



PREDIT 2002-2006
Groupe 1
Mobilité, territoires et développement durable

SIMAU^RIF :

**Modèle dynamique de SIMulation de l'interAction
Urbainisation-transports en Région Ile-de-France**

Application à la Tangentielle nord

Référence : n° 05 MT 6008

Rapport intermédiaire de la 2^{ème} phase



Octobre 2005

Institut d'Aménagement et d'Urbanisme de la Région d'Ile-de-France (IAURIF)
15 rue Falguière 75740 Paris cedex 15
Tél. 01.53.85.53.85 - Télécopie 01.53.85.76.02
E-mail : dgcomm@iaurif.org –http : www.iaurif.org
Directeur Général : François DUGENY
Directeur du Département Transports et Infrastructures : Alain MEYERE

Université de Cergy-Pontoise
Laboratoire THEMA
33 boulevard du Port 95011 Cergy-Pontoise Cedex
Tél. 01.34.25.63.37 - Télécopie 01.34.25.62.33
Directeur : Régis RENAULT

Equipe :

Pr André DE PALMA, THEMA

Dany NGUYEN-LUONG, IAURIF

Kiarash MOTAMEDDI, THEMA

Nathalie PICARD, THEMA

José MOYANO, adpC

Pr Paul WADDELL, Université de Washington,

Fanny CHAUCHARD-LEFEVRE, IAURIF, élève de 3ème année de l'ENSAE

Hakim OUARAS, THEMA

PLAN

1.	Introduction	6
2.	Méthodologie du projet	7
2.1.	Caractéristiques essentielles du projet	7
2.2.	Activités et tâches de la première phase	8
2.3.	Programme à réaliser pour la deuxième phase	8
2.4.	Programme complémentaire : un nouvel UrbanSim	9
3.	Finalisation de la base de données d'entrée	11
3.1.	Variables de desserte spatiale en transports en commun	12
3.2.	Variables de temps de parcours vers Châtelet-les-Halles	13
3.2.1.	Temps de parcours en VP	13
3.2.2.	Temps de parcours en TC	14
3.3.	Variable distance à Châtelet-les-Halles	14
3.4.	Variable de contraintes environnementales	15
3.4.1.	Les sites inscrits et classés	15
3.4.2.	Les zones de bruit aéroportuaires	15
3.5.	Variables de contraintes politiques locales	16
3.6.	Variable development_type	18
3.7.	Variable Politique de la Ville	20
3.8.	Variables de voisinage	21
4.	Recalage du modèle de demande de l'IAURIF et de METROPOLIS	23
4.1.	Les réseaux	23
4.1.1.	Routes	23
4.1.2.	Transports en commun	25
4.2.	Les 3 étapes du modèle de demande	26
4.2.1.	Génération	26
4.2.2.	Distribution	27
4.2.3.	Choix modal	31
4.3.	Recalage de METROPOLIS	36
4.3.1.	Rappel du calage semi-agrégé du projet QUATUOR	36
4.3.2.	Présentation des nouvelles données de comptage	37
4.3.3.	Estimation des fonctions de congestion	41
4.3.4.	Calage global du modèle	46
4.3.5.	Prise en compte du modèle d'affectation en TC	46
4.4.	Programmation	46
5.	Constitution des bases de données de calage	47
5.1.	Rappel : les 4 modèles d'UrbanSim à estimer	47
5.1.1.	Modèle de localisation résidentielle des ménages	47
5.1.2.	Modèle de localisation des emplois	47
5.1.3.	Modèle de l'évolution du MOS	48
5.1.4.	Modèle du prix du foncier	48

5.2.	Base de calage du modèle d'évolution du MOS	49
5.2.1.	Contraintes environnementales	49
5.2.2.	Contraintes politiques locales.....	50
5.2.3.	Statistiques descriptives sur l'évolution des cellules	50
5.2.4.	Récapitulation : variables de la base de calage	52
5.3.	Base de calage du modèle de localisation résidentielle.....	54
5.3.1.	Statistiques descriptives sur la mobilité résidentielle entre 1990 et 1999	55
5.3.2.	Variables sur les caractéristiques socio-économiques des ménages	58
5.3.3.	Variables sur les indicateurs d'accessibilité.....	58
5.3.4.	Récapitulation : variables de la base de calage	60
5.4.	Base de calage du modèle de localisation des emplois	64
5.4.1.	Statistiques descriptives sur la mobilité des emplois entre 1990 et 1999	64
5.4.2.	Récapitulation : variables de la base de calage	65
5.5.	Base de calage du modèle de prix	65
6.	Résultats des estimations de deux modules	
	d'UrbanSim	67
6.1.	Calage du modèle d'évolution du MOS.....	67
6.1.1.	Le logit multinomial.....	67
6.1.2.	Etude des corrélations entre les variables explicatives	69
6.1.3.	Sélection des variables explicatives	71
6.1.4.	Résultats du calage	73
6.1.5.	Validité du modèle	75
6.1.6.	Le pourcentage de prédictions correctes	76
6.1.7.	Amélioration proposée du modèle	77
6.2.	Calage du module de localisation des ménages	80
6.2.1.	Le logit multinomial conditionnel.....	80
6.2.2.	Sélection des alternatives	81
6.2.3.	Sélection des variables explicatives	81
6.2.4.	Résultats du calage	81
6.2.5.	Validité du modèle	83
6.2.6.	Le pourcentage de prédictions correctes	83
7.	Un nouvel UrbanSim	84
7.1.	Méthodologie	84
7.2.	Nouvelle architecture	84
7.3.	Le module de localisation résidentielle des ménages au niveau communal	84
7.3.1.	Modèle de localisation en trois étapes.....	84
7.3.2.	Méthode d'estimation.....	86
7.3.3.	Base de calage	86
7.3.4.	Résultats du calage	90
7.3.5.	Validité du modèle	93
7.4.	Le module de prix du foncier au niveau communal.....	96
7.4.1.	Statistiques descriptives sur la base initiale	97
7.4.2.	Méthode d'imputation des prix manquants.....	98
7.4.3.	Base de calage	101
7.4.4.	Résultats du calage	103
7.5.	Modèle de localisation des emplois au niveau communal	104
7.6.	Nouveau modèle de demande : le modèle EVA	106
7.6.1.	Présentation théorique	106

7.6.2.	Bases de calage.....	109
7.6.3.	Résultats du calage.....	111
7.6.4.	Application du modèle.....	115
7.7.	Perspectives.....	115
8.	Conclusions.....	116
Annexe 1	: Programme Scilab permettant de calculer les variables de voisinage.....	117
Annexe 2	: Méthode d'imputation des revenus aux ménages du Recensement général de la population.....	118
Annexe 3	: Quelques résultats d'estimation des fonctions de congestion de METROPOLIS.....	122

1. Introduction

L'Institut d'Aménagement et d'Urbanisme de la Région d'Ile-de-France (IAURIF) et le laboratoire THEMA de l'Université de Cergy-Pontoise ont engagé en juin 2003 une recherche sur l'interaction urbanisation-transport. Cette recherche vise à mettre au point un modèle dynamique et intégré de l'interaction urbanisation-transports sur la région d'Ile-de-France. Ce modèle, appelé *SIMAU*RIF, utilise trois modèles existants :

- UrbanSim, un modèle d'urbanisation (un « land-use model »), développé par l'Université de Washington à Seattle (Etats-Unis),
- METROPOLIS, un modèle routier de trafic dynamique développé par le THEMA,
- Le modèle global de l'IAURIF sur Davisum pour les graphes réseaux, les matrices de demande de déplacements origine-destination et la partie affectation en transports en commun.

Il est prévu d'appliquer l'outil *SIMAU*RIF sur le territoire desservi par la future Tangentielle nord entre Sartrouville et Noisy-le-Sec.

Ce projet de recherche a été programmé sur deux phases. Un **rapport n°1**¹ de la première phase a été remis au PREDIT le 14 octobre 2004.

Le présent rapport constitue le rapport intermédiaire de la seconde phase.

Le rapport final de cette deuxième phase est prévu pour le deuxième trimestre 2006.

¹ Il est recommandé de relire le rapport n°1 avant de lire celui-ci.

2. Méthodologie du projet

2.1. Caractéristiques essentielles du projet

Le projet *SIMAU*RIF présente 10 caractéristiques inédites en France :

- Les dimensions de l'aire d'étude : 12 000 km², 11 millions d'habitants, 5 millions d'emplois. A notre connaissance, jamais UrbanSim n'a été appliqué à une telle échelle.
- L'utilisation d'un maillage en cellules carrées de 50 000 cellules (exactement 49 236) de 500 mètres de côté couvrant l'ensemble de la région d'Ile-de-France.
- L'utilisation de la technique de micro-simulation : le modèle intégré tournera sur des données désagrégées (ménages et emplois).
- Le bouclage automatique entre un modèle de trafic à 5 étapes (METROPOLIS) et un modèle d'urbanisation (UrbanSim).
- L'utilisation de la simulation dynamique, le pas de temps étant un an pour UrbanSim et de trois ans pour METROPOLIS, la période de simulation s'étalant de 1999 à 2026.
- Une approche globale de l'interaction urbanisation-transport, tenant compte des contraintes environnementales et politiques, de l'accessibilité et des prix du foncier, et bénéficiant de l'expertise des urbanistes, économistes et démographes de l'IAURIF.
- Une forte dimension spatiale grâce au système d'information géographique de l'IAURIF permettant de travailler à un niveau infra-communal.
- Des modèles mathématiques sophistiqués basés sur des modèles de choix discrets (logits).
- Une disponibilité de données exceptionnelle à l'IAURIF mais imparfaite.
- Des méthodes innovantes de fusion de bases de données provenant de diverses sources afin de résoudre le problème des données manquantes.

2.2. Activités et tâches de la première phase

La première phase a été consacrée aux développements méthodologiques et au recueil de données sur la région d'Ile-de-France.

Plusieurs activités et tâches ont ainsi été réalisées ou entamées. **Elles sont décrites de manière détaillée dans le rapport n°1 :**

- Présentation du projet de la Tangentielle nord
- Revue de la littérature sur les modèles d'urbanisation et historique
- Etude théorique d'UrbanSim (les 9 modèles dont 3 modèles logit et 1 modèle de régression linéaire)
- Une première prise en main d'UrbanSim sur une étude de cas américaine (Eugène-Springfield)
- Adaptation du modèle de demande de l'IAURIF dans METROPOLIS
- Conception de l'architecture du modèle intégré *SIMAUURIF*
- Première phase de la constitution d'une base de données d'entrée en 50 000 cellules de l'Ile-de-France
- Première phase de la constitution de bases de données de calage à partir de différentes enquêtes

2.3. Programme à réaliser pour la deuxième phase

Un programme de travail a été proposé au PREDIT en janvier 2005. Il s'articule autour de 13 activités ou tâches. Elles ne sont pas réalisées de manière séquentielle mais en parallèle.

- 1) Finalisation de la base de données d'entrée
- 2) Finalisation des bases de données de calage
- 3) Calage de deux modules d'UrbanSim au niveau communal en vue d'une première approche sommaire
- 4) Calage des quatre modules principaux d'UrbanSim au niveau du maillage en 49 236 cellules
- 5) Paramétrage de l'ensemble d'UrbanSim dans la structure de données primaire d'UrbanSim.
- 6) Recalage et intégration du modèle de demande de l'IAURIF dans METROPOLIS
- 7) Prise en compte du modèle d'affectation en TC de l'IAURIF sous Davisum
- 8) Mise au point de l'interface entre UrbanSim et METROPOLIS
- 9) Recalage de METROPOLIS
- 10) Calage de l'ensemble UrbanSim/Davisum/METROPOLIS
- 11) Elaboration des scénarios de la Tangentielle nord
- 12) Simulation des scénarios de la Tangentielle nord
- 13) Rédaction des rapports

Dans ce rapport intermédiaire, on fera souvent référence à la dernière étude réalisée par l'équipe de Paul Waddell sur la région de Puget Sound. Deux rapports sont disponibles en téléchargement sur le site www.urbansim.org :

- « Database development for the Puget Sound Region . Technical report CUSPA-04-01 » - 17 janvier 2004
- « Model Estimation for the Puget Sound Region. Technical Report CUSPA-04-02 ». 28 août 2005.

Comparons les chiffres clés de Puget Sound avec ceux de l’Ile-de-France :

	Puget Sound	Ile-de-France
Aire	Non cité mais recalculé par nous : 2250 km ² :	12 000 km ²
Maillage	100 000 cellules de 150 m de côté	50 000 cellules de 500 m de côté
Ménages	1,3 millions en 2000	4,5 millions en 1999
Emplois	Non cité	5,04 millions en 1999
Période de calage	1995-2000 (semble-t-il ?)	1990-1999

Tableau 1 - Comparaisons Puget Sound - Ile-de-France

Il faut avouer que les deux rapports sur Puget Sound sont quelque peu décevants car ils manquent de transparence sur des questions essentielles : taille des bases de données de calage, méthodes d’estimation des modèles, modèle de trafic utilisé, etc ... En revanche, des pans entiers des deux rapports sont d’un intérêt très faible pour nous comme la description du synthétiseur de population qui fait l’objet de quasiment 50% du premier rapport ou de fastidieux tableaux de résultats d’estimation logit sans commentaires dans le deuxième rapport.

Les deux rapports permettent cependant de fournir quelques points de référence avec notre recherche.

2.4. Programme complémentaire : un nouvel UrbanSim

Le Professeur Paul Waddell a expertisé notre projet au début de son année sabbatique en France. Il a relevé des inadaptations de la version actuelle d’UrbanSim dont l’architecture et la structure de données sont basées sur un maillage de l’aire d’étude en cellules carrées de taille fixe (500 * 500 m dans le cas de l’Ile de France). En effet, exceptés le Mode d’Occupation du Sol (380 000 îlots couvrant la région), le Recensement général de la population (ménages et emplois localisés à l’îlot INSEE) et quelques couches du SIGR (Politique de la ville, Enquête Régionale Emplois, Plans Locaux d’Urbanisme ...), les autres données sont connues au mieux à la commune. Des méthodes de répartition de ces données à la cellule ont été décrites dans le rapport n°1 (pages 84 à 86). Mais il est évident que la répartition de données, disponibles initialement à la commune, sur des cellules crée une distorsion des données brutes, voire une perte d’information. Ce problème de l’inadaptation du maillage en cellules aux données disponibles n’est pas spécifique à la région d’Ile-de-France mais existe évidemment pour tous les cas d’étude. On peut cependant tempérer cet inconvénient en rappelant :

- l’intérêt du maillage dans le cadre de la théorie des automates cellulaires (voir paragraphe 3.8 sur le « Jeu de la vie »),
- l’utilisation de la puissance des SIG pour répartir de manière intelligente les données brutes de la commune vers les cellules

- la disponibilité de quatre données fondamentales (la couche MOS, la couche des îlots INSEE, l'Enquête régionale emplois 2000 et le carroyage de l'Enquête Globale Transports), qui, elles, sont connues à un niveau infracommunal, ce qui permet de les mettre en correspondance avec le maillage sans trop de distorsion.

Durant son séjour d'un an en France, Paul Waddell, assisté du THEMA, s'est efforcé de concevoir une nouvelle version d'UrbanSim afin de s'affranchir d'une structure de données à la cellule.

Les trois grands principes d'un nouvel UrbanSim sont les suivants :

- possibilité de travailler dans un découpage d'unités de taille variable, permettant d'utiliser des découpages existants : départements, communes, îlots. En revanche, on perd le concept de voisinage propre qui fait l'originalité de l'actuel UrbanSim (voir paragraphe 3.8).
- exploitation optimale des données brutes sans création de distorsion induite par un redécoupage à la cellule
- modèles économétriques de type nested logit (logit emboîtés) et non plus multinomial logit.

Donc parallèlement à la réalisation du programme proposé au Predit, l'équipe a travaillé sur une nouvelle architecture d'UrbanSim. Des estimations du module de localisation des ménages et du module du prix du foncier au niveau d'un découpage variable (communes) ont été réalisées. Ce travail correspond en fait à la tâche n°3 qui a été replacée dans un contexte plus large. La description de ce travail fait l'objet de la partie 7. Ce programme complémentaire est plus un programme de recherche qu'une application, il ne constitue donc pas le cœur du projet *SIMAU*RIF mais permet de préparer l'avenir. Il est prévu une implémentation du nouvel UrbanSim en langage Python et non en Java comme pour l'actuel UrbanSim. La mise à disposition de cette version n'est cependant pas envisagée avant fin 2006.

3. Finalisation de la base de données d'entrée

La base de données d'entrée est celle sur laquelle UrbanSim tournera la première année de la période de calage. Elle est à distinguer des bases de données de calage² à partir desquelles sont estimés les 4 modèles principaux d'UrbanSim. *SIMAU*RIF utilisera la version 2.3 d'UrbanSim datant de juillet 2004.

Rappelons que la structure de données d'UrbanSim comporte quatre types de tables :

- données
- constantes
- paramètres
- scénarios

Il est important d'avoir à l'esprit cette structure de données pré-établie d'UrbanSim, car c'est elle qui guide la méthodologie de construction de la base de données d'entrée puis celle des bases de calage.

Les tables « paramètres » seront remplies une fois que le calage des modèles d'UrbanSim sera effectué, les tables « Scénarios » seront constituées lors de l'élaboration des scénarios de l'application du modèle intégré à la Tangentielle nord. Les tables « constantes » seront décrites dans le rapport final.

On s'intéresse donc ici aux tables « données ».

Les trois tables les plus importantes sont *gridcells.tab*, *households.tab* et *jobs.tab*. Les tables *households.tab* et *jobs.tab* ont déjà été construites lors de la première phase. L'avantage de notre projet sur celui de Puget Sound est que nous disposons des Recensements généraux de la population et par conséquent, il n'est nul besoin d'utiliser un programme de synthèse de population.

Rappelons que la période de calage choisie est 1990-1999. Les bases de données sont donc constituées à 2 années : 1990 et 1999.

Il restait à compléter la table *gridcells.tab*, dont nous rappelons les variables dans le tableau 2.

Comme cela avait été suggéré dans le rapport n°1, cette table a été adaptée au cas de l'Ile-de-France. De nouvelles variables ont été ajoutées, en 1990 et en 1999. Elles concernent :

- la desserte spatiale en transports en commun (page 87 du rapport n°1),
- les temps de parcours vers Châtelet-les-Halles en VP et en TC (qui serviront aussi dans la table secondaire « *Zones.tab* », page 94 du rapport n°1),
- la distance à vol d'oiseau entre une cellule et Châtelet-les-Halles,
- les contraintes environnementales (page 88 du rapport n°1),
- les contraintes politiques locales (page 99 du rapport n°1),
- une variable à compléter : *DEVELOPMENT_TYPE* (page 89 du rapport n°1),

² Pour simplifier, on parlera de « base de calage » à la place de « base de données de calage ».

- la Politique de la Ville
- enfin, la série de nouvelles variables dites « de voisinage » (page 97 du rapport n°1).

Champ		Type	Unité	Valeur
GRID_ID	1	Entier		De 1 à 49236
RELATIVE_X	2	Entier		Lambert II
RELATIVE_Y	3	Entier		Lambert II
ZONE_ID	4	Entier		De 1 à 572
CITY_ID	5	Entier		Code INSEE
COUNTY_ID	6	Entier		N° département
PLAN_TYPE_ID	7	Entier		0, 1 ou 2
IS_OUTSIDE_URBAN_GROWTH_BOUNDARY	8	Entier		0 ou 1
FRACTION_RESIDENTIAL_LAND	9	Flottant		De 0 à 1
RESIDENTIAL_UNITS	10	Entier		Nb logements
COMMERCIAL_SQFT	11	Entier	m2	
INDUSTRIAL_SQFT	12	Entier	m2	
GOVERNMENTAL_SQFT	13	Entier	m2	
COMMERCIAL_IMPROVEMENT_VALUE	14	Entier	Euros	
INDUSTRIAL_IMPROVEMENT_VALUE	15	Entier	Euros	
GOVERNMENTAL_IMPROVEMENT_VALUE	16	Entier	Euros	
NONRESIDENTIAL_LAND_VALUE	17	Entier	Euros	
RESIDENTIAL_IMPROVEMENT_VALUE	18	Entier	Euros	
RESIDENTIAL_LAND_VALUE	19	Entier	Euros	
YEAR_BUILT	20	Entier	Années	
DISTANCE_TO_ARTERIAL	21	Entier		
DISTANCE_TO_HIGHWAY	22	Entier		
PERCENT_WATER	23	Entier		de 0 à 100
PERCENT_STREAM_BUFFER	24	Entier		
PERCENT_FLOODPLAIN	25	Entier		
PERCENT_WETLAND	26	Entier		
PERCENT_SLOPE	27	Entier		
PERCENT_OPEN_SPACE	28	Entier		
PERCENT_PUBLIC_SPACE	29	Entier		
PERCENT_ROADS	30	Entier		
DEVELOPMENT_TYPE	31	Entier		Formule

Tableau 2 : Caractéristiques d'une cellule

3.1. Variables de desserte spatiale en transports en commun

Par défaut dans UrbanSim, il n'y a aucune variable concernant les transports en commun. Trois nouvelles variables ont donc été ajoutées :

- Nombre de stations de métro par cellule
- Nombre de gares par cellule
- Nombre de stations de tramway par cellules

Ces variables ont été construites par intersection entre les couches correspondantes du SIGR (système d'information géographique régional) et du maillage en 50 000 cellules.

3.2. Variables de temps de parcours vers Châtelet-les-Halles

On a souhaité prendre en compte la notion de « distance au centre de Paris », qui reste, malgré le développement du polycentrisme de la région depuis 40 ans, une référence intemporelle à cause du poids prépondérant de la capitale en terme d'attractivité.

Cette « distance » d'une cellule à la cellule contenant la station Châtelet-les-Halles a été calculée comme temps de parcours en voiture particulière (VP) en heures creuses et en transports en commun (TC).

3.2.1. Temps de parcours en VP

Pour les temps VP en heures creuses, c'est le modèle METROPOLIS qui a été utilisé dans le découpage en 606 zones du modèle de trafic de l'IAURIF. On a simplement lancé un calcul de plus court chemin en temps à vide (considéré comme période en heures creuses) sur le graphe du réseau routier de l'IAURIF depuis chaque centroïde des 606 zones vers le centroïde de la zone de Châtelet. Pour calculer ensuite les temps de cellule à cellule, on a utilisé un coefficient correcteur fonction des distances du centroïde à Châtelet (d_1) et d'une cellule à Châtelet (d_2) qui sont connues (calcul de distance en coordonnées Lambert II). Finalement, le temps de transport moyen en VP t_2 pour aller d'une cellule à Châtelet est calculé de la façon suivante :

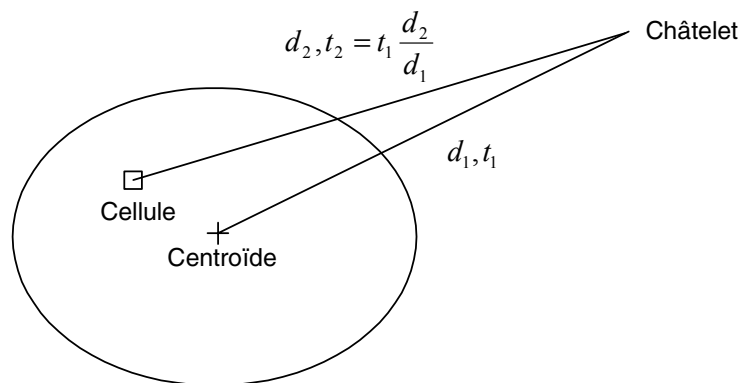


Figure 1 – temps de parcours en VP

Remarquons que l'on aurait dû calculer plutôt des temps de parcours en heure de pointe mais le problème est qu'au moment du calcul, la mise à jour de la matrice origine-destination en 606 zones n'était pas encore effectuée et le modèle METROPOLIS pas encore recalé. Depuis, le modèle de demande en 3 étapes a été recalé au niveau des étapes de la génération, de la distribution et du choix modal à partir de l'Enquête Globale Transports 2001-2002 (voir partie 4.2).

3.2.2. Temps de parcours en TC

Les temps en transport en commun pour aller d'une cellule à Châtelet sont calculés par le modèle TC de l'IAURIF basé sur le logiciel Davisum. Dans ce modèle, les horaires Ile-de-France sont disponibles sur deux heures de pointe du matin (7h – 8h59). Le modèle calcule le temps de transport t_0 pour aller du centroïde à la gare de rabattement la plus pratique³ pour rejoindre ensuite Châtelet ainsi que le temps de transport total t_2 ⁴ pour aller du centroïde d'une zone considérée à Châtelet. Pour simplifier le calcul, il a été considéré que la gare de rabattement affectée par le logiciel Davisum était la gare la plus proche du centroïde. Comme les distances du centroïde à la gare de rabattement (d_0) et du centroïde à Châtelet (d_1) sont connues, le temps de rabattement t_1 pour aller de centroïde à une cellule est ajusté et le temps t_3 pour aller d'une cellule à Châtelet est calculé de la façon suivante :

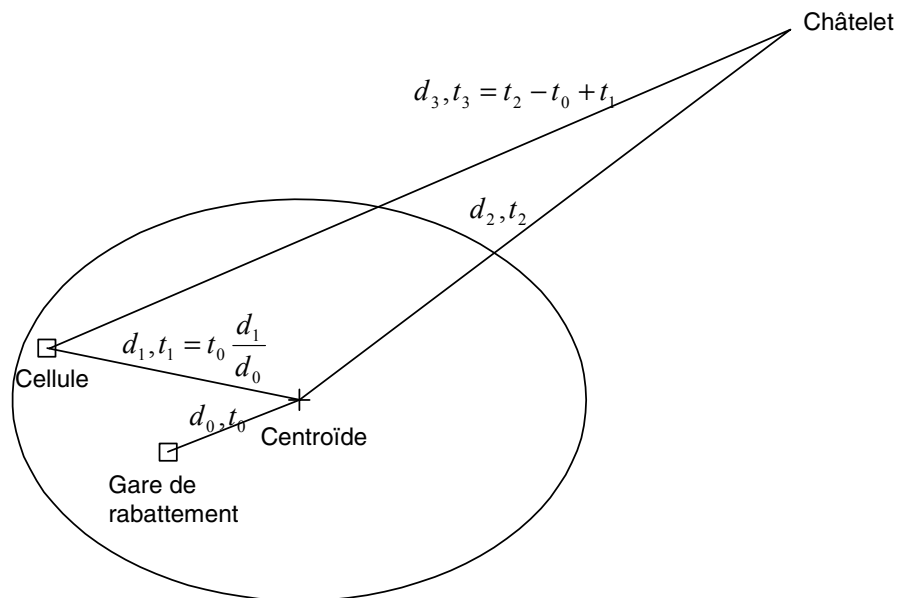


Figure 2 - temps de parcours en TC

3.3. Variable distance à Châtelet-les-Halles

Un calcul de distance en coordonnées Lambert II a été effectué. Chaque cellule est caractérisée par sa distance en kilomètres à Châtelet-les-Halles.

³ Cette gare de rabattement est choisie en fonction de sa proximité avec le centroïde et des horaires des trains.

⁴ Le temps de transport total t_2 est la somme du temps de transport pour aller du centroïde à la gare de rabattement et du temps de transport de la gare de rabattement à Châtelet.

3.4. Variable de contraintes environnementales

Les contraintes environnementales de type « open_space » et « public_space » font référence à des concepts typiquement américains⁵ qui n'ont pas à être pris en compte dans le cas de la région d'Ile-de-France. On a décidé de les remplacer par deux contraintes plus conformes au cas français : la contrainte des sites inscrits et classés et la contrainte des zones de bruit aéroportuaires.

3.4.1. Les sites inscrits et classés

Le SIGR dispose de deux couches qui inventorient

- d'une part les sites et monuments historiques en Ile-de-France et leurs périmètres de protection (rayon de 500 m) qui sont définis par la loi du 31/12/1913 pour les monuments historiques et la loi du 02/05/1930 pour les sites
- d'autre part les autres types de protection recensés sont les Zones de Protection du Patrimoine Architectural Urbain et Paysager (ZPPAUP, loi du 07/01/1993).

Les couches correspondantes sont des couvertures de polygones représentant les périmètres de protection des monuments inscrits (BUFMONUI, 647 polygones) et les périmètres de protection des monuments classés (BUFMONUCL, 458 polygones).

Ces couches sont intersectées avec le maillage. La couche résultante fournit pour chaque cellule la surface occupée par un périmètre protégé. Si cette surface occupe plus de 50% de la cellule, la cellule est considérée comme figée ou non urbanisable.

3.4.2. Les zones de bruit aéroportuaires

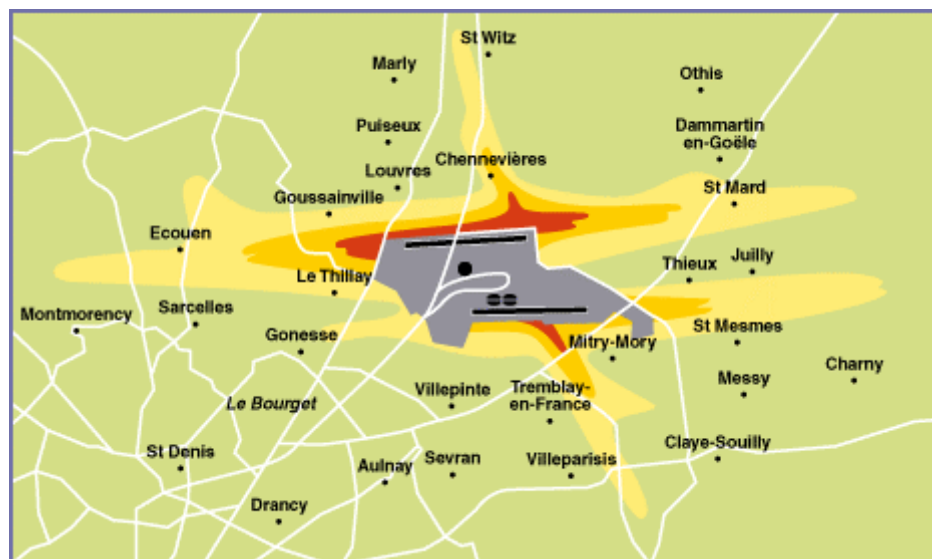
Le Plan d'Exposition au Bruit est un document d'urbanisme créé dans le but d'éviter d'exposer de nouvelles populations aux nuisances de bruit liées aux aéroports. Il intègre les prévisions de trafic aérien à un horizon de 10 à 15 ans et est donc destiné à maîtriser l'urbanisation autour de Roissy et Orly. L'actuel PEB a été élaboré en 1989 et est encore en vigueur, même s'il faut savoir que le nouveau PEB est en cours de révision.

Il y a trois niveaux de zones de bruit :

- En zones de bruit fort ou intense (zone 96 et plus), il est interdit de construire ou de densifier l'habitat existant.
- En zone de bruit sensible (zone 89-96), seules peuvent être autorisées, sous certaines conditions, des constructions individuelles non groupées ainsi que des logements collectifs.
- Enfin, dans la dernière zone (zone 78-89), toute construction sera autorisée à condition d'être isolée acoustiquement.

⁵ Le concept d'« open space » s'oppose à « developed space ». Les urbanistes américains sont confrontés au problème de préservation de l'open space, sachant que l'urbanisation se fait au détriment de cet espace.

En revanche, ces différentes zones peuvent accueillir les équipements liés à l'activité aéronautique et les emplois correspondants moyennant une isolation acoustique adaptée. Le Plan d'Exposition au Bruit est annexé au plan d'occupation des sols des communes concernées.



Source : ADP - 1989

En jaune : zone 78-89
En orange : zone 89-96
En rouge : zone > 96

Figure 3 - PEB de Roissy-CDG

Ces zones de bruit sont représentées dans une couche du SIGR. On a retenu la « zone 96 et plus » et intersecté avec le maillage. Les cellules à l'intérieur de cette zone sont considérées comme non urbanisables.

3.5. Variables de contraintes politiques locales

Elaboré par les communes, le Plan Local d'Urbanisme (PLU), anciennement appelé Plan d'Occupation des Sols (POS), est un document d'urbanisme et de planification. Il définit les orientations d'urbanisme de la commune. L'IAURIF dispose de la base de données numérisées des POS/PLU appelée BD-POS et gérée par la DREIF sur l'ensemble de la région. La BD-POS est une représentation numérique du zonage des POS et PLU établie à partir des documents approuvés conservés par les Directions Départementales de l'Équipement.

La couverture dans le SIGR s'appelle Parpos. Chaque parcelle comporte un attribut appelé « libellé abrégé ». S'il s'agit d'un POS, les modalités possibles sont U, NA, NAa, NB, NC, ND, ZAC, infrastructure, autre. S'il s'agit d'un PLU, les modalités possibles sont U, AU, A, N, autre.

Rappel : Le POS définit de manière détaillée les usages du sol à l'échelle de la commune et précise les droits de construire affectés à chaque parcelle pour une période de cinq ans. On distingue deux grandes familles d'affectation des sols : les zones urbaines (U) et les zones naturelles (N). Parmi les zones naturelles sont distinguées : des zones NA, zones naturelles d'urbanisation future sur lesquelles il est prévu que la construction de logements collectifs sera autorisée, et des zones NB sur lesquelles sera autorisée la construction de logements

individuels. Par ailleurs, chaque parcelle se voit attribuer un coefficient d'occupation du sol (COS), coefficient d'autorisation de densité dont le calcul intègre la surface du bâti et le nombre d'étages. Le POS est opposable au tiers.

On effectue l'intersection entre Parpos et le maillage. Puis pour chaque cellule, on a calculé la surface affectée à chacune des modalités.

Cette couche de données est extrêmement puissante. Elle permettra de tenir compte des zones non urbanisables à un niveau infracommunal.

Il existe aussi d'autres documents territoriaux de planification spatiale, créés au cours des 10 dernières années dans le cadre de politiques nationales. Leur mise en œuvre appelle des prescriptions territorialisées. Ainsi, de nombreuses intercommunalités avec des compétences en matière d'urbanisme et d'aménagement ont vu le jour.

On peut citer entre autres :

- la Politique de la Ville : Zone de Redynamisation Urbaine (ZRU), Zone Franche Urbaine (ZFU), Grande Politique de la Ville (GPV) ...
- la politique d'habitat : le prêt à taux 0 qui encourage la péri-urbanisation en incitant à l'achat de maisons individuelles (70% des logements financés en grande couronne)

Ces politiques témoignent de l'importance de la planification urbaine en France et en particulier en Ile-de-France. L'urbanisation n'est pas purement mécanique, elle n'est pas laissée au libre marché mais dépend fortement des politiques locales, régionales et gouvernementales. Il suffit de se rappeler de quelques exemples historiques :

- la politique urbaine menée à Paris sous l'impulsion de Napoléon III et du baron Haussmann dans la moitié du XIX^{ème} siècle
- la politique de construction des « grands ensembles » après la Seconde Guerre Mondiale
- la politique de création des Villes nouvelles dans les années 60
- la politique de décentralisation depuis le début des années 80

Plus récemment (octobre 2005), le Gouvernement a sélectionné trois sites prioritaires pour accueillir des « opérations d'intérêt national » (OIN) dans le cadre de son projet de loi sur l'engagement national pour le logement. Ces trois sites (Seine Amont -94, Seine Aval -78, Massy-Saclay -91) devraient pouvoir accueillir plusieurs milliers de logements nouveaux. Le Préfet d'Ile-de-France pourra alors prendre le pilotage des opérations et se substituer au maire pour délivrer des permis de construire. Autre exemple récent concernant les emplois: deux « Pôles de compétitivité » en Ile-de-France ont été lancés en juillet 2005 par le Gouvernement. Ils ont pour objectif de favoriser la croissance par l'innovation, lutter contre les délocalisations, favoriser l'emploi, accroître l'attractivité du territoire. Ces deux pôles concernent d'une part les neurosciences sur Paris et ses alentours, d'autre part les logiciels et systèmes informatiques complexes sur le plateau de Saclay (Essonne). Ces exemples montrent clairement que l'on atteint vite les limites de la modélisation dans un contexte de forte influence des pouvoirs publics, aussi bien centralisés que locaux.

3.6. Variable development_type

Par défaut, UrbanSim propose 25 types urbains différents, allant du résidentiel bas à la zone industrielle.

A chaque cellule est attribué un type urbain. Un type est fonction des variables tels que le nombre de logements, la surface au sol, l'espace non résidentiel, le % d'espace non urbanisable et la mixité. Pour attribuer le type 25 (non urbanisable), une autre formule a été écrite, en fonction des pourcentages d'eau, de zones humides, de zones inondées, de pente et d'espace public.

codes	Libellés
1	BOIS
2	GRANDES CULTURES
3	AUTRES CULTURES
4	EAU
5	AUTRE RURAL
6	PARCS ET JARDINS
7	SPORT (ESPACES OUVERTS)
8	TERRAINS VACANTS
9	HABITAT INDIVIDUEL
10	HABITAT COLLECTIF
11	HABITAT AUTRE
12	ACTIVITES SECONDAIRES
13	ACTIVITES TERTIAIRES
14	SPORT (CONSTRUIT)
15	EQUIPEMENT D'ENSEIGNEMENT
16	EQUIPEMENT DE SANTE
17	CIMETIERES
18	AUTRES EQUIPEMENTS LOCAUX, ADMINISTRATION
19	GRANDS EQUIPEMENTS
20	TRANSPORTS
21	CHANTIERS

Tableau 3 - 21 postes du MOS

Après avoir réfléchi avec des urbanistes de l'IAURIF et compte tenu des données disponibles, il a été décidé de ne pas garder la formule par défaut ni de l'adapter, mais de définir les types urbains en fonction de l'occupation du sol (« land-use ») indépendamment de la 3^{ème} dimension. Il nous a semblé que la définition de 25 types différents était inappropriée dans un maillage en 50 000 cellules. Cela signifierait le calcul d'une matrice de transition du type urbain de dimension 25*25 entre 1990 et 1999, donc on se retrouverait avec une matrice creuse compte tenu de la période relativement courte de 10 ans. Remarquons que sur le cas Puget Sound, aucune donnée chiffrée n'est donnée à propos de la matrice de transition en 25 types. Peu de crédit doit donc être accordé à ce qui a été fait sur ce cas américain.

L'avantage dans notre projet est que nous disposons de l'inventaire exhaustif et détaillé du Mode d'Occupation du Sol (MOS) de 1990 et 1999 en 21 postes sur l'ensemble de la région.

Ainsi, le modèle de développement immobilier d'UrbanSim (« development type model ») sera adapté à notre cas : ce sera en fait un modèle de l'évolution du MOS.

La variable `development_type` a été construite en deux étapes.

La *première étape* a consisté à effectuer un regroupement de ces 21 postes et à définir une typologie en 5 types urbains agrégés :

- **Type 1 « habitat »** = pourcentage de la surface de la cellule occupée par de l'habitat

où surface habitat = surface habitat individuel (poste 9) + surface habitat collectif (poste 10) + surface habitat autre (poste 11)

- **Type 2 « activités & équipements »** = pourcentage de la surface de la cellule occupée par 'activités & équipements'

où surface activités & équipements = surface activités secondaires (poste 12) + surface activités tertiaires (poste 13) + surface sport (construit) (poste 14) + surface équipement d'enseignement (poste 15) + surface équipement de santé (poste 16) + surface autres équipements locaux, administration (poste 18) + surface grands équipements (poste 19)

- **Type 3 « espaces publics couverts »** = pourcentage de la surface de la cellule occupée par 'espaces publics couverts'

où surface espaces publics couverts = surface parcs et jardins (poste 6) + surface sport (espaces ouverts) (poste 7) + surface cimetière (poste 17)

- **Type 4 « réseaux »** = pourcentage de la surface de la cellule occupée par les réseaux de transport et les réseaux fluviaux

où surface réseaux = surface eau (poste 4) + surface transports (poste 20)

- **Type 5 « vacant »** = pourcentage de la surface de la cellule considérée comme vacant

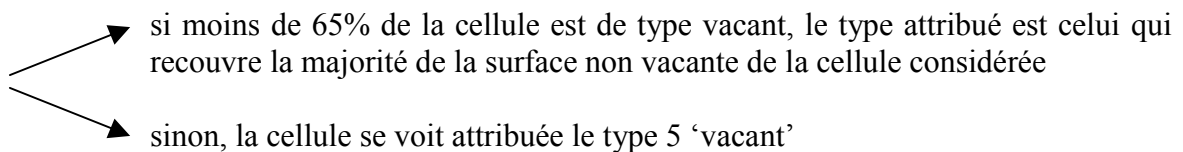
où surface vacant = surface bois (poste 1) + surface grandes cultures (poste 2) + surface autres cultures (poste 3) + surface autre rural (poste 5) + surface terrains vacants (poste 8) + surface chantiers (poste 21)

Une *deuxième étape* a consisté à déterminer une règle d'affectation d'un type à chaque cellule. Plusieurs possibilités ont été envisagées. La plus simple consiste à attribuer à chaque cellule le type qui recouvre le maximum de la surface de la cellule considérée. Ainsi par exemple, une cellule dont la surface est composée à 60% par le type 1 'habitat' et à 40% par le type 4 'réseaux' est considérée comme étant de type 1. Le problème de cette règle du maximum est que certaines cellules se voient attribuées en 1999 le type 'vacant' alors qu'elles ont pu connaître entre 1990 et 1999 un début d'urbanisation. Par exemple, une cellule dont la surface est composée en 1999 à 60% par le type 5 'vacant' et à 40% par le type 1 'habitat' serait toujours classée comme vacante alors qu'elle est déjà fortement urbanisée.

Cette règle du maximum dépend aussi de la taille des cellules. Avec un maillage plus fin, une zone urbanisée nouvelle aurait pu être prise en compte. En conclusion, cette règle du maximum ne permet donc pas de tenir compte d'un début d'urbanisation. C'est pourquoi une autre règle d'attribution a été retenue. Elle a pour but d'augmenter le nombre de cellules qui ont changé de type entre 1990 et 1999 (voir statistiques descriptives en 5.2.3).

Un type urbain est alors affecté de la façon suivante :

- si le type attribué à l'aide de la règle du maximum est différent du type 5 'vacant', alors c'est ce type qui est attribué à une cellule
- sinon la règle de décision suivante est appliquée :



3.7. Variable Politique de la Ville

On a souhaité identifier les secteurs à l'intérieur des communes qui présentent un niveau social peu élevé. Il est clair qu'un ménage aisé ne choisira pas ces secteurs comme lieu de résidence. Mais pour toute commune, même celle ayant un déficit d'image, il y a des zones résidentielles assez recherchées. Par exemple, à Aulnay-sous-Bois (93), le sud de la commune est plutôt résidentiel. Le SIGR de l'IAURIF a l'avantage de disposer de la couche numérisée des Zones Urbaines Sensibles (ZUS), Zones Franches Urbaines (ZFU) et Zones de Redynamisation Urbaine (ZRU). Ces zones ont été définies dans le cadre de la Politique de la Ville et sont caractérisées par un niveau social peu élevé et un déficit d'image. Cette couche de données est très intéressante car elle localise à un niveau infracommunal les zones dites sensibles, ce qui évite d'assimiler toute la commune à une commune « sensible ». Dans toutes communes, y compris celles ayant un déficit d'image, il existe des zones résidentielles recherchées. Par exemple, à Aulnay-sous-Bois, le sud de la commune est pavillonnaire et loin de l'image quelque peu dégradée que l'on peut avoir à tort de la commune.

Cette couche a donc été croisée avec le maillage en cellules carrées. Chaque cellule comporte alors la part de surface affectée aux ZUS+ZFU+ZRU. Si cette part est supérieure ou égale à 20% (par analogie avec le taux minimum de logements sociaux exigé imposé par la loi SRU), on attribue à la cellule le type 1 (déficit d'image, niveau social peu élevé), sinon 0. Après traitement, il y a 568 cellules codées 1, soit 1,15% des cellules.

3.8. Variables de voisinage

L'idée ici est similaire à celle du jeu de la vie de Conway (1970) (voir encadré) : l'évolution et l'attractivité d'une cellule tant en emplois qu'en habitation dépendent de celles des cellules qui lui sont voisines.

Le jeu de la vie

Le jeu de la vie est un automate cellulaire dont les règles ont été définies par J. Conway en 1970. L'état de l'automate à l'étape n est uniquement fonction de son état à l'étape $n-1$. L'évolution de l'état d'une cellule dépend de l'état de ses 8 plus proches voisins.

Dans l'automate de Conway, les règles sont les suivantes :

- une cellule vide à l'étape $n-1$ et ayant exactement 3 voisins sera occupée à l'étape suivante (naissance liée à un environnement optimal)
- une cellule occupée à l'étape $n-1$ et ayant 2 ou 3 voisins sera maintenue à l'étape n sinon elle est vidée (destruction par désertification ou surpopulation)

C'est l'analogie entre ces règles et certains critères d'évolution de populations de bactéries qui a conduit à donner à cet automate le nom de jeu de la vie. Tobler en 1979 a été le premier à suggérer d'appliquer la théorie des automates cellulaires à la dynamique spatiale. Aujourd'hui, elle semble tout à fait applicable grâce à l'avènement des systèmes d'information géographique.

Pour chacune des 49236 cellules, différentes variables de voisinage ont été calculées. Par exemple, la variable de voisinage associée à la variable 'surface commerciale' est la somme de la surface commerciale des 8 cellules adjacentes à la cellule considérée et de la surface commerciale de celle-ci :

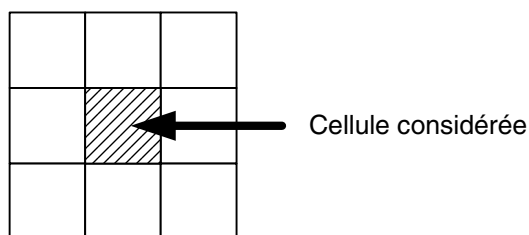


Figure 4 -les cellules

Le schéma ci-dessous permet de comprendre comment les variables de voisinage associées à différentes cellules se recouvrent :

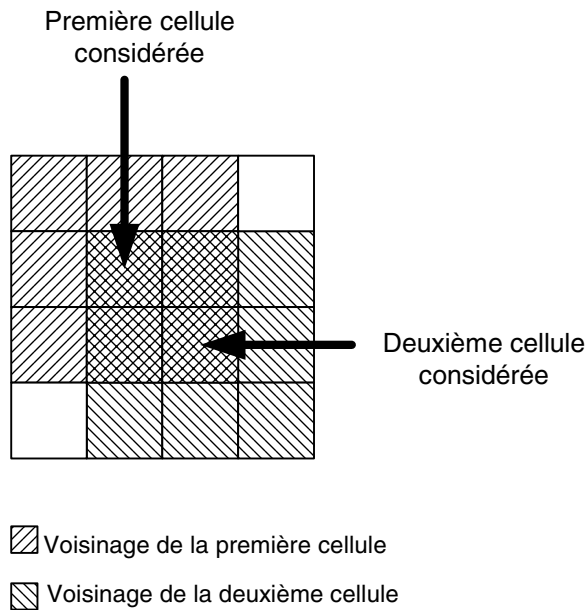


Figure 5 - calcul des variables de voisinage pour les cellules

Les variables de voisinage associées aux variables suivantes ont été calculées :

- 'residential_improvement_value' = valeur foncière du bâti résidentiel,
- 'residential_land_value' = valeur foncière du terrain,
- 'commercial_sqft' = surfaces de commerce,
- 'industrial_sqft' = surfaces d'activités secondaires,
- 'governmental_sqft' = surfaces d'activités tertiaires (bureaux),
- 'residential_units' = nombre de logements,
- 'commercial_improvement_value' = loyers des surfaces commerciales,
- 'industrial_improvement_value' = loyers des surfaces d'activités secondaires,
- 'governmental_improvement_value' = loyers des bureaux,
- 'nonresidential_land_value' = valeur foncière totale du non résidentiel,
- 'nb_nodes_subway' = nombre de stations de métro,
- 'nb_nodes_train' = nombre de gares,
- 'nb_nodes_tramway' = nombre d'arrêts de tramway.

Ces variables ont été calculées à partir d'un programme écrit sous Scilab qui est donné en annexe 1. A ce stade, toutes les variables caractéristiques d'une cellule ont été calculées. Pour attribuer un nom aux variables de voisinage, on a simplement adjoint un « v » à la première lettre du nom de la variable de la cellule. Par exemple, la variable de voisinage du nombre de logements est appelée « vresidential_units » (ou encore « vresu » dans le modèle de localisation des ménages). Les tables gridcells de 1990 et 1999 sont prêtes.

4. Recalage du modèle de demande de l'IAURIF et de METROPOLIS

4.1. Les réseaux

Il est prévu de réaliser le bouclage entre UrbanSim et METROPOLIS sur un pas de trois ans. Le calage du modèle de trafic (demande et affectation) est effectué sur la période 1990 – 1999, avec deux points intermédiaires : 1993 et 1996.

En simulation, le modèle de trafic tournera tous les trois ans à partir de 1999 jusqu'à l'année 2026, UrbanSim tournera chaque année.

4.1.1. Routes

Au 15 septembre 2005, les graphes des réseaux routiers de quatre années (1990, 1999, 2005 et 2026) ont été mis à jour pour être importés dans METROPOLIS. Cette mise à jour est réalisée d'abord sur le logiciel ArcGis, puis les graphes (arcs, nœuds, mouvements tournants) sont mis au format texte adéquat pour être lus par METROPOLIS. Afin de garder une cohérence entre les identifiants des arcs entre chaque horizon, la méthode a consisté à partir d'un réseau routier maximal en 2026 puis à supprimer au fur et à mesure des arcs ou modifier des caractéristiques d'arcs pour construire les réseaux des années 2005, puis 1999, puis 1990.

Réseau routier 2026 :

L'IAURIF dispose dans son propre modèle de trafic du réseau à l'horizon 2015 qui était celui inscrit au SDRIF de 1994. Ce graphe comporte environ 16700 arcs monodirectionnels et 5330 nœuds. Les mouvements interdits sont décrits dans un fichier texte.

Nous avons décidé de retenir pour 2026 tous les projets inscrits au SDRIF 1994, à l'exception de :

- réseau sous-terrain (MUSE, périphérique sud en tunnel, la voie Roissy Express)
- 3 infrastructures qui ont été officiellement abandonnées : le prolongement d'A16 de la Francilienne à A86, la voie rapide A6-RN6 dans l'Essonne et le prolongement d'A103 en Seine-et-Marne.

Tous les élargissements prévus au SDRIF de 1994 sont pris en compte.

Il est clair que le réseau retenu n'est pas réaliste. Cependant, d'un point de vue pratique, il est plus facile de supprimer des arcs que d'en rajouter. De plus, même si la révision du SDRIF est en cours, le SDRIF de 1994 est toujours en vigueur à la date du 15 septembre 2005. Lorsque l'on en sera au stade de construction des scénarios, ce réseau 2026 sera étudié et modifié de nouveau.



Figure 6 - Réseau routier du SDRIF de 1994

Réseau routier 2005 :

C'est le réseau actuel. Il est donc parfaitement connu. En partant du réseau 2026, il a suffi de supprimer tous les arcs qui n'existent pas aujourd'hui et requalifier des arcs aux caractéristiques actuelles (par exemple, le boulevard des Maréchaux sud mis à 2*1 voies à cause des travaux sur le tramway des Maréchaux T3, et le mouvement tournant A86->A6 vers Paris est autorisé).

Réseau routier 1999 :

Les arcs suivants ont été supprimés ou requalifiés par rapport au réseau 2005 :

- suppression des extrémités est et ouest de la Francilienne au nord
- mise à 2*3 voies du boulevard des Maréchaux sud
- suppression de la déviation de Jours-Ponchartrin
- mise à 2*3 voies de la RN6 à Maisons-Alfort
- suppression du prolongement d'A115 à Taverny
- suppression du Boulevard Intercommunal du Paris (BIP)
- mise à 2*2 voies de l'A86 au niveau du cimetière de Thiais
- mise à 2*2 voies de la Francilienne à Evry
- mise à 2*2 voies de l'A13
- suppression du mouvement tournant A86->A6 vers Paris

Réseau routier 1990 :

Les arcs suivants ont été supprimés ou requalifiés par rapport au réseau 1999 :

- suppression de la RD60 à Créteil
- suppression de la totalité de la Francilienne au nord

- mise à 2*1 voies de la RD307 à Bailly/Rocquencourt
- suppression du tronçon d'A86 à Drancy
- suppression de l'A14
- suppression du tronçon d'A86 à Rueil-Malmaison
- suppression du BIP ouest
- suppression de la RN184 à Saint-Ouen-l'Aumône
- suppression de l'A5
- suppression de l'A86 à la Croix de Berny
- mise à 2*3 voies de l'A12
- suppression de l'arrivée d'A16 au nord de l'Ile-de-France
- suppression de la déviation de Houdan
- suppression de la liaison A86-A1
- suppression de la rocade du parc Disneyland Paris

Il peut subsister encore des inexactitudes dans la construction de ces graphes. La mise à jour des réseaux routiers à différents horizons est une tâche permanente.

4.1.2. Transports en commun

Réseau TC 2026 :

Il comporte l'ensemble des projets inscrits au SDRIF de 1994, dont Orbitale et les Tangentielles ferrées. Là aussi, comme pour le réseau routier, il est préférable de garder un réseau exhaustif, quitte à supprimer des projets lors de la phase de définition des scénarios.

Le graphe comprend le réseau lourd (lignes SNCF et RATP) ainsi que l'ensemble du réseau de bus RATP et les principales lignes de bus OPTILE (222 lignes). L'ensemble est constitué d'environ 5200 arcs bidirectionnels et 4200 noeuds, auxquels s'ajoutent 1500 arcs de correspondance. Il y a sur le réseau lourd 230 missions à l'heure de pointe, chaque mission étant caractérisée par une fréquence et des temps de parcours inter-gares.

Réseau TC 2005 :

C'est le réseau actuel.

Réseau TC 1999 :

On supprime par rapport au réseau 2005 :

- EOLE, en particulier les gares Magenta et Haussman, mais on rétablit la ligne train de banlieue Paris Est
- La gare de Liesses
- La ligne La Verrière – La Défense

Réseau TC 1990 :

On supprime par rapport au réseau 1999 :

- les deux stations de la ligne 1 du métro à la Défense
- la station Saint-Denis-Université
- METEOR
- OrlyVal
- Le tramway T1
- Le TVM
- Le tramway T2

- La gare de Saint-Denis (RER D)
- Les gares de la Plaine Voyageurs-Stade de France (RER B) et de Bibliothèque-François-Mitterrand, ex Boulevard Masséna (RER C) sont redéplacées
- Les gares de Chessy, Val d'Europe, Bussy-Saint-Georges, Cergy-le-Haut et Neuville Université sur le RER A
- La gare Roissy-CDG2

4.2. Les 3 étapes du modèle de demande

Le modèle de demande de l'IAURIF en 3 étapes (génération, distribution, choix modal) a été décrit de manière théorique dans le rapport n°1 (pages 52 à 63 et annexe 3 pages 131-132).

Cette partie décrit les simplifications opérées par rapport à la description théorique et les résultats du calage du modèle de demande à partir de l'EGT 2001-2002.

Trois motifs de déplacements ont été retenus :

- Domicile → Travail et Affaires professionnelles
- Domicile → Achats hebdomadaires et exceptionnels
- Autres motifs

Le choix de ces trois motifs est cohérent avec les données requises par UrbanSim par défaut. Comme il n'y a pas de données sur les effectifs scolaires et les attracteurs de type loisirs dans UrbanSim, on n'a pas distingué les motifs Ecole et Loisirs.

La période 6h – 9h59 (donc 4 heures le matin) a été choisie. C'est l'heure d'arrivée du déplacement (et non l'heure de départ).

4.2.1. Génération

Le principe est le suivant : on calcule des taux de mobilité par motif, par couronne et par émission ou attraction, pour l'ensemble des modes VP et TC. Un taux de mobilité est le rapport d'une part entre le nombre total des déplacements effectués entre 6h et 9H59 par motif et par couronne d'origine ou de destination (Paris, Petite couronne, Grande couronne) calculé dans l'EGT 2001-2002, et d'autre part entre les variables émetteurs ou attracteurs du déplacement :

- Domicile → Travail et Affaires professionnelles
Emission : Nombre d'actifs
Attraction : Nombre d'emplois
- Domicile → Achats hebdomadaires et exceptionnels
Emission : Population
Attraction : Surfaces de vente (de plus de 1000 m2)
- Autres motifs
Emission : Population + Emplois
Attraction : Population + Emplois

Nous obtenons donc 3*3*2 coefficients appelés taux de mobilité sur la période de pointe du matin :

Taux de mobilité pour Emission D->T+Aff pro sur la période de pointe par couronne

P	PC	GC
0,45	0,47	0,50

Taux de mobilité pour Attraction D->T+Aff pro sur la période de pointe par couronne

P	PC	GC
0,53	0,55	0,52

Taux de mobilité pour Emission D->Achats sur la période de pointe par couronne

P	PC	GC
0,0051	0,0102	0,0183

Taux de mobilité pour Attraction D->Achats sur la période de pointe par couronne

P	PC	GC
0,0090	0,0142	0,0191

Taux de mobilité pour Emission T+aff pro+Achats -> D et Reste -> Reste sur la période de pointe par couronne

P	PC	GC
0,0999	0,1681	0,2297

Taux de mobilité pour Attraction T+aff pro+Achats -> D et Reste -> Reste sur la période de pointe par couronne

P	PC	GC
0,1432	0,1560	0,2139

Puis pour chacune des 572 zones internes du modèle de trafic en 606 zones (572 zones internes+34 zones externes), le taux correspondant est multiplié par les variables données ci-dessus et par couronne pour obtenir l'émission et l'attraction de chaque zone. Pour les 34 zones externes, un calcul différent est effectué et sera décrit dans le rapport final.

On obtient donc deux fichiers :

- émissions par zone et par motif
- attractions par zone et par motif

4.2.2. Distribution

La matrice de référence motif origine - motif destination de déplacements de l'EGT 2001-2002 pour l'ensemble des modes VP et TC sur la période 6h – 9h59 est la suivante :

O/D	VP+TC			
	D	T+aff Pro	Achats	Autres
D		2 673 638	145 431	
T+aff Pro	54 590			
Achats	31 181		2 789 081	
Autres				

5 693 921

La distribution est calculée par motif de déplacement et pour l'ensemble des modes VP et TC.

Contrairement à ce qui a été décrit dans le rapport n°1, on a simplifié la forme de la fonction de distribution. Elle n'est pas de la forme $f(d_{ij}) = c_{ij}^\alpha \exp(-\beta c_{ij})$ mais de la forme $f(d_{ij}) = 1 / c_{ij}^\alpha$, qui est aussi une fonction décroissante.

Il s'agit alors de caler ce paramètre α pour chacun des 3 motifs. Ce paramètre sera réutilisé dans la formule de l'accessibilité (paragraphe 5.3.3) pour le modèle de localisation résidentielle des ménages.

Pour cela, on procède à une exploitation spécifique de l'EGT 2001-2002 en analysant la distribution des déplacements par tranche de temps pendant la période 6h – 9h59. Le choix des tranches de temps se fait de manière empirique.

Le coût choisi est le temps de déplacement à vide en VP calculé par METROPOLIS.

L'ajustement des nuages de points par des fonctions polynomiales inverses donne les résultats suivants :

- Domicile → Travail et Affaires professionnelles

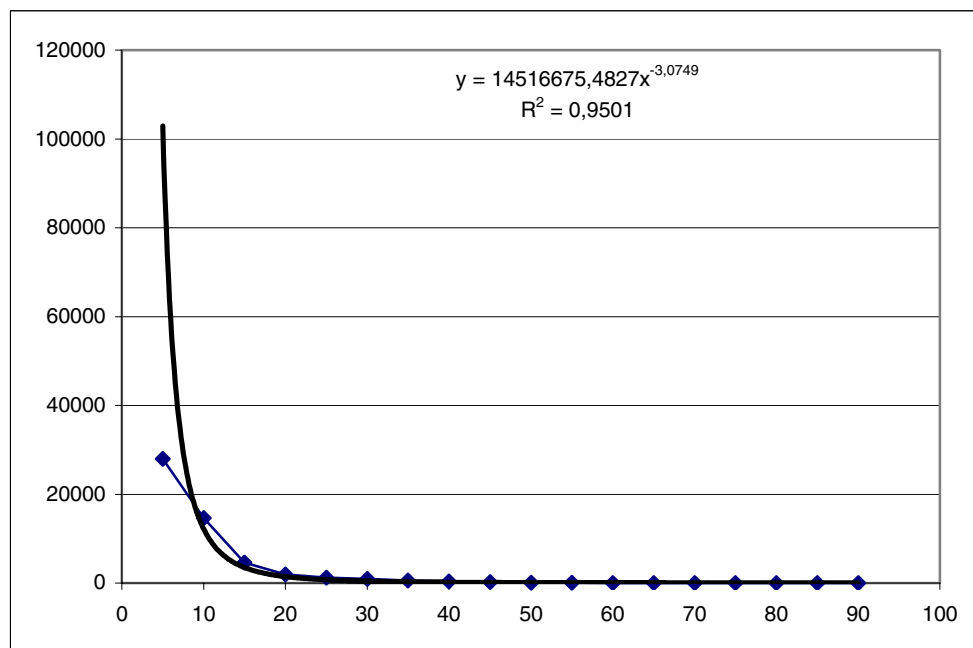


Figure 7 - distribution des temps pour le motif Domicile – Travail+Aff Pro

$$f(t_{ij}) = 14516675 * t_{ij}^{-3,075}$$

Dans le programme informatique de la distribution, on a imposé un maximum à 35000 pour ce motif, car le pic à 5 minutes de la fonction est trop élevé par rapport à la valeur réelle.

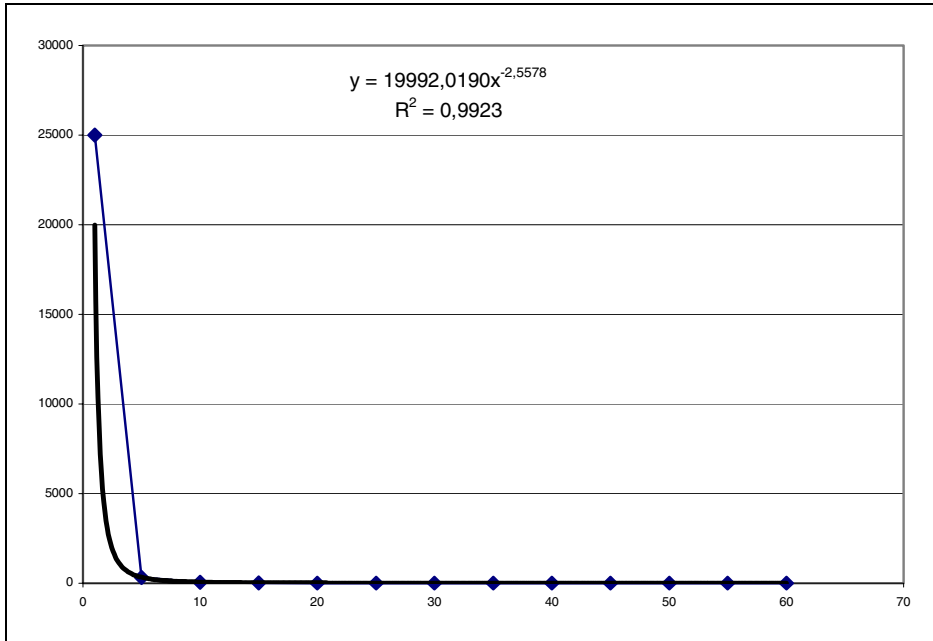


Figure 8 - distribution des temps pour le motif Domicile - Achats

- Domicile → Achat

$$f(t_{ij}) = 19992 * t_{ij}^{-2,558}$$

- Autres motifs

$$f(t_{ij}) = 82684 * t_{ij}^{-2,338}$$

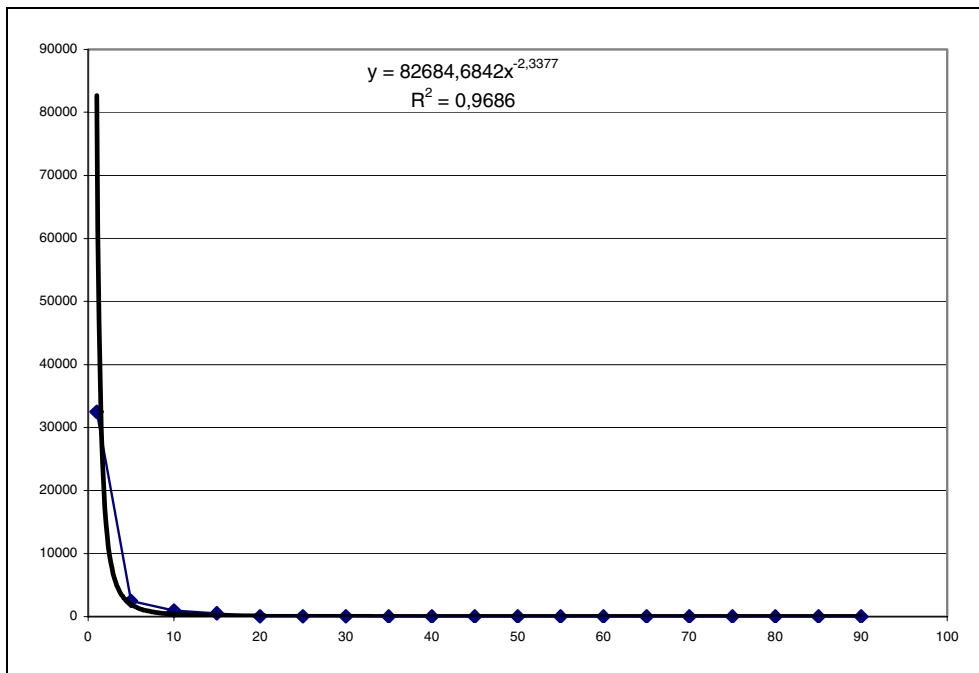


Figure 9 - distribution des temps pour les autres motifs

Deux indices permettent de contrôler les résultats de la distribution :

- le nombre des déplacements intra-zonaux par motif, que nous comparons au nombre de déplacements intra-communaux calculé par l'EGT
- la matrice des déplacements entre les 8 départements, que nous comparons avec la matrice 8*8 issue de l'EGT

Les intrazonaux, comparé aux intracommunaux

	Intrazonaux calculés	Intracommunaux EGT
D->T+Aff pro	209 000	211 000
D->Achats	31 000	58 000
Autres	890 000	949 000
Total	1 120 000	1 218 000

Tableau 4 - comparaison intrazonaux - intracommunaux

La matrice de référence département à département tous motifs

EGT 2001	6h à 9h59			VP+TC	Tous motifs				
	75	93	94	92	95	78	91	77	
O/D	75	93	94	92	95	78	91	77	
75	570	64	44	125	12	26	16	10	867
93	144	361	34	34	17	6	8	28	632
94	140	34	315	55	3	5	19	9	581
92	195	28	19	373	8	39	20	5	689
95	79	45	5	57	336	26	3	3	554
78	69	11	9	87	20	521	22	1	739
91	68	8	33	43	3	15	468	6	644
77	66	54	28	17	6	4	23	463	661
	1 332	605	488	791	405	641	579	526	5 367

Modèle	6h à 9h59			VP+TC	Tous motifs				
	75	93	94	92	95	78	91	77	
O/D	75	93	94	92	95	78	91	77	
75	666	51	40	87	8	9	6	4	870
93	145	302	48	43	46	7	5	28	624
94	134	47	278	40	7	9	27	22	565
92	175	25	30	390	23	47	21	5	716
95	66	57	9	66	353	34	3	10	597
78	61	11	11	96	35	501	19	6	740
91	60	10	52	41	4	25	399	27	618
77	52	44	40	21	16	10	36	419	638
	1 359	546	508	783	493	641	517	521	5 367

Ratios OVD	6h à 9h59			VP+TC	Tous motifs				
	75	93	94	92	95	78	91	77	
75	0,86	1,26	1,11	1,43	1,49	2,99	2,68	2,81	1,00
93	0,99	1,20	0,72	0,79	0,38	0,78	1,41	1,02	1,01
94	1,05	0,73	1,13	1,39	0,47	0,60	0,70	0,39	1,03
92	1,11	1,16	0,65	0,96	0,34	0,84	0,97	1,03	0,96
95	1,21	0,79	0,59	0,86	0,95	0,78	0,89	0,31	0,93
78	1,13	0,97	0,81	0,91	0,57	1,04	1,15	0,21	1,00
91	1,13	0,80	0,63	1,04	0,74	0,60	1,17	0,23	1,04
77	1,27	1,21	0,70	0,84	0,38	0,36	0,62	1,11	1,03
	0,98	1,11	0,96	1,01	0,82	1,00	1,12	1,01	1,00

La comparaison entre les intrazonaux et les intracommunaux montre une concordance acceptable. De même, la matrice de ratios entre la matrice de référence de l'EGT et la matrice simulée montre des coefficients proches de 1 pour les grands flux. On peut donc considérer que le calage de la distribution est satisfaisant.

4.2.3. Choix modal

Le principe du calage est le suivant. Pour chacun des trois motifs, une base de calage est construite. Une base est constituée d'observations tirées de l'EGT 2001-2002. Ces observations sont des déplacements motorisés (VP ou TC) effectués pendant la période du matin choisie : 6h-9h59 (heure de départ). Chaque déplacement comporte comme attributs :

- la zone origine (dans le découpage du modèle en 606 zones)
- la zone destination
- le mode choisi
- le temps de parcours

Nous avons choisi d'élaborer un modèle de choix modal le plus simple possible, ce qui explique que seules les variables de temps de parcours et de localisation des zones origine et destination par couronne sont prises en compte.

Il faut ensuite attribuer à chaque déplacement le temps du mode alternatif à la période de pointe. Ce temps est calculé par le modèle multimodal de trafic de l'IAURIF avec d'anciennes matrices calées en 1996.

Théoriquement, il faudrait boucler l'étape de choix modal avec l'affectation. Mais en pratique, on ne le fera pas car cela dépasse le cadre du projet.

Un autre problème se pose : il concerne la variable DDU (temps déclaré). Pour une même origine-destination, il existe une forte disparité des temps déclarés, qui s'explique par trois facteurs :

- un temps déclaré est un temps perçu et non réel, dépendant des caractéristiques de l'individu.

- le temps déclaré par un individu comprend le temps de rabattement et le temps de diffusion : ces deux temps introduisent une grande variabilité dans le temps total pour une même origine-destination car ils sont en réalité particuliers à chaque déplacement. Les temps de parcours modélisés sont quant à eux des *temps moyens* de centroïde à centroïde.
- Les temps déclarés que l'on a retenus sont compris entre 6h00 et 9h59. Sur les flancs de cette période de 4 heures, la circulation pour le mode VP est moins dense que en pleine heure de pointe. Ceci introduit donc une variabilité des temps déclarés pour une même origine-destination.

Exemples flagrants extraits de l'EGT 2001-2002 illustrant ce problème des temps déclarés à l'heure de pointe du matin :

- Pavillons-sous-Bois vers Paris 1^{er} **en VP** : 45 mn (contre 97 mn en temps simulé)
- Meudon vers Paris 15^{ème} (quartier Montparnasse⁶) **en VP** : 7 mn (contre 38 mn en temps simulé)
- Paris 12^{ème} vers Paris 15^{ème} **en TC** : 75 mn (contre 45 mn en temps simulé)
- Saint-Denis vers Paris 16^{ème} **en TC** : 85 mn (contre 47 mn en temps simulé)

Les temps simulés sont des temps moyens de zone à zone alors que les temps déclarés sont des temps individuels. Pour pouvoir comparer des temps moyens VP/TC, il faut donc retenir parmi les temps déclarés pour un mode que ceux qui sont "moyens". Il convient donc de procéder à ce stade à un filtrage du fichier : retenir seulement les observations dont le temps déclaré dans un mode donné ne diffère que de -20 mn ou +20 mn et de -30% ou +30% du temps simulé pour le même mode :

$$\Delta_{absolu} \in [-20mn, +20mn]$$

$$\Delta_{relatif} \in [-30\%, 30\%]$$

Finalement, 3 bases de calage sont construites, une par motif.

Motif	Nombre d'observations
D-> T + Aff pro	1586
D-> Achats	40
Autres	892

Tableau 5 - nombre d'observations pour le calage du choix modal

⁶ Rappel : Paris est découpé en 80 zones (quartiers) dans le modèle de l'IAURIF

La faiblesse de l'échantillon pour le motif D->Achats rend l'estimation peu fiable. Cependant, ce motif ne concerne qu'à peine 31 000 déplacements motorisés, donc cela n'a quasiment aucune conséquence négative sur le modèle global.

Deux variables explicatives ont été ajoutées au modèle : la disposition d'un parking gratuit à destination DPARK (correspondant à la variable IPGT de l'EGT 2001-2002) et le taux de motorisation TAUXM par couronne tiré de l'EGT 2001-2002.

	Paris	Petite couronne	Grande couronne
DPARK (zone destination)	0,05	0,60	0,90
TAUXM (zone origine)	0,56	0,96	1,28

Les résultats des calages sont les suivants, avec les notations suivantes :

T_{vp} : temps en VP de i à j

T_{tc} : temps en TC de i à j

DPARK : paramètre de disposition d'un parking gratuit à destination (entre 0 et 1)

TAUXM : taux de motorisation par couronne d'origine

P : probabilité de prendre la VP pour aller de la zone i à la zone j

- Domicile → Travail et Affaires professionnelles

$$P = \frac{1}{1 + \exp[-(-2,485 - 0,033 T_{vp} + 0,011 T_{tc} + 2,915 \text{ DPARK} + 1,694 \text{ TAUXM})]}$$

- Domicile → Achats

$$P = \frac{1}{1 + \exp[-(4,426 - 0,151 T_{vp} + 0,028 T_{tc})]}$$

- Autres motifs

$$P = \frac{1}{1 + \exp[-(-1,691 - 0,036 T_{vp} + 0,040 T_{tc} + 1,238 \text{ DPARK} + 0,129 \text{ TAUXM})]}$$

De même que pour la distribution, les résultats du choix modal au niveau des intrazonaux et des focalisations en 8*8 sont comparés avec les intracommunales de l'EGT et les matrices de référence pour chacun des deux modes.

Les intrazonaux par mode et tous motifs, comparé aux intracommunales

	Intrazonaux calculés	Intracommunales EGT
VP	519 000	1 004 000
TC	570 000	213 000
Total	1 089 000	1 217 000

Tableau 6 - comparaison intrazonaux - intracommunales

Les matrices de référence département à département tous motifs par mode

En VP

EGT 2001-2002

OD	75	93	94	92	95	78	91	77	
75	146	26	19	46	6	14	10	4	272
93	30	262	24	10	13	5	5	22	371
94	38	24	238	24	2	4	17	7	353
92	64	17	13	248	6	31	17	4	401
95	13	34	3	30	269	23	2	3	376
78	18	5	5	39	14	431	20	1	534
91	19	5	26	28	3	11	397	6	495
77	8	43	19	8	5	2	19	376	480
	336	415	347	433	318	521	488	424	3 281

Modèle

OD	75	93	94	92	95	78	91	77	
75	139	30	24	47	9	10	8	5	272
93	48	170	34	30	39	9	7	25	363
94	45	34	159	28	8	11	26	22	332
92	58	18	22	213	20	40	20	7	397
95	29	46	10	52	245	33	6	12	431
78	28	12	12	74	34	351	22	9	541
91	28	10	43	34	7	27	286	29	464
77	21	36	34	17	18	13	38	306	483
	395	355	338	494	379	493	412	416	3 317

Ratios

OD	75	93	94	92	95	78	91	77	
75	0,95	1,17	1,25	1,03	1,36	0,68	0,77	1,22	1,00
93	1,62	0,65	1,45	2,87	2,99	1,92	1,39	1,13	0,98
94	1,18	1,43	0,67	1,17	5,02	2,72	1,50	3,19	0,94
92	0,89	1,06	1,69	0,86	3,23	1,28	1,15	1,59	0,99
95	2,26	1,36	3,21	1,70	0,91	1,43	2,48	4,65	1,15
78	1,56	2,31	2,16	1,90	2,32	0,81	1,08	7,96	1,01

91	1,50	1,98	1,67	1,22	2,42	2,35	0,72	4,60	0,94
77	2,55	0,84	1,82	2,18	3,86	6,52	1,96	0,81	1,01
	1,18	0,86	0,97	1,14	1,19	0,95	0,84	0,98	1,01

En TC

EGT 2001-2002

	75	93	94	92	95	78	91	77	
75	425	39	25	79	6	11	6	5	596
93	114	100	11	23	4	1	2	6	261
94	102	11	77	32	2	1	2	2	228
92	131	11	7	126	2	8	3	1	288
95	67	11	2	27	67	3	0	0	178
78	51	6	3	48	5	90	2	-	206
91	50	2	7	15	0	3	71	-	149
77	58	11	9	9	1	2	3	87	181
	997	190	141	359	87	120	91	102	2 087

Modèle

OD	75	93	94	92	95	78	91	77	
75	522	26	21	44	4	4	3	2	627
93	101	130	17	15	10	3	2	6	284
94	94	16	118	15	2	3	6	5	258
92	122	9	11	173	6	12	5	2	339
95	43	15	4	17	109	7	2	2	197
78	41	4	4	26	7	150	4	2	238
91	39	4	13	11	2	4	114	4	189
77	38	13	11	8	3	3	5	114	195
	998	217	198	309	143	186	140	136	2 328

Ratios

OD	75	93	94	92	95	78	91	77	
75	1,23	0,67	0,83	0,56	0,69	0,39	0,51	0,41	1,05
93	0,89	1,31	1,55	0,66	2,41	2,53	0,77	1,00	1,09
94	0,92	1,50	1,54	0,46	1,29	1,82	2,99	2,59	1,13
92	0,93	0,84	1,67	1,37	3,83	1,43	1,56	1,38	1,18
95	0,64	1,31	1,51	0,65	1,62	2,05	4,40	6,16	1,11
78	0,80	0,75	1,22	0,55	1,31	1,67	1,82	-	1,16
91	0,78	1,47	1,89	0,72	6,62	1,27	1,60	-	1,27
77	0,65	1,19	1,15	0,82	2,31	2,17	1,62	1,31	1,08
	1,00	1,14	1,41	0,86	1,64	1,55	1,55	1,33	1,12

Globalement, le calage est correct. On pourrait améliorer le modèle de choix modal en prenant en compte une variable de pénibilité. De plus, il faudrait corriger le temps de parcours intrazonal en TC qui est manifestement sous-estimé (il est calculé comme le minimum divisé par 2 des temps interzonaux).

4.3. Recalage de METROPOLIS

Le modèle METROPOLIS avait été calé dans le cadre du projet QUATUOR. Puisque ce calage a été fait de manière semi-agrégée, nous avons proposé au Predit, dans le cadre de SIMAURIF, d'affiner ce calage en utilisant des données de comptages routiers récentes.

Dans l'attente de ce calage, METROPOLIS a été utilisé pour calculer les variables de temps de parcours en VP depuis chaque cellule vers Châtelet en 1990 et 1999 (TVP90 et TVP99) en heures creuses, c'est-à-dire sur un réseau non chargé, donc ne nécessitant pas un calage du réseau après affectation de la matrice origine-destination.

4.3.1. Rappel du calage semi-agrégé du projet QUATUOR

Le processus du calage consistait à effectuer un nombre assez important de simulations et à retenir uniquement la simulation qui minimise la différence entre les données observées et les données simulées. Les données observées sont tirées des observations de l'enquête globale transports (EGT) de l'année 1991. Elles représentent la distribution du nombre de déplacements en fonction de la durée du déplacement. Etant donné que les paramètres comportementaux sont estimés à partir de l'enquête MADDIF, nous avons calibré deux variables : la demande et la capacité des voies de circulation.

L'objectif du calage consistait à reproduire

- le nombre de déplacements donné par l'IAURIF entre 7h30 et 8h30
- la distribution des temps de déplacement par tranche 15 minutes durant la journée
- le temps de déplacement moyen pour les déplacements de la journée.

Pour atteindre cet objectif, une batterie de simulations est effectuée qui fait évoluer d'une part la demande totale et d'autre part les capacités des routes de façon uniforme c'est-à-dire multiplier les capacités des routes par le même facteur. La demande est multipliée par des coefficients allant de 3 à 5 avec un pas de 0,25. Les demandes totales simulées varient de 2910000 (=3x970000) usagers à 4850000 (5x970000) usagers. Les sous-matrices varient également uniformément puisqu'elles sont multipliées par le même coefficient. En résumé, 63 simulations sont effectuées pour déterminer le scénario de référence. Le scénario final ou calibré nécessite de multiplier la demande par un coefficient égal à 3,75 (soit une matrice totale de 3 637 500 usagers) et la capacité initiale par un coefficient égal à 0,90.

Rappels des résultats du scénario de référence

L'outil de simulation METROPOLIS permet de calculer différentes mesures d'efficacité. Le tableau ci-dessous décrit les résultats du scénario calibré. En moyenne, sur l'ensemble de la matinée, le temps de déplacement est égal à 32,50 minutes soit un temps de déplacement supplémentaire de 8,16 minutes par rapport au temps de déplacement à vide. Le coût du délai (ou de « deshorage ») moyen est égal à 2,45€ soit presque le double du temps de congestion qui est de 1,76€.

Mesures d'efficacité	Valeurs
Temps de parcours [min]	32,50
Temps de parcours à vide [min]	24,34
Temps de parcours de congestion [min]	8,16
Délai d'arrivée précoce [min]	28,28
Délai d'arrivée tardif [min]	19,07
Vitesse moyenne [km/h]	28,90
Part des arrivées précoces [%]	47,25
Part des arrivées tardives [%]	33,15
Coût moyen [€]	9,47
Coût du temps de parcours à vide [€]	5,26
Coût du temps d'attente [€]	1,76
Coût moyen de deshorage [€]	2,45
Volume total des déplacements [million de véh.km]	63,70
Distance moyenne parcourue [km]	17,51
Nombre d'arcs utilisés par individu	17,61
Indice de congestion	28,85

Tableau 7 : mesures agrégées de l'efficacité du trafic

4.3.2. Présentation des nouvelles données de comptage

Nous avons utilisé des comptages effectués par le SIER. Au total nous disposons de données sur 24 tronçons d'autoroutes et de routes nationales de la région pour la journée du jeudi 5 octobre 2000 où il n'y avait pas de pluie ou de phénomène météorologique particulier. Le Tableau 8 présente la distribution géographique des stations de comptage sur le réseau ainsi que le pourcentage des observations utilisables et la longueur de routes couverte.

Chaque station porte un nom (celui de la route), une direction et l'information de situation sur une bretelle d'entrée ou de sortie. Le nombre de voies et la date de comptage sont connus. Pour chaque période de 6 minutes, on dispose du nombre de voitures passées, le taux d'occupation, la vitesse moyenne et la longueur moyenne des voitures passées.

Route	Nombre de stations	Nombre de stations utilisées	% des observations valables	Longueur couverte
A1	87	25	8,24%	11,000 km
A10	27		47,82%	
A103	5		4,42%	
A104	107		18,24%	
A106	16		23,39%	
A12	39		6,20%	
A126	2		23,96%	
A13	75	46	16,30%	22,613 km
A14	108		29,26%	
A15	29		72,56%	
A3	112		35,12%	
A4	129	54	20,92%	20,970 km
A5AB	89		55,29%	
A6	103		39,53%	
A6A	44		15,27%	
A6B	55		11,61%	
A86E	151		20,42%	
A86I	149		16,09%	
N104	186	74	13,36%	35,850 km
N118	34		19,71%	
N13	16		37,40%	
N286	10		10,04%	
N315	20		23,02%	
N406	42		19,99%	
Somme	1635		23,67%	

Tableau 8 - Distribution géographique des points de comptage

La Figure 10 présente la distribution des stations selon le nombre de voies de la route. La majorité des points se trouvent sur les routes à deux voies. Notons que le nombre de voies est compté séparément pour chaque sens. Les tronçons à une voie sont majoritairement des bretelles et ceux de 3 voies ou plus se trouvent majoritairement sur les voies principales.

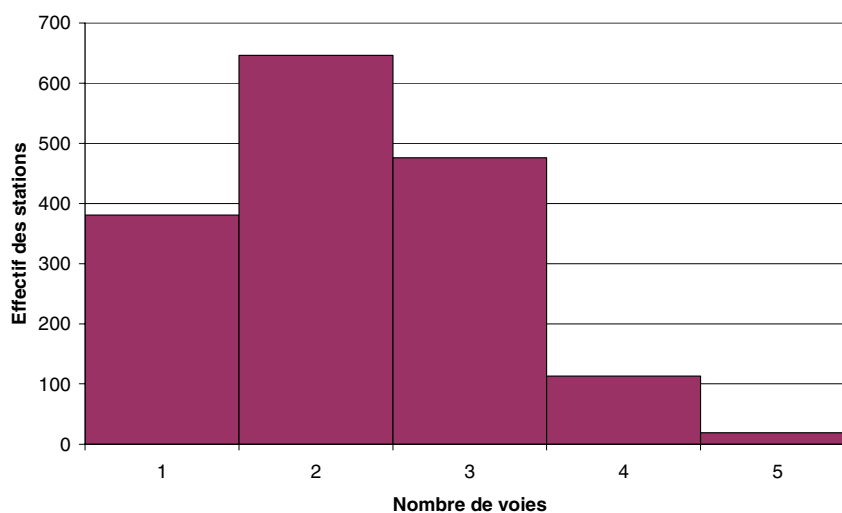


Figure 10- la distribution des stations selon le nombre de voies

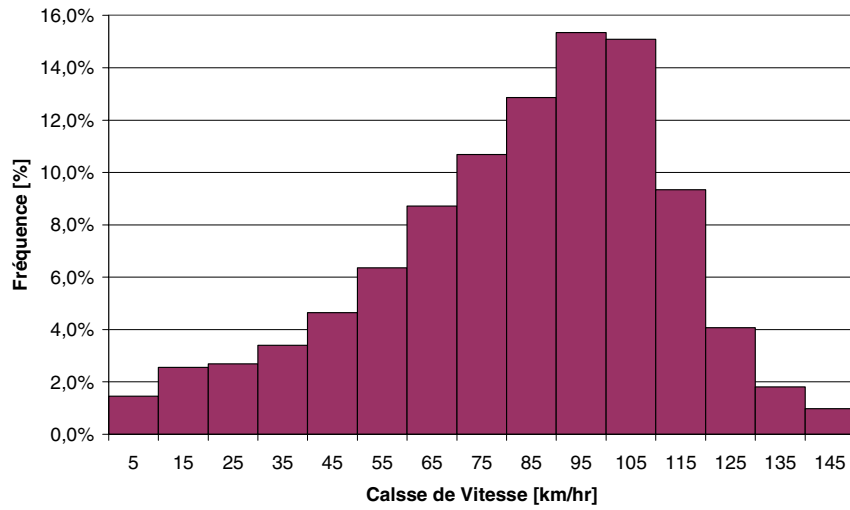


Figure 11- la distribution de la vitesse mesurée

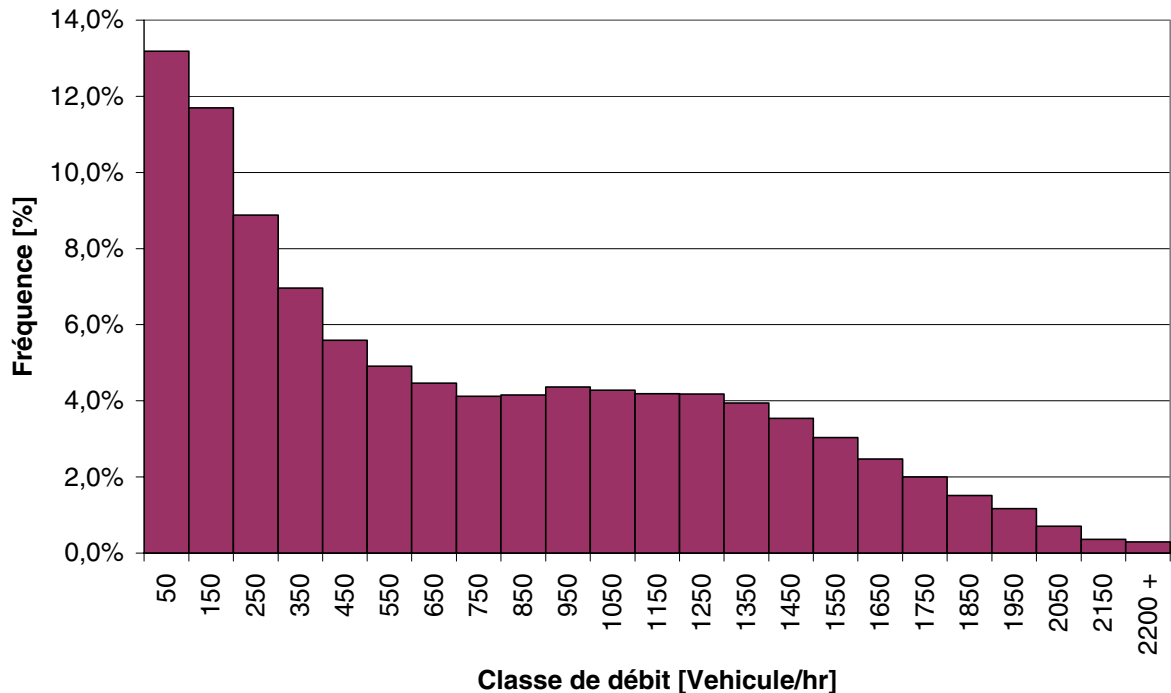


Figure 12- La distribution des débits observés

Parmi les observations, 24% ne comportent pas la donnée de vitesse mesurée et 0,11% ont des vitesses supérieures à 150 km/h. En dehors de ces deux catégories, la Figure 11 présente la distribution des vitesses observées par classe de 10 km/h. Le mode est à 95 km/h et la moyenne à 82,16 km/h.

23% des observations manquent de valeur pour le débit passant. Comme présentée sur la Figure 12, seulement 0,3% des observations donnent des débits supérieurs à 2200 véhicules

par heure et par voie. Le mode est à la classe de [0,100) et la moyenne des débits observés est égale à 173 Vé/h.

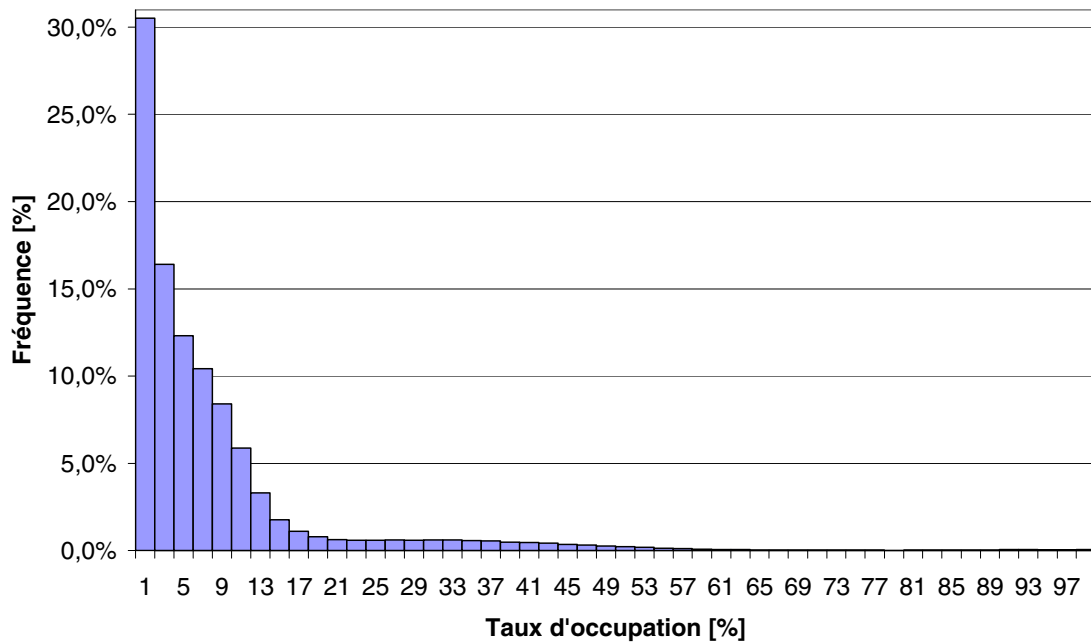


Figure 13 - la distribution des taux d'occupation observés

Les taux d'occupation observés sont présentés à la Figure 13. Ils dépassent rarement les 50%, le mode est à 1% et la moyenne à 8%.

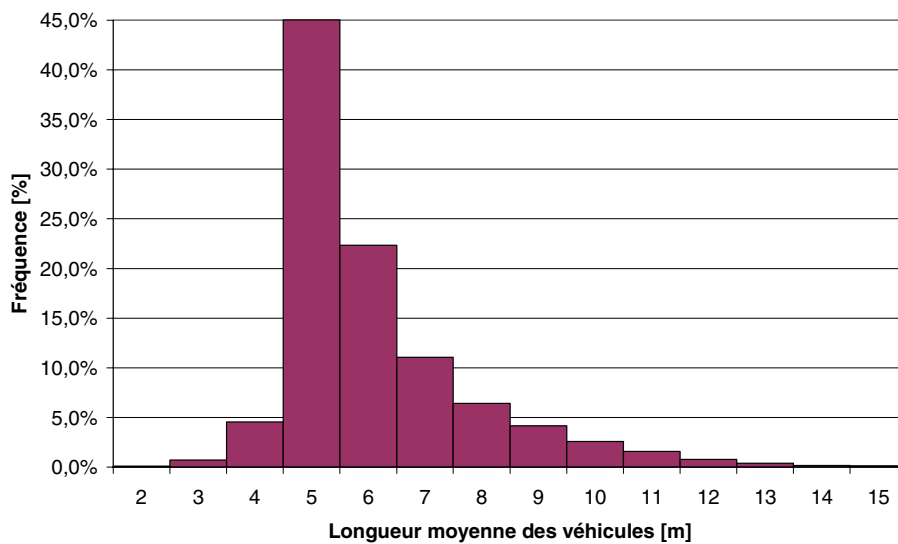


Figure 14 - la distribution de longueur moyenne des véhicules

La longueur moyenne des véhicules est donnée par la Figure 14. La médiane qui est à 5 mètres correspond à la longueur moyenne des voitures particulières.

Localisation des points de comptage

Nous disposons des plans schématiques du SIER présentant les points de comptage. Ces cartes nous donnent le kilométrage et la localisation de ces stations par rapport aux intersections importantes. Grâce à ces deux indices, nous avons localisé les stations sur les arcs du graphe codé. Etant donné que dans le codage du réseau, un échangeur est représenté comme un seul nœud, les stations qui se trouvent entre une sortie et une entrée sont ignorées. Les bretelles sont considérées comme les mouvements tournants et par conséquent, nous ne les prenons pas en compte dans les estimations des fonctions de congestion.

4.3.3. Estimation des fonctions de congestion

Il est nécessaire de rappeler succinctement les points les plus importants de l'ingénierie du trafic.

Description des variables de trafic

Notation :

N : le flux de trafic sur une période T

n : le nombre de voitures présentes sur un tronçon de route

n_l : le nombre de voies d'une route

q : le débit de trafic au moment t au point x de la route

k : la concentration du trafic

u : la vitesse

x : la coordonnée sur la route

D : la distance entre deux repères

τ : le taux d'occupation

\bar{L} : la longueur moyenne des voitures passées

d : la longueur de capteur (généralement 1 mètre)

Flux

Le flux de trafic représente le nombre de voitures passées dans un intervalle donné de temps. Le flux divisé par la durée de cet intervalle étant le nombre de voitures passant par unité de temps est appelé le débit. L'unité de temps est généralement une heure et le débit est exprimé en véhicules par heure.

$$q = N/T$$

Concentration

La concentration mesure le nombre de voitures présentes sur un kilomètre et une voie d'une route. Elle est mesurée instantanément.

$$K = n/D$$

Vitesse

C'est la distance parcourue dans l'unité de temps par un véhicule. Elle est généralement exprimée en km par heure. La vitesse instantanée d'un véhicule est définie comme :

$$U = dx / dt$$

Il existe deux manières différentes pour calculer la vitesse moyenne d'une voiture ou un ensemble de voitures qui donnent des résultats différents :

Vitesse moyenne dans le temps

La vitesse moyenne dans le temps est calculée comme la moyenne arithmétique des vitesses observées pour un ensemble des voitures ou pour une voiture sur différentes périodes.

$$\bar{u}_t = \frac{1}{N} \sum_{i=1, N} u_i$$

Vitesse moyenne d'espace

Pour cette vitesse moyenne, deux définitions sont couramment utilisées dans la littérature. La première est attribuée à Wardrop (1952). La vitesse moyenne d'espace, u_s , sur une section prédéfinie de longueur D , est la distance D divisée par la moyenne des temps nécessaires pour traverser la section.

$$\bar{u}_s = \frac{D}{\bar{t}} = \frac{D}{\frac{1}{N} \sum_{i=1, N} t_i}$$

La vitesse moyenne d'espace peut être calculée comme une moyenne harmonique des vitesses des véhicules comme suit :

$$\bar{u}_s = \frac{1}{\frac{1}{N} \sum_{i=1, N} \frac{1}{u_i}} = \frac{1}{\frac{1}{N} \sum_{i=1, N} \frac{t_i}{D}} = \frac{D}{\frac{1}{N} \sum_{i=1, N} t_i}$$

Le seul inconvénient de cette seconde relation est qu'elle ne prend pas en compte la longueur de la route. Par ailleurs, si les vitesses des véhicules ne sont pas constantes dans l'espace, la vitesse moyenne dans le temps est supérieure à la vitesse moyenne d'espace. En particulier, lorsque les variations des vitesses individuelles sont faibles, il est possible d'établir une relation entre les deux types de mesure des vitesses moyennes (Wardrop, 1952) par la relation suivante :

$$\bar{u}_t = \bar{u}_s + \frac{\sigma_s^2}{\bar{u}_s}$$

La variable σ^2 qui désigne l'écart-type de la distribution des vitesses spatiales est égale à $\sqrt{\sum k_i (u_i - \bar{u}_s)^2 / K}$.

Certaines études dans le domaine ont montré que les deux vitesses ne sont pas égales (voir Wardrop, 1952 et Gerlough et Huber, 1975) et que la différence est de l'ordre de 12% à 13% entre les deux vitesses moyennes. Dans le cas des routes périurbaines, cette dispersion est faible.

Taux d'occupation

Cette mesure ponctuelle donne le pourcentage du temps qu'il y ait une voiture sur le capteur. Pour les routes à plusieurs voies, il est divisé par le nombre de voies. Nous avons :

$$\tau = \frac{t_p}{T n_l} = \frac{\sum (L_i + d) / u_i}{T n_l} = \frac{N(\bar{L} / \bar{u}_l + d / \bar{u}_s)}{T n_l} \approx \frac{N(\bar{L} + d)}{T n_l \bar{u}_l} = k(\bar{L} + d)$$

L'équation fondamentale du flux de trafic

Pour N véhicules, avec une distance spatiale de s_i et une distance temporelle de h_i qui passent pendant une période T par une section, nous avons :

$$q = \frac{N}{T} = \frac{N}{\sum h_i} = \frac{N}{\left(\frac{\sum s_i}{u_s} \right)} = \frac{N}{\sum s_i} u_s = k u_s$$

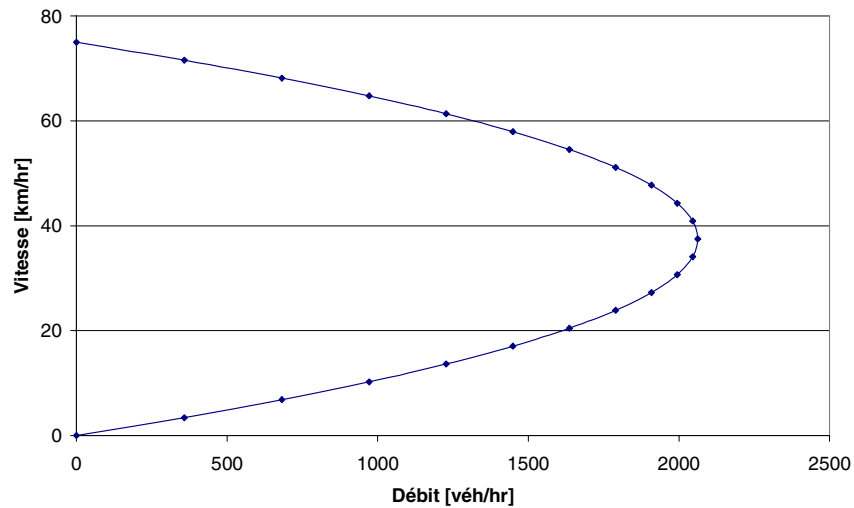


Figure 15 - relation débit - vitesse

Cette relation n'est pas monotone et ne donne pas une vitesse unique pour une valeur donnée du débit. Par contre les données expérimentales montrent une relation plus claire entre la vitesse et la concentration. D'ailleurs, la concentration moyenne étant le nombre des véhicules présents sur l'arc par unité de longueur est compatible avec le modèle classique de goulot d'étranglement et cette valeur est connue à chaque instant dans METROPOLIS.

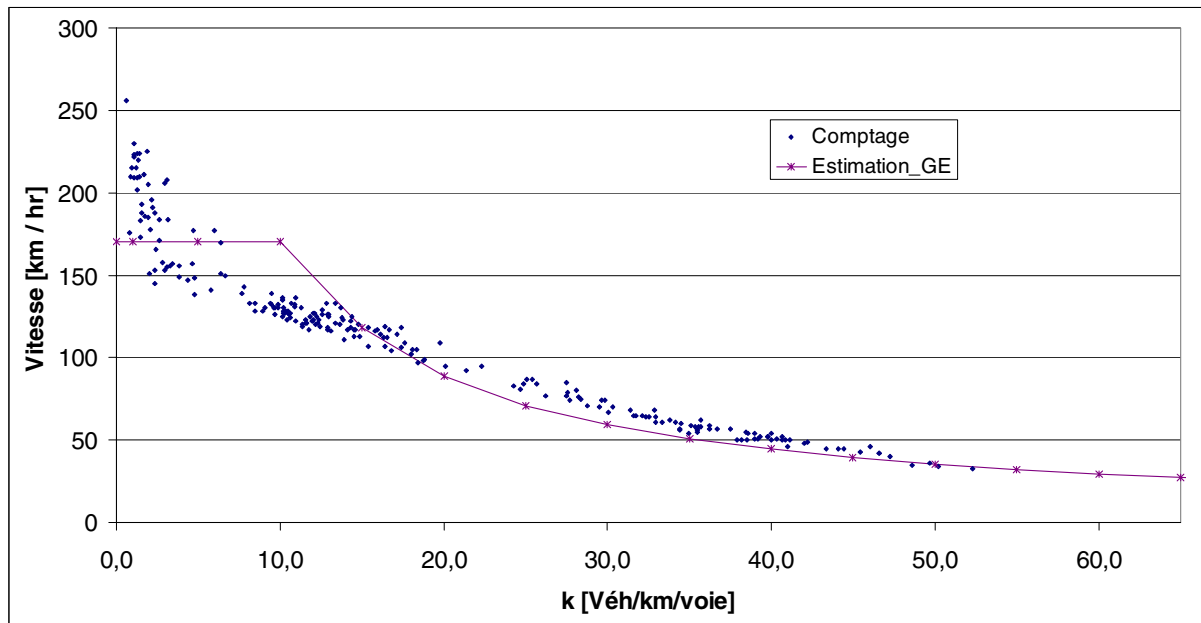


Figure 6 - relation vitesse - concentration

Types de fonction de congestion

La fonction de congestion est censée donner le temps pour traverser un arc en fonction du débit passant ou du nombre de véhicules présents sur l'arc et ses caractéristiques. METROPOLIS permet d'introduire une large famille de fonctions possibles tant qu'elles sont monotones et décroissantes. La formulation cohérente de ces fonctions a fait l'objet de plusieurs recherches (voir Carey). La fonction de type goulot d'étranglement a une forme simple et elle est théoriquement acceptable pour une affectation dynamique. En plus, nous avons essayé plusieurs formes fonctionnelles et étudiée la qualité d'ajustement qu'elles offrent par rapport aux données de comptage.

Les caractéristiques de chaque arc sont la vitesse à vide, la capacité, la longueur et les paramètres supplémentaires en fonction des nécessités de la fonction choisie. Dans le codage initial du réseau par l'IAURIF, les capacités sont établies par rapport à celle de référence soit 2000 véh/hr/file. Pour les voies urbaines, ces capacités sont multipliées par des facteurs correctifs afin de prendre en compte le temps perdu dû aux feux rouges et aux contraintes environnementales de circulation. Ces coefficients sont légèrement modifiés afin d'ajuster au mieux les résultats de l'affectation statique. Les vitesses à vide étaient choisies entre 25 et 100 km/hr en fonction du type de route, du nombre de voies et de la concentration urbaine dans son entourage.

Les données de comptage nous fournissent la vitesse mesurée à un ou plusieurs points sur l'arc par période de 6 minutes. Nous supposons que la vitesse moyenne sur l'arc est égale à celle calculée à ce point ou à la moyenne des points se trouvant sur cet arc. Cette hypothèse est raisonnablement valable pour une distance de 500 mètres avant et après la station de comptage. Nous disposons également du nombre de véhicules passant sur cette période ainsi que leur longueur moyenne et le taux d'occupation d'ensemble des capteurs d'une station. Notons que les données dont nous disposons sont des moyennes sur les périodes de 6

minutes. Nous sommes en train d'estimer les fonctions sur les données déjà agrégées. Les résultats seront présentés dans le rapport final.

Modèle	Formulation du modèle	Paramètres à estimer
Goulot d'étranglement (Bottleneck)	$u = \begin{cases} u^{\max} & \text{si } k < \widehat{k} \\ \frac{C}{k} & \text{si } \widehat{k} \leq k \end{cases}$	u^{\max}, C
Modèle de stabilité locale d'écoulement (SLE)	$u = \begin{cases} u^{\max} & \text{si } k < \widehat{k} \\ \frac{1}{\delta} \left(\frac{1}{k} - \frac{1}{k_j} \right) & \text{si } \widehat{k} \leq k \leq k_j \end{cases}$	$u^{\max}, 1/\delta, k_j$
Modèle de Greenberg	$u = \begin{cases} u^{\max} & \text{si } k < \widehat{k} \\ c \cdot \ln \left(\frac{k_j}{k} \right) & \text{si } \widehat{k} \leq k \leq k_j \end{cases}$	u^{\max}, c, k_j
Modèle de Greenshields	$u = u^{\max} \left(1 - \frac{k}{k_j} \right)$	u^{\max}, k_j
Modèle de Drew	$u = u^{\max} \left[1 - \left(\frac{k}{k_j} \right)^{1/2} \right]$	u^{\max}, k_j

Tableau 9 - les formes fonctionnelles à estimer pour la fonction de congestion

Considérations économétriques et choix du modèle

Afin d'utiliser les méthodes d'estimation linéaire, nous avons procédé à des changements de variable de k à $1/k$, $\ln(1/k)$ ou $k^{1/2}$ selon la forme fonctionnelle. De même, pour les deux dernières fonctions, nous avons ajouté un palier de $u=u^{\max}$ pour les concentrations les plus faibles. La concentration critique, c'est-à-dire lorsque le changement de régime apparaît, est déterminée par tâtonnement de sorte qu'elle maximise la qualité d'ajustement représentée par r^2 .

$$r^2 = 1 - \frac{\sum (u^{\text{modèle}} - u^{\text{observée}})^2}{\sum (u^{\text{observée}})^2}$$

La méthode de minimisation de la somme d'écart au carré utilisée pour l'estimation est basée sur l'hypothèse d'homoscédasticité qui suppose l'uniformité de la variance de la variable en fonction de différentes valeurs de variable explicative. Dans le cas des vitesses, nous observons une plus grande variabilité des vitesses pour les plus faibles concentrations. Ceci est dû à la liberté des usagers dans leur choix de la vitesse quand il n'y a pas de congestion. Quand la concentration augmente, cette liberté diminue et les usagers sont obligés de rouler à une vitesse d'équilibre pour l'ensemble des véhicules. Pour résoudre cette incohérence, nous divisons les deux cotés de l'équation par k et nous estimons, en fait, u/k .

Résultats d'estimation

Le tableau présenté en annexe 3 présente une partie des résultats pour certains arcs et pour les différents types d'arc. Etant donnée la simplicité de la fonction de goulot d'étranglement et la cohérence théorique de ce modèle, nous nous basons sur ce type de fonction. Les résultats montrent une différence non négligeable entre les résultats des différents points de même arc ou des différents arcs de même type. Nous utilisons les valeurs moyennes pour nos simulations. Les capacités et les vitesses à vide des arcs pour lesquelles nous ne disposons pas de comptage seront égales aux moyennes du même type d'arc.

4.3.4. Calage global du modèle

Le calage global du modèle en utilisant les fonctions estimées est en cours. Il utilise aussi d'autres données telles que les données des temps de déplacement issues de l'EGT 2001-2002, les mesures de temps de parcours dans Paris effectuées par les Services techniques de la Mairie de Paris etc. .

4.3.5. Prise en compte du modèle d'affectation en TC

Nous avons décidé d'utiliser le modèle d'affectation TC du logiciel Davisum utilisé à l'IAURIF. La dernière version de Davisum comprend le moteur de METROPOLIS. En fait, *SIMAUURIF* devient progressivement le logiciel intégrant UrbanSim, Davisum et METROPOLIS.

L'IAURIF dispose d'un réseau TC mis à jour en 2001 avec la description des missions sur deux heures de pointe du matin. Ce réseau comprend l'ensemble du réseau lourd, le réseau de bus RATP et 165 lignes du réseau de bus OPTILE.

On a opté pour la méthode suivante : on fera tourner Davisum indépendamment et on récupèrera uniquement les matrices de temps généralisés et les logsommes (indicateurs d'accessibilité) qui seront injectés comme entrées dans UrbanSim. Cette partie sera détaillée dans le rapport final.

Jusqu'à maintenant, le modèle TC a été utilisé pour calculer les variables de temps de parcours en transports en commun depuis chaque cellule vers Châtelet en 1990 et 1999 : TTC90 et TTC99.

4.4. Programmation

Les 3 étapes ont été programmées dans l'environnement Windev et intégrées à METROPOLIS.

5. Constitution des bases de données de calage

5.1. Rappel : les 4 modèles d'UrbanSim à estimer

Les quatre modèles principaux d'UrbanSim ont été présentés de manière détaillée dans la partie 5.3 du rapport n°1. Les trois premiers sont des modèles de choix discrets (logit), le quatrième est un modèle de régression linéaire.

5.1.1. Modèle de localisation résidentielle des ménages

Grâce à un logit multinomial, ce modèle prédit la probabilité qu'un ménage nouveau ou déplacé se localise dans une cellule donnée. Dans un tel modèle, des variables caractéristiques du ménage interagissent avec les caractéristiques de l'occupation du sol, de l'environnement et des transports.

Concrètement, dans la structure de données d'UrbanSim, ce modèle est formalisé par 6 tables qu'il faut construire :

- Table « ANNUAL_HOUSEHOLDS_CONTROL_TOTALS »
- Table « ANNUAL_RELOCATION_RATES_FOR_HOUSEHOLDS »
- Table « ANNUAL_RELOCATION_RATES_FOR_HOUSEHOLDS »
- Table « HOUSEHOLD_CHARACTERISTICS_FOR_HLC »
- Table « HOUSEHOLD_CHARACTERISTICS_FOR_HT »
- Table « HOUSEHOLD_LOCATION_CHOICE_MODEL_SPECIFICATION »

5.1.2. Modèle de localisation des emplois

Ce module calcule la probabilité qu'un emploi nouveau (créé dans le module de transition économique) ou relocalisé (mobilité économique) soit localisé dans une cellule donnée.

De même façon que le modèle précédent, un modèle logit multinomial est utilisé. Pour le caler, les emplois localisés durant la période de calage sont observés.

Le modèle repose sur 7 tables :

- Table « EMPLOYMENT_ADHOC_SECTOR_GROUP_DEFINITIONS »
- Table « EMPLOYMENT_ADHOC_SECTOR_GROUP »
- Table « EMPLOYMENT_SECTORS »
- Table « EMPLOYMENT_ADHOC_SECTOR_GROUP »
- Table « EMPLOYMENT_EVENTS »
- Table « EMPLOYMENT_LOCATION_CHOICE_MODEL_SPECIFICATION »

- Table « SQFT_FOR_NON_HOME_BASED_JOBS »

5.1.3. Modèle de l'évolution du MOS

On a substitué au module déterminant la probabilité de réalisation de programmes immobiliers un module de l'évolution du MOS. La variable DEVELOPMENT_TYPE a été définie comme étant le type urbain d'occupation du sol en 5 classes (voir paragraphe 3.6). Le modèle consiste alors à estimer la probabilité de transition d'un type à l'autre.

Les tables à construire dans la structure de données d'URBANSIM sont :

- Table « DEVELOPMENT_TYPE_GROUP_DEFINITIONS »
- Table « DEVELOPMENT_TYPE_GROUPS »
- Table « DEVELOPMENT_TYPES »
- Table « DEVELOPMENT_CONSTRAINT »
- Table « DEVELOPMENT_EVENTS »
- Table « DEVELOPMENT_CONSTRAINT_EVENT »
- Table « DEVELOPMENT_HISTORY »
- Table « DEVELOPER_MODEL_SPECIFICATIONS »
- Table « PLAN_TYPES »
- Table « TRANSITION_TYPES »
- Table « RESIDENTIAL_LAND_SHARE_MODEL_SPECIFIC_VARIABLES »
- Table « RESIDENTIAL_LAND_SHARE_MODEL_SPECIFICATION »
- Table « RESIDENTIAL_UNITS_FOR_HOME_BASED_JOBS »
- Table « LAND_USE_EVENTS » :

5.1.4. Modèle du prix du foncier

Les prix sont déterminés par une approche en termes de prix hédoniques.

L'équation hédonique est :

$$P_{it} = \alpha + \delta \left(\frac{V_i^s - V_{it}^c}{V_i^s} \right) + \beta X_{it}$$

avec V_i^s le taux de vacance structurel et V_{it}^c le taux de vacance courant.

Le modèle de prix d'UrbanSim est aussi basé sur une structure de données pré-établie. Il s'agit alors de se conformer à cette structure et construire les tables suivantes :

- Table « LAND_PRICE_MODEL_SPECIFIC_VARIABLES »
- Table « LAND_PRICE_MODEL_SPECIFICATION » :

Un autre modèle de prix du foncier, dit semi-hédonique, a été conçu par Paul Waddell et le THEMA pour le projet d'un nouvel UrbanSim. Il est décrit dans la partie 7.4.

Au 15 septembre 2005, les deux premiers modules ont été estimés. Le troisième est en cours. Le modèle de prix n'a pas encore été étudié. Il est particulièrement complexe à estimer à l'échelle des cellules de 500 mètres de côté car il requiert des données complémentaires qui ne sont pour l'instant pas encore collectées.

Avant de présenter les résultats des estimations des deux premiers modèles, il est nécessaire de décrire les bases de données de calage ainsi que quelques statistiques descriptives sur les phénomènes d'évolution du MOS et de mobilité résidentielle.

5.2. Base de calage du modèle d'évolution du MOS

Il s'agit de construire une base de données de calage⁷ comportant l'évolution d'un échantillon de cellules entre 1990 et 1999 selon l'occupation du sol en 5 types urbains.

Il existe deux séries de contraintes à l'évolution de l'occupation du sol d'une cellule : des contraintes environnementales et des contraintes politiques locales.

5.2.1. Contraintes environnementales

Par exemple, si une cellule est située en zone inondable, elle est non constructible et aura donc peu de chances de voir son type urbain évoluer. Par conséquent, les cellules situées en zones contraintes sont considérées comme figées et sont exclues de la base de calage. Les variables de contraintes environnementales ont été décrites dans le rapport n°1 (pages 87-89) et dans le présent rapport (paragraphe 3.4).

Une cellule sera exclue si elle remplit une des contraintes suivantes :

- plus de 50% de la surface de la cellule est composée d'eau ('percent_water')
- plus de 25% de la surface de la cellule est composée de berges des cours d'eau ('percent_stream_buffer')
- plus de 50% de la surface de la cellule est composée de zones inondables ('percent_floodplain')
- plus de 50% de la surface de la cellule est composée de zones humides ('percent_wetland')
- le terrain qui compose la surface de la cellule est trop escarpé (pente supérieure à 9%⁸) ('percent_slope')
- plus de 50% de la surface de la cellule est composée de routes ('percent_roads')
- plus de 50% de la surface de la cellule est composée de zones protégées autour des sites inscrits et monuments ('percent_monuments')
- plus de 50% de la surface de la cellule est exposée à la zone de bruit 96 dB des aéroports ('percent_noise')

⁷ On parlera plus simplement de « bases de calage ».

⁸ A l'exception des cellules situées dans les environs de Montmartre à Paris ou à Meudon

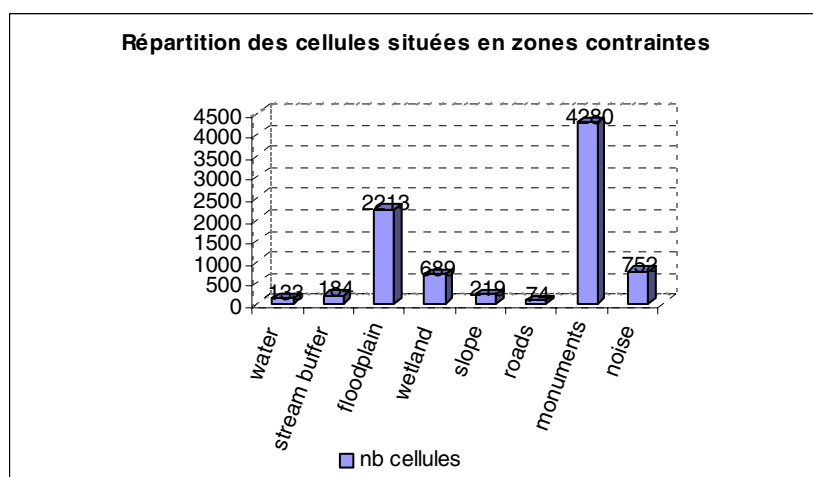


Figure 17 – répartition des cellules situées dans les zones de contraintes environnementales

La prise en compte de ces contraintes environnementales conduit à l'exclusion de 7367 cellules sur les 49236 cellules initialement présentes.

5.2.2. Contraintes politiques locales

On considère que les cellules dont plus de 50% de la surface est non urbanisable (« libellé abrégé » commençant par N, sauf NA et NB) sont à exclure de la base de calage. La prise en compte de cette contrainte aboutit à la suppression de 13548 cellules.

Au total, après avoir pris en compte les contraintes environnementales et les contraintes réglementaires des PLU, la base de calage est constituée de **31379 cellules** sur les 49236 cellules initialement présentes.

5.2.3. Statistiques descriptives sur l'évolution des cellules

Le tableau n°8 donne le nombre de cellules ayant transité du type i ($i = 1, \dots, 5$) pour l'année 1990 vers le type j ($j = 1, \dots, 5$) pour l'année 1999. Pour la typologie, on se reportera au paragraphe 3.6.

Exemple de lecture du tableau : 254 cellules sont passées du type 5 en 1990 au type 4 en 1999.

Ce qui nous intéresse plus particulièrement, ce sont les évolutions depuis le type « vacant » vers les types urbanisés 1, 2, 3 et 4.

Les premiers tests d'estimation du modèle d'évolution du MOS du type 5 vers chacun des quatre types 1 à 4 n'ont pas été très satisfaisants, à cause du nombre trop faible de cellules par transition $i \rightarrow j$. Ce faible nombre de cellules ayant évolué s'explique par la période

relativement courte (10 ans) pour observer des évolutions nettes de l'urbanisation. Les urbanistes raisonnent en général à des échelles de temps de 20 à 25 ans.

1990\1999	Type 1	Type 2	Type 3	Type 4	Type 5	Total
Type 1	3130	47	11	11	6	3205
Type 2	40	494	36	29	10	609
Type 3	285	45	903	19	36	1288
Type 4	28	52	40	417	88	625
Type 5	376	172	346	254	24504	26652
Total	3859	810	1336	730	24644	31379

Tableau 10 – effectifs des transitions observées

On a donc été amené à procéder encore à des regroupements. Deux nouveaux types urbains agrégés ont été définis :

- **'habitat'** : types 1+3. On considère que le type 3 'espace public ouvert' fait partie du type 1 'habitat'.
- **'activités'** : types 2+4. On considère que le type 4 'réseaux' fait partie du type 2 'activités & équipements'.

La figure n°2 donne les évolutions des cellules vacantes vers le type 'habitat' et vers le type 'activités'.

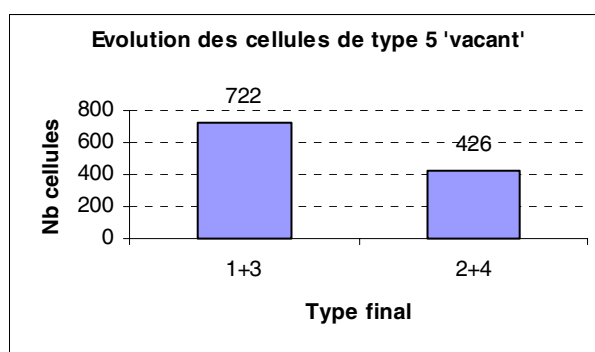


Figure 18 – évolution des cellules de type vacant

Au final, on a retenu uniquement les transitions du type 5 vers 1+3 et du type 5 vers 2+4. La base de données de calage du modèle d'évolution du MOS comporte alors **26 652 observations**, dont 1148 ayant évolué, soit 4,5%.

En ne travaillant que sur ces deux transitions, on fera l'hypothèse, en prédiction, qu'une cellule de type 1, 2, 3 ou 4 à l'année n ne changera plus de type à l'année n+1. Seules les cellules vacantes changeront de type.

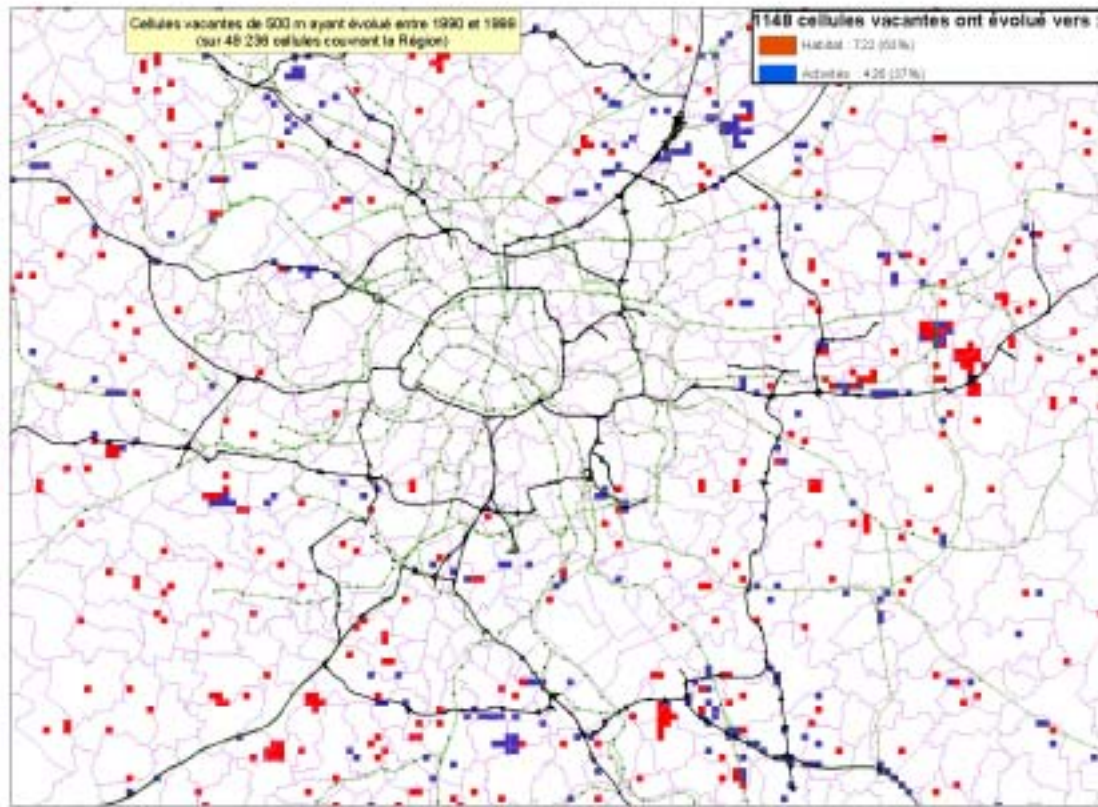


Figure 19 – Evolution des cellules vacantes entre 1990 et 1999

Cette carte montre clairement que les cellules vacantes qui se sont urbanisées vers de l’habitat ou de l’activité sont principalement localisées en grande couronne, dans les Villes nouvelles, le long des axes de transport et dans le secteur de Roissy.

5.2.4. Récapitulation : variables de la base de calage

Comme le modèle sera utilisé pour faire de la prédiction, les variables testées sont essentiellement les variables de la table gridcells pour l’année 1990. Ce sont les caractéristiques des cellules en 1990 qui ont motivé les décisions d’implantation d’habitats, de bureaux, d’équipements ... entre 1990 et 1999. Par contre, les variables d’accessibilité des deux années 1990 et 1999 sont prises en compte.

Un nombre important de variables (attributs d’une cellule), classées en 5 types, a été testé :

1) Des variables qui donnent la composition d’une cellule en 1990 :

- residential_units = nombre de logements
- commercial_sqft = surfaces commerciales
- industrial_sqft = surfaces d’activités secondaires
- governmental_sqft = surfaces d’activités tertiaires (bureaux)

2) Des variables qui donnent les conditions du marché d'une cellule en 1990 :

- residential_improvement_value = valeur foncière du bâti résidentiel
- residential_land_value = valeur foncière du terrain
- commercial_improvement_value = loyers des surfaces commerciales
- industrial_improvement_value = loyers des surfaces d'activités secondaires
- governmental_improvement_value = loyers des bureaux
- nonresidential_land_value = valeur foncière du non résidentiel

3) Des variables de distance en 1990 :

- distance_to_arterial = distance de la cellule à la route nationale la plus proche
- distance_to_highway = distance de la cellule à l'autoroute ou voie rapide la plus proche
- distance_a_Chatelet

4) Des variables de transports (temps et desserte spatiale) en 1990 et en 1999 :

- TVP = temps de parcours moyen en voiture (heures creuses) pour aller de la cellule à Châtelet
- TTC = temps de parcours moyen en transport en commun pour aller de la cellule à Châtelet
- nb_nodes_subway = nombre de stations de métro
- nb_nodes_train = nombre de gares
- nb_nodes_tramway = nombre de stations de tramway

5) Les variables de voisinage

Elles correspondent à la description faite en 3.4.1..

Au total, **39 variables** ont été testées. La base porte sur les 26 652 cellules de type vacant en 1990. Elles sont résumées dans le tableau suivant :

	Nom de la variable	Min	Max	Unités
1	residential_units	0	591	nombre
2	commercial_sqft	0	11 474	m ²
3	industrial_sqft	0	58 120	m ²
4	governmental_sqft	0	173 796	m ²
5	residential_improvement_value	0	620 728	milliers euros
6	residential_land_value	0	124 031	milliers euros
7	commercial_improvement_value	0	1 451	milliers euros
8	industrial_improvement_value	0	5 147	milliers euros
9	governmental_improvement_value	0	31 848	milliers euros
10	nonresidential_land_value	0	3 187	milliers euros
11	distance_a_Chatelet	6,52	94,58	km

12	distance_to_arterial	0	19,68	km
13	distance_to_highway	0	25	km
14	TVP90	0,24	2,18	heures
15	TTC90	0,47	4,11	heures
16	nb_nodes_subway90	0	0	nombre
17	nb_nodes_train90	0	2	nombre
18	nb_nodes_tramway90	0	0	nombre
19	TVP99	0,25	2,02	heures
20	TTC99	0,47	4,13	heures
21	nb_nodes_subway99	0	0	nombre
22	nb_nodes_train99	0	2	nombre
23	nb_nodes_tramway99	0	0	nombre
24	vresidential_units	0	7 002	nombre
25	vcommercial_sqft	0	28 801	m ²
26	vindustrial_sqft	0	688 754	m ²
27	vgovernmental_sqft	0	210 534	m ²
28	vresidential_improvement_value	0	12 392 355	milliers euros
29	vresidential_land_value	0	2 365 685	milliers euros
30	vcommercial_improvement_value	0	4 026	milliers euros
31	vindustrial_improvement_value	0	225	milliers euros
32	vgovernmental_improvement_value	0	41 148	milliers euros
33	vnonresidential_land_value	0	4 600	milliers euros
34	vnb_nodes_subway90	0	1	nombre
35	vnb_nodes_train90	0	3	nombre
36	vnb_nodes_tramway90	0	0	nombre
37	vnb_nodes_subway99	0	1	nombre
38	vnb_nodes_train99	0	3	nombre
39	vnb_nodes_tramway99	0	0	nombre

Tableau 11 - Les variables testées du modèle d'évolution du MOS

Les résultats de l'estimation du modèle d'évolution du MOS (variables significatives, coefficients, tests de validité) sont présentés dans la partie 6.1.

5.3. Base de calage du modèle de localisation résidentielle

Cette base est construite à partir de l'EGT 2001-2002. Les avantages de cette base par rapport au recensement général ont été présentés dans le rapport n°1 (pages 104-105). Le fichier Ménages de l'EGT comporte 11 152 ménages.

5.3.1. Statistiques descriptives sur la mobilité résidentielle entre 1990 et 1999

Des migrations de proximité

La base de données comporte 5497 ménages ayant déménagé au moins une fois sur la période 1991-2001. Pour simplifier, on assimile la période 1991-2001 entre deux EGT à la période de calage 1990-1999. Ainsi, 52,5% des individus ont déménagé. Parmi eux, ils ne sont que 40% à avoir changé de département et 24% à avoir changé de commune.

Remarquons que seule la dernière commune de résidence est enregistrée dans l'EGT (il en est de même dans le recensement sauf que l'on a l'information qu'au niveau départemental). En prenant une période de plusieurs années, il manque une information sur les déménagements précédant le dernier déménagement, ce qui conduit à sous-estimer la mobilité, cette sous-estimation étant d'autant plus marquée que la période prise en compte est grande (ce qui augmente la probabilité de déménagements multiples).

Anc. dépt \ Nouv. dépt	75	77	78	91	92	93	94	95	Total
75	12,6%	0,6%	0,6%	0,6%	2,5%	1,1%	1,4%	0,5%	19,9%
77	0,3%	6,5%	0%	0,2%	0,1%	0,3%	0,3%	0,1%	7,8%
78	0,4%	0,1%	8,3%	0,2%	0,5%	0,1%	0,2%	0,4%	10,2%
91	0,4%	0,4%	0,2%	6,1%	0,2%	0,1%	0,3%	0%	7,7%
92	0,9%	0,3%	0,7%	0,6%	7,8%	0,2%	0,6%	0,8%	11,9%
93	0,9%	1,1%	0,2%	0,2%	0,4%	6,4%	0,6%	0,9%	10,7%
94	0,7%	0,5%	0,2%	0,7%	0,5%	0,7%	6,3%	0,2%	9,8%
95	0,3%	0,2%	0,4%	0,1%	0,3%	0,3%	0,1%	6,2%	7,9%
Hors IDF	4,2%	1,2%	1,5%	1,1%	2,3%	1,2%	1,5%	1,3%	14,3%
Total	20,7%	10,9%	12,1%	9,8%	14,6%	10,4%	11,3%	10,4%	100%

Tableau 12 - matrice de transition des déménagements

Le Tableau 12 donne le détail des mobilités résidentielles inter-départementales observées dans l'échantillon, sans pondération. L'EGT ne permet pas de connaître les mouvements des résidents d'Ile-de-France vers la province, en particulier les franges de la région.

Les personnes qui résidaient anciennement à Paris sont majoritairement restées à Paris. Elles sont néanmoins environ 37% à avoir choisi de quitter Paris pour la Petite Couronne ou la Grande Couronne.

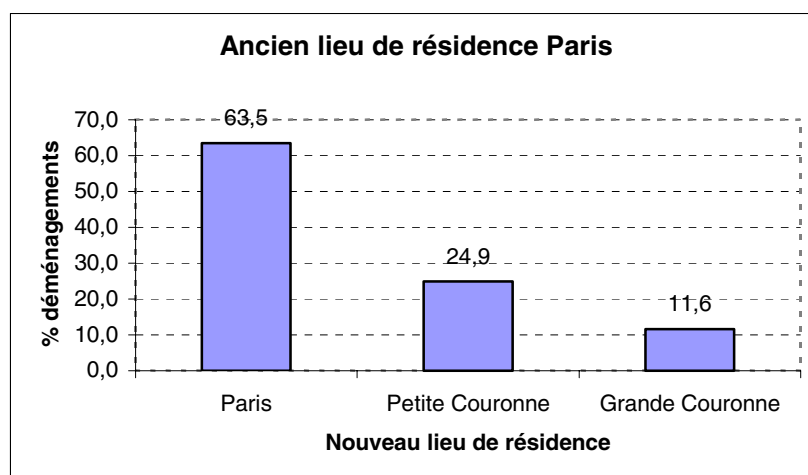


Figure 20 - couronne de destination pour ceux qui habitaient à Paris

La très grande majorité des résidents de la Petite Couronne est restée dans cette zone. Ils ne sont que 7,7 % à avoir rejoint Paris.

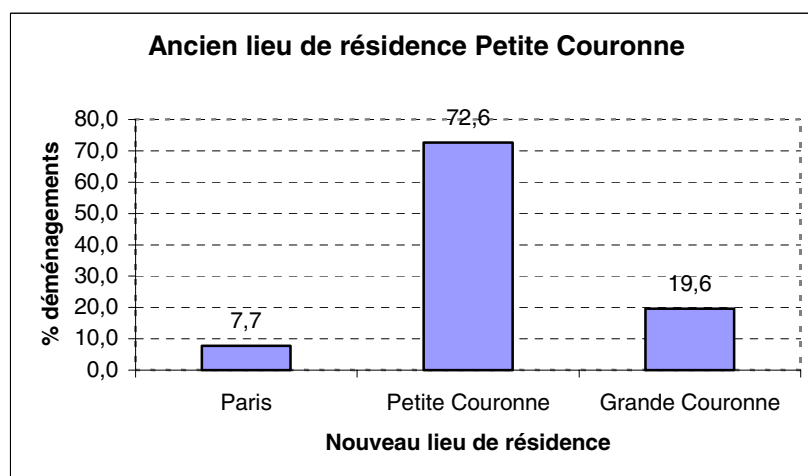


Figure 21 - couronne de destination pour ceux qui habitaient en Petite couronne

Ce phénomène est accentué pour les résidents de la Grande Couronne qui sont restés dans 88,1% des cas dans la Grande Couronne. Ils ne sont que 4,1 % à avoir quitté la Grande Couronne pour rejoindre Paris.

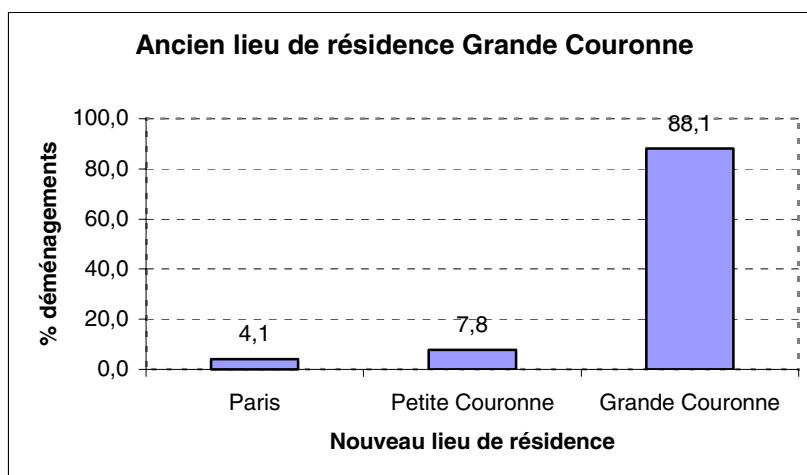


Figure 22 - couronne de destination pour ceux qui habitaient en Grande couronne

Des migrations très liées au cycle de vie

La plupart des individus qui quittent Paris ont entre 30 et 40 ans, âge auquel les personnes sont désireuses de fonder une famille.

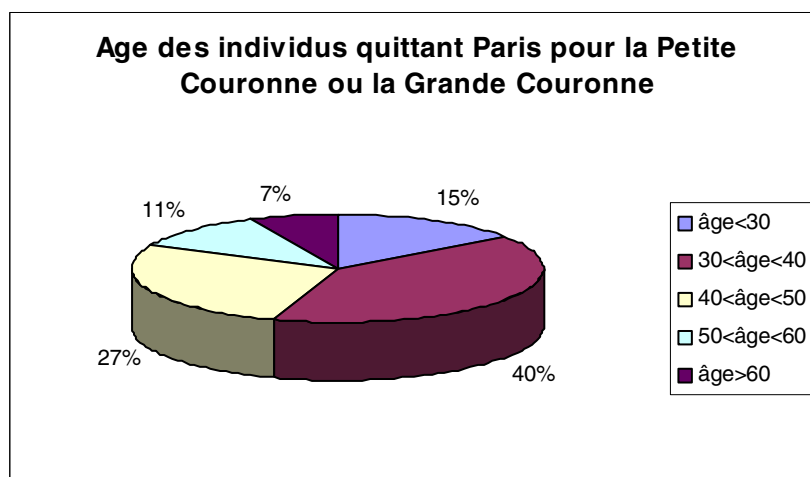


Figure 23 – distribution de l'âge de ceux qui partent de Paris

La plupart des individus qui s'installent à Paris sont jeunes : moins de 30 ans. Ils s'installent à Paris pour finir leurs études ou trouver un premier emploi.

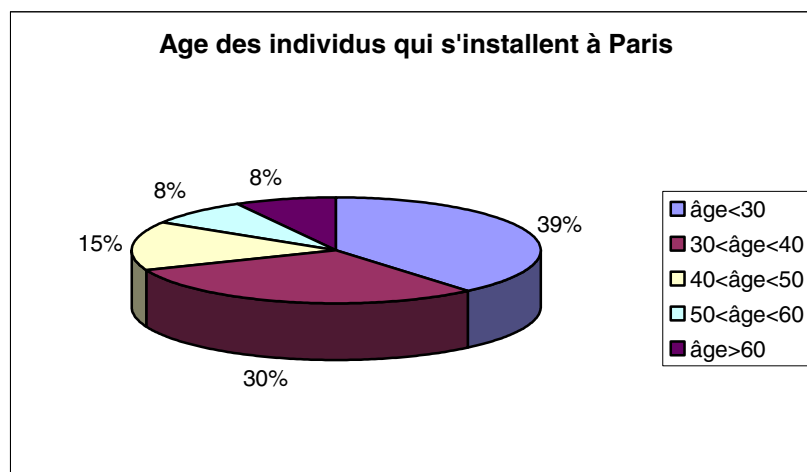


Figure 24 - distribution de l'âge de ceux qui emménagent à Paris

5.3.2. Variables sur les caractéristiques socio-économiques des ménages

Elles sont disponibles dans le fichier ménages de l'EGT 2001-2002. On a retenu les variables suivantes :

- Age du chef de ménage
- Nombre de personnes composant le ménage
- Nombre de personnes âgées de moins de 6 ans
- Revenus annuels du ménage par tranche
- Type d'habitation
- Indicatrice valant 1 si le ménage a changé de département et 0 sinon
- Indicatrice valant 1 si le ménage a changé de commune et 0 sinon
- Nombre d'actifs dans le ménage
- Nombre de véhicules dans le ménage

5.3.3. Variables sur les indicateurs d'accessibilité

Il est prévu que les indicateurs d'accessibilités (les logsommes) soient calculés par METROPOLIS. En attendant le calage de METROPOLIS, ces indicateurs ont été calculés indépendamment.

Ainsi, pour chacune des cellules, on calcule :

- EmploiVP = accessibilité à l'emploi calculé à partir des temps de parcours en voiture en heures creuses
- EmploiTC = accessibilité à l'emploi calculé à partir des temps de parcours en transport en commun
- ComVP = accessibilité aux commerces calculé à partir des temps de parcours en voiture en heures creuses
- ComTC = accessibilité aux commerces calculé à partir des temps de parcours en transport en commun

Pour cela, la formule de l'accessibilité proposée par Jean Poulit (1974) a été utilisée. Elle est basée sur la prise en compte du logarithme des biens offerts : $\log(Q_j)$

où Q_j représente soit les emplois, soit les commerces disponibles dans la cellule j , suivant la variable d'accessibilité souhaitée. $\log(Q_j)$ caractérise en fait (à un facteur près) l'intérêt que l'utilisateur attribue à la possibilité de pouvoir disposer à sa destination d'un marché d'emplois, de commerces ou de biens égal à Q_j . Il s'agit là du gain associé au déplacement. De ce terme positif, il convient de retirer le coût généralisé du transport (C_{ij}). La satisfaction ressentie par l'utilisateur s'écrit alors de la façon suivante :

$$S_{ij} = \lambda \log(Q_j) - C_{ij} \quad (1)$$

où :

- λ est un facteur de pondération qui vaut $\lambda = \frac{C_0}{\alpha}$, C_0 étant le coût d'une heure de transport⁹, et α est le coefficient exprimant la rapidité de l'effet d'atténuateur de la distance dans la loi de distribution des déplacements (voir paragraphe 4.2.2 pour les valeurs de α qui ont été calculées pour les motifs D->T+aff pro et D->Achats).
- C_{ij} est le coût généralisé du transport qui vaut $C_{ij} = C_0 t_{ij}$, t_{ij} étant le temps de transport nécessaire pour aller de la cellule i à la cellule j

L'équation (1) se réécrit de la façon suivante :

$$S_{ij} = \lambda \log \left(Q_j \exp \left[-\frac{C_{ij}}{\lambda} \right] \right) \quad (2)$$

Les résidents d'une cellule i , avant d'effectuer leur déplacement, prennent en compte l'ensemble des destinations potentielles qui leur sont offertes. Afin d'apprécier la satisfaction moyenne de la collectivité résidant en i , il convient d'étendre l'interprétation économique précédente à l'ensemble des liaisons issues de i .

Jean Poulit démontre que la satisfaction d'un résident de la zone i vis à vis de l'ensemble des biens offerts par l'ensemble des zones j qui entourent la zone i est égale à :

$$S_i = \lambda \log \left(\sum_j Q_j \exp \left[-\frac{C_{ij}}{\lambda} \right] \right) \quad (3)$$

Ainsi, l'utilisateur prend en considération l'ensemble des biens qui l'entourent, les pondère par l'effet atténuateur de la distance et en fait la somme. L'augmentation relative de cette somme est l'indicateur de satisfaction auquel il est sensible.

⁹ Le rapport « Boiteux II » (juin 2001) préconise les valeurs suivantes : 11,6€ pour le motif D->T et 6,4 € pour le motif D->Achats

C'est cette dernière formule qui a permis de calculer les variables d'accessibilité. Pour les deux variables d'accessibilité à l'emploi Q_j est le nombre d'emplois de la cellule j . Pour les deux variables d'accessibilité aux commerces, Q_j est le nombre de surfaces commerciales de la cellule j . Quant aux temps de transports t_{ij} , ils sont calculés par METROPOLIS pour les temps VP en heures creuses (donc sans affectation de matrice origine-destination) et par Davisum pour les temps TC¹⁰.

Plus S_i est grand, plus la zone i est accessible.

5.3.4. Récapitulation : variables de la base de calage

On peut les classer en 4 catégories :

1) Les variables disponibles dans gridcells :

- des variables qui donnent la composition d'une cellule (en 1990) :
 - residential_units = nombre de logements
 - commercial_sqft = surfaces commerciales
 - industrial_sqft = surfaces d'activités secondaires
 - governmental_sqft = surfaces d'activités tertiaires (bureaux)
- des variables qui donnent les conditions du marché d'une cellule (en 1990) :
 - residential_improvement_value = valeur foncière du bâti résidentiel
 - residential_land_value = valeur foncière du terrain
 - commercial_improvement_value = loyers des surfaces commerciales
 - industrial_improvement_value = loyers des surfaces d'activités secondaires
 - governmental_improvement_value = loyers des bureaux
- des variables de distance :
 - distance_a_Chatelet
 - distance_to_arterial = distance de la cellule à la route nationale la plus proche
 - distance_to_highway = distance de la cellule à l'autoroute ou voie rapide la plus proche
- des variables de transports (temps et desserte spatiale) pour l'année 1990 et l'année 1999 :

¹⁰ $\left[-\frac{C_{ij}}{\lambda} \right] = [-\alpha t_{ij}]$

- TVP = temps de parcours moyen en voiture (heures creuses) pour aller de la cellule à Châtelet
 - TTC = temps de parcours moyen en transport en commun pour aller de la cellule à Châtelet
 - nb_nodes_subway = nombre de stations de métro
 - nb_nodes_train = nombre de gares
 - nb_nodes_tramway = nombre de stations de tramway
- une variable de politique de la ville : appartenance à une ZUS/ZRU/ZFU.

2) Les variables de voisinage pour l'année 1990

Les variables de voisinage associées aux variables suivantes ont été calculées :

- 'residential_improvement_value' = valeur foncière du bâti résidentiel
- 'residential_land_value' = valeur foncière du terrain
- 'commercial_sqft' = surfaces de commerce
- 'industrial_sqft' = surfaces d'activités secondaires
- 'governmental_sqft' = surfaces d'activités tertiaires (bureaux)
- 'residential_units' = nombre de logements
- 'commercial_improvement_value' = loyers des surfaces commerciales
- 'industrial_improvement_value' = loyers des surfaces d'activités secondaires
- 'governmental_improvement_value' = loyers des bureaux
- 'nb_nodes_subway' = nombre de stations de métro
- 'nb_nodes_train' = nombre de gares
- 'nb_nodes_tramway' = nombre d'arrêts de tramway.

3) Des nouvelles variables d'accessibilité ont été calculées

- Accessibilité à l'emploi calculée à partir des temps de parcours en voiture en heures creuses
- Accessibilité à l'emploi calculée à partir des temps de parcours en transport en commun
- Accessibilité aux commerces calculée à partir des temps de parcours en voiture en heures creuses
- Accessibilité aux commerces calculée à partir des temps de parcours en transport en commun

4) Variables socio-économiques caractérisant les ménages

Elles sont tirées directement de l'EGT 2001-2002, les ménages étant été localisés dans le maillage :

- Age du chef de ménage
- Nombre de personnes composant le ménage
- Nombre de personnes âgées de moins de 6 ans
- Revenus annuels du ménage par tranche

- Type d'habitation
- Nombre d'actifs dans le ménage
- Nombre de véhicules dans le ménage
- Indicatrice valant 1 si le ménage a changé de département et 0 sinon
- Indicatrice valant 1 si le ménage a changé de commune et 0 sinon

Pour la variable Revenu, une moyenne par tranche a été calculée (il faudrait théoriquement connaître la distribution des revenus par tranche) :

Tranche	Revenu moyen
0	0 €
1	2 287 €
2	6 860 €
3	11 433 €
4	17 530 €
5	20 579 €
6	25 152 €
7	32 012 €
8	45 732 €
9	73 171 €
10	106 707 €

Tableau 13 - revenu moyen par tranche

Au total, la base de calage comporte **52 variables** (y compris la variable POIDS du ménage de l'EGT 2001-2002). Les noms de variables ont changé par rapport au modèle d'évolution du MOS. Cette base comprend 5497 observations, plus les alternatives (6 par observation), dont le tirage aléatoire est décrit dans la partie 6.2.1. Au total, la base comporte 38 479 lignes (5497*7).

	Nom de la variable	Min	Max	Unités
1	resu_99	0	65 464	Nombre
2	com_99	0	143 181	Nombre
3	ind_99	0	208 070	Nombre
4	gov_99	0	676 888	Nombre
5	resval_99	0	1 969 467	Milliers €
6	land_99	0	198 753	Nombre
7	comval_99	0	29 290	Milliers €
8	indval_99	0	19 792	Milliers €
9	govval_99	0	252 458	Milliers €
10	dart_99	0	18,8	Km
11	dhigh_99	0	24,9	Km
12	d_chatelet	0	92,6	Km
13	subway_99	0	5	Nombre
14	train_99	0	4	Nombre
15	tram_99	0	2	Nombre
16	subway_90	0	5	Nombre
17	train_90	0	4	Nombre
18	tram_90	0	0	Nombre

19	TTC_90	0,1	2,1	heures
20	TTC_99	0,	2,0	heures
21	TVP_90	0,1	21,2	heures
22	TVP_99	0,1	2,0	heures
23	ZUS	0	1	binaire
24	vresu_99	0	65 464	Nombre
25	vcom_99	0	190 027	Nombre
26	vind_99	0	1 219 762	Nombre
27	vgov_99	0	3 139 097	Nombre
28	vresval_99	0	9 394 736	Nombre
29	vland_99	0	988 790	Nombre
30	vcomval_99	0	40 682	Milliers €
31	vindval_99	0	101 077	Milliers €
32	vgovval_99	0	1 359 606	Milliers €
33	vsubway_99	0	21	Nombre
34	vtrain_99	0	8	Nombre
35	vtram_99	0	6	Nombre
36	vsubway_90	0	21	Nombre
37	vtrain_90	0	8	Nombre
38	vtram_90	0	0	Nombre
39	emploiTC	30,0	52,7	Nombre
40	emploiVP	40,4	54,8	Nombre
41	comTC	22,6	34,4	Nombre
42	comVP	28,5	35,8	Nombre
43	AGC	16	94	Nombre
44	MNPT	1	11	Nombre
45	MNPI	0	5	Nombre
46	MRTV	0	106 707	Nombre
47	GHA	1	7	Nombre
48	MNACT	0	4	Nombre
49	MNV4	0	4	Nombre
50	Meme_dep	0	1	Binaire
51	Meme_com	0	1	Binaire
52	POIDS	114	712	Nombre

Tableau 14 - variables testées du modèle de localisation résidentielle

Les résultats de l'estimation du modèle de localisation des ménages (variables significatives, coefficients, tests de validité) sont présentés dans la partie 6.2.

Notons enfin à titre de comparaison que dans l'étude Puget Sound, la base de calage du modèle de localisation résidentielle des ménages comportait 2364 observations, soit moins de deux fois que notre base.

5.4. Base de calage du modèle de localisation des emplois

Cette base n'a pas encore été construite à la date du 15 septembre 2005. On envisage d'utiliser les données de l'Enquête Régionale sur l'Emploi (ERE) qui sont connues à un niveau infracommunal. Les quatre dernières ERE ont eu lieu en 1990, 1994, 1997 et 2000, mais seule la dernière a été géolocalisée dans le SIGR.

Le fichier détaillé ERE donne la liste des établissements présents en Ile-de-France, en indiquant l'activité principale de cet établissement, son numéro de SIRET, son adresse et le nombre de salariés (y compris les entreprises mono-salariées). Un important travail de relocalisation de certains effectifs parisiens a été effectué, par recouplement avec d'autres fichiers. Les données n'ont été géolocalisées que pour les emplois salariés du secteur privé. Par contre, la base de données alpha-numérique comporte bien tous les emplois salariés, publics et privés, représentant plus de 92% des emplois, mais ne comptabilise pas les emplois de type libéral (moins de 8% des emplois). Notons que pour le modèle de localisation des emplois à la commune (présenté dans la partie 7.5) utilise une autre source de données, celle de l'UNEDIC.

La couche ERE2000 du SIGR est une couche de points comportant 189178 entreprises. L'intersection avec le maillage en 50000 cellules a été effectuée.

On utilisera le numéro de SIRET qui est une clé commune permettant de joindre l'ERE avec d'autres bases de données sur les emplois (BIL, ASTREE, SIRENE). Il sera alors possible de récupérer l'information d'année de création de l'établissement ou l'année d'emménagement.

Il est utile de rappeler que le modèle de localisation des emplois d'UrbanSim considère l'emploi comme agent économique et non l'entreprise. La base de calage sera donc constituée d'observations qui sont des emplois ayant été créés ou ayant déménagé entre 1990 et 1999. L'échantillon devra être représentatif suivant deux critères : la répartition par activité et la répartition par département.

On pense que la taille de l'échantillon pourrait tourner autour de 200 000 emplois. A titre de comparaison, l'étude de Puget Sound travaille sur un échantillon de 90 000 emplois.

Rappelons (voir page 98 du rapport n°1) que la table « employment_events » d'UrbanSim permet d'enregistrer des déménagements importants qui de toutes façons ne sont pas prévisibles par simulation. Par exemple, le regroupement à Guyancourt en 1998 de près de 8000 emplois du constructeur Renault.

5.4.1. Statistiques descriptives sur la mobilité des emplois entre 1990 et 1999

Entre l'ERE de 1990 et de 1994, les effectifs franciliens ont diminué de 123 000 salariés. Cette évolution négative est le résultat de mouvements de disparitions, de créations et de desserrement internes à la région ou vers les autres régions. Néanmoins, le bilan régional montre que les diminutions, notamment en zone centrale, ont été plus élevées que les augmentations constatées en grande couronne. Aussi, durant ces quatre premières années de la décennie précédente, la répartition des salariés s'est-elle modifiée, la part de Paris se

réduisant, tandis que celle des départements périphériques augmentait. Rappelons aussi que la conjoncture économique au niveau national a été extrêmement défavorable pendant ces années.

Entre l'ERE de 1994 et de 2000, l'emploi salarié a augmenté de nouveau, surtout à partir de 1997. Mais globalement, sur la période 1990-1999, selon le Recensement, la Région a perdu 34 000 emplois.

5.4.2. Récapitulation : variables de la base de calage

Elles seront présentées dans le rapport final de juin 2006. A priori, les variables testées seront les mêmes que celles du module de localisation des ménages. On envisage d'ajouter une variable géographique caractérisant la cellule par une valeur « d'image ». En effet, les entreprises dans les secteurs d'activités financières ou conseils recherchent des adresses de prestige, indépendamment de tout critère d'accessibilité et de prix des bureaux.

5.5. Base de calage du modèle de prix

Ce modèle est très important car il permettrait d'apporter des éléments de réponse à des questions récurrentes telles que :

- Quel est l'impact sur le foncier autour des gares de la mise en service d'une nouvelle infrastructure de transport en commun du type de la Tangentielle nord ?
- Le fonctionnement des marchés fonciers résidentiels peut-il expliquer le phénomène de péri-urbanisation, ou est-ce plutôt due à la carence de la desserte spatiale en transports en commun ?
- Comment les marchés fonciers résidentiels se structurent-ils dans l'espace et quels sont les facteurs qui déterminent leur extension territoriale ?

Le problème est que l'estimation de ce modèle nécessite un recueil de données spécifiques : les prix du terrain et du logement à un niveau infracommunal, autour des gares. On pourrait choisir comme gares ou stations celles de deux infrastructures avant et après leur mise en service entre 1990 et 1999 : le RER EOLE et le tramway T1. Les moyens possibles de recueil de données serait une enquête spécifique, ou d'utiliser la base de données désagrégées des Notaires ou encore la base de données de l'Observatoire régional du foncier. La base de données des Notaires dont on dispose aujourd'hui fournit seulement des informations agrégées au niveau communal.

Ce recueil n'était pas prévu dans le programme à réaliser. Pour l'instant, l'équipe ne dispose pas des moyens financiers pour le réaliser. Sachant que Réseau Ferré de France (RFF) a manifesté son intérêt pour ce module, l'équipe sollicitera un financement complémentaire de sa part.

Pour l'instant, la méthode proposée est la suivante : on estime un modèle de prix du foncier à la commune et on descendra au niveau de la cellule par une méthode adéquate qui reste à définir.

De nombreuses données sont disponibles à la commune en 1990 et 1999. Elles sont issues du Recensement et ont déjà été cartographiées par l'IAURIF dans le cadre d'autres études :

- le prix moyen au m² à la commune des terrains
- le prix moyen des appartements
- le prix moyen des maisons
- le taux de vacance des logements

Quant à la base de données agrégées des Notaires, elle comprend l'ensemble des transactions immobilières enregistrées par les notaires sur la période 1990-2003 mais ces transactions ne sont pas localisées à l'adresse. Seule la donnée de commune est fournie. Les autres données concernent les volumes et prix moyens par année, pour quatre types de logements :

- Maisons neuves
- Maisons anciennes
- Appartements neufs
- Appartements anciens

Dans la partie 7.4, nous avons décrit une méthode sophistiquée pour compléter cette base qui est incomplète.

Il existe cependant 3 sources de données sur les prix du logement à un niveau infracommunal :

- la base de données des Notaires au niveau du quartier : le CD-ROM « Biens »
- la base de données du GRECAM sur les logements neufs, individualisés à l'adresse et disponible à l'IAURIF
- la base de données de la FNAIM sur les logements anciens, non disponible à l'IAURIF.

Il faudra explorer ces trois sources.

6. Résultats des estimations de deux modules d'UrbanSim

Ce rapport intermédiaire présente les résultats des calages de 2 modules : le modèle d'évolution du MOS (terminé) et le modèle de localisation résidentielle des ménages (encore à approfondir).

6.1. Calage du modèle d'évolution du MOS

6.1.1. Le logit multinomial

Le but est d'estimer les probabilités de transition p_{ij} ($i = 1, \dots, 5$ et $j = 1, \dots, 5$) où p_{ij} est la probabilité qu'une cellule passe du type i en 1990 au type j en 1999. Par conséquent, 25 (5×5) probabilités de transition doivent être estimées. De plus, ces probabilités doivent respecter les 5 contraintes suivantes :

$$\begin{cases} p_{11} + p_{12} + p_{13} + p_{14} + p_{15} = 1 \\ p_{21} + p_{22} + p_{23} + p_{24} + p_{25} = 1 \\ p_{31} + p_{32} + p_{33} + p_{34} + p_{35} = 1 \\ p_{41} + p_{42} + p_{43} + p_{44} + p_{45} = 1 \\ p_{51} + p_{52} + p_{53} + p_{54} + p_{55} = 1 \end{cases}$$

Un des moyens de prendre en compte ces contraintes consiste à estimer 5 logits multinomiaux indépendants (voir encadré pour un rappel succinct sur ces modèles). Dans chacun des i logits ($i = 1, \dots, 5$) considérés, ce sont les cellules qui étaient initialement de type i qui sont sélectionnées : sachant qu'une cellule était de type i fixé en 1990, les probabilités p_{ij} sont estimées :

$$p_{ij} = \frac{\exp(X_i' \beta_{ij})}{\sum_{k=1}^5 \exp(X_i' \beta_{ik})}$$

où X_i est un vecteur contenant l'ensemble des variables explicatives du modèle pour la cellule i .

Une des difficultés à obtenir des estimations satisfaisantes provient du faible nombre d'observations de cellules qui ont changé de type pour chacun des 5 modèles logits (cf. figure 24).

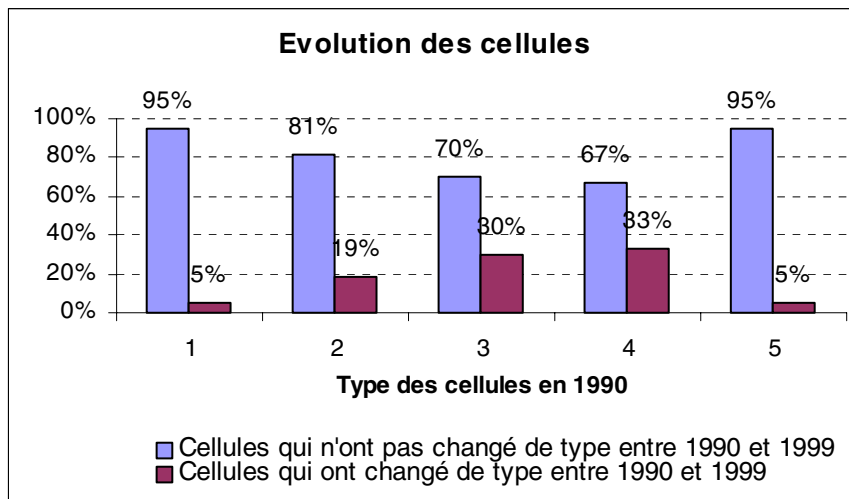


Figure 25 – transitions des cellules

Modèle Logit Multinomial

Définition :

A l'individu n et à chaque alternative $j, j = 1, \dots, J$, l'utilité suivante est associée :

$$U_{nj} = X'_n \beta_j + \varepsilon_{nj}$$

où :

- X'_n est un vecteur qui contient l'ensemble des variables explicatives du modèle,
- les perturbations $\varepsilon_{n1}, \dots, \varepsilon_{nJ}$ sont indépendantes et leur distribution est une distribution de Gumbel standard.

Il est supposé que l'individu n choisit l'alternative lui procurant l'utilité maximale. Après quelques étapes de calcul, il est montré que la probabilité que l'individu n choisisse l'alternative j prend la forme suivante :

$$P_{nj} = \frac{\exp(X'_n \beta_j)}{\sum_{k=1}^J \exp(X'_n \beta_k)} = \frac{1}{1 + \sum_{k \neq j} \exp[X'_n (\beta_k - \beta_j)]}$$

Identification :

Dans la mesure où les paramètres β_j n'apparaissent que sous la forme $\beta_j - \beta_k$, ils ne sont donc identifiés qu'à une constante additive près. Le changement de variable :

$$\beta_j \rightarrow \beta_j + c, \quad \forall j,$$

ne change pas la valeur des probabilités associées à chaque alternative. Par conséquent, il est d'usage de choisir une alternative j de référence pour laquelle β_j est fixé à zéro.

Seuls les résultats du modèle logit 5 (cellules de type vacant en 1990) sont présentés¹¹ d'une part parce que ce modèle met en évidence quelques résultats intéressants et d'autre part parce que c'est l'évolution des cellules vacantes qui est la plus intéressante dans un modèle d'urbanisation. Les types 1 & 3 ('habitat') et 2 & 4 ('activités') ont été regroupés pour l'année 1990 (voir paragraphe 5.2.3). De plus, il est intéressant d'effectuer ces regroupements parce que ce sont les évolutions vers de l'habitat et des activités qui intéressent les urbanistes. Par conséquent, les probabilités de transition suivantes sont estimées à l'aide du logiciel SAS¹² et de la procédure **proc logistic** (voir encadré pour plus de détails sur l'utilisation de cette procédure) :

- $p_{5,13}$ (transition du type 5 vers le type 1 + 3)
- $p_{5,24}$ (transition du type 5 vers le type 2 + 4)
- p_{55} (la cellule reste de type 5 'vacant')

La procédure LOGISTIC

Cette procédure permet d'estimer aisément différents modèles de choix discrets. Afin d'estimer un modèle logit multinomial, il faut ajouter l'option **link=glogit**. Si l'on souhaite obtenir le R^2 et le R^2 ajusté, il faut utiliser l'option **rsquare**. Enfin, l'option **output out=** suivie de l'option **predicted** permet de stocker dans une nouvelle table les probabilités estimées par le modèle.

La syntaxe est la suivante :

```
proc logistic data=nom_table_entree descending;

    model var_expliquee = var_explicative_1 var_explicative_2 /
    link=glogit rsquare ;

    output out=nom_table_sortie predicted=varname;

run;
```

6.1.2. Etude des corrélations entre les variables explicatives

L'étude des corrélations entre les variables explicatives est primordiale. Elle se pose aussi bien pour le modèle d'évolution du MOS que pour les modèles de localisation des ménages et des emplois. En effet, si différentes variables explicatives sont trop fortement corrélées, 3 types de problèmes peuvent se poser :

1. **Un problème d'estimation** : dans le cas d'une corrélation parfaite ou colinéarité (une variable est alors une combinaison linéaire d'autres variables explicatives), la matrice de variances-covariances n'est pas inversible et l'estimation se révèle impossible.

¹¹ L'alternative de référence est p_{55}

¹² Le programme permettant de procéder à ces estimations est donné en annexe 2

Ainsi, par exemple, il n'est pas possible d'intégrer dans le modèle toutes les variables qui donnent la composition d'une cellule. En effet, leur somme est une constante égale à la surface totale de la cellule et est colinéaire à la constante du modèle.

2. Un problème d'interprétation des résultats :

- Une corrélation très élevée entre deux variables explicatives ne sera pas immédiatement détectée car le modèle pourra quand même être estimé. Néanmoins, cette corrélation aura en général un impact immédiat sur la significativité des coefficients d'une des deux variables. Ainsi, même si les coefficients associés aux variables sont significativement différents de zéro, la corrélation biaise les estimations. Prenons l'exemple du nombre d'unités résidentielles et de la valeur foncière du bâti résidentiel. Comme ces deux variables sont très liées, la valeur foncière du bâti résidentiel capte une partie de l'influence du nombre d'unités résidentielles. L'effet du nombre d'unités résidentielles sur l'évolution du type d'une cellule est alors sous évalué. La corrélation génère alors un défaut de robustesse du modèle puisqu'il s'avère impossible de vérifier que chacune des variables explicatives est significative.
- Une corrélation très élevée engendre également des intervalles de confiance élevés et pose problème pour conclure quant à la significativité des variables explicatives envisagées.

3. Un problème de stabilité du modèle : les corrélations constatées dans la base de données ayant servi à la construction du modèle peuvent varier dans le temps et le modèle risque ainsi de ne pas être reproductible.

La PROC CORR du logiciel SAS a permis d'obtenir la matrice de corrélation des variables explicatives et de détecter les couples de variables dont la corrélation est très élevée (supérieure à 0,7) :

- ('v_residential_improvement_value', 'v_residential_land_value')
- ('v_commercial_improvement_value', 'v_commercial_sqft')
- ('v_industrial_improvement_value', 'v_industrial_sqft')
- ('v_governmental_improvement_value', 'v_governmental_sqft')
- ('v_residential_land_value', 'v_industrial_sqft')
- ('v_industrial_improvement_value', 'v_non_residential_land_value')
- ('industrial_improvement_value', 'industrial_sqft')
- ('residential_improvement_value', 'residential_land_value')
- ('governmental_improvement_value', 'governmental_sqft')
- ('residential_land_value', 'governmental_sqft')
- ('governmental_improvement_value', 'nonresidential_land_value')
- ('v_nb_nodes_train_90', 'v_nb_nodes_train_99')
- ('v_nb_nodes_subway_90', 'v_nb_nodes_subway_99')
- ('TVP_90', 'TVP_99')
- ('TTC_90', 'TTC_99')
- ('distance_a_chatelet', 'TVP_90')
- ('distance_a_chatelet', 'TVP_99')

Les 11 premiers cas sont des variables corrélées par construction. Par conséquent, nous gardons celles qui sont élaborées directement à partir des données disponibles de base. Les

quatre suivantes sont corrélées par le fait que l'état du système de transport en Ile-de-France n'a pas subi de modifications structurelles majeures en 10 ans. Nous utilisons alors celle qui représente mieux la situation au moment du choix des ménages. Le cas des deux dernières est plus difficile. Nous souhaiterions bien utiliser les méthodes économétriques qui nous permettent de distinguer les effets propres de chacune. Mais le nombre des observations dans l'Enquête Globale Transports n'est sans doute pas suffisant pour le modèle de localisation des ménages. Faute d'utiliser ces deux variables simultanément, nous arbitrerons sur des critères de bons sens. A ce titre, il est plausible que les ménages soient plus sensibles au temps de déplacement qu'à la distance à vol d'oiseau.

6.1.3. Sélection des variables explicatives

Il y a 39 variables à explorer (voir partie 5.2.4).

Pour plus de robustesse et de simplicité, toutes les variables disponibles ne doivent pas être introduites dans le modèle en même temps. La procédure dite du « stepwise » est un algorithme qui permet de sélectionner automatiquement les variables explicatives les plus pertinentes. Le principe de cette méthode est de partir du modèle contenant seulement la constante puis d'ajouter une à une les variables les plus significatives jusqu'à ce qu'aucune ne soit significative ou qu'il n'y en ait plus à inclure. Cette procédure autorise une variable déjà présente à l'étape n à en sortir à l'étape $n + 1$ si elle n'est plus significative. Cette procédure a permis de sélectionner les variables qui disposent a priori du plus grand pouvoir explicatif lorsqu'elles sont prises en compte simultanément. Le tableau 14 donne les résultats de la procédure du stepwise.

Résumé de la procédure dite du "stepwise"							
Itération	Effet		DF	Nb variables	Score Chi-Square	Wald Chi-Square	p-value
	Variable ajoutée	Variable supprimée					
1	v_industrial_sqft		2	1	3113,1089	.	<0,0001
2	residential_units		2	2	1359,6794	.	<0,0001
3	TVP_99		2	3	271,3436	.	<0,0001
4	distance_to_highway		2	4	123,8352	.	<0,0001
5	industrial_sqft		2	5	89,8033	.	<0,0001
6	commercial_improvement_value		2	6	66,9696	.	<0,0001
7	TTC_99		2	7	54,8426	.	<0,0001
8	nb_nodes_train_99		2	8	43,6351	.	<0,0001
9	v_residential_units		2	9	17,2661	.	0,0002

Tableau 15 - procédure dit du "stepwise"

La procédure dite du « stepwise » permet de classer les variables selon leur degré de significativité. Ici, la variable la plus significative est la variable de voisinage associée aux surfaces d'activités secondaires ('v_industrial_sqft'), suivie de la variable nombre d'unités résidentielles ('residential_units'), puis de la variable temps de parcours en voiture de la cellule à Châtelet 'TVP_99' etc. Enfin, la variable la moins significative est la variable de voisinage associée au nombre d'unités résidentielles 'v_residential_units'. Par conséquent, cette variable n'est pas retenue dans le modèle. La variable 'industrial_sqft' n'est pas retenue non plus car elle n'est pas significative pour la transition du type 5 vers le type 1 + 3 'habitat'.

Les variables d'accessibilité suivantes ont été exclues du modèle :

- nb_nodes_tramway_90 = nombre de stations de tramway en 1990
- nb_nodes_tramway_99 = nombre de stations de tramway en 1999
- nb_nodes_subway_90 = nombre de stations de métro en 1990
- nb_nodes_subway_99 = nombre de stations de métro en 1999
- v_nb_nodes_subway_90 = variable de voisinage du nombre de stations de métro en 1990
- v_nb_nodes_subway_99 = variable de voisinage du nombre de stations de métro en 1999
- v_nb_nodes_tramway_90 = variable de voisinage du nombre de stations de tramway en 1990
- v_nb_nodes_tramway_99 = variable de voisinage du nombre de stations de tramway en 1999

Le nombre de stations de métro et le nombre de stations de tramway n'ont pas d'influence ici sur l'évolution des cellules vacantes en 1990. Ceci s'explique par le fait que ces stations sont situées sur des cellules qui étaient déjà urbanisées en 1990 et qui par conséquent n'ont pas été considérées comme étant de type initial vacant.

Au total, sur les 39 variables disponibles, les 7 variables explicatives suivantes ont été retenues :

- distance_to_highway = distance de la cellule à l'autoroute ou voie rapide la plus proche
- nb_nodes_train_99 = nombre de gares en 1999
- TVP_99 = temps de parcours moyen en voiture (heures creuses) nécessaire pour aller de la cellule à Châtelet
- TTC_99 = temps de parcours moyen en transport en commun nécessaire pour aller de la cellule à Châtelet
- residential_units = nombre de logements
- commercial_improvement_value = loyers des surfaces commerciales
- v_industrial_sqft = variable de voisinage des surfaces d'activités secondaires

Dans son étude sur Puget Sound Region, l'équipe du Pr Waddell ne retenait également que 7 variables explicatives pour l'estimation du modèle logit multinomial du type 24 'vacant développable'. De plus, les variables sélectionnées étaient proches de celles qui ont été retenues ici puisqu'il y avait des variables de distance, des variables de voisinage et des

variables de composition de la cellule. Cependant, on peut regretter une forte opacité dans la présentation des données et des résultats dans le cas américain.

6.1.4. Résultats du calage

Les tableaux 15 et 16 donnent les résultats des estimations du modèle logit 5 ainsi que les p_value et les écart-types associés aux coefficients estimés des variables explicatives. La p_value s'interprète de la façon suivante :

Nota : Une variable est significative à un seuil de x% si et seulement si la p_value qui lui est associée est inférieure ou égale à x%.

Transition du type 5 'vacant' vers le type 1+ 3 'habitat'			
Variable	Estimation	Ecart-type	p_value
Constante	-1,8561	0,1661	<0,0001
Residential_Units	0,067	0,00283	<0,0001
Commercial_Improvement_value	0,00373	0,00225	0,0975
V_Industrial_sqft	6,43 ^E -06	1,04 ^E -06	<0,0001
TTC_99	-0,4504	0,0846	<0,0001
TVP_99	-1,2164	0,1852	<0,0001
Distance_to_highway	-0,0148	0,0085	0,0817
Nb_nodes_train_99	1,9621	0,5809	0,0007

Tableau 16 - résultat pour les transitions de type 5 à type 1+3

Toutes les variables explicatives retenues sont significatives au seuil de 10% . De plus, les coefficients estimés associés aux variables ont les signes attendus. Ainsi, les probabilités qu'une cellule de type 'vacant' passe à l'un des types agrégés 'habitat', 'activités', sont des fonctions :

- décroissantes du temps de parcours nécessaire en voiture pour rejoindre le centre de Paris (TVP_99)
- décroissantes du temps de parcours nécessaire en transport en commun pour rejoindre le centre de Paris (TTC_99)
- décroissantes de la cellule à l'autoroute ou voie rapide la plus proche (distance_to_highway)
- croissantes du nombre de gares en 1999 (nb_nodes_train_99)
- croissantes du nombre d'unités résidentielles (Residential_Units)
- croissantes des loyers de surfaces commerciales (Commercial_Improvement_Value)

- croissantes des surfaces d'activités secondaires adjacentes à la cellule (V_Industrial_sqft)

Transition du type 5 'vacant' vers le type 2+ 4 'activités'			
Variable	Estimation	Ecart-type	p_value
Constante	-2,014	0,1959	<0,0001
Residential_Units	0,0395	0,00388	<0,0001
Commercial_Improvement_value	0,00686	0,00195	0,0004
V_Industrial_sqft	0,000015	8,15 ^E -07	<0,0001
TTC_99	-0,6501	0,1193	<0,0001
TVP_99	-0,5536	0,2298	0,016
Distance_to_highway	-0,1648	0,0159	<0,0001
Nb_nodes_train_99	1,9621	0,5809	0,0007

Tableau 17 - résultats d'estimation pour la transition de type 5 à 2+4

R-square =0,0942
Max-rescaled R-square =0,2724

Ce modèle met en évidence l'importance des variables d'accessibilité : plus il est coûteux en temps de rejoindre le centre de Paris et moins il est probable que de l'habitat ou des bureaux s'installent dans la cellule considérée. De façon symétrique, plus il y a de gares ou plus la cellule est proche d'une autoroute ou d'une voie rapide, plus il est aisé de rejoindre le centre de Paris et plus la cellule est susceptible de s'urbaniser. Les loyers des surfaces commerciales jouent également un rôle non négligeable. En effet, lorsque ces loyers sont élevés, cela signifie que la cellule est confrontée à une forte demande de location et cela reflète un attrait élevé pour la zone considérée. Par conséquent, plus ces variables sont élevées et plus il est probable que des bureaux ou de l'habitat s'implantent dans la cellule. De même, plus le nombre d'unités résidentielles ou plus la surface d'activités secondaires proches sont élevées, plus la cellule est attractive et susceptible de ne pas rester vacante.

Hausse d'1 heure du temps de parcours en voiture jusqu'à Châtelet	
Odds Ratio	Estimation
$P_{55} / P_{5.13}$	1,7
$P_{55} / P_{5.24}$	3,4
$P_{5.13} / P_{5.24}$	2

Hausse d'1 heure du temps de parcours en transport en commun jusqu'à Châtelet	
Odds Ratio	Estimation
$P_{55} / P_{5.13}$	1,6
$P_{55} / P_{5.24}$	1,9
$P_{5.13} / P_{5.24}$	1,2

Tableau 18 - ratios de chance pour les temps de parcours VP et TC

Le Tableau 18 donne les valeurs des ratios de chance (Odds Ratio) lorsque les temps de parcours en voiture (respectivement en transport en commun) jusqu'à Châtelet augmentent d'une heure et confirment le rôle joué par les variables d'accessibilité. En effet, lorsque le temps de parcours en voiture nécessaire pour rejoindre le centre de Paris augmente d'une heure, la probabilité que la cellule reste vacante (P_{55}) est 1,7 fois plus élevée que la probabilité de transition vers le type 'habitat' ($P_{5,13}$). De même, la probabilité que la cellule reste vacante est 3,4 fois plus élevée que la probabilité de transition vers le type 'activités' ($P_{5,24}$). Il est également constaté que la probabilité de transition vers le type 'habitat' est 2 fois plus élevée que la probabilité de transition vers le type 'activités'. Ce dernier résultat met en évidence le fait que l'habitat a plutôt tendance à se développer à l'extérieur de Paris tandis que les bureaux s'implantent plutôt sur Paris même. Lorsque ce temps de parcours en transport en commun augmente d'une heure, les effets constatés sont similaires à savoir qu'il devient plus probable que la cellule reste vacante et que l'habitat se développe plus à l'extérieur de Paris que les bureaux.

6.1.5. Validité du modèle

Le R^2 ajusté (Max-rescaled R-square) : un premier indicateur de la validité du modèle difficilement interprétable...

Le logiciel SAS donne deux indicateurs de la validité du modèle qui sont le R^2 et le R^2 ajusté. Le R^2 est calculé de la façon suivante¹³ :

$$R^2 = 1 - \left(\frac{L(0)}{L(\hat{\theta})} \right)^{\frac{2}{n}}$$

où :

- $L(0)$ est la valeur de la vraisemblance¹⁴ lorsque le modèle n'est composé que de la constante,
- $L(\hat{\theta})$ est la valeur de la vraisemblance du modèle spécifié,
- n est le nombre d'observations total.

Si le modèle spécifié n'est pas meilleur que le modèle qui n'est composé que de la constante (i.e. pas de modèle), alors $L(0) = L(\hat{\theta})$ et $R^2 = 0$. Ceci est la valeur la plus faible que peut prendre le R^2 .

Inversement, si le modèle spécifié prévoit parfaitement les décisions de chaque individu, alors $L(\hat{\theta}) = 1$ puisque la probabilité d'observer les choix qui ont réellement eu lieu est de un. Par conséquent, la valeur maximale du R^2 est inférieure à un et vaut :

$$R_{\max}^2 = 1 - (L(0))^{\frac{2}{n}}$$

¹³ Cette méthodologie a été proposée par Cox et Snell (1989, pp. 208-209)

¹⁴ Rappel : la vraisemblance est le produit des probabilités des alternatives choisies par les individus

Le R^2 ajusté¹⁵ est calculé de façon à prendre comme valeur maximale la valeur un :

$$\tilde{R}^2 = \frac{R^2}{R_{\max}^2}$$

Le R^2 ajusté est plus facilement interprétable que le R^2 du fait de sa normalisation. Il prend des valeurs comprises entre 0 et 1. Lorsque le modèle spécifié n'apporte aucune information supplémentaire par rapport à un modèle qui n'est composé que de la constante, alors $\tilde{R}^2 = 0$. Inversement, quand le modèle spécifié prévoit parfaitement les choix des individus alors $\tilde{R}^2 = 1$.

Il est important de remarquer que le R^2 ajusté ne doit pas être interprété de la même façon que le R^2 utilisé habituellement dans les régressions linéaires. Le R^2 usuel s'interprète comme le pourcentage de variation de la variable dépendante qui est expliqué par le modèle estimé. Ici, le R^2 ajusté ne peut pas être interprété de façon intuitive pour des valeurs autres que zéro et un puisqu'il est calculé à partir d'un rapport de vraisemblance. Une variation du R^2 ajusté est liée une variation du rapport de vraisemblance $\frac{L(0)}{L(\hat{\theta})}$, variation qui est elle-

même difficilement interprétable. En conclusion, il faut retenir que le modèle estimé est d'autant meilleur que le R^2 ajusté est élevé et ne pas se servir de cet indicateur pour comparer deux modèles avec des valeurs de $L(0)$ différentes.

Le R^2 ajusté du modèle logit 5 estimé vaut 0,2724. C'est un résultat correct compte tenu du faible nombre d'observations pour les cellules ayant varié.

6.1.6. Le pourcentage de prédictions correctes

Il est bien connu qu'un modèle explicatif n'est pas forcément prédictif. Nous introduisons donc un deuxième indicateur de la validité du modèle : le pourcentage de prédictions correctes. Cette statistique est calculée en identifiant pour chaque cellule l'alternative correspondant à la probabilité la plus élevée estimée par le modèle. Cette alternative est ensuite comparée à celle qui est réellement observée pour la cellule. Le pourcentage d'alternatives égales aux observations donne le pourcentage de prédictions correctes.

Le tableau 5 donne le détail des prédictions correctes qui ont été obtenues à partir de l'estimation du modèle logit 5.

Transition	Nb observations	Nb prédictions correctes	Nb prédictions incorrectes
5.13	722	60	5.24 : 11 5.5 : 651
5.24	426	59	5.13 : 14 5.5 : 353
55	24504	24383	5.13 : 91 5.24 : 30

Tableau 19 - pourcentage de bonne prédiction

¹⁵ Ce R^2 ajusté a été proposé par Nagelkerke (1991)

Le pourcentage de prédictions correctes est de 96%. Néanmoins, parmi les cellules qui ont varié, le pourcentage de prédictions correctes n'est que de 10%. De plus, lorsque le modèle fournit des prédictions erronées, dans 87% des cas il prévoit à tort que la cellule ne va pas varier. Ce résultat, décevant, n'est néanmoins pas très surprenant compte tenu du faible nombre de cellules observées ayant varié. Dans son étude sur Puget Sound, l'équipe du Pr Waddell obtenait de meilleurs résultats mais aucune information est donnée sur la taille des bases de calage.

En sélectionnant les cellules observées qui ont varié, il est possible d'estimer à partir du même modèle les probabilités de transition conditionnelles au fait qu'une cellule ait varié :

$$\hat{P}(5j / j \neq 5) = \frac{\hat{P}_{5j}}{1 - \hat{P}_{55}}$$

Dans ce cas, le pourcentage de prédictions correctes est de 76%. Ce résultat confirme le fait que le modèle donne des résultats décevants à cause de la faible proportion de cellules observées qui ont varié.

En conclusion, même s'il est explicatif, le modèle proposé ne peut pas être utilisé à des fins de prévision puisque les transitions vers des états autre que l'état vacant ne sont correctement prédites par le modèle que dans 10% des cas.

A partir de ce constat, nous avons réfléchi à une amélioration du modèle, présentée dans le paragraphe suivant.

6.1.7. Amélioration proposée du modèle

Le modèle proposé fournit des résultats décevants très probablement à cause de la faible proportion de cellules qui ont varié par rapport aux cellules qui sont restées dans l'état vacant. Il est alors possible de se poser la question suivante : faut-il garder toutes les cellules invariantes pour estimer le modèle logit 5 ? On pense que non, et pour 2 raisons :

- 1) Une des hypothèses des modèles logit multinomiaux est l'indépendance des perturbations. Or, ici, il est probable que cette hypothèse ne soit pas vérifiée. En effet, il est possible qu'un facteur inobservé tel que le déficit d'image d'un secteur explique qu'un certain nombre de cellules restent vacantes. Dans ce cas, il n'y aurait pas indépendance ni des cellules ni des perturbations. Un des moyens de pallier le problème consiste à ne retenir qu'un certain nombre de cellules qui sont restées vacantes. Ces cellules doivent être sélectionnées de façon à ce qu'elles ne soient pas reliées par un facteur inobservable qui serait propre à une zone géographique. Pour cela, il suffit de retenir des cellules qui sont suffisamment éloignées les unes des autres par un tirage aléatoire équiprobable.
- 2) Sur les 24504 cellules qui n'ont pas varié, un grand nombre sont identiques. Par exemple, une cellule qui a 100% de sa surface occupée par des cultures, donc classée vacante, apparaît presque 2000 fois. C'est ce poids des cellules identiques qui ne permet pas de faire ressortir les cellules qui ont évolué.

On a donc procédé à un tirage aléatoire d'environ 10% de cellules parmi les 24504 cellules qui n'ont pas varié¹⁶, soit 2500 cellules. Au total, l'estimation du logit 5 porte maintenant sur 2500 cellules invariantes et 1148 cellules qui ont changé de type.

Les Tableau 20 et Tableau 21 donnent les **résultats de l'estimation de ce nouveau modèle**.

Transition du type 5 'vacant' vers le type 1+ 3 'habitat'			
Variable	Estimation	Ecart-type	p_value
Constante	0,1403	0,216	0,516
Residential_Units	0,1522	0,0065	<0,0001
Commercial_Improvement_value	0,00788	0,00703	0,262
V_Industrial_sqft	8,21 ^E -06	1,71 ^E -06	<0,0001
TTC_99	-0,2777	0,0982	0,0047
TVP_99	-1,5902	0,2408	<0,0001
Distance_to_highway	-0,0251	0,0109	0,0219
Nb_nodes_train_99	1,3901	0,9859	0,1585

Tableau 20 - les résultats de nouveau modèle (type 5-> 1+3)

Remarque : comme la modalité de référence est p_{55} , la probabilité estimée de transition du type 5 vers le type 1 + 3 pour une cellule i peut être calculée de la façon suivante

$$P_{i,5.13} = \frac{\exp(X_i' \beta_{5,13})}{\exp(X_i' \beta_{5,13}) + \exp(X_i' \beta_{5,24}) + 1} \text{ notée } p_{513}$$

où

$$p_{513} = \frac{\exp(-0.1403 + 0.1522 * residential_units + 8.206E-6 * v_industrial_sqft - 0.2777 * TC_99 - 1.5902 * tv_99 + 0.00788 * commercial_improvement_value - 0.0251 * distance_to_highway + 1.3901 * nb_nodes_train_99)}{\exp(-0.3088 + 0.1033 * residential_units + 0.000018 * v_industrial_sqft - 0.3837 * tc_99 - 0.8193 * tv_99 + 0.0116 * commercial_improvement_value - 0.1648 * distance_to_highway + 1.8614 * nb_nodes_train_99) + \exp(-0.1403 + 0.1522 * residential_units + 8.206E-6 * v_industrial_sqft - 0.2777 * TC_99 - 1.5902 * tv_99 + 0.00788 * commercial_improvement_value - 0.0251 * distance_to_highway + 1.3901 * nb_nodes_train_99) + 1}$$

¹⁶ Ce tirage aléatoire a été exécuté sous Scilab à l'aide de la fonction 'samwr'

Transition du type 5 'vacant' vers le type 2+ 4 'activités'

Variable	Estimation	Ecart-type	p_value
Constante	-0,3088	0,2316	0,1824
Residential_Units	0,1033	0,0073	<0,0001
Commercial_Improvement_value	0,0116	0,00696	0,0947
V_Industrial_sqft	0,000018	1,59E-06	<0,0001
TTC_99	-0,3837	0,1207	0,0015
TVP_99	-0,8193	0,2571	0,0014
Distance_to_highway	-0,1648	0,0168	<0,0001
Nb_nodes_train_99	1,8614	0,9778	0,057

Tableau 21 - les résultats de nouveau modèle (type 5-> 2+4)

R-square = 0.4426 Max-rescaled R-square=0.5464

Remarque : comme la modalité de référence est p_{55} , la probabilité estimée de transition du type 5 vers le type 1 + 3 pour une cellule i peut être calculée de la façon suivante

$$P_{i,5,24} = \frac{\exp(X'_i \beta_{5,24})}{\exp(X'_i \beta_{5,13}) + \exp(X'_i \beta_{5,24}) + 1} \text{ notée } p_{5,24}$$

où

$$p_{5,24} = \frac{\exp(-0.3088 + 0.1033 \cdot \text{residential_units} + 0.000018 \cdot \text{v_industrial_sqft} - 0.3837 \cdot \text{tc_99} - 0.8193 \cdot \text{tv_99} + 0.0116 \cdot \text{commercial_improvement_value} - 0.1648 \cdot \text{distance_to_highway} + 1.8614 \cdot \text{nb_nodes_train_99})}{(\exp(-0.1403 + 0.1522 \cdot \text{residential_units} + 8.206E-6 \cdot \text{v_industrial_sqft} - 0.2777 \cdot \text{TC_99} - 1.5902 \cdot \text{tv_99} + 0.00788 \cdot \text{commercial_improvement_value} - 0.0251 \cdot \text{distance_to_highway} + 1.3901 \cdot \text{nb_nodes_train_99}) + \exp(-0.3088 + 0.1033 \cdot \text{residential_units} + 0.000018 \cdot \text{v_industrial_sqft} - 0.3837 \cdot \text{tc_99} - 0.8193 \cdot \text{tv_99} + 0.0116 \cdot \text{comval_90} - 0.1648 \cdot \text{dhigh_90} + 1.8614 \cdot \text{train_99}) + 1)}$$

La probabilité p_{55} se déduit simplement du calcul des 2 autres probabilités :

$$p_{i,55} = 1 - p_{i,5,13} - p_{i,5,24}$$

Toutes les variables restent significatives au seuil de 10% pour le passage du type 5 au type 2 + 4 . Toutes les variables restent également significatives au seuil de 10% pour le passage du

type 5 au type 1 + 3 à l'exception des variables loyers des surfaces commerciales ('commercial_improvement_value') et nombre de gares ('nb_nodes_train_99'). Les coefficients estimés ont les signes attendus. Leur interprétation est similaire à celle du modèle précédent.

Le Tableau 22 donne le détail des prédictions correctes et incorrectes qui ont été obtenues à partir de l'estimation du nouveau modèle logit 5 et ce en utilisant ce modèle en prévision sur l'ensemble des cellules, soit 25562 cellules.

Transition	Nb observations	Nb prédictions correctes	Nb prédictions incorrectes
5.13	722	440	5.24 : 42 5.5 : 240
5.24	426	124	5.13 : 65 5.5 : 237
5.5	24504	23488	5.13 : 823 5.24 : 193

Tableau 22 - pourcentage de bonne prédiction

Le pourcentage de prédictions correctes est de 94%. Parmi les cellules qui ont varié, le pourcentage de prédictions correctes est de 50%. La méthode de tirage aléatoire proposée a permis d'améliorer considérablement les résultats. Par conséquent, **c'est ce modèle qui sera retenu pour faire de la prévision dans SIMAURIF.**

6.2. Calage du module de localisation des ménages

6.2.1. Le logit multinomial conditionnel

Le but est d'estimer la probabilité qu'un ménage *i* emménage dans une cellule *j*, connaissant les caractéristiques du ménage, les caractéristiques de la cellule *j* et les critères d'accessibilité et de temps de parcours.

La procédure MDC

Cette procédure permet d'estimer aisément différents modèles de choix discrets. Afin d'estimer un modèle logit multinomial conditionnel, il faut ajouter l'option **type=clogit**. L'option **choice=(choice_set)** permet d'indiquer l'ensemble des choix possibles pour les individus. Enfin, l'option **id** permet d'indiquer l'identifiant des individus.

La syntaxe est la suivante :

```
proc mdc data=nom_table;

    model var_expliquee = var_explicative_1 var_explicative_2/
    type=clogit choice=(choice_set);

    id num;
run;
```


6.2.2. Sélection des alternatives

La base de calage a été construite comme suit : pour chaque observation, on fait un tirage aléatoire équiprobable de 6 alternatives parmi les 9000 cellules de type 'habitat' en 1999 (type agrégé défini dans le modèle d'évolution du MOS). Ceci permettra de chaîner le modèle d'évolution du MOS et le modèle de localisation résidentielle des ménages.

En fait, une première base avait été construite en effectuant un tirage aléatoire de 9 alternatives par observation, pondéré (comme le demande la théorie) par le nombre de logements par cellule. Cette méthode est appliquée dans la construction de la base de calage pour l'estimation du modèle de localisation résidentielle à la commune (voir partie 7.3). Les premières estimations n'ont pas abouti à des résultats concluants : pas de convergence la plupart des cas, 2 à 4 variables au maximum incorporables par test, signes des coefficients contre-intuitifs, pas de robustesse des variables significatives si l'on combine avec d'autres variables, franchissement fréquent du seuil de significativité à 5%. C'est pourquoi la méthode de tirage aléatoire équiprobable a été utilisée pour créer les alternatives. Notons aussi que dans l'étude Puget Sound, le tirage aléatoire a été effectué sur les logements vacants par cellule.

Il y a 52 variables à explorer (voir partie 5.3.4). Les estimations économétriques sont réalisées sans utiliser la pondération POIDS.

6.2.3. Sélection des variables explicatives

En plus des variables prises isolément, les variables caractéristiques du ménage sont utilisées pour construire des effets croisés avec les caractéristiques des cellules. Par exemple, on pourra tester le fait que les ménages à haut revenu ont une faible probabilité à s'installer dans des cellules concernées par la Politique de la ville, en croisant la variable MRTV et ZUS.

6.2.4. Résultats du calage

Plusieurs dizaines de combinaisons de variables ont été testées, en particulier le croisement du nombre d'actifs dans le ménage avec le nombre d'unités résidentielles en voisinage, et le croisement de la variable revenu (classée en 3 types : faible, moyen élevé) avec la valeur foncière en voisinage ou la variable ZUS (zone urbaine sensible).

Les premières estimations ont mis en lumière le modèle suivant. Il reste cependant à approfondir, de nombreuses combinaisons de variables croisées n'ayant pas encore été explorées à la date du 15 septembre 2005. Mais on souhaite quoi qu'il en soit retenir un modèle simple, avec un nombre limité de variables utilisables en prédiction

Estimation du modèle de localisation des ménages

Variable	Estimation	Ecart-type	p_value
emploiTC	0,1202	0,0118	<0,0001
vresu_mnact	7,44E-06	1,58E-06	<0,0001
vtrain_99	0,0936	0,0202	<0,0001
vsubway_99	0,0445	0,0103	<0,0001
eleve_rev_vresval	2,50E-07	3,93E-08	<0,0001
TVP_99	-0,6071	0,1825	0,0009
vcom_99	4,91E-06	2,00E-06	0,014
vind_99	2,33E-06	1,42E-07	<0,0001
vgov_99	3,60E-07	1,23E-07	0,0033
ZUS_faible_rev	3,72E-07	6,21E-08	<0,0001
dhigh_99	-0,0269	0,008654	0,0019
dart_99	-0,0573	0,0102	<0,0001

Tableau 23 - résultat du modèle de localisation résidentielle

Le modèle présenté comporte donc 12 variables explicatives. Toutes ces variables sont significatives au seuil de 5% et ont les signes attendus. Ainsi, l'utilité d'un ménage qui s'installe dans une cellule de la région d'Ile-de-France est une fonction croissante de :

- l'accessibilité à l'emploi en TC
- du nombre de gares
- du nombre de stations de métro
- du nombre de commerces
- du nombre de bureaux et d'équipements

Symétriquement, l'utilité d'un ménage est une fonction décroissante :

- du temps de transport en voiture nécessaire pour rejoindre le centre de Paris
- de la distance à l'autoroute la plus proche
- de la distance à la voie rapide la plus proche

Ce modèle met également en évidence le fait que les ménages avec de faibles revenus s'installent souvent en zone urbaine sensible. Les ménages à haut niveau de revenu, quant à eux, ont une préférence pour les zones géographiques qui ont une valeur foncière élevée.

Enfin, les variables de transports (nombre de gares, de stations de métro et d'accessibilité en TC aux emplois) sont significatives, ce qui est très encourageant pour le modèle *SIMAU*RIF qui présuppose l'existence d'une interaction entre urbanisation et transport.

Deux variables caractéristiques des ménages ressortent : le revenu et le nombre d'actifs par ménage.

6.2.5. Validité du modèle

Le pseudo R^2 de MacFadden est égal à 0,2475. Il est acceptable. Dans l'étude de Puget Sound, le pseudo- R^2 n'est pas donné. Mais on a retrouvé une étude ancienne du Pr Paul Waddell présentant un pseudo R^2 de 0,13 et considéré comme correct. (" Firouzeh Nourzad. Incorporating Non-motorized Mode and Neighborhood Accessibility in an Integrated Land Use and Transportation Model System", *Transportation Research Record*, Mars 2002).

6.2.6. Le pourcentage de prédictions correctes

Il s'agit de faire tourner le modèle sur la base des 5497 observations et calculer le pourcentage de prédictions correctes agrégé au département.

Département où les ménages se sont installés	Nb d'emménagements observés	Nb d'emménagements prévus	% d'emménagements prévus
75	994	938	94,4
77	460	52	11,3
78	579	113	19,5
91	478	101	21,1
92	709	522	73,6
93	508	302	59,4
94	550	322	58,5
95	523	169	32,3

On constate que le taux de prédictions correctes est très satisfaisant sur Paris et la petite couronne. En revanche, le modèle reconstitue mal la mobilité résidentielle vers les départements de la grande couronne, même si une analyse succincte des prédictions incorrectes montre que l'équilibre au sein de la grande couronne est globalement respectée. C'est-à-dire que des ménages qui devaient se localiser par exemple dans le 77 se retrouvent par le modèle localisés dans le 78. Ceci nous incite à approfondir ces premiers résultats, en testant de nouvelles variables croisées et en ajoutant de nouvelles variables. Les résultats seront présentés dans le rapport final.

7. Un nouvel UrbanSim

7.1. Méthodologie

Devant l'inadaptation de la version d'UrbanSim ne travaillant qu'au niveau de cellules de taille fixe (500 * 500 m dans le cas de l'Ile de France), Paul Waddell a travaillé cette année sur une nouvelle version d'UrbanSim permettant de travailler au niveau d'unités de taille variable (le département, la commune et l'îlot INSEE dans le cas de l'Ile-de-France). Ce nouvel UrbanSim utilisera de façon optimale l'information initialement contenue dans les données. Par contre, elle ne prendra plus en compte la notion de voisinage.

Le travail présenté ici est essentiellement la **partie économétrique**, avec une analyse des variables explicatives plus approfondie que dans le travail sur l'actuel UrbanSim. Il reste à l'intégrer dans une architecture globale pour rendre les différents modèles prédictifs.

Pour la partie modèle de demande de déplacements, le THEMA a réfléchi à un nouveau modèle de demande, plus sophistiqué que celui de l'IAURIF. C'est le modèle EVA, présenté dans la partie 7.5.

7.2. Nouvelle architecture

La nouvelle architecture sera présentée dans le rapport final de juin 2006. Notons qu'elle sera fondée sur les principes suivants :

- La structure de données sera basée sur des découpages existants.
- Le modèle de développement urbain sera un modèle de choix de localisation de projets à l'îlot. Un projet est défini par son type urbain et par sa taille.
- Les modèles de localisation des ménages et des emplois seront des « Nested Multinomial Logit » ou Logit Multinomial Emboîté : choix de déménager ou pas, du département éventuellement, de la commune puis de l'îlot
- Le modèle de prix sera estimé au niveau de la commune.

7.3. Le module de localisation résidentielle des ménages au niveau communal

7.3.1. Modèle de localisation en trois étapes

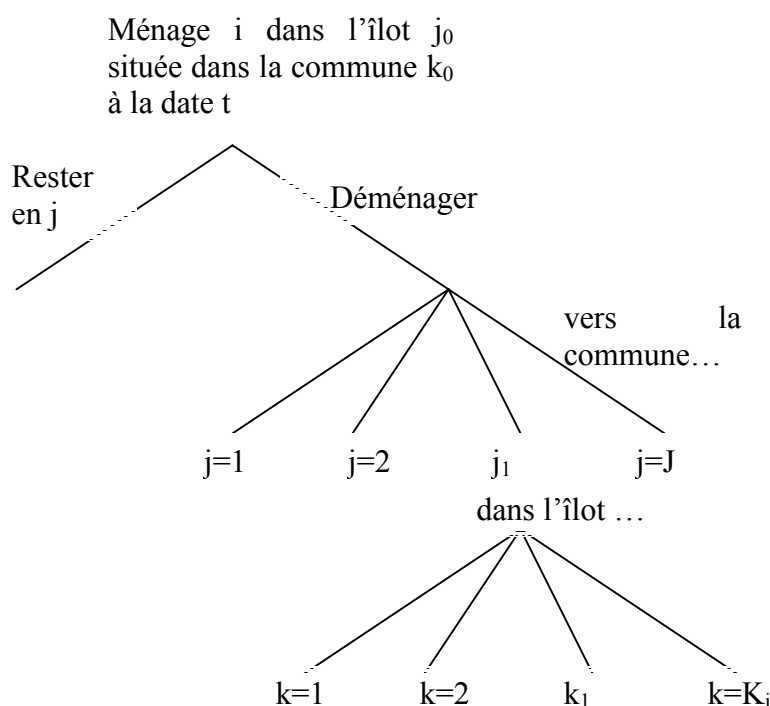
Il s'agit d'estimer :

- d'une part la probabilité de déménager au cours d'une période donnée en fonction des caractéristiques du ménage et des attributs de la localisation de départ
- d'autre part la localisation d'arrivée pour les seuls ménages qui déménagent en fonction de leurs caractéristiques et des attributs de la localisation d'arrivée

La localisation d'arrivée peut être définie à différents niveaux géographiques, en fonction de l'information disponible. L'objectif est d'optimiser l'utilisation de l'information contenue dans les données initialement disponibles. Dans cette optique, le choix de localisation résidentielle conditionnellement au fait de déménager est supposé se faire en deux étapes :

- Choix d'une commune en fonction des caractéristiques de la commune
- Choix d'un îlot dans la commune choisie en fonction des caractéristiques des îlots

Les alternatives offertes pour le choix de l'îlot dépendent de la commune choisie à l'étape précédente et cet effet est anticipé lors du choix de la commune. De même, la décision de déménager dépend des opportunités offertes au niveau communal (et donc aussi au niveau des îlots) en cas de déménagement. Il s'agit donc finalement d'estimer un modèle en 3 étapes, ce qui fait naturellement appel à un « Nested Multinomial Logit » (Logit Multinomial Emboîté).



Dans la version finale du nouvel UrbanSim, il s'agira d'estimer un modèle Multinomial Logit emboîté prenant en compte les liens statistiques entre les trois décisions. Pour l'instant, dans ce rapport intermédiaire, nous nous sommes concentrés sur la seconde étape.

La possibilité d'une estimation en trois étapes, avec définition des variables à plusieurs niveaux géographiques (à partir d'une information initiale qui peut être disponible à des niveaux géographiques différents pour les différentes variables), est en cours d'implémentation dans le nouvel Urbansim par le Pr Paul Waddell.

Les résultats présentés dans ce rapport intermédiaire se concentrent sur la seconde étape de l'estimation, à savoir l'estimation au niveau communal, à partir de variables définies au niveau communal, en se limitant aux ménages qui ont déménagé sur une période donnée.

7.3.2. Méthode d'estimation

Le calage global ne tient pas compte des variables d'accessibilité locale calculées automatiquement par UrbanSim pour chaque cellule carrée, ni des autres caractéristiques des cellules, mais il met l'accent sur les caractéristiques communales, au premier rang desquelles l'accessibilité calculée au niveau de la commune.

Comme nous nous sommes concentrés dans cette version sur la seconde étape du modèle (choix de la commune conditionnellement au fait de déménager), nous avons utilisé un modèle multinomial logit, toujours estimé à l'aide du logiciel SAS. La probabilité que le ménage i choisisse la commune j est calculée par la formule suivante :

$$P_j^i = \frac{\exp(V_j^i)}{\sum_k \exp(V_k^i)},$$

où l'utilité de déménager dans la commune j pour le ménage i est donnée par :

$$U_j^i = V_j^i + \varepsilon_j^i = Z_j^i \beta + P_j \gamma + \varepsilon_j^i.$$

Avec :

- Z_j^i : attributs de l'alternative j , éventuellement croisées avec des caractéristiques du ménage.
- P_j : prix moyen au m² du bâti dans la commune j . En prédiction (exécution de *SIMAU*RIF), il sera une sortie du modèle du prix du foncier. .

7.3.3. Base de calage

On considère dans la conception de ce nouvel UrbanSim que l'on n'est plus du tout dépendant de la structure de données de l'actuel d'UrbanSim. Ainsi, les variables de gridcells ne correspondent plus à celles que l'on va définir au niveau des communes, même si quelques variables d'occupation du sol ont été récupérées. La base de données caractéristiques des communes s'est avérée relativement facile à construire, sachant que parmi les données collectées en 49 236 cellules, beaucoup étaient au départ des données à la commune.

En ce qui concerne les caractéristiques socio-économiques des ménages, la source de données est le Recensement général de la population de 1999 et non l'EGT comme cela a été fait dans le modèle de localisation à la cellule de l'actuel UrbanSim.

Les déterminants des choix de localisation incluent des variables mesurant des effets croisés, construites à partir des caractéristiques des ménages et des caractéristiques communales.

Au final, la base comporte 589 355 ménages ayant déménagé en 1998. Le nombre d'alternatives étant beaucoup trop élevé (1300), nous avons procédé à un tirage aléatoire de 7 alternatives pour chaque ménage, en plus de l'alternative qu'il a choisie. La probabilité de chaque commune d'être tirée est proportionnelle au nombre de logements dans la commune, ce qui correspond à un tirage pondéré en fonction du nombre de logements. Cette méthode de tirage aléatoire est différente de celle réalisée pour la construction de la base de calage du modèle au niveau de la cellule (paragraphe 6.2.1).

Malgré cette réduction du nombre d'alternatives, la taille de la base de données nécessaire aux estimations dépasse 5 Giga-octets, ce qui conduit à des temps de calcul de plusieurs heures pour chaque estimation.

Une analyse de sensibilité ayant montré que les résultats ne sont pas significativement différents sur un échantillon de 20% de la population, nous avons effectué la plupart de nos estimations sur un tel sous-échantillon comportant 117 872 ménages.

Les prix estimés sont extraits de la base de donnée des notaires complétée comme expliqué à la section sur les prix, en sélectionnant les 5 200 observations de l'année 1998.

7.3.3.1. Variables comportant un « effet-taille »

Dans l'actuelle version d'Urbansim, toutes les cellules ont la même surface, ce qui implique l'absence d'effets-taille dans les données. Qui plus est, dans ce cas, les densités sont synonymes de quantités. Par exemple, le nombre de logements dans une cellule carrée de 500 m est synonyme de densité de logements sur cette surface. Il en va autrement lorsque l'unité d'observation est la commune, dont la superficie varie très sensiblement d'une commune à l'autre.

Afin d'assurer le plus de comparabilité possible entre les estimations au niveau communal et au niveau des cellules, toutes les variables susceptibles de comporter un effet-taille sont exprimées en termes de densité (quantités divisées par la surface de la commune), et une variable mesurant la taille de la commune est ajoutée à la liste des variables explicatives.

Nous décrivons ci-dessous les variables définies au niveau des ménages, qui serviront à calculer les effets croisés.

7.3.3.2. Variables caractéristiques des ménages

Comme dans tout modèle multinomial logit, les variables spécifiques au ménage ne peuvent pas être utilisées telles quelles pour expliquer les choix de localisation, puisqu'elles prennent les mêmes valeurs dans toutes les alternatives. En revanche, elles sont utilisées pour construire des effets croisés avec les caractéristiques des communes, ce qui reflète, par exemple, le fait que les ménages d'une personne et ceux de 5 personnes n'ont pas la même sensibilité à la densité résidentielle dans la commune.

La base de données étant tirée du Recensement, les variables ne portent pas les mêmes noms que dans l'EGT. Il y a plus de variables disponibles dans le RGP que dans l'EGT, les données sont exhaustives mais rappelons qu'il a deux défauts par rapport à l'EGT (comparaison déjà effectuée dans le rapport n°1, page 108) :

- il n'y a pas d'information sur le revenu du ménage. L'année dernière, dans la première phase de cette recherche, le revenu avait été reconstitué de manière statistique grâce à la base de données de la Direction générale des Impôts. Nous avons affiné cette méthode en utilisant une autre source de données : la base de données Budget des familles 2000, qui renseigne, pour plus de 10 000 ménages, sur les caractéristiques socio-économiques similaires à celles présentes dans le Recensement ainsi que sur le revenu disponible, transferts (allocations familiales, RMI, etc...) compris. L'imputation des revenus aux ménages du RGP est décrite en annexe 3.
- seul le département de la résidence précédente est connu, et non la commune.

Nom	Description
LIncome	Log du revenu du ménage (imputé, voir section sur le revenu)
Poor	Indicatrice = 1 si Income<20952, 0 sinon
MedInc	Indicatrice = 1 si 20952<Income<54176, 0 sinon
Rich	Indicatrice = 1 si Income>54176, 0 sinon
Foreign	Indicatrice = 1 si chef de ménage de nationalité étrangère, 0 sinon
AgeM	Age du chef de ménage, en années
SexeM	Sexe du chef de ménage (Indicatrice = 1 si femme, 0 sinon)
Young	Indicatrice = 1 si AgeM<36, 0 sinon
Medium	Indicatrice = 1 si 35<AgeM<61, 0 sinon
Old	Indicatrice = 1 si AgeM>60, 0 sinon
NPER	Nombre de membres du ménage, tronqué à 10
NPER1	Indicatrice = 1 si NPER=1, 0 sinon
NPER2	Indicatrice = 1 si NPER=2, 0 sinon
NPER3	Indicatrice = 1 si NPER>2, 0 sinon
NP3M	Nombre d'enfants de 3 ans ou moins dans le ménage
Child3	Indicatrice=1 si NP3M>0, 0 sinon
NP18M	Nombre d'enfants de 18 ans ou moins dans le ménage
NbCar	Nombre de voitures dans le ménage
NoCar	Indicatrice = 1 si NbCar=0, 0 sinon
Car1	Indicatrice = 1 si NbCar=1, 0 sinon
Car2	Indicatrice = 1 si NbCar>1, 0 sinon
NPAM	Nombre d'actifs dans le ménage, tronqué à 5
NPAM0	Indicatrice = 1 si NPAM=0, 0 sinon
NPAM1	Indicatrice = 1 si NPAM=1, 0 sinon
NPAM2	Indicatrice = 1 si NPAM>1, 0 sinon
FewerCars	Indicatrice = 1 si NPAM>NbCar, 0 sinon

Tableau 24 - variables définies au niveau des ménages

7.3.3.3. Variables caractéristiques des communes

Les variables communales susceptibles d'être utilisées sont décrites dans le tableau suivant :

Nom	Description
AreaKm2	Superficie en km ²
YrBuilt	Moyenne des années de construction dans la commune
DistArt	Distance moyenne à la rue centrale
DistHwy	Distance moyenne à l'autoroute
Subway	Nombre de stations de métro en 1999
Rail9	Nombre de gares en 1999
MMAN9	Mont Moy MA neuve 1999

MMMAA9	Mont Moy MA anc 1999
MMAPA9	Mont Moy AP anc 1999
MMAPN9	Mont Moy AP neuf 1999
Price	Moyenne pondérée des 4 prix précédents (voir section sur les prix)
LPriceHouse	Moyenne pondérée de MMMAA9 et MMMAN9
LPriceFlat	Moyenne pondérée de MMAPA9 et MMAPN9
EmpAg9	Emplois Agriculture 1999
EmpInd9	Emplois Industrie 1999
EmpCon9	Emplois Construction 1999
EmpTer9	Emplois Tertiaire 1999
EmpTot9	Emploi total 1999
Units9	Logements RP99
Density	Densité en logements = Units9 / AreaKm2
UnitsPrin9	Résidences principales RP99
UnitsSec9	Résidences secondaires RP99
UnitsVac9	Logements vacants RP99
UnitsInd9	Rés. princ. type maison individuelle ou ferme RP99
UnitsCol9	Rés. princ. dans un immeuble collectif RP99
UnitsOth9	Rés. princ. autres types RP99
Pop9	Population des ménages RP99
Hholds9	Total des ménages RP99
HH1Y9	Ménages 1 personne RP99
HHS2Y9	Ménages 2 personnes RP99
HH3Y9	Ménages 3 personnes RP99
HH4Y9	Ménages 4 personnes RP99
HH5Y9	Ménages 5 personnes RP99
HH6Y9	Ménages 6 personnes ou plus RP99
Car0Y9	Ménages sans voiture RP99
Car1Y9	Ménages 1 voiture RP99
Car2Y9	Ménages 2 voitures ou plus RP99
LndRur9	Surface RURAL Km ² 1999
LndVac9	Surface URBAIN OUVERT Km ² 1999
LndInd9	Surface HABITAT INDIVIDUEL Km ² 1999
LndCol9	Surface HABITAT COLLECTIF Km ² 1999
LndEmp9	Surface ACTIVITES Km ² 1999
TC	Accessibilité par transports publics (Log-Somme)
VP	Accessibilité par véhicule privé (Log-Somme)
Poor_M	% de ménages ayant Poor=1 dans la commune
MedInc_M	% de ménages ayant MedInc=1 dans la commune
Rich_M	% de ménages ayant Rich=1 dans la commune
Foreign_M	% de ménages ayant Foreign=1 dans la commune

Tableau 25 - variables définies au niveau communal

Attention : Les accessibilités nommées TC et VP ici ne sont pas définies de la même manière que dans la partie 5.3.3. Ici, elle sont calculées comme le temps moyen de déplacement à partir de la zone du découpage en 606 zones du modèle vers les autres zones, pondéré par le nombre de déplacements entre ces 2 zones (issu de la matrice origine-destination). Donc plus TC et VP sont grands, moins la zone est « accessible » (c'est l'inverse dans la formule de Jean Poulit, partie 5.3.3).

7.3.3.4. Estimation du modèle avec des variables croisées

Les variables testées sont issues de différents croisements des variables avec les caractéristiques des ménages. En particulier,

- Chaque indicatrice de type de ménage (selon la nationalité, le revenu et l'âge du chef de ménage) a été croisée avec le pourcentage de ménages de même type dans la commune concernée, afin d'évaluer le degré de mixité sociale ou au contraire de ségrégation sociale en Ile-de-France
- Les nombres de maisons et de logements collectifs dans la commune ont été croisés avec la taille du ménage afin de tester l'idée selon laquelle les familles nombreuses préfèrent plutôt vivre dans une maison individuelle entourée de maisons, alors que les familles réduites préfèrent les appartements ou accordent moins d'importance à être entourées de maisons
- Les prix communaux ont été croisés avec l'âge du chef de ménage et le revenu du ménage afin de permettre à l'élasticité-prix de varier d'un ménage à l'autre (en fonction du revenu ou de l'âge)
- Les variables d'accessibilité par mode ont été croisées avec le nombre de voitures dans le ménage, le nombre de personnes et le nombre d'actifs afin de traduire l'hétérogénéité des besoins pour chaque mode de transport
- Les variables d'accessibilité par mode ont été croisées avec le sexe du chef de ménage, puisque les femmes ont plus tendance que les hommes à emprunter les transports en commun et que les estimations montrent généralement que la valeur du temps diffère entre hommes et femmes

7.3.4. Résultats du calage

Les résultats de l'estimation du modèle de choix résidentiel sont reproduits dans le tableau 25. Le pouvoir explicatif de ce modèle est moyennement satisfaisant, puisque le pseudo-R² est de l'ordre de 23%. Il est à noter que l'estimation d'un tel modèle sur les données exhaustives du Recensement (totalité des ménages ayant déménagé en 1998) nécessite plus de 3 heures sur un ordinateur portable très puissant, alors que le même modèle est estimé en 3 minutes sur un échantillon à 20%. La raison principale à ce changement d'échelle dans les temps de calcul est le dépassement de la mémoire vive (2 Giga), qui augmente les temps de calcul en obligeant SAS à écrire sur le disque dur lors de son processus d'estimation.

Paramètre	Coefficient	Ecart-type	t	Pr> t
Même département	2.5439	0.004249	598.69	<.0001
Paris	-0.4562	0.0124	-36.89	<.0001
LPImput	-2.7686	0.0457	-60.52	<.0001
LPImput * (Age-20)/10	-0.0693	0.002107	-32.91	<.0001
LPImput * LIncome	0.2792	0.004545	61.43	<.0001
Rail9	-0.0122	0.001282	-9.53	<.0001
Subway	0.008806	0.000588	14.99	<.0001
TC	0.000704	0.000218	3.22	0.0013
TC * (Sexe=2)	-0.004648	0.000315	-14.74	<.0001

VP	0.0532	0.000472	112.74	<.0001
VP * NbCars	-0.0563	0.000340	-165.68	<.0001
DISTHWY	-2.033E-6	2.7631E-7	-7.36	<.0001
NPER1com	1.9562	0.0392	49.85	<.0001
NPER2com	0.9192	0.1390	6.61	<.0001
NPER3com	2.9836	0.0373	79.95	<.0001
NPAM0com	5.9145	0.1029	57.48	<.0001
NPAM1com	0.2705	0.0663	4.08	0.0044
NPAM2com	0.1529	0.0494	3.10	0.0020
young_m	0.4513	0.0600	7.53	<.0001
young_Com	4.5760	0.0611	74.95	<.0001
poor_m	-0.0316	0.0770	-0.41	0.6817
foreign_Com	6.0039	0.0743	80.78	<.0001
foreign_NoCom	-2.6130	0.0449	-58.17	<.0001
EMPTOT9	-7.702E-7	1.0753E-7	-7.16	<.0001
Density	-4.154E-6	4.7175E-7	-8.81	<.0001
LpopRP99	0.1076	0.002427	44.35	<.0001
delta_pop	0.0671	0.007606	8.83	<.0001

Tableau 26 - estimation de la demande de logement sur données exhaustives

Notons tout d'abord que le département de résidence précédent joue un rôle primordial pour expliquer les choix de localisation. Le fort coefficient de la variable « Même département » montre que les ménages ont une très forte tendance à rester dans le même département lorsqu'ils déménagent. Le fort coefficient négatif de la variable « Paris » montre que, **toutes choses égales par ailleurs**, les ménages préfèrent se localiser en banlieue plutôt qu'à Paris. Ce résultat est à relativiser en tenant compte du fait que l'accessibilité (au moins par les transports en commun) est bien meilleure à Paris, et que d'autres facteurs pris en compte dans la régression attirent les ménages vers Paris.

On note un effet négatif très prononcé du prix sur l'utilité retirée de la résidence dans une commune. Cet effet négatif est d'autant plus important que l'âge augmente, ce qui signifie que les personnes plus âgées sont plus sensibles aux prix du logement. En revanche, l'ampleur de cet effet négatif diminue avec les revenus. L'effet du prix sur la préférence pour une commune devient même positif lorsque le chef de ménage est très jeune et son revenu très élevé, ce qui signifie que, toutes choses égales par ailleurs, les ménages suffisamment riches et jeunes préfèrent habiter dans des communes où les prix du logement sont très élevés. Ce résultat, qui peut paraître contre-intuitif, suggère que les caractéristiques positives des communes, qui accroissent à la fois les prix dans une commune et la demande de logement à prix fixé ne sont pas suffisamment prises en compte dans ce modèle. Cette explication est renforcée par le fait que, lorsque l'on estime la demande de logement avec très peu de caractéristiques communales, l'effet-prix est positif pour tous les ménages.

Une augmentation des temps de déplacement moyens par transports en commun diminue sensiblement la demande de logement pour une commune pour les ménages dirigés par une femme (sexe=2), mais cet effet semble négligeable, et joue même en sens inverse pour les hommes. Le coefficient correspondant est toutefois non significatif dans les sous-échantillons à 20%, 5% et 1%.

Toutes choses égales par ailleurs (et en particulier à temps de déplacement fixé), les ménages préfèrent se localiser près des stations de métro (effet positif du nombre de stations de métro dans la commune), mais loin des gares ferroviaires (effet négatif du nombre de gares).

Les variations de temps de déplacement par véhicule particulier sont sans effet sur la localisation des ménages qui ont une seule voiture (la somme des coefficients de VP et de $VP \cdot NbCars$ n'est pas significativement différente de 0). Toutes choses égales par ailleurs, l'augmentation du temps de déplacement diminue la demande de logement des ménages ayant deux voitures ou plus, mais elle augmente la demande des ménages sans voitures. L'effet contre-intuitif des temps de déplacement par véhicule particulier pour les ménages ayant moins de 2 voitures est en partie contrebalancé par l'effet négatif de la distance à l'autoroute la plus proche.

Les variables de composition de la commune (% de ménages dans les différentes catégories) traduisent bien la tendance des ménages à se localiser à proximité de ménages semblables. Ainsi, les ménages ne comportant qu'un seul membre ont d'autant plus tendance à se localiser dans une commune que la fraction de ménages à un seul membre y est élevée. Il en va de même pour les ménages à deux membres et, surtout, pour les ménages à 3 membres et plus : les familles recherchent la proximité d'autres familles. On trouve un effet similaire, mais beaucoup plus marqué en ce qui concerne les familles sans aucun membre actif, ce qui peut refléter une concentration des chômeurs ou, plus vraisemblablement, des retraités. Cet effet de concentration reste positif, mais beaucoup moins marqué pour les ménages monoactifs, et encore moins marqué pour les ménages à plusieurs actifs. Cette dernière variable a même un effet négatif (peu ou pas significatif) dans les sous-échantillons. Il ne faut toutefois pas oublier que l'effet de concentration en termes de nombre d'actifs s'ajoute à l'effet de concentration en terme de nombre de membres du ménage, et que ces deux variables sont positivement corrélées.

Les communes abritant de nombreux ménages jeunes attirent l'ensemble des ménages, et cette attirance est beaucoup plus marquée pour les ménages qui sont eux-mêmes jeunes. En revanche, toutes choses égales par ailleurs, les ménages « vieux » ou d'âge moyen ne sont pas attirés par les autres ménages similaires. Les variables mesurant le pourcentage de ménages « vieux » ou d'âge moyen, et leurs effets croisés ont donc été exclues des régressions.

Etant donnée la liste des autres variables déjà prises en compte, la composition de la commune en terme de revenu des ménages semble sans effet sur la localisation des ménages riches, pauvres et de revenu moyen, mais un effet de concentration des ménages riches et pauvres peut être mis en évidence lorsque l'on ne contrôle pas, par exemple, la densité de population, la taille ou les variations de population.

Les ménages dirigés par un étranger ont une forte tendance à se localiser dans des communes comportant une très forte proportion d'étrangers, alors que les ménages dirigés par un français ont tendance à préférer les communes où la proportion d'étrangers est moindre.

Toutes choses égales par ailleurs, les ménages ont tendance à se localiser dans des « grosses » communes (en termes de nombre d'habitants), surtout si la population de ces communes croît. En revanche, les ménages préfèrent se localiser dans des communes moins denses (donc plus étendues, à nombre d'habitants fixé) et abritant moins d'emplois.

Lorsque le résidu de l'équation de prix est ajouté à la liste des variables explicatives, les paramètres estimés sont sensiblement les mêmes et le coefficient du résidu de l'équation de prix n'est pas du tout significatif (seuil critique de 28%), ce qui signifie que, avec les variables utilisées pour prédire les prix et les choix de localisation, il n'y a pas de problèmes d'endogénéité des prix. En d'autres termes, toute la corrélation entre les prix communaux du

logement et les choix de localisation passe par les variables explicatives utilisées dans ces deux régressions.

7.3.5. Validité du modèle

Pour évaluer le pouvoir prédictif du modèle, il est possible de comparer les demandes prédites pour chaque commune (somme des probabilités estimées pour chaque ménage) aux demandes observées (nombre de ménages ayant effectivement emménagé dans la commune). Il en résulte une très bonne qualité prédictive agrégée du modèle, puisque la corrélation entre la demande observée et la demande estimée est de 99,78%.

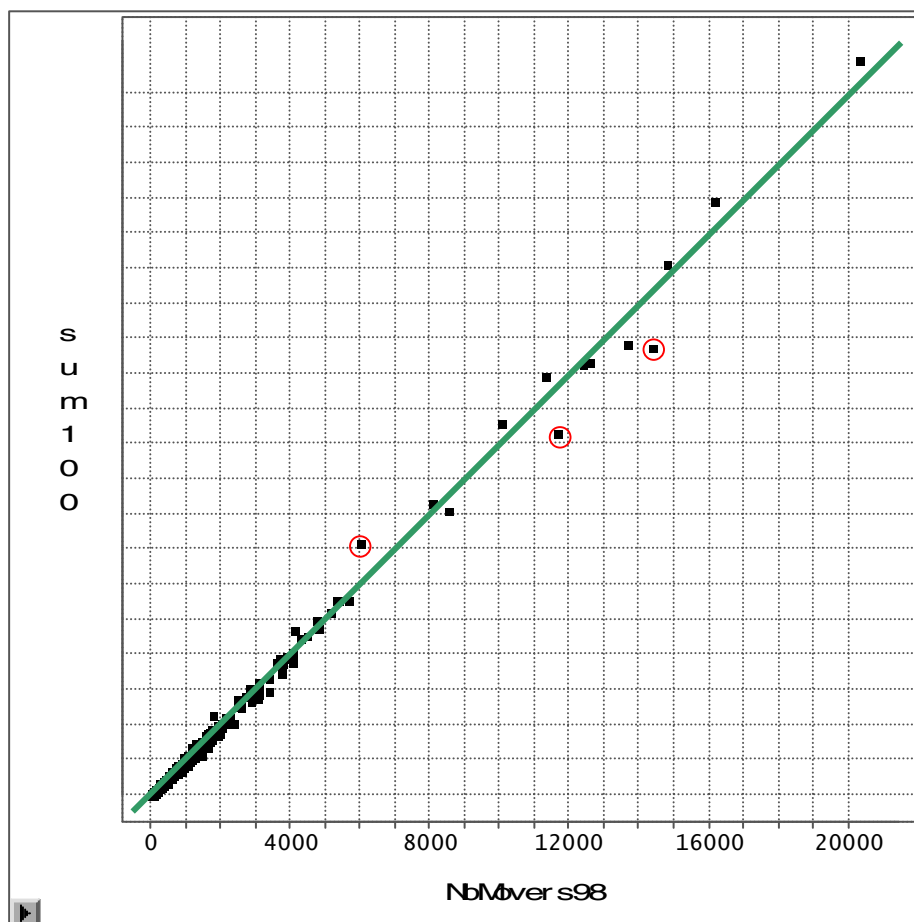


Figure 26 - Comparaison de la demande observée (NbMovers98) et prédite (sum100)

Ces bons résultats sont confirmés par la Figure 26, qui montre que les prédictions sont très proches des effectifs observés (diagonale) pour la plupart des communes. Les trois exceptions notables y sont entourées et correspondent aux arrondissements 17 et 19 (sous-estimation de la demande), 5 (surestimation de la demande).

Il semble à première vue que des résultats d'aussi bonne qualité soient obtenus sur des échantillons réduits, comme le montre la Figure 27 sur des échantillons à 1% (sum_1), 5% (sum5), 20% (sum20) et 100% (sum100). La corrélation entre la demande observée et la demande prédite dépasse en effet 99% sur chacun des sous-échantillons.

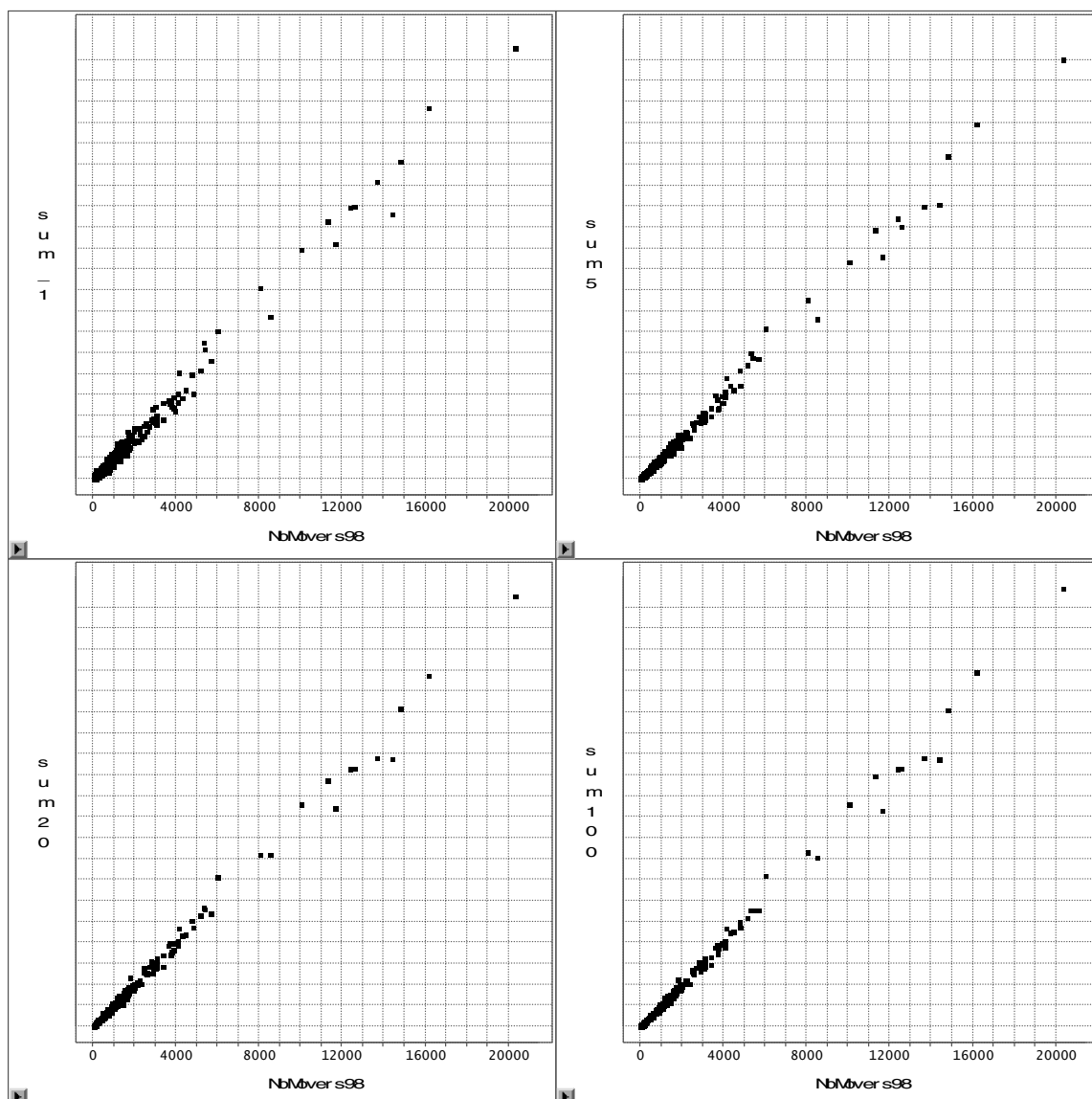


Figure 27 - Comparaison de la demande observée (NbMovers98) et prédite sur différents sous-échantillons

Mais la Figure 28 montre que cette première impression est quelque peu trompeuse. En effet, un zoom sur les communes pour lesquelles moins de 5000 habitants ont déménagé en 1998 (ce qui n'exclut que 17 communes dont 13 à Paris) montre clairement que la précision des estimations se dégrade sensiblement lorsque le taux d'échantillonnage descend à 5%, et encore plus à 1%. Cette dégradation de la qualité des estimations est particulièrement sensible dans les petites communes. Ainsi, dans les 1148 communes comptant moins de 1000 déménagements en 1998, la corrélation entre demande observée et demande prédite est de 99,4% sur l'échantillon total, 99,2% sur l'échantillon à 20%, 98,8% sur l'échantillon à 5% et 95,4% sur l'échantillon à 1%. Cette dernière descend à moins de 90% pour les 1054 communes comptant moins de 500 déménagements en 1998 et à 76% pour les 948 communes comptant moins de 200 déménagements en 1998 (qui représentent encore plus de 70% des communes d'Ile de France). Pour ces mêmes communes, la corrélation entre demande observée et demande prédite reste de 98% dans l'échantillon exhaustif et 97% dans l'échantillon à 20%. On voit ainsi une très forte dégradation de la précision des estimations en

dehors des plus grosses communes lorsque le taux d'échantillonnage descend en-dessous de 20%.

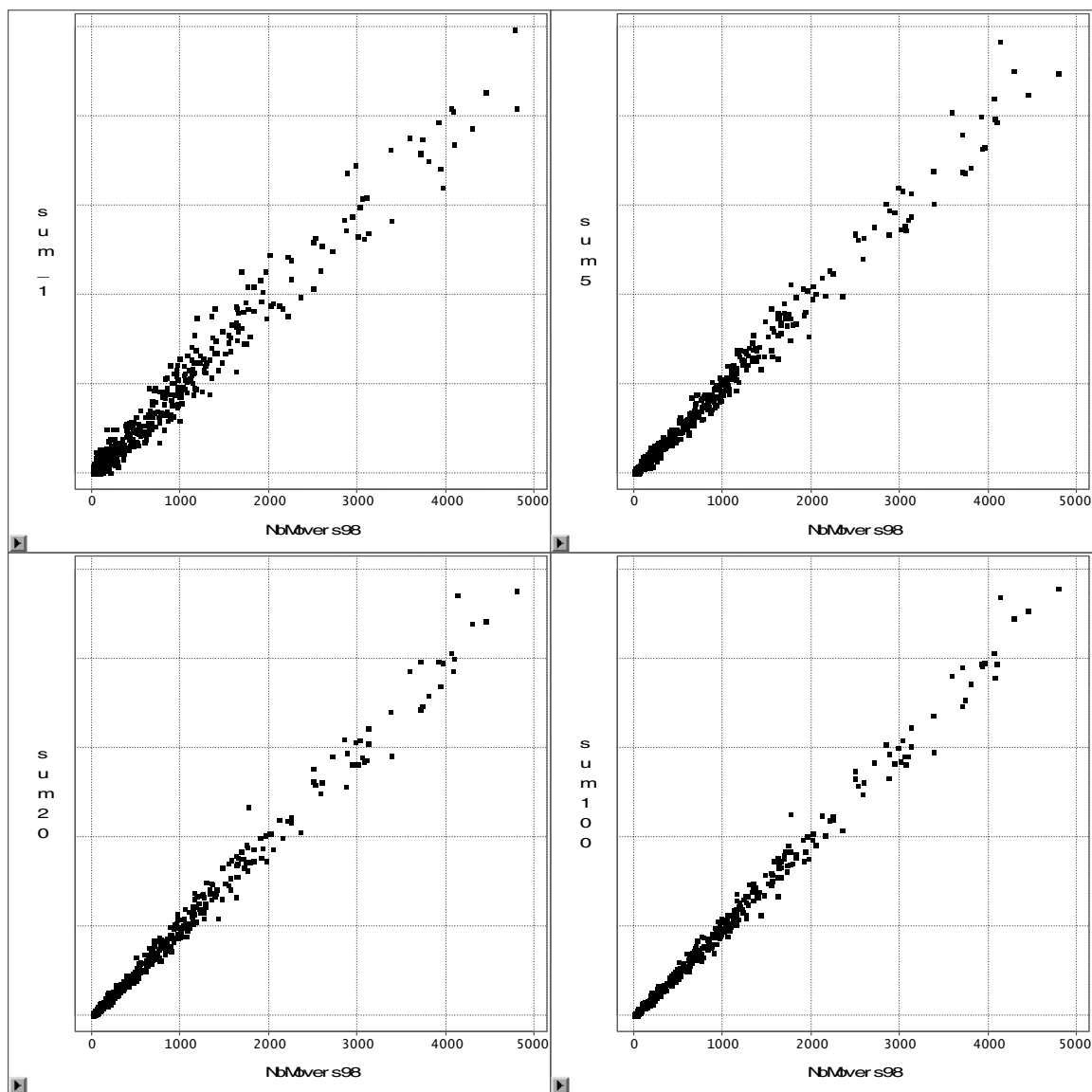


Figure 28 - Comparaison de la demande observée (NbMovers98) et prédite sur différents sous-échantillons, en excluant les « très grosses » communes

7.4. Le module de prix du foncier au niveau communal

Nous proposons l'équation semi-hédonique suivante :

$$P_{ilt} = \alpha + \lambda_1 S_{ilt} + \lambda_2 D_{ilt} + \beta X_{ilt}.$$

où P_{ilt} est le prix unitaire du terrain ou de logement de type i , dans la zone l , l'année t et X_{ilt} un vecteur d'attributs de la commune et S et D représentent l'offre et la demande. Les prix sont actualisés annuellement, en fin de période d'activité et utilisés ensuite comme références pour l'année suivante.

L'IAURIF dispose de la base de données agrégées des Notaires. Elle comprend l'ensemble des transactions immobilières enregistrées par les notaires sur la période 1990-2003 mais ces transactions ne sont pas localisées à l'adresse. Seule la donnée de commune est fournie. Les autres données concernent les volumes et prix moyens par année, pour quatre types de logements :

- Maisons neuves
- Maisons anciennes
- Appartements neufs
- Appartements anciens

ANNEE	Nombre de communes comportant des informations
1990	143
1991	143
1992	143
1993	143
1994	143
1995	143
1996	1 186
1997	1 213
1998	1 240
1999	1 263
2000	1 260
2001	1 244
2002	1 250
2003	1 111
Total	10 625

Tableau 27 - Nombre de communes comportant des informations, par année

En théorie, les données devraient porter sur 1300 observations (communes) par an, mais de nombreuses communes ne comportent, certaines années, aucune information dans la base. Les données manquantes sont surtout concentrées en début de période (avant 1996), puisque, en début de période, les données n'ont été fournies que pour 143 communes chaque année (cf Tableau 27).

Donc cette base de données comportait de nombreux « trous ». Nous avons élaboré une méthode statistique pour compléter cette base, en faisant appel à des techniques de données de

panel. Après avoir donné quelques statistiques descriptives sur la base initiale, nous présentons la méthode statistique.

7.4.1. Statistiques descriptives sur la base initiale

Une étude plus poussée montre que ces 143 communes comportant des données en début de période correspondent à l'ensemble des communes situées à Paris et en petite couronne.

Pour les années ultérieures (à partir de 1996), aucune information n'a été fournie pour environ 100 à 200 communes par an, toutes situées en grande couronne, mais ces communes varient d'une année à l'autre. Il s'agit typiquement de petites communes dans lesquelles aucune transaction immobilière n'a été effectuée au cours de l'année considérée.

Ainsi, 841 communes (soit 65% des 1300 communes) ne comportent aucune information sur les 6 premières années (1992 à 1995), mais comportent des informations sur toutes les années suivantes (1996 à 2003) ; 143 communes, soit 11% comportent des informations chaque année. Ces deux profils représentent plus de $\frac{3}{4}$ des communes ; 80 communes, soit 11% ne comportent aucune information sur les 6 premières années, ni sur la dernière, mais comportent des informations sur toutes les années intermédiaires (1996 à 2002). Les autres profils sont beaucoup moins courants.

Même lorsqu'il existe des informations pour une commune et une année données, l'ensemble des informations n'est pas toujours présent dans la base, comme en atteste le Tableau 4. Ainsi, sur les 10 625 communes-années comportant des informations sur au moins un type de transactions (appartement ou maison, neuf ou ancien), 10 470 comportent des informations sur les ventes de maisons anciennes, ce qui signifie que, lorsque des logements sont vendus dans une commune au cours d'une année, au moins l'un de ces logements est une maison ancienne dans 98,5% des cas. Ce pourcentage tombe à 53% pour les appartements anciens et à 31% pour les appartements neufs, ainsi que pour les maisons neuves.

Ces pourcentages varient assez peu dans le temps, comme en atteste le Tableau 28, mais ils varient beaucoup en fonction du département considéré, comme en atteste le Tableau 29.

Année	Au moins un type de transaction	Appartements neufs	Appartements anciens	Maisons neuves	Maisons anciennes
1990	143	123	139	59	114
1991	143	118	139	74	133
1992	143	116	142	76	131
1993	143	120	142	78	139
1994	143	126	141	87	135
1995	143	127	141	78	139
1996	1 186	332	537	342	1 172
1997	1 213	331	553	365	1 200
1998	1 240	345	583	353	1 231

1999	1 263	336	636	384	1 254
2000	1 260	330	648	405	1 254
2001	1 244	317	659	406	1 234
2002	1 250	318	656	375	1 243
2003	1 111	258	555	255	1 091
Total	10 625	3 297	5 671	3 337	10 470

Tableau 28 - Nombre de communes comportant des informations sur chaque type de transaction, par année

dept	Au moins un type de transaction		Appartements neufs		Appartements anciens		Maisons neuves		Maisons anciennes	
	Nbre	%	Nbre	%	Nbre	%	Nbre	%	Nbre	%
75	280	2.64	261	7.92	280	4.94	57	1.71	191	1.82
77	3867	36.40	456	13.83	1130	19.93	829	24.84	3856	36.83
78	1958	18.43	380	11.53	941	16.59	529	15.85	1945	18.58
91	1473	13.86	424	12.86	915	16.13	551	16.51	1468	14.02
92	504	4.74	458	13.89	503	8.87	271	8.12	497	4.75
93	560	5.27	468	14.19	559	9.86	351	10.52	551	5.26
94	658	6.19	528	16.01	641	11.30	383	11.48	649	6.20
95	1325	12.47	322	9.77	702	12.38	366	10.97	1313	12.54
Total	10625	100	3297	100	5671	100	3337	100	10470	100

Tableau 29 - nombre de communes comportant des informations sur chaque type de transaction, par département

Ainsi, par exemple, Paris intra muros compte 2,64% des communes-années pour lesquelles au moins une transaction a été enregistrée, mais ce pourcentage tombe à 1,7% et 1,8% respectivement en ce qui concerne les maisons, neuves et anciennes respectivement. A l'inverse, ce pourcentage monte à 7,9% pour les appartements neufs et 4,9% pour les appartements anciens. Cette sur-représentation des transactions sur appartements neufs (et dans une moindre mesure anciens) est aussi considérable dans les départements de petite couronne. A l'inverse, les transactions sur appartements, surtout neufs, sont sous représentées dans les départements de la grande couronne, surtout le 77 et le 78.

7.4.2. Méthode d'imputation des prix manquants

Afin de réduire l'impact de prix exceptionnellement bas ou, surtout, exceptionnellement élevés sur le calcul des moyennes, nous avons calculé les moyennes des logarithmes des prix observés par type de transaction, par commune et/ou par année.¹⁷ Cette méthode est d'autre part cohérente avec l'hypothèse que, d'une année à l'autre, d'une commune à l'autre et d'un type de logement à l'autre, les différences de prix sont approximativement proportionnelles. Par exemple, lorsque, sur l'ensemble de la période, le prix moyen des appartements anciens

¹⁷ Nous avons ensuite calculé l'exponentielle de la moyenne ainsi obtenue, afin de rendre le résultat comparable aux prix initialement collectés. Sauf indication contraire, les prix « moyens » correspondent donc à l'exponentielle de la moyenne des logarithmes de prix pour les observations considérées.

est de 181453€ à Paris contre 88151€ pour l'ensemble de l'Ile de France, il est plus réaliste de dire que les prix sont deux fois plus élevés à Paris que de dire qu'ils sont supérieurs de 10 000 € à Paris.

Les moyennes ont été systématiquement pondérées par le nombre de transactions correspondant au type de logement, à la commune et à l'année considérés.

Il en ressort, sur la période considérée, un prix moyen de 127 243€, soit 153 800€ pour les appartements neufs, 113 443€ pour les appartements anciens, 181 849€ pour les maisons neuves et 156 610€ pour les maisons anciennes.

Les prix moyens varient aussi beaucoup d'un département à l'autre : de 85 970€ en Seine-Saint-Denis à 150 115€ à Paris (voir Tableau 30). Enfin, l'évolution temporelle des prix diffère sensiblement en fonction du département.

Département	Prix moyen
75	150 115.3
77	108 977.7
78	147 788.5
91	112 904.1
92	148 542.5
93	85 970.39
94	114 040.5
95	108 006.9

Tableau 30 - prix moyens observés par département

Au vu des résultats de la section précédente, nous avons estimé, sur les données disponibles dans la base de données des Notaires, par moindres carrés pondérés, un modèle de régression linéaire du log du prix moyen de vente dans la commune considérée, pour l'année et le type de logement considéré.

Les variables explicatives sont :

- Une série de variables indicatrices (dummies) pour les 13 premières années (l'année 2003 sert de référence).
- Une série de dummies pour le département considéré (Paris sert de référence)
- Une série de dummies pour le type de logement considéré (les maisons anciennes servent de référence).
- Une variable de tendance linéaire par département, qui reflète le fait que, toutes choses égales par ailleurs, d'un département à l'autre, les prix ont tendance à croître plus ou moins vite sur la période considérée.
- Une variable de tendance linéaire par type de logement, qui reflète le fait que, toutes choses égales par ailleurs, sur la période considérée, les prix ont tendance à croître à une vitesse différente pour les appartements et pour les maisons, pour le neuf et pour l'ancien.
- Des variables croisées type de logement * département, qui reflètent le fait que, toutes choses égales par ailleurs, le rapport de prix entre appartements et maisons, ou entre neuf et ancien, varie d'un département à l'autre.

Le modèle est le suivant :

$$\log(P_{ijt}) = \beta_0 + \beta_{t1} \text{Année}_t + \beta_{d2} \text{Dept}_{id} + \beta_{j3} \text{Type}_j + \beta_{d4} \bullet t \bullet \text{Dept}_{id} + \beta_{j4} \bullet t \bullet \text{Type}_j + \beta_{jd5} \bullet t \bullet \text{Type}_j \bullet \text{Dept}_{id} + u_i + \varepsilon_{ijt}, \text{ où :} \quad (1)$$

- i représente la commune considérée
- j représente le type de logement considéré
- t représente l'année considérée ($t=0$ en 1990, $t=14$ en 2003)
- $\text{Dept}_{id}=1$ pour le département d où est située la commune i , 0 pour les autres
- u_i est une variable aléatoire d'espérance nulle et de variance σ_u^2 indépendante et identiquement distribuée (iid) d'une commune à l'autre
- ε_{ijt} est une variable aléatoire d'espérance nulle et de variance σ_ε^2 indépendante et identiquement distribuée (iid) d'une commune à l'autre, d'un type à l'autre et d'une année à l'autre, et indépendante de u_i

Log(Prix)	Coef.	Std.Err.	t	P>t
Constante	13.60567	.0070925	1918.32	0.000
Année				
1990	-.5340115	.0037128	-143.83	0.000
1991	-.4737941	.003514	-134.83	0.000
1992	-.4739327	.0032724	-144.83	0.000
1993	-.4696576	.0030436	-154.31	0.000
1994	-.4384186	.0028034	-156.39	0.000
1995	-.4878739	.0026677	-182.88	0.000
1996	-.4812077	.0023826	-201.97	0.000
1997	-.4828615	.0021802	-221.47	0.000
1998	-.4494522	.0019584	-229.50	0.000
1999	-.3773757	.0017385	-217.06	0.000
2000	-.2651915	.0015862	-167.18	0.000
2001	-.1862131	.0014594	-127.59	0.000
2002	-.0850129	.0013634	-62.35	0.000
Département				
77	-1.453042	.0085043	-170.86	0.000
78	-.9493633	.0084016	-113.00	0.000
91	-1.104245	.0085967	-128.45	0.000
92	-.6906736	.007245	-95.33	0.000
93	-1.419969	.0072199	-196.67	0.000
94	-1.131757	.0072237	-156.67	0.000
95	-1.232945	.0086471	-142.58	0.000
Type de logement				
Appart. neuf	-.6889251	.0075305	-91.49	0.000
Appart. ancien	-1.209798	.007102	-170.35	0.000
Maison neuve	.0011448	.0286979	0.04	0.968
Tendance par département				
Tendance, 77	-.0123964	.000492	-25.19	0.000
Tendance, 78	-.0150086	.0004639	-32.35	0.000
Tendance, 91	-.0250495	.0004986	-50.24	0.000
Tendance, 92	-.008621	.0002014	-42.81	0.000
Tendance, 93	-.0106961	.0002343	-45.65	0.000

Tendance, 94	-.0065662	.0002232	-29.41	0.000
Tendance, 95	-.0208042	.0005068	-41.05	0.000
Tendance par type				
Tendance, Appart. neuf	-.0128004	.0003101	-41.27	0.000
Tendance, Appart. ancien	-.0176807	.0002393	-73.88	0.000
Tendance, Maison neuve	-.0092657	.0006553	-14.14	0.000
Type * département				
Appart. neuf, 77	.5934904	.0078629	75.48	0.000
Appart. ancien, 77	.8326015	.0072256	115.23	0.000
Maison neuve, 77	.3840231	.0284074	13.52	0.000
Appart. neuf, 78	.4153248	.0078695	52.78	0.000
Appart. ancien, 78	.8362794	.0071096	117.63	0.000
Maison neuve, 78	.1616123	.0286556	5.64	0.000
Appart. neuf, 91	.4918475	.0080727	60.93	0.000
Appart. ancien, 91	.6670969	.007164	93.12	0.000
Maison neuve, 91	.230763	.0285019	8.10	0.000
Appart. neuf, 92	.4730439	.007479	63.25	0.000
Appart. ancien, 92	.599883	.0071058	84.42	0.000
Maison neuve, 92	.270307	.02894	9.34	0.000
Appart. neuf, 93	.6207704	.0076079	81.60	0.000
Appart. ancien, 93	.6928394	.0070475	98.31	0.000
Maison neuve, 93	.2194214	.0287463	7.63	0.000
Appart. neuf, 94	.6090374	.0075705	80.45	0.000
Appart. ancien, 94	.7219339	.0070581	102.28	0.000
Maison neuve, 94	.2462915	.0287035	8.58	0.000
Appart. neuf, 95	.4644343	.0080519	57.68	0.000
Appart. ancien, 95	.7245554	.0071757	100.97	0.000
Maison neuve, 95	.2475775	.0285606	8.67	0.000

Tableau 31 - résultats de l'estimation des prix pour imputation

Le pouvoir explicatif de ce modèle est satisfaisant, puisque $R^2=0,5269$, ce qui signifie que près de 53% des variations de prix observées (en log) sont expliquées par le modèle. L'écart-type des prix observés est de 0,4643, ce qui correspond à une variation de prix typique de l'ordre de $\exp(0,4643)-1=0,591$, soit 59%. En comparaison, l'écart-type des résidus ($u_i + \varepsilon_{it}$) est de 0,3194, ce qui correspond à une erreur typique sur les prix de l'ordre de $\exp(0,3194)-1=0,3763$, soit 38%. L'écart-type de u_i (variabilité inter-communale non expliquée) est légèrement supérieur à celui de ε_{it} (variabilité intra-communale non expliquée) : 0,274 contre 0,256.

7.4.3. Base de calage

L'équation (1) est utilisée pour imputer les prix manquants dans la base initiale, selon la formule :

$$\log(\hat{P}_{ijt}) = \hat{\beta}_0 + \hat{\beta}_{t1} \text{Année}_t + \hat{\beta}_{d2} \text{Dept}_{id} + \hat{\beta}_{j3} \text{Type}_j + \hat{\beta}_{d4} \bullet t \bullet \text{Dept}_{id} + \hat{\beta}_{j4} \bullet t \bullet \text{Type}_j + \hat{\beta}_{jd5} \bullet t \bullet \text{Type}_j \bullet \text{Dept}_{id} + \hat{u}_i, \text{ où :} \quad (2)$$

- les $\hat{\beta}_0, \hat{\beta}_{11}, \hat{\beta}_{d2}, \hat{\beta}_{j3}, \hat{\beta}_{d4}, \hat{\beta}_{j4}$ et $\hat{\beta}_{jd5}$ correspondent aux estimations des coefficients correspondants
- \hat{u}_i correspond à la moyenne pondérée des résidus $\hat{u}_{ijt} = \log(P_{ijt}) - \log(\hat{P}_{ijt})$ estimés pour les observations non manquantes dans la commune i .

Le terme \hat{u}_i traduit le fait que, pour des raisons inexplicables qui persistent dans le temps et concernent les 4 types de logements, les prix peuvent être systématiquement plus élevés dans une commune que ne le prédit le modèle. Dans ce cas, pour une commune donnée, l'effet-commune \hat{u}_i estimé sur la base des années-type de logement comportant des informations est ajouté aux prix imputés pour les années-type de logement ne comportant pas d'informations. Pour chaque année et chaque commune considérées, un indice des prix du logement du logement est calculé comme moyenne pondérée des quatre prix imputés définis ci-dessus (un par type de logement), les poids correspondant au nombre de transactions pour l'année et la commune concernée (poids nul lorsque aucune transaction n'a été enregistrée l'année considérée pour ce type de logement dans cette commune).

Pour chaque année, chaque commune et chaque type de logement considérés, le prix imputé correspond au prix observé lorsqu'il existe, et au prix prédit par le modèle dans le cas contraire.

Après imputation, les moyennes de prix sur toute la période par département sont très proches des moyennes observées, comme cela est mis en évidence dans le tableau 8.

Département	Prix moyen observé	Prix moyen imputé
75	150115.3	150115.3
77	107256.3	108977.7
78	147271.9	147788.5
91	113192.3	112904.1
92	148542.5	148542.5
93	85970.39	85970.39
94	114040.5	114040.5
95	108265.4	108006.9

Tableau 32 - prix observés et imputés : moyenne par département

Toutefois, en grande couronne (surtout dans les départements 77, 78 et 91) en début de période, les prix imputés sont en moyenne beaucoup plus élevés que les prix observés. Cela est dû au fait que les observations manquantes concernent surtout les plus petites communes du département, dans lesquelles d'une part les transactions sont beaucoup moins fréquentes (d'où les données manquantes) et d'autre part, les prix sont plus faibles que dans les plus grosses communes du département. La méthode d'imputation utilisée permet de corriger cet effet et d'obtenir des prix plus réalistes.

Au final, la base de données des Notaires a été complétée. Elle comporte 72 800 observations (1300 communes * 14 années * 4 types de transaction).

7.4.4. Résultats du calage

Le modèle de prix est un modèle semi-hédonique. L'équation structurelle de prix est uniquement fonction de l'offre S_j et de la demande D_j :

$$P_j = S_j \lambda_2 + D_j \lambda_3 + v_j$$

avec

- D_j : demande pour chaque commune, déduite du modèle de localisation des ménages
- S_j : offre

Les résultats sont reproduits dans le tableau suivant :

Paramètre	Coefficient	Ecart-type	t	Pr > t
Intercept	11.02668	0.12800	86.14	<.0001
LogS	-0.04791	0.02466	-1.94	0.0522
LogD100	0.09918	0.02244	4.42	<.0001
tt_DT	-0.00280	0.00085119	-3.28	0.0011
NPER1_M	5.09136	0.37884	13.44	<.0001
NPER2_M	1.87960	0.34135	5.51	<.0001
NPAM0_M	1.25241	0.30954	4.05	<.0001
NPAM1_M	0.82300	0.33762	2.44	0.0149
poor_m	-6.63187	0.50316	-13.18	<.0001
medInc_m	-4.54311	0.33102	-13.72	<.0001
foreign_m	1.58406	0.36279	4.37	<.0001

Tableau 33 - résultats d'estimation du modèle du prix

Les coefficients de l'offre et de la demande ont le bon signe, mais ne sont plus exactement opposés. L'amélioration des temps de déplacement (tt_DT) augmente significativement les prix..

Les prix sont très sensibles à la composition socio-économique de la commune. Ainsi, une augmentation de 10% de la part des ménages ne comportant qu'un seul membre augmenterait les prix de 50% ; de même, une augmentation de 10% de la part des ménages ne comportant que deux membres augmenterait les prix de 19%. Les prix sont aussi croissants lorsque la part des ménages comportant peu de membres actifs augmente. Les prix augmentent lorsque le pourcentage d'étrangers dans la commune augmente, mais il ne faut pas oublier que la notion d'étrangers considérée ici ne fait pas la distinction entre ressortissants de pays de l'OCDE et ressortissants de pays pauvres.

Enfin, les prix sont très négativement influencés par la part des ménages à revenus faibles ou moyens.

Ainsi, le modèle apparaît explicatif. Il s'agit maintenant de l'adapter de manière à ce qu'il soit prédictif en exécution dans le nouvel UrbanSim.

7.5. *Modèle de localisation des emplois au niveau communal*

Dans les derniers jours de préparation de ce rapport, nous avons réussi à effectuer les premières estimations de modèle de localisation des emplois à la commune. Ici, nous les présentons comme un premier aperçu. Les détails seront donnés dans le rapport final.

L'estimation de ce modèle est effectuée sur la base des emplois salariés préparée par l'UNEDIC. Cette base contient le nombre des entreprises et des emplois classifiés par secteur d'activité et taille d'entreprise pour chaque année et chaque commune. Nous avons agrégé les secteurs d'activité en 15 groupes présentés dans le Tableau 34.

Groupe	Secteurs
1	industries agricoles et alimentaires
2	industries des biens de consommation
3	industrie automobile
4	industrie des biens d'équipement
5	industrie des biens intermédiaires
6	Energie
7	Construction
8	Commerce
9	Transports
10	activités financières
11	activités immobilières
12	services aux entreprises
13	services aux particuliers
14	éducation, santé, action sociale
15	Administration

Tableau 34 - groupement des secteurs d'activité

La procédure d'estimation est en cours d'incorporation dans la nouvelle version d'UrbanSim et les estimations préliminaires sont effectuées dans ce cadre.

Le tableau 34 présente un résumé des résultats de ces estimations préliminaires. D'après ces résultats, les emplois dans les secteurs 1, 8, 9, 11, et surtout 13, 14 et 15 ont, toutes choses égales par ailleurs, plus de chances de se localiser à Paris qu'en petite ou grande couronne, contrairement aux autres secteurs. Conformément aux attentes, les emplois attirés par Paris sont principalement dans le tertiaire, avec en premier lieu l'administration et les services, mais aussi le commerce, les transports et de l'immobilier. A l'inverse, les emplois industriels ont plutôt tendance à se localiser en banlieue, ce qui est aussi conforme à l'intuition.

Certains secteurs industriels (1, 4 et 5), pour lesquels la localisation des emplois n'est pas primordiale, sont particulièrement sensibles aux prix. D'autre part, les activités liées aux transports (9) ou à l'énergie (7) se localisent facilement sur des terrains moins chers, contrairement aux activités immobilières (11) qui se développent naturellement là où la demande de logement est la plus forte, et corrélativement les prix les plus élevés. Certains autres effets positifs des prix sur la localisation des emplois plus contre-intuitifs peuvent refléter l'effet de variables corrélées aux prix, qui ne sont pas encore prises en compte dans cette version préliminaire du modèle, et ne manqueront pas d'être explorées dans le rapport final.

Les administrations et services, et les industries automobiles se localisent naturellement plus près des zones d'habitation ; en revanche les activités financières et le secteur d'énergie ne « font pas bon ménage » avec les logements.

Les secteurs d'énergie et les activités financières sont les plus sensibles à l'accessibilité en TC. En revanche, les industries des biens intermédiaires et les activités d'éducation, santé et les actions sociales sont les plus sensibles à l'accessibilité par la voiture.

Secteur	Coefficient	Dans Paris	Log prix	log nombre logements	Pauvre -M	TC	VP	Jeune
1	estimate	0,2161	-0,1104	0,1142	0,9311	0,0113	-0,0107	-4,0035
	<i>t-values</i>	3,5738	-2,1235	6,3161	2,2257	7,5637	-3,4470	-9,9182
2	estimate	-0,4250	0,5118	-0,1056	6,3295	-0,0021	-0,0128	0,6759
	<i>t-values</i>	-16,8930	17,6618	-11,1785	23,1305	-1,9873	-7,2794	3,0852
3	estimate	-3,8248	0,0117	0,2946	-5,0653	0,0035	0,0254	1,9322
	<i>t-values</i>	-21,3799	0,1628	11,5324	-10,6924	1,3181	6,3389	5,2710
4	estimate	-1,0133	-0,6907	0,0693	-4,5938	-0,0009	0,0179	-2,1659
	<i>t-values</i>	-26,3125	-19,1119	7,1470	-18,5268	-0,9631	9,5197	-11,8583
5	estimate	-0,5272	-0,4932	-0,0380	3,2599	0,0060	-0,0464	-4,6149
	<i>t-values</i>	-12,0574	-13,0214	-3,6535	12,6506	6,6827	-27,2895	-21,3541
6	estimate	-0,8118	-0,6715	-0,1219	-9,3415	-0,0907	0,0595	-2,2393
	<i>t-values</i>	-7,2619	-5,5585	-3,1785	-8,0477	-25,4257	10,8307	-2,6212
7	estimate	-0,1511	-0,5307	0,1820	1,1154	0,0114	-0,0045	-3,1536
	<i>t-values</i>	-5,1176	-18,0539	21,8123	5,3278	14,6083	-2,9980	-18,7303
8	estimate	0,0674	-0,3560	-0,0011	-1,9975	-0,0024	-0,0003	-0,8287
	<i>t-values</i>	4,1946	-23,1839	-0,2205	-15,8475	-4,7502	-0,3257	-8,0599
9	estimate	0,1479	-1,2023	-0,0287	-3,9974	0,0154	0,0305	-1,2567
	<i>t-values</i>	3,9647	-33,4156	-2,8651	-15,9472	15,9593	15,4320	-5,8321
10	estimate	-0,0210	0,0323	-0,2398	-6,1106	-0,0491	0,0361	4,4851
	<i>t-values</i>	-0,8024	1,0032	-22,2921	-19,9789	-35,4791	19,3318	19,3517
11	estimate	0,2904	0,8797	0,1815	7,4412	-0,0012	0,0081	-0,1436
	<i>t-values</i>	9,1265	23,9898	13,7022	19,1795	-0,7391	3,3736	-0,4741
12	estimate	-0,2714	0,1352	0,0553	-1,5615	-0,0116	0,0358	1,7827
	<i>t-values</i>	-24,0868	10,0232	12,3488	-12,8263	-23,0959	40,8667	19,3633
13	estimate	0,6511	0,6872	0,0143	2,2391	0,0253	-0,0104	1,7838
	<i>t-values</i>	29,9572	31,8248	1,9097	10,8235	35,0322	-7,5785	11,0922
14	estimate	0,7340	0,4744	0,1650	4,4670	0,0084	-0,0493	-2,1162
	<i>t-values</i>	27,8207	18,1251	18,4933	18,9907	10,2510	-33,8695	-10,9899
15	estimate	1,2886	-0,1787	0,2663	0,8586	-0,0090	-0,0233	1,2498
	<i>t-values</i>	34,3214	-4,8712	19,7954	2,5291	-6,7769	-10,3729	5,0911

Tableau 35 – Premiers résultats d'estimation

7.6. Nouveau modèle de demande : le modèle EVA

La méthode EVA, élaborée à l'Université de Dresde par Lohse, propose le choix simultané de la destination et du mode. Elle prend en compte les attributs socio-économiques des zones ainsi que ceux de l'utilisateur, ce qui est nouveau par rapport au modèle classique de l'IAURIF.

7.6.1. Présentation théorique

Génération des déplacements par EVA

L'approche EVA calcule l'émission et l'attraction de la même manière que le modèle classique. Seulement, à côté de l'attraction, pour certains motifs, il calcule le potentiel d'accueil des voyageurs, ce qu'on appellera une contrainte douce. Par exemple, pour le motif achats, un centre commercial peut accueillir jusqu'à un certain nombre de clients mais ce n'est pas sûr qu'on puisse atteindre cette capacité.

L'émission de déplacements pour chaque motif p par la zone i vaut :

$$N = \sum_i E_i \quad \text{où} \quad E_i = \sum_p TM_p \cdot PT_{ip} \cdot u_{ip}$$

avec

TM_p : taux de génération d'une unité du groupe p

PT_{ip} : Nombre des unités du groupe (motif) p

u_{ip} : part du déplacement intrazonal du groupe p dans la zone i

L'attraction de déplacements pour chaque motif avec les contraintes dures vaut :

$$A_j = \frac{\sum_p TM_p \cdot PT_{jp}}{\sum_{z=1,Z} \sum_p TM_p \cdot PT_{zp}} \cdot N$$

avec la contrainte douce

$$A_{\max j} = \frac{\sum_p U_{jp} \cdot TM_p \cdot PT_{jp}}{\sum_{z=1,Z} \sum_p TM_p \cdot PT_{zp}} \cdot N$$

avec

TM_p : taux d'attraction du motif p

PT_{jp} : volume d'attracteurs du motif p dans la zone j

U_{jp} : facteur de chargement de la zone j en motif p

A_j : attraction vers la zone j

$A_{\max j}$: la trafic maximum attiré vers la zone j

N : le volume total du trafic.

Choix simultané de destination et de mode dans EVA

Le nombre de déplacements émis par la zone i , E_i , est supposé être connu, au même titre que le nombre (contrainte dure) ou le nombre maximum (contrainte douce) de déplacements vers la zone j , A_j . La part du mode k dans les déplacements allant de la zone i à la zone j est calculée comme une fonction du coût du voyage. Cette fonction peut être représentée par la

probabilité conditionnelle d'accepter de voyager connaissant l'origine, la destination et le mode :

$$BW_{ijk} = P(W|E_i \cap A_j \cap M_k)$$

W : le déplacement de i vers j utilisant le mode k est accepté au regard des coûts générés.

La forme acceptable de la fonction des coûts générés dépend de la qualité d'ajustement obtenu et la flexibilité désirée des élasticités. Lohse suggère la transformation non-linéaire des coûts générés, w , qui requiert 3 paramètres additionnels E , F et G pour obtenir

$$BW = f(w) = \left[1 + \left(\frac{w}{F} \right)^G \right]^{\frac{E}{G}}$$

$$\varepsilon(w) = -E \cdot \frac{w^G}{F^G + w^G}$$

Dans leur article récent, Vrtic et al proposent de calculer BW par un modèle logit multinomial ou emboîté à estimer.

Pour distribuer le nombre total des déplacements N entre les relations particulières D_{ijk} , ils utilisent la formulation ci-dessous structurée sous forme du modèle de Bayes :

$$D_{ijk} = \frac{P(E_i \cap A_j \cap M_k | W)}{\sum_{i'=1, Z} \sum_{j'=1, Z} \sum_{k'=1, K} P(W | E_{i'} \cap A_{j'} \cap M_{k'})} \cdot N$$

$$D_{ijk} = \frac{P(E_i) \cdot P(A_j) \cdot P(M_k) \cdot P(W | E_i \cap A_j \cap M_k)}{\sum_{i'=1, Z} \sum_{j'=1, Z} \sum_{k'=1, K} P(E_{i'}) \cdot P(A_{j'}) \cdot P(M_{k'}) \cdot P(W | E_{i'} \cap A_{j'} \cap M_{k'})} \cdot N$$

On peut choisir n'importe quelle forme fonctionnelle pour le calcul de la probabilité, par exemple le modèle logit.

Dans le cas de contraintes rigides, les probabilités conditionnelles sont connues :

$$P(E_i | W) = \frac{E_i}{N}, P(A_j | W) = \frac{A_j}{N}, P(M_k | W) = \frac{M_k}{N}$$

Les ratios des probabilités conditionnelles et non conditionnelles définissent les facteurs d'équilibre initialement inconnus :

$$q_i = \frac{P(E_i)}{P(E_i|W)}, z_j = \frac{P(A_j)}{P(A_j|W)}, a_k = \frac{P(M_k)}{P(M_k|W)},$$

avec $P(E_i) = P(E_i|W).q_i$; $P(A_j) = P(A_j|W).z_j$; $P(M_k) = P(M_k|W).a_k$ on obtient

$$D_{ijk} = \frac{P(E_i|W).q_i.P(A_j|W).z_j.P(M_k|W).a_k.P(W|E_i \cap A_j \cap M_k)}{\sum_{i'=1,Z} \sum_{j'=1,Z} \sum_{k'=1,K} P(E_{i'}|W).q_{i'}.P(A_{j'}|W).z_{j'}.P(M_{k'}|W).a_{k'}.P(W|E_{i'} \cap A_{j'} \cap M_{k'})}.N$$

Avec la connaissance des probabilités $BW_{ijk} = P(W|E_i \cap A_j \cap M_k)$ et des probabilités conditionnelles $P(E_i|W) = \frac{E_i}{N}$, $P(A_j|W) = \frac{A_j}{N}$, $P(M_k|W) = \frac{M_k}{N}$, il est possible de déterminer les facteurs d'équilibre q_i , z_j , a_k et les probabilités $P(E_i)$, $P(A_j)$ et $P(M_k)$.

Après quelques transformations, on obtient le système d'équations avec contraintes :

$$D_{ijk} = BW_{ijk} \cdot \frac{E_i}{N} \cdot q_i \cdot \frac{A_j}{N} \cdot z_j \cdot \frac{M_k}{N} \cdot a_k \cdot f = BW_{ijk} \cdot f q_i \cdot f z_j \cdot f a_k$$

sous contraintes :

$$E_i = \sum_j \sum_k D_{ijk}$$

$$A_j = \sum_i \sum_k D_{ijk}$$

$$M_k = \sum_i \sum_j D_{ijk}$$

Le modèle EVA, qui est décrit de cette façon, est solvable par un processus itératif. Afin de résoudre le système d'équations, la transformation linéaire de la matrice BW est recherchée.

Pour les contraintes douces, le second ensemble de contraintes change comme suit :

$$A_{\max j} \geq A_j = \sum_i \sum_k D_{ijk}$$

Pour résoudre cette équation, nous utilisons une méthode numérique itérative de résolution d'un problème de point fixe :

$$fq_i = \frac{E_i}{\sum_j \sum_k BW_{ijk} \cdot fz_j \cdot fa_k}; fz_j = \frac{A_j}{\sum_i \sum_k BW_{ijk} \cdot fq_j \cdot fa_k}; fa_k = \frac{M_k}{\sum_i \sum_j BW_{ijk} \cdot fq_i \cdot fz_j}$$

Nous commençons par $D_{ijk}=BW_{ijk}, fq_i(0)=fz_j(0)=fa_k(0)=f(1)=1$ et calculons :

$$D_{ijk}(1) = D_{ijk}(0) \cdot \frac{E_i}{\sum_i \sum_k D_{ijk}(0) \cdot z_j(0) \cdot a_k(0)} \cdot \frac{A_j}{\sum_i \sum_k D_{ijk}(0) \cdot q_i(0) \cdot a_k(0)} \cdot \frac{M_k}{\sum_i \sum_k D_{ijk}(0) \cdot q_i(0) \cdot z_j(0)} \cdot f(0)$$

Les étapes suivantes calculent :

$$D_{ijk}(s+1) = D_{ijk}(s) \cdot \frac{q_i(s)}{\bar{q}_i(s)} \cdot \frac{z_j(s)}{\bar{z}_j(s)} \cdot \frac{a_k(s)}{\bar{a}_k(s)} \cdot f(s)$$

Avec

$$E_i(s) = \sum_j \sum_k D_{ijk}(s); A_j(s) = \sum_i \sum_k D_{ijk}(s); M_k(s) = \sum_i \sum_j D_{ijk}(s); N(s) = \sum_i \sum_j \sum_k D_{ijk}(s) ;$$

$$q_i(s) = \frac{E_i}{E_i(s)}; z_j(s) = \frac{A_j}{A_j(s)}; a_k(s) = \frac{M_k}{M_k(s)}; f(s) = \frac{N}{N(s)};$$

$$\bar{q}_i(s) = \frac{\sum_j \sum_k D_{ijk}(s) \cdot (z_j(s) + a_k(s))}{2 \cdot E_i(s)}$$

$$\bar{z}_j(s) = \frac{\sum_i \sum_k D_{ijk}(s) \cdot (q_i(s) + a_k(s))}{2 \cdot A_j(s)}$$

$$\bar{a}_k(s) = \frac{\sum_i \sum_j D_{ijk}(s) \cdot (q_i(s) + z_j(s))}{2 \cdot M_k(s)}$$

Nous pouvons également itérer entre les calculs de la matrice et l'affectation pour obtenir une solution mutuellement cohérente.

7.6.2. Bases de calage

Avant d'entamer l'estimation des paramètres déterminants du choix de destination et de mode de transport, nous avons procédé à la préparation des données de manière à les faire rentrer dans la procédure « proc MDC » de SAS.

Les données sont extraites de la base de l'EGT 2001-2002. Nous avons sélectionné un échantillon d'individus utilisant les deux modes, transport en commun (TC) et véhicule particulier (VP). Sur la base de cet échantillon, nous avons sélectionné certaines variables (caractéristiques) individuelles que nous avons jugé déterminantes dans le choix de la destination et du mode de transport.

Au cours de cette étape de préparation des données, nous avons simulé un certain nombre d'alternatives que nous supposons être parmi lesquelles les individus ont effectué leurs choix. Nous avons utilisé pour cela la loi uniforme.

Nous générons pour chaque individu 14 destinations en mode VP et 7 destinations en mode TC. Cela donne pour chaque individu un choix à effectuer parmi 22 alternatives.

La base finale contient au total 36 941 déplacements avec environ 60 variables, y compris les variables indicatrices.

Nous avons procédé à l'estimation sur trois sous-tableaux créés selon les motifs de déplacements retenus, à savoir :

- motif 1 : Domicile-travail et affaires professionnelles
- motif 2 : Domicile-Achats
- motif 3 : Autres

Variables

Notons que les variables des temps TC sont calculées par le modèle de l'IAURIF, sur un réseau TC de l'année 1996, à l'heure de pointe du matin. Par contre les variables de temps en VP sont calculées sur un réseau VP de l'année 1996, à l'heure de pointe du soir. La matrice résultante des temps a ensuite été transposée pour obtenir les temps à l'heure de pointe du matin.

L'ensemble des variables retenues dans le modèle est :

- T_VP : Temps de parcours, en minutes, en véhicule personnel (0 si mode = TC) ;
- T_PRKG : temps (mn) nécessaire pour trouver un parking pour stationner (0 si mode=TC) ;
- T_PAR : Temps effectif (mn) de parcours en transport en commun (0 si mode=VP) ;
- T_ATT : Temps total (mn) d'attente dans les gares et stations (0 si mode = VP) ;
- T_RAB : Temps total (mn) de rabattement au départ et à l'arrivée (0 si mode=VP) ;
- T_COR : Temps (mn) passé entre les correspondances (0 si mode=VP) ;
- AGE_T : Tranches d'âge (1=6 à 24 ans, 2=25 à 54 ans, 3=55 et plus) ;
- AGE_Ti : Dummy égale à 1 si pour un individu dont la tranche est *i* et 0 sinon ;
- TC : Dummy égale à 1 si le mode choisi est TC 0 sinon ;
- SEXE : Sexe de l'individu, égale à 1 pour homme et 2 pour femme ;
- HOM_TC : Dummy égale à 1 pour un homme qui choisit le TC et 0 sinon ;
- FEM_TC : Dummy égale à 1 pour une femme qui choisit le TC et 0 sinon ;
- AGTi_TC=AGE_Ti*TC variable de croisement entre les variables AGE_Ti et TC ;
- CSPi : Catégorie socioprofessionnelle de l'individu, égale à 1 pour un individu de catégorie *i* et 0 sinon (1= agriculteur, artisans, commerçant et chef d'entreprise, 2=cadre, professions libérales, professions intermédiaires ; 3= employés, ouvriers et étudiants ; 4= retraités, autres inactifs (référence))
- CSPi_TC = CSPi*TC ; (CSP4_TC : référence)
- REV_{*i*} : Revenu de l'individu, égal à 1 pour un individu à revenu dans la classe *i* et 0 sinon (1= bas revenu, 2=revenu moyen, 3=haut revenu) ;
- REV_{*i*}_TC = REV_{*i*}*TC ;
- D75 (O75), DPC (OPC) et DGC (OGC) : sont respectivement les départements des zones de destination (origine), Paris (référence), ceux de la petite couronne et ceux de la grande couronne ;
 - D75 (O75) = 1 si le département de destination (origine) est Paris, 0 sinon ; (référence)
 - DPC (OPC) = 1 si le département de destination (origine) est 92, 93 ou 94, 0 sinon ;
 - DGC (OGC) = 1 si le département de destination (origine) est 91, 95, 77 ou 78, 0 sinon ;
- AGTi_j=AGE_Ti*Dj : croisement entre la tranche d'âge *i* et les départements des zones de destination;

- $CSPi_Dj = CSPi * Dj$; $i=1,2,3,4$; $j= 75, PC, GC$;
- $REVi_Dj = REVi * Dj$; $i=1,2,3$; $j= 75, PC, GC$;
- $Oi_j = Oi * Dj$, ($i,j= 75, PC, GC$)
- POP : Population de la zone de destination ;
- ACTIF : Nombre d'actifs résidant dans la zone de destination en milliers d'individus;
- EMPLOI : Nombre d'emplois proposés dans la zone de destination en milliers d'individus;
- SUR_COM : La surface commerciale de la zone de destination en milliers de mètres carré.

7.6.3. Résultats du calage

Nous avons divisé par 1000 ces dernières variables pour éviter des coefficients minuscules. L'estimation est faite sur les trois sous-bases, cela nous permet d'interpréter les différents coefficients selon le motif de déplacement. La procédure proc MDC de SAS a été utilisée. Nous obtenons les résultats suivants.

Motif 1 : Domicile-Travail et affaires professionnelles

Pour ce motif, 6961 déplacements ont été choisis. Nous avons lancé la proc MDC pour tous les individus qui se sont déplacés de leur domicile vers le lieu de travail ou professionnel.

Le critère de McFadden LRI est de 0,3875 qui est une valeur satisfaisante.

Parameter	Estimate	Standard Error	t Value	Pr > t
T_VP	-0.0531	0.000814	-65.25	<.0001
T_PRKG	-0.1305	0.0116	-11.24	<.0001
TC_ATT	-0.0740	0.007313	-10.12	<.0001
TC_RAB	-0.0600	0.003844	-15.62	<.0001
TC_COR	-0.0170	0.009052	-1.88	0.0597
TC_PAR	-0.0547	0.002231	-24.53	<.0001
EMPLOI	0.0486	0.001000	48.63	<.0001
SUR_COM	0.001540	0.000373	4.13	<.0001
HOM_TC	-0.0691	2.2642	-0.03	0.9757
FEM_TC	0.5786	2.2643	0.26	0.7983
AGT2_TC	0.0797	2.2771	0.03	0.9721
AGT3_TC	0.1717	2.2781	0.08	0.9399
CSP1_TC	-1.8749	0.2763	-6.79	<.0001
CSP2_TC	-0.4267	0.1958	-2.18	0.0293
CSP3_TC	-0.1189	0.1962	-0.61	0.5444
REV1_TC	0.6710	0.1071	6.27	<.0001
REV2_TC	0.3459	0.0754	4.59	<.0001
AGT2_PC	-0.4649	0.2310	-2.01	0.0441
AGT2_GC	-0.9735	0.2781	-3.50	0.0005
AGT3_PC	-0.7361	0.2387	-3.08	0.0020
AGT3_GC	-0.9729	0.2857	-3.41	0.0007
CSP1_PC	-0.3615	0.2831	-1.28	0.2017
CSP1_GC	-0.4552	0.3300	-1.38	0.1677
CSP2_PC	-0.002442	0.2220	-0.01	0.9912
CSP2_GC	-0.1187	0.2667	-0.45	0.6563
CSP3_PC	-0.0719	0.2226	-0.32	0.7467
CSP3_GC	-0.1249	0.2667	-0.47	0.6396
REV1_PC	0.3758	0.1219	3.08	0.0020

REV1_GC	0.8916	0.1402	6.36	<.0001
REV2_PC	0.2099	0.0831	2.52	0.0116
REV2_GC	0.4472	0.0982	4.55	<.0001
O75_PC	0.0938	0.1026	0.91	0.3606
O75_GC	0.2089	0.1417	1.47	0.1406
OPC_75	0.1688	0.0772	2.19	0.0288
OPC_GC	-0.4237	0.0931	-4.55	<.0001

Tableau 36 - paramètres estimés pour le motif Domicile -> Travail+Aff Pro

Pour ce motif de destination Domicile-Travail et affaires professionnelles, les signes des coefficients pour les temps de déplacement, que ce soit en VP ou en TC, sont négatifs, donc conformes à l'intuition. Cela signifie que les gens évitent les zones pour lesquelles les temps de déplacement sont élevés. Donc nous dirons que les individus préfèrent travailler dans des zones proches de leurs habitations et les zones les plus accessibles.

On constate un coefficient de EMPLOI positif, donc les individus choisissent les zones qui offrent le plus d'emplois.

Les catégories CSP1 (agriculteur, artisans, commerçant et chef d'entreprise) et CSP2 (cadres, professions libérales, professions intermédiaires) préfèrent plus la VP que les TC, relativement à la CSP4, mais la CSP2 à un degré moindre. Le coefficient de CSP3_TC n'est pas significatif.

Les individus à revenus bas et moyen ont tendance à utiliser plus les TC que la VP, par rapport aux revenus élevés, et ceux à bas revenus préfèrent les TC plus que les autres.

Les tranches d'âges 2 (25 à 54 ans) et 3 (55 ans et plus) préfèrent travailler à Paris que dans les deux couronnes, (AGT2_PC= -0.4649, AGT2_GC= -0.9735, AGT3_PC= -0.7361, AGT3_GC= -0.9729). Plus on s'éloigne de Paris moins on est satisfait.

Le choix de destination selon les CSP n'est pas interprétable car tous les coefficients ne sont pas significatifs.

Enfin, les individus à bas revenus et moyen revenus préfèrent travailler plus dans les deux couronnes qu'à Paris par rapport aux gros revenus, mais les individus à bas revenus retirent plus d'utilité que ceux à revenus moyens.

Motif 2 : Domicile-Achats

Ce motif concerne 1391 déplacements. Les résultats sont les suivants :

Tous les coefficients des variables de temps de déplacement sont négatifs, ce qui veut dire que les individus préfèrent faire leurs courses à proximité de leurs habitations.

Le coefficient de SUR_COM (positif et significatif) indique que les individus font leurs achats dans des zones à forte densité en terme de surfaces commerciales.

Dans ce motif, le sexe détermine parfaitement le choix de mode. Les deux coefficients HOM_TC et FEM_TC sont négatifs et très significatifs, donc aussi bien les hommes que les femmes, préfèrent leurs véhicules personnels que les TC pour faire leurs achats.

Nous avons la même significativité que pour AGT2_TC et AGT3_TC, les tranches 25 à 54 ans et 55 ans et plus préfèrent utiliser la VP que les TC. De même, ces deux tranches préfèrent se rendre à Paris plutôt que vers les deux couronnes.

Comme dans le cas de motif domicile-travail, les individus à bas revenus et à revenus moyens préfèrent se rendre dans les deux couronnes qu'à Paris pour faire leurs achats. Cela s'explique par le fait que dans les deux couronnes, il y a des magasins de grandes surfaces dont les prix sont moins élevés que dans les petites boutiques de Paris.

Parameter	Estimate	Standard Error	t Value	Pr > t
T_VP	-0.1258	0.004419	-28.47	<.0001
T_PRKG	-0.1875	0.0376	-4.99	<.0001
TC_ATT	-0.0775	0.0343	-2.26	0.0240
TC_RAB	-0.0992	0.0163	-6.08	<.0001
TC_COR	-0.1134	0.0504	-2.25	0.0246
TC_PAR	-0.1418	0.0138	-10.25	<.0001
SUR_COM	0.0221	0.001065	20.77	<.0001
HOM_TC	-1.2688	0	-Infty	<.0001
FEM_TC	-0.7804	0	-Infty	<.0001
AGT2_TC	-1.0718	0	-Infty	<.0001
AGT3_TC	-1.1156	0	-Infty	<.0001
CSP1_TC	0.2096	1.2277	0.17	0.8644
CSP2_TC	-0.6556	0.4716	-1.39	0.1645
CSP3_TC	0.2616	0.3429	0.76	0.4455
REV1_TC	1.5700	0.4048	3.88	0.0001
REV2_TC	1.4217	0.3631	3.92	<.0001
AGT2_PC	-2.3158	0.5799	-3.99	<.0001
AGT2_GC	-2.4891	0.6260	-3.98	<.0001
AGT3_PC	-3.0564	0.5160	-5.92	<.0001
AGT3_GC	-3.6455	0.5429	-6.71	<.0001
CSP1_PC	0.5001	1.2758	0.39	0.6951
CSP1_GC	0.9826	1.5154	0.65	0.5167
CSP2_PC	-0.4465	0.5180	-0.86	0.3887
CSP2_GC	-0.3134	0.5840	-0.54	0.5915
CSP3_PC	0.0495	0.4191	0.12	0.9060
CSP3_GC	0.5175	0.4812	1.08	0.2822
REV1_PC	1.2844	0.4694	2.74	0.0062
REV1_GC	1.7690	0.5367	3.30	0.0010
REV2_PC	1.4197	0.4133	3.44	0.0006
REV2_GC	1.7262	0.4720	3.66	0.0003
O75_PC	0.4822	0.5212	0.93	0.3549
O75_GC	1.4388	0.6314	2.28	0.0227
OPC_75	0.4430	0.3918	1.13	0.2582
OPC_GC	-0.8177	0.3977	-2.06	0.0398

Tableau 37 - paramètres estimés pour le motif Domicile -> Achats

Motif 3 : Autres

Pour ce motif, nous avons tenu compte de toutes les caractéristiques des zones destinations, à savoir la population, nombre d'actifs résidents, nombre d'emplois proposés et la surface commerciale. Les coefficients estimés sont présentés dans le tableau suivant.

On constate que la majorité des coefficients sont significatifs, cela s'explique par le fait qu'on a plus d'observations par rapport aux autres motifs (au total 27300 déplacements).

Comme dans les deux autres motifs, les coefficients liés aux variables des temps de déplacements sont tous négatifs et largement significatifs. Donc, quel que soit le motif de déplacement, les individus ne retirent pas d'utilité des longs déplacements.

Parameter	Estimate	Standard Error	t Value	Pr > t
T_VP	-0.0673	0.000496	-135.72	<.0001
T_PRKG	-0.1409	0.005881	-23.95	<.0001
TC_ATT	-0.0411	0.003742	-10.98	<.0001
TC_RAB	-0.0553	0.002000	-27.65	<.0001
TC_COR	-0.0548	0.005299	-10.35	<.0001
TC_PAR	-0.0650	0.001243	-52.28	<.0001
POP	0.0419	0.003338	12.55	<.0001
ACTIF	-0.0423	0.006679	-6.33	<.0001
EMPLOI	0.008428	0.000788	10.70	<.0001
SUR_COM	-0.000637	0.000226	-2.82	0.0047
HOM_TC	-0.8283	3.2744	-0.25	0.8003
FEM_TC	-0.3552	3.2744	-0.11	0.9136
AGT2_TC	-0.8271	3.2750	-0.25	0.8006
AGT3_TC	-0.2441	3.2751	-0.07	0.9406
CSP1_TC	-1.0152	0.1429	-7.11	<.0001
CSP2_TC	0.4016	0.0524	7.67	<.0001
CSP3_TC	1.0899	0.0446	24.45	<.0001
REV1_TC	0.6568	0.0501	13.12	<.0001
REV2_TC	0.2728	0.0381	7.15	<.0001
AGT2_PC	-0.7248	0.0845	-8.58	<.0001
AGT2_GC	-0.2491	0.0883	-2.82	0.0048
AGT3_PC	-0.8316	0.0768	-10.83	<.0001
AGT3_GC	-0.3231	0.0787	-4.10	<.0001
CSP1_PC	-0.1240	0.1330	-0.93	0.3515
CSP1_GC	-0.1881	0.1502	-1.25	0.2106
CSP2_PC	0.2198	0.0635	3.46	0.0005
CSP2_GC	0.3709	0.0699	5.31	<.0001
CSP3_PC	0.5515	0.0554	9.96	<.0001
CSP3_GC	0.8196	0.0608	13.48	<.0001
REV1_PC	0.1465	0.0626	2.34	0.0192
REV1_GC	0.0321	0.0685	0.47	0.6391
REV2_PC	0.2633	0.0479	5.50	<.0001
REV2_GC	0.2802	0.0514	5.45	<.0001
O75_PC	0.1959	0.0634	3.09	0.0020
O75_GC	-0.0554	0.0650	-0.85	0.3939
OPC_75	0.4931	0.0624	7.91	<.0001
OPC_GC	-0.5453	0.0612	-8.91	<.0001

Tableau 38 - paramètres estimés pour les autres motifs

En ce qui concerne les variables caractéristiques des zones de destination, on observe que les individus préfèrent aller vers des zones peuplées, mais ayant peu d'actifs et de surfaces commerciales (POP=0.0419, ACTIF=-0.0423, SUR_COM=-0.000637).

Dans ce motif (Domicile-Achats), le sexe ne détermine pas le choix de mode. Les deux coefficients de HOM_TC et FEM_TC ne sont pas du tout significatifs. C'est le cas également des tranches d'âges, les coefficients de AGT2_TC et AGT3_TC ne sont pas significatifs.

La CSP3, contrairement à CSP1 et CSP2, préfère les TC à la VP.

Dans ce motif, comme dans le motif 1, la préférence pour les TC diminue avec le revenu (REV1_TC=0.6568, REV2_TC=0.2728), les individus à hauts revenus retirent moins d'utilité des TC que les individus à bas revenus.

Les CSP2 et CSP3 dont les coefficients sont positifs et significatifs préfèrent se diriger plus vers Paris que vers les deux couronnes.

7.6.4. Application du modèle

Ces travaux sur le modèle EVA en sont pour l'instant au stade de la recherche. Nous avons d'abord essayé de procéder à des analyses économétriques approfondies, d'expliquer les phénomènes de choix du mode et de la destination. La programmation des formules et son application sont en cours.

7.7. Perspectives

Dans la version actuelle du modèle UrbanSim (2.3), les choix des agents (individus, ménages, entreprises) s'effectuent au niveau des cellules carrées (en Ile-de-France, leur taille est de 500m * 500m). L'estimation des modèles économétriques avec UrbanSim 2.3 nécessite donc des données définies au niveau de ces cellules carrées. Ces données ne sont pas accessibles directement. Leur obtention nécessite un traitement statistique difficile, mais réalisable grâce au Système d'information géographique.

Pour pallier ce problème, une nouvelle version d'UrbanSim est en cours de réalisation par Paul Waddell. Celle-ci pourra être prise en considération dans le projet *SIMAU*RIF si elle est accessible dans les mois à venir. Cette nouvelle version décrit le comportement des agents à un niveau plus agrégé et variable (par exemple, communes ou polygones dans le cas de l'Ile-de-France) compatible avec toute la finesse des données brutes disponibles. Ceci ne remet en aucun cas en cause la possibilité d'étudier des effets de voisinage. Dans la version envisagée actuellement, les modèles de choix restent de type multinomial, mais une extension permettant d'utiliser des modèles emboîtés est aussi en cours d'étude.

Les estimations présentées dans ce rapport ont été effectuées pour ces deux versions d'UrbanSim (actuelle dans les parties 2 à 6, future dans la partie 7). Les résultats obtenus sont contrastés et mettent en lumière les différences entre ces deux approches telles qu'elles ont été pressenties par le concepteur d'UrbanSim.

8. Conclusions

Ce rapport intermédiaire a fait le bilan à mi-parcours de la seconde phase.

Il reste à faire :

- le calage des deux autres modules principaux d'UrbanSim au niveau de la cellule : le module de localisation des emplois et le module de prix du foncier qui nécessitera un recueil de données spécifique au niveau infracommunal.
- le paramétrage des 6 autres modules secondaires d'UrbanSim (voir partie 4.3 du rapport n°1).
- la mise au point de l'interface entre UrbanSim et METROPOLIS
- le calage de l'ensemble UrbanSim/Davisum/METROPOLIS, étape la plus complexe
- l'élaboration des scénarios de la Tangentielle nord
- la simulation des scénarios de la Tangentielle nord

L'équipe s'attachera en priorité en 2006 à terminer le projet *SIMAU*RIF avec la version actuelle d'UrbanSim.

On poursuivra parallèlement des recherches sur le nouvel UrbanSim, en collaboration avec Paul Waddell, aujourd'hui retourné aux Etats-Unis.

Annexe 1 : Programme Scilab permettant de calculer les variables de voisinage

```
// Pour tout effacer de la mémoire de scilab avant d'exécuter le programme
clear;

// pour augmenter la taille mémoire
stacksize(5e7);

// Pour importer un jeu de données au format .csv
fichier =D:\prg_scilab\var24.txt';
ini = fscanfMat(fichier);
fichier =D:\prg_scilab\xy.txt';
xy = fscanfMat(fichier);

// Pour sauver le jeu de données
save('D:\prg_scilab\var24.bin')
save('D:\prg_scilab\xy.bin')

// Pour charger le fichier de données
load('D:\prg_scilab\var24.bin')
load('D:\prg_scilab\xy.bin')

// Debut programme calcul des variables de voisinage
// x est un vecteur composé par les abscisses des cellules en coordonnées cartésiennes
x=xy(:,1);
// y est un vecteur composé par les ordonnées des cellules en coordonnées cartésiennes
y=xy(:,2);
// vini est une matrice composée des variables pour lesquelles on souhaite calculer les
variables de voisinage qui leur sont associées
vini = [];

for i=1:length(x),
    if int(i/100) == i/100 then disp('ligne n°'+sci2exp(i)); end;
    test1 = abs(x-x(i));
    test2 = abs(y-y(i));
    test3 = test1 + test2;
    testf = and([test1<=1,test2<=1,test3>0],'c');
    vini = [vini;testf * ini];
    fprintfMat('D:\prg_scilab\resultat_var1_1.txt',vini)
end;
```

Annexe 2 : Méthode d'imputation des revenus aux ménages du Recensement général de la population

Pour imputer les revenus, nous avons utilisé la base de données Budget des familles 2000, qui renseigne, pour plus de 10 000 ménages, sur les caractéristiques socio-économiques similaires à celles présentes dans le recensement ainsi que sur le revenu disponible, transferts (allocations familiales, RMI, etc...) compris. Ici encore, le pouvoir explicatif du modèle est satisfaisant, puisque le R^2 est de 60%. Les variables explicatives utilisées ont trait à :

- à la composition familiale
 - NPER=nombre de membres dans le ménage (maximum observé= 10)
 - NPER1 à NPER4 : indicatrices pour les tailles 1, 2, 3, 4 (effet non linéaire de la taille familiale)
 - NPAM=Nombre de personnes actives dans le ménage (maximum observé=5)
 - NPAM0 à NPAM2 : indicatrices pour le nombre d'actifs dans le ménage (effet non linéaire du nombre d'actifs)
- aux caractéristiques du chef de ménage
 - Foreign : Indicatrice = 1 si chef de ménage de nationalité étrangère, 0 sinon
 - IDF : Indicatrice = 1 si le ménage réside en Ile de France, 0 sinon
 - AGEM = Age du chef de ménage, en années
 - AGEM2 à AGEM4 : age du chef de ménage, élevé aux puissances 2, 3, 4 (effet non linéaire et non monotone de l'âge de chef de ménage)
- Variables croisées pour les caractéristiques de chef de ménage
 - Occupation : OCCUPCPRG=2 (chômeur),3 (étudiant), 5 (retraité), 8 (inactif), 9 (actif)
 - Niveau d'études : ETUDPR=0 (< certificat d'études), 1 (certificat d'études), 2 (CAP, BEP), 3 (BEPC), 4 (BAC technique), 5 (BAC), 6 (BAC + 2), 7 (>= BAC + 3), ETUDPRG=0,1,2,3,4 (BAC et +)
 - Sexe : Homme=1, Femme=2

Estimation du revenu des ménages

Nous avons choisi d'estimer le logarithme du revenu en fonction des caractéristiques du ménage, d'une part parce que ces caractéristiques ont plutôt un effet multiplicatif qu'additif et d'autre part parce que la qualité de la régression était bien supérieure (R^2 de l'ordre de 60% en log contre 40% en niveau). Les résultats de cette estimation sont reproduits dans le tableau 38.

Paramètre	Coefficient	Ecart-type	t	Pr> t
Constante	10.01515485	0.06910075	144.94	<.0001
NPAM	0.09007417	0.01886205	4.78	<.0001
NPAM=0	-0.34839717	0.04574487	-7.62	<.0001
NPAM=1	-0.15904342	0.02410953	-6.60	<.0001
NPAM>1	Référence			
NPER=1	-0.60733665	0.01991099	-30.50	<.0001
NPER=2	-0.23491093	0.01789212	-13.13	<.0001
NPER=3	-0.14149113	0.01801016	-7.86	<.0001
NPER=4	-0.08469468	0.01770273	-4.78	<.0001
NPER>4	Référence			
FOREIGN=1	0.21482435	0.02139571	10.04	<.0001
FOREIGN=2	Référence			
IDF=0	-0.20259883	0.01241638	-16.32	<.0001
IDF=1	Référence			
AGEM	0.04439184	0.00568993	7.80	<.0001
AgeM2	-0.00116625	0.00032208	-3.62	0.0003
AgeM3	0.00002119	0.00000700	3.03	0.0025
AgeM4	-0.00000014	0.00000005	-2.85	0.0044
AGEM*SEXEPR=1	-0.00141606	0.00070664	-2.00	0.0451
AGEM*SEXEPR=2	Référence			
OCCUPCPRG*ETUDPRG=(2,0)	-0.67502168	0.12976231	-5.20	<.0001
OCCUPCPRG*ETUDPRG=(2,1)	-0.67766782	0.09562021	-7.09	<.0001
OCCUPCPRG*ETUDPRG=(2,2)	-0.46534114	0.07492344	-6.21	<.0001
OCCUPCPRG*ETUDPRG=(2,3)	-0.63015061	0.06832823	-9.22	<.0001
OCCUPCPRG*ETUDPRG=(2,4)	-0.22169613	0.05525718	-4.01	<.0001
OCCUPCPRG*ETUDPRG=(3,2)	-0.67113779	0.29968183	-2.24	0.0251
OCCUPCPRG*ETUDPRG=(3,3)	-0.51350371	0.18625006	-2.76	0.0058
OCCUPCPRG*ETUDPRG=(3,4)	-0.54296189	0.05440472	-9.98	<.0001
OCCUPCPRG*ETUDPRG=(5,0)	0.13617803	0.17537466	0.78	0.4375
OCCUPCPRG*ETUDPRG=(5,1)	0.07553095	0.14878422	0.51	0.6117
OCCUPCPRG*ETUDPRG=(5,2)	0.30370857	0.13138942	2.31	0.0208
OCCUPCPRG*ETUDPRG=(5,3)	0.14602936	0.13094019	1.12	0.2648
OCCUPCPRG*ETUDPRG=(5,4)	0.57243746	0.11933567	4.80	<.0001
OCCUPCPRG*ETUDPRG=(8,0)	-0.60872894	0.12850675	-4.74	<.0001
OCCUPCPRG*ETUDPRG=(8,1)	-0.56242033	0.10877612	-5.17	<.0001
OCCUPCPRG*ETUDPRG=(8,2)	-0.53330103	0.08894081	-6.00	<.0001
OCCUPCPRG*ETUDPRG=(8,3)	-0.57490500	0.08038527	-7.15	<.0001
OCCUPCPRG*ETUDPRG=(8,4)	-0.09618671	0.08039084	-1.20	0.2315
OCCUPCPRG*ETUDPRG=(9,0)	-0.45115234	0.08866501	-5.09	<.0001
OCCUPCPRG*ETUDPRG=(9,1)	-0.60571527	0.06558489	-9.24	<.0001
OCCUPCPRG*ETUDPRG=(9,2)	-0.37400622	0.04990094	-7.49	<.0001
OCCUPCPRG*ETUDPRG=(9,3)	-0.41963242	0.04476026	-9.38	<.0001
OCCUPCPRG*ETUDPRG=(9,4)	Référence			
AGEM*ETUDPRG=0	-0.00831702	0.00216292	-3.85	0.0001
AGEM*ETUDPRG=1	-0.00578528	0.00163549	-3.54	0.0004
AGEM*ETUDPRG=2	-0.00712243	0.00125303	-5.68	<.0001
AGEM*ETUDPRG=3	-0.00262033	0.00128216	-2.04	0.0410
AGEM*ETUDPRG=4	Référence			
AGEM*OCCUPCPRG=2	0.00149243	0.00187593	0.80	0.4263

AGEM*OCCUPCPRG=3	0.01938166	0.00955118	2.03	0.0425
AGEM*OCCUPCPRG=5	-0.01263263	0.00300544	-4.20	<.0001
AGEM*OCCUPCPRG=8	-0.00171711	0.00213603	-0.80	0.4215
AGEM*OCCUPCPRG=9	Référence			
SEXEPN*ETUDPR=(1,0)	0.14225941	0.05080389	2.80	0.0051
SEXEPN*ETUDPR=(1,1)	0.23873780	0.03906305	6.11	<.0001
SEXEPN*ETUDPR=(1,2)	0.20300879	0.03657969	5.55	<.0001
SEXEPN*ETUDPR=(1,3)	0.21978438	0.03202330	6.86	<.0001
SEXEPN*ETUDPR=(1,4)	-0.16837379	0.03728581	-4.52	<.0001
SEXEPN*ETUDPR=(1,5)	-0.12737775	0.03583877	-3.55	0.0004
SEXEPN*ETUDPR=(1,6)	0.00231050	0.03420939	0.07	0.9462
SEXEPN*ETUDPR=(1,7)	0.22389688	0.03424535	6.54	<.0001
SEXEPN*ETUDPR=(2,0)	Référence			
SEXEPN*ETUDPR=(2,1)	Référence			
SEXEPN*ETUDPR=(2,2)	Référence			
SEXEPN*ETUDPR=(2,3)	Référence			
SEXEPN*ETUDPR=(2,4)	-0.36098170	0.06096085	-5.92	<.0001
SEXEPN*ETUDPR=(2,5)	-0.39901090	0.03750979	-10.64	<.0001
SEXEPN*ETUDPR=(2,6)	-0.13996318	0.03536218	-3.96	<.0001
SEXEPN*ETUDPR=(2,7)	Référence			

Tableau 39 - résultats de l'estimation de l'équation de revenus (en log)

On y constate que le revenu augmente de 9% par membre actif supplémentaire, auxquels s'ajoutent 16% à partir du 2^{ème} membre actif et 18% à partir du premier membre actif. A nombre de membres actifs constant, par rapport à un ménage de 5 membres ou plus, le revenu est réduit de 8,5% lorsqu'il n'y a que 4 membres, 14% lorsqu'il y a 3 membres, 23% pour 2 membres et 61% s'il n'y a qu'un seul membre. Toutes choses égales par ailleurs, le revenu du ménage est réduit de 21% lorsque son chef est étranger. Les ménages situés hors de l'Ile de France ont en moyenne un revenu inférieur de 20% par rapport à un ménage identique résidant en Ile de France. L'effet de l'âge du chef de ménage sur le revenu du ménage est représenté sur la figure 4. Il est maximal aux alentours de 70 ans et décroît très vite au-delà. Par rapport à un ménage dont le chef est centenaire, le revenu d'un ménage identique mais dont le chef n'aurait que 70 ans est multiplié par $\exp(1.3)=3.67$.

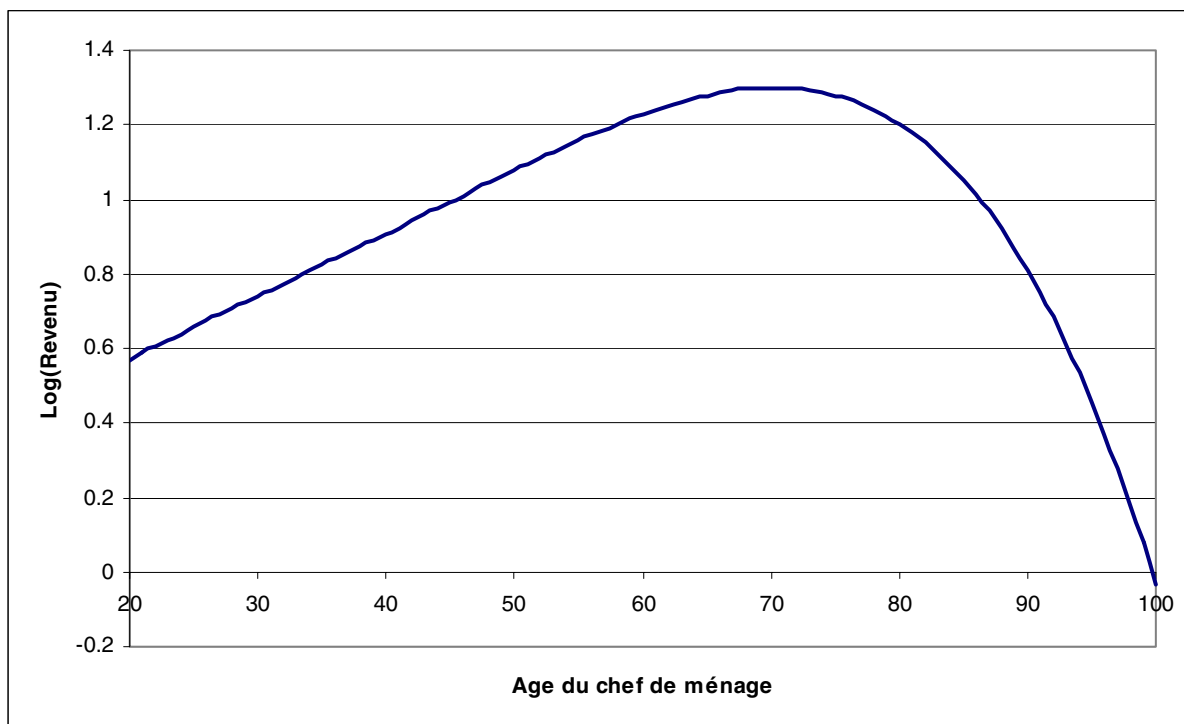


Figure 29 - effet de l'âge du chef de ménage sur le revenu du ménage (en logarithme)

Le revenu d'un ménage croît un peu moins vite (ou décroît plus vite au-delà de 70 ans) : 0.14% par an lorsque le ménage est dirigé par un homme plutôt que par une femme.

Par rapport aux ménages dont le chef a un diplôme au moins égal au baccalauréat et travaille, le revenu est plus élevé pour les retraités (à âge constant), surtout s'ils ont un niveau d'études CAP, BEP ou au moins égal au baccalauréat. Le revenu du ménage est évidemment plus faible lorsque son chef est chômeur ou inactif, et le revenu croît avec le diplôme du chef de ménage, mais d'une façon différente selon son occupation.

Le revenu croît plus vite avec l'âge (ou décroît moins vite après 70 ans) lorsque le chef de ménage est plus diplômé, la différence se faisant surtout sentir au niveau du BEPC.

Le revenu croît plus vite avec l'âge (ou décroît moins vite après 70 ans) lorsque le chef de ménage est étudiant, l'inverse se produisant pour les retraités.

Les différences hommes-femmes de revenus se font surtout ressentir pour les niveaux de diplôme élevés. Ainsi, le différentiel est de 22% en faveur des hommes pour les Bac+3 et plus, 14% pour les bac+2, (-0.399—0.127), 27% pour les niveaux baccalauréat, 19% pour les niveaux bac technique, 20% à 23% pour les niveaux certificat d'études, CAP, BEP, BEPC, mais seulement 14% pour les niveaux inférieurs au certificat d'études.

Annexe 3 : Quelques résultats d'estimation des fonctions de congestion de METROPOLIS

Type	Lien	Volume journalier	Débit max obs.	Capacité estimée pour goulot d'étranglement			Vitesse max obs.	Vitesse max estimée pour goulot d'étranglement		
				Min	Max	Moyenne		Min	Max	Moyenne
1	24494	32791	1360	978	978	978	125,0	80,0	80,0	80,0
	24495	37094	970	836	836	836	87,0	66,8	66,8	66,8
	34635	115252	2125	1807	1807	1807	109,0	81,0	81,0	81,0
	34636	116843	2330	1983	1983	1983	91,0	70,3	70,3	70,3
1 Total		301980	2330	836	1983	1401	125,0	66,8	81,0	74,5
2	23133	68035	2083	1512	1512	1512	117,0	108,3	108,3	108,3
	33436	256874	1777	1143	1266	1213	188,0	107,7	134,9	119,3
	34509	206071	1818	1302	1310	1306	112,0	96,2	98,3	97,2
	34510	334796	1928	1601	1651	1624	118,0	92,8	96,5	94,7
	34639	295178	2005	1598	1855	1685	109,0	80,6	88,6	84,5
	34640	215349	1840	1146	1396	1259	112,0	92,8	101,5	96,6
	34641	177739	2227	1897	1928	1912	139,0	96,6	106,7	101,6
	34642	178028	2163	1786	1904	1845	165,0	98,1	121,5	109,8
	34649	114234	1825	1529	1529	1529	107,0	85,0	85,0	85,0
	34650	103184	1900	1513	1513	1513	138,0	105,8	105,8	105,8
2 Total		4922372	2442	878	2227	1519	322,0	69,8	180,2	100,9
3	25903	375572	2187	1799	1839	1823	128,0	102,2	107,6	105,2
	25904	366144	2153	1232	1790	1646	151,0	100,2	133,4	109,8
	25970	309073	2303	1878	1974	1917	135,0	96,7	108,8	101,6
	26139	142558	1870	1120	1125	1123	105,0	69,3	71,9	70,4
	26144	112596	1755	1220	1539	1336	116,0	72,0	105,4	83,6
	26145	136785	1720	905	1077	1034	137,0	66,7	104,0	76,9
	26228	153054	2103	1675	1690	1683	126,0	89,5	110,5	100,0
	26229	130725	2160	1084	1320	1202	124,0	102,9	110,1	106,5
	26666	26340	1385	1316	1316	1316	115,0	103,0	103,0	103,0
	27042	28325	1485	1040	1040	1040	151,0	95,5	95,5	95,5
	27043	19409	1010	810	810	810	225,0	131,9	131,9	131,9
	27092	114724	1983	934	1202	1068	112,0	81,1	88,9	85,0
	27208	37907	1297	987	987	987	148,0	98,9	98,9	98,9
3 Total		3903194	2303	693	1974	1388	225,0	66,7	133,4	95,5

Type	Lien	Volume journalier	Débit max obs.	Capacité estimée pour goulot d'étranglement			Vitesse max obs.	Vitesse max estimée pour goulot d'étranglement		
				Min	Max	Moyenne		Min	Max	Moyenne
4	24682	35581	1685	1388	1388	1388	150,0	107,2	107,2	107,2
	24683	29192	1595	997	997	997	115,0	100,6	100,6	100,6
	24689	47021	1410	953	953	953	95,0	70,3	70,3	70,3
	25899	297619	2187	1388	1847	1622	129,0	88,0	93,3	90,1
	25900	282975	2117	1373	1386	1377	125,0	87,2	116,0	94,9
	26785	41914	1350	1005	1005	1005	108,0	80,4	80,4	80,4
	26786	25877	1310	723	723	723	91,0	53,7	53,7	53,7
	33332	122778	1885	495	1539	1164	118,0	50,7	105,7	74,8
	33333	143045	1740	1256	1530	1338	117,0	66,7	104,6	79,5
	33843	216437	2170	1323	1792	1614	178,0	68,8	111,8	99,4
	33844	274474	2465	1469	1931	1829	157,0	94,4	117,1	110,0
	33869	43163	1383	997	997	997	135,0	68,4	68,4	68,4
	33870	95695	2325	1563	1677	1620	148,0	81,2	86,6	83,9
	4 Total		1655771	2465	495	1931	1439	178,0	50,7	117,1
5	26343	40202	1257	1147	1147	1147	121,0	109,2	109,2	109,2
	26344	41448	1350	1190	1190	1190	120,0	109,2	109,2	109,2
	26560	45358	1503	928	928	928	87,0	57,2	57,2	57,2
	26561	46081	2225	1841	1841	1841	162,0	110,5	110,5	110,5
	27418	25290	1330	1138	1138	1138	117,0	102,3	102,3	102,3
	27421	19904	1115	770	770	770	179,0	85,3	85,3	85,3
5 Total		218283	2225	770	1841	1169	179,0	57,2	110,5	95,6
8	23143	292757	2950	1499	2252	1850	104,0	71,2	78,8	73,8
	23144	206063	2125	1421	1598	1549	100,0	68,2	72,6	70,7
	24223	56170	1517	1126	1126	1126	91,0	67,1	67,1	67,1
	24224	54326	1547	1126	1126	1126	92,0	65,5	65,5	65,5
8 Total		609316	2950	1126	2252	1609	104,0	65,5	78,8	71,3
9	26494	92779	2005	1391	1477	1434	113,0	66,5	102,2	84,4
	26765	46350	1940	1402	1402	1402	97,0	70,1	70,1	70,1
9 Total		139129	2005	1391	1477	1423	113,0	66,5	102,2	79,6
Grand Total		11750045	2950	495	2252	1444	322,0	50,7	180,2	93,9

