

L'avènement des méthodes de collectes numériques

La modélisation des transports a connu des évolutions significatives depuis les années 1950. Toutefois, l'autre composante des outils de mesure des déplacements, à savoir la méthode de recueil de données n'a pas connu « de rupture mais plutôt des évolutions graduelles » depuis les années 1970 (KONSTANTINOS, Décembre 2013). Ces évolutions graduelles sont le fruit de besoins d'ordre économique visant à réduire la taille de l'échantillon grâce à l'évolution des techniques d'échantillonnage. A cela s'ajoutent les évolutions sociétales qui entraînent un biais de représentativité parmi les populations refusant les enquêtes. Enfin, le besoin principal est la lutte contre la sélectivité de la mémoire humaine qui omet une partie des déplacements réalisés. En effet, entre 20 et 30% des déplacements ne seraient pas mesurés lors d'interviews (STOPHER, et al., 2007). Néanmoins, l'avènement des technologies de l'information et de la communication semblent ouvrir une porte vers d'autres solutions de recueil.

Ce paragraphe dresse un état de l'art des méthodes de collectes de déplacements. Nous distinguerons deux familles de méthodes de recueil de données : les méthodes actives pour lesquelles les unités (ménages/individus) enquêtées sont directement impliquées et les méthodes passives qui ne nécessitent aucune interaction entre unités enquêtées et enquêteurs. Les avancées technologiques ont permis un déploiement important des méthodes passives grâce à la multiplication de capteurs statiques et mobiles.

1.1 Les méthodes actives de recueil de données

Dans ce paragraphe, nous présentons succinctement les méthodes actives de recueil classiquement mises en œuvre. Une présentation rapide de ces méthodes nous permettra de nous interroger sur leurs avantages et leurs inconvénients pour répondre à la problématique de description fine des comportements des usagers des réseaux de transport.

1.1.1 Présentation des méthodes de recueil

En France, on recense un certain nombre de méthodes de mesure des déplacements que nous qualifierons d'actives car il s'agit d'enquêtes. Ces enquêtes sont réalisées au domicile des ménages, par téléphone ou encore au niveau de points fixes sur des itinéraires (« enquêtes cordons »). Toutes ces méthodes sont validées, standardisées et servent de base de travail pour les différents acteurs du transport car leurs limites sont connues et maîtrisées.

1.1.1.1 L'Enquête National Transport et Déplacements (ENTD)

En France, une seule enquête est réalisée à l'échelle nationale : l'Enquête Nationale sur les Transports et Déplacements (ENTD). Elle est reconduite tous les 10 ans environ. Plus précisément, les ENTD s'articulent autour de quatre thèmes principaux. Le premier est la mobilité régulière qui concerne surtout les déplacements habituels de nature contrainte (domicile-travail, domicile-étude et domicile-garde d'enfant). Deux parties de la mobilité sont traitées par rapport à la distance de l'activité au domicile ; la mobilité locale pour des déplacements vers une activité située dans un rayon de 80km à vol d'oiseau du domicile et les mobilités longue distance. Enfin, la connaissance du parc des véhicules et de leurs usages se décompose en différents axes. Elle recense une description des véhicules possédés par les ménages, la description de leurs usages ainsi que des informations sur la pratique de la conduite (détention de permis, pratiques et accidents de circulation). Une évaluation de l'accessibilité aux transports au commun (abonnement et réductions tarifaires) est également effectuée. Le caractère complet d'une telle enquête permette de traiter d'une vaste gamme de sujets allant de la concurrence entre modes de transport, la spécification en fonction des motifs, en passant par la consommation d'énergie et les déplacements touristiques. La dernière ENTD date de 2007-2008. Elle a été conduite par le ministère chargé des transports et l'INSEE (Institut National de la Statistique et des Etudes Economiques) qui assure la maîtrise d'œuvre. La maîtrise d'ouvrage est déléguée au SOeS (Service de l'Observation et des Statistiques prospectives). Les méthodes de recueil sont standardisées et garantissent ainsi des comparaisons dans le temps et dans l'espace (ARMOOGUM, et al., 2007-2008). Evidemment, la richesse d'une telle enquête à l'échelle nationale entraîne des coûts temporels. Par exemple, la réalisation de l'ENTD de 2008 s'étend de mai 2007 à avril 2008.

1.1.1.2 Les Enquêtes Ménages-Déplacements (EMD)

Les EMD (Enquêtes Ménages-Déplacements) sont une méthode de collecte classique et largement répandue pour l'étude des déplacements quotidiens des populations urbaines (ORTUZAR, et al., 2011). En France, les EMD existent depuis plus de trente ans. Les définitions, les concepts et la méthode de recueil ont peu évolué pendant cette période afin d'assurer la continuité des séries statistiques. Néanmoins, les enquêtes dans leur contenu se sont adaptées afin de tenir compte des avancées de la recherche en statistique, mais surtout des nouveaux besoins exprimés par les collectivités locales (CERTU, janvier 2013). De telles enquêtes sont standardisées par l'Etat français car elles constituent un outil essentiel pour

l'Etat dans son rôle d'évaluateur des politiques publiques. Cela garantit une méthode de recueil de données harmonisée, rigoureuse et pérenne car les données sont fiables et comparables dans le temps et dans l'espace.

Les enquêtes ménages sont un des outils essentiels à la réalisation d'études de déplacements décrivant la mobilité quotidienne des personnes habitant dans un périmètre donné. Ce type d'enquête est réalisable à plusieurs échelles spatiales du territoire : EMD pour les agglomérations, EDVM (enquêtes Déplacements Villes Moyennes) et EDGT (Enquêtes Déplacements Grand Territoire). En France, elles sont réalisées par les collectivités locales qui reçoivent une assistance de la part des CETE (Centres d'Etudes Techniques de l'Equipement). Des informations relatives à la maîtrise d'œuvre sont mises à disposition par exemple par le CERTU (Centre d'Etudes sur les Réseaux de Transports, l'Urbanisme et les constructions publiques). Rassemblées dans une base de données nationale, ces informations permettent des études et une évaluation des politiques de déplacement urbain (PDU).

La méthode « standard CERTU » repose sur quelques principes essentiels :

- Les enquêtes sont réalisées au domicile des ménages dont chaque personne âgée de plus de 5 ans est interrogée.
- Les recensements portent sur les déplacements effectués la veille du jour d'enquête.
- L'ensemble des paramètres de chaque déplacement est recueilli (motif, mode(s), origine, destination, heure de départ et heure d'arrivée).
- Les enquêtes portent sur un échantillon représentatif de la zone étudiée. Cet échantillon est tiré aléatoirement.

Une enquête « classique » contient quatre questionnaires portant sur le ménage, les personnes, les déplacements et les opinions. Des questions supplémentaires peuvent être ajoutées à l'enquête. Les enquêtes ménages-déplacements ont pour vocation de reconstituer la mobilité des habitants d'une zone lors d'un jour ouvrable « moyen », c'est-à-dire les mardis, mercredis et jeudis hors des périodes de vacances scolaires et jours fériés, de mi-octobre à mi-avril. Cela implique des contraintes fortes dans le déroulement de l'enquête. En effet, les actifs ne sont généralement présents à leur domicile qu'à partir de 18h30. Alors, la durée d'enquête sur le terrain est de 10 semaines en moyenne. La Figure 23 illustre le type de données disponibles pour chaque individu. On notera la richesse des informations fournies sur les activités réalisées, les modes de transport utilisés, ainsi que les horaires favorisés. On notera également que la dimension spatiale est plus problématique. En effet, la localisation exacte des individus n'est pas connue. Chaque activité est ainsi localisée approximativement dans une zone de déplacement, tandis que les trajets réellement effectués par les individus dans le réseau de transport ne sont pas recensés, ou avec un biais non négligeable (CERTU, janvier 2013).

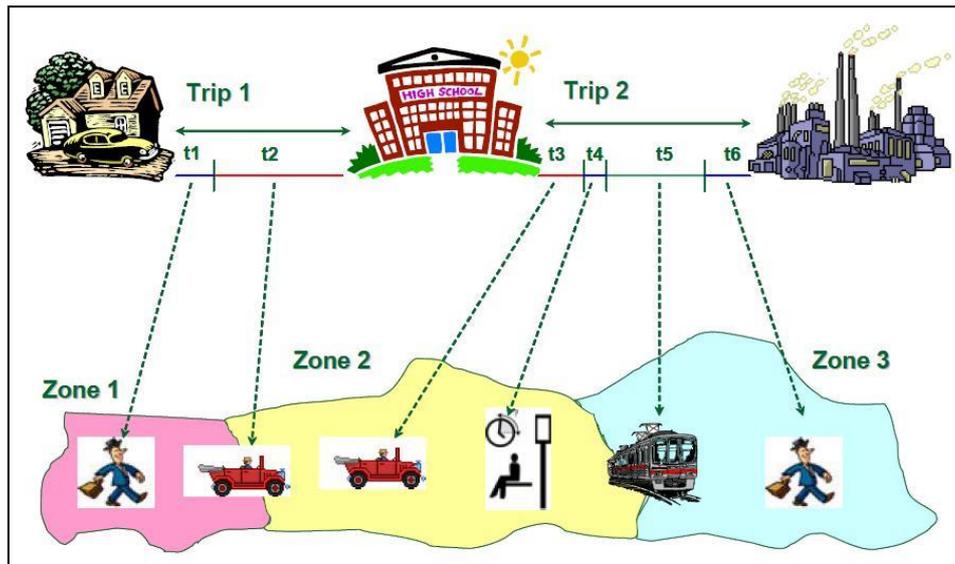


Figure 23 : schéma des informations fournies par les EMD

Une EMD particulière, appelée Enquête Globale Transport (EGT) a lieu tous les 10 ans sur les ménages franciliens. Elle est pilotée par le STIF (Syndicat des Transports de l'Île de France) et la DRIEA (Direction Régionale Interdépartementale de l'Équipement et de l'Aménagement). L'EGT a pour vocation de dresser une vue d'ensemble de la mobilité des franciliens, essentielle pour définir les politiques de déplacements adaptés aux besoins. De plus, cette enquête renseigne sur l'équipement des ménages ainsi que la détention d'un permis de conduire et/ou d'abonnement à des modes collectifs. Cependant, certaines facettes de cette enquête nécessitent des améliorations. A ce titre, la méthodologie d'enquête de l'EGT de 2010-2011 porte une attention particulière aux déplacements courts alors que les éditions antérieures présentaient quelques lacunes quant à l'observation de ces déplacements courts et/ou occasionnels qui n'étaient pas toujours mentionnés, mais aussi quant à la précision de l'itinéraire réellement parcouru. Cette nouveauté dans la méthodologie entraîne des difficultés d'interprétation d'évolution de certains indicateurs par rapport aux enquêtes précédentes (OMNIL, juillet 2012). Ce nouvel inconvénient introduit s'ajoute à différentes limitations de cette source pourtant nécessaire et très exploitée par les acteurs du transport. Tout d'abord, on peut noter des « imperfections » dans la construction de l'enquête. Par exemple, la distance euclidienne est appliquée au lieu de la distance réelle parcourue sur le réseau de transport. Ceci accroît les imprécisions sur les estimations vitesses de déplacement d'autant que les durées de temps de trajet sont déjà recueillies comme interprétations des enquêtés. De plus, la complexité du questionnaire le rend difficile à administrer, et les enquêteurs doivent faire face à une augmentation du refus de recevoir des enquêteurs à domicile. Ensuite, le dépouillement et l'analyse des données est un processus extrêmement long d'environ 18 mois en raison de la richesse des enquêtes mais aussi de leur construction et leur processus de recueillement « manuel ». Finalement, le coût (environ cinq millions d'euros pour l'actuelle EGT 2009-2011 en Île-de-France pour 18 000 ménages) est un réel

frein à la mise en place d'enquêtes de cette envergure. Ce coût est dû en grande partie à l'intervention humaine. En effet, les interviews sont en face-à-face et durent 1h30 en moyenne, les enquêteurs se déplacent à domicile et leur taux de productivité journalier est inférieur à un. Ce coût élevé a pour conséquence une fréquence de réalisation faible d'une enquête, tous les dix ans en moyenne.

1.1.1.3 Les enquêtes Origine-Destination

Les enquêtes O-D ont été très utilisées dans les années 1960 et 1970 dans les grandes aires urbaines des pays industrialisés et dans les quelques grandes villes des pays émergents. Ce type d'enquête est principalement employé pour deux applications : la prévision de trafic et l'alimentation des modèles activités-centrés lorsque les informations sont suffisamment fines et que les données sur l'infrastructure des activités (heure d'ouverture des lieux d'activités) sont disponibles. Dans le premier cas d'application, l'objet principal d'intérêt est le trajet entre une origine et une destination, plus que les composantes comportementales associées, d'où le terme enquête O-D. De manière générale, l'échantillon de données doit avoir les caractéristiques suivantes :

- Recueil de données basées sur « un trajet », celui-ci étant relié aux modes de transports empruntés, aux lieux accédés, au temps de parcours, à la longueur du trajet, etc.
- L'inclusion de tous les modes de transports, même les modes doux
- Des mesures au niveau très désagrégé, voir individuel si possible, des motifs de déplacements
- Une couverture temporelle la plus large possible (au moins une journée de 24h)
- Un recueil des informations de mobilité de tous les membres d'un ménage
- Une qualité des données suffisamment robuste pour une utilisation désagrégée

1.1.1.4 Les enquêtes cordons

Ce type d'enquêtes a pour vocation la mesure de flux autour d'une zone. L'étendue de l'enquête est délimitée par une ligne imaginaire appelée « cordon ». Il s'agit le plus souvent des limites communales, ou de quartiers. Cette bordure extérieure, ou cordon externe est choisie en fonction des besoins de l'enquête. Le but de cette délimitation est de pouvoir quantifier spatialement des variables telles que la population ou le nombre d'emplois (ORTUZAR, et al., 2011).

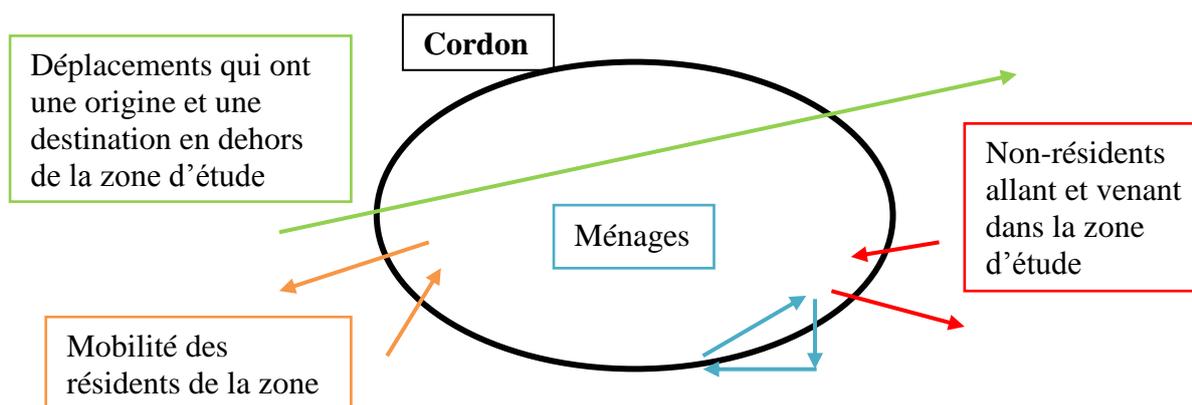


Figure 24 : représentation des éléments observés lors d'une enquête cordon

Etudier uniquement les déplacements à l'intérieur du cordon n'est pas suffisant car, comme indiqué sur la Figure 24, des déplacements sont susceptibles de traverser la zone d'étude. Pour une étude complète, les données suivantes sont nécessaires :

- Une enquête ménage déplacement sur les ménages de la zone d'étude
- Des données cordons, c'est-à-dire sur les déplacements qui franchissent le cordon, qu'ils soient entrants ou sortants de la zone.
- Des données de comptage de trafic
- D'autres informations comme des données comportementales, d'usage des sols, sur les infrastructures du territoire étudié, etc.

Ce type de recueil de données présente les mêmes limites que les méthodes citées précédemment en termes de coûts, de représentativité des échantillons et de leur caractère statique. De plus, par essence, elles sont principalement utilisées pour des comptages avec le minimum d'informations concernant les attributs des déplacements. Cette méthode ne permet pas de recueillir de connaissance sur les déplacements internes à la zone d'étude. Comme les enquêtes Origine-Destination, cette méthodologie est souvent déployée en complément d'autres enquêtes.

1.1.2 Des méthodes adaptées aux nouveaux besoins ?

Le caractère actif de l'enquête vient de l'interaction directe entre un enquêteur et un enquêté. Dès lors, la mise en place de telles enquêtes est lourde, coûteuse en temps et en ressources. Un compromis est à trouver entre la quantité d'informations à questionner, l'étendue de la zone d'enquête, la durée de l'enquête, la fréquence de renouvellement de l'enquête et bien sûr des ressources disponibles pour sa réalisation.

Néanmoins, les questionnaires de ces enquêtes sont très complets et permettent d'avoir accès à un très large panel de caractéristiques d'une population. De plus, la désagrégation jusqu'au niveau de l'individu est atteignable. Ces enquêtes constituent une source très riche d'informations sur les usagers des transports. Notamment la pratique du

réseau de transport et les composantes personnelles de choix des attributs des déplacements sont relatés. Ainsi, le point fort de telles sources est la possibilité qu'elles offrent de mettre en relation des facteurs individuels avec les déplacements qui ont effectivement été réalisés. Cette méthode de collecte permet notamment d'accéder à une catégorisation des activités ainsi qu'à des caractéristiques individuelles de démographie sociale, à l'instar des méthodes passives classiques préférées pour les études à objectifs larges (JIANG, et al., 2012).

Cependant, certaines limitations de cette méthode de recueil peuvent la rendre mal ou non adaptée aux nouveaux besoins de modélisation. Premièrement, le point discriminant est le besoin d'accéder aux caractéristiques d'une population de manière quasi-exhaustive. En effet, cette connaissance est nécessaire pour l'analyse des facteurs déterminant les comportements de mobilité. Or, pour réduire les coûts budgétaires et temporels, l'échantillon d'enquêtés doit être réduit au maximum, tout comme la fréquence des enquêtes. Ne serait-il pas intéressant d'avoir des informations moins complètes sur un plus grand nombre d'enquêtés et plus fréquemment ? Deuxièmement, les échantillons de ces enquêtes sont constitués à partir des ménages recensés en tant que résidents d'un périmètre. Les segments d'individu « de passage » sur une zone tels que des touristes ou des excursionnistes n'apparaissent que dans les enquêtes menées dans leur zone de résidence. De plus, les évolutions sociétales vont biaiser la « population enquêtable » par le fait que la volonté de répondre à une enquête est en forte baisse, et ainsi ce sont généralement les personnes les moins mobiles qui s'avèrent être les plus disponibles pour des enquêtes (KONSTANTINOS, Décembre 2013). Enfin, pour des raisons de coût, les tailles des échantillons sont réduites ainsi que le périmètre d'enquête. Ainsi, les méthodes actives, souvent restreintes dans l'espace, imposent alors des limitations lors des études des caractéristiques d'un système urbain dans son ensemble, notamment pour l'identification des « clusters » de mobilité (YUAN, et al., 2012). Un compromis est donc à trouver entre l'accès à des variables explicatives des comportements de déplacements et des contraintes liées à la taille de l'échantillon, sa représentativité et la fréquence de renouvellement de l'enquête (NATURE, 2008). Troisièmement, une grande majorité de notre compréhension des systèmes urbains provient de ces méthodes actives considérées comme traditionnelles. Bien que ces méthodes permettent d'accéder à des informations très détaillées sur les comportements urbains, les difficultés de mises à jour des jeux de données limitent la possibilité d'une observation dynamique et continue des phénomènes engendrés par les déplacements urbains (READES, et al., 2007). De plus, la procédure même d'enquête « porte à porte » ne permet pas à cette méthode de collecte de sortir de son caractère statique. Ainsi, des phénomènes évoluant dans une plage temporelle et spatiale faible, à l'échelle de la journée par exemple, telle que la congestion, ne peuvent être recensés. Enfin, les enquêtes ne peuvent pas s'affranchir de la « sélectivité » de la mémoire humaine. En effet, il est rare d'avoir accès à des itinéraires précis entre deux activités, même pour des trajets quotidiens, et les variations sont souvent oubliées. Ainsi, les trajets à caractère exceptionnel, tels que ceux pour motifs achats ou loisirs, ou encore les trajets secondaires d'une chaîne, sont assez difficiles à recenser. De plus, la différenciation du jour de la semaine ou de la saisonnalité ne fait généralement pas partie des enquêtes classiques. Par exemple, les EMD ont pour but de représenter les mobilités pour un

jour ouvré moyen. Ainsi, les comportements inhabituels, ou les faibles variations par rapport à des itinéraires habituellement empruntés sont très difficiles à mesurer.

La construction de nouveaux modèles de demande a entraîné des évolutions des méthodes d'enquêtes (BONNEL, 2004). Tout d'abord, afin de capturer le jeu de contraintes et d'opportunités qui structurent la mobilité quotidienne, il a fallu prendre en compte l'environnement social et spatial des individus. Alors, les méthodes de recueil des activités ont été développées afin d'alimenter des approches centrées sur les activités (AXHAUSEN, 1995). Ensuite, est apparue la méthode de préférences révélées afin d'estimer l'impact de facteurs qui influencent la manière de se déplacer et qui s'avéraient très difficiles à mesurer (STOPHER, 1998). Elle permet de quantifier de manière subjective des facteurs tels que le confort, la propreté, le design ainsi que la perception de la valeur du temps pour un usager grâce à des scénarii hypothétiques (ORTUZAR, et al., 1994). Or, un phénomène similaire à la sélectivité de la mémoire humaine face à des événements vécus ne peut-il pas se produire sur une projection erronée de réaction dans une situation hypothétique ? Enfin, dans le but d'évaluer finement l'incidence de la modification de l'offre de transport sur les choix des caractéristiques du déplacement, ces dernières doivent être observées finement, spatialement et temporellement. L'apport d'un repérage spatial fin des données est illustré par le modèle MADITUC construit à partir des données des enquêtes régionales origines-destinations de Montréal et de Toronto (CHAPLEAU, et al., 1997).

Les évolutions des modèles à 4 étapes sont de plus en plus sensibles au manque de précision, particulièrement quant à la localisation des activités. De même, la privatisation croissante des infrastructures (péages) entraîne une exigence croissante par rapport à la précision et à la qualité des informations sur la mobilité des personnes. Alors, au milieu des années 1990, une nouvelle forme de questionnaires est proposée pour lutter contre la mémoire « sélective » des enquêtés qui oublient un certain pourcentage des déplacements effectués et sont moins enclins à répondre à de telles enquêtes. Ainsi, l'apparition de nouvelles technologies telles que le GPS (Géolocalisation Par Satellite) ont permis de se poser la question suivante : pourrait-on remplacer les « enquêtes ménages déplacements » par d'autres méthodes de recueil pour pallier aux inconvénients présentés jusqu'alors ? (KONSTANTINOS, Décembre 2013).

1.2 Les capteurs numériques de déplacements

1.2.1 Vers des méthodes de recueil passives ?

Certains concepteurs d'enquêtes de mobilité ont fait le choix d'enquêtes assistées par la technologie GPS pour assurer la continuité des enquêtes classiques en place depuis plus d'un demi-siècle. En effet, les systèmes GPS engagent de réduire les coûts d'enquête tout en permettant un accès automatique à l'origine et la destination d'un déplacement, aux horaires précis de départ et d'arrivée, à la localisation des activités et à la longueur des trajets (WOLF, juillet 2000). En 1996 a lieu l'étude-pilote de Lexington, Kentucky, commanditée par l'« US Department of Transportation ». L'objectif poursuivi était de mesurer l'écart entre les pratiques de déplacement réelles et celles reportées dans les EMD. Des ménages ont été

équipés de GPS couplés de PDA installés en voiture pour enregistrer les déplacements sur une période de 6 jours et ont parallèlement participé à une enquête classique. Les résultats montrent qu'un nombre supérieur de déplacements ont été enregistrés via la solution GPS. De plus, les déplacements courts ont été mieux détectés. (WAGNER, 1997). Ainsi, dans la continuité des enquêtes actives, les premières études pilotes ont nécessité l'accord des ménages d'être équipés de GPS et de dispositifs techniques. La rupture avec les méthodes traditionnelles passe par la pacification des enquêtes, c'est-à-dire en s'affranchissant de toutes interactions entre un enquêteur et un enquêté. Alors plusieurs questions se posent. Par exemple, comment identifier le motif de déplacement ? Les SIG semblent apporter des éléments de réponses (WOLF, et al., 2008). Une enquête à Cincinnati en 2009 sur environ 5000 ménages équipés de GPS portables a montré qu'il est possible d'identifier avec un degré de précision de 96% le mode utilisé et une similarité des résultats de 90% a été obtenue en comparant ces résultats avec ceux obtenus par les questionnaires on-line remplis par les mêmes enquêtés (STOPHER, et al., février 2012).

Dans la suite de ce paragraphe, nous présenterons quelques méthodes de recueil de données dites passives décrites dans la littérature. En effet, la métrologie des déplacements peut se faire via diverses technologies de captures que nous diviserons en deux groupes : les systèmes « fixes » et les systèmes « mobiles ». Nous faisons le choix de faire porter l'adjectif « fixe » à un système de capture si le déclenchement de capture se produit en un lieu fixe (péage, tronçon de routes, intersection). La détection qui produit la collecte d'information peut se faire via un matériel embarqué (carte à puce) ou non (boucles électromagnétiques). Selon les dispositifs de détection de présence en un point fixe, il est possible d'associer un ensemble de points mesurés à un individu et donc de produire des trajectoires individuelles. A contrario, un système de capture est considéré « mobile », si le déclenchement de la mesure ne dépend pas d'un lieu particulier. Ces systèmes permettent de générer des successions de mesures dont le nombre et les intervalles de temps inter- mesure varient selon les mécanismes déployés. Un système de capture « mobile » peut être utilisé comme un système de capture « fixe » si l'on décide de déclencher toutes les mesures en des points fixés par avance. Dans tous ces mécanismes de collecte, les informations d'instant de capture et de localisation sont disponibles à minima.

1.2.2 Les systèmes de capture « fixes »

On peut classer la famille des systèmes de capture « fixes » en deux sous-familles. Par analogie avec l'approche Eulérienne en mécanique des fluides, une première sous-famille de capteurs fixes permet de mesurer le flux (de déplacements) en un point. La deuxième sous-famille est naturellement basée sur une approche Lagrangienne qui permet l'observation d'un écoulement entre deux points fixes.

1.2.2.1 Approche « Eulérienne »

Nous débuterons la présentation des systèmes de capture fixes par les réseaux de capteurs qui sont insérés dans l'infrastructure de transport et ne nécessitent pas d'équipements particuliers pour la détection des usagers. On peut assimiler cette première méthode de capture à une approche « Eulérienne », dans le sens où l'on observe un flux en un point de l'espace. La nature même de ces technologies de capteurs confère le caractère passif à la méthode de recueil. L'application largement majoritaire de ces systèmes de capture est la mesure de flux de trafic en un point du réseau de transport (ROSE, May 2006). Ils sont largement répandus sur les autoroutes ou les routes à forte densité de voyageurs afin d'observer le trafic en temps réel et sont, dans la grande majorité des cas, propriétés des entités opérant les portions de voies surveillées. Il s'agit de technologies intrusives car les capteurs sont placés sur ou sous les portions des voies. On peut citer comme technologies de capteurs les tubes pneumatiques, les capteurs piézoélectriques ou encore les boucles électromagnétiques. Cette dernière est la technologie la plus déployée. Selon les portions considérées, l'espacement entre deux boucles successives varie. Elles sont principalement utilisées pour remonter des informations d'occupation de la route et de flux, mais aussi pour produire des vitesses moyennes sur des segments. Des études se sont concentrées sur l'identification des véhicules à partir des boucles pour produire des temps de parcours point-à-point (ABDULHAI, et al., 2003). Cependant, cette identification est très délicate. De plus, bien qu'étant une technologie mature et approuvée, les boucles électromagnétiques demandent un coût de maintenance et de génie civil lourd qui ne permet pas un déploiement à large échelle sur le réseau de transport routier. Intéressons-nous alors aux méthodes non-intrusives basées sur des observations à distance. On peut citer le comptage manuel qui permet d'obtenir des informations complémentaires concernant le nombre de passagers dans les véhicules, de piétons ou de cyclistes, les capteurs infrarouges (actifs et passifs), les radars micro-ondes, les systèmes acoustiques ou encore les caméras vidéo. Ce dernier système permet également d'identifier le véhicule grâce à la reconnaissance des plaques d'immatriculation (LAPI : Lecture Automatique de Plaques d'Immatriculation). Il est donc possible de calculer les temps de parcours entre deux lieux équipés du système. Ces équipements électroniques disposent d'un algorithme de reconnaissance automatique de plaques (WASSON, et al., June 2008).

Ainsi, ces technologies de capteurs fixes sont matures et ont prouvé leur performance. Néanmoins, elles ont un inconvénient principal : leurs coûts d'installation et de maintenance. De plus, la défaillance des équipements et la sensibilité de certains systèmes aux mauvaises conditions météorologiques diminuent la fiabilité des mesures réalisées. Qui plus est, elles permettent la mesure des déplacements routiers exclusivement, et sur une partie seulement du réseau de transport (certaines portions des réseaux autoroutiers). En zone urbaine, certaines intersections munies de stops ou de feux en sont équipées, mais la différenciation entre les types de véhicules reste à faire pour isoler les transports en commun des flux de voitures individuelles. Enfin, le principe de fonctionnement des technologies de capteurs fixes produit majoritairement des données agrégées. Ainsi, elles ne permettent pas

l'accès à des trajectoires individuelles ni à aucune caractéristique des individus en déplacement.

On notera qu'une approche « Eulérienne » est très adaptée au comptage en un point mais nécessite des traitements supplémentaires pour produire des informations critiques pour les exploitants telles que la vitesse moyenne ou le temps de parcours d'un tronçon. En effet, la nature intrinsèque des technologies de capteurs présentées jusque-là ne permet pas d'identification immédiate des usagers/véhicules en écoulement sur un axe de transport. Une approche « Lagrangienne » est à favoriser pour produire ces informations car par nature elle permet de construire des ensembles de trajectoires. Le paragraphe suivant présente les systèmes de captures « fixes » dont le déclenchement de la mesure se produit grâce à un équipement embarqué qui peut contenir un identifiant unique de l'utilisateur et/ou du véhicule.

1.2.2.2 Approche « Lagrangienne »

Les systèmes de capture « fixes » qui permettent une approche « Lagrangienne » requièrent une paire constituée d'un réseau de capteurs de localisation fixes et d'un équipement embarqué contenant un système d'identification par usager/véhicule. Le déclenchement de la mesure se fait lorsque l'équipement embarqué se trouve au niveau d'un capteur fixe et permet l'identification du détenteur de l'équipement. Parmi ces équipements, on peut citer les cartes à puce d'abonnement à un réseau de transport. Par exemple, en Ile-de-France, les abonnés du réseau de transport en commun ferré valident leur admission dans le réseau en portant leur carte à puce (Pass Navigo) au niveau des barrières de péages des stations. Ainsi, chaque validation est horodatée et localisée à la station pour un identifiant d'abonné. Il est possible alors de reconstruire des volumes par paire de stations ainsi que des temps de parcours (BERTINI, et al., 2003). Un procédé équivalent est envisageable pour les barrières de péages routiers via le télépéage. Or, ces abonnés sont des usagers réguliers, ce qui peut entraîner un biais de représentativité. Il est faible dans le cas des transports en commun francilien mais ce n'est pas le cas pour les réseaux autoroutiers. En effet, la majorité des abonnés des télépéages sont des professionnels à forte mobilité et une partie des résidents à proximité d'une section payante d'autoroute sans itinéraire alternatif. D'autres technologies d'identification peuvent être envisagées au niveau des barrières de péages comme la reconnaissance de plaque ou encore un suivi d'une carte à puce fréquemment utilisée par son détenteur : sa carte de crédit. En effet, une étude sur la mobilité a été réalisée en suivant l'activité de retraits bancaires par carte de crédit aux Etats-Unis (BROCKMANN, et al., 2006). Néanmoins, un tel système de capture ne peut être envisagé en France sans l'accord de toutes les personnes enquêtées car l'exploitation de données à caractère personnelle n'est pas légale en France sans l'accord signé des individus ayant pris connaissance du cadre d'utilisation de ces informations dont ils sont les propriétaires, comme nous le verrons par la suite (paragraphe 2.2) Or dans ce cas l'adjectif "passif" n'aurait plus lieu d'être.

1.2.3 Les systèmes de capture « mobiles »

Le déclenchement d'une mesure par des systèmes de capture "mobiles" peut se produire en tous les points du territoire. En pratique, ces points ne peuvent donner lieu à une mesure car les technologies de capteurs « mobiles » sont soumises aux conditions de transmission du signal requise pour le fonctionnement du système de mesure. Il existe deux systèmes principaux pour la production de mesure de déplacements via un système de capture mobile : le système basé sur la technologie GPS (Global Positioning System) et le système basé sur les réseaux de téléphonie mobile. Dans les deux cas, le principe est basé sur la communication radio entre un réseau fixe (antenne) et un équipement terminal mobile qui se déplace avec l'unité à observer (véhicule, individu). Des mesures horodatées et géo-localisées sont remontées avec un pas de temps fonction de la technologie et pas nécessairement constant.

Les systèmes de capture basés sur la technologie GPS (Global Positioning System) permettent une localisation « outdoor » avec une précision à la dizaine de mètres. Le terminal portatif peut être embarqué dans un véhicule ou directement sur l'utilisateur. Le premier cas se matérialise par des flottes de véhicules équipés (taxi, véhicules de location, véhicules produits pour le système Vehicle Infrastructure Integration (VII) ...). On citera comme cas d'application la mesure de vitesse moyenne sur des portions de route pour la détection d'incident ou de congestion. Le deuxième cas est l'utilisation de téléphones mobiles équipés d'A-GPS (Assisted-GPS) comme dans l'étude menée par le Transportation Development Center du Canada en 2005 (KIRK, et al., 2005). Dans les deux cas, la taille de l'échantillon est limitée soit par le nombre de véhicules de la flotte, soit par le nombre de téléphones équipés et dont les utilisateurs ont accepté de partager les informations relatives à leur localisation. Le déterminisme de l'échantillon créé par cette méthode introduit un biais. Ces méthodes posent les mêmes questions sur la taille de l'échantillon et la minimisation du biais que les enquêtes actives. De plus, par le fait du fonctionnement de la technologie GPS, au moins trois satellites, voire quatre dans le cas du A-GPS, sont nécessaires et les consommations de batterie réduisent considérablement les périodes d'étude. Les remontées de traces GPS et leur exploitation pour produire de l'information sur le trafic routier est une technologie mature sous réserve des limites précédemment mentionnées. Néanmoins, il s'agit également d'une technologie qui ne s'applique pas à tous les modes, notamment pour les transports en communs ferrés avec des sections souterraines. En effet, la couverture GPS des infrastructures souterraines, des tunnels, voire même de certains bâtiments n'est pas technologiquement fonctionnelle.

Les technologies des capteurs embarqués sur la personne ne cessent de se répandre. Accéléromètres, gyroscopes, capteurs « d'ambiance », etc. sont autant d'outils de mesure disponibles sur des téléphones mobiles qui ne quittent presque jamais leur propriétaire. Cela constitue une source de données exploitables pour mesurer l'activité humaine, de la marche à l'immobilité, du déplacement à bord d'une voiture au déplacement à bord d'un train de banlieue (BAO, et al., 2004). Cependant, le taux d'équipement de la population en « capteurs portables » sophistiqués et la segmentation des personnes en possédant introduisent un biais.

Existe-t-il des technologies de capteurs plus largement répandues dans la population ? On peut tout simplement citer les téléphones portables dont les Smartphones sont la dernière évolution. En effet, même si les mobiles de dernière génération hébergent GPS, accéléromètres, etc. ; ils ont été développés à l'origine pour communiquer en tout point d'un réseau de télécommunication, comme les mobiles des anciennes générations. Ce sont des objets de la vie courante qui nous suivent dans tous nos déplacements. Ainsi, avec des méthodes adéquates de recueil de données et de traitement, connectés à un réseau cellulaire, les téléphones portables forment un système de capture des déplacements. Ce constat constitue l'essence du travail de thèse présenté. Dans la suite, nous présenterons les différents types de mesures réalisables à partir du réseau de téléphonie mobile selon la localisation du point de collecte dans l'architecture du réseau de télécommunications.

1.2.3.1 Les points de collecte d'un réseau de téléphonie mobile

Par construction des réseaux de télécommunication sans fil, la connaissance de la position approximative des téléphones mobiles est imposée. Cette connaissance permet d'une part de respecter des standards de sécurité (appels d'urgence) et d'autre part d'assurer l'acheminement des communications en tout point d'un territoire couvert. Pour assurer la fonctionnalité de communication, l'architecture du réseau est composée de plusieurs éléments entre lesquelles des informations transitent. Ainsi, il existe plusieurs points de collecte de données radio-mobiles. Selon les points de collecte, les données sont de nature différente. Dans cette section, nous présenterons une version simplifiée de l'architecture réseau afin de décrire les parties du réseau de télécommunication dont sont communément extraites « des données brutes » nécessaires au bon fonctionnement du réseau mobile. Puis, nous ferons un état de l'art des méthodes qui permettent de déterminer une mesure de déplacement à partir de ces informations « brutes ».

Tout d'abord, présentons un schéma très simplifié d'un réseau de télécommunication dont la vocation première est d'illustrer la nature des informations disponibles au niveau des points de collecte. Une description détaillée de l'architecture du réseau de téléphonie et son fonctionnement est l'objet principal du paragraphe 2. Un réseau de télécommunication est composé d'une partie mobile, un ensemble de téléphones cellulaires, et d'une partie fixe (voir Figure 27). Pour faciliter la compréhension, nous avons divisé cette partie fixe en deux entités. Chacune de ces entités correspond à une fonctionnalité et peut contenir un ou plusieurs éléments physiques du réseau. La première entité fonctionnelle est en charge de la qualité du canal de propagation ouvert entre un mobile et le réseau. Pour assurer sa fonction, les informations qui circulent entre elle et le mobile sont des mesures physiques de puissance de signal reçu/émis, des mesures de rapport signal sur bruit, de temps de propagation, d'orientation du signal reçu ou encore de charge au niveau de l'antenne. La deuxième entité a trois missions principales. Premièrement, elle a en charge la gestion de la mobilité, pour assurer que tout mobile en tout point du réseau peut recevoir ou émettre une communication. Ensuite, elle s'occupe du routage vers les différents réseaux mobiles et fixes. Enfin, elle gère les autorisations de connexion au réseau ou aux services demandés. Ces autorisations

dépendent de l'abonnement souscrit par l'utilisateur. Les informations concernant les abonnés (type d'abonnement, nom, adresse de facturation, etc.) se trouvent dans une troisième entité fonctionnelle non représentée. De plus, pour effectuer la facturation à l'abonné, les volumes (en minutes, en nombre de SMS ou en octets) consommés par les abonnés pendant la période de facturation sont stockés dans cette entité non représentée. Ces informations sont protégées par la CNIL (Commission Nationale de l'Informatique et des Libertés). A ce titre, elles sont propriété uniquement de l'abonné et ne sont pas légalement exploitables sans l'accord de l'abonné (hors facturation). Néanmoins, comme aucune information de localisation n'y est enregistrée, elles ne présentent aucun intérêt pour la mesure de déplacements.

Il y a deux points de collectes distincts, que nous noterons P et S (Figure 25). Le point P permet de recueillir des mesures physiques à propos de la qualité du canal de propagation aérien établie entre le mobile et le premier élément fixe du réseau. Le point de collecte S situé entre les entités fonctionnelles 1 et 2 de la partie fixe du réseau cellulaire permet d'accéder à des informations sur la nature de l'opération à réaliser pour permettre de satisfaire une demande du mobile. Il s'agit soit d'une demande d'établissement de communication (appel entrant ou sortant), d'une mobilité dans le réseau de télécommunication ou d'une connexion/déconnexion du réseau. Ces messages sont appelés événements de signalisation. De manière simplifiée, ils contiennent l'information de l'opération à réaliser. Les mécanismes de signalisation seront détaillés dans le paragraphe 2.

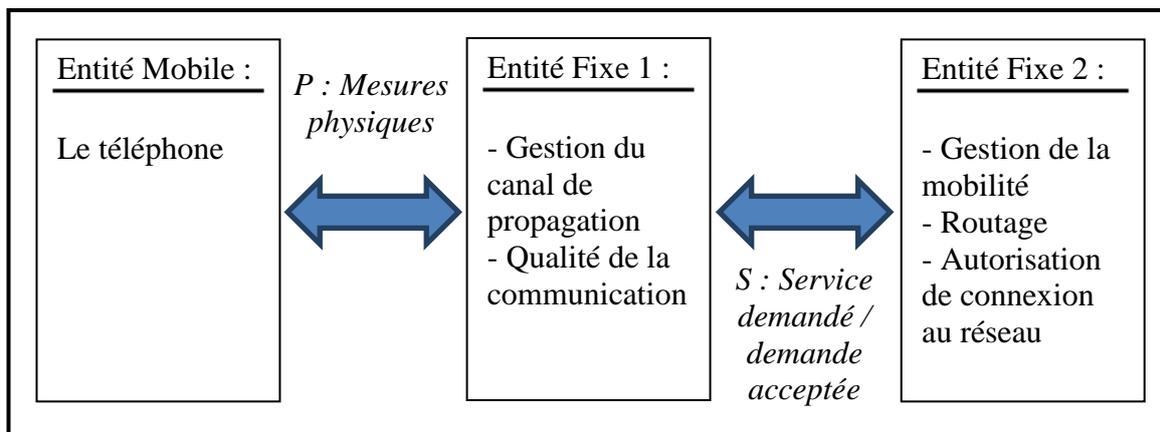


Figure 25 : points de collecte de données disponibles dans un réseau de télécommunications mobiles au travers d'une représentation fonctionnelle simplifiée

1.2.3.2 Traitement des données collectées en P

Les méthodes d'exploitation des données collectées en P sont nombreuses. Nous proposons de classer ces données en deux familles, les mesures de propagation radio et les mesures de charge. Dans les deux cas, ces mesures permettent soit une analyse en différé pour comprendre un dysfonctionnement, pour observer les variations de la demande ou encore pour prédire les évolutions à apporter pour améliorer la qualité de service, soit en temps réel pour le routage d'un mobile vers une antenne voisine moins chargée ou pour offrir un canal de propagation de meilleure qualité. Ces données de nature différente sont générées dans un but commun : garantir une qualité de communication optimale. Pourtant, dans le domaine de la mesure des déplacements, leur traitement mène vers deux applications clairement séparées. D'une part, la localisation précise spatialement et temporellement d'un terminal mobile s'obtient en traitant les mesures de propagations et d'autre part, les mesures de charge permettent l'observation de dynamiques de population à des échelles spatiales et temporelles plus étendues.

Chronologiquement, les mesures de propagation ont été les premières à être exploitées pour déterminer des positions de mobiles avec une précision spatiale fine en deçà de la cellule (périmètre couvert par une antenne). En effet, dès 1998, des techniques de localisation voient le jour (DRANE, et al., April 1998). Elles exploitent soit le temps de propagation, soit l'écart de temps à l'arrivée (TDOA Time Difference of Arrival), soit l'angle d'arrivée, soit la phase de porteuse ou encore le Timing Advance (TA). Ces mesures sont communiquées entre le mobile et le premier élément du réseau fixe. Une autre méthode très répandue pour extraire des informations de localisation à partir d'un réseau cellulaire est le « GSM fingerprinting ». De manière générale, le « fingerprinting » est une méthode basée sur l'exploitation des mesures de puissance de signal radio reçu, en un point donné, par un émetteur dont l'emplacement est connu (OTSASON, et al.). Selon les systèmes et l'environnement de propagation (indoor ou outdoor), ces méthodes permettent une précision de localisation de l'ordre de la dizaine de mètres. Ainsi, il est possible de mesurer avec précision le déplacement d'un mobile entre deux instants et obtenir alors des trajectoires individuelles. Cependant, les mesures de puissance de signal ou de temps de propagation ne sont disponibles qu'au niveau de la première interface aérienne entre un mobile et le premier élément du réseau cellulaire, la cellule. A titre d'illustration, en France, un réseau cellulaire compte des dizaines de milliers de cellules. Or, la collecte d'informations au niveau de cette interface nécessite un déploiement de sonde de collecte au niveau de chaque antenne. Ainsi, la méthode d'estimation de la position requiert le traitement d'un volume de données extrêmement conséquent, voire limitant, pour une exploitation au-delà de la dizaine de cellules.

Pour la deuxième famille de données, les méthodes de mesure de charge reposent sur le fait que le « trafic télécom » varie selon le lieu et l'instant de la journée. L'unité traditionnelle de mesure de trafic est l'Erlang qui représente le volume de trafic voix ou data transitant sur un canal de transmission. L'Erlang (symbole E) est une unité sans dimension à considérer comme une mesure statistique de la charge d'un élément. Il s'agit d'une

représentation de l'usage de la ressource et dépend donc du nombre de communications ainsi que de leurs durées. La mesure d'un Erlang au niveau d'une antenne est l'équivalent d'un usager en communication pendant une heure dont le mobile est resté connecté à cette antenne. Par exemple, si pendant une heure, 100 appels de 5 minutes chacun ont transités par l'antenne a , alors il sera mesuré $(100 \times 5) / 60 = 8.33E$ au niveau de a . En exploitant la distribution d'une mesure de charge normalisée d'un ensemble de cellules couvrant une zone d'intérêt, il est possible de faire ressortir les dynamiques de ce lieu. En effet, en 2005, Ratti introduit le concept de « Mobiles Landscapes » permettant la visualisation des dynamiques urbaines à partir de mesure de charge (RATTI, et al., 2006). Il montre alors l'intérêt d'utiliser les mesures de charge des cellules, déjà exploitées et donc accessibles, par les opérateurs de téléphonie, pour l'observation des déplacements dans la ville de Milan. Des applications possibles de ces traitements sont l'estimation des flux entrants et sortants d'une agglomération, l'observation des migrations pendulaires et des schémas de mobilité en fonction du moment de la semaine (jours travaillés et week-ends), et l'analyse des points critiques de l'utilisation d'une aire urbaine ainsi que des comportements de mobilité et la diffusion de la population lors d'évènements particuliers. La technologie de collecte est telle qu'une exploitation en temps réel est permise. Cependant, l'hypothèse sous-jacente à cette étude est que l'activité au niveau de chaque antenne est proportionnelle au nombre de personnes présentes dans la zone couverte. Ceci n'est vérifié que lorsque tous les abonnés de la zone communiquent à intervalles réguliers et de manière uniforme. Or, les schémas de communication dépendent des caractéristiques individuelles (âge, catégorie socio-professionnel, etc.) et de l'activité pratiquée (travail, récréation, etc.) (YUAN, et al., 2012). Bien que la capacité d'une antenne soit connue par dimensionnement, l'hétérogénéité des pratiques de téléphonie sur une zone ainsi que l'évolution de ces dernières font perdre toute signification à l'utilisation moyenne de la ressource radio pour une zone donnée. Ainsi, pour donner plus de signification aux mesures de charge, des historiques individuels d'évènements de communication (appels, SMS) sont nécessaires vers une mesure plus précise des déplacements (RATTI, et al., 2006).

1.2.3.3 Traitement des données collectées en S

Les données collectées en S (voir Figure 25) contiennent principalement des informations relatives à l'action à réaliser par le réseau pour assurer son fonctionnement. Il existe deux familles de données collectées en S traitées dans la littérature : les actions demandées par l'abonné, à savoir les communications, et les actions nécessaires au réseau pour assurer l'itinérance.

Dans la littérature les communications sont majoritairement exploitées au travers de bases de données appelées CDR pour Calling Detail Record. Ces CDR contiennent en général et à minima les champs suivants : un identifiant de l'abonné, le moment de début de communication, la durée, le type (voix, SMS/MMS, data) et un identifiant de l'antenne à laquelle est connecté le terminal mobile au moment de la communication. Cette source peut être fournie par des opérateurs de télécommunication. Depuis plus de 10 ans, des chercheurs

de domaines variés démontrent l'envergure des champs d'applications d'une telle source. En effet, en agrégeant dans l'espace l'ensemble des traces individuelles, horodatées et localisées au niveau de la zone de couverture d'une antenne, il est possible de visualiser les rythmes d'une ville ou d'un territoire (RATTI, et al., 2006). Ainsi, les CDRs sont adaptés pour des applications d'aide à l'aménagement urbain car ils permettent notamment la détection de zones denses (RUBIO, et al., 2013). De plus, à partir de cet ensemble de trajectoires individuelles partielles, il est possible de mesurer d'une part des déplacements individuels, mais surtout des flux de déplacements représentés dans des matrices Origine-Destination (WHITE, et al., 2002). L'étude des flux peut être approfondie jusqu'à l'inférence des modes de transport (WANG, et al., September 2010) ou la segmentation pour des types de voyageurs particuliers tels que les touristes (AHAS, et al., 2008). Les études sur le tourisme à partir d'un tel jeu de données sont un complément extrêmement intéressant aux enquêtes actives sur le tourisme car ces dernières sont difficiles à mettre en place étant donné la non-permanence des enquêtés sur une zone et le biais induit par les personnes acceptant de répondre à une enquête en vacances. De plus, pour des enquêtes sur le tourisme international, des coûts additionnels sont générés par la nécessité d'enquêteurs multilingues. À cheval sur l'urbanisme et l'étude des mobilités, ces données permettent de mesurer l'impact de la morphologie urbaine sur la mobilité (KANG, et al., 2012). Ainsi, elles ont permis de construire différents modèles de mobilité (GONZALEZ, et al.), (BAGROW, et al., 2009). Enfin, l'usage du réseau de téléphonie et l'activité pratiquée étant corrélés, et chaque élément des trajectoires individuelles localisées, le traitement des CDRs s'applique également pour faire ressortir des schémas d'activités quotidiens (CALABRESE, et al., 2013). Ratti et son équipe font évoluer les « mobiles landscapes » vers des « Activity-aware map » démontrant ainsi l'étendue des champs d'applications accessibles par l'exploitation des CDRs. Néanmoins, une telle source de données n'est pas sans désavantage. Premièrement, par construction, les données d'un CDR ne contiennent aucune information sur l'individu excepté un identifiant anonymisé. En comparaison aux données disponibles via des enquêtes actives, les CDRs manquent de caractéristiques sociodémographiques permettant notamment de comprendre les comportements de mobilités et les schémas d'activités en fonction de caractéristiques individuelles. Cette source est ainsi qualifiée d'incomplète au sens des informations contenues (FARRAHI, et al., 2010). Deuxièmement, une trace n'est déclenchée que lors d'un événement de communication. Pour la majorité de la littérature décrite dans ce paragraphe, les communications sont de nature vocale et proviennent de réseaux de télécommunications de technologie 2G. Ces événements peuvent être qualifiés de rares par comparaison avec des traces GPS par exemple (FARRAHI, et al., 2010). Ainsi, cette source est incomplète dans le sens où les trajectoires présentent un fort degré d'incertitude quant à la détection complète des tous les événements de mobilité. En effet, la connaissance du comportement du téléphone (mobile/immobile) en dehors des événements de communication n'est accessible qu'au travers d'hypothèses. Troisièmement, ajoutée à l'incertitude en dehors des points constituant l'ensemble des trajectoires discrètes des individus, il existe une imprécision non négligeable sur la localisation. En effet, un événement d'un CDR est horodaté et localisé à l'échelle d'une cellule du réseau de télécommunication. La dimension

d'une cellule varie d'une centaine de mètres de rayon à une dizaine de kilomètres selon la densité de population du territoire qu'elle couvre. De plus, en fonction de la charge du réseau et des conditions locales de propagation, le réseau de télécommunication peut faire le choix de router un mobile situé dans une cellule vers une cellule voisine. Dans la littérature, des exemples de « CDRs améliorés » sont proposés afin de remédier à certaines lacunes précédemment énoncées. Des entreprises comme AirSage ont développé un algorithme déposé WiSE (Wireless Signal Extraction) qui traite les remontées du réseau cellulaire Verizon et estime la position des mobiles (latitude, longitude) avec une précision de 300 mètres. Ils proposent alors des CDRs avec une précision spatiale améliorée (WANG, et al., September 2010). Cependant, le problème de la détection erronée des déplacements courts (entre deux cellules voisines) demeure car elle est la conséquence des fluctuations de la couverture radio (CALABRESE, et al., 2013). De la même manière, afin de quantifier, voire de corriger les carences des trajectoires en données sociodémographiques, des études avec des CDRs enrichis de l'âge et du genre du propriétaire du mobile ont été mises en place en Chine sur un large échantillon de la population (YUAN, et al., 2012). De telles études ont permis de montrer les corrélations entre mobilité et usage de la téléphonie selon des critères individuels. Cependant, la législation quant à la protection des données individuelles diffère selon les pays et une telle étude n'aurait pu être menée en Europe, et particulièrement en France, sans l'accord de tous les abonnés inclus dans l'étude. Parallèlement, des modèles de la distribution de temps entre deux appels pour un abonné ont été proposés afin de prédire les schémas de communication des usagers. A ce titre, sont proposés dans la littérature des lois exponentielles (YANG, et al., 2009), mais aussi des lois de puissance (GONZALEZ, et al.). Cette inimité vient du nombre important de paramètres qui influencent les schémas de communication. En effet, les caractéristiques individuelles telles que l'âge, le genre, etc., mais aussi l'activité pratiquée, déterminent la fréquence, la durée, le moment des communications tout autant que des composantes extérieures telles que le moment de la journée, le jour de la semaine, etc. Ainsi, pour plus de précision, les modèles de distribution de l'instant entre deux appels d'un usager devraient considérer chacun de ces facteurs. Cependant, en essence, les CDRs ne contiennent aucune information personnelle sur l'abonné. Il apparaît alors difficile de corriger les imperfections des trajectoires individuelles, même en exploitant les modèles du temps entre deux communications selon des critères affinés, car il n'est pas possible d'accéder à des critères sur l'individu dont la trajectoire est incomplète. Afin de constituer une base contenant des informations sur les individus, le MIT a mis en place une enquête mixte active et passive (EAGLE, et al., march 2006). L'idée sous-jacente à cette enquête mixte était de construire une base de données en collectant diverses informations depuis des téléphones mobiles habituellement transportés par leur propriétaire mais également depuis une enquête en ligne afin de valider les informations remontées via les réseaux de communication. L'étude se veut complète, car elle contient à la fois des informations quasi-continues sur la localisation des individus enquêtés et l'usage du terminal mobile en terme d'applications et de communication. Un tel jeu de données est approprié pour la mesure de systèmes sociaux complexes, à travers la reconnaissance de schémas d'interactions sociales dans les activités quotidiennes, dans l'inférence des relations sociales,

dans l'identification de lieux ayant une forte importance sociale et pour les modèles des rythmes d'organisation sociale. Un échantillon fût composé de 100 personnes du MIT Media Laboratory et du MIT Sloan Business School ayant accepté de participer à cette enquête sur 9 mois. Ils se sont vu remettre un téléphone Nokia 6600 équipé du logiciel Context. Ce logiciel permet la collecte des logs de communication, des appareils Bluetooth à proximité, des identifiants des antennes cellulaires et du statut du téléphone (RAENTO, et al., 2005). Le jeu de données passives a été constitué à partir de remontées depuis le réseau de téléphonie et du réseau Bluetooth pour permettre une localisation plus précise à proximité d'un environnement contenant des appareils Bluetooth, notamment dans les bâtiments. Lorsque le mobile ne se trouve pas dans un périmètre couvert par un réseau Bluetooth, la précision spatiale est alors à l'échelle d'une cellule. De plus, le fonctionnement classique d'un réseau cellulaire ne permet d'avoir un identifiant de cellule généré que lors d'une communication. Pour assurer la continuité des données, le logiciel embarqué sur les 100 mobiles enquêtés contient un protocole spécifique qui déclenche l'enregistrement de l'identifiant de cellule à laquelle il est connecté dès que cet identifiant est rencontré pour la première fois. Dès lors, le MIT dataset contient les trajectoires quasi-continues de chaque individu enquêté. Le jeu de données anonymisé correspond à environ 450 000 heures d'informations sur les positions des 100 individus et sur la manière dont ils utilisent leur téléphone pour communiquer mais aussi pour les applications disponibles sur le terminal. Ainsi, ce jeu permet de s'affranchir des biais relatifs au caractère épars et discret des CDRs et d'accéder à la compréhension des usages individuels en fonction des schémas de déplacement. De plus, de nombreux travaux se sont basés sur ces données disponibles en ligne. Notamment, il a été montré la faisabilité de réduire des bruits liés aux oscillations entre cellules et aux délais temporels de transfert intercellulaire (BAYIR, et al., 2009). Des méthodes de clustering de cellules permettent d'augmenter la robustesse du jeu de données. Cependant, les biais de représentativité et de taille de l'échantillon dû au caractère actif de l'étude réapparaissent. Cet exemple illustre bien le compromis entre taille de l'échantillon et données accessibles. Nous avons cité des exemples de travaux réalisés à partir de CDR pour illustrer l'étendue des mesures réalisables à partir de ces comptes rendus de communications. Néanmoins, le paragraphe 2.2 décrivant le cadre légal précise que l'exploitation de telles bases n'est pas autorisée en France sans l'accord des abonnés ayant laissé une trace de communications dans le CDR. Ainsi, dans cette thèse nous exploiterons les communications depuis une source temps réel qui permet de supprimer la phase de stockage et de respecter le cadre légal (paragraphe 2.2).

Cependant, les actions permettant d'assurer la fonction d'itinérance des réseaux de télécommunications sont également assimilables à des capteurs de déplacements. Il s'agit de messages de mobilité entre les unités spatiales du réseau cellulaire, à savoir les cellules, ou un regroupement de ces dernières. Les applications pour la mesure des déplacements sont différentes de celles mentionnées précédemment. En effet, dès 1994, les informations de changement de cellules en communication, appelées handover, et les changements de zones de localisation (regroupement logique de cellules) sont exploitées pour la mesure de flux de trafic routier (LINAARTZ, October 1994). En France, une application de mesures de trafic a

été opérée dans le corridor du Rhône (YGNACE, et al., 2000) (YGNACE, 2001). Le sujet de la mesure du trafic à l'aide de messages de mobilité générés par le réseau de téléphonie est récurrent depuis cette première preuve de faisabilité (BAR-GERA, 2007), (CACERES, et al., 2010). Ces mesures de flux routier permettent l'estimation de vitesse grâce à des capteurs virtuels placés le long d'un axe de transport avec le même niveau de précision que des boucles électromagnétiques par exemple. Des offres basées sur ces technologies sont disponibles depuis 2012. En effet, l'opérateur de services d'information trafic routier Médiamobile, fût le premier à proposer un service qui intègre des données de téléphonie mobile fournies par l'opérateur Orange comme source d'information trafic. Depuis les années 2000, les exploitants routiers peuvent accéder à des sources d'informations trafic qui s'affranchissent de matériel supplémentaire à déployer sur les routes. Ainsi, par opposition aux boucles électromagnétiques, aux caméras de surveillance de trafic et aux systèmes de reconnaissance de plaques, on cherche à exploiter les « capteurs mobiles ». Ce terme désigne un système permettant de recueillir des informations directement issues des véhicules traceurs via la méthode dite « Floating Car Data » (FCD) ou à partir d'appareils nomades, comme des téléphones mobiles, présents à bord et circulant dans le flot de la circulation via la méthode « Floating Mobile Data » (FMD) (RECLUS, 6 décembre 2011). Les méthodes basées sur des capteurs mobiles présentent les avantages suivants. Les informations sur les conditions de circulation sont disponibles en temps réel, les données de vitesse recueillies permettent de produire des indicateurs de vitesse moyenne, de temps de parcours, d'état de trafic, etc. Ainsi, la détection et la connaissance des congestions sont plus précises (début, longueur et fin de bouchons). Il est d'ailleurs envisagé de construire de l'information dédiée à des catégories d'usagers du réseau routier (poids lourds bus, taxis, etc.) De plus, cette source d'informations permet de compléter des données recueillies sur le terrain. Ainsi, il est envisageable d'exploiter ces données en temps différé afin de produire par exemple des matrices Origine-Destination. Néanmoins, des informations sur les volumes et la concentration (débit, taux d'occupation) ne sont pas encore constructibles à partir de ces méthodes. En effet, la représentativité statistique varie au cours de la journée et de la semaine. Cette variation est bien plus significative avec la méthode FCD qu'avec la méthode FMD. De plus, la composition exacte de flot de véhicule n'est pas connue. Et surtout, il n'existe pas encore de normalisation des aspects liés à la qualité et à la fiabilité des informations produites. Comparons maintenant ces deux méthodes. Tout d'abord, malgré sa démocratisation et la précision spatiale qu'elle permet d'atteindre, la méthode FCD basée sur la technologie GPS présente des inconvénients. En effet, le nombre de véhicules équipés de GPS communiquant ses informations de position avec les fournisseurs de service d'information trafic est bien souvent réduit à des flottes de véhicules comme les véhicules de messagerie ou les taxis. De plus, le système radio permettant de transmettre ces informations est plus onéreux, car dédié à cette application, alors que le système basé sur le réseau de téléphonie est naturellement en fonctionnement. La méthode FMD exploite deux procédés existants pour estimer la localisation à partir du réseau cellulaire: la triangulation avec des données collectées en P (voir Figure 25) et la méthode dite « Floating Mobile Data » qui exploite des événements de changement de cellules en communication (handover) que l'on

peut collecter en S. Ainsi, dans le cas de la deuxième approche, la taille de l'échantillon est réduite aux mobiles en communication suffisamment longtemps pour générer un minimum de 3 handovers. La Figure 26 illustre le procédé.

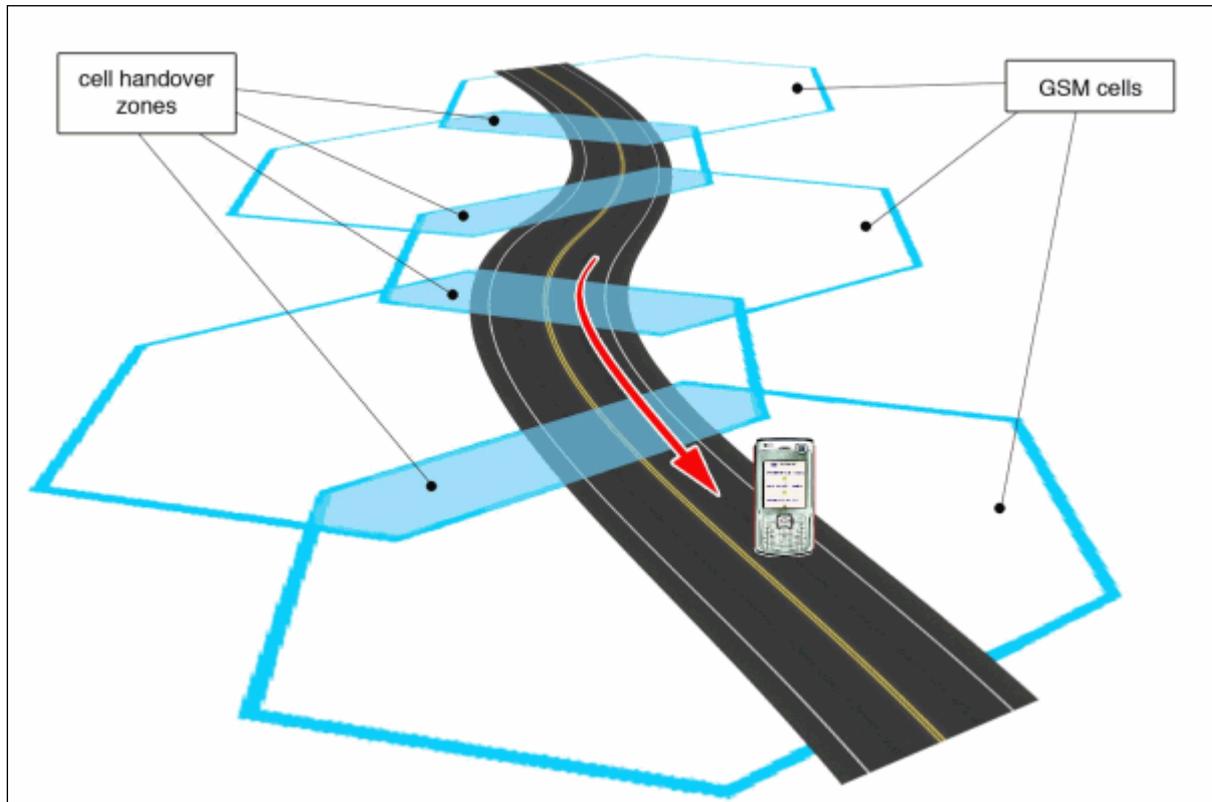


Figure 26 : principe du FMD¹¹

1.3 Conclusion

Les enquêtes passives traditionnelles sont efficaces, standardisées et permettent d'observer les évolutions des mobilités en fonction des caractéristiques individuelles. Néanmoins, leur caractère statique est un frein pour l'observation fine dans le temps, ainsi elles présentent des limites pour la mesure de comportements localisés étroitement dans le temps et dans l'espace et/ou de comportements sporadiques. Les technologies numériques semblent prometteuses pour de telles mesures. Néanmoins, leur caractère passif et le cadre légal qui protège les données personnelles, empêchent l'accès à des informations sociodémographiques sur les individus. Une autre limitation des enquêtes traditionnelles est la précision spatiale. La portée des déplacements est souvent approchée à la distance à vol d'oiseau et l'accès à l'itinéraire exact n'est pas toujours évident. De plus, en raison des coûts

¹¹ : <http://www.transport-intelligent.net/technologies/localisation-78/article/les-donnees-des-telephones-mobiles>

de mise en place de telles enquêtes, le périmètre d'étude doit être restreint, ce qui empêche à la fois l'étude de phénomènes à caractère occasionnel, mais aussi l'étude sur un vaste territoire. Les coûts réduits proposés par les enquêtes passives et le déploiement à l'échelle d'un grand territoire d'un réseau de capteurs déjà en place promettent de lever cette limitation. Cependant, il apparaît que dans le cas d'un réseau de capteurs fixes, le périmètre d'étude est borné par l'emplacement des capteurs qui sont fixes. Cela n'est pas le cas pour les réseaux de capteurs mobiles qui assurent un périmètre d'étude à l'échelle d'un territoire. L'intérêt des concepteurs d'enquêtes de recourir à la technologie GPS depuis une dizaine d'année est-il annonciateur de la disparition des enquêtes traditionnelles ou d'un enrichissement de ces dernières en couplant les deux méthodologies ?

Nous avons vu qu'à partir d'un objet de consommation du quotidien, le téléphone portable, des chercheurs sont parvenus à observer la mobilité sur des territoires, au travers de rythmes de présence ou même des déplacements sur un axe de transport. Il existe différentes méthodes basées sur la téléphonie mobile selon le point de collecte sur un réseau de télécommunication. Notamment, l'approche désagrégée basée sur l'étude des événements de signalisation présente l'avantage d'être le reflet d'un comportement d'usage du réseau de téléphonie lorsqu'un abonné mobile pratique un territoire couvert par ce réseau. On se situe alors aux abords de la perception de l'utilisateur. De plus, les équipements de remontées sont naturellement en place pour assurer le fonctionnement de ces réseaux. Enfin, les solutions basées sur les réseaux de télécommunications sont disponibles partout et pour un très large nombre d'individus passivement enquêtés, lorsque les technologies GPS ne sont pas encore opérationnelles en indoor (bâtiments, tunnels, etc.) et ne concernent que des segments particuliers d'enquêtés.

Parmi, les solutions de captures de mobilité sur un territoire basées sur le réseau radio-mobile, nous avons choisi dans cette thèse de travailler avec celles qui exploitent la signalisation des mobiles. Nous justifions le fait de favoriser les données collectées en S par rapport aux données disponibles en P par deux points. Tout d'abord, le coût de traitement informatique des volumes que l'on peut collecter en P restreint considérablement la dimension des périmètres d'étude. Ensuite, bien qu'offrant a priori une précision spatiale moindre, les données collectées en S vont dans le sens d'une représentation de la perception de l'utilisateur. Quant au choix du type d'événements de signalisation à utiliser pour mesurer des déplacements, nous combinons les deux familles que nous avons identifiées dans la littérature en considérant aussi bien les événements de communications que les événements de mobilité. Ainsi, on peut assimiler le système de capture que l'on considère comme un réseau de capteurs fixes (déclenchement d'évènement de mobilité LAU et HO, etc.) couplé à des capteurs mobiles (communications, etc.). Cela permet d'avoir un ensemble de points de mesure horodatés précisément (sur déclenchements d'évènements), localisés spatialement à l'échelle de la zone couverte par une antenne, mais contenant également des informations de contexte sur la manière dont les réseaux de transport et de télécommunication sont pratiqués.

Dans cette thèse, nous traiterons de la mesure de déplacements à partir de ces événements de signalisation. Alors, nous consacrerons la section suivante à l'introduction du fonctionnement d'un réseau de télécommunications axé autour des procédures de