



RÉPUBLIQUE FRANÇAISE
Ministère de l'écologie et du
développement et de
l'aménagement durables

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

SIMAU^RIF :

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

Application à la Tangentielle nord

Rapport final de la 2^{ème} phase



Novembre 2007

IAURIF n° 5.03.013
DRAST n°05MT6008

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 : Arnaud LEFRANC

adpC

6 rue de la Reine, 1330 Rixensart, Belgique.

E-mail : info@adpc.be

Equipe :

Dany NGUYEN-LUONG, IAURIF

Kiarash MOTAMEDDI, THEMA

Hakim OUARAS, THEMA

Nathalie PICARD, THEMA

Michel FERNANDES, stagiaire à l'IAURIF, élève de 2ème année de l'ISIMA

Conseiller scientifique : **André DE PALMA, THEMA**

Sommaire

Remerciements	9
1. Rappel du projet <i>SIMAU</i>RIF	11
2. Portage d'UrbanSim 2 à OPUS/UrbanSim 4.....	15
2.1. Historique.....	15
2.2. Justification du portage	15
2.3. Problèmes rencontrés	16
2.3.1. Problèmes informatiques.....	17
2.3.2. Problèmes de portage de la structure des données	18
2.3.3. Problèmes de portage de la structure des modèles.....	19
2.3.4. Problèmes d'estimation des modèles	19
2.4. Conséquences sur la méthodologie du projet.....	20
2.5. Adaptation du modèle de localisation des ménages à UrbanSim 4.....	21
2.6. Adaptation du modèle d'évolution du MOS à UrbanSim 4.....	22
3. Modèle de choix de localisation des emplois	25
3.1. Statistiques descriptives sur la mobilité des emplois entre 1990 et 1999	25
3.2. Construction de la base de données de calage.....	27
3.2.1. La source : l'Enquête Régionale sur l'Emploi	27
3.2.2. Etapes de construction de la base de calage.....	28
3.3. Critères de choix de localisation des entreprises.....	31
3.4. Variables de la base de calage.....	32
3.5. Résultats du modèle	34
3.6. Validation du modèle	37
3.6.1. Pseudo- R^2 de McFadden	37
3.6.2. Pourcentage de prédictions correctes	37
4. Modèle de prix de l'immobilier résidentiel.....	39
4.1. Rappel du modèle dans UrbanSim.....	39
4.2. Statistiques descriptives sur les prix de l'immobilier entre 1990 et 1999.....	40
4.3. Estimation d'un modèle calé sur deux cas : le RER E et le Tramway T1	41
4.3.1. Base de données des Notaires géolocalisées	43
4.3.2. Analyse descriptive des données.....	44
4.3.3. Diagnostic.....	46
4.3.4. Modélisation.....	49
4.4. Estimation d'un modèle au niveau communal	54
4.4.1. Données de prix.....	54
4.4.2. Modèle implémenté dans <i>SIMAU</i> RIF	55
5. Interface simplifiée entre UrbanSim et le modèle de trafic.....	57
5.1. Simplification de la prise en compte de l'interaction.....	57
5.2. Test de modification des réseaux de transport en 1996 et 1993.....	58
5.2.1. Réseau routier.....	58
5.2.2. Réseau de transports en commun	58
5.3. Implémentation de l'interaction	59
6. Calage global de <i>SIMAU</i>RIF	61
6.1. Principes généraux	61

6.1.1. Le calage global	61
6.1.2. Installation de la base « Paris »	62
6.1.3. Principe d'appel d'une variable.....	62
6.2. Base de données d'entrée 1990	63
6.2.1. La table des ménages households.....	64
6.2.2. La table des emplois jobs	64
6.2.3. La table « development_event_history »	65
6.2.4. La table des « events »	66
6.3. Les tables de type « control totals »	66
6.3.1. Table « ANNUAL_HOUSEHOLDS_CONTROL_TOTALS »	67
6.3.2. Table « ANNUAL_EMPLOYMENT_CONTROL_TOTALS »	68
6.4. Exécution de <i>SIMAU</i> RIF de 1990 à 1999	68
6.5. Résultats de l'exécution sur la période 1990-1999	70
6.5.1. Le modèle de choix de localisation des ménages.....	71
6.5.2. Le modèle de choix de localisation des emplois	72
6.5.3. Le modèle d'évolution du MOS	72
6.5.4. Le modèle de prix.....	73
6.6. Résultats de calage	74
7. Préparation des bases de données à l'horizon 2026	75
7.1. Ce qui s'est passé entre 1999 et 2006	75
7.1.1. Population.....	75
7.1.2. Emplois.....	76
7.1.3. Occupation du sol.....	77
7.1.4. Prix de l'immobilier résidentiel.....	77
7.2. Les réseaux entre 1999 et 2026	78
7.2.1. Réseau routier.....	78
7.2.2. Réseau de transports en commun	80
7.3. Les perspectives à l'horizon 2026	83
7.4. Les tables de type « control totals »	84
7.4.1. Table « ANNUAL_HOUSEHOLDS_CONTROL_TOTALS »	84
7.4.2. Table « ANNUAL_EMPLOYMENT_CONTROL_TOTALS »	86
7.5. Les tables de types « rates » et « constants »	87
7.6. Les tables de type « events»	87
8. Application à la Tangentielle Nord.....	89
8.1. Dernières nouvelles sur le projet de la Tangentielle Nord	89
8.2. Diagnostic territorial	90
8.2.1. Aire d'influence élargie.....	90
8.2.2. Population et emplois	92
8.2.3. L'occupation du sol.....	94
8.2.4. Le marché immobilier	95
8.2.5. Synthèse	96
8.3. Elaboration de scénarios.....	97
8.4. Simulations.....	99
8.5. Résultats	99
8.5.1. La population.....	99
8.5.2. Les emplois	101
8.6. Conclusions	102
9. Conclusions et perspectives	105

Bibliographie.....	111
Annexe 1 : Résumés d'articles publiés	113
Annexe 2 : Un modèle explicatif de choix de localisation résidentielle des ménages.....	115
Annexe 3 : Un autre modèle de choix de localisation des emplois et un modèle de choix de localisation des entreprises.....	119

Remerciements

Nous tenons à remercier celles et ceux qui nous ont accompagnés dans l'élaboration de ce projet. En premier lieu le Professeur Paul Waddell de l'Université de Washington pour son investissement dans ce projet notamment en passant une année sabbatique en France, ainsi que Hana Sevcikova qui a effectué également un séjour à Paris et qui nous a accompagnés dans le lancement d'UrbanSim 4. Ensuite, nous remercions la DRAST, en particulier Monsieur Gérard Brun. Egaleme nt la DREIF et RFF, les co-financeurs du projet, en particulier Messieurs François Bertrand, Arnaud Gbahoue-Archambault, Eric Ladegaillerie et Philippe Ayoun. Enfin, tous les membres du comité de pilotage du projet.

Nous remercions également, Sophie Dantan et Laurent Wagner, étudiants à l'Université de Cergy-Pontoise, pour leur assistance.

1. Rappel du projet *SIMAU*RIF

L'Institut d'Aménagement et d'Urbanisme de la Région d'Ile-de-France (IAURIF), en partenariat avec le laboratoire THEMA de l'Université de Cergy-Pontoise et la société adpC (Belgique), a commencé en juin 2003 une recherche sur l'interaction transport-urbanisation subventionnée par le Predit 3, la Région d'Ile-de-France, la DREIF et RFF. 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 l'Etat de Washington à Seattle (Etats-Unis),
- **METROPOLIS**, un modèle de trafic routier développé par adpC et adapté à l'Ile-de-France par l'Université de Cergy-Pontoise grâce à un financement du Predit 2 ,
- **Le modèle global de demande de déplacements**, développé par l'IAURIF, avec les logiciels MinUTP, Arc-Info/ArcGis puis **Davisum**.

Il était prévu d'appliquer le modèle au territoire desservi par la future « Tangentielle nord » entre Sartrouville et Noisy-le-Sec.

Ce projet est véritablement un projet pionnier en France dans le domaine de la modélisation de l'interaction transport-urbanisation, thème de recherche interdisciplinaire, prolifique aux Etats-Unis et en pleine croissance dans le reste du monde. Nous avons eu pour ambition de dépasser la description théorique et de développer une application concrète. Le projet *SIMAU*RIF est ainsi cité dans de nombreuses publications internationales comme étant « LE » projet de modélisation de l'interaction transport-urbanisation de Paris. L'autre projet en France qui utilise également UrbanSim est celui du LET à Lyon (projet SIMBAD).¹

Le projet *SIMAU*RIF est ambitieux car il vise à modéliser de manière quasi-complète les interactions entre système de transport et système urbain, et partant, à produire un véritable outil de micro-simulation utilisable à des fins de prédiction. Il est bien connu que la mise en service d'une infrastructure de transport procure un avantage immédiat aux usagers qui correspond essentiellement à des gains de temps. Les évaluations classiques du type coûts-avantages permettent de rendre compte de la mesure des effets directs d'une infrastructure. Mais on ne sait pas encore bien évaluer les effets indirects des infrastructures sur l'attractivité du territoire desservi, sur le développement économique local, sur les prix de l'immobilier, etc. *SIMAU*RIF permettrait de mesurer, en plus des effets directs, les conséquences indirectes de la mise en service d'une infrastructure de transport tels que la localisation de nouvelles entreprises, l'évolution de la demande de transport ou encore les valorisations foncière et immobilière qui ne sont en effet jamais prises en compte dans les évaluations socio-économiques. Or une plus-value bénéficie au particulier mais aussi à l'Etat grâce à la taxe fiscale incluse dans les frais de notaire lors de la mutation et grâce à la taxe foncière. Cependant, le projet *SIMAU*RIF n'a pas la prétention de révolutionner le calcul économique en intégrant les effets indirects dans la classique analyse coûts-avantages. Il propose par une approche territoriale et une utilisation de données désagrégées de simuler l'occupation des sols qui évolue en partie en fonction de l'offre de transport sur le moyen et le long termes.

¹ Mentionnons aussi trois autres modèles développés en France et subventionnés par le Predit qui permettent de simuler l'interaction transport-urbanisation mais sur le principe des systèmes multi-agents, et non des systèmes désagrégés et de choix de localisation comme *SIMAU*RIF : MIRO (Modélisation Intra-urbaine des Rythmes Quotidiens, développé par le Laboratoire d'Informatique de l'Université de Franche-Comté), ILOT (Interactions Localisations Transport, développé au LET) et Mobisim (développé par ATN).

La recherche avait été envisagée en deux phases, les résultats de la première phase conditionnant le lancement d'une seconde phase. Un rapport final de la première phase a été remis au Predit en septembre 2004 (voir [1]), qui a déterminé le lancement de la seconde phase. Pour la deuxième phase, un rapport intermédiaire a été remis en octobre 2005 (voir [2]). A partir de septembre 2004, nous avons accueilli pendant un an le concepteur du modèle UrbanSim, le Professeur Paul Waddell à l'IAURIF. Ce séjour a permis d'aborder des particularités du cas de l'Ile-de-France et d'étudier ensemble les adaptations nécessaires d'UrbanSim pour l'y appliquer. Cela a imposé des changements méthodologiques importants et a été une des causes qui ont donné lieu aux modifications apportées à UrbanSim version 4. Ensuite la recherche a dû être interrompue pendant un an pour raisons administratives. Elle a repris en septembre 2006 pour se terminer en décembre 2007. En tout, la recherche aura duré trois ans et demi, en mobilisant une équipe relativement restreinte pour un projet de cette envergure.

Cette application d'UrbanSim en Île-de-France s'est avérée plus complexe que prévue. Les difficultés viennent en premier lieu du fait qu'au début du projet, l'état des connaissances et des expériences de modélisation ne permettaient pas d'avoir un cadre bien défini pour la mise au point de ces modèles. Cela nous a obligés à développer des modèles pratiquement sur mesure dans chaque cas. C'est pour cette raison que la philosophie d'UrbanSim a évolué au fil des années. D'un « package » prêt à utiliser, le logiciel s'est transformé en une plateforme ouverte pour permettre la construction des modèles taillés sur mesure. Notre projet, un des premiers réalisés en dehors des Etats-Unis, a contribué à mettre en évidence les particularités de chaque aire d'étude en terme de structure urbaine, de comportements de mobilité ou de disponibilité des données. Le modélisateur doit tenir compte du fait que l'évolution d'une métropole telle que la région Ile-de-France est fonction de sa dimension historique et de sa densité importante qui sont sûrement différentes de celle d'une ville moyenne américaine ayant des constructions beaucoup plus récentes et une organisation plus simple.

Nous attirons ici également l'attention sur la différence de disponibilité des données et de découpage géographique entre les cas français et américain. Les règles d'accès aux données sont différentes, parfois plus strictes et contraignantes en France qu'aux Etats-Unis et cela nous a imposé de traiter les problèmes de manière différente. Nous avons été amenés au cours de ce projet à essayer plusieurs types de modèle afin d'en trouver un de qualité explicative suffisante compte tenu des données disponibles mais surtout de qualité prédictive. Cette démarche a été très « gourmande » en temps de travail.

Ce rapport final de la deuxième phase n'entend pas se substituer aux deux rapports précédents, disponibles en ligne sur le site du Predit dans la rubrique GO1. **La compréhension du présent rapport nécessite comme pré-requis la lecture des deux précédents.**

Rappelons le bilan à mi-parcours de la phase 2 (octobre 2005) :

- La quasi-totalité des tables principales a été construite dans la structure de données d'UrbanSim version 2 sur la base d'un maillage de la région en 50 000 cellules carrées de 500 m de côté, et en particulier les trois principales tables : gridcell, households, jobs.
- Deux modèles d'UrbanSim sur les quatre ont été estimés : le modèle de choix de localisation résidentielle des ménages et le modèle d'évolution du mode d'occupation du sol (MOS).
- Les modules de génération, distribution et choix modal du modèle global de l'IAURIF ont été recalés à partir de l'Enquête Globale Transports 2001-2002 dans le découpage en 606 zones de l'IAURIF (dont 80 zones pour Paris et 34 zones externes).
- Un prototype d'interface a été développé permettant de relier UrbanSim version 2 au modèle METROPOLIS et au modèle global de demande de déplacements.

Entre septembre 2006 et décembre 2007, nous avons poursuivi et terminé le projet en réalisant les travaux suivants, qui sont décrits dans le présent rapport :

- Portage des bases de données et des modules estimés depuis la version 2 vers la nouvelle plateforme OPUS/UrbanSim 4
- Construction de la base de données de calage et estimation du modèle de choix de localisation des emplois
- Construction de la base de données de calage et estimation du modèle de prix de l'immobilier
- Prise en compte de l'interaction transport-urbanisation dans la nouvelle plateforme OPUS/UrbanSim 4. Par manque de temps, nous n'avons pas réussi à adapter le prototype de l'interface informatique développé en 2005 à la nouvelle version d'UrbanSim basée sur langage Python. Le prototype de l'interface informatique développé en 2005 a donc été abandonné.
- Calage global et validation qui consistent en l'assemblage de tous les modèles et leur exécution enchaînée sur la période 1990-1999
- Préparation des données Ile-de-France pour des simulations sur la période 1999-2026
- Construction des scénarios de la Tangentielle nord, simulations et résultats
- Conclusions et perspectives

Un dernier message : cette recherche a été menée avec une démarche davantage pratique que théorique, l'objectif étant de parvenir à un outil utilisable en Ile-de-France, et non à produire encore un rapport de recherche bibliographique ou de réflexions théoriques. Ce rapport a donc l'ambition de servir de **guide d'utilisation** pour des praticiens développant un projet analogue à *SIMAU*RIF. Il est rédigé avec une terminologie technique propre à la modélisation, aux bases de données, au langage Python et au logiciel UrbanSim. Que le lecteur non initié en soit averti ...

2. Portage d'UrbanSim 2 à OPUS/UrbanSim 4

2.1. Historique

Depuis le début de la recherche en juin 2003, le logiciel UrbanSim a beaucoup évolué : il est passé de la version 2 à la version 4.1 aujourd'hui. Les principales nouveautés de la dernière version sont :

- une architecture plus ouverte, permettant une plus grande flexibilité dans la gestion des données et l'ajout de variables ainsi qu'une meilleure adaptation des données brutes disponibles. Par exemple, UrbanSim 4 offre la possibilité de travailler dans d'autres découpages que celui d'un maillage en cellules carrées, tels qu'un découpage à la commune ou à l'îlot.
- un nouvel algorithme pour le module de développement urbain. Ce n'est plus un modèle de transition comme dans la version 2, mais un modèle de choix de développement de projet, tel qu'il peut être perçu par un aménageur.
- de nouveaux outils d'estimation économétrique de modèles, les procédures étant quasi-automatiques.
- un nouveau langage de développement plus ouvert, Python, à la place de Java. Le code source d'UrbanSim 4 est plus accessible et plus facilement modifiable pour les utilisateurs, contrairement à celui d'UrbanSim 2 en Java dont la modification nécessitait une connaissance plus approfondie en informatique.
- des performances de calculs améliorées, permettant des gains de temps en simulation considérables, dues à l'utilisation des bibliothèques NumPy et SciPy associées à Python qui permettent de manipuler des matrices de grande taille de manière très efficace, alors que dans UrbanSim 2, le calcul de matrices était codé en Java.
- Son intégration dans une plateforme globale de simulation appelée OPUS (Open Platform for Urban Simulation). UrbanSim 4 est un des « packages » d'OPUS, qui comporte deux autres ensembles de « packages » : un modèle de simulation de la demande basée sur les activités et un modèle d'affectation (ils sont en cours d'intégration). OPUS est le résultat d'une collaboration internationale d'équipes de recherche (Université de Washington, Université de Toronto, Université Technique de Berlin, Institut Technologique de Zurich) travaillant sur la modélisation urbaine et la modélisation de trafic, et souhaitant mettre à disposition dans le monde du logiciel libre un ensemble d'outils de simulation prêts à utiliser.

2.2. Justification du portage

D'après les nombreux messages envoyés à la liste de diffusion d'UrbanSim, tous les utilisateurs dans le monde entier (et ils sont nombreux : Seattle, Detroit, Honolulu, Houston, Phoenix, San Francisco, Burlington, Durham, Melbourne, El Paso, Amsterdam, Tel Aviv, Zurich, Pescara, Lyon) se sont posé la question vers la fin de l'année 2006 : faut-il passer à la nouvelle version UrbanSim 4 ?

Pour notre part, les réticences étaient évidentes :

- Nous avons passé quasiment deux ans sur la version 2 développée en langage Java.
- Il nous faudrait estimer un nouveau modèle de développement urbain, alors que nous avons passé plusieurs mois à construire la base de calage d'évolution du Mode d'Occupation du Sol (MOS) et à estimer un modèle de transition.

- Il nous faudrait faire l'apprentissage du langage Python, ce qui ne semblait pas une mince affaire compte tenu de la pression des délais d'achèvement de notre projet.
- La version 4 en octobre 2006 était clairement une version bêta-test, en perpétuel développement, encore sujette à de nombreux « bugs » et dépourvue d'une documentation complète. De plus, la documentation (voir [5]) mélange les concepts de la version UrbanSim 2 et de la nouvelle version, sans que l'on sache clairement ce qui est nouveau.

Malgré ces inconvénients, nous avons décidé en octobre 2006 de passer à la nouvelle version, convaincu par les arguments suivants du Pr Paul Waddell :

- la version 4 prenait en compte les spécificités de l'Ile-de-France :
 - o Il était possible d'avoir des unités géographiques non uniformes tels que des communes ou des îlots.
 - o La disponibilité d'un modèle de choix emboîté (choix de la commune puis de la cellule ou de l'îlot). Ce modèle nous semblait le modèle le mieux adapté compte tenu, d'une part de la perte d'information lorsqu'on désagrège les données au niveau de cellule, d'autre part correspond au processus de décision du choix de localisation.
 - o La possibilité d'estimer un modèle de localisation des établissements et non des emplois
 - o Le modèle de développement urbain basé sur la localisation des projets était plus réaliste, mieux ajustable en fonction du rythme d'urbanisation. Etant donné le faible effectif des cellules ayant subi une évolution importante de leur occupation du sol, le modèle de transition était encore moins valable pour le cas parisien que pour les cas américains.
- il nous a clairement affirmé que la version précédente en Java ne serait plus maintenue, ni même téléchargeable sur le site UrbanSim.org. L'utilisation de la nouvelle version devenait par conséquent obligatoire.
- il nous a également déclaré que certains algorithmes d'UrbanSim 2 souffraient de problèmes dits de « dispersion », et que ces problèmes de fond avaient été résolus dans UrbanSim 4.
- les bases de données construites dans le format d'UrbanSim 2 n'auraient pas à être modifiées, ou de façon très mineure, pour fonctionner dans UrbanSim 4.
- on y gagnerait en temps de développement pour le reste des modules à estimer car UrbanSim 4 offre des outils d'estimation automatique de logit et de régression linéaire. Il serait inutile d'utiliser un logiciel externe de type SAS qui demande des mois de travail de calage. En restant dans UrbanSim 4, les calages de modèle peuvent être effectués en un jour, toujours selon le Pr Paul Waddell.
- on y gagnerait en temps de calculs sur un ordinateur standard.

L'affaire se présentait plutôt bien.

2.3. Problèmes rencontrés

Il nous semble utile d'exposer les difficultés rencontrées dans l'utilisation de la nouvelle version. Nous avons rencontré des problèmes importants de tous ordres, et pas seulement d'ordre informatique. Il semble que depuis le début, presque tous les utilisateurs ont subi les mêmes déboires que nous, si l'on en croit les nombreux messages reçus sur la liste de diffusion. Les causes sont doubles : une version en bêta-test et une documentation toujours en retard par rapport aux nouveaux développements.

Les phénomènes impliqués dans l'évolution urbaine sont beaucoup plus variés et divers et par conséquent moins connus que ceux qui touchent au domaine des transports seul. A cela s'ajoutent les différences entre le cas des villes américaines et des villes européennes en termes de structure historique et de politiques d'aménagement. De même, les données disponibles étaient bien différentes dans le cas de l'Ile-de-France par rapport aux applications précédentes d'UrbanSim aux Etats-Unis. Tenant compte de ces difficultés, UrbanSim et dans un cadre plus général, la plateforme OPUS, s'est donné comme vocation d'être plutôt une boîte à outils pour développer des modèles convenant à chaque cas métropole. Ceci nous a imposé de revoir une partie importante des méthodes déjà implémentées dans UrbanSim version 2.

2.3.1. Problèmes informatiques

L'installation d'OPUS/UrbanSim (version Windows²) et de l'exemple Eugene demandent déjà une persévérance à toute épreuve. Disons qu'une fois cette épreuve initiatique surmontée, qui peut prendre plusieurs jours suite à des problèmes de compatibilité ou de conflits entre les versions de Python/NumPy/SciPy/MySQL/MySQL Control Center/Windows et des problèmes de définitions des variables d'environnement de Windows (PYTHONPATH, etc), on a enfin pu faire tourner l'exemple Eugène.

Un **premier problème** a été le processus de déboguage. Le moteur de simulation d'UrbanSim 4 est conçu de façon à créer un nouveau processus pour chaque année simulée. Le but de cette manœuvre est de contourner un « bug » de gestion de la mémoire, présent dans la version 2.4 de Python, en forçant la libération de l'ensemble des ressources. Cette manœuvre met en déroute la plupart des débogueurs, rendant encore plus ardu l'examen détaillé du code. Lorsqu'un processus rencontre une erreur, c'est l'ensemble des processus qui s'arrête, avec autant de messages d'erreurs différents. Il est alors difficile d'identifier la source de l'erreur. C'est en échangeant de nombreux emails avec Hana Sevcikova, de l'Université de Washington, que nous avons pu avancer petit à petit mais ce fut laborieux. Nous avons eu aussi la chance, par rapport à d'autres utilisateurs dans le monde, de recevoir la visite d'Hana Sevcikova à l'Université de Cergy-Pontoise au cours du mois de mai 2007, qui nous a grandement aidés dans le déboguage de nos fichiers.

² Car il existe aussi la version sous Linux. Il semble qu'elle présente encore plus de « bugs » que la version Windows. Témoin cette réponse d'un développeur d'UrbanSim reçu le 9 août 2007 sur la liste de diffusion, suite à une question d'un utilisateur à propos d'un crash lors de l'installation de la version Linux d'OPUS/UrbanSim :

« I'm guessing this is a matplotlib problem. To see if this is the case, try running just the one unit test and see whether the same thing happens. To do this connect to the directory opus_core/indicator_framework/image_types and execute python matplotlib_chart.py. If you get a crash then that's a good sign that the problem is with that combination of the version of matplotlib, Python 2.4, and a 64 bit machine. We have mostly been working on 32 bit machines, but we do have one 64 bit linux machine. So I just tried running this on the 64 bit machine, using Python 2.5 (which comes with Fedora 7), and matplotlib 0.90.0, and it works OK. So you could try this again but with Python 2.5 and matplotlib 0.90.0. There were also a couple of problems that turned up with trying to run with Python 2.5 that we fixed since the stable release, so get the latest code out of the subversion repository as well. However, I just tried running all_tests on the 64 bit machine, and there are some other unit tests that fail because it is returning 64 bit instead of 32 bit integers. (The computation is actually OK - it's just that the tests are checking for 32 bit ints.) We will try to get these all fixed, but as far as I can tell, they are just a problem with the unit tests needing to be generalized to allow for both 32 and 64 bit machines, not actually indicating a bug in the rest of the code. » Sans commentaires ...

Le **deuxième problème** a concerné les formats utilisés et les conversions successives. Il y a le format de départ (Excel ou texte), puis la conversion en format MySQL, puis la conversion en format binaire flt requis par UrbanSim 4. Par exemple, l'exemple Eugène ne fournit les données qu'en format binaire, donc il est impossible de lire directement le contenu des tables. Si l'on veut les lire, il faut les convertir dans l'autre sens, c'est-à-dire de flt à MySQL, en utilisant le script « export_cache_to_mysql.py » disponible dans le répertoire python24/site_packages/paris/tools. Mais cette démarche n'est décrite nulle part dans la documentation, alors que c'est un passage obligé.

Le **troisième grand problème** a été l'apprentissage de Python. C'est un langage assez utilisé dans le milieu de la recherche, de par sa syntaxe claire, sa sémantique puissante et ses bibliothèques conséquentes. Il permet notamment d'utiliser divers paradigmes de programmation : programmation structurée, orientée objet, fonctionnelle (lambda calcul). Il autorise également quelques manipulations impossibles à réaliser d'ordinaire (hormis avec LISP et quelques autres langages marginaux), telles que l'interprétation (c'est-à-dire la compilation de code à la volée), le passage d'arguments de fonction par nom, l'utilisation d'une fonction en tant que valeur, etc... Mais c'est un langage incompréhensible à celui qui n'a pas une expérience préalable d'un langage tel que Java. Dans une équipe classique de recherche, on ne trouve pas de compétence spécifique en Python.

Le **quatrième grand problème** a été le contexte d'utilisation d'UrbanSim 4. Cette version qui est en plein développement depuis juillet 2006 à l'Université de Washington a connu des mises à jour très fréquentes entre janvier et juillet 2007. Le site UrbanSim.org proposait quasiment chaque semaine une nouvelle version à télécharger ainsi que la documentation. Certaines fonctions très attendues par les utilisateurs n'ont été disponibles qu'en avril, par exemple la variable dite « expression ». Une expression est une chaîne de caractères qui peut se substituer à une variable. Elle est définie dans les tables « spécifications ». C'est le système OPUS/UrbanSim 4 qui gère l'analyse syntaxique et en déduit s'il s'agit d'une variable existante ou d'une expression. Dans ce dernier cas, l'expression sera convertie automatiquement en variable anonyme, destinée à gérer les calculs. Avant la mise à disposition des « expressions », on a été amené à écrire des scripts spécifiques en Python qui ont permis de créer une classe pour chaque variable, ce qui nous a pris là aussi beaucoup de temps.

2.3.2. Problèmes de portage de la structure des données

Sur le site UrbanSim.org, il existe un document censé décrire les étapes nécessaires pour migrer d'une base de données développée sous UrbanSim 2 ou 3 à une base de données sous UrbanSim 4 : « Migrating from UrbanSim 3 to UrbanSim 4 ».

Or ce document ne décrit nullement les étapes à suivre mais contient une simple compilation de commentaires généraux et d'explications spécifiques à quelques tables. Par exemple, on peut y lire que la table « annual_employment_controls_totals » n'utilise plus la variable « total_employment », sans plus d'explications. Nous avons alors comparé méticuleusement les structures de données d'UrbanSim 3 et d'UrbanSim 4. Contrairement à ce que nous avait évoqué le Pr Waddell, les structures diffèrent significativement : des tables ont disparu, d'autres sont fusionnées, de nouvelles tables apparaissent. A l'intérieur des tables, de nouvelles variables apparaissent, d'autres disparaissent. En pratique, nous avons donc dû procéder de la manière suivante : nous avons strictement copié la structure de données de l'Ile-de-France sur celle de l'exemple Eugène, quitte à créer des fichiers vides. Mais même en procédant comme cela, nous avons rencontré des problèmes d'exécution sur nos données Ile-de-France. Il s'avérait qu'aucune table d'UrbanSim 4 ne devait être vide, sinon le programme s'arrêterait brutalement en laissant des

messages d'erreurs mystérieux. Nous avons donc dû dans nos premiers essais créer des tables comprenant des données fictives.

2.3.3. Problèmes de portage de la structure des modèles

La structure des modèles d'UrbanSim 4 a été complètement modifiée par rapport à la version 2, comme le montre la figure 1.

UrbanSim 4 comporte un ensemble de modèles prédéfinis, chacun étant implémenté comme étant un dérivé d'un modèle du « package » OPUS core (voir [5], chapitre 7).

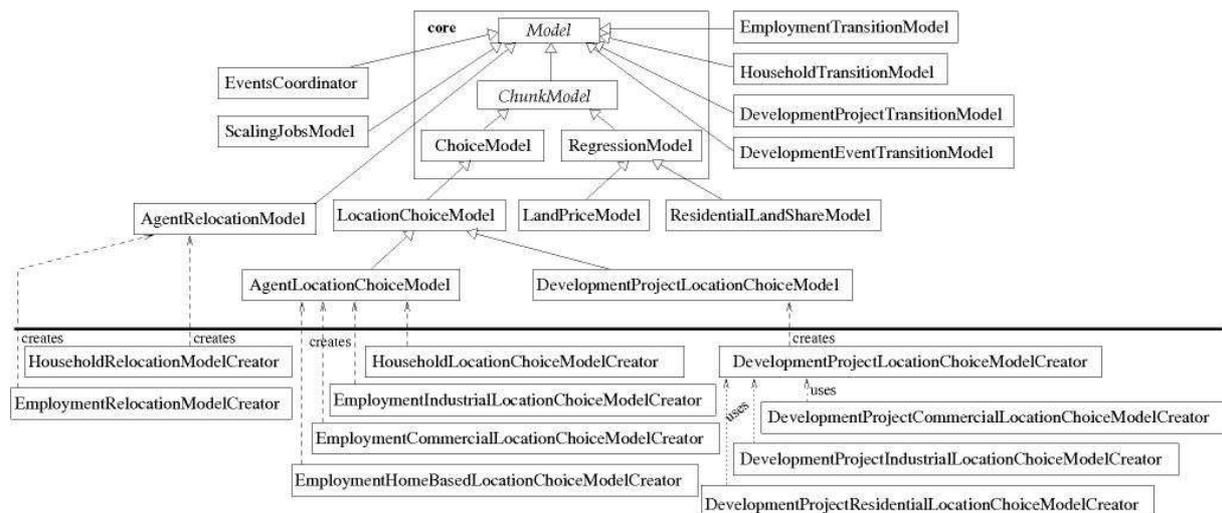


Figure 1 : Les modèles d'UrbanSim 4 (extrait de [5] page 101)

La philosophie d'UrbanSim 4 est de fournir un macro-langage permettant à l'utilisateur de programmer lui-même ses modèles basés soit sur une régression logistique soit sur une régression linéaire, toutes les deux étant implémentées dans le package « OPUS core ». A l'inverse, la version précédente proposait une boîte noire de modules pré-établis.

La grande nouveauté est le modèle de développement urbain qui est décomposé en trois sous-modèles : un pour les projets d'habitat, un pour les projets de commerces, un pour les projets industriels. En ce qui concerne les projets de type « governmental » et que nous avons remplacé par le type « bureaux » (voir [1] page 84), les concepteurs d'UrbanSim ont considéré que ces projets n'étaient pas localisables par modélisation car les choix d'implantation relèvent uniquement de la décision politique. Le sous-modèle correspondant n'est donc pas disponible. Or dans notre cas, il est indispensable car nous l'avons remplacé par la donnée de « bureaux ».

2.3.4. Problèmes d'estimation des modèles

Nous avons essayé d'utiliser l'outil d'estimation automatique des logits pour le modèle de localisation des emplois, car paraît-il, le travail aurait duré moins d'un jour. Ces outils se trouvent dans le répertoire /myworkspace/paris/estimation. Ils portent le nom « nommodèle_specification.py » (par exemple, « HLCM_specification.py »).

Malheureusement, nous ne sommes pas parvenus à utiliser cet outil. La documentation ne nous paraît pas suffisamment claire en ce qui concerne ce sujet. L'autre difficulté d'utilisation d'UrbanSim concerne la construction des variables. Finalement, il est plus avantageux de faire les tests avec le logiciel SAS et d'implémenter les modèles optimisés dans UrbanSim.

Quant au modèle de développement urbain, comme il l'a été dit, il a complètement changé entre la version 2 et la version 4. Si l'on avait dû estimer un nouveau modèle, cela aurait pris sans doute plusieurs mois. Le plus difficile n'est pas tant la phase d'estimation économétrique que la construction de la base de données de calage. Il faut recenser un maximum de « projets de développement » (les agents qui sont les polygones du MOS ayant évolué), en faire une typologie, les localiser dans le maillage et y attribuer la variation de nombre de logements ou d'espace d'activité. Le nouveau modèle de développement urbain se divise en deux modèles distincts. Le premier génère des « projets de développement », matérialisant les besoins de l'espace étudié en ce qui concerne les logements et les différentes activités. Il s'appuie pour cela sur l'existant, le disponible, ainsi que sur des valeurs de contrôle (taux de vacance structurelle des logements, par exemple). Les projets générés ne sont pas localisés. C'est le second modèle qui est chargé de les placer géographiquement, il se comporte en pratique comme un modèle de localisation classique.

En août 2007, nous étions contraints de ne pas utiliser ce nouveau modèle. Nous étions condamnés à adapter d'une manière sans doute très simplifiée le modèle d'évolution du mode d'occupation du sol (MOS) de la version 2 d'UrbanSim déjà estimé en 2005.

2.4. Conséquences sur la méthodologie du projet

Nous n'avions pas prévu en juin 2003 l'évolution du logiciel d'UrbanSim dans sa version actuelle. On peut regretter que la tâche de portage d'UrbanSim 2 (ou 3) à UrbanSim 4 ait été sous-estimée par l'équipe même de l'Université de Washington et que les utilisateurs ont été bercés d'illusions sur la facilité du travail. La ré-ingénierie du logiciel a en fait bouleversé la structure de données initiale, et l'apprentissage du nouveau langage Python a demandé un investissement non négligeable, sans compter les aléas d'utilisation d'une version en perpétuel développement.

Comme le lecteur a dû s'en rendre en compte en parcourant la partie 2.3, les problèmes informatiques nous ont quelque peu éloignés pour un temps de la modélisation de l'interaction transport-urbanisation appliquée au cas de l'Ile-de-France qui est le cœur même de notre projet. En pratique, partant au départ sur l'idée d'un simple assemblage des modèles déjà estimés, on en est arrivé à la fin à faire du développement informatique très spécialisé qui était hors programme. En conséquence, compte tenu du manque d'information exhaustive sur les espaces d'activité disponible par type d'activité, nous avons dû procéder à une simplification majeure : les espaces de trois types d'activités (commerces, industries et bureaux) ont été regroupés en une seule (par exemple pour les commerces nous ne disposions que de la surface de vente des grands enseignes). Ceci nous a obligés à reprendre la table gridcell et les deux modèles déjà estimés en 2005 (modèle de localisation des ménages et modèle d'évolution du MOS) :

- Dans la table gridcell, on a créé deux nouvelles variables :

$ACTIVITE_SQFT = COMMERCIAL_SQFT + INDUSTRIAL_SQFT + GOVERNMENTAL_SQFT$

$ACTIVITE_IMPROVEMENT_VALUE = COMMERCIAL_IMPROVEMENT_VALUE + INDUSTRIAL_IMPROVEMENT_VALUE + GOVERNMENTAL_IMPROVEMENT_VALUE$

- Les estimations des deux modèles en 2005 ont fait ressortir des variables concernant les activités. Par exemple, la variable `COMMERCIAL_IMPROVEMENT_VALUE` dans le modèle

d'évolution du MOS et la variable VCOM dans le modèle de localisation des ménages. Conséquence : il a fallu ré-estimer ces modèles en testant les variables ACTIVITE_SQFT et ACTIVITE_IMPROVEMENT_VALUE, ainsi que les deux variables de voisinage associées. Les résultats sont présentés dans les paragraphes 2.5 et 2.6.

Au final, le regroupement dès le départ de ces trois activités en une seule rend cohérent les modèles de localisation avec le modèle d'évolution du MOS qui, lui, ne distingue pas les types d'activités.

2.5. Adaptation du modèle de localisation des ménages à UrbanSim 4

L'estimation de ce modèle a été refaite en prenant en compte la nouvelle variable $vact_{99} = vcom_{99} + vind_{99} + vgov_{99}$ (ces trois variables interviennent dans le modèle estimé en 2005, voir [2] page 82).

Estimation du modèle de localisation des ménages			
Variable	Estimation	Ecart-type	p_value
emploiTC	0,1217	0,0118	<0,0001
vresu_mnact	8,67E-06	1,58E-06	<0,0001
vtrain_99	0,0595	0,0201	0,0031
eleve_rev_vresval	2,31E-07	3,93E-08	<0,0001
TVP_99	-0,7104	0,1817	<0,0001
vact_99	1,24E-06	9,36E-08	<0,0001
ZUS_faible_rev	3,97E-07	6,32E-08	<0,0001
dhigh_99	-0,0286	0,008654	0,0010
dart_99	-0,0675	0,0103	<0,0001

Tableau 1 : Résultats mis à jour du modèle de localisation résidentielle

On remarque que la variable $vsubway_{99}$ n'intervient plus. C'est ce modèle à 9 variables qui sera implémenté dans UrbanSim 4.

Nous rappelons que ce modèle est un modèle simple, estimé dans le but de pouvoir être implémenté et utilisé en prédiction dans SIMAURIF, au moins dans une première expérimentation. Les variables testées étaient *par défaut* celles proposées par UrbanSim, auxquelles on a ajouté quelques variables de transports en commun et quelques variables spécifiques à l'Ile-de-France telles que les zones urbaines sensibles (ZUS). Mais on est bien conscient que de nombreux facteurs ne sont pas pris en compte ici et dont on sait qu'ils prennent de plus en plus de poids dans le choix résidentiel. Ils concernent les facteurs de qualité de vie au quotidien : présence d'espaces verts, proximité d'écoles, présence de commerces de proximité, propreté de la ville.

En plus de ce modèle implémenté dans UrbanSim, une étude complémentaire et plus riche a été réalisée pour comprendre au mieux le phénomène de localisation des ménages. Elle est présentée en annexe 2. Si nous en avons les moyens, nous prévoyons d'intégrer dans le futur ces résultats dans UrbanSim afin d'améliorer nos prédictions. Elle est présentée en annexe 2.

2.6. Adaptation du modèle d'évolution du MOS à UrbanSim 4

Rappelons que le principe consistait à considérer la cellule comme un agent et à déduire l'évolution de son mode d'occupation du sol selon ses caractéristiques actuelles. On ramenait ainsi l'évolution du MOS (Mode d'Occupation du Sol) à un choix discret. Les différents types de MOS avaient été regroupés en 3 types généraux : le type vacant, le type habitat et le type activité. Les cellules non urbanisables (à cause des caractéristiques du terrain ou de la législation) n'étaient pas prises en compte par ce modèle.

Nous avons ré-estimé ce modèle en prenant en compte les nouvelles variables regroupant les trois activités (« commercial », « industrial », « governmental »).

Transition du type 'vacant' vers le type 'habitat'			
Variable	Estimation	Ecart-type	p_value
Constante	-0,1337	0,2175	0,5388
Residential_Units	0,1515	0,0065	<0,0001
V_Activity_sqft	6,93 ^E -06	1,80 ^E -06	<0,0001
TTC_99	-0,2795	0,0984	0,0045
TVP_99	-1,6012	0,2416	<0,0001

Tableau 2 : Les résultats de nouveau modèle (type vacant-> habitat)

R-square = 0.4418 Max-rescaled R-square=0.5454

Il n'y a plus que 5 variables explicatives contre 8 dans l'ancien modèle (voir [2] page 73).

Transition du type 'vacant' vers le type 'activités'			
Variable	Estimation	Ecart-type	p_value
Constante	-0,2896	0,2324	0,2129
Residential_Units	0,1005	0,0073	<0,0001
Activity_Improvement_value	0,000666	0,000232	0,0947
V_Activity_sqft	0,000014	1,67E-06	<0,004
TTC_99	-0,3974	0,121	0,001
TVP_99	-0,7955	0,2565	0,0019

Distance_to_highway	-0,1636	0,0168	<0,0001
Nb_nodes_train_99	1,8242	0,9652	0,058

Tableau 3 : Les résultats de nouveau modèle (type vacant-> activité)
R-square = 0.4418 Max-rescaled R-square=0.5454

Toutes les variables restent significatives au seuil de 10%. Les coefficients estimés ont les signes attendus..

3. Modèle de choix de localisation des emplois

La spécification d'un modèle dépend essentiellement de l'usage que l'on veut en faire. Dans notre cas, il doit être implémentable dans UrbanSim 4. Le modèle de localisation des emplois a donc été estimé sur le même principe que le modèle de localisation des ménages : un modèle simple, c'est-à-dire :

- comportant au plus une douzaine de variables explicatives
- implémentable directement dans UrbanSim
- simplifié par rapport à un modèle dit « explicatif » qui serait intéressant pour expliquer la mobilité des entreprises ou des emplois mais ne serait pas applicable en prédiction car il comporterait trop de variables dont certaines non projetables.
- décliné par type d'emplois (a priori 8 types)

Rappelons aussi que l'agent économique dans ce modèle est l'emploi et non l'entreprise. Or c'est sans doute l'entreprise qui est l'entité la plus pertinente car il existe de nombreuses enquêtes sur les critères de choix de localisation auprès des entreprises (voir partie 3.3). Un modèle de localisation des entreprises serait plus à même de rendre compte de la concentration des emplois et fournirait donc de meilleures prédictions.

Nous avons donc réfléchi à un modèle de localisation des entreprises, qui est présenté brièvement en annexe 3. Mais cette réflexion dépasse le cadre du projet *SIMAU*RIF qui consiste avant tout à appliquer un outil existant à l'Ile-de-France. Toutefois, un modèle de localisation des établissements tel que présenté en annexe 3 serait plus à même de rendre compte de la concentration des emplois et donc de fournir de meilleurs prédictions. Dans cette annexe 3, on trouvera aussi les résultats d'un autre modèle de localisation des emplois à visée plus explicative que prédictive.

Avant de décrire la construction de la base de données de calage, les variables utilisées et les résultats du modèle, nous commençons par rappeler quelques données fondamentales sur l'emploi en Ile-de-France.

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

Entre l'Enquête Régionale sur l'Emploi (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 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. En revanche, entre l'ERE de 1994 et de 2001, l'emploi salarié a augmenté de nouveau, surtout à partir de 1997.

Selon le Recensement, sur la période 1990-1999, la Région a perdu 34 000 emplois, Paris à lui seul en a perdu 215 000 et la banlieue en a gagné environ 180 000. Les activités ont continué à se desserrer géographiquement vers l'ouest parisien et la grande couronne. C'est une tendance historique. En grande couronne, ce sont les Villes nouvelles qui enregistrent les hausses d'emplois importantes (surtout Marne-la-Vallée grâce à Disneyland Paris, et Saint-Quentin-en-Yvelines grâce à l'implantation du Technocentre de Renault et du siège social de Bouygues), puis les zones autour des aéroports de Roissy et d'Orly, ainsi que les communes situées le long

de l'axe sud (Massy, Palaiseau, Saclay, les Ulis ...). On observe un développement très marqué d'un « croissant sud » le long de la Francilienne entre les Villes nouvelles de Sénart et de Saint-Quentin-en-Yvelines qui englobe le plateau de Saclay. Seul le Val d'Oise connaît une stagnation entre 1990 et 1999. 75% de l'emploi salarié de la grande couronne est concentré sur 11% des communes.

Si on regarde par secteur, on peut noter les éléments marquants suivants :

- Les créations d'emplois ont eu lieu seulement dans le secteur tertiaire. Mais Paris en a perdu beaucoup (125 000).
- 71 000 emplois supplémentaires ont été créés dans le secteur Services aux entreprises (emplois dits « péri-productifs ») entre 1990 et 1999.
- Perte de 308 000 emplois dans l'industrie et la construction
- 72 000 emplois supplémentaires dans l'administration
- Perte de 25 000 emplois dans le commerce

La géographie des emplois montre également des images contrastées selon les secteurs d'activités. Il est donc nécessaire dans la modélisation de distinguer des secteurs d'activités.

Enfin, pour faire le lien avec l'interaction transport-urbanisation, comment s'explique le desserrement des emplois du centre vers la périphérie ?

- le développement des réseaux de transport permet d'atteindre des destinations plus lointaines pour des temps de déplacements identiques. Avec l'apparition de rocade routières (A86 et Francilienne), le desserrement ne se coule plus seulement le long des axes radiaux mais aussi le long des axes en rocades.
- Le desserrement de l'habitat favorise aussi le desserrement des emplois. En effet, les migrations résidentielles ont entraîné dans leur sillage des activités économiques, commerciales, de services aux particuliers et des emplois publics (collectivités territoriales, services de l'Etat, fonction publique hospitalière, éducation). Mais l'émiettement spatial des emplois n'a pas atteint celui de l'habitat. L'emploi reste polarisé et non dispersé.

Les spécialistes de l'emploi à l'IAURIF affirment qu'il est quasiment impossible de modéliser finement les choix de localisation des emplois. Les très nombreux travaux de prospective sur l'emploi ne font jamais appel à de la modélisation mais à des analyses seulement descriptives et qualitatives, et, partant, à des propositions de scénarios du futur, sans qu'aucune ne sorte du lot dans un contexte teinté d'incertitude. Avant même de commencer l'effort de modélisation, nous pensions qu'ils avaient en partie raison (on aurait pu écrire : « nous allons prouver le contraire car nous avons une confiance indéfectible en la puissance de la modélisation »). Ce serait tromper le lecteur que d'affirmer le contraire lorsque l'on connaît les limites de la modélisation. Les phénomènes sont très complexes, très fortement dépendants du type d'activité, de la conjoncture économique et de la politique gouvernementale. De nombreuses inconnues demeurent quant au devenir de pôles d'emplois régionaux tels que La Défense. En effet, le Conseil régional d'Ile-de-France actuel préconise de maintenir le volume actuel de surface de bureaux tandis que l'Etat et le Conseil général des Hauts-de-Seine envisagent une forte augmentation du parc de bureaux. A court terme (2013), il est prévu 450 000 m² de bureaux supplémentaires, 100 000 m² de locaux pour le MEDAD et 100 000 m² affectés à l'habitat (1400 logements). En revanche, à plus long terme, c'est l'incertitude. De plus, les stratégies de localisation peuvent varier dans le temps et de nombreux facteurs sont tout à fait imprévisibles. Il suffit de se projeter 25 ans en arrière. Qui aurait pu imaginer l'impact des nouvelles technologies sur les mutations du travail (augmentation de la productivité, apparition de nouveaux secteurs d'activités, disparition d'autres secteurs) et sur les délocalisations.

La description du travail qui est faite dans la partie 3.2 a pour but de montrer que notre modélisation utilise des données très riches et inédites, mais que l'objectif est d'aboutir à un modèle simple, validé en moyenne et utilisable dans UrbanSim, qui soit capable au moins de reconstituer les grandes tendances observées tel que le desserrement des emplois et mettre en lumière quelques variables transports comme critères de choix de localisation.

3.2. Construction de la base de données de calage

3.2.1. La source : l'Enquête Régionale sur l'Emploi

La base a été construite à partir de l'Enquête Régionale sur l'Emploi (ERE). Les deux dernières ERE ont été réalisées en 1998 et 2001. Le choix de cette base est justifié par le fait que la base de 2001 est géolocalisée sur le SIGR. Il existe d'autres bases de données sur les emplois (par exemple BIL, ASTREE, SIRENE) mais non géolocalisées.

Le fichier alpha-numérique ERE 2001 donne la liste des établissements présents en Ile-de-France, en indiquant l'activité principale de cet établissement (code APET en 712 postes, regroupables en 16 postes puis en 8 postes, voir tableau 4), son numéro de SIRET, son adresse et le nombre de salariés (y compris les entreprises mono-salariées). Le travail de géolocalisation a été réalisé en 2002. Un important travail de relocalisation de certains effectifs parisiens avait été effectué, par recoupement 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). Au final, la couche ERE2001 du SIGR est une couche de points comportant 187 415 entreprises. Pour retrouver le numéro de la cellule dans le maillage g1 en 50 000 cellules à partir des coordonnées (X,Y) dans le système Lambert II, on applique la formule suivante :

$$g1-id = \text{Ent}[(X-534\ 000)/500] + 312 * \text{Ent}[(Y-46000)/500]$$

(on aurait pu aussi procéder en croisant la couche ERE 2001 et la couche du maillage en 50000 cellules)

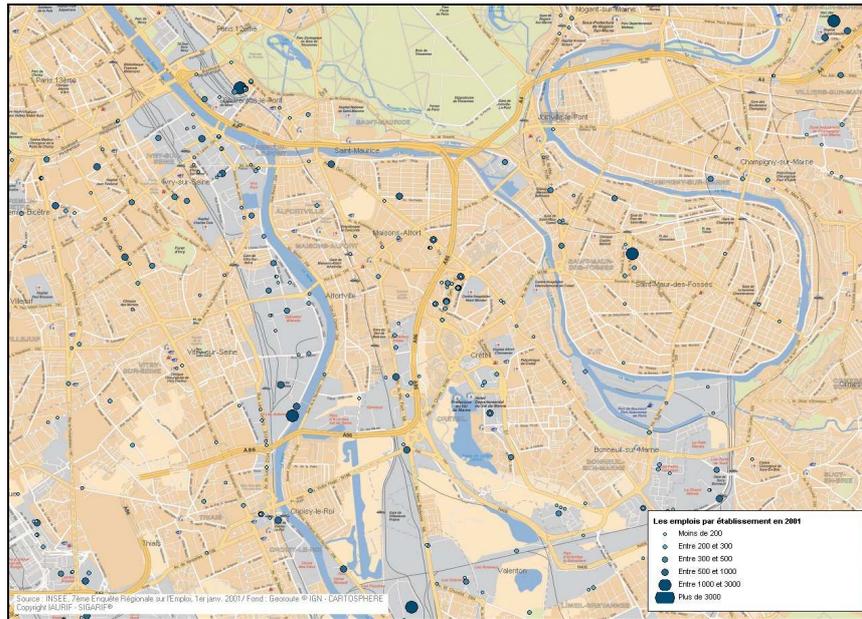


Figure 2 : Extrait de la couche ERE 2001 géolocalisée

Il est utile pour la suite d'expliquer la différence entre les identifiants SIREN et SIRET d'une part, et entre une entreprise et un établissement d'autre part. Les personnes physiques ou morales qui constituent les entreprises se voient attribuer un identifiant de 9 chiffres, le SIREN, qu'elles conservent pendant toute leur vie pour les entrepreneurs individuels ou pendant toute leur durée de vie pour les sociétés, même si leur activité change de nature ou d'endroit. Chacun des établissements où s'exerce l'activité de l'entreprise reçoit un numéro SIRET qui reprend les 9 chiffres du SIREN, complétés de 5 autres. Un SIRET caractérise ainsi un lieu d'activité et l'entreprise à laquelle il appartient : si l'établissement passe sous le contrôle d'une autre entreprise, il recevra un SIRET complètement nouveau, y compris les 9 premiers chiffres qui identifient l'entreprise dont il dépend dorénavant. Si l'activité de l'établissement change à l'intérieur de l'entreprise, son SIRET ne changera pas. Si l'entreprise déménage son établissement, celui qui sera installé à la nouvelle adresse aura un nouveau SIRET. La partie SIREN de ce nouveau SIRET sera identique au SIRET précédent, que le nouvel établissement ait ou non la même activité.

Rappelons enfin que le modèle de choix de localisation des emplois d'UrbanSim considère l'emploi et non l'entreprise comme agent économique. La base de calage est donc constituée d'observations qui sont des emplois *ayant été créés ou ayant déménagé* entre 1998 et 2001. La représentativité de l'échantillon sera basée sur deux critères : la répartition par activité et la répartition par département.

3.2.2. Étapes de construction de la base de calage

La base a été construite en 10 étapes. Il nous semble utile de détailler ces étapes de manière complètement transparente, pour montrer que ce projet se base sur des données riches et inédites, que la qualité des résultats d'un modèle dépend avant tout de la qualité des données d'entrée, et parce que cette phase accapare une large part du temps de travail.

On travaille d'abord sur les entreprises et établissements. Ce n'est qu'à la fin que l'on passera aux emplois.

1) Fusion des fichiers alpha-numériques ERE de 1998 et de 2001

Cette fusion se fait en utilisant comme clé le SIRET (14 chiffres).

Création de 2 attributs booléens : « 1998 » et « 2001 » (0 : n'existe pas cette année là, 1 : existe)

Création d'un attribut booléen : « Entreprise » (0 : établissement, 1 : entreprise)

Sur les 401 838 entreprises et établissements, 154 786 ont leurs attributs « 1998 » et « 2001 » à 1. Il reste donc 247 052 observations (dont 30 sont mis au rebut car le champ « nom d'entreprise » n'est pas renseigné).

2) Suppression des entreprises ayant disparu

83 756 observations ont pour valeurs : « 1998 » = 1, « 2001 » = 0, « entreprise » = 1. Ce sont soit des entreprises ayant déposé le bilan, soit des entreprises ayant déménagé (il y a alors deux observations dans la base, la première est à supprimer).

Par exemple, voici un cas de déménagement :

N°SIRET	Nom société	1998	2001	Entreprise
586 478 921 26945	EXEMPLE1	1	0	1
586 478 921 12693	EXEMPLE1	0	1	1

Il reste 163 266 observations (247052 – 83756).

3) Séparation entre entreprises ayant été créées ou déménagées et autres cas (création d'établissements, fermeture d'établissements)

89433 observations ont pour valeurs : « 1998 » = 0, « 2001 » = 1, « entreprise » = 1. Ce sont soit des entreprises nouvellement créées, soit des entreprises ayant déménagé.

Il reste 73833 observations qui concernent les établissements.

4) Suppression des établissements ayant disparu

33942 observations ont pour valeurs : « 1998 » = 1 et « 2001 » = 0

Il reste 39891 établissements (73833-33942) ayant été créés ou ayant déménagé.

5) Croisement du fichier des établissements avec la base de données géolocalisées ERE 2001

Sur les 39891 observations, 26506 sont géolocalisées. Les effectifs vont de 1 à 10699 salariés, totalisant 986 572 emplois.

6) Filtrage sur l'effectif

Nous décidons de ne retenir que les établissements ayant un effectif compris entre 10 et 999 salariés. En dessous de 10 salariés, on considère que le choix de localisation ne répond pas à des critères modélisables. Au-delà de 1000, on se rapproche d'établissements et de sièges sociaux d'un grand groupe. Il nous a semblé préférable pour ces cas d'utiliser la table « employment_events » d'UrbanSim (voir [1], page 98) qui permet d'enregistrer des déménagements de grands établissements. D'une part, il est évident que ces déménagements ne sont pas prévisibles par simulation dans un découpage cellulaire, d'autre part, on en a recensés moins de 20 entre 1990 et 1999, donc la table « employment_events » n'est pas difficile à remplir manuellement. Par exemple, le regroupement à Guyancourt en 1998 de près de 8000 emplois du constructeur Renault.

En restreignant donc l'effectif à la tranche 10-999, il reste 11853 établissements, soit 679324 emplois.

7) Croisement du fichier des entreprises avec la base de données géolocalisées ERE 2001

Sur les 89433 entreprises, 57221 sont géolocalisées. Si c'est une entreprise créée entre 1998 et 2001, elle a un effectif encore faible. Si c'est une entreprise (donc sans établissement, donc une petite ou moyenne entreprise) ayant déménagé, elle a sans doute un effectif inférieur à 1000.

Ainsi, sur les 57221 entreprises, 46356 ont entre 1 et 9 salariés, et 10763 ont entre 10 et 999 salariés, totalisant 580524 emplois.

8) Regroupement des fichiers entreprises et établissements

Le regroupement donne lieu à une base de 1 259 848 emplois (679324+580524) ayant été créés ou ayant déménagé entre 1998 et 2001.

9) Extraction d'un échantillon représentatif

Notre base n'est pas nécessairement représentative par construction. Nous devons en extraire un échantillon représentatif selon deux critères : le type d'emploi en 8 secteurs et la répartition géographique par département.

Nous avons défini notre propre typologie d'emploi en 8 secteurs à partir de la nomenclature classique de l'INSEE en 16 postes. Ci-dessous le tableau de correspondance :

INSEE	libellé	8 postes
1	agriculture	1
2	industries agricoles	
3	industrie biens consommation	2
4	industrie automobile	
5	industrie biens équipement	
6	industrie biens intermédiaires	
7	énergie	3
8	construction	
10	transports	
9	commerce	4
11	activités financières	5
12	activités immobilières	
13	service aux entreprises	6
14	service aux particuliers	
15	éduc, santé, action sociale	7
16	administration	8

Tableau 4 : Définition des 8 secteurs

La base source, l'ERE alphanumérique de 2001, comporte 2 181 769 emplois. On en tire une matrice A en deux dimensions : en 8 secteurs et 8 départements.

De la base construite géolocalisée, on en tire une matrice B en deux dimensions : en 8 secteurs et 8 départements.

On extrait alors de la matrice B une matrice B' proportionnelle à la matrice A en utilisant comme pivot la plus petite valeur de la matrice B. Au final, la matrice B' comporte **464 519 emplois**.

Cette méthode permet de générer une matrice de coefficients correcteurs 8*8 qui sont appliqués en calculant pour chaque observation un poids. Ce poids est égal à l'effectif total de l'entreprise ou établissement multiplié par le coefficient (secteur, département).

Un programme écrit en SAS utilisant la procédure SelectSurvey permet enfin de générer un fichier d'emplois (et non plus d'entreprises) en utilisant la variable poids.

Département	Secteur								Total
	1 agric	2 industrie	3 énergie	4 commerce	5 finances	6 services	7 éducation	8 administr	
75	839	8 886	7 079	12 159	18 485	56 817	17 859	13 657	135 781
77	527	4 091	3 706	4 597	1 066	6 499	3 611	5 293	29 390
78	414	12 996	3 833	4 516	1 789	10 051	4 517	7 069	45 185
91	356	5 085	4 512	7 435	731	11 052	3 883	6 701	39 755
92	1 072	13 156	6 811	8 863	8 310	40 520	6 474	7 643	92 849
93	245	5 229	6 834	5 388	1 719	8 843	5 025	9 455	42 738
94	302	2 797	4 521	5 547	2 031	9 516	7 574	7 365	39 653
95	288	3 027	3 017	15 938	1 188	6 421	4 014	5 275	39 168
Total	4 043	55 267	40 313	64 443	35 319	149 719	52 957	62 458	464 519

Tableau 5 : Répartition de l'échantillon d'emplois en 8 secteurs et en 8 départements (ERE 2001)

A titre de comparaison, la base de calage de l'étude de référence Puget Sound réalisée par le Pr Waddell comportait 90 000 emplois, soit 5 fois moins que notre base.

10) Tirage des alternatives

On complète la base en tirant 5 cellules alternatives pour chaque secteur, sauf le secteur 6 (deux alternatives seulement), à cause de la limitation à environ 450 000 lignes de la table lue par la procédure MDC (logit multinomial) de SAS. La méthode de tirage aléatoire simple a été utilisée. Pour chaque observation, on fait un tirage aléatoire équiprobable de 5 alternatives parmi les 9000 cellules de type 'habitat' ou de type 'activité' 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 des emplois.

3.3. Critères de choix de localisation des entreprises

De nombreuses études ont essayé de mieux connaître les attentes et les processus de décision des entreprises en matière de localisation de leurs bureaux. Ces études se basent en général sur des enquêtes auprès des établissements. Citons l'étude du BIPE réalisée pour le compte de la DREIF en 2002 sur les choix de localisation des bureaux des entreprises de plus de 300 emplois. Les critères qui ressortent sont les suivants :

- La desserte par les réseaux de transports reste le critère primordial dans les choix de localisation de bureaux, les transports en commun étant jugés plus importants que la desserte routière.
- D'autres critères sont plus secondaires : qualité du cadre de vie, sécurité des biens et des personnes, proximité d'établissements d'enseignement secondaire et supérieur
- Les 2/3 des établissements recherchent un quartier périphérique avec des densités moyennes à élevées mais ne souhaitent pas localiser leurs bureaux dans les centres-villes qui présentent un caractère trop contraignant.
- Plus de la moitié des établissements recherchent des quartiers mêlant habitat et activités.
- Les grandes entreprises préfèrent implanter leurs bureaux dans de petits ensembles immobiliers (moins de 10000 m²) plutôt que dans de grands centres d'affaires.

Du point de vue des opérateurs immobiliers, quelques critères généraux ressortent de manière unanime :

- du point de vue des opérateurs de bureaux, quelque soit la taille de l'entreprise et le secteur d'activité, le critère de la desserte est primordial dans les choix d'implantation d'une opération immobilière. Pour les opérations de bureaux, la présence de transports en commun de type métro ou RER doit permettre l'accès. A l'inverse, pour les opérations de locaux d'activités, la desserte routière prime sur la desserte par les TC, le critère étant celui d'une desserte immédiate, c'est-à-dire un accès direct par échangeur à une voie rapide. Pour les opérateurs de logement, le critère de la desserte retenu est celui de la distance aux zones d'emplois par les TC exprimée en temps.

- Au-delà de la desserte, c'est la proximité des zones d'emplois qui est recherchée dans les choix d'implantation d'opérations.
- Les professionnels sont attentifs au choix de la collectivité locale qui sera leur partenaire dans l'élaboration d'une opération.

Du point de vue des entreprises du tertiaire, en ce qui concerne la demande de bureaux, il apparaît qu'une part majoritaire de cette demande provient des grandes entreprises. Ces dernières années, sur le marché immobilier de bureaux, les mouvements se sont concentrés principalement sur le Quartier Central des affaires et le fameux Triangle d'or « Paris Centre Ouest, croissant d'or Neuilly/Boulogne/Issy, la Défense ». Des tentatives s'amorcent au nord, notamment dans la Plaine Saint-Denis, ainsi qu'au sud dans le pôle Orly-Rungis. Enfin une couronne tertiaire se profile le long du périphérique sud à la limite extérieure de Paris. De manière générale, le marché de bureaux est très fortement soumis à des effets de concentration (on parle de concentration au-dessus de 100 000 m² de bureaux calculés sur la base d'un ratio de 20 m²/emplois de bureaux). De tels effets de concentration, peuvent être obtenus en prédiction dans un modèle de localisation des établissements tel que celui que nous envisageons d'implémenter un jour, mais pas dans un modèle de localisation des emplois tel que celui implémenté actuellement dans UrbanSim.

Les grandes entreprises s'intéressent à la desserte par les transports en commun. Les liaisons avec les aéroports constituent également un critère de desserte important. Enfin, un critère est de plus en plus mis en avant : celui de l'image du territoire.

Enfin, les choix de localisation suivent des cycles liés au niveau des loyers. Les mouvements de décentrage suivent les hausses de loyers et inversement, il y a retour vers le centre lorsqu'il y a diminution des loyers. Les grandes entreprises convoitent aussi des emplacements dans les Villes nouvelles car les loyers sont peu élevés.

3.4. Variables de la base de calage

C'est la même base que celle utilisée dans le modèle de localisation des ménages, sauf que les variables d'activités ont été additionnées et que les variables d'accessibilités aux commerces n'ont pas été retenues. Ainsi, on voit bien que les variables testées correspondent aux critères de choix exprimés lors des enquêtes auprès des entreprises souhaitant déménager ou auprès des promoteurs.

On peut classer les variables en 3 catégories :

1) Les variables disponibles dans gridcells :

- des variables qui donnent la composition d'une cellule (en 1999) :
 - residential_units = nombre de logements
 - activites_sqft = surfaces d'activités (commerciales, industrielles, bureaux)
- des variables qui donnent les conditions du marché d'une cellule (en 1999) :
 - residential_improvement_value = valeur foncière du bâti résidentiel
 - residential_land_value = valeur foncière du terrain

- `activites_improvement_value` = loyers des surfaces d'activités³
- 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 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
- une variable de politique de la ville : appartenance à une ZUS/ZRU/ZFU.

2) Les variables de voisinage pour l'année 1999

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
- '`activites_sqft`' = surfaces d'activités
- '`residential_units`' = nombre de logements
- '`activites_improvement_value`' = loyers des surfaces d'activités
- '`nb_nodes_subway`' = nombre de stations de métro
- '`nb_nodes_train`' = nombre de gares
- '`nb_nodes_tramway`' = nombre d'arrêts de tramway.

3) Les variables d'accessibilité

- 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

Au total, la base de calage comporte **24 variables**.

	Nom de la variable	Min	Max	Unités
1	<code>resu_99</code>	0	65 464	Nombre
2	<code>act_99</code>	0	1 028 139	Nombre
3	<code>resval_99</code>	0	1 969 467	Milliers €
4	<code>land_99</code>	0	198 753	Nombre

³ En finalisant ce rapport, nous avons pris connaissance de nouvelles données concernant les loyers et les prix de bureaux issues de la côte Callon que nous exploiterons dans les développements futurs du modèle.

5	actval_99	0	301 540	Milliers €
6	dart_99	0	18,8	Km
7	dhigh_99	0	24,9	Km
8	d_châtelet	0	92,6	Km
9	subway_99	0	5	Nombre
10	train_99	0	4	Nombre
11	tram_99	0	2	Nombre
12	TTC_99	0,1	2,0	heures
13	TVP_99	0,1	2,0	heures
14	ZUS	0	1	binaire
15	vresu_99	0	65 464	Nombre
16	vact_99	0	4 548 886	Nombre
17	vresval_99	0	9 394 736	Nombre
18	vland_99	0	988 790	Nombre
19	vactval_99	0	1 501 365	Milliers €
20	vsubway_99	0	21	Nombre
21	vtrain_99	0	8	Nombre
22	vtram_99	0	6	Nombre
23	emploiTC	30,0	52,7	Nombre
24	emploiVP	40,4	54,8	Nombre

Tableau 6 : Variables testées du modèle de localisation des emplois

Les résultats de l'estimation du modèle de choix de localisation des emplois par secteur (variables significatives, coefficients, tests de validité) sont présentés dans la partie suivante.

3.5. Résultats du modèle

Plusieurs dizaines de combinaisons de variables ont été testées.

Les huit modèles suivants, un par secteur, ont été retenus.

Secteur	Variable	Paramètre estimé	Sum of tValue	p_value
1 (agricul.)	1. McFadden	43%		
	2. N of Obs	4043		
	act_99	0,000011	12,13	< 0,001
	actval_99	-0,000028	-8,26	< 0,001
	d_chatelet	0,038632	4,51	< 0,001
	dart_99	-0,045764	-3,44	0,001
	dhigh_99	0,038776	3,12	0,002
	emploiTC	-0,096903	-5,34	< 0,001
	TC_99	-0,432891	-5,23	< 0,001
	TV_99	-5,293186	-10,62	< 0,001
	vact_99	0,000002	13,74	< 0,001
	vtrain_99	0,376822	18,24	< 0,001
	zus	0,845512	6,27	< 0,001

Secteur	Variable	Paramètre estimé	Sum of tValue	p_value
2 (industrie)	1. McFadden	40%		
	2. N of Obs	55267		
	act_99	0,000020	88,67	< 0,001
	actval_99	-0,000046	-53,71	< 0,001
	d_chatelet	0,051528	24,81	< 0,001
	dart_99	-0,033978	-10,00	< 0,001
	dhigh_99	-0,013657	-4,03	< 0,001
	emploiTC	0,113832	25,31	< 0,001
	emploiVP	0,362753	21,08	< 0,001
	subway_99	0,132600	8,69	< 0,001
	TC_99	0,209785	18,04	< 0,001
	vact_99	0,000001	16,81	< 0,001
	vactval_99	0,000002	15,05	< 0,001
	vresu_99	0,000025	30,65	< 0,001
zus	-0,110939	-2,74	0,006	

Secteur	Variable	Paramètre estimé	Sum of tValue	p_value
3 (énergie)	1. McFadden	37%		
	2. N of Obs	40313		
	act_99	0,000015	58,02	< 0,001
	actval_99	-0,000041	-41,78	< 0,001
	d_chatelet	0,023605	9,27	< 0,001
	dart_99	-0,075236	-17,54	< 0,001
	dhigh_99	-0,071855	-15,76	< 0,001
	emploiVP	0,267554	13,24	< 0,001
	TV_99	-0,392340	-2,64	0,008
	vact_99	0,000001	23,31	< 0,001
	vactval_99	0,000002	14,64	< 0,001
	vresu_99	0,000020	21,41	< 0,001
	vtrain_99	-0,019272	-2,59	< 0,001
	zus	0,105856	2,42	0,015

Secteur	Variable	Paramètre estimé	Sum of tValue	p_value
4 (commerces)	1. McFadden	32%		
	2. N of Obs	64443		
	act_99	0,000014	70,28	< 0,001
	actval_99	-0,000030	-39,20	< 0,001
	d_chatelet	0,008753	4,61	< 0,001
	dart_99	-0,018174	-6,13	< 0,001
	dhigh_99	-0,123642	-34,67	< 0,001
	emploiTC	0,027101	6,74	< 0,001
	emploiVP	0,029801	1,98	0,047
	subway_99	0,400478	29,69	< 0,001
	TC_99	-0,196757	-14,09	< 0,001
	TV_99	-1,053861	-9,37	< 0,001
	vact_99	0,000001	20,87	< 0,001
	vresu_99	0,000010	14,27	< 0,001
vtrain_99	0,033394	5,83	< 0,001	
zus	0,155076	4,59	< 0,001	

Secteur	Variable	Paramètre estimé	Sum of tValue	p_value
5 (finances)	1. McFadden	69%		
	2. N of Obs	35319		
	act_99	0,000002	4,74	< 0,001
	actval_99	0,000003	2,27	0,023
	d_chatelet	0,113100	24,21	< 0,001
	dart_99	-0,176310	-21,64	< 0,001
	dhigh_99	-0,084810	-9,48	< 0,001
	emploiTC	0,172922	17,72	< 0,001
	emploiVP	0,431511	11,89	< 0,001
	subway_99	0,087392	4,12	< 0,001
	TC_99	0,043464	3,83	< 0,001
	TV_99	-2,780817	-11,77	< 0,001
	vact_99	0,000001	10,10	< 0,001
	vactval_99	0,000003	11,36	< 0,001
	vresu_99	0,000043	33,65	< 0,001
	vtrain_99	0,337710	36,17	< 0,001
zus	0,848177	14,64	< 0,001	

Secteur	Variable	Paramètre estimé	Sum of tValue	p_value
6 (serv. aux entreprises)	1. McFadden	59%		
	2. N of Obs	149719		
	act_99	0,000012	55,77	< 0,001
	actval_99	-0,000022	-31,14	< 0,001
	d_chatelet	0,075989	37,51	< 0,001
	dart_99	-0,100080	-29,47	< 0,001
	dhigh_99	-0,043823	-12,57	< 0,001
	emploiVP	0,361798	22,87	< 0,001
	subway_99	0,138510	11,18	< 0,001
	TC_99	0,030366	3,33	0,001
	TV_99	-3,156016	-27,25	< 0,001
	vact_99	0,000001	26,44	< 0,001
	vactval_99	0,000002	12,24	< 0,001
	vresu_99	0,000047	68,90	< 0,001
	vtrain_99	0,183257	33,51	< 0,001
	zus	-0,189389	-5,51	< 0,001

Secteur	Variable	Paramètre estimé	Sum of tValue	p_value
7 (éducation)	1. McFadden	30%		
	2. N of Obs	52957		
	act_99	-0,000001	-2,10	0,036
	actval_99	0,000007	7,19	< 0,001
	d_chatelet	0,047921	22,93	< 0,001
	dart_99	-0,115595	-31,20	< 0,001
	dhigh_99	0,066391	23,44	< 0,001
	emploiTC	0,099555	21,99	< 0,001
	subway_99	0,109714	8,05	< 0,001
	TC_99	-0,336951	-23,91	< 0,001
	TV_99	-3,125182	-25,78	< 0,001
	vresu_99	0,000049	67,96	< 0,001
	vtrain_99	0,086788	15,33	< 0,001
	zus	0,483818	14,63	< 0,001

Secteur	Variable	Paramètre estimé	Sum of tValue	p_value
8 (administ.)	1. McFadden	24%		
	2. N of Obs	62458		
	act_99	-0,000006	-20,39	< 0,001
	actval_99	0,000022	26,02	< 0,001
	d_chatelet	0,031407	16,18	< 0,001
	dart_99	-0,069824	-22,55	< 0,001
	emploiTC	0,056059	13,03	< 0,001
	emploiVP	0,185755	12,07	< 0,001
	subway_99	0,072970	5,46	< 0,001
	TC_99	-0,851501	-39,82	< 0,001
	TV_99	-0,851566	-7,59	< 0,001
	vact_99	-0,000001	-10,52	< 0,001
	vactval_99	0,000001	7,86	< 0,001
	vresu_99	0,000026	35,23	< 0,001
	vtrain_99	0,116665	21,93	< 0,001
zus	0,330221	11,13	< 0,001	

Tableau 7 : Variables retenues du modèle de localisation des emplois

3.6. Validation du modèle

3.6.1. Pseudo-R² de McFadden

Tous les pseudo-R² sont supérieurs à 30% sauf pour le secteur 8. Rappelons que cet indicateur de la qualité d'ajustement ne couvre pas l'intervalle [0-1] et qu'il est toujours faible. Il est communément admis qu'un pseudo-R² d'un logit supérieur à 30% montre que le modèle est très satisfaisant.

Par ailleurs, les coefficients ne présentent pas de signes contre-intuitifs. Par exemple, la propension à se localiser dans une cellule augmente bien quand l'accessibilité aux emplois augmente ou quand le nombre de gares augmente.

3.6.2. Pourcentage de prédictions correctes

Comme pour les modèles de localisation des ménages et d'évolution du MOS, nous introduisons un deuxième indicateur de la validité du modèle : le pourcentage de prédictions correctes en partant de la base de calage de 464519 observations, tous secteurs confondus. 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.

Département où les emplois se sont installés	Nb d'emplois observés	Nb d'emplois prévus	% prévus
75	135 781	125 259	92%
77	29 390	23 506	80%
78	45 185	35 100	78%
91	39 755	32 645	82%
92	92 849	98 656	106%
93	42 738	54 058	126%
94	39 653	45 297	114%
95	39 168	31 987	82%

Tableau 8 : Taux de prédictions par département

On constate que le taux de prédictions correctes est moyen. Il y a une sous-estimation sur Paris et la grande couronne et une sur-estimation sur les départements de la petite couronne. Même avec des résultats très agrégés, on se rend compte que le modèle est moyennement satisfaisant.

4. Modèle de prix de l'immobilier résidentiel

4.1. Rappel du modèle dans UrbanSim

Aux Etats-Unis, depuis 50 ans, les espaces urbains sont désinvestis par la puissance publique. Les agglomérations deviennent des patchworks où alternent espaces totalement privatisés et espaces négociés entre investisseurs privés et collectivités publiques. Ces dernières n'ont pas toujours la capacité d'imposer le bien commun de l'urbanité aux groupes de pression privés. Ainsi, le prix du terrain a en général plus de valeur que le prix de ce que l'on bâtit par-dessus. On peut y construire de l'immobilier résidentiel, des commerces, des activités de type industriel ou des bureaux, selon les projets des investisseurs privés.

Il y a ainsi quatre variables dans la table gridcell correspondantes à la valeur (en terme de capital) pour chacun de ces postes :

- residential_improvement_value
- commercial_improvement_value
- industrial_improvement_value
- government_improvement_value

Ces quatre variables, attributs d'une cellule, ainsi que les quatre variables de voisinage, sont susceptibles d'expliquer les phénomènes de localisation, d'évolution du MOS et des prix du foncier.

Le modèle de prix d'UrbanSim 4 est un modèle hédonique de prix du terrain de type régression linéaire. Il produit en sortie les variables « residential_land_value » et « non_residential_land_value » qui sont mises à jour dans la table gridcell chaque année. Connaissant le nombre de résidences et les surfaces du non-résidentiel de chacun des 3 autres postes (commerces, industries, bureaux) pour l'année initiale de la simulation, qui vont évoluer chaque année au cours de la simulation, UrbanSim peut calculer les 4 variables de type improvement_value chaque année plus les 4 de voisinage. Ces variables vont alimenter ensuite les deux modèles de localisation et le modèle d'évolution du MOS si elles interviennent comme variables explicatives. En fait, suite au regroupement des trois activités en une seule (voir partie 2.4), il ne reste plus que quatre variables.

Il s'avère que pour notre cas de la région d'Ile-de-France, nous ne pouvons pas appliquer ce principe propre à l'économie de l'immobilier aux Etats-Unis car nous disposons de très peu de données sur les valeurs foncières. En revanche, il existe une base de données exhaustive sur les ventes immobilières par les particuliers : la base BIEN des Notaires (Base d'Informations Economiques Notariales). Celle-ci a l'immense avantage d'être géolocalisée. On a donc adapté le modèle pré-établi dans UrbanSim 4 pour le cas de l'Ile-de-France. On a donc estimé un modèle linéaire de prix de l'immobilier résidentiel. Pour ce qui est du prix du non résidentiel, c'est-à-dire les activités, nous avons utilisé un pis-aller, qui sera expliqué dans la partie 4.6⁴. En sortie du modèle, la variable calculée est le prix moyen d'un logement par cellule. Cette variable va être écrite chaque année dans l'attribut residential_land_value contenue dans la table gridcell (voir paragraphe 6.1.1.2 sur l'interaction au sein d'UrbanSim).

⁴ En décembre, nous avons commencé à travailler sur un modèle de prix des bureaux à partir de données issues de la Cote Callon.

On cherche à modéliser l'interaction entre les prix des logements, les modèles de localisation et le modèle de trafic. Il est clair que le prix des logements est un des facteurs explicatifs de la mobilité résidentielle.

Un parc immobilier vaste et fluide facilite la mobilité des ménages. Il permet aux individus de saisir pleinement les opportunités de travail qu'ils peuvent rencontrer au cours de leur carrière. Une insuffisance de logements disponibles pousse les prix de l'immobilier à la hausse. Une hausse du prix de l'immobilier engendre un large étalement des grandes communes vers la périphérie. Le prix d'un logement dépend aussi de son accessibilité aux transports.

Mais la hausse des prix n'est pas la seule raison de l'étalement urbain. Il faut aussi admettre que tout n'est pas modélisable en ce domaine. Les politiques foncières telles que le prêt à taux zéro ont beaucoup contribué à l'étalement urbain. Il est difficile d'introduire au cours de la simulation dans le temps un paramètre modélisant le prêt à taux zéro.

En réalité, nous avons mené deux recherches en parallèle pour modéliser le prix d'un logement : l'une au niveau du découpage cellulaire en 50000 cellules sur un territoire similaire à celui de la Tangentielle nord (territoires du RER E et du Tramway T1), l'autre au niveau communal (1300 communes en Ile-de-France). Après avoir rappelé quelques statistiques sur le prix du logement entre 1990 et 1999, nous présentons succinctement ces deux recherches.

4.2. Statistiques descriptives sur les prix de l'immobilier entre 1990 et 1999

La période 1991- 1997 est une période de baisse globale des prix de l'immobilier, après une bulle spéculative entre 1989 et 1991. La période entre 1997 et 2002 a été une période de reprise de l'immobilier, qui n'a pas cessé depuis, avec des taux de croissance à 2 chiffres. Ces évolutions de prix sont complètement déconnectées de l'évolution des prix à la consommation, la correction effectuée sur les prix est donc marginale.

La figure 3 montre une image générale du marché immobilier sur la période 1990-2003 par département, avec des tendances qui apparaissent nettement : un sommet en 1991 et un creux vers le milieu de 1997. Les évolutions sont plus marquées à Paris qu'en banlieue.

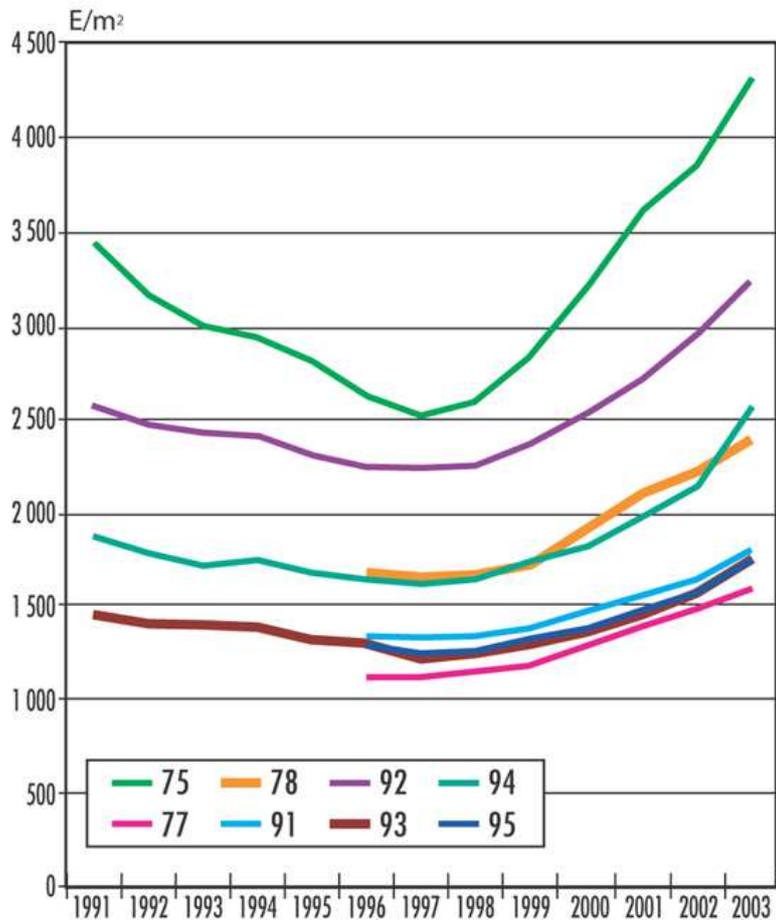


Figure 3 : Evolution des prix/m² par département, tous logements vendus confondus
 (Source: « Les Cahiers de l'Habitat de l'IAURIF » supplément Habitat n°40 – février 2006)

Les prix en €/m² varient considérablement selon les départements. Paris vaut plus du double de la Seine-Saint-Denis en 1999. Les Yvelines sont plus chères que le Val de Marne, ce qui montre que l'appartenance à l'ouest parisien est plus déterminante que la proximité de la capitale. Il existe d'importants écarts au sein d'un même département, reflétant la composition sociale et la richesse relative de la population.

Les écarts de prix entre communes continuent à s'accroître. Ils expriment surtout la qualité d'une localisation (proximité des transports, services et équipements, environnement, espaces verts, écoles). Mais c'est surtout l'adresse, l'image sociale de la commune ou du quartier qui jouent un rôle crucial dans la valeur d'un logement. Ce facteur est quasiment impossible à modéliser. L'augmentation des prix accroît l'effet de sélection des acquéreurs, filtrant l'installation dans les communes les plus chères pour n'y laisser venir que les ménages disposant de revenus élevés. Ainsi se renforce peu à peu la ségrégation sociale.

4.3. Estimation d'un modèle calé sur deux cas : le RER E et le Tramway T1

Cette étude économétrique a fait l'objet d'un rapport commandé par Réseau Ferré de France en 2006 (voir [3]). Le champ de cette étude était cependant plus large qu'une contribution au projet *SIMAU*. Il s'agissait d'essayer de quantifier la création et la récupération éventuelle de la **rente foncière créée par l'amélioration d'une desserte** résultant, soit d'une infrastructure de

transport nouvelle, soit de l'amélioration d'une infrastructure existante (par exemple au droit d'un échangeur nouvellement créé ou d'une nouvelle station de métro).

Dans les méthodes classiques d'évaluation socio-économiques des projets d'infrastructures de transports, la valorisation foncière induite par un projet n'est jamais prise en compte. On peut penser que les propriétaires de logements situés dans l'aire d'influence des gares du RER E et des stations du T1 ont vu leur logement prendre de la valeur après la mise en service de ces infrastructures, et peut-être même un ou deux ans avant par anticipation. Mais d'une part ces gains ne sont pas quantifiables tant que les logements ne sont pas vendus, d'autre part il est difficile de connaître l'aire d'influence exacte d'une gare. D'où la nécessité de recueillir des données sur les ventes (ou transactions).

Il est clair qu'il existe d'autres impacts que la valorisation de l'immobilier résidentiel. Les commerces situés dans les quartiers de gare pourront voir leur chiffre d'affaires augmenter après la réalisation (peut-être diminuer pendant la durée des travaux) et leurs locaux valorisés. De même, les loyers de bureaux pourront augmenter plus qu'ailleurs après la mise en service.

Il est communément admis que de telles plus-values ne doivent pas être prises en compte au titre des avantages d'un projet, car elles peuvent être interprétées comme une internalisation par les propriétaires d'avantages déjà pris en compte dans l'évaluation, tels que les gains de temps. Le problème de la récupération de ces rentes par la puissance publique ne s'en pose pas moins.

La difficulté majeure, bien connue dans ce genre d'exercice, est donc de pouvoir isoler l'effet propre de la mise en service de l'infrastructure de TC parmi de nombreux autres effets induits par la politique locale, le contexte macro-économique et éventuellement la spéculation immobilière. Un des moyens d'y remédier est de choisir des territoires traversés par des transports en commun où il se produit a priori moins de changements importants provoqués par les politiques locales, c'est-à-dire des territoires faiblement dynamiques. En ce qui concerne le contexte macro-économique et la spéculation immobilière (en particulier celle de la période 1989-1991), il a fallu trouver une méthode pour corriger les séries temporelles de prix de manière à pouvoir les comparer sur des années différentes, par exemple 3 ans avant et 3 ans après la mise en service de l'infrastructure en réajustant les indices du logement.

Deux infrastructures ont été choisies pour caler le modèle : le RER E (EOLE) et le tramway T1 Saint-Denis – Bobigny.

- Le RER E a été mis en service en juillet 1999. Il comporte deux branches : la branche Haussmann St Lazare - Chelles-Gournay et la branche Haussmann St Lazare - Villiers-sur-Marne (prolongé jusqu'à Tournan en décembre 2003). On compte en tout 21 gares, dont deux dans Paris. En heures de pointe, la fréquence est de 7 à 8 mn sur la première branche et de 15 mn sur la deuxième.
- Le tramway T1 a été mis en service en décembre 1992. Ses caractéristiques principales étaient les suivantes : 9 km de lignes, 21 stations, capacité de 2300 par heure par sens, une fréquence de 4 mn à la pointe (7 mn en heures creuses), une vitesse commerciale de 16,5 km/h et des correspondances avec les lignes 5, 7, 13, lignes de banlieue nord SNCF et une trentaine de lignes de bus. Il a été prolongé en décembre 2003 jusqu'à Noisy-le-Sec, offrant 5 stations supplémentaires et une nouvelle correspondance avec le RER E.

Ces deux cas nous paraissaient correspondre à nos besoins. De plus, ils ont l'avantage de desservir en partie le territoire traversé par la future Tangentielle nord.

4.3.2. Analyse descriptive des données

Les données sont analysées au niveau désagrégé, c'est-à-dire au niveau des transactions mêmes. Cette analyse désagrégée permet de prendre en compte finement la localisation précise du logement par rapport à la ligne et aux gares, mais aussi les caractéristiques du logement et de son environnement immédiat.

Les questions que l'on s'est posé sont, entre autres :

- Observe-t-on des prix plus élevés à l'intérieur des zones d'influence des gares, par rapport au reste de la zone d'étude, et à la zone témoin ?
- Quels sont les taux de croissance des prix observés depuis 7 ans (3 ans avant la mise en service et 4 ans après), et peut-on définir une segmentation de ces taux ou des prix selon le niveau d'offre ?
- Quelles principales autres variables explicatives peut-on identifier ?

L'analyse descriptive des données est spatio-temporelle :

- **Spatialement**, elle prend en compte le concept de proximité, représentée par la distance à la gare ou station la plus proche par création de couronnes autour des gares ou stations. Une couronne est affectée à un des trois types de zones :
 - une zone d'influence proche
 - une zone intermédiaire (a priori une zone d'influence éloignée)
 - une zone témoin (a priori zone où l'influence de la gare ou de la station ne joue pas)

La méthodologie est inédite dans le sens où **l'on s'affranchit des limites communales** pour élaborer ce diagnostic.

Les zones d'influence des gares ou arrêts autres que ceux d'EOLE et du tramway T1 ont été exclues de la zone témoin.

- **Temporellement**, ce sont les évolutions des données de transactions avant et après la mise en service de l'infrastructure qui sont décrites.

Des couronnes (ou buffers) centrées sur chacune des 13 gares EOLE ont été créées :

- de rayons⁵ multiples de 100 mètres entre 0 et 800 mètres
- puis de rayons multiples de 200 mètres entre 800 et 3000 mètres.

⁵ La distance à vol d'oiseau (ou euclidienne) minore la distance effective qui est une distance (dite de Manhattan) le long d'un réseau rectiligne.

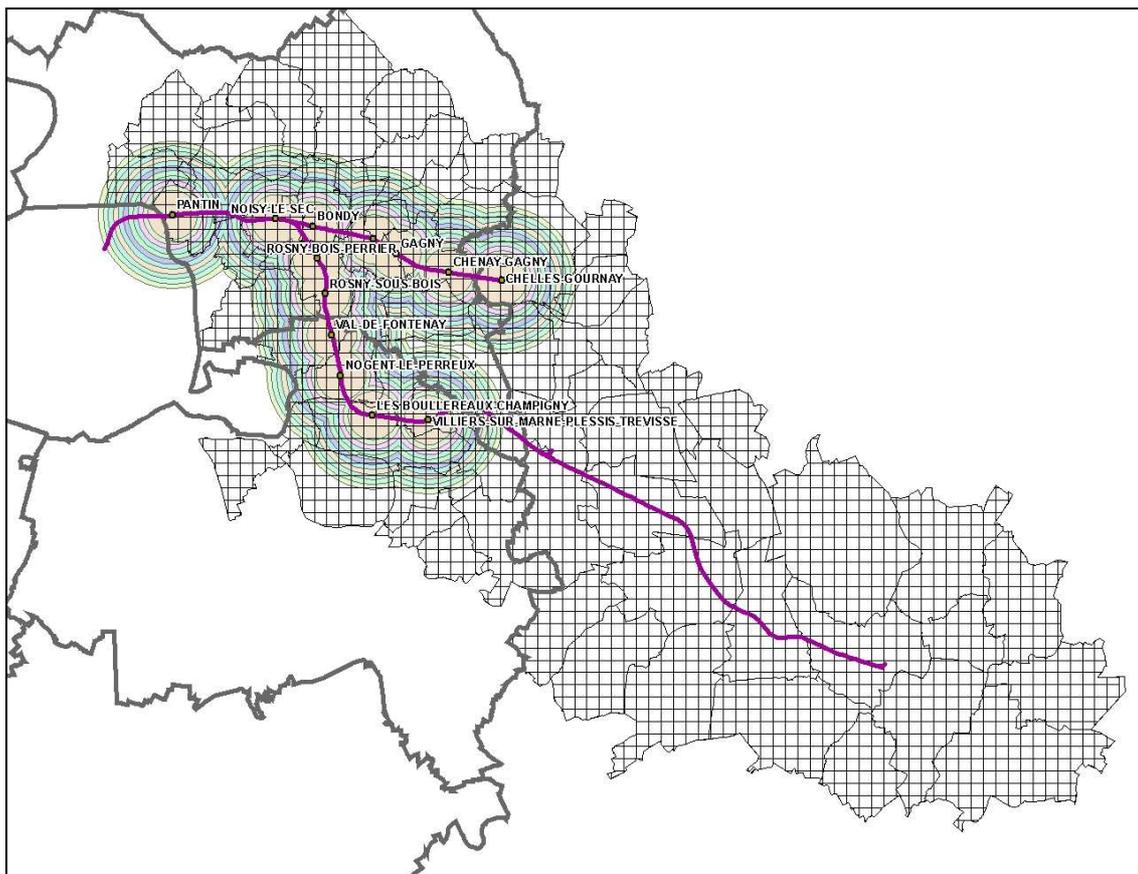


Figure 5 : Buffers centrés autour des gares d’EOLE, maillage en cellules carrées de 500 m

Le problème a consisté à déterminer la zone d’influence qui est celle pour laquelle des variations de prix ressortent de manière plus marquée que dans la zone témoin. Finalement, le découpage utilisé retenu est constitué de trois zones : une zone d’influence proche, une zone d’influence éloignée et la zone témoin. Cette zone témoin est définie comme le reste de l’aire d’étude excluant la zone d’influence de toutes les gares (tant celles d’EOLE que des autres lignes ferrées et des stations de métro).

Une analyse des densités de mutations par rapport à une zone de 3000 mètres autour des gares aboutit au graphique suivant :

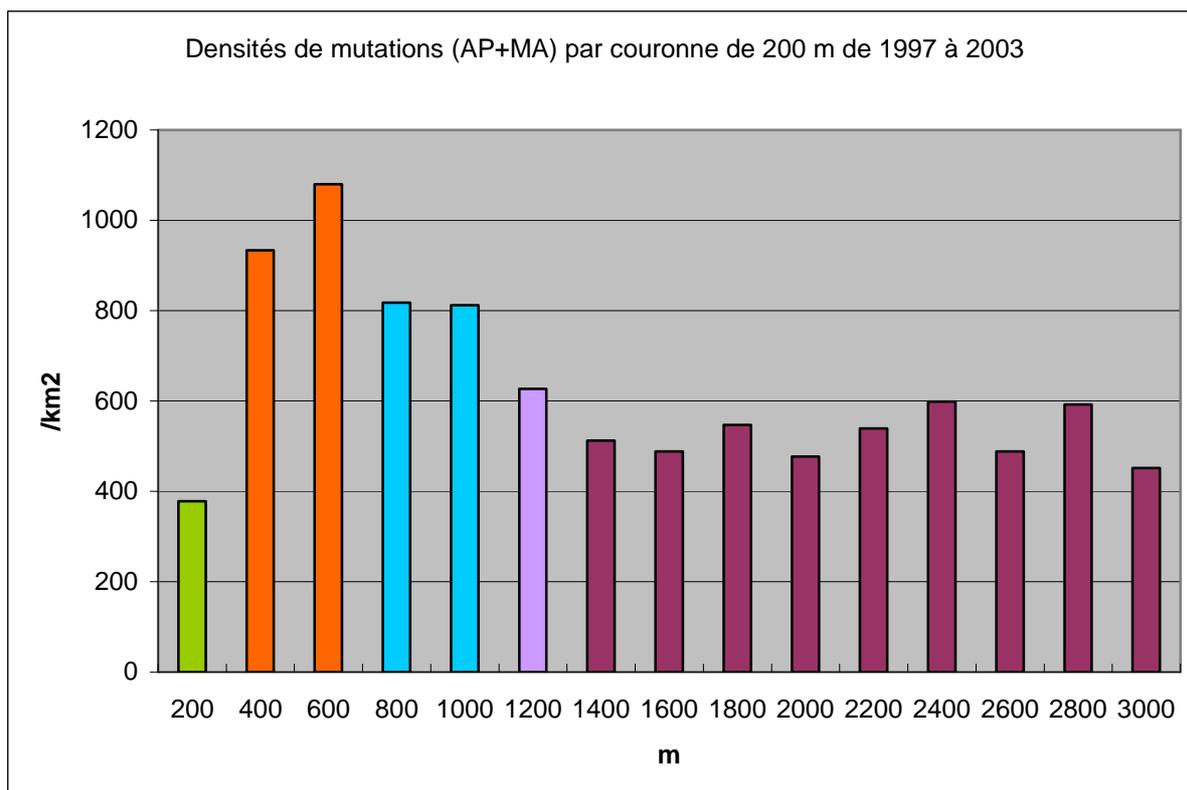


Figure 6 : Densités de mutations par couronne

Il apparaît ainsi que la zone d'influence est à moins de 1200 m des gares. On distingue alors une zone d'influence proche, scindée en deux couronnes, et une zone d'influence éloignée appelée zone intermédiaire, elle aussi scindée en deux couronnes :

- 0-200 m
 - 200 – 600 m
- } Zone d'influence proche
- 600 – 1000 m
 - 1000 – 1200 m
- } Zone intermédiaire
- > 1200 m
- } Zone témoin

La zone d'influence élargie des autres gares RER a été également fixée à 1200 mètres et celle des stations de métro à 600 mètres. Sur l'aire d'étude d'EOLE, il y a 13 gares EOLE, 34 autres gares RER ou train de banlieue (dont les 6 gares après Villiers-sur-Marne jusqu'à Tournan) et 19 stations de métro.

4.3.3. Diagnostic

On a calculé les taux d'évolution annuel des prix moyens au m² par couronne puis par zone, par type de logement et par période.

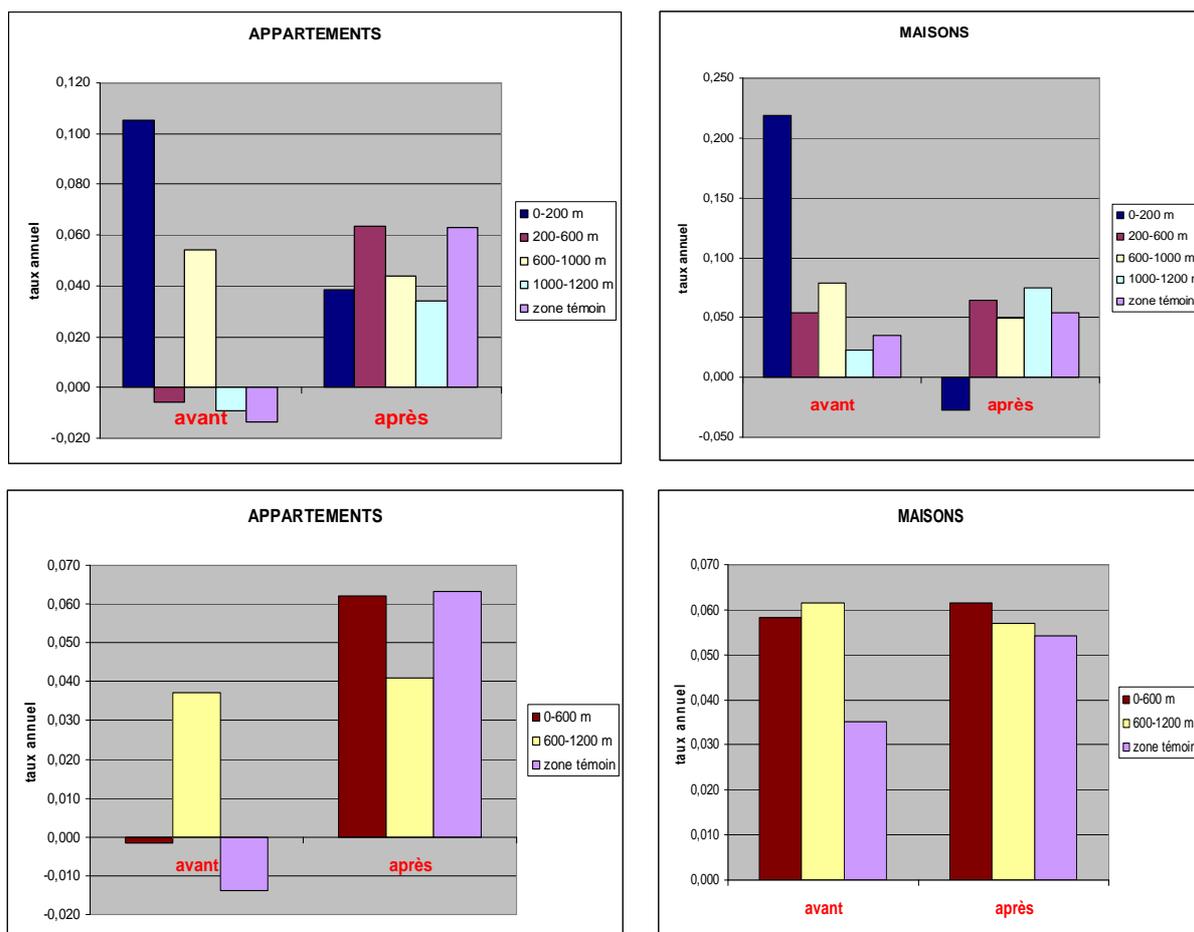


Figure 7 : Taux d'évolution annuel des prix moyens au m² par couronne puis par zone, par type de logement et par période.

Entre 1997 et 1999, on ne trouve pas de relation claire entre la variation annuelle des prix et la distance aux gares. En revanche, après la mise en service d'EOLE, pour les appartements, on constate que le rythme annuel d'augmentation (pendant 4 ans) des prix la plus importante se trouve dans la couronne 200-600 mètres. Mais cette couronne ne se distingue pas de la zone témoin.

En ce qui concerne les maisons, la zone d'influence proche se distingue bien de la zone témoin après la mise en service d'EOLE. Il semblerait donc que la mise en service d'EOLE, toutes choses égales par ailleurs, ait eu un impact sur l'augmentation des prix des maisons dans un rayon de 1200 m, en particulier dans la couronne 200-600 mètres et l'auréole 1000-1200 mètres. Lorsqu'on regroupe les couronnes deux à deux pour reconstituer les zones d'influence, l'augmentation du prix moyen dans la zone intermédiaire est plus lente que dans la zone d'influence proche.

Nous avons également calculé les différentiels de prix moyen au m² des appartements et des maisons à proximité d'une gare EOLE par rapport au prix moyen au m² de la zone témoin

APPARTEMENTS

	1997	1998	1999	2000	2001	2002	2003
A moins de 600 m d'une gare	9,9%	14,2%	12,6%	12,0%	12,7%	9,9%	12,2%
De 600 à 1200 m d'une gare	13,6%	20,5%	25,6%	22,6%	19,4%	18,3%	15,5%

MAISONS

	1997	1998	1999	2000	2001	2002	2003
A moins de 600 m d'une gare	9,1%	9,9%	14,1%	13,5%	17,1%	16,2%	17,3%
De 600 à 1200 m d'une gare	9,1%	13,3%	14,7%	10,5%	15,2%	12,9%	15,9%

Tableau 10 : Différentiel de prix moyen au m² des maisons à proximité d'une gare EOLE par rapport au prix moyen au m² de la zone témoin (en € constants 97)

Ces deux tableaux montrent que les prix de la zone 600-1200 m sont les plus élevés par rapport à la zone témoin pour les appartements quelle que soit l'année. Pour les maisons, les prix sont les plus élevés dans un rayon de 600 m à partir de l'an 2000, soit après la mise en service d'EOLE.

En résumé, pour le cas du RER E, l'analyse des données a fait ressortir les éléments principaux suivants :

- **Il se vend plus d'appartements au km² à proximité des gares qu'ailleurs.** Pour les maisons, cela est moins évident mais il y a sans doute un effet de composition de l'habitat individuel/habitat collectif qui n'est pas uniforme sur le territoire.
- **Il y a bien une différenciation locale du prix moyen au m² des logements.** Il n'y a pas de gradient continu descendant des prix en fonction de la distance mais on observe tout de même des couronnes proches où les prix des maisons augmentent plus vite que dans la zone témoin située au-delà de 1200 mètres. L'impact sur les prix ne se produit pas à proximité immédiate des gares mais dans à une distance entre 200 et 600 mètres.
- **Il n'y a pas de lien de causalité évident entre augmentation des prix et mise en service d'EOLE,** juste des corrélations plus ou moins positives qui ne sont pas complètement contre-intuitives. On ne peut pas dire qu'il y a eu une spéculation immobilière autour des gares. Ainsi, il semble que l'effet du RER est plus important sur les volumes de ventes que sur les prix.
- L'aire d'étude est très étendue et hétérogène. Les moyennes calculées sur l'ensemble des gares masquent des situations très disparates entre les différentes gares. Le contexte local est un facteur déterminant de l'impact sur les prix de l'immobilier résidentiel : environnement urbain et social, revenu moyen des ménages, politique communale volontariste (par exemple nombre de ZAC), jeu des acteurs locaux ... Autour de chaque gare se constitue un micro-marché foncier plus ou moins actif, chacun ayant sa spécificité. L'effet sur les prix de l'immobilier de la mise en service d'EOLE est optimisé par la synergie d'autres effets concomitants.

La même analyse a été effectuée sur les **données du tramway T1**. Ce cas s'est avéré beaucoup plus difficile à analyser ou interpréter que celui d'EOLE. On observe globalement plus de ventes de logements le long du T1 dans un corridor proche et il n'y a pas eu d'impact significatif sur les prix. Quelques hypothèses peuvent être formulées :

- Il y a peut-être eu un impact sur les prix mais cet impact est complètement absorbé dans une conjoncture très défavorable à l'immobilier durant la période 1990-1996.
- S'il y a un léger impact positif sur les prix, il est extrêmement faible voire non significatif. Peut-être est-dû au fait que les avantages procurés par le T1 sont très faibles en ce qui concerne les gains d'accessibilité. Le T1 joue essentiellement un rôle de cabotage et de rabattement sur le métro. Il se substitue à des lignes de bus existantes qui assuraient déjà le même service avec une qualité moindre. Il a certes généré une meilleure qualité de vie urbaine par la requalification de l'axe mais ces gains sont peut-être compensés par de nouvelles nuisances induites par le bruit et les vibrations du tramway.
- Il se vend apparemment plus de logements dans la zone d'influence proche qu'au-delà. Peut-être est-ce dû au fait que les vendeurs ne supportent pas le bruit du tramway même s'ils y ont gagné en diminution de la pollution et du bruit routier. En tous cas, la raison n'est pas dans une quelconque recherche de plus-value de la part des propriétaires car la baisse des prix n'est pas moins marquée à proximité que dans la zone témoin.
- C'est peut-être aussi dû à la structure de l'habitat à proximité des stations où les petits appartements dont le taux de rotation (« turn-over ») est plus élevé sont plus nombreux que les grands appartements et les maisons.
- S'il y a une plus-value immobilière suite à la mise en service du T1, on ne sait pas quelle est la part due au gain d'accessibilité et la part due au gain d'aménagement et environnemental (insertion paysagère, diminution de la pollution).

4.3.4. Modélisation

Le but est d'estimer un modèle hédonique de prix conformément au modèle « Land Price Model » d'Urbansim 4.

Rappel : La méthode hédonique d'évaluation marchande de biens est couramment utilisée dans l'analyse économique appliquée depuis plusieurs années, en particulier dans le marché du logement.

Son point de départ repose sur la constatation qu'un bien n'est pas recherché pour lui-même, mais pour les quantités de différentes caractéristiques qui le définissent. Cette méthode suppose que le prix d'un bien sur le marché est égal à la rémunération qu'accorde le marché aux caractéristiques possédées par ce bien.

Il s'agit donc de rechercher les caractéristiques fondamentales des logements considérés, pour obtenir une fonction de prix qui ne dépende que de ces caractéristiques. C'est le modèle de Rosen (1974).

Soit z un logement. Rosen considère que ce bien est entièrement décrit par un nombre n de caractéristiques :

$$z = (z_1, z_2, \dots, z_n)$$

Le logement est vendu au prix $p(z)$: c'est la fonction de prix hédonique. Sous l'hypothèse d'équilibre entre l'offre et la demande du marché de chaque caractéristique, il s'écrit :

$$p(z) = f(z_1, z_2, \dots, z_n)$$

Dans le modèle considéré, le prix s'écrit :

$$p(z) = a \times f(z_1, z_2, \dots, z_n) + \varepsilon$$

où ε est l'erreur commise.

Pour $f(z)$, Rosen préconise l'utilisation de trois types de relations : linéaire, logarithmique, semi-logarithmique.

Par régression, on obtient le prix marginal de chaque caractéristique, c'est-à-dire le prix associé à la variation marginale de la caractéristique considérée, les quantités des autres caractéristiques étant données :

$$\forall i, p_i = p(z) \times z_i$$

Ce modèle est très controversé, car l'attrait d'un logement n'est pas facilement mesurable, et dépendra des priorités des acheteurs. De plus, la théorie hédonique ne fait pas intervenir le problème de l'offre et la demande dans la détermination des prix, ce qui l'affaiblit d'autant plus.

Comme il l'a été dit, dans le contexte de l'Ile-de-France, nous modélisons le prix des logements et non les prix du terrain. Or il est connu que la valeur d'un logement est due à des attributs dits intrinsèques (surface, parking, salle de bains, etc) et des attributs extrinsèques (proximité d'une gare, d'un espace vert, d'une école, etc). Or le modèle hédonique de prix du terrain peut fonctionner dans Urbansim car ce sont essentiellement des variables extrinsèques qui contribuent au prix au mètre carré d'un terrain. En revanche, pour le prix d'un logement, ce sont les variables intrinsèques qui priment.

Dans un premier temps, nous avons essayé d'estimer un modèle hédonique de prix moyen d'un logement (appartement et maisons confondus) situé sur le territoire d'EOLE, chaque logement étant localisé dans le maillage en 50000 cellules.

Le modèle hédonique est basé sur l'hypothèse que les prix sont des indicateurs de l'adéquation entre l'offre et la demande, entre les différentes zones et les différents types de logement. Les prix expliquent en partie les choix de localisation des ménages. Généralement, des prix trop élevés en centre-ville poussent les ménages à chercher de nouvelles localisations en périphérie.

Les hypothèses suivantes sont prises :

- L'ensemble des agents est preneur de prix et les ajustements sont faits en réponse à la demande et l'offre agrégées.
- Les préférences et les déséquilibres entre offre et demande se reportent dans la valeur de l'immobilier.
- Il existe un taux de vacance structurel de long terme pour chaque type de bâti et la relation entre le taux de court terme et le taux de long terme influence le prix du bâti.
- Le prix du bâti est déterminé par une approche en termes de prix hédoniques. Cela permet de prendre en compte les interactions entre la demande et l'offre. Du fait que la régression hédonique utilise des variables actualisées par les différents modules d'UrbanSim, on peut également actualiser les prix de cette façon. En plus de cette détermination hédonique des prix relatifs, le prix absolu s'ajuste aux variations de l'offre et de la demande. Ces fluctuations sont reliées à la différence entre le taux de vacance structurel V^s_i du type de bâti i et le taux de vacance courant V^c_{it} à une date t .

L'équation hédonique est alors :

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

où P_{ilt} est le prix unitaire du bâti de type i , dans la cellule l , l'année t et X_{ilt} un vecteur d'attributs de la localisation. Les prix sont actualisés annuellement, en fin de période d'activité et utilisés ensuite comme références pour l'année suivante.

Les attributs testés X_{ilt} sont des variables extrinsèques (celles de gridcells et les variables transport utilisées dans les autres modèles de localisation des ménages ou des emplois) mais il n'y a aucune variable intrinsèque du logement. Evidemment, le résultat ne faisait aucun doute : l'estimation n'a rien donné. Donc il est impossible d'estimer un modèle de prix du logement à la cellule basé uniquement sur des variables extrinsèques.

Dans un deuxième temps, nous avons estimé un modèle hédonique des prix de logement en prenant en compte les variables intrinsèques et extrinsèques. Cependant, nous étions conscients que ce modèle ne serait pas utilisable dans UrbanSim. Mais notre objectif était d'estimer un modèle explicatif pour mesurer la part des variables extrinsèques, en particulier les variables transport dans le prix total du logement. Les appartements et les maisons ont été étudiés séparément.

La variable expliquée est le prix de vente du logement, exprimé en euros constants, c'est-à-dire que l'effet de l'inflation sur la période de temps considérée est éliminé.

Les variables testées sont :

- surface habitable
- année de vente
- ancienneté
- nombre de garages
- distance à quatre types de projets : centre commercial, infrastructure sportive, lycée, espace vert
- distance à un hôpital
- distance à une Zone Urbaine Sensible
- distance à une voie ferrée
- pourcentage de HLM dans la commune
- pourcentage d'habitations collectives **dans une cellule de 500 m²** autour du logement
- distance au métro
- distance à la gare la plus proche, et mention du fait que ce soit une gare EOLE
- vente effectuée avant ou après la mise en service du RER E
- temps de trajet et nombre de correspondances vers Saint-Michel, Saint-Lazare, La Défense.
- Pour les appartements, l'étage apparaît aussi, et pour les maisons, la surface totale.

Les variables de distance sont discrétisées en trois ou cinq intervalles.

Nous présentons ci-après les résultats dans le cas des **appartements**.

Pour étudier les facteurs qui expliquent le prix des appartements, on introduit une à une les variables explicatives dans la régression linéaire.

Le modèle est amélioré si on utilise une transformation de Box Cox pour le prix :

$$BC = \frac{1}{\lambda} \times (\text{prix}^\lambda - 1)$$

Le coefficient λ est choisi de façon à minimiser l'erreur dans la régression. Ici, il vaut 0.4.

La base comporte 57 218 observations.

1^{er} modèle : les conséquences du RER E ne sont pas prises en compte dans l'explication du prix des appartements.

Plusieurs variables ne sont pas significatives. C'est le cas de l'étage, de l'année 1996, et des distances aux espaces verts et aux infrastructures sportives. Le modèle arrive à expliquer **76% du prix**.

Variable	Résultat estimé des paramètres	Erreur std	Valeur du test t	Pr > t
Intercept	214.86523	0.33473	641.90	<.0001
SURFH	1.59229	0.00497	320.45	<.0001
ANC	36.42827	0.36436	99.98	<.0001
GARAG	9.42955	0.18590	50.72	<.0001
POURHLM_COMMUNE	-0.75589	0.00935	-80.84	<.0001
POURHAB_COLL_CELLULE	-0.24840	0.00507	-48.98	<.0001
An1997	-7.16706	0.45953	-15.60	<.0001
An1998	-7.94854	0.38871	-20.45	<.0001
An1999	-4.51046	0.35919	-12.56	<.0001
An2001	5.62362	0.35756	15.73	<.0001
An2002	12.71820	0.35437	35.89	<.0001
An2003	24.39941	0.34890	69.93	<.0001
d_ccom1	-3.01889	0.34476	-8.76	<.0001
d_ccom2	-4.61117	0.30294	-15.22	<.0001
d_lycee1	2.47115	0.25732	9.60	<.0001
d_lycee2	0.68684	0.24543	2.80	<.0001
d_hopit1	4.05203	0.35942	11.27	<.0001
d_hopit2	4.48454	0.27957	16.04	<.0001
d_zus1	-10.07916	0.32847	-30.69	<.0001
d_zus2	-2.75560	0.38522	-7.15	<.0001
d_gare1	14.09869	0.68162	20.68	<.0001
d_gare2	12.76887	0.28700	44.49	<.0001
d_gare3	10.14868	0.26148	38.81	<.0001
d_gare4	3.50642	0.35682	9.83	<.0001
d_metro1	9.81686	0.74101	13.25	<.0001
d_metro2	10.41689	0.39115	26.63	<.0001
d_metro3	5.67847	0.52031	10.91	<.0001
d_metro4	14.70437	0.78586	18.71	<.0001

Tableau 11

La variable expliquée est ici le prix transformé par la méthode de Box-Cox, qu'on note BC. Sa valeur de référence est 214,865 ce qui correspond à un prix de 70 490 €, soit une erreur de 20% par rapport au prix attendu. Cette différence est due au fait que la régression n'explique pas entièrement le prix des appartements.

Dans ce modèle, une augmentation de surface de 1m², par rapport à la surface moyenne, entraîne une hausse de 1 593 du BC, c'est-à-dire de 1 299 € du prix.

Le logement de référence est vendu en 2000. Si le même logement est vendu 3 ans plus tard, BC vaut 24,4 points de plus, et le prix augmente donc de 21 477 €.

La proximité d'un centre commercial déprécie la valeur du logement, tout comme la proximité d'une ZUS, le pourcentage de HLM et d'habitations collectives de la commune. La distance à un hôpital est considérée comme un avantage, surtout si l'hôpital se situe de 200 à 400m de l'appartement.

Les appartements proches d'une gare se vendent plus cher.

2^{ème} modèle : on ajoute les variables liées à la mise en service du RER E

Les variables ajoutées sont : le fait que la vente ait eu lieu avant ou après la mise en service du RER, le fait que la gare la plus proche soit une gare EOLE ou non, les temps de trajet et nombres de correspondances vers Saint-Michel, Saint-Lazare et La Défense.

En réalité, le nombre de correspondances vers Saint-Michel, Saint-Lazare et La Défense sont corrélées entre elles et avec la variable avant/après la mise en service. Il faut donc les exclure. La question « la gare la plus proche est-elle EOLE ? » apparaît non significative.

Les temps de transports influent de façon différente sur l'appréciation des prix des logements : pour Saint-Michel et La Défense, la rapidité de transport est valorisée, mais l'effet semble être contraire pour Saint-Lazare, ce que nous ne pouvons expliquer.

En ajoutant les variables transport liées à EOLE, le modèle explique **78 % du prix**, soit deux points de plus, ce qui est minime.

La mise en service du RER E semble avoir un effet positif sur les prix des logements, puisque le coefficient de la variable avant/après est positif.

Variable	Résultat estimé des paramètres	Erreur std	Valeur du test t	Pr > t
Intercept	172.47382	1.11160	155.16	<.0001
SURFH	1.58214	0.00484	326.86	<.0001
ANC	37.25037	0.35535	104.83	<.0001
GARAG	9.74893	0.18180	53.62	<.0001
POURHLM_COMMUNE	-0.59537	0.00969	-61.41	<.0001
POURHAB_COLL_CELLULE	-0.19266	0.00538	-35.78	<.0001
An1997	-5.08005	0.60675	-8.37	<.0001
An1998	-5.86975	0.55788	-10.52	<.0001
An1999	-3.83194	0.41869	-9.15	<.0001
An2001	5.06885	0.34988	14.49	<.0001
An2002	12.27737	0.34678	35.40	<.0001
An2003	24.01070	0.34153	70.30	<.0001
d_ccom1	-3.00709	0.33601	-8.95	<.0001
d_ccom2	-4.67888	0.29496	-15.86	<.0001
d_lycee1	3.23699	0.25338	12.78	<.0001
d_lycee2	1.38452	0.24263	5.71	<.0001
d_hopit1	3.02817	0.35138	8.62	<.0001
d_hopit2	4.02944	0.27425	14.69	<.0001
d_zus1	-7.65954	0.32676	-23.44	<.0001
d_zus2	-2.95661	0.37667	-7.85	<.0001
d_gare1	9.36974	0.66974	13.99	<.0001
d_gare2	10.36355	0.29193	35.50	<.0001
d_gare3	6.48154	0.26913	24.08	<.0001
d_gare4	1.10200	0.35213	3.13	<.0001
d_metro1	10.07969	0.72291	13.94	<.0001
d_metro2	8.16035	0.38386	21.26	<.0001
d_metro3	2.59641	0.51042	5.09	<.0001

d_metro4	13.04650	0.76695	17.01	<.0001
AVANT_APRES	25.83411	0.59787	43.21	<.0001
TEMPS_ST_LAZARE	3.80240	0.09147	41.57	<.0001
TEMPS_ST_MICHEL	-2.79727	0.13390	-20.89	<.0001
TEMPS_LA_DEFENSE	-0.88734	0.06985	-12.70	<.0001

Tableau 12

En se fiant à ce modèle, le fait d'être vendu après la mise en service du RER E fait augmenter le prix du logement de base de 22 847 € (BC augmente de 25,83). En réalité, le résultat est probablement biaisé du fait de la hausse des prix de l'immobilier sur la période étudiée.

En conclusion, ce modèle explicatif permet de montrer que la contribution de la mise en service du RER E dans le prix des logements existe bien mais est plutôt faible. Les variables retenues ne permettent pas d'implémenter un modèle de prix tel qu'il est conçu dans UrbanSim 4.

Nous avons donc estimé un nouveau modèle de prix du logement dans le découpage communal.

4.4. Estimation d'un modèle au niveau communal

Les estimations basées sur les territoires du RER E et du T1 se basent sur un nombre limité de communes qui ne sont pas représentatives de toute la région. Or nous avons besoin d'un modèle valable pour la région entière. C'est pourquoi nous estimons un modèle de prix du logement au niveau communal avec la base de Notaires agrégée au niveau communal.

4.4.1. Données

de prix

Nous disposons de données de prix de vente moyens des logements par commune d'après la base de données des notaires. Tous les prix correspondent, pour chaque commune, à une moyenne des prix de vente des biens correspondants, pondérée par le nombre de biens de ce type vendus dans cette commune au cours de l'année. Cette série de prix comporte différentes lacunes.

- Elle ne porte que sur les transactions (ventes), pas sur les locations.
- Les données ne comportent aucune information sur la surface des logements ou leur nombre de pièces, et l'information sur l'ancienneté des logements est limitée (neuf/ancien).
- Il s'agit du prix de vente des logements, pas de prix au m². Or, les logements sont plus grands en banlieue. Plus généralement, les différences intercommunales de prix dans cette base reflètent autant des différences de surfaces que des différences de prix au m².

C'est sur ces prix que nous avons travaillé, faute de disposer à l'époque de meilleures données de prix. Nous disposons de données annuelles sur la période 1990 (1996 en grande couronne) à 2003. Nous avons utilisé ces années pour combler les quelques « trous » dans la série et obtenir des prix pour chacune des 1300 communes d'Ile-de-France en 1998.

	Tout logement	Appartements			Maisons		
		Neuf	Ancien	Tous	Neuf	Ancien	Tous
Moyenne	116099	95987	69853	77977	163541	138889	141853
Ecart – type	44584	43339	31570	30311	116128	90676	89841
Nombre	1300	1300	1300	1300	1300	1300	1300
<i>Moyenne pondérée*</i>	<i>118695</i>	<i>152149</i>	<i>94666</i>	<i>100843</i>	<i>284179</i>	<i>254185</i>	<i>258717</i>
<i>Ecart – type pondéré</i>	<i>3100940</i>	<i>5289370</i>	<i>3161191</i>	<i>3193116</i>	<i>16168887</i>	<i>16102623</i>	<i>15733518</i>
<i>Somme des poids</i>	<i>5082655</i>	<i>5082655</i>	<i>5082655</i>	<i>5082655</i>	<i>5082655</i>	<i>5082655</i>	<i>5082655</i>
Tous	100%	64%	72%	77%	48%	68%	67%
Appart., neuf	75%	100%	80%	86%	75%	81%	83%
Appart. ancien	91%	88%	100%	96%	64%	74%	76%
Tous appart.	92%	88%	99%	100%	67%	77%	79%
Maison, neuf	67%	80%	75%	76%	100%	72%	76%
Maison, ancien	79%	78%	87%	86%	63%	100%	99%
Toutes maisons	81%	82%	90%	89%	70%	99%	100%

*Les chiffres en italique représentent les valeurs calculées en pondérant par l'effectif communal.

Tableau 13 : Corrélation entre les prix des différents biens selon la Base de Données des Notaires (1998)

Lorsque l'on se concentre sur les prix de vente, la corrélation entre les prix des maisons et ceux des appartements est de l'ordre de 75% à ancienneté identique. C'est plus que la corrélation entre le prix des maisons neuves et anciennes (72%), mais moins qu'entre appartements neufs et anciens (80%). Il semble par ailleurs que les appartements neufs soient valorisés comme les maisons anciennes (corrélation de 81%), alors que la corrélation entre le prix des appartements anciens et celui des maisons neuves n'est que de 64%.

L'ensemble des résultats suggère que le niveau local (communal) du coût du logement peut difficilement se résumer à quelques chiffres, mais devrait au contraire différencier location et achat, appartements et maisons, ainsi que neuf et ancien. Il faudrait donc proposer des modifications du modèle de prix tel qu'il est actuellement conçu dans UrbanSim 4. Ceci permettrait alors d'implémenter le modèle plus sophistiqué présenté en annexe 2 de choix de localisation des ménages qui, lui, distingue les appartements des maisons.

4.4.2. Modèle implémenté dans SIMAURIF

Les prix des logements dépendent de plusieurs variables liées à l'offre ou la demande de logements, elles-mêmes fonction des caractéristiques socio-économiques de la population et des attributs de la commune (ou plus finement de la cellule) où se situe le logement. Parmi les variables retenues dans notre modèle, citons une indicatrice de Paris (qui indique si la commune ou la cellule est localisée à Paris), le temps de trajet moyen en véhicule particulier pour accéder aux emplois à partir du lieu concerné, l'effectif total d'emploi local et le pourcentage de ménages à revenu moyen.

Trois estimations ont été menées : tous logements confondus, appartements, maisons. Le tableau 14 présente les résultats d'estimation des trois modèles des prix de logements (en logarithme). Le R² est de 53% pour le prix moyen des logements, 62% pour les appartements et 70% pour les maisons.

Variable	Logements		Appartements		Maisons	
	Coeff	t-Stat	Coeff	t-Stat	Coeff	t-Stat
Constante	12.8461	65.36	12.4362	76.70	13.0972	72.88
Paris	0.1483	1.63	-0.0472	-0.63	0.2747	3.30
Temps moyen en voiture privée	-0.0008	-1.54	-0.0018	-3.94	-0.0015	-3.08
Total des emplois (/1000)	-0.0010	-0.99	0.0024	2.89	0.0029	3.11
Log du total des emplois	0.0204	3.71	0.0089	1.97	0.0295	5.88
% ménages à revenu moyen	-2.1624	-16.48	-1.8436	-17.03	-2.2916	-19.10
% ménages avec faible revenu	-2.9731	-16.05	-3.2086	-20.99	-3.6871	-21.76
% ménages sans actif	0.9149	2.64	1.3059	4.57	1.3615	4.30
% ménages avec un seul actif	0.4902	2.35	0.8818	5.11	0.8070	4.22
% ménages avec un seul membre	-0.0875	-0.42	0.7494	4.32	1.1503	5.98
% mén. à chef avec un âge moyen	-0.2071	-1.05	-0.4376	-2.68	-0.6086	-3.37
% mén. à chef avec un âge >60	-0.0614	-0.23	-0.7490	-3.40	-0.6340	-2.60
% ménages à chef étranger	1.1972	4.78	1.3319	6.44	1.6740	7.30
% des écoles en ZEP	-0.0846	-1.97	-0.0991	-2.80	-0.0872	-2.22
% de zone franche dans la commune	-1.1984	-4.21	-0.8850	-3.77	-0.1765	-0.68

Tableau 14 : Estimation des prix des logements

Le prix des logements dépend du temps moyen d'accès aux destinations, du nombre d'emplois dans la commune et de la structure socio-économique de la population. Une baisse de 10 minutes du temps moyen de déplacements augmente de 1,5 à 1,8% le prix des maisons et appartements. Une hausse de 10% de la population à bas revenu cause une baisse de 30% du prix des logements. Une même baisse de la proportion de la population à revenu moyen entraîne une baisse du prix de 21%. La fraction des ménages sans ou avec un seul actif a un effet positif sur le prix.

C'est le modèle tous logements confondus qui a été implémenté dans *SIMAU*RIF. On peut remarquer qu'aucune variable de transport en commun n'apparaît dans ce modèle à la commune.

5. Interface simplifiée entre UrbanSim et le modèle de trafic

Il nous a été impossible dans les délais impartis du projet et suite au basculement sur la plateforme OPUS/UrbanSim 4 de redévelopper une interface conviviale de gestion du bouclage automatique entre UrbanSim 4, le modèle de trafic Davisum-METROPOLIS et le modèle de demande de l'IAURIF. A notre grand regret, nous avons donc dû faire un sacrifice : l'abandon du prototype de l'interface que nous avons développé dans la première phase du projet en 2004-2005. Nous avons peut-être aussi consacré trop de temps à l'estimation indépendante de chaque modèle d'UrbanSim.

Mais nous n'avons à aucun moment perdu le fil directeur du projet : simuler l'interaction transport-urbanisation. Nous avons cherché comment fonctionnait l'interaction dans OPUS/UrbanSim 4. Des scripts sont disponibles mais uniquement en interfaçage avec le logiciel de trafic Emme2 et donc dans un format spécifique. Sans doute cette interface a-t-elle été développée dans le cadre d'une étude aux Etats-Unis (Puget Sound ou Salt Lake City). Elle n'est pas ré-utilisable dans notre cas. Nous avons été amenés à écrire notre propre script.

5.1. Simplification de la prise en compte de l'interaction

Nous avons bien pris en compte les modifications de l'offre de transport dans le temps et leurs effets sur l'urbanisation (évolution de l'occupation du sol, relocalisation des ménages et des emplois). Mais cette interaction transport-urbanisation a été modélisée hors phénomène de congestion sur le réseau routier. Rappelons (voir [2], page 60) que les variables transport utilisées dans tous les modèles d'UrbanSim ont été calculées en heures creuses, tant les variables d'accessibilité (de type « Poulit ») que les variables de temps généralisés. Cela devait être provisoire, le temps de recalculer les matrices de demande et le logiciel METROPOLIS, de procéder aux affectations et de refaire alors toutes les estimations économétriques en utilisant des variables en heures de pointe, donc prenant en compte la congestion et le choix de l'heure de départ. Au final, ces variables provisoires en heures creuses seront des variables définitives.

En pratique, on ne réalisera pas donc pas d'affectation ni sur le modèle METROPOLIS pour la route ni sur le modèle Davisum pour les transports en commun, et on n'utilisera pas le modèle de demande de l'IAURIF pour produire les matrices origine-destination par mode. On utilisera Davisum-METROPOLIS pour calculer les variables transport sur des réseaux routiers **non chargés** et de transport en commun pour chaque année de modification de l'offre. Cela autorise finalement à ne pas développer d'interface. Toutes les opérations sur les modèles seront réalisées manuellement en dehors d'UrbanSim 4. Elles produiront des données d'accessibilité et de temps de parcours nécessaires à chaque pas de temps, qui seront alors importées dans UrbanSim 4 au format adéquat. En revanche, UrbanSim 4 tournera automatiquement dans un bouclage annuel sans intervention manuelle de l'opérateur. Les ménages et les emplois seront relocalisés chaque année mais ces nouvelles répartitions ne serviront pas à mettre à jour l'étape de génération du modèle global de déplacements.

Evidemment, cette simplification extrême est loin d'être satisfaisante mais la réalité de la pratique s'est imposée. Il est toujours possible d'arrêter l'exécution d'UrbanSim chaque année, de récupérer les sorties (population, emplois et commerces par zone du modèle), de faire tourner les trois étapes du modèle de demande de déplacements, les affectations sur le réseau routier (c'est inutile en transport en commun), récupérer les matrices de temps, calculer les variables

transport (accessibilités, nombre de nœuds, temps d'accès à Châtelet, distances), les réinjecter dans UrbanSim et enfin faire tourner une année de plus. Cette procédure manuelle prendrait au minimum une vingtaine de jours pour la période 1999-2026.

5.2. Test de modification des réseaux de transport en 1996 et 1993

Nous avons donc introduit les modifications de l'offre de transport de manière exogène dans UrbanSim. Les graphes de réseaux routiers et de transports en commun ont été construits pour deux années, 1993 et 1996, dans les logiciels METROPOLIS et Davisum.

5.2.1. Réseau routier

On supprime sur le graphe de 1999 puis sur celui de 1996 puis sur celui de 1993 les arcs suivants :

1996 par rapport à 1999	A86 entre Bobigny et Drancy
	A86 entre Nanterre et Rueil Malmaison
	Prolongement d'A115 à Taverny
	La RN315 à Gennevilliers
	Le BIP à Ermont
	A104 entre la RD10 et la RD9 (dans le Val d'Oise)
	A104 entre la RD26 et la RD10
	A104 dans la vallée de l'Isle-Adam
	La liaison entre RN6 et la RN19 à Créteil
1993 par rapport à 1996	A5
	A16 jusqu'à la RN1
	A14
	A86 (tunnel de la Croix de Berny)
	A104 : liaison entre la RN1 et la RD26
	A105 à Melun
	Déviations de Houdan
	Boulevard urbain à Gennevilliers
	RN406 à Créteil
	Déviations de Limay
BIP est	
1990 par rapport à 1993	A86 à Fresnes
	A86 de la RN305 à l'A6
	A86 de la RN410 à la RN2
	A86 : liaison vers A1
	A104 : de A105 à A5
	RN12 : déviation de Saint-Appoline (dans les Yvelines)
RD60 à Créteil	

Tableau 15 : Modifications de l'offre routière de 1990 à 1999

5.2.2. Réseau de transports en commun

On supprime sur le graphe de 1999 puis sur celui de 1996 puis sur celui de 1993 les nœuds et les arcs suivants :

1996 par rapport à 1999	METEOR (ligne 14)
	Station Bibliothèque-François-Mitterrand (RER C)
	Ajouter la gare Massena (RER C)
	Gare Val d'Europe (RER A)
	Gare Stade de France-Saint-Denis (RER D)
	T2
	Station Saint-Denis-Université (ligne 13)
1993 par rapport à 1996	Tunnel du RER D entre gare du Nord et Châtelet-les-Halles
	Gare Cergy-le-Haut
	Gare Neuville-Université
	Gare Roissy-Charles-de-Gaulle 2
1990 par rapport à 1993	TVM
	T1

Tableau 16 : Modifications de l'offre routière de 1990 à 1999

Puis nous avons calculé au niveau de chaque cellule les variables de transport qui apparaissent dans les deux modèles de localisation : variables d'accessibilités « Poulit » (les logsum), variables de nombres de gares et de stations de métro, temps d'accès à Châtelet en voiture, distances aux autoroutes et routes nationales. Des fichiers correspondant à chaque année (1993, 1996) sont créés au format texte. Nous ne mettons pas à jour la table « travel_data » qui normalement doit être générée automatiquement par UrbanSim 4 après chaque exécution de la partie Davisum-METROPOLIS (rappelons que cette table contient les logsum par mode).

5.3. Implémentation de l'interaction

Nous avons écrit un script Python qui permet à UrbanSim 4 de lire ces fichiers au cours de la simulation, dont voici le canevas :

```
'import': {
    'opus_core.model': 'Model',
    'paris.models.transport_model': 'TransportModel'
},
```

L'instruction d'import fait appel à la classe TransportModel, contenue dans le module transport_model du package paris.models. C'est elle qui implémente la logique et assure les échanges de données.

```
'init': {
    'name': 'TransportModel'
    'arguments': {
        'data_dir': getRepertoireDonneesExternes()
    },
},
```

On note que la fonction d'initialisation ne compte qu'un seul paramètre, qui est le répertoire où sera stockée la hiérarchie des données. Ce répertoire sera de la forme « UrbanSim_cache/paris/data ». Lors de l'initialisation du modèle, il cherchera à lire, dans ce répertoire, un fichier nommé « liste_vars.txt ».

Le contenu de ce fichier déterminera quelles variables le modèle va mettre à jour au cours de la simulation. Le format de ce fichier est de la forme suivante :

```
float gridcells.tvp
float gridcells.emploi_tc
int    gridcells.railway_stations
int    gridcells.subway_stations
```

Le premier champ donne le type de la donnée. Ce type servira lors de la conversion du format externe vers le format matriciel de NumPy. La chaîne fournie est convertie, grâce à l'évaluation dynamique de code, en type Python. Dans ce langage, comme en Java notamment, les types sont des objets à part entière, et donc facilement manipulables. Il était possible d'utiliser une table de hachage indexée sur les noms de types, mais cela aurait nécessité des vérifications supplémentaires ; nous laissons les exceptions de Python signaler toute erreur de format.

Ces exceptions seront interceptées et illustrées de messages d'erreur explicites. Des tests exhaustifs ont permis d'identifier les différents problèmes possibles, permettant ainsi d'aiguiller l'utilisateur en conséquence.

Lecture et mise à jour des données

```
'run': {
  'arguments': {
    'dataset_pool': 'dataset_pool',
    'simulated_year': 'year'
  }
}
```

La méthode de traitement possède, elle aussi, un argument unique. Ici, c'est l'année actuellement simulée qui est donnée comme information d'entrée au modèle. Cette valeur sera utilisée pour identifier le fichier contenant les données à importer.

Les fichiers sont organisés en arborescence. Le premier niveau indique l'année. Le nom du fichier donne la table et la variable concernées. On aura, en reprenant nos exemples, une hiérarchie de la forme :

```
data/1993/gridcells.tvp.txt
data/1993/gridcells.emploi_tc.txt
data/1993/gridcells.railway_stations.txt
data/1993/gridcells.subway_stations.txt
```

Puis :

```
data/1996/gridcells.tvp.txt
```

Et ainsi de suite...

Les informations sont ordonnées de façon croissante selon l'identifiant de la table considérée.

Un fichier absent signifie qu'une variable ne doit pas être mise à jour pour les années concernées.

Un message d'avertissement signalera alors ce fait, pour éviter toute mauvaise manipulation involontaire.

Le script s'appelle « transport_model.py » et se trouve dans le répertoire python24/site_packages/lib/paris/models.

6. Calage global de *SIMAU*RIF

6.1. Principes généraux

6.1.1. Le calage global

6.1.1.1. Un problème d'assemblage

Chacun des modules a été estimé indépendamment les uns des autres. A la limite, il serait possible de valoriser du point de vue de la recherche chaque estimation (voir en annexe 1 les résumés des articles publiés). Mais le but premier n'était pas de publier les résultats de chacun des modules. Il s'agit de réaliser un « calage global », tâche autrement plus difficile. En quoi consiste-t-il ? Dès juin 2003, lors de l'élaboration de la méthodologie du projet, nous avons eu l'intuition de l'importance de cette « phase de calage global ». Elle consiste, en pratique, en l'assemblage de tous les modules estimés et en l'exécution sur la période 1990-1999 de l'ensemble dans une boucle de rétroaction annuelle (1 an pour UrbanSim et 3 ans pour le modèle de trafic), sachant que les sorties d'un modèle une année peuvent être les entrées d'un autre modèle l'année suivante. L'objectif de ce calage global est de reconstituer la situation observée en 1999, c'est-à-dire la répartition géographique des populations et des emplois à un niveau infracommunal, le mode d'occupation du sol en trois postes agrégés et les prix moyen des logements au niveau communal.

6.1.1.2. L'interaction au sein d'UrbanSim

L'intérêt premier d'UrbanSim est l'interaction entre les différents modèles internes. On s'est demandé si cette interaction était déjà implémentée. On peut regretter que cet aspect *pourtant fondamental* soit très mal expliqué dans la documentation, voire pas du tout. C'est une approche empirique qui nous a permis de comprendre progressivement comment les sorties d'un modèle alimentent les entrées d'un autre et comment les variables de gridcells et les tables households et jobs sont mises à jour annuellement. Des points sont restés obscurs jusqu'au dernier moment, par exemple : en sortie du modèle de prix de l'immobilier résidentiel, on obtient chaque année le prix moyen d'un logement dans chaque cellule, qui va être écrite dans la variable `residential_land_value` de gridcell. Cette sortie est nécessaire pour calculer la variable « `vresidential_improvement_value` » qui intervient dans le modèle de localisation des ménages (`vresidential_units * residential_land_value`). Faut-il écrire une expression dans le fichier « `specification` » ou faut-il coder « en dur » ce produit dans un script Python ?

6.1.1.3. Une compétence nécessaire : la maîtrise de Python

Autant les phases précédentes du projet demandaient des compétences pluridisciplinaires (modélisation de trafic, économétrie, système d'information géographique, gestion de bases de données, connaissance du fonctionnement de la région Ile-de-France), autant il s'avère que cette phase de calage global dans le nouvel UrbanSim requiert principalement des compétences en programmation en langage Python, comme l'a montré le paragraphe 2.3.

6.1.2. Installation de la base « Paris »

L'installation de la base de données que nous avons appelée « Paris » (au même niveau que l'exemple « Eugène ») est organisée autour de quatre répertoires :

- Le premier, Python24/, contient l'ensemble du code de Python, et du système de simulation, incluant toutes les librairies, de Numpy jusqu'à Paris.
- Le second, UrbanSim_extra/, contient l'ensemble des documents concernant le développement de la simulation, des échantillons d'individus particuliers aux résultats des estimations.
- Dans UrbanSim_cache/ est contenu le cache initial de la simulation, c'est-à-dire les données de l'année initiale 1990 (le cache est plus rapide d'accès que la base MySQL d'UrbanSim 3 et supplée à l'insuffisance de mémoire vive RAM)
- Un dernier répertoire est utilisé, UrbanSim_output/, contenant les données de sortie de la simulation 1990-1999.

6.1.3. Principe d'appel d'une variable

Le principe d'appel d'une variable en Python est le suivant : le nom de chaque variable d'un modèle donné est défini dans la table « spécification » (par exemple la table « household_location_choice_model_specification »).

Il y a 3 cas possibles :

- 1) La variable est une **variable simple** définie au niveau d'une cellule, elle est donc directement disponible (par exemple dans la table gridcell). Il suffit d'appeler cette variable dans la table spécification. Exemple : paris.gridcell.emploi_tc
- 2) La variable est une **variable de voisinage**. Elle n'est pas en lecture directe dans une table. C'est dans la table UrbanSim_constants qu'est défini le nombre de cellules à prendre en compte autour d'une cellule. Par exemple, en prenant walking_distance=500, on considère les 8 cellules voisines autour de chaque cellule, sachant qu'une cellule a 500 m de côté. On doit alors écrire un script qui permet de faire le calcul de la somme de la variable pour les 9 cellules.

Exemple : le script « industrial_sqft_within_walking_distance.py » est écrit comme suit :

```
from UrbanSim.gridcell.abstract_within_walking_distance import
abstract_within_walking_distance

class
industrial_sqft_within_walking_distance(abstract_within_walking_distance):
    """Sum over c in cell.walking_radius, c.industrial_sqft."""
    _return_type = "int32"
    dependent_variable = "industrial_sqft"
```

Les scripts se trouvent dans le répertoire /python24/lib/site-packages/paris, et à l'intérieur dans différents sous-répertoires.

- 3) La variable est une **variable croisée**. Il faut aussi écrire un script pour calculer cette variable. Par exemple, dans le modèle de localisation des ménages, rappelons qu'il y a 3

variables croisées. Les variables croisées utilisées mettent en relation des cellules cibles potentielles avec les ménages considérés comme mobiles. Ce croisement génère une matrice, qui peut devenir très volumineuse si l'on prend en compte trop de possibilités d'implantation. Chaque ligne représente un ménage, et chaque élément de cette ligne représente une cellule à considérer. Néanmoins, une même cellule peut apparaître dans plusieurs lignes : c'est le tirage aléatoire qui déterminera quels choix seront considérés pour chaque agent.

Prenons pour exemple la variable ZUS_faible_revenu. Elle appartient à l'ensemble croisé gridcell par household, sa définition devra donc se trouver dans un répertoire de même nom, tel que households_x_gridcell, dans le répertoire paris. Elle dépend des données des deux tables parents :

```
def dependencies(self):  
    return ["paris.gridcell.is_sensible", "UrbanSim.household.is_low_income"]
```

Elle est formée à partir de deux variables booléennes, la résultante sera donc un indicateur.

```
def compute(self, dataset_pool):  
    return self.get_dataset().multiply("is_low_income", "is_sensible")
```

Le croisement s'effectue à l'aide de la méthode multiply de la classe InteractionDataset, et revient à utiliser un ET logique.

Au final, le script « low_income_if_sensible.py » s'écrit en Python comme suit :

```
from opus_core.variables.variable import Variable  
  
class low_income_if_sensible(Variable):  
    """Interaction between low income and sensible gridcell. """  
    _return_type="bool_"  
    pgc_is_sensible = "is_sensible"  
    hh_low_income = "is_low_income"  
  
    def dependencies(self):  
        return ["paris.gridcell."          + self.pgc_is_sensible,  
                "UrbanSim.household."     + self.hh_low_income]  
  
    def compute(self, dataset_pool):  
        return self.get_dataset().multiply( self.hh_low_income,  
                                             self.pgc_is_sensible)
```

Remarque : pour les variables de voisinage et croisées, il est possible aussi d'utiliser une expression, au sens d'UrbanSim 4. Ainsi, certaines variables ont été définies par des expressions.

Comme pour la structure de données, nous avons copié-collé les deux principaux scripts d'exécution de l'exemple Eugène :

- baseline.py, dans le répertoire python24/lib/site_packages/paris/configs, qui enchaîne les modules à exécuter
- run_simulation_cache.py, dans le répertoire /myworkspace/paris/tools.

6.2. Base de données d'entrée 1990

Comme il l'a été écrit dans le rapport [1], les bases de données d'entrée ont été construites à deux années : 1990 et 1999. La base 1999 sert à deux niveaux : d'abord à comparer les résultats de la simulation (calage) sur la période 1990 à 1999 avec la situation observée, ensuite à démarrer la simulation future sur la période 1999-2026.

La base de 1990 est la base de l'année de départ du calage.

Les 3 tables principales sont gridcells, households, jobs.

Pour gridcells, il n'y a pas eu de problème particulier pour les construire sur les deux années lors de la 1^{ère} phase du projet en 2004. Pour les deux autres, on a rencontré quelques problèmes pour construire les tables de 1990, que nous décrivons succinctement dans les paragraphes suivants.

6.2.1. La table des ménages households

Pour households, seule la table de 1999 avait été construite lors de la 1^{ère} phase (voir [1], page 92). Nous avons à cette époque reporté la construction de la table de 1990. On s'y est attelé cette année, mais nous avons rencontré un imprévu : le fichier exhaustif du recensement de 1990 est disponible à l'IAURIF mais l'attribut îlot n'est pas renseigné, contrairement à 1999. On n'a donc pas pu affecter chaque ménage de 1990 à sa cellule carrée. Il a fallu procéder à une méthode rudimentaire de distribution des ménages en fonction de la densité de population qui est un des attributs de la couche des îlots MOS du SIGR, puis par croisement entre les îlots MOS et le maillage, on a affecté chaque ménage à une cellule.

La table households 1990 contient 4 232 526 ménages (10 952 000 individus), avec quatre attributs :

- le nombre d'individus dans le ménage
- le nombre d'actifs dans le ménage
- le nombre de voitures dans le ménage
- la cellule du lieu de résidence

6.2.2. La table des emplois jobs

En 2004, on avait rapidement construit des tables emplois 1990 et 1999 (voir [1] page 93) en utilisant une règle d'affectation simpliste : à partir du fichier du nombre d'emplois à la commune, les emplois ont été répartis sur les cellules de chaque commune en fonction de leur surface d'activité. Plus précisément, on affectait par cellule les emplois tertiaires sur les postes du MOS de type commerces et bureaux, et les autres emplois (agriculture, industrie, construction) sur les postes activités secondaires du MOS. On a ajouté à la fin les emplois à domicile.

On a affiné la méthode en utilisant à l'origine des données infracommunales et non des données communales. L'IAURIF dispose de la couche des emplois à l'îlot INSEE en 16 postes sur Paris, la petite couronne et les communes importantes de la grande couronne en 1990. En revanche, cette couche n'existe pas en 1999, l'INSEE n'ayant pas réalisé de travail d'îlotage des emplois comme elle l'avait fait en 1990.

Pour 1990, on a donc croisé la couche des îlots INSEE avec le maillage puis généré un fichier d'emplois sur l'Île-de-France, chaque emploi étant renseigné par son secteur d'activité en 8 postes (on a utilisé la correspondance 16->8, voir tableau 4). On a pu ainsi localiser dans chaque cellule les emplois de Paris, de la petite couronne et de quelques communes de la grande couronne. Pour le reste de la grande couronne, on a utilisé le fichier d'emplois à la commune de l'INSEE en 1990 en 8 secteurs. Pour passer de la commune à la cellule carrée, on a utilisé les 6 postes de type activité du MOS dans la nomenclature en 21 postes :

- poste 12 : activités secondaires
- poste 13 : activités tertiaires
- poste 15 : équipement d'enseignement
- poste 16 : équipement de santé
- poste 18 : autres équipements locaux, administration

- poste 19 : grands équipements

La correspondance entre secteurs d'emplois et postes du MOS n'est pas parfaite, on a opté pour la suivante :

Secteurs	MOS 21
Agriculture et industrie	12
Energie, commerce, finances, services, administration	13+16+18+19
Education	15

Tableau 17 : Correspondances entre secteurs et MOS en 21 postes

Puis la part d'emploi par secteur à la commune est affectée à la cellule comprise dans la commune.

En revanche, pour 1999, on n'a pas pu faire mieux qu'en partant d'un fichier d'emplois à la commune.

La table jobs 1990 contient 5 075 974 emplois, avec cinq attributs :

- numéro de l'emploi
- numéro de la cellule
- home_based (toujours à 0)
- secteur_id (de 1 à 8)
- building_type (toujours à 1)

6.2.3. La table « development_event_history »

Cette table est nouvelle par rapport à UrbanSim 2 et ne doit pas être confondue avec la table « development_events_exogenous » (voir paragraphe suivant).

Cette table est indispensable au nouveau modèle de développement urbain d'UrbanSim 4. Elle est utilisée par le « developer project transition model ». Dans la documentation (voir [5], page 136), la structure de cette table est décrite. Nous l'avons adaptée en regroupant les 3 activités en une seule, en pensant que la plateforme UrbanSim 4 pourrait gérer les nouveaux noms de variables. Or, a posteriori, nous avons découvert que les variables de la table étaient codées « en dur » dans les programmes scripts Python d'UrbanSim ! Autrement dit, nous avons dû entrer dans les scripts et renommer une variable d'activité, à chaque fois qu'elle apparaissait, par son nouveau nom « activity ». Comme programme soit disant « ouvert et flexible », cela laisse vraiment à désirer ... Encore une fois, c'est l'accumulation de ces problèmes informatiques qui a occasionné des pertes de temps considérables dans notre projet.

Les variables de la table sont donc les suivantes :

- grid_id
- scheduled_year
- starting_development_type_id
- residential_units
- activity_sqft
- residential_improvement_value
- activity_mprovement_value

Le contenu de cette table est la base de données de calage du modèle d'évolution du MOS (voir [2] page 51). On considère qu'il y a deux types de « projet », dans le sens d'UrbanSim 4 : la transition d'une cellule vacante vers le type habitat et la transition d'une cellule vacante vers le

type activité. La table contient un historique de 1029 observations (647 de type 1, 382 de type 2). Pour les 4 dernières variables de la table, on a récupéré les valeurs correspondantes dans la table gridcell de 1999.

6.2.4. La table des « events »

Rappelons qu'il existe un moyen dans UrbanSim de prendre en compte des événements exceptionnels qu'aucun modèle, aussi sophistiqué soit il, ne peut prédire (voir [1] pages 49 et 98). Il suffit de définir dans les fichiers de type « exogenous events » les nombres d'emplois, de ménages, le changement de type urbain et l'année pour les cellules où se produisent ces événements.

C'est un travail assez lourd et fastidieux. Il faut par exemple recenser sur la période 1990-1999 toutes les entreprises de plus de 1000 salariés qui ont déménagé ou qui ont été créées. Puis il faut localiser la cellule d'emménagement à partir de l'adresse postale. La table « employment_events » a ainsi été construite avec les données suivantes :

ENTREPRISE	COMMUNE	DATE CREATION	EFFECTIF 2005
CALYON	COURBEBOIE	1997	2000
SAMSI SECURITE GESTION	PARIS	1991	1319
CAISSE EPARGNE PREV IDF	PARIS	1991	1000
THALES COM SA	COLOMBES	1991	3700
TECHNIP France	COURBEBOIE	1993	1450
CS SYSTEME INFORMATION	LE PLESSIS ROBINSON	1993	1300
JOHNSON CONTROLS AUTOMOTIVE	OSNY	1996	1500
BOUYGUES TP	GUYANCOURT	1996	1100
SIEMENS VDO AUTOMOTIVE	RAMBOUILLET	1996	1000
EQUANT France SA	PARIS	1996	1500
RENAULT FLINS	BOULOGNE-BILLANCOURT	1996	5800
RENAULT SANDOUVILLE	BOULOGNE-BILLANCOURT	1996	6000
RENAULT DOUAI	BOULOGNE-BILLANCOURT	1996	6000
NOVARTIS PHARMA SAS	RUEIL-MALMAISON	1996	1200
EADS SECURE NETWORKS	GUYANCOURT	1998	1100
RAIL RESTAURATION	PARIS	1997	1135
LOGICACMG	BOULOGNE-BILLANCOURT	1997	1700
NEUF CEGETEL	BOULOGNE-BILLANCOURT	1997	1500
CEGELEC HOLDINGS	NANTERRE	1998	1173
ALTIS SEMICONDUCTEURS	CORBEIL ESSONNE	1999	2116

Tableau 18 : Entreprises de plus de 1000 salariés ayant déménagé ou ayant été créées entre 1990 et 1999

Nous n'avons pas construit les 2 autres tables de type « event ».

6.3. Les tables de type « control totals »

Actuellement, les prévisions démographiques et d'emplois sont effectuées de manière exogène, il n'y a pas dans UrbanSim de modèle démographique. Ceci nécessite de faire des projections dans le futur pour toute la période à simuler par l'intermédiaire de tables dites « control totals ». Ces tables donnent le nombre **total** annuel de ménages par type et d'emplois par secteur en Ile-de-France pour chaque année de la période de calage (voir [1], page 95). Les chiffres sont connus pour 1999 grâce au Recensement général de la population. Pour les années 1991 à 1998,

nous avons deux points fixes pour les emplois grâce à l'enquête régionale emplois de 1993 et 1997. Pour la population, l'INSEE a fourni des estimations de population pour chaque année. Au final, nous avons construit ces deux tables en utilisant les points fixes et par interpolation linéaire.

Un autre aspect important à développer est d'intégrer un modèle démographique dans tout le système.

6.3.1. Table « ANNUAL HOUSEHOLDS CONTROL TOTALS »

YEAR	TOTAL_NUMBER_OF_HOUSEHOLDS
1990	4 232 526
1991	4 262 532
1992	4 292 751
1993	4 323 185
1994	4 353 834
1995	4 384 700
1996	4 415 785
1997	4 447 091
1998	4 478 618
1999	4 510 369

Tableau 19 : Nombre de ménages entre 1990 et 1999

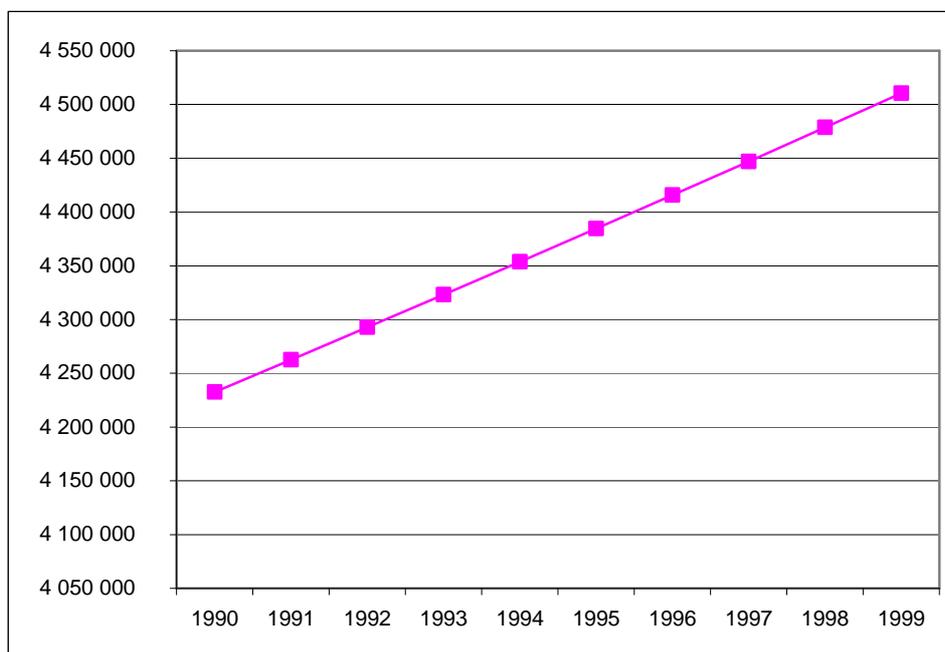


Figure 8 : Nombre de ménages entre 1990 et 1999

6.3.2. Table « ANNUAL_EMPLOYMENT_CONTROL_TOTALS »

YEAR	Agriculture	Industrie	Energie	Commerce	Finances	Services	Education	Administration	Total
1990	95 254	849 496	818 171	618 898	271 188	1 439 506	639 250	344 211	5 075 974
1991	93 349	815 516	785 444	621 374	282 036	1 452 462	645 003	368 306	5 063 489
1992	90 548	758 430	738 318	619 509	284 856	1 448 104	641 133	383 038	4 963 937
1993	86 927	690 171	694 018	612 075	276 310	1 416 246	628 951	398 360	4 803 059
1994	85 188	641 859	645 437	607 179	273 547	1 400 667	623 920	410 311	4 688 108
1995	83 484	609 766	619 620	613 251	276 283	1 406 270	642 637	441 084	4 692 395
1996	82 649	585 376	601 031	618 770	287 334	1 407 535	649 064	478 576	4 710 335
1997	81 823	567 814	595 021	624 339	301 701	1 408 943	681 517	521 648	4 782 805
1998	81 005	556 458	589 071	636 825	319 803	1 417 115	721 045	563 380	4 884 701
1999	81 989	548 711	579 006	655 915	358 882	1 431 842	781 069	604 581	5 041 995

Tableau 20 : Nombre d'emplois entre 1990 et 1999 par secteur (estimation IAURIF)

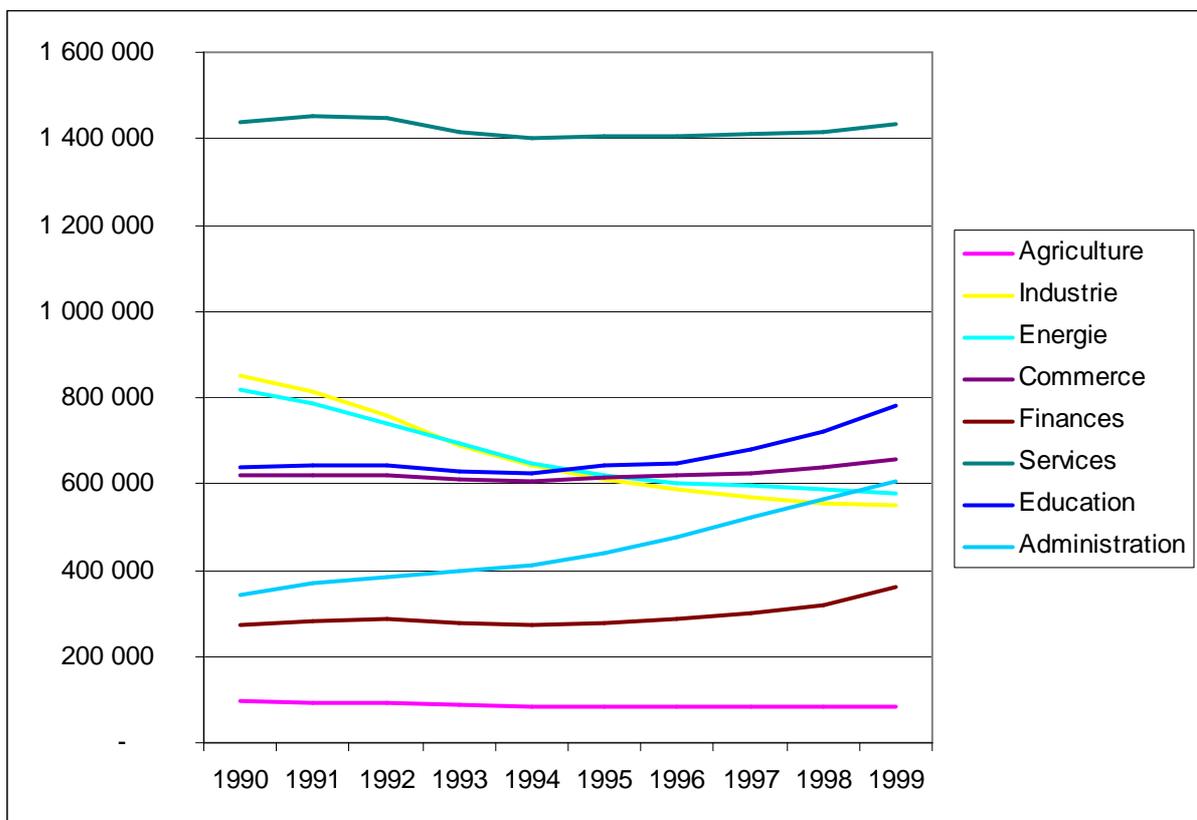


Figure 9 : Evolution des emplois entre 1990 et 1999 par secteur (estimation IAURIF)

6.4. Exécution de SIMAURIF de 1990 à 1999

On trouvera ci-dessous le canevas d'un script Python de calage global entre 1990 et 1999 dans UrbanSim 4 (extrait de notre fichier baseline.py) qui enchaîne les différents modules à exécuter : les 4 modules d'UrbanSim et le modèle de trafic réduit à sa plus simple expression (script de prise en compte de l'offre de transport).

```
'base_year': 1990,
```

```

        'years': (1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998,
1999),
        'models': [
            # mise a jour des variables d'accessibilité
            'transport_model',
            # -- pas necessaire si pas de modeles de developpement,
            # mais initialisent la liste des cellules modifiees (pour le
            residential_land_share_model)
            'prescheduled_events',
            'events_coordinator',
            # active le modèle LandPriceModel qui doit prédire la valeur foncière
            # des espaces résidentiels et des espaces d'activités
            'land_price_model',

            # active le DevelopmentProjectTransitionModel qui crée des projets de
            # différents types
            #'development_project_transition_model',

            # active le DevelopmentLocationChoiceModel pour les projets d'habitat.
            # Il place les projets d'habitat nouvellement créés dans les cellules
            # susceptibles d'en accueillir
            #'residential_development_project_location_choice_model',

            # active le DevelopmentLocationChoiceModel pour les projets de centres
            # commerciaux. Il place les projets de centres commerciaux nouvellement
            # créés dans les cellules susceptibles d'en accueillir
            #'commercial_development_project_location_choice_model',

            # active le DevelopmentLocationChoiceModel pour les projets
            # d'activités. Il place les projets d'activités nouvellement créés dans
            # les cellules susceptibles d'en accueillir
            #'industrial_development_project_location_choice_model',

            # on a adapté notre modèle de transition du MOS pour remplacer les
            # quatre précédents : ce modèle est propre à notre étude et remplace les
            # 3 modules précédents
            'evolution_mos_model',

            # active le DevelopmentEventTransitionModel qui créé des événements
            # urbains à partir des projets donnés
            'development_event_transition_model',

            # active le EventsCoordinator qui met à jour les développements
            # annuels
            'events_coordinator',

            # active le ResidentialLandShareModel qui prédit la part de surface
            # résidentielle
            'residential_land_share_model',

            # active le HouseholdTransitionModel qui simule les naissances et les
            # décès dans la population des ménages : on ne l'utilise pas
            #'household_transition_model',

            # active le EmploymentTransitionModel qui simule les créations et les
            # disparitions d'emplois : on ne l'utilise pas
            #'employment_transition_model',

```

```

# active le AgentRelocationModel où les agents sont des ménages. Il
détérmine quels ménages vont déménager l'année en cours.
    'household_relocation_model',

# active le AgentLocationChoiceModel où les agents sont des ménages.
Il place les ménages en train de déménager vers d'autres cellules.
    'household_location_choice_model',

# active le AgentRelocationModel où les agents sont des emplois. Il
détérmine quels emplois vont déménager l'année en cours.
    'employment_relocation_model',

# active le AgentLocationChoiceModelMember où les agents sont des
emplois qui sont contenus dans la table 'job_building_types'. Pour
chaque type d'emplois, il place les emplois en cours de localisation
dans d'autres cellules. Une exception est faite pour le type
« bureaux » pour lequel le ScalingJobsModel est activé.
    {'employment_location_choice_model':{
'group_members': '_all_'}}},

# active le DistributeUnplacedJobsModel qui place de manière aléatoire
les emplois pas encore localisés dans d'autres cellules
    'distribute_unplaced_jobs_model'
],

```

La configuration de la machine est de 2 Go de RAM.

Le temps d'exécution d'une simulation de 1990 à 1999 est d'environ 7 heures.

6.5. Résultats de l'exécution sur la période 1990-1999

Le résultat principal qui nous intéresse est la répartition géographique des ménages et des emplois en 1999. Ce sont les fichiers households et jobs produits par UrbanSim en fin de simulation (année 1999), à ne pas confondre avec les fichiers households et jobs qui ont été construits pour être utilisés à l'année de départ de la simulation 1999-2026. Viennent ensuite l'évolution des prix de l'immobilier et le mode d'occupation du sol en postes agrégés (urbain, rural, vacant) en 1999.

Les résultats de chaque modèle sont différents de ceux qui sont produits lorsque l'on fait tourner chaque modèle seul sur les données de calage. En effet, dans le cadre d'une exécution de l'ensemble des modules, il y a d'une part une interaction entre les modèles car les sorties des uns sont les entrées des autres, d'autre part il y a une boucle dans le temps.

OPUS/UrbanSim 4 offre trois possibilités de calculs d'indicateurs :

- un ensemble de scripts permettant le calcul d'indicateurs. Un exemple peut être trouvé dans le répertoire opus_core.indicator_framework.make_indicators_example. Une page HTML de résultats appelée « indicator_results.html » est générée automatiquement.
- L'utilisation de l'interface GUI dotée d'un éditeur graphique pour spécifier et générer un indicateur. Pour l'ouvrir, il faut lancer le script /paris/indicators/indicator_gui.py
- une troisième méthode consiste à exporter le cache résultant d'une simulation vers une base de données. Il est alors possible d'opérer des requêtes plus complexes sur l'ensemble des attributs primaires.

Comme l'avait indiqué le Pr Waddell au commencement de notre recherche en 2004 (voir [1] page 49⁶), il ne faut pas s'attendre à des résultats extraordinaires.

Les tables households, jobs et gridcells sont mises à jour chaque année par UrbanSim. Les résultats de la dernière année (1999) sortis au niveau cellulaire sont agrégés au niveau départemental pour les deux modèles de localisation et à la commune pour le modèle d'évolution du MOS et le modèle de prix de l'immobilier. Une bonne reconstitution au niveau départemental ou communal ne signifie pas une bonne reconstitution au niveau cellulaire, mais une mauvaise reconstitution à un niveau agrégé condamne la validité du calage.

Le recalage a consisté à ajouter aux fonctions d'utilité une constante d'ajustement par département afin d'égaliser les totaux par département (voir paragraphe 6.6) Ces constantes d'ajustement seront conservées lors de l'étape de simulation à l'horizon futur. Nous présentons ci-après les résultats de la simulation sur la période 1990-1999 **avant recalage par ces constantes**.

6.5.1. Le modèle de choix de localisation des ménages

Dép.	Observé (INSEE)		Simulé (SIMAURIF)		Différence en 1999	
	1990	1999	1990	1999	absolue	relative
75	1 095 108	1 110 912	1 094 871	1 288 454	177 542	16%
92	590 946	624 926	590 853	646 750	21 824	3%
93	508 631	524 387	508 580	549 893	25 506	5%
94	471 503	499 404	471 385	501 416	2 012	0%
77	369 604	432 351	369 687	386 386	- 45 965	-11%
78	459 396	503 096	459 049	455 824	- 47 272	-9%
91	379 272	420 603	379 542	394 751	- 25 852	-6%
95	358 231	394 690	358 559	359 558	- 35 132	-9%
Région	4 232 691	4 510 369	4 232 526	4 583 032	72 663	2%

Tableau 21 : Nombre de ménages par département

Les différences en 1999 sont faibles dans la petite couronne. En grande couronne, elles sont déjà plus élevées, de l'ordre de 9%, avec une sous-estimation systématique. Le département où la simulation donne un résultat très moyen est la capitale, même si l'évolution va dans le bon sens (accroissement du nombre de ménages entre 1990 et 1999).

⁶ « Ainsi, l'exemple de l'application Eugène-Springfield montre que la dynamique de l'urbanisation (localisation des ménages et des activités) n'est pas simple à reconstituer, sur un territoire 70 fois plus petit que la Région d'Ile-de-France. On peut donc nourrir de petites inquiétudes et s'attendre à être confronté à de réelles difficultés de calage au cours de l'année 2 de notre recherche. » On avait écrit cela en septembre 2004 ...

6.5.2. Le modèle de choix de localisation des emplois

Dép.	Observé (INSEE)		Simulé (SIMAURIF)		Différence en 1999	
	1990	1999	1990	1999	absolue	relative
75	1 777 679	1 600 815	1 777 679	1 671 814	70 999	4%
92	770 370	815 471	770 370	757 496	- 57 975	-7%
93	480 675	483 999	480 675	496 403	12 404	3%
94	465 236	472 447	465 236	478 727	6 280	1%
77	393 069	388 947	393 069	407 228	18 281	5%
78	494 002	504 469	494 002	511 785	7 316	1%
91	328 262	401 397	328 262	339 527	- 61 870	-15%
95	366 681	374 450	366 681	379 015	4 565	1%
Région	5 075 974	5 041 995	5 075 974	5 041 995	-	0%

Tableau 22 : Nombre d'emplois par département

Globalement, les résultats sont tout à fait acceptables, sauf peut-être dans l'Essonne.

6.5.3. Le modèle d'évolution du MOS

En ce qui concerne la transition vers l'habitat, nous présentons comme indicateur le nombre de logements par département.

	1990	1999	Différence relative
75	1 323 919	1 339 435	1%
77	421 323	459 357	9%
78	504 236	549 629	9%
91	413 215	451 687	9%
92	647 672	698 148	8%
93	550 030	605 877	10%
94	500 558	551 603	10%
95	385 388	422 661	10%

Tableau 23 : Nombre de logements par département

SIMAURIF surestime une augmentation du nombre de logements dans tous les départements. Cette évolution est plus importante dans les départements de la Seine-Saint-Denis, du Val-de-Marne et du Val-d'Oise.

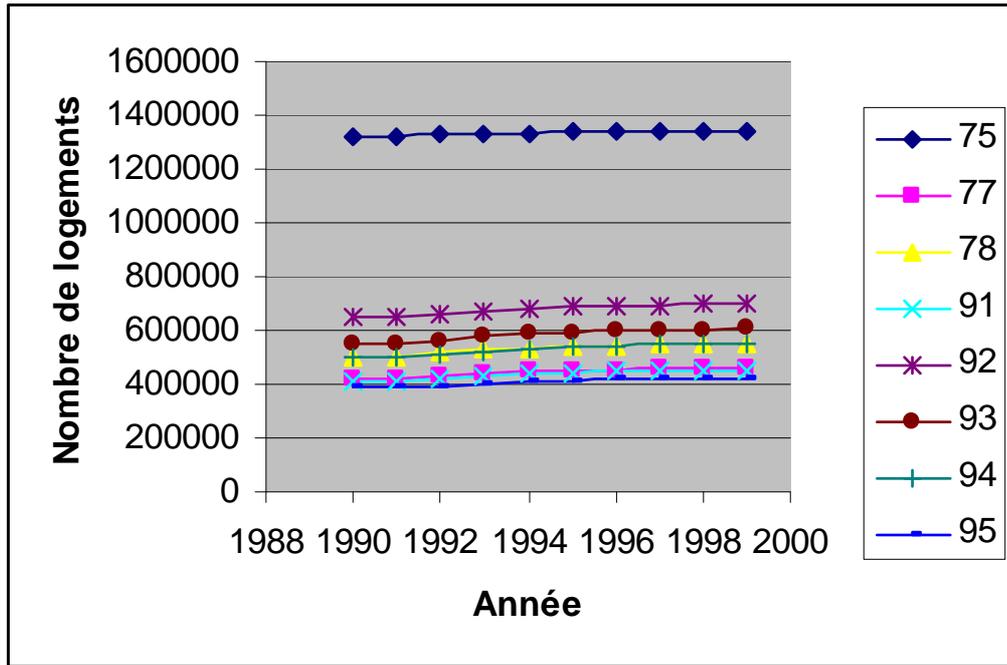


Figure 10 : Evolution du nombre de logements par département simulé par SIMAURIF

6.5.4. Le modèle de prix

Le graphique suivant montre la distribution des écarts relatifs sur les 1300 communes d'Ile-de-France entre le prix moyen d'un logement simulé par SIMAURIF (sans distinction entre appartement et maison et sans distinction de l'ancienneté) et le prix moyen observé dans la base de données communales des Notaires en 1999.

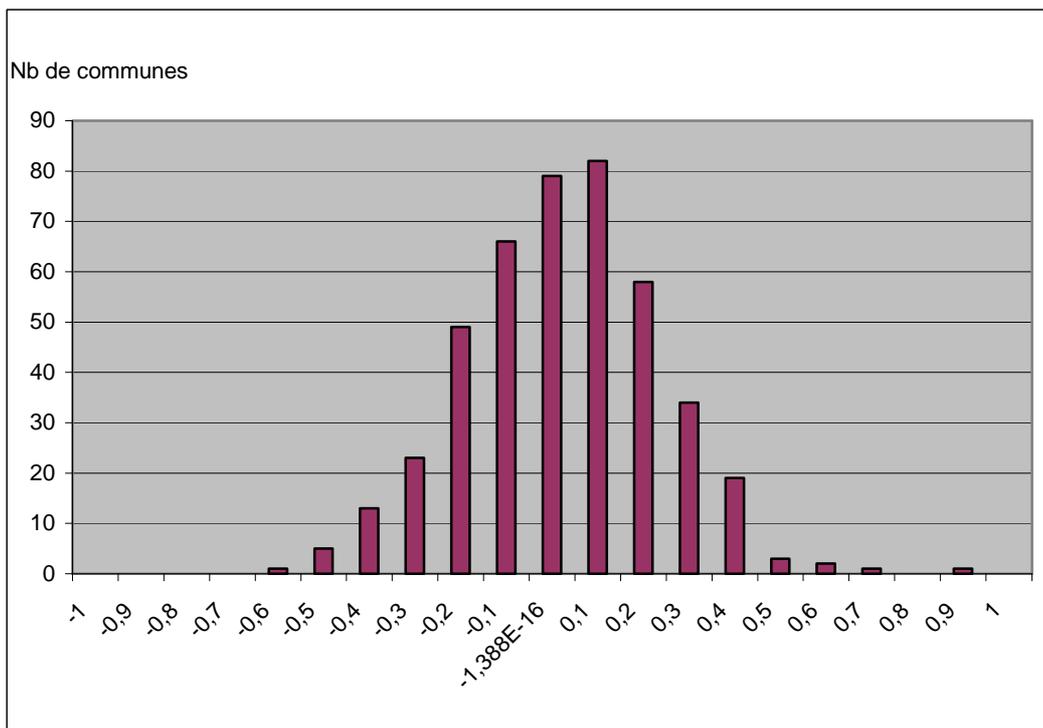


Figure 11 : Distribution des écarts relatifs entre prix simulés et prix observés sur les 1300 communes

Ce graphique montre des résultats tout à fait acceptables.

6.6. Résultats de calage

Nous calons le modèle au niveau départemental. Après analyse des premiers résultats, nous avons décidé de rajouter des constantes pour les indicatrices département pour favoriser ou défavoriser tel ou tel département. Par exemple, si nous observons une surestimation de 4% du nombre de ménages qui se localisent à Paris, nous ajoutons un coefficient négatif pour l'indicateur Paris, ce qui réduira l'attractivité de Paris, en d'autres termes, l'utilité calculée par le modèle de localisation des ménages sera moindre.

Nous avons procédé de cette manière pour les 4 modèles.

Nous présentons ci-après le résultat du calage pour le modèle de localisation des ménages, obtenu après une trentaine de simulations.

	constante	Pop 1999	Différence (simulé-observé)	Ecart relatif
75	-55	1 100 623	-10289	-1%
77	90	418 100	-14251	-3%
78	110	499 005	-4091	-1%
91	0	420 269	-334	0%
92	-17	669 381	44455	7%
93	-125	563 221	38834	7%
94	-125	512 252	12848	3%
95	100	398 528	3838	1%
Ménage > 2 membres	-30			

Tableau 24 Calage du modèle de localisation des ménages

(la dernière indicatrice a été introduite pour réduire le nombre de ménages excessif qui se localisent à Paris).

Avec cette méthode empirique, nous avons pu améliorer le résultat en passant d'une erreur globale de 17% à seulement 7%.

Il est possible d'améliorer encore les résultats mais pour des raisons de contrainte de délais, nous avons décidé de rester à ce niveau de précision.

7. Préparation des bases de données à l'horizon 2026

On prépare une base de données de référence (scénario de base) pour les simulations sur la période 1999-2026.

Au-delà de 2007, il faut faire des hypothèses sur la date de mise en service des infrastructures de transport et sur l'évolution démographique et économique de la Région. Pour la période 1999-2008, nous l'avons déjà vécu, les tendances et les évolutions sont connues. Nous en faisons quelques brefs rappels dans la partie 7.1.

7.1. Ce qui s'est passé entre 1999 et 2006

La source de données est l'enquête annuelle de recensement de 2004, 2005 et 2006 de l'INSEE, et les couches MOS de 1999 et 2003⁷ du Système d'information géographique régional.

7.1.1. Population

Rappelons qu'entre 1990 et 1999, le taux de croissance du nombre de ménages franciliens (+0,67 % par an) apparaît deux fois plus rapide que celui de la population (+0,30 %). Cette croissance plus rapide du nombre de ménages que de la population s'est poursuivie au cours des années suivantes. La Région d'Ile-de-France a gagné, entre 1999 et 2006, plus de 530 000 habitants pour atteindre 11,49 millions contre 10,725 millions en 1999.

Le nombre de logements en Ile-de-France est estimé à 5,2 millions à la date du 1er juillet 2004. Les résidences principales constituent 91 % du parc de logements. Celui-ci a augmenté d'environ 30 000 par an, soit 0,6% depuis 1999.

Depuis 1999, la population progresse fortement dans la petite couronne, notamment grâce au dynamisme des communes situées à la périphérie de la capitale. Paris a stoppé la baisse de sa population observée entre 1990 et 1999 (-30000) et affiche même une légère croissance de sa population (+28000). Le département des Hauts-de-Seine enregistre la hausse démographique la plus importante de la région.

En grande couronne, la population croît à peu près au même rythme qu'à la fin des années 1990. La Seine-et-Marne se distingue, avec en moyenne, une croissance de sa population de 0,9 % par an.

Le phénomène de desserrement de la population se poursuit mais à un rythme plus lent. En 2005, la grande couronne représente 44 % de la population régionale, 37 % pour la petite couronne et 19 % pour Paris, comme en 1999. Trente ans auparavant, la répartition montrait un poids plus faible pour la grande couronne (37 %) contre 40 % pour la petite couronne et 23 % pour Paris (voir tableau 25). On observe des mouvements de retour progressif vers Paris des ménages monoparentaux, des seniors, des jeunes étudiants, mais aussi des ménages cherchant un environnement scolaire rassurant à partir du collège.

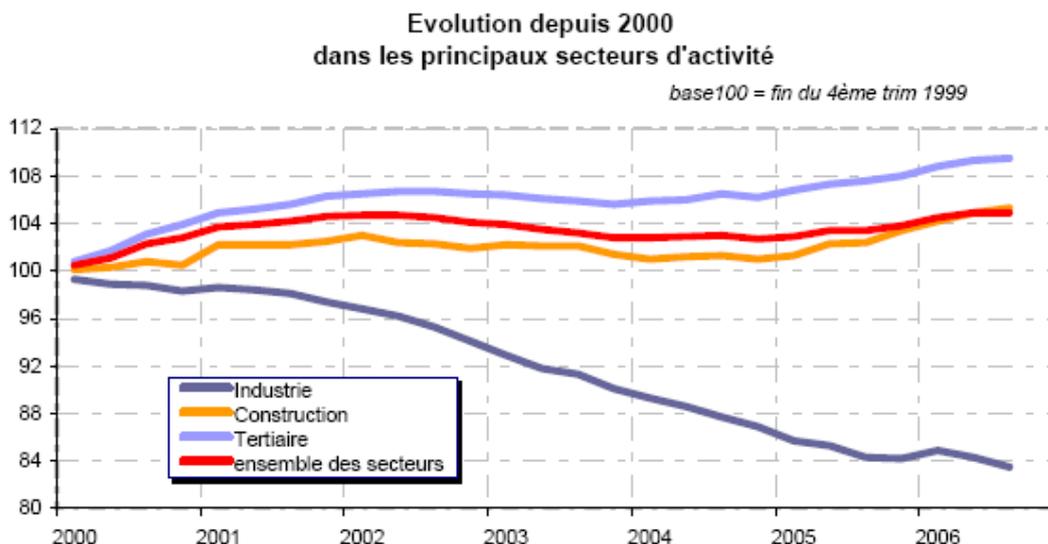
⁷ La couche MOS 2008 n'est pas encore disponible.

	1975		1999		2005		Différence 2005/1999
75	2 300	23%	2 125	19%	2 154	19%	28
92	1 439	15%	1 429	13%	1 517	13%	88
93	1 322	13%	1 383	13%	1 459	13%	76
94	1 216	12%	1 227	11%	1 279	11%	52
77	756	8%	1 194	11%	1 261	11%	67
78	1 082	11%	1 354	12%	1 395	12%	41
91	923	9%	1 134	10%	1 188	10%	54
95	841	9%	1 106	10%	1 148	10%	43
Région	9 879	100%	10 952	100%	11 399	100%	447

Tableau 25 : Evolution de la population (en milliers) par département entre 1975 et 2005 (source INSEE)

7.1.2. Emplois

Comme le montre la figure 12, l'emploi de type industriel continue à baisser fortement (ces emplois sont robotisés ou délocalisés dans les pays à faible coût du travail). C'est l'emploi tertiaire qui progresse nettement. Ces tendances, observées entre 1990 et 1999, se poursuivent donc.



Source : Unédic, données cvs, hors agriculture

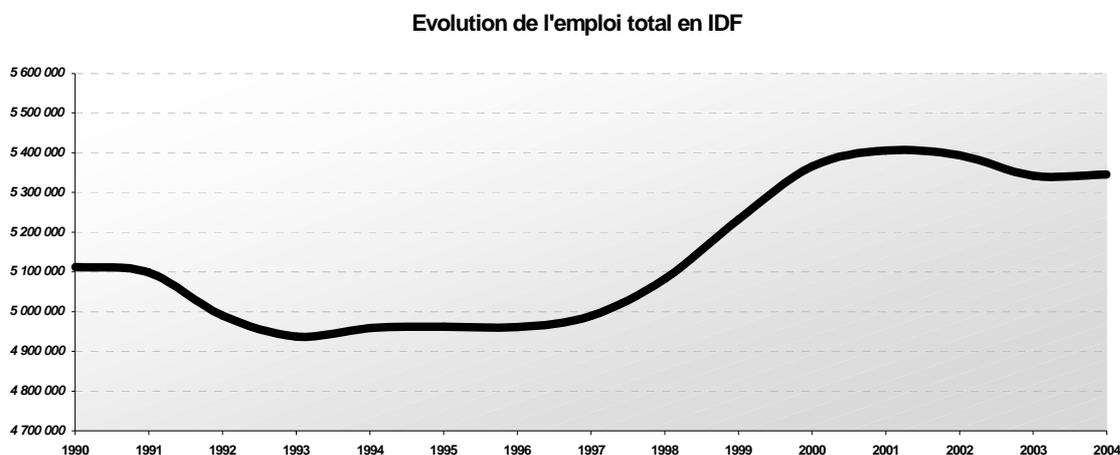
Figure 12 : Extrait de la note « Ile-de-France- Faits et chiffres, n°140 – janvier 2007 » de l'INSEE

	En 2005								IdF
	75	77	78	91	92	93	94	95	
Agriculture	1 372	6 428	3 779	2 630	865	503	892	2 335	18 804
Industrie	101 712	62 121	102 342	51 169	106 328	55 486	37 602	51 585	568 345
Construction	31 330	31 633	31 007	29 077	32 664	38 127	32 683	24 780	251 301
Commerce	171 394	67 317	74 808	68 121	94 961	86 363	75 774	62 016	700 754
Services	1 347 743	269 067	324 283	281 650	638 957	367 140	363 490	285 109	3 877 439
Total	1 653 551	436 566	536 219	432 647	873 775	547 619	510 441	425 825	5 416 643

Tableau 26 : Répartition de l'emploi en 2005 par secteur et par département

Cependant, entre 1999 et 2006, on a constaté un ralentissement du desserrement de l'activité et un recentrage de la construction de bureaux sur la zone dense.

Globalement, on a assisté à une phase de stabilité de l'emploi depuis 2001 qui a succédé à une forte croissance (1997-2000). La répartition de l'emploi est stable sur le territoire francilien depuis 2000 (Paris 31,2% / PC 35,1% / GC 33,7%). De nouveaux besoins (emploi, formation) s'expriment dans les services à la personne.



Source: INSEE, estimations annuelles d'emploi au 31 décembre

Figure 13 : Evolution de l'emploi total de 1990 à 2004

7.1.3. Occupation du sol

La comparaison des MOS 1999 et 2003 a fait ressortir les éléments principaux suivants :

- Le rythme d'urbanisation s'est ralenti entre 1999 et 2003 par rapport à la période 1990-1999. L'urbanisation est mesurée par la consommation des espaces naturels et agricoles (donc hors chantiers et terrains urbains vacants). Cette mesure montre un sensible ralentissement de ces espaces entre 1999 et 2003. Les postes les plus touchés sont l'habitat individuel et les jardins/parcs/terrains de sports/espaces récréatifs ouverts (moins de création de golfs), mais les autres postes de type urbain sont également concernés par ce ralentissement global : habitat collectif, équipements, activités. Seul le poste transport n'a pas ralenti, à cause notamment de la poursuite de deux grands chantiers, le TGV Est et Roissy.
- L'urbanisation nouvelle se situe essentiellement le long de la Francilienne et au-delà.
- Les Villes nouvelles polarisent relativement moins qu'avant cette urbanisation.
- Les bourgs et les villages reçoivent une part importante de la croissance générale.

7.1.4. Prix de l'immobilier résidentiel

Entre 1999 et 2007, les prix de l'immobilier en Ile-de-France ont augmenté chaque année. Dans la capitale, l'augmentation des prix a atteint 12,5% en 2006. Les évolutions ont été à peu similaires en banlieue. En 2005, les prix ont même dépassé les records atteints en 1991. Les écarts de prix entre communes continuent à s'accroître.

Mais une étude récente du cabinet d'étude Bipe montre que les prix de l'immobilier résidentiel dans l'ancien en France s'orientent vers une baisse de 3% en 2008, après avoir conclu l'année 2007 sur une hausse de 2%. Pour un autre cabinet d'analyses, Precepta, la baisse pourrait même

atteindre 18% d'ici à 2010. Ces analyses sont confirmées par une étude de la HSBC (novembre 2007) qui parle d'un "ajustement du marché" : le marché décélère mais on n'anticipe pas de ralentissement brutal. Cette étude annonce comme le BIPE que les prix devraient baisser de 3 % au premier semestre 2008.

En 2007, quelques grandes tendances sont intéressantes à relever : les stocks de logements neufs continuent de croître, les mises en chantier reprennent "mais les perspectives se dégradent avec la moindre capacité d'achat des ménages", peut-on lire dans l'étude. Des ménages qui se remettront avec difficulté de la poursuite de la hausse des taux d'emprunts, d'ici la fin de l'année.

Ainsi, parmi les facteurs susceptibles d'entraîner le marché à la baisse et qui ne sont pas pris en compte dans notre modèle, il y a la hausse continue des taux d'intérêt depuis la fin 2005, qui est l'une des causes du tassement de la production de crédit immobilier constaté en 2007, mais aussi la crise des « sub-primes » partis des Etats-Unis en 2007.

Élément supplémentaire pouvant contribuer à une baisse, l'allongement de la durée moyenne des emprunts "atteint visiblement ses limites", selon l'étude du Bipe, ce qui ne permet plus de compenser une éventuelle hausse des prix.

Les prix les plus élevés persistent dans Paris, où le m² vaut en moyenne 5 970 euros, soit une hausse de 8,3% sur un an. Dans le 7^{ème} arrondissement, le plus cher de la capitale, le prix moyen atteint 8 829 euros le m² (+9,3%), selon l'indice Notaires/Insee.

7.2. Les réseaux entre 1999 et 2026

7.2.1. Réseau routier

On supprime du graphe de 2026 des arcs pour construire les réseaux routiers tous les 3 ans (voir [2], page 23). Au début du projet, en 2003, on avait envisagé de prendre des périodes de 5 ans (voir [1], page 108) mais finalement on a opté pour 3 ans. Rappelons que le graphe du réseau routier de l'IAURIF comporte environ 16 000 arcs monodirectionnels et 5000 nœuds.

Il est inutile de recenser les projets d'élargissement car on ne tient pas compte de la congestion. Le réseau de référence chaque année est un réseau a minima compte tenu des grands principes actuels sur la limitation de la réduction du trafic routier.

2023 par rapport à 2026	A184 (Francilienne ouest)
	Liaison RN2-RN3
2020 par rapport à 2023	Tunnel RN13 à Neuilly-sur-Seine
	Prolongement du BIP est jusqu'à la RN1
2017 par rapport à 2020	Prolongement d'A12
	C5 (contournement nord de Melun)
2014 par rapport à 2017	Prolongement d'A16 jusqu'à la Francilienne
	Boulevard urbain de la Vallée de la Mauldre
	Rocade RN36 en Seine-et-Marne
2011 par rapport à 2014	A86 ouest (section autorisée aux poids-lourds)
	Liaison Meaux-Roissy
2008 par rapport à 2011	Liaison A4-RN4

	Echangeur A14-RN13
2005 par rapport à 2008	A86 ouest (tunnel)
	RN 14 : 2*2 V entre Puisseux et Magny-en-Vexin (voie express)
	Déviation sud-ouest de Meaux
	RN12 : élargissement à 2*3 V
2002 par rapport à 2005	BIP ouest entre rue d'Ermont et RD109
	RN 2 entre A104 et la limite départementale du 77 (voie express)
1999 par rapport à 2002	BIP ouest : section A15 - rue d'Ermont
	A115 : prolongement à Taverny
	Francilienne nord : section entre Fontenay-en-Parisis et Louvres
	Tronc commun A6-Francilienne dans le sens Paris-province
	RN 12 : déviation de Jouars-Ponchartrin
	RN 14 : déviation de Saint-Clair-sur-Epte
	A86 à Saint-Denis : doublement du viaduc
	N105 : accueil de A5 vers Melun

Tableau 27 : Modifications de l'offre routière de 1999 à 2026

Nous n'avons pas pris en compte les projets suivants :

- doublement du tronc commun A4-A86
- doublement du tronc commun A4-A104
- la déviation de la RN6 à Villeneuve-Saint-Georges en souterrain
- C13-F13 dans le nord des Yvelines
- A126
- Tous les projets départementaux d'aménagement ou de déviation routière telles que la déviation nord de Maisse dans l'Essonne, la desserte du SENIA par la rue du Bas Marin à Orly et Thiais, la déviation de la RD307 à Saint-Nom-la-Bretèche ... (on peut recenser ainsi une trentaine de projets départementaux).

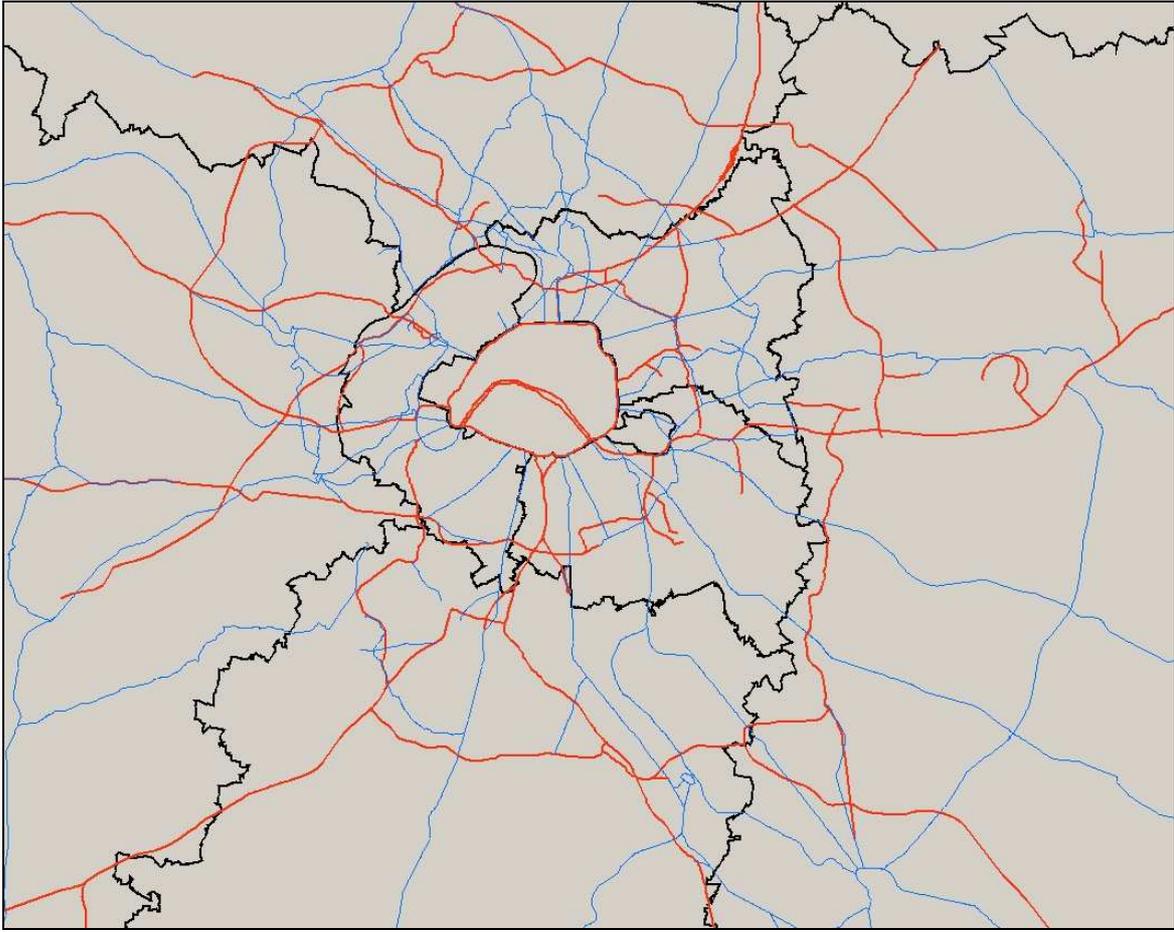


Figure 14 : Réseau routier de voies rapides et routes nationales en 2026

7.2.2. Réseau de transports en commun

On supprime du graphe de 2026 des arcs et des nœuds pour construire les réseaux de transport en commun tous les 3 ans (voir [2], page 25). Ce réseau de référence est un réseau à minima. L'objectif est de simuler les effets de la Tangentielle nord sur l'urbanisation. Compte tenu du temps des procédures administratives et des ressources financières très militées, nous avons opté pour un scénario de base réaliste, en partant du constat qu'il faut 30 ans depuis le moment des études amont pour qu'un projet voie le jour en Seine-Saint-Denis, 25 ans dans le Val-de-Marne, l'Essonne et la Seine-et-Marne, 15 ans dans les Yvelines, 10 ans dans les Hauts-de-Seine, 5 ans à Paris ...

Rappelons que le graphe du réseau de transport en commun de l'IAURIF 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 10000 arcs monodirectionnels et 4000 nœuds, auxquels s'ajoutent 1350 arcs de correspondance. Il y a sur le réseau lourd 220 missions à l'heure de pointe, chaque mission étant caractérisée par une fréquence et un temps de parcours intergares.

2023 par rapport à 2026	Prolongement du métro à Petit Bagneux (L4)
	Prolongement du métro à Mairie d'Aubervilliers (L12)
	Prolongement du métro au Bourget (L7)
2020 par rapport à 2023	Gare Evangile (RER E)
	Barreau de Gonesse (liaison RER D – RER B)
	Gare Carrefour Pompadour (RER D)
2017 par rapport à 2020	Prolongement du T1 à Montreuil
	Prolongement du TVM jusqu'à la gare Champigny-Boullereaux (RER E)
2014 par rapport à 2017	Tangentielle nord Sartrouville-Noisy-le-Sec
	CDG Express
	Prolongement du métro de Porte de la Chapelle à Proudhon-Gardinoux (L12)
	Tramway Châtillon-Viroflay
	Prolongement du T1 à Asnières-Gennevilliers-Les Courtilles
	Prolongement du T3 jusqu'à la Porte de la Chapelle
2011 par rapport à 2014	Tramway Saint-Denis - Epinay Villetaneuse/Stains (« TRAM'Y »)
	Prolongement du métro à Mairie de Montrouge (L4)
	Prolongement du T2 d'Issy-Val de Seine à Porte de Versailles (connexion avec le T3)
	Prolongement du T4 à Clichy-sous-bois
	Tramway RN7 Villejuif - Athis-Mons
	Tramway RN1 Saint-Denis – Sarcelles
	Site propre bus entre Pompadour et Sucy-Bonneuil
	Prolongement du T2 à Bezons
2008 par rapport à 2011	Prolongement du métro à Asnières-Gennevilliers-Les Courtilles (L13)
	Prolongement du métro à Créteil sud (L8)
2005 par rapport à 2008	Prolongement du TVM jusqu'à la gare Croix de Berny (RER B)
	Tramway des Maréchaux T3
	Ligne directe Ermont-Eaubonne – Saint-Lazare
	Prolongement de Meteor jusqu'à Olympiades
	Tramway Aulnay-Bondy T4
2002 par rapport à 2005	Grande Ceinture Ouest (Saint-Germain en Laye GC - Noisy le Roi)
	Prolongement du T1 à Noisy-le-Sec
	Prolongement de Meteor jusqu'à Saint-Lazare
	Prolongement d'EOLE à Tournan
1999 par rapport à 2002	EOLE jusqu'à Chelles et Villiers-sur-Marne
	Bus en site propre St-Quentin-en-Yvelines – Massy (1 ^{ère} phase)

Tableau 28 : Modifications de l'offre de TC de 1999 à 2026

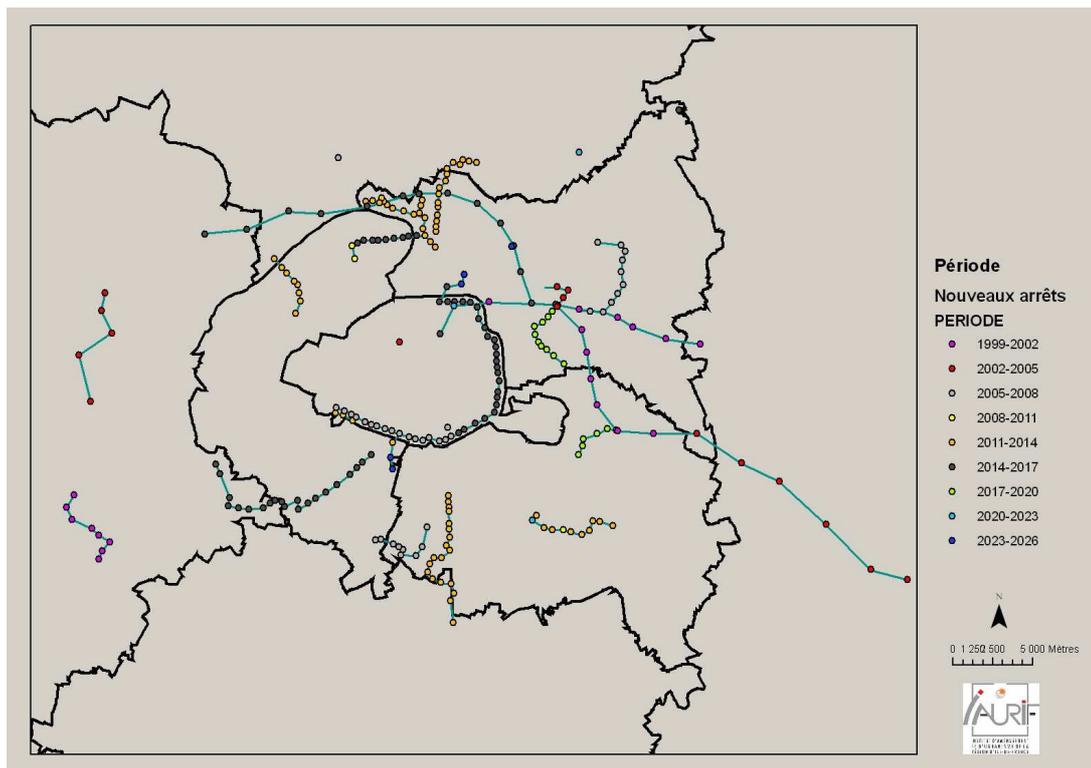


Figure 15 : Nouveaux nœuds du réseau TC à l'horizon 2026

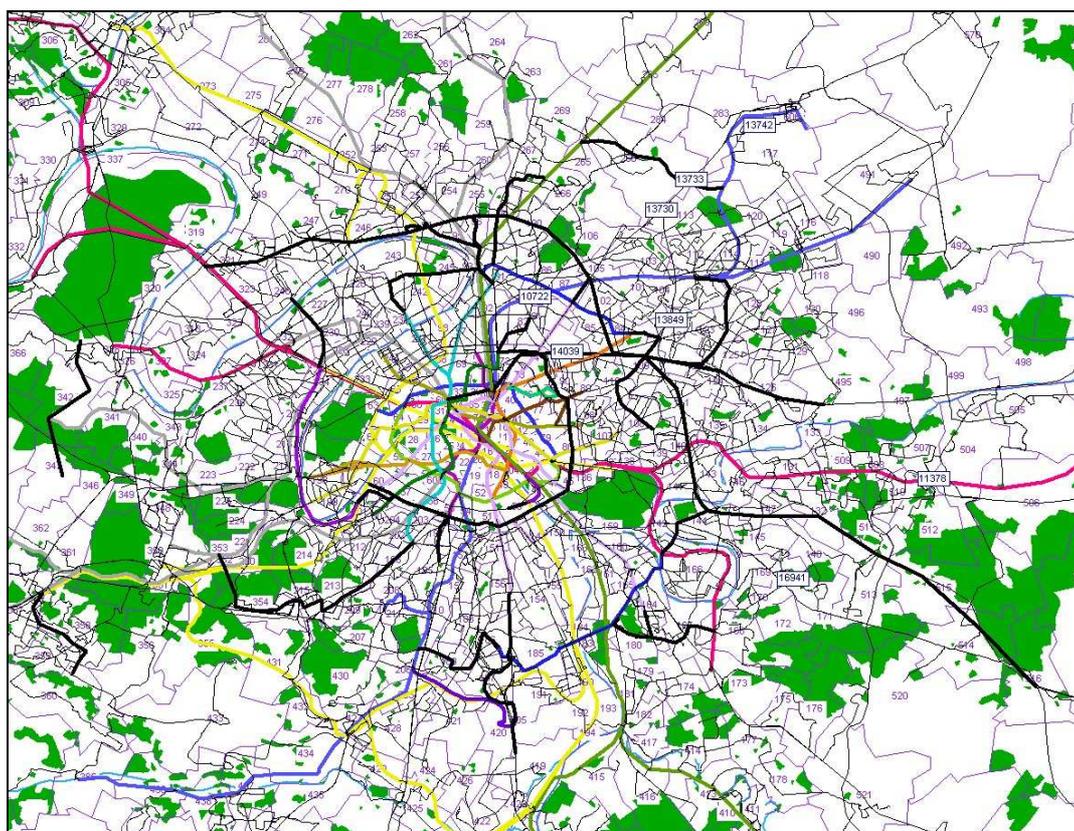


Figure 16 : Réseau TC du modèle à l'horizon 2026 (logiciel Davisum) sur le zonage en 606 zones
En noir : les nouveaux projets entre 1999 et 2026

Nous n'avons pas pris en compte les projets suivants :

- Arc-Express (ou Métrophérique), le projet de métro en rocade en proche banlieue
- La Tangentielle sud Massy-Evry en tram-train

- Les autres projets de prolongements de métro
- Le prolongement du RER E à La Défense (qui suppose la construction d'un tunnel entre la gare Saint-Lazare et la Défense)
- Le prolongement de Meteor au-delà de la gare Saint-Lazare
- La gare TGV à Orly
- Les projets de transport en commun en site propre en grande couronne

En tout, il y a 20 graphes à construire (1 par mode tous les 3 ans entre 1999 et 2026).

On fait tourner les modèles METROPOLIS pour les routes et Davisum pour les transports en commun pour calculer les matrices de temps en heures creuses (donc non issues d'une affectation). Ces matrices sont lues par UrbanSim grâce au script Python que nous avons écrit, puis elles sont importées dans la table « travel_data » tous les trois ans, et c'est UrbanSim qui calcule (grâce à l'« accessibility model ») les variables d'accessibilité en partant des données d'emplois qu'il répartit chaque année dans le maillage.

7.3. Les perspectives à l'horizon 2026

Habituellement, les perspectives d'accroissement de la population et des emplois à un horizon long terme sont données par le Schéma Directeur Ile-de-France. Or, en novembre 2007, il y a une grande incertitude concernant le devenir du projet de nouveau Schéma directeur Ile-de-France élaboré par l'IAURIF pour le Conseil régional d'Ile-de-France. L'Enquête publique est en cours au moment où ces lignes sont écrites mais l'Etat a émis des réserves importantes sur ce projet, en particulier sur son manque d'ambition en terme de croissance économique. Le projet de SDRIF 2007 prévoit la création de 60 000 logements par an jusqu'en 2030, soit 1.5 millions de logements supplémentaires à cet horizon. Mais il ne donne pas de cadrage plus précis ni sur la population ni sur les emplois. Dans ce contexte, nous avons préféré élaborer dans le cadre de ce projet nos propres hypothèses démographiques et socio-économiques à l'horizon 2026. Ces hypothèses sont basées sur un scénario intermédiaire entre ce que Raymond Aron appelle une « économie de bien-être » plutôt préconisée par le projet de SDRIF 2007 et une « économie de puissance » plutôt préconisée par le Gouvernement actuel. Il y a néanmoins des hypothèses certaines : la taille des ménages va continuer à décroître, avec de plus en plus de ménages d'une personne, et le vieillissement de la population va s'accroître. En ce qui concerne les emplois, les effectifs vont poursuivre leur baisse dans le secteur industriel et leur hausse dans les secteurs des services à la personne et aux entreprises et le secteur financier. L'économie industrielle se recompose: diminution des effectifs franciliens, fermeture des sites de production, augmentation des licenciements, externalisation des fonctions support de l'entreprise (RH, R&D, logistique, marketing...)

De nombreuses autres incertitudes demeurent, essentiellement pour la donnée d'emplois. Formulons-les sous forme de quelques questions :

- Les projets de tours à Paris se concrétiseront-ils ?
- Quelle suite sera donnée aux projets de villes Ecopolis préconisés par Jacques Attali dans son rapport sur la « Libération de la croissance française » (2008), sur le modèle de la ville Mountain View au sud de San Francisco ? Une ville Ecopolis est une ville d'au moins 50 000 habitants intégrant haute qualité environnementale et nouvelles technologies de communication. Ces villes pourraient servir notamment à faire progresser massivement l'urbanisme général, au service d'objectifs de croissance durable tels que des infrastructures haut débit, des transports technologiquement à la pointe et des constructions « durables ».

- Le projet de 3^{ème} aéroport dans le bassin parisien resurgira-t-il ?
- Quel sera le poids des pouvoirs publics dans 20 ans dans l'aménagement du territoire, la politique foncière et la politique du logement ? En 2007, on voit se mettre en place des Opérations d'Intérêt National (OIN)⁸ et des Etablissements Publics Fonciers départementaux (EPF). Leurs interventions auraient un impact sur les opérations immobilières. Quels seront leurs rayons d'action ?
- Qu'en sera-t-il des six pôles de compétitivité en Ile-de-France et combien d'emplois vont être créés ?

7.4. Les tables de type « control totals »

Rappelons que ces tables donnent le nombre **total** annuel de ménages par type et d'emplois par secteur en Ile-de-France pour chaque année de la période de simulation (voir [1], page 95). C'est *SIMAU*RIF qui simule la répartition géographique de ces données chaque année.

7.4.1. Table « ANNUAL HOUSEHOLDS CONTROL TOTALS »

La croissance plus rapide du nombre de ménages que de la population observée entre 1990 et 2006 devrait perdurer au cours des prochaines années. Elle proviendra pour 80 % du vieillissement de la population francilienne avec l'arrivée à la retraite des générations du baby-boom, les retraités formant pour l'essentiel des ménages d'une ou deux personnes.

Selon le contexte démographique global, la croissance en volume du nombre de ménages sera plus ou moins forte : elle devrait être comprise entre 0,67 et 1,4 million, selon que la population francilienne se stabilise voire diminue légèrement à l'horizon 2026, ou au contraire croît de nouveau beaucoup plus rapidement, sachant que dans un scénario « médian », l'Île-de-France compterait pratiquement autant de ménages supplémentaires en 2026 (+1 million par rapport à 1999) que d'habitants supplémentaires (+1,2 million).

Il faut savoir que l'INSEE réalise des projections de population à long terme. Ces projections s'appuient sur la connaissance du passé. Elles permettent d'apprécier comment évoluerait une population sous un ensemble de conditions fixées a priori. Une projection n'est pas une prévision. C'est le résultat mécanique d'hypothèses formulées sur les différents phénomènes démographiques. L'ensemble des hypothèses constitue un scénario. Le principe est de projeter la population d'une zone constituée de communes entières, par vieillissement et application de quotients. A partir des données issues des derniers recensements et de l'état civil, l'INSEE calcule des quotients de fécondité, de mortalité et de solde migratoire en fonction des hypothèses choisies en utilisant le modèle Omphale. L'INSEE insiste bien sur le fait que les projections sont très sensibles aux hypothèses.

En 1999 (année de départ de la simulation), la population était de 10,95 millions d'habitants. Nous partons sur une projection de **12 millions d'habitants en 2026** et une taille moyenne de ménages de 2.1, contre 2.3 en 2004 et 2.4 en 1999. L'INSEE a donné une estimation de la

⁸ En 2005, le premier ministre et le préfet d'Ile-de-France, en concertation avec les collectivités locales, ont posé les conditions de trois grandes Opérations d'Intérêt National d'urbanisme sur les territoires de "Seine-amont", "Massy, Saclay, Versailles, Saint Quentin-en-Yvelines" et "Seine-aval". Par exemple, pour le 2^{ème} OIN, la réalisation d'un grand pôle scientifique devrait permettre la création annuelle de 3.300 nouveaux emplois et 2.600 logements au-delà des besoins du renouvellement du parc actuel, qui est de 2.400 logements par an.

population francilienne en 2004 à 4,738 millions. Nous prenons cette valeur comme point fixe dans l'interpolation linéaire.

**YEAR TOTAL_NUMBER_OF
_HOUSEHOLDS**

1999	4 510 369
2000	4 555 003
2001	4 600 079
2002	4 645 601
2003	4 691 573
2004	4 738 000
2005	4 777 978
2006	4 818 293
2007	4 858 949
2008	4 899 947
2009	4 941 291
2010	4 982 985
2011	5 025 030
2012	5 067 429
2013	5 110 187
2014	5 153 305
2015	5 196 787
2016	5 240 637
2017	5 284 856
2018	5 329 448
2019	5 374 416
2020	5 419 764
2021	5 465 494
2022	5 511 611
2023	5 558 116
2024	5 605 014
2025	5 652 307
2026	5 700 000

Tableau 29 : Projections du nombre de ménages

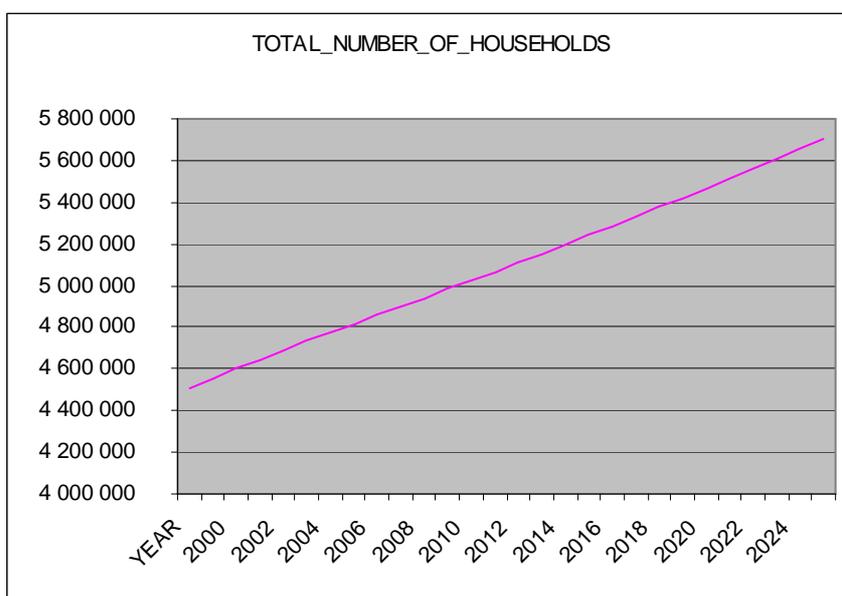


Figure 17 : Projections du nombre de ménages

7.4.2. Table « ANNUAL EMPLOYMENT CONTROL TOTALS »

En 1999 (année de départ de la simulation), le nombre d'emplois était de 5,04 millions . Nous partons sur une projection de **5,58 millions d'emplois en 2026**, soit une augmentation de 10,7% en 27 ans. Cette augmentation nette est le résultat de baisse d'effectifs dans certains secteurs (industrie par exemple) et d'augmentation dans d'autres secteurs (services à la personne et aux entreprises, financier par exemple).

YEAR	Agriculture	Industrie	Energie	Commerce	Finances	Services	Education	Administration	Total
1999	81 989	548 711	579 006	655 915	358 882	1 431 842	781 069	604 581	5 041 995
2000	81 914	536 577	583 090	662 242	365 779	1 437 743	783 519	604 411	5 055 275
2001	81 840	524 711	587 202	668 630	372 808	1 443 668	785 977	604 240	5 069 077
2002	81 766	513 107	591 344	675 080	379 972	1 449 617	788 443	604 070	5 083 400
2003	81 691	501 760	595 515	681 592	387 274	1 455 592	790 917	603 900	5 098 241
2004	81 617	490 664	599 715	688 167	394 716	1 461 590	793 398	603 730	5 113 598
2005	81 543	479 813	603 945	694 806	402 301	1 467 614	795 887	603 560	5 129 469
2006	81 469	469 202	608 205	701 508	410 032	1 473 662	798 384	603 390	5 145 852
2007	81 395	458 826	612 495	708 275	417 912	1 479 735	800 889	603 220	5 162 746
2008	81 321	448 679	616 815	715 108	425 943	1 485 833	803 401	603 050	5 180 150
2009	81 247	438 757	621 165	722 006	434 128	1 491 957	805 922	602 880	5 198 062
2010	81 173	429 054	625 546	728 971	442 471	1 498 105	808 450	602 710	5 216 481
2011	81 099	419 566	629 959	736 002	450 974	1 504 279	810 987	602 541	5 235 406
2012	81 025	410 288	634 402	743 102	459 640	1 510 478	813 531	602 371	5 254 837
2013	80 952	401 214	638 876	750 271	468 473	1 516 703	816 083	602 201	5 274 774
2014	80 878	392 342	643 382	757 508	477 476	1 522 954	818 643	602 032	5 295 215
2015	80 804	383 665	647 920	764 815	486 652	1 529 230	821 212	601 862	5 316 160
2016	80 731	375 181	652 490	772 193	496 004	1 535 532	823 788	601 693	5 337 611
2017	80 658	366 884	657 092	779 642	505 535	1 541 860	826 372	601 523	5 359 567
2018	80 584	358 771	661 727	787 163	515 250	1 548 214	828 965	601 354	5 382 027
2019	80 511	350 837	666 394	794 756	525 152	1 554 595	831 565	601 184	5 404 994
2020	80 438	343 078	671 094	802 422	535 243	1 561 002	834 174	601 015	5 428 467
2021	80 365	335 491	675 828	810 163	545 529	1 567 435	836 791	600 846	5 452 447
2022	80 292	328 072	680 594	817 978	556 013	1 573 894	839 417	600 676	5 476 936
2023	80 219	320 817	685 395	825 869	566 698	1 580 380	842 050	600 507	5 501 934
2024	80 146	313 722	690 229	833 835	577 588	1 586 893	844 692	600 338	5 527 443
2025	80 073	306 784	695 097	841 879	588 687	1 593 433	847 342	600 169	5 553 465
2026	80 000	300 000	700 000	850 000	600 000	1 600 000	850 000	600 000	5 580 000

Tableau 30 : Projections du nombre d'emplois par secteur

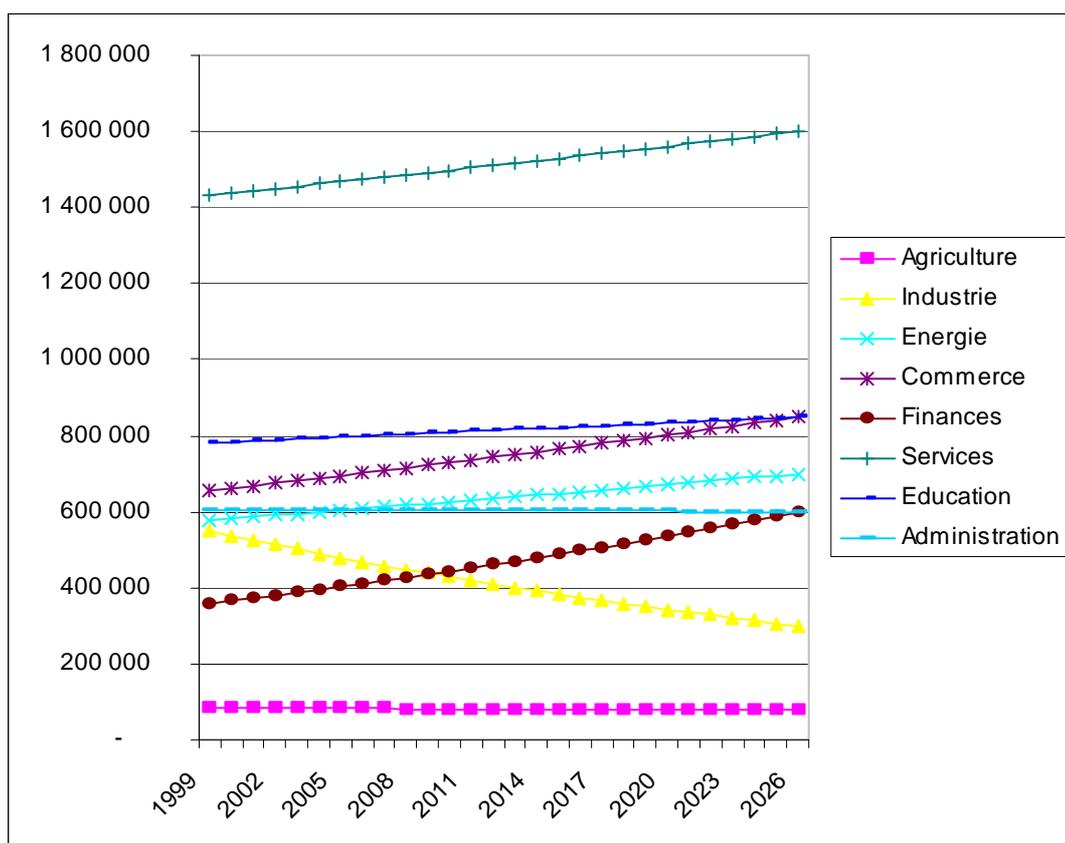


Figure 18 : Projections du nombre d'emplois par secteur

7.5. Les tables de types « rates » et « constants »

Pour les tables de taux et de constantes (voir [1] pages 95 et 103), on fait l'hypothèse que les comportements ne changent pas. Donc les taux observés sur la période de calage 1990-1999 sont appliqués à la période de simulation 1999-2026.

ANNUAL RELOCATION RATES FOR JOBS	0,2
ANNUAL RELOCATION RATES FOR HOUSEHOLDS	0,15
SAMPLING RATES	0,5
WALKING DISTANCE CIRCLE RADIUS	500

Tableau 31 : Quelques constantes

7.6. Les tables de type « events »

Nous n'utilisons pas ces tables, contrairement à la simulation sur la période de calage. La documentation à ce sujet est très incomplète. Dans le chapitre 9 du manuel de référence (voir [5]), seule la table « development_events_exogenous.tab » est décrite. Mais les tables « employment_events.tab » et « land-use_events.tab » qui existaient dans la version 2 ne sont même pas évoquées, alors qu'elles sont maintenues dans la nouvelle version.

8. Application à la Tangentielle Nord

8.1. Dernières nouvelles sur le projet de la Tangentielle Nord

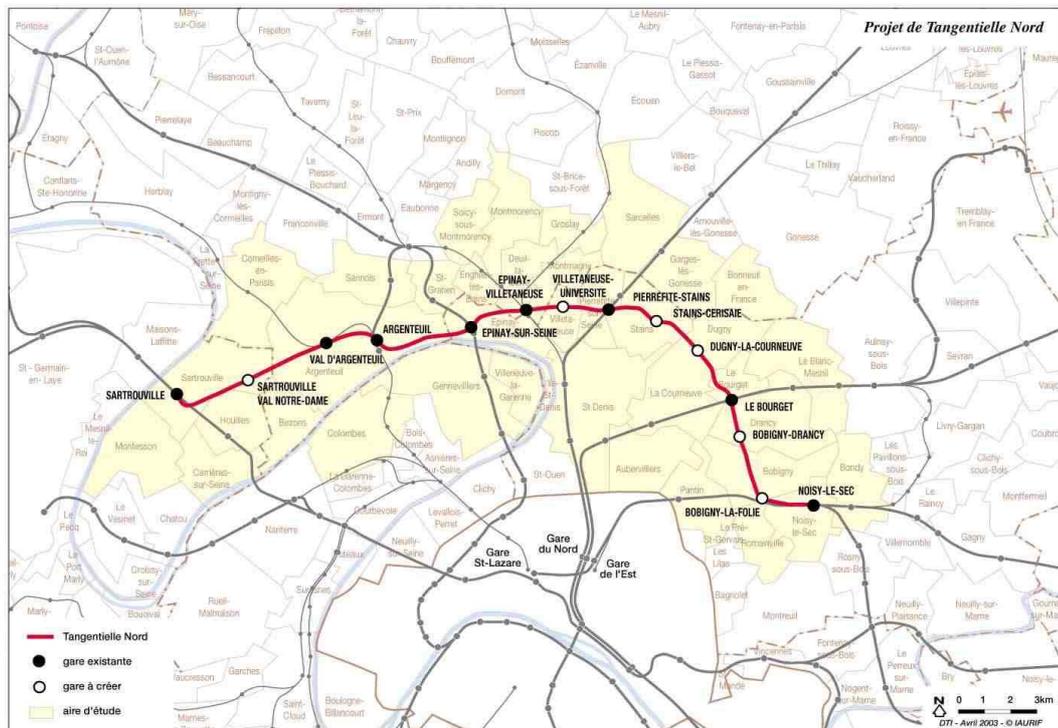


Figure 19 : Plan de situation de la Tangentielle Nord

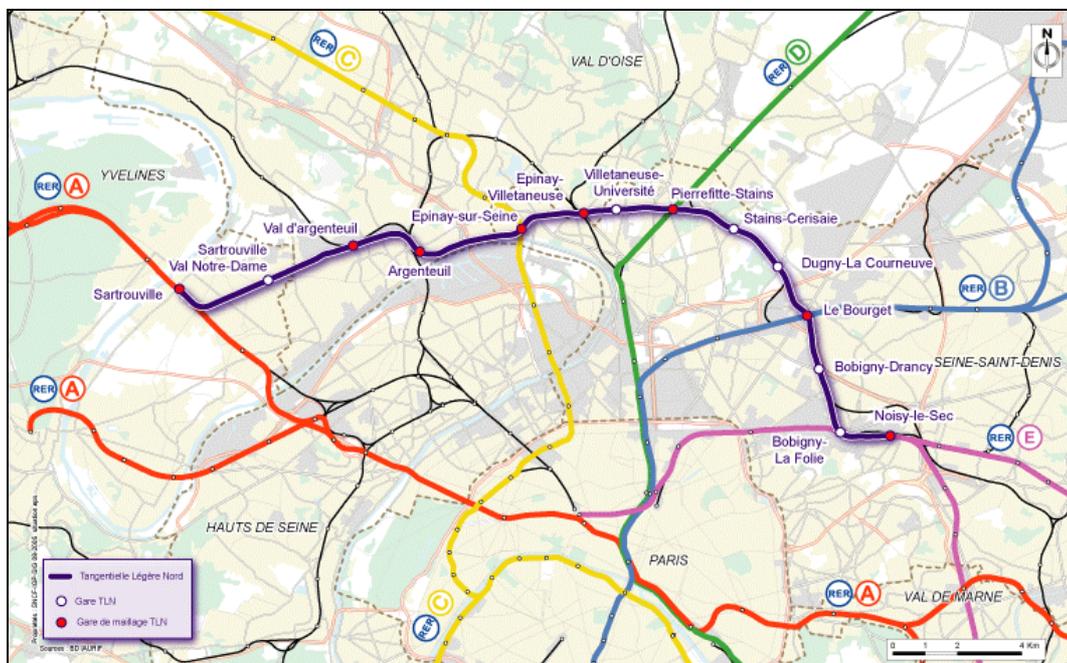


Figure 20 : Maillage de la Tangentielle Nord avec les 5 lignes de RER et les Transiliens

Le projet consiste en la création, sur 28 kilomètres, de deux voies dédiées, contiguës aux voies de la ligne actuelle dite “Grande Ceinture”. 6 nouvelles gares seront créées et 8 gares existantes

réaménagées pour optimiser les correspondances avec les 5 lignes RER, les lignes “Transilien” provenant de Paris Nord et Paris Saint-Lazare, ainsi qu'avec le tramway T1 (Noisy-le-Sec/Saint-Denis) et le futur Tramway Saint-Denis - Epinay-Villetaneuse. La Tangentielle Nord sera la première rocade ferroviaire en banlieue.

Il est prévu une mise en service progressive : en 2014, la section Epinay-sur-Seine – Le Bourget, puis en 2016, toute la ligne de Sartrouville à Noisy-le-Sec sera mise en service (dans le premier Schéma de principe, la mise en service était envisagée pour 2009).

La fréquence des trains sera élevée, avec un passage toutes les 5 minutes aux heures de pointe et toutes les 10 minutes aux heures creuses.

La vitesse moyenne sera d'environ 50 km/h. Le temps de trajet entre Sartrouville et Noisy-le-Sec sera de 35 minutes. La ligne sera ouverte de 5h à 24h.

En ce qui concerne le matériel roulant, il sera de type tram-train, c'est-à-dire de type urbain, plus léger, plus court et plus silencieux qu'un RER. La capacité de chaque tram-train sera de 500 voyageurs et pourra être portée à 750 voyageurs dans le futur.

L'enquête publique a eu lieu en 2006.

8.2. Diagnostic territorial

8.2.1. Aire d'influence élargie

L'aire d'influence élargie de la Tangentielle nord est constituée de 38 communes.

INSEE	Nom commune	Pop 1999	Empl 1999
78124	Carrières-sur-Seine	12 050	3 209
78311	Houilles	29 634	5 782
78418	Montesson	13 750	4 865
78586	Sartrouville	50 219	12 232
92025	Colombes	76 757	25 568
92036	Gennevilliers	42 513	30 265
92078	Villeneuve-la-Garenne	22 349	8 304
93001	Aubervilliers	63 136	25 843
93007	Le Blanc-Mesnil	46 936	13 366
93008	Bobigny	44 079	25 986
93010	Bondy	46 826	11 487
93013	Le Bourget	12 110	7 504
93027	La Courneuve	35 310	14 213
93029	Drancy	62 263	14 164
93030	Dugny	8 641	1 604
93031	Epinay-sur-Seine	46 409	8 960
93039	L'Île-Saint-Denis	6 810	1 620
93053	Noisy-le-Sec	37 312	10 131
93055	Pantin	49 919	23 386
93059	Pierrefitte-sur-Seine	25 816	4 670
93063	Romainville	23 779	7 820
93066	Saint-Denis	85 832	48 949
93072	Stains	32 839	6 776
93079	Villetaneuse	11 376	4 161

95018	Argenteuil	93 961	30 777
95063	Bezons	26 263	8 312
95088	Bonneuil-en-France	774	518
95176	Cormeilles-en-Parisis	19 643	3 743
95197	Deuil-la-Barre	20 160	3 218
95210	Enghien-les-Bains	10 368	4 677
95268	Garges-lès-Gonesse	40 058	8 194
95288	Groslay	7 385	1 414
95427	Montmagny	13 090	2 046
95428	Montmorency	20 599	5 477
95555	Saint-Gratien	19 226	3 368
95582	Sannois	25 349	5 002
95585	Sarcelles	57 871	14 584
95598	Soisy-sous-Montmorency	16 802	3 809
		1 258 214	416 004

Tableau 32 : Population et emplois des 38 communes en 1999

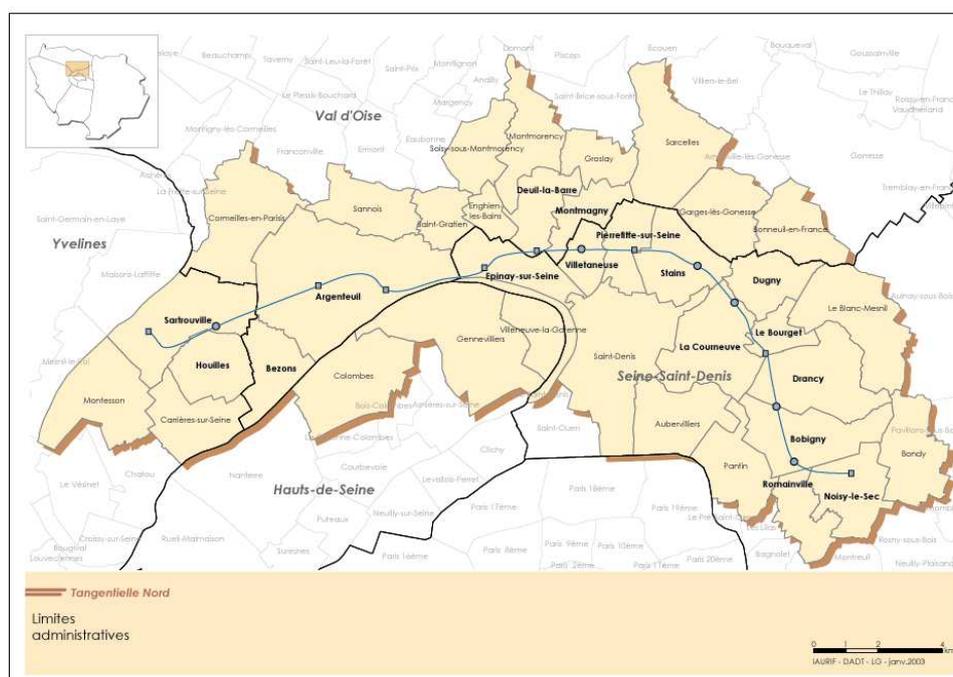


Figure 21 : Aire d'étude (carte extraite de [9])

L'aire d'étude est constituée de **853 cellules** de 500 mètres de côté. La simulation sera exécutée sur l'ensemble des 50 000 cellules du découpage, mais les résultats seront analysés sur l'aire d'étude.

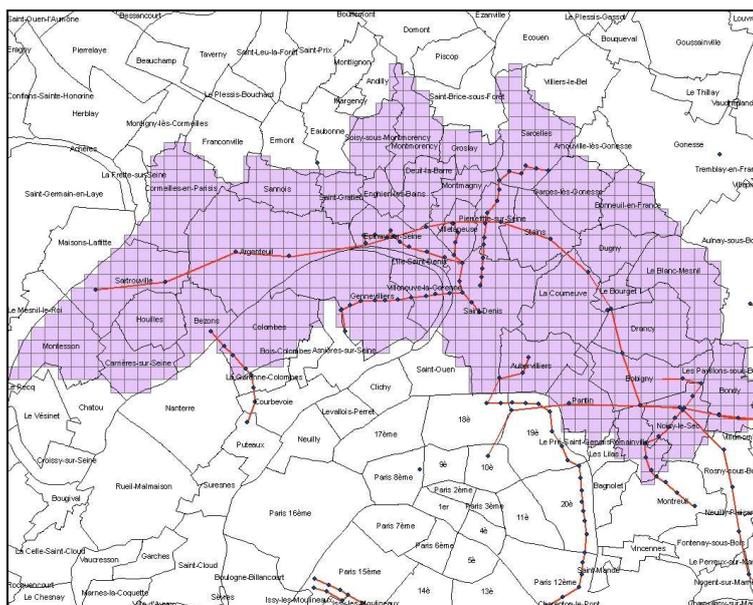


Figure 22 : Les 853 cellules du territoire de la Tangentielle nord

8.2.2. Population et emplois

L'aire d'étude comporte environ **1 300 000 habitants**. Cette population est stable mais le solde migratoire est négatif (- 9 % en moyenne).

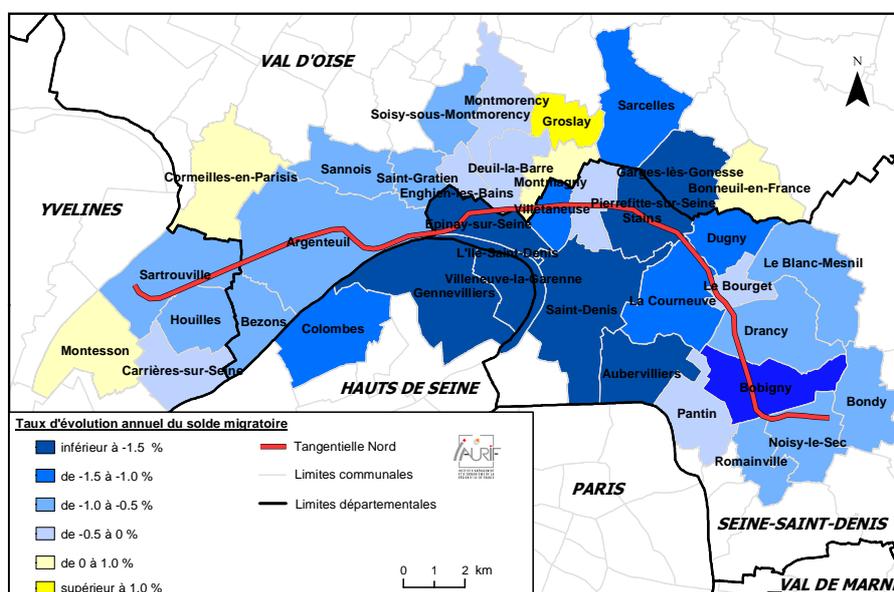


Figure 23 : Taux d'évolution annuelle du solde migratoire (carte extraite de [9])

NOMBRE D'EMPLOIS ENTRE 1990 ET 1999

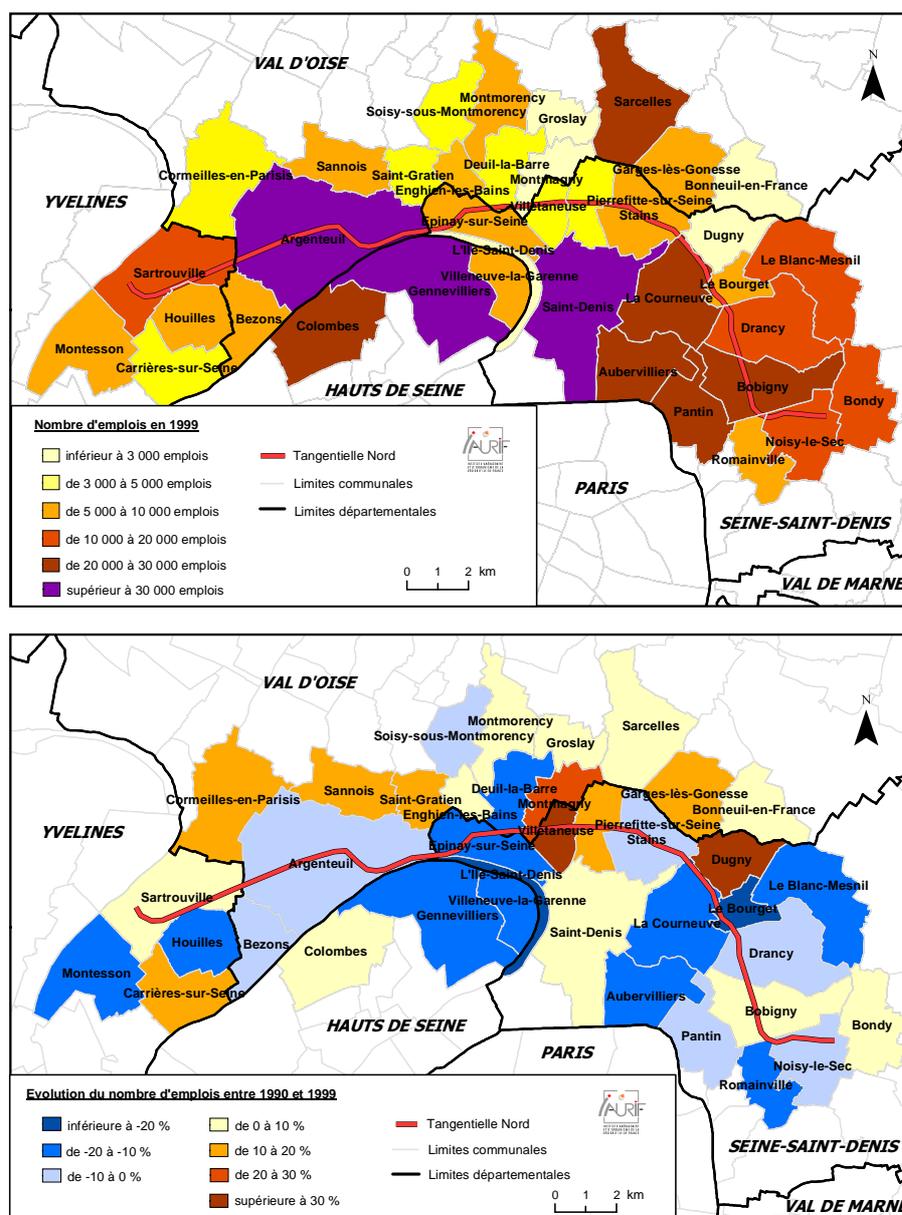


Figure 24 : Nombre d'emplois entre 1990 et 1999 (carte extraite de [9])

Le revenu moyen des ménages est faible, le taux de chômage élevé. Le taux de logement social est élevé (39 % contre 30 % en moyenne en petite couronne), avec des disparités très fortes (Dugny : 77 %, Enghien : 7 %). On trouve aussi une forte population étrangère (19 % contre 14 % en petite couronne) et un faible niveau de qualification (ouvrier : 24 % contre 18 % en petite couronne).

L'aire d'étude comporte en 1999 environ **416 000 emplois**

Le nombre d'emplois a baissé de 5 500 entre 1991 et 1998 (entre les deux ERE). Soulignons cependant que les emplois publics constituent 25% du total et qu'ils ont augmenté de 10 000 entre 1991 et 1998.

Les deux principaux pôles d'emplois sont Bobigny (30 000 emplois) et Argenteuil (26 000).

Les activités traditionnelles de ce territoire sont la métallurgie (5 000 emplois), le matériel de transport (3 700 emplois) et l'industrie chimique (3 200 emplois).

Le bâtiment représente 14 000 emplois (en hausse de 13% entre 1998 et 2000), les activités commerciales 32 000 emplois et l'immobilier et services, 20 000 emplois.

De grandes entreprises (Aventis, Eurocopter, Alstom, Bosch, ELM Leblanc, Dassault...) sont implantées. On a observé une diminution des emplois dans les grands établissements. On trouve de nombreuses Zones d'Activités Economiques (65) et de nombreux sites vacants. Ce territoire est en concurrence avec des bassins d'activités concurrents : la Plaine-Saint-Denis, la Boucle de Gennevilliers et le Sud de Roissy CDG.

On peut cependant recenser des leviers économiques. D'abord des centres d'enseignement supérieur : l'Université Paris XIII (Villetaneuse et Bobigny), l'Université Paris VIII (Saint-Denis), le CNAM et l'ISMCM-CESTI. Ensuite des filières porteuses : la santé (Bobigny, Noisy-le-Sec), l'image (Epinay, Plaine-Saint-Denis) et l'aéronautique (Le Bourget, Argenteuil). Enfin, l'immobilier d'entreprise a du potentiel : 55 ha commercialisables de ZAE (mais commercialisation lente), 70 000 m² de bureaux commercialisables (mais marché peu actif), et 90 000 m² de locaux d'activités commercialisables, dont 60 % d'entrepôts.

8.2.3. L'occupation du sol

L'habitat individuel constitue 32 % de la surface, l'habitat collectif 12 %, les activités 14 %, les transports (routes, fer, aéroports) 12 % et les espaces ouverts 25 %.

500 ha ont été urbanisés entre 1994 et 1999, soit 2,5 % de l'aire d'étude.

Les postes les plus consommateurs d'espaces sont les activités, équipements et habitat.

La mutation des espaces a concerné 11 % de la surface de l'aire d'étude.

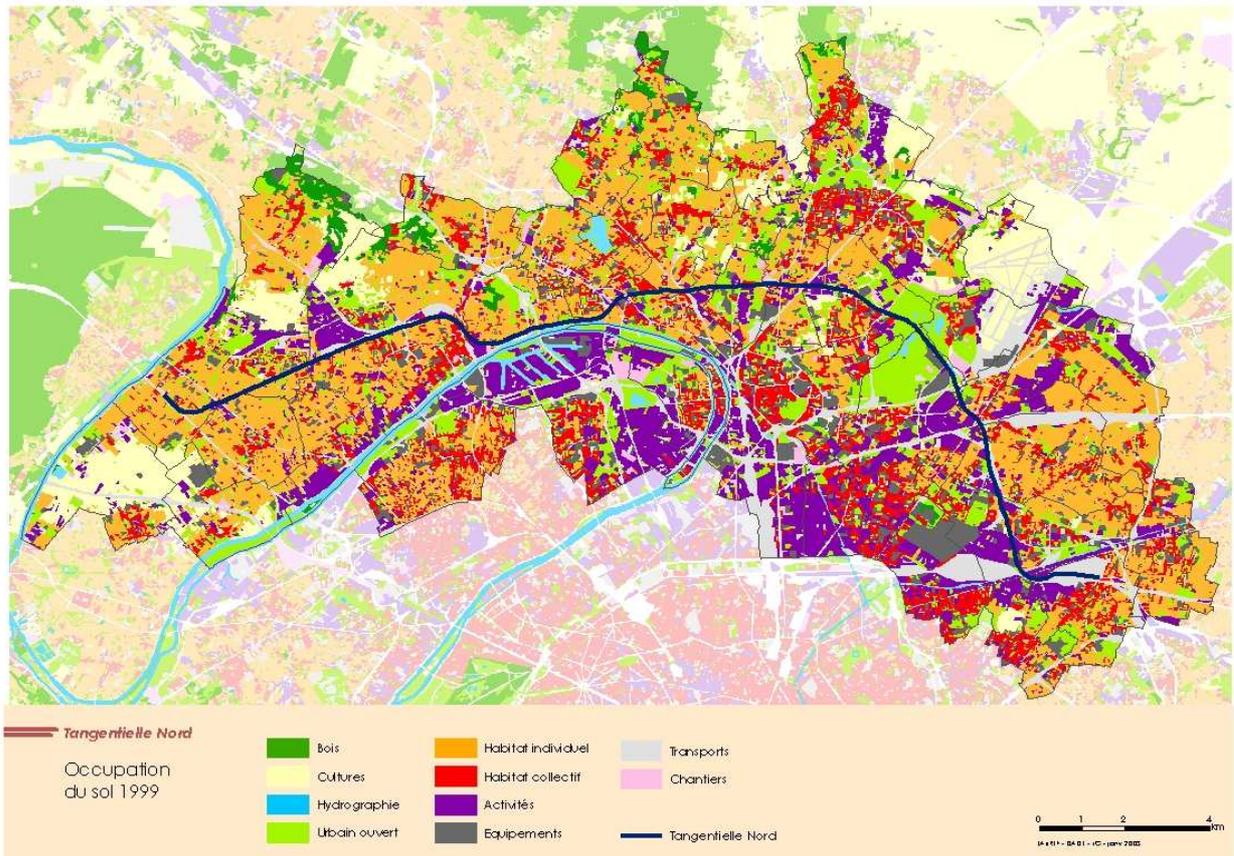


Figure 25 : MOS en 1999 (carte extraite de [9])

8.2.4. Le marché immobilier

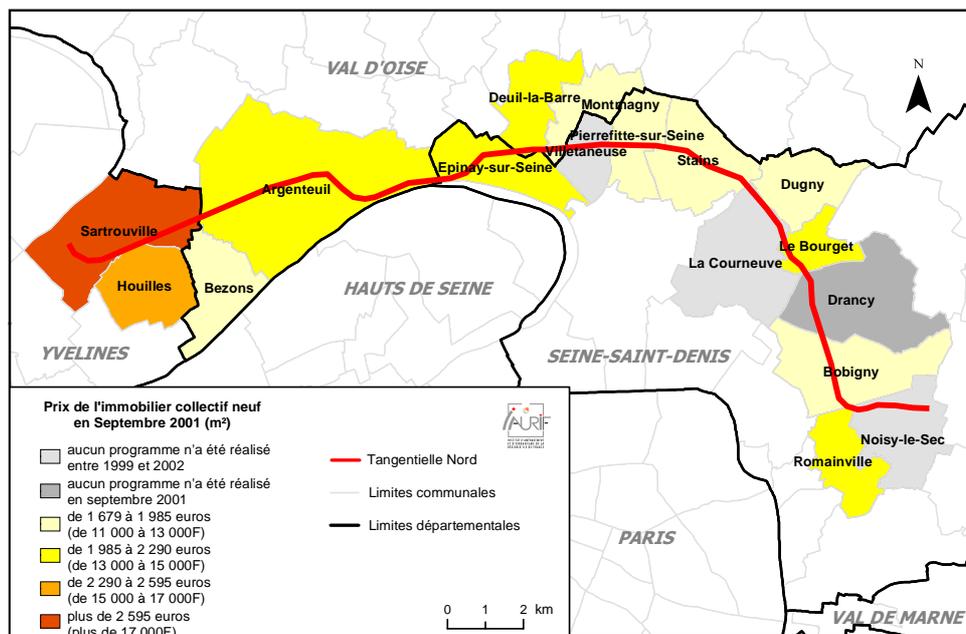


Figure 26 : Prix de l'immobilier neuf en septembre 2001(/m2) (carte extraite de [9])

Le marché est peu dynamique. Le collectif domine (3/4 des logements offerts). Les prix sont 20 % en-dessous de la moyenne de petite couronne. Il y a des secteurs plus dynamiques (Houilles, Sartrouville, Argenteuil et Romainville) et des écarts de prix importants entre Sartrouville, Houilles (environ 2800 €/m²) et Stains, Dugny (environ 1 700 €/m²).

Le parc de logement est très contrasté, avec 71 % de collectif et 29 % d'individuel. Le pavillonnaire est plus ancien que le collectif. Le parc social constitue 42 % du parc total.

Il y a des secteurs en difficultés, de grands ensembles du parc social, des copropriétés en difficulté, des cités jardins et des quartiers pavillonnaires en déclin (Argenteuil, La Courneuve, Drancy).

La Politique de la ville permet à toutes les communes d'avoir un Contrat de ville (sauf la commune du Bourget). 5 Grands Projets de Ville (GPV) sont concernés par la Tangentielle nord. 7 communes sont concernées par les Zones franches urbaines. Enfin 7 communes sont concernées par des Opérations programmées par l'amélioration de l'habitat (OPAH).

8.2.5. Synthèse

Ce territoire de 38 communes comporte un certain nombre d'atouts pour son **développement économique** :

- la pluri-fonctionnalité des sous-territoires desservis
- des pôles attractifs
- des filières établies ou émergentes
- des potentiels de développement importants : importance des surfaces affectées aux activités, importance des terrains mutables, un bassin de main d'œuvre important
- une possibilité de maillage sur les axes radiaux par des infrastructures de rocade nord.

En revanche, les faiblesses sont nombreuses :

- un secteur relativement déprimé, une image de secteur dévalorisé, un secteur mal identifié du point de vue économique
- des facteurs défavorable au renouvellement urbain : un tissu qui se renouvelle au coup par coup, un foncier qui reste cher
- des équipements majeurs non desservis directement
- une faiblesse au niveau des communes situées sur la section centrale de la Tangentielle nord
- une structuration traditionnellement très radiale et un morcellement institutionnel
- des sites autour des gares en général assez contraints
- une concurrence avec des secteurs à proximité plus attractifs ou en développement
- une main d'œuvre sous qualifiée

De même, en ce qui concerne **l'attractivité résidentielle**, le territoire présente aussi des atouts :

- proximité des grands centres d'emplois et d'équipements
- des tissus d'habitat de qualité
- des atouts environnementaux
- des situations d'insularité
- des potentialités de renouvellement dans les tissus existants
- des prix qui restent faibles
- des actions fortes d'amélioration du tissu d'habitat : Contrat de ville, GPV, objectif 2

Mais là aussi, les faiblesses sont un véritable frein au choix de localisation des ménages sur ce territoire :

- faible mixité sociale
- manque de logements pour les couches moyennes ou aisées de la population
- un cadre de vie peu attractif : environnement, services, image du territoire
- une faible valorisation du bâti
- un manque de dynamisme de la construction

La Tangentielle nord offre des **opportunités** pour favoriser le développement économique et augmenter l'attractivité résidentielle :

- amélioration du maillage du territoire, du développement des pôles et une mise en réseau des pôles majeurs
- amélioration de l'accès à l'emploi et aux équipements
- ouverture du territoire à un bassin de main d'œuvre plus large
- amélioration du niveau de formation par l'accessibilité aux équipements de formation
- renforcement de l'attractivité résidentielle du territoire, en particulier à proximité des gares
- renouvellement urbain aux abords des gares
- redynamisation dans les tissus urbains

Enfin, c'est un territoire caractérisé par l'importance des actions déjà engagées depuis 10 ans : le CPER, des Fonds européens de développement régional (FEDER), contrats Région-Département, des actions de l'EPA Plaine de France, une anticipation des collectivités pour valoriser l'apport de la tangentielle (communes, intercommunalité, GEP 93, EPA, GPV, étude bruit de la Région). Il n'en reste pas moins qu'en réalité, il y a beaucoup de velléités et que la mise en service a encore été reportée à 2014 pour la 1^{ère} section.

8.3. Elaboration de scénarios

Deux scénarios sont testés :

- le scénario de référence sans Tangentielle nord
- le scénario avec la Tangentielle nord (mise en service au cours de la période 2014-2017)

Entre ces deux scénarios, seules les variables transports changent. Elles sont mises à jour après l'année de mise en service de la Tangentielle nord, donc aux années 2017, 2020, 2023 et 2026 sur le découpage en 50 000 cellules :

1) Nombre de gares par cellule : cela concerne seulement 14 cellules (sur les 50 000).

On peut s'interroger ici sur la sensibilité d'un modèle à la modification infiniment faible d'une variable. A ce stade, on s'est inquiété de savoir si les résultats des deux scénarios montreraient des différences.⁹

⁹ Evoquons une petite analogie avec le phénomène de la Résonance Magnétique Nucléaire, toute proportion gardée. Ce phénomène n'est possible que parce qu'il existe une différence d'énergie entre la population des spins parallèles et la population des spins anti-parallèles (une différence qui porte sur 6 atomes par rapport à 100 000). A l'époque de la découverte, de nombreux théoriciens pensaient que cette différence n'était pas mesurable. Ils se sont trompés, ce qui a valu le prix Nobel de physique à Purcell et Bloch en 1952.

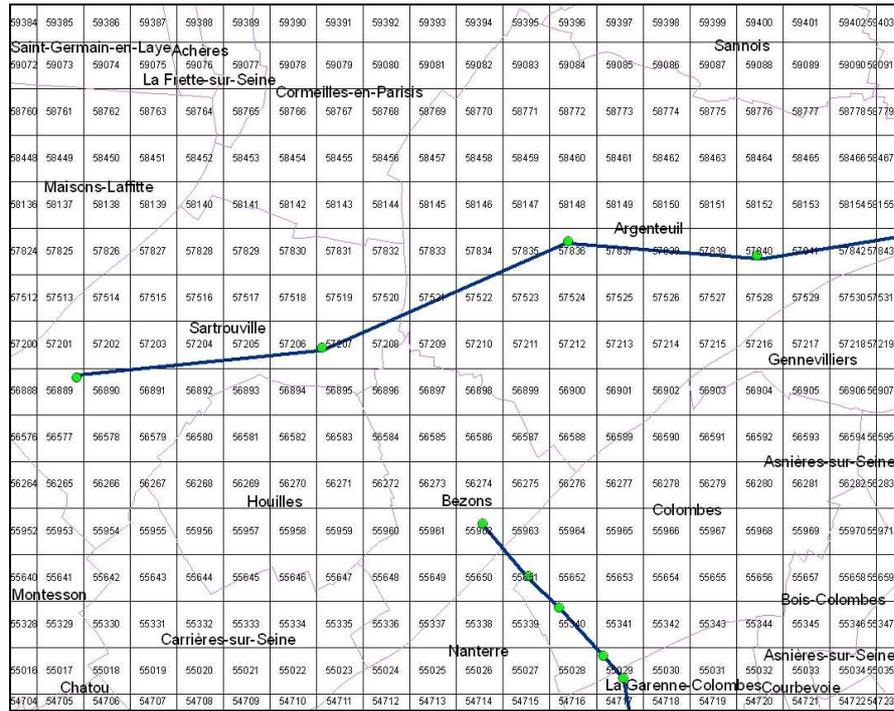


Figure 27 : Numérotation du maillage à l'ouest de la Tangentielle nord

2) L'accessibilité « emploiTC ». Rappelons ici la formule, expliquée dans le rapport [2] page 59 :

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

avec Q_j : nombre d'emplois par zone j

$$\lambda = 2,86$$

C_{ij} : temps de déplacement en TC entre les zones i et j

3) Le temps d'accès à Châtelet en TC

Il est important de noter que les différences entre les deux scénarios ne portent que sur l'offre de transport. Ce qui signifie que si les simulations montrent par exemple un gain en emplois sur le territoire entre les deux scénarios à l'horizon futur, ce gain sera un minimum car il sera uniquement imputé aux gains d'accessibilité locale et régionale et au gain de temps vers Paris. Or on a vu dans le diagnostic territorial combien l'engagement des acteurs locaux pour accompagner la mise en service de la Tangentielle nord est important pour optimiser l'investissement transport.

Il aurait été intéressant de construire d'autres scénarios pour tester la sensibilité des paramètres d'offre telles que la fréquence et la vitesse ou d'autres scénarios basés sur des politiques d'aménagement ou foncières différentes.

8.4. Simulations

Un nouvel ordinateur a été utilisé pour les simulations. Il est équipé d'un processeur Core Duo 2,4 GHz et de 4 Go de RAM.

La durée d'une simulation sur la période 1999-2026 est d'environ 7 heures.

En pratique, ces simulations sont lancées en fin de journée, les résultats attendus pour le lendemain matin. Evidemment, les « plantages » étaient fréquents au début ... On laisse le lecteur imaginer le pire, on se serait crû revenir à l'informatique des années 70.

8.5. Résultats

Le résultat principal qui nous intéresse est la répartition géographique des populations et des emplois en 2026, en particulier dans le territoire élargi desservi par la Tangentielle nord et plus précisément dans les environs immédiats des gares (les 9 cellules autour de chaque gare).

SIMAU met à jour chaque année les tables households, jobs et gridcells. Nous avons extrait de ces tables en 2026 les résultats sur le territoire de la Tangentielle nord.

8.5.1. La population

Nous avons cartographié les résultats sur les 853 cellules du territoire.

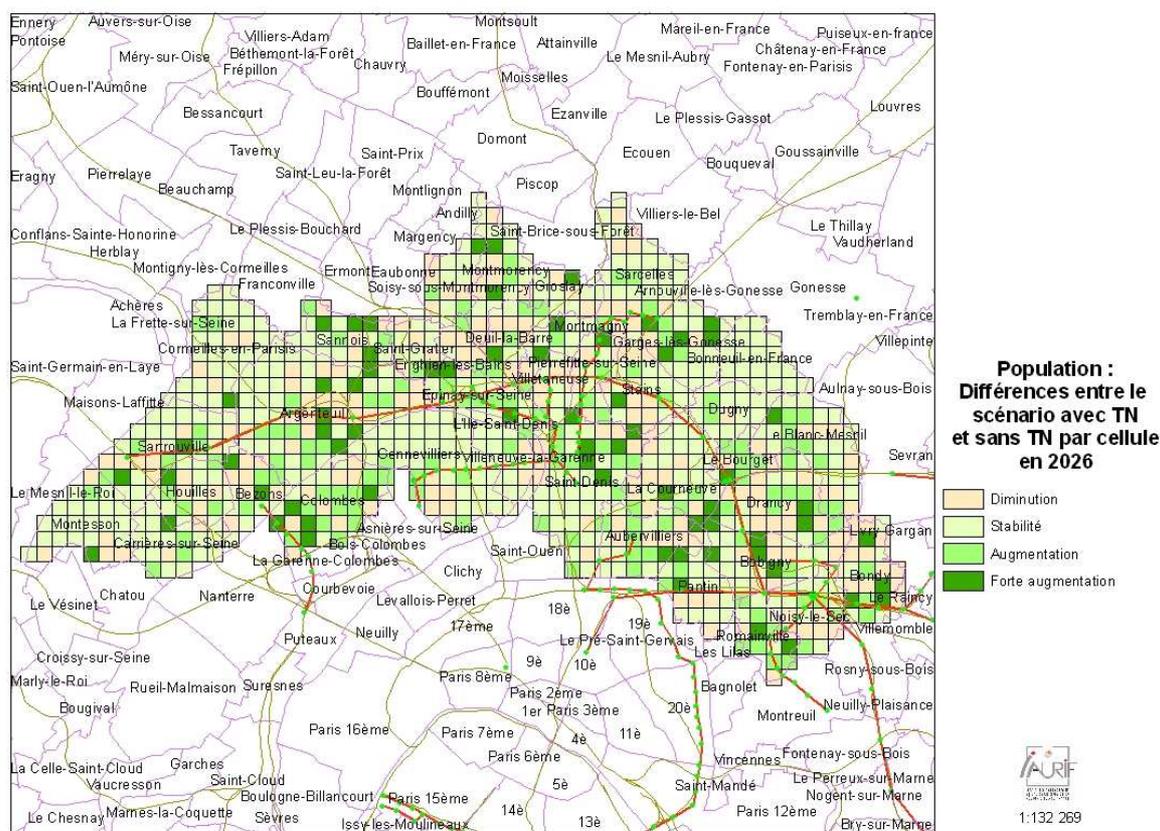


Figure 28

Globalement, la population sur l'ensemble de la région était de 10,95 millions d'habitants en 1999 et de 12 millions en 2026 (hypothèse exogène). Sur le territoire de la Tangentielle nord, la population en 1999 était d'environ 1,26 millions d'habitants.

Dans le scénario sans Tangentielle nord, *SIMAU*RIF trouve une population en 2026 de 1,520 millions sur le territoire, soit une augmentation de 260 000 habitants en 27 ans. Dans le scénario avec la Tangentielle nord, *SIMAU*RIF trouve une population de 1,522 millions, soit **2000 habitants de plus** seulement.

Empressons nous d'affirmer que cette différence de 2000 habitants est extrêmement faible par rapport à l'augmentation tendancielle de la population (2000/260000 <1%). Il est possible que *SIMAU*RIF soit suffisamment précis pour détecter cette différence. Mais il est plus probable que l'on se trouve dans la marge d'erreur du modèle. Donc en réalité on ne peut pas conclure de manière tranchée sur l'attractivité résidentielle.

De toutes façons, si le modèle est fiable, la faiblesse de ce surplus de population ne doit pas surprendre. Ce territoire n'est au départ pas très attractif du point de vue résidentiel. Ce n'est pas la simple arrivée de la Tangentielle nord qui va modifier radicalement l'image du territoire. D'autres actions concomitantes sont nécessaires afin d'attirer de nouveaux ménages. Il faut améliorer le cadre de vie au quotidien : créer des espaces verts, des commerces de proximité, revaloriser le bâti, requalifier des axes urbains, etc.

Nous avons quand même voulu savoir comment se répartissent ces 2000 habitants supplémentaires. D'abord, c'est un solde net, résultant de départs et d'arrivées. Ensuite, la figure 28 montre les différences en quatre tranches de population entre les deux scénarios. On constate que les cellules dans lesquelles la population a augmenté ne se trouvent pas toutes dans le corridor immédiat de la Tangentielle nord ou à proximité des 14 gares comme on aurait pu l'espérer. Cependant, on perçoit que ces cellules se trouvent plutôt dans les communes les plus denses du territoire.

8.5.2. Les emplois

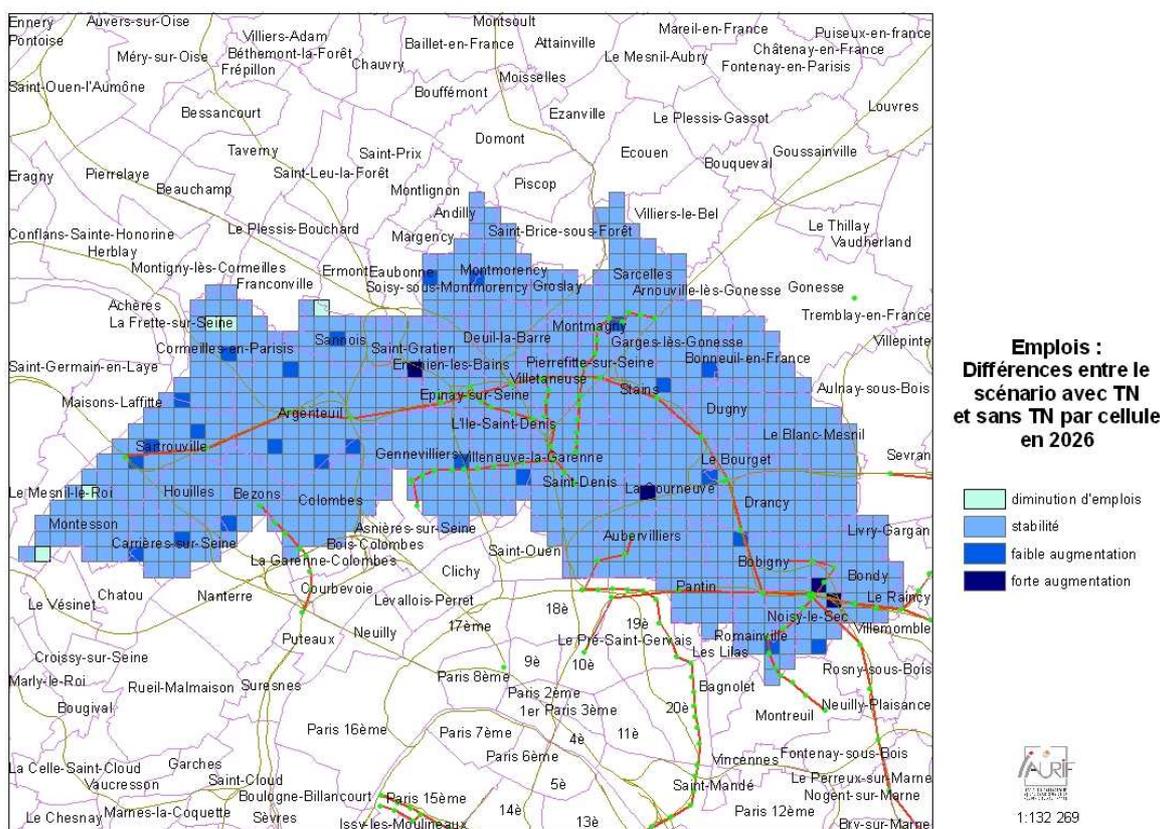


Figure 29

Globalement, le nombre d'emplois sur l'ensemble de la région était de 5,042 millions d'habitants en 1999 et de 5,58 millions en 2026 (hypothèse exogène). Sur le territoire de la Tangentielle nord, ce nombre en 1999 était d'environ 416 000 emplois.

Dans le scénario sans Tangentielle nord, *SIMAU* trouve un nombre d'emplois en 2026 de 508 000 sur l'aire d'étude, soit une augmentation de 92 000 emplois en 27 ans. Dans le scénario avec la Tangentielle nord, *SIMAU* trouve un nombre d'emplois de 518 000, soit **10 000 emplois de plus** par rapport au scénario sans Tangentielle nord.

1^{er} élément : on est cette fois plutôt rassuré sur le fait que *SIMAU* arrive à sortir une différence entre les scénarios moins marginale que pour la population, que cette différence est positive donc intuitivement va dans le bon sens. Mais cette différence de 10 000 emplois reste encore faible au regard de l'augmentation tendancielle des emplois entre 1999 et 2026 dans le scénario de référence ($10000/92000=11\%$). On peut donc se demander si ce résultat est attribuable à la marge d'erreur du modèle *SIMAU* ou bien si, le modèle étant suffisamment fiable et performant, on a mis le doigt sur un phénomène très intéressant, à savoir le faible impact de la Tangentielle nord sur le développement économique sans actions concomitantes (urbaines, environnementales, paysagères, foncières, fiscales, etc) de valorisation de l'infrastructure.

2^{ème} élément : Comment se répartissent ces 10 000 emplois supplémentaires ? La carte montre les différences en quatre tranches entre les deux scénarios. On constate qu'il y a très peu de

cellules dans lesquelles l'emploi a augmenté (28 cellules exactement, sur 863). Les augmentations se localisent principalement aux extrémités de la ligne, du côté de Sartrouville-Argenteuil et du côté du Bourget-Bobigny-Drancy-Noisy-le-Sec. Les très fortes augmentations se localisent dans 4 cellules seulement, dont deux situées aux environs immédiats de la gare de Noisy-le-Sec. Les quartiers de la gare de Noisy-le-Sec apparaissent comme le nouveau pôle économique du territoire, ce qu'aucun modèle aurait pu prévoir auparavant. La desserte par la Tangentielle nord de cette gare permet de mailler avec le T1, avec son prolongement vers Montreuil, avec le RER E et avec le tram-train T4, créant un effet multiplicateur de réseau. Ce gain d'accessibilité locale et régionale a contribué très fortement dans le choix de localisation des emplois à proximité de la gare.

8.6. Conclusions

C'est la première fois qu'un modèle donne d'une part la répartition géographique des populations et des emplois qui ne sont plus des données exogènes mais des sorties, d'autre part parvient à mettre en évidence des différences entre scénarios sur ces variables de population et d'emplois. Jusqu'à maintenant, on raisonnait, dans les modèles de trafic classiques, à demande constante pour chaque scénario. Mais il faut rester très prudent sur les résultats car les différences calculées par *SIMAU*RIF entre les deux scénarios peuvent être imputées aux incertitudes des données d'entrée et de calculs et se retrouver à l'intérieur de la marge d'erreur.

Les sources d'incertitudes sont multiples :

- incertitude des données d'entrée (valeurs manquantes, perte d'information ou distorsion par croisement de données géoréférencées, biais par fusion de bases de données différentes)
- erreur systématique, par exemple due à la procédure de sélection des alternatives qui n'est pas parfaitement aléatoire
- incertitude des modèles prédictifs : les variables retenues peuvent s'avérer insuffisantes (par exemple il manque les facteurs de qualité de vie dans le modèle de localisation des ménages), le type de modèle lui-même n'est peut-être pas le mieux adapté (probit ou mixed logit plutôt qu'un multinomial logit), ou même le modèle de choix discret n'est qu'une première approximation
- incertitude sur les paramètres généraux : les « total controls », les taux de déménagement, etc
- incertitude due au générateur de nombres aléatoires

Les résultats présentés (différences entre les deux scénarios) sont malheureusement invérifiables, même quand on sera en 2026. En effet, le scénario de référence (celui sans Tangentielle nord) ne verra pas le jour, normalement ...

On peut quand même tirer des enseignements très intéressants pour le cas de la Tangentielle nord. S'ils s'avèrent exacts, les résultats des simulations montrent que l'arrivée de la Tangentielle nord induira des augmentations relativement faibles d'emplois et surtout de population, toutes choses égales par ailleurs. On ne peut donc pas conclure que la mise en service, *seule*, de la Tangentielle nord créera des effets indirects importants sur l'essor économique et l'attractivité résidentielle d'un territoire comportant de nombreuses faiblesses et à l'image quelque peu dévalorisée. Autrement dit, l'étude montre qu'il y a un risque de ne pas exploiter au mieux l'occasion donnée à ce territoire. Il faudra une forte volonté politique et encore beaucoup d'investissements pour accompagner l'arrivée de la Tangentielle nord et pour maximiser les retombées économiques de sa mise en service.

Ces conclusions rejoignent celles des urbanistes qui ont travaillé de manière qualitative sur le diagnostic territorial de la Tangentielle nord (voir rapport [9]).

Les propositions pour accompagner l'investissement de la Tangentielle nord ne manquent pas :

- renforcer les pôles et favoriser une synergie entre ces pôles
- revitaliser le secteur central
- favoriser l'implantation de grands équipements aux abords des gares
- créer un lieu d'échange et de coordination sur le développement du grand territoire
- donner une ambition forte à certains secteurs : des lieux d'intérêt régional
- mettre en place des outils d'interventions foncières et opérationnelles
- favoriser des opérations nouvelles d'habitat aux abords des gares
- favoriser le renouvellement dans le tissu urbain
- améliorer le cadre de vie au sens large : environnement, emplois, services
- améliorer l'habitat ancien
- ouvrir un partenariat d'investissement sur le territoire

9. Conclusions et perspectives

Au terme de 3 ans et demi de travail, ce projet de recherche innovant et complexe permet de tirer de nombreux enseignements :

1) De la théorie à la pratique : un passage très complexe

Dans un modèle d'urbanisation (« Land use model ») qui est un assemblage de plusieurs modèles, chacun de ceux-là pourrait faire l'objet d'une thèse (voir annexe 1 des résumés d'articles publiés sur un aspect particulier du projet *SIMAU*RIF). Nous avons essayé d'implémenter au cours de notre recherche un aspect particulier : l'interaction entre modèle d'urbanisation et modèle de trafic. Une revue de la littérature a montré que les thèmes de l'interaction transport-urbanisation, des effets redistributifs indirects des transports et des évaluations socio-économiques des transports ne traitent dans leur quasi-totalité que d'aspects théoriques. Lorsqu'il s'agit de passer à la pratique, il n'y a plus grand monde. Nous avons pu acquérir de l'expérience dans le développement d'un modèle intégré urbanisation-transport qui est d'une complexité insoupçonnable au départ, due à la prise en compte de la temporalité et d'une multitude de facteurs en jeu.

2) L'élaboration d'une méthodologie : un exemple à suivre

Il nous semble que ce projet a été mené avec une méthodologie qui pourrait faire école. Cette méthodologie repose sur les principes suivants : la prise en compte des contraintes environnementales (zones inondables, zones de protection des monuments, ZNIEFF, etc) et de politique locale (Plans Locaux d'Urbanisme), une volonté de donner une cohérence entre les entrées/sorties des différents modèles, une adaptation de la structure de données d'UrbanSim au cas de l'aire d'étude (en l'occurrence la Région d'Ile-de-France), une volonté d'exploiter au maximum les données au niveau infracommunal, la construction de bases de données désagrégées exhaustives et non synthétisées comme aux Etats-Unis, un souci permanent d'analyser les données de manière descriptive avant de passer à la modélisation, une base de données unique pour estimer les différents modèles de choix discret, un calage de modèle indépendamment les uns des autres puis un calage global, la validation de chaque modèle et la validation de l'ensemble, la distinction claire entre une période de calage et une période de simulation du futur.

3) Les bases de données infracommunales : un effort sans précédent de construction

Les données sont les plus riches possibles. Partant du principe anglo-saxon « Garbage in, garbage out », nous avons fourni un effort inédit pour construire les bases de données, et ce grâce à la puissance d'un Système d'information géographique. Les données ont ainsi été préparées dans un découpage de la région en carrés de 500 mètres de côté (50 000 cellules), ce qui ne s'est jamais vu en Ile-de-France. Les bases de données ne sont néanmoins pas complètes pour l'implémentation d'autres fonctionnalités de l'outil (par exemple les surfaces bâties qui doivent être extraites du plan cadastral). Mais la question fondamentale est de savoir si ce découpage fin est vraiment pertinent, et si oui, est-il rentable ? Nous répondons oui, sauf pour le modèle de prix. Il a été impossible d'estimer un modèle de prix du logement dans un découpage aussi fin et sur une aussi large aire d'étude. En revanche, pour les autres modèles, le découpage en cellules carrées a permis deux apports intéressants : la prise en compte de la notion de voisinage ou proximité et la possibilité de sortir des résultats à l'échelle infracommunale (autour des gares de

la Tangentielle nord par exemple). La fiabilité des résultats reste cependant à démontrer à cette échelle. Il y a sans doute une nécessité d'estimer des modèles plus fins, comportant plus de variables que celles que l'on a retenues, pour que les résultats soient eux-mêmes exploitables à un niveau plus fin (on a parfois l'impression de vouloir mesurer la taille d'une bactérie avec un double-décimètre ...). Mais on est ensuite confronté au problème d'élaborer des projections de ces nouvelles variables. On en arrive à devoir faire compromis, pas toujours satisfaisant intellectuellement.

4) Estimation économétrique des modèles : modèle explicatif contre modèle prédictif

D'abord, il a fallu estimer chaque modèle de manière indépendante. La réalité de la pratique a obligé à forger un compromis : pour développer un outil, il vaut mieux se contenter d'un modèle simple mais prédictif, plutôt qu'un modèle explicatif sophistiqué mais inapplicable en prévision dans la version actuelle d'UrbanSim. Il faut reconnaître que nous avons consacré trop de temps à l'estimation indépendante de chaque modèle. Ces modèles plus fins et plus sophistiqués n'ont pas été implémentés dans la version actuelle de *SIMAU*RIF, mais nous envisageons de le faire dans des développements futurs.

Rappelons un passage d'anthologie du roman « Vendredi ou les limbes du Pacifique » de Michel Tournier (1967) : après avoir trouvé un secours moral dans la lecture du chapitre du déluge et de l'arche, Robinson entreprend, pour rejoindre le Chili, la construction d'un bateau qu'il appelle "L'évasion". Il consacre beaucoup d'énergie et d'imagination à construire ce bateau. Mais, au moment où il est presque achevé, il se rend compte qu'il lui sera impossible de le mettre à flot à cause de son poids. Par analogie, si on met au point un modèle explicatif trop sophistiqué, il ne sera pas possible de l'appliquer en prévision car les variables ou les données à utiliser ne seront pas disponibles. De toutes façons, ce ne sont pas tant les résultats des estimations qui sont importantes mais la méthodologie avec laquelle elles sont menées.

5) Projet dépendant d'un logiciel, UrbanSim, pur produit de la recherche.

La version OPUS/UrbanSim 4 relève d'une logique tout à fait différente de la version 2. Pour faire court, c'est « macrolangage contre boîte noire ». En tant qu'utilisateur, il est clair que notre préférence va pour une boîte noire comportant des modèles pré-établis interconnectés entre eux, qu'il suffirait d'estimer sans aucune programmation supplémentaire. On se concentrerait sur notre métier qui est de la modélisation et non de la programmation. Il s'est avéré que le choix d'UrbanSim a été tout à fait pertinent au départ mais que le changement de version en cours de projet a eu pour conséquence un travail d'adaptation et de simplification de nos données et de nos modèles qui n'était pas prévu (par exemple regroupement des activités en une seule). De plus, la documentation de la nouvelle version n'est pas du tout au point. Il apparaît au travers des messages envoyés à la liste de diffusion des utilisateurs d'UrbanSim que la nouvelle version a fortement déstabilisé les projets en cours et que tous les utilisateurs ont été bloqués sur des problèmes purement informatiques.

Plus généralement, l'architecture du logiciel est tout à fait pertinente et logique. Elle prend bien en compte la localisation des ménages et des emplois, les prix du foncier et les projets de développement urbain. Mais c'est l'interaction entre ces quatre composants qui est mal programmée par l'équipe de développement de l'Université de Washington. De plus, le choix du langage Python lors de la ré-ingénierie du logiciel en 2006 n'a pas été le meilleur choix car il nécessite de faire appel à un spécialiste de ce langage.

A la date de fin novembre 2007, à notre connaissance, aucun projet hormis aux Etats-Unis utilisant UrbanSim n'est encore terminé. Serions-nous les premiers ? Il faudra publier pour le montrer.

6) Problème de la transposition d'un modèle conçu pour des villes américaines au cas français

Cette recherche a montré qu'il est difficile de transposer un modèle qui simule le développement urbain dans un contexte donné (Etats-Unis) à un autre contexte très différent (France). Le modèle UrbanSim est conçu pour simuler des métropoles américaines, caractérisées par une urbanisation non contrôlée qui s'étale, une grande consommation d'espaces, des densités urbaines relativement faibles et un système de transport essentiellement routier privilégiant la vitesse et la capacité. En France, le développement des villes se fait davantage par une organisation efficace de l'espace que par une extension du réseau routier. Par exemple, UrbanSim ne prévoit pas dans sa structure initiale de données sur les transports en commun ! Donc l'adaptation au cas de l'Ile-de-France a été rendue possible au prix d'adaptations parfois très lourdes et dévoreuses en temps de travail, et parfois très simplificatrices et intellectuellement peu satisfaisantes.

7) Programmation de l'interface gérant l'interaction transport-urbanisation : un objectif partiellement atteint

Ce volet de *SIMAU*RIF est tout à fait compréhensible d'un point de vue conceptuel. Mais là encore, c'est le passage à la pratique qui est complexe. Il s'agit d'automatiser l'exécution des 4 étapes d'un modèle de trafic, d'exporter les résultats dans la base de données du modèle d'urbanisation, d'enchaîner l'exécution des différents modules du modèle d'urbanisation, et enfin d'exporter les résultats de ces modules dans la base de données de génération du modèle de trafic. Puis de refaire tourner sans s'arrêter l'ensemble sur 27 années, en modifiant l'offre de transport par période (de trois ans par exemple). Nous admettons que nous n'avons que partiellement atteint cet objectif. L'interface de couplage entre les trois ensembles de modèles a été simplifiée jusqu'à sa plus simple expression : un script écrit en Python permettant de prendre en compte l'offre de transport de manière exogène et seulement sur des réseaux à vide (c'est-à-dire sans affectation de trafic). Il faudrait reprendre le prototype d'interface développé en 2005 et l'adapter à la nouvelle version d'UrbanSim.

8) Nécessité d'une bonne connaissance du fonctionnement de la région d'Ile-de-France

Ce projet de recherche ne se résume pas à de l'économétrie. Pour le réussir, il faut une vision globale des interactions entre transports et urbanisation, une parfaite compréhension des processus de développement et d'aménagement de l'aire d'étude, une parfaite connaissance des projets tant urbains que de transports, une parfaite connaissance de la géographie de l'emploi, de l'habitat et de l'évolution de l'occupation du sol, et enfin une expérience des jeux d'acteurs. Un grand projet urbain ne voit le jour qu'au bout d'une génération au moins, comme la Défense, la Plaine Saint-Denis, les Villes nouvelles. D'autres grands projets portés par les décideurs politiques sont en cours, qu'il faut savoir d'une manière ou d'une autre prendre en compte dans les scénarios (par exemple, le territoire Seine-Amont-Orly-Rungis, les pôles de compétitivité, le secteur Seine Rive-Gauche à Paris).

9) L'approche territoriale et la simulation pour évaluer les effets indirects des transports sur l'urbanisation : *SIMAU*RIF le permet en partie.

Les effets directs d'une infrastructure de transport se mesurent essentiellement en gain de temps. L'approche micro-économique permet de formaliser un taux de rentabilité interne d'un projet. En revanche, pour évaluer des effets indirects, la mise en équations n'est pas possible. Il y a en effet une multitude de relations entre l'attractivité résidentielle d'un territoire, l'accessibilité transport, les prix (du foncier, de l'immobilier, des commerces, des bureaux), le développement économique. Seule l'approche territoriale, la simulation et un ensemble de modèles de choix discrets peuvent tenter de quantifier une partie de ces effets. Pour la 1^{ère} fois en France, un outil de simulation a pu prévoir une différence de répartition des populations et des emplois à un horizon futur entre deux scénarios ne différant que par une offre de transport nouvelle. En revanche, *SIMAU*RIF simule les prix de l'immobilier résidentiel à la commune et ne permet donc pas d'évaluer la fameuse rente foncière qui ne peut se calculer qu'à un niveau infracommunal.

10) Nécessité d'une équipe pluridisciplinaire et de moyens appropriés

Ce thème de recherche est réellement pluri-disciplinaire. Autant la modélisation de trafic ne nécessite que des compétences d'ingénieur-modélisateur, autant la modélisation de l'interaction transport-urbanisation doit faire appel à des compétences beaucoup plus larges : économètre, démographe, économiste, urbaniste, informaticien. On s'en est bien rendu compte au cours de notre recherche : une seule compétence manque et tout est dépeuplé. Malgré une équipe restreinte, nous avons pu terminer ce projet. Si l'objectif est d'aboutir à un outil utilisable, il faut aussi prendre garde à ne pas laisser le projet être accaparé par l'informatique et par l'économétrie. C'est le modélisateur qui doit rester maître du projet et limiter les ambitions des informaticiens et des économètres. On perçoit bien ce qui se passe depuis un an dans l'équipe du Pr Paul WADDELL à l'Université de Washington. Il semble que les informaticiens ont pris le dessus sur les urbanistes, d'où l'orientation actuelle du logiciel qui devient essentiellement un projet universitaire de développement pour des étudiants en thèse d'informatique (PhD in Computer science). Bien sûr, si le projet est plus orienté vers la recherche académique, dans ce cas, le modélisateur peut se mettre plus en retrait. Il faut donc être clair au niveau des objectifs.

Perspectives

A court terme, il nous paraît intéressant d'utiliser *SIMAU*RIF pour évaluer un autre projet d'envergure tel que le projet Arc-Express (rocade en métro en proche couronne). La base de données au niveau régional est déjà construite, le modèle est déjà calé sur la période 1990-1999, même s'il faut encore parfaire ce calage. Il n'y a plus qu'à construire les scénarios, les graphes et lancer les simulations. Par ailleurs, il faudra valoriser le projet *SIMAU*RIF au niveau international auprès de la communauté des chercheurs par des communications (TRB, WCTR, ERSA, ...) et plus près de nous, auprès des utilisateurs potentiels en Ile-de-France (DREIF, STIF, SNCF, RATP).

A moyen terme, il nous semble que si l'on veut développer un outil combinant un modèle d'urbanisation et un modèle de trafic qui soit convivial et prêt à utiliser, il vaut mieux partir de zéro (sauf peut-être pour la 4^{ème} étape, l'affectation des matrices sur les réseaux, qui est parfaitement implémentée dans les logiciels du commerce), plutôt que d'essayer d'interfacer des logiciels existants. Désormais, nous maîtrisons l'architecture globale d'un outil combiné, avec

ses différentes briques. Il est préférable de programmer soi-même ces modules dans un langage informatique simple (Visual Basic ou langage C par exemple), une fois qu'ils ont été estimés de manière économétrique, plutôt que de dépendre d'un logiciel qui est un pur produit du monde universitaire, et de programmer la boucle de rétroaction entre les trois ensembles de modèles (modèle d'urbanisation, modèle de demande de déplacements et modèle d'affectation de trafic à l'heure de pointe) dans un cadre non contraint et adapté spécifiquement au cas de la Région d'Ile-de-France.

A plus long terme, il faudrait envisager d'interconnecter à ce modèle intégré transport-urbanisation un modèle d'évaluation environnementale (calcul d'émission et de diffusion de polluants, calcul d'émission de bruit) pour pouvoir évaluer des indicateurs de développement durable.

Quoi qu'il en soit, ce projet aura indéniablement permis de faire avancer l'état de la pratique de la modélisation de l'interaction transport-urbanisation en France.

Bibliographie

- [1] IAURIF-THEMA/T&R. « *SIMAU*RIF – Rapport final de 1^{ère} phase. (sept. 2004) » - Pour le compte de la DRAST/PREDIT, N°03MT11
- [2] IAURIF-THEMA/T&R. « *SIMAU*RIF –Rapport intermédiaire de 2^{ème} phase (oct. 2005) » Pour le compte de la DRAST/PREDIT, N°05MT6008
- [3] NGUYEN-LUONG Dany. IAURIF. « Contribution au projet *SIMAU*RIF. Recueil et analyse de données de prix de l'immobilier résidentiel en Ile-de-France. Cas du T1 et d'EOLE. Estimation d'un modèle de prix hédonique » (déc. 2006). – Pour le compte de RFF.
- [4] DE PALMA André, MOTAMEDI Kiarash, MOYANO José, PICARD Nathalie. THEMA/T&R. « Décisions de (dé)localisation des emplois en Ile-de-France.» (déc. 2007). – Pour le compte de la DREIF.
- [5] CUSPA. « Opus : The Open Platform for Urban Simulation and UrbanSim Version 4. Reference Manual and Users Guide. The UrbanSim Project.» (27 mars 2007).
- [6] DE PALMA André, MOTAMEDI Kiarash, PICARD Nathalie, WADDELL Paul. «A model of residential location choice with endogenous housing prices and traffic for the Paris region Discrete choice models with capacity constraints: An empirical analysis of housing market of greater Paris region», European Transport, 2005.
- [7] DE PALMA André, MOTAMEDI Kirash, PICARD Nathalie, WADDELL Paul. «Accessibility and environmental quality: inequality in the Paris housing market», European Transport, 2007.
- [8] DE PALMA André, MOTAMEDI Kiarash, NGUYEN-LUONG Dany, OUARAS Hakim, PICARD Nathalie, Development of a dynamic integrated Land use – Transportation model Application to the Paris Region (second phase), présenté à la 11^{ième} Conférence mondiale de recherche en transports (WCTR), Berkley, USA, 2007.
- [9] GORRAND François, NAVARRE Danièle. IAURIF. « Diagnostic territorial de la Tangentielle Nord » (nov. 2005). – Pour le compte du STIF.

Annexe 1 : Résumés d'articles publiés

Nous présentons ici les résumés de trois articles de recherche publiés dans le cadre de ce projet.

1) Journal of Urban Economics 62 (2007) 204–230

Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater Paris region

Résumé

Les modèles de choix discrets sont basés sur l'idée que chaque agent peut choisir parmi l'ensemble des alternatives, librement et indépendamment des autres usagers. Mais cela n'est pas le cas dans certaines situations. En particulier, des limites et interactions peuvent intervenir quand la quantité disponible d'un produit est inférieure à la demande de ce type de produits. comme conséquence, certains individus peuvent renoncer à leur choix préféré. de Palma A., Picard N. et Waddell ont développé une méthodologie pour tenir compte de ces contraintes et l'ont appliqué au choix de localisation résidentiel où ils ont montré que l'existence de ces contraintes peut biaiser les choix actuels. L'analyse présentée par ces auteurs fournit des développements théoriques et élabore un procédure itérative afin d'estimer la demande en présence des contraintes de capacité. L'application empirique est basée sur le modèle de choix de localisation résidentielle développé et estimé pour la région d'Île-de-France et généralisé pour intégrer les contraintes de capacité.

2) European Transport \ Trasporti Europei n. 36 (2007): 47-74

Accessibility and environmental quality: inequality in the Paris housing market

Abstract

In this paper we examine empirically the market for local amenities in the Paris metropolitan region. We find first that there is considerable inequality in the spatial distribution of these local amenities, including accessibility, environmental and social indicators. We use a spatial representation and Lorenz curves to examine the degree of inequality in these amenities, and this provides evidence that some amenities (or disamenities) are much more inequitably distributed than others. The most extremely unequally distributed amenities are noise (due to its concentration near airports), "Redevelopment Areas", presence of water (lakes and rivers) and forests, and presence of train and subway stations. Some indicators, such as the "Poult accessibility" measure, were by contrast remarkably constant over the region. We recognize that local amenities should be capitalized into the housing market, and explore the willingness to pay of households for these amenities within the Paris region using alternative specifications of a location choice model. One of the core questions we examine is the spatial scale of the amenity effects and how this is captured in a location choice context. By estimating models at both a commune and at a grid cell level, we obtain new insights into how households in the Paris region trade off amenities against each other and against housing cost. We find that the residential location choice model fits the data moderately better at the smaller scale of the grid cell compared to the commune.

3) European Transport \ Trasporti Europei n. 31 (2005): 67-82

A model of residential location choice with endogenous housing prices and traffic for the Paris region

Abstract

There is a growing interest in the development and the use of large-scale planning models. In this paper, we describe the first step of a project to integrate UrbanSim, a dynamic microsimulation land use model, and METROPOLIS, a dynamic traffic model. This is the first attempt, to our knowledge, to integrate a dynamic land use model and a dynamic traffic model. We briefly describe the two models and propose a unified framework for their integration. Within this integrated framework we develop a model of residential location choice, with endogenous housing prices and traffic. The study area for this research is the Ile-de-France (Paris region), for which we provide empirical results.

Annexe 2 : Un modèle explicatif de choix de localisation résidentielle des ménages

Parallèlement à l'implémentation du modèle dans UrbanSim, une étude complémentaire et plus riche a été réalisée pour expliquer au mieux le phénomène de localisation des ménages. Pour cela deux estimations ont été réalisées, une au niveau de la commune, l'autre au niveau de la cellule. Les pseudo- R^2 de McFadden sont de 0.2414 et 0.2272, respectivement (voir [7]).

Les résultats sont les suivants :

Variable	Estim. à la commune		Estim. à la cellule	
	Coefficient	t-stat	Coefficient	t-stat
Même dept. que celui avant déménagement	2.53879	275.91	2.53494	272.47
Paris	-0.25412	-10.10	-0.16871	-10.29
Log(Prix d'appartement)	0.00644	0.20	0.02178	0.94
Log(Prix d'appartement)* (Age-40)/10	-0.02976	-1.74		
Log(Prix d'appartement)* (Log(Revenu)- $\overline{\text{Log(Income)}}$)	0.40919	8.20	0.24849	5.84
Log(Prix d'appartement)* (Nb membres de Men. - 1)	-0.04236	-2.40		
Log(Prix de maison)	0.08735	3.70	0.13934	7.41
Log(Prix de maison)* (Age-40)/10	-0.10359	-8.78	-0.11371	-20.90
Log(Prix de maison)* (Log(Revenu)- $\overline{\text{Log(Income)}}$)	0.18693	5.78	0.14874	5.17
Log(Prix de maison)* (Nombre de Men. membres - 1)	-0.10932	-8.38	-0.13038	-20.43
Nombre stations Métro au voisinage			-0.00412	-2.37
Nombre stations Métro dans la commune / cell	0.00518	4.59	0.00146	0.21
Nombre stations Train au voisinage			0.01326	3.85
Nombre stations Train dans la commune / cell	-0.00940	-4.86	0.00616	0.60
Temps moyen en TC partant de (TC) [hr]	0.02483	0.80	0.04007	1.45
TC*(Dummy femme) [hr]	-0.37377	-8.28	-0.29400	-6.56
Distance à l'autoroute la plus proche [km]	-0.00594	-3.10	-0.00146	-0.74
Distance à la route nationale [km]	-0.00798	-2.98	-0.01500	-5.87
Distance à Châtelet (Paris centre) [km]	0.00167	2.94	-0.00054	-1.06
% ménages avec 1 membre * Men avec 1 membre	1.87670	20.62	2.27023	30.73
% ménages avec 2 membres* Men. avec 2 membres	1.33649	4.20	1.77322	10.03
% ménages avec 3+ membres* Men. avec 3 et+ membres	2.22967	22.16	2.07516	27.98
% Men. sans membre actif * sans membre dans le Men.	7.79690	33.91	5.38190	33.49
% Men. avec 1 membre actif * 1 membre actif dans Men.	-0.93134	-5.84	0.50248	4.58
% Men. avec 2+ membres actifs * 2 et+ actifs dans Men.	1.66425	14.83	0.40430	5.01
%Men à chef jeune	0.45006	3.01	1.02140	10.86
% Men. à chef jeune * chef jeune dans Men.	4.77740	28.34	3.24295	29.59
% Men. à chef à âge moyen * chef à âge moyen	-0.69337	-4.16	-0.16208	-1.51
% Men. riche * Men. Riche	3.28038	28.36	2.87636	36.58
% Men. à revenu moyen * Men. à revenu moyen	1.62120	10.11	1.92296	17.00
% Poor Men. * poor h	0.45889	3.53	1.06997	12.11
% Men. à chef étranger * Men. à chef étranger	5.04570	26.74	4.57810	37.26
% Men. à chef étranger * Men. à chef français	-1.99493	-17.44	-1.30553	-16.71
% surface dans Zone Sensible * Men. Riche	0.67811	4.97	-0.22815	-4.28
% surface dans Zone Sensible * Men. à revenu moyen	-0.08340	-0.78	-0.19330	-5.16
% surface dans Zone Sensible * Men. Pauvre	0.43164	4.45	0.16457	4.99

Tableau 33 : Modèle explicatif de localisation des ménages

Notes : “%” représente la proportion.

“N. d’enfants” compte les membres de la famille âgés de 11 ans ou moins.

Quelques variables ont été omises pour une meilleure présentation.

Visiblement, d’après le coefficient de la variable « Même dept. que celui avant déménagement », les ménages préfèrent se localiser dans le même département que celui dans lequel ils étaient avant le déménagement. Paris n’est pas attractive dans les deux niveaux d’estimation. Ils préfèrent habiter loin des autoroutes et routes nationales mais pas trop loin de Châtelet. Chaque

type de ménages cherche à se localiser dans une commune ou cellule qui contient le plus de ménages de son type. Ainsi, la probabilité d'un ménage à un seul membre de se localiser dans une commune (cellule) augmente quand la proportion de ménages avec un seul membre est élevée. De même, la probabilité de se localiser pour un jeune ménage dans une commune (cellule) augmente quand la proportion de jeunes ménages augmente.

Annexe 3 : Un autre modèle de choix de localisation des emplois et un modèle de choix de localisation des entreprises

Estimation du nombre d'emplois à la cellule

Dans les ERE, les emplois sont (presque) tous localisés à la commune, alors que la géolocalisation à la cellule est absente en 1997 et très partielle en 2001. Un effort particulier de géolocalisation a été fourni à Paris et dans les communes limitrophes, ainsi que pour les grands établissements et ceux de plus de 10 salariés.

Nous ne connaissons donc qu'une partie de l'effectif de chaque cellule. Dans nos estimations, le nombre d'emplois par cellule sert de proxy pour le nombre de places pouvant être occupées par des emplois. Cette information doit être incluse dans les estimations afin de pouvoir comparer les estimations faites à la cellule et à la commune.

Partant de ces constatations, nous estimons des régressions qui donnent, pour chaque commune, le nombre d'emplois non géolocalisés ($NEmp_s^n$) en fonction du nombre des emplois géolocalisés ($NEmp_s^g$), du nombre d'établissements géolocalisés ($NEtb_s^g$) et de l'occupation du sol (MOS_m , en 21 postes) dans chaque commune.

$$NEmp_s^n = Cte + \sum_{m=1}^{21} a_{m,s} * MOS_m + b_s * NEmp_s^g + \sum_{z=1}^3 c_s^z * NEmp_s^g * I^z + \sum_{t=1}^3 \sum_{z=1}^3 d_{s,t}^z * NEmp_t^g * I^z + e_s * NEtb_s^g + \sum_{z=1}^3 f_s^z * NEtb_s^g * I^z + \sum_{t=1}^3 \sum_{z=1}^3 g_t^z * NEtb_t^g * I^z + \sigma * \varepsilon$$

Afin de tenir compte de l'effort différencié de géolocalisation selon la taille de l'établissement et sa zone géographique, nous avons ajouté à la liste des variables explicatives les nombres d'emplois géolocalisés et d'établissements géolocalisés par classe de taille t (1 pour petit, 2 pour moyen et 3 pour grand), et nous avons estimé des coefficients différents pour ces variables en fonction de la zone géographique (I^z est l'indicatrice de la zone géographique : $z=1$ pour Paris, $z=2$ pour les communes limitrophes de Paris et $z=3$ pour le reste des communes de la petite couronne). La grande couronne sert alors de référence.

Nous avons ensuite appliqué les coefficients de cette régression aux cellules afin d'obtenir le nombre total d'emplois par cellule.

Toutes les estimations sont faites par secteur. Nous avons testé deux formulations : l'une en niveau (nombre d'emplois) et l'autre en logarithme. Nous avons aussi testé l'utilité d'estimer deux modèles distincts : pour Paris et les communes limitrophes d'une part et pour le reste des communes d'autre part.

Le tableau 34 présente le R^2 ajusté des régressions pour les différents modèles et les différents secteurs. La distinction de Paris et des zones limitrophes donne des résultats mitigés, ainsi que l'estimation d'un modèle en logarithme plutôt qu'en niveau. Par souci de simplicité, et en l'absence de supériorité marquée des modèles alternatifs, nous avons donc finalement opté (voir tableau 35) pour une spécification en niveau sur l'ensemble des communes (une régression par secteur).

Secteur	Zone	Logarithme	Niveau
1	Total	58,0%	55,7%
1	Paris et limitrophe	58,9%	57,7%
1	Autre	38,8%	37,6%
2	Total	56,9%	42,7%
2	Paris et limitrophe	56,1%	39,2%
2	Autre	63,4%	78,7%
3	Total	77,0%	85,4%
3	Paris et limitrophe	77,0%	86,5%
3	Autre	60,3%	44,6%
4	Total	56,0%	60,0%
4	Paris et limitrophe	53,7%	56,9%
4	Autre	56,5%	70,5%
5	Total	77,1%	80,4%
5	Paris et limitrophe	76,1%	77,8%
5	Autre	34,6%	77,0%
6	Total	83,7%	76,2%
6	Paris et limitrophe	82,7%	73,9%
6	Autre	73,0%	54,5%
7	Total	84,2%	82,1%
7	Paris et limitrophe	83,6%	82,8%
7	Autre	67,1%	68,1%
8	Total	71,1%	79,0%
8	Paris et limitrophe	67,5%	81,6%
8	Autre	17,2%	31,6%

Tableau 34 : Emploi géolocalisé et non géolocalisé, par secteur

Secteur		1	2	3	4	5	6	7	8
Constante	Coeff	-1,2382	7,8980	4,6030	4,9019	-1,1224	-9,4926	-22,2053	-3,2392
	t Stat	<i>(-1,70)</i>	<i>(3,07)</i>	<i>(0,98)</i>	<i>(3,21)</i>	<i>(-0,83)</i>	<i>(-0,77)</i>	<i>(-1,94)</i>	<i>(-0,28)</i>
Superficie de MOS type 1 En Km2	Coeff	0,8735	0,0232	0,1964	0,0370	-0,0018	-1,5040	-1,2126	-0,5297
	t Stat	<i>(11,05)</i>	<i>(0,08)</i>	<i>(0,38)</i>	<i>(0,22)</i>	<i>(-0,01)</i>	<i>(-1,11)</i>	<i>(-0,98)</i>	<i>(-0,42)</i>
Superficie de MOS type 2 En Km2	Coeff	0,0658	0,1530	-0,5428	-0,2104	-0,1726	-0,6893	2,8032	2,4474
	t Stat	<i>(0,79)</i>	<i>(0,52)</i>	<i>(-1,01)</i>	<i>(-1,20)</i>	<i>(-1,12)</i>	<i>(-0,49)</i>	<i>(2,16)</i>	<i>(1,84)</i>
Superficie de MOS type 3 En Km2	Coeff	2,4845	14,1796	46,8672	4,2905	5,4991	25,2543	34,4908	-27,4421
	t Stat	<i>(0,91)</i>	<i>(1,46)</i>	<i>(2,65)</i>	<i>(0,75)</i>	<i>(1,09)</i>	<i>(0,54)</i>	<i>(0,81)</i>	<i>(-0,63)</i>
Superficie de MOS type 4 En Km2	Coeff	2,0224	-3,6583	-11,0653	-2,2871	-3,6634	40,2853	33,0626	5,6246
	t Stat	<i>(1,13)</i>	<i>(-0,57)</i>	<i>(-0,95)</i>	<i>(-0,60)</i>	<i>(-1,10)</i>	<i>(1,32)</i>	<i>(1,18)</i>	<i>(0,20)</i>
Superficie de MOS type 5 en Km2	Coeff	-1,6667	9,3041	3,8389	3,9368	0,9509	-18,1546	-41,2138	58,4687
	t Stat	<i>(-1,14)</i>	<i>(1,79)</i>	<i>(0,41)</i>	<i>(1,28)</i>	<i>(0,35)</i>	<i>(-0,73)</i>	<i>(-1,80)</i>	<i>(2,51)</i>
Superficie de MOS type 6 en Km2	Coeff	-2,3318	2,7488	2,4670	0,3466	-2,4945	131,5961	8,7640	44,9881
	t Stat	<i>(-1,37)</i>	<i>(0,45)</i>	<i>(0,22)</i>	<i>(0,10)</i>	<i>(-0,79)</i>	<i>(4,56)</i>	<i>(0,33)</i>	<i>(1,67)</i>
Superficie de MOS type 7 en Km2	Coeff	-4,6107	-5,8776	-11,5951	-8,8795	4,4437	-19,1641	-85,6672	36,4848
	t Stat	<i>(-1,85)</i>	<i>(-0,67)</i>	<i>(-0,72)</i>	<i>(-1,70)</i>	<i>(0,97)</i>	<i>(-0,46)</i>	<i>(-2,18)</i>	<i>(0,92)</i>
Superficie de MOS type 8 en Km2	Coeff	-15,6944	-19,6564	-132,4617	-5,7100	-21,4493	-195,2481	-5,0853	-295,1922
	t Stat	<i>(-2,46)</i>	<i>(-0,86)</i>	<i>(-3,20)</i>	<i>(-0,43)</i>	<i>(-1,81)</i>	<i>(-1,79)</i>	<i>(-0,05)</i>	<i>(-2,90)</i>
Superficie de MOS type 9 en Km2	Coeff	5,6902	13,6709	67,7702	6,0850	12,4380	-15,7887	6,7061	-85,7107
	t Stat	<i>(8,07)</i>	<i>(5,39)</i>	<i>(14,21)</i>	<i>(4,13)</i>	<i>(9,41)</i>	<i>(-1,28)</i>	<i>(0,60)</i>	<i>(-7,60)</i>
Superficie de MOS type 10 en Km2	Coeff	33,4560	49,3300	298,4988	24,7600	59,0203	-45,5569	204,7763	-174,7374
	t Stat	<i>(9,69)</i>	<i>(4,10)</i>	<i>(12,81)</i>	<i>(3,43)</i>	<i>(8,73)</i>	<i>(-0,72)</i>	<i>(3,34)</i>	<i>(-3,11)</i>
Superficie de MOS type 11 en Km2	Coeff	55,5565	51,1022	131,8141	-87,5023	118,7340	797,0227	398,5688	3120,3298
	t Stat	<i>(2,33)</i>	<i>(0,60)</i>	<i>(0,84)</i>	<i>(-1,72)</i>	<i>(2,67)</i>	<i>(1,93)</i>	<i>(1,05)</i>	<i>(8,16)</i>
Superficie de MOS type 12 en Km2	Coeff	3,8896	75,9260	69,9360	14,9343	6,1800	-78,0532	-97,5648	-106,1287
	t Stat	<i>(1,84)</i>	<i>(7,89)</i>	<i>(4,51)</i>	<i>(2,99)</i>	<i>(1,55)</i>	<i>(-2,11)</i>	<i>(-2,95)</i>	<i>(-3,20)</i>
Superficie de MOS type 13 en Km2	Coeff	-28,5468	49,2256	640,8508	-35,9198	-108,7148	249,8239	867,0750	-13,1775
	t Stat	<i>(-1,88)</i>	<i>(0,87)</i>	<i>(5,24)</i>	<i>(-1,11)</i>	<i>(-3,63)</i>	<i>(0,88)</i>	<i>(3,67)</i>	<i>(-0,05)</i>
Superficie de MOS type 14 en Km2	Coeff	-6,5015	17,7611	-35,9046	-8,7466	-7,2428	87,2084	-87,5780	-177,7352
	t Stat	<i>(-0,73)</i>	<i>(0,56)</i>	<i>(-0,62)</i>	<i>(-0,46)</i>	<i>(-0,44)</i>	<i>(0,57)</i>	<i>(-0,63)</i>	<i>(-1,25)</i>
Superficie de MOS type 15 en Km2	Coeff	31,4865	-110,9122	-242,1911	-35,2209	12,3547	1087,2493	3147,0516	186,2918
	t Stat	<i>(2,62)</i>	<i>(-2,58)</i>	<i>(-3,13)</i>	<i>(-1,38)</i>	<i>(0,55)</i>	<i>(5,18)</i>	<i>(16,02)</i>	<i>(0,93)</i>
Superficie de MOS type 16 en Km2	Coeff	3,6807	-117,7064	-112,4678	-80,4057	63,0658	-126,0288	-1571,3273	384,1052
	t Stat	<i>(0,29)</i>	<i>(-2,58)</i>	<i>(-1,38)</i>	<i>(-3,07)</i>	<i>(2,65)</i>	<i>(-0,58)</i>	<i>(-5,89)</i>	<i>(1,90)</i>

Secteur		1	2	3	4	5	6	7	8
Superficie de MOS type 17 en Km2	Coeff t Stat	-5,1217 (-0,59)	-39,9172 (-1,30)	32,0660 (0,58)	-72,6044 (-3,95)	-10,4980 (-0,66)	270,9332 (1,82)	10,8938 (0,08)	161,6615 (1,18)
Superficie de MOS type 18 en Km2	Coeff t Stat	14,9358 (4,80)	16,9101 (1,54)	71,5490 (3,53)	0,1109 (0,02)	24,4534 (4,15)	324,8891 (5,90)	35,1380 (0,70)	423,2657 (8,40)
Superficie de MOS type 19 en Km2	Coeff t Stat	-17,6493 (-3,76)	18,1546 (1,08)	88,9342 (2,93)	0,0000 (-0,00)	2,2356 (0,26)	-48,2305 (-0,60)	64,7086 (0,87)	-69,1304 (-0,92)
Superficie de MOS type 20 en Km2	Coeff t Stat	2,3985 (1,21)	0,0764 (0,01)	53,0125 (4,05)	18,0662 (3,76)	0,5553 (0,15)	77,4376 (2,22)	48,5844 (1,57)	-27,0523 (-0,86)
Superficie de MOS type 21 en Km2	Coeff t Stat	5,1053 (1,41)	-8,6783 (-0,67)	-28,9873 (-1,24)	10,3165 (1,28)	3,0812 (0,46)	86,7584 (1,37)	-7,3819 (-0,13)	110,8327 (1,92)
Nombre d'emplois géolocalisés de même secteur	Coeff t Stat	6,1501 (0,96)	-8,7944 (-3,98)	-56,9123 (-5,57)	4,7633 (3,51)	2,6478 (0,32)	-58,0712 (-3,93)	228,0344 (6,79)	817,0506 (17,43)
Nombre d'emplois géolocalisés de même secteur * Paris	Coeff t Stat	-32,4647 (-2,46)	6,2409 (0,60)	20,4817 (1,17)	-7,2591 (-2,17)	-2,3489 (-0,28)	145,1781 (6,22)	-296,9946 (-8,63)	-783,1933 (-16,14)
Nombre d'emplois géolocalisés de même secteur * Limitrophe	Coeff t Stat	-45,0593 (-2,15)	9,8301 (0,85)	38,9616 (1,98)	-18,7612 (-1,72)	42,3944 (4,94)	139,5575 (4,03)	-50,1037 (-1,07)	-388,3435 (-4,39)
Nombre d'emplois géolocalisés de même secteur * PC	Coeff t Stat	15,7331 (1,49)	-7,7553 (-1,34)	171,0826 (8,27)	6,1833 (1,49)	25,7285 (2,70)	-149,3342 (-3,84)	25,3346 (0,59)	-14,3281 (-0,22)
Nombre d'emplois géolocalisés tous secteurs, grands étab. * Paris	Coeff t Stat	-1,0109 (-1,00)	3,9147 (0,99)	-17,2161 (-2,53)	2,0099 (0,73)	-2,6221 (-1,24)	-49,0016 (-2,36)	153,0989 (9,18)	-15,4571 (-0,75)
Nombre d'emplois géolocalisés tous secteurs, grands étab. * Limitrophe	Coeff t Stat	1,3147 (0,40)	-14,9455 (-1,05)	22,4935 (1,03)	13,4436 (1,79)	44,6954 (7,05)	240,7147 (3,94)	152,0044 (2,91)	350,8437 (6,17)
Nombre d'emplois géolocalisés tous secteurs, grands étab. * PC	Coeff t Stat	8,8788 (4,67)	-3,0859 (-0,43)	30,3498 (2,58)	14,3480 (3,70)	-4,7117 (-1,42)	-53,5569 (-1,61)	77,9475 (2,71)	-95,3020 (-2,87)
Nombre d'emplois géolocalisés tous secteurs, étab. moyens * Paris	Coeff t Stat	1,2107 (1,17)	-5,5124 (-1,94)	-27,2982 (-6,17)	2,1571 (1,51)	-7,5942 (-5,52)	-5,2249 (-0,37)	-26,5744 (-2,27)	53,5161 (4,79)
Nombre d'emplois géolocalisés tous secteurs, étab. moyens * Limitrophe	Coeff t Stat	-0,6485 (-0,62)	-2,1147 (-0,52)	-1,1798 (-0,18)	1,2475 (0,57)	3,6671 (1,83)	41,1261 (1,47)	-6,0516 (-0,41)	146,5566 (7,07)
Nombre d'emplois géolocalisés tous secteurs, étab. moyens * PC	Coeff t Stat	1,2963 (1,65)	8,0381 (2,80)	-42,5741 (-5,77)	2,1286 (1,26)	-3,0311 (-1,65)	113,2772 (3,91)	46,7592 (3,44)	-31,0123 (-2,34)
Nombre d'emplois géolocalisés tous secteurs, petits étab. * Paris	Coeff t Stat	20,4423 (2,47)	-110,2680 (-2,14)	41,4624 (0,66)	-41,4014 (-2,32)	28,8037 (0,99)	1263,8616 (6,60)	-73,8641 (-0,48)	263,1537 (1,92)
Nombre d'emplois géolocalisés tous secteurs, petits étab. * Limitrophe	Coeff t Stat	52,7688 (1,67)	162,5565 (1,55)	25,2492 (0,12)	-48,7693 (-0,70)	-65,9255 (-1,23)	-812,7587 (-1,45)	-634,2393 (-1,34)	-796,8586 (-1,70)

Secteur		1	2	3	4	5	6	7	8
Nombre d'emplois géolocalisés tous secteurs, petits étab. * PC	Coeff	43,2205	50,8498	-135,6658	-51,5961	-7,0519	215,9504	191,0899	376,0406
	t Stat	(1,85)	(0,63)	(-0,89)	(-1,08)	(-0,17)	(0,57)	(0,52)	(1,04)
Nombre d'étab. géolocalisés de même secteur	Coeff	2,4632	1,9278	8,8880	3,4577	8,1391	18,7721	6,4216	18,4422
	t Stat	(3,14)	(5,38)	(15,95)	(6,39)	(11,93)	(13,67)	(1,64)	(1,85)
Nombre d'étab. géolocalisés de même secteur * Paris	Coeff	-3,3051	-1,8018	-8,6626	-3,7357	-8,0557	-18,6407	-6,1794	-21,6693
	t Stat	(-4,12)	(-4,71)	(-15,36)	(-6,41)	(-11,50)	(-12,88)	(-1,57)	(-2,15)
Nombre d'emplois géolocalisés de même secteur * Limitrophe	Coeff	-45,0593	9,8301	38,9616	-18,7612	42,3944	139,5575	-50,1037	-388,3435
	t Stat	(-2,15)	(0,85)	(1,98)	(-1,72)	(4,94)	(4,03)	(-1,07)	(-4,39)
Nombre d'étab. géolocalisés de même secteur * PC	Coeff	-3,9677	-1,4637	-9,6981	-1,4940	-8,0526	-11,9627	-8,1056	13,9266
	t Stat	(-4,41)	(-3,20)	(-14,14)	(-2,40)	(-10,57)	(-6,63)	(-1,85)	(1,26)
Nombre d'étab. géolocalisés tous secteurs, grands étab. * Paris	Coeff	1,4731	-4,7336	84,8030	-0,5694	9,0900	67,3177	-208,6347	-10,0002
	t Stat	(0,57)	(-0,40)	(4,65)	(-0,09)	(1,64)	(1,25)	(-5,10)	(-0,18)
Nombre d'étab. géolocalisés tous secteurs, grands étab. * Limitrophe	Coeff	-3,6119	25,4611	-59,9940	-31,4714	-76,9751	-695,1534	-289,5408	-528,4352
	t Stat	(-0,59)	(1,04)	(-1,48)	(-2,24)	(-6,67)	(-6,15)	(-2,93)	(-5,01)
Nombre d'étab. géolocalisés tous secteurs, grands étab. * PC	Coeff	-14,6483	17,6784	-60,2882	-23,9497	-0,6967	255,0651	-316,5514	171,4110
	t Stat	(-3,78)	(1,26)	(-2,47)	(-2,97)	(-0,10)	(3,63)	(-5,14)	(2,80)
Nombre d'étab. géolocalisés tous secteurs, étab. moyens * Paris	Coeff	-0,0639	0,7359	2,3113	0,1068	0,6977	-4,1442	1,3621	-3,1060
	t Stat	(-0,82)	(3,90)	(6,28)	(1,08)	(4,76)	(-3,91)	(1,76)	(-4,22)
Nombre d'étab. géolocalisés tous secteurs, étab. moyens * Limitrophe	Coeff	0,1284	0,4335	1,7037	0,2571	-0,1175	1,2204	0,8129	-9,1604
	t Stat	(1,21)	(0,98)	(2,17)	(1,07)	(-0,61)	(0,52)	(0,48)	(-4,31)
Nombre d'étab. géolocalisés tous secteurs, étab. moyens * PC	Coeff	-0,0831	-0,5715	1,6860	-0,0703	0,3477	-3,5047	-1,5720	0,4940
	t Stat	(-1,48)	(-2,83)	(3,37)	(-0,53)	(2,95)	(-2,66)	(-1,75)	(0,53)
Nombre d'étab. géolocalisés tous secteurs, petits étab. * Paris	Coeff	-0,0629	0,2551	-0,4880	0,1051	-0,1929	-3,6590	0,0667	-0,4488
	t Stat	(-2,37)	(1,82)	(-2,89)	(1,97)	(-2,22)	(-5,89)	(0,12)	(-1,03)
Nombre d'étab. géolocalisés tous secteurs, petits étab. * Limitrophe	Coeff	-0,2392	-0,7188	-0,9993	0,0785	0,1689	3,3838	1,7917	2,8462
	t Stat	(-2,24)	(-2,10)	(-1,54)	(0,36)	(0,96)	(1,84)	(1,04)	(1,87)
Nombre d'étab. géolocalisés tous secteurs, petits étab. * PC	Coeff	-0,1672	-0,2262	-0,0692	0,0508	-0,1388	-2,5986	-0,2488	-2,2960
	t Stat	(-2,10)	(-0,90)	(-0,14)	(0,34)	(-1,04)	(-2,05)	(-0,20)	(-2,01)

Tableau 35 : Déterminants du nombre d'emplois non géolocalisés

Le tableau 35 s'appuie sur une classification MOS de l'occupation de sols en 21 postes.

Codes 3	libellés	Codes 21	Libellés
1	Rural	1	Bois
		2	Grandes cultures
		3	Autres cultures
		4	Eau
		5	Autre rural
2	Urbain ouvert	6	Parcs et jardins
		7	Sport (espaces ouverts)
		8	Terrains vacants
3	Urbain construit	9	Habitat individuel
		10	Habitat collectif
		11	Habitat autre
		12	Activités secondaires
		13	Activités tertiaires
		14	Sport (construit)
		15	Equipements d'enseignement
		16	Equipements de santé
		17	Cimetières
		18	Autres équipements locaux, administrations
		19	Grands équipements
20	Transports		
21	Chantiers		

Tableau 36 : Découpage MOS en 21 postes et en 3 postes

Estimation de modèle de localisation d'emploi à la cellule

Plusieurs dizaines de combinaisons de variables ont été testées. Seules les variables significatives au seuil de 10% ont été retenues. Les huit modèles suivants, un par secteur, ont été retenus.

Secteur	Variable	estimation	p-value
Secteur 1 Agriculture, industries agricoles	Pseudo-R ² de McFadden	27%	
	Nb d'observations	1110	
	Population	0,00011944	<.0001
	Log(Nb d'emplois non géolocalisés, même secteur)	-0,01988586	0,0931
	Log(Nb d'emplois géolocalisés même secteur)	-0,00640399	<.0001
	Log(Nb d'emplois secteur 1)	0,87059504	<.0001
	Log(Nb d'emplois secteur 2)	-0,13737332	<.0001
	Log(Nb d'emplois secteur 4)	-0,08507114	0,0003
	Log(Nb d'emplois secteur 6)	-0,30204367	<.0001
	Log(Nb d'emplois secteur 8)	-0,13752693	<.0001
	Log de prix des logements	-0,48746282	<.0001
	Nb de gares SNCF et RER	0,23971558	0,001
	Temps moyen TC partant de la zone correspondante	-0,01351769	0,0233
	Petite couronne	-0,05263057	<.0001
	Paris	2,09067371	<.0001

Tableau 37 : Secteur 1 (Agriculture, industries agricoles)

Dans ce secteur, la première variable déterminante est le nombre d'emplois dans ce secteur même. Donc les entreprises agricoles s'installent dans les zones à forte présence d'agriculteurs. En revanche, dans une zone où la présence d'emplois d'autres secteurs est forte, cela réduit la probabilité de s'y installer. La variable « Appartient à Paris » a un coefficient positif, cela n'est pas contre-intuitif car une grande partie des entreprises agro-alimentaires sont localisées à Paris. On comprend facilement qu'il y a un plus fort flux en termes de localisation dans les industries agro-alimentaires que dans l'agriculture.

Secteur	Variable	estimation	p-value
	Pseudo-R ² de McFadden	18%	
	Nb d'observations	5677	
	Population	8,0149E-05	<.0001
	Log(Nb d'emplois non géolocalisés même secteur)	0,00687493	<.0001
	Log(Nb d'emplois géolocalisés même secteur)	-0,00031412	<.0001
	Log(Nb d'emplois secteur 2)	0,4146018	<.0001
	Log(Nb d'emplois secteur 3)	0,13916763	<.0001
	Log(Nb d'emplois secteur 4)	-0,07264064	<.0001
	Log(Nb d'emplois secteur 5)	-0,094365	<.0001
	Log(Nb d'emplois secteur 6)	-0,22862027	<.0001
	Log(Nb d'emplois secteur 7)	-0,0271575	0,0194
	Log(Nb d'emplois secteur 8)	-0,13108809	<.0001
Secteur 2	Zone franche	0,44943093	0,0278
Industrie	Zone urbaine sensible	0,47967932	<.0001
	Log prix de logement	-0,201064	<.0001
	Nb d'arrêts de métro	0,09352112	<.0001
	Nb de gares SNCF et RER	-0,15380859	0,0002
	Distance à l'autoroute la plus proche	0,11697403	<.0001
	Temps moyen TC partant de la zone correspondante	0,0064312	0,0195
	Temps moyen VP partant de la zone correspondante	-0,01731042	0,0001
	Nombre de logements au voisinage	1,1678E-05	<.0001
	Distance à Châtelet	-0,03261029	<.0001
	Petite couronne	0,66259495	<.0001
	Paris	0,78607143	<.0001
	Surface d'activité au voisinage	1,3082E-07	0,0014

Tableau 38 : Secteur 2 (Industrie)

Comme dans le secteur 1, les entreprises industrielles se localisent dans les zones qui offrent le plus d'emplois dans ce secteur, c'est l'effet généré par la concentration dans les zones industrielles. Les entreprises voient leur utilité augmenter en s'installant dans les zones franches urbaines, on comprend que c'est à cause des avantages accordées pour se localiser dans ces zones. Les coefficients de « Appartient à la petite couronne » et « appartient à Paris » sont positifs, donc le secteur s'installe plus à Paris et en petite couronne qu'en grande couronne.

Secteur	Variable	estimation	p-value
	Pseudo-R ² de McFadden	15%	
	Nb d'observations	24289	
	Population	0,00010207	<.0001
	Log(Nb d'emplois géolocalisés même secteur)	-0,00020125	<.0001
	Log(Nb d'emplois secteur 1)	0,08432608	<.0001
	Log(Nb d'emplois secteur 2)	-0,02930412	<.0001
	Log(Nb d'emplois secteur 3)	0,46720984	<.0001
	Log(Nb d'emplois secteur 4)	-0,0412367	<.0001
	Log(Nb d'emplois secteur 5)	-0,03400703	<.0001
	Log(Nb d'emplois secteur 6)	-0,22822168	<.0001
	Log(Nb d'emplois secteur 7)	-0,04064038	<.0001
	Log(Nb d'emplois secteur 8)	-0,16003151	<.0001
Secteur 3	Zone franche	0,31502967	0,0014
	Zone urbaine sensible	0,34815645	<.0001
Energie, construction, transports	Vocation tertiaire	-0,11921692	<.0001
	Log des prix des logements	-0,24499591	<.0001
	Nb arrêts de métro	0,04185486	<.0001
	Nb de gares SNCF et RER	-0,03456116	0,0501
	Distance à l'autoroute la plus proche	0,09255975	<.0001
	Temps moyen TC partant de la zone correspondante	0,00525361	<.0001
	Temps moyen VP partant de la zone correspondante	-0,06395855	<.0001
	Accessibilité VP	-0,04125273	<.0001
	Distance à Châtelet	-0,02755147	<.0001
	Petite couronne	1,08631491	<.0001
	Paris	1,10743044	<.0001
	Surface d'activité	-9,4765E-07	<.0001
	Surface d'activité au voisinage	-6,6055E-07	<.0001
	Valeur totale du bâti dédié à l'activité	2,1024E-06	<.0001

Tableau 39 : Secteur 3 (Energie, construction, transports)

Pour ce secteur, énergie, construction et transport, on a la même interprétation que dans les deux précédents. On trouve des coefficients particulièrement forts pour les variables « Paris » et « petite couronne », les sièges sociaux des entreprises de ce secteur sont largement implantés dans Paris et la petite couronne. Il y a des coefficients négatifs pour les emplois des autres secteurs, mis à part secteur 1, ce qui veut dire ces entreprises ne s'installent pas dans les mêmes zones que les entreprises des autres secteurs.

Secteur	Variable	estimation	p-value
	Pseudo-R ² de McFadden	19%	
	Nb d'observations	2697	
	Population	0,0001095	<.0001
	Log(Nb d'emplois géolocalisés même secteur)	-0,00042468	<.0001
	Log(Nb d'emplois secteur 2)	-0,06005198	<.0001
	Log(Nb d'emplois secteur 3)	0,08901212	0,0001
	Log(Nb d'emplois secteur 4)	0,45577962	<.0001
	Log(Nb d'emplois secteur 5)	-0,1191542	<.0001
	Log(Nb d'emplois secteur 6)	-0,18451048	<.0001
Secteur 4	Log(Nb d'emplois secteur 8)	-0,14153574	<.0001
Commerce	Zone franche	1,31480401	<.0001
	Zone urbaine sensible	0,38954451	0,0082
	Vocation tertiaire	-0,12165833	0,0153
	Log des prix des logements	-0,16057594	0,0309
	Temps moyen TC partant de la zone correspondante	0,01021734	0,004
	Temps moyen VP partant de la zone correspondante	-0,06058691	0,0103
	Accessibilité VP	-0,04165561	0,0103
	Petite couronne	0,9352023	<.0001
	Paris	1,13143872	<.0001
	Valeur totale du bâti dédié à l'activité	6,2578E-07	<.0001

Tableau 40 : Secteur 4 (Commerce)

La concentration des commerces dans des centres commerciaux fait que les entreprises de ce secteur préfèrent se localiser dans des zones commerciales, c'est ce que nous indique un coefficient positif pour les emplois de ce secteur. On trouve des coefficients assez élevés pour les variables zones franches, Paris et petite couronne.

Secteur	Variable	estimation	p-value
	Pseudo-R ² de McFadden	22%	
	Nb d'observations	4838	
	Population	9,3481E-05	<.0001
	Log(Nb d'emplois géolocalisés même secteur)	-8,3758E-05	<.0001
	Log(Nb d'emplois secteur 1)	0,1029198	<.0001
	Log(Nb d'emplois secteur 3)	-0,04860543	0,0241
	Log(Nb d'emplois secteur 4)	-0,04447063	0,0002
	Log(Nb d'emplois secteur 5)	0,39893379	<.0001
Secteur 5	Log(Nb d'emplois secteur 6)	-0,17220746	<.0001
	Log(Nb d'emplois secteur 7)	-0,06311889	<.0001
Activités financières, immobilières	Log(Nb d'emplois secteur 8)	-0,12819417	<.0001
	Vocation tertiaire	-0,1451416	0,0007
	Distance à l'autoroute la plus proche	0,05675508	0,0015
	Distance à la route nationale la plus proche	-0,07409339	0,0009
	Distance à Chatelet	-0,02176532	0,0003
	Petite couronne	1,69026455	<.0001
	Paris	1,54269679	<.0001
	Surface d'activité	-5,2958E-07	0,0227
	Surface d'activité au voisinage	-1,5546E-06	<.0001
	Valeur totale du bâti dédiée à d'activité	4,4912E-06	<.0001

Tableau 41 : Secteur 5 (Activités financières, activités immobilières)

La même interprétation est valable pour ce secteur, activités financières et immobilières.

Secteur	Variable	estimation	p-value
	Pseudo-R ² de McFadden	14%	
	Nb d'observations	31608	
	Population	0,00010671	<.0001
	Log(Nb d'emplois non géolocalisés même)secteur	4,7912E-05	0,0425
	Log(Nb d'emplois géolocalisés même secteur)	-6,3213E-05	<.0001
	Log(Nb d'emplois secteur 1)	0,05924289	<.0001
	Log(Nb d'emplois secteur 2)	-0,0115878	0,0174
	Log(Nb d'emplois secteur 3)	0,06700263	<.0001
	Log(Nb d'emplois secteur 4)	-0,02413265	<.0001
	Log(Nb d'emplois secteur 6)	0,10104786	<.0001
	Log(Nb d'emplois secteur 7)	-0,07075891	<.0001
	Log(Nb d'emplois secteur 8)	-0,10410382	<.0001
	Vocation tertiaire	-0,07591248	<.0001
Secteur 6	Log de prix des logements	0,11336454	<.0001
Services	Nb de gares SNCF et RER	0,03436279	0,0096
	Distance à l'autoroute la plus proche	0,05679917	<.0001
	Distance à la route nationale la plus proche	-0,04718165	<.0001
	Temps moyen TC partant de la zone correspondante	-0,00524029	<.0001
	Temps moyen VP partant de la zone correspondante	-0,0253678	0,0101
	Accessibilité VP	-0,01125894	0,0832
	Distance à Châtelet	-0,01949996	<.0001
	Petite couronne	1,12675906	<.0001
	Paris	1,12149111	<.0001
	Surface d'activité	-7,8275E-07	<.0001
	Surface d'activité au voisinage	-8,1969E-07	<.0001
	Valeur totale du bâti d'édifiée à l'activité	2,6718E-06	<.0001

Tableau 42 : Secteur 6 (Services aux entreprises, services aux particuliers)

La concentration dans les mêmes zones pour ce secteur, services aux entreprises et services aux particuliers, est moins forte par rapport aux autres secteurs, emplois secteur 6 = 0.101. Relativement aux autres secteurs, les entreprises de ce secteur s'implantent dans des zones à prix de logement élevé.

Secteur	Variable	estimation	p-value
	Pseudo-R ² de McFadden	16%	
	Nb d'observations	3505	
	Population	0,00013144	<.0001
	Log(Nb d'emplois non géolocalisés même secteur)	-0,00205818	<.0001
	Log(Nb d'emplois géolocalisés même secteur)	-0,00076083	<.0001
	Log(Nb d'emplois secteur 1)	0,04488213	0,0138
	Log(Nb d'emplois secteur 2)	-0,05757212	<.0001
	Log(Nb d'emplois secteur 4)	-0,05671904	<.0001
	Log(Nb d'emplois secteur 5)	0,02989832	0,0749
	Log(Nb d'emplois secteur 6)	-0,22068926	<.0001
	Log(Nb d'emplois secteur 7)	0,46668277	<.0001
Secteur 7	Log(Nb d'emplois secteur 8)	-0,13518926	<.0001
Education, santé, action sociale	Zone urbaine sensible	-0,35209115	0,0071
	Nb d'arrêts de métro	0,05450439	0,0316
	Nb de gares SNCF et RER	0,18185425	<.0001
	Distance à l'autoroute la plus proche	0,0346736	0,0216
	Distance à la route nationale la plus proche	-0,05003381	0,0125
	Temps moyen TC partant de la zone correspondante	-0,00745915	0,0198
	Temps moyen VP partant de la zone correspondante	-6,6527E-06	0,0521
	Distance à Châtelet	-0,01951699	<.0001
	Petite couronne	1,03925474	<.0001
	Paris	1,03675806	<.0001
	Surface d'activité	-7,9531E-07	0,0219
	Surface d'activité au voisinage	-7,232E-07	<.0001
	Valeur totale du bâti dédié à l'activité	1,9997E-06	<.0001

Tableau 43 : Secteur 7 (Education, santé, action sociale)

Les centres de ce secteur, éducation, santé et action sociale, s'implantent dans des zones bien desservies par le train et le métro, et proche des routes nationales. Mais ils s'implantent loin des autoroutes, sans doute à cause des nuisances sonores.

Secteur	Variable	estimation	p-value
	Pseudo-R ² de McFadden	17%	
	Nb d'observations	2268	
	Population	0,00011527	<,0001
	Log(Nb d'emplois non géolocalisés même secteur)	-0,00139967	<,0001
	Log(Nb d'emplois géolocalisés même secteur)	-0,00026006	<,0001
	Log(Nb d'emplois secteur 2)	-0,03744983	0,0359
	Log(Nb d'emplois secteur 4)	-0,06375019	0,0001
	Log(Nb d'emplois secteur 6)	-0,20564369	<,0001
Secteur 8	Log(Nb d'emplois secteur 7)	-0,08919255	<,0001
Administration	Log(Nb d'emplois secteur 8)	0,2711602	<,0001
	Log(Nb de gares SNCF et RER)	0,08456421	0,0687
	Distance à l'autoroute la plus proche	0,04691083	0,0118
	Temps moyen TC partant de la zone correspondante	-0,0081439	0,0662
	Distance à Châtelet	-0,02183963	0,0009
	Petite couronne	0,97452037	<,0001
	Paris	1,5820506	<,0001
	Surface d'activité au voisinage	-8,4971E-07	<,0001
	Valeur totale du bâti dédié à l'activité	2,3631E-06	<,0001

Tableau 44 : Secteur 8 (Administration)

Comme tous les autres secteurs, les administrations se localisent dans des zones à forte concentration d'employés administratifs. Ils se localisent principalement à Paris et en petite couronne, dans des cellules comportant un grand nombre de gares, loin des autoroutes et avec un temps moyen en TC le plus faible.

Tenant compte du fait que nous disposions des données concernant les établissements, nous avons décidé d'effectuer une étude à ce niveau. En effet, une étude de localisation par emplois souffre de deux lacunes :

- premièrement, les emplois vont être distribués sur la région d'une manière plus uniforme que la réalité et cela à cause d'atomisation des établissements qui sont éclatés en plusieurs emplois.
- deuxièmement, un tel modèle reviendra à distribuer les emplois arrivants en fonction d'effectif déjà présent sur la cellule ou commune. Une telle distribution donne des valeurs élevées pour les indicateurs de qualité de régression qui en réalité ne représentent pas assez bien l'évolution de nombre d'emplois à la cellule ou commune.

Un modèle de choix de localisation des établissements

Nous présentons ici quelques pistes de réflexion dans le développement d'un modèle de localisation des entreprises ou établissements (et non des emplois) en mettant l'accent sur le niveau géographique utilisé et les questions d'échantillonnage.

« Taille » des alternatives

Dans ce modèle de localisation, les établissements nouvellement créés ou délocalisés choisissent un lieu pour s'installer. Chaque établissement choisit alors entre un très grand nombre d'alternatives, correspondant aux différents locaux disponibles, ou susceptibles de l'accueillir. La difficulté réside dans la détermination des locaux disponibles. Nous ne disposons pas de données sur l'espace d'activité, qu'il soit libre ou occupé, mais nous pouvons supposer que l'espace disponible est proportionnel à l'espace occupé, et par conséquent au nombre d'emplois présents à la période initiale (1997).

On note I le nombre total de locaux disponibles en Ile-de-France, C_j le nombre de locaux disponible dans la commune j et G_k le nombre de locaux disponibles dans la cellule k . La probabilité que l'établissement h choisisse les locaux i est donnée par la formule du Logit Multinomial :

$$P_i^h = \frac{\exp(V_i^h)}{\sum_{i'=1}^I V_{i'}^h},$$

où V_i^h représente le bénéfice espéré des locaux i pour l'établissement h . On suppose, en première approximation, que tous les locaux situés dans la même cellule k ont les mêmes caractéristiques observables et procurent donc le même bénéfice espéré. Il en résulte que tous les locaux situés dans la même cellule k ont la même probabilité, notée P_i^h , d'être choisis par l'établissement h . La probabilité que l'établissement h choisisse la cellule k vaut alors :

$$P_k^h = G_k P_i^h = \frac{G_k \exp(V_k^h)}{\sum_{k'=1}^K \sum_{i' \text{ in } k'} V_{i'}^h} = \frac{\exp(V_k^h + \log(G_k))}{\sum_{k'=1}^K \exp(V_{k'}^h + \log(G_{k'}))}.$$

De la même façon, la probabilité que l'établissement h choisisse la commune j vaut :

$$P_j^h = C_j P_i^h = \frac{C_j \exp(V_j^h)}{\sum_{j'=1}^J \sum_{i' \text{ in } j'} V_{i'}^h} = \frac{\exp(V_j^h + \log(C_j))}{\sum_{j'=1}^J \exp(V_{j'}^h + \log(C_{j'}))}.$$

Il est donc indispensable d'inclure le log de la « taille » des communes ou des cellules dans la liste des variables explicatives des modèles de choix de localisation.

Nous voyons que la variable de taille doit être incluse en logarithme. Nous avons envisagé et testé plusieurs variables pour rendre compte de la « taille » de l'alternative : le nombre d'emplois, le flux d'emplois (total et/ou par secteur), et la superficie de la commune (totale ou par type d'occupation de sol).

Le nombre et le flux d'emplois sont bien connus par commune, mais pas par cellule. Comme nous ne disposons d'aucune information sur la distribution infra-communale des emplois en 1997, il nous est impossible de connaître les flux d'emplois à la cellule entre 1997 et 2001. En revanche, nous pouvons calculer le nombre d'emplois à la cellule, en nous basant sur des régressions de nombres d'emplois non géolocalisés sur le nombre d'emplois géolocalisés et les caractéristiques de la commune.

Echantillonnage des alternatives

Etant donné le grand nombre d'alternatives (1300 communes, 50 000 cellules, des millions de locaux), il est impossible d'estimer le modèle de localisation avec l'ensemble des alternatives. Ce problème est récurrent dans la littérature sur les choix de localisation, et une réponse générale est apportée dans le cas du Logit Multinominal. Sous l'hypothèse IIA (Indépendance des Alternatives Non Pertinentes), il est possible d'estimer ce modèle en tirant au hasard un petit nombre d'alternatives pour chaque établissement (en plus de l'alternative qu'il a effectivement choisie).

L'échantillon d'alternatives doit alors être représentatif d'ensemble des unités de choix, dans notre cas les locaux disponibles. Dans notre modèle, nous avons supposé que tous les locaux d'une même cellule étaient semblables et avaient donc la même probabilité d'être choisis. Tirer au hasard les locaux avec une probabilité uniforme revient alors à tirer au hasard les cellules ou les communes avec une probabilité proportionnelle à leur « taille ». La variable de taille que nous pouvons utiliser pour calculer les probabilités d'apparition est le nombre total d'emplois.

La variable de pondération doit être la même pour les cellules et pour les commune. Plus précisément, la somme des poids des cellules qui se trouvent dans une commune doit être égale au poids de cette dernière. Cette règle est délicate à mettre en œuvre car certaines cellules sont à cheval sur plusieurs communes. Par souci de simplification, nous considérons que chaque cellule est liée à une seule commune (celle dans laquelle se trouve le centre de gravité de la cellule). La différence entre les pondérations calculées directement pour les communes et celles calculées par somme sur les cellules est relativement minime. Par exemple, pour les 20 arrondissements de Paris, la différence est de l'ordre de 0,2%.

Estimations par secteur

Les tableaux suivants présentent les résultats des estimations. Nous constatons que Paris attire surtout les nouveaux établissements des secteurs 3 (Energie, construction, commerces) et dans une moindre mesure ceux du secteur 1 (agro-alimentaire), mais pas ceux du secteur 4 (transport), qui sont déjà installés à Paris de longue date. Ce résultat pour l'agro-alimentaire n'est pas contre-intuitif lorsque l'on pense qu'il s'agit surtout de maisons-mères et non de lieux de production agricole. Les communes limitrophes ainsi que la Défense sont dépréciées toutes choses égales par ailleurs pour la plupart des secteurs. Ces zones ont accueilli de nombreux établissements par le passé, et le manque de place n'est pas propice à l'installation de nouveaux établissements. Le peu d'attrait des villes nouvelles (pour l'ensemble des secteurs sauf celui de transport) s'explique plus par le prestige ou des effets de mode que par le manque de place, et il est probablement plus conjoncturel que structurel.

Le nombre d'habitants dans la commune constitue un effet taille très significatif dans la plupart des secteurs. Il s'avère très substituable avec la densité de population, puisque, dans chaque secteur, une seule de ces deux variables a un impact significatif sur la localisation des établissements.

Les établissements ont tendance à se localiser soit dans des communes à revenus bas soit dans des communes à revenus élevés, les communes à revenus intermédiaires étant délaissées sur la période.

Les établissements se localisent plutôt dans des communes « jeunes », caractérisées par une forte proportions de ménages dont le chef de ménage a moins de 35 ans, et dans une moindre mesure, entre 35 et 65 ans. Les retraités n'attirent pas les emplois !

La proximité à Châtelet (« centre » de l'Ile-de-France) est dévalorisée dans les secteurs 5 (services immobiliers et financiers) et surtout 3 (commerces). Les secteurs 2 et 6 à 8 sont par ailleurs sensible à la proximité des grands axes (Routes Nationales ou autoroutes).

Les communes rapidement accessibles par TC ou surtout par VP attirent les emplois dans tous les secteurs sauf l'agro-alimentaire. La spécificité de ce secteur est toutefois à relativiser car les variables de temps de trajet sont corrélées à la distance à Châtelet et à l'accessibilité (qui améliore l'attractivité des communes dans le cas du secteur 1).

Des prix supérieurs pour les bureaux ou des logements attirent les établissements, ce qui reflète principalement un effet de prestige.

Une forte densité des emplois dans la commune attire les nouveaux établissements dans les secteurs 2 et 3, alors que l'effet inverse s'observe dans les secteurs 5 et 6 pour la densité des emplois au niveau des zones ou des pôles d'emploi. Il semble donc qu'on assiste à une concentration géographique des établissements à une échelle fine (commune), mais à une dispersion à une échelle plus large (zone ou pôle d'emploi).

Les nouveaux établissements sont globalement attirés par les communes comportant une plus grande fraction d'habitants dans les CSP qu'ils sont susceptibles d'employer.

Les effets négatifs du nombre total d'emplois sont à relativiser par l'effet souvent positif du nombre d'emplois dans les différents secteurs. Dans tous les secteurs, on assiste à une concentration de l'emploi sectoriel.

Des estimations préliminaires ont montré que la localisation des petits établissements est un peu mieux expliquée que celle des établissements de taille moyenne ou surtout celle des grands établissements.

Secteur		1	2	3	4	5	6	7	8
La commune est-elle à Paris	Coeff t Stat	2,1445 (1,37)		0,1442 (2,41)	-1,9232 (-6,04)				
Commune limitrophe de Paris	Coeff t Stat	-0,3211 (-2,99)	-0,1041 (-1,81)	-0,1429 (-5,91)		-0,0111 (-0,17)	0,0162 (0,56)	-0,0227 (-0,38)	0,0931 (1,15)
La Défense	Coeff t Stat			-0,2045 (-4,28)		-0,0157 (-0,15)	-0,2604 (-5,83)	0,0264 (0,22)	0,2525 (1,78)
Ville nouvelle	Coeff t Stat	-0,2915 (-2,21)	-0,1498 (-2,39)	-0,1412 (-5,02)	0,1590 (2,21)	-0,0294 (-0,37)	-0,1244 (-4,03)	-0,1579 (-2,35)	-0,2401 (-3,03)
Population de commune en 99 (en millier)	Coeff t Stat	0,0085 (8,11)		0,0036 (14,15)	0,0070 (10,15)	0,0040 (7,06)	0,0035 (12,19)	0,0047 (8,30)	
Densité de population en 99 commune (Pers. / m2)	Coeff t Stat		16,0993 (5,68)				0,8235 (0,60)		14,1133 (4,49)
% des ménages avec revenu élevé	Coeff t Stat	2,1850 (2,05)		1,7098 (7,66)		3,5395 (5,77)	3,1164 (13,61)	3,2744 (6,17)	
% des ménages avec revenu bas	Coeff t Stat	3,7719 (2,73)	0,7297 (2,29)	3,1969 (11,43)		2,6617 (3,18)	2,7578 (9,09)	2,8981 (4,31)	
% des ménage dont l'age de PR inférieur à 35 ans	Coeff t Stat	2,5514 (1,80)	2,9331 (3,86)	0,5987 (3,28)		1,4412 (3,17)	1,6658 (8,78)	1,2754 (2,96)	
% des ménage dont l'age de PR entre 35 et 65 ans	Coeff t Stat	3,9371 (1,90)	5,5359 (5,24)	0,6003 (2,99)		-3,0029 (-5,01)	-0,1676 (-0,82)		
% des ménages avec enfants	Coeff t Stat	-2,1634 (-1,71)							
% des ménages sans actif	Coeff t Stat	3,0384 (1,46)	4,1523 (4,06)						
Distance du centre commune à Châtelet	Coeff t Stat			-0,0049 (-6,29)		-0,0053 (-2,28)			
Distance de la route nationale (km)	Coeff t Stat		0,000027 (2,97)				0,000005 (1,14)		
Distance de l'autoroute (km)	Coeff t Stat							0,000009 (1,47)	0,000027 (3,83)

Secteur		1	2	3	4	5	6	7	8
Temps de trajet moyen en TC (Heure)	Coeff t Stat		-0,0340 (-1,17)		-0,0409 (-1,01)		-0,0287 (-2,16)		-0,0255 (-0,70)
Temps de trajet moyen en VP (Heure)	Coeff t Stat			-0,6766 (-10,04)		-0,3889 (-2,02)	-0,3537 (-4,55)	-1,1041 (-5,69)	-0,9540 (-4,30)
Accessibilité aux commerces (Poulit)	Coeff t Stat	0,0455 (1,68)							
Nb arrêts de métro	Coeff t Stat		0,0243 (5,04)		-0,0186 (-3,05)	0,0107 (1,91)	-0,0038 (-1,59)		0,0093 (1,46)
Nombre de gares SNCF et RER	Coeff t Stat			0,0104 (2,94)		0,0212 (2,31)	0,0145 (4,32)	0,0239 (2,70)	0,0176 (1,72)
Log de prix de bureau	Coeff t Stat		0,2678 (3,03)			0,2116 (2,22)	0,3382 (9,44)		0,2537 (2,48)
Logarithme de prix de logement (Emp./m2)	Coeff t Stat	0,2308 (1,66)		0,2602 (4,92)		0,2550 (1,84)			
Densité d'emploi à la commune 1997	Coeff t Stat		16,6091 (6,30)	14,1139 (9,80)					-12,9414 (-2,95)
Densité d'emploi au pôle d'emploi 1997	Coeff t Stat					-117,4519 (-3,93)	8,0918 (0,62)		
Densité d'emploi à la zone d'emploi 1997	Coeff t Stat						-9,3713 (-0,54)		
Spécialisation d'emploi à la commune 1997	Coeff t Stat		-1,5751 (-4,65)		-1,3805 (-3,48)				-1,3001 (-3,59)
Spécialisation d'emploi au pôle 1997	Coeff t Stat				-0,5661 (-1,50)				-0,6417 (-1,67)
Spécialisation d'emploi à la zone 1997	Coeff t Stat							2,1248 (1,86)	
% d'agriculteurs parmi travailleurs	Coeff t Stat	2,4723 (1,78)			1,9106 (2,46)				

Secteur		1	2	3	4	5	6	7	8
% artisans et commerçants parmi travailleurs	Coeff t Stat		1,5785 (3,91)	1,2198 (7,39)	1,4880 (3,17)	0,9107 (1,85)	0,8188 (4,89)		
% cadre parmi travailleurs	Coeff t Stat			-1,5292 (-11,93)	-1,4398 (-2,89)		-0,9018 (-7,64)	-0,9734 (-2,34)	
% profession intermédiaire parmi travailleurs	Coeff t Stat			-0,5822 (-4,00)					
% employé parmi travailleurs	Coeff t Stat		-0,6811 (-2,26)						
% ouvriers parmi travailleurs	Coeff t Stat		0,9355 (3,18)						
% niveau étude primaire parmi travailleurs	Coeff t Stat	3,1826 (3,72)		0,2372 (1,40)				1,2710 (2,50)	
% niveau étude collège parmi travailleurs	Coeff t Stat	1,8043 (2,77)	0,7822 (2,74)			1,7815 (5,57)		0,9704 (2,32)	
% niveau étude lycée parmi travailleurs	Coeff t Stat		1,3886 (3,57)		0,6228 (1,30)				
% niveau étude supérieure parmi travailleurs	Coeff t Stat	0,9972 (1,40)			-0,6280 (-1,49)			0,9069 (1,83)	
Log de nombre d'emplois Total	Coeff t Stat	-0,8158 (-10,81)	-1,0496 (-9,44)	-1,0321 (-26,92)	-0,3837 (-4,42)	-1,0252 (-9,01)	-0,9684 (-25,70)	-1,0366 (-12,10)	-0,6977 (-7,77)
Log de nombre d'emplois secteur 1	Coeff t Stat	0,2040 (4,55)	0,0656 (2,97)	0,0400 (4,08)	0,0906 (3,28)	0,0731 (2,49)	0,0706 (6,78)	0,0759 (2,85)	0,0361 (1,48)
Log de nombre d'emplois secteur 2	Coeff t Stat		0,2102 (7,81)	0,0638 (7,82)		0,0613 (2,76)	0,0487 (4,96)		
Log de nombre d'emplois secteur 3	Coeff t Stat		0,3985 (8,54)	0,4587 (27,19)	0,1160 (2,47)	0,0758 (1,45)	0,1864 (9,72)	0,0950 (2,24)	
Log de nombre d'emplois secteur 4	Coeff t Stat		-0,0353 (-2,10)	-0,0127 (-1,78)	0,1467 (6,61)				

Secteur		1	2	3	4	5	6	7	8
Log de nombre d'emplois secteur 5	Coeff t stat	0,0560 (1,39)		0,0150 (1,65)		0,2472 (9,56)	0,0780 (8,04)	0,0705 (2,98)	
Log de nombre d'emplois secteur 6	Coeff t stat		0,1454 (3,24)	0,1303 (7,65)		0,1105 (2,13)	0,2120 (11,74)	0,1176 (2,60)	0,1315 (2,32)
Log de nombre d'emplois secteur 7	Coeff t stat	0,1326 (2,97)	0,0594 (2,18)	0,0854 (7,49)	-0,0925 (-3,01)	0,2288 (6,68)	0,0682 (5,88)	0,2864 (9,29)	0,1931 (5,43)
Log de nombre d'emplois secteur 8	Coeff t stat	0,1149 (1,99)		0,0776 (6,20)	-0,0588 (-1,88)	0,0958 (2,53)	0,0727 (5,36)	0,2301 (6,69)	0,2202 (6,57)
Nombre d'emplois en commune secteur 1 (millier)	Coeff t stat	-0,2676 (-1,39)	-0,2243 (-2,17)	-0,1474 (-3,54)	-0,5787 (-5,06)	-0,2740 (-2,49)	-0,3581 (-8,65)	-0,2499 (-2,34)	
Nombre d'emplois en commune secteur 2 (millier)	Coeff t stat				0,0170 (2,06)		0,0202 (5,96)		0,0329 (3,03)
Nombre d'emplois en commune secteur 3 (millier)	Coeff t stat		0,0205 (2,19)			0,0255 (2,96)	0,0141 (3,65)		
Nombre d'emplois en commune secteur 4 (millier)	Coeff t stat	0,0248 (1,83)		0,0097 (3,02)					
Nombre d'emplois en commune secteur 5 (millier)	Coeff t stat	0,0294 (3,53)	-0,0112 (-2,39)	-0,0155 (-9,06)	0,0089 (2,73)	0,0167 (4,48)	0,0026 (1,75)	0,0118 (3,85)	0,0345 (4,94)
Nombre d'emplois en commune secteur 6 (millier)	Coeff t stat	-0,0185 (-3,08)	-0,0080 (-2,40)						-0,0226 (-5,00)
Nombre d'emplois en commune secteur 7 (millier)	Coeff t stat			-0,0176 (-6,82)		-0,0498 (-7,41)	-0,0102 (-4,36)	-0,0278 (-4,43)	-0,0167 (-2,63)
Nombre d'emplois en commune secteur 8 (millier)	Coeff t stat	-0,0279 (-2,43)	-0,0195 (-3,31)	-0,0085 (-3,27)		-0,0211 (-3,02)	-0,0089 (-3,50)	-0,0272 (-4,35)	
Nombre total d'emplois au pôle secteur 1 (millier)	Coeff t stat		-0,2227 (-5,24)		0,1065 (2,44)		0,0736 (4,22)	0,1304 (3,16)	
Nombre total d'emplois au pôle secteur 2 (millier)	Coeff t stat	-0,0134 (-1,37)	0,0329 (5,01)	-0,0022 (-4,35)	0,0237 (3,83)				

Secteur		1	2	3	4	5	6	7	8
Nombre total d'emplois au pôle secteur 3 (millier)	Coeff t stat	0,0320 (2,61)	-0,0098 (-2,02)		-0,0272 (-4,63)		-0,0025 (-1,45)		0,0050 (2,12)
Nombre total d'emplois au pôle secteur 4 (millier)	Coeff t stat	-0,0312 (-1,98)			0,0486 (7,81)				
Nombre total d'emplois au pôle secteur 5 (millier)	Coeff t stat	-0,0171 (-1,94)	0,0123 (1,90)			0,0102 (3,04)			-0,0153 (-3,19)
Nombre total d'emplois au pôle secteur 6 (millier)	Coeff t stat		-0,0098 (-3,85)					-0,0024 (-3,35)	
Nombre total d'emplois au pôle secteur 7 (millier)	Coeff t stat	-0,0184 (-3,08)	0,0162 (5,19)					-0,0092 (-3,03)	
Nombre total d'emplois au pôle secteur 8 (millier)	Coeff t stat					-0,0048 (-2,27)	-0,0032 (-1,62)		
Nombre d'emplois à la zone secteur 1 (millier)	Coeff t stat		0,1141 (5,65)	0,0688 (9,22)	-0,1172 (-4,00)			-0,0350 (-1,20)	
Nombre d'emplois à la zone secteur 2 (millier)	Coeff t stat	-0,0113 (-2,38)	-0,0152 (-5,27)	-0,0052 (-9,93)	-0,0157 (-5,66)		-0,0052 (-1,88)		0,0078 (2,67)
Nombre d'emplois à la zone secteur 3 (millier)	Coeff t stat				0,0184 (7,16)		0,0011 (0,55)	-0,0102 (-4,54)	-0,0177 (-5,57)
Nombre d'emplois à la zone secteur 4 (millier)	Coeff t stat						-0,0051 (-2,57)	0,0127 (3,94)	0,0242 (5,52)
Nombre d'emplois à la zone secteur 5 (millier)	Coeff t stat		-0,0085 (-1,53)				-0,0033 (-1,24)		0,0174 (4,57)
Nombre d'emplois à la zone secteur 6 (millier)	Coeff t stat	0,0044 (2,00)	0,0053 (2,40)				0,0016 (1,03)		
Nombre d'emplois à la zone secteur 7 (millier)	Coeff t stat					-0,0081 (-2,98)	-0,0028 (-1,08)	0,0173 (5,26)	
Nombre d'emplois à la zone secteur 8 (millier)	Coeff t stat					0,0096 (2,80)	0,0072 (3,42)		

Tableau 45- Déterminants de la localisation des établissements

Indicateurs de qualité globale des régressions

Le tableau 46 met en évidence une assez médiocre qualité des régressions, puisque le Pseudo-R² de McFadden varie de 3% à 8% selon les secteurs. Toutefois, d'autres indicateurs de qualité tels que Estrella ou Cragg-Uhler montent à 30% environ. Il convient de ne pas accorder trop d'importance à ces indicateurs, pas plus qu'au pourcentage de prédictions correctes. De façon générale, il n'existe pas d'indicateur vraiment satisfaisant pour mesurer la qualité d'un modèle dans le cas des choix discrets, et des Pseudo-R² inférieurs à 10% sont courants dans le cas du Logit Multinomial.

Le faible nombre d'itérations nécessaires à la convergence montre qu'il n'y a pas de problèmes de multicollinéarité entre les explicatives et que la précision des estimations est satisfaisante (sauf peut-être pour les secteurs 2 et 6).

Secteur	1	2	3	4	5	6	7	8
Nb Observations	1945	8343	41465	4407	7150	46666	6083	4395
Nb Itérations	5	86	5	4	5	66	5	4
Log Likelihood	-4080,9	-17812,5	-90827,9	-9506,1	-15471,1	-103938,7	-13274,3	-9723,3
AIC	8227,82	35705,00	181725,87	19068,37	31010,26	207963,31	26614,58	19502,69
Schwarz Criterion	8411,73	35986,16	182028,01	19247,31	31244,01	208339,60	26836,12	19681,56
Likelihood Ratio (R)	795,23	2795,94	9297,51	1282,62	1984,70	7027,56	1464,67	793,03
Upper Bound of R (U)	8957,06	38420,93	190953,38	20294,99	32926,97	214904,87	28013,25	20239,72
Estrella	34,83%	29,39%	20,54%	25,97%	24,90%	14,20%	21,91%	16,81%
Adjusted Estrella	32,37%	28,65%	20,40%	24,96%	24,13%	14,03%	21,01%	15,70%
Aldrich-Nelson	29,02%	25,10%	18,32%	22,54%	21,73%	13,09%	19,41%	15,29%
Cragg-Uhler 1	33,56%	28,48%	20,09%	25,25%	24,24%	13,98%	21,40%	16,51%
Cragg-Uhler 2	33,90%	28,76%	20,29%	25,51%	24,48%	14,12%	21,61%	16,68%
McFadden's LRI	8,88%	7,28%	4,87%	6,32%	6,03%	3,27%	5,23%	3,92%
Veall-Zimmermann	35,32%	30,55%	22,29%	27,44%	26,45%	15,93%	23,62%	18,61%

Tableau 46 : Déterminants de l'évolution des effectifs des établissements

