MOTORIZED INDIVIDUAL MOBILITY IN COMMUTING TRIPS: MODAL PREFERENCE OR CONSTRAINED MODE CHOICE? A MACHINE LEARNING APPROACH

Rémy LE BOENNEC, VEDECOM Institute, Versailles, FRANCE
Fouad HADJ SELEM, VEDECOM Institute
Ghazaleh KHODABANDELOU, University of Paris-Est Créteil Val de Marne

OUTLINE OF THE PRESENTATION

1. Motivations
2. Research assumptions
3. Related literature
4. Data
5. Method
6. Perspectives

1. MOTIVATIONS

MOTIVATIONS (1/2)

- The **Ile-de-France region** in France is specific regarding **mobility behaviours**: the modal share of private car in commuting trips is 42% *(Source: French National Transport Survey 2008)*
- The French region is the most endowed in public transit (PT) with the perspective of the **Grand Paris Express** to optimize trips from suburb to suburb within 5 years *(regional railway network)*
- **Low-density areas** in the edge of the region could not see credible alternatives to private vehicle (PV)

MOTIVATIONS (2/2)

- **Where public transit is available,** private vehicle =
  - A **constrained** mode choice *(if PT=no competitive alternative)*
  - Or a modal **preference** *(<> psycho- or sociological factors)*
- A **transport policy** issue:
  - If modal **preference for PT** > the Grand Paris Express should be relevant
  - If modal **preference for PV** > incentives towards ridesharing

2. RESEARCH ASSUMPTIONS

RESEARCH ASSUMPTIONS (1/2)

- **The idea** = to compare travel times:
  - Using public transit (PT) or private car (PV)
  - Between living and working places in the Ile-de-France region
- **The research assumption**:
  - **Comparable travel times** to commute choosing PT or PV should be associated with comparable modal shares for both transport modes from origin to destination
  - AND a **higher travel time** using one of the transport modes should be associated with a higher modal share for the other transport mode

RESEARCH ASSUMPTIONS (2/2)

- **A regional science issue**: within the region, specific mobility behaviours should be revealed in:
  - (1) Paris
  - (2) the suburban areas and
  - (3) the periurban areas *(given heterogenous human density levels)*

- **An academic stake** (Barthelemy 2016): to assist urban economics practitioners' for the calibration of households' utility function

3. RELATED LITERATURE

RELATED LITERATURE (1/2)

- **Abundant literature** addressing mode choice (1/2):

  - The **mode** itself (from private vehicle to others): *public transit* (Kamruzzaman et al 2015, Shaaban and Kim 2016, Shen et al 2016), *active modes* (Lee et al 2015, Braun et al 2016, Cooper 2017, Ton et al 2018), *ride-and carsharing* (de Luca and Di Pace 2015, Lalou and Winter 2017, Bulteau et al 2019)

  - The **trip purpose**: *commuting* (Cao 2015, Heinen 2016, Franco 2017, Ko et al 2019), *accompanying children* (He and Giuiliano 2017, Liu et al 2018, Ferenchak et al 2019)

  - **Socioeconomic features** of the mobility users: *gender* (Abasahl et al 2018), *age* (Kamargianni et al 2015, *adolescents*), *employment status* (*students*: Shaaban and Kim 2016, Zhou 2016, Haggar et al 2019)

  - **Attitudes** as psycho-sociological factors: Lind et al 2015, Pike and Lubell 2016, Munoz et al 2016, Klinger 2017, Lanzini and Kahn 2017, Prato et al 2017, Vinayak et al 2018

  - **Mobility habits** (Klinger 2017, Bouscasse et al 2018) and **activities** conducted during the trip (Malokin et al 2015, Malokin et al 2019)

- **Abundant literature** addressing mode choice (2/2):
  - **Weather conditions**: Liu et al 2015, Anta et al 2016, Manoj and Verma 2016
  - **Land-use interactions and location choice issues**: Manoj and Verma 2016, Boulange et al 2017, Sun et al 2017, Helbich 2017, Bhat et al 2017, Ding et al 2018

- A **recurrent limitation**: available data is old (last national household travel survey in 2008 in France) and scattered (sociodemographic data, road counting data, GPS tracking…)

> A room for innovative methods to estimate human mobility flows

4. DATA

DATA (1/3)

- **Entropy** = a machine-learning tool that uses multisource data to merge it and estimate:
  - Mobility **flows**
  - Associated transport **modes**
  - Associated trip **purposes**
- The input **data** come from:
  - **Points of interest** (POI)
  - **GPS** tracking for public transit (PT) and private vehicles (PV)
  - **Ticketing** (PT)
  - Road **radars**
  - **Survey** data (origin-destination surveys, Household Travel Surveys (HTS), road data couting)
  - **Sociodemographic data** from the National Institute for Statistics and Economic Studies (INSEE)

> Mobility flows are estimated for origin-destination couples in the Ile-de-France region

- **Home-to-work** is the only trip purpose considered (30% of the total trip purposes *(Source : French National Transport Survey 2008)*

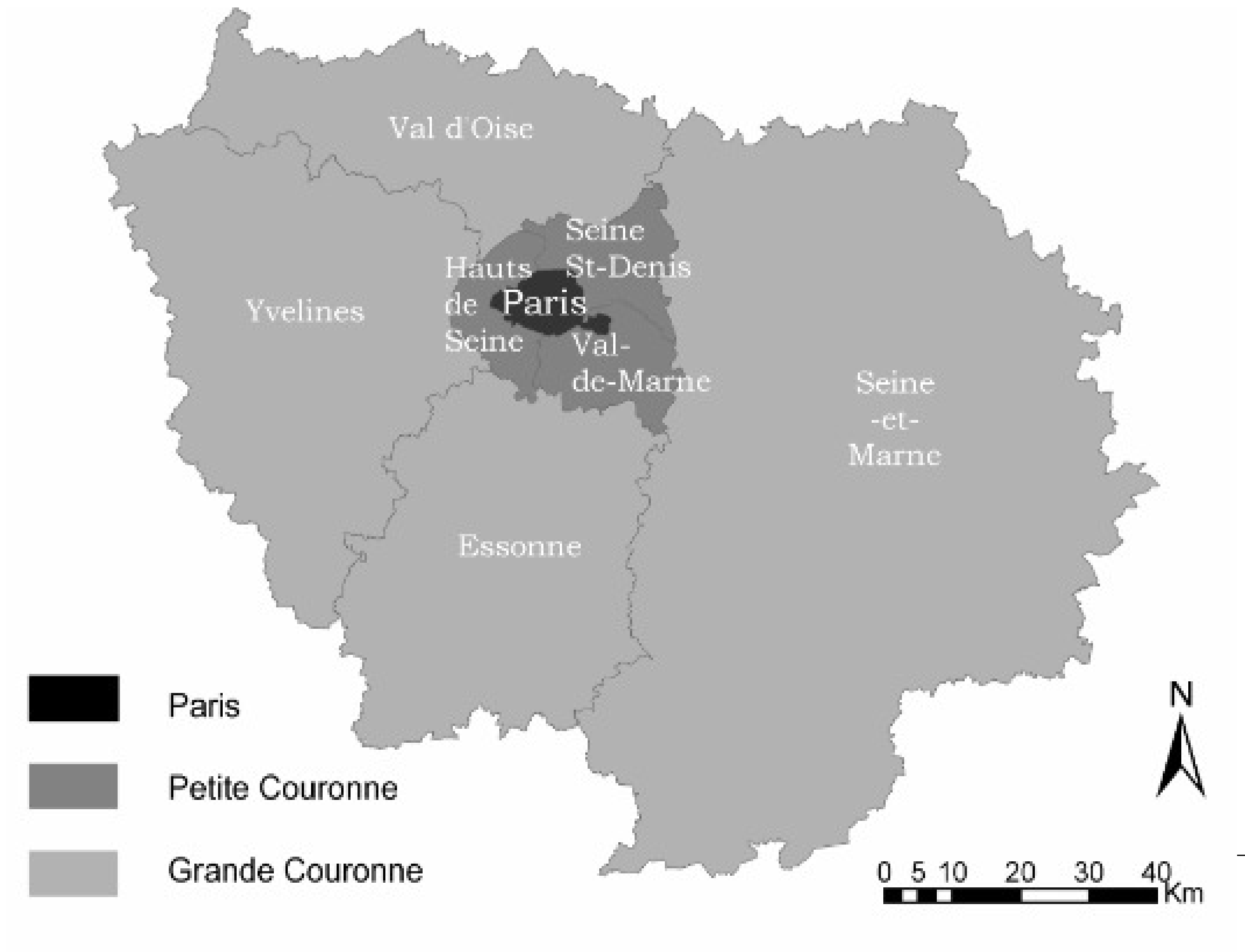
- **Travel time data** comes from:

  - Either **theoretical data** > Considers a speed limit of the vehicles in every road section (PV), without slow-down or congestion

  - Either (declared) **survey data (HTS)** > Realistic travel times, acurate distances (does not consider the centroid of the municipality like in the previous method)

DATA (3/3)

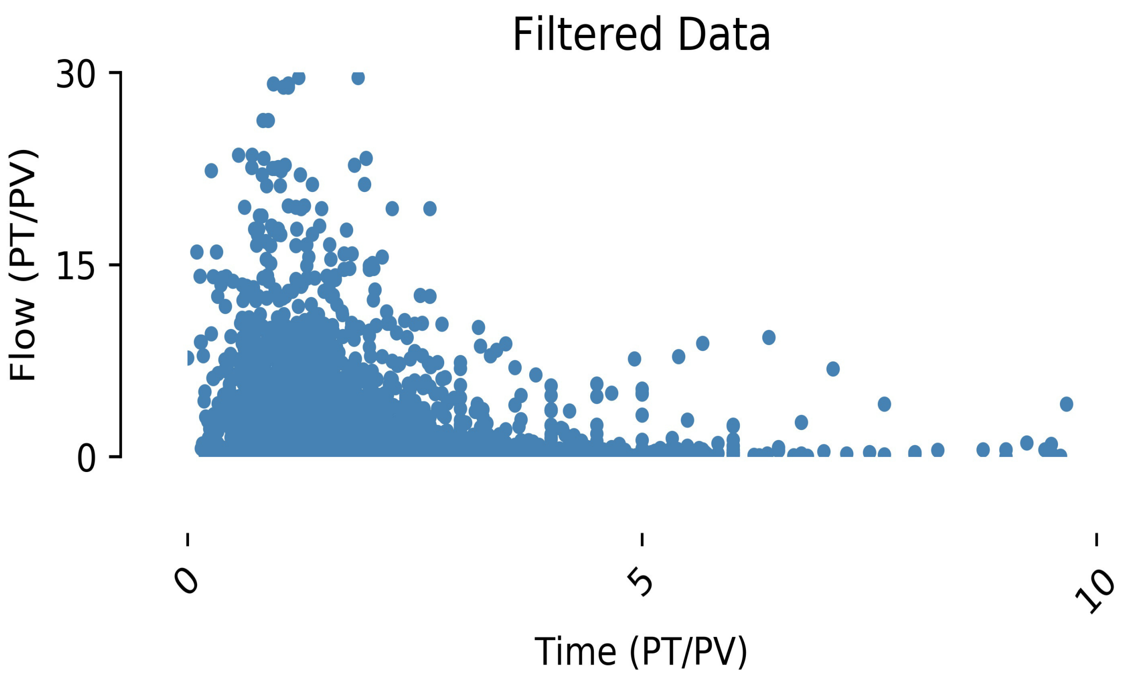
The Ile-de-France region : (1) Paris inner city *(« Paris »)*, (2) Close suburbs *(« Petite couronne »)* and (3) The rest of the metropolitan area *(« Grande couronne »)*(*Source : Halbert 2008*)



5. METHOD

METHOD (1/4)

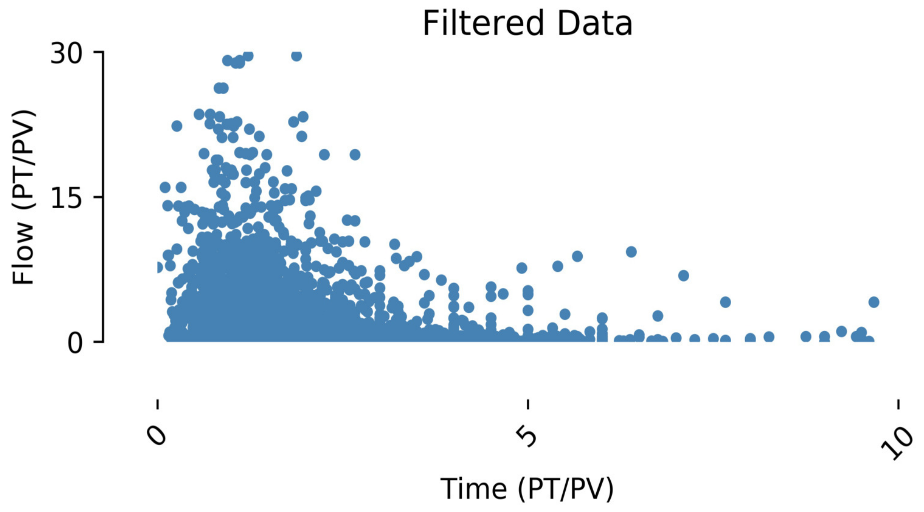
- **Research strategy=clustering:** to compare clusters of origin-destination couples obtained by the *K-means* method and a *Gaussian Mixture Model* (GMM)

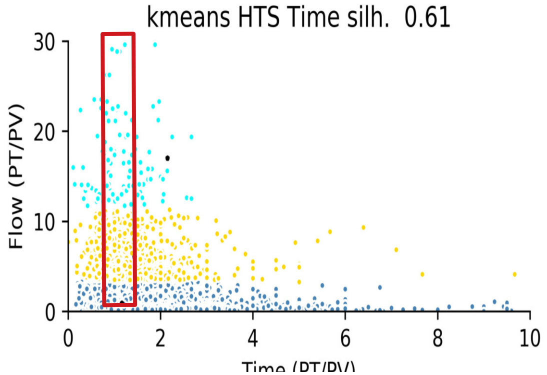


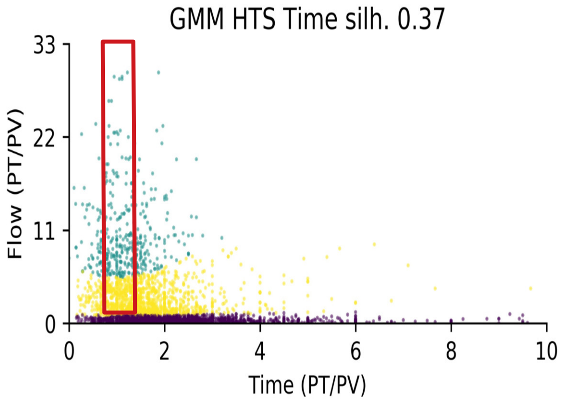
METHOD (2/4)

Both methods are applied to estimate associated **mobility flows.** Coming back to our research assumption we search to reveal, in a **time band** for travel time ratios around 1:

- Either a « *modal indifference* » (if $Time_{PT}$ # $Time_{PV}$ and $Flow_{PT}$ # $Flow_{PV}$)
- Either a *modal preference for PV* (if $Time_{PT}$ # $Time_{PV}$ but $Flow_{PT}$ << $Flow_{PV}$)
- Either a *modal preference for PT* (if $Time_{PT}$ # $Time_{PV}$ but $Flow_{PT}$ >> $Flow_{PV}$)



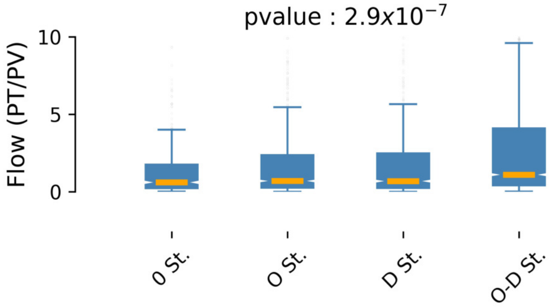




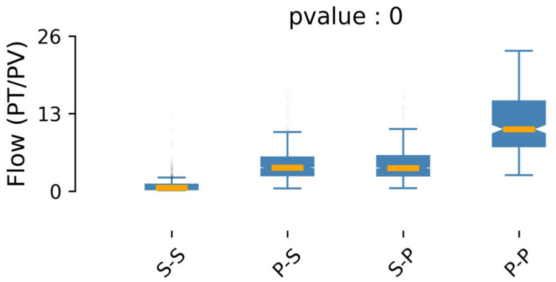
METHOD (3/4)

- Need to add **covariates** in our dataset to better qualify the relationship between time ratios and relative flows *(based on mode choice literature)*

- **Analysis of Variance (ANOVA)** applied to the first two covariates :
  - Presence of a railway station in the municipality
  - Type of area (Paris/ Metropolitan area)



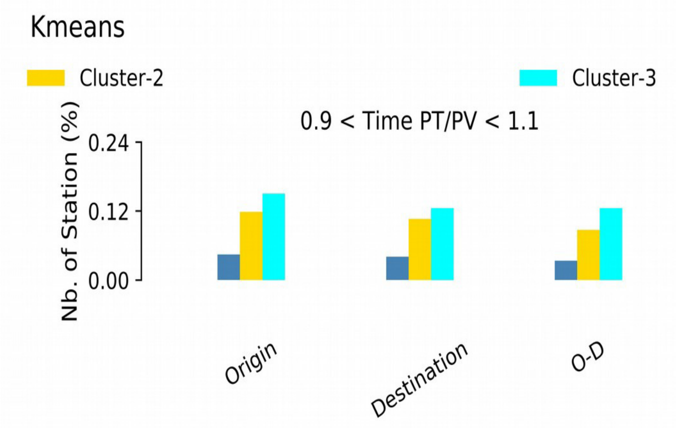
	Median	Mean	Std. Deviation
0 Station	0.60	1.84	2.94
O Station	0.69	2.23	3.86
D Station	0.68	2.25	3.90
OD Station	1.1	2.97	4.01

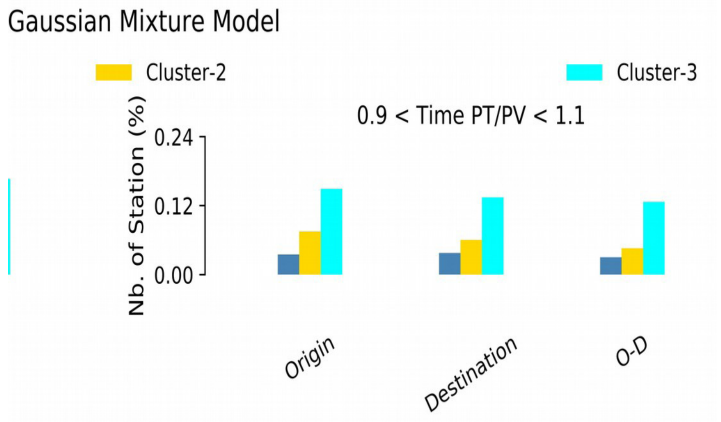


	Median	Mean	Std. Deviation
Subs.-Subs.	0.57	0.86	0.95
Paris-Subs.	3.94	4.64	2.97
Subs.-Paris	3.89	4.68	2.99
Paris-Paris	10.41	11.98	6.01

METHOD (4/4)

To describe Cluster *i*, *i=1,2,3*, we consider the relationship between the number of railway stations and the relative flows *(if comparable travel times using PT or PV)*





6. PERSPECTIVES

PERSPECTIVES (1/2)

- **Work in progress**: incorporate additional covariates to better qualify the modal preference for one or the other transport mode in a linear regression model, using:
  - **Geographical** data: presence of a car-park in the station in origin/ in destination/ both; Number of parking spaces
  - **Sociodemographic data**: age, gender, motorization rate, occupation *(R-squared>0.4)*

> Associate specific values of the covariates to Cluster *i, i=1,2,3* > What values associated to modal preference for PV or for PT?

PERSPECTIVES (2/2)

- Better use our **typology of areas** > finer results within the Ile-de-France region, distinguishing between:
  - The inner city of **Paris**
  - Close **suburbs**
  - The rest of the **Metropolitan area**

- An original approach, merging and estimating complete data from incomplete multisource data

> The method could be replicated to areas where the modal share of PT is lower than in the Ile-de-France region (if available) to identify specific brakes and incentives for modal shift (French metropolises)

Thank you!

Together to accelerate the mobilities of tomorrow

remy.leboennec@vedecom.fr