



PREDIT 2002-2007
Groupe 1
Mobilité, territoires et développement durable

Projet *SIMAU*RIF
(SIMulation de l'interAction Urbainisation-transports
en Région Ile-de-France.) :

Perfectionnement et valorisation

Réf : n 06MTE053



INSTITUT
D'AMÉNAGEMENT
ET D'URBANISME

ÎLE-DE-FRANCE

Juillet 2008

IAU-îdF n°5.07.003
DRAST n°06MTE053

Institut d'Aménagement et d'Urbanisme d'Île-de-France (IAU-îdF)

15 rue Falguière 75740 Paris cedex 15
Tél. 01.77.49.53.85 - Télécopie 01.77.49.76.02
E-mail : dgcomm@iau-idf.fr –http : www.iau-idf.fr
Directeur général : François DUGENY
Directeur du Département Mobilité et Transport : Alain MEYERE

Dany NGUYEN-LUONG, ingénieur chargé d'études

Email : dany.nguyen-luong@iau-idf.fr

Tél : 01.77.49.77.44

Remerciements

Nous tenons à remercier les personnes ayant permis le cofinancement de cette étude :

- Mme Régine BREHIER (Directrice de la DRAST),
- M. Yves CROZET (Professeur, Président du GO1 du Predit, LET).
- M. Gérard BRUN (Chargé de mission à la DRAST),

Un grand merci à nos partenaires du Laboratoire THEMA de l'Université de Cergy-Pontoise qui nous ont accompagnés tout au long de ces quatre années de recherche :

- M. André DE PALMA (Professeur, Université de Cergy-Pontoise)
- M. Kiarash MOTAMEDI (Doctorant)
- M. Hakim OUARAS (Doctorant)
- Mme Nathalie PICARD (Maître de conférences, Université de Cergy-Pontoise)

Un grand merci également à M. Paul WADDELL (Professeur, Université de Washington) pour tous ses conseils.

Sommaire

1. Objet de l'étude	9
2. Synthèse du projet SIMAURIF.....	11
<i>Comment est né le projet SIMAURIF ?</i>	11
<i>Quelle est l'opportunité du projet SIMAURIF ?</i>	11
<i>Quelle est la traduction des termes « land use model » et « LUTI model » ?</i>	14
<i>Quels sont les deux modèles qui ont été couplés ?</i>	14
<i>Quel est l'intérêt d'utiliser un modèle dynamique plutôt qu'un modèle classique statique ?</i>	15
<i>Pourquoi avoir choisi UrbanSim ?</i>	15
<i>Comment se positionne le projet SIMAURIF dans la recherche académique ?</i>	17
<i>Que cherche-t-on à modéliser exactement ?</i>	17
<i>Quelles sont les différences entre les projets SIMAURIF, SIMBAD, MobiSim, MIRO et ILOT ?</i>	20
<i>Quelles sont les applications terminées d'UrbanSim dans le monde ?</i>	20
<i>Quel est l'indicateur d'accessibilité « log sum » utilisé dans SIMAURIF ?</i>	20
<i>Pourquoi avoir choisi trois motifs de déplacements seulement ?</i>	23
<i>Comment le modèle de prix d'UrbanSim a-t-il été adapté dans le projet SIMAURIF ?</i>	23
<i>Quel a été le déroulement chronologique du projet ?</i>	25
<i>Quelle est l'échelle d'étude ?</i>	26
<i>Qu'entend-on par période de calage et période de simulation ?</i>	26
<i>Quelles sont les bases de données d'entrée utilisées ?</i>	27
<i>Quelles sont les bases de données d'estimation utilisées ?</i>	27
<i>Comment le problème des données manquantes a-t-il été résolu ?</i>	28
<i>Quelle est la structure de données d'UrbanSim ?</i>	28
<i>Comment UrbanSim a-t-il été adapté au cas de l'Ile-de-France ?</i>	28
<i>Comment les contraintes environnementales et politiques ont-elles été prises en compte ?</i>	29
<i>Quelles ont été les difficultés rencontrées dans l'estimation des modèles partiels ?</i>	29
<i>Comment UrbanSim prend-il en compte des événements imprévisibles, par exemple le déménagement de grands groupes industriels ?</i>	30
<i>Quelle est la différence entre spécifier, estimer et caler un modèle ?</i>	30
<i>Comment valider chaque modèle partiel ?</i>	30
<i>En quoi consiste le calage global d'un modèle intégré ?</i>	30
<i>Quelle est la pertinence du découpage en 50 000 cellules carrées ? Et si oui, est-il rentable ?</i>	31
<i>En quoi consiste le projet de Tangentielle Nord ?</i>	33
<i>Quelle est l'aire d'étude de la Tangentielle Nord et les données globales du territoire desservi ?</i> ..	33
<i>Quelle sont les atouts et faiblesses de ce territoire ?</i>	34
<i>Quelles sont les hypothèses socio-économiques à l'horizon 2026 ?</i>	37
<i>Combien de temps dure une simulation ?</i>	38
<i>Des tests de sensibilité ont-ils été effectués ?</i>	38
<i>Quels sont les résultats de l'application de SIMAURIF au cas de la Tangentielle Nord ?</i>	38
<i>Quel est le niveau de fiabilité du modèle SIMAURIF et quelles sont les sources d'incertitude ?</i>	40
<i>Les résultats seront-ils vérifiables ?</i>	41
<i>Quels sont les enseignements de ces résultats sur la Tangentielle Nord ?</i>	41
<i>Quels sont les moyens humains nécessaires pour mettre au point un modèle intégré ?</i>	42
3. Portage des données et des modèles d'UrbanSim 2 à OPUS/UrbanSim 4.....	43
3.1. Historique	43
3.2. Justification du portage	43
3.3. Problèmes rencontrés lors du portage.....	44
3.4. Résultats du portage	47
3.5. Conclusion du chapitre	49
4. Estimation d'un modèle de prix de l'immobilier de bureaux.....	50
4.1. Rappel sur le modèle de prix du foncier d'UrbanSim	50
4.2. Géographie succincte de l'immobilier d'entreprise	51
4.3. Les données utilisées	52

4.4. Spécification du modèle.....	52
4.4.1. Modèle de Rosen.....	52
4.4.2. Quelques particularités sur le modèle spécifié.....	53
4.5. Résultats du modèle.....	54
4.6. Conclusion du chapitre.....	57
4.7. Bibliographie.....	57
5. Développement de l'interface entre Urbansim et Davisum-METROPOLIS.....	59
5.1. Interface entre UrbanSim 2 et METROPOLIS.....	59
5.1.1. Gestion des paramètres.....	59
5.1.2. Gestion avancée des paramètres.....	60
5.1.3. Gestion de l'exécution.....	61
5.2. Interface entre OPUS/UrbanSim 4 et Davisum-METROPOLIS.....	61
5.2.1. Prise en compte des variables transport.....	62
5.2.2. Lancement de <i>SIMAU</i> RIF.....	63
5.3. Conclusion du chapitre.....	65
6. Valorisation de la recherche.....	66
7. Conclusions.....	67
Annexe : les 2 articles.....	71

1. Objet de l'étude

L'Institut d'Aménagement et d'Urbanisme d'Île-de-France (IAU-îdF, anciennement IAURIF), en partenariat avec le laboratoire THEMA de l'Université de Cergy-Pontoise, a réalisé entre 2003 et 2007 une recherche sur la modélisation de l'interaction transport-urbanisation. Cette recherche, subventionnée par le Predit 3, le Conseil régional d'Île-de-France, la DREIF et RFF, visait à mettre au point un modèle dynamique et intégré opérationnel de l'interaction urbanisation-transports sur la région d'Île-de-France et à son application à une étude de cas, la Tangentielle Nord entre Sartrouville et Noisy-le-Sec.

Ce modèle, appelé *SIMAU*RIF (*SIM*ulation de l'*inter*Action Urbanisation-transport en Région d'*Île-de-France*), utilise deux modèles existants :

- Urbansim, un modèle d'urbanisation (un « land-use model »), développé par l'Université de l'Etat de Washington à Seattle (Etats-Unis), basé sur la théorie des modèles de choix discrets.
- Davisum-METROPOLIS, un modèle de prévision de trafic incluant le modèle global de demande de déplacements développé par l'IAU-îdF, avec les logiciels MinUTP, Arc-Info/ArcGis puis Davisum. La partie affectation routière sur le logiciel METROPOLIS a été développée par le laboratoire THEMA de l'Université de Cergy-Pontoise grâce à un financement du Predit 2.

La présente étude, commencée au début de l'année 2007, consiste en un « perfectionnement et une valorisation de l'outil *SIMAU*RIF ». Elle a été menée parallèlement à la deuxième phase du projet *SIMAU*RIF (recherche n°5MT6008) qui comportait deux tâches principales : le calage global du modèle intégré et l'application au cas de la Tangentielle Nord. Le rapport final de la deuxième phase a été remis au Predit en décembre 2007.

Le présent rapport comporte trois parties¹ :

- L'approfondissement de deux volets techniques de la deuxième phase : d'une part le **portage des données et des modules d'Urbansim 2 à OPUS/Urbansim 4**, d'autre part l'**estimation d'un modèle de prix de l'immobilier de bureaux**,
- Le développement d'**une interface informatique entre OPUS/Urbansim 4 et Davisum-METROPOLIS**,
- La **valorisation de la recherche** sous forme d'articles, de communications, de séminaires, etc.

Avant de traiter ces trois parties, nous présentons une longue synthèse du projet *SIMAU*RIF qui se veut avant tout vulgarisatrice et vise à faire connaître le projet en dehors du milieu académique. Elle décrit les enseignements principaux du projet sous forme de questions-réponses. Pour plus d'informations techniques, le lecteur pourra consulter les trois rapports de 2004, 2005 et 2007, disponibles sur le site Internet de l'IAU-îdF www.iau-idf.fr dans la rubrique Bibliothèque ou directement via les liens suivants :

¹ Conformément à la proposition d'étude adressée au Predit, datée du 13 novembre 2006

1) *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 1ère année (2004)

http://www.iau-idf.fr/exl-doc/IA46029-1_OUV00001603.pdf

2) *SIMAU*RIF. Modèle dynamique de simulation de l'interaction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle Nord. Rapport intermédiaire de la deuxième phase (2005)

http://www.iau-idf.fr/exl-doc/IA46029-2_OUV00002471.pdf

3) *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 2ème phase (2007)

http://www.iau-idf.fr/exl-doc/IA48950_OUV00005358.pdf

2. Synthèse du projet *SIMAU*RIF

Comment est né le projet *SIMAU*RIF ?

L'idée du projet *SIMAU*RIF est née lors d'un voyage d'étude aux Etats-Unis effectué par l'auteur en 1999. En étudiant les modèles de prévision de trafic utilisés outre-atlantique, il a découvert que les modèles dits « land-use » (MEPLAN, DRAM-EMPAL, ITLUP, POLIS, TRANUS) étaient utilisés depuis longtemps dans plusieurs métropoles américaines. Les recherches en cours concernaient les modèles dits « LUTI » (land use transportation integrated) qui visent à coupler un modèle « land use » et un modèle de prévision de trafic. Ces recherches étaient motivées par un contexte réglementaire, les deux lois sur l'air ISTEA (1991) et TEA21 (1998), qui rendait obligatoire la prise en compte de l'interaction urbanisation-transport dans l'élaboration des plans de transports régionaux. Par conséquent, on a vu aux Etats-Unis au cours des 15 dernières années se développer une recherche prolifique sur les modèles intégrés urbanisation-transport ainsi que la mise à disposition d'outils opérationnels. De retour en France, l'auteur a maintenu une veille technologique sur les modèles intégrés développés aux Etats-Unis. C'est en 2002 que le logiciel UrbanSim est apparu suffisamment au point pour envisager son utilisation en France. Ce logiciel est un pur produit du monde universitaire. Son développement a commencé en 2000 par l'équipe du Pr Paul Waddell du CUSPA (Center for Urban Simulation and Policy Analysis) de l'Université de Washington à Seattle. En septembre 2003, l'IAU île-de-France et le laboratoire THEMA (équipe du Pr André de Palma) se sont associés pour mettre au point un modèle intégré opérationnel sur la région d'Île-de-France. Le laboratoire THEMA avait déjà conçu et développé un modèle routier dynamique d'affectation appelé METROPOLIS dans le cadre du Predit 2, tandis que l'IAU-îdF avait son propre modèle de trafic multimodal à quatre étapes classique. La collaboration entre ces deux partenaires apparaissait donc judicieuse. De plus, une synergie naît toujours d'une collaboration entre chercheurs et praticiens.

Le projet *SIMAU*RIF est donc véritablement un projet pionnier en France dans le domaine de la modélisation de l'interaction urbanisation-transport. 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).

Quelle est l'opportunité du projet *SIMAU*RIF ?

Les modèles « LUTI » (land-use transport integrated models) font aujourd'hui l'objet de beaucoup d'attention, aussi bien de la part de la communauté des chercheurs (LVMT, INRETS, LET) que des services techniques du MEEDDAT (ex-Ministère de l'Équipement), exploitants, gérants et autorités organisatrices de transports (RATP, SNCF, RFF, OPTILE, STIF), bureaux d'études et de quelques entreprises privées de transport (Veolia, Cofiroute). Il y a trois raisons principales qui fondent l'opportunité du projet *SIMAU*RIF :

D'abord, **première raison** : il est maintenant reconnu que la modélisation de trafic classique en milieu urbain, du type modèle statique à quatre étapes, est complètement dépassée. Née après la Seconde Guerre Mondiale aux Etats-Unis dans le but de planifier des systèmes de transport routier, la méthode de prévision de trafic classique a atteint ses limites aujourd'hui et ne répond plus à la demande des décideurs, même si elle est encore largement utilisée par défaut en France et aux Etats-Unis pour l'évaluation des politiques de transport et l'évaluation socio-économique d'un projet d'infrastructure de transport. D'une part, les modèles de trafic statiques ont très peu évolué depuis 50 ans. Les quatre

étapes (génération, distribution, choix modal, affectation) sont toujours présentes. Il y a bien quelques avancées dans le milieu de la recherche académique : modèle conjoint distribution-choix modal, modèle basé sur les activités, modèle logit emboîté de choix modal, etc. Mais en pratique quotidienne, le modèle à quatre étapes séquentielles demeure, et la rétroaction entre les quatre étapes, lorsqu'elle est envisagée, se fait presque toujours en mode manuel. D'autre part, un modèle de prévision de trafic permet essentiellement de calculer, à partir de données socio-économiques exogènes fixées à un horizon futur et supposées implicitement indépendantes de l'évolution de l'offre de transport, des gains de temps procurés par une nouvelle infrastructure. En économie des transports, il est habituel de monétariser ces gains de temps grâce à la valeur du temps, calculer le surplus actualisé pour ensuite aboutir à un taux de rentabilité immédiat et un taux de rentabilité interne pour l'ensemble de la collectivité, compte tenu du montant de l'investissement, de l'année envisagée de mise en service, de la longévité de l'infrastructure et du taux d'actualisation qui est une valeur tutélaire. Or, les enquêtes globales de transports (EGT) réalisées en Ile-de-France montrent que ces gains de temps n'existent pas car ils sont convertis graduellement en gains de distance par les usagers grâce à l'amélioration de la performance des réseaux. Ceci a pour conséquence un phénomène, que l'on observe dans toutes les métropoles mondiales : l'étalement urbain (« urban sprawl »). La conjecture de Zahavi émise en 1980, qui stipule que les budgets temps de transport sont stables, se trouve ainsi confirmée pour la région parisienne depuis 25 ans. Cette conjoncture, même si elle a été remise en question récemment pour d'autres métropoles, en particulier aux Etats-Unis, aura mis finalement presque 20 ans pour diffuser dans les esprits des modélisateurs français. Par conséquent, il faut aujourd'hui repenser la méthode même d'évaluation socio-économique des transports qui ne repose quasiment que sur des gains de temps (les gains environnementaux sont marginaux). Il faut reconnaître à M. Jean Poulit d'avoir essayé depuis dix ans de renouveler l'approche en proposant de calculer des indicateurs de gains d'accessibilité et de cartographier les distributions spatiales de ces bénéfices. Mais au final, cette nouvelle approche ne fait que représenter autrement des gains de temps, et ce ne sont donc toujours que les avantages *directs* de l'infrastructure qui sont mis en avant pour évaluer le projet. Mais ces avantages directs sont globaux et ne disent pas qui sont les gagnants et qui sont les perdants. Pour l' élu local qui est directement concerné par l'infrastructure nouvelle car elle dessert son territoire d'administration, si les usagers peuvent aller plus loin, cela devient un effet indésirable car l' élu local préfère que les usagers « consomment sur place » et, au contraire, n'aillent pas chercher ailleurs ce qu'ils peuvent trouver à proximité (emplois, commerces, écoles, etc). Cette nouvelle approche par les gains d'accessibilité peut donc être contre-productive au niveau local, même si elle est intéressante pour évaluer un bilan global au bénéfice de l'ensemble de la collectivité.

Ensuite, la **deuxième raison** de l'insuffisance des modèles classiques à quatre étapes est qu'ils ne permettent pas d'évaluer les effets *indirects* des transports dans la durée. D'aucuns ont parlé d'effets structurants des transports. Quels sont ces effets indirects ? Ils sont de trois ordres : le développement économique générateur d'emplois, l'attractivité résidentielle des ménages et la plus-value foncière. Ces effets indirects n'apparaissent pas le jour de l'inauguration de l'infrastructure, mais des années après, parfois dix à quinze après la mise en service en ce qui concerne le développement économique. Ces trois effets sont très parlants aux fameux décideurs politiques. En effet, leurs critères de décision principaux concernent les impacts qu'aura le projet sur l'urbanisation et sur l'attractivité de leur territoire (commune, intercommunalité, département). Ils se demandent : Combien de nouvelles entreprises vont s'implanter près de l'infrastructure ? Quelles sont les retombées économiques en terme de taxe foncière ? Combien de ménages nouveaux, si possible à hauts revenus, vont s'installer près de ma nouvelle gare ? Quelle plus-value mon administration va-t-elle pouvoir récupérer si le foncier et l'immobilier prennent de la valeur après l'ouverture de l'infrastructure ? Il est bien fini le temps où ingénieur et économiste pouvaient convaincre les décideurs de l'intérêt du projet de transport en leur apportant sur un plateau le taux de rentabilité interne de cette infrastructure, quitte à l'augmenter en jouant sur le taux d'actualisation ou sur la valorisation des gains

environnementaux, ou pire encore sur le montant de l'investissement. Des grands maîtres de l'économie des transports (Quinet et Walliser) ont bien avant nous, en 1990, parlé des « splendeurs et misères du calcul économique ». Loin de nous l'idée que le calcul économique est impuissant mais il est insuffisant dans le contexte actuel. Il faut se résigner et repenser la méthode. Nous proposons de revenir à une analyse multicritères qui ne cherche pas à se substituer à la méthode coûts-bénéfices mais à la compléter.

Enfin, la **troisième raison**, suggérée dans la réponse à la première question, est que la France est en retard sur les Etats-Unis, et même par rapport à certains pays européens (Pays-Bas, Ecosse, Suède, Allemagne, Belgique) dans le développement des modèles « land-use ». Il n'existe pas en France de logiciel d'urbanisation, alors qu'aux Etats-Unis, il y en a plusieurs, comme on l'a évoqué. En France, il y a eu une seule tentative pour appliquer un modèle d'urbanisation. Ce fut Tranus appliqué à Lyon à la fin des années 90 par le CERTU et la tentative s'est soldé par un échec (problème de ressources humaines et de calage). Il peut y avoir plusieurs explications à ce retard français :

- l'absence de loi comme aux Etats-Unis obligeant à l'intégration urbanisation – transports dans un document de planification de type Contrat de Plan ou de Projets, Plan de déplacements urbains (PDU) ou Plan régional de la qualité de l'air (PRQA).
- l'absence de logiciel de modélisation de l'urbanisation accessible, compréhensible, et suffisamment complet, destiné aux professionnels, alors qu'il existe plusieurs logiciels commerciaux de simulation de trafic (Davisum, Cube et Emme/2 pour les trois principaux utilisés en France). Pour les praticiens qui seraient un peu familiarisés avec les jeux vidéo ludiques, il serait souhaitable de trouver sur le marché un logiciel comparable au jeu SimCity mais à destinée professionnelle.
- les évidentes difficultés méthodologiques de la modélisation, ne serait-ce que parce qu'il est extrêmement difficile de dissocier les effets purement transports des autres effets sur l'urbanisation.
- le manque de polyvalence des ingénieurs de trafic, économètres, urbanistes, démographes et autres informaticiens. Les premiers n'ont en général qu'une approche mathématique et informatique de la simulation de trafic, les seconds vont consacrer tout leur temps à estimer des modèles explicatifs sans avoir une vision globale du projet et des interactions, les troisièmes n'ont en général qu'une approche qualitative des problèmes basée sur une longue expérience et pour qui les transports ne sont qu'une branche de l'urbanisme, les quatrièmes raisonnent essentiellement en termes de soldes migratoire et naturel au niveau d'une commune (avec des paramètres tels que le taux de natalité, de décès, d'immigration, l'espérance de vie, ...) et leurs modèles démographiques de type âge-cohorte ne prennent pas en compte les variables d'accessibilité et de desserte par les transports, enfin pour les derniers, et pour paraphraser Clémenceau, la « modélisation est une chose trop grave pour la confier à des informaticiens ». La barrière reste encore étanche entre ces cinq différentes spécialités.

C'est à partir de ce triple constat que se fonde l'opportunité de la modélisation intégrée occupation des sols – transports en France. Les données zonales de l'étape de génération ne sont plus des hypothèses exogènes mais sont calculées par UrbanSim qui répartit chaque année les populations et les emplois sur l'aire d'étude à partir de totaux. Ainsi, pour évaluer l'intérêt d'un projet d'infrastructure de transport (par exemple le projet de rocade de métro en proche couronne de Paris), ce type de modèle permet justement d'évaluer, en plus des avantages directs (gains de temps), les effets indirects de l'infrastructure sur l'occupation des sols en prenant en compte le court, le moyen et le long terme. Il permet donc par rapport à la modélisation de trafic pure des évaluations dans les deux sens, à savoir les impacts des politiques de transport sur l'urbanisation, et réciproquement les effets des politiques urbaines sur la mobilité. Par exemple, quels sont les effets à long terme sur la relocalisation des entreprises et des ménages de l'instauration d'un péage cordon en proche couronne autour de Paris ? Quels sont les effets d'une politique de densification autour des gares sur les comportements de mobilité, en particulier sur le choix modal et les distances parcourues ? Ce type d'évaluation est

très demandé après le Grenelle de l'environnement qui a fixé la priorité aux critères de développement durable des transports.

Quelle est la traduction des termes « land use model » et « LUTI model » ?

Une revue de la littérature montre des traductions diverses en français. On trouve ainsi pour « land use model » : modèle d'occupation du sol, modèle de localisation, modèle urbain, modèle d'urbanisation. Le premier peut être confondu avec Mode d'Occupation du Sol (MOS) qui est l'inventaire numérisé de la couche d'occupation du sol du Système d'information géographique régional (SIGR) de l'IAU-îdF. « Modèle de localisation » évoque les deux modèles partiels que sont le modèle de choix de localisation résidentielle des ménages et le modèle de choix de localisation des entreprises mais ignore les deux autres modèles qui font partie intégrante d'un modèle land-use : un modèle de prix et un modèle de développement du type urbain ou un modèle de localisation de projet urbain. Les prix du foncier et/ou de l'immobilier résidentiel, de bureaux, d'activités et de commerces sont déterminants dans les choix de localisation, tandis que le modèle de développement du type urbain agit sur ces choix comme producteur de contraintes de capacité (un ménage ne peut emménager dans un quartier s'il n'y a plus de logement disponible ou de projet de construction résidentielle). Ce sont ces 4 modèles partiels qui forment un ensemble indissociable, interagissant les uns sur les autres. Le terme « localisation » semble donc trop restrictif et de plus peut évoquer d'autres notions qui n'ont rien à voir tels que les systèmes de localisation GPS (Global Positioning System). Quant à « modèle urbain », c'est un terme consacré chez les urbanistes qui recouvre des concepts de formes urbaines (ville monocentrique, polycentrique, étalée, dense, compacte, linéaire, ...). C'est pour cette raison que notre choix de terminologie se porte au final sur « modèle d'urbanisation ». Il évoque de plus un aspect dynamique que l'on ne ressent pas dans « modèle urbain ». C'est aussi le terme « modèle d'urbanisation » qui a été utilisé dans les années 60 par l'IAURP (devenu IAURIF en 1976 puis IAU-îdF en 2008). Rendons hommage à l'IAURP qui a été le précurseur des recherches sur les « land use model » en France (voir Cahier de l'IAURP, vol 11, juillet 1968) en s'inspirant des théories développées aux Etats-Unis sur la croissance urbaine (modèle de Lowry, d'Alonso, etc).

Quant au « LUTI model » (land use transport integrated model), on a relevé diverses traductions : modèle intégré urbanisation-transport, modèle intégré occupation des sols-transport, modèle interactif urbanisation-transport. On utilisera indifféremment par la suite ces expressions, et de manière abrégée, « modèle intégré » ou « modèle interactif ».

Quels sont les deux modèles qui ont été couplés ?

Nous avons utilisé deux modèles existants :

- le modèle de trafic **Davisum-METROPOLIS**. Ce modèle multimodal comporte trois parties : un modèle classique de prévision de la demande de déplacements sous forme de matrices origine-destination (génération, distribution, choix modal) en 606 zones, un modèle d'affectation statique en transport en commun (Davisum) et un modèle d'affectation dynamique routière (METROPOLIS).
- le **modèle d'urbanisation UrbanSim**. Ce modèle américain a connu entre 2000 et 2008 de nombreuses évolutions et plusieurs phases de réingénierie. C'est un modèle de type micro-simulation, qui comporte quatre principaux modèles partiels (ou sous-modèles) : un modèle de choix de localisation résidentielle des ménages, un modèle de choix de localisation des emplois, un modèle de choix de développement urbain (dans la dernière version OPUS/UrbanSim, c'est un modèle de choix de localisation de projets) et enfin un modèle de prix du foncier. Les trois premiers modèles sont des

modèles de choix discrets, le quatrième est un modèle de prix hédonique. UrbanSim exploite largement la théorie des modèles de choix discrets conçue par McFadden, qui a reçu en récompense le prix Nobel d'économie en 2000.

En principe, on cherche à modéliser dans une approche holistique ou systémique les interactions suivantes : la localisation des ménages et des emplois génère des déplacements qui sont affectés sur le réseau multimodal de transport. Les coûts généralisés de transport et les mesures d'accessibilité sont, en retour, injectés dans l'expression de l'utilité de chaque modèle de choix. Les disponibilités foncières et immobilières ainsi que leurs prix sont modifiés à chaque horizon temporel. Les réseaux de transport sont mis à jour à chaque pas temporel. Cette boucle de rétroaction annuelle doit tourner pendant 25 à 30 ans.

Quel est l'intérêt d'utiliser un modèle dynamique plutôt qu'un modèle classique statique ?

Le logiciel Davisum est un logiciel du commerce, développé et commercialisé par la société allemande PTV. Développé par le laboratoire THEMA de l'Université de Cergy-Pontoise, le logiciel METROPOLIS est quant à lui un pur produit de la recherche. C'est en 2003 que la société PTV a racheté METROPOLIS et l'a intégré à son produit phare Davisum. Par rapport à un modèle d'affectation routier statique, METROPOLIS présente trois aspects novateurs : d'abord, le modèle est dynamique. Il produit ainsi des indicateurs d'accessibilité qui prennent en compte l'évolution de l'état de trafic sur la période simulée (par exemple une journée). Ensuite, il intègre un modèle de choix de l'heure de départ basé sur l'heure de désir d'arrivée et le temps estimé du déplacement, ainsi qu'un modèle de choix modal qui est cependant plutôt rudimentaire. Enfin, il offre une représentation microscopique de la demande à partir d'une matrice statique. Cette représentation permet de tenir compte de l'hétérogénéité de la demande.

Le processus itératif entre la modélisation de l'urbanisation et la modélisation de trafic permet de prendre en compte des décisions individuelles qui se situent à des échelles de temps très différentes :

- le très court terme pour les choix de destination
- le court terme pour les choix d'heure de départ
- le court-moyen terme pour les choix d'itinéraire et de mode de transport
- le moyen-long terme pour les choix de localisation résidentielle ou de localisation d'activités.
- le long terme pour l'évolution des types urbains de l'occupation du sol.

Le choix de localisation résidentielle des usagers de la voiture dépend des conditions de circulation sur les trajets domicile-travail à l'heure de pointe, donc en période de congestion. METROPOLIS permet de prendre en compte les décisions à plusieurs échelles de temps, contrairement à un modèle statique, sans toutefois aller jusqu'à modéliser de façon simultanée le choix de localisation résidentielle et les choix de l'heure de départ, de mode et d'itinéraire. Mais il faut reconnaître que l'affectation dynamique est un luxe pour un modèle intégré. Si l'on dispose d'un modèle statique performant, ce sera déjà bien. D'ailleurs, pour la partie TC, c'est un modèle classique d'affectation sur Davisum qui a été utilisé.

Pourquoi avoir choisi UrbanSim ?

L'équipe dirigée par le Professeur Waddell à l'Université de Washington a commencé à développer UrbanSim en 1999. Compte tenu de son importance stratégique, l'équipe a reçu en 2001 une

subvention de 5 millions de dollars de l'Etat fédéral américain pour poursuivre son développement. Immédiatement, il a suscité l'intérêt de toute la communauté scientifique des modélisateurs de trafic.

UrbanSim présentait de prime abord de nombreux avantages pour notre projet :

- Il est l'aboutissement de nombreuses recherches commencées dès les années 60, d'abord en Amérique latine, puis aux Etats-Unis. Le Professeur Waddell a su tirer des expériences de l'utilisation d'autres logiciels existants tels que MEPLAN, Tranus, MetroSim, ITLUP, DRAM/EMPAL pour concevoir son logiciel. Il intègre donc toutes les fonctionnalités d'un logiciel de prévision de l'occupation du sol, en particulier un modèle de prix du foncier et des modèles de choix discret de localisation des ménages et des entreprises.
- Le logiciel a été conçu dès le départ pour fonctionner en interaction avec différents modèles externes classiques de prévision de trafic. De plus, en post-traitement, les résultats peuvent être exportés dans un système d'information géographique pour être visualisés, ce qui constitue un élément de communication et de diffusion incontournable aujourd'hui.
- Le logiciel fonctionne sur le principe de la micro-simulation : chaque ménage, chaque emploi et chaque unité géographique (cellule de 500 mètres de côté) sont représentés dans le modèle, chacun avec ses caractéristiques et ses comportements. Toutes les hétérogénéités sont ainsi prises en compte. Le modèle intégré va faire passer chaque agent dans le processus de simulation (par exemple probabilité pour un ménage et pour un emploi de se localiser dans telle cellule, probabilité pour une cellule vacante de passer à la catégorie habitat). En modélisation de trafic, on peut faire le parallèle avec les modèles de micro-simulation de type VISSIM (logiciel commercialisé par PTV).
- L'équipe a su fédérer l'ensemble des utilisateurs potentiels dans une communauté « virtuelle » en créant une liste de diffusion internationale et un site Internet constamment mis à jour (www.urbansim.org). Chaque mois, des nouveaux utilisateurs rejoignent cette communauté. On trouve sur le site de très nombreux articles, des rapports d'étude et de la documentation en ligne sur le logiciel.
- L'équipe du Professeur Waddell a opté pour la stratégie du logiciel dit « libre » (sous licence publique GNU) pour la diffusion d'UrbanSim. Celui-ci est développé entièrement en langage Java (langage « libre » lui aussi) et en langage Python dans sa dernière version. Le code source est disponible de sorte que tout utilisateur peut l'utiliser, le modifier et le redistribuer. UrbanSim est gratuit et téléchargeable sur Internet. Il est décrit comme étant un outil moderne, c'est-à-dire modulaire, ouvert et robuste. On exposera dans la partie 3.3 les grandes difficultés que nous avons rencontrées dans l'utilisation d'UrbanSim.

Il faut reconnaître que c'est la gratuité de ce logiciel qui a été l'élément déclencheur de notre recherche en 2002. Il s'avère que le Professeur Waddell a été mis au courant dès le début de notre recherche et de notre intérêt pour son logiciel. Il a accepté de nous apporter toute l'aide possible, tant au niveau de conseils méthodologiques qu'en adaptant, si nécessaire, le code informatique d'UrbanSim au cas de l'Ile-de-France. En effet, UrbanSim a été conçu dans le contexte du système et de l'économie urbaine américaine. Les variables expliquant tel phénomène aux Etats-Unis (par exemple le prix du terrain pour le choix résidentiel, le taux de motorisation pour le choix modal) ne sont peut-être pas discriminantes dans le cas de la région Ile-de-France. De même la variable d'accessibilité par le mode véhicule particulier apparaît sur-représentée dans les différents modules d'UrbanSim, ce qui est compréhensible dans le cas américain où le véhicule particulier est le mode prépondérant. Mais on ne trouve pas de variables décrivant l'accessibilité et la desserte par les transports en commun. Il a fallu remédier à cette lacune lors de l'adaptation au cas de l'Ile-de-France.

Invité par l'IAU-îdF et l'Université de Cergy-Pontoise, le Pr Paul Waddell a fait une présentation d'UrbanSim lors de la 5ème Journée Transports organisée par l'Université de Cergy-Pontoise le 24 octobre 2002. L'application d'UrbanSim sur une aire aussi grande que la région d'Ile-de-France

devenait pour lui une première et une référence. Séduit par le « challenge » que constituait notre projet de recherche, il nous a fait l'honneur de rejoindre notre équipe dans le cadre d'une année sabbatique subventionnée partiellement par la National Science Foundation (NSF). Entre septembre 2004 et juin 2005, Paul Waddell a ainsi travaillé à l'IAU-îdF et à l'Université de Cergy-Pontoise en tant que chercheur associé. Sa participation au projet a constitué un encouragement fort.

Comment se positionne le projet SIMAURIF dans la recherche académique ?

Nous considérons que la recherche théorique sur l'interaction urbanisation-transport restera sans lendemain si des efforts importants de modélisation ne sont pas engagés. On peut s'interroger longtemps en effet sur l'existence ou non d'une relation entre urbanisation et transports, mais cela mènerait à des discussions sans fin sur la nature des relations (directes, indirectes, de concomitance, de congruence, etc) entre système de transport et système urbain. Nous avons préféré laisser de côté ce débat pour modéliser sans état d'âme. Notre objectif était donc de mettre au point un modèle intégré opérationnel, ou du moins un prototype, et de l'appliquer à une étude de cas : le territoire desservi par la Tangentielle Nord entre Sartrouville et Noisy-le-Sec.

Evidemment, nous n'ignorons pas les nombreuses recherches menées avant nous dans le domaine de l'économie urbaine, l'économie spatiale et la nouvelle économie géographique par d'éminents chercheurs. Loin de prétendre à l'exhaustivité, citons en quelques uns trouvés dans une littérature abondante : Von Thünen (1826) sur la rente foncière, Losch (1954) sur les choix de localisation, Lowry (1963) et Alonso (1964) sur la croissance urbaine, Krugman (1985) sur la localisation des entreprises, Fujita et Thisse (1989) sur la dynamique urbaine et les mécanismes d'enchères foncières, Arnott (2001) sur les villes monocentriques et polycentriques. Mais le projet *SIMAURIF* n'a pas cherché à « faire de la littérature » en plus mais bien à aboutir à un outil *opérationnel* de modélisation intégrée, dynamique, désagrégée de l'interaction transport-urbanisation, quitte à transgresser les principes de la recherche académique. C'est volontairement que nous n'avons pas inséré de bibliographie dans ce rapport pour montrer que nos travaux n'ont pas été influencés par des recherches antérieures. Pour cette raison, cette recherche est réellement inédite en France.

Que cherche-t-on à modéliser exactement ?

Ce projet de modélisation intégrée est une première en Ile-de-France, rendu encore plus ardu par la dimension de l'aire d'étude (12 000 km²) et la complexité des interactions spécifiques au cas français (figure 1). Il est évident que l'on ne peut pas modéliser l'ensemble de ces interactions. On choisit de modéliser les 4 phénomènes principaux suivants :

- le choix de localisation résidentielle des ménages
- le choix de localisation des emplois
- le choix de développement du type urbain ou le choix de localisation de projet urbain
- le modèle de prix de l'immobilier

On cherche parmi l'ensemble des variables pouvant influencer ces phénomènes à faire ressortir, lorsque les variables sont réellement déterminantes, celles qui ont trait au transport : temps de parcours en mode VP (véhicule particulier) et en mode TC (transport en commun), accessibilité locale aux TC, accessibilité régionale (« log sum »), distance aux voies rapides et aux routes nationales, distance au centre de Paris. Par ailleurs, la figure 1 a un grand défaut : elle ne montre pas la dimension du temps. Or ce sont justement les effets à différents termes qu'un modèle intégré essaie de simuler. Donc on simule sur une longue période, en l'occurrence de 1999 à 2026, avec un pas d'un an pour le modèle de trafic et un pas de trois ans pour UrbanSim (figure 2). C'est en ce sens que le

modèle intégré est dit « dynamique », alors que dans un modèle statique, le temps n'est pris en compte qu'une seule fois de manière exogène à l'année de mise en service du projet.

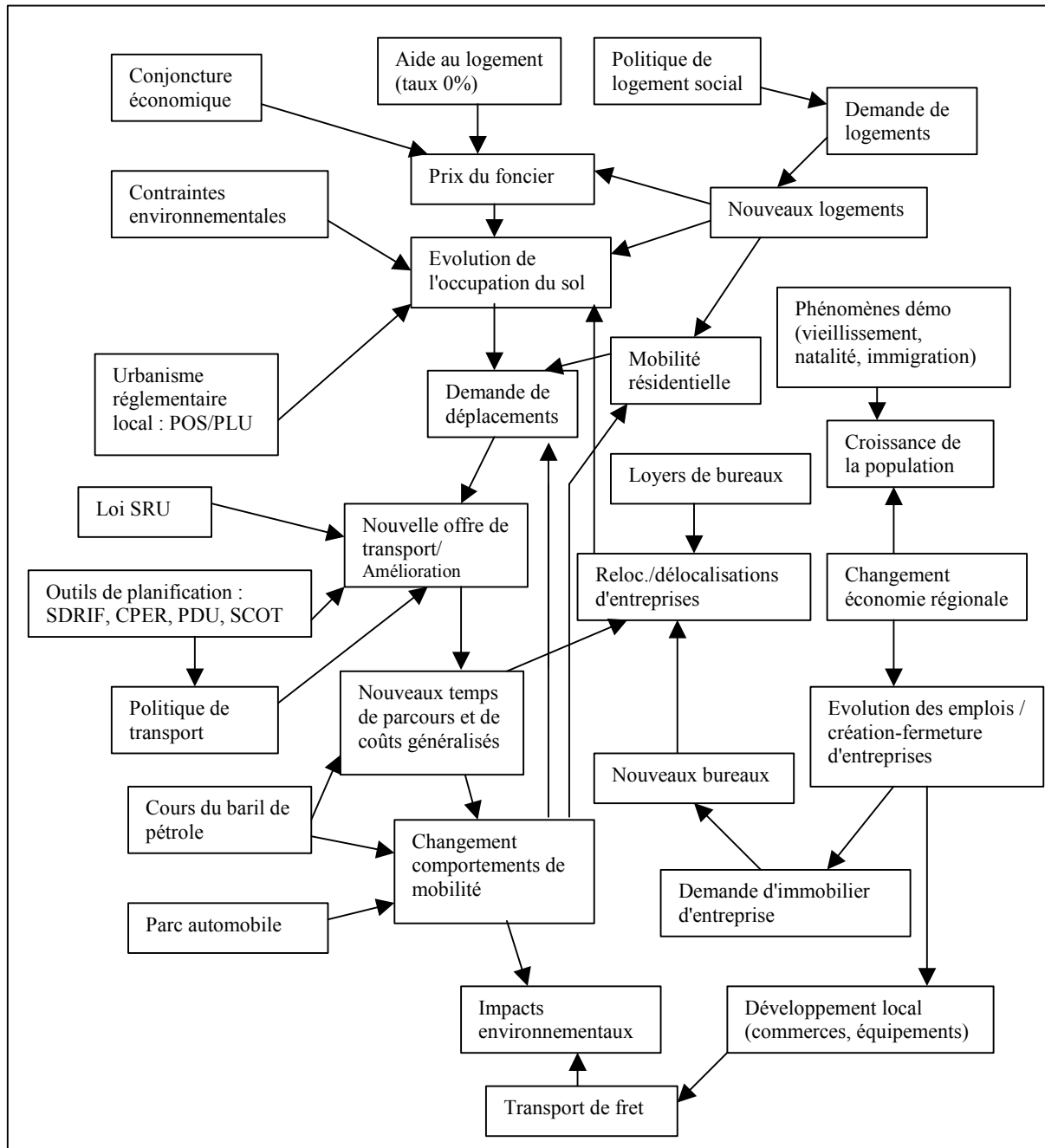


Figure1 : Schéma des interactions complexes entre système urbain et système de transports en Ile-de-France

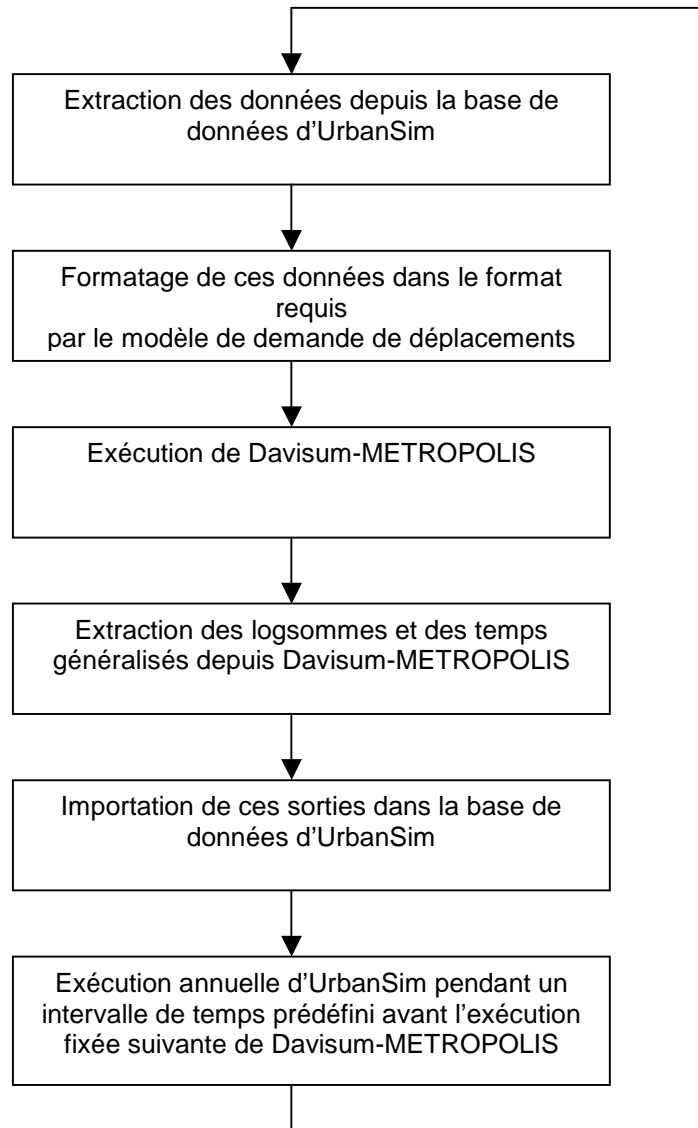


Figure 2 : Schéma du bouclage dans le temps entre Davisum-METROPOLIS et UrbanSim produisant la dynamique du système

En résumé, un modèle intégré doit prendre en compte au moins les facteurs suivants :

- les choix de localisation résidentielle des ménages
- les choix de localisation des entreprises
- les marchés fonciers et immobiliers (résidentiel, bureaux, commerces)
- les choix de localisation des projets immobiliers ou de développement du type urbain ou d'évolution du mode d'occupation du sol
- les comportements de mobilité (choix de la destination, du mode)
- les gains d'accessibilité
- le temps

La complexité vient du fait qu'il faut relier ces modèles partiels, sachant que les sorties des uns peuvent être les entrées des autres, puis les assembler pour une exécution automatisée dans une boucle de rétroaction temporelle.

Les objectifs des modèles LUTI peuvent être doubles :

- calculer des indicateurs de développement durable
- contribuer à l'évaluation socio-économique des infrastructures de transport.

Quelles sont les différences entre les projets SIMAURIF, SIMBAD, MobiSim, MIRO et ILOT ?

Ces cinq projets sont réalisés dans le cadre du Predit. Ils visent tous à simuler l'interaction transport-urbanisation mais sur des principes différents : *SIMAURIF* et *SIMBAD* (Simuler les MoBilités pour une Agglomération Durable) optent pour une modélisation désagrégée, dynamique et basée sur la théorie des choix discrets. Les trois autres se basent sur les systèmes multi-agents issus de l'approche automate cellulaire mais s'en différencient par la prise en compte de la confrontation des règles de décisions des agents. *MIRO* (Modélisation Intra-urbaine des Rythmes Quotidiens) est développé par le Laboratoire d'Informatique de l'Université de Franche-Comté, *ILOT* (Interactions Localisations Transport) et *SIMBAD* par le LET à Lyon, et *Mobisim* par ATN puis repris par le laboratoire Théma-Besançon. *SIMBAD* cherche à calculer des indicateurs de développement durable, *SIMAURIF* cherche plutôt à calculer les avantages indirects des infrastructures. La grande originalité de *SIMAURIF* est le traitement de données désagrégées à une échelle infracommunale.

Quelles sont les applications terminées d'UrbanSim dans le monde ?

SIMAURIF est le premier projet terminé dans le monde, en dehors des Etats-Unis, à utiliser la version OPUS/UrbanSim. Aux Etats-Unis, seules trois villes sont allés, semble-t-il, au bout de la démarche : Puget Sound, Salt Lake City et le Comté de Volusia (Floride). Il existe de nombreux projets de par le monde, mais aucun à notre connaissance n'est encore terminé en juin 2008 : Detroit, Honolulu, Houston, Phoenix (Arizona), San Francisco, Burlington (Vermont), Durham (Caroline du Nord), Melbourne, El Paso (Texas), Amsterdam, Tel Aviv, Zurich, Pescara et Turin (Italie).

Quel est l'indicateur d'accessibilité « log sum » utilisé dans SIMAURIF ?

Décrivons d'abord en quelques mots le modèle logit multinomial. Il fournit une expression simple et facilement interprétable des probabilités de choix. C'est ce qui explique sa popularité, augmentée encore depuis l'attribution du prix Nobel d'économie à McFadden en 2000.

A l'individu n et à chaque alternative j , $j = 1, \dots, J$, on associe une utilité aléatoire :

$$U_{nj} = V_{nj} + \varepsilon_{nj} \quad \forall j$$

où V_{nj} est l'utilité représentative, connue par le chercheur, et où les perturbations $\varepsilon_{n1}, \dots, \varepsilon_{nJ}$ sont des variables aléatoires i.i.d. (indépendantes et identiquement distribuées) dont la distribution est une distribution de Gumbel standard (ou loi des valeurs extrêmes), dont la fonction de répartition et la densité sont :

$$F(\varepsilon_{nj}) = \exp(-\exp(-\varepsilon_{nj}))$$

$$f(\varepsilon_{nj}) = F'(\varepsilon_{nj}) = \exp(-\varepsilon_{nj})F(\varepsilon_{nj})$$

L'hypothèse d'indépendance des erreurs est une hypothèse restrictive qui n'a plus cours dans le modèle logit emboîté que l'on trouve dans la version OPUS/UrbanSim. Néanmoins, cette hypothèse d'indépendance peut aussi s'interpréter comme le résultat naturel d'un modèle bien spécifié. Le chercheur a à sa disposition suffisamment d'information pour bien spécifier V_{nj} de telle sorte que la portion d'utilité inobservée ε_{nj} est essentiellement un bruit blanc.

On suppose que l'individu n choisit l'alternative lui procurant l'utilité maximale et que les alternatives sont indépendantes les unes des autres (hypothèse « IIA : indépendance of irrelevant alternatives »), contrairement au modèle probit. En conséquence, on peut démontrer que la probabilité que cet individu choisisse l'alternative i est :

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$

En général, l'utilité représentative, V_{nj} , est linéaire en les paramètres :

$$V_{nj} = \beta' x_{nj}$$

où x_{nj} est le vecteur des variables observées qui sont reliées à l'alternative j. Dans ce cas, la probabilité que l'individu n choisisse l'alternative i prend la forme suivante :

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}}$$

McFadden a démontré en 1974 que la fonction log vraisemblance associée à ces probabilités était globalement concave en les paramètres β .

Le terme $I_n = \ln\left(\sum_{j=1}^J e^{V_{nj}}\right)$ est généralement appelé le "log sum".

Il représente à une constante près l'espérance d'utilité de l'individu n.

Dans le projet *SIMAURIF*, on a été confronté pour chacun des modèles partiels au problème de l'estimation sur un sous-ensemble d'alternatives. L'estimation d'un modèle logit standard s'effectue habituellement par des méthodes de maximum de vraisemblance standards. Néanmoins dans certains cas, le nombre d'alternatives auquel un agent fait face est tellement élevé que l'estimation des paramètres du modèle peut se révéler très coûteuse voire impossible. Avec un modèle logit, l'estimation peut être mise en oeuvre sur un sous-ensemble des alternatives sans entraîner des problèmes d'inconsistance. Supposons par exemple que l'on examine un modèle où les agents ont le choix entre 100 alternatives. Il est alors possible de procéder à une estimation sur un sous-ensemble de 10 alternatives seulement pour chaque agent faisant partie de l'échantillon étudié. Le sous-ensemble examiné contient alors l'alternative choisie par l'agent plus 9 autres alternatives qui ont été sélectionnées de façon aléatoire parmi les 99 restantes. Si toutes les alternatives ont la même probabilité d'être sélectionnées dans le sous-ensemble, alors la procédure d'estimation sur le sous-ensemble est équivalente à l'estimation sur l'ensemble complet des alternatives. Si en revanche les alternatives ont une probabilité inégale d'être sélectionnées, il est alors nécessaire de mettre en oeuvre des procédures d'estimation un peu plus complexes.

SIMAURIF met en oeuvre le concept d'accessibilité qui est à l'interface entre système de transport et système urbain. Elle est mesurée par le log sum. Concrètement, pour chacune des cellules, on calcule dans *SIMAURIF* les quatre log sum suivants :

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

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

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

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

où :

- λ est un facteur de pondération qui vaut $\lambda = \frac{C_0}{\alpha}$, C_0 étant le coût d'une heure de transport², et α est le coefficient exprimant la rapidité de l'effet d'atténuateur de la distance dans la loi de distribution des déplacements (les valeurs de α qui ont été calculées pour les motifs D->T+aff pro et D->Achats sont dans le rapport 2005).
- C_{ij} est le coût généralisé du transport qui vaut $C_{ij} = C_0 t_{ij}$, t_{ij} étant le temps de transport nécessaire pour aller de la cellule i à la cellule j

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

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

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

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

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

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

² Le rapport « Boiteux II » (juin 2001) préconise les valeurs suivantes : 11,6€ pour le motif D->T et 6,4 € pour le motif D->Achats

C'est cette dernière formule³ qui a permis de calculer les variables d'accessibilité. Pour les deux variables d'accessibilité à l'emploi, Q_j est le nombre d'emplois de la cellule j . Pour les deux variables d'accessibilité aux commerces, Q_j est le nombre de surfaces commerciales de la cellule j . Quant aux temps de transports t_{ij} , ils sont calculés par METROPOLIS pour les temps VP en heures creuses (donc sans affectation de matrice origine-destination) et par Davisum pour les temps TC⁴.

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

Il est intéressant de remarquer que les modèles logit sont autant à la mode depuis le milieu des années 90 que l'étaient les méthodes de recherche opérationnelle dans les années 60-70 (programmation linéaire, optimisation, etc) et que les problèmes de modélisation de la croissance d'une métropole se posaient déjà il y a 50 ans. Seule leur mise en équation et les moyens informatiques étaient différents.

Pourquoi avoir choisi trois motifs de déplacements seulement ?

Ce choix des trois motifs (Domicile-Travail+Affaires professionnelles, Domicile-Commerce et Autres) est cohérent avec la structure de données pré-établie d'UrbanSim. Dans la table gridcell, les attributs par défaut concernent les emplois et les commerces, mais il n'y a pas d'attribut concernant les motifs Ecole, Santé, Loisirs. Le modèle de demande de l'IAU-îdF a donc été recalé à partir de l'EGT 2001-2002 sur ces trois motifs. On trouvera dans le rapport 2005 les paramètres estimés des modèles de génération (régression linéaire), distribution (modèle gravitaire classique) et choix modal (logit binomial, en l'occurrence bimodal).

Comment le modèle de prix d'UrbanSim a-t-il été adapté dans le projet SIMAURIF ?

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. Parmi ces caractéristiques, on trouve le gain d'accessibilité et le gain environnemental (requalification de la voirie urbaine, baisse du trafic routier, etc) procurés par une nouvelle infrastructure de transport, qui induit une plus-value immobilière pour le propriétaire.

L'intérêt de mesurer des plus-values immobilières est expliqué dans le rapport Boîteux II (2001), pages 69 : « Reste une question à évoquer qui est celle de la création et de la récupération éventuelle de la rente foncière créée par l'amélioration de la 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). Il est clair, et communément admis, que de telles plus-values ne doivent pas être prises en compte au titre des avantages d'un projet, car

³ Le passage de la formule (2) à (3) peut surprendre du point de vue mathématique, la fonction log n'étant pas linéaire. En fait la formule (3) ne s'obtient pas en additionnant sur l'indice j , mais par la transposition de l'expression de la satisfaction ressentie par l'utilisateur pour une O-D donnée à celle ressentie par l'utilisateur pour l'ensemble des destinations depuis la zone i .

⁴
$$\left[-\frac{C_{ij}}{\lambda} \right] = \left[-\alpha t_{ij} \right]$$

elles peuvent être interprétées comme une internalisation par les propriétaires des alentours d'avantages déjà pris en compte dans l'évaluation, tels les gains de temps. Le problème de la récupération de ces rentes par la puissance publique ne s'en pose pas moins ; il sort de l'objet du présent rapport. »

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). Une description technique de ce modèle est présentée dans la partie 4.3.1.

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 assez 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.

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.

Dans UrbanSim, le marché de la demande immobilière est déterminé à chaque localisation ainsi que les acteurs et les processus de choix qui influencent la configuration de l'espace urbain et le prix sur le marché immobilier.

Le modèle de prix du foncier simule les prix du sol de chaque cellule en suivant les changements des caractéristiques des localisations au cours du temps. Pour chaque zone, la valeur du sol dépend de caractéristiques extrinsèques (voisinage, accessibilité, distances aux aménités, etc). La fonction de prix hédoniques du marché est estimée sur la base des caractéristiques et des valeurs d'une seule

année. Le modèle de prix du sol simule donc la valeur du sol à chaque fin d'année après l'exécution de l'ensemble des autres modèles. Ces prix influencent alors les choix de localisation des agents et les comportements des promoteurs pour l'année suivante.

Dans *SIMAU*RIF, ce n'est pas le prix du foncier mais le prix de l'immobilier que nous avons essayé de modéliser. Pour chaque zone, la valeur du m² de logement dépend des caractéristiques intrinsèques du logement et de caractéristiques extrinsèques (voisinage, accessibilité, distances aux aménités, etc). La difficulté de modélisation ne vient pas des bases de données qui existent, en particulier la base BIEN des notaires qui est géolocalisée, mais de la multitude de facteurs, dont ceux dans lesquels interviennent les pouvoirs publics (dispositifs d'aide tel que le prêt à taux 0, taux d'intérêt, etc) et ceux qui ne sont pas modélisables (image d'un quartier, vue sur un monument classé, etc), sans parler des phénomènes d'anticipation des vendeurs qui faussent l'analyse désagrégée des prix. Quant aux facteurs qui nous intéressent, les facteurs transport, ils ont un effet sur les prix, mais cet effet est difficile à quantifier précisément car il est dilué dans une multitude d'autres effets et dans un contexte général de forte augmentation des prix au cours des dix dernières années.

Dans le modèle UrbanSim, 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 logement (appartement, maison, ancien, nouveau) 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_i^s du type de bâti i et le taux de vacance courant V_{it}^c 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.

Quel a été le déroulement chronologique du projet ?

Le déroulement chronologique reflète la méthodologie de projet.

- Revue de la littérature sur les modèles de trafic, les modèles d'urbanisation et les modèles intégrés
- Etude théorique d'UrbanSim (les 4 modèles principaux et 5 modèles secondaires non évoqués dans cette étude)

- Une première prise en main d'UrbanSim sur une étude de cas américaine (Eugène-Springfield)
- Conception de l'architecture du modèle intégré *SIMAU*RIF
- Recalage du modèle de demande de déplacements de l'IAU-îdF en 606 zones à partir de l'Enquête Globale Transports 2001 et pour 3 motifs de déplacements
- Constitution d'une base de données géolocalisées et désagrégées d'entrée en 50 000 cellules carrées sur l'Île-de-France
- Constitution de bases de données géolocalisées et désagrégées d'estimation à partir de différentes sources de données (enquêtes ménages, recensements, enquêtes emplois, mode d'occupation du sol, base des Notaires, etc)
- Estimation économétrique et calage des quatre modules principaux d'UrbanSim au niveau du maillage en 50 000 cellules sur la période 1990-1999
- Paramétrage de l'ensemble d'UrbanSim dans la structure de données primaire d'UrbanSim.
- Recalage de METROPOLIS à partir d'une nouvelle base de comptages
- Recalage de la partie affectation en TC du logiciel Davisum de l'IAU Île-de-France
- Mise au point de l'interface entre UrbanSim et Davisum-METROPOLIS
- Construction des graphes transport sur la période 1990-1999
- Assemblage puis calage de l'ensemble UrbanSim/Davisum-METROPOLIS sur la période 1990-1999
- Construction des graphes transport sur la période 1999-2026
- Elaboration des scénarios de la Tangentielle Nord
- Simulation des scénarios de la Tangentielle Nord sur la période 1999-2026
- Rédaction des rapports

Quelle est l'échelle d'étude ?

Par défaut, UrbanSim fonctionne sur des unités géographiques appelées cellules. Ce sont des carrés de 500 mètres de côté qui partitionnent les 12 000 km² de l'Île-de-France. Pour le modèle de trafic, c'est un découpage en 606 zones (dont 80 à Paris et 34 zones externes). Pour certains modèles partiels qui n'ont pu être estimés au niveau de la cellule, on est passé par l'intermédiaire de la commune.

Qu'entend-on par période de calage et période de simulation ?

La période de calage est la période 1990-1999. On estime et on cale chaque modèle partiel sur cette période, puis le modèle intégré est calé globalement sur cette période. Davisum-METROPOLIS tourne par pas d'un an, UrbanSim tourne aux années 1990, 1993, 1996, 1999. Les résultats de l'exécution sont comparés aux données observées en 1999. Il y a donc une différence avec les modèles de trafic purs pour lesquels on parle d'une *année* de calage (ou de référence) alors que pour les modèles intégrés, il faut parler de *période* de calage. Cela suppose aussi qu'il faut construire les bases de données à deux années : 1990 et 1999. C'est un double travail. On peut légitimement critiquer le choix d'une période trop courte (10 ans) pour caler le modèle intégré. On aurait pu prendre la période 1982-1999 mais on aurait été confronté au problème de disponibilité de données pour l'année 1982. Une fois le modèle intégré calé, on peut le faire tourner sur la période de simulation 1999-2026 avec les mêmes pas de temps. Les résultats en 2026 sont ensuite cartographiés, analysés, interprétés.

Quelles sont les bases de données d'entrée utilisées ?

Il faut distinguer les bases de données d'entrée du modèle intégré des bases de données d'estimation qui servent à estimer les différents modèles partiels. En entrée du modèle UrbanSim, il y a les 3 tables principales : gridcell, households, jobs. Ces tables comportent des données exhaustives désagrégées et sont construites pour deux années : 1990 (année de base) et 1999 (année cible). Les bases de données source sont géolocalisées.

La table gridcell comporte pour chacune des 50 000 cellules de 500 mètres de côté une cinquantaine d'attributs (voir rapport 2004), par exemple le nombre de logements, la valeur de ces logements, la distance au centre de Paris, la surface en zone inondable, la surface en zone urbaine sensible, etc

La table households comporte 4 232 526 lignes correspondant au nombre de ménages en 1990. Chaque ménage est renseigné par quelques caractéristiques (par exemple, taille, tranche de revenu, cellule de résidence, etc)

La table jobs comporte 5 075 974 lignes correspondant au nombre d'emplois en 1990. Chaque emploi est renseigné par sa catégorie (on en a distingué 8) et par sa cellule de localisation.

Comment ces trois tables ont-elles été construites ? Donnons juste un aperçu : une grande partie de la table gridcell a été construite grâce au système d'information géographique régional (SIGR) de l'IAU-îdF et ses très nombreuses couches référencées sur la couche MOS (Mode d'Occupation du Sol). Cette couche MOS comporte plus de 380 000 polygones répartis en 83 postes différents. L'intersection entre cette couche et le maillage en 50 000 cellules permet d'extraire des informations d'occupation du sol pour chaque cellule.

La table households est basée sur le fichier du Recensement général de la population. Celui-ci est exhaustif, désagrégé. L'avantage de la France sur les Etats-Unis est qu'en France nous disposons des Recensements généraux de la population et par conséquent, il n'est nul besoin d'utiliser un programme de synthèse de population comme aux Etats-Unis.

La table jobs est également basée sur le fichier du Recensement général de la population. Celui-ci n'est connu qu'à l'îlot INSEE (chaque emploi est localisé à l'îlot INSEE). Pour identifier le carré de 500 mètres où se trouve l'emploi, on a dû effectuer une intersection spatiale entre la couche îlot INSEE et le maillage.

Quelles sont les bases de données d'estimation utilisées ?

Il y avait une nécessité à travailler au niveau de la cellule, donc à utiliser des bases de données infracommunales géolocalisées.

Pour le modèle de localisation résidentielle des ménages, on a utilisé l'EGT 2001-2002. Cette enquête comporte un carroyage de 300 mètres de côté sur la région. Une correspondance a été trouvée entre ce carroyage et le maillage en 50 000 cellules.

Pour le modèle de localisation des emplois, on disposait de l'Enquête Régionale Emplois (ERE) de 1997 et 2001. Celle de 2001 est entièrement géolocalisée.

Pour le modèle de prix, on a utilisé la base de données des Notaires BIEN géolocalisées. Le calage du modèle a d'abord été effectué au niveau cellulaire sur deux territoires tests, celui du RER E et celui du tramway T1, mais la qualité de ces tests était faible. Au final, le modèle a été estimé sur des données classiques de prix moyens à la commune.

Pour le modèle de choix de développement urbain, c'est la base de données EVOLUMOS de l'IAU-îdF qui donne l'inventaire numérisé de l'occupation du sol en 1990 et 1999.

Pour recalibrer le modèle de demande de déplacements (les trois premières étapes), c'est l'EGT 2001-2002 qui a été exploitée.

Pour recalibrer les modèles d'affectation sur METROPOLIS et sur Davisum, on a utilisé des bases de données de comptages provenant du SIER et des Conseils généraux, des « serpents de charge » et des résultats d'enquêtes origine-destination aux gares provenant de la SNCF et de la RATP.

Comment le problème des données manquantes a-t-il été résolu ?

Il y a deux méthodes : fusion de bases de données et construction de données par modélisation. Par exemple, pour imputer un revenu à chaque ménage (la donnée de revenu n'est pas disponible dans le RGP), on a utilisé la base de données de la Direction générale des impôts (DGI). Autre exemple : il y avait des communes pour lesquelles la donnée de prix moyen de vente d'appartement était inconnue. On a créé un modèle linéaire pour combler les prix des communes manquantes.

Quelle est la structure de données d'UrbanSim ?

La structure de données d'UrbanSim comporte quatre types de tables :

- données
- constantes
- paramètres
- scénarios

Les trois tables de données principales sont : gridcells, households et jobs.

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

Les tables « paramètres » ont été remplies une fois que le calage des modèles d'UrbanSim a été effectué, les tables « Scénarios » ont été constituées lors de l'élaboration des scénarios de l'application du modèle intégré à la Tangentielle Nord. Les tables « constantes » comportent les constantes telles que le taux moyen de déménagement, le rayon de voisinage d'une cellule, etc

Comment UrbanSim a-t-il été adapté au cas de l'Ile-de-France ?

Il a fallu adapter UrbanSim à deux niveaux : la structure de données qui est pré-établie et la spécification des modèles partiels.

Pour la structure de données, certains attributs environnementaux ont été remplacés, d'autres ajoutés (par exemple, les variables zone inondable et zone de protection ZNIEFF). On a aussi ajouté des variables sur les transports en commun qui sont cruellement absentes dans UrbanSim. D'autres variables concernant l'environnement urbain ont été ajoutés (par exemple l'appartenance d'une cellule à une zone urbaine sensible).

On a aussi simplifié les activités en les regroupant toutes en une seule classe.

Quant à la partie modèles, le modèle de prix du foncier a été transformé en un modèle de prix de l'immobilier résidentiel. Le modèle spécifié a une forme log-linéaire et non pas linéaire comme c'est le cas par défaut.

Le modèle de choix de développement urbain est un modèle de transition d'un type à un autre. Le principe de ce modèle 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.

Ainsi on voit bien que la transposition d'UrbanSim en Ile-de-France n'était pas immédiate.

Comment les contraintes environnementales et politiques ont-elles été prises en compte ?

Elles ont été prises en compte dans la table gridcell. Les contraintes environnementales sont par exemple la surface de la cellule incluse dans une zone inondable ou une zone de protection des monuments. Les contraintes politiques peuvent être d'ordre local ou régional. La base des PLU (Plan local d'urbanisme, ex POS) a été croisée spatialement avec le maillage en 50 000 cellules. Chaque cellule est aussi renseignée par son appartenance à une zone urbanisable selon le SDRIF de 1994. Ensuite, la prise en compte de ces contraintes se fait au moment du choix des alternatives lors de l'estimation des modèles logit. Ainsi, une cellule non urbanisable car ayant trop de contraintes est d'emblée exclue de l'ensemble des alternatives sur lequel s'effectue le tirage aléatoire.

Quelles ont été les difficultés rencontrées dans l'estimation des modèles partiels ?

La principale difficulté réside dans la construction des bases de données d'estimation qui est une tâche laborieuse et ingrate car peu reconnue et très gourmande en temps de travail. Ensuite, l'autre problème rencontré est l'impossibilité d'implémenter des modèles trop sophistiqués dans UrbanSim. En fait, au départ, on a estimé les modèles partiels indépendamment les uns des autres, avec une démarche économétrique rigoureuse et autant de variables qu'il était possible, quitte à en modéliser de nouvelles pour satisfaire nos besoins. Ces estimations étaient chronophages, et au final quasiment impossible à implémenter dans la structure pré-établie d'UrbanSim. De plus, on perdait la vision globale du modèle intégré et les interactions existantes entre les différents modèles. On est alors passé à une méthode d'estimation plus systématique, en utilisant une base de variables *commune* à tous les modèles partiels, en nombre limité, dans le but d'aboutir à un modèle utilisable en prédiction et non plus un modèle explicatif. Parmi les variables exogènes testées, on essaie de faire ressortir des variables transport (accessibilités locales et régionales en VP et en TC, temps d'accès au centre de Paris) mais la base de variables est incomplète en ce qui concerne les coûts de transport.

La troisième difficulté a concerné le modèle de choix de développement urbain. En 2005, nous avons estimé un modèle de transition (voir rapport 2005) conformément à la version d'UrbanSim 2. En 2007, la nouvelle version d'UrbanSim a vu se modifier radicalement ce modèle en un modèle de choix de localisation de projet urbain. Nous avons choisi de conserver le premier modèle, mais dans la logique d'UrbanSim, le nouveau modèle est plus pertinent car il permet de simuler des contraintes chiffrées de capacité pour les deux modèles de localisation et distingue le promoteur immobilier comme nouveau agent économique. C'est en général ce modèle qui est le moins bien traité dans les modèles intégrés.

La quatrième difficulté a consisté à se prémunir des problèmes d'auto-corrélation spatiale (tests de Durbin et Watson, de Box-Pierce), d'endogénéité des variables, de colinéarité des variables et d'hétéroscédasticité⁵ (tests de White, de Goldfeld et Quandt, de Breusch et Pagan). La tâche de détection et de résolution de ces problèmes incombait au spécialiste en économétrie.

Dernière difficulté : l'échelle d'étude. On a essayé d'estimer les modèles partiels au niveau de la cellule, ce qui était inhabituel et un peu déstabilisant pour les économètres habitués à travailler au niveau communal.

⁵ L'analyse des résidus en modélisation linéaire consiste à évaluer comment les résidus sont distribués en fonction des valeurs prédites de la variable dépendante. Si la dispersion des résidus n'est pas homogène, on parle d'hétéroscédasticité.

Comment UrbanSim prend-il en compte des événements imprévisibles, par exemple le déménagement de grands groupes industriels ?

UrbanSim a prévu ces cas. Il existe un objet appelé « event » 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'était pas difficile à remplir manuellement (par exemple, le regroupement à Guyancourt en 1998 de près de 8000 emplois du constructeur Renault).

Quelle est la différence entre spécifier, estimer et caler un modèle ?

La spécification d'un modèle est le choix d'une forme fonctionnelle appropriée de modèle mathématique. Par exemple un logit standard, un logit emboîté, une régression linéaire, une régression log-linéaire, etc

Une fois le modèle spécifié, on estime les paramètres (les coefficients des variables explicatives) avec un logiciel d'économétrie de type SAS.

Enfin, on cale le modèle en comparant les résultats aux données observées, puis en ajoutant par exemple une variable dummy par département ou en utilisant des matrices de recalage par département à appliquer aux matrices origine-destination par mode pour ajuster le modèle. Cette phase de calage fait plutôt appel à des méthodes empiriques et au savoir-faire.

Comment valider chaque modèle partiel ?

Il y a trois indicateurs : le pseudo R^2 de McFadden pour un logit, le R^2 pour un modèle hédonique et le nombre et pourcentage de prédictions correctes sur la période dite de calage 1990-1999.

En quoi consiste le calage global d'un modèle intégré ?

C'est d'abord une question d'assemblage.

Chacun des modules a été estimé indépendamment les uns des autres. A la limite, il est possible de valoriser du point de vue de la recherche chaque estimation. Mais le but premier n'était pas de publier les résultats de chacun des modèles partiels. Il s'agissait 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 modèles partiels 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 partiel une année peuvent être les entrées d'un autre modèle la même année ou l'année suivante. L'objectif de ce calage global était 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.

Le deuxième aspect du calage global est l'interaction entre les différents modèles au sein d'UrbanSim. 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.

Enfin, il est apparu en fin de projet qu'une compétence manquait à l'équipe : la maîtrise du langage 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'avérait que cette phase de calage global dans la dernière version OPUS/UrbanSim requerrait principalement des compétences en programmation en langage Python.

Formellement, le calage global consiste à résoudre un problème d'optimisation : $\text{minimum}[(\text{observé} - \text{simulé})^2]$. On peut imaginer une procédure automatique de calage qui mettrait en œuvre une boucle de rétroaction avec critère d'arrêt à partir du principe des moindres carrés ou à partir du maximum de vraisemblance. De plus, il y a différentes échelles de calage : cellule, commune, zonage du modèle de trafic, département.

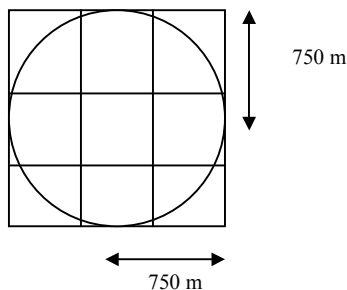
Concrètement, nous calons le modèle intégré au niveau départemental (de même d'ailleurs que le modèle de demande de déplacements). Après analyse des premiers résultats, nous avons décidé de rajouter des constantes pour les indicatrices département pour favoriser ou défavoriser tel ou tel département. 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'indicatrice 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 partiels d'UrbanSim (voir article 1 en annexe).

Avec cette méthode empirique, nous avons pu améliorer les résultats. Il faut savoir qu'une simulation de SIMAURIF sur la période 1990-1999 dure environ 4 heures sur un PC standard et qu'un bon calage nécessite à notre avis une centaine de simulations. Il était encore possible d'améliorer ces résultats mais pour des raisons de contrainte de délais, nous avons décidé de ne pas aller au delà de ce niveau de précision.

Quelle est la pertinence du découpage en 50 000 cellules carrées ? Et si oui, est-il rentable ?

Partant du principe anglo-saxon « Garbage in, garbage out », nous avons fourni dans ce projet 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 le découpage de la région en mailles carrés de 500 mètres de côté. En fait, nous n'avons pas choisi ce découpage. C'est une structure géographique imposée par UrbanSim dans ses versions 2 et 3. 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 et la possibilité de sortir des résultats à l'échelle infracommunale (autour des gares de la Tangentielle Nord par exemple). Le voisinage d'une cellule est constitué des huit cellules voisines. Par exemple, UrbanSim calcule automatiquement pour chaque cellule la surface commerciale totale comprise dans les huit carrés voisins (avec une approximation sur la diagonale des quatre carrés aux coins assimilée à 750 mètres).⁶



Ce calcul automatique par UrbanSim permet de caractériser chaque cellule par une sorte de valeur d'accessibilité locale. Plus cette valeur est élevée, plus elle permet d'atteindre des opportunités dans son voisinage. Mais il faut rester vigilant lorsqu'on introduit ces variables de voisinage de en pas générer des problèmes d'autocorrélation spatiale.

Il y a cependant un problème que l'on ne peut ignorer : le biais d'agrégation. Lorsqu'on intersecte la couche des îlots MOS de population et la couche du maillage en 50 000 cellules pour affecter les populations sur les cellules, il se produit une perte d'information ou un biais ou une distorsion. De plus, la fiabilité des résultats reste 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. Mais on est ensuite confronté au problème d'élaborer des projections de ces nouvelles variables. On en arrive à devoir faire un compromis, pas toujours satisfaisant intellectuellement.

Il faut souligner qu'au niveau cellulaire, seuls des modèles simples et prédictifs peuvent être estimés de manière fiable. Si l'on cherche à estimer des modèles explicatifs, la modélisation au niveau cellulaire ne donne pas des coefficients de régression très acceptables. C'est pour cette raison que la dernière version OPUS/UrbanSim permet de travailler sur des découpages de taille variable autres que le maillage en cellules carrées, par exemple directement sur le découpage communal ou le découpage en îlot MOS, mais en contrepartie on perd les avantages de la notion de voisinage.

⁶ Un calcul simple de surfaces montre que la surface des 9 carrés est supérieure de $4/\pi - 1$ soit 28% par rapport au disque de rayon 750 m. On peut imaginer un maillage en hexagones de 250 m de côté. Topologiquement parlant, un maillage en carrés et un maillage en hexagones sont équivalents, sauf que le second est une meilleure approximation de disques. Un calcul de surfaces montre que l'hexagone central plus les six qui l'entourent ont une surface inférieure de $(1 - \frac{3\sqrt{3}}{2\pi})$ soit 17% seulement par rapport au disque qui circonscrit les 7 hexagones.

En quoi consiste le projet de Tangentielle Nord ?

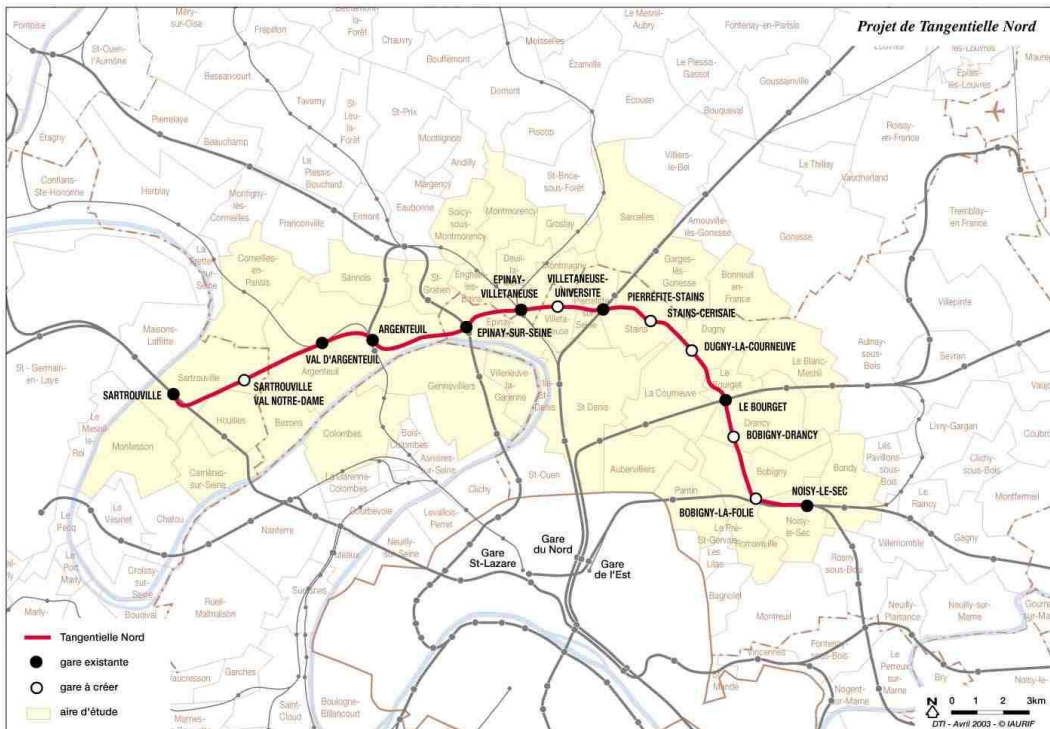


Figure 3 : Plan de situation de la Tangentielle Nord

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). Le lancement officiel de ce projet a eu lieu le 30 juin 2008.

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.

Quelle est l'aire d'étude de la Tangentielle Nord et les données globales du territoire desservi ?

L'aire d'influence élargie de la Tangentielle Nord est constituée de 38 communes, de **853 cellules** de 500 mètres de côté. Les simulations ont été exécutées sur l'ensemble des 50 000 cellules du découpage, mais les résultats ont été analysés sur l'aire d'étude.

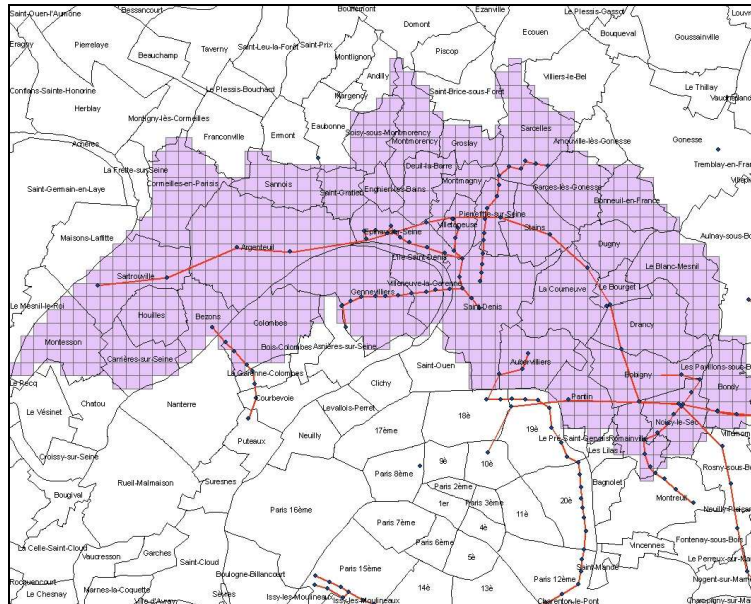


Figure 4 : Les 853 cellules du territoire de la Tangentielle Nord

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

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

En ce qui concerne 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 %.

Le marché immobilier est peu dynamique. Le collectif domine (3/4 des logements offerts). Les prix sont 20 % en-dessous de la moyenne de petite couronne. 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).

Quelle sont les atouts et faiblesses de ce territoire ?

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.

Quels scénarios ont été élaborés puis testés par l'outil SIMAURIF ?

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.⁷

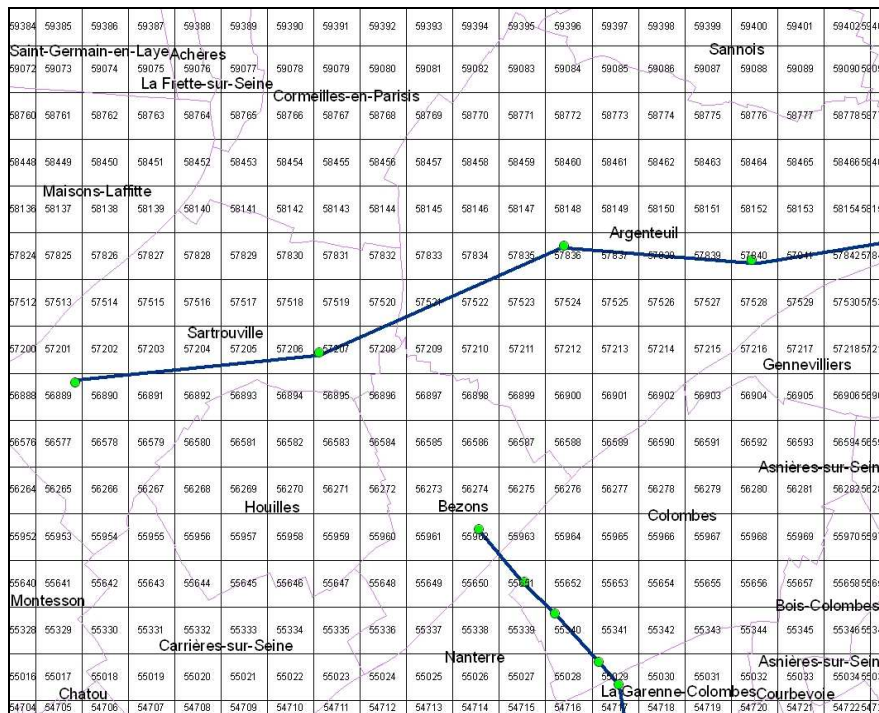


Figure 5 : Numérotation du maillage à l'ouest de la Tangentielle Nord

2) **L'accessibilité « emploiTC » :**

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

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

⁷ 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.

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. Les contraintes environnementales et de politique locale (PLU par exemple) ont déjà été prises en compte dans les bases de données et elles sont communes aux deux scénarios. 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 sait 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 tels que la fréquence et la vitesse ou d'autres scénarios basés sur des politiques d'aménagement ou foncières différentes. On pourrait aussi construire des scénarios basés sur des combinaisons d'offres (par exemple Tangentielle Nord + prolongement du T1 à Montreuil) pour mesurer les effets de réseau.

Quels sont les réseaux routier et de transport en commun sur la période de simulation ?

L'IAU-îdF dispose dans son propre modèle de trafic d'un réseau routier et d'un réseau de transport en commun. Le graphe du réseau routier comporte environ 16700 arcs monodirectionnels et 5330 nœuds pour l'année 1999. Les mouvements interdits sont décrits dans un fichier texte.

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

Nous avons codé les réseaux tous les 3 ans à partir de 1999 jusqu'à 2026 en prenant des hypothèses réalistes de mise en service des infrastructures. Ces réseaux sont décrits dans le détail dans le rapport de 2007.

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

Quelles sont les hypothèses socio-économiques à l'horizon 2026 ?

La période de simulation est 1999-2026. La population totale de la région a été fixée à 12 millions d'habitants à 2026 (hypothèse basse), et la taille moyenne de ménages à 2.1, contre 2.4 en 1999. Le nombre d'emplois a été fixé à 5,58 millions. Rappelons qu'en 1999 (année de départ de la simulation), la population était de 10,95 millions d'habitants, le nombre d'emplois de 5,04 millions (en 2007, la population est estimée à 11,6 millions d'habitants). L'augmentation nette de 10,7% du nombre d'emplois en 27 ans 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).

Combien de temps dure une simulation ?

Les simulations sont effectuées sur l'ensemble de la région mais on focalise les résultats sur le territoire d'étude.

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

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

Des tests de sensibilité ont-ils effectués ?

Il faudrait soumettre le modèle intégré à une batterie de tests de sensibilité. Cela est prévu pour la suite.

Quels sont les résultats de l'application de SIMAURIF au cas de la Tangentielle Nord ?

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).

SIMAURIF 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.

Nous avons cartographié les résultats sur les 853 cellules du territoire pour la population (figure 6) puis les emplois (figure 7).

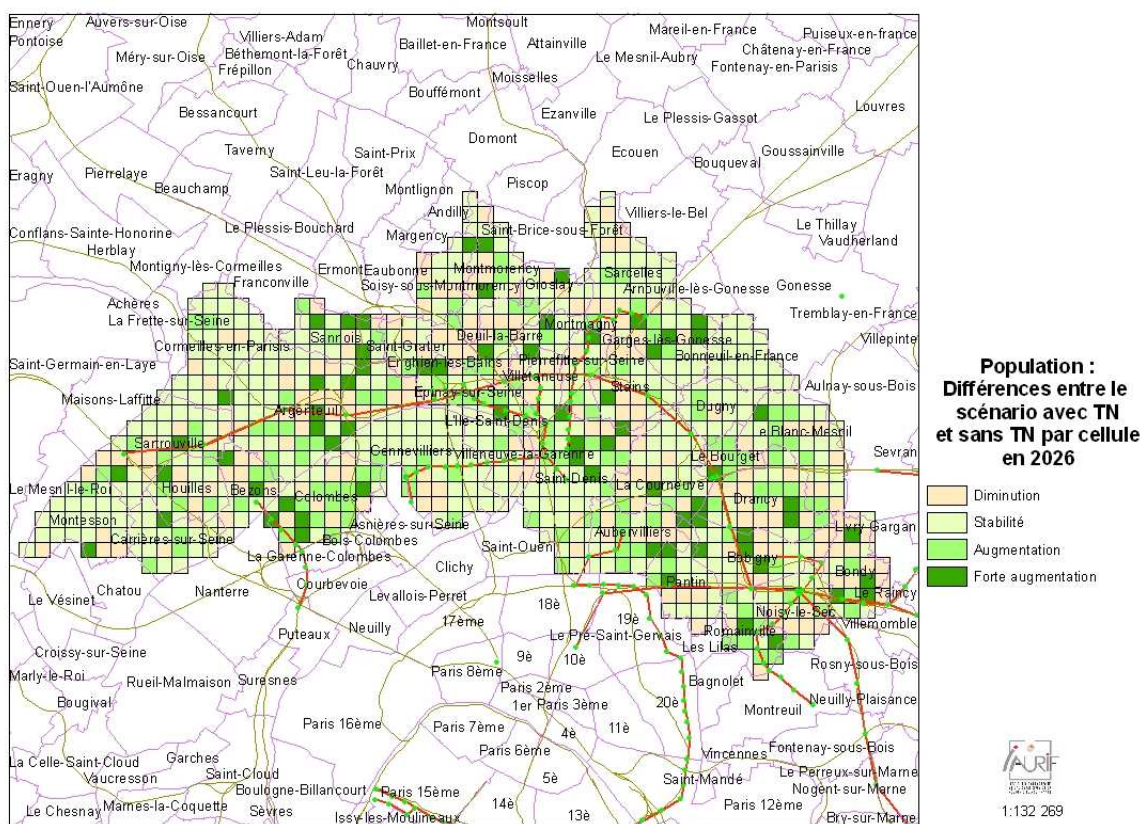


Figure 6

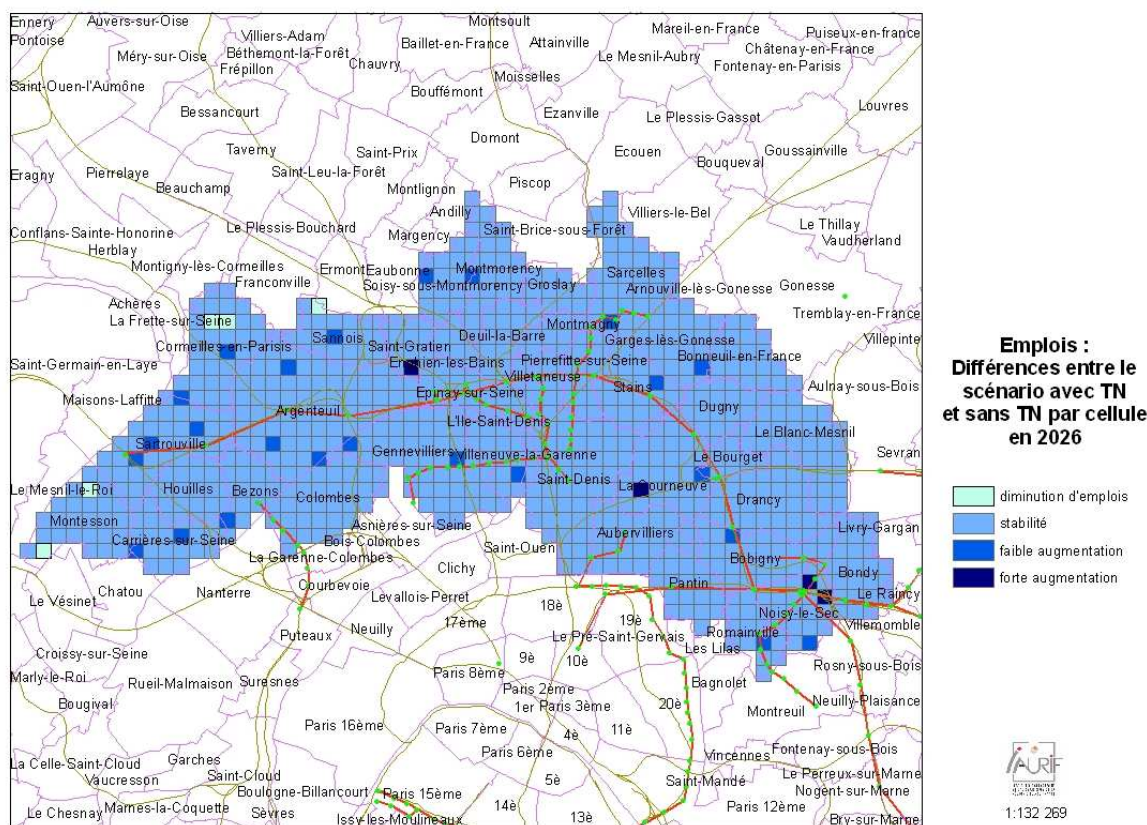


Figure 7

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, *SIMAURIF* 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, *SIMAURIF* 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, voire non significative, par rapport à l'augmentation tendancielle de la population ($2000/260000 < 1\%$). Il est possible que *SIMAURIF* 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 toute façon, si on suppose que 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, d'autres mesures d'accompagnement sont nécessaires afin d'attirer de nouveaux ménages (sous-entendu à revenus élevés). 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 6 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.

Voyons maintenant les emplois. Sur le territoire de la Tangentielle Nord, le nombre d'emplois en 1999 était d'environ 416 000 emplois.

Dans le scénario sans Tangentielle Nord, *SIMAU*RIF 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*RIF 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*RIF 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*RIF 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 figure 7 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 deux extrémités de la ligne, du côté de Sartrouville- Argenteuil et du côté de Noisy-le-Sec. On observe aussi des augmentations localisées d'emplois à proximité des gares de Bobigny, du Bourget et de Stains, mais très peu d'impact sur le tronçon central de la Tangentielle Nord qui dessert le secteur le plus défavorisé du territoire. On aurait pu s'attendre à des augmentations d'emplois plus importantes autour du pôle de la gare du Bourget mais il faut savoir que le prolongement de la ligne 7 du métro n'est pris en compte dans la simulation que dans la dernière période 2023-2026. 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 et de maillage. Ce gain d'accessibilité locale et régionale a contribué très fortement dans le choix de localisation des emplois à proximité de la gare.

Enfin, en ce qui concerne les prix de l'immobilier, *SIMAU*RIF n'a pas été en mesure de produire des résultats fiables au niveau cellulaire mais seulement au niveau communal. Ce volet sur les prix est donc à perfectionner.

***Quel est le niveau de fiabilité du modèle SIMAU*RIF et quelles sont les sources d'incertitude ?**

C'est **la première fois** en France qu'un modèle produit, 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, ce qui ne permettait pas de mesurer des différentiels de population ou d'emplois. Pour la première fois, un modèle permet de mesurer l'attractivité résidentielle et l'attractivité économique qui peuvent donc être de nouveaux critères d'aide à la décision dans une démarche d'analyse multi-critères, au même titre qu'un taux de rentabilité interne.

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. Comme en traitement du signal, le bruit dû à la nature probabiliste des modèles peut être plus grand que le signal à mesurer.

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 tels que la proximité d'écoles 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 seront-ils vérifiables ?

Les résultats présentés (différences entre les deux scénarios de population et d'emplois) 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 pourra seulement comparer en 2026 le nombre de ménages et le nombre d'emplois simulés dans l'aire d'étude avec les nombres réels mais pas les volumes supplémentaires dus à l'attractivité de la Tangentielle nord. Remarquons aussi que nous n'avons pas ici comparé les résultats de prévision de trafic en 2026 sur la Tangentielle Nord avec ceux des études de trafic antérieures de la fin des années 90 qui envisageaient une date de mise en service en 2009.

Quels sont les enseignements de ces résultats sur la Tangentielle Nord ?

On peut 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. Remarquons que le gain d'accessibilité n'est pas si important a priori car sur les 14 gares, 8 existent déjà. 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, les simulations montrent 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 dans des projets urbains pour accompagner l'arrivée de la Tangentielle Nord et pour maximiser ses retombées économiques.

Ces conclusions rejoignent celles des urbanistes qui ont travaillé de manière qualitative sur le diagnostic territorial de la Tangentielle Nord.

Les propositions pour accompagner l'investissement de la Tangentielle Nord ne manquent pas. Maintenant que son lancement a été officiellement annoncé le 30 juin 2008, il est temps de :

- 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
- coordonner à l'échelle des gares les projets urbains et leur articulation avec la Tangentielle Nord
- 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 sous l'égide d'une structure spécifique ayant un rôle de coordination et de suivi

Quels sont les moyens humains nécessaires pour mettre au point un modèle intégré ?

Le projet a nécessité une équipe de trois personnes équivalent plein-temps pendant 3 ans et demi, aux compétences multiples, plus un stagiaire par an à l'IAU île-de-France. On a aussi bénéficié de la disponibilité et de la richesse des données de l'IAU-îdF, qui ne sont pas intégrées dans le budget total du projet. Les compétences minimales et permanentes qu'il faut réunir dans une équipe sont : un spécialiste en modélisation, un économètre, un analyste-programmeur connaissant le langage Python, un géomaticien gestionnaire de bases de données. Mais il est important de souligner que si l'objectif est d'aboutir à un outil opérationnel, c'est le modélisateur qui doit rester maître du projet et canaliser les exigences des autres membres de l'équipe. Sinon le risque est de ne travailler que sur des estimations économétriques indépendantes les unes des autres ou que sur des aspects purement théoriques ou informatiques, et au final de ne pas atteindre l'objectif de mise au point d'un modèle intégré opérationnel. L'approche purement économétrique ne permet pas d'aboutir à un outil opérationnel de modèle intégré car il y a trop de variables explicatives et les variables transport ne sont pas assez discriminantes. Il est donc vital pour la réussite du projet que le praticien ne se laisse pas débordé par les chercheurs puristes versés dans les débats théoriques ou à la recherche d'une solution ultime. Il faut aussi insister sur l'importance de bien connaître l'aire d'étude du point de vue de l'urbaniste, du démographe et de l'économiste avant de se lancer dans la modélisation. Des discussions avec ces spécialistes sont indispensables pour connaître leurs visions et essayer de les « mettre en équations » dans nos modèles.

3. Portage des données et des modèles d'UrbanSim 2 à OPUS/UrbanSim 4

Ce portage avait pour but de perfectionner le modèle intégré *SIMAU*RIF. La description est quelque peu technique car elle utilise un vocabulaire propre aux informaticiens.

3.1. 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 OPUS/UrbanSim 4 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, OPUS/Urbansim 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 localisation de projet urbain, tel qu'il peut être vu par un aménageur.
- de nouveaux outils d'estimation économétrique de modèles, les procédures étant quasi-automatiques.
- la mise à disposition du modèle logit emboîté en plus du logit multinomial
- un nouveau langage de développement plus ouvert, Python, à la place de Java. Le code source d'Urbansim 4 est accessible et modifiable, contrairement à celui d'Urbansim 2 en Java.
- 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) qui comporte deux autres ensembles de modèles : un modèle de simulation de la demande basée sur les activités et un modèle d'affectation. 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.

3.2. 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 : fallait-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 fallait estimer un nouveau modèle de développement urbain, alors que nous avons passé plusieurs mois à construire la base de données de l'évolution du Mode d'Occupation du Sol (MOS) et à estimer un modèle de transition.
- Il nous fallait faire l'apprentissage du nouveau 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.

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

- il nous a clairement affirmé que la version précédente en Java ne serait plus maintenue, ni même téléchargeable sur le site Urbansim.org.
- 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 d'estimation. En restant dans Urbansim 4, les estimations de modèle pouvaient être effectuées en un jour, toujours selon le Pr Waddell.
- on y gagnerait en temps de calculs sur un ordinateur standard.

3.3. Problèmes rencontrés lors du portage

Nous avons rencontré des problèmes importants de tous ordres, et pas seulement d'ordre informatique. Il semble que depuis début 2007, selon les messages reçus sur la liste de diffusion, presque tous les utilisateurs ont subi les mêmes déboires que nous, à cause d'une version en bêta-test et d'une documentation toujours en retard par rapport aux nouveaux développements.

3.3.1. Problèmes informatiques

L'installation d'OPUS/Urbansim (version Windows⁸) et de l'exemple Eugene demandent déjà une persévérance à toute épreuve. Disons qu'une fois cette épreuve initiatique surmontée, qui pouvait

⁸ Car il existe aussi la version sous Linux. Il semble qu'elle présentait 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 (un collègue de Lyon) 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

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 débogage. 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 voire abstrus l'examen détaillé du code. Lorsqu'un processus rencontre une erreur fatale, 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 à Paris au cours du mois de mai 2007, qui nous a grandement aidés dans le débogage de nos fichiers.

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 !

Le **troisième grand problème** a été l'apprentissage de Python. C'est un langage de plus en plus populaire 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 était en plein développement depuis juillet 2006 à l'Université de Washington a subi des mises à jour très fréquentes (et cela continue en 2008). 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 2007, 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

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. »

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.

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

3.3.3. Problèmes d'estimation des modèles

Nous avons essayé d'utiliser l'outil d'estimation automatique des logits pour le modèle de localisation des emplois. 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 est très incomplète à ce sujet. Finalement, nous avons procédé comme nous l'avions fait pour le modèle de localisation des ménages, c'est-à-dire en utilisant la proc MDC du logiciel SAS. Cela nous a pris plusieurs semaines de travail.

Quant au modèle de développement du type urbain, comme il l'a été dit, il a complètement changé entre la version 2 et la version 4. Si l'on devait estimer un nouveau modèle, cela aurait pris sans doute plusieurs mois de travail. Le plus difficile n'est pas tant la phase d'estimation économétrique que la construction de la base de données de calage. Il aurait fallu recenser un maximum de « projets de développement » (les agents), en faire une typologie, les localiser dans le maillage, ce qui nous aurait demandé plusieurs semaines de travail. Le nouveau modèle de localisation de projet 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 choix de localisation classique.

En août 2007, nous avons décidé de ne pas utiliser ce nouveau modèle. D'une manière sans doute très simplifiée, nous étions condamnés à adapter le modèle de transition du mode d'occupation du sol (MOS) de la version 2 d'Urbansim déjà estimé en 2005. 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 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) étaient d'un type particulier et n'étaient pas prises en compte par ce modèle.

3.4. Résultats du portage

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 bibliothèques, 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.
- Un dernier répertoire est utilisé, /Urbansim_output, contenant les données de sortie de la simulation 1990-1999.

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=750 m, 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.

3.5. Conclusion du chapitre

Au 1^{er} semestre 2007, les problèmes informatiques nous avaient quelque peu éloignés pour un temps de la modélisation de l'interaction transport-urbanisation. Heureusement, la plupart des problèmes ont été résolus vers le milieu de l'année 2007. Nous avons réussi, non sans difficultés, le portage de nos données et de nos modèles partiels d'UrbanSim 2 à OPUS/UrbanSim 4. Cela nous a permis de consacrer la fin de l'année 2007 au calage global et à l'application au cas de la Tangentielle nord.

4. Estimation d'un modèle de prix de l'immobilier de bureaux

4.1. Rappel sur le modèle de prix du foncier d'UrbanSim

Le modèle de prix dans UrbanSim est un modèle de prix du foncier (du terrain). Il faut savoir qu'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 monétaire (en terme de capital) pour chacun de ces postes :

- residential_improvement_value
- commercial_improvement_value
- industrial_improvement_value
- government_improvement_value

Le principe d'UrbanSim est de calculer le prix moyen du foncier de chaque cellule chaque année, puis de calculer en fonction de ce prix et des surfaces bâties la valeur totale de résidences, de locaux commerciaux, de locaux industriels et de bureaux. UrbanSim suppose que le prix du terrain est le même pour les résidences et toutes les activités (commerces, industries, bureaux).

Ces quatre variables de type « improvement », plus les variables de voisinage, vont alimenter ensuite les deux modèles de localisation et le modèle d'évolution du MOS si elles interviennent comme variables explicatives.

Il s'avère que pour le 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 ventes de terrain, du moins insuffisamment pour estimer un modèle fiable. Théoriquement, il nous faut modéliser quatre variables de prix : le prix de l'immobilier résidentiel, le prix de surface commerciale, le prix de surface d'activités industrielles et le prix de l'immobilier de bureau. Dans le prototype d'outil *SIMAU*RIF terminé en décembre 2007, nous avons simplifié le problème en regroupant les activités de commerce, d'industrie et de bureau en une seule activité. Puis nous avons estimé un modèle de prix de l'immobilier résidentiel en utilisant la base BIEN des Notaires (Base d'Informations Economiques Notariales). Celle-ci a l'immense avantage d'être géolocalisée. Ce modèle est décrit dans le détail dans le rapport 2007⁹.

Nous proposons dans la présente étude de perfectionnement du modèle intégré *SIMAU*RIF d'estimer un des trois autres modèles de prix : le prix de l'immobilier de bureau. Ce modèle ne sera pas implémenté dans la version actuelle de *SIMAU*RIF car l'architecture d'OPUS/UrbanSim ne le permet

⁹ En réalité, nous avons mené deux recherches en parallèle pour modéliser le prix d'un logement en distinguant les appartements des maisons : 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).

pas de manière aisée sauf à développer de véritables programmes en script Python. Cette partie est donc davantage de la recherche académique.

4.2. Géographie succincte de l'immobilier d'entreprise

L'Ile-de-France dispose d'une offre en immobilier d'entreprise diversifiée. Schématiquement, il y a 3 types d'immobilier d'entreprise :

- dominante tertiaire
- dominante mixte
- dominante activités et logistique

Sur le marché des bureaux, il faut distinguer les zones à très forte densité de bureaux et le marché des bureaux « élargi », moins dense et pouvant accueillir d'autres catégories d'immobilier de type locaux mixtes.

Le marché des bureaux est traditionnellement concentré sur l'ouest parisien, la partie médiane et le sud des Hauts-de-Seine s'est diffusé à l'est de Paris (Seine RG – Tolbiac), en petite couronne (Saint-Denis, abords du périphérique) et dans quelques zones d'activité économique qui se sont tertiarisées (à Vélizy, dans la zone de la SILIC à Rungis, à Paris nord II au sud de Roissy et dans la zone de Courtaboeuf). Des ensemble de bureaux sont également localisés dans les quartiers des Villes nouvelles.

La localisation des projets montre la poursuite du recentrage de la construction neuve en petite couronne, amorcée au milieu des années 90 (nord des Hauts-de-Seine, Saint-Denis, Montreuil, nord de la Seine-Amont). Pour l'avenir, le rééquilibrage centre-périphérie ne semble pas se dessiner et seuls quelques sites de grande couronne maintiennent leur position (Saint-Quentin-en-Yvelines, Massy, secteur de Val d'Europe). Un tiers de la production de bureaux en blanc concerne des réhabilitations dans les secteurs les plus valorisés du parc (à Paris-ouest et la Défense)

Pour les locaux mixtes ou locaux d'activités en blanc, ce marché est principalement localisé le long d'A86 de Nanterre à Colombes, sur la Plaine Saint-Denis, Vélizy, Rungis, Créteil et sur des pôles plus périphériques comme certains parcs d'activité des Villes nouvelles à Cergy-Pontoise, à Saint-Quentin-en-Yvelines, Evry, Marne-la-Vallée (secteurs 1 et 2), dans les zones d'activités de Courtaboeuf (les Ullis, Villebon, Villejust) et de Paris Nord II.

Les transactions récentes concernent le parc de la SILIC et le parc des Nations à Roissy, les communes de Clichy et Asnières.

Le développement spatial des ZAE montre un déséquilibre entre l'est et l'ouest. A l'ouest, davantage concerné par le marché des bureaux, les ZAE sont principalement localisées dans le nord des Hauts-de-Seine , la vallée de la Seine, les Villes nouvelles de Saint-Quentin-en-Yvelines et de Cergy-Pontoise.

Ailleurs en Ile-de-France, 3 axes se dessinent :

- axe nord entre Cergy-Pontoise et Roissy, englobant Argenteuil, Gennevilliers, Saint-Denis
- axe sud entre Saint-Quentin et Sénart en passant par Saclay et Massy
- axe est de Massy à Marne-la-Vallée qui couvre Orly, Créteil, Bonneuil , Ivry et Vitry.

On observe aussi le développement d'une offre de bureaux ou locaux mixtes dans les zones d'activités les plus dynamiques de la région. Les nouveaux pôles tertiaires de l'est sont circonscrits à quelques sites en banlieue et à l'axe de Fontenay – Val d'Europe. Concernant les zones d'activité, malgré un stock très important, l'offre semble mal structurée. La question du renouvellement d'une offre de terrains de qualité, bien desservis, se pose dans un contexte d'affaiblissement à terme des capacités d'accueil des Villes nouvelles.

Les commercialisations sont très liées aux cycles économiques. Elles ont connu des fluctuations importantes au fil des années. Depuis une dizaine d'années, les $\frac{3}{4}$ des transactions sont concentrées sur une dizaine de secteurs : les Villes nouvelles, Roissy, Marne-la-Vallée, la Francilienne sud (jusqu'à Evry et Sénart), Massy-Orly-Créteil, Saint-Quentin-en-Yvelines et Cergy-Pontoise.

L'existence de ces fluctuations est confirmée par les études annuelles du Crédit Foncier. La dernière en date (mars 2008) lance un véritable appel à la vigilance aux professionnels de l'immobilier. Elle livre une vision très pessimiste du marché de la location de bureaux en Ile-de-France.

L'année 2007 aura enregistré un haut niveau de transactions avec 2,7 millions de mètres carrés loués par des entreprises en Ile-de-France (contre 2,9 en 2006). Mais la crise s'est fait sentir au dernier trimestre. La location de bureau a accusé un repli de 17 % par rapport au dernier trimestre de 2006. La décrue s'est amplifiée début 2008 avec une chute de 70 % des transactions au premier trimestre en regard du premier trimestre 2007. Le taux de vacance réel est de 8 % à 10 % au lieu des 5 % habituellement annoncés. Cette faiblesse est due à l'atonie de la création d'emplois de bureau.

4.3. Les données utilisées

Nous avons trouvé une base de données de prix du m² de bureau : la Cote Callon (2001). Cette base recense des prix au m² pour les bureaux anciens pour deux catégories : bureaux rénovés et non rénovés. Son grand inconvénient est de ne porter que sur environ 300 communes (plus de 5000 habitants) sur les 1300 communes d'Ile-de-France. Cette limitation nous impose un nombre limité de variables explicatives.

4.4. Spécification du modèle

Cette partie a été réalisée en collaboration avec Kiarash Motamedi, Doctorant au laboratoire Thema de l'Université de Cergy-Pontoise.

4.4.1. Modèle de Rosen

Rosen (1974) présente la première interprétation économique de tels modèles. Le prix hédonique est ainsi présenté comme la contribution des caractéristiques dans la constitution du prix d'équilibre d'un bien résultant de l'action de forces concurrentielles sur le marché. Ces hypothèses ont fait l'objet de critiques mais les travaux de Rosen restent toujours la référence en la matière.

Rosen démontre que sous ces hypothèses, la courbe de la fonction de prix hédonique est tangente aux courbes de :

- l'inverse de la fonction de demande compensée (représentant le prix maximum que le consommateur est disposé à payer pour une unité de plus d'une caractéristique),
- l'inverse de la fonction d'offre compensée (représentant le prix minimum pour lequel le producteur est prêt à vendre une unité de plus d'une caractéristique).

L'équilibre sur le marché de bien différencié est atteint en égalisant l'offre et la demande. Il démontre les difficultés de calcul de cet équilibre dans un horizon temporel de court terme. A long terme, en considérant la possibilité d'introduction de réorganisation des firmes sur le marché, il montre que la fonction de prix hédonique est complètement déterminée par l'offre.

Ensuite, Rosen propose une procédure à deux étapes pour obtenir économétriquement les fonctions de demande et d'offre des biens différenciés. La première étape consiste à estimer une fonction hédonique de prix dont la dérivée par rapport à chaque caractéristique donne un point des courbes de l'inverse des fonctions de demande et d'offre compensées de cette caractéristique.

Dans la deuxième étape, il faut résoudre un système d'équations à deux équations pour chaque caractéristique,

$$\begin{cases} p_i(X) = D^i(X, Y_1) & \text{demande} \\ p_i(X) = O^i(X, Y_2) & \text{offre} \end{cases}$$

Y_1 et Y_2 représentent les variables exogènes décrivant le revenu et les préférences du consommateur pour la première et le profit et les paramètres de la fonction de production pour la deuxième. Si les producteurs ou les consommateurs sont homogènes, la fonction hédonique représentera la fonction d'offre ou celle de demande. Si ces deux ensembles ne sont pas homogènes (ce qui est normalement le cas), la fonction des prix hédoniques estimée à la première étape doit être nécessairement non-linéaire. Pour une application de modèle de Rosen au marché de logement, nous renvoyons le lecteur au papier de Witte, A. D., H. J. Sumka and H. Erekson (1979).

Brown, J. N. and H. S. Rosen (1982) démontrent l'importance de la forme fonctionnelle du modèle hédonique dans l'efficacité de la procédure à deux étapes de Rosen. La question d'identification du modèle hédonique étant soulevée, nous pouvons nous référer aux travaux suivants à ce sujet : Tinbergen (1956), Epple (1987), Ekeland, I., J. J. Heckman and L. Nesheim (2004).

Nous estimons ici un modèle qui donne le logarithme du prix du m^2 de bureau dans la commune, $Ln(P_c)$, en fonction des caractéristiques de la commune (vecteur X_c , de composantes X_c^k , $k = 1, \dots, K$) :

$$Ln(P_c) = Cte + \beta X_c + \varepsilon_c = Cte + \sum_{k=1}^K \beta^k X_c^k + \varepsilon_c. \quad (1)$$

Dans cette équation, chaque coefficient β^k du vecteur β représente l'effet marginal de la caractéristique communale X_c^k sur le logarithme du prix. Dans une telle spécification log-linéaire, les effets des différentes caractéristiques sur le logarithme du prix sont supposés s'additionner, ce qui revient à supposer que les différentes caractéristiques communales ont un effet multiplicateur sur les prix. L'équation (1) est en effet équivalente à :

$$P_c = \exp(Cte + \beta X_c + \varepsilon_c) = \exp(Cte) \prod_{k=1}^K \exp(\beta^k X_c^k) \exp(\varepsilon_c). \quad (2)$$

L'effet d'une variable indicatrice¹⁰ X_c^k est alors de multiplier le prix¹¹ par un facteur $\exp(\beta^k)$.

4.4.2. Quelques particularités sur le modèle spécifié

Concernant la forme fonctionnelle du modèle de prix hédonique, les formes logarithmiques et semi-logarithmiques offrent une facilité d'estimation. De plus, Malpezzi, S. (2003) en dénombre d'autres

¹⁰ $X_c^k = 1$ lorsque la condition correspondante est vérifiée dans la commune c , $X_c^k = 0$ dans le cas contraire.

¹¹ Le prix au m^2 ou, ce qui revient au même dans un modèle log-linéaire, le prix total à surface fixée.

avantages. Premièrement, l'impact marginal d'une caractéristique sur la valeur du bien semble être plutôt proportionnel à la taille et à la qualité du bien. Deuxièmement, Les coefficients estimés dans une telle spécification sont plus faciles à interpréter. Dans le cas d'une spécification semi-logarithmique,

$$\ln p = \alpha + \sum_i \beta_i x_i + \varepsilon ,$$

le coefficient β_i peut être interprété approximativement comme la variation en pourcentage de la valeur du bien pour une unité supplémentaire de la variable x_i , toute chose égale par ailleurs. Nous rappelons que dans le cas des variables muettes, cette variation est plus exactement $\exp(\beta_i)-1$, voir Halvorsen, R. and R. Palmquist (1980).

Dans le cas d'une spécification logarithmique,

$$\ln p = \alpha + \sum_i \beta_i * \ln(x_i) + \varepsilon ,$$

le coefficient β_i représente la valeur de l'élasticité ou la variation en pourcentage de la valeur du bien due à 1% d'augmentation de la caractéristique x_i .

D'autres formes fonctionnelles plus élaborées dont le degré de non-linéarité est estimé de manière endogène sont appliquées depuis les années 80 (voir Linneman, P. (1980) comme une première application et Cavailhès, J. (2005) pour une application récente). Ces méthodes se basent sur la transformation Box-Cox. En plus la fonction de prix peut ne pas être monotone par rapport à certaines variables telles que la distance aux infrastructures de transports. Dans ces derniers cas, nous sommes amenés à estimer la fonction sur plusieurs intervalles.

Ici, nous disposons d'une base de données assez riche en terme de données locales, en revanche assez pauvre en terme de caractéristiques intrinsèques des biens immobiliers étudiés. C'est dû au fait que l'échelle d'étude se limite à la commune et par conséquent, le nombre d'observations est de 300 avec les données de la Cote d'Azur. Le nombre limité des observations est une contrainte importante qui nous empêche d'étudier en détail les déterminants du prix de l'immobilier de bureaux. Nous nous contentons donc de modèles simples du prix de l'immobilier de bureau. En revanche, le fait de travailler dans le cadre d'un modèle intégré qui comprend également des modèles de localisation des ménages, des établissements et des projets de construction nous offre l'opportunité de travailler en détail sur les déterminants de la demande d'immobilier et le consentement à payer des établissements pour les différentes caractéristiques locales et des aménités urbaines.

4.5. Résultats du modèle

Le tableau 1 présente les déterminants des prix des bureaux au niveau communal. Le fort pouvoir explicatif des régressions permet d'être confiants dans la pertinence des prix de bureaux imputés pour les communes n'apparaissant pas dans la Cote d'Azur. Un R^2 de l'ordre de 80% signifie une corrélation de l'ordre de 90% (racine carrée de 0.8) entre la variable observée et la variable prédite, ce qui est excellent.

La spécificité de Paris semble bien expliquée par les caractéristiques observables, puisque l'indicatrice de Paris n'est presque jamais significative. La seule exception est le prix minimal des bureaux à Paris. Toutes choses égales par ailleurs, le prestige de Paris intra muros et la rareté des terrains justifient que les prix des bureaux neufs ne descendent jamais aussi bas à Paris que dans des communes semblables. Ainsi, le prix minimal d'un bureau neuf à Paris est supérieur de 31,5% à ce

qu'il serait dans une commune semblable mais située en dehors de Paris.¹² En revanche, le fait que les bureaux anciens ou les plus prestigieux bureaux neufs soient si chers à Paris semble tout à fait s'expliquer par la meilleure accessibilité ou les caractéristiques de la population locale.

De façon similaire, les villes nouvelles semblent souffrir d'une image négative qui ne subsiste toutes choses égales par ailleurs que pour l'ancien non rénové, c'est-à-dire pour les bureaux dans le plus mauvais état. Le prix maximal dans une ville nouvelle pour les bureaux dans l'ancien non rénové est ainsi inférieur de 11% ($\exp(-0,11975) = 0.8871 = 1 - 0,1129$) à ce qu'il serait dans une commune similaire, mais qui ne serait pas dans une ville nouvelle.

En résumé, l'image de marque de Paris et des villes nouvelles ne semble avoir un impact excessif sur les prix des bureaux (au-delà de ce qui peut s'expliquer par l'accessibilité ou les caractéristiques de la population) que dans les cas extrêmes, à savoir les bureaux les plus prisés à Paris et les moins prisés dans les villes nouvelles.

A l'inverse, le surcoût des bureaux à la Défense ne peut pas s'expliquer par l'accessibilité ou la composition de la population. L'image de la Défense est valorisée pour elle-même, dans les prix de tous les types de bureaux. Elle se traduit par une augmentation de tous les prix de plus de 50% par rapport à une commune similaire qui ne serait pas située à la Défense ($\exp(0,415)=1,514$).

Nous nous gardons bien d'interpréter en termes de causalité le coefficient positif et très significatif de la fraction des emplois dans le secteur 6 dans toutes les régressions de prix des bureaux. Selon ce coefficient, une augmentation de 10% de la part de l'emploi dans le secteur des services est associée à une augmentation des prix d'environ 5% ($0,5 \cdot 0,1$). Toutefois, un tel coefficient peut traduire deux effets très différents. D'une part, la spécialisation dans le secteur tertiaire peut renchérir les prix car elle améliore le prestige de la commune. Mais d'autre part, il y a beaucoup d'autres sources de prestige pour une commune, et ce prestige peut à la fois augmenter directement les prix (à composition de l'emploi fixée) et attirer les emplois tertiaires (à prix fixé). Nous pouvons donc affirmer qu'il existe une forte corrélation entre la composition de l'emploi et les prix, ce qui est très utile pour imputer un prix par m² dans les communes où le prix des bureaux n'est pas observé directement. Mais nos régressions ne doivent pas être utilisées pour prédire l'impact qu'une modification exogène de la structure de l'emploi pourrait avoir sur les prix futurs. Des réserves similaires s'appliquent à l'interprétation des coefficients suivants.

Dans l'ancien non rénové, les prix semblent dépendre plus de la densité de population que de la densité d'emploi. Ceci suggère que les prix dans l'ancien rénové sont plus sensibles au prestige local (qui attire la population) qu'au niveau d'emploi. L'ampleur de l'effet des densités de population et d'emploi sur les prix des bureaux anciens rénovés ne diffèrent pas significativement. A l'inverse, les prix des bureaux neufs sont plus sensibles à la densité d'emplois qu'à la densité de population. Sur la base de cet ensemble de résultats, on peut penser qu'une forte densité d'emplois correspond à une augmentation de la demande qui tire vers le haut les prix, ce qui touche surtout les bureaux neufs, alors qu'une forte densité de population constitue plus un indicateur de prestige de la commune, qui affecte plutôt les prix des bureaux dans l'ancien que dans le neuf.

Les prix des bureaux diminuent tous significativement lorsque la fraction de pauvres augmente dans la commune. Cet effet peut traduire aussi bien une causalité directe (la proximité de pauvres réduit la valeur des bureaux) qu'une simple corrélation, dans le sens où les mêmes caractéristiques inobservables peuvent à la fois attirer les pauvres (ou, de façon équivalente, éloigner les riches) et dévaloriser les bureaux.

¹² $\exp(0,27396)=1,315$.

Les prix sont beaucoup élevés dans les communes où les bâtiments sont plus anciens (un peu moins de 1% d'augmentation de l'ensemble des prix pour un an de plus en moyenne). Cet effet est en accord avec la constatation selon laquelle les quartiers les plus anciens (par exemple les quartiers du centre de Paris) sont aussi les plus prestigieux, donc les plus chers.

Les prix sont plus élevés lorsque la commune comporte une plus grande fraction de sa surface en parcs et jardins. Cet effet peut traduire à la fois l'attrait des parcs et jardins et un effet d'éviction : les surfaces accueillant des espaces verts ne peuvent pas accueillir de bureaux, ce qui réduit l'offre et fait monter les prix. Une augmentation de 10% de la surface en parcs et jardins se traduirait par une augmentation de l'ordre de 6% des prix.

Chaque arrêt de métro supplémentaire se traduit par une augmentation de l'ensemble des prix de bureaux de l'ordre de 1,8%. Des régressions préliminaires ont montré que, toutes choses égales par ailleurs, le nombre de gares (trains de banlieue et grandes lignes) n'avait pas d'effet significatif sur les prix de bureaux.

Etat du bureau		Bureau ancien non renové		Bureau ancien renové		Bureau neuf		
		Min	Max	Min	Max	Min	Max	Moyenne
R²		79,25%	75,40%	82,91%	80,55%	78,23%	79,69%	80,21%
Constante	Coeff	6,04950	6,53047	6,83908	7,22997	7,09495	7,52461	7,30426
	t_Stat	55,44	53,29	78,30	73,51	67,67	75,09	74,09
La commune est-elle Paris	Coeff	0,21812	0,09586	0,16303	-0,03085	0,27396	-0,01600	0,07826
	t_Stat	1,69	0,66	1,58	-0,27	2,21	-0,14	0,67
La commune est-elle La Défense	Coeff	0,46015	0,47732	0,41505	0,47721	0,47092	0,51653	0,49203
	t_Stat	3,95	3,65	4,45	4,54	4,20	4,83	4,67
La commune est-elle dans une Ville nouvelle	Coeff	-0,08002	-0,11975	-0,03291	-0,02673	-0,03084	-0,03780	-0,06804
	t_Stat	-1,68	-2,24	-0,86	-0,62	-0,67	-0,87	-1,58
Fraction des emplois en secteur 6 (services)	Coeff	0,52281	0,69135	0,37613	0,54967	0,43965	0,59050	0,57692
	t_Stat	3,61	4,25	3,24	4,21	3,16	4,44	4,41
Densité d'emplois (commune, 1000/km2)	Coeff	0,00715	0,00759	0,00797	0,01029	0,00809	0,01063	0,00942
	t_Stat	2,93	2,77	4,08	4,67	3,44	4,74	4,27
Densité de population (commune, 1000/km2)	Coeff	0,01186	0,01314	0,00758	0,01057	0,00582	0,00999	0,00880
	t_Stat	3,94	3,89	3,15	3,90	2,01	3,62	3,24
Fraction des ménages avec revenu bas	Coeff	-0,62989	-0,58851	-0,69646	-0,78141	-0,64585	-0,78406	-0,70846
	t_Stat	-3,25	-2,70	-4,48	-4,47	-3,46	-4,40	-4,04
Age moyen des logements	Coeff	0,00756	0,00839	0,00842	0,00906	0,00834	0,00931	0,00924
	t_Stat	4,63	4,58	6,44	6,15	5,31	6,20	6,26
% de surface couverte par des parcs et jardins	Coeff	0,64061	0,80492	0,47044	0,50273	0,56844	0,54116	0,57615
	t_Stat	2,54	2,84	2,33	2,21	2,34	2,33	2,52
Nombre d'arrêts de métro	Coeff	0,01824	0,01899	0,01736	0,01951	0,01523	0,01879	0,01726
	t_Stat	3,04	2,82	3,61	3,61	2,64	3,41	3,18
Temps de trajet en TC (IAU, heure)	Coeff	-0,03301	-0,03105	-0,03862	-0,03588	-0,03832	-0,03471	-0,02568
	t_Stat	-1,47	-1,23	-2,14	-1,77	-1,77	-1,68	-1,26
Temps de trajet en VP (IAU, heure)	Coeff	-0,39571	-0,28698	-0,36779	-0,36613	-0,33050	-0,36975	-0,37749
	t_Stat	-5,09	-3,29	-5,92	-5,23	-4,43	-5,18	-5,38

Tableau 1 : Estimation des prix au m² des bureaux, selon leur état

Les estimations montrent que, toutes choses égales par ailleurs, les prix des bureaux sont beaucoup plus sensibles aux temps de trajet par véhicule particulier que par les transports en commun. Les coefficients de ces derniers (TC) sont en effet à peine significatifs et environ 10 fois moins élevés que ceux des premiers (VP). Ils montrent qu'une augmentation d'une heure des temps de trajet vers la commune aux heures de pointe se traduirait par une diminution des prix de l'ordre de 30 à 40% dans le cas des VP, contre 3% dans le cas des TC.

4.6. Conclusion du chapitre

Nous avons présenté des modèles hédoniques pour le prix des bureaux au niveau communal. Ces modèles se basent sur les données communales et pas sur les données individuelles au niveau des biens. Cette limitation du pouvoir explicatif de notre modèle est dû à la faiblesse du nombre d'observations et par le manque de variables représentant les caractéristiques intrinsèques du bien immobilier étudié. Dans le modèle de prix des bureaux, nous n'avons pas une représentation explicite de l'offre et de la demande mais nous avons réussi à inclure la variable de densité d'emplois. Cette variable avec un effet positif sur le prix peut partiellement compenser cette lacune. Malgré ces difficultés, nous estimons que les résultats obtenus représentent une approximation suffisamment précise. Nous envisageons d'implémenter à l'avenir dans le modèle intégré ce modèle de prix des bureaux, si possible en collaboration avec l'équipe de développement d'UrbanSim aux Etats-Unis.

4.7. Bibliographie

- BAYER, P., F. FERREIRA, and R. McMILLAN (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," National Bureau of Economic Research, Inc.
- BROWN, J. N., and H. S. ROSEN (1982): "On the Estimation of Structural Hedonic Price Models," *Econometrica*, 50, 765-768.
- CALLON, E. (1998): *Valeurs Vénales Immobilières*. Sugères, Fr.: Editions Callon.
- (2001): *Valeurs Vénales Immobilières*. Sugères, Fr.: Editions Callon.
- CAVAILHES, J. (2005): "Le Prix Des Attributs Du Logement," *ÉCONOMIE ET STATISTIQUE*, 23.
- COURT, A. T. (1939): "Hedonic Price Indexes with Automotive Examples," in *The Dynamicso F Automobiled Emand,S Tatement of the Problem*. New York: General Motors Corporation, 98-119.
- EKELAND, I., J. J. HECKMAN, and L. NESHEIM (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112, S60-S109.
- GOODMAN, A. C. (1998): "Andrew Court and the Invention of Hedonic Price Analysis," *Journal of Urban Economics*, 44, 291-298.
- HALVORSEN, R., and R. PALMQUIST (1980): "The Interpretation of Dummy Variables in Semilogarithmic Equations," *The American Economic Review*, 70, 474-475.
- LANCASTER, K. J. (1966): "A New Approach to Consumer Theory," *The Journal of Political Economy*, 74, 132-157.
- LINNEMAN, P. (1980): "Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing Market," *Journal of Urban Economics*, 8, 47-68.
- MALPEZZI, S. (2003): "Hedonic Pricing Models: A Selective and Applied Review," in *Housing Economics and Public Policy: Essays in Honor of Duncan MacLennan.*, ed. by A. O'SULLIVAN, and K. Gibb. Oxford, UK.: Blackwell.
- MICHELANGELI, A. (2005): "Modélisation Hédonique Pour L'évaluation Des Politiques Publiques," Cergy: Université de Cergy-Pontoise, 227.

- MUSGRAVE, J. C. (1969): "The Measurement of Price Changes in Construction," *Journal of the American Statistical Association*, 64, 771-786.
- MUTH, R. F. (1966): "Household Production and Consumer Demand Functions," *Econometrica*, 34, 699-708.
- SHEPPARD, S., C. PAUL, and S. M. EDWIN (1999): "Chapter 41 Hedonic Analysis of Housing Markets," in *Handbook of Regional and Urban Economics*: Elsevier, 1595-1635.
- WAUGH, F. (1929): *Quality as a Determinant of Vegetable Prices: A Statistical Study of Quality Factors Influencing Vegetable Prices in the Boston Wholesale Market*. New York: Columbia University Press.
- WAUGH, F. V. (1928): "Quality Factors Influencing Vegetable Prices," *Journal of Farm Economics*, 10, 185-196.
- WITTE, A. D., H. J. SUMKA, and H. EREKSON (1979): "An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets," *Econometrica*, 47, 1151-1173.

5. Développement de l'interface entre Urbansim et Davisum-METROPOLIS

Au cours des deux premières années du projet, une première maquette informatique a été développée sur la plateforme d'atelier de génie logiciel WinDev de PC Soft par la société adpC. Le portage de UrbanSim 2 à la plateforme OPUS/UrbanSim en 2007 a eu pour conséquence l'abandon de cette interface et le développement d'une interface simplifiée mais opérationnelle.

5.1. Interface entre UrbanSim 2 et METROPOLIS

Cette interface graphique conviviale entre la première version d'UrbanSim et le logiciel METROPOLIS a été développée au cours de la première année du projet avec le logiciel WinDev. Elle est présentée ci-après.

5.1.1. Gestion des paramètres

L'écran principal de l'interface permet de visualiser et de modifier les paramètres de base du projet (onglets « Project », « Scenario », « Base Year », « Travel Model »)

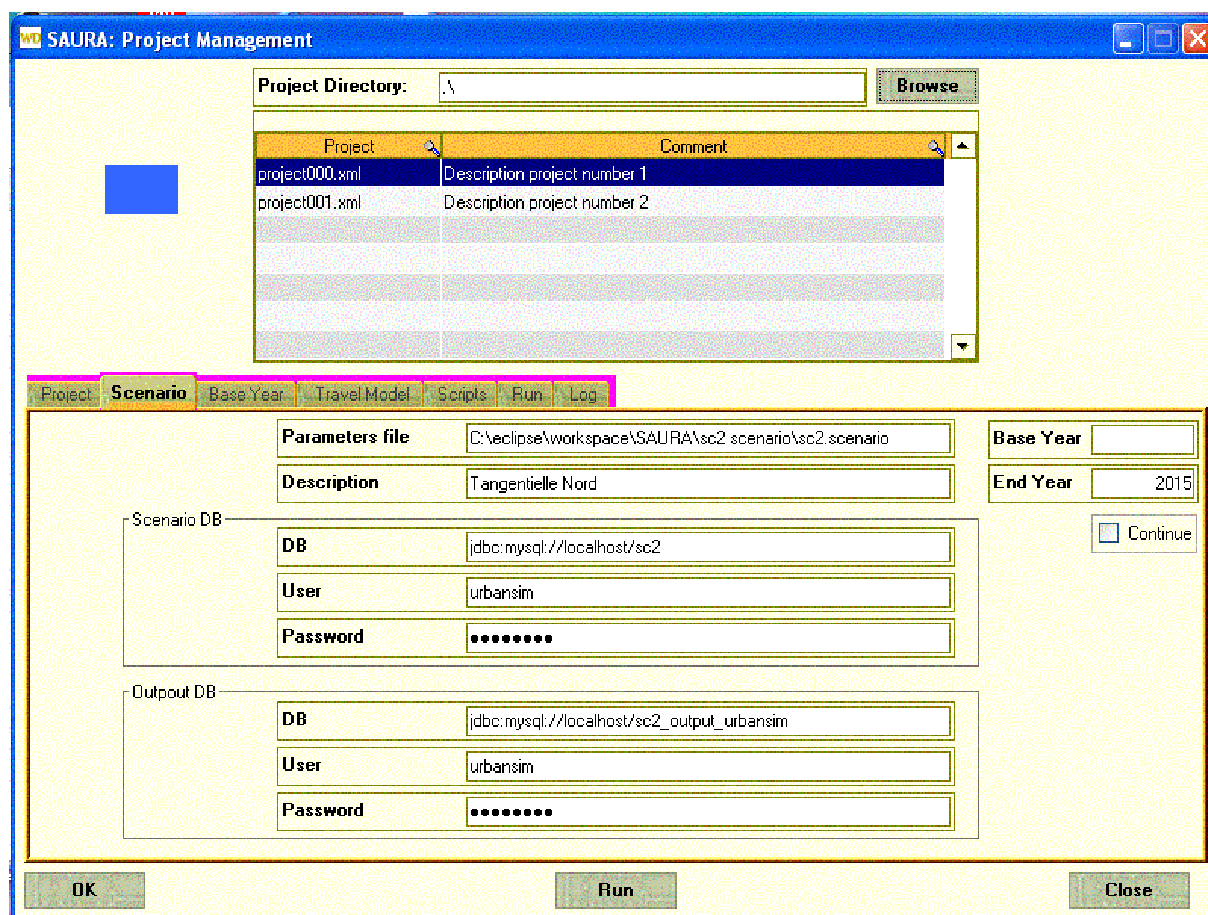


Figure 8 : Fenêtre de gestion des paramètres du modèle intégré

5.1.2. Gestion avancée des paramètres

L'écran principal de l'interface permet de visualiser et de modifier des fichiers « script », ceux-ci sont principalement de type :

- « xml » standard
- « sql » au format des commandes Mysql acceptant des paramètres de substitution au format « \$nom_du_paramètre », « nom_du_paramètre » étant un paramètre défini dans les scripts du projet ou du scénario
- « txt » (édition de fichier texte standard : txt, bat, ...)

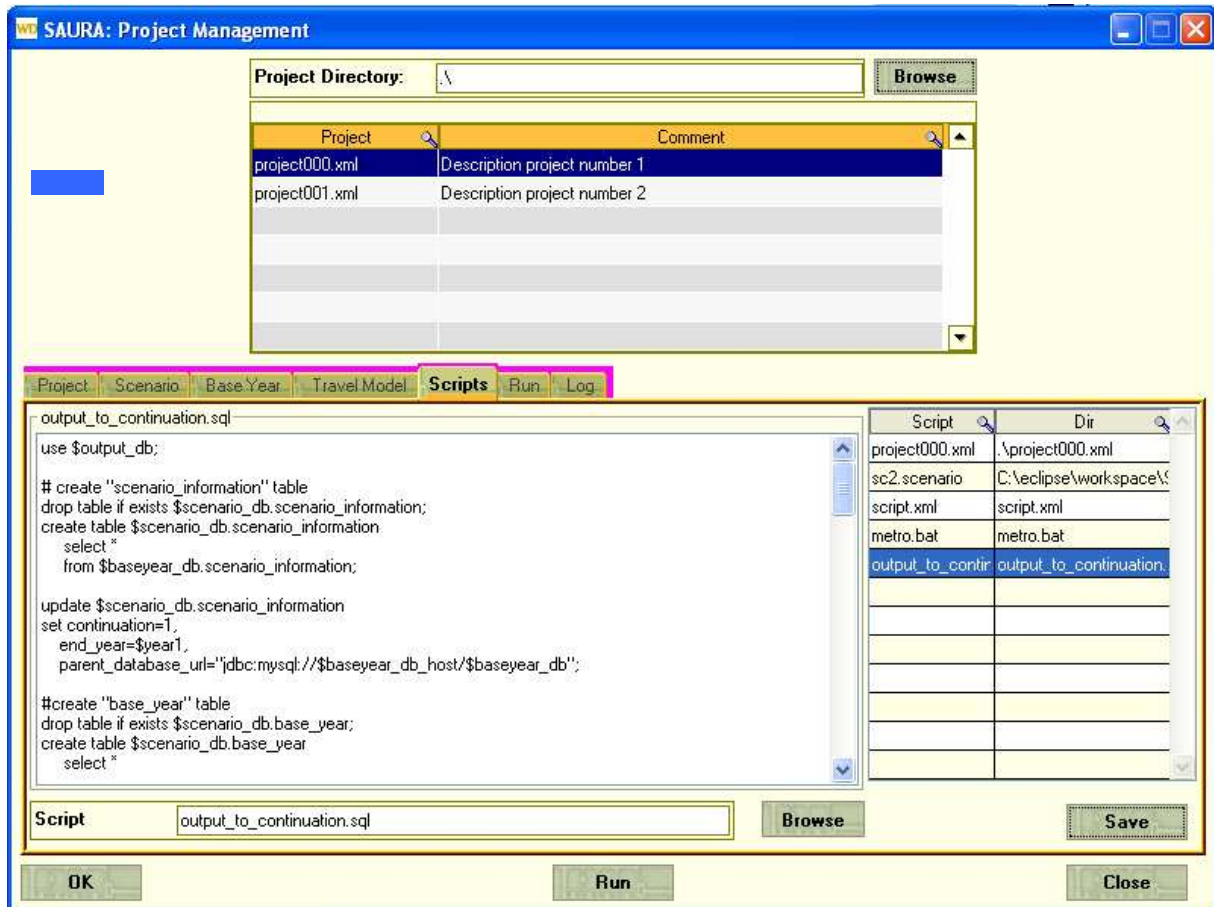


Figure 9 : Fenêtre de la gestion avancée des scripts

5.1.3. Gestion de l'exécution

L'écran principal de l'interface permet de visualiser les étapes d'exécution et leur déroulement (onglet «Run »)

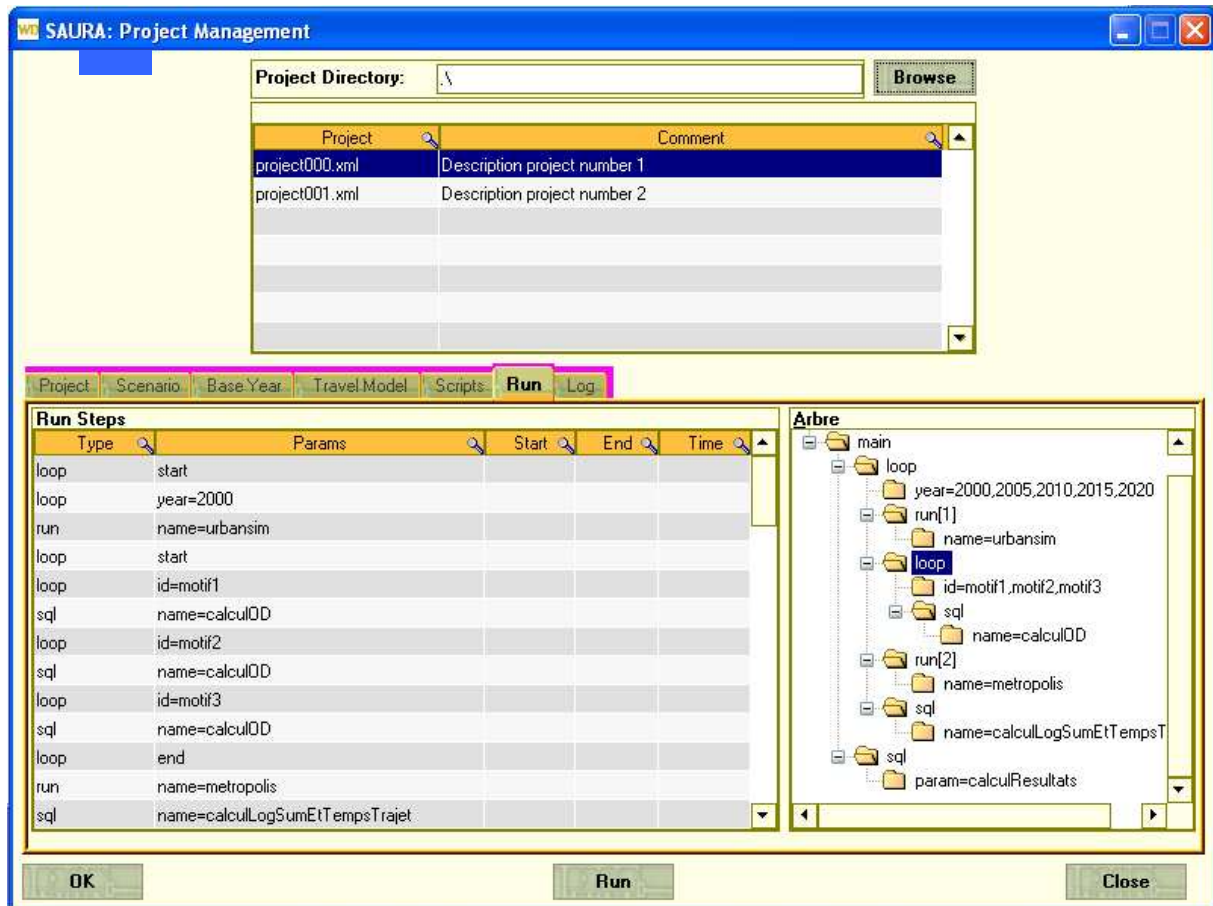


Figure 10 : Fenêtre de la gestion d'exécution des modules

L'onglet « Log » permet de visualiser les messages des différentes étapes.

Malheureusement cette interface n'est plus adaptée à cause du portage d'Urbansim 2 à Urbansim 4.

5.2. Interface entre OPUS/UrbanSim 4 et Davisum-METROPOLIS

Nous avons choisi de développer un script en Python qui permet de prendre en compte l'évolution de l'offre de transport tous les 3 ans de manière exogène, les manipulations se faisant à l'extérieur de *SIMAUURIF*.

Il était important, avant même de disposer d'une interface complète de gestion du bouclage automatique entre Urbansim et Davisum-METROPOLIS (y compris le modèle de demande de déplacements), que nous testions manuellement l'interaction transport-urbanisation dans Urbansim 4. 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.

Puis nous avons calculé au niveau de chaque cellule les variables de transport qui apparaissent dans les deux modèles de localisation : variables d'accessibilités (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) ont été créés au format texte. Pour l'instant, nous n'avons pas mis 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).

5.2.1. Prise en compte des variables transport

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 est 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 a donc 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
etc
```

Puis :

```
data/1996/gridcells.tvp.txt
Et ainsi de suite jusqu'à 1999
```

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.

5.2.2. Lancement de *SIMAU*RIF

Finalement, nous avons écrit un script de lancement de *SIMAU*RIF enchaînant tous les modèles.

Voici le canevas de ce script :

```
'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 essaye d'adapter notre modèle de transition du MOS pour remplacer les quatre précédents : ce  
modèle est propre à notre étude  
  #'evolution_mos_model',  
  
# active le DevelopmentEventTransitionModel qui crée 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 détermine 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 détermine 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'
```

```
],
```

5.3. Conclusion du chapitre

Dans sa dernière version, l'interface entre UrbanSim 4 et Davisum-METROPOLIS est un script Python qui gère les sorties du modèle de trafic et les importe en entrée d'UrbanSim. Cette interface n'est pas une interface conviviale comme l'était la maquette réalisée la première année. Il est prévu de consacrer un peu plus de temps à l'amélioration de cette interface par la suite, mais il est vrai que nous sommes tributaires des évolutions du logiciel UrbanSim qui est en perpétuel développement à l'Université de Washington.

6. Valorisation de la recherche

Ce travail de valorisation apparaît aujourd'hui utile car le projet suscite un intérêt croissant dans la communauté des chercheurs et des praticiens. Le thème de l'interaction transport-urbanisation est un véritable centre d'intérêt dans la recherche sur les transports car c'est un domaine interdisciplinaire, au carrefour de la modélisation de trafic, de la modélisation urbaine, de la planification, de l'analyse de la mobilité et du comportement, de l'économétrie, de l'économie, de la démographie, de la géographie, des sciences régionales, de l'urbanisme, de l'informatique avancée et des systèmes complexes.

Deux articles (« papers ») ont été rédigés en anglais : "*SIMAU*RIF (SIMulation of the interAction between land Use and transport in the Region Paris Ile-de-France) : the whole calibration and the application to the North Tangential railway project" et "An integrated land use-transport model for the Paris Region (*SIMAU*RIF) : ten lessons learned after four years of development". Ils se trouvent en annexe.

Le premier article décrit la procédure de calage global du modèle intégré et les résultats des simulations sur le cas de la Tangentielle Nord. Il fait suite à deux articles de 2005 et 2007 qui décrivent les résultats des trois premières années de la recherche (objectifs de la recherche, recueil de données et résultats des premiers modèles partiels) et qui ont été présentés à quatre conférences : TRB en 2005 (Washington), ERSA en 2005 (Amsterdam), WCTR en 2007 (San Francisco) et le Séminaire francophone de socio-économie des transports en 2007 (San Francisco).

Le deuxième article expose les dix enseignements principaux tirés du projet *SIMAU*RIF.

Pour valoriser la recherche, nous avons soumis les deux articles de 2008 à sept conférences en Europe en 2008 et début 2009 :

- Conférence « Modelling the Spatial Economy » du 3 au 5 juillet 2008 (Kiel)
- Conférence Kuhmo Nectar (Conference and Summer School on Transport and Urban Economics) du 30 juin au 4 juillet 2008 (Amsterdam)
- Colloque ERSA (European Regional Science Association) du 27 au 31 août 2008 (Liverpool)
- Conférence ETC (European Transport Conference) du 6 au 8 octobre 2008 (Leeuwenhorst, Pays-Bas)
- 18ème Conférence européenne des utilisateurs PTV Vision User Group Meeting (UGM) du 21 au 22 Octobre 2008 (Karlsruhe)
- Conférence Internationale «Transportation and Land Use Interaction» du 23 au 25 octobre (Bucarest)
- Conférence mondiale TRB du 11 au 15 janvier 2009 (Washington DC)

Les articles ont tous été retenus pour communication aux six premières conférences. Pour TRB, l'« abstract » a été soumis en juin (réponse en septembre). Il est prévu un intervenant de l'IAU-îdF ou de l'Université de Cergy-Pontoise à chacune de ces conférences.

Paradoxalement, le projet *SIMAU*RIF est plus connu hors de France qu'en France. Exceptée une présentation le 15 janvier 2008 au Réseau des Economistes des Transports (RET), aucune autre communication en Ile-de-France n'a été faite. On envisage au second semestre 2008 de présenter le projet aux utilisateurs potentiels en Ile-de-France (STIF, DREIF, SNCF, RFF, RATP, Ville de Paris).

7. Conclusions

Le projet *SIMAU*RIF est riche d'enseignements.

- 1) Le projet *SIMAU*RIF signe un nouveau paradigme dans les études de modélisation de trafic. Ce sont trois nouvelles approches qui sont ainsi mises en pratique dans ce projet :
 - l'approche **intégrée** : couplage entre un modèle de trafic et un modèle d'urbanisation. Mais il faut envisager l'intégration à un niveau encore plus large, avec en aval un modèle d'émission et de diffusion de polluants et de bruit, et en amont un modèle socio-démographique. La lucidité et l'expérience nous font simplement dire qu'il faudra encore beaucoup de temps avant de voir un modèle universel de simulation des interactions entre transport, urbanisation et environnement, surtout quand on sait que le modèle de trafic statique à quatre étapes a très peu progressé en pratique depuis 50 ans, et cela risque de continuer.
 - l'approche **infracommunale** : les études à l'échelle de l'adresse postale, îlot, quartier doivent se généraliser désormais. Avec l'avènement des Systèmes d'information géographique et les outils dérivés tels que les géocodeurs, il n'y a plus d'excuse aujourd'hui à continuer à faire des études de modélisation au niveau communal.
 - l'approche **désagrégée** : les agents sont représentés un par un avec leurs caractéristiques et avec leurs probabilités de choix discrets.

- 2) La qualité des modèles dépend évidemment de la qualité des estimations économétriques, et aussi, on ne le dira jamais assez, de la qualité des données. Un effort sans précédent a donc été fourni pour construire les bases de données d'entrée et les bases de données d'estimation des modèles.

- 3) Tout au long du projet, nous avons été confrontés à deux méthodologies concurrentes : celle des chercheurs et celle des praticiens. Les chercheurs-théoriciens travaillent chaque modèle partiel de manière théorique, sans souci de parcimonie dans le choix des variables et indépendamment les uns des autres jusqu'à en oublier l'objectif d'intégration de tous les modèles. Cela a pour conséquent de produire de l'incohérence entre les données voire une incompatibilité de fonctionnement en prédiction du modèle intégré, même si chaque modèle partiel est rigoureusement traité et fournit des explications intéressantes sur le phénomène en question. L'autre méthodologie est celle du praticien-ingénieur-modélisateur qui a une vision globale du modèle intégré et qui doit arbitrer entre la qualité économétrique et le réalisme qui se traduit par la possibilité d'implémenter ou non le modèle. Cette deuxième méthodologie a pour conséquence l'obligation de prendre des hypothèses parfois très simplificatrices, peu satisfaisantes intellectuellement, mais l'avantage de permettre l'implémentation donc d'aboutir au final à un outil opérationnel. C'est une réelle gageure que de réaliser l'alliance idéale entre le chercheur, puriste par nature, et le praticien, pragmatique par expérience. Autrement dit, si l'objectif est de mettre au point un outil opérationnel dans des délais imposés, le praticien doit canaliser, contenir, recadrer en permanence les ambitions théoriques du chercheur.

- 4) L'application du modèle intégré *SIMAU*RIF sur le cas de la Tangentielle Nord est un succès. Elle a donné des résultats intéressants à l'horizon 2026 (12 ans après la mise en service) sur l'attractivité économique et résidentielle du territoire desservi par l'infrastructure. Ces résultats rejoignent globalement l'analyse qualitative des urbanistes, sans toutefois être concluante quant au thème de la rente foncière. L'outil *SIMAU*RIF peut donc se positionner comme un nouvel outil d'aide à la décision dans le cadre d'une analyse multi-critères. Il évalue l'impact

d'une infrastructure de transport sur la relocalisation des ménages et des entreprises sous contraintes d'espace disponible et sous condition de concurrence dans le marché immobilier.

- 5) La modélisation de l'interaction transport-urbanisation est un thème situé au croisement de multiples disciplines : mathématiques appliquées, économétrie, économie des transports, urbanisme, sociologie, démographie, économie urbaine, géographie, informatique. Notre recherche a permis un dialogue fructueux entre des experts de chaque discipline.

En ce qui concerne la phase de perfectionnement de *SIMAUURIF* qui a fait l'objet de la présente étude, nous avons réalisé le portage des données et des modèles de la version UrbanSim 2 à la version OPUS/UrbanSim 4, malgré de nombreuses difficultés d'ordre informatique. Nous avons estimé un sous-modèle du modèle de prix, celui du modèle de prix de bureaux, mais qui n'est pas encore implémenté dans l'outil. Enfin, nous avons mis au point une interface assez rudimentaire mais qui fonctionne efficacement entre le modèle de trafic et le logiciel UrbanSim. L'utilisation de cette interface nécessite la connaissance du langage Python. Il reste à rendre cette interface conviviale, mais cela dépendra aussi des développements en cours du produit UrbanSim à l'Université de Washington.

En ce qui concerne la valorisation du projet, nous avons essayé de faire connaître ce projet en dehors de nos frontières. Deux articles en anglais (en annexe) ont ainsi été proposés pour communication à plusieurs conférences européennes en 2008 et 2009 et ils ont tous été acceptés. Il est prévu également à partir du second semestre 2008 de mieux faire connaître l'outil aux utilisateurs potentiels en Ile-de-France (STIF, DREIF, SNCF, RFF, RATP, Ville de Paris).

Quelles sont enfin les perspectives de développement futur du projet *SIMAUURIF* ?

A court terme, il faudrait améliorer encore le calage global et effectuer des tests de sensibilité aux paramètres. Il est prévu de mieux modéliser aussi les impacts des infrastructures sur les prix de l'immobilier. Une étude est actuellement en cours sur le cas du tramway T3 à Paris, après celles sur le RER E et le tramway T1 (par l'IAU-îdF), le tramway T2 (par l'INRETS) et les zones 30 de Paris (collaboration CERNA/IAU-îdF). Une synthèse de ces cinq études en 2009-2010 permettra de conclure quant à l'intérêt d'évaluer l'impact d'une infrastructure sur les prix de l'immobilier à un niveau infracommunal. Il semble intéressant ensuite d'appliquer *SIMAUURIF* à d'autres projets d'envergure en Ile-de-France tels que le projet Arc-Express (rocade de métro en proche couronne), le projet de Tangentielle sud-ouest, et un scénario de péage routier autour de Paris. Les bases de données sont construites, le calage global du modèle intégré est réalisé, même s'il est encore perfectible, les graphes des réseaux transport pour le scénario de base sont construits (un tous les trois ans de 1999 à 2026). Il suffit juste de définir les scénarios, construire les réseaux correspondants puis exécuter le modèle intégré pour voir les effets à long terme de ces projets d'infrastructure ou d'une politique de transport sur l'urbanisation. La proposition de recherche PLAINSUDD soumise à l'appel d'offres ANR « Ville Durable » en mai 2008, en collaboration avec le LET (leader, le Pr Bonnafous), Vinci-Cofiroute et Geomod a été retenue (11 projets sur 44 soumis ont été sélectionnés). Une nouvelle recherche sur trois ans va pouvoir s'engager et permettre de réaliser les développements à court terme sur *SIMAUURIF*.

A moyen terme, il serait intéressant de connecter le modèle intégré en aval à des modèles environnementaux (pollution, bruit) dans le but de calculer des indicateurs de développement durable et en amont à un modèle socio-économique qui alimenterait automatiquement les tables des totaux de population et d'emploi pour chaque année de la période de simulation.

A long terme, on pourrait envisager de ne plus dépendre du logiciel UrbanSim qui a nécessité des adaptations importantes pour le cas de l'Ile-de-France, et développer notre propre modèle français d'urbanisation. En ce sens, il est regrettable que la France ne dispose pas de modèle LUTI et que l'on soit obligé d'utiliser le modèle américain UrbanSim. Il faudrait que la France ait son propre modèle LUTI d'ici 2015, qui soit applicable aux villes françaises et qui soit une référence dans le monde, au même niveau que le logiciel américain UrbanSim. Rappelons que l'équipe du Pr Waddell a reçu des subventions de plusieurs millions de dollars de l'Etat fédéral, d'entreprises privées (IBM) et de l'Etat de l'Oregon pour développer le modèle américain UrbanSim. L'outil est en perpétuel développement et amélioration par des étudiants en PhD. Il faut donc être conscient que le développement d'un tel modèle en France nécessitera des moyens financiers très importants, de l'ordre de 1.5 à 2 millions d'euros sur 4 ans, sans compter la maintenance ultérieure. Il faudra sans doute mutualiser les moyens des laboratoires de recherche et des bureaux d'études, et peut-être coopérer avec des partenaires européens francophones.

Dès maintenant, nous proposons de donner à ce projet le nom de code suivant : ULTISIM (Urban Land - Transport Interaction Simulation). Nous espérons que le projet *SIMAU*RIF servira de base de lancement et ouvrira une nouvelle ère de créativité et d'innovation, comme l'ont permis les modèles de trafic 50 ans auparavant.

Annexe : les 2 articles

ARTICLE 1

***SIMAU*RIF (SIMulation of the interAction between land Use and transport in the Region Paris Île-de-France) : the whole calibration and the application to the “North Tangential” railway project**

**André DE PALMA(*), Kiarash MOTAMEDI (*),
Dany NGUYEN-LUONG (**), Hakim OUARAS (*), Nathalie PICARD (*)**

() University of Cergy-Pontoise - THEMA
95011 Cergy-Pontoise Cedex, FRANCE.*

*(**) Institut d'Aménagement et d'Urbanisme île-de-France (IAU-îdF)
15 rue Falguière, 75740 Paris Cedex 15, France.*

Abstract : This paper describes the second and last part of a land use transport interaction (LUTI) model for the Paris Region. It uses two existing models: UrbanSim (a land use model) and Davisum-METROPOLIS (a dynamic traffic model) including a travel demand model. The whole was assembled and run together. We managed to calibrate it on the period 1990-1999 even though this step can still be improved.

*SIMAU*RIF has been tested on a railway project : the “North Tangential” which is scheduled to open in 2016. The study area is located in the north of Paris and comprises 853 square cells. *SIMAU*RIF could measure in 2026 the indirect benefits of the infrastructure : the number of households and jobs attracted by the North Tangential and their localization around the stations. Interesting results and maps are presented and show that *SIMAU*RIF can be reliable at the expense of some improvements.

Keywords: transportation systems, land use modeling, land use transport interaction model, LUTI, UrbanSim, Davisum, Metropolis

1. INTRODUCTION

*SIMAU*RIF is a land-use – transport interaction (LUTI) model for Paris region that has been developed since 2004 by the Institute for Urban Planning and Development of Paris Ile-de-France Region (IAU île-de-France), in collaboration with the THEMA laboratory of the University of Cergy-Pontoise. The project was completed in december 2007 after almost four years. It is based on two existing models : OPUS/UrbanSim - a land use model developed by the University of Washington, and Davisum-METROPOLIS - the transport model including the travel demand model developed by IAU île-de-France.

*SIMAU*RIF is the first project using OPUS/UrbanSim in the world, outside the U.S., to be completed. Although there are many urban areas in the world that are developing integrated land use-transport models using OPUS/UrbanSim (e.g., Seattle; Detroit; Honolulu; Houston; Phoenix, Arizona; San Francisco; Burlington, Vermont; Durham, North Carolina; Melbourne; El Paso, Texas; Amsterdam; Tel

Aviv; Zurich; Pescara and Turin, Italy; Lyon, France), none of these cases has resulted in a completed model to date.

Before exposing the whole calibration step and the simulations results on the North Tangential, we present some background about the *SIMAU*RIF project and the main results of the two first years.

2. BACKGROUND

In France, there are many classical transport models, of varying degrees of sophistication, which make use of commercial software such as PTV's Davisum or Citilab's Cube, but which remain basically a four-step model. By contrast, there are no examples of integrated land use-transport models in France. By contrast, in the U.S., this kind of model is more common, in part, because of laws such as ISTEA and TEA-21, which require regional transport plans to take into account the interactions between land use and transport. So, in the U.S., one finds applications of land use model software like MEPLAN, DRAM/EMPAL, TRANUS, and, more and more, UrbanSim. In 2004, we decided to choose UrbanSim for our project. The main reason was that it contained interesting capabilities versus the other packages and above all, it was free (GNU license). The aim was to develop an operational model that could test an isolated transport scheme or simulate transport policies or land use policies and show the effects on urban development and on mobility behavior.

The second impetus for our project was that in France, in the field of socio-economic assessment of infrastructures, the cost-benefit method often does not satisfy decision makers, even though many scientists and engineers find this method to be robust and powerful. To most decision makers, terms such as "internal benefit-cost ratio," "time savings," or "externalities" of an infrastructure project have little meaning. So, our project had another objective : to allow one to measure the indirect effects of providing infrastructure, and not only the direct effects (e.g., time savings and environmental savings). What are these indirect effects? Residential attractiveness (e.g., the number of households that will move near an infrastructure project ten years after its opening), economic development (e.g., the number of jobs attracted around or along a new infrastructure project ten years after the project is built), and, finally, the increase in land prices or housing prices inducted by the new infrastructure. So we aimed to develop a model that would be able to quantify these indirect effects in the mid and long terms and to include them as new criteria, similar to an internal benefit-cost ratio in a multi-criteria approach.

Finally, it is well known that the long term analysis of the transportation system is not possible without considering its interaction with land use and urban form. The integrated land use transportation model is a necessary tool to evaluate the effects of the policies on local and global pollution which is a global concern nowadays.

3. MAIN RESULTS OF THE TWO FIRST YEARS

Realized by a team of four persons, the project lasted 3 years and a half. The methodology was based on the construction of data bases at the grid level (original in France), on the use of disaggregated data to estimate the different models and on the use of discrete choice models theory. Two existing software were assembled : the land use model OPUS/UrbanSim (developed at the University of Washington by Pr Waddell's team) and the forecasting traffic model Davisum-METROPOLIS including the travel demand model developed by IAU île-de-France.

The first year of this project (2004) was devoted to data collection and input and calibration data bases construction based on a very fine zoning of Paris Region. UrbanSim uses grid cells as the data and simulation unit. We have for the first time in France built a huge database in a grid of 50,000 cells covering the entire Paris Region, with each cell being a 500 meter by 500 meter square. We took into account the environmental constraints (flood zones, protection zones around monuments, etc) and political constraints, and we exploit at best the data at the smallest level (parcel data) thanks to a geographical information system (GIS).

It was difficult to transpose the pre-defined data structure of UrbanSim to the case of Paris Region. UrbanSim was designed to simulate American metropolitan areas, which are typically characterized by an uncontrolled urbanization, which spreads over a big area, but results in relatively low urban densities and is characterized by a transport system, dominated by roads with relatively high speed and capacity. In France, by contrast, city development is more controlled by government plans and intervention, which is characterized by an efficient organization of space and not so much by adding to the road network. For example, in the pre-defined data structure of UrbanSim, there is no place to integrate public transport variables. So we had to work hard to adapt UrbanSim to the case of the Paris Region, sometimes resulting in simplifications that were not intellectually satisfying.

In the second year (2005), each submodel was estimated independently (residential location choice model, jobs location choice model, development type choice model, hedonic land price model, and the four classical steps of the demand and assignment models). As described in [1], [4] and [5], these two first years (data base imputation and submodels estimations) demanded the most significant efforts.

4. THE WHOLE CALIBRATION

After the independent estimation of each sub-model, we assembled all the models and ran the whole together on OPUS/UrbanSim platform. The step of the whole calibration is very tedious and very time-consuming. An execution during the period 1990-1999 lasts four hours on a standard PC (CPU 3.4 GHz, RAM 2 GB) and we had to run it about fifty times to calibrate it.

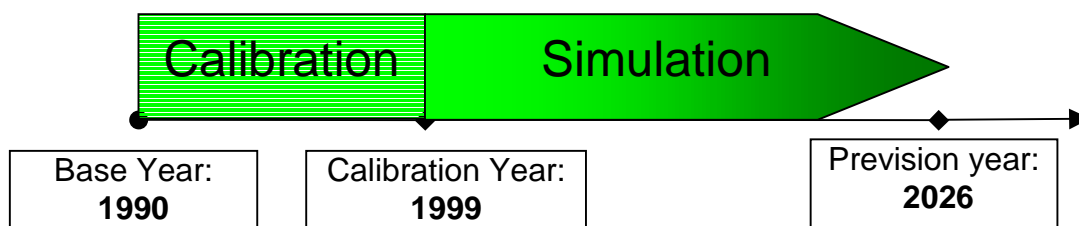


Figure 1

Figure 1 depicts two periods : the calibration period (1990-1999) necessary to validate the model, and the simulation period (1999-2026).

To make the calibration, we compared the 1999 socioeconomic and demographic forecasts of *SIMAURIF* to the 1999 observed data published by INSEE at an aggregated level. So we ran *SIMAURIF* from 1990 to 1999 and calculated some indicators that we compared to the same observed indicators. We used two levels of calibration : municipality (or city) and county levels. Calibration consists in adding dummy constants per county in the utility functions.

Let's remind a few key figures about Paris Region, namely Ile-de-France Region, which embraces Paris and its suburbs. The city of Paris has about 2 million inhabitants, on a regional total of 11.3 million. The total number of jobs is 5.1 million. It covers 4,610 sq. miles (12,000 sq. km). There are 3

administrative institutions : 1 “region”, 8 “departments” (counties) and 1300 “communes” (municipalities).

The simulations were run with a one-year step for UrbanSim and a three-years step for Davisum-METROPOLIS. So the accessibilities in UrbanSim are updated every three years to reflect changes in the spatial distribution of population and employment.

The means to couple UrbanSim and the regional travel demand model involved creating a script in Python language to extract data from the UrbanSim output database, reformatting of these data to the format required by the trip generation step in Davisum-METROPOLIS, execution of the travel model, extraction of logsums and travel times and exporting them into the UrbanSim scenario database, and then running UrbanSim each year until the next scheduled travel model run. This process was completely automated by a script that runs UrbanSim from the base year of 1990 to the end year of 1999.

Municipality level

To evaluate the goodness of *SIMAU*RIF, we compared the simulated values of main indicators to their observed values in 1999. To do this, we used the formula of the relative difference and the absolute difference which are given by

$$relative_difference = \frac{Simulated - Observed}{Observed}$$

$$absolute_difference = Simulated - Observed$$

We defined the classes of 0.05 between -0.5 and 0.5. The values of classes represent the relative difference between the simulated and observed data of indicators calculated by this equation. According to the classes, we computed the number of cities which have the relative difference equal to the value of one class (frequency). The following results are given *before* calibration.

Households

We ran *SIMAU*RIF from 1990 to 1999 and we computed the number of households in each municipality, we compared these values with the observed ones by computing the relative difference. Table 1 presents the results.

Classes	Frequency	%
-0.5	5	0.4%
-0.4	2	0.2%
-0.3	13	1.0%
-0.2	43	3.3%
-0.1	148	11.4%
0	282	21.7%
0.1	319	24.5%
0.2	202	15.5%
0.3	122	9.4%
0.4	79	6.1%
0.5	37	2.8%
or more...	48	4%

Table 1- Number of households, relative difference between simulated and observed data.

About 67% of municipalities have 15% (in absolute value) of the difference relative between the simulated and observed number of households. More than 75% (986 municipalities) have less than 20%.

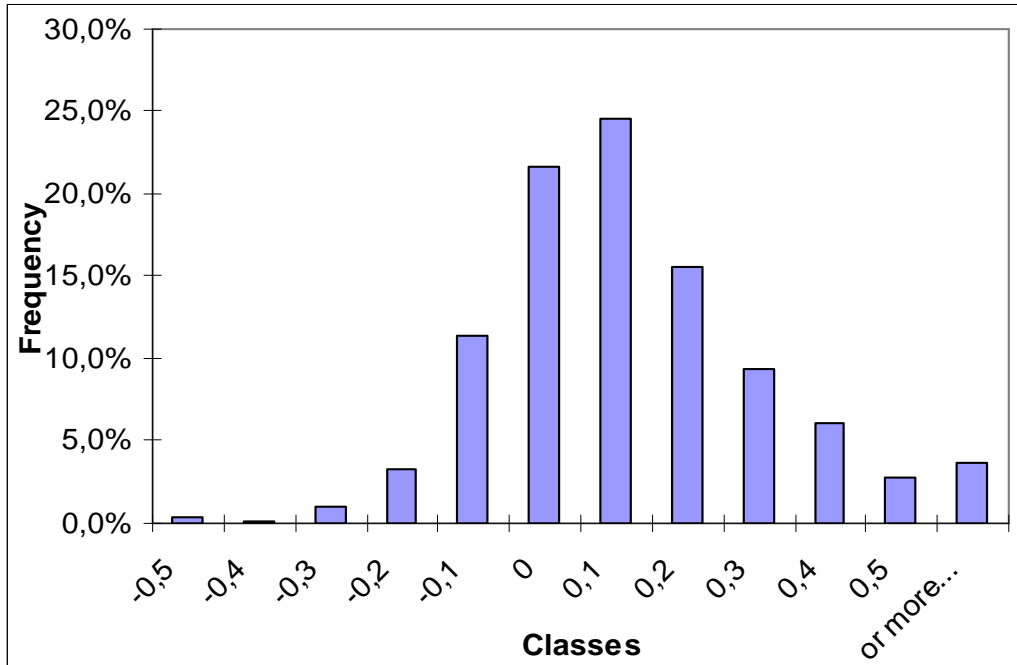


Figure 2- Number of households, relative difference between simulated and observed data.

About 16% of municipalities are underestimated. SIMAURIF has estimated the same number of households in 159 cities (22%). 62% are overestimated.

Population

As well as for the number of households, we computed the population indicator for each municipality. Then, we compared the results to the observed data by computing the relative difference.

Classes	Frequency	%
-0.5	5	0.4%
-0.4	9	0.7%
-0.3	20	1.5%
-0.2	128	9.8%
-0.1	248	19.1%
0	330	25.4%
0.1	254	19.5%
0.2	143	11.0%
0.3	76	5.8%
0.4	33	2.5%
0.5	22	1.7%
or more...	32	2.5%

Table 2- Population, relative difference between simulated and observed data

About 66% of municipalities have a relative difference between -15% and 15%.

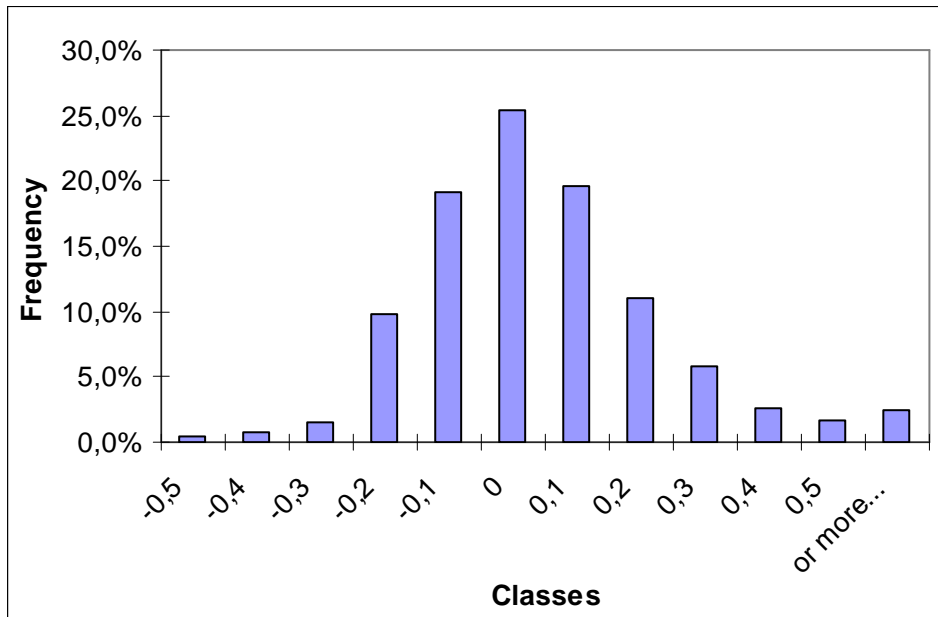


Figure 3- Population, relative difference between simulated and observed data

In this case, 46% of cities are underestimated. 12.2% have the same value between estimated and observed data, and 41.3% are higher than can be observed.

Jobs

The number total of jobs has decreased between 1990 and 1999. There are 5 180 541 jobs in the entire region in 1990 and only 5 041 995 in 1999.

*SIMAU*RIF uses 8 types of jobs. The total employment is the sum of the employment of the 8 types.

In the Figure 4 below, we present the simulated data as a function of observed data by using a logarithmic scale.

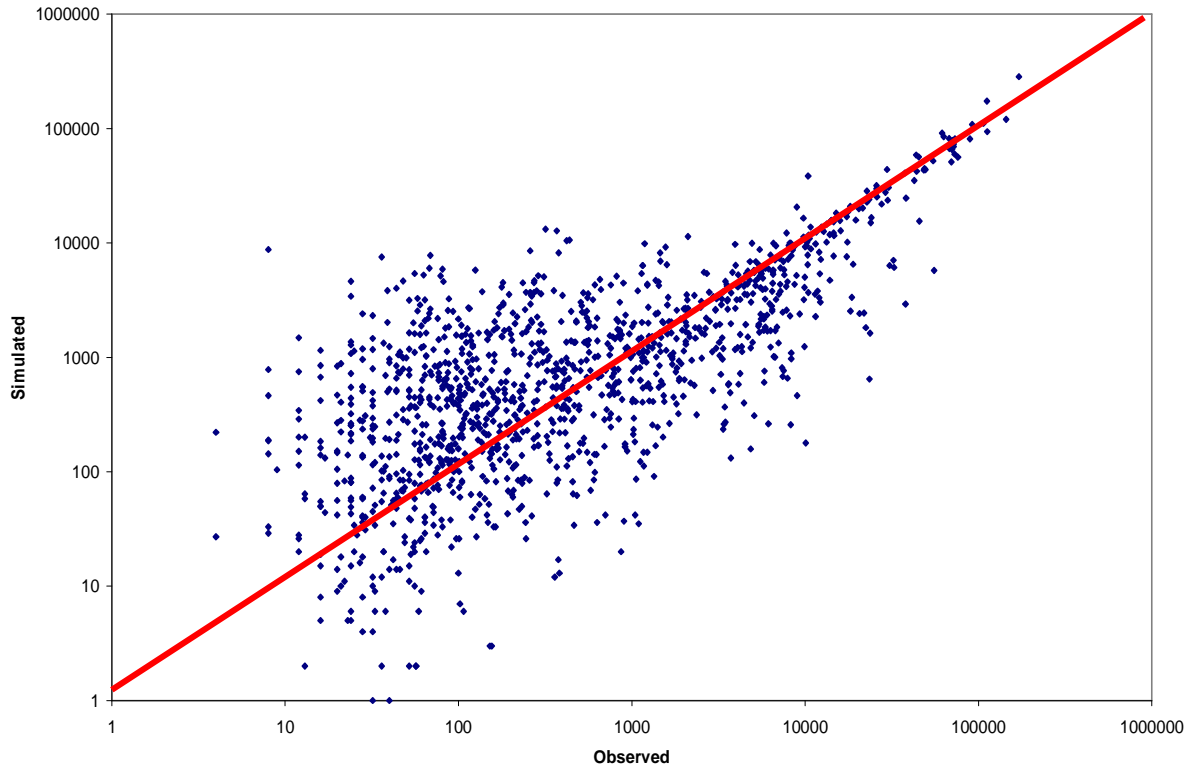


Figure 4- Number of jobs, simulated function of observed data.

Figures 5 show the relative difference between the simulated results and the observed data. The left one represents the municipalities with less than 5000 jobs and the right one the municipalities with number of jobs greater than 5000.

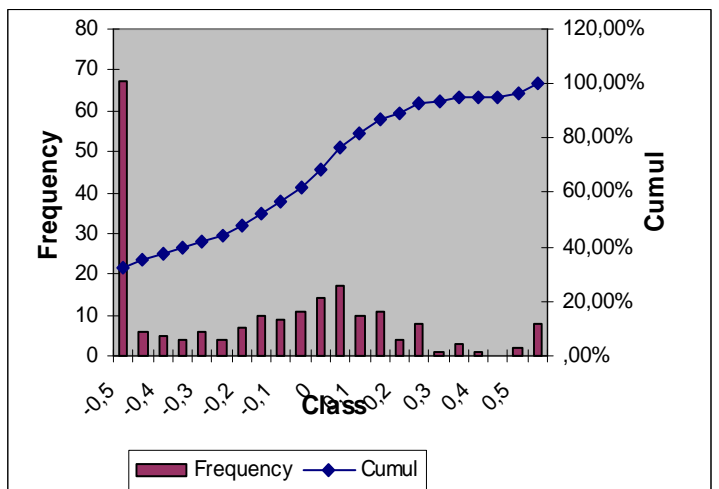
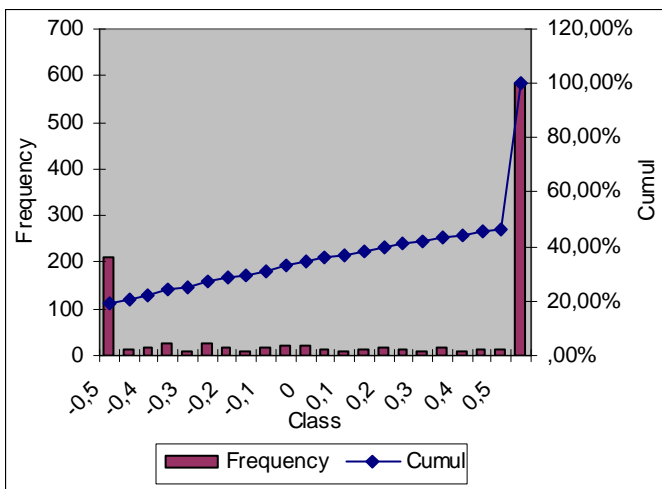


Figure 5- Number of jobs, relative difference for municipalities with less and more than 5000 jobs.

SIMAURIF overestimates the number of jobs in the municipalities having a small number of jobs and sub-estimate the jobs of the municipalities having a high number of jobs.

Dwelling Price

We used a hedonic estimation to compute the dwelling price. We didn't succeed to estimate a price model at the cell level. So the estimation was made at the municipality level by using the notary database which contains the average of dwelling price. The price was estimated for three types of dwelling : the houses ($R^2 = 70\%$), the flats ($R^2 = 60\%$) and the two types combined together ($R^2 = 53\%$). The model implemented in *SIMAU*RIF is the last one.

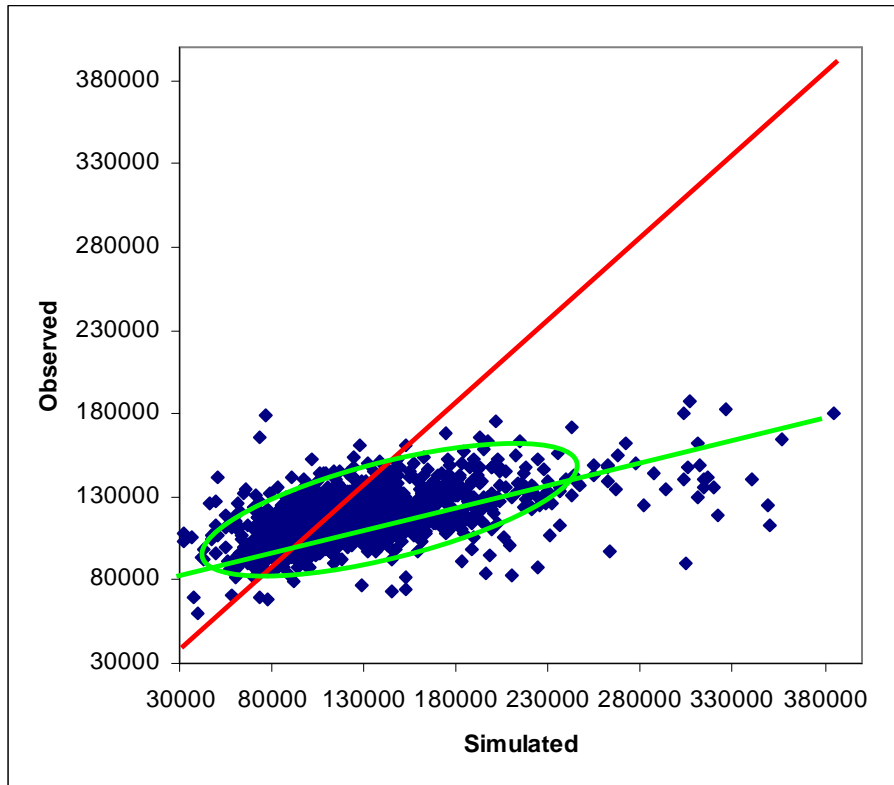


Figure 6- Correlation between simulated and observed data of dwelling price

The coefficient of correlation between the observed and simulated data is about 60%. Due to a weak R^2 , the model sub-estimates the price in the municipalities with high dwelling price. Let's note that the estimation of this model is made at the municipality level and the computing in *SIMAU*RIF is made at the cell level according to the UrbanSim software. To use the results of this model, we had to disaggregate the results at cell level. So it is a new source of error.

In terms of relative difference, more than 80% of the municipalities have a relative difference between -30% and $+30\%$ (see Figure 7).

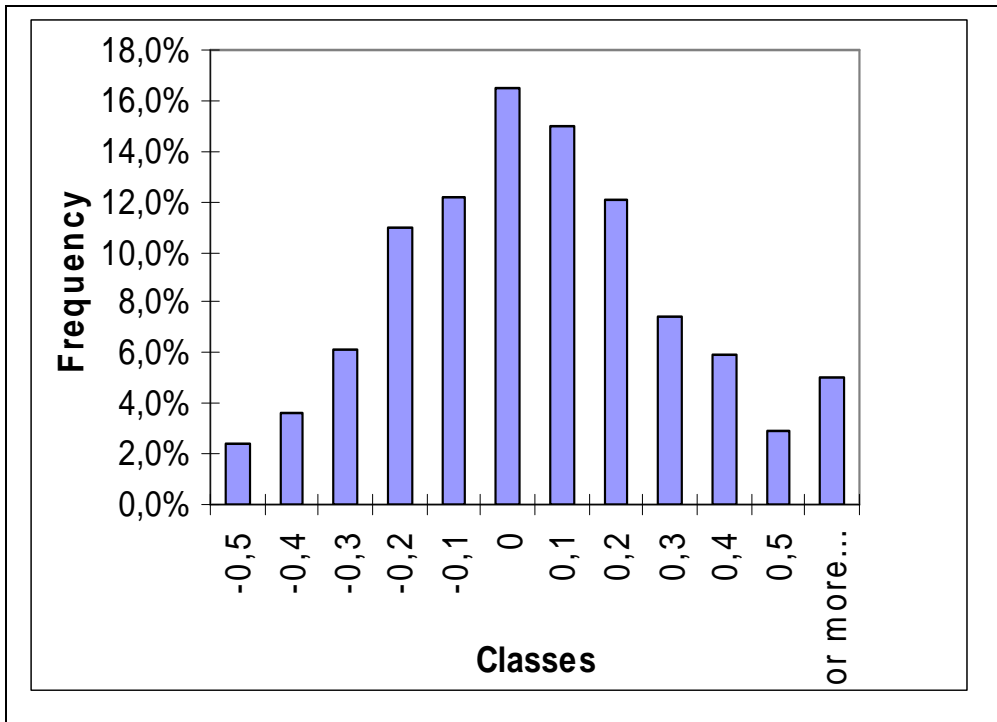


Figure 7- Dwelling price, relative difference between simulated and observed data

Residential units

We aimed by this indicator to illustrate one of the results of the development model.

Figure 8 represents the number of dwellings (residential units) simulated by SIMAURIF according to the number of dwellings observed in each municipality.

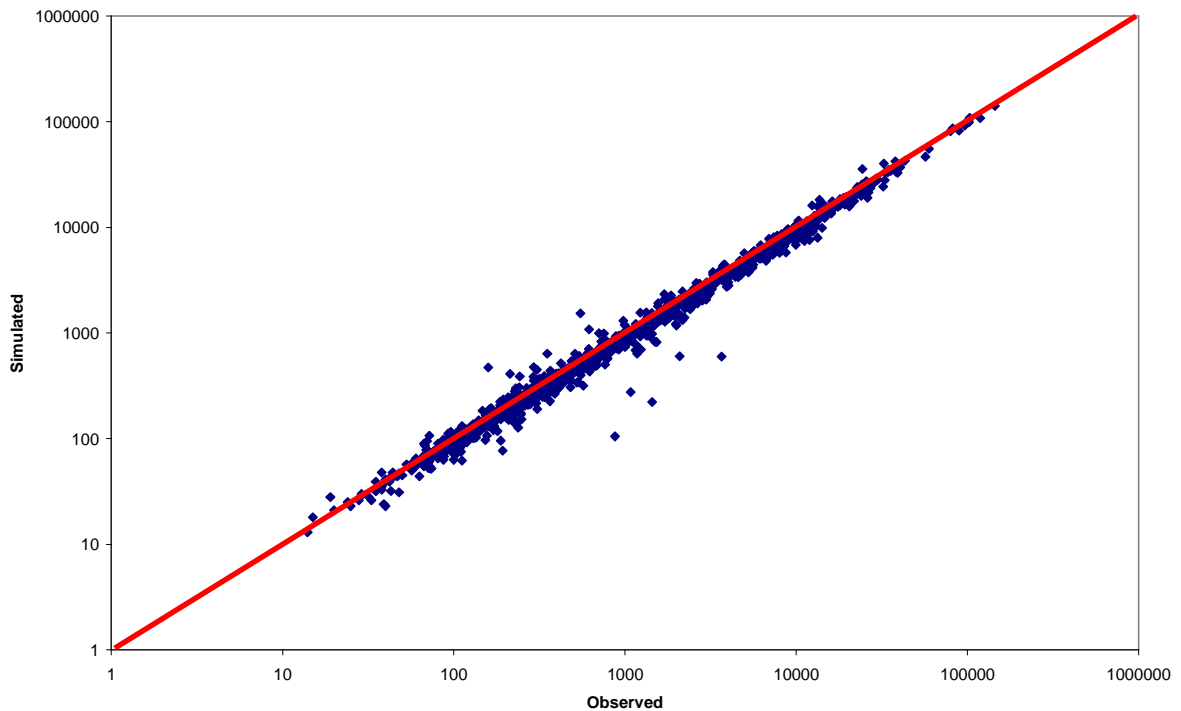


Figure 8- Correlation between simulated and observed residential units

We got a correlation of 99.6% between the simulated and observed data.

Classes	Frequency	%
-0.5	6	0.5%
-0.4	7	0.5%
-0.3	17	1.3%
-0.2	62	4.8%
-0.1	241	18.5%
0	476	36.6%
0.1	243	18.7%
0.2	138	10.6%
0.3	64	4.9%
0.4	26	2.0%
0.5	6	0.5%
or more...	14	1.1%

Table 3- Residential units, relative difference between simulated and observed data

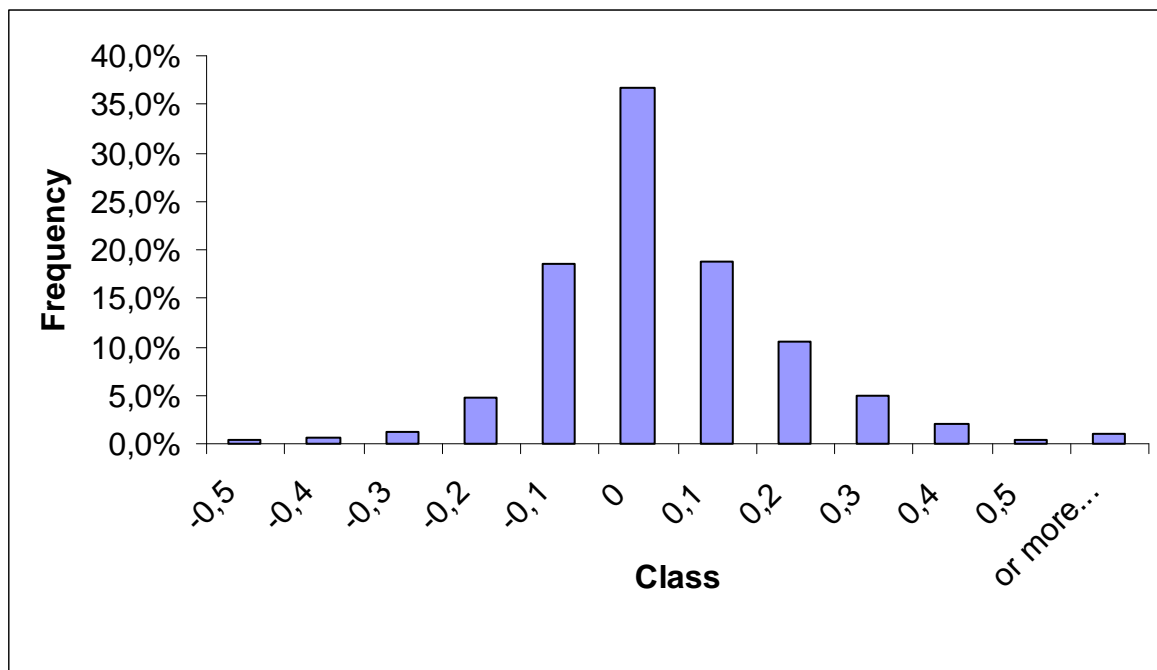


Figure 9- residential units, relative difference between simulated and observed data

About 74% of municipalities have a difference less than 0.1 in absolute value. *SIMAU*RIF simulated approximately the same value than the observed values for 36% (476) of municipalities.

County level

At the county level, it is easy to get the differences because there are only 8 counties. At this level, we aggregated the data and calculated the differences between the simulated and the observed data. We presented the differences in terms of number of households, population and number of jobs in each county, always *before* calibration.

Households

Table 4 shows the absolute and relative differences of number of households by county between the simulated and observed data in 1999.

County	Simulated	Observed	Absolute difference	Relative difference
75	1184665	1110912	73753	7%
77	428165	432351	-4186	-1%
78	507562	503096	4466	1%
91	416218	420603	-4385	-1%
92	615113	624926	-9813	-2%
93	535589	524387	11202	2%
94	492187	499404	-7217	-1%
95	384423	394690	-10267	-3%

Table 4- Number of households, relative difference between simulated and observed data.

The changes in the spatial distribution of households from 1990 to 1999 show an over-growth in Paris and least discrepancies in the suburb.

Population

We computed below the relative difference for the population in each county.

County	Simulated	Observed	Absolute difference	Relative difference
75	2575376	2077806	497570	24%
77	1077365	1175254	-97889	-8%
78	1266531	1328456	-61925	-5%
91	1034562	1106128	-71566	-6%
92	1421491	1399672	21819	2%
93	1320828	1355810	-34982	-3%
94	1172976	1198617	-25641	-2%
95	977462	1083005	-105543	-10%

Table 5- Population, relative difference between simulated and observed data.

The great difference is found for Paris. It is more than what we observed with the households. This means that *SIMAU*RIF locates more of large households in Paris.

Ratio of number of households by population

Table 6 shows the ratio of number of households by population by county. This ratio allows to present the differences in terms of structure of the households.

County	Simulated	Observed	Relative difference
75	0.46	0.53	-16%
77	0.39	0.36	7%
78	0.40	0.37	5%
91	0.40	0.38	5%
92	0.43	0.44	-3%
93	0.40	0.38	4%
94	0.42	0.41	1%
95	0.39	0.36	7%

Table 6- Ratio of number of households by population.

The ratio of the households by population is relatively great (in absolute value) in Paris. This means that the model locates less households than the observed data, relatively to the population. The reason is that *SIMAU*RIF locates in Paris more of large households.

Average of households size

Table 7 shows the average of households size. This average is calculated by dividing the total population by the number of households in each county.

County	Simulated	Observed	Difference relative
75	2.17	1.87	14%
77	2.51	2.71	-8%
78	2.49	2.63	-6%
91	2.48	2.62	-6%
92	2.31	2.23	3%
93	2.46	2.58	-5%
94	2.37	2.40	-1%
95	2.53	2.73	-8%

Table 7- Average of households size

Even for this indicator, we notice the great difference in Paris. It confirms that *SIMAU*RIF locates in Paris too much of large households.

Jobs

We present below the results of the employment model *before* and *after* calibration.

Results *before* calibration

We computed the absolute and relative differences for the number of jobs in each county.

County	Simulated	Observed	Absolute difference	Relative difference
75	1815350	1600815	214535	13%
77	329243	388947	-59704	-15%
78	447315	504 469	-57 154	-11%
91	328502	401 397	-72 895	-18%
92	786497	815 471	-28 974	-4%
93	490550	483 999	6 551	1%
94	474999	472 447	2 552	1%
95	369539	374 450	-4 911	-1%

Table 8- Number of jobs, relative difference before calibration

So, we got relatively bad results for Paris and the far suburb (77, 78 and 91 counties).

Results after calibration

We did some calibration by adding dummies for each county (without Paris) as explanatory variables during the estimation step. Furthermore, to reduce the weight of Paris, we added in the function of the utility a dummy "Paris" which is equal to 1 if the location county is Paris, 0 otherwise. We assigned a negative value for this dummy. Then we ran again the whole system and got the following results for a value equal to -100 :

County	Simulated	Observed	Absolute difference	Relative difference
75	1676809	1600815	75994	5%
77	399821	388947	10874	3%
78	503893	504 469	-576	0%
91	334857	401 397	-66 540	-17%
92	786497	815 471	-28 974	-4%
93	490550	483 999	6 551	1%
94	474999	472 447	2 552	1%
95	374569	374 450	119	0%

Table 9- Number of jobs, relative difference after calibration with négative constant for dummy Paris.

So, after adding the constants we got best results. Paris is still too attractive and the county 91 is the less attractive.

At that stage, we decided to stop the calibration process, after two months of hard work.

5. APPLICATION: Simulation from 1999 to 2026

Before presenting the results of comparison of two scenarios, with and without North Tangential, we provide below some general results about running *SIMAU*RIF on the whole Paris region for the scenario with North Tangential. The *SIMAU*RIF forecast used a total population of 12 millions while the jobs control total is 5.58 millions. Like for the calibration period. The simulations were run from 1999 to

2026 with a one-year step for UrbanSim and a three-years step for Davisum-METROPOLIS. The run time for a single scenario using a standard PC is about 7 hours.

General results

Figure 10 shows how will evolve the concentration of the population in the whole region between 1999 and 2026.

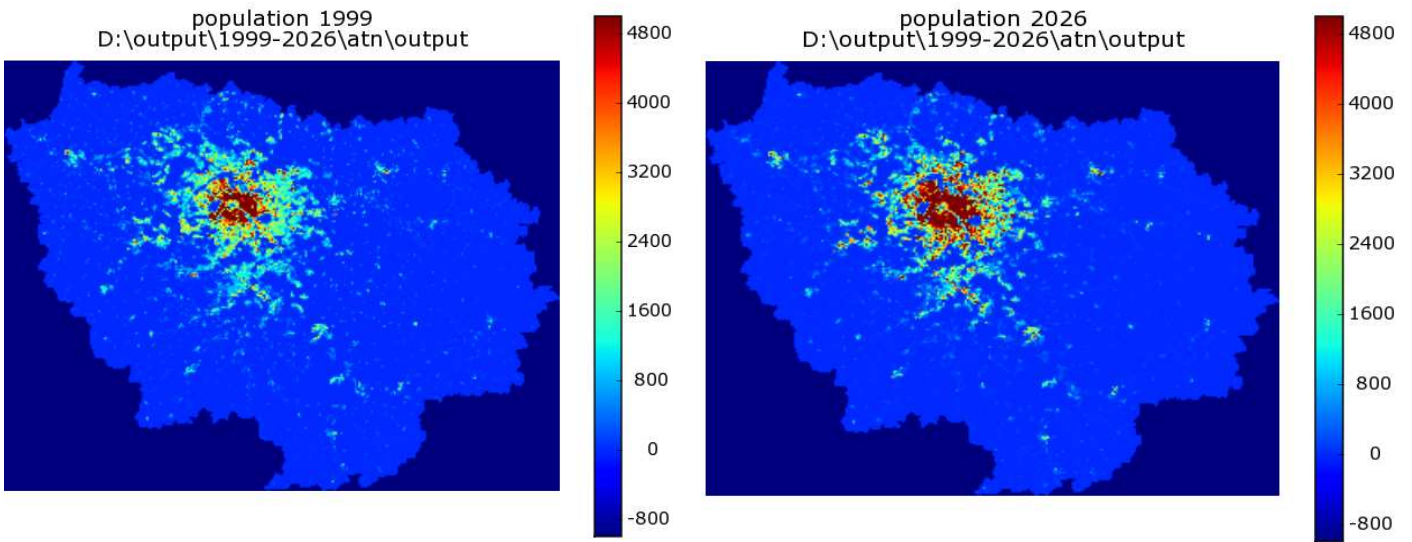


Figure 10- Results of SIMAURIF: Distribution of the population in 2026

The circle of concentration of the population is growing from year 1999 to 2026.

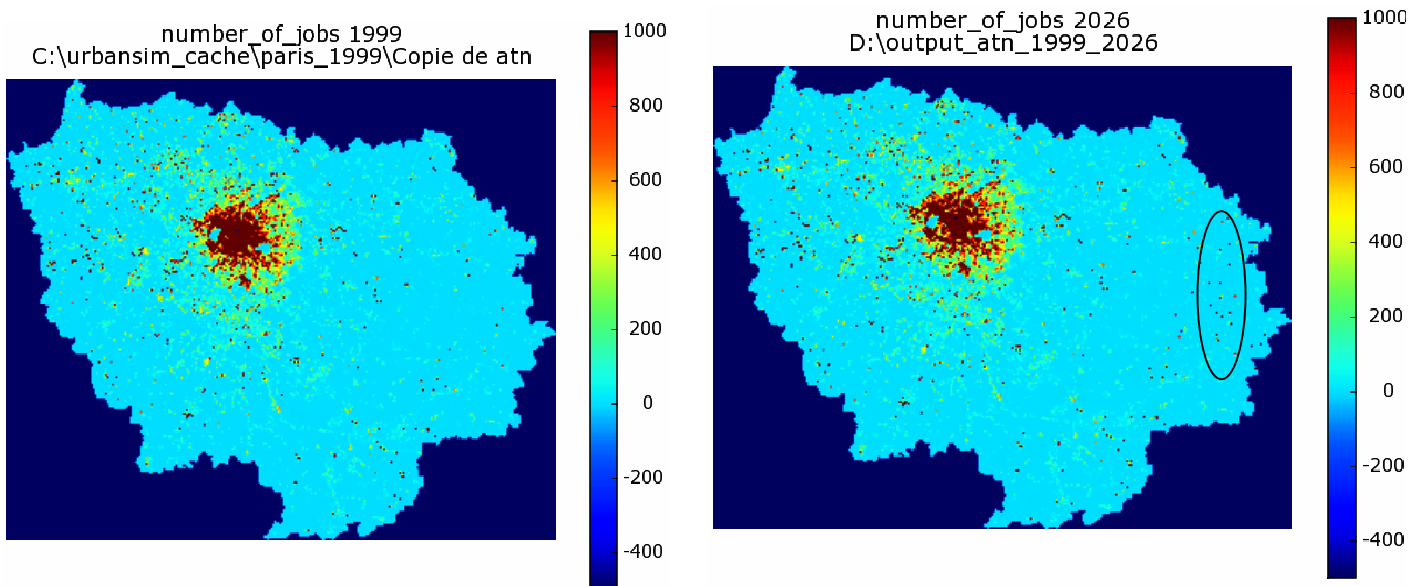


Figure 11- Results of SIMAURIF: Evolution of the distribution of jobs in 2026

Under the assumptions about the jobs [6], we can see that the structure of jobs generally speaking remains the same between 1999 and 2026 in the entire region. But there are two highlights. The first

one is that the number of jobs will decrease in Paris and around. The second point is the relocalization of jobs in the east of the Region (zone encircled by the ellipse).

Application to North Tangential (NT)

SIMAURIF was applied to assess the effects of the North Tangential, a project of tram-train in the north of Paris, which opening is scheduled in 2016.

In the two scenarios examined, there were several common assumptions. These include the total population and employment in the region for each year, the model coefficients and land use plan assumptions. The only differences between the two scenarios concern the accessibility and transport variables for 2017, 2020, 2023 and 2026.

A study area was defined as a corridor of 853 squared cells and 38 municipalities. Table 11 shows the total population and employment in 1999 and the simulated results from the residential location and employment location models for the two scenarios in 2026.

	Population		Employment	
1999	1.260		0.416	
	Population		Employment	
	Without NT	With NT	Without NT	With NT
2026	1.520	1.522	0.508	0.518

Table 11

The North Tangential allows to attract 2000 inhabitants more in 2026. This difference between the two scenarios is extremely weak, probably not significant irrelevant to the trend ($2000/260000 < 1\%$).

As far as concerned the employment, *SIMAURIF* finds 10 000 jobs more in the scenario with the NT irrelevant the scenario without NT. This time, the difference of 10 000 jobs should be significant even though it is weak. It shows that the impact of North Tangential on the economic development is not important if no other accompaniment policy is proposed (urban and development projects, landscape, housing rehancement, ...)

How are these 10 000 jobs distributed ? The spatial distribution of jobs in 2026 at the cell level is displayed in Figure 12. New jobs are mainly located at the two extremities of the line (Sartrouville-Argenteuil and Noisy-le-Sec stations). So the model has highlighted the fact that the neighbourhoods of Noisy-le-Sec station become the new economic pole of the area. The North Tangential allows to connect with the tramway T1. with its extension to Montreuil. with the RER E (express railway) and with the tram-train T4, bringing about an effect of interconnection. This gain of local and regional accessibility contributes much to the jobs location choice near the station.

For the first time in France, an integrated model could forecast the spatial repartition of the population and of the jobs at the cell level at the horizon year 2026 and distinguish a difference of repartition between the scenario with and the scenario without the infrastructure. The GIS allows to create maps, once grid cells results have been transferred, and to show interesting results around the stations. On the other hand, *SIMAURIF* was not able to simulate the housing price in 2026 at the cell level but only at an aggregate level.

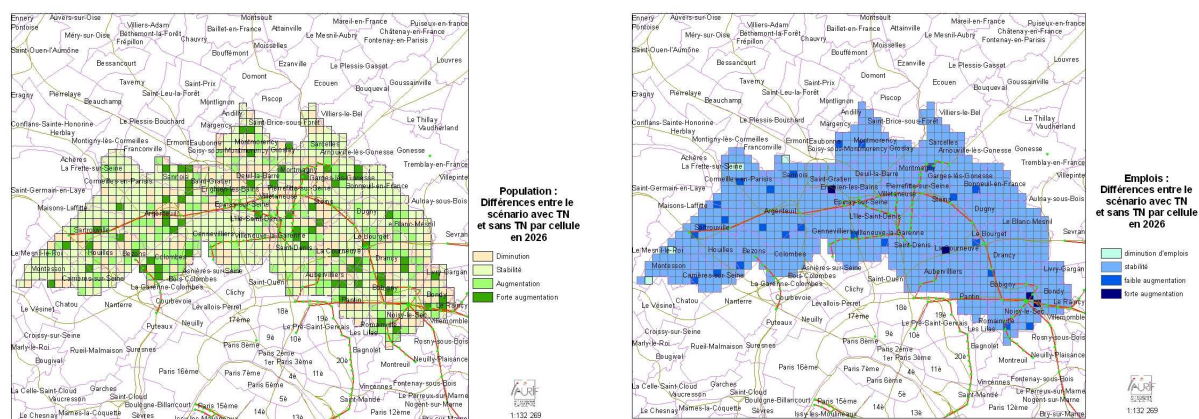


Figure 12

6. PERSPECTIVES

In the short term, it seems worthwhile to apply *SIMAURIF* to other important transport planning studies in the Paris Region. such as the Arc-Express project (tangential subway in the near suburb). In the current context when it's recommended to enrich the assessment methods, an integrated tool such as *SIMAURIF* can be used in a multi-criteria analysis to evaluate the indirect socio-economic benefits of an infrastructure investment, in complement with the conventional cost benefit analysis, and allow to classify transports schemes according to these new criteria. The database for the region is already built, the model is calibrated (even though the calibration could always use improvements), the transport model networks for the base scenario have already been built (one for every three years from 1999 to 2026). You just have to define the scenarios, build the corresponding modeling networks and run the model. Furthermore, we'll have to enhance the value of this project to the academic community and also to other potential users in the Paris Region.

In the mid term, it would be interesting to connect the integrated model to an environmental assessment model (pollutants emission and dispersion model, noise propagation model) in order to calculate sustainable development indicators, and upstream, to couple a macro socio-economic model which can feed with projections of population and employments the totals controls tables for each year of the period of simulation.

This research has also allowed to experiment the last version of *OPUS/UrbanSim* and to learn about the software developments in Python language, about the specifications of the different sub-models and their interaction, and finally about the necessary adaptations for its application to the case of Paris region. The three published reports will serve as a reference to develop a French land use transport interaction model recognized at the international level. But this is a long term project which requires important investments and probably a national cooperative R&D program. We hope our project will open an era of creativity and innovation, like in the early days of travel modeling forty years ago.

Bibliography

- [1] de Palma, A. Nguyen-Luong, D. Motamedi, K. Picard, N. Moyano, J. (2005). Development of a dynamic integrated Land use – Transportation model. *45th Congress of the European Regional Science Association (ERSA) 23-27 August 2005, Vrije Universiteit Amsterdam*.
- [2] de Palma, A. and Marchal, F. (2002). Real cases applications of the fully Dynamic METROPOLIS Tool-box: an Advocacy for Global Large-scale Mesoscopic Transportation Systems. *Networks and Spatial Economics*. **Vol. 2** No. 4. Kluwer Academic Publishers. pages 347-369.
- [3] Waddell, P. Borning, A. Noth, M. Freier, N. Becke, M. and Ulfarson, G. (2003) Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. *Networks and Spatial Economics*. **Vol. 3** No. 1. pages 43-67.
- [4] Waddell, P. Sevcikova, H. Socha, D. Miller, E. and Nagel, K. (2005) "Opus: An Open Platform for Urban Simulation". The Computers in Urban Planning and Urban Management Conference. London. U.K.

The three reports in French are downloadable :

- [5] *SIMAUURIF*. Modèle dynamique de simulation de l'interaction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle Nord. Rapport final de la 1ère année (2004)
http://www.iau-idf.fr/exl-doc/IA46029-1_OUV00001603.pdf
- [6] *SIMAUURIF*. Modèle dynamique de simulation de l'interaction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle Nord. Rapport intermédiaire de la deuxième phase (2005)
http://www.iau-idf.fr/exl-doc/IA46029-2_OUV00002471.pdf
- [7] *SIMAUURIF*. Modèle dynamique de SIMulation de l'interAction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle Nord. Rapport final de la 2ème phase (2007)
http://www.iau-idf.fr/exl-doc/IA48950_OUV00005358.pdf

Acknowledgements

We would like to acknowledge French Ministry of Ecology and sustainable development (Land transportation research and development program, PREDIT) for supporting two projects. The first one *SIMAUURIF* (Ref. 05MT6008) and the second one entitled "GESTION DU TRANSPORT ET DE LA MOBILITÉ DANS LE CADRE DU CHANGEMENT CLIMATIQUE" (Ref. 07MTS063). Also we would like to thank DREIF (Governmental department of transport in Paris region) and RFF (French Railways Network) who provided co-funding.

ARTICLE 2

An integrated land use-transport model for the Paris Region (*SIMAU*RIF) : Ten lessons learned after four years of development

Submission date : June 4, 2008

Word count : 3398

Dany NGUYEN-LUONG

*Institut d'Aménagement et d'Urbanisme île-de-France (IAU-îdF)
15 rue Falguière, 75740 Paris Cedex 15, France.*

dany.nguyen-luong@iau-idf.fr

Tel : (33) 01.77.49.77.44

Fax : (33) 01.77.49.77.69

Abstract

After four years of developing of an integrated land use-transport model for the Paris Region, we have learned a number of lessons. This paper presents ten of these lessons from the point of view of a practitioner, after presenting the background of the project, and then proposes a few perspectives.

Keywords: transport systems, land use modelling, integrated land use-transport model, LUTI, simulation, calibration, UrbanSim, Davisum, Metropolis

1. INTRODUCTION

*SIMAU*RIF is an integrated land use-transport model for the Paris Region that has been developed over the last four years by the Institute for Urban Planning and Development of the Paris Ile-de-France Region (IAU île-de-France), in collaboration with the THEMA laboratory of the University of Cergy-Pontoise. We completed the project in December 2007. It is based on two existing models : OPUS/UrbanSim - a land use model developed by the University of Washington, and Davisum-METROPOLIS - the transport model including the travel demand model developed by IAU île-de-France. This project was funded via a grant by PREDIT (Inter-ministerial transport research and innovation programme), DREIF (Governmental department of transport in Paris region) and RFF (French Railways Network).

SIMAURIF is the first project using OPUS/UrbanSim in the world, outside the USA, to be completed. Although there are many urban areas in the world that are developing integrated land use-transport models using OPUS/UrbanSim (e.g., Seattle; Detroit; Honolulu; Houston; Phoenix, Arizona; San Francisco; Burlington, Vermont; Durham, North Carolina; Melbourne; El Paso, Texas; Amsterdam; Tel Aviv; Zurich; Pescara and Turin, Italy; Lyon, France), none of these cases has resulted in a completed model to date.

This paper is addressed to Metropolitan Planning Organizations as well as to research laboratories that are using or contemplating using OPUS/UrbanSim. The reader can be reassured : there are no mathematical formulas in this paper. We present the main lessons of such a project from the point of view of a modeller-practitioner, not a researcher. So we are aware this paper is not very academic, but it will be quite objective and will reflect the reality of the daily practise. We hope it will be useful to the research community because it contains an abundance of practical advice and raises some legitimate concerns about developing integrated land use-transport models.

Before presenting the ten lessons, we present some background about the *SIMAURIF* project and the context of transport modelling in France.

2. BACKGROUND

In France, there are many classical transport models, of varying degrees of sophistication, which make use of commercial software such as PTV's Davisum or Citilab's Cube. On the other hand, there are no examples of integrated land use-transport models in France. By contrast, in the U.S., this kind of model is more common, in part, because of laws such as ISTEA and TEA-21, which require regional transport plans to take into account the interactions between land use and transport. So, in the U.S., one finds applications of land use model software like MEPLAN, DRAM/EMPAL, TRANUS, and, more and more, UrbanSim. In 2003, we decided to choose UrbanSim for our project. The main reason was that it contained interesting capabilities versus the other packages and above all, it was free (GNU licence). The aim was to develop an operational model that could test an isolated transport scheme or simulate transport policies or land use policies and show the effects on urban development or on mobility behaviour.

The second impetus for our project was that in France, in the field of socio-economic assessment of infrastructures, the cost-benefit method often does not satisfy decision makers, even though many scientists and engineers find this method to be robust and powerful. To most decision makers, terms such as "internal benefit-cost ratio," "time savings," or "externalities" of an infrastructure project have little meaning. So, our project had another objective : to allow one to measure the indirect effects of providing infrastructure, and not only the direct effects (e.g., time savings and environmental savings). What are these indirect effects? Residential attractiveness (e.g., the number of households that will move near an infrastructure project ten years after its opening), economic development (e.g., the number of jobs attracted around or along a new infrastructure project ten years after the project is built), and, finally, the increase in land prices or housing prices inducted by the new infrastructure. So we aimed to develop a model that would be able to quantify these indirect effects and to include them as new criteria, similar to an internal benefit-cost ratio in a multi-criteria approach.

Next, we present our ten lessons.

3. TEN LESSONS

Lesson 1 : From theory to practice : a difficult journey

A review of the literature in 2003 and again in 2007 concerning integrated land use-transport models showed that the literature about this kind of model is typically very theoretical (e.g., focusing on aspects such as localisation models, hedonic land price models, activity-based models, etc.), but nothing was found about practical development and applications. And even if an application is presented in a paper or in a report, no details are provided about the construction of the databases or the whole calibration, which are usually the most difficult steps. Actually, in the beginning of such a project, most people underestimate the difficulties involved in developing these models. The more you advance in the project, the more you find out about complexities that you did not foresee in the beginning. So, the first lesson is : if you underestimate the difficulties in a project, you will be obliged, by the end of the project, to simplify or to sacrifice some tasks and/or to take short cuts because of the deadlines.

Lesson 2 : A methodology to follow

We have developed a methodology that should gain widespread acceptance. This methodology is based on the following principles : Take into account environmental constraints (e.g., flood zones, historic preservation areas, etc.) and political constraints. Maintain a consistency between inputs and outputs of the different models. Do not hesitate to adapt the data structure of OPUS/UrbanSim to the situation at hand in the study. Exploit, to the maximum extent possible, fine-grained geographic data, thanks to a geographical information system (GIS). Avoid simulating data or synthesizing missing data that will be used for calibration. Make a descriptive analysis of the data before using it for modelling. Build a single calibration database for all the models. Keep in mind that after the independent estimation of each sub-model, there will be a step of assembling all the models together and calibrating the whole model that will be tedious and very time-consuming (e.g., a simulation lasts seven hours on a standard PC and you should run the model at least one hundred times). Last, distinguish clearly the calibration period and the simulation/application period.

Lesson 3 : It's worth building a huge database at the grid level

The saying "Garbage in, garbage out" is well known. It's interesting to note that almost all modelling studies conducted by researchers present extraordinary results, despite the fact that the data used in the models are barely described, and/or the quality of the data is often mediocre. In our project, we have, for the first time in France, built a huge database in a grid of 50,000 cells covering the entire Paris Region, with each cell being a 500 meter by 500 meter square. But is this grid relevant, and if so, is it worth the trouble, given that its construction is very time consuming? We answer, "yes," except in the case of the land price model. We were unable to estimate a land price model at the cell level, due to the large size of the study area (about 12,000 sq km). On the other hand, the grid allowed us to incorporate two interesting elements : the possibility to take into account the notion of a neighbourhood (each cell interacts with its eight neighbours) and the possibility of outputting results at the cell level (around stations, for example). Of course, the reliability of the results at that level remains to be shown due to the margins of error.

Lesson 4 : Explanative model versus predictive model : finding the compromise

Our experience with the project shows that you are better off building a simple model, with few variables (ten at the most), than building a very sophisticated model, with plenty of variables. This is especially true if you included simulated variables, perhaps because you felt they were important, but, since they were not available, you decided to create them by estimating another sub-model. The problem is that, once you've estimated this explanative model, you are likely to find out it is impossible to use it in a simulation. In our project, we made the mistake of spending too much time in the estimation of each model, and, in the end, we were unable to implement the models in UrbanSim, so we had to go back and re-estimate simpler models.

In the famous novel by Michel Tournier (1967) entitled, *Vendredi ou les Limbes du Pacifique* ("Friday"), there is a passage where, after reading the Bible, Robinson decides to join Chili by building a boat he calls "The Escape". Robinson devotes all his energy and imagination to building this boat. But, the day the boat is finished, Robinson suddenly realizes the boat will never float because it's too heavy. The same lesson applies to modelling : if your model is too sophisticated, you won't be able to apply it in forecasting because the variables which were chosen for estimation cannot be forecasted.

Lesson 5 : UrbanSim, which was designed for American metropolitan areas, has to be converted to the French context

Our project has shown that it's difficult to convert a model which simulates urban development in the North American context to another very different context. UrbanSim was designed to simulate American metropolitan areas, which are typically characterised by an uncontrolled urbanization, which spreads over a big area, but results in relatively low urban densities and is characterized by a transport system, dominated by roads with relatively high speed and capacity. In France, by contrast, city development is more controlled by government plans and intervention, which is characterized by an efficient organization of space and not so much by adding to the road network. For example, in the data structure of UrbanSim, there is no place to integrate public transport variables. So we had to work hard to adapt UrbanSim to the case of the Paris Region, sometimes resulting in simplifications that were not intellectually satisfying.

Lesson 6 : Developing an interface to manage the interaction between the transport model and the land use model : a partially reached objective

This part of *SIMAU*RIF is easily understandable. But here again, it's the passage from theory to practise which is complex. We wanted to automate the following steps : execution of the classic "four-step" transport model, exporting the results of the transport model into the UrbanSim input database, running the four components of UrbanSim, exporting the results in the UrbanSim database to the trip generation step of the travel demand model, then running this cycle each year during the period of simulation, by modifying the road and public transport networks every three years (all with the push of a button!). We concede that we have only partially reached this objective. We had developed, during the second year of the project, a prototype to interface between UrbanSim 2 and Davisum-Metropolis. But the last year, we had to abandon this interface because of the re-engineering of UrbanSim that was part of its Version 4 update. So the current interface has been simplified to its simplest form : a script written in Python language, which takes into account every three years the supply of transport and only for off-peak/unconstrained conditions (i.e., without assigning the origin-destination matrices).

The transport variables are computed externally and are then formatted into text files which are read by the script.

Lesson 7 : It's vital to have an excellent knowledge of the study area (development, demography, economy, transport, environment)

To make the project successful, it's necessary to have a global vision of the interactions between transport and urbanization, a good understanding of the development process occurring in the study area, a "perfect" knowledge of the urban projects and transports projects, an excellent knowledge of the geography of the employment and population, a "perfect" knowledge of the land use evolution, and, finally, a good understanding of the game theory that governs the actors or stakeholders in the system. In other words, the risk in conducting a project such as this is to think that it is simply an exercise in econometrics and computer science, and that you could simply estimate the logit models for Paris Region in the same way that you would for San Francisco.

Lesson 8 : *SIMAU*RIF allows one to measure some indirect effects (e.g., residential attractiveness and economic development caused by an infrastructure), but not yet the housing prices.

*SIMAU*RIF was applied to assess the effects of the North Tangential, a proposed tram line in the northern suburb of Paris, which is scheduled to open in 2016. For the first time in France, an integrated model was used to forecast the spatial distribution of population and jobs at the cell level for a horizon year of 2026, and to distinguish a difference in the distribution of populations and jobs between the scenario with and the scenario without the infrastructure. The GIS allows one to create maps, once grid cells results have been transferred, and to show interesting results around the stations. On the other hand, *SIMAU*RIF was not able to simulate housing prices in 2026 at the cell level, but only at an aggregate level. So a tool like *SIMAU*RIF can be used in a multi-criteria analysis to evaluate the indirect socio-economic effects of an infrastructure project.

Lesson 9 : UrbanSim, which is a product of the research world, is so difficult to master that we do not recommend it for use in France if you have to get the model up and running in a short time frame

The latest version of OPUS/UrbanSim, Version 4, has a very different logic from that used in versions 2 and 3. In short, Version 4 is about scripting and macrolanguage, whereas the previous versions took the approach of a "black box" or hidden system. For a new user, the "black box" approach is preferable, since it relies on pre-established and inter-connected models, which you just have to estimate, without worrying about additional implementation issues. In earlier versions of the software, the practitioner could concentrate on his job, which is modelling, not implementation. In the beginning of the project, we were really enthusiastic about the choice of UrbanSim : the architecture looked really interesting and relevant. It was comprised of four clear blocks : a residential location choice model, an employment location choice model, a land price model, and a development type choice model. But the shift to a new version in 2007 meant that we had to adapt and simplify all the databases we had built in the previous version. Moreover, the documentation wasn't updated at the same speed as the software. It appeared, based on various messages sent to the UrbanSim users' mailing list, that this new version of UrbanSim had really destabilized a number of current projects around the world, and

that the users were having many difficulties because of purely software-specific problems, not stable release and enigmatic warnings or error messages.

Another important problem is that the interaction between the four main components of UrbanSim is not, in our opinion, clearly designed. Finally, the choice of the Python language for the new version was not, according to us, the best choice because it requires one either to learn this new language or to find a computer programmer, who would work on only this language.

Lesson 10 : It's necessary to put together an inter-disciplinary team and to allocate adequate resources

Classical transport modelling requires only abilities in modelling and applied mathematics. By contrast, modelling the interaction land use and transport requires one to have an additional skill set : econometrics, demography, economics, geography, urban planning, complex systems, mobility behaviour, and computer science. A project like *SIMAUURIF* takes at least three years and a team of at least four people. So you have to put together a stable team for that period. If the objective is to result in an operational tool, you also have to make sure that the project is not dominated by econometrics experts or computer experts. The modeller has to remain the head of the project and limit the ambitions of his colleagues. It seems to us that the current UrbanSim development team at the University of Washington is now controlled more by econometrics experts and computer experts than by urban planners and modellers. As a result, OPUS/UrbanSim has become a project for computer science PhD students. But to our knowledge, the development team is working hard on addressing most of the concerns presented in this paper through new software enhancements (e.g., development of a graphical user interface and a tutorial).

4. PERSPECTIVES

In the short term, it seems worthwhile to apply *SIMAUURIF* to other important transport planning studies in the Paris Region, such as the Arc-Express project. The database for the region is already built, the model is calibrated (even though the calibration could always use improvements), the transport model networks for the base scenario have already been built (one for every three years from 1999 to 2026). You just have to define the scenarios, build the corresponding modelling networks and run the model. Furthermore, we'll have to enhance the value of this project to the academic community and also to other potential users in the Paris Region.

In the mid term, it seems that if you want to get a fully customized piece of software and, above all, one that you can control easily without using Python, it is probably better to develop your own land use model. You might choose to keep the global architecture of UrbanSim (without dropping any sub-models, even the most difficult one, the Development Project Location Choice Model which aims to limit the capacity of the other location choice models) and the grid data structure, both of which are really quite relevant, and then implement your own model in a classical procedural language (e.g., Visual Basic, C). You can also program the travel demand model (the first three steps) and program easily the feedback between your own software and an assignment software package, such as Davisum-METROPOLIS.

In the long term, it would be interesting to connect the integrated model to an environmental assessment model (pollutants emission and dispersion model, noise propagation model) in order to

calculate sustainable development indicators, and upstream, to couple a macro socio-economic model which can feed with projections of population and employments the totals controls tables for each year of the simulation period.

Despite the difficulties encountered in this project and any shortcomings with the current *SIMAU*RIF model, this project has allowed us to advance the state-of-the-practise in the modelling of the interaction between land use and transport in France.

References

- *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 2ème phase* . IAU île-de-France, Paris. www.iau-idf.fr/exl-doc/IA48950_OUV00005358.pdf .
Accessed Dec. 19, 2007.

- *SIMAU*RIF. *Modèle dynamique de simulation de l'interaction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle Nord. Rapport intermédiaire de la deuxième phase*. IAU île-de-France, Paris.
www.iau-idf.fr/exl-doc/IA46029-2_OUV00002471.pdf . Accessed Nov. 9, 2005.

- *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 1ère année*. IAU île-de-France, Paris.
www.iau-idf.fr/exl-doc/IA46029-1_OUV00001603.pdf . Accessed Nov. 4, 2004.

Acknowledgements

I wish to thank Mark Moran from the Metropolitan Washington Council of Governments (MWCOCG) and Paul Waddell from the Center for Urban Simulation and Policy Analysis (CUSPA) at the University of Washington for their comments and corrections. I wish also to thank my partners from the University of Cergy-Pontoise, France : André de Palma, Kiarash Motamedi, Hakim Ouaras and Nathalie Picard.