

Quel apport des données issues des nouvelles technologies à la connaissance des mobilités ?

Présentation par : Aurélie Bousquet, Cerema Territoires et Ville

Co-auteurs des travaux (par ordre alphabétique) :

Alice Charpe, Cerema Ouest,
Julien Harache, Cerema Ouest,
Maxime Le Corre, Cerema Est,
Maria Tébar, Cerema Nord-Picardie.

Typologie des données quantitatives de mobilité : définitions

Collecte active (via un questionnaire)

Collecte passive

Représentativité
forte

Enquêtes à échantillons représentatifs

Échantillons représentatifs d'une population, obtenus par **sondage aléatoire**. Généralisation à toute la population grâce aux propriétés asymptotiques des estimateurs (loi des grands nombres...).

Collectes passives exhaustives

Observations **tendant vers l'exhaustivité** du flux **en un point du réseau**.

Représentativité
faible

Enquêtes à échantillons non-représentatifs

Variabilité des échantillons (parfois garantie par une méthode par quotas), **sans garantie de représentativité** de la population

Collectes passives non-exhaustives

Observations qui peuvent être nombreuses, mais **pas de maîtrise qualitative de l'échantillon**.

Données massives



**1. Enquêtes à échantillons
représentatifs**

3. Collectes passives
exhaustives

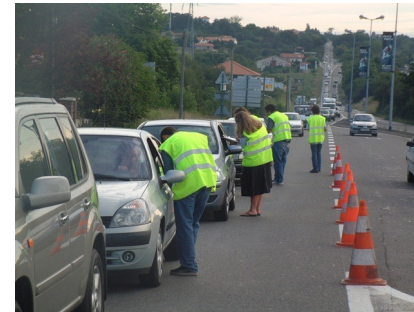
2. Enquêtes à échantillons
non-représentatifs

4. Collectes passives
non-exhaustives

Auprès des résidents d'un territoire



Auprès des usagers d'un axe / d'un arrêt



Crédit photo : Cerema



Enquête à Montpellier – Crédit photo : France 3

Grandes caractéristiques

1. Enquêtes à échantillons représentatifs

3. Collectes passives exhaustives

2. Enquêtes à échantillons non-représentatifs

4. Collectes passives non-exhaustives

	EMC ²	Enquêtes OD (routières / TC à bord ou aux arrêts)	Recensement de la population
Enrichissement sur les individus / les foyers / les véhicules	●	●	●
Connaissance modes	●	●	●
Connaissance motifs	●	●	●
Représentativité	● (taux échantillonnage : 1 à 3 %)	● (taux échantillonnage : 10 à 25 %)	●
Mesure d'évolutions	●	●	●
Producteur données	Collectivités / État / Cerema	Collectivités / État / Cerema	INSEE
Précision spatiale OD	IRIS voire (X,Y)	Commune, voire adresse	Commune
Autres	Attention : pas d'intégration des résidences collectives dans l'échantillon	Enquêtes routières fournissent aussi le taux d'occupation des véhicules	Attention : pas des déplacements mais des potentiels !



1. Enquêtes à échantillons
représentatifs

3. Collectes passives
exhaustives

**2. Enquêtes à échantillons
non-représentatifs**

4. Collectes passives
non-exhaustives

Grandes caractéristiques

1. Enquêtes à échantillons représentatifs

3. Collectes passives exhaustives

2. Enquêtes à échantillons non-représentatifs


4. Collectes passives non-exhaustives

	Enquêtes de préférences déclarées
Enrichissement sur les individus / les foyers / les véhicules	●
Connaissance modes	●
Connaissance motifs	●
Représentativité	● Variabilité, mais pas de représentativité
Mesure d'évolutions	● Forte variabilité des méthodes
Producteur données	Collectivités / État / Cerema / Chercheurs
Précision spatiale OD	Selon besoins
Autres	Questionnaires « à la carte » selon les problématiques visées


• Objectif


- Tester les effets de modifications hypothétiques de l'offre de transport sur les comportements, en particuliers de choix modal ;
- Application à une population synthétique d'un modèle* de comportement des individus calibré à partir de l'enquête.


Option 1 - Voiture





Option 2 - Train





 17 minutes de voiture


 14 minutes de train


 Le trajet coûte 2.40 € (essence et péage).

 Un aller simple plein tarif coûte 3.00 €

 Compte tenu du trafic, il est possible que vous preniez un retard de 5 minutes.

 Les trains passent tous les quarts d'heure.

 Les trains passent à horaires fixes.

 Il est possible que le train arrive avec un retard de 15 minutes.

* Généralement, modèles de choix discrets :

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



1. Enquêtes à échantillons
représentatifs

**3. Collectes passives
exhaustives**

2. Enquêtes à échantillons
non-représentatifs

4. Collectes passives
non-exhaustives

Dispositifs de collecte

1. Enquêtes à échantillons représentatifs

3. Collectes passives exhaustives

2. Enquêtes à échantillons non-représentatifs

4. Collectes passives non-exhaustives

Comptages routiers, ponctuels, directionnels ou en entrée/sortie, permanents ou temporaires

Boucles électromagnétiques



Crédit photo : Accor solutions

Transactions péage



Crédit photo : Le blog auto

Pneumatiques



Crédit photo : Nouvelles Pulications

Radar



Crédit photo : Sferiel

Lecture automatique de plaques d'immatriculation



Crédit photo : Pixalarm

Comptages montées/descentes dans les transports en commun

Comptages manuels / automatiques

Données billétiques



Grandes caractéristiques

1. Enquêtes à échantillons représentatifs

3. Collectes passives exhaustives

2. Enquêtes à échantillons non-représentatifs

4. Collectes passives non-exhaustives

	Mesures ponctuelles de flux / vitesses (route)	Lecture de plaques d'immatriculation	Transactions péage	Billétique TC
Enrichissement sur les individus / les foyers / les véhicules	●	●	●	●
Connaissance modes	●	●	●	●
Connaissance motifs	●	●	●	●
Représentativité	●	● (sauf masquage de plaque)	●	●
Mesure d'évolutions	●	●	●	●
Producteur données	Collectivités / État / Cerema	Collectivités / État / Cerema	Sociétés Concessionnaires d'Autoroutes	Exploitants réseaux TC
Précision spatiale OD	Sans objet	Sans objet, points de passage uniquement		
Autres	Erreur de 5 à 15 % selon les matériels	Peut fournir des volumes par entrée/sortie et des vitesses (si itinéraire évident)		Fraude importante constatée sur certains réseaux (plus de 30 %)



1. Enquêtes à échantillons représentatifs

3. Collectes passives exhaustives

2. Enquêtes à échantillons non-représentatifs

4. Collectes passives non-exhaustives

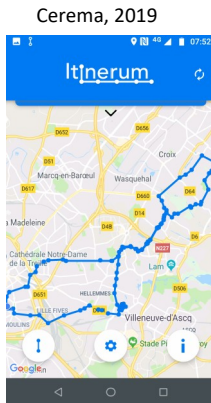
Données massives

Dispositifs de collecte

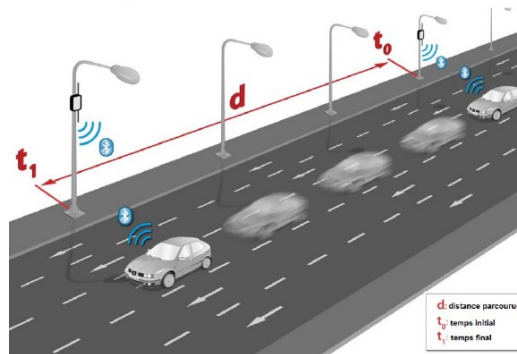
1. Enquêtes à échantillons représentatifs	3. Collectes passives exhaustives
2. Enquêtes à échantillons non-représentatifs	4. Collectes passives non-exhaustives

Données massives
← Big-data →

Traces GPS
d'applications smartphone



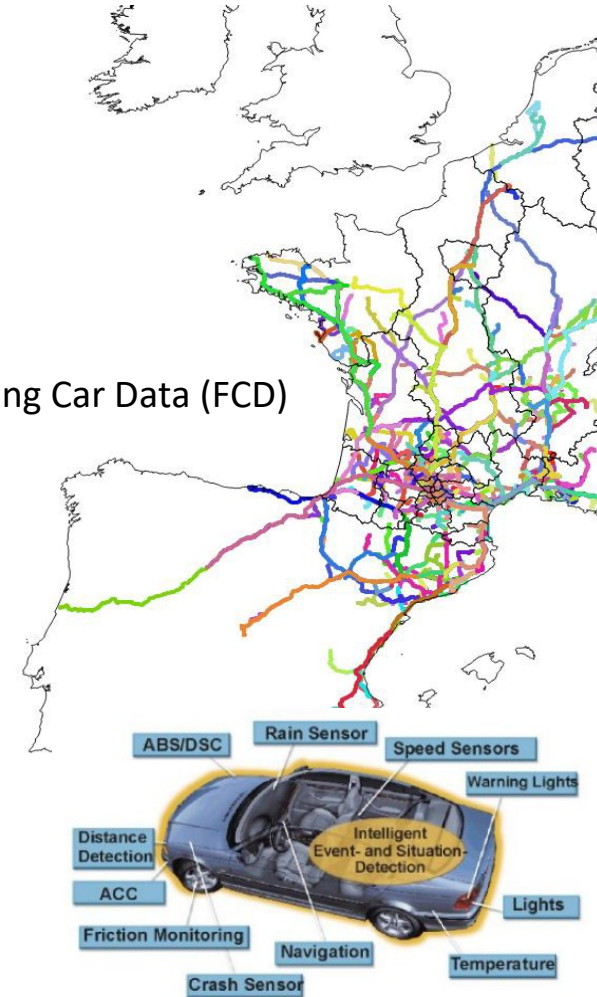
Bluetooth



Floating Mobile Data (FMD)



Floating Car Data (FCD)



Grandes caractéristiques

1. Enquêtes à échantillons représentatifs

3. Collectes passives exhaustives

2. Enquêtes à échantillons non-représentatifs

4. Collectes passives non-exhaustives

	Bluetooth	FMD	FCD	Traces GPS applis smartphone
Enrichissement sur les individus / les foyers / les véhicules	●	●	●	●, ● ou ● selon les sources
Connaissance modes	●	● Routier/ferroviaire éventuellement	●	● à ● selon les sources
Connaissance motifs	●	● Envisageable d'identifier le domicile et le lieu d'activité régulière		
Représentativité	●	●	● à ● selon les sources	●
Mesure d'évolutions	●	● Sur - d'1 an	● Sur - d'1 an (sauf vitesse)	● Sur - d'1 an
Producteur données	Collectivités / État / Cerema	Opérateurs téléphonie	Constructeurs automobiles, agrégateurs de données	Agrégateurs de données, citoyens
Précision spatiale OD	● Points de passage uniquement	●	●	●
Autres	Peut fournir des volumes par entrée/sortie et des vitesses (si itinéraire évident), adressage MAC dynamique compromet son utilisation	Voir diapos suivantes		Applis spécifiques pour les enquêtes ou réutilisation de données d'applis de guidage ou d'achat

Floating Mobile Data (FMD)

1. Enquêtes à échantillons représentatifs

3. Collectes passives exhaustives

2. Enquêtes à échantillons non-représentatifs

4. Collectes passives non-exhaustives

- **Quelles sont les données collectées ?**
 - Données de localisation liées aux évènements recensés par les antennes (appels, messages, data, « Ping »)
- **Comment découpe-t-on les déplacements ?**
 - Temps de « stationnarité » a minima de 15 min → on perd les petits déplacements.
 - Algorithme de traitement complexe pour lier localisation et déplacements.
- **Quelle est la précision spatiale de localisation des origines et des destinations ?**
 - Flux à grandes échelles, pas de flux internes à une agglomération ;
 - Précision liée à la localisation des antennes + problématique de « frontière » entre « zones-antennes ».
- **Quels sont les biais de collecte spécifiques à cette source ?**
 - Échantillonnage temporel irrégulier et hétérogène ;
 - Sous-représentation des personnes âgées ;
 - Données d'un seul opérateur : pas de commercialisation d'agrégats aujourd'hui.
- **Comment les données collectées sont-elles généralisées à la population ?**
 - Traitement algorithmique propre à chaque fournisseur et rarement transmis ;
 - Redressement par rapport à la part de marché, mais pas nécessairement sur l'âge.
- **De quelles informations dispose-t-on sur les individus et leur foyer ?**
 - Lieux d'activité et de résidence : en fonction de l'opérateur, ± fiable.

- **Quelles sont les données collectées ?**
 - Données de localisation horodatées fournies par un système GNSS (via un système d'aide à la conduite, intégré ou non au véhicule) fournissant :
 - des vitesses et donc 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 max, généralement < 10 m) ;
- **Comment découpe-t-on les déplacements ?**
 - Temps de « stationnarité » variable : nombreuses limites, car une grande partie des traces se termine sur une vitesse non nulle ;
 - Possibilité d'utiliser des algorithmes + poussés (par ex. pour filtrer les pauses PL).
- **Quels sont les biais de collecte spécifiques à cette source ?**
 - Pour les VL :
 - Vitesses légèrement au-dessus de la moyenne (= problème en situation fluide) ;
 - Classes de distances (donc OD) non-représentatives ;
 - Pour les PL : a minima, les classes de distance (donc OD) ne sont pas représentatives.

- **Comment généraliser les données collectées à l'ensemble de la population ?**
 - Vitesse : sans redressement, si l'échantillon est assez important
 - Attention : en fluide, veiller à ce que la source ne soit pas trop biaisée sur les comportements de conduite (ex : dispositif de signalement des radars ayant tendance à être utilisé plus fortement par les personnes qui roulent au-dessus des limites de vitesses) ;
 - 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.
- **De quelles informations dispose-t-on sur les véhicules ?**
 - FCD « classique » :
 - direction (cap) et vitesse instantanée du véhicule ;
 - 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.

Conclusions générales

- **Les données massives actuellement disponibles ne sont jamais complètement représentatives d'un flux, d'une population résidente ou d'une population présente sur un territoire :**
 - Elles ne peuvent donc pas être analysées directement via des tris à plats, comme le serait une enquête classique par sondage aléatoire.
 - Elles doivent être croisées avec d'autres données et des hypothèses doivent être faites pour compenser la non-représentativité : ça revient à construire un petit modèle.
 - Elles doivent être utilisées sur leur champ de pertinence, à bien définir en fonction du processus de production de la donnée (qui doit donc être connu !)*.
- **Pour leur utilisation en modélisation :**
 - Les données massives ne contiennent pas ou peu de données caractérisant les individus / les foyers / les véhicules. Elles ne permettent donc pas de lier les caractéristiques individuelles aux comportements de mobilité.
 - Leur utilisation est toujours complexe et limitée, à très fort caractère expérimental.
- **Leur détention par des opérateurs privés ne permet pas de garantir pérennité et comparabilité spatio-temporelle des données.**
- **Les coûts ne sont pas nécessairement plus faibles pour des données massives que pour des collectes par sondage sur un champ à peu près équivalent.**

* cf. liste de questions à se poser pour utiliser des données de mobilité

https://www.cerema.fr/system/files/documents/2019/09/cerema_grille_analyse_sources_donnees.pdf

Pour aller plus loin...

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

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

Utilizing 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>

Projet de modèle sur Gournay-en-Bray

<https://www.cerema.fr/fr/actualites/cerema-adapte-ses-outils-analyse-prospective-deplacements>

Enquête cordon transfrontalière du franco-valdo-genevois

<https://www.cerema.fr/fr/actualites/cerema-suit-transfrontaliers-trace-gsm>

Document complémentaire à ce diaporama (avec plus de détails sur les différentes sources)

<https://www.cerema.fr/fr/actualites/quel-apport-donnees-issues-nouvelles-technologies>



Contacts :

- modelisation-deplacements@cerema.fr
- observation.mobilite@cerema.fr