

Quel apport des données issues des nouvelles technologies à la modélisation des transports ?

Contributeurs (par ordre alphabétique) :

Aurélie Bousquet, Cerema Territoires et Ville,

Alice Charpe, Cerema Ouest,

Julien Harache, Cerema Ouest,

Maxime Le Corre, Cerema Est,

Maria Tébar, Cerema Nord-Picardie.

Introduction - Typologie des données quantitatives d'observation des comportements de mobilité et des flux

1. Enquêtes à échantillons représentatifs
2. Enquêtes à échantillons non-représentatifs
3. Collectes passives exhaustives
4. Données massives

Conclusion - Utilisation des données massives pour modéliser les mobilités



Introduction

Typologie des données quantitatives
d'observation des comportements de
mobilité et des flux

Typologie des données quantitatives

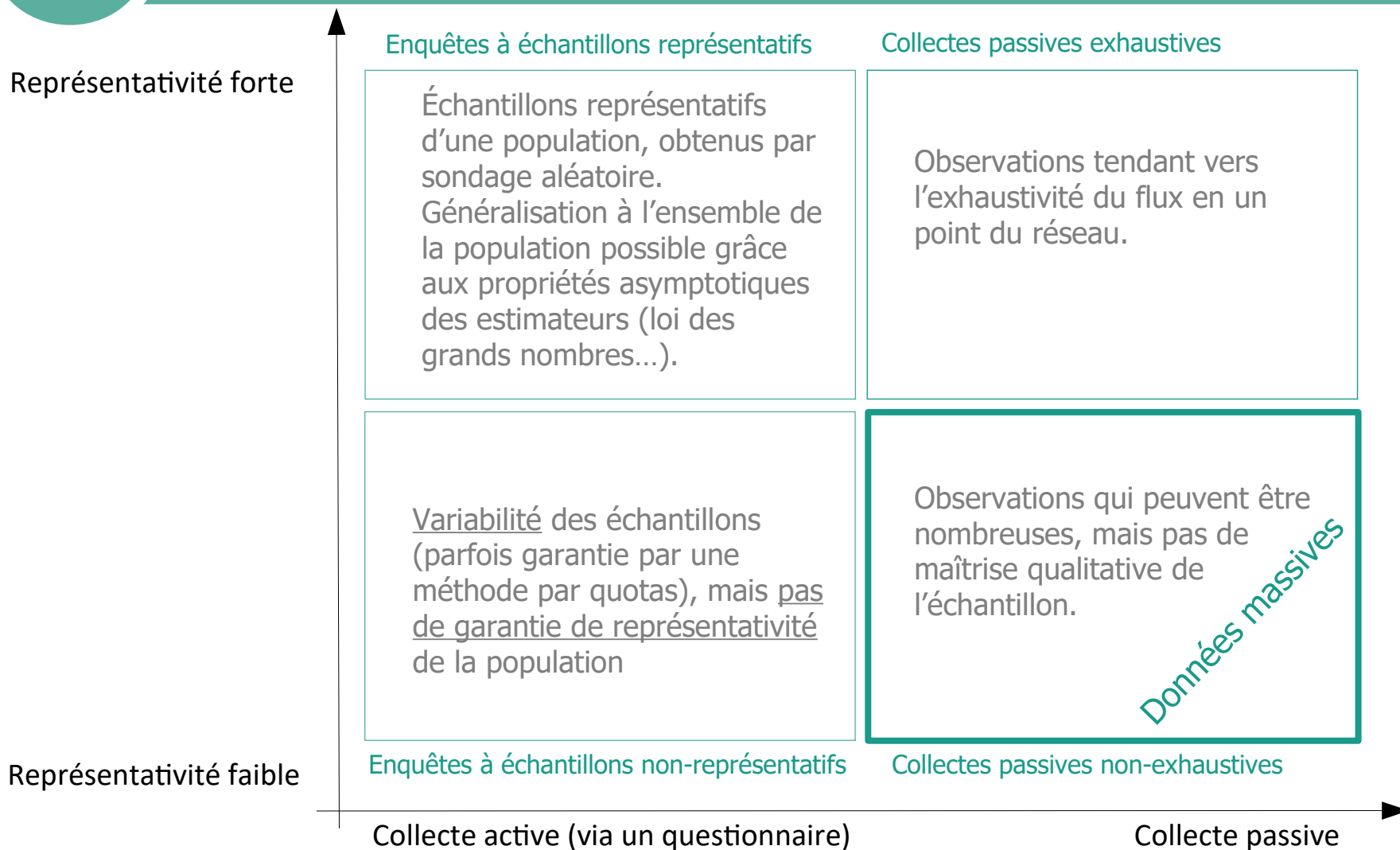
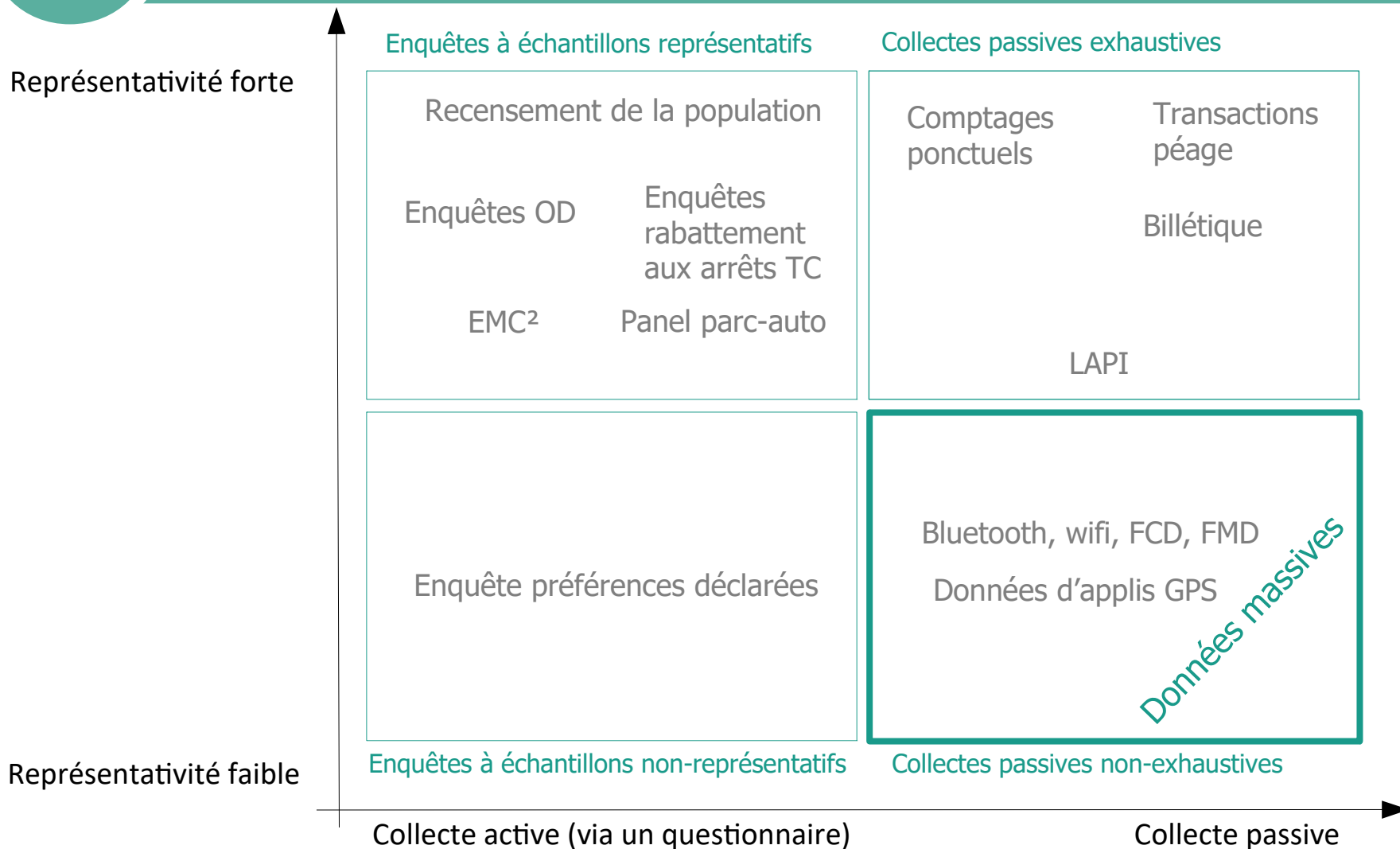


Illustration sur les collectes de données quantitatives de mobilité les plus courantes





1. Enquêtes à échantillons représentatifs

- **Fiche d'identité**
 - Co-producteurs : collectivités et Cerema ;
 - Enquête des résidents d'un territoire en face-à-face et / ou par téléphone ;
 - Photographie à l'instant « t » de leurs déplacements sur l'ensemble de la journée (jours de semaine + option week-end).
- **Quelle est la précision spatiale de localisation des origines et des destinations ?**
 - A minima l'IRIS, voire géolocalisation en (x,y) pour les grands générateurs de déplacements.
- **Peut-on connaître les motifs et les modes de déplacements ?**
 - Oui, avec beaucoup de détails.
- **Quels sont les biais de collecte spécifiques à cette source ?**
 - Biais classiques liés aux enquêtes statistiques (biais enquêteurs, non-réponse, etc.), connus et maîtrisés ;
 - Pas d'intégration des résidences collectives dans l'échantillon ;
- **Comment généraliser les données collectées à l'ensemble de la population ?**
 - 1 à 3 % de la population enquêtée ;
 - Redressement de la non-réponse à partir de la base de logements de la Direction Générale des Impôts ;
 - Redressement sur le recensement de la population.



EMC²
Enquête Mobilité
Certifiée Cerema

- **Peut-on mesurer des évolutions ?**
 - Oui (méthodologie stable), mais sur des périodes de 10 ans sauf option « Fréquence + ».
- **De quelles informations dispose-t-on sur les individus et leur foyer ?**
 - Individus :
 - âge, genre,
 - PCS, statut d'occupation, niveau de scolarité,
 - permis de conduire, abonnement TC,
 - adresse précise du lieu de travail ou d'études.
 - Ménages :
 - taille du ménage,
 - type de logement,
 - nombre de véhicules,
 - lieu et type de stationnement à domicile et au travail.

- **Fiche d'identité**

- Producteurs : collectivités ou services de l'État ;
- Enquêtes réalisées en bord de route par interview des conducteurs ;
- Questionnaires courts pour ne pas trop gêner l'écoulement du trafic : 30 sec à 1 min.

- **Quelle est la précision spatiale de localisation des OD ?**

- Généralement à la commune pour la France, sur un zonage agrégé à l'étranger ;
- Parfois, collecte infra-communale pour les zones proches du poste d'enquête.



- **Peut-on connaître les motifs de déplacements ?**

- Oui, toujours collectés pour les VL, selon les postes pour les PL.

- **Quels sont les biais de collecte spécifiques à cette source ?**

- Collecte entre 7 heures et 19 heures classiquement : on manque les OD de nuit ;
- Non-réponse, notamment conducteurs ne parlant pas français : + difficile pour les enquêteurs.

- **Comment généraliser les données collectées à l'ensemble de la population ?**

- 10 à 25 % du trafic enquêté ;
- Redressement sur des comptages manuels sur la période d'enquête (types détaillés de véhicules) et automatiques sur 2 semaines (VL et PL).

- **Peut-on mesurer des évolutions ?**
 - Oui, on actualise habituellement un poste tous les 10 ans.
- **De quelles informations dispose-t-on sur les véhicules / les individus et leur foyer ?**
 - VL (selon les enquêtes), informations classiquement collectées :
 - taux d'occupation,
 - âge, sexe, PCS, communes de résidence et de travail du conducteur
 - PL (selon les enquêtes), informations classiquement collectées :
 - marchandises, tonnage, volume chargé

- **Fiche d'identité**

- Producteurs : collectivités ou exploitants des réseaux TC ;
- Enquêtes réalisées à bord des transports collectifs (toutes les courses d'une ligne ou un échantillon) ;
- Questionnaires courts pour capter les OD courtes : 30 sec maxi ;
- Connaissance des arrêts de montée / descente / correspondance, des lignes empruntées / des modes de rabattement à l'origine et à la destination.

- **Quelle est la précision spatiale de localisation des OD ?**

- La commune, parfois collecte infracommunale.

- **Peut-on connaître les motifs de déplacements ?**

- Oui.

- **Quels sont les biais de collecte spécifiques à cette source ?**

- Sous-représentation des OD courtes : usagers qu'on n'a pas le temps de capter ;
- Période horaire limitée : pour des raisons de coût.

- **Comment généraliser les données collectées à l'ensemble de la population ?**

- Taux de sondage très variable selon les enquêtes ;
- Redressement sur des comptages manuels sur la période d'enquête et éventuellement automatiques (billétique...), à la course et à la ligne.



Enquête à Montpellier – Crédit photo : France 3

- **Peut-on mesurer des évolutions ?**
 - Oui, selon les réseaux, on actualise les enquêtes tous les 3 à 10 ans ;
- **De quelles informations dispose-t-on sur les individus et leur foyer ?**
 - Selon les enquêtes : âge, sexe, PCS, communes de résidence et de travail, type de titre de transport, fréquence d'utilisation.

Autres dispositifs d'enquête à échantillons représentatifs

- **Enquêtes en gare / aux points d'arrêt**
 - Producteurs : collectivités ou services de l'État ;
 - Pour connaître les pratiques de rabattement sur le point d'arrêt
 - Enquêtes par sondage aléatoire avec contrôle de l'échantillon ;
 - Environ 40 % de taux de sondage observés sur des petites gares, très dépendant du terrain.
- **Recensement de la population de l'INSEE**
 - Producteur : INSEE
 - Actualisé tous les ans mais comparaison de deux vagues nécessite un intervalle de 5 ans (rotation des communes enquêtées) ;
 - Commune de résidence et commune de travail / étude : pas des déplacements mais des potentiels de déplacements ;
 - Enquêtes par sondage aléatoire avec contrôle de l'échantillon ;
 - Taux de sondage :
 - questions principales : 40 % à 100 % selon les communes ;
 - questions complémentaires (dont lieu de travail/d'étude et mode de transport utilisé) : 20 à 40 % selon les communes.



2. Enquêtes à échantillons non-représentatifs

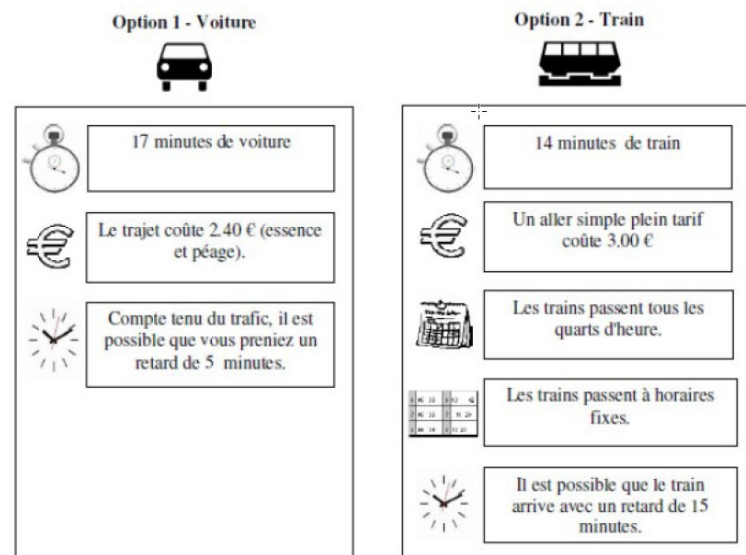
Enquêtes de préférences déclarées

- **Fiche d'identité**

- Producteurs : collectivités ou services de l'État ;
- Pour tester les effets de modifications hypothétiques de l'offre de transport sur les comportements, en particuliers de choix modal ;
- L'unité d'observation est l'individu, soit en relation à un territoire de résidence, soit en relation à un lieu ou à un réseau de transport qu'il fréquente ;
- Pas d'exigence de représentativité, mais besoin de variabilité pour balayer au maximum les différents types de comportements ;
- Questions très variables d'une enquête à l'autre, pas de dispositif standardisé.

- **Comment généraliser les données collectées à l'ensemble de la population ?**

- Application à une population synthétique d'un modèle de comportement des individus calibré à partir de l'enquête réalisée. Généralement, utilisation de modèles de choix discrets*.



* <https://www.cerema.fr/fr/actualites/modeles-choix-modal-edition-guide-methodologique>



3. Collectes passives exhaustives

- **Comptages routiers, ponctuels ou directionnels**
 - Producteurs : gestionnaires des réseaux routiers ;
 - Permanents en section : boucles électromagnétiques,
 - Temporaires : radars, pneumatiques, comptages manuels, vidéo, etc.
- **Comptages montées/descentes dans les transports en commun**
 - Producteurs : collectivités ou exploitants des réseaux TC ;
 - Comptages manuels ou capteurs.

Boucles électromagnétiques



Crédit photo : [Accor solutions](#)

Pneumatiques



Crédit photo : [Nouvelles Pulications](#)

Radar



Crédit photo : [Sferiel](#)

- **Fiche d'identité**

- Producteurs : collectivités, concessionnaires ;
- Données de validation / paiement sur un réseau donné
- Soit en un point de passage, soit en entrées/sorties selon les dispositifs et les réseaux ;
- Quasi-exhaustives (sauf fraude).

- **Peut-on connaître les motifs de déplacements ?**

- Non.

- **Peut-on mesurer des évolutions ?**

- Oui, données fortement dynamiques, possibilité de mesurer des évolutions de jour en jour, entre saisons différentes, d'une année sur l'autre ;
- En théorie : suivre un véhicule ou un usager sur plusieurs jours ;
- → les contraintes d'anonymisation rendent généralement cette information indisponible.

- **De quelles informations dispose-t-on sur les individus et leur foyer ?**

- Titre de transport utilisé. Aucune autre information en général (contraintes d'anonymisation).



Crédit photo : Le blog auto



Crédit photo : Actu Maint

Fiche d'identité

- Producteurs : collectivités, services de l'État ;
- Lecture vidéo de plaques d'immatriculation fournissant :
 - Volumes en un point donné (comptage) ;
 - Entrées/sorties d'un réseau ou d'un périmètre ;
- Quasi-exhaustives (sauf plaque illisible) :
 - Redressement sur des comptages.

Quels sont les biais de collecte spécifiques à cette source ?

- Sous-estimation du transit : quand une plaque est illisible en entrée ou en sortie, le véhicule est considéré en échange.

Peut-on connaître les motifs de déplacements ?

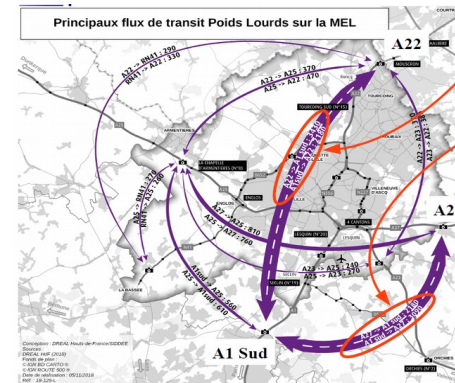
- Non, connaissance impossible.

Peut-on mesurer des évolutions ?

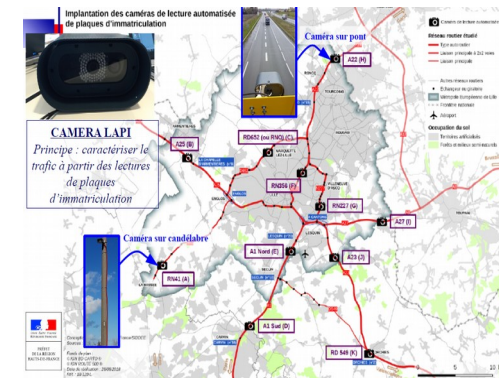
- Oui, si on situe la caméra au même endroit.

De quelles informations dispose-t-on sur les véhicules ?

- Pays d'immatriculation du véhicule ;
- Type de véhicule : grand PL, petit PL, VUL, VP, moto ;
- Avec dispositif spécifique (RGPD et autorisation CNIL) : possibilité d'apparier avec des données issues du fichier SIV (vignette, motorisation, etc.).



Source : Dreal HF



Source : Dreal HF



4. Données massives

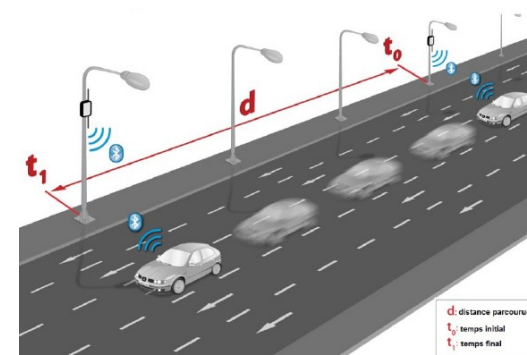
- **Fiche d'identité**

- Producteurs : collectivités ou services de l'État
- Données d'adresses MAC horodatées : distinction équipement (smartphone, oreillette, casque audio, etc.) ;
- Dispositif « bord de route » :
 - antenne directionnelle ou omnidirectionnelle ;
 - rayon de capture paramétrable ;
- Peut fournir :
 - Volumes en un point donné (comptage) : bord de route ou comptage à bord des TC ;
 - Entrées/sorties d'un réseau ou d'un périmètre.



- **Peut-on connaître les motifs et les modes de déplacements ?**

- Motifs : origines-destinations inconnues : pas possible d'inférer des motifs ;
- Modes : pas de détermination possible sur la base des temps de parcours ; localisation (route, centre commercial, etc.) et paramétrage des dispositifs, éventuellement.



- **Quels sont les biais de collecte spécifiques à cette source ?**
 - Paramétrage complexe
 - Taux d'équipement (difficulté de faire le lien appareils → personnes → véhicules) et nécessité que le Bluetooth soit activé
 - Nouvelle norme Bluetooth : détection dépend de la génération des équipements
 - A venir : adressage MAC dynamique
- **Comment généraliser les données collectées à l'ensemble de la population ?**
 - Vitesse : pas de redressement
 - Volumes :
 - Redressement sur comptages automatiques, à réaliser en parallèle
 - Taux de pénétration très variable autour de 10 % tous équipements (doubles comptes)
 - possible de redresser une campagne « cordon » (entrée/sortie)
- **Peut-on mesurer des évolutions ?**
 - Vitesse : oui, si on situe les équipements au même endroit ;
 - Volumes : difficile en raison de l'évolution de l'équipement électronique de la population : variation du taux de pénétration, évolution des normes et technologies.
- **De quelles informations dispose-t-on sur les individus et leur foyer ?**
 - Pas d'enrichissement possible

- **Fiche d'identité**

- Producteurs : opérateurs de téléphonie ;
- Données de localisation liées aux événements recensés par les antennes (appels, messages, data, « Ping ») : conservées 1 an ou commande anticipée.

- **Comment découpe-t-on les déplacements ?**

- Temps de « stationnarité » = arrêt. A minima 15 min → pas ou peu de déplacements pour motifs brefs ;
- Algorithme de traitement complexe pour lier localisation et déplacements.

- **Quelle est la précision spatiale de localisation des origines et des destinations ?**

- Liée à la localisation des antennes + problématique de « frontière » entre « zones-antennes » ;
- Flux à grandes échelles, à minima d'une ville.

- **Peut-on connaître les motifs et modes de déplacements ?**

- Motifs : éventuellement travail/études sans distinction et domicile, sur la base d'un algorithme d'apprentissage ;
- Modes : distinction mode ferroviaire/routier envisageable sur de grandes distances, très spécifiques aux cas d'études.



- **Quels sont les biais de collecte spécifiques à cette source ?**
 - Traitement algorithmique propre à chaque fournisseur et rarement transmis ;
 - Redressement de la population sur la base de la part de marché de l'opérateur
 - Échantillonnage temporel irrégulier et hétérogène ;
 - Données provenant d'un seul opérateur : pas de commercialisation d'agrégats entre opérateurs à ce stade.
- **Comment généraliser les données collectées à l'ensemble de la population ?**
 - Par rapport au taux de pénétration, mais pas nécessairement de redressement sur l'âge notamment (quid des personnes âgées ?).
- **Peut-on mesurer des évolutions ?**
 - Données conservées un an : analyses saisonnières, semaine/week-end ;
 - Évolutions rapides des algorithmes et du fonctionnement des réseaux :
 - incertain, au-delà d'un an ;
 - Nécessité de choisir le même opérateur.
- **De quelles informations dispose-t-on sur les individus et leur foyer ?**
 - Offre hétérogène selon les opérateurs :
 - Données socio-démographiques : incertain ;
 - Lieux d'activité et de résidence : en fonction de l'opérateur, \pm fiable.

- **Fiche d'identité**

- Producteurs : opérateurs privés spécifiques ;
- Données de localisation horodatées fournies par un système GNSS (via un assistance d'aide à la conduite, intégré ou non au véhicule, ou une application smartphone) fournissant :
 - des vitesses et des congestions ;
 - des itinéraires et donc des flux OD.

- **Quelle est la précision spatiale de localisation des OD ?**

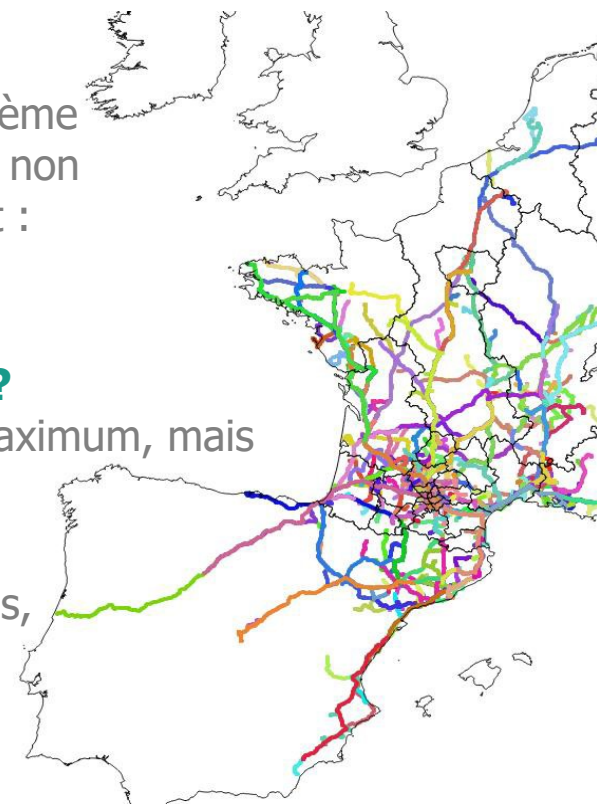
- Précision liée à la fiabilité des données GPS (25 m au maximum, mais généralement inférieure à 10 m).

- **Comment découpe-t-on les déplacements ?**

- Temps de « stationnarité » = arrêt : nombreuses limites, car une grande partie des traces se termine sur une vitesse non nulle ;
- Possibilité d'utiliser des algorithmes plus poussés (par exemple, pour filtrer les pauses pour les PL).

- **Peut-on connaître les motifs de déplacements ?**

- Non, sauf si historique plus long permettant par déduction de repérer le domicile et éventuellement le travail (R&D + contraintes d'anonymisation).

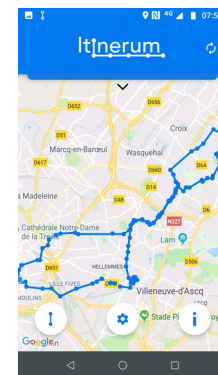


- **Quels sont les biais de collecte spécifiques à cette source ?**
 - Pour les véhicules légers :
 - Vitesse légèrement au-dessus de la moyenne (problématique en situation fluide) ;
 - Motifs de déplacement et classes de distance non-représentatifs (moins de déplacements courts pour motif personnel) ;
 - Pour les poids-lourds :
 - A minima, les classes de distance ne sont pas représentatives.
- **Comment généraliser les données collectées à l'ensemble de la population ?**
 - Vitesse : sans redressement, si l'échantillon est assez important ;
 - Flux OD : un redressement est indispensable, méthodologie à définir au cas par cas.
 - Si flux trop peu représentatifs, le redressement ne suffira pas à corriger les OD.
- **Peut-on mesurer des évolutions ?**
 - Vitesse : Oui, si la taille de l'échantillon le permet ;
 - Flux OD : Difficilement, le taux d'équipement évoluant vite...
- **De quelles informations dispose-t-on sur les véhicules ?**
 - FCD « classique » :
 - direction (cap) et vitesse instantanée du véhicule ;
 - l'identifiant unique propre au véhicule ;
 - xFCD : FCD « classique » + données disponibles via le bus CAN du véhicule :
 - ID-xFCD : xFCD + informations sur le véhicule et ce qu'il mesure.

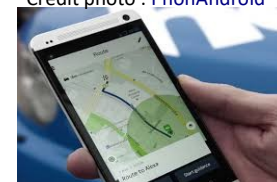


• Fiche d'identité

- Selon l'appli : localisations obtenues à l'aide du GPS du téléphone, des antennes Wifi ou des antennes du réseau de téléphonie mobile ;
- Applis d'enquêtes :
 - Producteurs : collectivités ou Cerema ;
 - Recrutement des individus dans le cadre d'une enquête ;
 - Interactions avec l'utilisateur possibles ;
 - Anonymisation à faire par le développeur de l'appli.
- Applis de guidage/achat, dont on réutilise les données a posteriori :
 - Producteurs : opérateurs privés spécifiques ;
 - Collecte des traces sans action spécifique de l'utilisateur, pas de recrutement ;
 - Anonymisation par le développeur de l'appli ou l'opérateur téléphonique.



Crédit photo : PhonAndroid



• Comment découpe-t-on les déplacements ?

- Trace croisée avec un fichier « arrêts » (applis d'enquête) ;
- Fichier « déplacements » directement saisi (applis d'enquête) ;
- Trace seule : algorithme à développer a posteriori (applis d'enquête + guidage).

• Quelle est la précision spatiale de localisation des origines et des destinations ?

- Si GPS : Itinéraire et OD précis, si antennes/wifi : précision comparable aux FMD ;

• Peut-on connaître les motifs et modes de déplacements ?

- Selon l'appli : possibilité d'avoir une interprétation du mode / du motif.

- **Quels sont les biais de collecte spécifiques à cette source ?**
 - Nécessite de disposer d'un smartphone / d'un forfait « data » / de savoir l'utiliser ;
 - Perte de données possible : panne de batterie, désactivation du GPS, volontaire ou non ;
 - Différences de tracking selon le système d'exploitation du téléphone, sa version, la marque et le modèle ;
 - Les applications doivent être mises à jour régulièrement ;
 - Biais spécifiques aux applis d'enquêtes :
 - Acceptabilité : côté intrusif du tracking ;
 - Recrutement complexe : pas de méthode éprouvée à ce jour ;
 - Biais spécifiques aux applis de guidage/achat :
 - Souvent spécifiques à un certain type d'utilisateurs, pas représentatif de la population ;
- **Comment généraliser les données collectées à l'ensemble de la population ?**
 - Pas de méthode de redressement validée à ce jour : besoin d'autres données
- **Peut-on mesurer des évolutions ?**
 - Applis d'enquête : envisageable, si recrutement comparable ;
 - Applis de guidage/achat : difficile, les usages des applis évoluant vite
- **De quelles informations dispose-t-on sur les individus et leur foyer ?**
 - Applis d'enquêtes (selon l'appli) : questionnaire complémentaire possible
 - Applis de guidage/achat : marché peu connu...



Conclusion

Utilisation des données massives pour
modéliser les mobilités

- **Matrice interne sur un macro-zonage, pour le calage du choix de destination**
 - FMD : test à venir sur le modèle allégé de Vendée (Cerema – Département 85) ;
- **Matrice interne pour alimenter un modèle de choix d'itinéraire (zonage + fin)**
 - FMD : test sur le modèle de Gournay-en-Bray (Cerema – Région Normandie) ;
<https://www.cerema.fr/fr/actualites/cerema-adapte-ses-outils-analyse-prospective-deplacements>
 - FCD : test à venir pour calcul matrice PL dans le cadre du projet MMUST (AGAPE – Cerema – LISER – MMTP Luxembourg – Universités Namur et Liège, financement FEDER).
- **Matrice interne additionnelle pour les non-résidents / touristes**
 - FMD : ventilation potentielle des flux totaux OD entre résidents et non-résidents ;
→ Pas d'expérimentation connue à ce jour en France.
- **Matrice externe (échange / transit)**
 - FCD/LAPI : Enquête transit sur le périphérique de Rennes et cordon rocade ouest (Cerema - DREAL Bretagne - Rennes Métropole) ;
 - FMD/LAPI : Construction d'une matrice OD externe sur l'agglomération genevoise (Cerema - Groupement Local de Coopération Transfrontalière) ;
 - <https://www.cerema.fr/fr/actualites/cerema-suit-transfrontaliers-trace-gsm>.

Calage du choix d'itinéraire

- **En temps de parcours**
 - Route : données Bluetooth, FCD ;
→ Fréquemment réalisé.
- **En débits ponctuels**
 - Calage réalisé classiquement à partir de données de comptages automatiques ;
 - A ce stade, les données massives reposent dans la majorité des cas sur des données « classiques » pour le redressement ;
→ Pas d'expérimentation connue à ce jour en France.
- **En débits pour des entrées/sorties**
 - Calage réalisé classiquement à partir de données LAPI, de transactions péage, de billétique ;
 - Route : possibilité d'utiliser également le Bluetooth ;
→ Pas d'expérimentation connue à ce jour en France.
- **En poids relatifs des itinéraires**
 - Potentiel pour FCD et traces GPS pour calibrer le choix des itinéraires dans un modèle ;
 - Les problèmes de représentativité doivent être résolus au préalable ;
→ Pas d'expérimentation connue à ce jour en France.

Autres usages envisageables

- **Adaptation temporelle des résultats des modèles**

- Les enquêtes par sondage ne permettent généralement pas de mesurer des variations en fonction du jour ou de la saison (pointe du vendredi soir pour le ferroviaire, flux saisonniers...) ;
- Les données massives « NTIC » permettent de mesurer des évolutions, notamment de flux OD, mais problématique de représentativité de la population.
- Combiner les deux sources => transposer les résultats du modèle à d'autres périodes que sa période de calage ;
→ Pas d'expérimentation connue à ce jour en France.

- **Construction de modèles désagrégés**

- Les données massives pourraient être utilisées pour calibrer des modèles désagrégés liant les caractéristiques des individus à leurs comportements : déduire des « préférences » ;
- Les données massives « NTIC » offrent, à ce jour, peu d'informations sur les individus (notamment pour des raisons d'anonymisation) ;
→ Potentiel à long terme et à ce stade incertain.

Conclusions générales

- **Les données massives ne peuvent pas être analysées directement via des tris à plats. Elles doivent être croisées avec d'autres données pour corriger leurs biais de représentativité. Elles peuvent donc permettre de caler des modèles, avec différents niveaux de complexité.**
 - Pas ou peu de nouvelles variables sur les comportements de déplacement ou les caractéristiques des individus → ne modifient donc pas les grands principes statistiques et algorithmiques des modèles ;
 - Les méthodes d'intégration dans les modèles sont en cours de test et présentent encore un très fort caractère **expérimental**.
- **L'évaluation de la pertinence d'une source de données massives nécessite :**
 - Une évaluation théorique des principes de collecte et de redressement (grille de critères en cours de consolidation par le Cerema) ;
 - Une comparaison sur un territoire test (et éventuellement un redressement) avec des données plus classiques (EMC², enquêtes OD, comptages...), à refaire pour chaque source, chaque changement de méthode... → Il est crucial de continuer à disposer de sources plus classiques.
- **Les coûts ne sont pas nécessairement plus faibles pour des données massives que pour des collectes par sondage sur le même champ.**

Pour aller plus loin...

FCD :

Représentativité des origines - destinations des usagers utilisant un système GPS

Maxime Le Corre, Gilles Bedat, Marie Giraud – Présentation et article pour le congrès ATEC 2018

FMD :

Données de téléphonie mobile pour la connaissance de la mobilité : enseignements de trois expérimentations

Alice Charpe, Julien Harache, Maxime Le Corre, Olivier Richard, Wilfried Raballand

Présentation et article pour le congrès ATEC 2019

Utilising mobile network data for transport modelling

Catapult Transport Systems, Department for Transport of the United Kingdom

<https://www.gov.uk/government/publications/mobile-phone-data-in-transport-modelling>



Contact : modelisation-deplacements@cerema.fr