

# La téléphonie mobile un marqueur de présence

---

*Dans ce chapitre est présenté un nouveau modèle de répartition dynamique de la population à partir de données téléphoniques. Nous tentons de répondre à un besoin toujours plus marqué de modéliser finement les déplacements de la population. Cette modélisation constitue l'élément essentiel pour permettre d'analyser et comprendre les besoins en services de mobilité. L'utilisation de la téléphonie comme marqueur de présence apparaît alors comme une nouvelle approche, permettant d'imager autrement la distribution spatio-temporelle de population. La première partie de ce chapitre est destinée à présenter de façon détaillée le modèle proposé. L'accent sera mis sur la façon dont s'articulent dans ce modèle les données téléphoniques d'une part, et la structure du territoire d'autre part, à travers les données géographiques et socio-économiques. Dans la deuxième partie du chapitre, une phase de validation étayée sur un cas d'étude réel est présentée associant une collectivité et un opérateur téléphonique, une force pour cette étude.*

## Sommaire

---

<b>2.1</b>	<b>Introduction</b>	<b>40</b>
<b>2.2</b>	<b>La téléphonie mobile : une nouvelle source de données</b>	<b>41</b>
2.2.1	Structure et fonctionnement d'un réseau cellulaire	41
2.2.2	Quelles données sont disponibles?	45
<b>2.3</b>	<b>Un nouveau modèle de distribution de la population</b>	<b>47</b>
2.3.1	Mobilité et caractérisation socio-économique du territoire	48
2.3.2	Modèle de distribution spatio-temporelle des mobiles	55
<b>2.4</b>	<b>Résultats et analyse</b>	<b>58</b>
2.4.1	Environment d'analyse	58
2.4.2	Constitution des données de référence	60
2.4.3	Expérimentation et résultats	64
2.4.4	Construction d'une <i>Signature Territoriale</i>	70
2.4.5	Généralisation de la <i>Signature Territoriale</i>	73
<b>2.5</b>	<b>Synthèse</b>	<b>73</b>

---

## 2.1 Introduction

Dans le souci de fournir des informations de plus en plus précises et à un coût maîtrisé, de nouvelles sources d'information sont à inventer pour localiser des individus et identifier les flux de mobilité au fil du temps. Une meilleure prise en compte des déplacements des habitants d'un territoire doit ainsi permettre de déployer des offres de transports en commun de meilleure qualité, ou encore d'aménager des voiries dédiées aux cyclistes, de construire des stations de vélo ou de voitures partagées. A l'heure où les voitures électriques constituent une nouvelle façon de se déplacer, de grands investissements en matière d'infrastructures de recharges sont à prévoir. Pour accompagner toutes ces évolutions, il apparaît indispensable d'améliorer notre connaissance de la mobilité des personnes.

Les téléphones portables constituent sans aucun doute une source de données avérée pour un retour informatif sur les déplacements de personnes. Les données de réseau de téléphonie mobile : appels entrants et sortants pour chaque cellule, les transferts intercellulaires (handover), etc. fournissent de nouveaux marqueurs de la mobilité. Leur pertinence est d'autant plus forte que le téléphone portable est devenu en quelques années omniprésent [Rheingold 2002]. L'International Telecommunication Union (ITU), agence des nations unies en charge des Technologies de l'Information et de la Communication (TIC) indique que le nombre d'abonnés dans le monde est passé de moins d'1 milliard à plus de 5 milliards au cours de ces dernières années. La figure 2.1 représente cette tendance au niveau mondial. Par ailleurs, toujours selon l'ITU, en 2010 90% de la population mondiale était couverte par un signal cellulaire contre 61% en 2003. Nous assistons bien à une évolution rapide qui a des répercussions fortes sur notre vie quotidienne : travail, loisirs, déplacements, etc. [Campbell 2008]

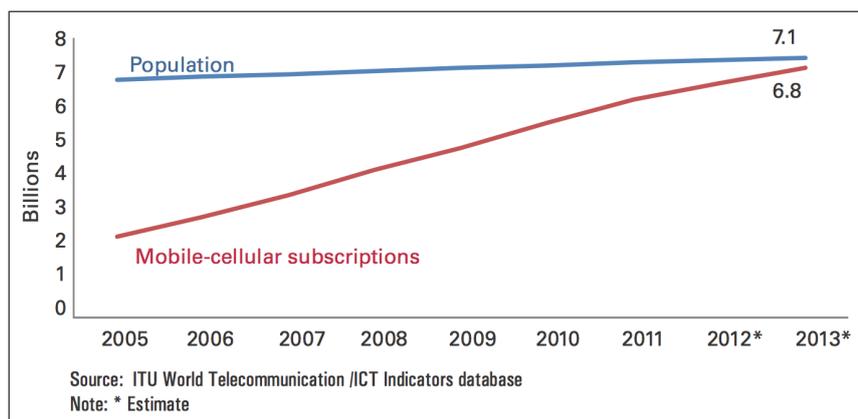


FIGURE 2.1 – Evolution du nombre d'abonnements téléphoniques dans le monde

Alors que jusqu'à présent le déploiement des réseaux mobiles s'appuyait sur une

connaissance des déplacements de la population, les téléphones portables peuvent maintenant également améliorer la compréhension sur la façon dont les personnes se déplacent. La téléphonie sans fil devient alors un traceur de mobilité, disponible tout au long de la journée, tous les jours de la semaine, pour des jours travaillés ou de vacances.

Dans ce chapitre nous proposons d'élaborer une caractérisation spatio-temporelle du terrain basée sur une analyse statistique des données de déplacements et l'usage de la téléphonie mobile. Un modèle de distribution de la population "multicritère" est ainsi présenté. Il tient compte à la fois de critères géographiques, sociologiques, événementiels, et temporels. Ce modèle est complété par un tableau de bord permettant d'imager la répartition dynamique des individus à travers une carte géographique. Nous visons à fournir un nouveau moyen d'appréhender la mobilité tout au long de la journée. Un cas d'étude a permis de valider la modélisation sur des données réelles appartenant au département du Territoire de Belfort, en France, et à la société Orange.

Enfin, on notera que le modèle que nous proposons a servi à alimenter divers travaux de recherche, pour simuler le déplacement d'individus afin de valider des protocoles de communication inter-véhiculaires ou véhicule à infrastructure [Ait Ali 2010a] [Ait Ali 2010b], pour élaborer un nouveau modèle de simulation de la mobilité [Joumaa 2009a] [Joumaa 2009b] ou encore pour proposer une nouvelle approche de configuration des réseaux cellulaires eux-mêmes, pour le développement de la 4G notamment [Chariete 2013].

La section suivante nous permet de présenter les éléments clés du fonctionnement d'un réseau de téléphonie mobile et d'identifier les éléments disponibles pour une modélisation de la mobilité. Ensuite, la section 2.3 est consacrée à la présentation de notre modèle, basé en premier lieu sur la téléphonie mobile mais intégrant également une connaissance fine de la caractérisation spatio-temporelle du terrain. Enfin, la troisième section fournit une analyse des résultats dans un contexte réel, résultats qui ont été validés dans le cadre du projet *Territoire Mobile*, labellisé par le pôle de compétitivité Véhicule du Futur, de 2006 à 2009.

## 2.2 La téléphonie mobile : une nouvelle source de données

### 2.2.1 Structure et fonctionnement d'un réseau cellulaire

La principale caractéristique d'un système radio mobile est de permettre aux personnes de se déplacer librement tout en assurant une continuité de service sur l'ensemble du territoire couvert. Pour répondre à ce besoin l'opérateur répartit des stations de base (émetteurs radio) sur le territoire à couvrir par lesquelles les appels

ou les données pourront être transférés vers les mobiles. Lorsqu'un mobile établit une communication, un événement de localisation est généré sur la station de base qui gère la communication.

**Définition 1 (Station)** *Une station, également appelée station de base, est un relai radio permettant d'acheminer toute forme de communication avec un mobile. A une station est toujours associée une zone de couverture qui dépend des caractéristiques de la station (paramétrage de l'antenne) et de l'environnement (relief, construction, etc.).*

**Définition 2 (Site)** *Un site est le lieu géographique repéré par ses coordonnées où est placée une station. Un même site peut regrouper plusieurs stations permettant de couvrir des orientations différentes, de proposer diverses technologies (UMTS, GSM) et éventuellement plusieurs opérateurs.*

**Définition 3 (Cellule)** *On appelle cellule la zone de couverture radio d'une station. Il s'agit de la portion de territoire sur laquelle les communications peuvent être acheminées par la station associée.*

Pour permettre aux personnes d'être continuellement connectés lorsqu'elles se déplacent, la majeure partie du territoire est couverte par plusieurs cellules. Afin de calculer, visualiser et gérer les recouvrements entre cellules, les opérateurs utilisent des méthodes et des logiciels de planification de réseaux. En particulier, ils ont recours à des modèles de propagation des ondes radio leur permettant de déterminer en chaque lieu quelles sont les antennes susceptibles d'acheminer une communication, et avec quelle probabilité. C'est ce que nous appellerons la probabilité de prise de communications (PPC)<sup>1</sup>. La Figure 2.2 montre sur un schéma théorique comment ces superpositions cellulaires opèrent, avec sous forme maillée les PPC de trois antennes.

La Figure 2.3 est une illustration de l'enchevêtrement réel de cellules où chaque couleur représente une cellule et l'intensité de la couleur est le reflet de l'intensité de la probabilité de prise de communication (PPC) de la station.

Lorsqu'un mobile se déplace, il détecte successivement les stations de base en fonction de la couverture de chacune d'elles, de leur niveau de saturation et de paramètres sur les antennes qui gèrent l'admission des terminaux. Ce déplacement a un effet réduit sur le réseau si le téléphone est en veille (pas d'appel, pas de transfert de données) car celui-ci fait une mise à jour de la localisation par groupe de cellules (Location Area - LA). Par contre dans le cas où le téléphone est en cours de communication, tout changement de cellule implique un dialogue avec le

---

1. Ce concept est la propriété d'Orange et ne sera pas détaillé dans cette thèse.

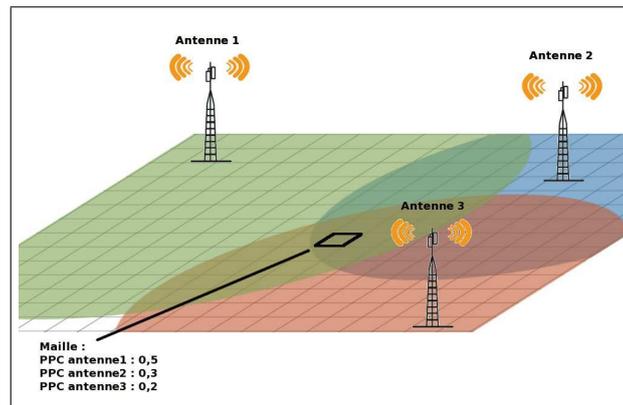


FIGURE 2.2 – Maillage et PPC

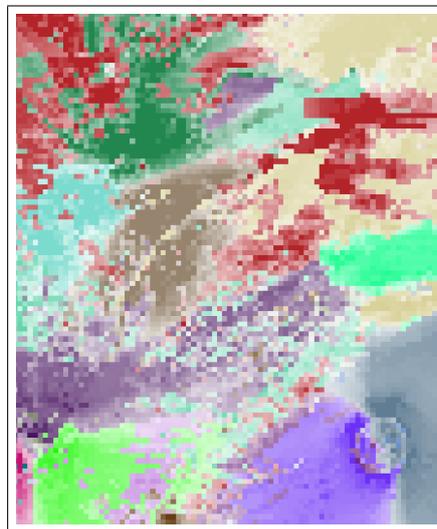


FIGURE 2.3 – Couverture radio

réseau pour changer la connexion du terminal d'une station à une autre (transfert intercellulaire ou Handover).

**Définition 4 (Handover)** *On appelle Handover ou Handoff (HO) le transfert intercellulaire d'un mobile.*

Plusieurs événements sont générés sur le réseau lorsqu'un mobile en cours de communication se déplace de façon à le raccrocher à la station la plus adaptée au contexte. Ainsi un terminal peut être tracé au sein de plusieurs cellules en fonction de sa mobilité en cours d'appel et générer un ensemble d'événements sur le réseau. Les paragraphes suivants illustrent ce phénomène au travers d'un exemple en deux étapes :

1. La personne 1 initie, depuis la cellule A, un appel à la personne 2 présente dans la cellule B (Figure 2.4). Un appel sortant est comptabilisé pour la station ou antenne A, et un appel entrant est comptabilisé pour la station B.

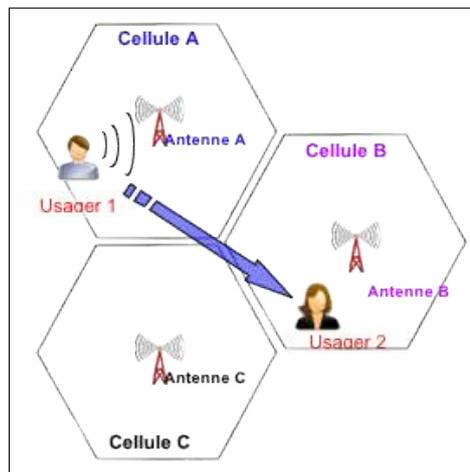


FIGURE 2.4 – Génération d'un événement de communication

2. La personne 1 se déplace alors, pendant la communication, dans la cellule C. Le mobile de la personne 1 doit donc opérer un transfert intercellulaire (Figure 2.5). Ce type d'événement fait comptabiliser un HO sortant pour la cellule A et un HO entrant pour la cellule C.

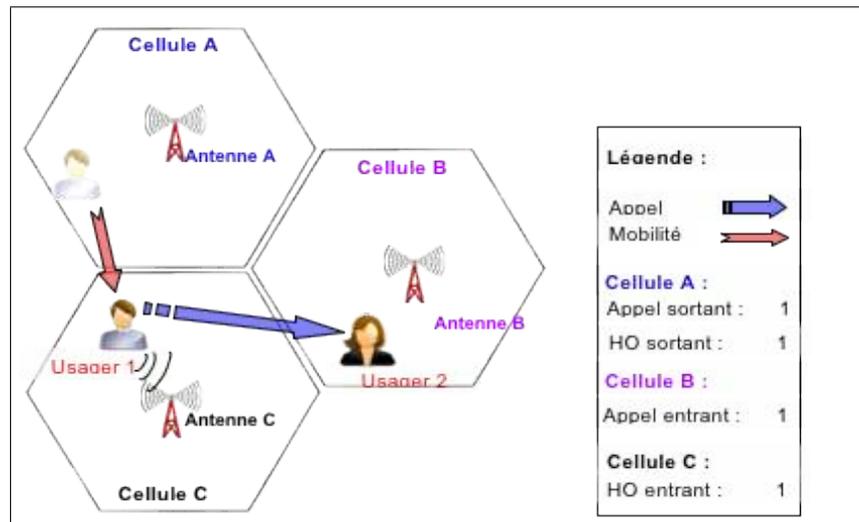


FIGURE 2.5 – Génération d'un événement de transfert intercellulaire

Le HO permet de localiser la présence de personnes, par contre ce n'est pas un indicateur de la direction des flux. Lorsqu'un HO entrant est comptabilisé pour une cellule, nous pouvons affirmer qu'une personne est entrée dans celle-ci, mais nous ignorons sa provenance. De même, si nous comptabilisons un HO sortant pour une cellule, nous pouvons affirmer qu'une personne a quitté cette cellule, mais nous ignorons sa destination. Bien entendu, le réseau identifie le besoin de transfert intercellulaire mais cette information n'est pas stockée, seuls les événements anonymes d'entrée ou de sortie de cellules sont archivés.

Nous ne détaillerons pas ici les mécanismes impliqués dans le choix de la station de rattachement, ni dans le déclenchement d'un transfert intercellulaire. Le lecteur souhaitant plus de détails pourra se référer au livre [Tabbane 2002].

### 2.2.2 Quelles données sont disponibles ?

Tout ce qui se passe sur un réseau cellulaire n'est pas enregistré pour des questions de technologie, de volume de données, de stratégie opérateur, de protection des libertés (CNIL), etc. Pour des raisons de limitation de ressources sur le lien radio, il n'est pas possible de connaître à tout moment la station de rattachement de chaque téléphone. Toutefois les opérateurs captent et conservent les événements suivants :

- les communications entrantes (information de communications) : ce sont les appels/SMS/transfert de données reçus par la personne se trouvant dans la cellule, c'est un marqueur de présence.
- les communications sortantes (information de communications) : ce sont les

appels/SMS/transfert de données émis par la personne se trouvant dans la cellule, c'est un marqueur de présence.

- les "handover" (information de signalisation) : c'est un marqueur de mobilité étant donné qu'il s'agit de communications ayant débuté dans une cellule et qui se poursuivent dans une autre cellule.

Les données de trafic et de signalisation se présentent en général sous la forme d'une matrice individus/variables où les individus sont les stations, caractérisées par leur identifiant, assurant la couverture de la zone étudiée, et les variables sont les événements. Dans notre travail sur l'utilisation de ces données pour modéliser la mobilité, nous avons regroupé les événements par 1/4 d'heure de 6h00 à 23h45, soit 71 plages horaires. Ce pas est tout à fait ajustable pour augmenter ou diminuer la précision temporelle du modèle. Voici la liste des données utilisées par le modèle que nous proposons :

- Le nombre d'appels sortants. Celui-ci indique le nombre de personnes se trouvant dans la cellule  $c$  durant le quart d'heure  $t$  ayant initié une communication depuis leur téléphone mobile.
- Le nombre d'appels entrants. Celui-ci indique le nombre de personnes se trouvant dans la cellule  $c$  durant le quart d'heure  $t$  ayant reçu une communication sur leur téléphone mobile.
- Le nombre d'appels total = le nombre d'appels entrants + le nombre d'appels sortants.
- Les handover sortants. Ils indiquent le nombre de personnes quittant la cellule  $c$  pour une cellule voisine durant le quart d'heure  $t$ .
- Les handover entrants. Ils indiquent le nombre de personnes entrant dans la cellule  $c$  depuis une cellule voisine durant le quart d'heure  $t$ .
- La somme et la différence des handover sortants et entrants. Ces calculs seront utiles pour la détection de pôles plus ou moins attracteurs sur la zone géographique étudiée.

De plus, pour faire le lien entre les communications et le territoire, nous utiliserons une matrice de couverture radio pour l'ensemble des antennes du territoire étudié. Les données de couverture radio utilisées ici sont des données maillées à un pas de 25m. Cette matrice contient pour chaque maille la valeur de puissance reçue et provenant de l'antenne considérée. Elles permettent de localiser les zones couvertes par chaque antenne et de connaître la PPC associée. Enfin cette matrice permet par superposition de faire le lien entre les cellules et les données géographiques (parcelles, communes, bâtiments, routes) du territoire.

## 2.3 Un nouveau modèle de distribution de la population

Dans cette section nous présentons notre approche permettant de modéliser les déplacements, qui s'appuie sur trois éléments essentiels :

- l'utilisation des réseaux téléphoniques sans fil
- les caractéristiques du sursol
- les objectifs de déplacements des personnes

Ce modèle a été mis au point et expérimenté dans le cadre d'un projet labellisé par le Pôle de Compétitivité "Véhicule du Futur" [Lamrous 2008] avec comme partenaire le SMTC Belfort, la société Orange, l'Agence d'Urbanisme et les collectivités territoriales du Territoire de Belfort. La vocation du projet est de proposer une modélisation et une vision via un logiciel informatique des déplacements de l'ensemble de la population d'un territoire afin de planifier des services de transport en commun. Cette approche constitue ainsi une nouvelle façon d'appréhender l'environnement urbain. Ce modèle se place dans la catégorie des modèles basés sur l'activité couplé avec des données issues des réseaux de téléphonie cellulaire.

Afin de permettre de visualiser les résultats, nous avons développé la plateforme logicielle "geoLogic" proposant un tableau de bord ayant pour fond une carte géographique (cf. Figure 2.6) à laquelle se superpose plusieurs calques de couleur pour scruter les déplacements des individus. Ce tableau de bord permet de visualiser les déplacements sur le territoire couvert sur une journée complète au quart d'heure près et à un pas de 25 mètres. Tous les moyens de zoom et de sélection sont disponibles à l'utilisateur. Une présentation plus détaillée de geoLogic est proposée dans le chapitre 4. Nous utiliserons des sorties graphiques de geoLogic pour illustrer des composants du modèle de mobilité de ce chapitre.

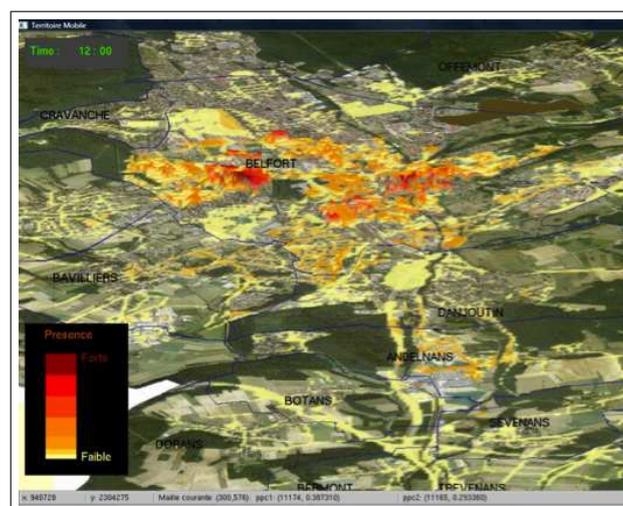


FIGURE 2.6 – Cartographie temporelle de la distribution de la population belfortaine

Dans les projets issus de la littérature, le niveau de précision est celui de la couverture radio, or ces tailles peuvent varier de façon significative. Alors que dans un contexte urbain dense, elles peuvent couvrir un rayon de 300m, elles peuvent hors agglomération s'étendre à des rayons de plusieurs kilomètres. Selon les besoins, ce niveau de découpage du territoire peut être suffisant, mais dans beaucoup de situations un niveau plus fin est nécessaire. L'introduction ici d'une connaissance de la topographie permettra d'estimer de façon probabiliste comment les mobiles se répartissent à l'intérieur des zones de couverture des antennes considérées, grâce aux données socio-économiques et géographiques.

### 2.3.1 Mobilité et caractérisation socio-économique du territoire

Les informations sur le trafic téléphonique reportent le volume global de trafic écoulé (communications et signalisations) sur chacune des cellules. A ce stade la connaissance du trafic est imprécise et macroscopique puisque les cellules de l'opérateur ne permettent pas de savoir où le mobile se trouve à l'intérieur de celle-ci. Les procédés de localisation que nous proposons permettent de calculer la distribution non uniforme des appels à l'intérieur de chacune des cellules en exploitant toutes les informations spatiales et temporelles disponibles. En général, de telles données décrivant la géographie d'un territoire sont disponibles dans les services des collectivités traitant des Systèmes d'Information Géographique (SIG). En France, ces données sont souvent issues de l'IGN qui géo-référence de nombreuses caractéristiques ; elles sont complétées par les collectivités elles-mêmes qui disposent d'une connaissance très précise de leur territoire. Notons également qu'à l'heure du développement de l'Open Data, des outils tels que OpenStreetMap constituent une nouvelle source de données en plein essor.

A partir des données géographiques et socio-économiques, l'objectif est d'établir un profil d'attraction temporelle propre à chaque maille du territoire. Ce profil ne peut être envisagé que dans un contexte dynamique. A chaque maille est ainsi attribué un vecteur décrivant son attraction au cours du temps (poids d'attraction spatio-temporelle). De cette manière, il est possible d'associer différents profils pour différents types de jours (semaine, weekend, vacances, etc.). La Figure 2.7 décrit le processus complet permettant d'aboutir à cette pondération. Ainsi, à partir d'une description vectorielle du territoire, une première étape consiste à appliquer une grille de maillage à ces données. Chaque maille contient alors la proportion que représente chaque classe de terrain qui la compose. Dans la figure chaque carré de couleur en haut à droite indique la classe dominante. Un profil horaire à trois états est alors associé aux classes de sursol, pour décrire en fonction du temps si une classe est attractive de façon forte (Heure Pleine - HP), faible (Heure Creuse - HC) ou non attractive ( $\emptyset$ ). Ce profil est alors couplé à des données statistiques issues de l'INSEE pour attribuer un poids d'attraction spatio-temporelle final à chaque maille. Ce poids est donc variable dans l'espace et dans le temps.

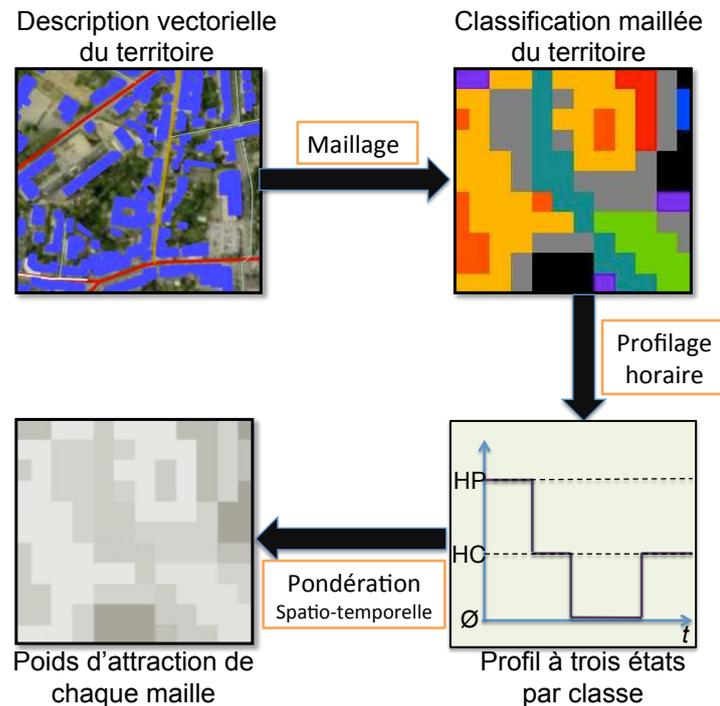


FIGURE 2.7 – Process complet de la pondération spatio-temporelle du territoire

Les sections suivantes décrivent de façon détaillée ce processus qui mène à la description du territoire. Nous présenterons ensuite (cf. section 2.3.2) comment cette description est couplée au réseau de communications (PPC) pour distribuer les mobiles dans les mailles du territoire.

### 2.3.1.1 Données spatiales socio-économiques et géographiques

Le territoire est caractérisé par l'information dite de sursol qui indique l'occupation du sol (route, bâti, eau, etc.). Une approche hiérarchique nous permet de structurer cette information en différentes classes de terrain.

Le terrain est dans un premier temps divisé en grandes catégories (cf. Figure 2.8) : commerce, activité, habitat, etc. Seuls les lieux pouvant significativement définir une concentration de population seront importants dans la suite du travail mais tous les types sont pris en compte. Ces catégories sont couramment utilisées dans les études traitant de la mobilité. C'est le cas des enquêtes ménages/déplacements par exemple, qui recensent les motifs de déplacement directement liés à ces catégories. La catégorie *commerce* contient ainsi toutes activités non professionnelles de type achat, loisir, etc. La catégorie *activité* contient toutes activités professionnelles ou liées à la scolarité. Et la catégorie *voirie* concerne toutes activités liées aux déplacements.

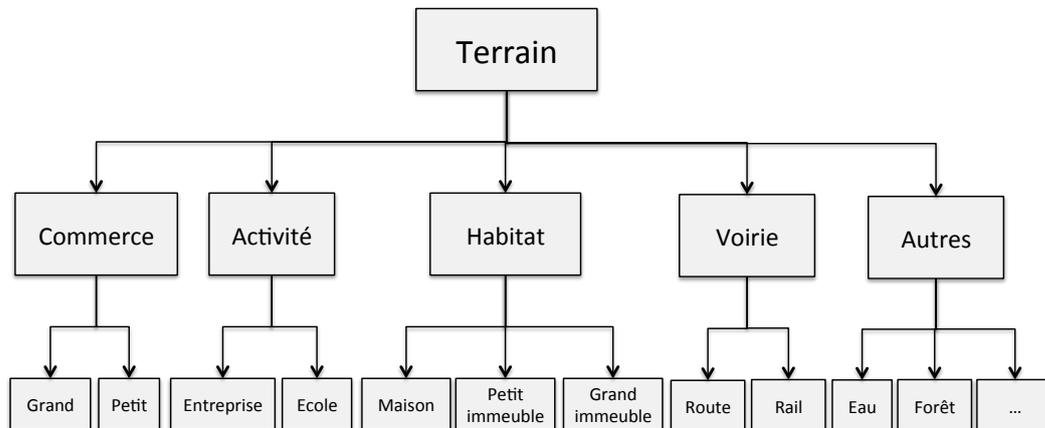


FIGURE 2.8 – Organisation du sursol

Chaque catégorie est subdivisée en plusieurs classes afin d'apporter plus de finesse lors de la distribution des mobiles. L'habitat sera par exemple décomposé en maison, petit immeuble et grand immeuble, ce qui permettra lors d'affectation de personnes au domicile d'affecter plus ou moins de personnes selon la taille du bâtiment. De la même façon, les commerces sont distingués selon leur taille : un petit commerce de proximité n'aura pas le même pouvoir d'attraction qu'un centre commercial.

Ainsi, la description du territoire selon cette arborescence est tout d'abord disponible sous forme vectorielle, puis la première étape du traitement est le maillage et la classification de ces données. Cette opération relève d'un géo-traitement à base d'intersection de polygones. Il ne sera pas détaillé ici mais est publié dans [Lamrous 2008]. Voyons à présent comment est attribué un profil horaire à chaque classe de terrain, soit comment inclure une dimension temporelle aux données spatiales.

### 2.3.1.2 Un profil temporel des classes de terrain à trois états

Selon ses caractéristiques socio-économiques, un lieu attire des personnes différemment à chaque moment de la journée et d'un jour à l'autre. Pour rendre compte de ce phénomène nous procédons à un découpage du temps en plages horaires qui sera associé au découpage géographique du territoire en mailles de l'étape précédente. Une bonne connaissance de l'occupation du sursol du territoire nous permettra ainsi de distribuer les mobiles rattachés à un relai téléphonique différemment selon l'heure de la détection. Nous pouvons citer l'exemple d'un plan d'eau urbain interdit à la baignade et à la navigation, ou encore d'un établissement scolaire pouvant recevoir des personnes en journée mais pas la nuit. Nous élaborons ainsi un schéma temporel de caractérisation du sursol.

Chaque classe élémentaire de terrain représentée Figure 2.8 a un pouvoir d'attraction à trois états qui dépend du jour et du moment de la journée. Nous définissons un tel profil horaire pour une classe de la manière suivante : chaque plage horaire  $\Delta t$  (généralement de 15 minutes) peut prendre la valeur 0 lorsqu'un nombre non significatif de personnes peut se trouver dans cette classe, 1 lorsque la classe est considérée en heure creuse (faiblement fréquentée) et 2 lorsqu'un nombre important de personnes peut se trouver dans cette classe. Cette approche à trois niveaux (absence, faible présence, forte présence) apporte une grande souplesse dans la construction et dans l'évolution de la matrice d'appartenance horaire. Il s'agit donc d'un compromis entre une approche binaire du tout ou rien et une approche continue qui complexifierait la construction de ces profils.

	6h	6h15	6h30	6h45	7h	...	
Petits commerces	0	0	0	0	0	0	Commerce
Grands commerces	0	0	0	0	0	0	
Entreprise	0	0	0	0	0	0	Activité
Ecole	0	0	0	0	0	0	
Maisons <10m	2	2	2	2	2	2	Habitat
Petits immeubles < 20m	2	2	2	2	2	2	
Grands immeubles > 20m	2	2	2	2	2	2	
Routes	1	1	1	1	1	1	Voie
Rail	0	0	0	0	1	1	
Jardins / Parcs	0	0	0	0	0	0	Autres
Forêt	0	0	0	0	0	0	
Eau	0	0	0	0	0	0	

FIGURE 2.9 – Table de profilage temporel du sursol

Un extrait de la table de profilage temporel du sursol à trois états est donné Figure 2.9 de 6h à 7h (en colonnes). Chaque ligne correspond à une classe de terrain pour un lundi d'une période travaillée. Le profil change d'un jour à l'autre de la semaine et éventuellement d'une semaine à l'autre pour refléter au plus près la réalité.

Plus formellement, à la table de profilage temporel est associée la matrice  $A$ , qui est définie de la façon suivante :

$$A = (a_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$$

avec :

$p$  : le nombre de classes élémentaires de sursol

$n$  : le nombre de plages horaires dans une journée

$$a_{ij} = \begin{cases} 2 & \text{si la classe } i \text{ est considérée en} \\ & \text{HP à la période } j \\ 1 & \text{si la classe } i \text{ est considérée en} \\ & \text{HC à la période } j \\ 0 & \text{sinon (pas de personne affectée)} \end{cases} \quad (2.1)$$

Cette étape nous a permis de définir une matrice décrivant le profil temporel à trois états de chaque classe de terrain. Il s'agit maintenant de convertir cette matrice en poids d'attraction spatio-temporelle variables dans le temps pour chaque classe, avant d'en déduire un poids d'attraction pour chaque maille.

### 2.3.1.3 Un poids d'attraction spatio-temporelle pour le territoire

Nous abordons maintenant la phase qui permet de définir la façon dont la nature du sursol influence la répartition des mobiles sur le territoire. Cette répartition se déroule en deux temps. Tout d'abord, il s'agit de déterminer un poids d'attraction global pour chaque classe définie dans le tableau 2.8; puis une deuxième étape consiste à distribuer les poids globaux sur les plages de la journée en tenant compte des trois niveaux possibles de présence de personnes (absence, faible présence, forte présence).

**Répartition globale des poids de chaque classe de terrain** Pour procéder à cette étape nous avons considéré les données statistiques fournies par l'INSEE pour la France sur la période 2007-2009, afin de définir une estimation du temps moyen passé dans chaque classe. Remarquons que ces poids constituent une donnée d'entrée de l'algorithme, et peuvent être modifiés, pour tenir compte de caractéristiques propres à chaque territoire.

La période de nuit n'est pas prise en compte, toutes les personnes sont considérées appartenant à la classe habitat. Pour la période de la journée (6h - 0h), nous considérons le pourcentage de temps moyen passé par catégorie (ramené à l'ensemble de la population) tel que décrit dans le tableau 2.1. La première ligne du tableau issue de l'INSEE donne la répartition entre les catégories commerce, activité, habitat, voirie et autre. On remarque que 18% du temps (catégorie autre) ne peut être attribuée à une classe identifiée. On choisit alors de répartir cette proportion de temps entre les classe identifiées au prorata de celles-ci, ce qui donne la nouvelle répartition de la deuxième ligne du tableau. La répartition au prorata est arbitraire mais neutre; elle pourrait être faite autrement selon le territoire. Cette répartition est définie pour un jour travaillé (hors vacances scolaires).

Pour définir la proportion de temps passé dans chaque classe de sursol du

	Commerce	Activité	Habitat	Voirie	Autres
Enquête brute	25%	25%	22%	10%	18%
Enquête modifiée	30,49%	30,49%	26,83%	12,2%	0%

TABLE 2.1 – Temps moyen passé par catégorie

Catégorie	% catégorie	Classe de sursol $g$ pour <i>ground</i>	% classe	Répartition globale $\omega_g$
Commerce	30,49%	petit commerce	30%	9,15%
		grand commerce	70%	21,34%
Activité	30,49%	entreprise	67%	20,43%
		école	33%	10,06%
Habitat	26,83%	maison	6,25%	1,68%
		pt immeuble <20m	31,25%	8,38%
		gd immeuble >20m	62,5%	16,77%
Voirie	12,2%	route	75%	9,15%
		rail	25%	3,05%

TABLE 2.2 – Répartition de la population par thème

tableau 2.8, les catégories sont elles-mêmes réparties selon des proportions guidées par des statistiques de l'INSEE. Le tableau 2.2 synthétise cette répartition. Les deux premières colonnes qui décrivent les catégories sont directement issues du tableau 2.1. Les colonnes suivantes concernent les classes de sursol et donnent, d'abord la répartition interne entre les classes de sursol dans chaque catégorie, puis la répartition globale par rapport à l'ensemble du temps passé qui croise les catégories et les classes de sursol. Prenons l'exemple de la catégorie *activité* qui représente 30,49% du temps passé. 33% de ce temps est attribué à la classe école, et 67% à la classe entreprise, soit 33% de 30,49% = 10,06% de temps pour la classe école et de même 20,43% de temps pour la classe entreprise. La répartition attribuée aux classes de sursol provient d'une estimation de la proportion de population concernée ; elle doit être affinée par territoire d'étude.

La répartition du temps global par classe de sursol est ainsi donnée dans la dernière colonne du tableau 2.2. Il s'agit des poids  $\omega_g$  de chaque classe de sursol  $g$  qu'il faut à présent répartir dans le temps.

**Répartition temporelle de chaque classe de terrain** La dernière étape de la caractérisation temporelle des classes de terrain consiste à déterminer le poids d'attraction de chacune d'elles tout au long de la journée pour un type de jour donné. C'est ce poids final qui permettra de distribuer statistiquement les mobiles à

l'intérieur de la cellule radio. Cette étape consiste donc à répartir les poids calculés globalement pour chaque classe sur l'ensemble des plages horaires de la journée, en s'appuyant sur la matrice  $A$  qui décrit le profil des classes en trois états.

Soient les définitions suivantes :

$G$  l'ensemble des classes de sursol ( $G$  pour *ground*)

$g$  une classe de sursol,  $g \in G$

$\omega_g^{hp}$  le poids de la classe  $g$  en heure pleine

$\omega_g^{hc}$  le poids de la classe  $g$  en heure creuse

$\omega_g$  le poids global de la classe  $g$  calculé précédemment

$\omega_g^0 = 0$  le poids de la classe  $g$  en heure d'absence (ce poids peut être mis à une valeur différente de 0 si on veut traduire une très faible présence).

$n_g^{hp}$  le nombre de plages horaires en heure pleine pour la classe  $g$

$n_g^{hc}$  le nombre de plages horaires en heure creuse pour la classe  $g$

$n_g^0$  le nombre de plages horaires d'absence pour la classe  $g$

Pour une classe donnée  $g$ , la matrice  $A$  contient les valeurs 0, 1 ou 2 selon les plages horaires. Le poids  $\omega_g$  est alors réparti linéairement sur l'ensemble des plages absence, HC et HP de la journée, avec une proportion  $x$  à définir entre HC et HP selon le territoire.

Calculons à présent les poids  $\omega_g^{hp}$  et  $\omega_g^{hc}$  de la classe  $g$  en heure pleine et en heure creuse respectivement. Ils vérifient :

$$\begin{cases} \omega_g^{hc} n_g^{hc} + \omega_g^{hp} n_g^{hp} + \omega_g^0 n_g^0 = \omega_g \\ \omega_g^{hp} = x * \omega_g^{hc} \end{cases} \quad (2.2)$$

Avec  $\omega_g^0 = 0$ , il s'en suit que les poids d'attraction des classe de sursol sont donnés par :

$$\begin{cases} \omega_g^{hc} = \frac{\omega_g}{x * n_g^{hp} + n_g^{hc}} \\ \omega_g^{hp} = x * \frac{\omega_g}{x * n_g^{hp} + n_g^{hc}} \end{cases} \quad (2.3)$$

Chaque maille  $m$  est à présent caractérisée par un vecteur d'attraction  $\omega_m = [\omega_m^1, \dots, \omega_m^n]$ , construit à partir des classes de sursol couvrant la maille en respectant la proportion de couverture de chacune. Chaque composante  $\omega_m^t$ , avec  $t \in [1..n]$ , est définie par :

$$\omega_m^t = \sum_{g \in G} \omega_g^t \cdot p_{m,g} \quad (2.4)$$

avec :

$\omega_m^t$  le poids d'attraction spatio-temporelle de la maille  $m$  à l'instant  $t$

$\omega_g^t$  le poids de la classe de sursol  $g$  à l'instant  $t$

$$\omega_g^t = \begin{cases} \omega_g^{HP} & \text{si } a_{g,t} = 2 \\ \omega_g^{HC} & \text{si } a_{g,t} = 1 \\ \omega_g^{\emptyset} & \text{si } a_{g,t} = 0 \end{cases} \quad (2.5)$$

$p_{m,g}$  la proportion de classe de sursol  $g$  dans la maille  $m$

L'étape suivante de notre modèle consiste à distribuer les mobiles sur un territoire.

### 2.3.2 Modèle de distribution spatio-temporelle des mobiles

Dans cette partie nous rassemblons les éléments présentés jusqu'ici, les données téléphoniques et les poids d'attraction spatio-temporelle, pour aboutir à la répartition des mobiles sur l'ensemble du territoire. La Figure 2.10 illustre ce qui est présenté dans cette section.

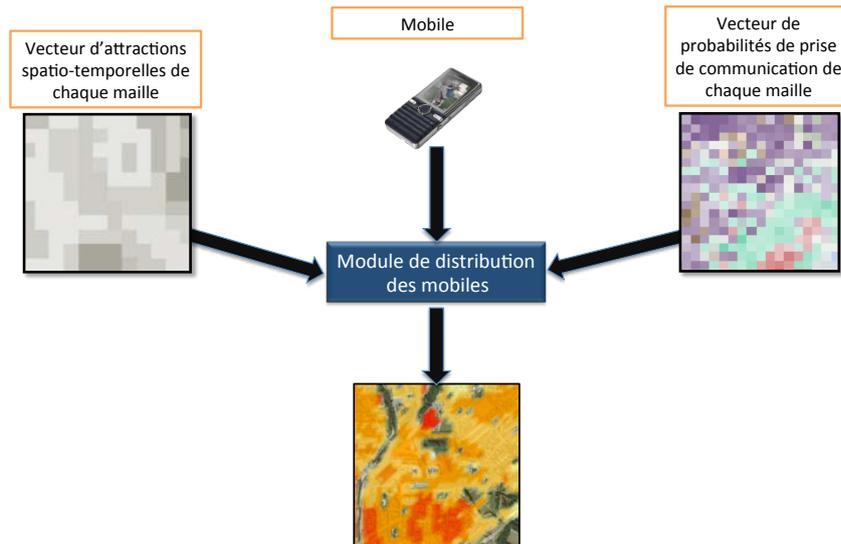


FIGURE 2.10 – Module de distribution des mobiles

Nous procéderons à une répartition en deux temps. Dans un premier temps les poids d'attraction spatio-temporelle ne sont pas intégrés. Ceci permettra de vérifier qu'une répartition des mobiles sans prise en compte du territoire a une forte tendance à être homogène. Puis dans un deuxième temps les poids d'attraction seront intégrés pour réaliser une distribution des mobiles plus réaliste.

### 2.3.2.1 Modèle de distribution de mobiles selon la cartographie des Probabilités de Prise de Communication (PPC)

Comme nous l'avons vu, les cellules radio sont découpées en mailles carrées qui définissent le niveau de précision que l'on souhaite atteindre. Ainsi, pour chaque maille est générée une structure multicouche qui intègre toutes les cellules qui la recouvre ainsi que les PPC associées. De la même façon, la journée est découpée en plages horaires de  $\Delta t$ . Pour l'application numérique nous fixons le pas de maillage  $pdm$  à 25m de côté et le pas de temps  $\Delta t$  à 15 minutes. La répartition des mobiles sur les mailles d'une cellule radio pendant l'intervalle  $\Delta t$  est faite proportionnellement aux PPC de chaque maille. Ainsi la densité des mobiles sur chaque maille  $m$  est déterminée en fonction de la représentation de la PPC de la maille relativement à l'ensemble des PPC de toutes les mailles  $m'$  de la cellule. Pour une station  $s$  sur une maille  $m$  à l'instant  $t$  (pour simplifier la notation,  $t$  sera assimilée à la période de temps  $\Delta t$ ) nous avons la répartition suivante :

$$n_{m,s}^t = \frac{ppc_{m,s}}{\sum_{m' \in C_s} ppc_{m',s}} \cdot n_s^t \quad (2.6)$$

avec :

$n_{m,s}^t$  le nombre de mobiles rattachés à la station  $s$  réparti sur la maille  $m$  à l'instant  $t$

$C_s$  l'ensemble des mailles appartenant à la zone de couverture de la station  $s$  i.e. l'ensemble des mailles du territoire pour lesquelles la PPC associée à  $s$  est non nulle

$ppc_{m,s}$  la probabilité lorsqu'un mobile est sur la maille  $m$  que son appel soit acheminé par la station  $s$

$ppc_{m',s}$  la probabilité lorsqu'un mobile est sur la maille  $m'$  que sa communication soit acheminée par la station  $s$

$n_s^t$  le nombre de mobiles rattachés à la station  $s$  à l'instant  $t$

Remarquons dans le modèle (2.6) que les PPC sont invariantes dans le temps, seule l'apparition des mobiles est variable.

Nous avons ainsi défini la distribution des mobiles à l'intérieur d'une cellule radio indépendamment des propriétés spatio-temporelles du sursol. Ce même traitement est appliqué à l'ensemble des cellules, puis nous définissons pour chaque maille un cumul de présences issues de l'ensemble des cellules qui contiennent cette maille i.e. pour lesquelles la PPC est non nulle. Ainsi pour chaque maille  $m$  à l'instant  $t$  nous obtenons la répartition globale de l'ensemble des mobiles de la façon suivante :

$$n_m^t = \sum_{s \in S_m} n_{m,s}^t \quad (2.7)$$

avec :  $S_m$  l'ensemble des stations couvrant, PPC non nulle, la maille  $m$ .

La Figure 2.11 illustre la répartition des mobiles sur un territoire sans prise en compte des propriétés du sursol. Nous apercevons les limites d'une telle approche qui est basée uniquement sur les caractéristiques du réseau à travers la PPC. En effet, même si en ingénierie de réseaux on cherche à avoir une bonne qualité de signal dans les lieux de forte fréquentation, la réciproque n'est toutefois pas vraie. Ce n'est pas parce qu'il y a un bon signal (une bonne PPC) en un lieu donné que nécessairement il y a beaucoup de personnes en ce lieu.

La Figure 2.11 illustre bien ce phénomène. On aperçoit notamment que, sans prise en compte du sursol, le modèle distribue autant de mobiles dans un plan d'eau que dans les habitations voisines, de façon très uniforme.

### 2.3.2.2 Modèle de distribution spatio-temporelle intégrant les propriétés du sursol

Le modèle vu jusqu'ici ne tenait compte que de la couverture radio. L'étape suivante consiste à intégrer dans le calcul les vecteurs de poids d'attraction spatio-temporelle  $\omega_m$  de chaque maille tels que définis dans la section 2.3.1.3.

En appliquant cette pondération au modèle de distribution des mobiles précédent, nous obtenons pour la maille  $m$  à l'instant  $t$  l'affectation suivante :

$$n_m^t = \sum_{s \in S_m} \left( \frac{ppc_{m,s} \cdot \omega_m^t}{\sum_{m' \in C_s} ppc_{m',s} \cdot \omega_{m'}^t} \cdot n_s^t \right) \quad (2.8)$$

L'intégration des poids d'attraction spatio-temporelle avec le modèle de distribution de mobiles a pour effet de réorganiser la distribution des mobiles de façon plus réaliste. Pour s'en rendre compte la Figure 2.12 présente le résultat de l'application

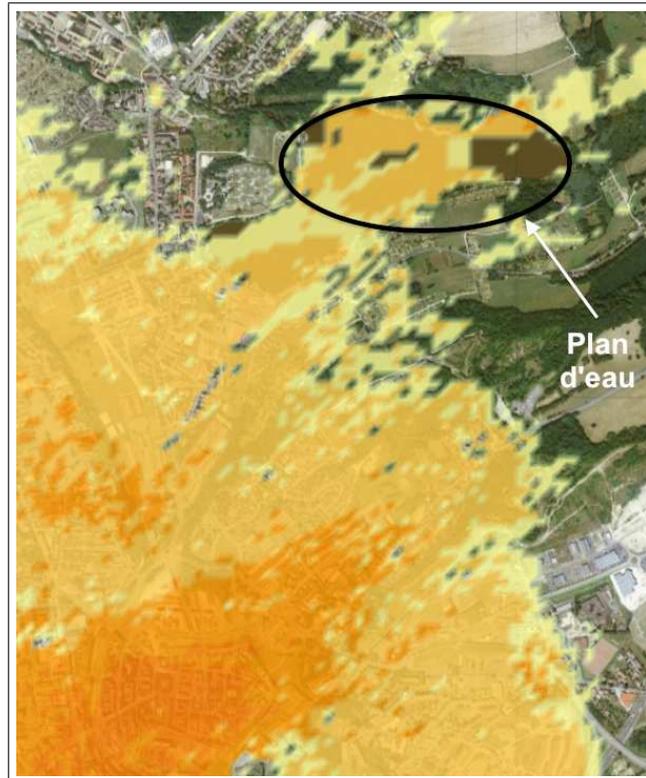


FIGURE 2.11 – Distribution des mobiles à un instant  $t$  sans prise en compte des propriétés du sursol

de ce modèle, à gauche sans prise en compte de ces poids et à droite en les intégrant. En reprenant notre exemple précédent, nous pouvons ainsi remarquer que la partie de la population qui était affectée à un plan d'eau lorsque seules les PPC étaient considérées se retrouve réaffectée dans des mailles plus appropriées.

La visualisation de la répartition des mobiles avec et sans prise en compte des poids d'attraction spatio-temporelle donne une nouvelle visibilité de la présence des personnes. Toutefois, une analyse numérique des résultats doit permettre de vérifier cette approche. Ce point est abordé dans la section suivante.

## 2.4 Résultats et analyse

### 2.4.1 Environnement d'analyse

Une bonne analyse de la mobilité nécessite de disposer d'un outil efficace pour représenter les résultats des calculs sur les déplacements. Une plate-forme logicielle capable de gérer à la fois des données spatiales telles que des données géographiques

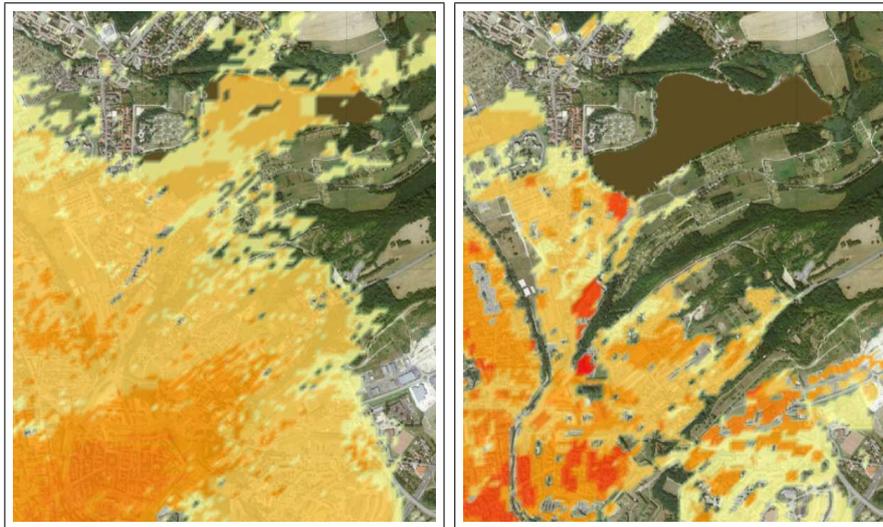


FIGURE 2.12 – Distribution des mobiles à un instant  $t$  sans et avec prise en compte des poids d'attraction spatio-temporelle

ou de localisation, ainsi que des données temporelles reflétant la mobilité apparaît alors comme un élément essentiel.

De nombreux outils existent dans le domaine des SIG (Systèmes d'Information Géographiques) pour représenter des données géo-localisées. Toutefois la prise en compte simultanément de l'espace et du temps est rare.

Afin de permettre une grande souplesse d'utilisation nous avons donc développé notre propre plate-forme de représentation des résultats. Elle dispose des fonctionnalités habituellement présentes dans les outils SIG classiques, telles que l'affichage de fonds de carte ainsi que de données vectorielles ou ponctuelles de type shapefile (cf. Figure 2.14). Il est par ailleurs possible de se déplacer dans cet environnement, de zoomer sur un lieu précis, etc. Chaque objet chargé, tel que les limites de communes ou les contours des bâtiments, constitue un calque qu'il est possible d'afficher ou masquer selon les besoins.

Les données peuvent être statiques ou dynamiques. Alors que les éléments topographiques tels que les axes routiers ou les bâtiments ne varient pas dans le temps, mais leurs propriétés quant à elles peuvent évoluer au cours de la journée. En particulier un supermarché sera toujours au même endroit mais aura un pouvoir d'attraction différent à 9h, 16h ou 23h et cela suivant le jour considéré. De même la localisation de la population, tant au niveau macroscopique, sous forme de masse, qu'au niveau microscopique, sous forme ponctuelle ou de trajectoire, varie dans le temps. Pour rendre compte de cette dynamique l'application que nous proposons permet de définir la journée de l'étude et le moment de la journée à observer. Enfin il est possible de faire défiler le temps entre 2 plages horaires, de faire varier la vitesse de

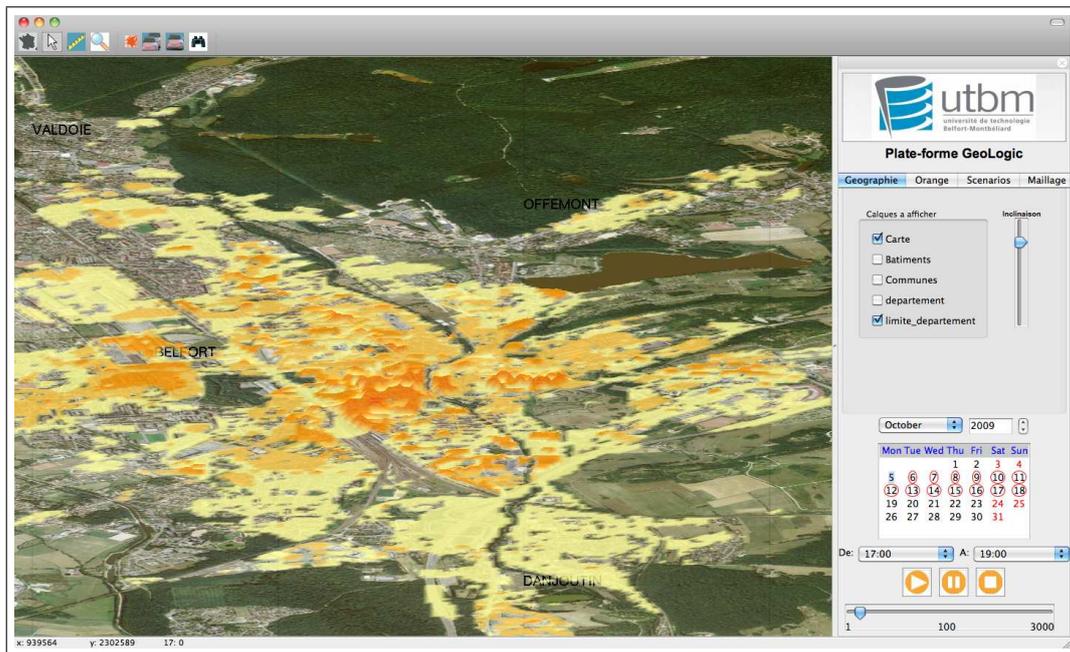


FIGURE 2.13 – Plate-forme geoLogic



FIGURE 2.14 – Informations vectorielles : SIG shapefile

défilement du temps, ou encore de mettre l'application en pause.

## 2.4.2 Constitution des données de référence

Les données utilisées pour constituer la base de référence sont l'Enquête Ménages/Déplacements (EMD) ainsi que les données INSEE du territoire. Les résultats de l'enquête effectuée en 2004/2005 (enquête redressée) fournissent une base de données contenant notamment les 435 360 enregistrements correspondant aux réponses apportées par chaque personne enquêtée sur les déplacements de la veille, redressées à l'ensemble de la population. Les déplacements contenus dans la base sont ceux ayant pour origine ou destination une zone prédéfinie de la CAB et ayant été effectués en semaine. Ainsi l'EMD nous permet de connaître les déplacements des habitants du territoire étudié tout au long de la journée.

L'EMD est donc riche d'informations sur la manière dont la population se déplace sur le territoire. Les traitements que nous avons effectués sur ces données ont deux objectifs : d'une part fournir des résultats visualisables dans le logiciel développé afin de se rendre compte de la façon dont la population évolue (avec une information variable dans l'espace et dans le temps à différente échelle spatiale i.e. pour une zone particulière ou pour tout le territoire de l'EMD) et d'autre part de constituer la base de référence pour comprendre le lien pouvant exister entre déplacements et activité téléphonique.

### 2.4.2.1 Agrégation temporelle

L'enquête réalisée contient l'horodatage précis des déplacements de la population sondée. Dans le cadre du projet nous avons échantillonné les journées en périodes de 15 minutes qui correspondent à une précision temporelle exploitable pour l'analyse des déplacements dans le cadre d'une étude sur le transport urbain. Une première étape consiste donc à affecter à chacun des 435 360 enregistrements la plage horaire (le 1/4 d'heure) lui correspondant pour pouvoir agglomérer ces affectations et disposer d'une information globale par période.

### 2.4.2.2 Agrégation géographique

Une classe de déplacements est un ensemble zone d'origine, zone de destination, plage horaire, motif de déplacement au sens de l'EMD. Le motif est facultatif car les déplacements pourront être distingués ou non selon le motif. Il faut donc calculer pour chaque classe ainsi définie le nombre d'enregistrements correspondant. Ce calcul répond donc à la question : Combien de personnes sont parties de la zone  $A$  pour aller à la zone  $B$  au cours du 1/4 d'heure  $t$  (éventuellement pour le motif  $m$ ) ?

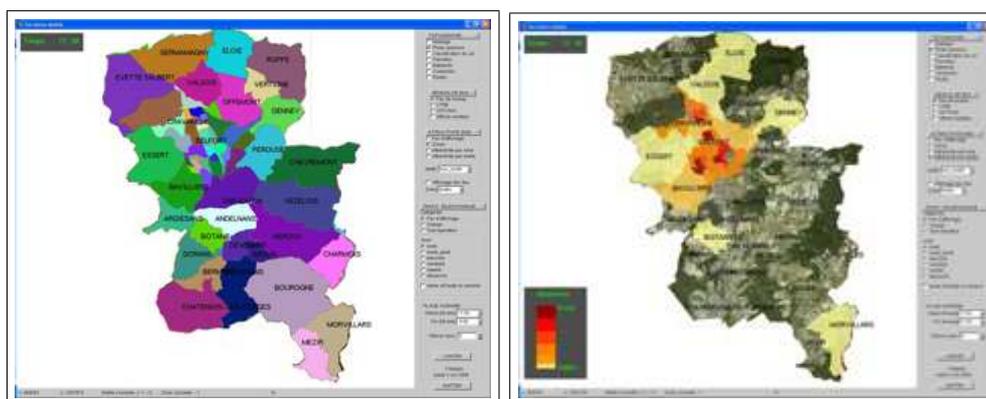


FIGURE 2.15 – Zonage et attractivité des zones à 17h

La Figure 2.15 illustre à gauche le découpage du territoire en différentes zones

(chaque couleur représente une zone). A droite est représentée l'attractivité selon l'EMD de chaque zone. Un niveau de couleur allant du blanc au rouge traduit le niveau d'attractivité de la zone. Ainsi les zones rouges foncées indiquent des lieux à forte attractivité.

### 2.4.2.3 Calcul de la différence cumulée au cours de la journée

L'évolution de la population dans les zones au cours de la journée est une information utile que l'on peut calculer à partir de l'évolution de l'attractivité de chaque zone au cours du temps. Pour cela, il faut prendre en considération non seulement les flux entrants et sortants à l'instant considéré, mais y ajouter également tous les flux entrants opérés depuis le début de la journée (cumul des arrivées) et retrancher de la même manière la somme des flux sortants depuis le début de la journée (cumul des départs). Le point de référence temporel considéré est à 6h du matin.

Cette information s'exprime de la façon suivante : à un moment donné, est-ce que la zone s'est remplie (teintes rouges) ou vidée (teintes bleues) depuis 6h00 du matin par rapport à la population initiale ? Et quel est le solde en nombre de personnes arrivées ou parties ?

Les teintes claires expriment un faible solde dans un sens ou dans un autre, soit un équilibre des arrivées et des départs. Les teintes foncées présentent toutes un solde positif dans les rouges et négatif dans les bleues.

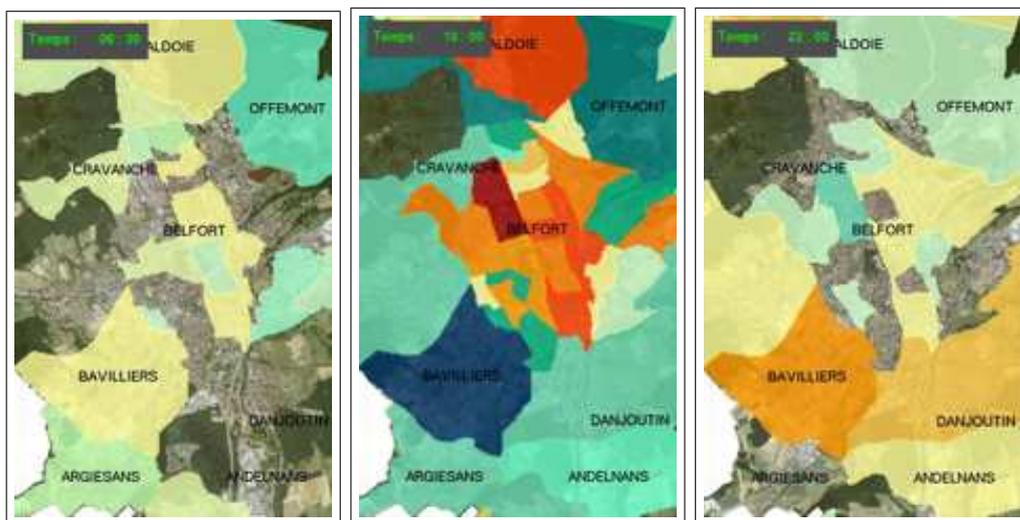


FIGURE 2.16 – Différence cumulée à 6h30, 10h et 22h

La Figure 2.16 présente la différence cumulée à 6h30, 10h et 22h. Alors que le matin à 6h30 il n'y a encore eu que très peu de mouvements, à 10h nous observons

que les zones industrielles et de centre ville se sont particulièrement remplies au détriment des zones périurbaines. A 22h la tendance se rapproche de la situation de 6h30, marquant un retour au domicile.

#### 2.4.2.4 De la différence cumulée à la présence

Dans le cadre de cette section nous nous intéressons à la mise au point d'un procédé pour fournir une information pertinente sur la localisation de la population au cours de la journée.

Afin de faire un parallèle entre les données téléphoniques et l'enquête ménages déplacements, nous devons donc traduire les données de l'EMD en termes de population présente à chaque instant et en chaque endroit du territoire représenté par les zones. Ce calcul qui fournit une information utile en soi sera également la base de comparaison utilisée pour valider la localisation identifiée par les téléphones portables, ces résultats étant construits de manière totalement indépendante.

Pour ce traitement nous considérons la différence cumulée présentée en 2.4.2.3. Pour chaque zone nous ajoutons la population résidante issue de l'INSEE et nous obtenons ainsi une approximation du nombre de personnes présentes chaque 1/4 d'heure de la journée. Il n'y a en effet que les déplacements effectués entre minuit et 6h00 qui ne sont pas pris en considération. L'information produite revient donc à reprendre les résultats de la différence cumulée translatés du nombre d'habitants de chaque zone à tout moment de la journée.

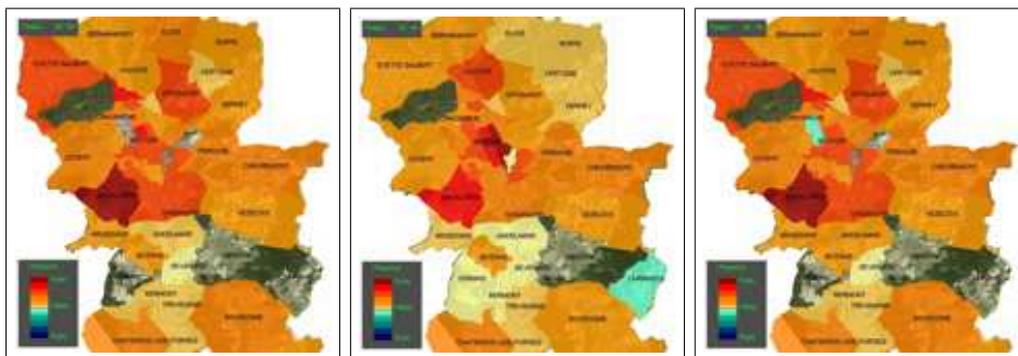


FIGURE 2.17 – Présence à 3 périodes de temps

Les résultats de présence sont représentés Figure 2.17 pour des calculs effectués à 6h, 10h et 22h. La carte de gauche représentant le résultat à 6h00 correspond au recensement INSEE ramené aux zones de l'EMD. A 10h00 nous pouvons observer la nouvelle répartition de la population en fonction des activités de chacun. Enfin la carte de droite montre la répartition à 22h00 qui est semblable à celle observée le matin à 6h00 : la population a presque repris sa distribution initiale.

### 2.4.3 Expérimentation et résultats

La comparaison des résultats téléphoniques avec l'EMD a pour but d'établir le lien pouvant exister entre l'activité téléphonique, la mobilité des personnes et le territoire. A cette fin nous proposons ici de comparer la répartition spatiale des mobiles d'après notre modèle exposé en section 2.3.1 à l'EMD, ce qui doit nous permettre d'identifier dans quelle mesure les personnes utilisant des téléphones portables sont représentatives de l'ensemble de la population selon le moment de la journée et le lieu. Le principe de validation emploie les zones de l'enquête ménage : pour une zone donnée, l'expérimentation doit montrer comment la présence des personnes localisées par les téléphones portables reflète la présence des personnes issues de l'enquête sur cette même zone. Il faut donc dans un premier temps agglomérer dans les zones les présences détectées via les téléphones sur l'ensemble des mailles du territoire par l'application de notre modèle de répartition spatio-temporelle du trafic téléphonique. Nous pouvons alors comparer les tendances des courbes issues de l'EMD et de la localisation.

Les résultats de localisation doivent être observés quasi obligatoirement dans des contextes de variation spatiale et temporelle. Par exemple, comment évolue au cours d'une journée travaillée la présence de population sur les zones industrielles ? Ou sur les zones commerciales les samedis ? Etc. Il faut donc à la fois regarder le lieu et le moment.

Nous présentons ci-après les comparaisons entre les courbes de présences de l'enquête EMD et les courbes de présences issues du modèle de répartition du trafic téléphonique (les identifiants des zones sont ceux issus de l'enquête). Nous montrons ces comparaisons sur 4 cas de figure représentatifs des phénomènes que nous avons observé sur l'ensemble des 47 zones.

- 2 zones urbaines : zone d'activité industrielle (Techn'hom) et zone du Centre Ville avec de nombreux petits commerces et habitat de type immeuble.
- 2 zones périurbaines : Andelnans et Bavilliers qui sont deux villages avec population résidente (500 habitants à Andelnans et 5400 à Bavilliers) et quelques commerces.

Les courbes sur les 4 graphiques (Figures 2.18 à 2.21) indiquent :

- En rouge, les résultats de l'enquête EMD, c'est-à-dire les données de référence. L'ordonnée tient compte de la population qui habite sur les lieux observés donc le point départ dépend de la zone ; par exemple 0 pour la zone d'activité et 3000 pour la zone centre ville. La zone d'activité fournit une très bonne référence de traçage de la mobilité puisqu'elle n'inclut pas de résidents.
- En bleu, les résultats de notre modèle de distribution spatio-temporelle, c'est-

à-dire les données à valider. Les données sont multipliées par 100 pour que l'amplitude de ces courbes soient dans le même champ de lecture que les courbes de référence et faciliter la comparaison (sauf les données d'Andelnans qui sont multipliées par 10). Le point initial est toujours à 0 c'est-à-dire que l'on n'inclut pas la population résidante dans la mesure.

Une moyenne mobile sur 8 valeurs est calculée sur chaque courbe pour donner les tendances et faciliter l'analyse.

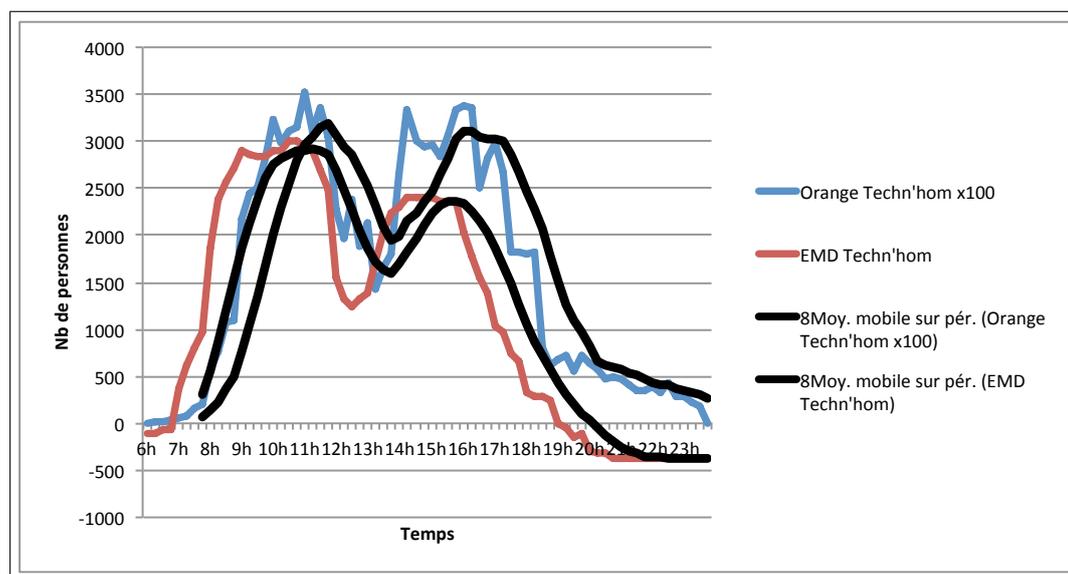


FIGURE 2.18 – Comparaison EMD / modèle en zone d'activité industrielle sans habitat

La zone d'activité sans résident initial montre Figure 2.18 deux courbes dont les tendances sont bien corrélées. Il en est de même pour la zone de centre ville (cf. Figure 2.19). Les autres zones qui sont essentiellement des zones d'habitation ont des courbes qui semblent suivre une corrélation en opposition de phase tel qu'illustré en Figures 2.20 et 2.21. Pour mieux observer l'inversion de phase des zones d'habitation, nous avons retourné les courbes correspondantes. Les Figures 2.22 et 2.23 montrent cette symétrie des données issues du modèle par rapport à l'axe  $f(x) = 150$  et  $f(x) = 2500$  respectivement.

Des types de zones apparaissent donc selon l'impact du sursol, résidentiel ou pas, et avec activités économiques ou pas. On peut ainsi identifier deux catégories de zones, caractérisées par les points suivants :

- Sur les 2 zones urbaines, les courbes sont très corrélées : la présence des personnes distribuées par le modèle suit étroitement celle des personnes vues par l'EMD. Il s'agit principalement de zones non habitées (Techn'hom) ou

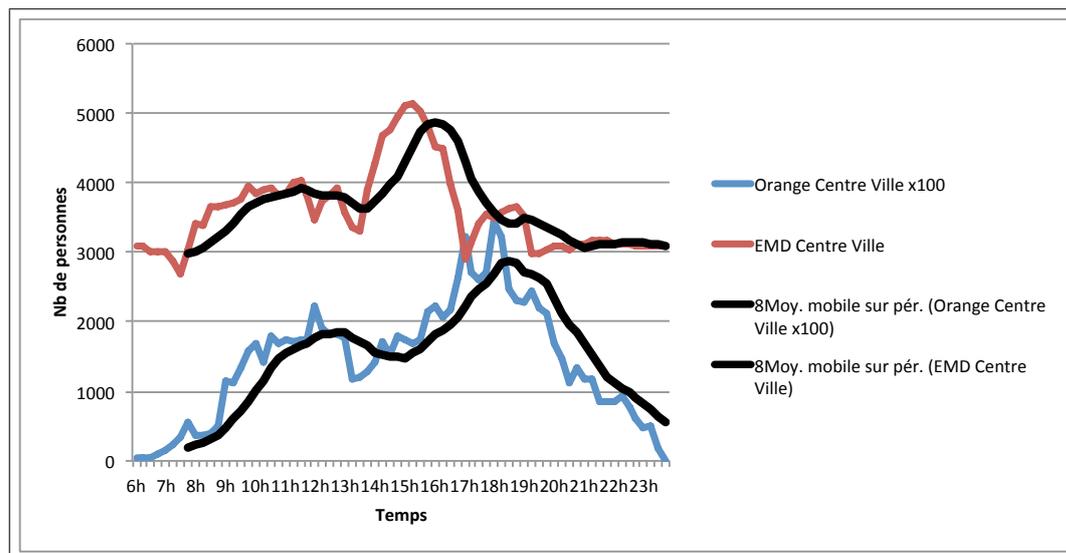


FIGURE 2.19 – Comparaison EMD / modèle en centre ville avec habitats et commerces

faiblement habitées (centre ville) où les personnes se rendent pour travailler ou faire leurs courses du fait d'une typologie d'activités économiques. L'emploi des téléphones mobiles est bien corrélé à la présence des personnes en déplacement sur ces zones.

- Sur les 2 zones péri-urbaines, les courbes sont inversées ou corrélées en inversion de phase. Quand les personnes sont chez elles à des heures de faible mobilité (dans la matinée, à l'heure du déjeuner et le soir après 21h30), elles utilisent peu leur téléphone mobile ; dans les périodes de transitions où des déplacements sont effectués (arrivées pour midi, arrivées en fin de journée entre 17h00 et 21h30), les personnes utilisent leur téléphone mobile. Les deux courbes sont harmonieusement inversées. Là aussi le périmètre horaire propice aux déplacements est détectable par l'activité téléphonique.

Notons que les 4 zones présentent un glissement sur les pics issus de notre modèle par rapport à ceux de l'EMD. Ce glissement est homogène : il se produit vers 12h-12h30 et vers 17h30-19h30 et toujours en décalage d'une heure plus tard avec notre modèle (2h pour la zone centre ville). Ce qui est intéressant dans ce phénomène est qu'il est visible quel que soit le type de zones, on peut donc supposer que le modèle pourrait être calibré. Cependant, il s'agit de périodes à mobilité forte. Côté trafic téléphonique cela signifie un fort usage du téléphone sur ces périodes. Côté EMD cela signifie plus d'incertitude sur les résultats enregistrés pendant l'enquête. En effet, lors de l'enquête, les personnes enquêtées s'expriment sur ce qu'elles ont fait comme déplacement la veille. Ainsi, de façon assez conventionnelle, pour prendre un exemple, une personne aura déclaré qu'elle part du travail à 17h-17h30 et rentre à son domicile. L'enquête traduit cela par le fait de marquer une zone d'origine à

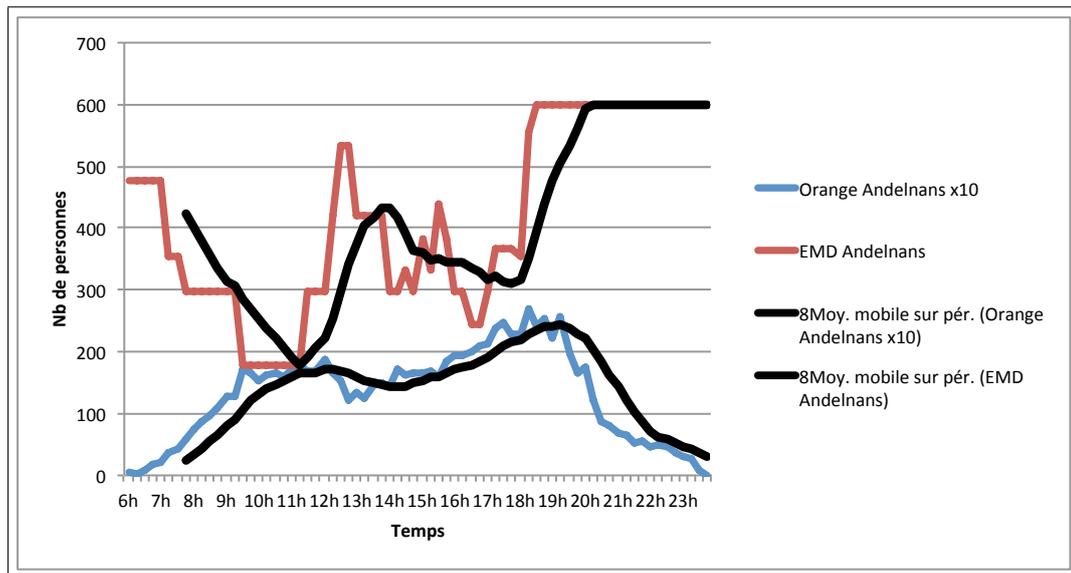


FIGURE 2.20 – Comparaison EMD / modèle en secteur périurbain 1 avec habitat et très peu de commerce

17h (zone de travail) et une zone de destination à 17h30 (zone du domicile). En pratique, l'enquête exclut donc très facilement une période de temps, entre le travail et le domicile, pendant laquelle il y a une forte variation d'un jour à l'autre, la plus forte en réalité. Un jour, la personne va rentrer à cette heure-là ; le lendemain, 1h30 plus tard car elle aura fait des courses ; le surlendemain, 30 min plus tard car elle aura travaillé un peu plus, etc. Ces phénomènes sont visibles lorsque les contraintes de mobilité sont souples, en général quand on sort du travail plutôt que quand on y va, soit plutôt aux heures où nous avons observé un glissement des résultats du modèle par rapport à l'EMD. On ne peut donc pas en conclure définitivement qu'il faut calibrer le modèle mais plutôt que c'est la partie la plus stochastique du comportement humain et qu'une enquête sur une journée, comme c'est fait pour l'EMD, ne traduit pas forcément correctement cette dimension aléatoire des comportements. Ces hypothèses demanderaient à être vérifiées sur d'autres villes, pour ce qui est des données issues de l'EMD, et sur plusieurs semaines pour ce qui est de données issues de la téléphonie.

Par ailleurs, nous voyons qu'une simple lecture naïve de l'activité téléphonique ne permet pas en général de conclure sur le nombre de personnes présentes dans une zone. Et pour cause, les téléphones ne sont pas utilisés de façon linéaire tout au long de la journée. Il est vraisemblable que nous appelons plus facilement à 12h qu'à 23h. De même le lieu semble avoir un effet direct sur l'utilisation du téléphone portable. En zone très résidentielle, on favorisera sans doute le téléphone fixe étant à la maison, alors que dans les commerces ou au travail, le portable sera privilégié.

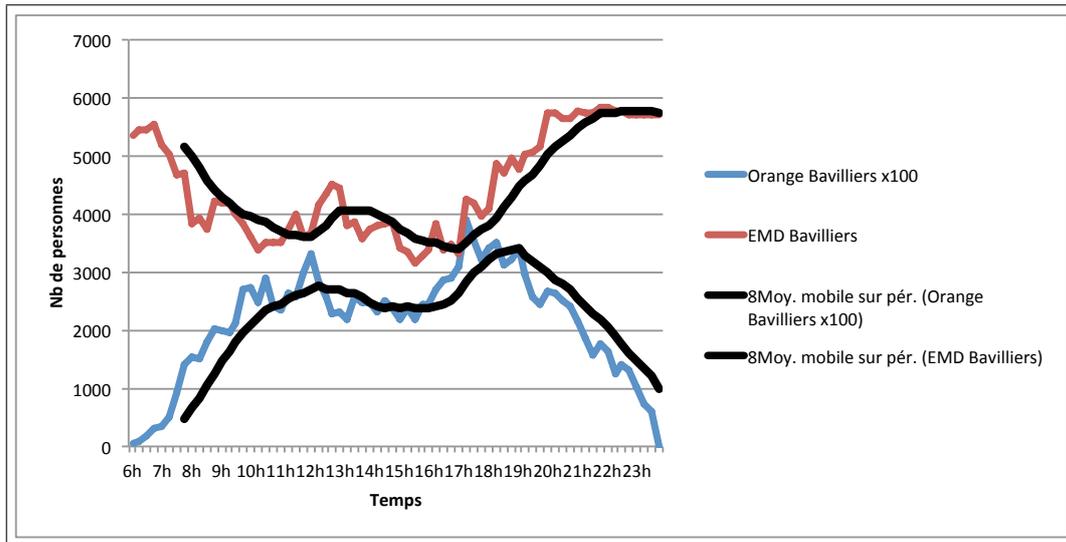


FIGURE 2.21 – Comparaison EMD / modèle en secteur périurbain 2 avec habitat et très peu de commerce

Il apparaît alors nécessaire de ne pas se contenter de traduire l'activité téléphonique directement en niveau de présence mais de déterminer la représentativité de l'activité téléphonique dans un contexte spatio-temporel.

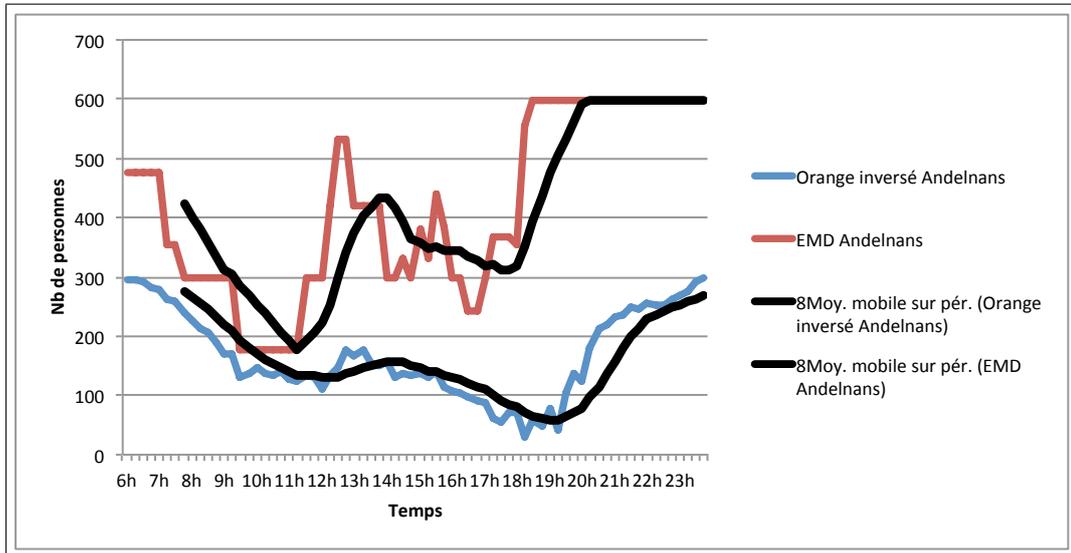


FIGURE 2.22 – Symétrie du modèle en secteur périurbain 1 avec habitat et très peu de commerce

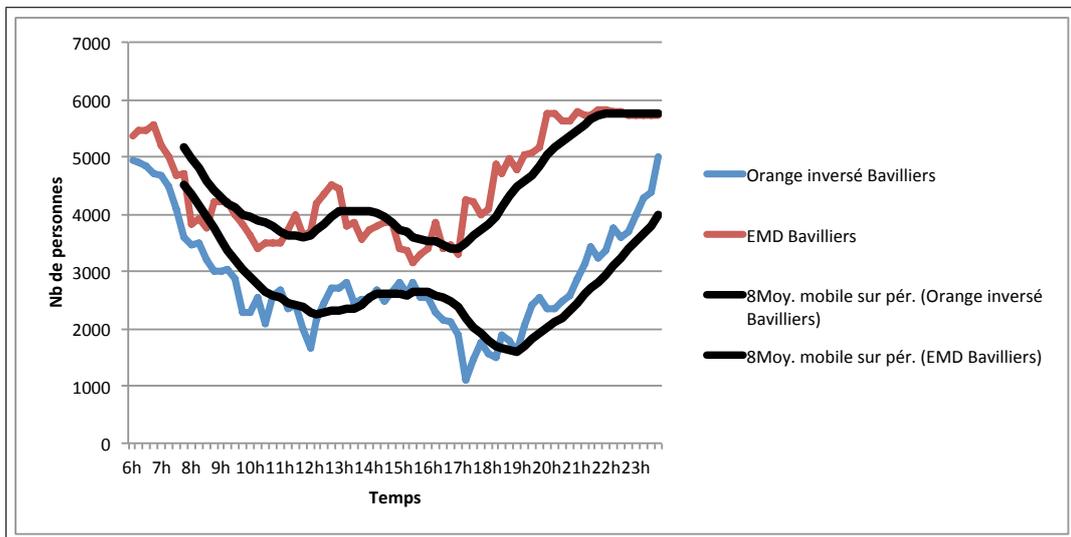


FIGURE 2.23 – Symétrie du modèle en secteur périurbain 2 avec habitat et très peu de commerce

#### 2.4.4 Construction d'une *Signature Territoriale*

Les résultats précédents ont illustré le fait que la façon de téléphoner dépend directement du type de lieu (l'espace) et du moment (la journée, l'heure). Il apparaît donc fondamental d'étudier le niveau de stabilité des communications, non pas au cours de la journée mais d'une semaine à l'autre. En effet, si l'utilisation du téléphone varie de façon significative pour un même type de jour, dans la même zone et au même moment de la journée, aucune conclusion sur la représentativité des communications ne peut être tirée. A l'échelle de l'ensemble des personnes, l'activité téléphonique serait chaotique et ne laisserait aucun espoir d'en tirer une information sur la répartition de la population. Au contraire, si pour une même zone, le même jour et au même moment, nous pouvons observer un comportement stable d'une semaine à l'autre, cela montre que la façon de téléphoner est directement et étroitement liée au lieu. A ce titre les communications constitueraient un bon marqueur de la présence dans une zone.

Pour vérifier ceci, nous introduisons la notion de *Signature Territoriale*. Nous la définissons comme étant le coefficient multiplicateur qui, pour chaque jour type et à chaque moment, permet de passer du nombre d'événements téléphoniques au nombre de présents issus de l'EMD. Ce coefficient devrait certainement être important le soir par exemple, étant donné qu'un appel à 22h représente certainement beaucoup plus de monde présent qu'un appel à 12h. Il s'obtient naturellement par la division euclidienne du nombre de personnes comptées par l'EMD par le nombre d'événements téléphoniques.

**Définition 5 (Signature Territoriale)** *On appelle Signature Territoriale d'une zone la matrice caractéristique permettant d'identifier pour chaque type de jour et chaque horaire le nombre de personnes que représente un événement téléphonique.*

Afin d'étudier la stabilité et donc la pertinence de cette signature, reprenons les zones utilisées précédemment. Nous allons nous focaliser sur les journées connues pour être à la fois semblables et de plus forte mobilité, le mardi et le jeudi [Jeannic 2011]. Par ailleurs, pour renforcer cette validation, nous comparons ces journées au cours de deux semaines différentes. La figure 2.24 illustre les résultats obtenus pour les zones que nous avons identifiées comme étant en phase (centre ville et zone d'activité industrielle). Les courbes hautes concernent le centre ville et les courbes basses la zone d'activité industrielle.

En abscisse est représentée l'heure de la journée, et en ordonnée le coefficient permettant de passer du nombre de personnes issu du modèle proposé au nombre de personnes issu de l'EMD. Concernant les courbes de la zone d'activité industrielle (courbes basses), on constate qu'elles sont à 0 passé 19h. Etant donné qu'il s'agit d'une zone sans habitat, cela signifie que, d'après l'EMD, il y a plus de personnes

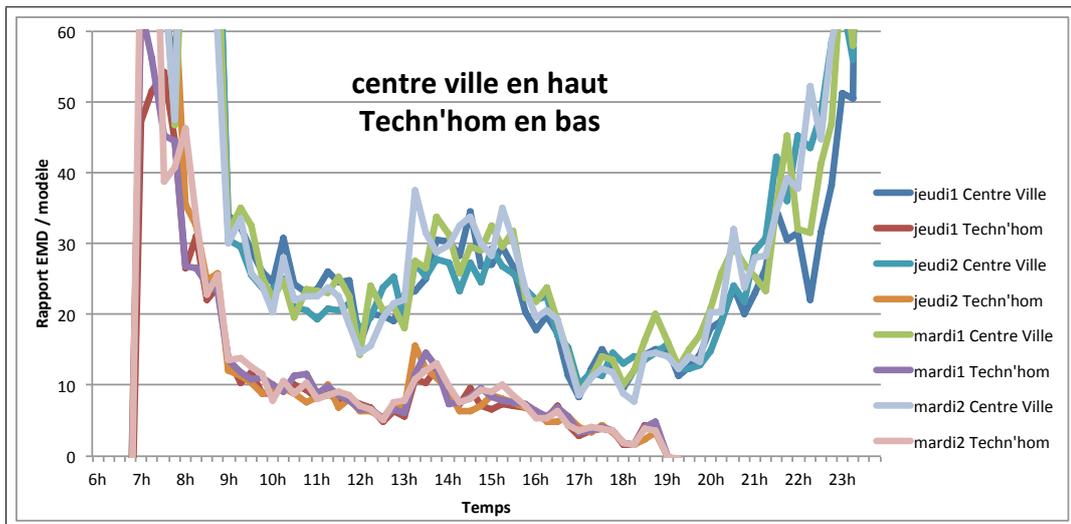


FIGURE 2.24 – Signature Territoriale pour le centre ville (courbes hautes) et la zone d’activité industrielle (courbes basses) - pour deux mardis et jeudis différents

qui sont parties depuis le début de la journée (6h du matin) que de personnes qui sont arrivées. L’EMD indique donc un nombre négatif de personnes passé 19h. Cela montre les limites de ces enquêtes, qui sont la meilleure référence que l’on ait aujourd’hui mais qui comportent de nombreuses incertitudes quant à leurs résultats.

Les 4 courbes de chaque zone se révèlent être très fortement semblables. De plus, d’une zone à l’autre les courbes sont bien distinctes. La signature se révèle être un identifiant propre à chaque zone et stable d’un jour à l’autre.

Nous avons de la même façon construit les signatures correspondant aux zones périurbaines. Le résultat observable sur la Figure 2.25 révèle une tendance très semblable à ce que nous venons de commenter. La signature permet d’identifier clairement chaque zone, avec en haut les courbes de la signature de Bavillers, et en bas celles d’Andelnans.

Nous identifions donc très nettement une tendance stable propre à chaque zone pour les 4 jours observés, quel que soit le type de zone. A ce titre nous montrons ici que la façon dont nous avons utilisé les événements du réseau cellulaire conduit à un marqueur fiable de la présence des personnes. De plus, l’observation généralement faite dans les études de transport que le mardi et le jeudi ont des profils de mobilité très semblables est ici également vérifiée.

Enfin, la lecture détaillée des graphiques fournit des indicateurs sur la représentativité des communications mobiles. Un pic bas à 12h précise au centre ville (Figure 2.24) montre que, proportionnellement au nombre de personnes présentes, on appelle beaucoup plus. Le ratio descend alors à 15 ce qui signifie qu’un événement généré

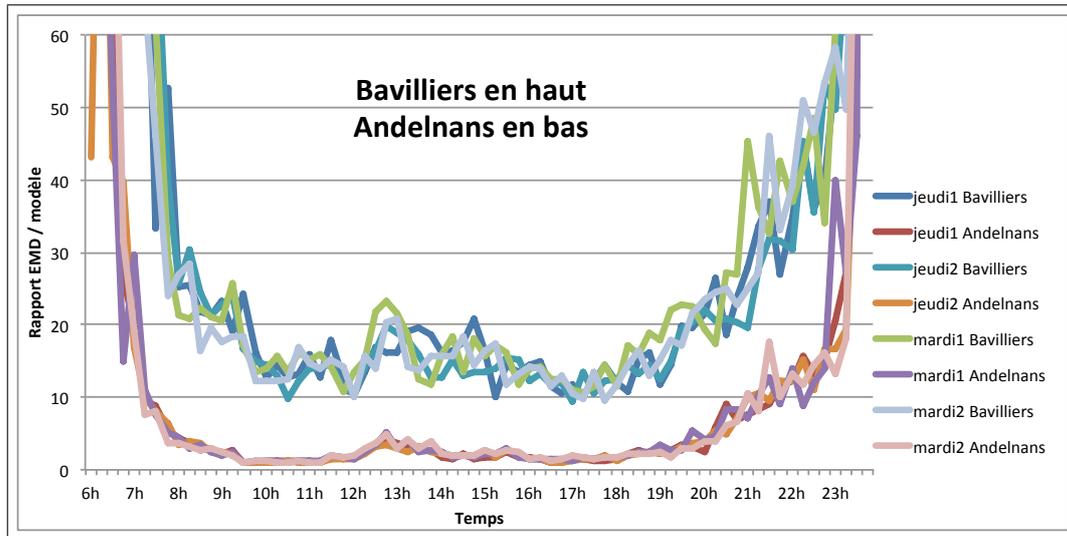


FIGURE 2.25 – Signature Territoriale pour les zones périurbaines - pour deux mardis et jeudis différents

au cours du 1/4 d'heure correspond à 15 personnes présentes. Au contraire, un pic haut vers 15h avec un ratio de 35 signifie qu'un événement téléphonique représente 35 personnes présentes. On constate des pics bas également à 17h et 18h mais pas entre les deux. Sur la zone d'activité le pic de 12h est moins marqué : on téléphone de manière plus étalée dans le temps avec un léger pic à 12h30. On peut également constater un pic haut entre 13h et 14h sur la zone d'activité. Il s'agit d'un moment où le nombre de communications par personne diminue, c'est l'heure du déjeuner. L'ensemble de ces observations demanderait bien entendu à être confronté à des relevés sur plusieurs semaines mais nous ne disposons pas d'autres relevés actuellement.

La comparaison des deux profils de zone permet d'identifier deux tendances distinctes. Alors que les zones de centre ville et d'activité industrielle ont une évolution forte entre 12h et 14h (pic bas suivi de pic haut), ce phénomène est nettement moindre dans les zones périurbaines. De plus, alors que dans les zones périurbaines le niveau de la signature est très semblable le matin et l'après-midi, dans les zones d'activité ce niveau est différencié, nettement plus élevé le matin qu'en fin d'après-midi. Cela montre que dans les zones d'activité l'utilisation du mobile est très dépendante de l'heure et du couplage entre l'heure et l'activité territoriale. Dans les zones d'habitation, nous observons une tendance beaucoup plus linéaire en journée. Entre 9h et 20h, l'utilisation du téléphone se fait de façon assez uniforme, quelle que soit l'heure.

L'étude approfondie de ces *signatures territoriales* se révèle être riche de sens, et fournit de nombreuses informations sur le fonctionnement du territoire.

### 2.4.5 Généralisation de la *Signature Territoriale*

Nous avons montré que la façon d'utiliser le téléphone mobile est globalement très stable d'une semaine à l'autre pour une même zone, au même jour et au même moment. Par contre, elle peut différer significativement d'une zone à l'autre. Nous avons ainsi pu quantifier pour chaque zone le nombre de personnes représentées par l'activité téléphonique. La question qui se pose est de savoir dans quelle mesure cette signature peut être déduite de la constitution de la zone, en termes de données de sursol et socio-économiques. Est-il possible d'identifier des caractéristiques territoriales qui permettent d'estimer la signature correspondante ?

Malgré nos efforts dans ce sens, il apparaît que cette signature territoriale suit des règles particulièrement subtiles que nous n'avons pas encore clairement identifiées. Nous avons eu recours à une approche à base de régression linéaire et d'un estimateur des moindres carrés pour tenter d'identifier la manière dont contribue chaque classe de terrain à la pondération globale des zones. Or nous avons constaté que la façon de communiquer via les mobiles n'est pas toujours directement corrélée à la constitution territoriale des zones (habitat, voirie, commerces, etc.) mais qu'elle intègre d'autres facteurs, peut-être sociaux. Il apparaît ainsi que l'usage des mobiles dans les habitations diffère d'un quartier à un autre. De la même façon les téléphones sont utilisés différemment dans les entreprises selon les zones de localisation des entreprises concernées ; là encore il faudrait regarder d'autres facteurs tels que la nature de l'activité de l'entreprise.

Une orientation qui nous semble prometteuse est le recours aux techniques de classification (souvent appelées clustering) telles que l'analyse en composante principale (ACP) afin de regrouper les zones ayant des profils semblables. Nous laissons cette partie en perspective pour des travaux futurs.

## 2.5 Synthèse

Le modèle proposé dans ce chapitre s'inscrit dans un triptyque téléphonie - mobilité - territoire. Les travaux présentés sont novateurs à plusieurs titres, tout d'abord, de part les données utilisées. Contrairement à ce que font les outils classiques de modélisation de la mobilité à partir d'enquêtes et d'estimations, notre approche vise à fournir des résultats qui sont des mesures de déplacements issus du terrain. Ensuite, ces travaux sont novateurs à travers l'unification de données de déplacement et de données du territoire. Contrairement à ce qui commence à être fait à travers l'utilisation des données issues de la téléphonie mobile, le modèle que nous proposons s'appuie sur la localisation des pôles attracteurs/générateurs de mobilité constitués des habitations, des commerces, des entreprises, des écoles etc., tout ce qui constitue une activité humaine, et à ce titre est un modèle de la catégorie activité-centrée.

Dans une première partie a été présenté le fonctionnement général d'un réseau de téléphonie mobile. L'objet de cette partie était de fournir un éclairage sur les données issues de cette technologie qui peuvent être utiles à une modélisation de la mobilité. Deux types de données ont été identifiés, les données de couverture radio et les données de trafic. Alors que les données de couverture radio s'attachent à définir la portion de territoire couverte par chaque antenne, avec une probabilité de prise de communication pour chaque maille (PPC), les données de trafic rendent compte du nombre d'événements enregistrés par chaque station, événements liés à des communications. Les données de couverture sont des données statiques, fixées par la topographie et le paramétrage des stations. Les données de trafic sont quant à elles dynamiques, variant au cours de la journée et d'un jour à l'autre.

La deuxième partie de ce chapitre contient la description du modèle proposé pour une répartition spatiale et temporelle des personnes. Ce modèle qui vise à affiner les connaissances actuelles sur la mobilité regroupe des données téléphoniques et des données de description du territoire. Il s'appuie sur la notion de classe d'attraction qui, à chaque classe de sursol, attribut un vecteur dynamique de pondérations dépendant du type de jour et du moment de la journée. La mise en correspondance de la cartographie des PPC, de la cartographie des poids d'attraction spatio-temporelle, et du trafic téléphonique dans notre modèle a donné lieu à une distribution spatio-temporelle des personnes. La mise en oeuvre du modèle proposé dans la plateforme geoLogic a donné lieu au développement d'un logiciel d'analyse visuelle de la mobilité. Il offre la possibilité d'étudier des moments et des lieux précis, en continu ou en arrêt sur image sur la carte. Il est ainsi possible d'en tirer des informations utiles pour l'aménagement d'un service. Cela permet par exemple de fixer les fréquences de passage des bus, ou encore de déterminer les meilleurs emplacements pour des arrêts des bus. Sur le territoire considéré, l'étude est basée sur la présence de personnes détectées par leurs communications téléphoniques au cours de la journée.

Enfin, afin d'expérimenter notre modèle, une étude comparative est proposée. Ainsi la distribution des personnes issue de notre modèle est mise en regard de données de localisation construites à partir d'une enquête ménages/déplacements, ainsi que de données statistiques issues du recensement de l'INSEE. Deux types de zones territoriales ont ainsi été identifiés, avec deux types d'utilisation des téléphones portables bien distincts. D'une part il y a les zones de forte attractivité, tel que le centre ville, ou les zones d'activité. Il s'agit des zones particulièrement intéressantes à étudier car elles engendrent une forte mobilité. D'autre part il y a les zones de type résidentiel. On y retrouve les quartiers d'habitation ou les communes périphériques. Ces zones faiblement attractives sont souvent plus difficiles à étudier. Elles suivent un schéma de présence et de mobilité plus variable. L'approche proposée a permis de montrer que dans les zones de forte attractivité la présence de personnes à partir d'enquêtes est étroitement corrélée avec l'activité téléphonique. Dans ce cas, l'activité téléphonique est un bon marqueur direct de la population présente. Dans les zones d'habitation par contre, l'activité téléphonique s'est révélée être en opposition

de phase. Paradoxalement, plus il y a de personnes dans ces zones, moins il y a d'activité téléphonique via les mobiles. Le phénomène mis en évidence est qu'une forte fréquentation de ces zones signifie qu'il s'agit de résidents, qui utilisent de façon préférée le téléphone fixe, et ne génère ainsi que peu d'activité téléphonique sur le réseau mobile. Une étude complémentaire s'avère alors nécessaire pour permettre d'effectuer la transcription de l'activité téléphonique en nombre de personnes présentes. Pour étudier de façon plus fine le lien existant entre l'usage du téléphone et la présence de personnes, le concept de signature territoriale a été proposé. Il a permis de mettre en évidence ce lien, de façon différenciée pour chaque zone étudiée. Cette signature a également permis de vérifier le très bon niveau de stabilité de l'indicateur téléphonique d'une semaine à l'autre, ce qui renforce l'intérêt du modèle que nous avons proposé.

Une modélisation précise des déplacements est un élément essentiel pour guider une décision en matière d'aménagement territorial. Mais souvent, savoir comment une population se déplace n'est pas suffisant. En s'appuyant sur des données de mobilité, la seconde partie de ce mémoire est consacrée à la définition et l'utilisation d'algorithmes permettant d'optimiser les aménagements au sein d'un territoire. Le cas d'étude réel présenté concerne le déploiement d'un service d'auto-partage pour une collectivité. Mais pour introduire cette seconde partie, le chapitre suivant s'attache tout d'abord à présenter les éléments permettant de comprendre les métaheuristiques combinatoires en contexte multiobjectif.

