



Combining usual roadside survey and cellular phone data to produce accurate transit and exchange OD matrices - Application to the cross-border Geneva 4-step model

Pierre-Antoine Laharotte, Wilfried Raballand

► To cite this version:

Pierre-Antoine Laharotte, Wilfried Raballand. Combining usual roadside survey and cellular phone data to produce accurate transit and exchange OD matrices - Application to the cross-border Geneva 4-step model. 47th European Transport Conference, Oct 2019, Dublin, Ireland. hal-02317382

HAL Id: hal-02317382

<https://hal.science/hal-02317382>

Submitted on 16 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



COMBINING USUAL ROADSIDE SURVEY AND CELLULAR PHONE DATA TO PRODUCE ACCURATE TRANSIT AND EXCHANGE OD MATRICES – APPLICATION TO THE CROSS-BORDER GENEVA 4-STEP MODEL

LAHAROTTE Pierre-Antoine,
RABALLAND Wilfried,
CEREMA Centre-Est, Bron, France

1. INTRODUCTION

Knowing the mobility of people is a major challenge to fuel traffic or travel models in order to assess town policies. This knowledge of mobility involves traffic surveys to collect the origins and destinations of users. Owing to the cost of surveys and the emergence of electronic toll collection, it is no longer possible to carry out field surveys on French motorways.

Faced with these difficulties, the use of data from on-board technologies is a good way to explore mobility, since it can be used to reconstruct traces of vehicles. They have been assessed in order to estimate the demand for travel on the network [McNeill et al (2017), Bonnel et al (2015)] or to better understand local dynamics [Wang et al (2017), McNeill et al (2017), Nabavi et al (2015), Caceres et al (2008)]. Usually, in the field of transport, digital data is shared between:

- data from on-board geolocation systems (Floating Car Data - FCD),
- data from mobile telephony (Floating Mobile Data - FMD) based on a network of antennas (Edge, 2G, 3G, 4G, Bluetooth, WIFI, etc.).

FCD data turns out to be very efficient in reconstructing traffic dynamics [Herring et al (2010), Vanajakshi et al (2009)], especially the speed of the vehicles. Nevertheless, the volumes collected by OD (Origin-Destination) from these tracks require significant corrections aiming to offset the focus made on specific population strata. FMD data do not suffer from such a bias in representativeness and can be used to reconstruct OD matrices from antenna to antenna [Caceres et al (2008), Michau et al (2014)]. However, the projection of these OD flows onto specific zoning and fine reproduction of secondary flows still arise some difficulties [Caceres et al (2008), Raballand (2018)]. The ability of these new data sources to be used instead of conventional roadside surveys requires further methodological refinement before they can be implemented fully.

In order to guarantee operational implementation, we introduce a hybrid approach, combining usual surveys and digital collections provided in the form of OD matrices. The aim is to build an OD matrix of transit or exchange trips on the perimeter of Geneva's travel model.

2. METHODOLOGY

2.1 General principle of the introduced protocol

The proposed protocol aims at carrying out a traffic survey picking up the transit and exchange flows over a wide targeted area: the Geneva model area. The usual protocol consists of positioning survey stations on each injection point over the perimeter of the targeted area, i.e. any of the road infrastructures allowing the access to the area. Each conventional survey station consists of roadside investigators interviewing users, and automatic vehicle counting to correct the survey bias.

We propose to set up conventional survey stations as far as possible, i.e. mainly on national and local roads. Then we deal with the motorway cases by setting up digital stations (see Fig. 1). Digital stations are synthetic stations, reconstructed from mobile phone data collections and rectified from automatic counts and number plate readings (NPR).

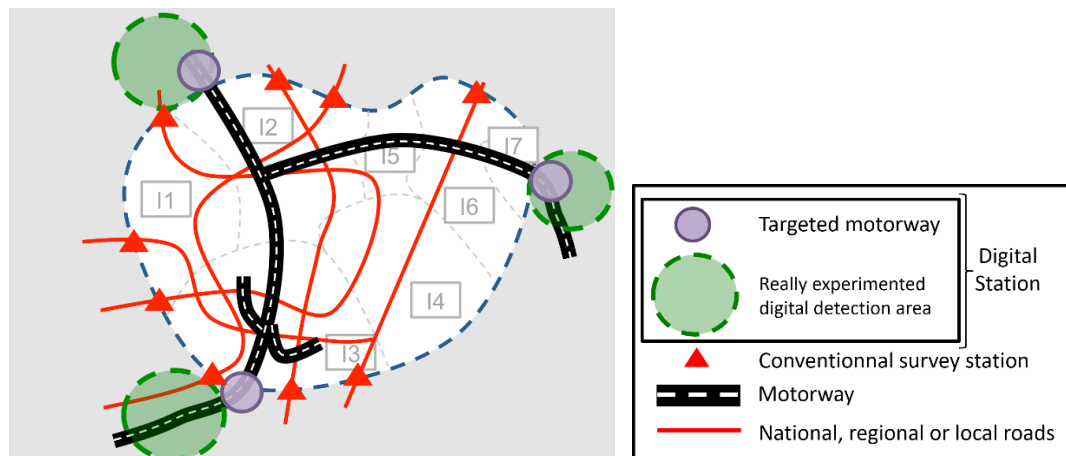


Figure 1: Protocol layout

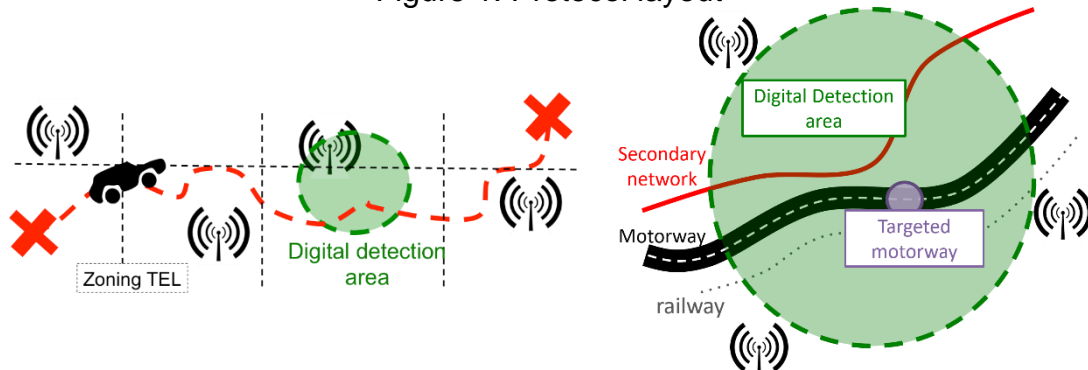


Figure 2: mobile phone detection process with digital survey

Figure 3: Digital survey detection bias

2.2 Specific features of data collected from mobile telephony

The collection process of mobile telephony data requires to define:

- a *digital detection area* to select all the trips to store within the digital OD matrix (see Fig. 2, green area);

- a *slack zoning*, known as *TEL zoning*, guaranteeing a good fit between how the area is divided up and the positioning of the mobile telephone antennas;
- a *minimum period of immobility* to characterize the beginning and the end of a trip, in order to set its origin and destination zone.

Such process enables to track mobile phone users and retrieve OD matrices for any timestamp on a specific zoning. Nevertheless, the antenna networks were not designed for this purpose, some bias are therefore introduced:

- the digital detection zone does not fit only the targeted motorway area: it equally picks up motorway users as users with other modes (trains) or on nearby roads (secondary network), as represented in Fig. 3;
- *TEL zoning* is dependent on the layout of the antenna network and does not coincide with the administrative zoning, which requires defining flow distribution rules to move from one zone to the other.

The final reconstitution of a digital survey station requires the specific features of these digital data collections to be integrated.

2.3 Theoretical concept to achieve a hybrid road survey

Assuming the OD matrices resulting from conventional roadside survey as the reference outputs, some adjustments are required on the digital OD matrices. It is indeed necessary to tackle the issue of:

- mixed mode and mixed road category within the data resulting from digital surveys;
- zoning differences between digital surveys (*TEL zoning*) and conventional roadside survey (administrative zoning).

2.3.1 Mode and road category discrimination process applied to OD flows resulting from digital stations

The targeted motorway OD matrix $M_{digit-veh}^{motorway}(od)$ results of the subtraction from the raw OD matrix $M_{digit-pers}(od)$ of:

- a matrix composed of OD flows observed on the surrounding of the motorway, but driven with alternative modes (ex: train): $M_{digit-pers}^{other mode}(od)$;
- a matrix composed of OD flows observed on the surrounding roads of the motorway: $M_{digit-veh}^{other roads}(od)$. These flows need to be removed as they are iterated within the global study: they appear as well within the conventional survey as within the digital survey.

a / Assessment of the OD matrix associated to alternative modes

It is easier to collect available road counts about the roads passing through the targeted area than to collect exhaustive and accurate data about alternative modes. It is assumed that the alternative modes matrix $M_{digit-pers}^{other mode}(od)$ results from the difference between the raw matrix $M_{digit-pers}(od)$ and an assessment of the road matrix $M_{digit-pers}^{road}(od)$, as related in Eq.1. The resulting matrix

$M_{digit-pers}^{other mode}(od)$ is punctually compared to available alternative modes data. The assessment of the road matrix $M_{digit-pers}^{road}(od)$, expressed in persons, results from a Fratar type algorithm [Fratar (1954)] applied to the raw matrix $M_{digit-pers}$. The margins are assigned with the combination of roadside counts surrounding the digital detection areas. This relationship is described by Eq. 2, where $\lambda_{occupancy}$ is the average vehicle occupancy rate and $\sigma(od)$ is an adjustment coefficient. Thanks to the road count, the unit is switched from person to vehicles.

$$(1) \quad M_{digit-pers}^{other mode}(od) = M_{digit-pers}(od) - M_{digit-pers}^{road}(od)$$

$$(2) \quad M_{digit-pers}^{road}(od) = \lambda_{occupancy} M_{digit-veh}^{road}(od) = \sigma(od) M_{digit-pers}(od)$$

b / Identification of the iterated flows within the road matrix

The resulting road matrix $M_{digit-veh}^{road}$, expressed in vehicles, is composed of the motorway flows ($M_{digit-veh}^{motorway}$) and parasite flows ($M_{digit-veh}^{other roads}$), assumed to be already stored within conventional roadside surveys (T_{survey}^k), as related in Eq. 3. The relationship between the iterated flows and flows resulting from the roadside surveys is described in Eq. 4 and Figure 4, where $\alpha_k \times T_{survey}^k$ represent the proportion of flows collected both by the conventional roadside survey station k and a digital station and C is a change of basis matrix to convert the administrative zoning into a TEL zoning.

$$(3) \quad M_{digit-veh}^{road}(od) = M_{digit-veh}^{motorway}(od) + M_{digit-veh}^{other roads}(od)$$

$$(4) \quad M_{digit-veh}^{other roads}(od) = C \left(\sum_k \alpha_k T_{survey}^k(o'd') \right)$$

In practice the C and α_k contributions are directly assessed through the implementation of a simulator. The simulator aims to reproduce the detection process of the digital stations. It applies this detection process to the OD trips from the conventional roadside surveys, previously assigned to the road network. Finally, the simulator is composed of:

- a detection area associated to any digital survey and depicted as a circular area defined by its radius R ;
- a road assignment process enabling to make assumptions about the possible route for each OD relationship from the conventional roadside surveys.

The matrix $M_{digit-veh}^{other roads}$ results from the intersection between the simulated detection area and the possible routes. The optimum radius R is a parameter to adjust thanks to digital survey results and further available data such as number plate readings.

c / Conversion of the motorway matrix from TEL zoning to administrative zoning

The motorway OD matrix $M_{digit-veh}^{motorway}$ is built on the TEL zoning. In order to ensure the communication with OD matrices resulting from conventional surveys, it must be allocated on the administrative zoning. It is equivalent to apply an inverse basis change C^{-1} , described in concrete terms in paragraph 2.3.2.

2.3.2 Principle for converting TEL zoning to administrative zoning

The inconsistency between the coverage topology of the mobile phone network and the administrative division of the area may lead to certain flow distribution biases, especially when a high-density zone is divided into several parts. The transition between the mobile phone collection zoning and the fine zoning of the model is achieved through a breakdown method based on

- explanatory data on the generation (population) and attractiveness (employments) specific to the fine zones;
- significant divergences of division affecting highly urbanized and/or adjacent areas.

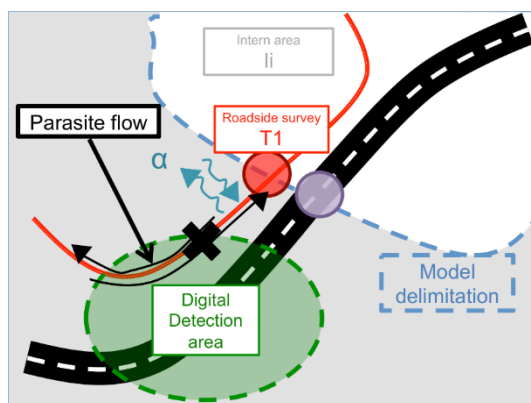


Figure 4: digital detection area outside the model delimitation

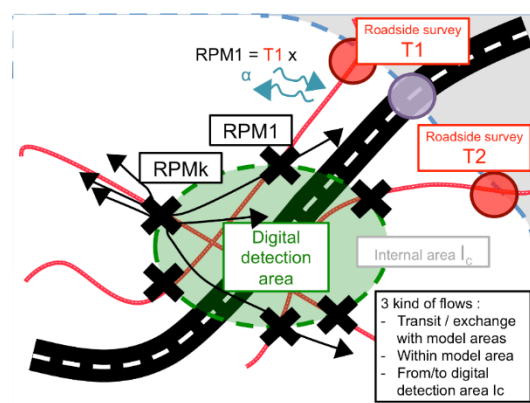


Figure 5: digital detection area inside the model delimitation

2.4 Treatment of additional imperatives

2.4.1 Non-optimal location of the digital stations

As tracking the mobile phone user is not the main purpose of the phone network antennas, the location and orientation of the antennas could not match with the optimal location of the digital survey stations. Some digital stations could have to be located within the perimeter of the study. It involves that some internal flows are detected by the digital station. Most of these flows are properly assigned to internal flows within the digital matrix. As the study only focuses the external flows, they are removed. Nevertheless, to ensure the results and improve the simulator calibration, NPR have been set up on the perimeter of the concerned detection areas.

2.4.2 Partial coverage of the area by the operator's antenna network

The method set out in 2.3 assumes the continuity of coverage of the mobile phone network, but mobile phone operators can only track their users in their respective countries. Carrying out a crossing-border survey, *i.e.* on a perimeter covering 2 countries, requires performing 2 joint, simultaneous surveys, and therefore using data from two mobile phone operators. As the Geneva model is a crossing border model, a French operator (Bouygues Telecom) and a Swiss operator (Swisscom) have been asked to carry out a digital survey within their respective territories.

To ensure the continuity of the crossing-border flows within the data, digital stations have been set up at the border by both phone operators. A former border survey is available to check the consistency of the results. Any digital station at the border can pick up movements between the two countries. The connection between the 2 digital surveys can be made, after adjustment of the digital OD matrices, by:

- applying a probabilistic assignment according to a gravity model,
- or using the distribution from cross-border travels surveys.

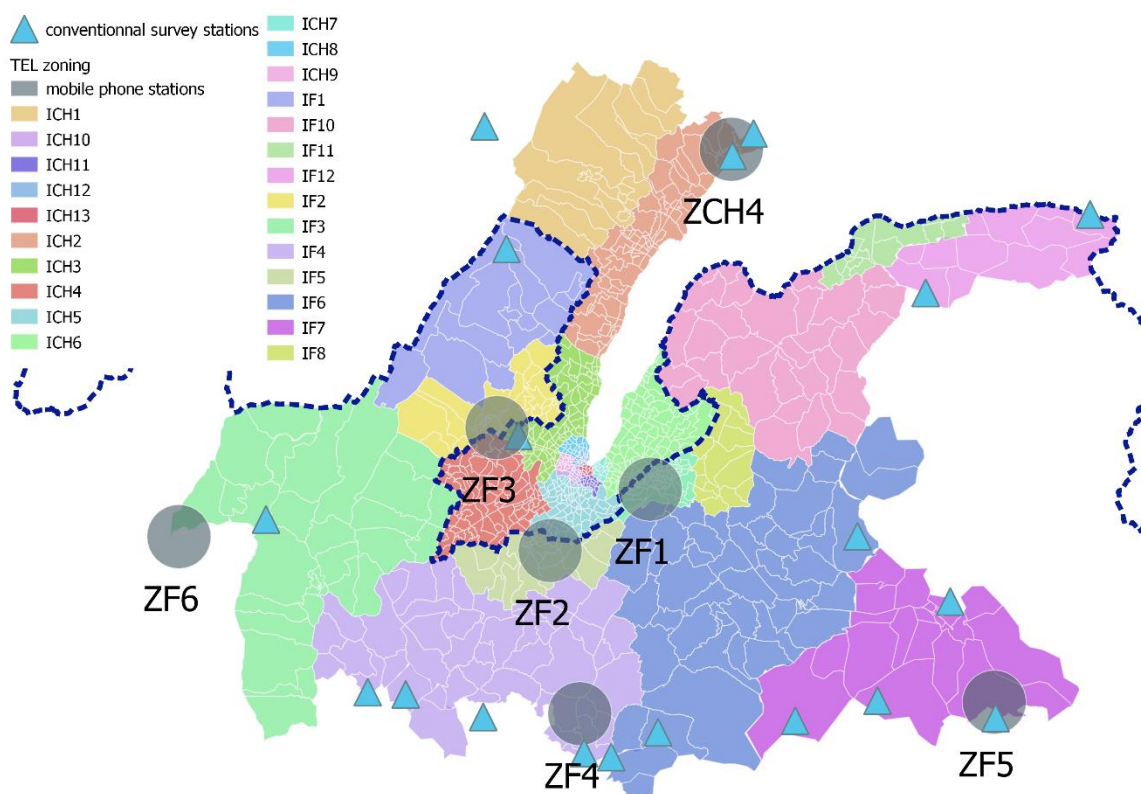


Figure 6: Implementation of the full survey on the case of study of Geneva

3. OPERATIONAL IMPLEMENTATION AND RESULTS

The methodology introduced in section 2 has been implemented on the Geneva's model perimeter to retrieve the exchange and transit flows driving through the area. It requires to set up:

- a conventional roadside survey composed of 19 stations;
- a digital survey composed of 6 stations (3 stations at the border, 3 stations on the perimeter) on the French side;
- a digital survey composed of 4 stations (3 stations at the border, 1 station on the perimeter) on the Swiss side.

3.1 Analysis of the raw digital surveys 'outputs

The preliminary analysis of the digital raw OD matrices $M_{digit-pers}$ provided by the phone operators highlights 2 sources of inconsistency or bias:

- the flows of crossing-border peoples and its distributions provided by the digital surveys do not match the automatic counts, nor the trends given by the former conventional survey at the border, assumed to be a reference (see Table 1);
- the spatio-temporal instability of the TEL zoning involves some mistakes in the distribution of the flows over the zoning (see Table 2). Especially 2 kind of misattribution are identified:
 - (a) in high-density areas, it is expected to get a tightened grid, but the current algorithms of the phone operators do not enable to reach this level of accuracy for any area: some generated trips are inverted between 2 close areas, as illustrated in table 2 (a) for areas IF6 & IF7 during Morning Peak Period;
 - (b) because of the presence of local roads with high traffic at the border of some TEL zones, the volume of generated trips could be overestimated, as shown in table 2 (b) on zones ICH1 & ICH2.

Both issues are tackled by altering the raw OD matrices according to automatic traffic counts and socioeconomic data (population, job etc ...) from INSEE.

Station at the border	French digital survey	Swiss digital survey	Crossing-border counts -reference
Station #1	7,150	4,200	10,200
Station #2	6,000	4,900	14,750
Station #3	700	3,050	4,900

Table 1: crossing-border flows of people according to the kind of survey - Morning Peak Period

Case	Area ID	Number of generated trips (exchange & transit) given by the digital survey	Population of the administrative area	Trips per inhabitant ratio
(a)	IF6	4,850	50,700	0.09

MPP	IF7	7,450	25,100	0.3
(b)	ICH1	1,550	17,600	0.09
Day	ICH2	2,400	81,800	0.03

Table 2: Impact of the instability of TEL zoning

3.2 Isolation of the motorway flows

French station (as origin)	Total flow (raw data)	flows assigned to other roads	Flows assigned to railway
Station #4	13 757	1 864	NA
Station #5	4 058	1 086	NA
Station #6	3 090	270	52

Table 3: Impact of the process of mode / road type discrimination – number of people - Morning Peak Period (MPP)

The application of the mode and road type discrimination process to the current case study enables to split the original flow provided by phone operators into 3 clusters : the alternative mode (ex: railway) flow, the motorway flow and the flow driven on local roads (see Table 3).

As, in our case study, the alternative mode flow is assumed to mainly result from trains, this flow has been compared, on the proper areas, to a rough estimation of users on the TGV line LYRIA (Paris-Geneva) for peak hours. It results in a difference of MAPE=55% in the direction of Geneva and MAPE=19% in the reverse direction. The gap may be explained by the rough assessment and definition of the alternative mode.

The discrimination according to the driven road type is made thanks to a simulation of the digital detection area and the road assignment of the conventional survey flows. The main parameter of the simulator is the radius R of the digital detection area, which has been assessed around 4000m on the station #4 thanks to a complimentary NPR survey.

3.3 Building the exchange and transit OD matrix

Once the motorway OD matrices are built on both sides of the border (French and swiss), the junction can be made on the 3 digital stations at the border thanks to 3 specific probabilistic matrices. Any probabilistic matrix describes the probability $p(o,d)$ of a crossing-border trip to begin in an zone o and to end in the zone d. The probabilistic matrices result from the combination of a former (2011) crossing-border survey and the current socioeconomic data from INSEE. The use of a former survey enables to refine any gravity model with local and specific trends. The junction is made between the both motorway matrices with respect to the flow conservation assumption.

OD matrix	Destination			
Origin	F (outside)	F (model)	CH (model)	CH (outside)
F (outside)	1050	1750	1500	50

F (model)	1200	-	6550	100
CH (model)	550	1750	-	3050
CH (outside)	50	100	2900	-

Table 4: Final aggregated exchange and transit matrix - Light Vehicles – Morning Peak Hour (MPH)

3.4 Digital survey versus conventional roadside survey

To better feature the digital station abilities, one location at the border (station #3) is covered with 2 kind of survey: a conventional survey and a digital survey. The figure 7 illustrates the differences between the conventional roadside survey results and the outputs of the digital survey before/after adjustments. Only the French side of the survey is shown, it means that it figures out the distribution of the flows crossing the border through the station #3 to reach Switzerland and coming from a French area (inner area IF_x or outer area ZF_i). Flows are aggregated by origin.

It appears that the main trends are well identified by the mobile phone data (main flows on IF1, IF2 and IF3), but the residual flows are significant. These residual flows are due to the instability of the TEL zoning. Furthermore, as mentioned in previous works [Raballand (2018)], the assessment of the lowest flow by mobile phone data is inconsistency.

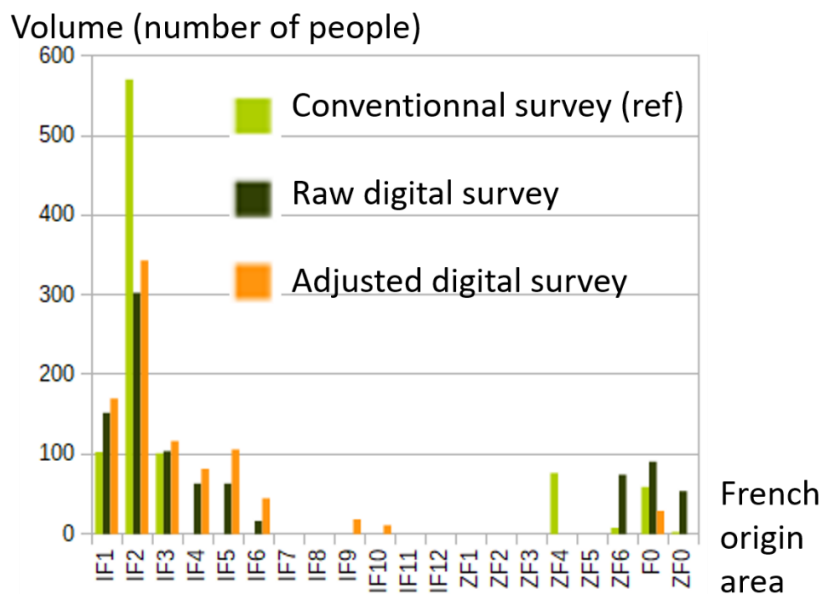


Figure 7: Flow distribution over the French zones for digital station #3 - MPH

4. CONCLUSIONS AND PERSPECTIVES

This work develops a method of using mobile phone commercial data to analyse the mobility of people in an area. In particular, the approach is appropriate to meet the specific needs of carrying out a traffic survey by



substituting "digital" stations for field survey stations that are not feasible on the motorway. In the future, work to make the method more reliable and expand the use of mobile data is needed to confirm the effectiveness and robustness of this innovative approach.

5. BIBLIOGRAPHY

Bonnel, P., Hombourger, E. Olteanu-Raimond, A.-M., Smoreda, Z. (2015) Passive mobile phone dataset to construct origin-destination matrix: potentials and limitations, *Transportation Research Procedia*, 00 (11) 381-398.

Caceres, N., Wideberg, J. P., & Benitez, F. G. (2008), Review of traffic data estimations extracted from cellular networks, *IET Intelligent Transport Systems*, 2(3), 179-192.

Fratrar, T. J. (1954), Vehicular trip distribution by successive approximations, *Traffic Quarterly*, 8(1).

Herring, R., Hofleitner, A., Abbeel, P., & Bayen, A. (2010), Estimating arterial traffic conditions using sparse probe data, *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on* (pp. 929-936).

McNeill, G., Bright, J., & Hale, S. A. (2017) Estimating local commuting patterns from geolocated Twitter data, *EPJ Data Science*, 6(1), 24.

Michau, G., Nantes, A., Chung, E., Abry, P., & Borgnat, P. (2014), Retrieving dynamic Origin-Destination matrices from bluetooth data, *Transportation Research Board 93rd Annual Meeting*, 12-16 January 2014, Washington, DC.

Nabavi Larijani, A., Olteanu-Raimond, A.-M., Perret, J., Brédif, M., Ziemlicki, C. (2015), Investigating the mobile phone data to estimate the origin destination flow and analysis: case study: Paris region, *Transportation Research Procedia*, 00 (6) 64-78.

Raballand, W., (2018). Assessing the relevance of mobile phone data to estimate origin-destination matrices, *Proceedings of the 7th Transport Research Arena*, Vienna, Austria. DOI: 10.5281/zenodo.2241628.

Vanajakshi, L., Subramanian, S. C., & Sivanandan, R. (2009), Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses. *IET intelligent transport systems*, 3(1), 1-9.

Wang, Z., He, S. Y., & Leung, Y. (2017) Applying mobile phone data to travel behaviour research: A literature review, *Travel Behaviour and Society*.