

La population synthétique localisée Un pré-requis pour la modélisation

Module D1 – Population et logements

01. Contexte et état de l'art
02. Données disponibles
03. Méthodologie
04. Résultats
05. Limites et conclusion

Partie 1

Contexte et état de l'art

- 01. Contexte et état de l'art**
02. Données disponibles
03. Méthodologie
04. Résultats
05. Limites et conclusion

Contexte et positionnement

Modéliser les comportements individuels



Les modèles LUTI se fondent généralement sur

- des approches économiques ou issues de la recherche en transport
- des modèles robustes et anciens : 4 étapes, interaction spatiale, fonctions d'utilité, etc.
- des approches agrégées (groupes d'individus)
- des échelles spatiales larges (commune, aire urbaine, etc.)

Mais de nouvelles approches émergent, fondées sur de nouvelles questions

- 1 La description des comportements individuels est devenue centrale pour comprendre le choix et les décisions des individus dans leur stratégie de mobilité quotidienne et résidentielle.
- 2 Il existe un intérêt croissant pour les 'déplacements courts' notamment dans les espaces centraux et les nouveaux quartiers résidentiels périurbains (gentrification, piétonisation, mobilité douce, etc.).
- 3 Ce double constat incite les chercheurs à modéliser les dynamiques de mobilité à partir des acquis de la micro-simulation, des systèmes multi-agents (MAS) ou des modèles basés sur les activités (ABM).
- 4 Les approches MAS ou ABM doivent être nourries avec des données individuelles, c.à.d des données désagrégées au niveau des individus.

Un problème : accéder à ces données

- ? Compte tenu des législations nationales et de la protection de la vie privée, les données sont rarement accessibles à l'échelle des individus

En France par exemple, l'information est agrégée par groupes de plusieurs centaines de personnes, pour que les habitants ne puisse être ni isolés ni identifiés (CNIL).

Contexte et positionnement

Une solution : générer une population synthétique



Pour pallier ce problème, une solution consiste à générer une population dite synthétique

Population synthétique =

Population artificielle composée d'individus auxquels sont associées des caractéristiques individuelles (niveau n), et construite à partir des données connues au niveau agrégé $n+1$ du recensement

→ Comment faire ?



Contexte et positionnement

Un état de l'art (rapide)



La méthode IPF



L'*Iterative proportional fitting* (IPF) a été introduite par Beckman (1996) pour créer des agents à partir de deux sources de données disponibles aux Etats-Unis



Arentze *et al.* (2007) ont adapté cette méthode à la problématique du transport à partir de données néerlandaises équivalentes

Autres méthodes



Münnich and Schürle (2003) utilisent des simulations de Monte-Carlo pour générer une population synthétique à partir d'une enquête spécifique (German Microsensus)



La méthode a été modifiée par Alfons *et al.* (2010) et appliquée à l'échelle de l'Union européenne à partir des fichiers *Statistics on income and living conditions* (SILC)

→ Ces méthodes dépendent fortement de la structure des données d'entrée et ne peuvent pas être appliquées directement au cas français

Méthodes françaises



Mathis *et al.* (2006) ont développé une méthode fondée sur des probabilités conditionnelles, mais elle ne tient pas compte de la structure spatiale et résidentielle



Banos *et al.* (2010) se basent sur une analyse en composantes principales à partir d'une EMD, données qui ne sont pas disponibles pour toutes les agglomérations françaises

→ Un développement spécifique est nécessaire avec trois objectifs

- une comptabilité avec les données françaises
- une reproductibilité pour toutes les agglomérations
- une échelle fine pour prendre en compte les déplacements courts

Contexte et positionnement

Les objectifs du Module D1



La méthode doit permettre de lier

- les caractéristiques des agents (et des ménages)
- les caractéristiques spatiales de la ville et sa structure résidentielle (immeubles, appartements, maisons individuelles, etc.)

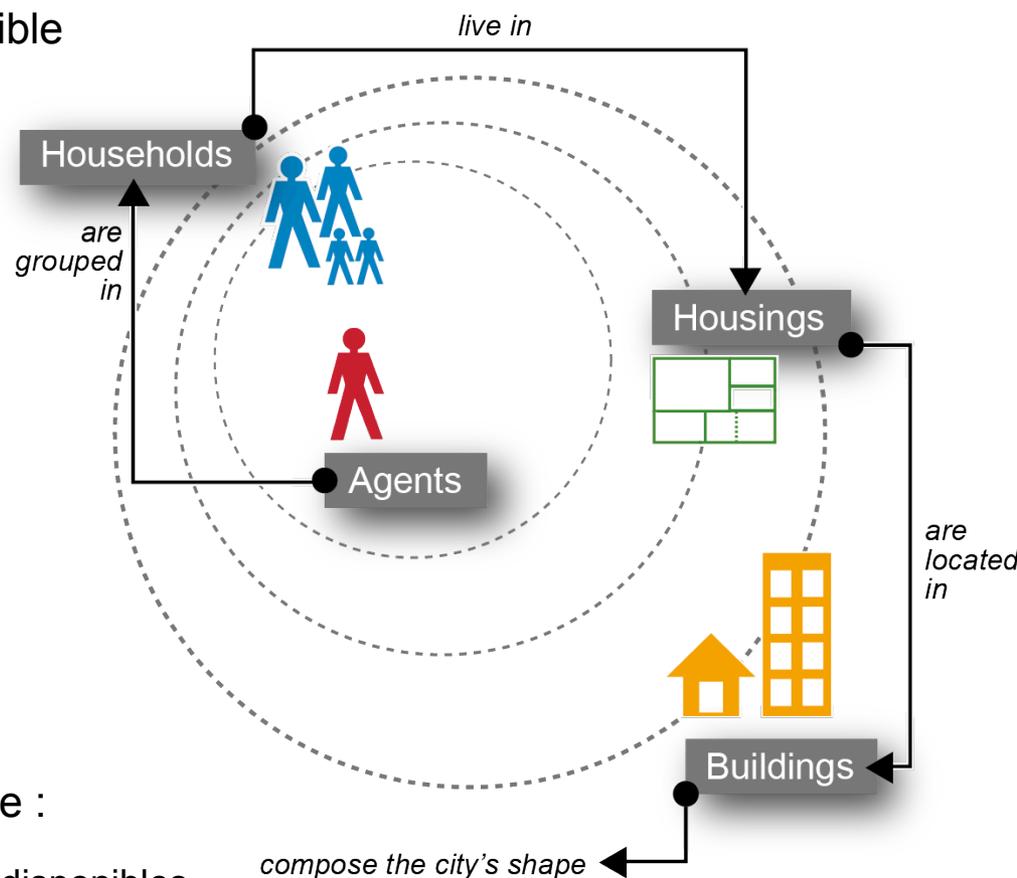
La méthode doit être reproductible
pour chaque agglomération.
Elle a été testée sur trois
villes différentes

- Besançon
- Strasbourg
- Lille

Pour chacune de ces villes,
le schéma conceptuel est
identique :

La méthode doit assurer une
correspondance maximale entre :

- le niveau agrégé des données disponibles
- le niveau synthétique de la population d'agents, de ménages et de logements contenues dans les bâtiments qui composent la forme urbaine



Partie 2

Données disponibles

01. Contexte et état de l'art
- 02. Données disponibles**
03. Méthodologie
04. Résultats
05. Limites et conclusion

Données disponibles

Les données socio-démographiques



Plusieurs jeux de données peuvent être téléchargés sur le site de l'INSEE. Ils regroupent les informations en cinq fichiers essentiels :

- **Population**
Nombre d'habitants par sexe, âge et nationalité (Français, étrangers), population en ménage ou non, etc.
- **Famille**
Population en ménage par âge, personnes seules par âge, par statut marital (seul, marié, veuf, divorcé) ;
- **Formation**
Population scolaire par âge, diplômes par sexe et par âge, etc.
- **Activités**
Population active, type d'activité par sexe, par âge et par catégorie sociale, mode de transport utilisé pour travailler, etc.
- **Logements**
Nombre de logements et statut (principal, secondaire, vacant), nombre de pièces, type (maison, appartement), surface, année d'achèvement, occupation (propriétaire, location, social), etc.

→ Les fichiers sont disponibles au format tableur, sans relation ni géométrie.

Données socio-démographiques Communes et IRIS



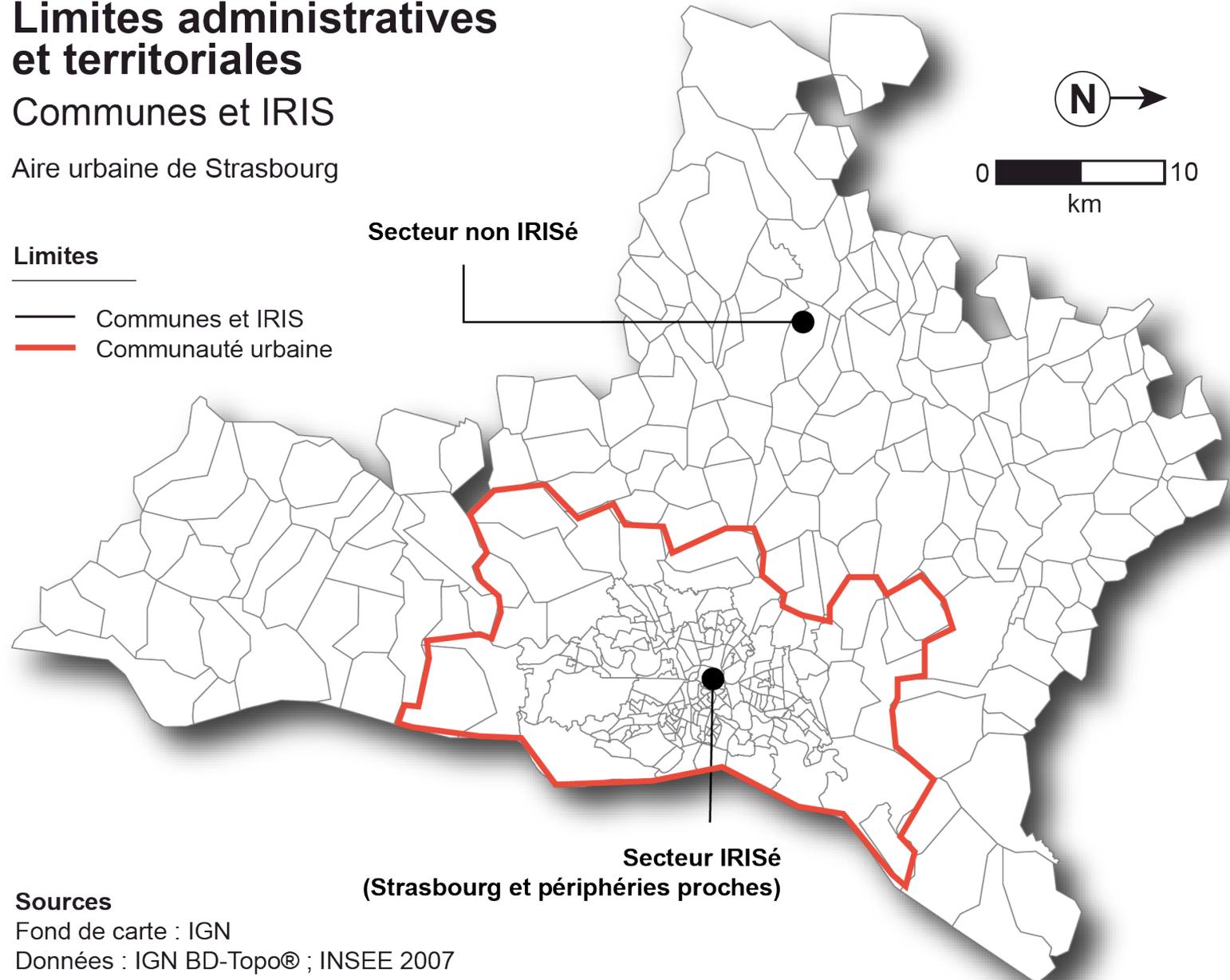
Limites administratives et territoriales

Communes et IRIS

Aire urbaine de Strasbourg

Limites

- Communes et IRIS
- Communauté urbaine



Sources

Fond de carte : IGN
Données : IGN BD-Topo® ; INSEE 2007

G. Rochet, J. Aubé, J.P. Antoni © ThéMA (2011)

Données disponibles

Les données socio-économiques



Certaines informations sur la structure des emplois, les entreprises, les commerces, les administrations, etc. sont disponibles dans le fichier SIRENE (INSEE).

SIRENE indique notamment

- L'activité économique des établissements (NACE Code) |— Z localisation ?
- Leur nombre d'employés
- D'autres informations sur demande
- Leur localisation indiquée par

- Nom de ville
- Nom de rue
- Numéro dans la rue

XY localisation?

Des informations
exploitables « facilement » :

Batch Geocoding
de Google Maps API

Vitrine ?

**Cartographie
tridimensionnelle par points**
de la structure socio-économique

OUI

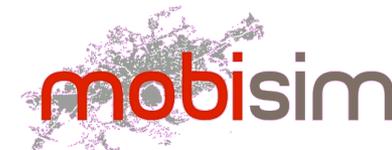
NON

Rez-de-chaussée

Autres étages

Données disponibles

Les données géographiques



Des informations sur les bâtiments et la forme urbaine sont disponibles dans la BD Topo (INSEE) au format SIG (shapefiles)

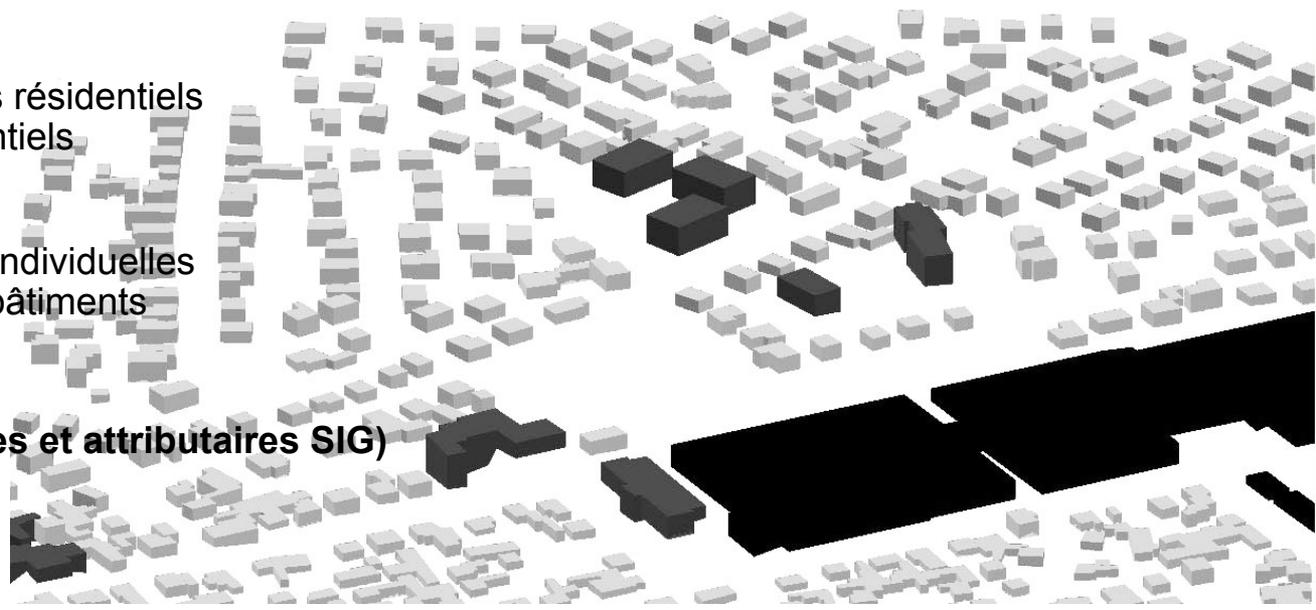
La BD Topo indique notamment :

- Les bâtiments qui composent la forme urbaine
- Leur localisation (XYZ)
- Leur géométrie (forme, taille et hauteur)
- Leur fonction (résidentielle, industrielle, administrative, etc.)

Cette information peut être exploitée « facilement pour distinguer » :

- Les bâtiments résidentiels et non résidentiels
- Les maisons individuelles et les autres bâtiments

(Requêtes spatiales et attributaires SIG)



Données géographiques

Discrimination des maisons individuelles (1/2)

Commune de Strasbourg
(quartiers Meinau et Elsau)

Requête effectuée avec ArcGIS
pour le bâti inférieur ou égal à 10 mètres
de hauteur avec une surface comprise
entre 20 et 500 m².

Batiments considéré comme

- Maisons individuelles
- Immeubles



Sources

Google Maps et Google Streetview (photographies)
Données : IGN BD-Topo® (données vectorielles)
G. Rochet, J. Aubé, J.P. Antoni © Théma (2011)

Données géographiques

Discrimination des maisons individuelles (2/2)



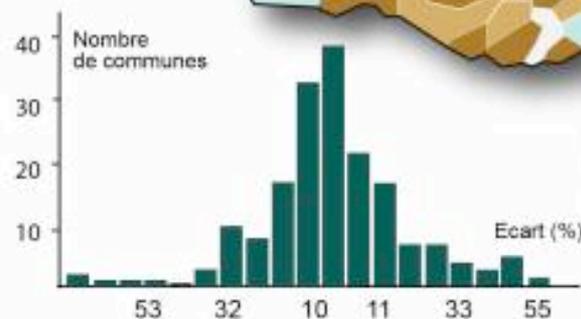
Discrimination des maisons individuelles

Ecart aux données INSEE

Aire urbaine de Strasbourg

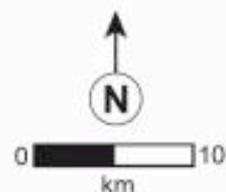
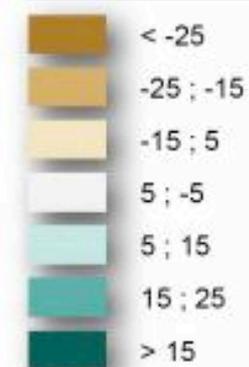
Requête effectuée avec ArcGIS pour le bâti inférieur ou égal à 10 mètres de hauteur avec une surface comprise entre 20 et 500 m².

Histogramme des fréquences



Ecart avec les données INSEE

En pourcentage par commune



Limites territoriales

Aire urbaine —

Communauté urbaine —

Sources

Fond de carte : IGN

Données : IGN BD-Topo® ; INSEE 2007

G. Rochet, J. Aubé, J.P. Antoni

© ThéMA (2011)

Partie 3

Méthodologie

01. Contexte et état de l'art
02. Données disponibles
- 03. Méthodologie**
04. Résultats
05. Limites et conclusion

Méthodologie

Caractérisation des agents et des logements



Le principe général se fonde sur des probabilités conditionnelles et s'exécute en deux temps :

- **L'indentification de probabilités conditionnelles discrètes**
Ces probabilités peuvent être déterminées par croisement des fréquences dans les fichiers de données disponibles
- **Une loi de distribution aléatoire basée sur ces probabilités**
Elle permet d'affecter une valeur particulière à chaque agent à l'intérieur de sa classe

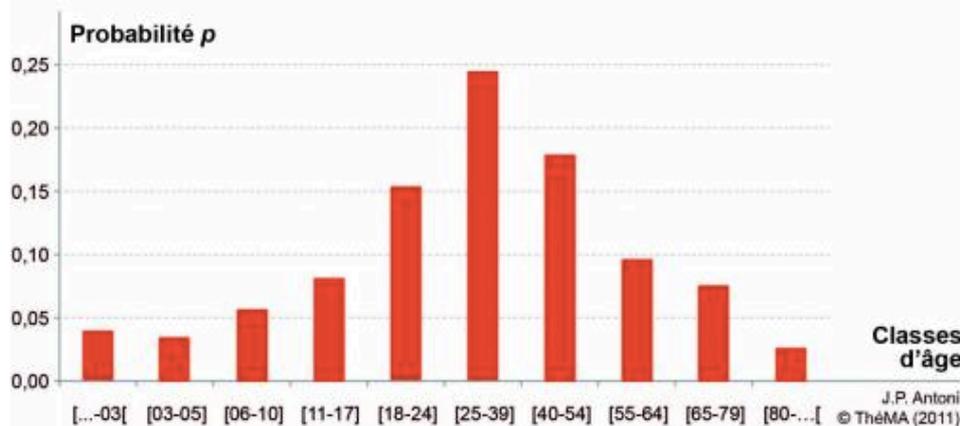
Exemple:

Age	[...-03[[03-05[[06-10[[11-17[[18-24[[25-39[[40-54[[55-64[[65-79[[80-...[Σ
Fréquence n	5186	4760	7532	10452	19843	31554	23298	12544	9984	3444	128598
Probabilité p	0,0403	0,0370	0,0586	0,0813	0,1543	0,2454	0,1812	0,0975	0,0776	0,0268	1

Distribution des probabilités

Classes d'âge de la population masculine

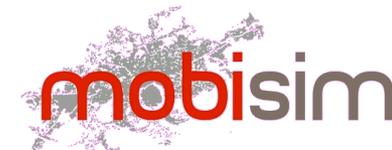
Commune de Strasbourg (Insee, 2011)



- **La méthode est reproduite** pour chacune des caractéristiques des agents et des logements
- Des erreurs (différence entre les données sources et la population synthétique) apparaissent (probabilités conditionnelles)
- **Ces erreurs sont généralement très faibles**

Méthodologie

Principe des affectations agent/ménage/logement



Affectation des agents dans les ménages

● Génération des couples

D'après la littérature et les statistiques de l'INSEE, elle se fonde sur **une minimisation de l'âge et du niveau de formation entre un homme et une femme**

Les couples sont créés jusqu'à atteindre le nombre adéquat d'agents en ménage

● Affectation des logements

Les enfants (< 17 ans) sont affectés aux "couples" ou "familles mono-parentales" selon :

- l'âge de l'enfant et l'âge de primo-nuptialité des mères
- l'âge moyen dans les fratries

Les enfants sont affectés jusqu'à atteindre le nombre adéquat de ménages "avec enfant(s)" et/ou qu'il n'y ait plus d'enfants à affecter

Affectation des ménages dans les logements

● Création des logements dans les bâtiments

- **Les maisons individuelles** sont distinguées des autres bâtiments (requête SIG)
- Les maisons individuelles ne contiennent qu'**un seul logement**
- Le nombre de logements dans les autres bâtiments **est déterminé par leur volume**
- **Le volume non résidentiel** (commerces, activités, etc.) est soustrait au RDC (si vitrine) ou dans les autres étages (pas de vitrine)

● Affectation des ménages dans les logements

- Uniquement si le logement est un logement principal
- Minimisation de la différence entre la taille du ménage et du logement (pondération par le revenu)

Méthodologie

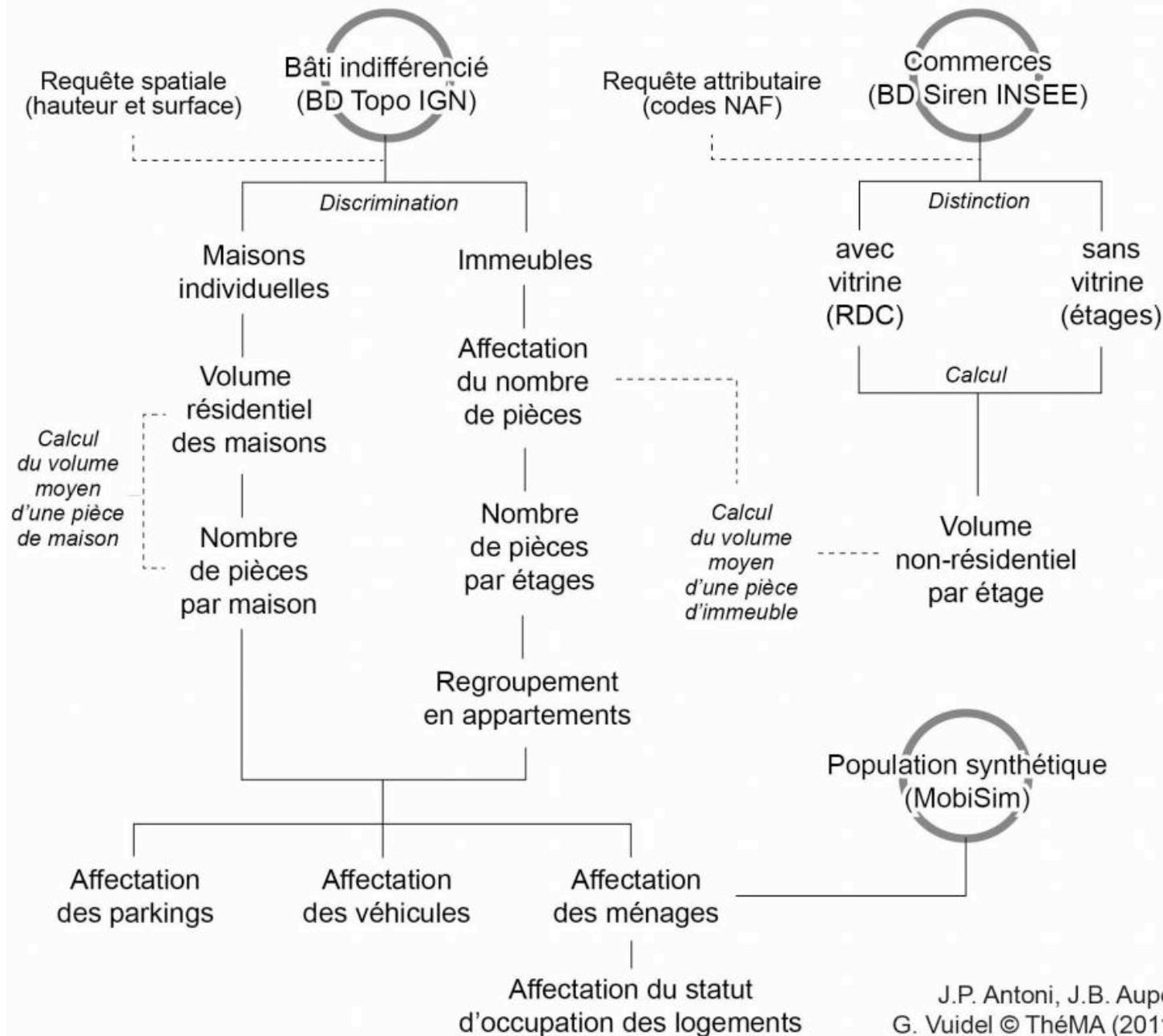
Caractérisation des logements



Module D1
Population
et logements

Avril 2012

Schéma méthodologique de principe



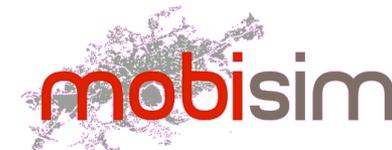
Partie 4

Résultats

01. Contexte et état de l'art
02. Données disponibles
03. Méthodologie
- 04. Résultats**
05. Limites et conclusion

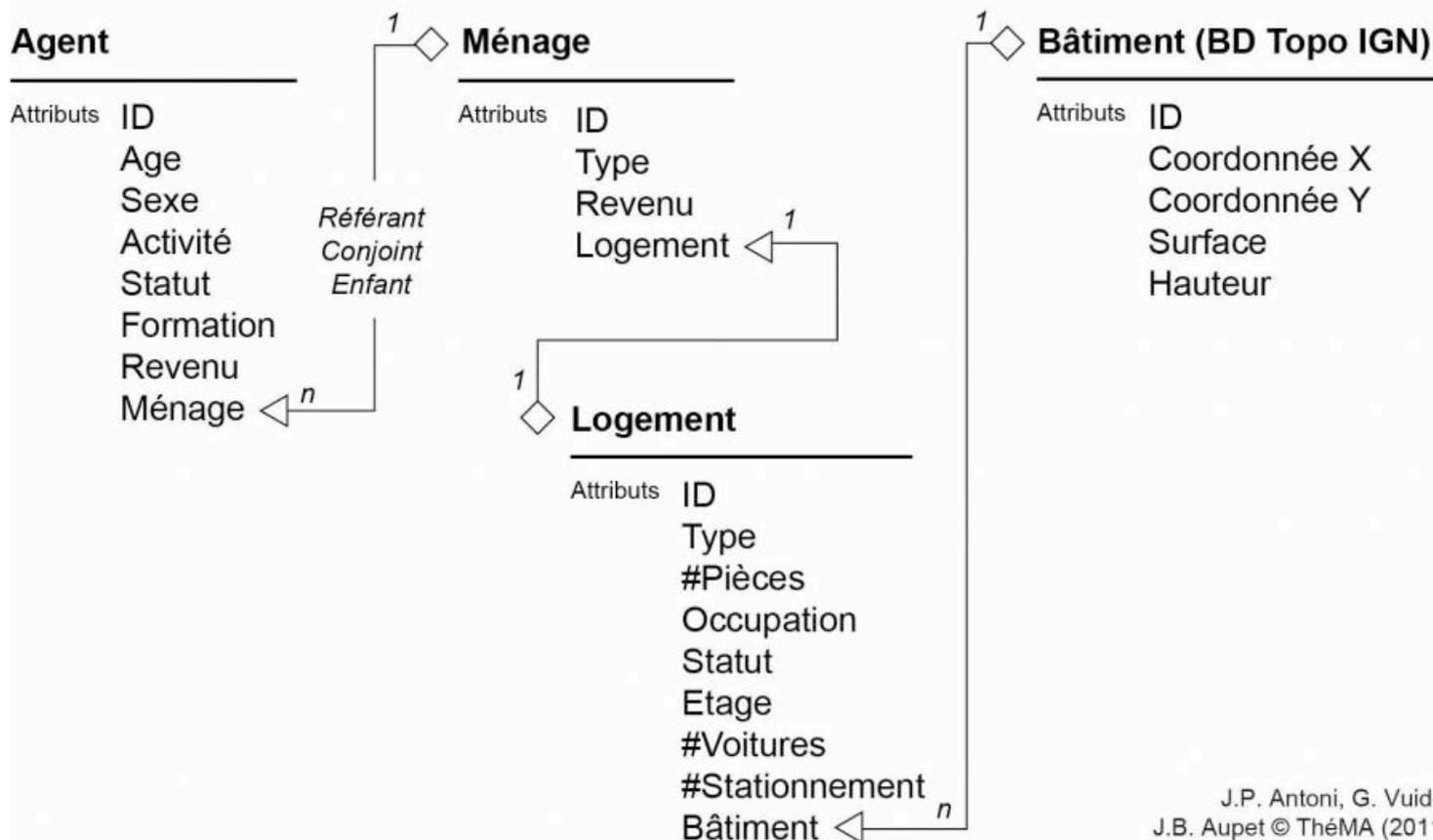
Résultat

Modèle logique (type SGBDR)



Population synthétique localisée

Modèle conceptuel de données



Résultats

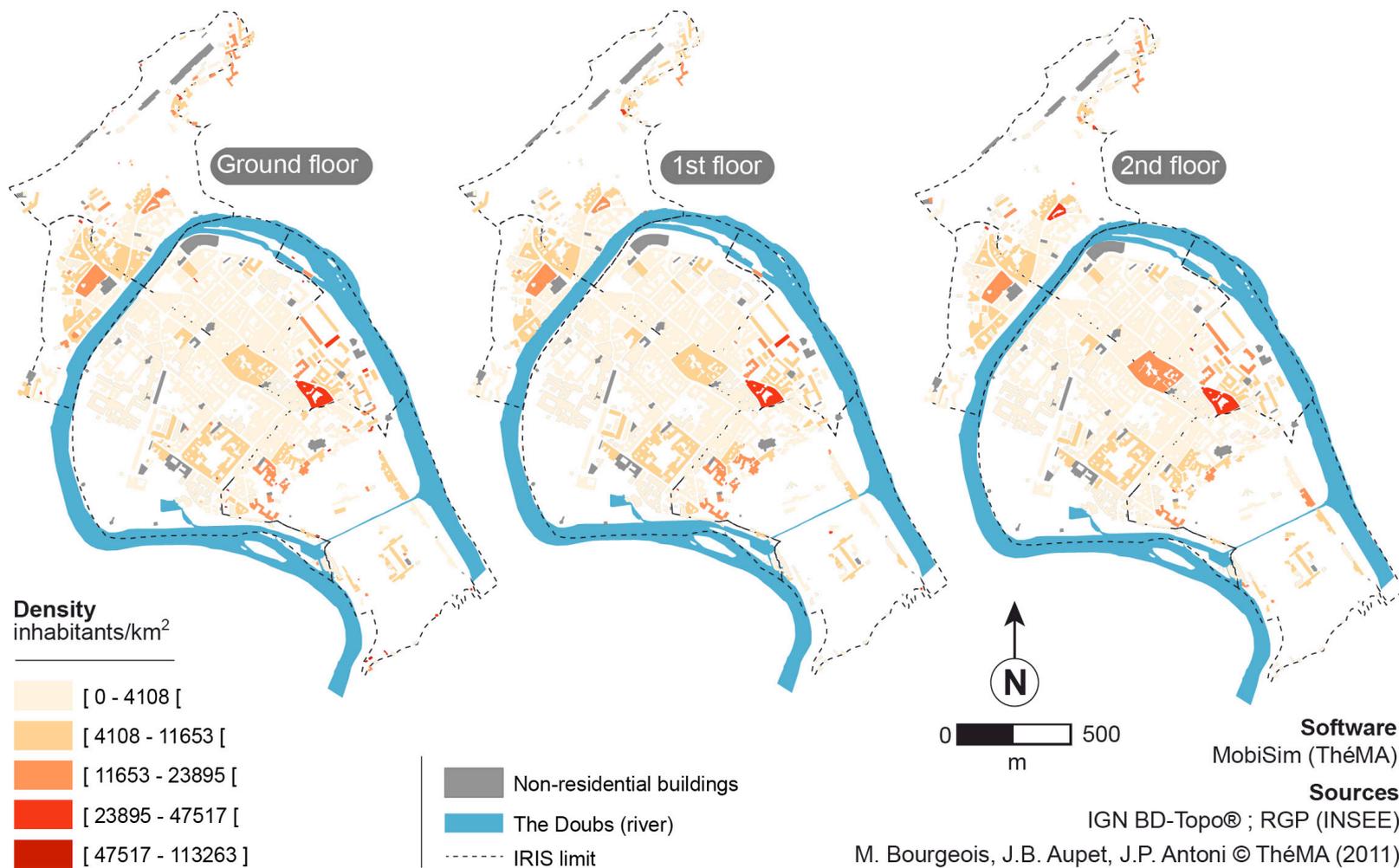
Cartographie de la population étage par étage (1/2)



Generating a synthetic population

Assignment of the agents in the buildings by floor

Downtown and district Battant (Besançon)



Résultats

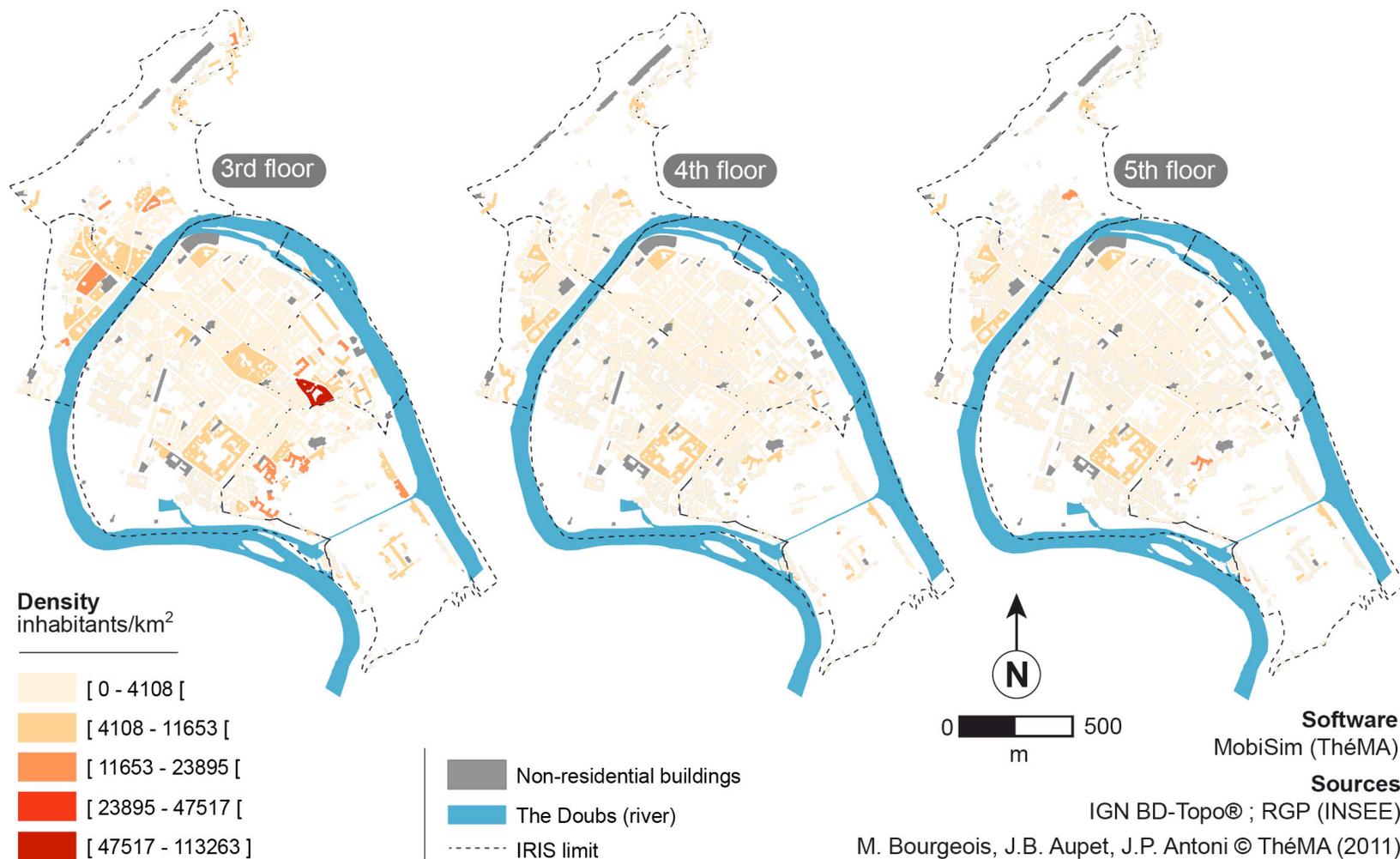
Cartographie de la population étage par étage (2/2)



Generating a synthetic population

Assignment of the agents in the buildings by floor

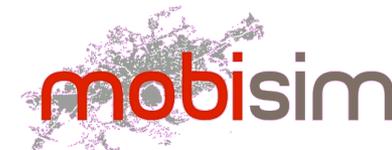
Downtown and district Battant (Besançon)



Avril 2012

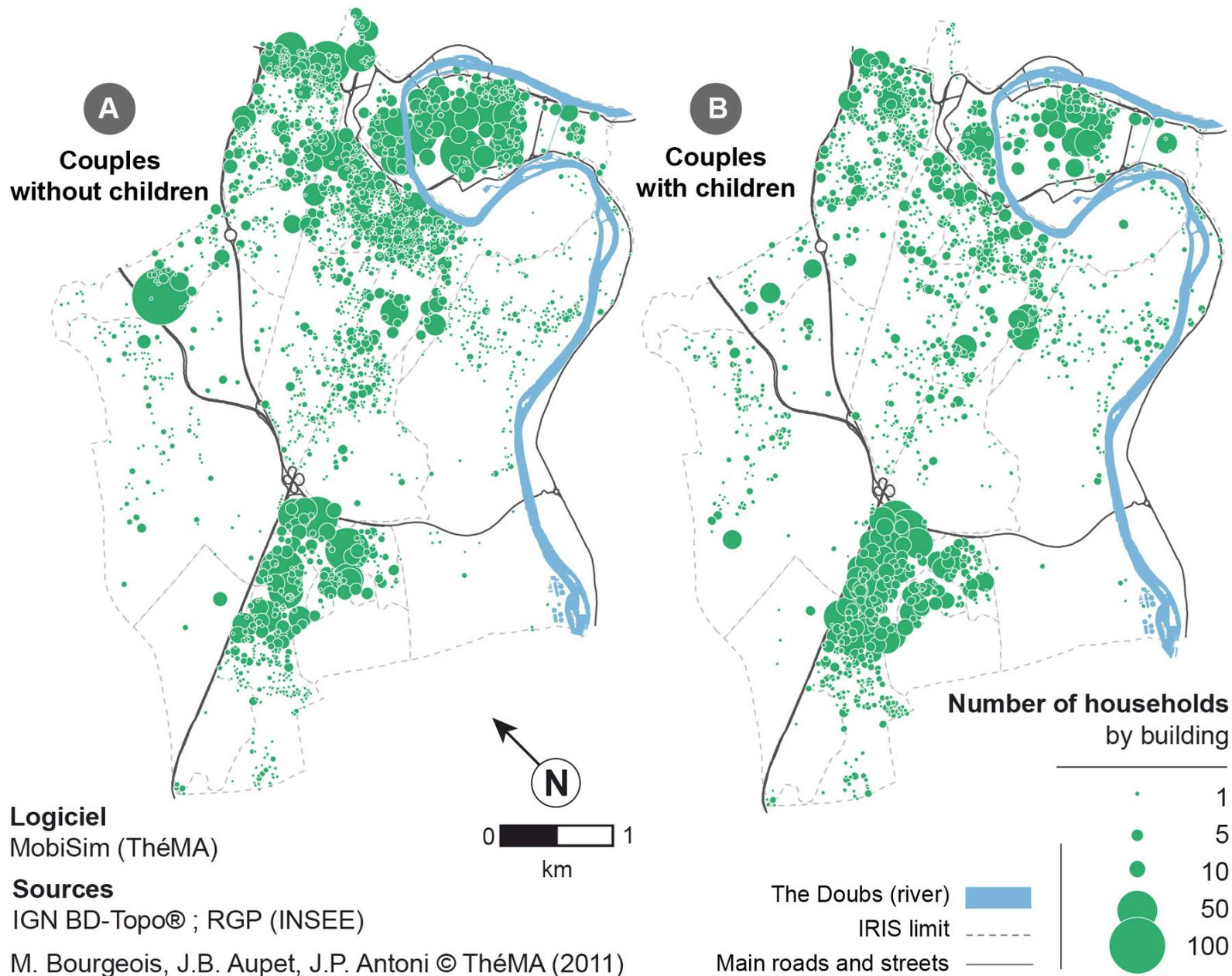
Résultats

Cartographie des structures familiales



Composition of the households

Downtown and quartier Planoise (Besançon)



Résultats

Cartographie des niveaux de formation

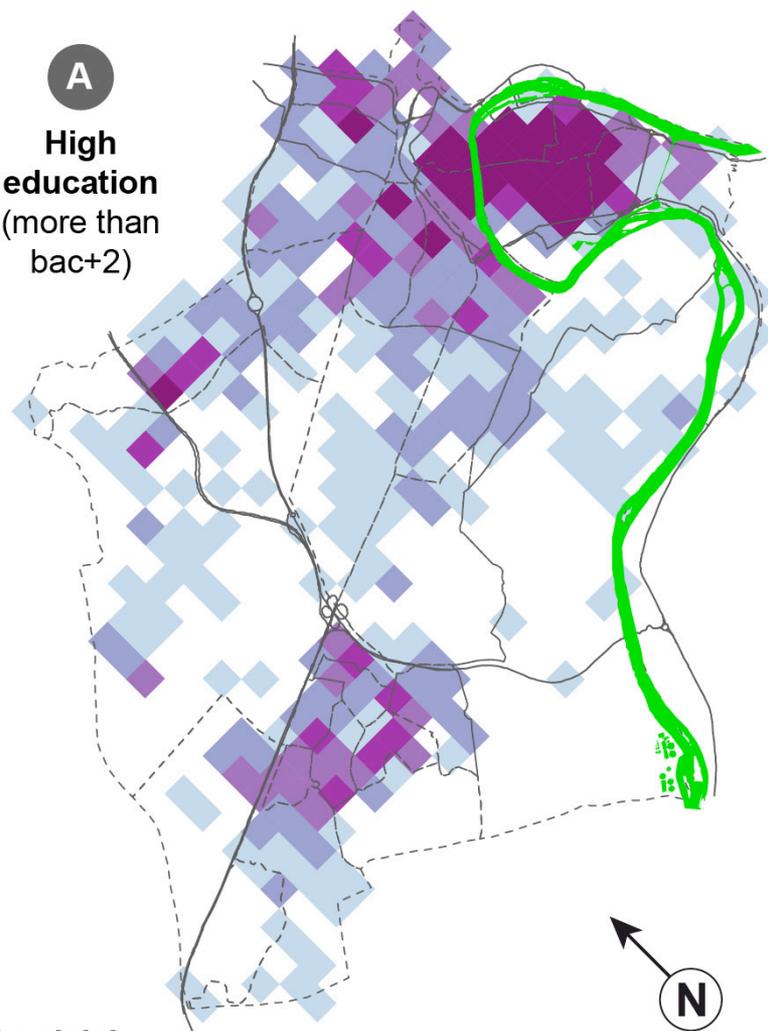
Education of the agents

Downtown and quartier Planoise (Besançon)

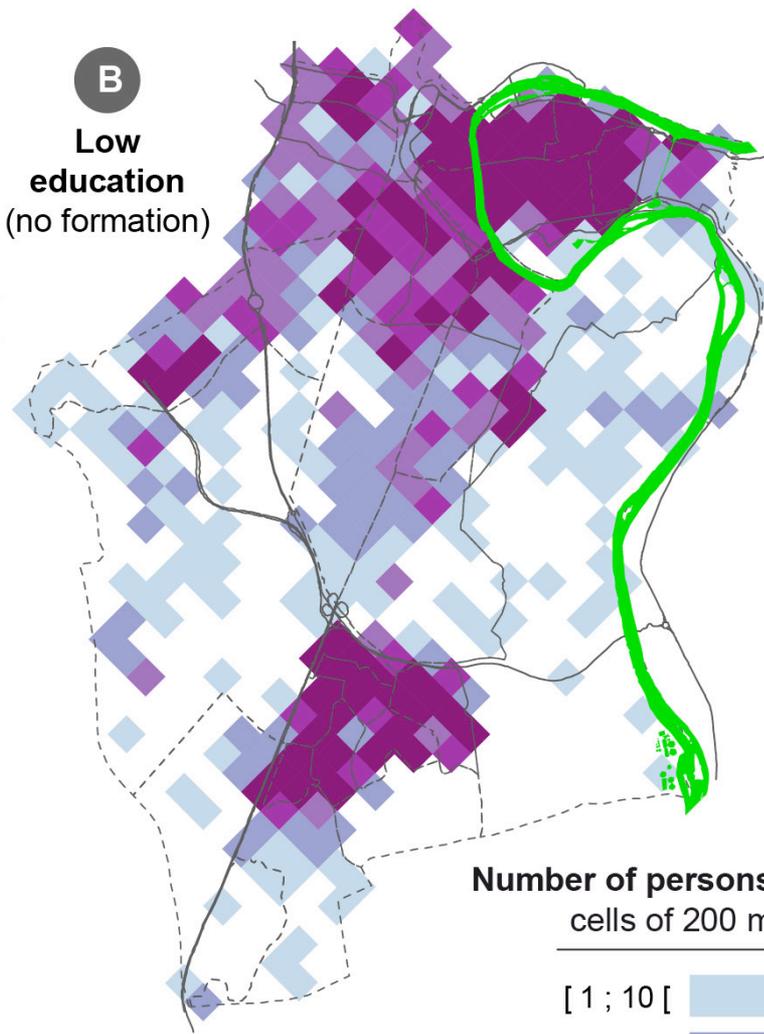
Module D1
Population
et logements

Avril 2012

A
**High
education
(more than
bac+2)**



B
**Low
education
(no formation)**



**Number of persons
cells of 200 m**

[1 ; 10 [

[10 ; 50 [

[50 ; 100 [

[100 ; 200 [

> 200

The Doubs (river)

IRIS limit

Main roads and streets

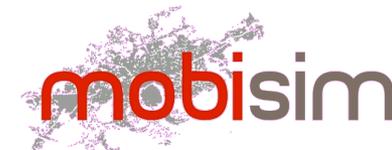
Partie 4

Limites et conclusion

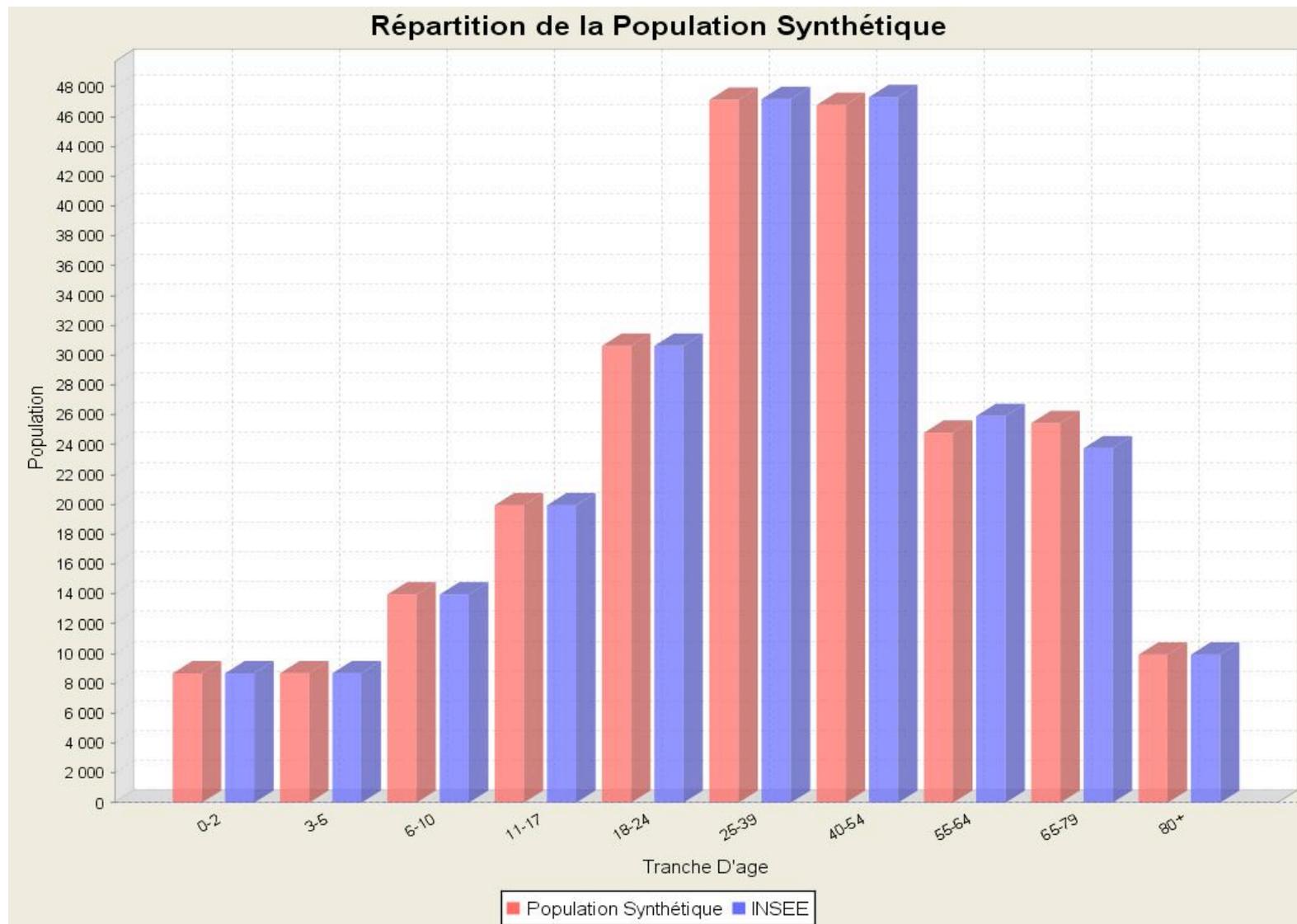
01. Contexte et état de l'art
02. Données disponibles
03. Méthodologie
04. Résultats
- 05. Limites et conclusion**

Limites et conclusion

Des résultats plutôt convaincants (1/4)



Répartition des agents par âge
L'exemple de Besançon (copie de la fenêtre de contrôle MobiSim)



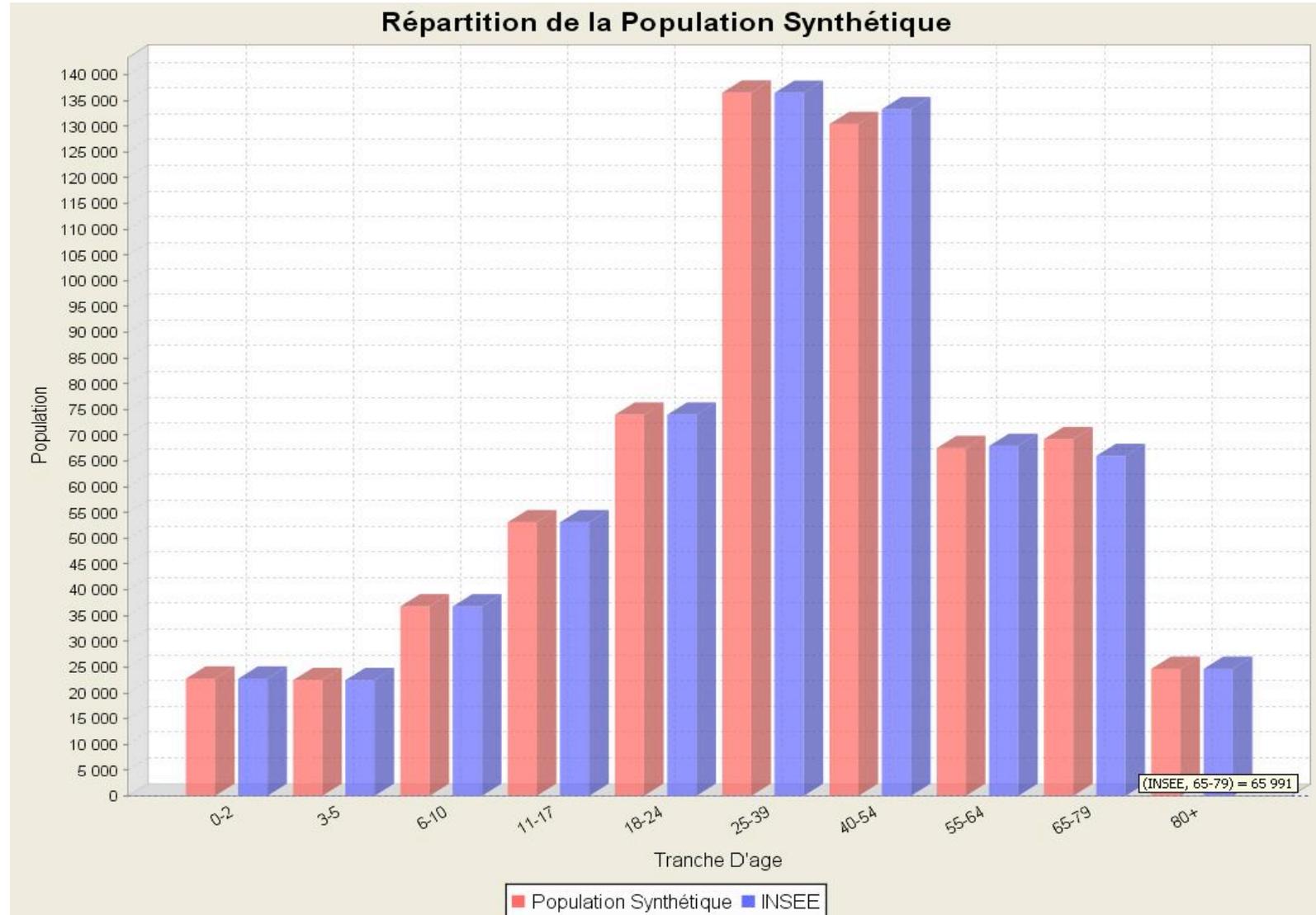
→ Des résultats qui correspondent globalement à la réalité

Limites et conclusion

Des résultats plutôt convaincants (2/4)



Répartition des agents par âge
L'exemple de Strasbourg (copie de la fenêtre de contrôle MobiSim)



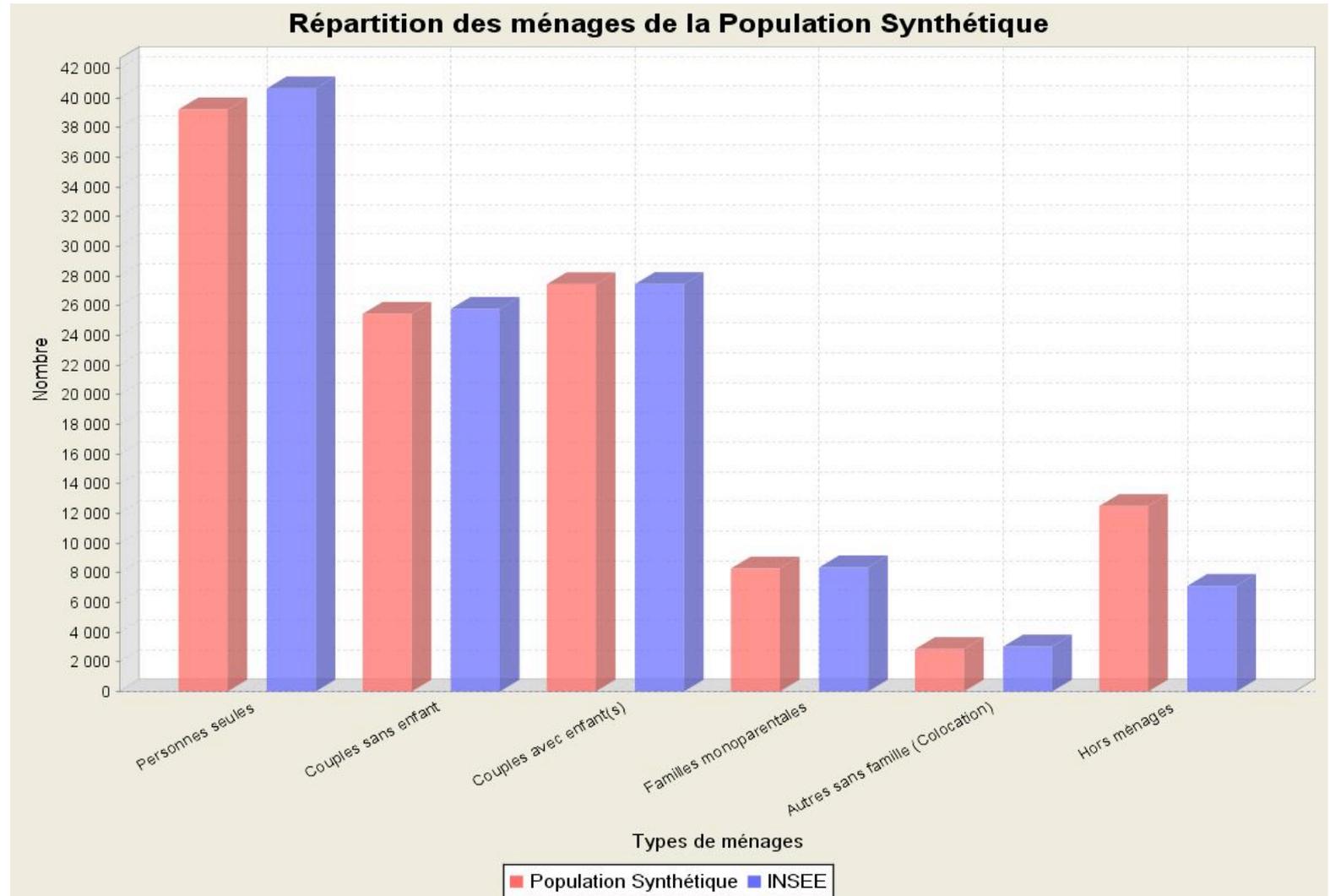
→ Des résultats qui correspondent globalement à la réalité

Limites et conclusion

Des résultats plutôt convaincants (3/4)



Répartition des ménages par type
L'exemple de Besançon (copie de la fenêtre de contrôle MobiSim)



→ Des résultats qui posent question dans certains cas

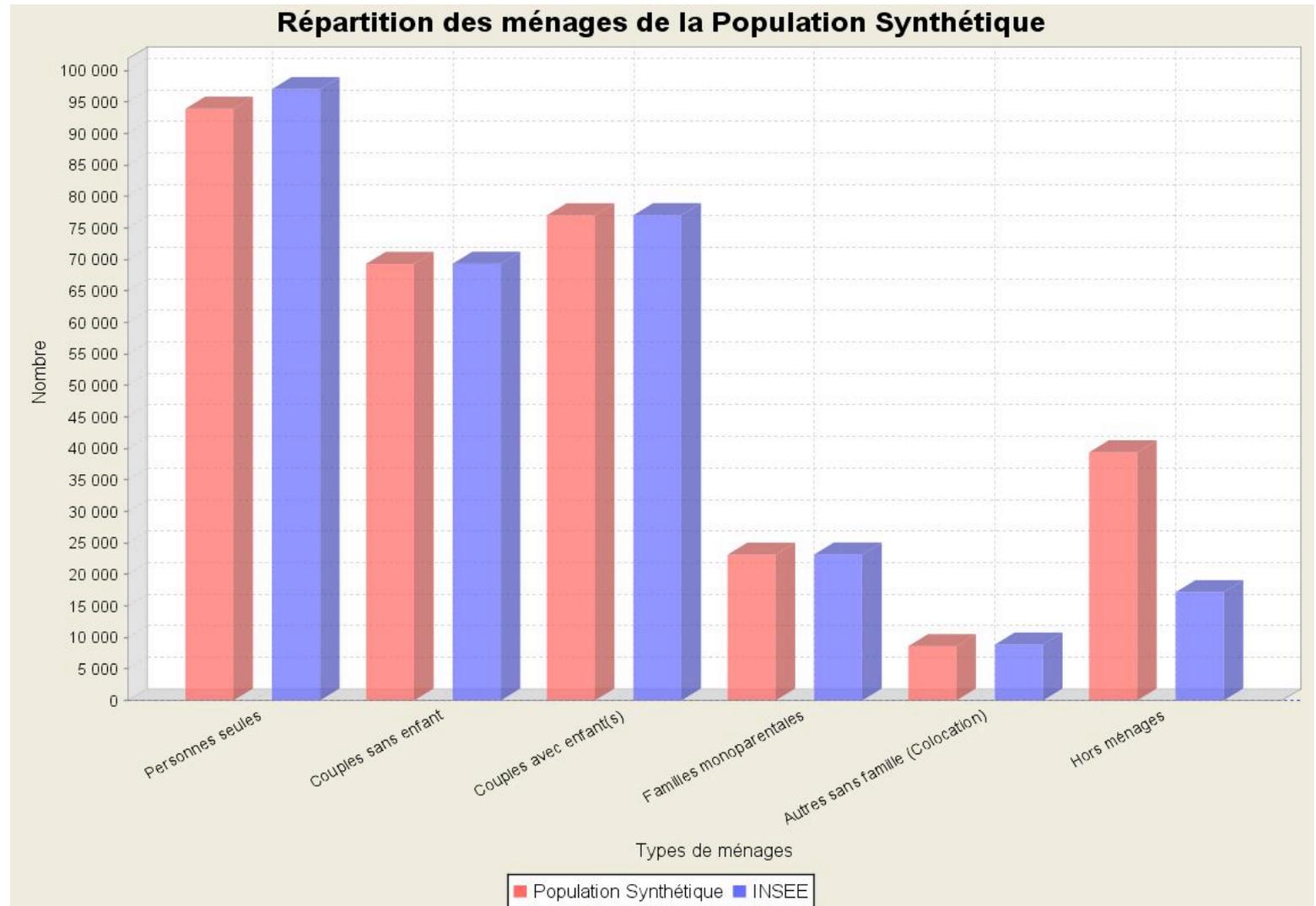
Avril 2012

Limites et conclusion

Des résultats plutôt convaincants (4/4)



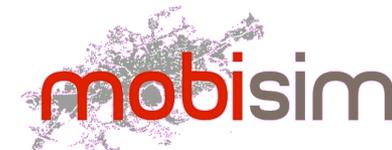
Répartition des ménages par type
L'exemple de Strasbourg (copie de la fenêtre de contrôle MobiSim)



→ Des résultats qui posent question dans certains cas

Limites et conclusion

Un intérêt pour la modélisation désagrégée



Malgré une approximation nécessaire dont les écarts à la réalité sont difficilement quantifiables au niveau désagrégé, la méthode offre de nouvelles perspectives

Une méthode limitée par la précision des données d'entrée

- Les données fournies par les sources officielles (IGN, INSEE, ...) ne sont pas toujours cohérentes : on observe des différences **d'un fichier à un autre**
- **Les dates des données** (recensements, relevés, etc.) ne sont pas homogènes. Des erreurs apparaissent notamment quand les données démographiques et géographiques sont croisées
- **La généralisation des données géographiques** diffère d'une ville à l'autre. Des erreurs apparaissent durant la caractérisation des bâtiments
- **Les retraités et les étudiants** ne sont pas clairement identifiés dans les données de départ. Des erreurs apparaissent durant l'affectation dans les ménages

Un nouveau cadre pour la modélisation LUTI individu-centrée

- **Les données désagrégées** peuvent être introduites dans un modèle de trafic ; mais ce modèle (quatre étapes par exemple) doit être modifié
- Il peut être couplé avec une approche « **activity-based** »
- **Les déplacements courts** peuvent être introduits dans le modèle
- **Les comportements individuels** peuvent également être introduits