Colloque de la chaire Mobilité Polytechnique Montréal
19-20 mai 2022

How comparable is travel demand estimated from automatic fare collection, large scale origin-destination survey and household travel survey? An empirical investigation in Lyon

Patrick Bonnel, Oscar EGU
Université de Lyon, ENTPE, LAET, Lyon

UNIVERSITÉ DE LYON

ENTPE
L'école de l'aménagement durable des territoires

LAET
LABORATOIRE AMÉNAGEMENT ÉCONOMIE TRANSPORTS
TRANSPORT URBAN PLANNING ECONOMICS LABORATORY

Keolis

Research objectives

- Traditional travel surveys offer rich semantic data, but only one or few travel day every 5-10 years with limited sample size
- Origin-destination surveys offer high sample size, but for a single day every 4-5 years in Lyon with limited semantic data
- Smart card data offer continuously high volume of data but with poor semantic

What is the comparability of these data sources if we want to combine them to enrich the data

Smart card data for Lyon conurbation

- Lyon conurbation (1.3 million inhabitants) transit network transaction only at vehicle boarding (including transfer)
  - In average 1.5 million trip-legs a day
  - Smart card (80% of validation, same Id over a long period)
  - Magnetic paper ticket (20% of validation, without Id)
- AVL (Automatic vehicle location)
- Automated passenger counting system (bus, tramway, subway)

Origin-Destination survey for Lyon conurbation

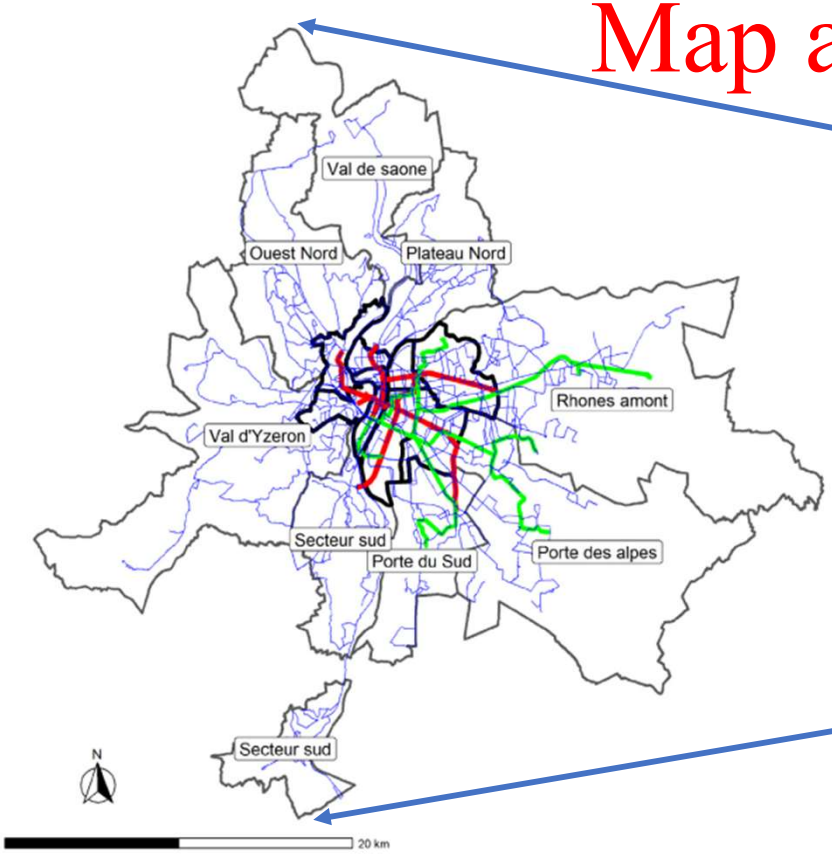
- Origin-destination surveys are performed on a public transport route basis all along the year
- All routes are surveyed at least once every 5 years
- Bus routes (about 100) all individuals on all services during a day (no sampling)
- Subway (4 lines), Tramway (5 lines) random sampling of about 25-35% of individuals during a day
- Limited semantic (O and D of trip-leg and of trip at stop level; connection before/after; fare type; purpose; few demographics)

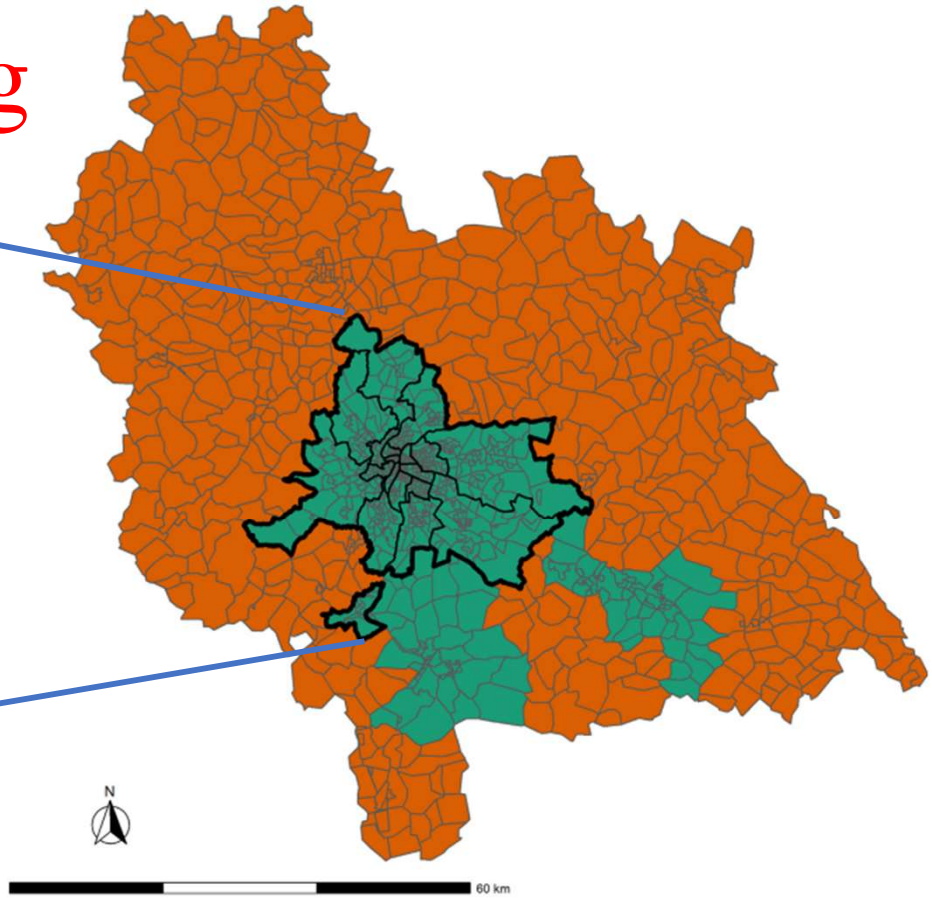
All O-D surveys of a 5-year period (2013/17) are used to build an origin-destination matrix of public transport trips with stop level zoning

Household travel survey for Lyon conurbation

- Every 10 years, nearly 1% stratified sampling, face-to-face + telephone
- Last survey in 2015 according to CEREMA standard of about 16.000 households/28.000 individuals
- Zoning system for stratification (169 zones) and for O-D coding (1290 zones)
- About 100.000 surveyed trips (HTS survey area) but only 10.570 in public transport for the Greater Lyon area (TCL)
- Very rich semantic with socio-demographic of household and individuals; car availability + car characteristics; detailed trip/trip-leg characteristics

Map and zoning





Metro lines | Tramway lines | Bus lines | Districts of Lyon and Villeurbanne

Study Area | Face-to-face | Phone

Study area (Greater Lyon PT area)

Household travel survey area

Smart card data processing

Data correction and imputation

- Missing data imputation + deduplication
- Transfer identification to transform trip-legs into trips (rules from literature)
- Destination inference rules only for smart card data (same Id): 80.8% success (trip chaining method)
- Fraud (or non-validation) represents 21% of total transit trips

Smart card data expansion

- Transit trips with alighting location: ≈ 50% of total transit trips
- Automated passenger counting system is the base for expansion
- Expansion with non uniform scaling factors because fraud and non-validation are non uniform; id for alighting imputation
- Definition of control node: bus or tramway route + subway station passenger counting (155 control nodes)
- Definition of expansion factors at itinerary (same O-D + same transfer) level

$$\sum_{i \in I} B_{ni} \alpha_i t_i = \Delta_n \quad \forall n \in N$$

I set of itineraries (53.000) ; n control node

B_{ni} = 1 if node n belong to itinerary i

t_i flow on itinerary i

α_i expansion factor

Smart card data "ground truth" validation

	Smart card data route uniform expansion factors	Smart card data itinerary expansion factors	O-D survey	Household travel survey (HTS)
Trip legs (million)	1.55	1.56	1.51	1.11
Trips (million)	1.11	1.10	1.16	0.80
Bus trip legs (%)	34	41	39	43
Tramway trip legs (%)	20	23	22	21
Subway trip legs (%)	46	37	39	36

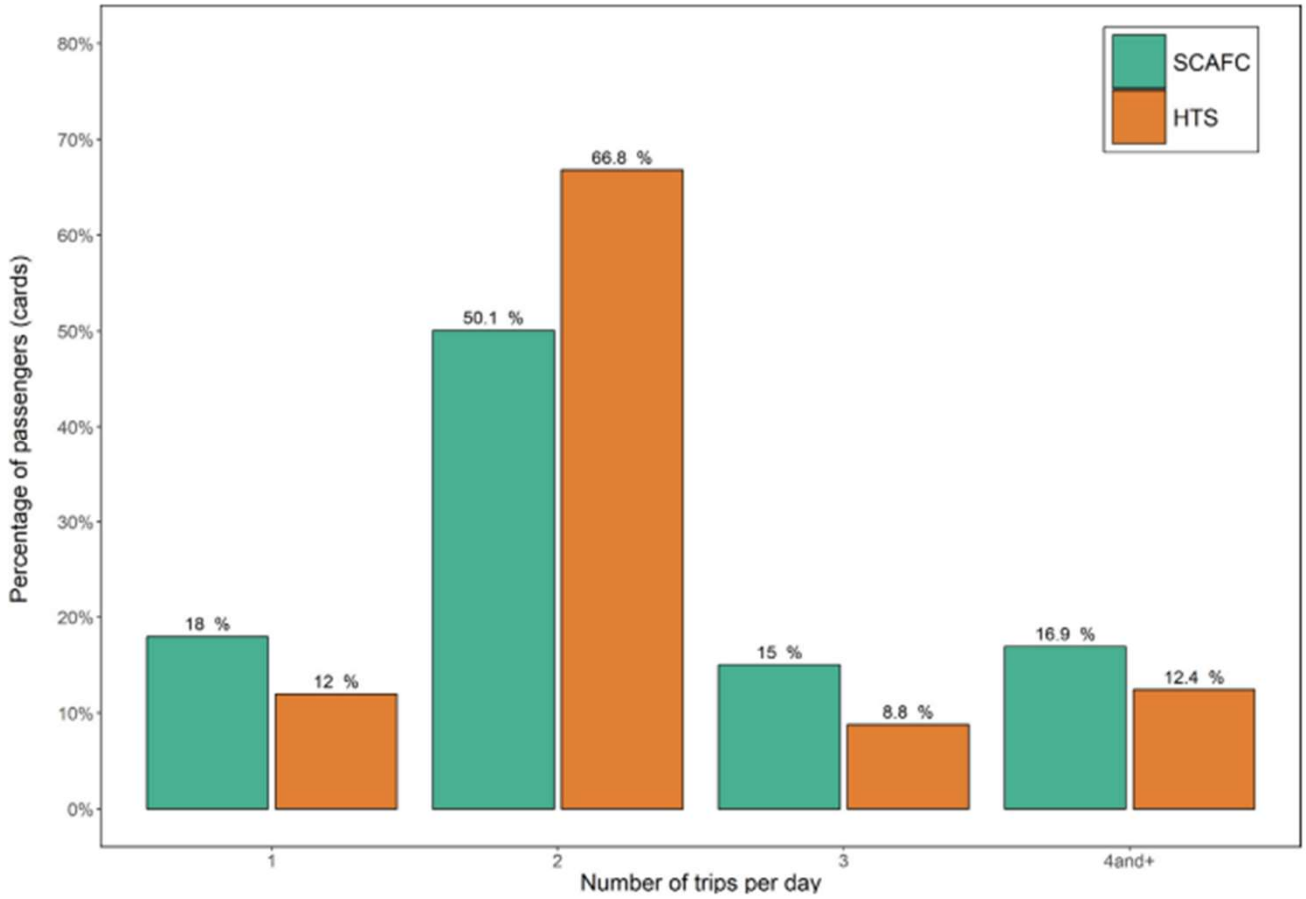
- Much less trip-legs and trips in household survey compared to smart card data and O-D survey which appears much more coherent
- Fine spatial expansion factors using itinerary increase data quality vs route uniform expansion factors

Distribution of trips among individuals (cards)

Less single trip or 3 trips+ with household travel survey than smart card

Half individuals made 2 trips a day from smart card, but 2/3 for HTS

Peak periods are stronger for HTS than other data sources

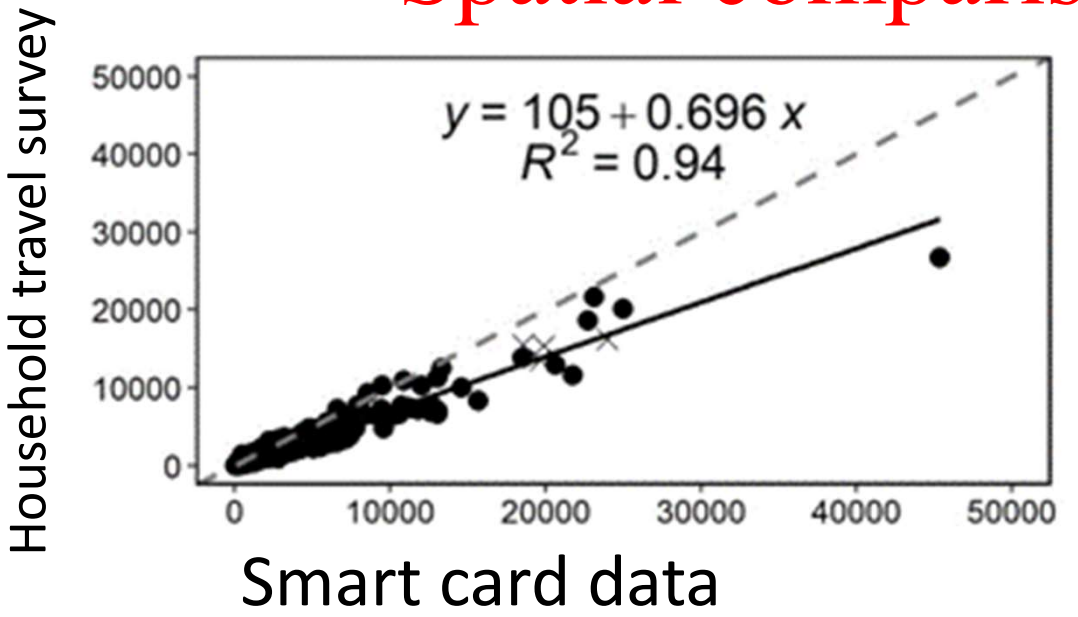


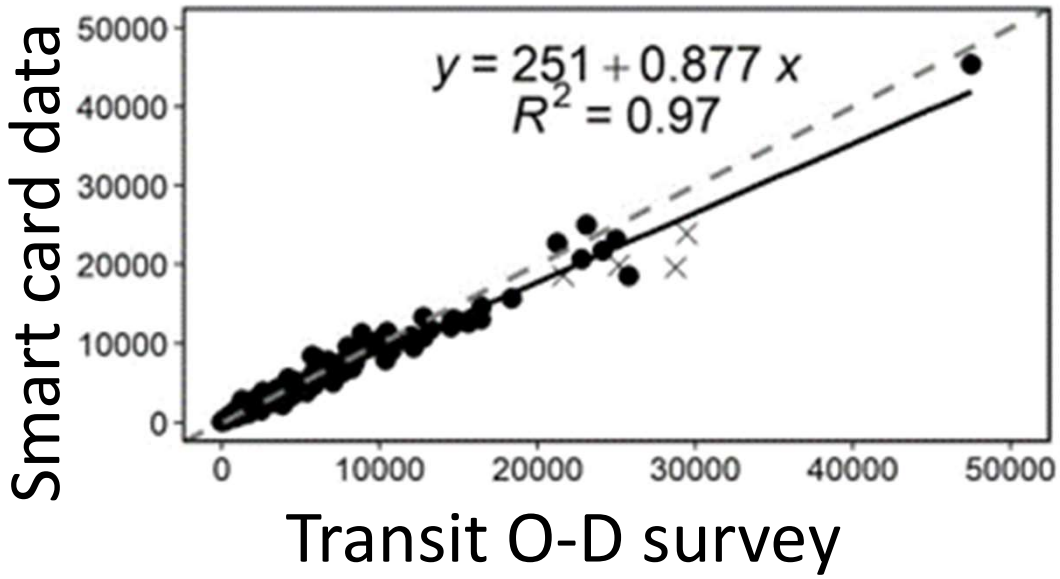
Macro-spatial distribution of trips

Origin	Destination	Smart card	O-D survey	HTS
Central area	Central area	58	61	57
Central area	Peripheral ring	14	13	15
Peripheral ring	Central area	15	12	14
Peripheral ring	Peripheral ring	13	14	15

Similar spatial distribution at macro level

Spatial comparison with 18 zones





- Spatial comparison at O-D level (18 zones)
- Smart card data are much more coherent with transit O-D survey, than with household travel survey (21% of error for the comparison smart card – OD survey and 40% for OD survey – HTS ($\text{SMAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{\frac{1}{2}|y_{1,i} - y_{2,i}|}{(y_{1,i} + y_{2,i})}$)

Synthesis-1-

Smart card data represent a high potential for public transport analysis but

- Destinations need imputation methods
- Fraud, non validation, magnetic ticket and trip without destination imputation might represent half public transport trips
- O-D matrices build from smart card data need expansion factors
- Counting data (like Automated passenger counting system) allow non uniform expansion factors
- Non uniform expansion factors based on itineraries improve O-D matrix quality

Synthesis-2-

- If HTS is necessary for rich semantic, public transport O-D matrices are under-estimated

- Smart card data allows to build dynamic O-D matrices and to update O-D matrices over time