramètre du temps d'attente initial n'est pas significativement non nul pour la valeur la plus faible de disponibilité de la voiture.

Il est à noter que, conformément à ce que l'analyse des résultats du modèle 2 pouvait laisser supposer, la segmentation partielle sur les motifs est une hypothèse réaliste : un modèle calibré selon une segmentation complète motifs x captivité a été testé, et n'aboutit qu'à une valeur de \bar{p}^2 ajusté de 0,48.

6.2.3.4 Modèle 4 : transformation des variables

Le modèle 4 correspond à l'introduction de transformations Box-Cox sur les variables de niveau de service du modèle 2.

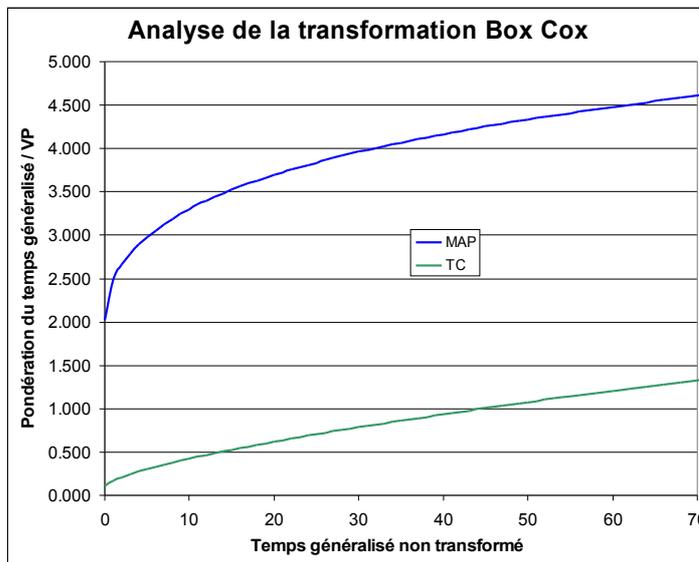
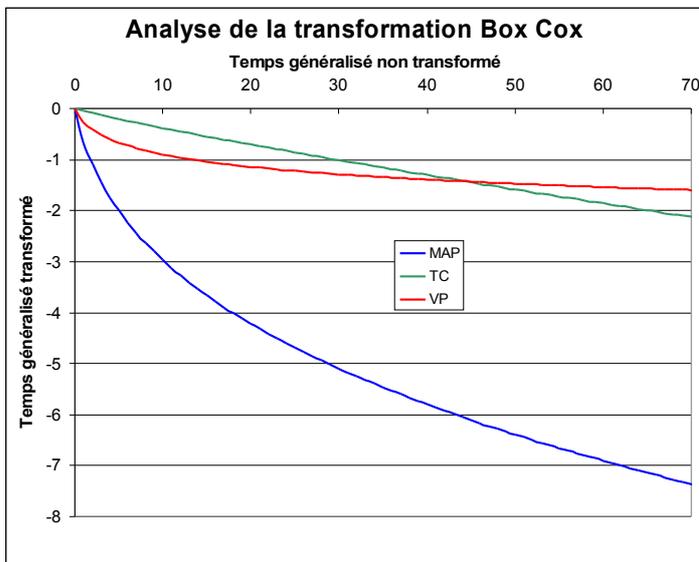
Les résultats obtenus sont les suivants :

Paramètres / constantes	Modèle 2		Modèle 4	
	Valeur	t-tests	Valeur	t-tests
ASC_TC_DA	-1.55	-16.38	-4.15	-13.50
ASC_TC_SUP_DT_SEC_SECONDAIRE	-1.96	-23.21	-4.54	-15.04
ASC_TC_PRIMAIRE	0.00		-2.79	-5.55
ASC_VP_SUP	-2.68	-16.84	-4.90	-17.21
ASC_VP_PRIMAIRE_DT_DA	-0.835	-15.73	-3.06	-12.63
ASC_VP_SEC	-0.389	-5.94	-2.59	-10.36
ASC_VP_SECONDAIRE	-3.39	-34.97	-6.20	-22.83
BETA_ACC_TIME_TC_DA_DT	-0.0671	-15.64	2.42	9.04
BETA_ACC_TIME_TC_PRIMAIRE	-0.204	-10.63	6.78	4.10
BETA_ACC_TIME_TC_SEC_SUP_SECONDAIRE	-0.0375	-8.62	1.39	7.19
BETA_IN_TIME_TC_DA_SUP	-0.0245	-9.30	1.00	
BETA_IN_TIME_TC_DT_SEC	-0.0317	-10.50	1.09	8.94
BETA_IN_TIME_TC_PRIMAIRE_SECONDAIRE	0.00		0.00	
BETA_IWT_DA_SEC_PRIMAIRE	-0.0999	-11.90	3.48	7.55
BETA_IWT_SUP_DT_SECONDAIRE	0.00		0.349	2.43
BETA_STATIO_FORTE_DA_SEC_PRIMAIRE	-1.60	-31.01	-1.52	-28.78
BETA_STATIO_FORTE_DT	-1.14	-11.63	-1.14	-11.12
BETA_STATIO_FORTE_SUP	-0.768	-3.48	-0.741	-3.32
BETA_STATIO_FORTE_SECONDAIRE	0.00		0.107	0.52
BETA_STATIO_MOY_DT_DA_PRIMAIRE_SEC	-0.552	-9.64	-0.498	-8.51
BETA_STATIO_MOY_SUP_SECONDAIRE	-0.241	-1.54	-0.184	-1.17
BETA_TIME_MAP_PRIMAIRE	-0.0630	-12.28	1.00	
BETA_TIME_MAP_autres	-0.0870	-38.34	1.23	0.00
BETA_TIME_VP_DA_SUP	-0.0192	-3.15	1.00	
BETA_TIME_VP_SECONDAIRE	0.00		0.0100	276
BETA_TIME_VP_SEC_DT	-0.0341	-5.24	1.23	4.22
BETA_TIME_VP_PRIMAIRE	0.00		0.00	
alpha_MAP			-0.839	-7.83
alpha_TC			-0.0473	-2.46
alpha_VP			-0.375	-7.46
lambda_map			0.306	8.73
lambda_tc			0.865	9.28
lambda_vp			0.00	152
MU_ALTERVP	1.61	26.34	1.70	26.06
LogV		-11105.603		-10936.152

Résultat final de l'estimation du modèle 4 ; Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Le modèle 4 induit une amélioration très modérée de la log-vraisemblance, le $\bar{\rho}^2$ ajusté passant de 0,38 (pour le modèle 2) à 0,39. Les constats sur les valeurs des paramètres sont très similaires à ceux du modèle 2, avec toutefois une amélioration notable : les paramètres du temps d'attente initial des motifs **DT**, **DE_SUP** et **DE_SECONDAIRE** sont significativement non nuls, contrairement à leur estimation dans le modèle 2. Comparé au temps de parcours à bord du véhicule des motifs **DA** et **DE_SUP**, il est pondéré d'un facteur environ égal à 3.

Les graphiques ci-dessous permettent de visualiser l'effet de la transformation Box-Cox :



Si pour un temps généralisé de quelques minutes, la pondération de la MAP par rapport à la VP est de l'ordre de 2, elle atteint un rapport de 4 à 1 pour un temps généralisé de 30 min. La pondération des TC par rapport à la VP croît quant à elle plus ou moins linéairement, de 0,1 pour les valeurs de temps généralisés faibles à 1,5 pour un temps généralisé de 70 minutes. Le temps passé à bord des TC devient plus pénalisant que celui passé en voiture à partir de trajets de 45 minutes de temps généralisé.

En conclusion, ce modèle 4 met en lumière des mécanismes de pondération du temps intéressants, même s'il n'apporte pas suffisamment d'amélioration du modèle pour être réellement pertinent.

6.2.3.5 Modèle 5 : introduction de constantes d'alternatives permettant de capter les préférences individuelles

Enfin, dans le modèle 5, la stabilité des paramètres de niveau de service du modèle 1 est testée en introduisant une grande quantité de constantes d'alternatives, afin de prendre en compte l'effet que pourraient avoir les préférences individuelles sur les valeurs des variables (détection de l'endogénéité, Cf. § 5.3). Cette comparaison doit tenir compte des facteurs d'échelle introduits par la formulation « nested logit » du modèle.

Des constantes d'alternatives ont été testées dans ce modèle pour l'ensemble des valeurs prises par :

- les motifs de déplacement,
- la captivité à la voiture,
- les CSP,
- les niveaux d'études,
- les habitudes d'utilisation des différents modes de transport,
- le sexe,
- la possession d'un abonnement de transport.

Les résultats sont les suivants :

Modèle	Modèle 1	Modèle 5	Comparaison paramètres	
	Valeur	Valeur	Modèle 1	Modèle 5
ASC_TC_DE_SUP	-1.66	-3.84	-2.77	-4.42
ASC_TC_DE_PRIMAIRE	-2.05	-3.47	-3.42	-3.99
ASC_TC_DE_SECONDAIRE	-1.33	-3.23	-2.22	-3.71
ASC_TC_DT_DA_SEC	-2.00	-4.00	-3.34	-4.6
ASC_VP_DE_SUP	-2.40	-2.99	-2.4	-3.44
ASC_VP_DE_PRIMAIRE	-1.32	-1.18	-1.32	-1.18
ASC_VP_DE_SECONDAIRE	-3.25	-2.68	-3.25	-2.68
ASC_VP_DT_DA_SEC	-0.686	-1.43	-0.69	-1.43
ASC_CSP0_TC		0.727		
ASC_CSP1_VP		11.00		
ASC_occupation_principale3_VP		0.316		
ASC_occupation_principale9_VP		0.461		
ASC_derniere_ecole0_TC		-0.285		
ASC_derniere_ecole5_TCVP		-0.219		
ASC_TC_METRO		0.349		
ASC_abo_tc_inter_nonvalid		-1.84		
ASC_abo_tc_non		-0.558		
ASC_abo_tc_urb_nonvalid		-1.09		
ASC_reseau_urbain0		-6.27		
ASC_reseau_urbain1		2.72		
ASC_reseau_urbain2		1.63		
ASC_train1		0.854		
ASC_VP_conducteur1		1.51		
ASC_VP_conducteur345		-0.815		
ASC_VP_passager0		1.50		
ASC_VP_passager1		1.12		
ASC_segment1_VP		-1.19		
ASC_segment3_VP		0.129		
BETA_ACC_TIME_TC	-0.0513	-0.0405	-0.0857	-0.0466
BETA_IN_TIME_TC	-0.0207	-0.00766	-0.0346	-0.0088
BETA_IWT	-0.0317	-0.00766	-0.053	-0.0088
BETA_STATIO_FORTE	-1.41	-1.37	-1.41	-1.37
BETA_STATIO_MOY	-0.503	-0.419	-0.503	-0.419
BETA_TIME_MAP	-0.0827	-0.0912	-0.138	-0.105
BETA_TIME_VP	-0.0166	-0.0164	-0.0166	-0.0164
MU_ALTERVP	1.67	1.15		
LogV	-11279.544	-7846.989		

Résultat final de l'estimation du modèle 5 ;Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Les constantes d'alternatives retenues ci-dessus sont celles qui ressortent au final comme étant significatives (à partir des résultats des t-tests). Les notations utilisées sont les suivantes :

- **Motif** : idem modèle 1
- **CSP** : catégories socioprofessionnelles (0 = pas de réponse ; 1 = agriculteur)
- **Occupation principale** : 3 = apprentissage ; 9 = autre
- **Dernière école fréquentée** (0 = en cours de scolarité ; 5 = supérieur BAC+3 et plus)
- **Sexe**
- **METRO** : 1 si le trajet correspondant à l'OD de l'usager utilise le métro, 0 sinon
- **abonnement** : possession d'un abonnement TC urbain ou interurbain, valide le jour du déplacement ou non
- **reseau_urbain, train, VP_conducteur, VP_passager** : fréquence de l'utilisation de ces modes de transport en semaine (0=pas de réponse ; 1=tous les jours ; 5=jamais)
- **segment** : disponibilité de la voiture, idem modèle 3

Le modèle 5 est plus performant que les autres modèles : son $\bar{\rho}^2$ ajusté vaut 0,56 à comparer à 0,37 pour le modèle 1 et 0,46 pour le modèle 3. Les résultats sont donc très significatifs.

Les valeurs retenues comme significativement non nulles dans ce modèle donnent un aperçu intéressant des facteurs de premier ordre :

- En ce qui concerne la catégorie sociale de l'individu (activité, type d'emploi, niveau de diplôme), la plupart des constantes non nulles correspondent à des valeurs de non réponse ainsi qu'à certaines situations particulières (agriculteurs, apprentissage).
- L'absence de possession d'un abonnement TC, ou la possession d'un abonnement TC non valide, ressortent également fortement.
- Les habitudes de déplacement entrent dans la fonction d'utilité exclusivement pour les usagers n'ayant pas répondu ou les usagers utilisant très fréquemment un mode de transport : il n'y a donc pas de gradualisme de la constante en fonction du niveau d'habitude.
- Enfin, la disponibilité de la voiture et l'utilisation du métro dans le chemin TC sont également conservés.

Il ressort également de ce modèle 5 que les paramètres de niveau de service VP sont stables, mais *a contrario* que les paramètres de niveau de service TC et MAP évoluent considérablement avec l'intégration de ces constantes d'alternatives.

6.2.3.6 Discussion à partir des résultats de ces 5 modèles, et choix du modèle à retenir

Les 5 modèles estimés présentent différentes caractéristiques en ce qui concerne les tests statistiques et les parts modales modélisées. La comparaison de ces éléments offre des clés de compréhension pour retenir une formulation finale. Toutefois, il convient de rappeler que l'expérience et le bon sens du modélisateur jouent également un rôle crucial.

L'examen séparé de ces différents modèles apporte donc différentes conclusions :

- La structure « nested logit » reste nécessaire pour tous les modèles.
- La segmentation complète selon les motifs (**modèle 2**) peut être abandonnée.
- La transformation Box-Cox des niveaux de service (**modèle 4**) n'est pas essentielle.

- La segmentation totale selon la disponibilité de la voiture joue un rôle décisif dans la reproduction des choix effectués : mais peut-on l'utiliser sans modèle d'équipement automobile ? Il pourrait également être intéressant d'améliorer la définition de cette notion, notamment pour les motifs de déplacement « école primaire » et « école secondaire ».
- d'autres préférences apparaissent que celles liées au motif et à la disponibilité de la voiture, en particulier les habitudes (pas de réponse et utilisation très fréquente des TC) et l'absence de possession d'un abonnement valide. Celles-ci influencent fortement l'estimation des paramètres. Doit-on conserver les résultats de cette formulation, même si elle oblige à la formulation d'hypothèses « manuelles » pour appliquer le modèle ? En particulier, l'intégration des habitudes n'est aucunement opérationnelle dans le cadre d'un modèle prédictif, alors qu'elle peut être utile pour expliquer les comportements actuels.

Il convient donc d'approfondir la comparaison des trois modèles suivants : **modèle 1** (modèle simple non segmenté), **modèle 3** (modèle segmenté selon la captivité) et **modèle 5** (modèle non segmenté avec de nombreuses constantes d'alternatives). L'élasticité agrégée (Cf. définition § 7.2) du modèle aux différentes composantes du niveau de service est un indicateur qui permet d'effectuer des comparaisons intéressantes :

Élasticités agrégées aux variables :	Modèle 1	Modèle 3	Modèle 5
TIME_MAP	-1,088	-1,042	-1,053
TIME_VP	-0,021	-0,056	-0,022
STATIO_FORTE	-0,063	-0,051	-0,060
STATIO_MOY	-0,012	-0,009	-0,010
ACC_TIME_TC	-0,530	-0,317	-0,380
IN_TIME_TC	-0,283	-0,144	-0,114
IWT	-0,163	-0,125	-0,040

Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Ainsi, il apparaît que :

- L'élasticité au temps de parcours en voiture est plus élevée pour le **modèle 3**, tandis que l'élasticité au stationnement est cohérente pour les trois modèles.
- Le **modèle 5** donne des valeurs d'élasticité proches de celles du **modèle 1**, sauf pour les temps de parcours TC pour lesquels les divergences sont très élevées. Pour ces variables, les **modèles 3 et 5** sont relativement cohérents avec des élasticités plus faibles que celles du **modèle 1**.
- Les trois modèles donnent le même résultat d'élasticité au temps en marche à pied.

Le **modèle 1**, dont l'efficacité a été mise en évidence par rapport à un modèle de simple reproduction des parts modales, sert ici de témoin pour évaluer la pertinence d'un modèle complexe par rapport à un modèle très simple.

En ce qui concerne le **modèle 3**, la différence tient à la segmentation selon le niveau de captivité aux alternatives à la voiture. D'un point de vue comportemental, le **modèle 3** est nettement préférable puisque le calcul des utilités est cohérent avec la situation des individus. Néanmoins, il présente l'inconvénient de rendre nécessaire une décomposition en amont du choix de motorisation des ménages. Ainsi, la question de la comparaison de ces deux modèles se pose selon les termes suivants : le **modèle 1** non segmenté contient-il les effets des décisions d'équipement automobile ?

Le **modèle 3** fait diminuer les élasticités aux temps TC par rapport au **modèle 1** conformément à l'intuition. Comme le choix de motorisation résulte lui-même d'un arbitrage individuel entre des contraintes économiques (coût d'acquisition d'un véhicule) et des

contraintes de déplacement, le choix modal est en partie effectué dès la décision de l'achat ou non d'un véhicule. En revanche, la segmentation conduit cette fois à une augmentation de la sensibilité au temps VP, ce qui est assez peu intuitif. On en conclut que l'absence de segmentation ne permet pas de « compenser » l'absence de prise en compte de la captivité par une sensibilité plus forte à toutes les variables de niveau de service.

Face à de telles informations, comment le modélisateur peut-il alors choisir s'il vaut mieux segmenter ou non ?

Pour répondre, une démarche complémentaire peut être menée, fondée sur la réalisation de simulations de scénarios hypothétiques : il s'agit de tester des scénarios très volontaristes, dont la conséquence devrait être une diminution de l'équipement automobile des ménages. Cette démarche permet dans le même temps de tester les limites d'applicabilité du modèle en termes de variation des variables explicatives.

- Dans le premier scénario (scénario TC++), on suppose ainsi qu'un développement considérable de l'offre TC est pris en compte, permettant à tous les usagers de disposer d'une alternative TC de bon niveau de service : les temps d'attente initiale et de trajet terminal valent au plus 5 minutes (à chaque extrémité), et le temps de parcours (trajet à bord du véhicule + correspondances) est au maximum identique à celui de la voiture.
- Dans le deuxième scénario (scénario VP--), on suppose qu'une politique très contraignante pour la voiture particulière est mise en œuvre : la contrainte de stationnement est étendue à l'ensemble des zones denses, et des mesures de ralentissement des trajets en voiture sont prises de façon à ce que le temps total du trajet en TC soit au maximum 50 % plus élevé que celui en voiture.

Les résultats sont les suivants :

Mode de transport	Parts modales observées	Parts modales du modèle 1		Parts modales du modèle 3	
		Scénario TC++	Scénario VP--	Scénario TC++	Scénario VP--
	Observées				
MAP	16.6 %	15.7 %	18.2 %	15.8 %	21.3 %
TC	8.6 %	16.5 %	10.2 %	13.6 %	15.5 %
VP	74.8 %	67.8 %	71.6 %	70.6 %	63.2 %

Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Le scénario TC++ conduit à une baisse de 7 points de la part modale VP pour le **modèle 1** et de 4,2 points pour le **modèle 3**, tandis qu'elle baisse de respectivement 3,2 points et 11,6 points dans le scénario VP--. Ainsi, le **modèle 3** est moins sensible que le **modèle 1** à une évolution de l'offre TC, mais la différence est légère ; *a contrario*, le **modèle 3** est très nettement plus sensible à une évolution de l'offre VP que le **modèle 1**. L'hypothèse selon laquelle le modèle non segmenté « contiendrait » les effets de décision d'équipement automobile des ménages est ainsi contredite par ces résultats, ce qui conduit à préférer le **modèle 3** au **modèle 1**.

Le **modèle 5** diffère quant à lui du **modèle 1** par l'intégration d'un nombre important de constantes d'alternatives, incluant en particulier les abonnements TC et les habitudes modales. Ainsi, les termes de la comparaison entre les deux modèles sont sensiblement identiques à ceux de la comparaison entre les **modèles 1 et 3** : le **modèle 1** contient-il des effets d'habitudes modales et de choix d'abonnement ? En termes de sensibilité, la seule différence significative concerne la sensibilité à l'offre TC qui diminue nettement dans le **modèle 5** par rapport au **modèle 1**, notamment en ce qui concerne le temps d'attente initial auquel le modèle devient presque insensible. Cela signifie entre autres que les constantes modales jouent un rôle très important dans les choix modaux, ce qui est peu souhaitable dans

le cadre d'un modèle prédictif. Ainsi, dans la mesure où d'une part l'objectif visé est de construire un modèle prédictif, où d'autre part aucun modèle de prévision de la composition des classes de population utilisées par le **modèle 5** ne vient alimenter le modèle de choix modal, ces résultats conduisent plutôt à rejeter le **modèle 5**.

Le **modèle 3** est ainsi retenu comme base de travail. Il reste encore à diagnostiquer dans quelle mesure le **modèle 3** peut être amélioré.

Pour cela, la qualité de la reproduction des données agrégées permet de mettre en évidence d'éventuels domaines pour lesquels le **modèle 3** est peu performant. En outre, la comparaison des résultats agrégés du **modèle 3** avec le **modèle 5** est intéressante, car le **modèle 5**, même s'il n'a pas été retenu, présente de grandes qualités en termes d'ajustement statistique qui peuvent fournir des indications quant au potentiel d'amélioration du **modèle 3**. Dans l'exemple, le croisement de la classe de distance du déplacement et de la typologie des zones d'origine et de destination (urbain, périurbain, rural) fournit les informations les plus intéressantes :

Typologie de l'OD	Classe de distance	Nb Interviews	Modèle	%MAP	%VP	%TC	% erreur
periurbain_periurbain	< 2 km	1587	Obs	28%	66%	6%	
			M3	32%	61%	7%	9,4%
			M5	31%	63%	6%	7,2%
periurbain_urbain	< 2 km	2900	Obs	38%	54%	8%	
			M3	40%	52%	8%	3,1%
			M5	41%	51%	8%	5,3%
rural_rural	< 2 km	521	Obs	17%	80%	3%	
			M3	24%	73%	3%	14,4%
			M5	24%	74%	2%	14,9%
rural_urbain	< 2 km	1459	Obs	44%	47%	9%	
			M3	41%	51%	8%	6,6%
			M5	43%	47%	10%	1,9%
urbain_urbain	< 2 km	2388	Obs	57%	36%	7%	
			M3	49%	43%	8%	16,5%
			M5	52%	40%	8%	11,7%
	< 5 km	936	Obs	4%	63%	33%	
			M3	6%	71%	23%	19,4%
			M5	5%	67%	28%	9,0%
	de 5 à 10 km	250	Obs	1%	70%	29%	
			M3	0%	77%	23%	12,8%
			M5	0%	74%	26%	8,0%
Autre (classes pour lesquelles l'erreur est < 5%)		15582	Obs	4%	88%	8%	
			M3	4%	88%	8%	2,0%
			M5	4%	88%	8%	1,6%
Total		25598	Obs	17%	74%	9%	
			M3	16%	76%	8%	5,2%
			M5	17%	74%	9%	3,9%

Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

Les deux modèles reproduisent globalement de manière très satisfaisante les choix agrégés selon ces segments, avec un pourcentage global d'erreur de respectivement 5,2 % pour le **modèle 3** et 3,9 % pour le **modèle 5**. Pour ces deux modèles, une erreur systématique apparaît sur la reproduction de la concurrence modale MAP/VP pour les déplacements courts (< 2 km de distance à vol d'oiseau) : les modèles ont tendance à sous-estimer la part de la MAP pour les déplacements en lien avec le milieu urbain et à la surestimer si le déplacement n'est pas en lien avec le milieu urbain.

Diverses explications peuvent être émises à propos de ce biais systématique. Il pourrait être relié au choix de ne pas conserver les déplacements intrazonaux, ce qui entraîne des différences de traitement entre les milieux périurbains et ruraux, pour lesquels les zones sont assez vastes, et le milieu urbain où les zones sont plus petites. Ainsi, la part MAP serait artificiellement faible en milieu périurbain et rural. Les différences de taille de zone peuvent

également introduire des biais d'agrégation : les déplacements entre les zones de grande taille auront un écart type de temps de déplacement plus important que les zones de petite taille. Les choix modaux ne sont donc pas correctement caractérisés par la moyenne des temps de parcours en MAP pour des classes de population qui peuvent représenter une part importante de l'échantillon. Comme le choix modal n'est pas linéaire, la moyenne des choix modaux n'est donc pas nécessairement proche du choix modal de l'utilisateur moyen. Au vu de ces éléments, il apparaît qu'une nette amélioration du modèle passe par une réflexion sur la prise en compte de la MAP. Il est notamment possible de s'interroger sur la pertinence de conserver la MAP, alors qu'il relève de logiques très différentes des autres modes.

Une faiblesse du modèle quant à la reproduction de la part modale TC sur les déplacements de plus de 2 km en milieu urbain apparaît également, alors qu'il s'agit du type de déplacements pour lesquels cette part est la plus importante. Néanmoins, les constantes d'alternatives du **modèle 5** ne corrigent que partiellement cette sous-estimation. En fait, le **modèle 5** et le **modèle 3** font exactement le même type d'erreur : une réelle amélioration du modèle passe donc par d'autres modifications que l'introduction d'une meilleure reproduction des préférences des individus.

Ainsi, des améliorations du **modèle 3** sont possibles et les critères de progrès apparaissent clairement. Les erreurs sont néanmoins dans des ordres de grandeur largement acceptables, aucune part modale ne s'écartant de plus de 10 points (et généralement là où les échantillons sont faibles) de la part de référence : **la démarche de construction conduit donc ici à retenir le modèle 3 comme un outil satisfaisant pour effectuer des prévisions de partage modal selon différentes politiques d'offre de transports.**

Toutefois, dans la mesure où ce modèle utilise en entrée des segments de captivité, il faudra être en mesure de prévoir l'évolution de ces segments dans le temps (éventuellement en fonction de la nouvelle offre de transports), soit à l'aide d'hypothèses exogènes (par ex. en raisonnant par analogies), soit à l'aide d'un modèle de prévision de l'évolution de ces segments.

Rappelons enfin qu'il s'agissait ici de construire un modèle de choix modal dans le but d'étudier un certain nombre de politiques liées à l'offre de transports. Ce modèle ne serait donc pas adapté pour étudier des politiques liées à l'urbanisation, par exemple à la localisation des différents groupes sociaux : dans ce cas, une segmentation totale du modèle selon ce critère devrait être mise en œuvre, en utilisant une démarche similaire à ce qui a été présenté précédemment. Ainsi, on voit qu'il ne peut y avoir de modèle générique valable pour tout type d'étude (d'ailleurs, le modèle construit dans cet exemple n'a aucune autre valeur que son intérêt pédagogique) : chaque étude nécessite donc bien un modèle spécifique, et la mise en œuvre d'une démarche de construction spécifique qui pourra s'inspirer du processus intellectuel présenté ici.

6.3 Différents problèmes spécifiques

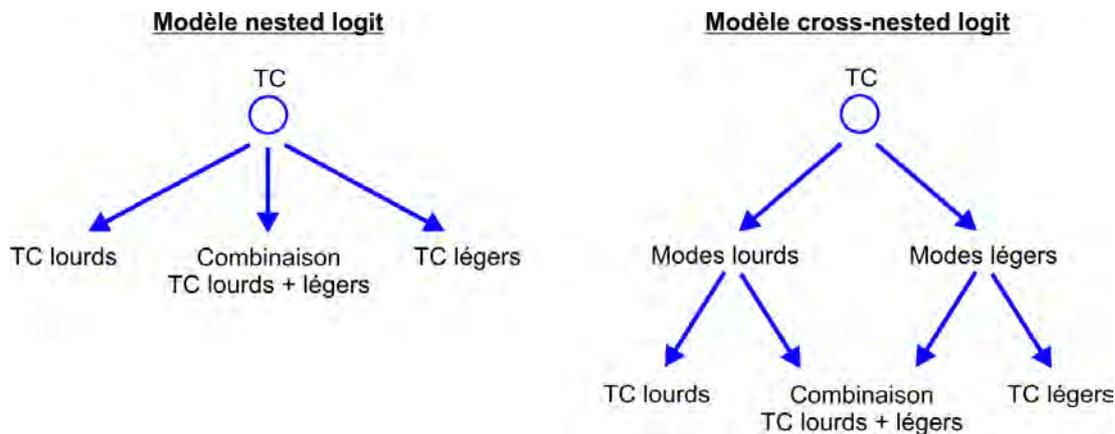
6.3.1 Distinction modes légers / modes lourds de transports en commun

Lorsque le réseau de transports en commun étudié comporte des modes lourds du type métro, tramway, voire dans une certaine mesure des BHNS¹⁹, il peut être jugé utile de chercher à différencier les différents modes TC dès le modèle de choix modal (cette différenciation devra également être effectuée dans l'affectation, mais cette étape ne sera pas traitée ici). Cette différenciation a du sens si l'on suspecte que les comportements des usagers face à un même jeu de concurrence en termes de niveaux de services (temps, coûts) diffèrent selon le mode TC. Plusieurs éléments peuvent expliquer cette hétérogénéité des comportements : on peut citer par exemple l'effet d'image, le confort, ou la régularité (qui n'aurait pas pu être prise en compte dans la définition de l'utilité classique faute de données).

Afin d'effectuer cette différenciation, deux pratiques peuvent être utilisées :

19 Bus à haut niveau de service

- Conserver un seul mode TC dans le modèle de choix modal, mais introduire un bonus spécifique pour les modes lourds dans la fonction d'utilité ou une pondération du temps différente (en lien avec l'affectation). L'inconvénient de cette méthode est que la répartition entre les différents sous-modes TC est assurée au niveau de l'affectation, ce qui pourra rapidement conduire à l'adoption de paramètres du coût généralisé différents de ceux utilisés pour le choix modal.
- Considérer les TC lourds et les TC légers comme des modes à part entière (éventuellement au sein d'un nid de sous-modes TC), dont les paramètres sont à estimer. La combinaison de ces différents sous-modes TC doit également être considérée comme un mode au même titre que les autres. Selon les résultats des tests IIA (Cf. § 5.2), on aboutirait ainsi à un modèle nested ou cross-nested logit dont la structure serait la suivante :



Remarque : il est possible, par extension, de différencier les différents modes lourds entre eux (métro, tramway, BHNS ...) par le même type de méthode.

Pour mettre en place cette méthode, il est nécessaire de vérifier la disponibilité du sous-mode puis d'obtenir les valeurs de coût généralisé pour chacun des sous-modes disponibles. Une solution sera de coder plusieurs sous-réseaux de TC : un sous-réseau modes légers (dans lequel les modes lourds ne sont pas accessibles), un sous-réseau modes lourds (dans lequel les modes légers ne sont pas accessibles), et un réseau complet où les combinaisons sont possibles. Concernant ce dernier, il est important de préciser qu'un paramétrage spécifique du modèle d'affectation, qui favorise les combinaisons de modes, devra probablement être recherché. Il faut théoriquement définir un modèle d'affectation permettant d'obtenir exclusivement des itinéraires combinant modes légers et modes lourds, ce qui nécessite un paramétrage très délicat au vu de la faible quantité de données disponibles. Une autre alternative serait de sortir les coûts généralisés des différentes combinaisons de modes dans le cas d'algorithmes multi-chemins comme les stratégies optimales (Cf. § 4.2).

Une possibilité consisterait à utiliser un algorithme d'énumération des chemins puis de répartition stochastique. Cette méthode reste toutefois très peu utilisée et nécessitera très probablement la rédaction d'un programme de recherche de chemins ad hoc (peu de logiciels sur le marché disposant de tels outils). Cette méthode présente donc l'avantage de poser moins de problèmes de cohérence avec l'affectation (sauf pour le mode constitué de la combinaison des sous-modes TC), mais sa mise en œuvre est lourde en pratique (plusieurs réseaux différents, donc plusieurs affectations différentes à réaliser, etc.). Par ailleurs, la taille des échantillons dans les EMD sera souvent insuffisante pour estimer proprement le modèle de choix modal : une collecte de données complémentaires, par exemple de préférences déclarées, sera inévitable.

Sur l'exemple de l'agglomération rouennaise, la première méthode a été testée pour la différenciation du métro avec le reste des TC, à partir du **modèle 1** (même si ce n'est pas le modèle qui a finalement été jugé le plus satisfaisant, il a le mérite d'être plus simple que les autres modèles, et permet donc de réaliser des interprétations plus aisées). Les résultats obtenus sont les suivants :

Paramètres / constantes	Modèle 1		Modèle 1 + Bonus métro	
	Valeur	t-tests	Valeur	t-tests
ASC_MAP	0.00	fixed	0.00	fixed
ASC_TC_DE_ETU	-1.66	-12.59	-1.78	-13.00
ASC_TC_DE_PRIMAIRE	-2.05	-12.52	-2.20	-13.02
ASC_TC_DE_SECONDAIRE	-1.33	-14.89	-1.49	-15.88
ASC_TC_DT_DA_SEC	-2.00	-21.85	-2.21	-23.01
ASC_VP_DE_ETU	-2.41	-16.56	-2.41	-16.45
ASC_VP_DE_PRIMAIRE	-1.32	-11.97	-1.35	-12.21
ASC_VP_DE_SECONDAIRE	-3.25	-34.09	-3.26	-34.03
ASC_VP_DT_DA_SEC	-0.686	-13.54	-0.727	-14.35
BETA_ACC_TIME_TC	-0.0513	-15.13	-0.0534	-15.12
BETA_IN_TIME_TC	-0.0207	-11.31	-0.0208	-11.55
BETA_IWT	-0.0314	-7.50	-0.0212	-5.24
BETA_STATIO_FORTE	-1.41	-31.74	-1.39	-31.32
BETA_STATIO_MOY	-0.504	-9.39	-0.505	-9.40
BETA_TIME_MAP	-0.0826	-35.77	-0.0841	-36.54
BETA_TIME_VP	-0.0166	-3.52	-0.0150	-3.18
bonus_métro			0.413	8.21
MU_ALTERVP	1.67	24.35	1.58	24.87
LogV		-11279		-11244

Source : modèle de choix modal développé sur l'agglomération de Rouen par les auteurs.

On obtient, grâce au **modèle** avec bonus métro, une log-vraisemblance de -11 244 (contre -11 279 pour le **modèle 1**), ce qui traduit une amélioration du modèle, quoique relativement modérée. Il est également intéressant d'observer la capacité de ces modèles à reconstituer le nombre de déplacements en mode métro :

	Observé	Modèle 1	Modèle 1 + bonus métro
Nombre de déplacements avec métro	824	597	730

La distinction modes légers / modes lourds apporte donc très clairement un gain dans la reconstitution des déplacements en métro, même si le résultat n'est pas encore parfait. Une amélioration du modèle d'affectation serait ici nécessaire.

La mise en œuvre de la méthode 2 fournirait de façon certaine de meilleurs résultats, mais elle nécessiterait un travail spécifique sur le paramétrage du modèle d'affectation pour la reconstitution du mode « combinaison métro + bus », qui n'a pas été mené ici.

En tout état de cause, il semble donc que pour les deux méthodes proposées, le principal effort porte sur le modèle d'affectation. Au niveau du choix modal, il est donc probablement préférable de se concentrer sur la plus simple des deux (c'est-à-dire la première) si les résultats obtenus sont jugés suffisants.

6.3.2 Effet du prix du déplacement

Le prix du déplacement est une des variables déterminantes du choix de mode qui vient immédiatement à l'esprit au moment du calibrage du modèle. Dans la pratique, l'estimation de ce paramètre peut toutefois s'avérer assez complexe. En effet, la calibration simultanée de paramètres liés au coût et au temps pose inévitablement la question de la calibration de la valeur du temps. Or celle-ci est souvent variable selon les individus et les déplacements, donc l'introduction d'une variable de coût unique génère un problème de variation de perception qu'il est nécessaire de corriger. Dans la littérature, ce problème est traité, soit en segmentant le modèle (Cf. § 5.5), soit en ayant recours à des modèles logit mixtes (famille des modèles *generalized extreme values*) introduisant des distributions de la valeur du temps au sein même de la fonction d'utilité, soit les deux (voir encadré ci-dessous). Ainsi, de nombreux auteurs (Cf. notamment [DE PALMA et FONTAN., 2001], [HESS et al., 2004], et [AXHAUSEN et al., 2008]) se sont penchés sur cette question et sont parvenus à calibrer des distributions de valeur du temps à l'aide de modèles logit mixtes, en mettant en évidence notamment le lien entre la valeur du temps et le revenu, la distance, ainsi que les motifs de déplacement. Ces méthodes sont reproductibles localement.

Modèle logit mixte

Dans un modèle logit mixte utilisant une distribution de la valeur du temps, l'utilité est alors fonction notamment du temps et du coût combinés de la manière suivante :

$$U_{i,m} = \beta_{\text{coût}} \times (\text{Coût}_{i,m} + f(\mu, \sigma) \times \text{temps}_{i,m}) + \beta_x \times X_{i,m} + \varepsilon_{i,m}$$

où :

$f(\mu, \sigma)$: valeur du temps dont la distribution (de paramètres μ et σ) est à estimer.

La distribution doit être choisie *a priori* (la distribution la plus couramment utilisée pour les valeurs du temps est la distribution lognormale), et l'estimation des paramètres de la distribution s'effectue ensuite par le logiciel de calibrage à l'aide de tirages aléatoires. Cette procédure peut donc s'avérer assez chronophage, mais est indispensable dans le cas où l'étude porte précisément sur l'influence de ce paramètre, notamment en présence de péages autoroutiers sur le territoire.

Toutefois, l'estimation correcte de ce paramètre peut localement poser problème :

- dans le cas des TC urbains, le prix est souvent constant quel que soit le déplacement (pour des aires urbaines hors Île-de-France). Il n'y a donc aucune variabilité dans les données observées, ce qui empêche d'estimer le paramètre associé au coût de manière spécifique à ce mode. Par ailleurs, la prise en compte des abonnements est une problématique à part, qui a été évoquée au § 6.2.3.
- Dans le cas des TC périurbains ou interurbains (ex : TER), les prix sont différents selon les OD, mais les échantillons d'une EMD peuvent s'avérer insuffisants pour le calibrage. Par ailleurs, la recomposition des montants réellement acquittés par les usagers peut s'avérer très complexe, en particulier lorsque l'exploitant pratique le *yield management*²⁰, qui peut générer des prix différents pour un même train et pour une même OD.

20 Système de gestion ayant pour objectif de proposer une variation des prix en fonction des places disponibles.

- Dans le cas du mode VP, l'utilisation d'une variable de coût définie comme un coût kilométrique introduit souvent une corrélation entre ce paramètre et celui associé au temps VP, ce qui peut fausser leur estimation (Cf. § 5.1.2 qui traite le problème des corrélations). La prise en compte de la variation de la consommation de carburant en fonction de la vitesse pourrait cependant permettre de limiter ce problème. En tout état de cause, même si l'on parvient à calibrer cette élasticité par rapport au coût global, il sera toutefois difficile d'isoler l'effet spécifique lié au carburant (ce qui pose donc des problèmes pour une utilisation en prospective), le prix du carburant ayant en effet le plus souvent été constant pendant toute la période d'enquête.

Ainsi, si la méthode de calibrage de la valeur du temps à l'aide de modèles logit mixtes mentionnée plus haut est impossible à mettre en œuvre, ou si la politique à tester ne porte pas explicitement sur le coût du déplacement, il est possible d'intégrer tout de même cette variable, à l'aide de l'une ou l'autre des méthodes ci-dessous :

- Choisir *a priori* une formulation de coût généralisé (ce qui impose de postuler également une valeur du temps – pour ce faire il est possible de s'aider du tableau ci-dessous adapté au milieu urbain), puis estimer, lors du calibrage du modèle de choix de mode, un paramètre global lié à ce coût généralisé. Notons qu'il sera alors impossible d'isoler la sensibilité propre des différentes composantes de ce coût généralisé : il s'agit donc plutôt « d'afficher » que le modèle est bien sensible au temps et au coût de transport.

Motif du déplacement	France entière	Île-de-France
Professionnel	17,5	22,3
Domicile-travail/études/garderie	10	12,6
Autres (achats, soin, visites, loisir, tourisme, etc.)	6,8	8,7
Sans détail du motif	7,9	10,7

Source : *l'évaluation socioéconomique des investissements publics – septembre 2013*

- Intégrer ce paramètre par l'intermédiaire de son élasticité obtenue sur une autre étude, comme explicité au § 7.3. Cette seconde méthode est théoriquement préférable à la première, dans le sens où elle permet de mieux maîtriser l'élasticité de chaque variable au sein du modèle, mais sa mise en œuvre peut s'avérer impossible car il faut disposer d'une valeur de référence calculée dans des conditions « compatibles » avec le modèle que l'on cherche à estimer ici. Dans le cas de l'élasticité au coût VP, les corrélations avec le temps risquent de poser problème. *A contrario*, l'introduction d'une élasticité au coût du billet ou de l'abonnement TC devrait être envisageable par ce moyen.

6.3.3 Nouvelles pratiques / nouveaux services de mobilité

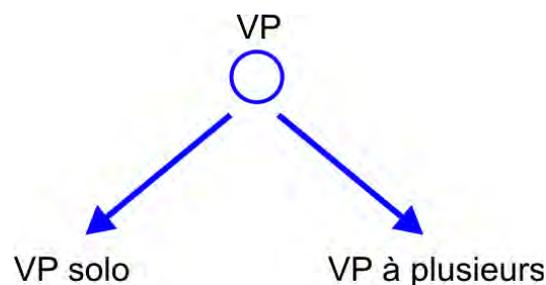
Si les pratiques de modélisation sont fortement centrées sur les modes VP (éventuellement VP passager / VP conducteur), TC et marche à pied, cinq autres comportements de mobilité émergents ont été identifiés : le covoiturage, l'intermodalité VP/TC, l'usage du vélo (incluant les vélos en libre-service), le transport à la demande et l'autopartage.

Ces modes ne sont aujourd'hui pas ou peu intégrés dans les modèles. Il est donc trop tôt pour conseiller des pratiques en se basant sur des retours d'expérience. Par contre, différentes pistes existent pour envisager leur prise en compte, qui sont exposées ici.

6.3.3.1 La montée en puissance du covoiturage

L'idée du covoiturage séduit. Selon une enquête conduite en Essonne, [PREDIT, 2010], 40 % des personnes envisagent d'y recourir et 90 % considèrent qu'il s'agit d'une solution écologique à développer dans leur département. Il existe principalement deux types de politiques publiques liées au covoiturage qu'il est intéressant d'évaluer à l'aide d'un modèle multimodal de déplacements : la création de parkings de covoiturage et la réservation de certaines voies pour les véhicules à plusieurs occupants. Ainsi, en restreignant la portée du terme covoiturage à la population la plus sensible à de telles politiques (exclusion du covoiturage de membres de la famille et des accompagnements), plus de 7 %²¹ de la mobilité quotidienne est effectuée par cette alternative d'après l'Enquête Nationale Transport et Déplacements de 2008 (ENTD, 2008).

Mettre au point un outil de calcul du choix modal tenant compte du covoiturage reste toutefois un défi important : l'état des pratiques met en évidence l'intérêt potentiel d'une décomposition VP solo / VP à plusieurs plutôt que de celle, VP conducteur / VP passager, usuellement retenue en France²². Dans l'optique de l'évaluation de parkings de covoiturage, une différenciation du mode VP à plusieurs selon l'appartenance ou non au même ménage sera nécessaire. La structure du modèle pourra, en fonction des résultats des tests IIA (Cf. § 5.2), soit intégrer ces deux modes de façon indépendante, soit les rassembler dans un nid « VP » suivant le schéma ci-dessous :



Cette décomposition VP solo / VP à plusieurs, couplée avec un bouclage complet du modèle (entre choix modal et affectation), ouvre la possibilité de calculs de temps de parcours distincts en fonction du nombre de passagers à bord du véhicule, permettant ainsi l'évaluation des voies de covoiturage. Outre ces éventuels gains de temps, la fonction d'utilité associée à ce mode devra intégrer au minimum :

- une quantification de l'économie réalisée (partage des coûts entre les covoitureurs) et du gain de confort (devenir passager au lieu de conducteur),
- une estimation des pertes de temps impliquées par ce mode (détour pour atteindre l'autre covoitureur, temps d'attente du covoitureur...),
- et enfin, des variables liées à la perte de fiabilité/liberté induite par ce mode par rapport au mode VP solo.

Il manque toutefois des éléments factuels sur la valorisation de ces éléments par les usagers pour mener cette démarche à son terme. Les enquêtes de préférences déclarées, telles que celle conduite en Essonne, [PREDIT, 2010], sont un moyen intéressant de combler ces lacunes.

21 Ce pourcentage inclut l'ensemble des déplacements, y compris ceux de plus de 80 km pour lesquels la part du covoiturage est plus importante.

22 Il reste toutefois des usages très différents des modes VP à plusieurs, selon par exemple qu'il s'agit des usagers d'une même famille ou des covoitureurs dans un sens strict. Néanmoins, l'estimation des valeurs prises par les variables explicatives de différents types d'alternatives VP à plusieurs est très difficile avec les bases de données disponibles (quel allongement du temps de parcours si les passagers ne sont pas des voisins ou de la famille du conducteur ?). Il est apparu au groupe de travail qu'il était impossible d'effectuer une analyse plus poussée entre les modes VP à plusieurs et VP passager en l'état de la connaissance.

En termes de typologie d'usagers, l'ENTD montre que les usagers prêts à covoiturer (hors ménage) sont majoritairement des actifs et/ou des jeunes (20-29 ans) : ce sont des pistes intéressantes qui pourront donc être testées au niveau de la segmentation du modèle.

Enfin, notons que la complexité des chaînes d'activités est souvent identifiée parmi les freins à l'usage du covoiturage. L'amélioration à terme de la modélisation du covoiturage dans le choix modal passera donc probablement par une modélisation de la demande à l'aide des schémas d'activités.

6.3.3.2 L'intérêt croissant pour les parcs relais (*park and ride* = P+R)

L'intégration des déplacements intermodaux VP+TC dans les modèles de choix de mode est un enjeu qui émerge. D'après l'ENTD 2008, les déplacements intermodaux représentent 2,4 % des déplacements courte distance, et les déplacements spécifiques VP conducteur / TC urbains ne représentent que 0,2 % des déplacements. Malgré cette faible proportion, l'étude de ces déplacements est intéressante car elle permet l'évaluation des P+R (dimensionnement, choix d'implantation) qui sont bien identifiés comme des infrastructures de transport à part entière et utiles au report modal (Cf. [SAUER, 2010]).

Aujourd'hui, sauf dans de rares exceptions, ces déplacements ne sont pas réellement intégrés aux modèles de choix modal. Les tentatives de prise en compte les plus fréquentes en France s'opèrent plutôt au niveau de l'affectation (VP ou TC selon les cas), avec l'utilisation de connecteurs qui permettent de reproduire les coûts généralisés du déplacement restant sur le mode complémentaire (respectivement TC, et VP), et paramétrés empiriquement de façon à reproduire la fréquentation observée des P+R. Cette méthode s'avère excessivement limitée (et risquée) pour l'étude de nouveaux P+R, car il n'est pas du tout certain que le paramétrage des connecteurs soit adapté pour l'ensemble de l'aire d'étude. De plus, la sensibilité à ce paramètre n'étant pas maîtrisée, les résultats pourront s'avérer incohérents.

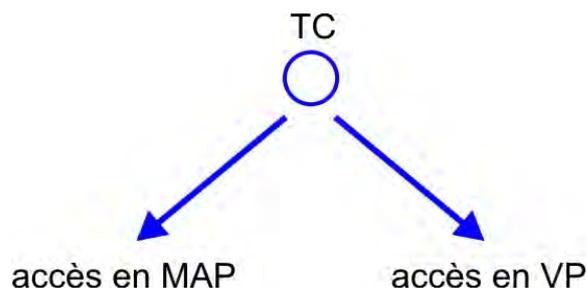
En revanche, un traitement spécifique de ce mode VP+TC dans le modèle de choix modal, comme un mode à part entière, est à privilégier.

Pour ce faire, la littérature montre l'intérêt de décomposer la fonction d'utilité du mode VP+TC en trois composantes :

- l'utilité du mode VP (Domicile → P+R), qui intégrerait les éléments classiques de l'utilité VP ;
- l'utilité spécifique du P+R, composée du coût du stationnement au P+R, temps de correspondance VP → TC, et dans l'idéal, d'éléments relatifs à la « visibilité », la facilité d'accès, et l'information des voyageurs ;
- l'utilité du mode TC (P+R → Destination), composée des composantes classiques de l'utilité TC.

Le calcul de l'utilité globale du mode VP+TC se ferait alors par la somme de ces trois composantes.

La structure du modèle pourra, en fonction des résultats des tests IIA (cf § 5.2), soit intégrer le mode VP+TC comme un mode indépendant des autres, soit le considérer comme une variante du mode MAP+TC en adoptant la structure emboîtée suivante²³ :



23 L'indépendance du mode VP+TC et du mode VP devra également être examinée.

En termes de segmentation de la demande, il pourra être utile d'effectuer des tests à partir des résultats des enquêtes disponibles (EMD et ENTD) qui montrent que :

- les P+R intéressent surtout des déplacements entre la périphérie et le centre/hypercentre. Ce sont donc souvent des déplacements plus longs (20 %) que la moyenne ;
- l'une des extrémités du déplacement est le plus souvent le domicile et l'autre extrémité le travail (mais ce deuxième constat est plus contrasté selon les agglomérations et le jour de la semaine) ;
- les usagers des P+R sont le plus souvent seuls dans leur véhicule (80 à 90 % des déplacements) et sont majoritairement des femmes (60 à 70 % des usagers).

Les enquêtes standard risquent néanmoins de ne pas fournir des échantillons suffisants, et il faudra donc éventuellement sur-échantillonner ce mode. Ceci impose le recours à une technique particulière pour l'intégration de ces données (Cf. § 3.3.1). Enfin, dans le cas des villes moyennes où il n'y a pas forcément encore de P+R dans la zone d'étude, il pourra être intéressant de recourir à des enquêtes de préférences déclarées pour quantifier l'usage de ces infrastructures.

6.3.3.3 La difficile prise en compte du vélo

Le terme « vélo » renferme des réalités très différentes : si le vélo est la propriété de l'utilisateur, ou s'il s'agit d'une location de longue durée, on est plutôt dans le mode vélo « classique », sans contrainte sur le déplacement. En revanche, s'il s'agit de vélos en libre-service (locations de courte durée), il faut y ajouter des contraintes en termes d'OD et de stationnement. Les comportements de ces deux sous-modes vélo en termes d'usage et de choix modal sont donc différents, et un traitement à part semble indispensable.

D'après les EMD, la part modale du vélo dans les grandes agglomérations françaises est proche de 3 % en moyenne. Elle est néanmoins assez variable selon les agglomérations, avec par exemple 0,5 % à Marseille et 7,6 % à Strasbourg.

Aujourd'hui, aucune expérience pratique d'intégration des vélos en libre-service dans un modèle de choix modal n'existe en France. En revanche, certains modèles intègrent déjà le mode vélo « classique », mais on retrouve systématiquement une description de l'offre très simple (vitesse constante sur un réseau marche à pied ou VP) et une intégration dans le modèle limitée à un paramètre de temps et une constante modale (souvent mal paramétrés car les échantillons des déplacements vélo des EMD sont assez faibles) : en conséquence, la capacité de ces modèles à évaluer une politique de transports donnée en faveur du mode vélo est quasi nulle. La prise en compte de ce mode n'est donc là que pour traiter de l'exhaustivité des déplacements.

L'intérêt d'intégrer ce mode dans un modèle est de pouvoir réellement tester une politique telle que la création d'itinéraires cyclables ou la modération généralisée de la vitesse en ville. Cela exige un premier traitement du point de vue de l'affectation : pour ce faire, des démarches cherchant à prendre en compte les spécificités de l'affectation vélo à l'aide du concept de « cyclabilité » émergent (Cf. [SAUVANET, 2011] ou [PREDIT, 2011]).

L'utilité (et le choix d'itinéraire) du mode vélo pourra ainsi être définie à partir des éléments suivants :

- la décomposition de la distance de déplacement différenciée par niveau de cyclabilité (dépendant de la vitesse et du niveau de trafic automobile, ainsi que de la présence ou non d'itinéraires cyclables – éventuellement différenciés eux-mêmes selon leur qualité) ;
- le relief ;
- les pénalités aux carrefours converties en distance de la même façon que pour la marche à pied (Cf. § 4.4).

Là encore, la demande pourra être segmentée à partir des enseignements des enquêtes disponibles sur ce mode (principalement les EMD). Celles-ci nous apprennent que :

- En termes de sur-représentation par rapport au poids dans la population : le cycliste est plutôt un homme, élève, étudiant ou exerçant une profession libérale ou de cadre (avec le niveau d'études correspondant). A l'inverse, les femmes, les personnes de 65 ans ou plus et les employés sont sous-représentés.
- Les plus fortes parts modales vélo enregistrées sont liées aux déplacements ayant pour origine le domicile et pour destination l'université, la promenade ou le travail habituel. Inversement, les plus faibles parts modales sont enregistrées pour les motifs accompagnement, achats (grandes surfaces), et affaires professionnelles.
- L'usage du vélo est plus développé pour les résidents des centres des agglomérations, et il est plus fréquent dans le cas de déplacements internes au centre-ville des agglomérations (3 % de part modale) et beaucoup plus faible sur les trajets liant le centre et la périphérie (1 %). Les déplacements à vélo sont donc de courte distance (85 % se font à moins de 4 km).

En tout état de cause, l'intégration du mode vélo au modèle de choix modal passe par un sur-échantillonnage au moment de l'EMD ou une enquête spécifique (ex : enquête de préférences déclarées). Là encore, la combinaison de ces données hétérogènes doit faire l'objet de précautions particulières (Cf. § 3.3).

6.3.3.4 Transport à la demande et autopartage, des modes marginaux

Au vu de leur part modale de l'ordre de 0,02 % (ENTD 2008), le nombre de déplacements semble aujourd'hui insuffisant pour que la prise en compte de services de transport à la demande (TAD) dans les modèles de choix modal soit possible. L'utilisation discrétionnaire de ce mode ne peut se modéliser à l'aide de comportements moyens et les résultats produits resteront dans la marge d'erreur du modèle. En outre, la clientèle cible du TAD correspond à une population captive, qui ne peut se déplacer de manière autonome, et qui n'a donc pas accès à un véhicule personnel.

Le cas de l'autopartage est d'une nature différente : il intervient principalement en amont du choix de mode, sur la décision de possession d'une voiture. Son intégration cohérente dans un modèle multimodal est donc reliée à l'existence d'un modèle d'équipement automobile du ménage (Cf. § 6.1 et 6.2). Comme pour les autres modes émergents, cette pratique est encore trop peu connue pour être mesurée correctement par les enquêtes standard et l'estimation des fonctions d'utilité ne pourra être réalisée qu'à l'aide d'un sur-échantillonnage (enquête auprès des usagers de services d'autopartage relativement simple à conduire) ou d'une enquête de préférences déclarées.

6.4 Valider le modèle

La vérification de la cohérence du modèle d'un point de vue statistique à l'aide des indicateurs décrits au § 5. ne constitue pas à elle seule une validation du modèle à proprement parler. Idéalement, il serait souhaitable de vérifier que le modèle reproduit de façon statistiquement satisfaisante les comportements individuels observés sur une autre base de données que celle qui a servi au calage. Cette méthode étant la plupart du temps impossible à mettre en œuvre, une alternative consiste à vérifier que le modèle reproduit de façon fiable un certain nombre de comportements observés au niveau agrégé et réagit correctement à certains stimuli (voir la partie sur l'analyse de sensibilité, § 7.2).

6.4.1 Au niveau désagrégé

La démarche « idéale » pour valider le modèle au niveau désagrégé consiste à n'estimer le modèle que sur une partie de l'échantillon total, afin d'en réserver une seconde partie pour la validation : la validation consiste alors simplement à simuler le modèle obtenu sur ce

second échantillon et à en observer les résultats à l'aide des indicateurs généraux classiques qui permettent de juger de la qualité générale du modèle (\bar{p}^2). Le sous-échantillon réservé à la validation est dans l'idéal compris entre 15 et 25 % de l'échantillon total.

De plus, il est important de prévoir une analyse spécifique des individus pour lesquels le modèle reproduit mal le comportement de choix de mode, afin de vérifier qu'il n'y a pas certains comportements systématiques que le modèle serait incapable de reproduire.

Enfin, l'estimation du modèle sur les données de préférences révélées de deux horizons (par ex. l'EMD actuelle et l'EMD précédente) apporte un éclairage sur l'hypothèse selon laquelle les fonctions d'utilité sont stables dans le temps. Cela nécessite néanmoins de reconstruire l'offre d'une situation passée relativement éloignée (une dizaine d'années), ce qui a des conséquences en termes de coût.

6.4.2 Au niveau agrégé

En premier lieu, il est nécessaire de vérifier que les parts modales par segment sont bien conformes à l'enquête. Cela sera en principe toujours le cas, mais cette vérification simple peut permettre d'éviter les erreurs grossières.

Ensuite, décomposer la contribution à la log-vraisemblance selon le mode choisi est utile car cela permet d'examiner si la qualité de la reproduction des comportements est similaire selon les alternatives.

Enfin (et surtout), tous les autres tests de comparaison des parts modales entre le modèle d'une part et les observations d'autre part (en précisant la fourchette de valeurs des observations), sont importants :

- Parts modales par période horaire : dans le cas où le modèle de choix modal ne différencie pas les périodes horaires (dans ce cas, celles-ci sont donc différenciées simplement lors de la formulation des différents coûts généralisés), une vérification importante consiste à s'assurer que le modèle reconstitue bien les parts modales par période horaire. Si ce n'est pas le cas, il faudra envisager, soit de retravailler les coûts généralisés si l'erreur provient d'une mauvaise caractérisation de ceux-ci, soit de segmenter le modèle de choix de mode par période horaire (cela revient à considérer que les comportements des individus face à un même choix d'alternatives diffèrent selon les périodes horaires).
- Parts modales par classe de distance,
- parts modales pour l'ensemble des déplacements émis et attirés par zone,
- parts modales pour les OD significatives et/ou sur un macrozonage,
- parts modales par segments de population non utilisés pour l'estimation du modèle, en particulier :
 - parts modales par typologie de zone (ex. : périurbain vers centre-ville),
 - parts modales par CSP.

Au fur et à mesure de l'amélioration du modèle, il sera utile de croiser ces différents éléments, par exemple la typologie des zones et la classe de distance, afin de répondre aux deux objectifs de bien connaître les limites du modèle et de déterminer les pistes d'amélioration ultérieures.

7. Faire des simulations avec un modèle de choix modal

7.1 Méthodes d'application du modèle calibré

7.1.1 Présentation des deux méthodes disponibles

L'estimation du modèle, décrite dans les chapitres 2 à 6, permet d'évaluer la sensibilité du choix modal aux différents facteurs causaux. Son application pour tester différentes politiques publiques ne saurait être immédiate. Elle nécessite d'une part l'identification de l'ensemble de la population potentiellement sensible à ces politiques et d'autre part la connaissance pour chaque individu des valeurs prises par les variables socioéconomiques entrant dans la fonction d'utilité.

Certes, l'accès aux données individuelles du recensement de la population est à ce titre un apport considérable. Cependant, il apparaît très vite qu'il ne saurait fournir l'intégralité des informations nécessaires à l'application du modèle de choix modal : il ne contient pas d'éléments sur les déplacements *stricto sensu*, et ne permet donc pas d'évaluer les niveaux de service. En outre, les destinations des déplacements peuvent évoluer avec les politiques à tester et il convient d'en tenir compte.

Ainsi, l'application du modèle requiert l'usage de méthodes simplificatrices. Deux possibilités sont disponibles : l'application sur données agrégées ou l'utilisation d'un échantillon de la population (incluant la génération de « populations synthétiques »).

Indépendamment de la méthode retenue pour l'application du modèle, la définition des politiques et/ou des évolutions prévisibles du territoire à tester interfèrent avec la phase d'estimation du modèle : il n'y a aucune garantie que les variables utiles pour évaluer ces effets aient été introduites au cours d'une procédure d'estimation si celle-ci a été menée dans le seul but d'identifier les facteurs déterminants du choix modal. Il y a donc un fort intérêt à anticiper les politiques à tester dès la conception du modèle, en vue de l'adapter spécifiquement pour la simulation de celles-ci. Cela peut concerner :

- Une nouvelle variable, par exemple un élément du coût généralisé comme le coût du billet TC. L'analyse statistique peut tout à fait conduire à des résultats contre-intuitifs (mauvais signe de la variable) ou à l'absence de significativité (Cf. § 6.3.2). Dans ce cas, il existe des techniques pour forcer l'intégration de la variable (Cf. § 7.3).
- Une nouvelle segmentation du modèle. Dans ce cas, la méthode statistique de test de différence de perception (Cf. § 5.5) fournit l'intégralité des informations nécessaires.

L'**utilisation d'un échantillon de population** est très intuitive. Cette méthode suppose de disposer d'une estimation de toutes les valeurs individuelles des variables entrant dans les fonctions d'utilité pour un échantillon représentatif de la population sensible aux politiques à tester. Les probabilités d'usage de chaque mode peuvent être calculées au niveau individuel puis agrégées par une simple somme.

Parmi les différentes possibilités de mise en œuvre (Cf. par ex. [ORTUZAR et WILLUMSEN, 2011]), les deux techniques situées aux extrêmes du champ des possibles permettent d'exposer les limitations inhérentes à la méthode.

Une méthode « naïve » consiste à utiliser directement l'échantillon de l'enquête de préférences révélées. Dans ce cas, le taux de sondage conditionne les analyses possibles des résultats du modèle. Cette technique est très utile pour évaluer des politiques globales, telles que le renchérissement d'un mode de transport, ou une augmentation généralisée de la fréquence. Ainsi, il s'agira de la méthode à mettre en œuvre pour étudier le comportement du modèle. *A contrario*, une agrégation des résultats au niveau des OD ne pourra se faire que sur de grandes typologies de zones (par ex. urbain/périurbain/rural) ; dans tous les cas, les

OD obtenues ne sauraient être suffisamment détaillées pour effectuer des affectations. En outre, comme les valeurs des variables sont obtenues par l'enquête, tester l'effet d'une politique ciblée d'amélioration des niveaux de service (par ex. une ligne de métro) donnera par nécessité des résultats imprécis et incohérents.

Pour remédier à ces difficultés, il est possible de générer une population synthétique, dont les choix seront nécessairement examinés à chaque étape du modèle de déplacement, au moins jusqu'à l'étape précédant l'affectation. Cette méthode n'est pas encore entrée dans les pratiques : elle est considérablement plus lourde à mettre en œuvre que l'application du modèle sur données agrégées. Par ailleurs, étant fondée sur des principes de micro-simulation, elle comporte les faiblesses usuelles liées à ces méthodes, notamment lorsqu'il s'agit de techniques Monte Carlo (Cf. [ORTUZAR et WILLUMSEN, 2011]), à savoir le risque d'une forte variation des résultats en fonction des différentes répliques du modèle. En définitive, la pertinence de cette méthode dépend de deux facteurs : le nombre de répliques réalisées d'une part et le niveau d'agrégation utilisé pour l'analyse d'autre part (forte variation possible des résultats sur les analyses de détail). *In fine*, la génération d'une population synthétique ne permet qu'un dépassement partiel des défauts de la méthode « naïve ». Néanmoins, l'immense gain en flexibilité du modèle, permettant des analyses hors de portée des modèles agrégés, devrait conduire à l'adoption progressive de la génération de populations synthétiques.

L'application du modèle sur données agrégées consiste :

- à identifier des groupes d'individus de valeurs d'utilité homogènes ;
- à déterminer la valeur moyenne prise par chaque variable entrant dans les fonctions d'utilité pour chacun de ces groupes ;
- à obtenir une estimation des parts modales agrégées pour chaque groupe par l'application du modèle estimé sur les valeurs moyennes des variables.

Cette méthode comporte un problème important : comme la fonction de choix modal n'est pas linéaire, il en résulte que le choix moyen de plusieurs individus est différent du choix de l'individu moyen (Cf. [ORTUZAR et WILLUMSEN, 2011]).

Néanmoins, cette erreur peut être minimisée selon [BEN-AKIVA et LERMAN, 1985] en utilisant un nombre (même réduit) de groupes d'individus. Comme il est en pratique impossible de classer la population selon la valeur de l'utilité, afin de minimiser la variance de la composante systématique, il conviendra de sélectionner les groupes sur la base de quelques variables jouant un rôle important dans le choix modal et à veiller à ce que l'agrégation traduise explicitement les différences de disponibilité des alternatives (par ex. entre les ménages ayant une voiture et ceux n'en disposant pas).

Ainsi, l'agrégation est généralement réalisée selon un découpage géographique très fin sur le cœur d'une agglomération et à l'échelle de la commune sur le reste du territoire, puis en fonction des résultats des tests de différences de perception selon les motifs, et selon différentes variables socio-économiques... Au regard des recommandations théoriques, deux critères semblent particulièrement importants pour l'identification des groupes d'individus :

- L'éloignement des arrêts TC : il s'agit d'un critère décisif dans les zones de taille importante et relativement mal desservies. Lorsqu'elle est disponible, cette variable peut en premier lieu être introduite dans les fonctions d'utilité au niveau individuel, ce qui améliorera grandement l'estimation. Ensuite, pour l'application du modèle, il convient de calculer sa moyenne pour chaque groupe d'individus : il apparaît alors qu'un travail préliminaire sur la segmentation selon cette variable, avec pour objectif de définir une éventuelle absence de disponibilité de l'alternative TC à partir d'un certain seuil de temps d'accès, peut être un second facteur d'amélioration du modèle. Cette variable n'est cependant pas recueillie systématiquement dans les enquêtes déplacements : il conviendra donc d'utiliser le niveau de zonage le plus fin possible et d'envisager l'usage de critères de

segmentation alternatifs tels que les habitudes de déplacements (fréquence d'usage des TC) ;

- la disponibilité de la voiture : elle devra systématiquement faire partie des critères de segmentation à tester. Néanmoins, si la quantité de voitures des ménages est un élément recueilli exhaustivement par le recensement de la population, il n'en est pas de même pour le permis de conduire. En outre, la part de chaque segment doit être estimée au niveau des OD et non plus de la zone de résidence, ce qui est particulièrement délicat pour les motifs secondaires. L'estimation de la population de chaque segment requiert donc un examen attentif, qui pourra passer par l'usage de modèles, par exemple de motorisation.

7.1.2 Illustrations

Sur l'exemple du modèle de Rouen, il est possible de mettre en œuvre les deux méthodes d'application exposées ci-dessus afin de montrer leurs potentialités.

Pour illustrer la méthode d'application sur données agrégées, on simule la mise en place d'une nouvelle offre de transports en commun dans l'agglomération rouennaise, selon plusieurs scénarios de desserte. Le modèle de choix modal utilisé ici est le modèle 1 (Cf. § 6.2.3), appliqué sur une matrice tous modes, différenciée par motif et issue d'une procédure gravitaire.

On ne s'intéresse ici qu'aux produits de l'étape de choix modal, qui permet d'estimer directement le nombre de voyageurs supplémentaires dans les TC.

Les scénarios testés sont les suivants : plusieurs combinaisons possibles (Cf. schémas) d'une nouvelle ligne (B), et d'une ancienne ligne (A), sont modélisées dans les scénarios S1, S2, et S3, et une optimisation de la nouvelle ligne en termes de fréquence et de temps de parcours est testée dans le scénario S3+. Les résultats sont les suivants :

Scénario	S0 - Fil de l'eau	S1	S2	S3	S3+
					
Report modal TC (depl)		2 557	2 086	2 156	2 941
Report <VP (depl)		-1 489	-1 166	-1 143	-1 760
% nvx usagers TC venant de la VP		58 %	56 %	53 %	60 %
% nvx usagers TC venant de la MAP		42 %	44 %	47 %	40 %

Le modèle de choix modal permet donc bien ici de discriminer les scénarios en termes de report modal, puisque la nouvelle ligne permet de capter selon les scénarios de 2000 à 3000 voyageurs provenant des autres modes de transport (avec une provenance depuis la VP et la MAP qui diffère selon les scénarios), et un effet des optimisations de fréquence et de vitesse commerciale qui est très net. On comprend facilement ici la potentialité d'un modèle de choix modal appliqué sur données agrégées.

Pour illustrer la méthode de l'application sur un échantillon de la population, une segmentation totale du modèle selon l'âge est mise en œuvre à partir du **modèle 3** (Cf. § 6.2.3). Les résultats obtenus, après regroupement des variables non significativement différentes et suppression des variables non significatives, sont les suivants :

Variable	Valeur	Variable	Valeur
ASC_TC_DE_SUP_12_age12345	-1.65	BETA_ACC_TIME_TC_123_age1235	-0.0434
ASC_TC_DE_SUP_3_age12345	-2.96	BETA_ACC_TIME_TC_123_age4	-0.0705
ASC_TC_DE_PRIMAIRE_123_age12345	-2.47	BETA_IN_TIME_TC_13_age1345	-0.0327
ASC_TC_DE_SECONDAIRE_123_age1245	-1.63	BETA_IN_TIME_TC_13_age2	-0.0143
ASC_TC_DT_DA_SEC_12_age12345	-2.02	BETA_IWT_23_age1	-0.0352
ASC_TC_DT_DA_SEC_3_age123	-2.98	BETA_IWT_23_age2345	-0.0585
ASC_TC_DT_DA_SEC_3_age45	-1.93	BETA_STATIO_FORTE_123_age1	-1.17
ASC_VP_DE_SUP_12_age12345	-3.30	BETA_STATIO_FORTE_123_age234	-1.62
ASC_VP_DE_SUP_3_age1345	-1.26	BETA_STATIO_FORTE_123_age5	-2.19
ASC_VP_DE_SUP_3_age2	-1.89	BETA_STATIO_MOY_123_age1234	-0.506
ASC_VP_DE_PRIMAIRE_123_age145	-1.36	BETA_TIME_MAP_12_age1234	-0.0857
ASC_VP_DE_PRIMAIRE_123_age23	-8.65	BETA_TIME_MAP_12_age5	-0.0697
ASC_VP_DE_SECONDAIRE_1_age1345	-11.3	BETA_TIME_MAP_3_age123	-0.101
ASC_VP_DE_SECONDAIRE_1_age2	-3.34	BETA_TIME_MAP_3_age4	-0.0679
ASC_VP_DE_SECONDAIRE_2_age12345	-3.24	BETA_TIME_MAP_3_age5	-0.0618
ASC_VP_DE_SECONDAIRE_3_age12345	-0.513	BETA_TIME_VP_13_age1	-0.143
ASC_VP_DT_DA_SEC_12_age15	-1.48	BETA_TIME_VP_13_age245	-0.0386
ASC_VP_DT_DA_SEC_12_age234	-2.45	BETA_TIME_VP_13_age3	-0.0723

Où :

- âge 1 = moins de 20 ans
- âge 2 = 20 – 40 ans
- âge 3 = 40 – 60 ans
- âge 4 = 60 – 80 ans
- âge 5 = plus de 80 ans

A partir de ce modèle, on obtient les parts modales par classe d'âge suivantes :

classe d'âge	échantillon	MAP	TC	VP
inf 20 ans	4434	29,1 %	21,5 %	49,5 %
20-40 ans	8095	14,0 %	8,0 %	78,0 %
40-60 ans	8833	11,0 %	4,7 %	84,2 %
60-80 ans	3898	18,6 %	4,2 %	77,2 %
sup 80 ans	338	29,8 %	5,9 %	64,3 %

Par application des hypothèses d'évolution de la population par groupe d'âge à l'horizon 2030 de l'INSEE (modèle Omphale), nous obtenons l'évolution prévisible des parts modales entre 2007 et 2030 :

	MAP	TC	VP
Part modale 2007	16,5 %	8,6 %	74,9 %
Part modale 2030	16,8 %	8,0 %	75,2 %

Le modèle simule donc bien ici les effets du vieillissement de la population grâce à la méthode de l'énumération. Il faut également souligner que le modèle de choix modal n'a pas besoin d'être inséré dans un modèle à 4 étapes complet pour faire ce genre de simulation : il est suffisant à lui seul pour réaliser ce type de projection, sauf variation significative du temps VP issu des itérations choix modal / affectation.

On voit également ici l'intérêt d'anticiper les politiques à tester dès la conception du modèle en vue de l'adapter spécifiquement pour la simulation de celles-ci.

7.2 Analyser les élasticités

L'élasticité permet d'exprimer dans quelle mesure la variation d'un facteur causal et la variation résultante d'une grandeur sont reliées. Ici, l'élasticité concerne le lien entre la probabilité de choisir une alternative et la valeur des variables explicatives, en particulier des variables de description du niveau de service. Dans la théorie des choix discrets, une distinction s'opère entre l'élasticité individuelle et l'élasticité agrégée.

L'**élasticité individuelle** est déterminée pour chaque situation de choix (et donc pour chaque individu) i se présentant. Elle se calcule selon la formule :

$$e_m(X, i) = \frac{\partial P_m(i)}{\partial X} \times \frac{X(i)}{P_m(i)} \quad \text{où } X \text{ est une variable entrant dans la composante}$$

systematique de l'utilité

La valeur de la dérivée dépend fortement de la structure du modèle (logit multinomial ou non) et de la formulation des utilités (utilité linéaire, variables intervenant dans les fonctions d'utilité d'une ou de plusieurs alternatives...). Les formules peuvent se simplifier selon certaines hypothèses, en distinguant deux situations :

- l'élasticité directe concerne la variation de l'usage d'un mode en réponse à une variation d'un facteur entrant dans sa fonction d'utilité ;
- l'élasticité croisée concerne la variation de l'usage d'un mode en réponse à la variation d'un facteur entrant dans la fonction d'utilité d'un autre mode.

Les formules suivantes concernent deux cas particulièrement utiles :

- le modèle logit multinomial avec des utilités potentiellement non linéaires pour les variables n'intervenant que dans l'utilité d'un seul mode, car il s'agit de l'usage le plus courant de ce type de modèle ;
- le modèle cross nested logit, avec les mêmes restrictions sur la formulation de l'utilité, car il s'agit d'un cas relativement général.

Elasticité directe pour un modèle multinomial, lorsque la variable entre dans l'utilité d'un seul mode :

$$e_m(X_m, i) = \frac{\partial U_m(i)}{\partial X_m} \times (1 - P_m(i)) \times X_m(i)$$

Elasticité croisée pour un modèle multinomial, lorsque la variable entre dans l'utilité d'un seul mode :

$$e_m(X_m', i) = -P_m(i) \times \frac{\partial U_m(i)}{\partial X_m'} \times X_m'(i)$$

Elasticité directe pour un modèle cross nested logit, lorsque la variable entre dans l'utilité d'un seul mode :

$$e_m(X_m, i) = \frac{\sum_{M'} P_{M'(i)} \times P_{m|M'(i)} \times \left[(1 - P_m(i)) + \left(\frac{1}{\mu_{M'}} - 1 \right) \times (1 - P_{m|M'(i)}) \right]}{P_m(i)} \times \frac{\partial U_m(i)}{\partial X_m} \times X_m(i)$$

où :

- M' : identifiant des nids
- $\mu_{M'}$: facteur d'échelle du nid M'
- $P_{M'}$: probabilité de choix du nid M'
- $P_{m|M'}$: probabilité de choix du mode m au sein du nid M'

Elasticité croisée pour un modèle cross nested logit, lorsque la variable entre dans l'utilité d'un seul mode :

$$e_m(X_m', i) = \left[-P_m'(i) + \frac{\sum_{M'} P_{M'(i)} \times P_{m'|M'(i)} \times P_{m|M'(i)} \times \left(\frac{1}{\mu_{M'}} - 1 \right)}{P_m(i)} \right] \times \frac{\partial U_m(i)}{\partial X_m'} \times X_m'(i)$$

L'**élasticité agrégée** correspond à la réponse d'un groupe à la variation du facteur causal. Elle est déterminée à partir de l'élasticité individuelle de chacune des situations de choix composant le groupe, à l'aide de la formule :

$$\overline{e_m(X, i)} = \frac{\sum e_m(X, i) \times P_m(i)}{\sum P_m(i)}$$

Dans une certaine mesure, le calcul des élasticités permet, comme cela a été présenté au § 6, de comparer différents modèles et ainsi d'aider à la recherche de l'outil le plus pertinent. Ce calcul peut également constituer un outil efficace, parmi d'autres, d'aide à la décision en ce qui concerne les politiques publiques de rééquilibrage des parts modales. Il fournit ainsi une première appréciation des leviers d'action disponibles.

La démarche du § 6 de démonstration sur l'exemple de l'aire urbaine de Rouen permet d'illustrer ce propos, sans toutefois prétendre à constituer une réelle identification des politiques publiques les plus pertinentes sur ce secteur, qui est un processus évidemment hors de portée de ce guide.

Les valeurs d'élasticité aux variables de niveau de service obtenues pour le modèle 3 sont rappelées sur le tableau suivant :

Paramètre	Élasticité directe	Élasticité croisée		
		VP	MAP	TC
Temps de parcours Marche à Pied	-1,042	0,731		0,282
Temps de parcours Voiture	-0,056		0,029	0,028
Contrainte de stationnement forte	-0,051		0,030	0,021
Contrainte de stationnement moyenne	-0,009		0,006	0,003
Temps d'accès aux arrêts TC	-0,317	0,086	0,035	
Temps de parcours de l'arrêt de départ à l'arrêt d'arrivée	-0,144	0,125	0,017	
Temps d'attente du premier maillon TC	-0,125	0,190	0,117	

Ces résultats sont à interpréter au vu des parts modales actuelles, soit 16,5 % pour la MAP, 74,9 % pour la VP et 8,6 % pour les TC. Ainsi, une diminution de 10 % du temps de parcours en marche à pied conduirait à une augmentation de 10,42 % de la part de la MAP, soit +1,8 points de part modale MAP. Une augmentation de 10 % des temps de parcours VP conduit quant à elle à une diminution de la part modale VP de 0,4 point. La différence d'effet sur les parts modales est donc nettement moins importante que les valeurs respectives d'élasticité de -1,042 et -0,056 ne le laisseraient supposer.

La définition précédente correspond à la notion d'« **élasticité point** » communément admise en économie. Toutefois, lorsque la valeur de la part modale ou de la variable tend vers 0, cette définition peut engendrer une valeur de l'élasticité qui peut s'avérer parfois trompeuse par rapport aux véritables effets attendus de la variable. Dans la mesure où ce sont les effets de telles ou telles politiques de transports (relativement marquées) sur les parts modales qui sont réellement intéressants pour la compréhension du modèle, on utilise souvent une autre définition de l'élasticité, appelée « **élasticité arc** », qui consiste à calculer l'élasticité non plus en un point unique, mais sur tout ou partie de la courbe. Cette deuxième définition est généralement plus utile pour comprendre les effets de chacune des variables du modèle. Cette élasticité est aussi plus facile à calculer, puisqu'elle ne nécessite pas le calcul des élasticités individuelles : il suffit de simuler le modèle en appliquant une variation (par exemple de +10 %) sur une des variables, et d'observer la variation des parts modales globales prévues par le modèle.

Il sera donc d'usage de présenter les variations de parts modales selon différents scénarios, qui comprendront :

- une augmentation/diminution de 10 % de la valeur de chaque variable continue,
- un croisement de l'augmentation/diminution de 10 % de la valeur de toutes les variables continues d'un même mode,
- un croisement des améliorations/dégradations des niveaux de service des différents modes,
- des déplacements de 10 % de la population de chaque segment vers les autres segments.

Il est à noter que l'étape de croisement est parfois superflue (elle ne sera donc réalisée qu'en cas de besoin bien précis).

Cependant, ces résultats n'exploitent pas complètement la richesse du modèle, étant donné qu'ils ne mettent pas en évidence les transferts modaux maximaux qui peuvent être tirés de différentes politiques. Il pourra donc être utile de définir, avec le maître d'ouvrage, des scénarios plus marqués, à l'image de ceux présentés au § 6.2.3 (exemple des scénarios VP-- et TC++).

7.3 Intégrer des variables supplémentaires

Les modèles de choix modal développés selon cette méthodologie ne sont pas adaptés pour représenter des ruptures comportementales ou des changements radicaux hors des plages de valeurs prises par les variables. Deux méthodes sont proposées pour remédier à cet état de fait :

1. si l'élasticité de la demande à la variable est largement étudiée par ailleurs, et si cette variable n'est pas correctement estimée par le modèle ou qu'elle a été rejetée car « déformant » l'estimation du modèle, il est possible de la réintroduire en estimant la valeur du paramètre bêta à l'aide de cette élasticité, en utilisant la formule :

$$\beta_k = \frac{E_{Xmk}^{Part(m)} \times \sum_i Pi(m)}{\sum_i Pi(m) \times (1 - Pi(m)) \times Ximk}$$

où :

- **i** : individu
- **Pi(m)** : probabilité que **i** choisisse l'alternative **m**
- **E** : élasticité agrégée

La nouvelle variable associée à β peut ainsi être ajoutée à la fonction d'utilité, et le modèle devra ainsi être à nouveau calibré, en prenant soin de fixer ce paramètre β . Cela engendrera nécessairement une modification de la valeur des autres paramètres, qu'il faudra surveiller avec attention. La mise en œuvre de cette méthode peut donc s'avérer délicate car il faut disposer d'une valeur de référence calculée dans des conditions (typologie de territoire, composantes de l'utilité, méthode d'estimation des coûts généralisés, etc.) exactement « compatibles » avec le modèle que l'on cherche à estimer ici.

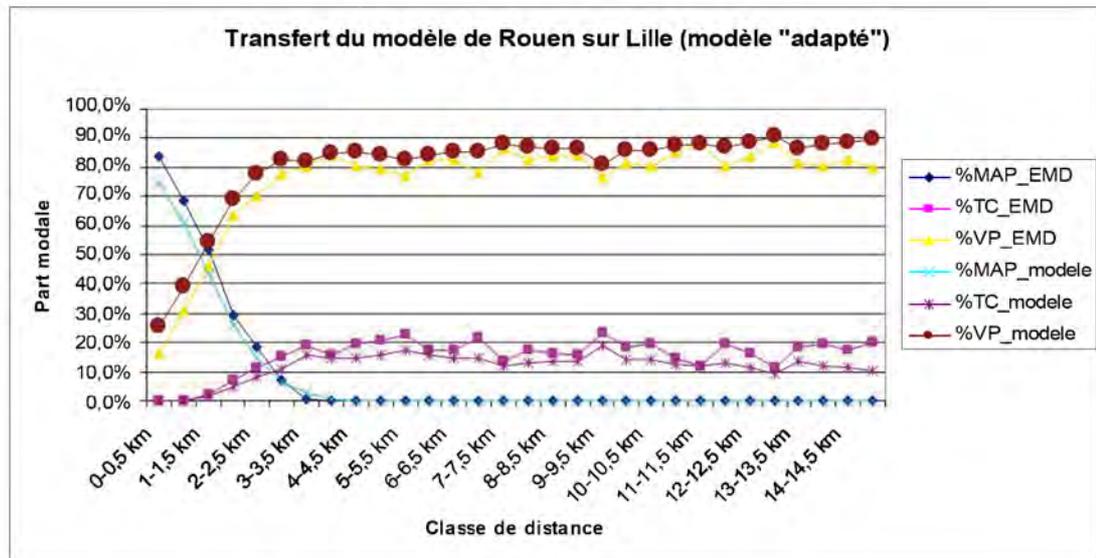
Des tests de sensibilité peuvent alors être effectués en cohérence avec les études d'élasticité.

2. Dans le cas contraire, il faut procéder à une enquête de préférences déclarées. Cela peut être utile notamment lorsque le temps de parcours d'un mode est amené à évoluer fortement et qu'on souhaite savoir s'il vaut mieux « créer » un nouveau mode ou tenir compte de l'amélioration du temps de parcours. Dans ce cas, il faut procéder à une estimation conjointe d'un modèle fondé sur l'enquête de préférences déclarées et l'enquête de préférences révélées initiale. (cf § 3.3.2)

7.4 Réutiliser un autre modèle

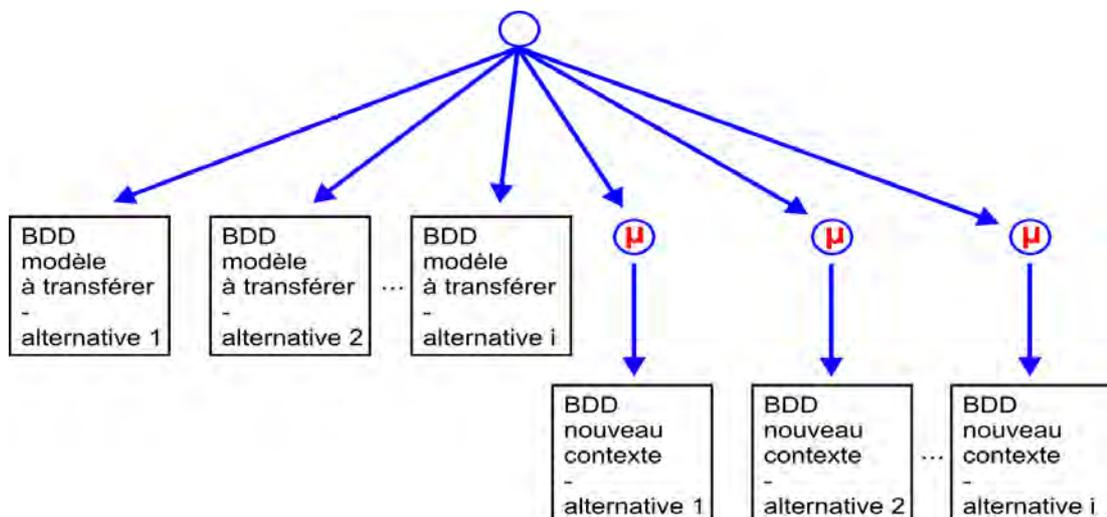
Au vu des difficultés pratiques de développement d'un modèle de choix modal sophistiqué (coût élevé, délais importants, besoin probable d'enquêtes spécifiques au-delà de l'EMD classique, recours nécessaire à du personnel très qualifié pour la construction et la réception du modèle), l'utilisation de modèles existants transférés sur d'autres contextes présente des avantages économiques certains.

Le graphique suivant montre la reproduction des parts modales en fonction de la distance dans le cas du transfert « naïf » (réutilisation du modèle à l'identique) du modèle de Rouen (modèle 1 – Cf. § 6.2.3) sur l'agglomération de Lille (dont la base de données de l'EMD a été renseignée par les caractéristiques des alternatives MAP, VP et TC à l'instar de ce qui a été fait sur l'agglomération rouennaise) :



Ce graphique tend à montrer que des résultats intéressants peuvent être obtenus par cette approche. Des améliorations seraient ici cependant souhaitables, en particulier pour les très courtes distances (erreur d'environ 10 points de part modale pour la MAP et la VP), et pour le mode TC (légère sous-estimation par le modèle).

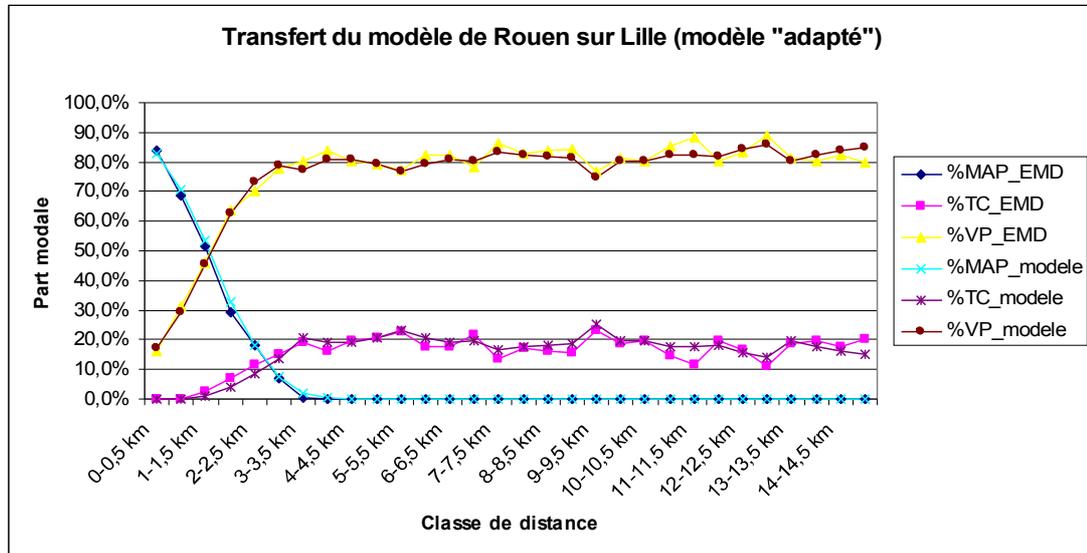
De manière générale (Cf. [NCHRP, 2012]), peu d'évidences empiriques soutiennent le postulat d'égalité des paramètres des modèles de choix modal entre différentes agglomérations. *A contrario*, il semble bien établi qu'il est possible d'améliorer très significativement le transfert du modèle en utilisant une technique de mise à jour d'un modèle existant à l'aide de données de préférences révélées correspondant au nouveau contexte, en se basant sur la méthode de combinaison d'enquêtes présentée au § 3.3, et en l'adaptant de la façon suivante :



Pour illustrer ce processus, le modèle de Rouen va ici être transféré sur le contexte lillois, en utilisant une toute petite partie de la base de données de l'EMD de Lille (tirage aléatoire de 1000 déplacements seulement). L'ensemble des variables du modèle sont communes aux deux enquêtes (excepté le facteur d'échelle du nid TC+MAP, et les constantes d'alternatives). Les résultats sont les suivants :

Paramètre	Valeur	t-tests
ASC_MAP	0.00	fixed
ASC_TC_DE_SUP_Rouen	-1.70	-14.62
ASC_TC_DE_SUP_Lille	-1.72	-3.30
ASC_TC_DE_PRIMAIRE_Rouen	-2.09	-14.06
ASC_TC_DE_PRIMAIRE_Lille	-2.09	-0.00
ASC_TC_DE_SECONDAIRE_Rouen	-1.37	-18.65
ASC_TC_DE_SECONDAIRE_Lille	-1.84	-7.28
ASC_TC_DT_DA_SEC_Rouen	-2.03	-33.72
ASC_TC_DT_DA_SEC_Lille	-1.85	-13.19
ASC_VP_DE_SUP_Rouen	-2.42	-16.85
ASC_VP_DE_SUP_Lille	-2.93	-4.17
ASC_VP_DE_PRIMAIRE_Rouen	-1.33	-12.27
ASC_VP_DE_PRIMAIRE_Lille	-1.51	-3.92
ASC_VP_DE_SECONDAIRE_Rouen	-3.26	-35.03
ASC_VP_DE_SECONDAIRE_Lille	-2.64	-8.91
ASC_VP_DT_DA_SEC_Rouen	-0.699	-16.19
ASC_VP_DT_DA_SEC_Lille	-0.990	-10.41
BETA_ACC_TIME_TC	-0.0481	-14.66
BETA_IN_TIME_TC	-0.0212	-12.07
BETA_IWT	-0.0327	-7.80
BETA_STATIO_FORTE	-1.39	-32.12
BETA_STATIO_MOY	-0.495	-9.73
BETA_TIME_MAP	-0.0831	-51.14
BETA_TIME_VP	-0.0174	-3.84
MU_Lille	1.23	13.37
MU_ALTERVP_Lille	2.04	3.29
MU_ALTERVP_Rouen	1.67	fixed

Si l'on simule maintenant sur la base complète de l'EMD de Lille le nouveau modèle obtenu, les résultats sont les suivants :



Les résultats sont nettement améliorés, notamment sur les courtes distances. Des erreurs subsistent néanmoins, en particulier pour les TC, mais les résultats sont globalement satisfaisants. Une telle technique pourrait donc être utilisée en cas d'absence de gros enjeu spécifique lié au modèle de choix modal (par exemple si le modèle ne sert qu'à fournir des matrices initiales pour l'affectation, et que le report modal n'est pas un des éléments discriminants dans l'analyse de scénarios).

Ces techniques de transfert des modèles présentent des avantages très importants dans le contexte français où de très nombreuses enquêtes sur les comportements de mobilité sont organisées dans différentes agglomérations selon une méthodologie identique : leur compilation fournirait des quantités d'informations considérables sur les pratiques modales trop faiblement échantillonnées sur une seule enquête (par ex. les trains régionaux, l'intermodalité VP/TC, ou le vélo), mais aussi augmenterait potentiellement les plages de valeurs prises par certaines variables (prix du carburant lors de la période d'enquête, prix des billets TC différents) ; elle permettrait également de mettre en évidence l'existence réelle de disparités régionales de comportements. La difficulté de cette approche consiste néanmoins à collecter les impédances des différentes alternatives à la date de l'enquête et selon des méthodes identiques.

8. Perspectives

Au vu de la complexité des comportements et des mathématiques sous-jacentes, cet ouvrage ne saurait prétendre à l'exhaustivité. Il cherche plutôt à donner les principales clés pour construire des modèles de choix modal robustes. Il s'est appuyé sur les pratiques existantes pour identifier les méthodes couramment utilisées les plus pertinentes, ainsi que pour définir les améliorations faisables. Par ailleurs, ce guide n'est pas complètement indépendant de l'ensemble de la chaîne de modélisation mise en œuvre : ainsi, l'usage croissant de la modélisation selon les boucles ou les schémas d'activité serait de nature à imposer d'étoffer la rédaction sur les spécifications possibles du modèle et les méthodes d'estimation.

Dans les prochaines années, l'évolution des pratiques devrait porter :

- sur la chaîne globale de la modélisation des déplacements : faut-il estimer la génération au niveau des ménages (ou des individus) ? Doit-on introduire une prise en compte des boucles de déplacements ? Y-aura-t-il des avancées dans l'usage des modèles de l'interaction transport-urbanisme ? La complexité croissante de l'usage des modèles conduira-t-elle au développement de l'utilisation de modèles selon les schémas d'activité ?
- sur les données disponibles, permettant notamment une meilleure connaissance des niveaux de service : localisation en longitude/latitude dans les EMD, prise en compte des big data, traces GPS...
- sur les enquêtes de préférences déclarées, voire les enquêtes de panels, restant encore trop rarement utilisées en France. Celles-ci imposent à la fois dans la construction des enquêtes et dans l'estimation des modèles des pratiques de modélisation qui n'ont été que très partiellement abordées dans ce guide.
- sur les spécifications de modèles de choix modal, qui sont en constante évolution : modèles MEV, modèles mixtes, modèles probit rendus possibles par l'amélioration des performances de calcul.

9. Annexes

9.1 Annexe 1 - modèles agrégés

On ne présentera ici que les modèles agrégés le plus souvent mis en œuvre et en particulier :

- les grilles de partage modal (modèle normatif) ;
- les courbes en S de partage modal (avec une application sous forme de modèle agrégé de type logit binomial).

Pour plus de détails sur les autres types de modèles, le lecteur pourra se référer à d'autres ouvrages, comme le guide [CERTU, 2003].

9.1.1 Grilles de partage modal

Le principe de grilles pour « modéliser » le choix modal est un principe simple et transparent mais réducteur. Il consiste à définir une norme de répartition modale par couple origine-destination des flux.

On effectue tout d'abord une typologie des zones du modèle sur des critères socioéconomiques et/ou des critères géographiques.

Le plus souvent les critères utilisés vont être liés à la localisation géographique des zones (centre/périphérie) et à l'offre de transports (desserte TC/TCSP, contraintes de stationnement...).

Cette typologie définit enfin des liaisons pour lesquelles à partir des données d'observation on attribue un pourcentage par mode de transport. Ces pourcentages peuvent être globaux ou définis par motif.

Pour être pertinent, et pour avoir des valeurs significatives sur les pourcentages, il convient de définir un nombre restreint de macro-zones.

Cette démarche d'utilisation de grilles de partage modal n'est pas récente et convient parfaitement aux démarches de modélisation du trafic automobile. En effet, si cette méthode est satisfaisante pour les déplacements en voiture qui représentent 60 % des déplacements (tous modes), elle n'est pas adaptée aux transports collectifs qui n'en représentent, en général, que 10 %.

9.1.2 Courbes de partage modal

Cette méthode consiste à donner la répartition d'usage entre deux modes (le plus souvent VP et TC) sous forme d'une courbe. En général, on essaie de lier le pourcentage d'utilisation d'un mode par rapport à l'autre à la différence (ou au rapport) des niveaux de service offerts par l'un et l'autre mode. Ces différences de niveaux de service sont souvent mesurées par des variables comme le temps de parcours, les coûts.... (on parle en général de coûts généralisés).

Ce type de courbe peut être obtenu de différentes manières :

- soit directement à partir des données d'observation (les données d'enquêtes ménages déplacements par exemple) ou déduites d'une répartition constatée (déduite par régression par exemple),
- comme résultat d'un modèle logit agrégé (dont les fonctions d'utilité dépendent en général uniquement des temps et des coûts).

Les modèles logit (dont une présentation plus détaillée est faite dans le chapitre 2) sont fondés sur le principe du « rational choice ». Entre plusieurs modes de transport, un individu (ou un groupe d'individus) choisira le mode le plus « avantageux » (le plus rapide ou le moins coûteux par exemple). Quand un individu a le choix entre deux modes de transport, on observe ainsi une courbe en forme de S mettant en relation la part d'individus choisissant le

mode de transport n° 1 et la différence (ou le rapport) entre les coûts généralisés de transport C1 et C2 des modes n° 1 et n° 2.

Les modèles de « rational choice » qui sont appliqués avec des approches agrégées s'écrivent de manière pratiquement systématique sous la forme de modèles logit multinomiaux :

$$P_k = \frac{\exp(-aC_{ij}^k)}{\sum_m \exp(-aC_{ij}^m)}$$

où :

- **P_k** : probabilité de choisir le mode **k** parmi **M** modes de transport
- C_{ij}^m : coût généralisé du trajet entre **i** et **j** avec le mode **m** sur la liaison **ij**.
- **a** : paramètre de calage.

Dans le cas classique où on a deux modes de transport en présence, on appelle ce modèle le modèle logit binomial, modèle qui peut aussi s'écrire sous la forme :

$$P_1 = \frac{1}{1 + \exp(-a(C_{ij}^2 - C_{ij}^1))}$$

Principes de calage : dans les modèles agrégés, le principe est de regrouper les données suivant un certain zonage de l'agglomération étudiée. On évalue alors pour les *n* couples origine-destination les parts modales et les coûts généralisés. Ces *n* observations permettent de caler le modèle.

Dans le cas du modèle logit binomial, le modèle peut se mettre sous la forme $\ln[(1/P_1) - 1] = -\lambda(C_{ij}^1 - C_{ij}^2)$

Il est possible de caler λ assez facilement par régression linéaire (dans le cas de logit multinomiaux le calage des paramètres est plus complexe et se base généralement sur la méthode du maximum de vraisemblance).

Ce type de modèle logit binomial agrégé a le mérite de la simplicité. Le risque est que l'écart de temps ou de coût ne soit pas le seul facteur explicatif des choix modaux.

Les modèles de type logit multinomial calibrés sur données désagrégées, qui sont l'objet de ce guide, sont en fait un cas relativement général par rapport aux deux méthodes précédentes :

- En ne retenant que des critères de localisation des zones origine et destination, et de variables socioéconomiques, ces modèles correspondent à une grille de partage modal.
- En ne retenant que des variables agrégées, ces modèles correspondent à une courbe de partage modal.

9.2 Annexe 2 - méthode de décomposition des flux VP en flux VPC/VPP

La méthode présentée ici a été développée par la DTer Ouest du Cerema pour le modèle Sim'44 au niveau de l'agglomération de Nantes.

La problématique du choix modal est effectuée par segment de population mais tous motifs confondus. Les motifs du type accompagnement ou covoiturage ne sont donc pas traités de façon spécifique à ce stade.

Les paramètres de choix de mode sont calés de façon agrégée sur les données globales disponibles. Quatre modes sont distingués :

- la marche à pied,
- le vélo,
- la voiture et la moto (conducteur ou passager, indifféremment),
- les transports en commun.

Par exemple, pour le segment primaire des actifs, on effectue un calage des paramètres du choix de mode pour les actifs motorisés d'une part, pour les actifs non motorisés d'autre part. Dans chacun de ces segments, on considère que le voyageur peut être soit conducteur, soit passager, mais évidemment dans une proportion très différente pour les motorisés et pour les non motorisés.

L'analyse de l'ENTD 2007-2008, pour les déplacements des résidents de Loire-Atlantique, permet d'évaluer la proportion moyenne de conducteurs et de passagers pour chacun des segments de population retenus dans le modèle, lors de leurs déplacements. Les parts observées sont les suivantes :

	Segment	Conducteur (VL et Moto)	Passager
Motorisés	Actifs urbains	95 %	5 %
	Actif ruraux	97 %	3 %
	Etudiants	73 %	27 %
	Inactifs urbains	90 %	10 %
	Inactif ruraux	72 %	28 %
	Moyenne	92 %	8 %
Non motorisés	Actifs urbains	63 %	37 %
	Etudiants	8 %	92 %
	Inactifs urbains	28 %	72 %
	Inactif ruraux	8 %	92 %
	Moyenne	33 %	67 %

Par ailleurs, la segmentation du modèle distingue également les résidents de Nantes, de Nantes Métropole et du reste du département de Loire-Atlantique.

Sur ces bases, une méthode de conversion des flux de voyageurs en mode individuel motorisé (VL ou moto) en nombre de véhicules a été établie.

On part donc des matrices OD de voyageurs en VL ou moto par segment de population. A chaque segment, à partir des éléments observés, on fixe une part de conducteurs et la part complémentaire de passagers. En travaillant OD par OD, on cumule sur l'ensemble des segments le nombre de voyageurs conducteurs d'une part et le nombre de voyageurs passagers d'autre part. Cela permet de calculer le taux d'occupation global de l'OD. On obtient *in fine* une matrice de passagers et une matrice de conducteurs, celle-ci étant bien entendu assimilable à une matrice de véhicules.

Il est possible d'affiner la méthode en n'associant dans les mêmes véhicules, par exemple, que les résidents d'un même secteur. C'est ce qui a été réalisé en associant entre eux d'une part les résidents de Nantes Métropole, d'autre part les autres habitants de Loire-Atlantique. A noter que, du fait que les déplacements des scolaires de 11-14 ans et de 15-17 ans sont pris

en compte dans Nantes Métropole mais pas dans le reste du département, on aboutit pour les déplacements en VL des résidents de ces deux territoires à un taux d'occupation différent (plus élevé pour Nantes Métropole du fait que les scolaires font partie des passagers, soit 1,25 *versus* 1,15 pour le reste du département).

Par ailleurs, pour répondre à une problématique de covoiturage domicile-travail, on peut envisager de procéder de la même façon à des regroupements d'utilisateurs par motif plutôt que par lieu de résidence, voire d'associer les deux critères.

Cette méthode permet de maîtriser, dans la construction d'un scénario prospectif, l'évolution du taux d'occupation des véhicules en jouant sur le taux de conducteurs, globalement ou au sein de chaque segment (actifs seuls ou autres). On peut ainsi obtenir un taux souhaité *a priori*, par exemple pour intégrer un certain niveau de pratique du covoiturage. Cependant, dans le cas où un accroissement important du taux d'occupation moyen des véhicules serait envisagé, il conviendrait de procéder à une nouvelle estimation des temps généralisés VL, tenant compte d'une répartition des coûts de transport portant sur un nombre plus élevé d'occupants.

10. Bibliographie

- [ANSELME et COMBES, 2012] ANSELME T., COMBES F. (2012), « Complex model structures versus fine disaggregation of travel demand : the specification of an urban mode choice model for the French urban area of Rouen », in *Proceedings of the European Transport Conference 2012*, Glasgow, UK.
- [AXHAUSEN et al., 2008] AXHAUSEN K.W., HESS S., KÖNIG A., ABAY G., BATES J.J., BIERLAIRE M. (2008), « Income and distance elasticities of values of travel time savings : New Swiss results », in *Transport Policy*, 15(3), p. 173-185.
- [BEN-AKIVA et LERMAN, 1985] BEN-AKIVA M., LERMAN S. (1985), *Discrete Choice Analysis : Theory and application to travel demand*, The MIT Press, 1985, 390 p.
- [BIERLAIRE, 2003] BIERLAIRE M. (2003), « BIOGEME : a free package for the estimation of discrete choice models », in *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.
- [BIERLAIRE et al., 2008] BIERLAIRE M., BOLDUC D., MCFADDEN D. (2008), « The estimation of generalized extreme value models from choice-based samples », in *Transportation Research Part B*, vol. 42, p. 381-394.
- [BONNEL, 2004] BONNEL P. (2004), *Prévoir la demande de transports*, Presse de l'Ecole nationale des ponts et chaussées.
- [CERTU, 1998] CERTU (1998), *Comportements de déplacement en milieu urbain : les modèles de choix discrets – vers une approche désagrégée et multimodale*.
- [CERTU, 2003] CERTU (2003), *Modélisation des déplacements urbains de voyageurs, Guide des pratiques*.
- [CERTU, 2005] CERTU (2005), *L'estimation de la densité par la méthode du noyau - Méthode et outils, Rapport d'étude*.
- [CERTU, 2009] CERTU (2009), *Contraintes de stationnement et pratiques modales - Méthodologie et étude des cas de Lille, Lyon et Montpellier, Rapport d'étude*.

-
- [DE JONG et GUNN, 2001]** **DE JONG G., GUNN H.F.** (2001), « Recent Evidence on Car Cost and Time Elasticities of travel Demand in Europe », in *Journal of Transport Economics and Policy*, 35(2), p 137-160.
- [DE PALMA et FONTAN., 2001]** **DE PALMA A., FONTAN C.** (2001), « Choix modal et valeurs du temps en Ile-de-France », in *Recherche Transports Sécurité*, vol. 71, p. 24-46.
- [DEPARTMENT FOR TRANSPORT, 2012]** **DEPARTMENT FOR TRANSPORT** (2012), *Transport analysis Guidance – UK WebTAG*. <http://www.dft.gov.uk/webtag/>
- [FHWA, 2007]** **FEDERAL HIGHWAY ADMINISTRATION** (2007), *Technical Solutions to Overcrowded Park and Ride Facilities - Final Report*, University Transportation Research Center – Region 2, May 2007, FHWA-NJ-2007-011.
- [FELDMAN et al., 2012]** **FELDMAN O., FORERO-MARTINEZ J., COOMBE D.** (2012), « Alternative gravity modelling approaches for trip matrix synthesis », in *Proceedings of the European Transport Conference*.
- [GARCIA CASTELLO et LEURENT, 2007]** **GARCIA CASTELLO F.J, LEURENT F.** (2007), *Avancées dans la modélisation des choix discrets de déplacement, une revue critique internationale*, France, ATEC-ITS.
- [GUEVARA et BEN-AKIVA, 2006]** **GUEVARA C.A., BEN-AKIVA M.** (2006), « Endogeneity in residential location choice models », in *Transportation Research Record*, vol. 1977, p. 60-66.
- [HALLDORSOTTIR et al., 2011]** **HALLDORSOTTIR K., CHRISTENSEN L., JENSEN T.C., PRATO C.** (2011), « Modeling mode choice in short trips – shifting from car to bicycle », in *Proceedings of the European Transport Conference 2011*, Glasgow, UK.
- [HASIAK et al., 2011]** **HASIAK F., MERLE N., VERRY D.** (2011), « Parking policies and modal choice : a disaggregated approach applied to French cities », in *Proceedings of the European Transport Conference 2011*, Glasgow, UK.
- [HESS et al., 2004]** **HESS S., BIERLAIRE M., POLAK J.W.** (2004), *Estimation of value-of-time using Mixed Logit models*, TRB 2005 Annual Meeting.

- [KARASMAA, 2003] **KARASMAA N.** (2003), *The transferability of travel demand models. An analysis of transfer methods, data quality and model estimation*, Dissertation for the degree of Doctor of Science in Technology. in Helsinki University of Technology, Transportation Engineering Publication.
- [LEURENT et LIU, 2008] **LEURENT F., LIU K.** (2008), « On seat congestion, passenger comfort and route choice in urban transit : a network equilibrium assignment model with application to Paris », Paper submitted for 88th TRB meeting, Paris, 17 p.
- [LEURENT et al., 2011] **LEURENT F., BENEZECH V., SAMADZAD M.** (2011), « A Stochastic Model of Trip End Disaggregation in Traffic Assignment to a Transportation Network », Paper presented at the XIVth Conference of the EWGT, Poznan, September 2011.
- [MEDDE, 2007] (2007), *Instruction relative aux méthodes d'évaluation économique des investissements routiers interurbains.*
- [NCHRP, 2010] **NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM Synthesis 406** (2010), *Advanced practices in travel forecasting*, Transportation Research Board.
- [NCHRP, 2012] **NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM Report 716** (2012), *Travel Demand Forecasting : Parameters and Techniques*, Transportation Research Board.
- [ORTUZAR et WILLUMSEN, 2011] **ORTUZAR J., WILLUMSEN L.** (2011), *Modelling transport*, John Wiley & Sons.
- [PONS, 2011] **PONS D.** (2011), *Mise en place d'enquêtes par préférences déclarées dans le cadre de projets d'étude relatifs au secteur des transports de personnes*, Thèse de l'Université Lumière Lyon II.
- [PREDIT, 2010] **PREDIT** (2010), *Développer une infrastructure de covoiturage ?*
- [PREDIT, 2011] **PREDIT** (2011), *Plusieurs Options de Rabattement ou Transfert Vélo et Réseaux de Transport - Approche multi-aspect des diverses formules d'intermodalité.*
- [SANTOSO et TSUNOKAWA, 2009] **SANTOSO D.S., TSUNOKAWA K.** (2009), « Economic Perspective of Transferability of Mode Choice Models », in *Journal of the Eastern Asia Society of Transportation Studies*, vol. 8.

-
- [SAUER, 2010] SAUER S. (2010), *Assessing Park-and-Ride Impacts – Preliminary Investigation* – Caltrans Division of Research and Innovation.
- [SAUVANET, 2011] SAUVANET G. (2011), *Recherche de chemins multiobjectifs pour la conception et la réalisation d'une centrale de mobilité destinée aux cyclistes*, Thèse de doctorat en informatique de l'Université François Rabelais – Polytech Tours.
- [SETRA, 2008] SETRA (2008), *Accessibilité des territoires et des services- notions et représentations*, Rapport d'étude.
- [SETRA, 2009] SETRA (2009), « Modélisation de la variation annuelle des trafics routiers – Calcul des facteurs de concentration », Note d'information, série Transports, n° 4.
- [SETRA, 2013] SETRA (2013), *Sources de données flottantes et mobiles liées au trafic*, Rapport d'étude, novembre 2013.
- [YÁÑEZ et al., 2009] YÁÑEZ M.F., RAVEAU S., ROJAS M., ORTÚZAR J. de D. (2009), « Modelling and forecasting with latent variables in discrete choice panel models », in *Proceedings of the European Transport Conference*.

11. Liste des sigles

Nom	Description
ASC	Constantes de modes ou de segments (<i>Alternatives specific constants</i>)
CEREMA	Centre d'Études et d'Expertise sur les Risques, l'Environnement, la Mobilité, et l'Aménagement
CSP	Catégorie Socio Professionnelle
EMD	Enquête Ménages Déplacements
ENTD	Enquête Nationale Transports et Déplacements
Enquêtes RP et SP	<p>Enquête RP = enquête de préférences révélées (<i>revealed preferences</i>) : enquêtes sur les préférences des individus révélées par leur comportement actuel</p> <p>Enquête SP = enquête de préférences déclarées (<i>stated preferences</i>) : enquête sur les préférences des individus confrontés à des situations de choix hypothétiques</p>
FCD	<i>Floating Car Data</i> : données enregistrées dans les systèmes électroniques embarqués dans le véhicule et permettant de recueillir des informations (localisation, vitesse, accélération, sens du déplacement...) issues de capteurs proprioceptifs. Elles sont géolocalisées et récupérées de façon anonyme et à intervalle régulier à partir des véhicules traceurs.
GPS	<i>Global Positioning System</i>
HCM	<i>Highway Capacity Manuel</i>
IIA	Propriété d'indépendance vis à vis des alternatives non utilisées (<i>Independance of Irrelevant Alternatives</i>)
MAP	Marche à Pied
MNL	Modèle logit multinomial
NL	Modèle nested logit
OD	origine-destination
P+R	Parking relais (<i>Park and ride</i>)
SIM	Système d'Information Multimodale
SIG	Système d'Information Géographique
SIREDO	Système Informatisé de Recueil de Données, pour les données de trafic routier
TAD	Transport à la demande
TC	Transports en commun
VDT	Valeur du Temps – Consentement des usagers à payer un service pour gagner du temps dans leurs déplacements
VP	Voiture particulière

Table des matières

1.Introduction.....	5
1.1 Contexte de création de ce guide.....	5
1.2 Champs d'application.....	5
1.3 Démarche adoptée.....	6
2.Principes de base.....	8
2.1 Choix des alternatives.....	8
2.2 Théorie des choix discrets.....	9
2.2.1Principe de l'utilité aléatoire.....	9
2.2.2Cas du modèle logit multinomial (MNL).....	10
2.2.2.1Hypothèses du modèle logit multinomial (MNL).....	10
2.2.2.2Propriété d'indépendance vis-à-vis des autres alternatives (propriété IIA).....	11
2.2.3Autres types de modèles.....	12
2.3 Quelle formulation de l'utilité ?.....	13
2.4 Disponibilité des alternatives.....	15
3.Bases de données.....	16
3.1 Bases de données disponibles.....	16
3.1.1Recensement de la population.....	16
3.1.2Enquêtes « voyageurs ».....	16
3.1.3Enquêtes ménages / déplacements (EMD).....	17
3.2 Avec quelles bases de données estimer le modèle ?.....	17
3.3 Combinaison d'enquêtes.....	18
3.3.1Sur-échantillonnage d'un mode de transport.....	18
3.3.2Enquête complémentaire de type préférences déclarées.....	19
3.3.3Combinaison de plusieurs enquêtes de préférences révélées.....	20
3.4 Les bases de données sur l'offre de transports.....	20
4.Du bon calcul du coût généralisé.....	23
4.1 Périodes de modélisation.....	23
4.2 Coût généralisé TC.....	24
4.2.1Traitement des trajets terminaux.....	24
4.2.2Calcul des itinéraires TC.....	25
4.3 Coût généralisé VP.....	26
4.4 Coût généralisé MAP.....	29
4.5 Cas particulier des coûts généralisés intrazonaux.....	30
4.6 Incidence de la congestion sur le calibrage du modèle.....	31
5.Les outils statistiques disponibles pour estimer le modèle de choix modal.....	32
5.1 Les indicateurs statistiques de base.....	32
5.1.1Indicateurs généraux.....	32
5.1.1.1 ρ^2 , ρ^2 , et ρ^2 ajusté.....	33
5.1.1.2Ratio de vraisemblance (LR).....	34
5.1.2Indicateurs spécifiques au niveau des variables.....	35
5.1.2.1Signe (et valeur) des paramètres.....	35
5.1.2.2t tests asymptotiques.....	35
5.1.2.3Corrélations entre paramètres.....	36
5.1.3Récapitulatif.....	38
5.2 La propriété IIA.....	38

5.3 L'endogénéité.....	39
5.4 Le biais d'agrégation.....	40
5.5 La segmentation.....	40
6.Processus de construction d'un modèle.....	43
6.1 Problème préliminaire : le modèle dans son univers d'utilisation.....	43
6.2 Proposition d'un processus pratique de construction du modèle.....	45
6.2.1Principes de construction du modèle.....	45
6.2.2Détection des valeurs aberrantes dans les bases de données.....	46
6.2.3Mise en œuvre du processus.....	48
6.2.3.1Modèle 1 : modèle simple.....	48
6.2.3.2Modèle 2 : modèle segmenté selon les motifs de déplacement.....	55
6.2.3.3Modèle 3 : modèle segmenté selon la captivité.....	57
6.2.3.4Modèle 4 : transformation des variables.....	59
6.2.3.5Modèle 5 : introduction de constantes d'alternatives permettant de capter les préférences individuelles.....	61
6.2.3.6Discussion à partir des résultats de ces 5 modèles, et choix du modèle à retenir.....	63
6.3 Différents problèmes spécifiques.....	68
6.3.1Distinction modes légers / modes lourds de transports en commun.....	68
6.3.2Effet du prix du déplacement.....	71
6.3.3Nouvelles pratiques / nouveaux services de mobilité.....	72
6.3.3.1La montée en puissance du covoiturage.....	73
6.3.3.2L'intérêt croissant pour les parcs relais (park and ride = P+R).....	74
6.3.3.3La difficile prise en compte du vélo.....	75
6.3.3.4Transport à la demande et autopartage, des modes marginaux.....	76
6.4 Valider le modèle.....	76
6.4.1Au niveau désagrégé.....	76
6.4.2Au niveau agrégé.....	77
7.Faire des simulations avec un modèle de choix modal.....	78
7.1 Méthodes d'application du modèle calibré.....	78
7.1.1Présentation des deux méthodes disponibles.....	78
7.1.2Illustrations.....	80
7.2 Analyser les élasticités.....	82
7.3 Intégrer des variables supplémentaires.....	85
7.4 Réutiliser un autre modèle.....	85
8.Perspectives.....	89
9. Annexes.....	90
9.1 Annexe 1 - modèles agrégés.....	90
9.1.1Grilles de partage modal.....	90
9.1.2Courbes de partage modal.....	90
9.2 Annexe 2 - méthode de décomposition des flux VP en flux VPC/VPP.....	91
10.Bibliographie.....	94
11.Liste des sigles.....	98

Multimodal mobility modelling

Designing a modal choice model - a modeller's guide

Mobility planning requires increasingly accurate simulation tools, able to take differences in individual mobility behaviour into account. To meet these objectives, transport demand analysis methods have been getting more and more sophisticated and complex in France since the early 2000s.

This publication, for all technicians developing or using multimodal mobility models, is based on feedback from Cerema staff who produce or follow up models in many areas in France. It gives an account of methodological advances in the field of discrete modal choice models and offers a number of practical recommendations to ensure better use of these models as tools for decision-making.

Modelización multimodal de los desplazamientos de viajeros

Concebir un modelo de elección modal – guía del modelizador

La planificación de los desplazamientos requiere herramientas de simulación cada vez más precisas, capaces por ejemplo de considerar las diferencias entre los comportamientos individuales de movilidad. Los métodos de análisis de la demanda de transporte son más detallados y complejos en Francia, desde principios de los años 2000, para responder a estos objetivos.

La presente obra, destinada a todos los técnicos que desarrollan o utilizan modelos multimodales de desplazamiento, se basa en el análisis de experiencia de los técnicos del Cerema, que producen o siguen modelos en numerosos territorios de Francia. Pretende rendir cuentas de los avances metodológicos en el ámbito de los modelos discretos de elección modal y propone unas cuantas recomendaciones prácticas con el fin de garantizar un mejor uso de esos modelos como herramientas de ayuda a la decisión.

© 2015 - Cerema

Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement, créé au 1er janvier 2014 par la fusion des 8 CETE, du Certu, du Cetmef et du Sétra.

Le Cerema est un établissement public à caractère administratif (EPA), sous la tutelle conjointe du ministère de l'Écologie, du Développement durable et de l'Énergie et du ministère du Logement et de l'Égalité des territoires. Il a pour mission d'apporter un appui scientifique et technique renforcé, pour élaborer, mettre en œuvre et évaluer les politiques publiques de l'aménagement et du développement durables, auprès de tous les acteurs impliqués (État, collectivités territoriales, acteurs économiques ou associatifs, partenaires scientifiques).

Toute reproduction intégrale ou partielle, faite sans le consentement du Cerema est illicite (loi du 11 mars 1957). Cette reproduction par quelque procédé que ce soit, constituerait une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal.

Dépôt légal : octobre 2015

ISBN : 978-2-37180-092-2

ISSN : 2276-0164

Prix : gratuit

Illustration de couverture : Fréquentation des transports collectifs urbains à Rouen.

Représentation des flux et des montées descentes aux arrêts.

Source Cerema DterNC mai 2015

Éditions du Cerema

Cité des mobilités

25 avenue François Mitterrand

CS 92803

69674 Bron Cedex

Bureau de vente

Cerema / Direction technique Territoires et ville

2 rue Antoine Charial

CS 33927

69426 Lyon Cedex 03 – France

Tél. 04 72 74 59 59 – Fax. 04 72 74 57 80

catalogue.territoires-ville.cerema.fr

La collection « Références » du Cerema

Cette collection regroupe l'ensemble des documents de référence portant sur l'état de l'art dans les domaines d'expertise du Cerema (recommandations méthodologiques, règles techniques, savoir-faire...), dans une version stabilisée et validée. Destinée à un public de généralistes et de spécialistes, sa rédaction pédagogique et concrète facilite l'appropriation et l'application des recommandations par le professionnel en situation opérationnelle.

Modélisation multimodale des déplacements de voyageurs

Concevoir un modèle de choix modal

La planification des déplacements requiert des outils de simulations de plus en plus précis, capables notamment de prendre en compte les différences de comportements individuels de mobilité. Les méthodes d'analyse de la demande de transports se sont enrichies et complexifiées en France depuis le début des années 2000 pour répondre à ces objectifs.

Le présent ouvrage, à destination de tous les techniciens développant ou utilisant des modèles multimodaux de déplacements, s'appuie sur les retours d'expériences des agents du Cerema qui produisent ou suivent des modèles sur de nombreux territoires en France. Il vise à rendre compte des avancées méthodologiques dans le domaine des modèles discrets de choix modal et propose un certain nombre de recommandations pratiques afin de garantir un meilleur usage de ces modèles comme outils d'aide à la décision.

Sur le même thème

Mobilité en transitions

Connaître, comprendre et représenter
2015

Modélisation des déplacements urbains de voyageurs

Guide des pratiques
2003

Comportements de déplacements en milieu urbain: les modèles de choix discrets

Vers une approche désagrégée et multimodale
1998

En téléchargement gratuit sur :
catalogue.territoires-ville.cerema.fr

Connaissance et prévention des risques - Développement des infrastructures - Énergie et climat - Gestion du patrimoine d'infrastructures
Impacts sur la santé - Mobilité et transports - Territoires durables et ressources naturelles - Ville et bâtiments durables

gratuit
ISSN : 2276-0164
ISBN : 978-2-37180-092-2

