L’utilisation des traces GPS et de Smartphones

Séminaire conjoint GT "Ville et Numérique" et Projet "Mobilités et Transitions Numériques de l'IFSTTAR"

4 Avril 2019

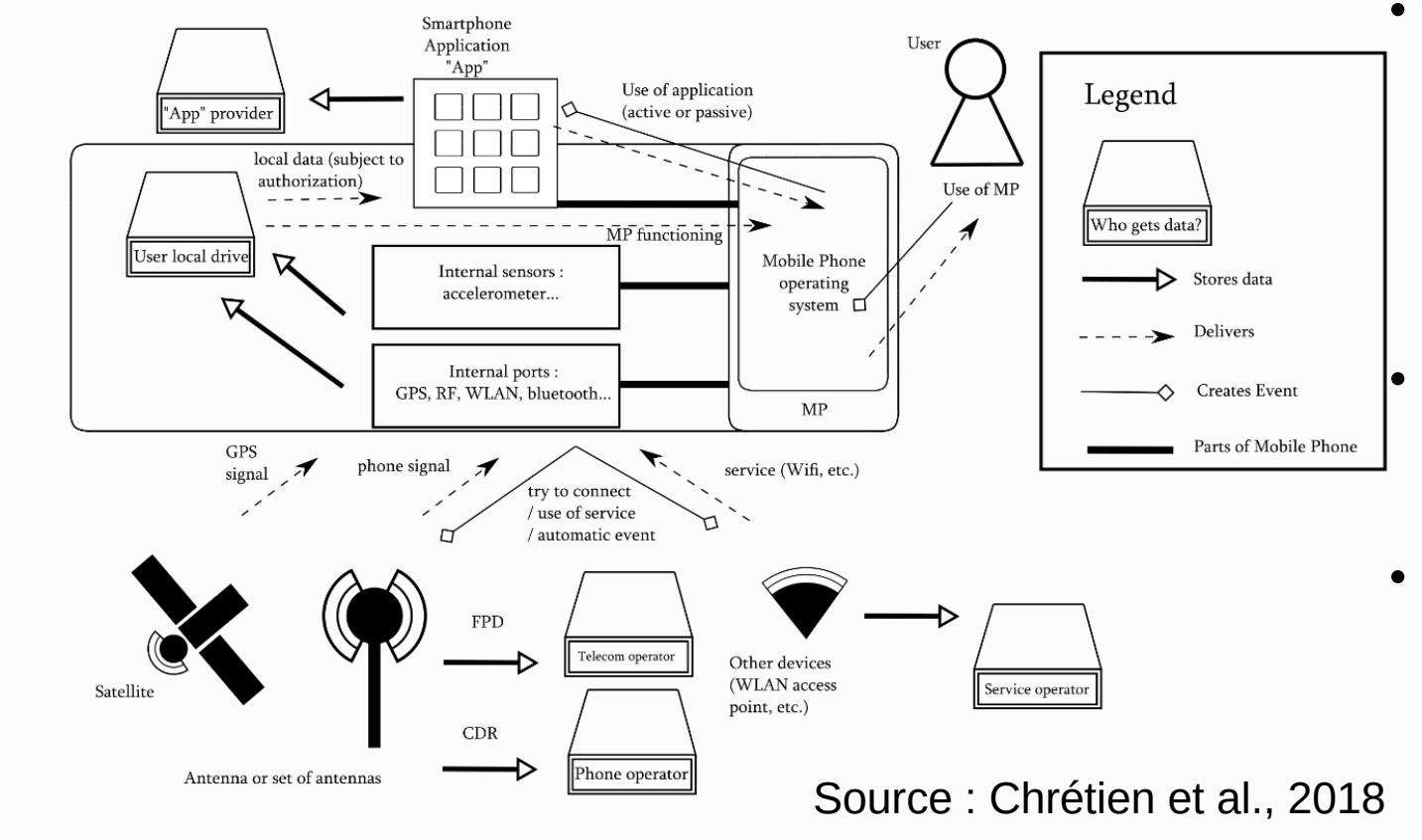
Julie Chrétien julie.chretien@enpc.fr

Florent Le Nechet florent.lenechet@u-pem.fr

Laurent Proulhac laurent.proulhac@enpc.fr

De quoi parle-t-on? – données de « tracking » issues d'applications

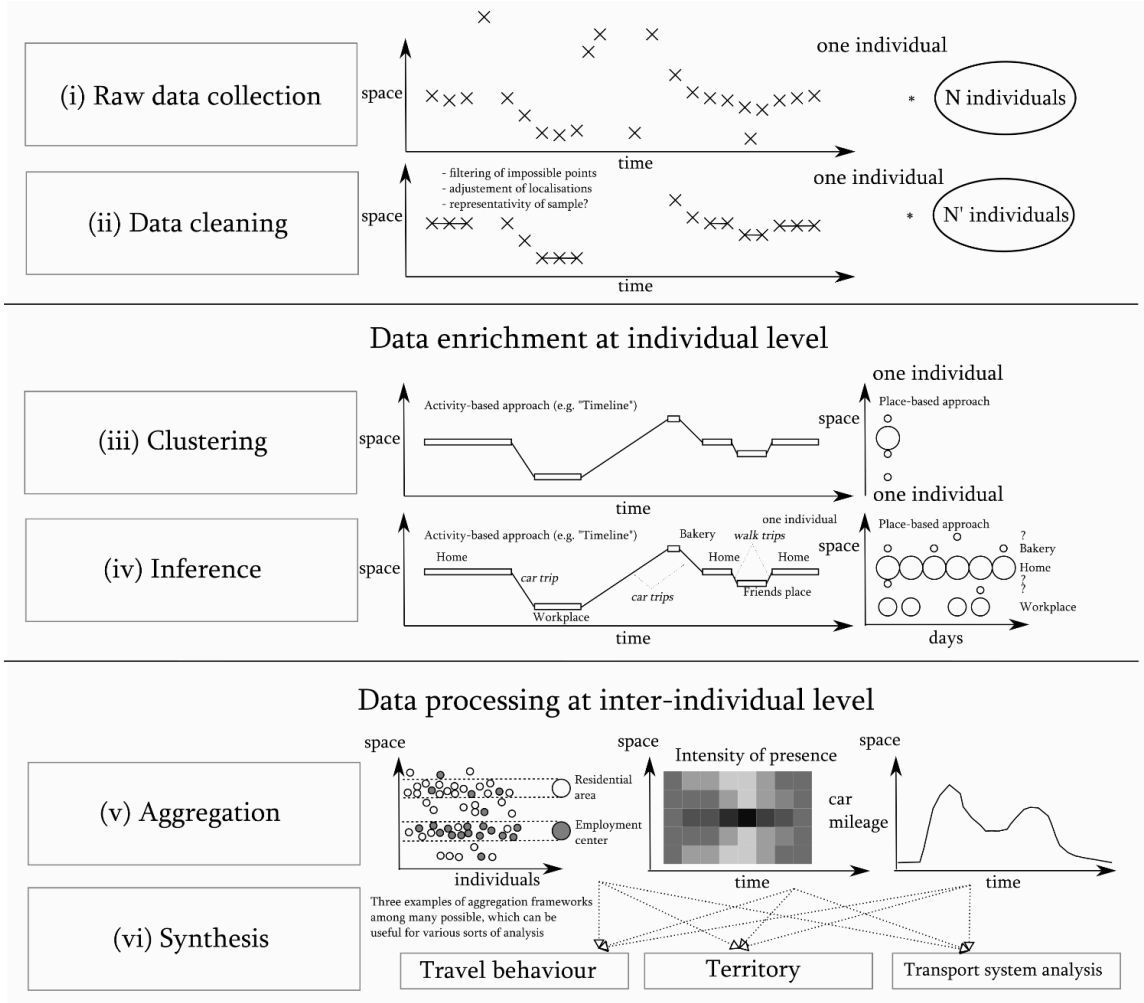
- Les téléphones alimentent beaucoup de bases de données internes et externes au téléphone…



Source : Chrétien et al., 2018

- Les applications smartphone peuvent combiner différents signaux
- Importance de la notion d' « événément »
- Taux de possession de smartphone chez les 14-69 ans : 83% (CREDOC, 2017)

De quoi parle-t-on? – « Timeline »



- La start up IT4PME (F. Lainée) et l'ENPC-LVMT (F. Leurent) sont partenaires au sein d'un projet de recherche « FUI Geolytics »
- Cette recherche se place en aval des premiers enrichissements effectués par IT4PME (mode de transport)
- 3000 téléphones, 1 million de points, 1 mois de données, début 2018

Source : Chrétien et al., 2018

Applications des données de téléphonie mobilité

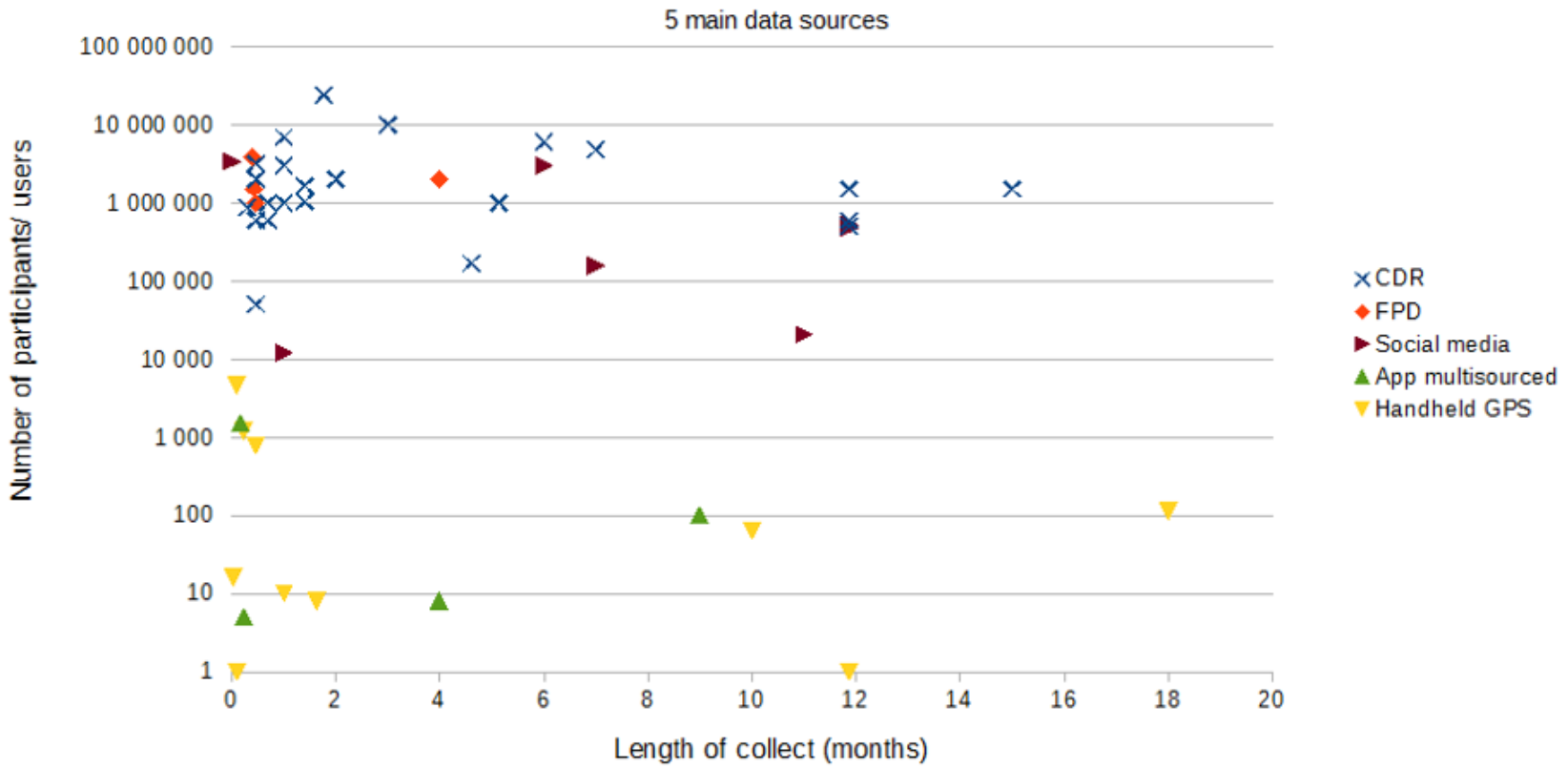
- Analyse des comportements de mobilité
  - Régularité du comportement de voyage
  - Mobilité des non-résidents
  - Liens entre mobilité physique, mobilité virtuelle et réseaux sociaux
- Analyse territoriale
  - Présence dans l'espace et le temps
  - Détection d'objets géographiques: centralités, limites de villes, réseaux de villes
  - Inférence des fonctions urbaines
- Analyse du système de transport
  - La caractérisation du trafic des éléments de transport: gares, segments et qualité de service
  - Caractérisation de la demande par Origine-Destination
  - Management du système de transport

Applications des données de téléphonie mobilité

- Attrait et enjeux spécifiques aux données issues d'applications smartphone :
  - Combiner précision Wifi GPS Antenne
  - Données « tous modes »
  - Longitudinalité de l'information
  - Question de la masse de données
  - Ajout d'informations autres que géolocalisation (ex. accéléromètre)
  - Acceptabilité par le développeur d'application / l'utilisateur
  - Défis techniques
- Extrait de Chrétien, J., Le Néchet, F., Leurent F., Yin B. (2018). Using mobile phone to observe and understand mobility behaviour, territories and transport usage. In, "Urban mobility and the smartphone, transportation, travel behaviour and public policy" by Aguilera, A. and Boutueil, V. (dir), Elsevier, 224 p.
  - (revue de littérature environ 80 articles étudiés)

Comparaison des données CDR / d'applications de recherché / d'applications commerciales

Figure 3: Number of participants per data set depending on the number of month of collected data
(source: review by the authors of 47 datasets mentioned in the literature.
Method: Were kept only cases where the original size of the dataset before filtering was mentioned.)



Comparaison des données CDR / d'applications de recherché / d'applications commerciales

	CDR / FPD	Research apps	Commercial tracking apps
Contrôle de la représentativité	non	oui	non
Taille de l'échantillon	~millions	~centaines	~milliers / centaines de milliers
Précision temporelle	faible	haut	haut
Précision spatiale	faible	haut	haut
Caractéristiques individuelles	non	oui	non
Durée de suivi	années	mois/ années	années

Données testées: la base “Geolytics”

- Issue de 2 applications smartphone :
  - Application “Geo4cast” (recherche d’itinéraire)
  - Autre application grand public
- En théorie :
  - Géolocalisation (ou “évènement”) toutes les 3 minutes
  - Prend la meilleure source de localisation: wifi>GPS>antenne
- En réalité :
  - Les utilisateurs éteignent leur téléphone
  - Téléphones donnent location seulement quand en mouvement
  - Les OS (Apple, Samsung) interagissent différemment avec les applications de tracking
  - Les données en sous-sol posent problème

➔ Moy.: 420 pts/pers, med.: 223 pts/pers (14 pts/jours/pers)

➔Filtrage : garde individus avec au moins 7 jours dotés de 5 points ou plus ➔ restent ~1900 individus

Traitement de l'information

- Comment qualifier l'information acquise ?
- Fréquent dans la littérature : inférence sur base de données externes (Fen Chong, 2012)

- L'algorithme IB-SMoT (*Intersection-Based Stops and Moves of Trajectories*)

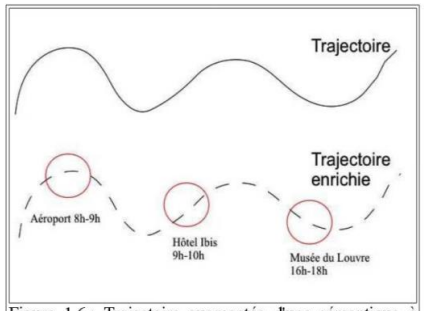
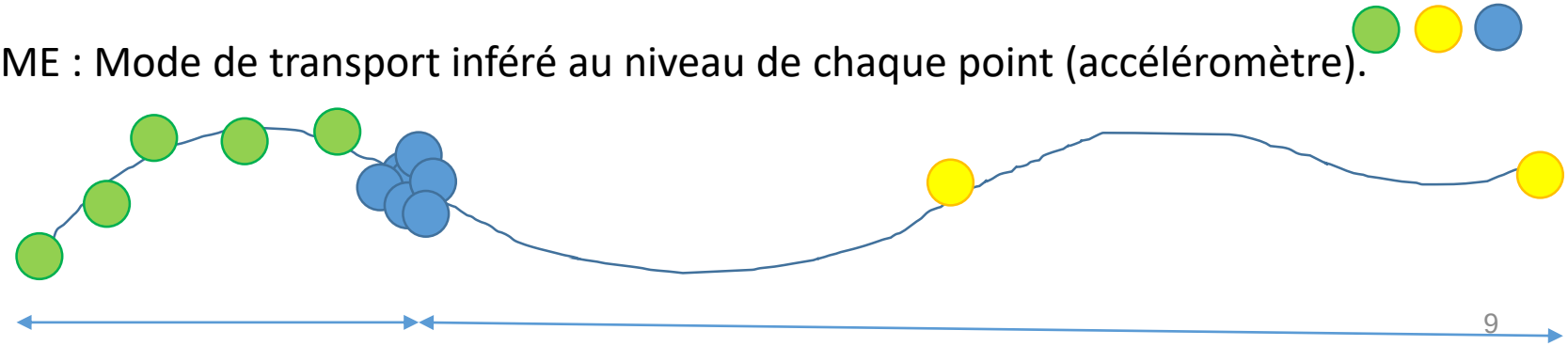


Figure 1.6 : Trajectoire augmentée d'une sémantique à l'aide des lieux et temps d'arrêt à partir de données de GPS
(Source : d'après Alvares et alii, 2007)

- Contrainte de projet : inférence uniquement sur base de la Timeline fournie : test de solutions génériques « simples »
- Un « log » toutes les 3 minutes

IT4PME : Mode de transport inféré au niveau de chaque point (accéléromètre).



LVMT : Déplacements inférés sur base de l'ensemble des traces

Objectifs et plan de travail

Objectifs	**Plan de travail**
Evaluer la représentativité du jeu de données du point de vue des "masses" (lieu de residence, etc.)	Etape 1 : identifier lieu de residence et lieu de travail
Inférer les trajets domicile-travail en cherchant un algorithme "optimal"	Etape 2 : identifier les trajets domicile-travail
Evaluer la représentativité du jeu de données du point de vue des mobilités (Enquête Globale Transport, 2010)	Etape 3 : comparaison avec l'Enquête Globale de Transport
Tester l'apport de telles données pour l'analyse de la mobilité (longitudinalité)	Etape 4 : Etudier des phénomènes qui ne peuvent pas être appréhendés avec les enquêtes classiques

Données de comparaison

	EGT 2010	ENTD 2009	Recensement 2012
Type de données de mobilité	Une journée de déplacement (semaine)	Un jour de semaine et un jour de weekend	Trajets domicile-travail
Nombre d'individus en Île-de-France	35175	?	Exhaustif
Filtres appliqués lors de l'étude	Limités aux 12-69 ans	Limités aux 12-69 ans	Limité aux gens qui travaillent hors de leur commune

Etape 1

Inférence du lieu d’habitation et du lieu de travail

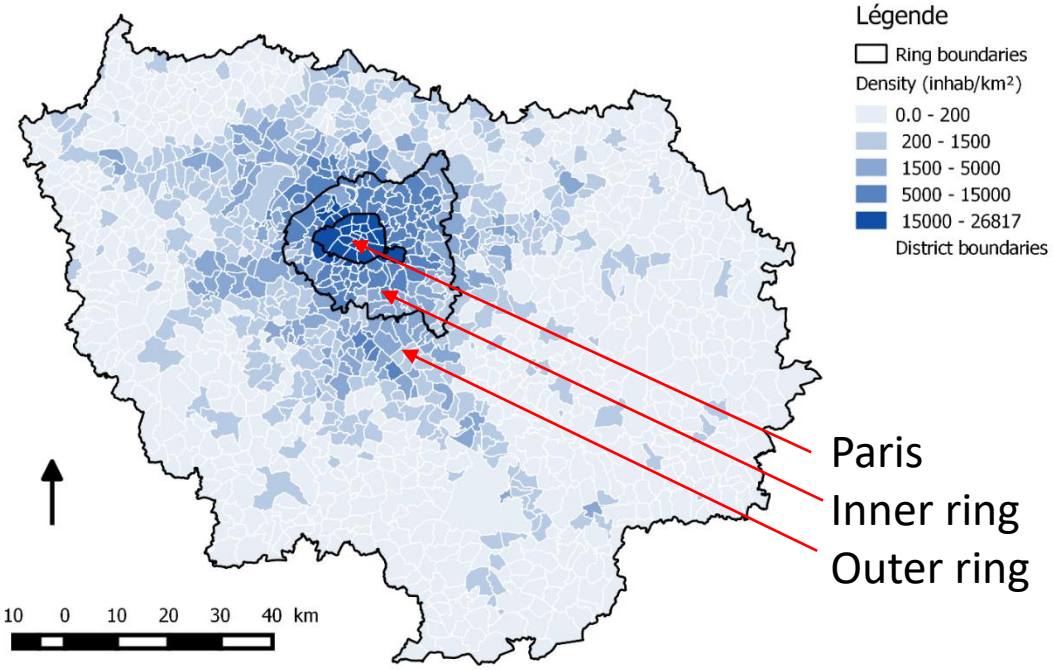
Inferring individual “clusters”

Points enriched with district information and other stable information

Aggregate successive points in same district

For each individual:
- list of districts where has been
- moments when has been there

Districts and “rings” in the Parisian Metropolitan Area



~1200 districts
Average area of a district: 9.4 km²

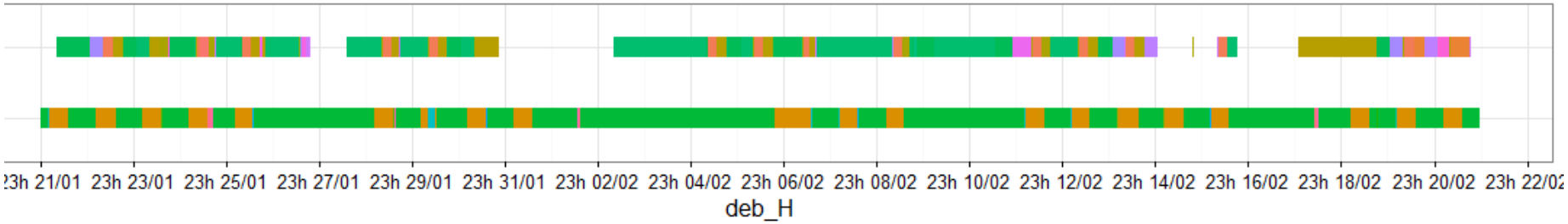
Inférer des “clusters” individuels

Points enrichis du code de la commune où se trouve

Agrégation des points successifs dans la même commune

Bases avec pour chaque individu
- liste des communes où s’est trouvé
- moments où s’y est trouvé

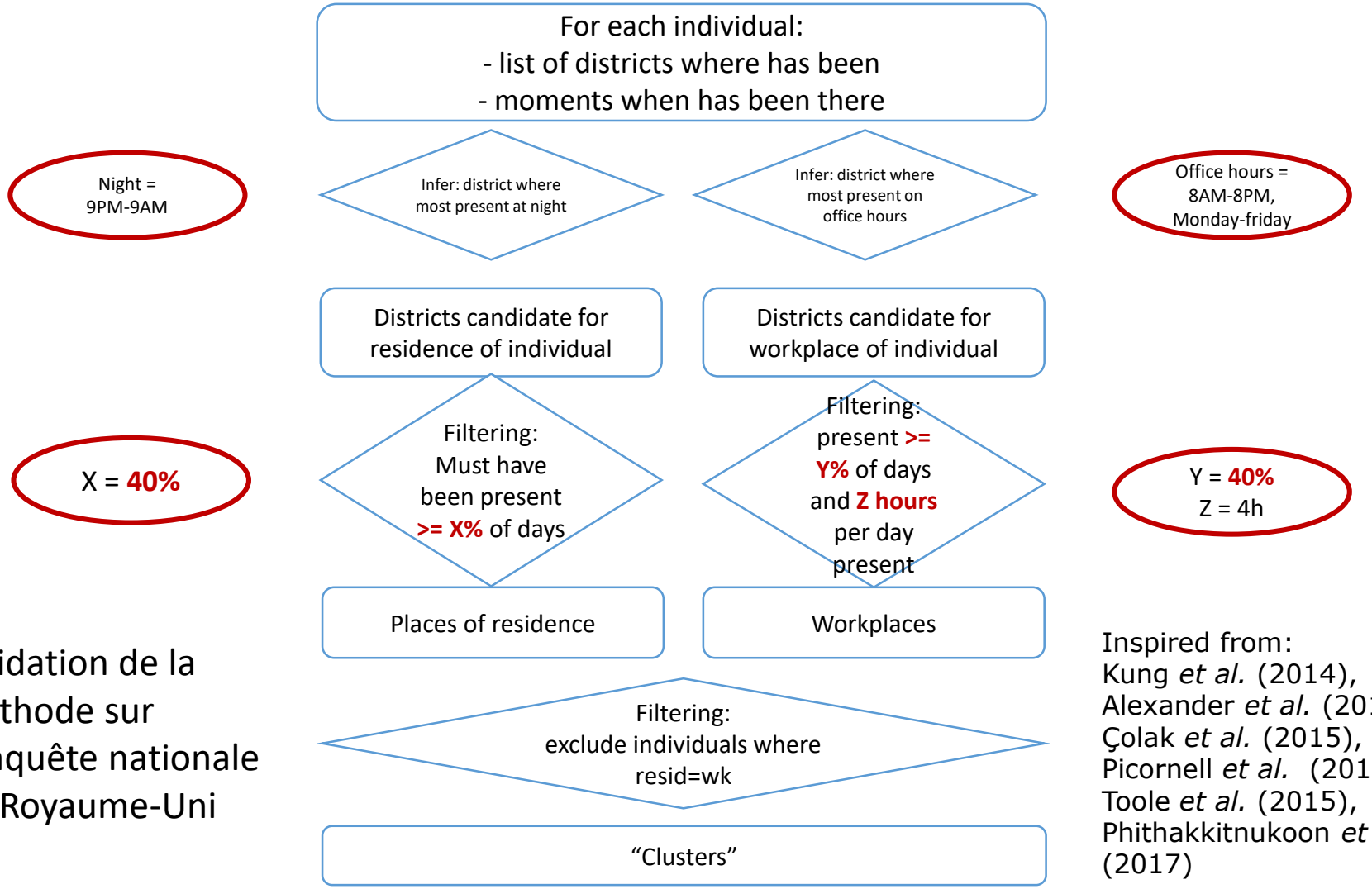
Résultat : « emploi du temps » de l’individu
Exemple de deux emplois du temps d’individus distincts



Couleur différente = commune différente

➔ Jeu : à quelle couleur correspond le domicile ? Le lieu de travail ?

Etiqueter les clusters : algorithme



Validation de la méthode sur l'enquête nationale du Royaume-Uni

Inspired from:
Kung *et al.* (2014),
Alexander *et al.* (2015),
Çolak *et al.* (2015),
Picornell *et al.* (2015),
Toole *et al.* (2015),
Phithakkitnukoon *et al.* (2017)

Les stocks de lieu de residence sont approximativement retrouvés, les lieux emplois moins bien

Home	French Census	Geolytics	Différence
Paris	23%	19%	4 pts
Inner ring	40%	34%	6 pts
Outer ring	37%	47%	10 pts

Work	French Census	Geolytics	Différence
Paris	32%	56%	24 pts
Inner ring	38%	32%	6 pts
Outer ring	28%	12%	16 pts

People working outside district of residence

Cluster filtering: x=40%, y=40%

Répartition des origines destination encore imparfaite

French Census (2012)

Home \ Work	Paris	Inner ring	Outer ring
Paris	16%	6%	1%
Inner ring	12%	24%	4%
Outer ring	6%	8%	23%

People working outside district of residence

Geolytics

Home \ Work	Paris	Inner ring	Outer ring
Paris	11%	7%	1%
Inner ring	20%	12%	2%
Outer ring	25%	13%	9%

Cluster filtering: x=40%, y=40%

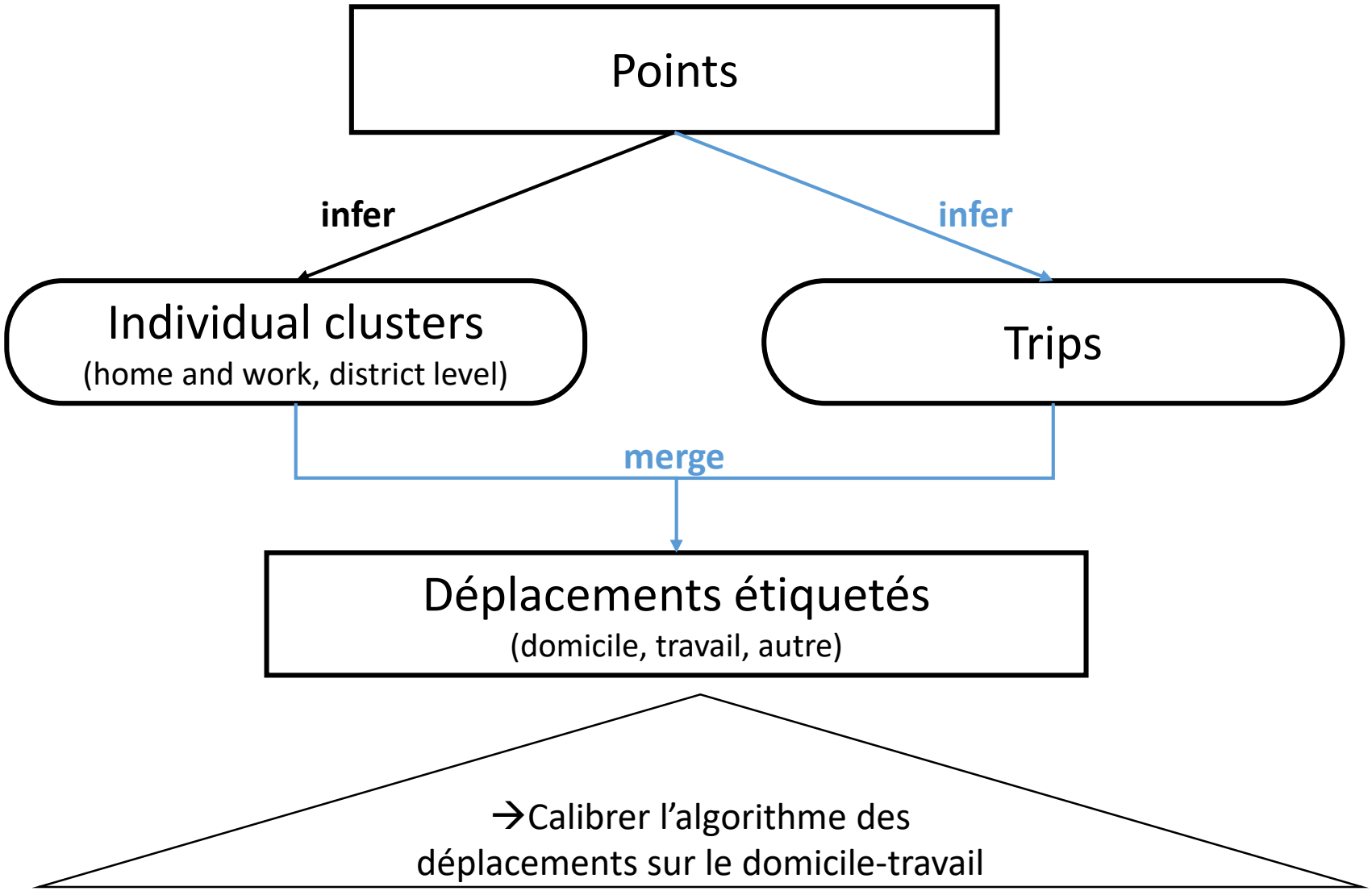
➔ Représentativité des OD imparfaite à ce stade

➔ On étudie les indicateurs OD par OD

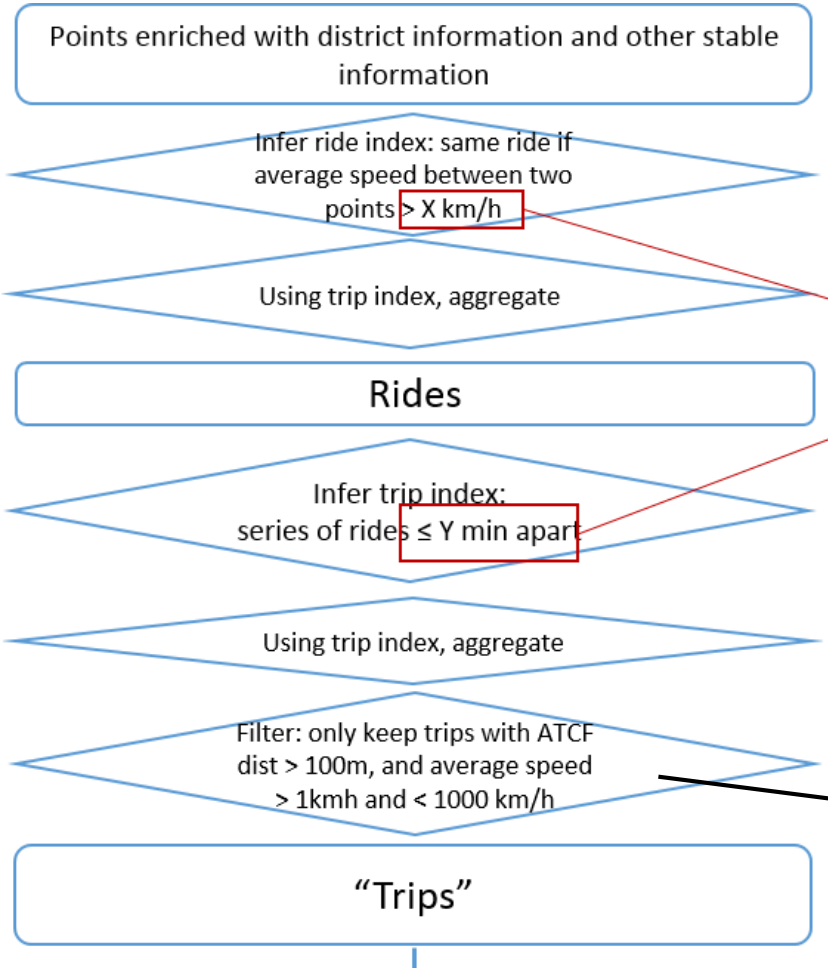
Etape 2

Inférer les déplacements

Workflow


Points
infer
infer
Individual clusters
(home and work, district level)
Trips
merge
Déplacements étiquetés
(domicile, travail, autre)
→Calibrer l'algorithme des
déplacements sur le domicile-travail

Inférer les déplacements: algorithme à deux paramètres



X, Y = VARY

X : vitesse entre deux points
Y : temps immobile entre deux trajets

Exclure déplacements dont :
- Durée >12h
- Durée <1 min
- portée/durée <1 km/h
- Portée/durée >1000 km/h
- Distance <100m

Calibration des paramètres (méthode)

Indicator with most variability: **time** 15%-150% (whereas distance: 0%-10%)

Example: Calculate time (by OD), norm L2

Home -> Work	**Trip-based (5, 3, 20)**	**EGT**
Paris -> Paris	16.5	31.8
Paris -> Inner Ring	33.6	45.4
Paris -> Outer Ring	52.8	62.2
Inner Ring -> Paris	29.8	47.1
Inner Ring -> Inner Ring	33	41
Inner Ring -> Outer Ring	42.6	48.7
Outer Ring -> Paris	37.5	68
Outer Ring -> Inner Ring	45.8	59.5
Outer Ring -> Outer Ring	28	32.3
Norm L2	**45**	**/**

Etape 3

Comparaison avec l’Enquête Globale de Transport

Répartition des motifs de déplacement

	Whole population		Population with stable workplace and full-time job	
	EGT	Geolytics	EGT	Geolytics
Home	38.3%	37.5%	36.1%	34.5%
Work	11.5 %	13.4%	26.9%	29.3%
Others	50.2%	49.1%	37.0%	36.2%
Total	100%	100%	100%	100%

Geolytics: movement-based algorithm (minimum speed = 2 km/h, maximum pause between rides = 15 minutes)
NB: for EGT, the purpose "Other work-related purpose" was classified as "Others"

Comparaison des temps de trajets par OD(minutes)

(H->W only)

EGT 2010

Home \ Work	Paris	Inner ring	Outer ring
Paris	32	45	62
Inner ring	47	41	49
Outer ring	68	60	32

Geolytics

Trip-based algorithm

Home \ Work	Paris	Inner ring	Outer ring
Paris	25	44	NA
Inner ring	42	43	NA
Outer ring	46	54	37

Comparaison des distances de trajets(km)

(H->W only)

Home \ Work	Paris	Inner ring	Outer ring
Paris	4.0	8.0	22.3
Inner ring	8.8	7.7	16.2
Outer ring	23.8	18.2	11.1

Home \ Work	Paris	Inner ring	Outer ring
Paris	3.2	8.0	NA
Inner ring	8.8	8.0	NA
Outer ring	24.6	17.4	9.2

Comparaison des parts modales (%)
(tous motifs)

- Au niveau global

	All trips	
	EGT	Geolytics
Train	9%	20%
Metro	6%	0%
Bus	5%	6%
Car or taxi	46%	13%
Walk	30%	28%
Others	4%	0%
Unidentified	0%	33%

Comparaison des parts modales (%)

(H->W only)

%train (pas métro!) parmi les modes identifiés

EGT 2010

Home \ Work	Paris	Inner ring	Outer ring
Paris	4%	31%	55%
Inner ring	40%	18%	22%
Outer ring	81%	42%	6%

Geolytics

Trip-based algorithm

Home \ Work	Paris	Inner ring	Outer ring
Paris	0%	18%	NA
Inner ring	38%	17%	NA
Outer ring	84%	40%	13%

%voiture (idem)

EGT 2010

Home \ Work	Paris	Inner ring	Outer ring
Paris	7%	21%	33%
Inner ring	18%	44%	69%
Outer ring	15%	51%	77%

Geolytics

Trip-based algorithm

Home \ Work	Paris	Inner ring	Outer ring
Paris	12%	23%	NA
Inner ring	17%	30%	NA
Outer ring	7%	25%	58%

Discussion des résultats

- Approche exploratoire sur des données peu nombreuses, ayant une faible granularité temporelle
- Au niveau global (niveau régional agrégé) : ces données spécifiques ne sont pas représentatives et ne peuvent être utilisées telles quelles pour un diagnostic territorial
- En regardant les déplacements domicile-travail, par OD : au contraire les résultats sont plutôt satisfaisant (notamment Inner Ring > Paris)
- Potentiel important de ces données (avec massification + usage de données externes)
- Développement d'un code R assez générique

Etape 4

Application à des variables indisponibles dans les EMD

Rares données sur les activités de loisirs “courtes”

Une demande d’information sur les loisirs :

- Une part croissante de notre mobilité y est liée
- Compréhension nécessaire aux “Activity-based models”

Limites de « l’offre » :

- Etudes de mobilité portent sur les « grands » loisirs (tourisme, séjours)
- Rareté des études sur la mobilité quotidienne non contrainte

Difficultés à étudier les loisirs hors tourisme avec les EMD :

- Fréquence hebdomadaire ou mensuelle
- Hétérogénéité inter et intra-individuelle des activités

Focus sur activités du soir à l'extérieur (ASE)

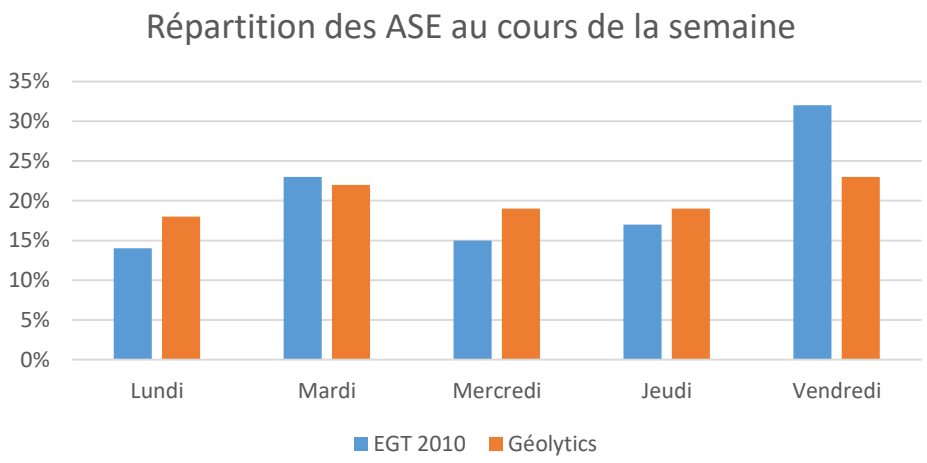
Loisirs = ASE = activités :
- Où les gens sont arrivés entre 19h et 22h
- Où ils sont restés entre 30 minutes et 4h30

Filtrage
- Individus dont la commune de résidence a été identifiée en Île-de-France

➔ ~1200 individus

Comparaison avec EGT 2010, individus entre 12 et 69 ans

Part de la population ayant fait un ASE en semaine :
- EGT 2010 : 8%
- Géolytics : 9,5%



Loisirs de semaine et excursions le weekend

-> Cumul ou compensation ?

Hypothèse sens commun :
« on n’a que 24h/ jours et 7 jours / semaine » -> compensation

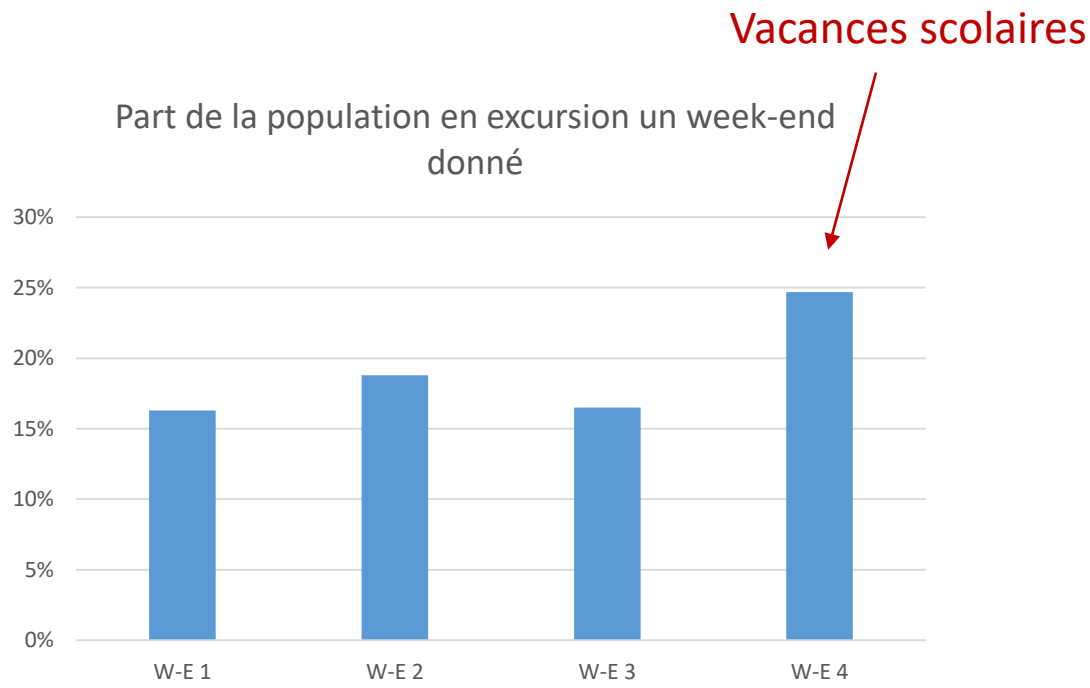
Hypothèse Bourdieu et littérature sur la consommation de loisirs :
« plus on fait de loisirs, plus on fait de loisirs » -> cumul

Identifier des “excursions” le weekend

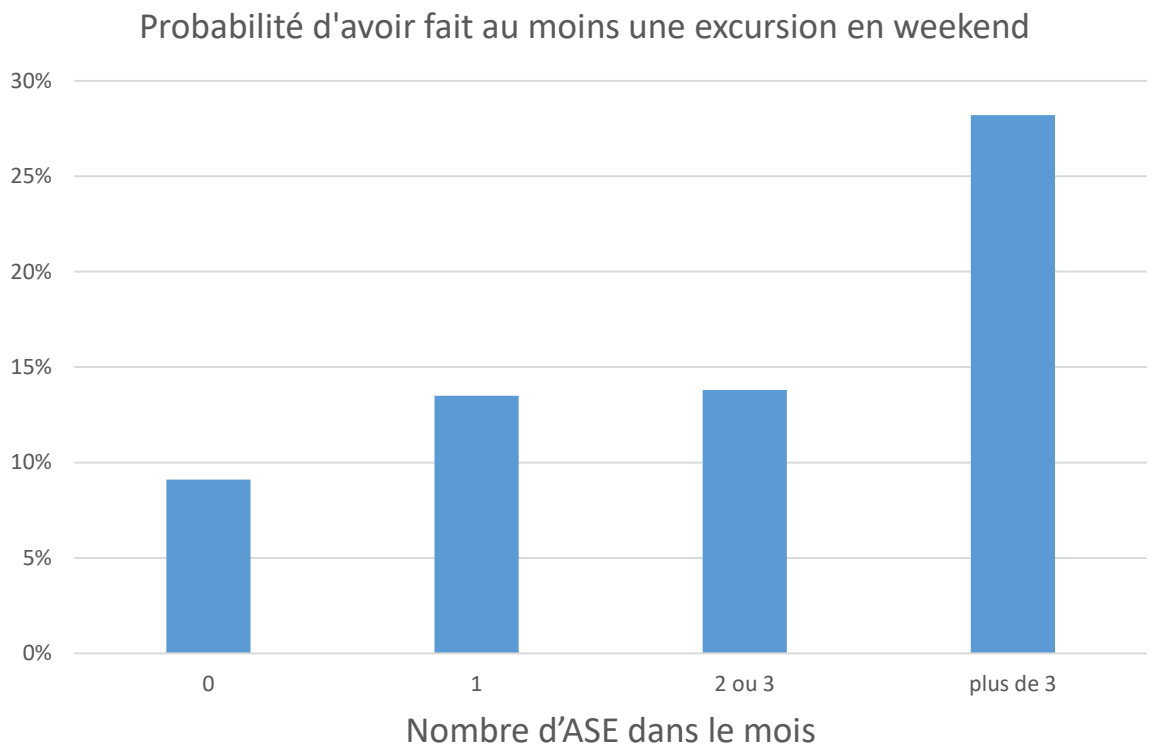
Gens sont en « excursion » si ils ont été absent de leur département de résidence de minuit à 9h du matin la nuit du samedi au dimanche (hors trou dans les données)

Filtres : idem que pour ASE

60 % des individus n’ont fait aucune excursion dans le mois



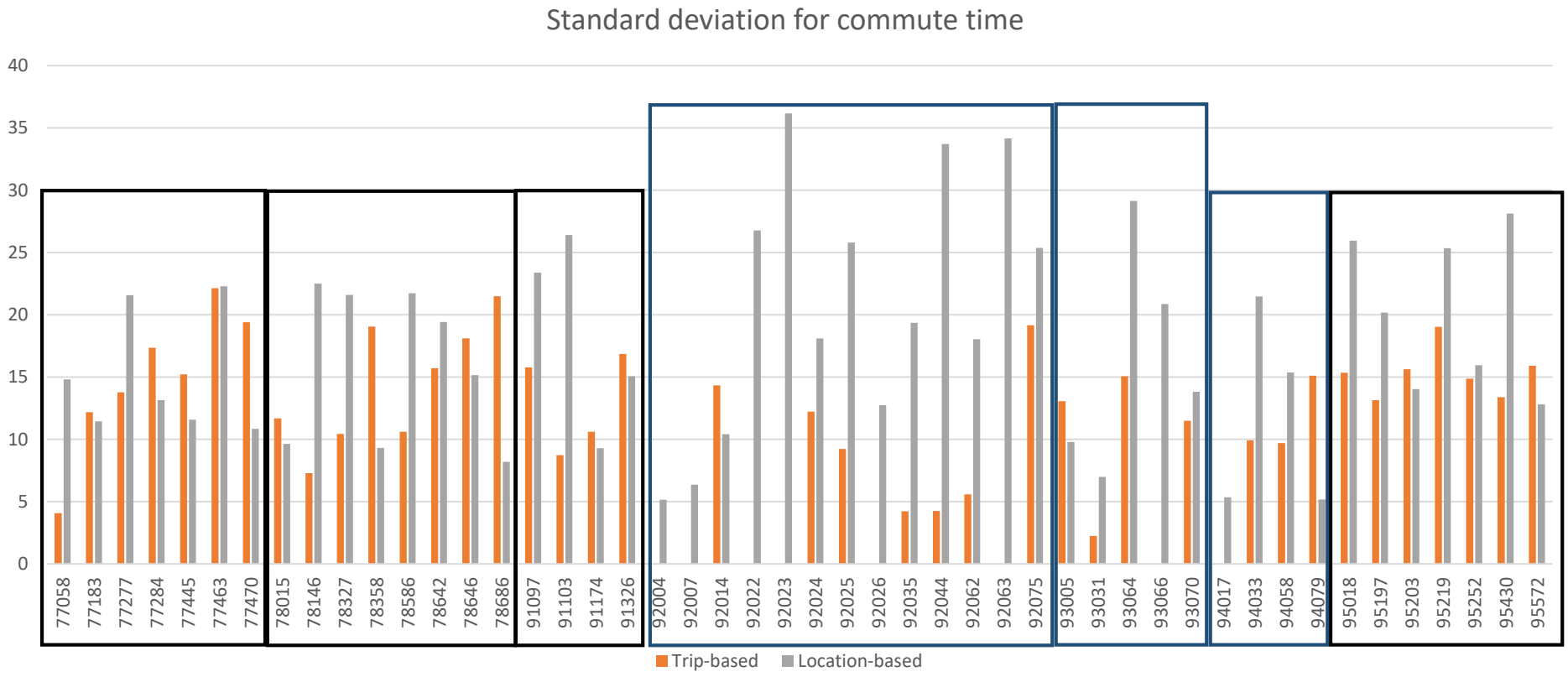
Forte correlation entre ASE en semaine et excursions



$p = 2^{E}\text{-}05$

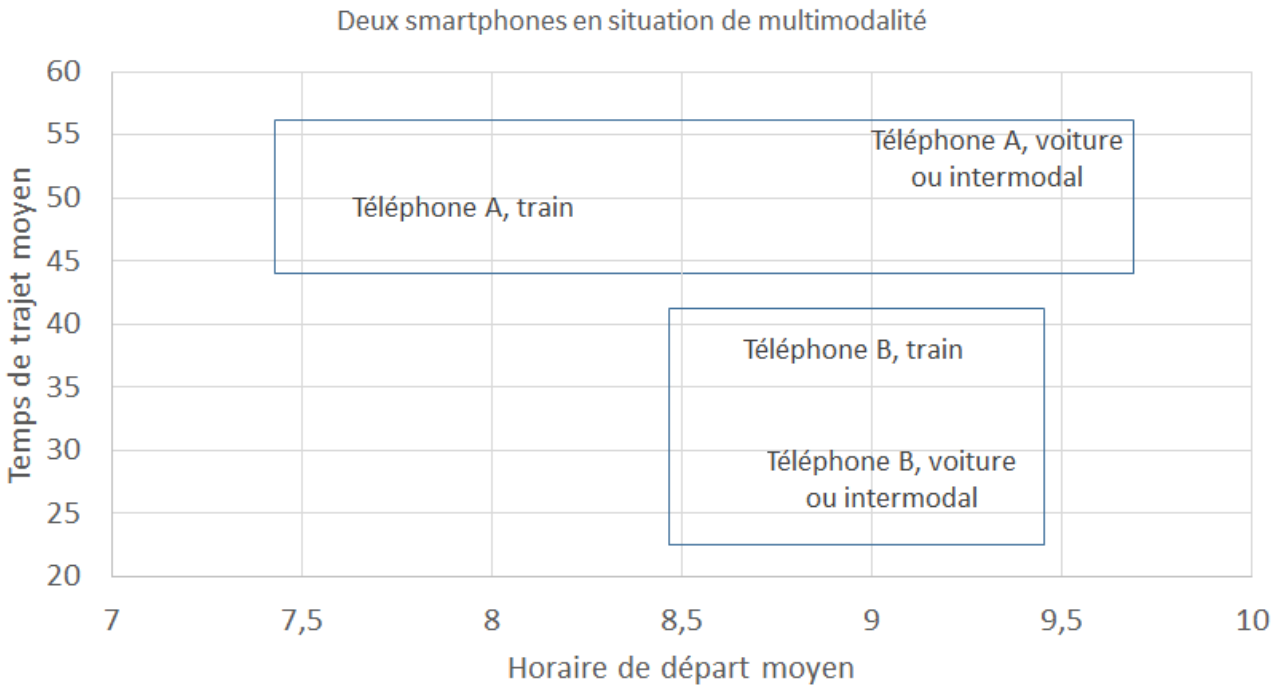
Variabilité des trajets domicile-travail

Trajets domicile-travail au niveau communal avec au moins 10 individus concernés



Focus sur des smartphones effectuant une même liaison : variabilité des trajets domicile-travail

- 7 smartphones effectuant une même liaison Outer Paris > Paris : CV du temps de trajet individuel d'environ [0,1 – 0,15] pour une heure de départ entre 7h et 8h ; d'environ [0,2 – 0,3] pour une heure de départ entre 8h et 9h
- Focus sur deux smartphones effectuant cette même liaison, en fonction du mode



Conclusion / discussion : des données qui sont…

- Sémantiquement « pauvres » (pas d'information sur les utilisateurs)
- Produites selon des processus non orientés vers l'analyse de la mobilité
  - Besoin d'inférer le mode de transport (certains points dur par exemple bus vs voiture?)
  - Besoin d'inférer lieu de résidence et lieu d'emploi, moins facilement les autres motifs d'activité
  - Inférer les caractéristiques individuelles est un enjeu actuel de la littérature scientifique
- … mais ont d'autres propriétés qui peuvent les rendre utiles pour l'analyse de la mobilité
  - Comportement longitudinal
  - Variation interjournalière des pratiques de mobilité
- Les données spécifiques utilisées pour cette recherche…
  - Ne peuvent pas être utilisées directement pour estimer des matrices OD
  - Reproduisent correctement certains indicateurs classiques pour certaines OD agrégées
  - Montrent un potentiel de cas de « massification » de ce type de données
  - Doivent et peuvent être traitées de manière compatibles avec le RGPD, du fait de leur potentiel de tracking personnel (De Montjoye et al., 2014)

Merci de votre attention

Julie Chrétien julie.chretien@enpc.fr
Florent Le Nechet florent.lenechet@u-pem.fr
Laurent Prouhlac laurent.proulhac@enpc.fr

Laboratoire Ville Mobilité Transport