Human and network mobility management using mobile phone data

Solohaja Rabenjamina

N°d'ordre NNT : 2023ISAL0063

THESE de DOCTORAT DE L'INSA LYON, membre de l'Université de Lyon

Ecole Doctorale EDA 512
InfoMaths

Spécialité/ discipline de doctorat :
Informatique

Soutenue publiquement le 29/09/2023, par :
Solohaja Rabenjamina

Human and Network Mobility Management using Mobile Phone Data

---

Devant le jury composé de :

FIORE, Marco	Directeur de Recherche, IMDEA Networks	Rapporteur
CONAN, Vania	Habilité à Diriger des Recherches, Thales	Rapporteur
CARNEIRO VIANA, Aline	Directeur de Recherche, INRIA	Examinatrice
HOTEIT, Sahar	Maître de Conférences, Université Paris Saclay	Examinatrice
SECCI, Stefano	Professeur des Universités, CNAM	Examinateur
RIVANO, Hervé	Professeur des Universités, INSA-Lyon	Directeur de thèse
STANICA, Razvan	Maître de Conférences HDR, INSA-Lyon	Co-directeur de thèse

Département FEDORA – INSA Lyon - Ecoles Doctorales

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **https://www.edchimie-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage secretariat@edchimie-lyon.fr	**M. Stéphane DANIELE** C2P2-CPE LYON-UMR 5265 Bâtiment F308, BP 2077 43 Boulevard du 11 novembre 1918 69616 Villeurbanne **directeur@edchimie-lyon.fr**
E.E.A.	**ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE** **https://edeea.universite-lyon.fr** Sec. : Stéphanie CAUVIN Bâtiment Direction INSA Lyon Tél : 04.72.43.71.70 secretariat.edeea@insa-lyon.fr	**M. Philippe DELACHARTRE** INSA LYON Laboratoire CREATIS Bâtiment Blaise Pascal, 7 avenue Jean Capelle 69621 Villeurbanne CEDEX Tél : 04.72.43.88.63 **philippe.delachartre@insa-lyon.fr**
E2M2	**ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION** **http://e2m2.universite-lyon.fr** Sec. : Bénédicte LANZA Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.e2m2@univ-lyon1.fr	**Mme Sandrine CHARLES** Université Claude Bernard Lyon 1 UFR Biosciences Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69622 Villeurbanne CEDEX **sandrine.charles@univ-lyon1.fr**
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTÉ** **http://ediss.universite-lyon.fr** Sec. : Bénédicte LANZA Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.ediss@univ-lyon1.fr	**Mme Sylvie RICARD-BLUM** Institut de Chimie et Biochimie Moléculaires et Supramoléculaires (ICBMS) - UMR 5246 CNRS - Université Lyon 1 Bâtiment Raulin - 2ème étage Nord 43 Boulevard du 11 novembre 1918 69622 Villeurbanne Cedex Tél : +33(0)4 72 44 82 32 **sylvie.ricard-blum@univ-lyon1.fr**
INFOMATHS	**INFORMATIQUE ET MATHÉMATIQUES** **http://edinfomaths.universite-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 infomaths@univ-lyon1.fr	**M. Hamamache KHEDDOUCI** Université Claude Bernard Lyon 1 Bât. Nautibus 43, Boulevard du 11 novembre 1918 69 622 Villeurbanne Cedex France Tél : 04.72.44.83.69 **hamamache.kheddouci@univ-lyon1.fr**
Matériaux	**MATÉRIAUX DE LYON** **http://ed34.universite-lyon.fr** Sec. : Yann DE ORDENANA Tél : 04.72.18.62.44 yann.de-ordenana@ec-lyon.fr	**M. Stéphane BENAYOUN** Ecole Centrale de Lyon Laboratoire LTDS 36 avenue Guy de Collongue 69134 Ecully CEDEX Tél : 04.72.18.64.37 **stephane.benayoun@ec-lyon.fr**
MEGA	**MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE** **http://edmega.universite-lyon.fr** Sec. : Stéphanie CAUVIN Tél : 04.72.43.71.70 Bâtiment Direction INSA Lyon mega@insa-lyon.fr	**M. Jocelyn BONJOUR** INSA Lyon Laboratoire CETHIL Bâtiment Sadi-Carnot 9, rue de la Physique 69621 Villeurbanne CEDEX **jocelyn.bonjour@insa-lyon.fr**
ScSo	**ScSo*** **https://edsciencessociales.universite-lyon.fr** Sec. : Mélina FAVETON INSA : J.Y. TOUSSAINT Tél : 04.78.69.77.79 melina.faveton@univ-lyon2.fr	**M. Bruno MILLY** Université Lumière Lyon 2 86 Rue Pasteur 69365 Lyon CEDEX 07 **bruno.milly@univ-lyon2.fr**

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Résumé

Au cours de la dernière décennie, nous avons observé une augmentation du nombre et de l'utilisation des téléphones mobiles, plus précisément des smartphones. Ce phénomène, couplé avec l'évolution des différentes technologies et des types d'échange, a entrainé une augmentation des volumes de données échangées par l'intermédiaire des réseaux mobiles des opérateurs téléphoniques. Chaque nouvelle génération de réseau mobile génère plus de données que sa prédécesseure. On estime que d'ici à 2027, 289 EB de données seront échangées par mois, dont 62% seront générées dans le réseau mobile 5G.

La disponibilité de cette immense masse de données a ouvert des perspectives d'études dans des domaines qui se basaient, originellement, sur des longs processus avec des données collectés individuellement. L'étude de la mobilité en est un exemple. Concernant ce dernier, les données mobiles, ont permit de mener des études sur une plus large population et des zones géographiques plus étendues.

Dans cette thèse, dans un premier temps, nous montrons que les événements décrits dans les données mobiles peuvent aussi être retrouvés dans d'autre sources de données. Pour cela, nous avons comparé les données mobiles avec des données provenant de capteurs de détection de présence humaine, par l'intermédiaire de différentes métriques. Cela nous a montré que les deux types de données sont assez bien corrélés. Cependant, des événements comme la synchronisation des pics de présence ou bien la fin de l'activité dans la zone d'étude ont une similarité moindre.

Dans un second temps, nous avons utilisé les données mobiles afin d'étudier l'impact des trois confinements, imposés par le gouvernement français, sur l'évolution de l'utilisation du sol dans la ville de Paris. Nous montrons que le premier confinement a radicalement impacté les habitudes de mouvement de la population et donc l'utilisation du sol, ce qui était un des objectifs de la mise en place des différentes restrictions. Quant aux deuxième et troisième confinement, leur impact sur l'utilisation du sol est moindre comparé au premier confinement.

Enfin, nous utilisons ces données mobiles dans le cadre de la reconfiguration du réseau mobile concernant la gestion de la micro mobilité des utilisateurs, aussi appelé handover. Tout d'abord, nous montrons que les différents eNodeB, constituant le réseau d'accès, peuvent avoir différents profiles et catégories. Les eNodeB peuvent avoir un profil source de mobilité ou bien destination. De plus, les eNodeB peuvent gérer un nombre d'utilisateurs différents, ce qui permet de les catégoriser. Par la suite, en distinguant les utilisateurs mobiles des utilisateurs statiques, nous pouvons faire de l'économie de ressources en reconfigurant le réseau pour ces derniers. En utilisant les différents profiles d'eNodeB, nous montrons que la reconfiguration dynamique du réseau permet aussi une économie des ressources utilisées pour les utilisateurs mobiles.

Abstract

Over the last decade, we have seen an increase in the number and use of mobile phones, more specifically smartphones. This phenomenon, coupled with the evolution of different technologies and types of services, has led to an increase in data exchanged through the mobile networks of telephone operators. Each new generation of mobile network generates more data than its predecessor. By 2027, 289 EB of data will be exchanged per month, 62% of which will be generated in the 5G mobile network alone.

The availability of this immense mass of data has opened up the prospect of studies that were originally carried out in long processes with individually collected data. The study of mobility is an example. For the latter, mobile data have made it possible to conduct studies on a larger population and larger geographical areas.

In this thesis, we first show that the events described in mobile data can also be found in other data sources. For this purpose, we compared mobile data with data from human presence detection sensors, using different metrics. This showed us that the two types of data correlate quite well. However, events such as the timing of presence peaks or the end of the activity, in the study area have a lower similarity.

In a second step, we used mobile data to study the impact of the three lockdowns, imposed by the French government, on the evolution of the land use in the city of Paris. We show that the first lockdown radically impacted the movement patterns of the population and thus the land use, which was the objective of the implementation of the different restrictions. As for the second and third lockdowns, their impact on the land use is less significant compared to the first lockdown.

Finally, we use this mobile data in the context of mobile network reconfiguration regarding the management of user micro-mobility, also called handover. First, we show that the different eNodeBs, constituting the access network, can have different profiles and categories. The eNodeBs may have a mobility source or destination profile. Moreover, the eNodeBs can manage a different number of users, which allows to categorize them. Then, by distinguishing mobile users from static users, we can save resources by reconfiguring the network for the latter. And, by using the different profiles of eNodeB, we show that the dynamic reconfiguration of the network also allows savings in the resources used for mobile users.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Context and motivation

Over the last decade, we observed an increase in the number and the use of mobile phones. This is coupled with the evolution of different technologies, and the increase of data volumes exchanged via the mobile network. Each new generation of mobile network produces more data than its predecessor. By 2027, an estimated 289 EB of data will be exchanged per month worldwide, 62% of which will be generated in 5G mobile networks.

The availability of this huge amount of data opens opportunities to study social and technical phenomena with new approaches. Mobile data is a source of information on human mobility. Indeed, with the number of mobile phone subscriptions exceeding the number of people on Earth, it becomes an unprecedented source of information which provies much broader insight than classical surveys [1]. Therefore, these data are more and more used as a proxy to study human mobility. With mobile phones becoming proxies for human presence, datasets collected by mobile operators outlined many interesting properties regarding human mobility and activity patterns, such as high periodicity [2, 3], shortest path routing [4, 5], or travel patterns that follow preferred locations [6, 7]. While these properties seem to be universal and have been observed in numerous studies over different geographical areas [8, 9, 10], mobile phone data have also proven useful in the study of specific events and situations. A few representative examples are the study of mobility patterns in urban regions during major cultural events [11], in urban planning [12], the analysis of migration patterns following important natural disasters [13], or the link between human mobility and disease propagation [14]. For example, the Covid-19 pandemic is an interesting event in terms of studying human mobility. Indeed, the implementation of various mobility restrictions with lockdowns or curfews modified the general movement habits of the population. The comparison of mobile data during a normal period and during a period of restriction can, among other things, indicate the effects that the imposed restrictions may have.

All these studies make the inherent assumption that mobile phone data represents a natural proxy for human presence and mobility in the studied area. However, there are two important factors with an impact on the accuracy of mobile phone data. First of all, the network operators can only collect mobile phone data when the user is active on the network, i.e. the mobile phone is exchanging data with the network. Of course, the probability of using a mobile phone is not uniformly distributed in space and time, as it can be easily seen in the operator data which classically shows

a much more important network presence during day time compared to night time [15]. Although this phenomenon is well known, it is rarely accounted for in practice, when mobile phone data is used.

The second issue is that the process used by network operators to collect this data is quite complex itself and imposes assumptions that were never properly validated. Practically, mobile phone data is collected by network probes on a per-user basis: when a user has a control plane interaction with the mobile core network, a timestamp, and his serving base station (so, indirectly, his approximate coordinates) are logged. However, this sampling is far from regular and, in case the position of the user changes, some assumptions need to be made regarding the time when this mobility took place [16]. Moreover, such individual user data presents serious privacy issues [17] and it is rarely used, and even more rarely shared, by the mobile operators. Instead, most of the datasets used in the literature are spatio-temporally aggregated, i.e. they indicate how many users were present in (or moved in/out) a geographical area of interest in a given time interval. This aggregation also implies some assumptions, once again related to the fact that the user sampling is not uniform, e.g. some user movements might be observed by the network with a certain delay, or the presence of some users might not be accounted for during certain time intervals.

While the bias of mobile phone data was studied for individual trajectory data [18], the impact of the assumptions behind the aggregated datasets, to our knowledge, was never measured. Of course, this is because it is relatively easy to find a few users ready to log their accurate location data and compare it with operator data, but it is much more complicated to recover ground truth information considering an entire geographical area.

On the operator side, mobile networks are complex systems which are mostly statically planned and dimensioned on a regular basis. These network configurations are based on limited knowledge of the environment, and on basic metrics like the peak of the users demand on the network or the estimated total number of mobile users. However, mobile services supported by the mobile network are not static. Users are mobile and their communication varies greatly over time and location. Therefore, an adaptation of the mobile network depending on the important spatiotemporal fluctuation of the user context should be implemented. Furthermore, recent years have witnessed an increase in the diversification of content consumed via the mobile network. A striking example is the case of video streaming, with the surge of new actors, where the high data rate compensates for packet loss. On the other hand, a traffic such as machine type communication requires reliable channel but it allows high delay and low bandwidth. The mobile network must therefore meet the growing heterogeneous requirements in terms of quality of service (QoS) and key performance indicators (KPIs). Considering these differences will result in several declinations in 5G that will correspond to different network configurations. Each variation has the objective to enforce some KPI, according to the service needs. The main declinations are: extreme mobile broadband (xMBB), ultra-reliable and low latency communication (URLLC) and massive machine type communication (mMTC).

In the presence of such strong (spatiotemporal) context and content diversity of traffic demands (QoS and KPI) of mobile networks, adding resources in the network alone cannot solve all future service needs. A more efficient use of network resource must also be carried out. The next generations of mobile networks must be flexible, and dynamically adapt their configurations to the spatiotemporal fluctuations of services with different QoS and KPI requirements. This better management of resources

in the mobile network must also concern mobility management. Indeed, some users can be mobile, and they may have different constraints compared to static users. Handover for example is a mobility management mechanism, which ensures the continuity of a communication despite the mobility of the user. The proper management of the latter can therefore influence the quality of experience (QoE) perceived by the user, but also resource saving.

In this thesis, we use different types of mobile data: user presence data and user mobility data, represented by handover data. We use data from two different French mobile network operators. As we mentioned earlier, despite the biases observed, mobile data are increasingly used in the study of human mobility. To validate the accuracy of these mobile data, we compare them with the data obtained from presence detection sensors. Then, we use these data to study human mobility, in particular to understand the effect of different mobility restriction measures in France during the Covid-19 pandemic. Finally, these mobile data, namely the handover data, allow us to propose an original design for network slicing on the control plane.

1.2 Contributions

The thesis covers the main following contributions:

- **Cross validation of mobile phone data with human presence detection sensor data.** We compare these two sources of data using several metrics on the data features in Chapter 3. We study the correlation between mobile data and sensor data that is directly collected in the field.

- **Impact quantification of lockdowns on the Parisian land use.** A study of the impact of the different lockdowns imposed in France, and in particular the city of Paris, is discussed in Chapter 4. With different levels of severity, the different lockdowns have had different impacts on the evolution of the type of area, defined by its land use.

- **Zone importance in epidemics propagation.** In Chapter 4, based on features extracted from mobile data, we propose a metric that represents the importance of an area in the epidemic spread. This metric is based on the different peaks of presence in an area, which we assume to be an indication of population mixing.

- **Handover profilling.** We show that an eNodeB can have different behaviors in terms of handover in Chapter 5. These different behaviors are based on the handover type that is handled by the eNodeB (i.e incoming or outgoing). The detection of these different behaviors is conducted with a metric indicating the difference between incoming and outgoing handovers, but also with an unsupervised clustering algorithm.

- **Handover categorisation.** In Chapter 5, we propose to categorize eNodeBs according to the number of handovers they handle. Like with handover profiling, we also use an unsupervised clustering algorithm.

- **Network slicing proposition on the control plane.** Network slicing is often implemented on the data plane side. In Chapter 6, we show the benefits

that slicing can offer on the control plane level, more specifically on mobility management. We also show the benefits we can bring by distinguishing into two different slices the mobile users and the static users.

- **Resource saving for micro mobility management.** We propose a different slice for static users, and another slice for mobile users in Chapter 6. We show that, for the static slice, the number of required ressources can be reduced. For the mobile slice, a dynamic configuration, during the day, can be set up. Considering a dynamic reconfiguration, we compare different scenarios and evaluate the possible resource savings that these can bring.

1.3 Organisation

The thesis is divided into seven chapters, the first of which is this introductory chapter. The remainder of this thesis is structured as follows. Chapter 2 covers the usage of mobile phone data in the literature. We start with biases identified in the mobile phone data and how literature tries to overcome these biases. Then, we spotlight some use cases of mobile phone data, beginning with using mobile phone data as human mobility proxies, and its usage in the context of network management and configuration.

Chapter 3 is dedicated to the assessment of the accuracy of mobile phone data by cross validation with human presence detection sensor data. Three main different metrics are used to compare both sources of data. A spatial analysis and a temporal analysis are carried out to verify the existence of a good correlation between the two datasets.

In Chapter 4, we analyze and compare the evolution of Parisian land use for a normal period and during the three lockdowns imposed in France during the covid-19 pandemic. The characterization of these land uses was obtained through the design of a signature for each area, but also with agglomerative hierarchical clustering for grouping areas with similar signatures in the same land use. We also define a new metric based on the ratio between the peaks present in the signatures and the average of the signatures, which we believe can be related to the spread of the virus.

Chapter 5 characterizes handovers. We describe our methodology to obtain the different profiles and categories of eNodeBs with respect to the behavior and the volume of handover they handle. Afterwards, we analyze the different profiles and categories.

Chapter 6 presents the benefits that slicing can bring to mobility management in terms of resource savings. First, we present the concept of network slicing. Then we propose the creation of a slice for static users and a different slice for mobile users. We present the resource savings that can be achieved by implementing a static slice. Then, for the mobile slice, we present three different scenarios for resource allocation, and the resource saving that each scenario can bring when using dynamic reconfiguration.

Finally, in Chapter 7, we draw our conclusions on our different contributions and provide some perspective for future research.

Chapter 2

Mobile data in the literature

The usage of mobile phone has been increasing for the last decade and will still increase [19]. In 2000, there were 738 million of mobile subscriptions worldwide and for 2020 the number has grown to 8.27 billion subscriptions [20]. Paired with the evolution of mobile network technologies, from 2G to 5G, this increase in mobile phone usage has produced a gigantic volume of data. In 2011 there were approximately 0.2 EB traffic generated per month, and in 2027 this number is estimated to 289 EB per month, 62% of generated traffic with 5G [19]. Each technology allows exchanging more and diverse data than its predecessor [21] and in 5G the explosion of the traffic is due to a massive 5G IoT ecosystem [22]. 3G, 4G and 5G are designed to allow exchange of data at different rates, contrary to 2G where only voice and text messages can be shared [23].

The rise of the amount of generated data, and consequently the increase of the amount of mobile phone metadata, which describes the activity of users seen by the mobile network, has created the opportunity to study different phenomena from a new angle [24]. Mobile phone data are increasingly used as human mobility proxy indicator [25]. It allows, for instance, to describe how certain events can influence human mobility and presence in some areas [26]. However, to describe real phenomena that occur in the real world, mobile data should contain information close to reality [26]. Indeed, the results of the various analyses of these mobile data can only be taken into account if they reflect reality. If not, the possible measures based on these results are not usable. Furthermore, with the fast-growing new generation network, mobile data is needed in order to make the network more intelligent and flexible [27]. Mobile data describes the behavior of traffic passing through the mobile network, and can also be used for traffic forecasting, so its use will be useful to adapt the network dynamically to the traffic.

This chapter is structured as follows. First, we will provide a literature review of work that assessed the accuracy of mobile data. The second part deals with different usages of mobile data as human mobility proxy. In the third part, we will discuss about the impact of the configuration and management of the mobility at the network side.

2.1 Mobile data quality assessment

Mobile phone data found numerous applications in the literature. Some of these applications are related to the management and configuration of mobile networks [28], while others use the mobile network information as a proxy for human presence [29]

and mobility [30]. In the former case, the use of mobile phone data does not introduce any bias, since it is precisely the network activity that counts: data regarding the mobile network usage is collected, analyzed and then used to propose new traffic-aware management solutions for the same network [31]. In the latter case, however, the activity of mobile users on the network is considered as a proxy for actual human presence in the corresponding area, leading to studies using mobile phone data to guide the evolution of the road infrastructure in a city [12] or to evaluate the potential benefits of a ride-sharing service [32]. The conclusions of these non-network related studies are deeply related to the assumption that mobile phone data gives an accurate picture of human presence and mobility.

2.1.1 Mobile data bias

User localisation

The bias introduced by mobile phone data in the reconstruction of individual trajectories was characterised a decade ago by Ranjan *et al.* [18]. This study aimed to evaluate the bias that Call Data Records (CDR) can present in terms of human localisation and presence. Indeed, CDR, which correspond to signalling data related to voice calls and SMS made by users, and stored by operators for billing purposes, are frequently used to infer human mobility. In [18], the authors conduct an analysis on a dataset that covers 7000 square miles of San Francisco Bay for over 1 million users during one month, July 2011. They use the Shannon entropy Eq.2.1 and the radius of gyration Eq.2.2 to estimate bias in CDR. The Shannon entropy is, here, the measure of the spread of a user's activity (i.e. voice, SMS or even data) over his location. In Eq.2.1 P_i represents the probability that an activity was observed at location i from the set of O locations the user can visit. The radius of gyration is used to quantify the extent of user trajectories. In Eq.2.2 $\overrightarrow{R_t}$ is the position of the user at time t and $\overrightarrow{R_{cm}}$ is the center of mass of all the locations where the user was seen, during L time steps.

$$\mathcal{S}^U = -\sum_{i=1}^{O} P_i \log_2(P_i), \tag{2.1}$$

$$\mathcal{R}_G = \sqrt{\frac{1}{L}\sum_{t=1}^{L}(\overrightarrow{R_t} - \overrightarrow{R_{cm}})^2}, \tag{2.2}$$

The authors showed that significant locations for the user (such as his home or workplace) are correctly retrieved by using mobile phone data, but the resulting trajectories present significant errors. This is a consequence of the spatio-temporal granularity of mobile phone data, where the user's location is associated to the position of his serving base station and the location sampling is irregular and dictated by the phone activity. However, this can be mitigated by completing the CDR with data access records. Indeed, the activities recorded in the CDR are initiated by the user and will therefore depend on the use of the latter, whereas with the data access records, which will record the session data, we are less dependent on the user because some data exchanges are automatically initiated in the background.

User trips

Caceres *et al.* [33] compared two types of Origin-Destination (OD) matrix on an hourly basis: an OD matix created with household travel surveys, coming from census data, and an OD matrix obtained with mobile phone data. The latter was generated with CDR and passive signalisation. The passive signalisation allows to have more events, which are not triggered by the user. Their study focuses on the region of Malaga in Spain, on an area of 1400 km^2 and about 200 000 people during two consecutive weeks in February 2015. To compare the two OD matrices, they used Pearson's correlation coefficient and the Mean Structural SIMilarity (MSSIM) index. Pearson's coefficient shows the degree of similarity in a quantitative point of view between the two OD representations. The MSSIM was initially developed for image comparison at pixel level. Therefore, instead of comparing statistically all the pixels of the two images together, MSSIM computes statistics on groups of pixels, then takes the mean. It is interesting because adjacent pixels tend to have strong dependencies. By using MSSIM, the authors assimilated the OD matrix as image. Their OD matrix gives good results on the accuracy of the origin and the destination, however it misses short trips. With a similar objective, Qiu *et al.* [34] collected vehicular user data to prove that mobile phone data estimates the travel time of a car with an error of 5-15%.

2.1.2 Aggregated mobile phone data and census data

Most of the research that tried to assess the accuracy of aggregated mobile phone data used census data as ground truth. Multiple independent studies showed significant matching between the night-time user distribution in mobile phone data (considered as a proxy for home locations) and national and local census data in different areas.

User localisation

Issacman *et al.* [30], with their work on the range of human mobility, wanted to compute the maximum distances that a user makes per day, what they call daily range. To that end, they used two months of CDR data that covered Los Angeles and New York. They showed that Angelenos and New Yorkers present different behaviors. To assess the results obtained with the CDR, they relied on volunteers. They had the CDRs of these volunteers and who also had to note their location each time they changed places, and the time they spent on these places. By comparing the results based on the volunteers' CDRs and the data that the volunteers recorded, they obtained median errors of less than 1.5 miles, which they considered satisfactory.

Douglass *et al.* [35] wanted to know if the mobile data could reliably represent the population distribution of an area. For this end, they use CDR data and Internet activity that covered the city of Milan. These data have a granularity of 10 minutes and the users localisation is referred to a 235m x 235m meter grid cells, unlike many works, where the location of users is associated with those of the antennas to which they are connected. To validate the population distribution seen by the mobile phone data, the authors compare it with high-resolution census. The result show that, in areas with a large population, the mobile data gives a good estimation with an error of about 200 persons.

Lenormand *et al.* [36] cross check three data sources: Twitter, census and mobile phone data from users' call record, from urban area of Madrid and urban area of

Barcelona. However their datasets do not cover the same period. To compare these three sources of data, they used Pearson correlation coefficient on three different metrics: *(i)* the spatial distribution of the density of population, only on Twitter and mobile data *(ii)* the temporal evolution of the population density (where the users are located depending on the hour of the day), only on Twitter and mobile data and *(iii)* the users mobility pattern based on OD matrix. Their results showed that Twitter and mobile data have a Pearson correlation coefficient of 0.9 on the density of population; both datasets show a similar temporal distribution as well. For OD matrix, they are similar for Twitter and mobile phone data, and with the census, the difference is also acceptable. The authors consider that some observed differences can be explained by the fact that the data was not collected during the same period.

Origin destination

Calabrese *et al.* [37] propose a method to estimate OD flows. For this purpose, they use mobile data, CDR and Internet connection data that cover over a million users in the Boston metropolitan area. Their method aimed to build a dynamic OD matrix to include details for each user trip, like the overall distance traveled or the number of paths taken. Their method gives two types of information. The first is the characterization of the trips and the second is their OD matrix. To validate this information, they compare the results with census data. For that purpose, they count for each cell the number of users during the night (6 pm to 8 am) then consider that the cells which have the highest number of users are the home cells. They compare the result with data from census, which describe population home distribution. They also compare their result with another census datasets which describes home work trips. These comparisons showed that, on average, there is a good linear relationship between census data and their method using mobile phone data.

Bekhor *et al.* [38] evaluate long-distance travel patterns at a nation range, in Israel, by tracking mobile phones positions. For this evaluation, they use the mobile phone position approximate to the position of the antenna to which the mobile phone was associated. To evaluate if there were biases in the mobile data, they performed on it home location detection. For the authors, with mobile data a home is where the mobile phone is during the night hours. Then they compare this distribution of home locations to the official statistics. They find a Pearson correlation coefficient of 0.86, which they consider acceptable. To validate the OD matrix constructed with mobile phone data, they compared it to a survey about national travel habit. This comparison showed that mobile phone data present more OD pairs than expected from the survey.

At a country level, mobile phone data was used to estimate the number and the destination distribution of people who left the affected area in the months following the 2010 Haiti earthquake [26]. To that end, the authors used the location of the mobile users, approximated to that of the antenna to which they were connected and found the evolution of these locations during the period of study. Their results showed a good matching with a large survey conducted by the United Nations.

Regarding mobility information, Schneider *et al.* [39] affirm the existence of patterns in the movement of people, and said that mobile phone data can capture them. These patterns called motifs represent the way in which people move around the study area, the traveled distance of a trip, the sequence of visited locations, and paths between visited locations. They extract mobility motifs from CDR data that cover Paris. To validate these motifs, they compare them with mobility motifs from travel

surveys. Their results showed that mobility motifs extracted from mobile phone data correspond to those obtained by using population surveys.

However, some survey-based studies also noticed significant differences with mobile phone data. For example, Tizzoni *et al.* [14] identified an overestimation in commuting flows inferred from mobile phone data with respect to national census data in France, Spain and Portugal. In the same lines, Wesolowski *et al.* [25] discovered that users who travel a lot also use their mobile phones more, introducing an overestimation in the average level of mobility of the population. In order to get this result, they use individual data extracted from CDR, provided by an operator in Kenya, which allowed them to define the mobility pattern. Then they compare this mobility pattern to the one extracted from travel survey collected in the same period.

2.1.3 Comparison with other data

Bachir *et al.* [40] proposed a methodology to infer dynamic Origin-Destination flows per transport modes using mobile network data. To that end, they used CDR, Internet session data, and location area data on Greater Paris region. The identification of the transport mode was done in two steps: *(i)* semi supervised learning by using hierarchical clustering on antenna sectors, in order to group sectors with similar transport usage, and *(ii)* Bayesian inference to determine the most probable transport mode associated to a mobile phone trajectory. After these two steps, the authors build the OD matrix by transport mode. They compare their result with two types of data, census data and travel card data. The validation of the result with census data is the same as previously mentioned works. The authors use Pearson correlation coefficient and normalized root mean square error (NRMSE) metrics to compare mobile data result with travel card data. These comparisons show a median Pearson correlation coefficient of 0.98 and the NRMSE median is 0.062.

Ma *et al.* [41] use a different method to assess mobile data accuracy in terms of highway mobility trajectory. For that end they collected car information by using a licence plate recognition system and find a strong correlation, up to 0.94, with mobile phone data extracted from CDR and Internet data usage.

2.1.4 Bias correction

A series of works have been conducted in order to solve the individual trajectory problem discussed above. Chen *et al.* [42] proposed to solve the problem of sparse and irregular in time CDR data, by using complex interpolation techniques. Indeed, as discussed earlier, CDR data depend on user activities, which are unequally distributed in time and space. The authors proposed a solution that allows to reconstruct the trajectory, thus to estimate the position of the user when the original mobile data does not provide this information. Their solution, called Context-enhanced Trajectory Reconstruction (CTR), is based on three habits of mobility that people have: *(i)* a person will spend a significant time, of the order of several hours, in a fixed place, *(*ii) during the night, a person is often at home, and *(iii)* there are regularities in the movements of a person, who tends to visit and return to the same place, which gives a periodic pattern to his movements. Using these three assumptions, the authors propose:

1. Reconstruction of night time trajectories: easily achievable with the knowledge of the location of user homes (extracted from the mobile data). The user home

is the location where the user spends the most time during the night. When a record is missing during a night for a user, the user's home location is assigned.

2. Reconstruction of the trajectory during the remaining time: based on the fact that there are regularities in the movement of users. To exploit this, they adapted the tensor factorization technique that allows them to deduce the missing values using their hyper-parameters.

3. Homogenisation of location: homogenize the locations estimated by the tensor factorization by associating them to the most appropriate cell tower location. This association is made to have the same location base as the rest of the data. Indeed, the location of users in the original dataset is approximated by the location of the cell towers.

The CTR trajectory reconstruction is acceptable with a median cell displacement of two cells, either for the nearest adjacent cell or the one cell further.

Another solution to solve individual trajectory problem is proposed by Bonnetain *et al.* [43]. They propose TRANSIT, which uses Network Signaling Data (NSD), instead of CDR, to detect user mobile phase and stationary phase, then infer and reconstruct the user trajectory in mobile phase. NSD has the advantage of having data generated at the level of the control plane: signaling data such as handovers, Location Area (LA) and Tracking Area (TA) updates, active paging, network attaches and detaches, and data connections, and also at data plane level with voice and SMS (the same information present in CDR). Therefore, NSD are more complete data, less sparse and less unequally distributed in time and space than CDR. Like CTR, TRANSIT also uses assumptions about people's mobility habits, thus it exploits the redundancy in the data. The authors proposed a trajectory reconstruction in two steps, detailed below: *(i)* detect when the user is static and when the user is mobile, and *(ii)* for the moments when the user is mobile, improve the estimation of the trajectory that the user has taken.

1. Static and mobile user event detection: first, tag antennas as static. An antenna is static for an user if the user is connected to this antenna for at least a certain cumulative time during the day. This user will be considered as static at that time when all his activity is associated with that antenna. Then, during the time when the user is not static, he is tagged as mobile.

2. Trajectory improvement for user mobile event: based on the habit of the user, for the same point of departure and point of arrival. For that, the authors carry out a DBSCAN clustering on all the paths having the same point of departure and point of arrival, and base on that to improve the trajectory through interpolation.

With their dataset, TRANSIT achieves 80% of precision and 190 m spatial accuracy in the estimation of the trajectories.

2.1.5 Privacy constraints

Individual user mobile phone data is rare and strongly regulated by privacy protection policies [17]. For instance, the General Data Protection Regulation (GDPR) [44] says that personal data shall be securely stored and not freely distributed. For more open distribution, data must first go through an *anonymization* process, which prevents

the data to be linked to an identified or identifiable natural person or data subject. Whereas, real *anonymization* is technically very challenging. For these reasons, most of the mobile phone data used in the literature takes the form of aggregated datasets, that indicate simply the number of users present in an area, or the number of users moving between two areas in a given time interval. Among other use cases, this type of data was heavily used recently to analyze human mobility during the Covid-19 pandemics [45].

2.2 Mobile data as human mobility proxy

We saw previously that mobile phone data has been used for over a decade as a human mobility proxy. In the following, we present a brief overview of how different fields of study use mobile phone data to describe some phenomena also called urban computing [46]. In the remainder of this part, we will focus on the transport field, study on marketing, land use and epidemics studies, focused on Covid-19.

2.2.1 Transport field

Traditionally, urban planning and transportation engineering have used household surveys or censuses and road surveys to study population mobility. This method has several disadvantages, the first being high cost and the time needed to carry out the surveys. Then, the representation is complex and it has difficulties to represent the evolution of the population, because it is difficult to update. Finally, it does not allow to put forward the behaviors of the population for particular periods like vacations. Mobile phone data allow to compensate these disadvantages, even if census and surveys will still be used if only to validate the results obtained with mobile phone data.

Transport mode detection using mobile phone data is still at its early stage [47], even if knowing transport mode can help to fluidify the movement of the population, for instance by adapting the frequency of every transport mode. Huang *et al.* [47] described methods that are frequently used to detect transport mode. There are three main methods: *(i)* rule-based heuristics (RBH), which consist in the comparison of extracted features to the closest transport mode feature, based on speed for example or the nearest transport infrastructure network, *(ii)* clustering, especially K-means and hierarchical agglomerative clustering, based on trip features, like speed or duration, and *(iii)* statistical analysis for driving, biking or walking transport mode, mainly based on hidden Markov model.

With the incentive of public services for the valorisation of public transport and therefore its adoption by the greatest number [48], they must be optimized in order to best satisfy the needs of users. Berlingerio *et al.* [49] propose a system, AllAboard, to identify optimal location for new transport routes, by also considering existing transit networks. They test their system on CDR data that cover the city of Abidjan, Ivory Coast. AllAboard proposed four new routes which allow an expected 10% reduction in travel time across the city of Abidjan.

2.2.2 Study on marketing

The great breakthrough of mobile phone, especially with smartphones, in the population has led to the generation of huge amounts of data. This data is a great source of

information to describe the habits of people. The marketing field will use these habits to propose a more accurate service in terms of location, for example by estimating the transit of potential customers on a given route [50]. It can also use such data to maximise the revenue of telecommunication industry. The latter can use mobile data to improve customers experience and their satisfaction, by optimizing the offer, churn cause identification, personalise advertisements, but also cell site optimization [51].

Manchanyake *et al.* [52] tried to predict the customer who will be inclined to increase their data plans, then proposed to upselling to a higher data plans. To that end, they used 4G CDR with data usage, but also extra data volume purchase history. They showed that a logical regression model is an accurate model to predict the customers who will potentially want to upgrade their plan.

Lin *et al.* [53] identify several customer behaviors and indicate marketing strategies for some of these behaviors. The authors used CDR data for their analysis. To identify different types of behavior, they used K-means clustering to group similar behaviors in the same cluster. They identity three groups: professional and international group, big local service consumer group and small usage group. They setup different suggestion strategies for each group. Some of their suggestions were implemented in the real world by a mobile network operator and they noticed an increase in new mobile customers.

2.2.3 Land use detection

Each plot of land can be defined by its physical characteristics. This will characterize its use, for example agricultural, park, water. In an urban area the combination of urban infrastructures and predominant activities of people at those locations can be assimilated to the socioeconomic function of that given area which we define here as land use. Knowing the land use in an urban area can provide a better understanding of the city by the different people who pass or live there. For example, for tourists, knowing the different composition of the neighborhoods can facilitate the planning of their trip. For the locals it gives an overview of the evolution of the activities in their city. Land use can help as well public urban planning. The knowledge of the different functions of each area allows to improve the different services that a metropolis can offer. For example, better interconnecting residential areas with activities areas through public transport. Or to improve the quality of life of the inhabitants by setting up, for example, green areas. Business and leisure can also take advantage of the knowledge of land use. Choosing strategic locations, for example between business and residential areas, can increase the number of people passing through.

Public authority establishes urban planning rules, which lead to zoning regulations, i.e. how an area is supposed to be used. However, cities evolve through time and the original zoning may not have to reflect the actual land use. Several methods exist to determine the land use at a specific time, like surveys, point of interest (POI) determination or remote sensing.

The survey is the original method. Inferring land use by conducting a survey presents the same advantages and disadvantages as mentioned above for other fields of study.

The POIs are geographical points defined by their coordinates and indicate the activity which is present at this specific point [54]. Therefore, POI can indicate the location of companies, universities, businesses, restaurants, multimodal exchange hubs, etc. An area where companies are in the majority can be qualified as an activity

zone, the case for example in technological parks or central business districts. Land use inference of an area using POI can be based on the concentration or frequency of the POI in question [55] or using machine learning techniques [56]. For residential areas, home localisations are used to define them. As a reminder, home localisations are often defined by the place where users are during the majority of nights.

The other traditional method to determine the land use is remote sensing technique. It is based on the physical characteristic of the land. It allows to define if, on a specific area, there is a building, road, tree, water present. This information alone cannot define the land use [57], for example, for a building, this cannot differentiate residential construction with others. Additional information is therefore necessary to define the function of the area. Information like the size of the parcel, the type of the building, etc. An example of study that uses remote sensing technique was conducted by Douglass *et al.* [35]. To estimate the population size and distribution over the city of Milan, they had to define the land characteristics. To that end, they divided Milan into 15m x 15m cells. Then, for each cell, they classified them into five categories of land usage: buildings, vegetation, water, road/pavement, and railroads. They used high resolution satellite image and Open Street Map (OSM) for the classification. Information given by OSM allowed them to extract the land usage of 41% of the cells. For the remaining 59%, they use the satellite image with 15m resolution, and Random Forest classification using land usage extracted from OSM as training data.

These different methods can be quite cumbersome to implement and update, need additional information to be more accurate and with the evolution of the city, may be rapidly not up to date. Whereas mobile data represent the activity pattern of users in a specific area, pattern which allows inferring the land use [58]. By updating the collected mobile data, it is easier to update the land use.

In the following, we distinguish several categories of methods to determine the land use from mobile data. Feature selection methods, supervised and unsupervised clustering are some of them.

Features selection, reduction, inference

Land use can be inferred from mobility patterns. Several information can be extracted from these mobility patterns, which can characterise a specific land use.

Reades *et al.* [59] show that some specific area can be characterised by some of its specific features, such as the eigenvalue and eignevector obtained as follows. To this end, they used aggregated CDR that covered the city of Roma, Italy, with a granularity of 15 minutes. They divided Roma into 2115 cells of 500m x 500m cells. From these data they build the Erlang signature, which represents for each cell the evolution of demand through time. Then, they used eigendecomposition process to extract patterns from these signatures and obtained eigenvalues and eigenvectors, which they called eigenplaces due to the link with geographical areas. Some of the eigenvectors show the similar characteristic of each cells, like circadian rhythm of the city. However, others can show specific features of a specific cell, for example, high value during late night and early morning represent nighttime activity for well-known nightlife cell. The authors showed that extracted features from mobile data can capture specific values that describe activity in a specific area. Nevertheless, they stated that, in their case, using eigenplaces alone cannot map land use on each cell, due to a high degree of overlap between activities.

Furno *et al.* [60] use a common tool in psychology research, Exploratory Factor Analysis (EFA). It consists in finding automatically latent factors that will explain the

correlation between the observed variables. Therefore, according to EFA the observed variable is a linear combination of factors:

$$\boldsymbol{X} = \boldsymbol{\Lambda F} + \boldsymbol{U} \tag{2.3}$$

In Eq.2.3, the observed variable $\boldsymbol{X}$ corresponds to a weighted common factor $\boldsymbol{F}$, a K $\times$ 1 vector, and unique factor $\boldsymbol{U}$, N $\times$ 1 vector. $\boldsymbol{\Lambda}$ represents an unknown N $\times$ K matrix of common factor pattern coefficients. The authors identified the most appropriate rotation of factors $\boldsymbol{\Lambda}$ and use EFA in order to discover land use in Paris. They use CDR data with one hour granularity aggregated to antenna level, and Voronoi tessellation obtained from antenna location to represent the spatial granularity of the land use. They compare the land use given by EFA with *Agglomerative Hierarchical Clustering* (AHC) using Pearson correlation as metric, average distance linkage and skewness minimization as the stopping rule. They found approximately the same geographical distribution of main land use, residential and business with the two methods. However with EFA, a limited number of factors describe these land use, whereas in AHC, there are more clusters to cover the same spatial distribution. Hence, in terms of land use detection, EFA detects the main macroscopic land use, as AHC but it can also detect information about more mixed land use. Indeed the main land use of an area is described by its most important factor, and if its others factors are not negligible, they can also describe its others land uses.

Mao *et al.* [61] use non negative matrix factorisation (NMF) to infer the land use. Prior, they create a timeseries signature based typical median week (i.e. median value of all 8 am on Monday for an area) from a version of z-score on original timeseries. Then they construct a matrix $\boldsymbol{A}$, where rows are timeseries signature for each zone. Moreover, they tried to approximate $\boldsymbol{A}$ by two other matrices:

$$\boldsymbol{A} \approx \boldsymbol{WH} \tag{2.4}$$

$\boldsymbol{H}$ is B x M matrix where B is number of land use to detect and M is the size of a signature. For a typical week of one hour granularity, this size is 168. Here $\boldsymbol{H}$ contains fundamental patterns that describe each land use. $\boldsymbol{W}$ is E x B matrix where E represents the number of zones to be characterised. $\boldsymbol{W}$ is a combination of weights that will be used to indicate to which likely land use a zone could belong. This is a typical non-NMF, because they try to approximate a high dimensional matrix with lower dimensional matrices with non-negativity constraints. To obtain $\boldsymbol{W}$ and $\boldsymbol{H}$, they use a gradient descent algorithm. To test their method, the authors used CDR data over Dakar, Senegal. Their typical average week signature represents the typical mean evolution of call volume. They compare the land use they obtained with known POIs. They found that there is a good correlation for business, restaurant, and shopping POIs with Business/Industrial/Commercial predicted land use. Also POIs like embassies, offices, and hotels are located in these land use. Their method can detect mixed land use because of the weights they have in the $\boldsymbol{W}$ matrix.

Supervised clustering

Supervised clustering techniques are also used to infer land use. Supervised methods use prior knowledge on different zones to train the algorithms and then classify the rest of zones based on extracted features from previous knowledge.

Toole *et al.* [58] used a supervised clustering technique to predict the dynamic land use across the city. To that end, they use enhanced CDR that covered the

metropolitan area of Boston. Unlike traditional CDR, the location is estimated by triangulation of the user. Then, they divided the metropolis in 200m x 200m cells and aggregated user activity at cell granularity (i.e. all user activities at a given time are summed). The activity pattern that they consider comes from all the events that occur in one hour interval. From this activity pattern, the authors create a timeseries for each cell. Then they use a z-score normalisation to consider the behavior of each cell instead of the volume of events. To infer the land use of each cell, they use Random Forest. For the training dataset, they labeled normalized timeseries of some cells with classes provided by the zoning regulations. Due to the predominance of residential areas, it is difficult to detect more specific land use, like parks or universities. However, they still have an overall accuracy of 54%, and affirm that using neuronal networks led to similar results.

Pei *et al.* [62] use a semi supervised technique to determine the land use. To that end, they worked on data aggregated at antenna level that cover Singapore at hourly basis. They use fuzzy c-mean (FCM) as a clustering algorithm. Prior to using FCM, they create a timeseries that combines a normalised pattern (i.e. evolution of the number of calls through time) and the total calling volume. This combination needs a weight value. The authors obtained this weight value through a training process using samples from known different land use areas and then maximise an objective function to determine it. Using FCM soft clustering, each data point is assigned a likelihood or probability score to belong to a cluster. Therefore, mixed land use could be detected. FCM generates a validation index during its execution. This index can be used to determine the number of output clusters. To validate their results, the authors take the data from the Singapore urban planning and the land use they have defined. They achieve about 58% accuracy compared to the urban planning land use map. They mention several possible explanations for their results: *(i)* the criteria for land use allocation by urban planners are not consistent *(ii)* the heterogeneity of land use, mixed land use *(iii)* the granularity at the antenna (antenna aggregation); the higher the antenna density, the more accurate the land use can be, because there is less likelihood that there is mixed land use.

Zhan *et al.* [63] compare supervised clustering and unsupervised clustering in inferring land use. The authors test their methodology on data from a social network, Twitter, on New York. They divided the area of study into 200m x 200m cells, and each cell contains aggregated user information. A location-based service provider affected to some cells an activity category e.g. home, work, shopping. Before using each clustering technique, they apply a dimension reduction to tackle the curse of dimensionality [64], the possible degradation of performance when using large number of features. The authors use Laplacian Score (LS) for feature reduction. Firstly, they computed the LS for all features, then keep features with the top R highest scores. For unsupervised clustering, they used K-means. As stopping rules they use two indices, Dunn's index (DI) and Xie and Beni's index (XB). For the supervised clustering they tested several algorithms: Naïve Bayes method, Support Vector Machine (SVM) and Random Forest. For the training data, a major part of their data from Twitter have labels on the type of land use, and they split in two parts the cells for each land use. The first is used as training data and the second as test data, together with cells which were not labeled. The Naïve Bayes classifier has a poor accuracy and SVM classifies all cell as residential. Hence, Random Forest algorithm is the one that is most suitable to detect land use among the three. By comparing the predicted land use with land use provided by New York City Department of City Planning (NYCDCP) on the same 200m x 200m grid, K-means reaches an accuracy of 65% and Random Forest

up to 78%. However, these values are not comparable, since supervised classification can detect additional land use. To detect the main land use, and if no ground truth data is available, unsupervised clustering is adequate.

Unsupervised clustering

K-means is one major technique used for unsupervised clustering algorithms. K-means is a method that seeks to minimise the intra-class inertia. At initialisation, the K-means algorithm will choose k random points that will be considered as the center of gravity of the k classes. Then it will compute the center of gravity $G(C_k)$ of each class C_k. For an element w, it computes the distance of this element to each center of gravity $G(C_k)$, in order to find the center of gravity closest to w. Next, it affects w to the nearest class. Then, it redoes the process until no change occurs. Each solution found is a local optimum for the minimisation of the intra-class inertia. The stopping criterion of the algorithm is either the intra-class inertia stops decreasing, or a number of iterations fixed by the user has been reached. At each initialisation, a different result will correspond. To obtain a stable partioning, it is advisable to carry out several executions with different initialisations. The user must set the number of classes in advance.

Calabrese *et al.* [65] use K-means in [59]. The authors use WiFi data from the network at the Massachusetts Institute of Technology (MIT) with more than 3000 access points (AP). Their objective is to build an overview of the use of rooms, buildings and spaces on the campus. For that end, they build eingenplaces, as explained above, then cluster them with K-means. To find the number of classes, they used the Silhouette index. They showed that clusters present a mixed usage, which can be explained by the fact that an AP can cover different sorts of areas at the same time, for instance an auditorium and a cafe. As for Soto *et al.* [66], they propose a methodology to detect land use for any type of data and provide an example using CDR data. They used antenna spatial granularity level. Each antenna has a signature which represents the evolution of the number of calls through time, each 5 minutes time slot. They proposed three types of aggregation to represent the antenna signature: *(i)* total aggregation, it corresponds to the average typical day (i.e. 24h timeseries), *(ii)* week-day and weekend aggregation, it is the concatenation of typical weekday and typical weekend, *(iii)* daily aggregation or typical week, it is the concatenation of typical Monday to typical Sunday. Each of these three levels of signature aggregation can be used to identify the land use. They use K-means on the signatures of all antenna to automatically detect the underlying land use. To obtain an appropriate number of clusters, they used clustering validity index values. To validate their method, they used data that cover Madrid. Due to the lack of information, at that time, on the land use, they checked one by one on a map the spatial distribution of the different clusters and noticed that each cluster covers different types of activities, hence land use. They also applied their methodology in Barcelona and obtained similar results. However, the three levels of aggregation do not give the same accuracy. Total aggregation is the less accurate since it attenuates the difference, sometimes important, that can have the weekday and the weekend. Daily aggregation and weekday and weekend aggregation present in many instances great similarity.

Dong *et al.* [67] use K-means to detect different land use. Instead of using the raw evolution of network events, possibly normalized, for example the use of signature like in some previous work, they use another semantic. They set up a combination

of metrics: real time volume of mobile phone in zone, inflow (i.e. number of mobile entering in the zone), outflow (i.e. number of mobile phone which exit the area), and incremental flow (i.e. inflow - outflow). They use this traffic semantic in addition to the geographical coordinates of antennas as an input to K-means. They use the geographical coordinates to have a homogeneous division of the geographical space. The authors set the number of clusters based on the number of traffic zone divisions that a city can have depending on its size and its population density. They compared their land use with the actual land use, collected by the authorities for general zoning purpose, of Beijing and found a match between the corresponding working areas and residential areas.

AHC is another unsupervised clustering technique that can detect the land use. At the begging of the algorithm, it starts with each element placed in a separate class, then the two most similar elements or classes are grouped together, until only one single class is obtained. Similarity between two elements is computed by using a specific metric.

To study the spatial distribution of the population, Lenormand *et al.* [36] use AHC to detect land use in Barcelona and Madrid. They use detected land use as validation for the temporal distribution activity pattern (i.e. timeseries describing how activity evolved through time). For their study, they used timeseries which represent the typical day from Monday to Thursday, concatenated with typical Friday, typical Saturday and typical Sunday. They used average linkage method and Pearson correlation coefficient as similarity metric. The authors also tested K-means but the silhouette index values, used to compute the number of clusters, were better with the AHC algorithm. They pinpointed three main land uses, from three clusters: residential, business and nightlife. The activity patterns for these clusters present some specificities, for instant, high activity during the night and more in particular during the weekends for clusters labeled as nightlife land use. They obtained the same three patterns for both Madrid and Barcelona.

Furno *et al.* [15] also use the principle of *mobile traffic signature*. They compared different methods to build and to cluster signatures to obtain accurate land use: Weekday-Weekend Signature (WWS), Typical Week Signature (TWS), Seasonal Communication Series (SCS) and their Median Week Signature (MWS). WWS is the concatenation of average typical weekday with the an average typical weekend, it is used in previously described [66] with K-means clustering. TWS corresponds to the typical average signature, described originally in [68] and used with K-means to detect land use. SCS, originally described in [69], uses the whole time sequence, normalized with standard score and relies on AHC with Pearson correlation coefficient as distance metric to detect land use. The method proposed by the authors, MWS, used standard score normalized typical median week (e.g. the median of all 8 am slots) and AHC with Pearson correlation coefficient as distance metric to detect land use. To compare the land use resulting from these different methods, the authors used different ground-truth data which contain different information on the land composition (i.e. parks, woods, university location, school location, business) and land use. The computation of ground-truth information can be assimilated to remote sensing technique or POI detection. The evaluation of the relevance of the results is done through metrics indicating the ground truth of the class and whether this actually corresponds to the clustering result. These metrics are: *(i)* density, which measures the frequency of presence of the ground-truth in the signature class, *(ii)* entropy, which corresponds to the dispersion of the ground truth in the class, *(iii)* coverage, representing the percentage of presence of the ground truth in the class, and *(iv)* F-

score value that indicates the quality of the class (signature) through a combination of entropy and coverage. The results of these metrics showed that the MWS method process is the best to detect underlying land composition and land use. Then, the authors used their methodology on 10 cities over France and Italy to detect land use. They successfully detect main land use, which are residential and office areas. They can also detect leisure and touristic areas.

Stopping rules

AHC algorithms build a series of cluster levels, by merging similar clusters in an iterative operation, from one element per cluster to a single cluster containing all elements. A dendrogram usually displays a graphical information of all clustering possibilities. The existence of all these cluster possibilities means that the AHC does not give us the optimal number of clusters. It is for this reason, also developed in the literature, that the AHC requires the use of a *stopping rule* to select the best aggregation level and thus the number of clusters. In [70] they studied 7 stopping rules, namely: Calinski-Harabasz (CH), Beale index, C-index, Duda-Hart index (DH), Silhouette index (SL), Hartigan index (HT), and Krzanowski-Lai index (KL). These different stopping rules give different scores to each cluster number. Then the authors developed the principle of *top-K index* to choose the best cluster number.

2.2.4 Epidemics

Contact tracing

Digital technologies were one of the main tools in the fight against the Covid-19 pandemic. This is especially true in the contact tracing step, a task for which many countries proposed specific smartphone applications [71]. Most of these solutions are based on Bluetooth ranging [72], with a lot of attention being given to privacy problems [73]. Because of these privacy issues, these applications cannot actually be used to understand the impact of the pandemic on human mobility.

Mobile phone data collected by network operators has been used in the past to understand the propagation and the impact of mobility on infectious diseases, such as malaria [74] or HIV [75]. With respect to the Covid-19 pandemic, several countries, especially in Asia, included data collected by mobile operators in the official tracing and monitoring operations [76, 77, 78]. Numerous academic studies also included mobile phone data, with different objectives. Oliver *et al.* [79] provide a detailed discussion of these works and the way they can be used by governments to estimate the effectiveness of sanitary control measures.

Impact of Covid-19 on nation-wide or regional-wide mobility

A number of studies focused on the impact of Covid-19 and associated measures on human mobility. Hara and Yamaguchi [80] observed a mobility reduction of 20% in Japan, as well as noting that people avoided traveling to densely populated areas. A significant drop in inter-metropolitan mobility was also observed in Finland [81]. A relation between inter-metropolitan mobility and Covid-19 propagation is observed by a study in Brazil [82]. As a matter of fact, mobility changes during the pandemic were observed globally, as shown by a study on 124 countries [83], and the intensity of these changes seems to be related to the development index of the countries, with a more significant mobility reduction in highly developed countries.

Similarly, in France, G. Pullano *et al.* [84] showed that 9 out 10 symptomatic cases of Covid-19 were undetected by the surveillance system in the 7 weeks following the end of the first lockdown in France. This analysis was performed at regional scale and using mobile phone data provided by Google Mobility Trends. In [85], the authors studied the impact and effectiveness of the first lockdown, when exploring different types and duration of social distancing interventions. Mobile phone data in the Ile de France region is used to deduce mobility reduction (one of the studied impacted factors) during the lockdown (70% of reduction observed by the analysis) and to model the effect of sanitary measures. However, all these studies focus on macroscopic mobility data, at a regional or even national level. Badr and Gardner [86] discuss at length the limitations of this country-wide analysis, showing that mobility is not always correlated with virus propagation at this large scale.

Most studies using mobile phone data are based on data coming from Apple and Google, the developers of the two mobile operating systems dominating the market [87]. However, the data provided through these means has a much lower resolution than the mobile phone data collected by the operators. For example, in [87], the authors observe a decrease of 80% in human mobility in the United Kingdom during Spring 2020 using data provided by Apple, but the resolution is country wide.

Impact of Covid-19 at metropolitan scale

A few studies used fine-grained spatial mobile phone data. For example, Li *et al.* [88] study the impact of the pandemics on smartphone usage, showing a decreased user engagement during the outbreak. Zhou *et al.* [89] use mobile phone data from the metropolitan area of Shenzhen to build a virus transmission model and simulate the impact of different scenarios of mobility restrictions. Zanella *et al.* [90] studied the impact of Covid-19 control measures between the beginning of the second wave with the second lockdown and the third wave after the third lockdown in mobile service usage (i.e. apps) in France at communes granularity (i.e. local administrative units). The mobile service, based on the IP session, was identified from the operator mobile data traffic by using Deep Packet Inspection and proprietary traffic classifiers deployed by the operator. They pinpoint the steady evolution of volume usage of some streaming apps like Youtube or Netflix despite different restrictions over the period of study. Whereas for some other streaming apps, like Disney+, volume usage was increasing during the two lockdowns. This shows that, despite some services being in the same category, their usage can differ. The authors also noticed that there was no significant change in terms of consumption of the major part of the mobile service, the change in data traffic during Covid-19 was mainly due to the popular streaming services. They also noted that the various restriction measures have decreased the contribution of urban centers in the total traffic demand. However, the share of these areas became important again when there were relaxations in the restrictions. Finally, Ayan *et al.* [91] use both aggregated and individual mobile phone data covering the city of Rio de Janeiro, in Brazil. Their spatial results focus on information at the level of district. The authors mostly centered their analysis on the unique individual mobility dataset at their disposal, allowing them to distinguish that 45% of the users entirely stopped their mobility during the Brazil lockdown, while 4% of the users still present a very high mobility index.

Romanillos *et al.* [92] used CDR data and signaling data that cover Madrid. Their study aimed to evaluate the distribution of the population on a typical day and depends on the land use of the area. They evaluate the impact of the lockdown

on this urban dynamic and population distribution on the land use. They use the land use of each entity of the study area as defined by official public documents. The authors have shown that activity zones, during periods with restrictions, stop attracting people during the day. Moreover, for residential areas, the number of people during the day is stable, unlike in periods where there are no restrictions, where the number of people during the day decreases.

2.3 Mobile data usage in network configuration

Mobile phone data is the reflection of the activity that takes place in the network. Analysing this data to propose new traffic-aware management and configuration for the same network, to serve current and future needs, are therefore relevant approaches. It is important to note that these solutions do not introduce any bias, unlike when using mobile phone data as human mobility proxy.

The network traffic analysis using mobile data for network configuration and management provides an opportunity to improve network performance, for example to have better Quality of Experience (QoE) of users, resource saving, or to help the evolution of network architectures.

2.3.1 Parameter tuning

We focus here on mobility, represented by user handover (i.e. change of serving cell for an active user), and QoE implications that this can represent. A good management of handover could increase the QoE of the user, especially since these events occur when the user is active on the network, i.e. communicating, for example making a call or exchanging data. A handover that does not perform well can lead to a degradation of the service or even its interruption [93] and therefore decreases the QoE. More detailed discussion on the handover process will be conducted in Appendix A.

Balachandran *et al.* [94] characterize the impact of network characteristics on user web browsing session QoE. To measure the QoE, they used different metrics like the number of pages a user clicked on, which they called session length, whether the user leaves the website after visiting the landing page, called abandonment rate, and incomplete download. To that end, they used UMTS mobile phone data from an US mobile network operator. These metrics are not accessible easily from the network operator perspective, since they are typically monitored from server-side or client-side. The authors had to infer QoE metrics from the collected data. They used URLs in HTTP traces to extract the metrics. From the analysis of the evolution of these metrics, they showed that inter-radio access technology (IRAT) handovers have a strong impact on QoE, for example all sessions with two IRAT handovers are abandoned. However, for the soft handovers this impact is fairly negligible. Decreasing the run time of IRAT handovers can therefore help to reduce the impact on the QoE.

Shafiq *et al.* [95] use the same approach and an equivalent mobile traffic dataset as [94], but they studied the effects on video streaming. They evaluate a vast number of network parameters that could affect the video streaming abandonment. For the case of handovers, all handovers types result in a decrease of user engagement. For soft handover, they showed that they even had a higher abandonment rate.

Ali *et al.* [96] proposed a new handover management solution that considers the QoE perceived by the user. Their objective is to choose the more suitable cell that provides an acceptable QoE for the user. They measured the QoE with two metrics:

(i) the probability of successfully downloading a file, and *(ii)* the file download time for completed downloads. Their proposal is based on a neural network that will learn from the previous measurements given by the network (measured by the user equipment) during a handover and output the effect on the metrics that represent the QoE. However, for the validation, they do not use real data, but a simulation. They showed that, with their solution, there was an improvement compared to the handover based on the signal strength.

Hashmi *et.al* [97] instead use handover as key performance indicator (KPI) to evaluate the QoE offered by the network in real time. To that end, the authors used LTE data of 7,000 cells for over 5 weeks duration and with one hour granularity. They used three QoE KPIs: *(*i) scheduled user throughput, *(ii)* inter-frequency handover success rate, and *(iii)* intra-frequency handover success rate. They consider that these KPIs, in the absence of user-level measurements, provide a fair representation of the average user QoE in the areas covered by the cells. Based on this, the authors proposed a framework using machine learning for real-time user QoE prediction. They tested several machine learning algorithms and showed that the generalized boosted trees (GBN) have the best prediction of the QoE for the metrics on scheduled user equipmet throughput and inter-frequency handover success rate, and less for the intra-frequency handover success rate. They estimated the lack of reliability of the prediction of the latter was due to the fact that it needs some mobility features of the user. The authors showed that the number of inter-frequency handovers influences the user QoE: the higher their number, the lower the QoE. With the scheduled throughput metrics, they indicated that the user QoE is affected by the cell load, and therefore proposed to focus on small cell deployment, and improving channel conditions, especially for users at the cell edge, to improve the user QoE.

2.3.2 Architectural evolutions

The network functions in the mobile architecture were traditionally linked to the underlying infrastructure, hence each function had its own dedicated equipment [98]. However, with the increase of the user demand, and the necessity of flexibility of the network, this kind of architecture is not very adaptable. Virtualizing these network functions is therefore a solution [99]. With virtualization, new paradigms have appeared. Software defined network (SDN), network function virtualisation (NFV) and cloud computing give new ways to use the network functions. The principle of the SDN is to decorrelate the control plane and the data plane. The SDN will manage the control plane, in other words the signaling in the network. A network function includes all operations related to configuration and lifecycle management to provide an end-to-end network service by satisfying the constraints related to this service. NFV is thus the virtualization of these functions, i.e. dissociate these network functions from the corresponding hardware, and hence to be able to run them on virtual machines [100]. These concepts are the basis of slicing approaches, which we will discuss in Chapter 6. Both SDN and NFV can use cloud computing services for the mobile network [101].

The impact on handovers brought by the cloud concept is addressed by Naboulsi *et al.* [102]. The authors stated that, by migrating some network functions from the radio access network (RAN) to the cloud with the Cloud RAN (CRAN) concept, the number of handovers performed decreases. We can deport the Baseband processing unit (BBU), which is traditionally installed in the same place as the remote radio head (RRH), i.e. antenna, therefore a RRH is associated with a BBU and inversely,

to a cloud. The latter is connected to the RRHs by optical fiber link. A handover is executed at the level of the user equipment (UE) when it changes cells and at the level of the BBU without participation of the RRH. Indeed, BBU is the RAN entity that performs the calculations and operations for the execution or not of the handover. With CRAN, several RRHs can be associated with a single BBU. When the load associated to all RRHs exceeds the capacity of the BBU, another BBU can be created. Also, with CRAN, when a UE changes its RRH and if the new one is connected to the same BBU as the RRH it has just left, there will be no handover linked to this mobility, contrary to the classic architecture. A handover linked to mobility can still exist when the target RRH is not connected to the same BBU as the source RRH. However, the CRAN will induce a new type of handover, which is the reconfiguration handover. This occurs when a RRH changes the associated BBU during a reconfiguration. The authors proposed dynamic BBU-RRH mapping methods based on time varying graph describing the behavior of the UE and then RRH clustering. RRH clustering aimed to reduce the number of handovers by using the resulted time varying graph. To evaluate their proposition, they implemented their algorithm using real network data covering the city of Dakar in Senegal. Their results showed that, compared to handover in a classical RAN, their proposal reduces the number of handovers by more than 30%. Thus, we can estimate that there is an improvement of user QoE, because according to previous studies, a decrease in handovers induces an increase of user QoE.

The use of CRAN allows to adapt the number of resources and computing power to the load of the network, and thus to save energy, because all the resources are not permanently deployed. Our work, in the next chapters, will go in this direction. We propose the concept of slicing, using the virtualization of network functions, for the management of handovers. This will reduce resource consumption, and also reduce the potential number of handovers by distinguishing between mobile and static users.

2.4 Conclusion

In this chapter, we discussed that most assessments of the accuracy of mobile data as a proxy for human mobility have been conducted primarily on a large scale. This assessment was based mainly on the use of census and surveys. Afterwards, we have also seen that these mobile data are used in several fields of study, to describe human mobility. Moreover, these data are also used to improve the configuration and management of the network, in the case of mobility.

Chapter 3

Comparison of user presence information from mobile phone and sensor data

3.1 Introduction

Accurate mobility models, providing realistic user presence information at a given space and time, are essential in networking research [103, 104]. These models are generally based on some kind of real data, used to extract important mobility properties. Two main data sources have been used in the literature recently to retrieve user presence and mobility information: mobile phone data collected by cellular operators [105] and presence detection sensor data [106]. The first can cover an important number of users (i.e. millions) over large geographical areas (i.e. country wide), while the second is generally focused on the study of a specific environment (i.e. a neighborhood or a building). However, as we discuss in Chapter 2, mobile phone data have biases, and presence detection sensor data have biases as well. The inherent biases of both these data sources are yet to be fully studied and understood. This raises questions regarding the quality of the mobility models produced using these methodologies.

Presence detection sensors are widely available on the market and they can be easily installed. They are based on different technologies, such as infrared, ultrasonic, magnetic, image processing, or ultra-wide band [107]. Different biases are also encountered in this case. First of all, presence detection sensors have a limited sensing range, which means some users might not be observed, despite being present in the area of study. Second, most presence detection sensors can not distinguish between two different persons, which leads to errors when such sensors are used to count the number of users present in an area, as some of them might be counted twice. Finally, if they are not well calibrated, presence detection sensors can also detect non-human presence of some wild or domestic animals.

In this work, we deploy human motion detection sensors in an area where we also have access to aggregated mobile phone data. We conduct a spatial analysis, where we deploy the sensors in nine different locations for a few weeks within the studied area, and a temporal analysis, where we collect 6 months of data in two specific locations. Since the biases of the two data sources are different, we believe that our results are a strong sign of the accuracy of the two datasets, which cross validate each other. This work was partially published in [108].

The remainder of this chapter continues with the details of the two datasets we used, then we present the metrics we used to compare them. After, we present results related to a sensor deployment campaign covering nine different locations in the studied area for periods of a few weeks. These results are complemented by a study on 6 months of data that are collected in two specific locations.

3.2 Datasets

Our objective is to assess the accuracy of aggregated mobile phone data and that of human presence detection sensor data by cross correlation. For this, we focus on one geographical area, for which we collect both mobile phone data and human presence detection data. We target a suburban area, with few residential buildings and mostly hosting industrial and commercial buildings. The choice of this area was guided by social acceptability constraints: the deployment of sensors to detect human presence was conducted by the city administration and the city population was informed through multiple channels. We note that we did not use any individual data for this study, and that the collection of aggregated data was conducted in cooperation with the local administration.

3.2.1 Mobile phone data

The area of study represents one mobile network cell of a nation-wide mobile operator. For this cell, we have access to aggregated mobile phone data for the period July 2020 - March 2021. This data gives, with a 30 minutes granularity, the number of users entering and exiting the area covered by the cell. The aggregated data is produced by the mobile operator following a methodology that is quite common in the datasets used in the literature [24]. Practically, a user is associated with the most recent serving base station he used. For a given 30 minutes interval, if a user connects to the serving base station covering the area of interest, after being previously associated with a different base station, a user entry in the area is counted. Reversely, if a user previously associated with the area of interest is now observed in a different cell, this counts as an exit from the area. In our study, we do not distinguish between entries and exits. As explained below, this is a consequence of the fact that our sensors do not measure the direction of travel. Therefore, we sum the number of entries in and exits from the area in one single measure: the number of users moving in the area of interest. The granularity of 30 minutes is clearly a limiting factor for the possible resolution of the study. However, most studies using mobile phone data (*e.g.* [12, 15]) use a temporal granularity of 30 minutes or more, which seems to be sufficient to describe the user mobility.

3.2.2 Human presence detection sensor data

For the second dataset, we also deployed motion detection sensors in the area of study. We used the sensors developed by Dahan *et al.* [109] for detecting urban mobility, which were tested and validated in real deployments. The motion detection component of these sensors uses a technology based on passive infrared receiver (PIR) and can detect movements over a distance of up to 7 meters, with a detection angle of 110 degrees. Its output is a binary response when a thermal motion is detected, and it can detect passing-by pedestrians, as well as vehicles. The sensors were deployed

solely on lamp posts situated outdoor. Of course, we note an obvious bias through the fact that a detected vehicle might transport several persons.

These sensors were deployed during two different periods: from July 2020 to September 2020 and from January 2021 to March 2021. In the first period, the sensors were deployed for relatively short time intervals (from a few days to a few weeks) in nine different locations in the area of interest, as shown in Fig. 3.1. During the second period, the sensors were deployed at only two locations, denoted as SJ161 and SJ214 in Fig. 3.1 (these two locations were also continuously monitored during the first time period).

Figure 3.1: The nine locations where sensors were deployed.

We note that the two periods for which we collected sensor data present two significant differences. The first one is obvious, in terms of season: the first collection period covers the summer holidays (with a high impact on this mostly industrial and commercial area), while the second collection period covers winter months. The second major difference is related to the Covid-19 sanitary measures: no particular measure was active during the first period, while a curfew at 7 pm was in place during the second collection period.

3.3 Metrics

In order to compare aggregated mobile phone data and human presence sensor data, we use a series of metrics, detailed in the following. We denote by $\mathcal{M}$ the mobile phone data time series, which can be divided in daily time series $\mathcal{M}_d$, which itself is formed of mobile data mobility measurements, m_i, recorded every 30 minutes. Similarly, for the sensor data, the time series at a given location l, denoted as $\mathcal{S}^l$, is divided in daily time series $\mathcal{S}^l_d$, formed of sensor data measurements aggregated over a 30 minutes interval, s^l_i.

3.3.1 Correlation coefficient

The most obvious metric to compare two time series is their correlation. In this work, we use the Pearson correlation coefficient computed on a daily basis, which indicates the linear correlation between two daily time series. This coefficient, representing the

ratio between the covariance of the two time series and the product of their standard deviations, is defined as follows:

$$\rho_{(\mathcal{M}_d,\mathcal{S}_d^l)} = \frac{cov(\mathcal{M}_d, \mathcal{S}_d^l)}{\sigma_{\mathcal{M}_d} \cdot \sigma_{\mathcal{S}_d^l}}, \tag{3.1}$$

A correlation coefficient close to 1 or -1 implies a strong relationship (positive or negative) between the two time series, whereas a $\rho_{(\mathcal{M}_d,\mathcal{S}_d^l)}$ close to 0 indicates that the two time series are unrelated and very different.

However, simply computing $\rho_{(\mathcal{M}_d,\mathcal{S}_d^l)}$ can not account for the fact that the two data sources might be slightly shifted in time. Indeed, a person entering or exiting the area might not be detected exactly at the same time by the mobile network and the motion detection sensors. For this reason, we also measure the Pearson correlation coefficient between the daily time series with a certain delay shift in one of the two time series. Practically, since our datasets have a 30 minutes granularity, when we introduce a delay of 30 minutes, we align m_i and s_{i-1}, when the delay is 1 hour, we align m_i and s_{i-2}, etc.

We denote the sensor collected time series at location l, shifted with a delay τ, as $\mathcal{S}_d^l(\tau)$. In this case, the Pearson correlation coefficient with a delay τ is defined as:

$$\rho^{\tau}_{(\mathcal{M}_d,\mathcal{S}_d^l(\tau))} = \frac{cov(\mathcal{M}_d, \mathcal{S}_d^l(\tau))}{\sigma_{\mathcal{M}_d} \cdot \sigma_{\mathcal{S}_d^l(\tau)}}. \tag{3.2}$$

Please note that the Pearson correlation coefficient is agnostic of the actual amplitudes of the two time series, hence reducing the bias introduced in the sensor data by vehicles possibly carrying multiple mobile users.

3.3.2 Peaks synchronisation

The detection of peaks in mobile phone data, and in mobility data in general, is a routine task. Peaks are often related to human activities, for example morning and evening mobility peaks are classically associated with commuting [14]. Similarly, detecting the location and the time of the most important mobility peaks in a given area is a common way to detect hotspots [29].

For this reason, we focus on peak synchronisation as a second metric in our study. First of all, we define a peak as a local maximum in a given time series, over a window of predefined size w. In other words, using the mobile phone data time series as an example:

$$P_{\mathcal{M}} = \{i | m_i \geq m_{i-k}, \forall k \in [-w, w]\} \tag{3.3}$$

In a similar way, we can define $P_{\mathcal{S}^l}$ as the peaks found in the sensor data at location l. We can now compute different peak synchronisation metrics, as follows. For each peak $p_m \in P_{\mathcal{M}}$, we compute $\delta_{p_m}(P_{\mathcal{S}^l})$ representing the time difference between p_m and the closest peak in the sensor data time series $\mathcal{S}^l$. Practically, if the peaks in the two time series are perfectly synchronised, $\delta_{p_m}(P_{\mathcal{S}^l}) = 0$. We note that the computation of this time difference between the peaks of two time series is not commutative, so we also compute $\delta_{p_s^l}(P_{\mathcal{M}})$. Finally, we compute $\delta_{p_m}(P_{\mathcal{S}})$, representing the time difference between p_m and the closest peak in all the sensor data time series, covering all the nine sensor locations. We also extract the two highest daily peaks in the different time series, denoted as $H_{\mathcal{M}}$ and $H_{\mathcal{S}^l}$, and compute the time difference between them using an analogous procedure to the one described above.

3.3.3 Start and end of the day

The night-time and day-time behavior is very different in mobile phone data [15]. Detecting home locations and work locations as the predominant user locations during night-time and, respectively, day-time, is standard procedure when working with mobile phone data [30, 38]. We therefore focus on detecting the start of the day (i.e. the beginning of human activities in the studied area) and the end of the day (i.e. the moment when most of the human activity in the area stops) in the two datasets.

To this end, we use the symmetric derivative, which is an approximation of the derivative on time series. Practically, we assume that the derivative of the time series is a good indicator for the start and the end of the day. More precisely, the start of the day is defined as the moment when the derivative has the highest value (we limit the studied interval for this purpose from 3 am to 12 pm), while the end of the day is defined as the moment when the derivative has the lowest value (we limit the studied interval for this purpose from 4 pm to 3 am).

This allows us to compute daily values for the start of the day in mobile phone data (denoted as $B_{\mathcal{M}}$) and in sensor data (denoted as $B_{\mathcal{S}^l}$). Similarly, we identify the end of the day $E_{\mathcal{M}}$ and $E_{\mathcal{S}^l}$. With this, we can compute δ_B and δ_E, the time difference between the start of the day, and respectively the end of the day, in the two data sources.

3.4 Spatial analysis

In this section, we discuss results related to the 3 months deployment in the summer 2020, using nine different sensor locations in the area of study, as shown in Fig. 3.1. To this end, we compare mobile phone data and sensor collected data by using the metrics described in Section 3.3.

3.4.1 Correlation coefficient

Correlation coefficient distribution

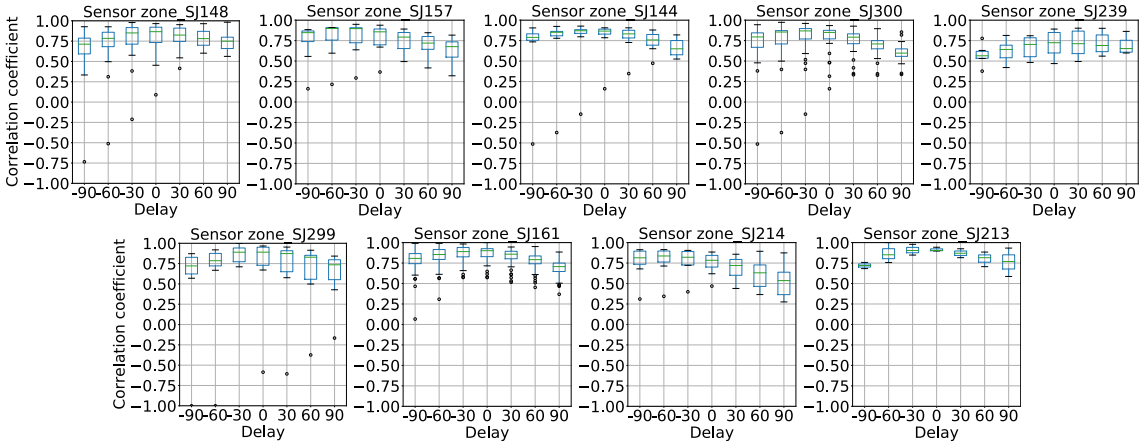


Figure 3.2: Distribution of the Pearson correlation coefficient depending on the delay shift between the two time series.

Fig.3.2 shows, for each of the nine sensor locations, the distribution of their daily correlation coefficient with mobile phone data, depending on the delay shift between

the two time series. We test several values for this delay, from -90 minutes to +90 minutes, with a step of 30 minutes.

We can see from the boxplots that, in general, at every location there is at least one delay shift where the median correlation coefficient is superior to 0.8, a sign of significant correlation between the two data sources. The only exception comes from location SJ239, which shows a much lower Pearson correlation coefficient compared to the other locations. As it can be seen in Fig.3.1, in this case the sensor is deployed in the parking place of a commercial area. Our intuition in this case is that the parking place has slightly different dynamics than the rest of the area of study, as people can enter and exit the area without necessarily visiting the parking place.

We also notice that the highest median correlation coefficient is usually obtained for a delay of -30 or 0 minutes, with specific patterns in SJ239 and SJ214. These fluctuations indicate that, depending on the location, a certain shift exists between the two data series. To better study this phenomenon, Fig.3.3 shows the histogram of the delay shift producing the highest daily correlation in each of the nine sensor locations.

Delay shift distribution

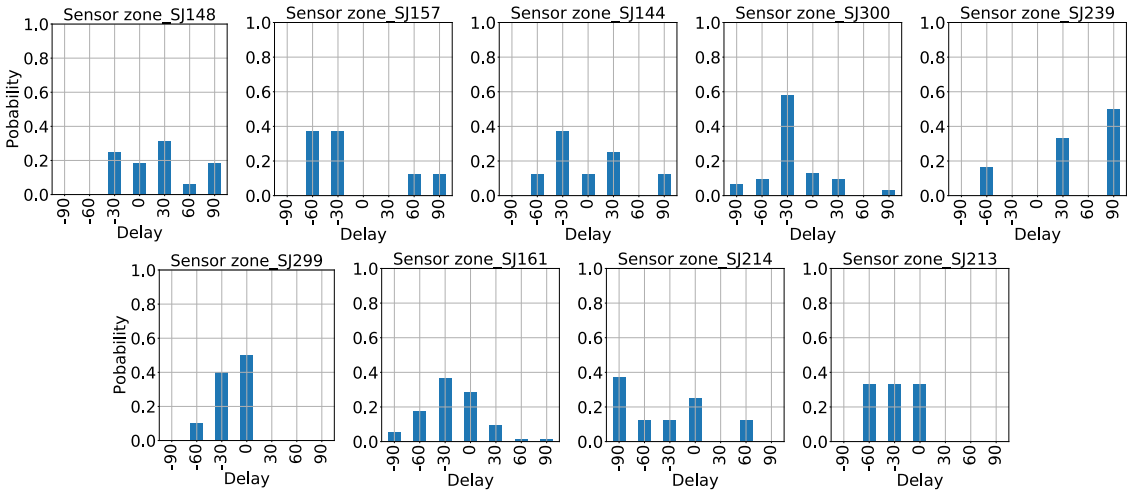


Figure 3.3: Distribution of the daily delay shift between the time series giving the highest Pearson correlation coefficient.

The first observation in Fig.3.3 is that the delay giving the best correlation shows important daily variations. For example, in location SJ148, the delay between the two time series varies from -30 to +90 minutes. Only one location, SJ300, has a delay value (-30 minutes) achieving the highest correlation for more than 50% of the days. The only location where the mobile phone data and the sensor data seem strongly synchronised (i.e. a 0 delay) is SJ299, situated in the south of the area of study. Finally, the specific pattern of the parking place location SJ239 is again visible, with a +90 minutes delay usually giving the best results, but also with days where a totally opposite behavior (-60 minutes delay) is observed.

Overall, our results confirm that the aggregated mobile phone data is strongly correlated with the human mobility measured by the sensors. Yet, a delay often occurs between the two time series. This can be explained by the pre-processing done on the mobile operator side in order to produce the aggregated dataset. Indeed, the mobility of a user can be observed with a certain delay on the mobile network,

since the user is not necessarily active on the mobile network when moving. Finally, some specific properties can be observed at different sensor locations, demonstrating the existence of a spatial diversity. However, it is striking to note the very high correlation between the two time series at practically all the locations, despite this spatial heterogeneity. We consider that this significant correlation allows for a cross correlation of the two data sources, showing that they accurately represent the user presence in the studied area.

3.4.2 Peaks synchronisation

We proceed by comparing the peaks in the mobile phone data and in the sensor data, using the metrics described in Section 3.3. Fig.3.4 shows the distribution of $\delta_{p_m}(P_{\mathcal{S}})$. Practically, for every peak detected in the mobile phone data, we find the closest peak in the sensor data, in any of the nine locations. We can notice once again a strong correlation, with a probability of more than 95% to find a peak in at least one sensor location within less than 1h with respect to mobile phone data. The delay observed in terms of correlation is also visible here, with most of the peaks being shifted by 30 minutes between the two data sources.

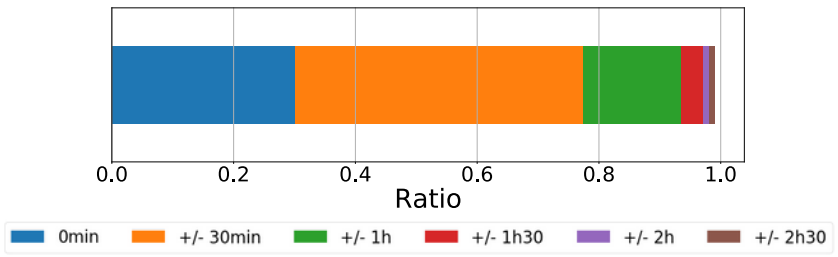


Figure 3.4: Distribution of $\delta_{p_m}(P_{\mathcal{S}})$.

All peaks

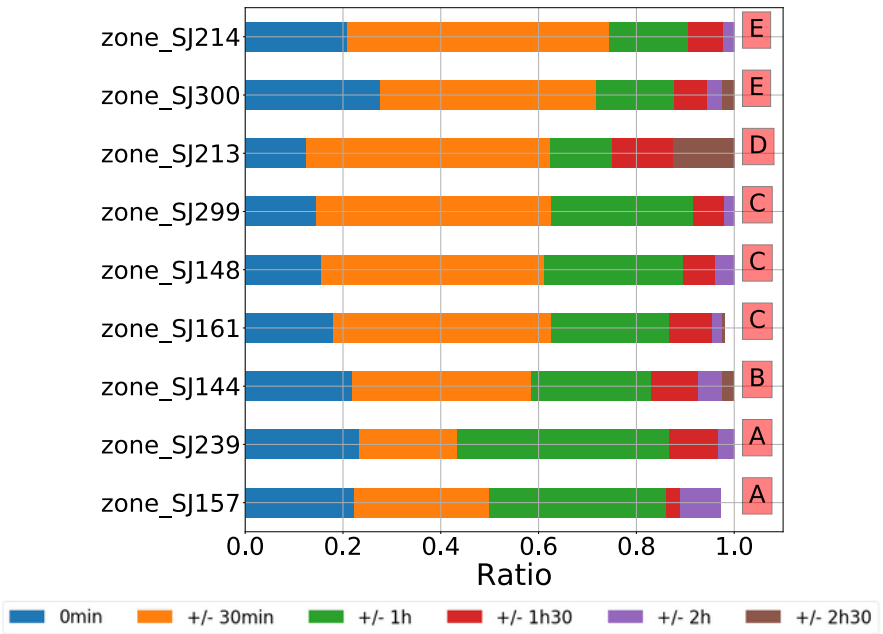


Figure 3.5: Distribution and clustering of $\delta_{p_m}(P_{\mathcal{S}^l})$.

A similar trend can be observed when we analyse the different locations individually, as shown by the distribution of $\delta_{p_m}(P_{\mathcal{S}^l})$ in Fig.3.5. In this case, we also run a clustering algorithm on the results of the nine locations in order to automatically detect locations with similar trends. From a practical point of view, the results at each location are seen as a vector of six elements and a hierarchical clustering algorithm is used to group the similar vectors together. The obtained classes are labeled as A-E in Fig.3.5.

For most of the locations, the probability to find a peak in less than 1h with respect to the mobile phone data is higher than 0.8. The only exception is SJ213, which shows a more reduced peak synchronisation level. For locations clustered in classes B to E, the most likely delay between the peaks is 30 minutes. However, a different pattern can be noticed for locations in class A, where a delay of 1h between the peaks is more probable. With respect to the clustering, the similarity in terms of peak synchronisation behavior does not seem to depend on geographical proximity. In fact, except for SJ148 and SJ161, clustered together in class C, the locations classified together are not relatively close geographically.

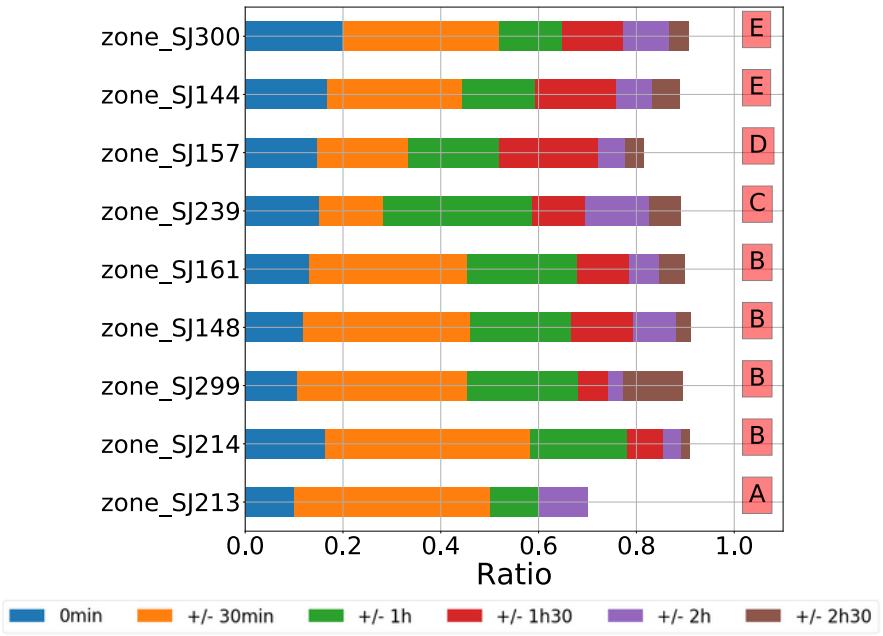


Figure 3.6: Distribution and clustering of $\delta_{p_s^l}(P_{\mathcal{M}})$.

The distribution of $\delta_{p_s^l}(P_{\mathcal{M}})$, shown in Fig.3.6, gives a different view. We recall that, in this case, we first detect the peaks in the sensor data and then find the closest match in the mobile phone data. The synchronisation between the peaks is much weaker in this sense, with a probability of around 0.6 of finding a corresponding peak in less than 1h. For the location SJ213, the delay between the peaks is higher than 2h30 (the limit we set in Fig.3.6) in more than 30% of the cases. These results are a consequence of the sensor data being more dynamic and presenting more peaks than the mobile phone data. Practically, whenever a peak shows up in the aggregated mobile phone data, it is likely that it also appears in the sensor data. However, local peaks in the sensor data are less likely to appear in the more spatially aggregated mobile phone data.

Two most important peaks

Since the number of peaks is different in the two data sources, we focus our analysis on the two most important daily peaks in each dataset. We choose this value because, generally, there are two main daily peaks observed in mobile phone data [15]. As detailed in Section 3.3, for each day we extract $H_{\mathcal{M}}$, the two highest peaks in the mobile phone data. We also extract $H_{\mathcal{S}}$, the two highest peaks in the sensor data, considering all the locations. The time difference between the peaks in $H_{\mathcal{M}}$ and those in $H_{\mathcal{S}}$ is presented in Fig.3.7, where we distinguish the results for the first peak and the second peak in $H_{\mathcal{M}}$.

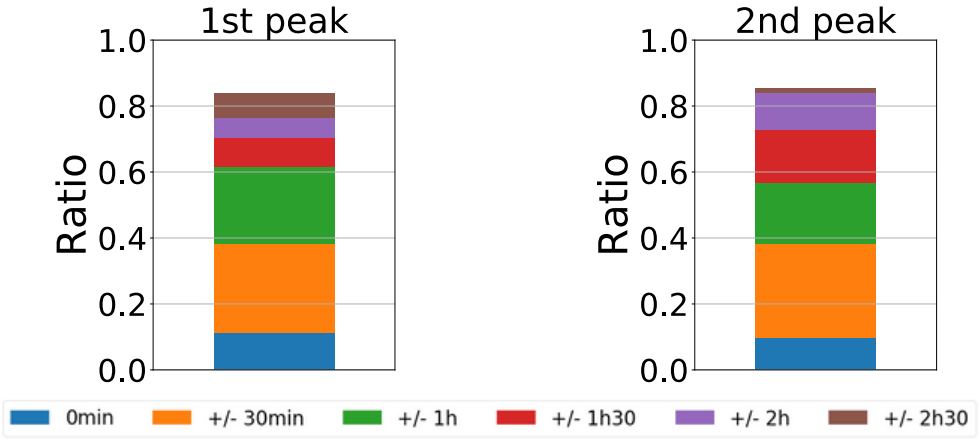


Figure 3.7: Time difference between the peaks in $H_{\mathcal{M}}$ and those in $H_{\mathcal{S}}$.

We notice that the time difference between the highest daily peaks in the two datasets is lower than 1h only 60% of the time. Moreover, almost 20% of the time, this time difference is higher than 2h30. The results are even more surprising when looking at the nine locations individually, in Fig. 3.8. With the exception of SJ213, where the peaks are very often in a 30 minutes range, all the other locations show a rather poor synchronisation with mobile phone data in terms of the most important peaks. The probability to have a time difference lower than 1h for the most important peak is generally below 0.5, while it is a little bit higher for the second most important daily peak. We also run the clustering algorithm on the nine locations based on these results, obtaining four classes (A-D) for the most important daily peak and three classes (1-3) for the second most important one.

These results indicate significant differences between the most important peaks in the mobile phone data and those detected by human presence sensors in the area of study. These peaks are rarely synchronised, as it can be seen by the reduced amount of blue color in Fig.3.7 and in Fig.3.8. Moreover, even when considering a 1h delay between the peaks in the two datasets, it is more likely not to find the peaks detected by the sensors in the mobile phone data. This raises doubts on the results of studies detecting significant peaks in mobile phone data since these peaks are not confirmed by our second data source.

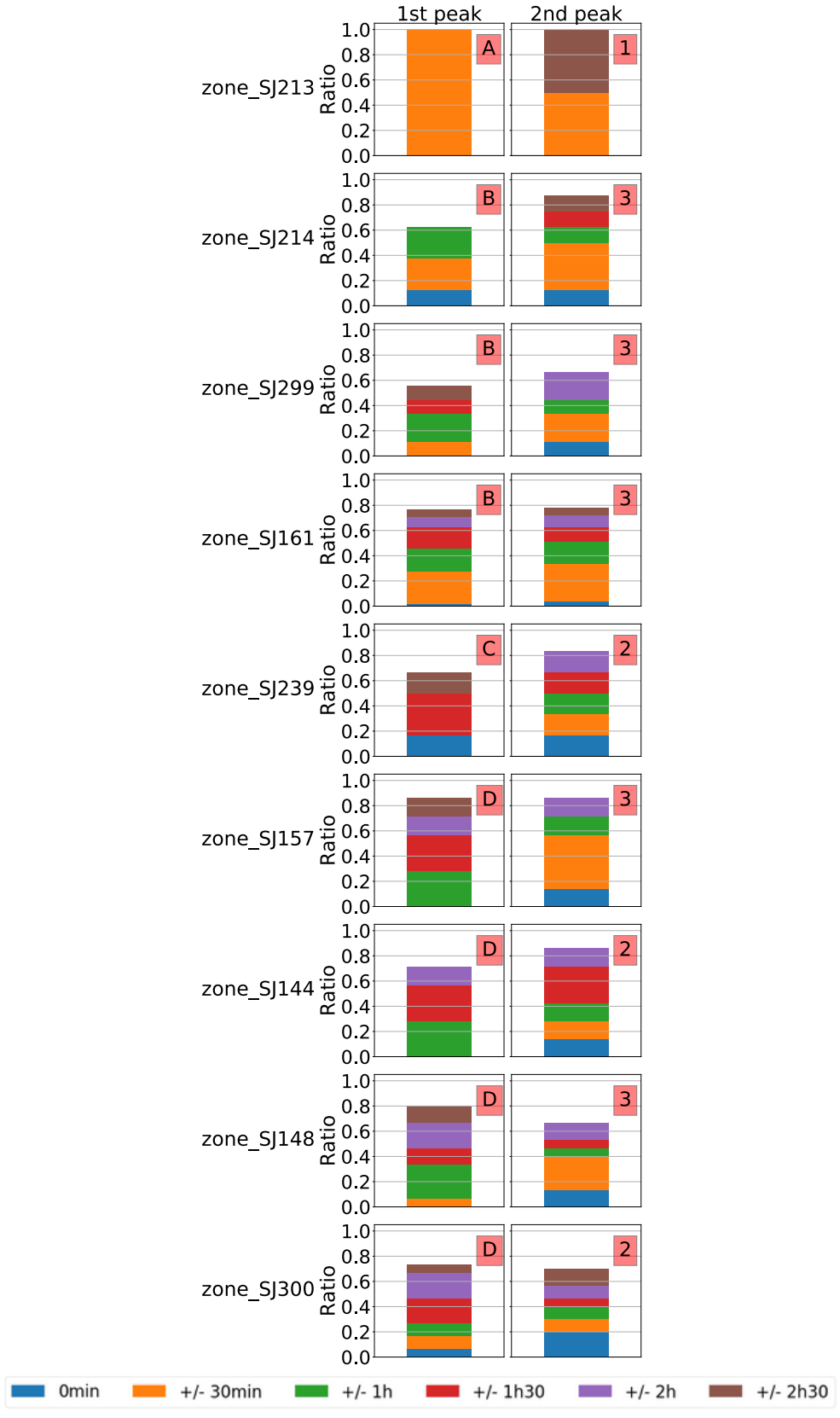


Figure 3.8: Time difference between the peaks in $H_{\mathcal{S}^l}$ and those in $H_{\mathcal{M}}$.

3.4.3 Start and end of the day

We wrap up this spatial analysis by comparing the start of the day and end of the day in the two datasets, following the methodology described in Section 3.3. Fig.3.9

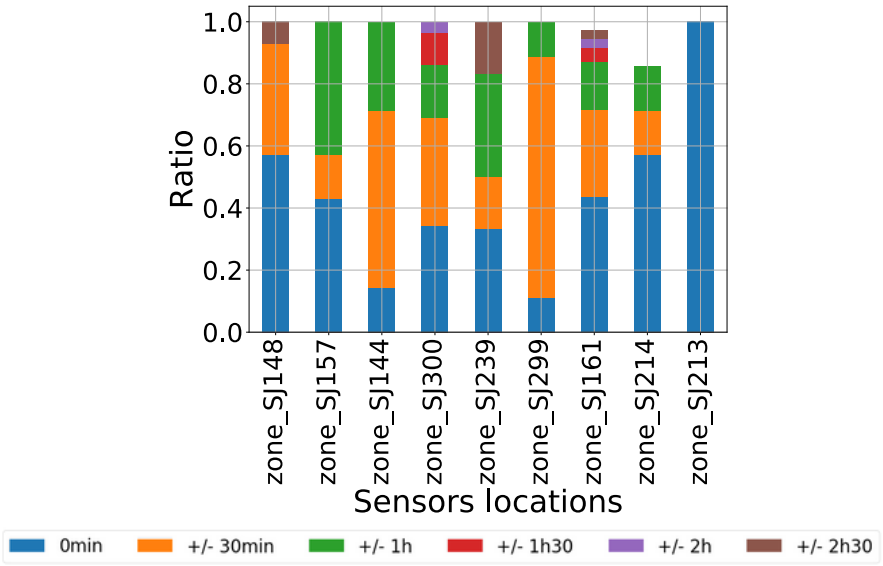


Figure 3.9: Start of the day time difference between mobile data and sensors

shows the synchronisation of the start of the day between the mobile phone data and the sensor data. We can see a good synchronisation, as all the nine locations have at least 80% of their days with an 1h synchronisation with the mobile data. Some locations show even a better synchronisation: SJ148, SJ299 and SJ213 have a 30 minutes synchronisation probability above 80%.

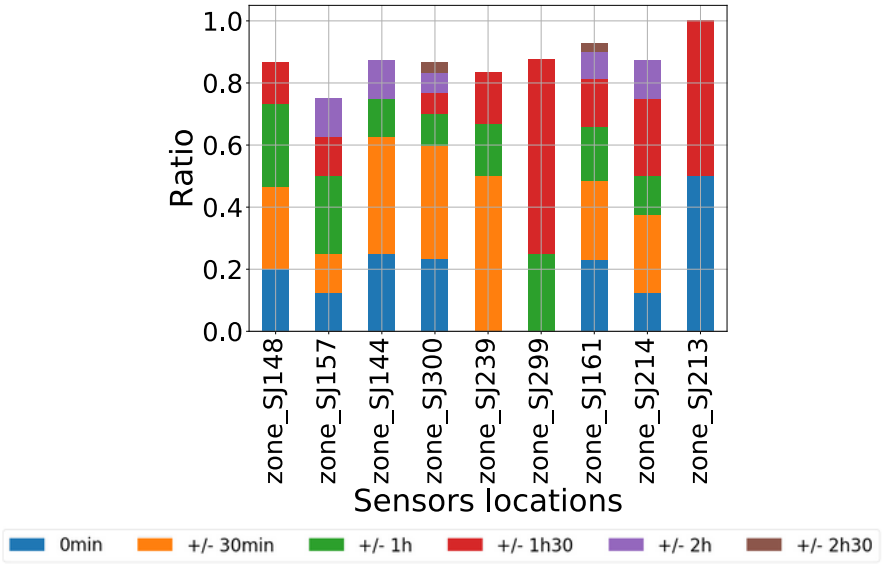


Figure 3.10: End of the day time difference between mobile data and sensors

On the other side, the synchronisation of the end of the day, presented in Fig.3.10, is much more problematic. The end of day synchronisation within an 1h interval has a probability of around 60% for most locations, with some locations, such as SJ299 or SJ157, reaching even lower results. Even considering the 2h30 limit used in our tests, the difference is more significant for around 15% of the days. As we will discuss later, this difference usually comes from a more significant evening activity detected by the sensors, which does not appear in the mobile phone data.

These results confirm an excellent correlation between the two data sources in the morning hours, after accounting for the 30 minutes delay between the datasets that

was observed by all our metrics. However, significant differences are noticed between mobile phone data and sensor data with respect to the end of the day, where the synchronisation between the two data series is weak. Studies such as those focusing on the evening commuting behavior [14] should be aware of these differences.

3.5 Temporal Analysis

In this section, we complement the summer 2020 results discussed in Section 3.4 with 3 months of data collected during winter 2021. For this period, we only collect data for two sensor locations: SJ161 and SJ214. With 6 months of data for these two locations, we conduct an in-depth temporal analysis.

3.5.1 Correlation coefficient

We represent the Pearson correlation coefficient per day of the week, for the two locations and the two time periods, in Fig.3.11. The figure only presents results for Monday, Thursday and Sunday, all the other days (including Saturday) being very similar to the Monday results.

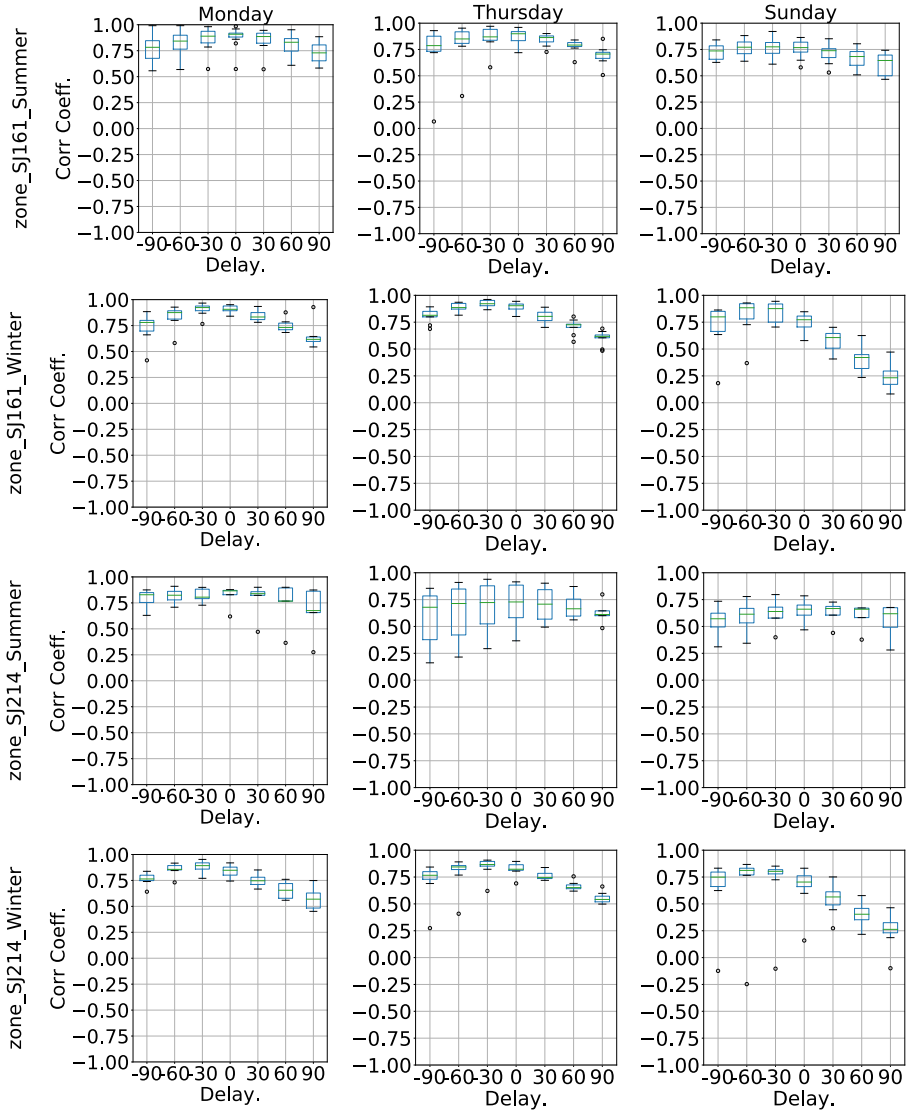


Figure 3.11: Distribution of the Pearson correlation coefficient per day of the week.

The results confirm the very strong correlation between the two time series, with a correlation coefficient superior to 0.9. There are, however, two exceptions to this. Thursdays during the summer period at location SJ214 show a median correlation coefficient below 0.75, a phenomenon no longer observed during the winter period. The second exception is represented by Sundays, when the correlation is much lower compared to the other days, especially during the summer period. We also observe the delay shift of 30 minutes between the two series, already noticed in Section 3.4.

3.5.2 Peaks synchronisation

All peaks

Regarding the peaks, Fig.3.12 and Fig.3.13 show the distribution of $\delta_{p_m}(P_{\mathcal{S}^l})$ per day of the week, for the two studied periods. We only show results for location SJ161, since the results for SJ214 are very similar. The main observation here is that the synchronisation of the peaks degrades during the winter period. For most of the days of the week (the only exception is the Tuesdays), the peak synchronisation probability with a 1h window reduces by 10% or more. We can also notice that, during both periods, the peak synchronisation is weaker towards the end of the week, and especially on Sundays during winter.

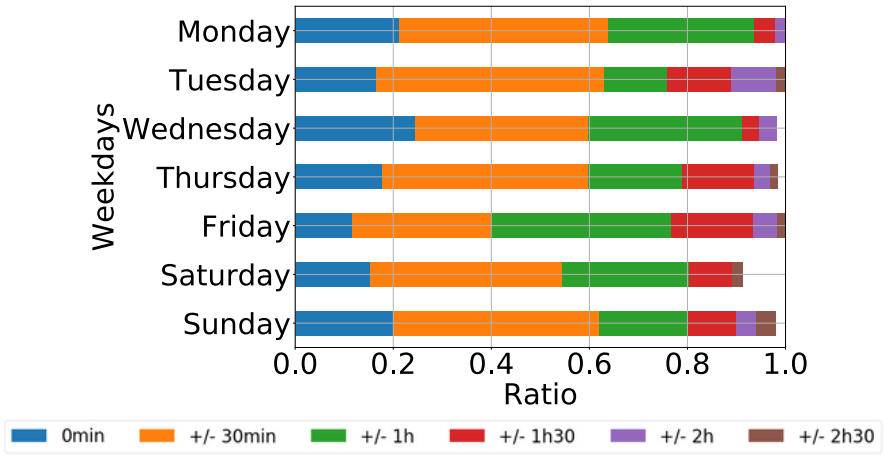


Figure 3.12: Distribution of $\delta_{p_m}(P_{\mathcal{S}^l})$ for location SJ161 during summer.

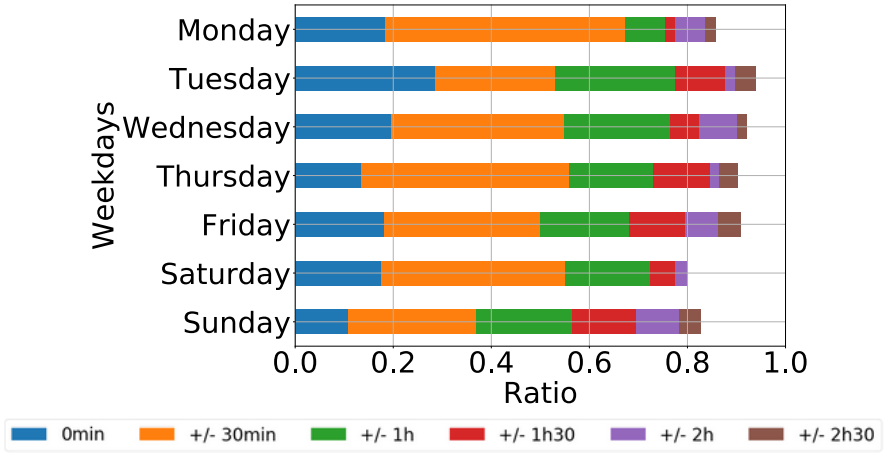


Figure 3.13: Distribution of $\delta_{p_m}(P_{\mathcal{S}^l})$ for location SJ161 during winter.

Most important peaks

Focusing on the most important peaks, Fig.3.14 shows the distribution of the two most significant daily peaks as a function of the hour of the day. We observe that data from the two locations is very consistent, and the peaks in the sensor data are much more uniformly spread over the entire day, while the mobile phone data generally presents peaks in the morning (8:30 am - 9 am) and the evening (5:30 pm - 6 pm). This underlines, once again, the poor correlation of the two data sources in detecting mobility peaks.

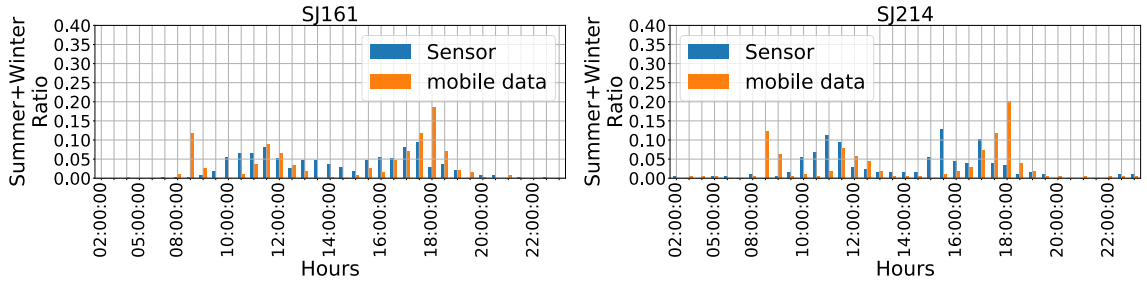


Figure 3.14: Hourly distribution of the most important peaks in mobile phone data and in sensor data.

However, we notice an important exception, presented in Fig.3.15. The distribution of the most significant peaks on Saturdays during the winter period is very similar in the two datasets, as shown in the figure for location SJ161.

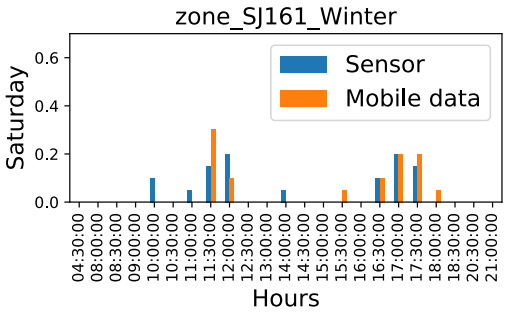


Figure 3.15: Most important daily peak at location SJ161 during the winter period.

3.5.3 Start and end of the day

As discussed in Sec. 3.4, the start of day generally presents a good synchronisation in the two datasets. However, as shown in Fig.3.16, some days of the week have a lower synchronisation than others, e.g. Saturdays for SJ161 and Mondays for SJ214. It is difficult to distinguish a general trend here, but overall the results seem to improve during the winter period.

This improvement is obvious for the end of the day results, in Fig.3.17, where the synchronisation is much more important during the winter, compared with the summer period. The explanation for this behavior is that, during the winter period, a local lockdown at 7 p.m was in place because of the Covid-19 pandemic. Therefore, the evening mobility observed in the sensor data during summer, which had a significant impact on the end of the day results, was no longer present during the winter time.

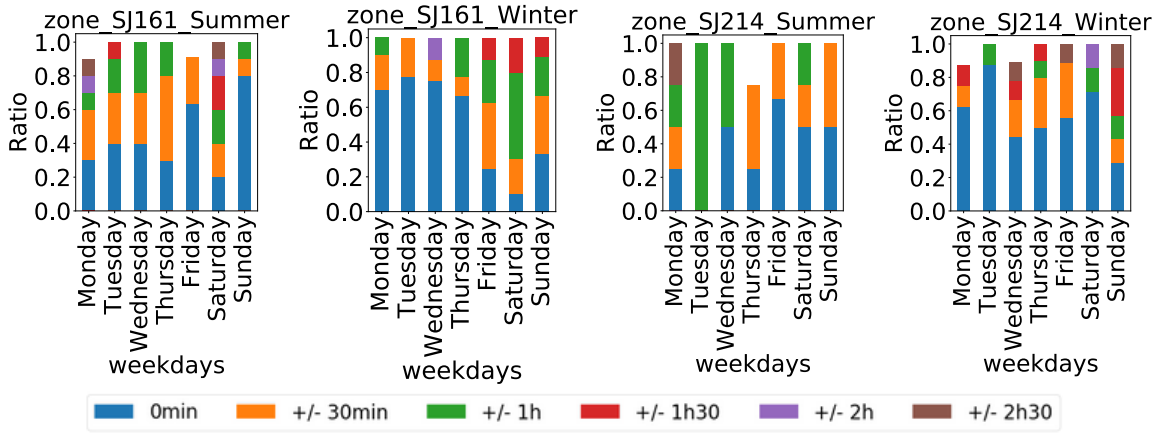


Figure 3.16: Synchronisation of the start of the day depending on the weekday between sensor and mobile data .

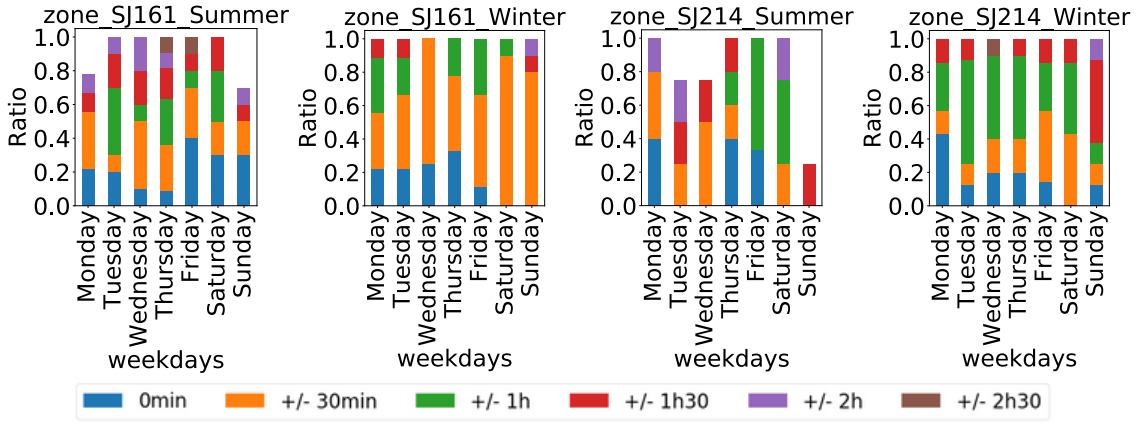


Figure 3.17: Synchronisation of the end of the day depending on the weekday between sensor and mobile data .

3.6 Conclusion

In this Chapter, we assess the accuracy of user presence data used in recent mobility modelling solutions, by comparing aggregated mobile phone data with data collected by motion detection sensors. We define original metrics to compare the two data sources and our results indicate that there is indeed a strong correlation between the two datasets, which validates the realism of the data. However, we also notice some important differences, which need to be accounted for in future studies based on mobile phone data: *i)* there is generally a delay in the range of 30 min between the two time series; *ii)* the most important peaks in the two time series are weakly synchronised, with mobile phone data apparently biased by the user activity on the mobile network; *iii)* the correlation between the two data sources is less important on evenings and on Sundays.

Chapter 4

Covid-19 impact on Parisian mobility behavior

4.1 Introduction

Mobile phone data is nowadays acknowledged as a common tool to study human mobility [110, 111, 112]. As expected, mobile phone data was extensively used in the fight against the Covid-19 pandemic. Some governments, mainly in Asian countries, included operator data in their contact tracing strategy [76, 77, 78] and most countries tried to assess the impact of sanitary measures, such as lockdowns and travel restrictions [85]. Many research teams specialized in epidemics also made extensive use of mobile phone data to parameterize their models [89]. However, most of the available studies focused on mobility at a large spatial scale, between different regions or different cities in a country.

In this chapter, we report a data-driven analysis focused on human mobility during the Covid-19 pandemic on a metropolitan scale. More precisely, we study the Paris metropolitan area and investigate how mobility patterns at an urban scale were affected by the pandemic, and especially by the harsh lockdown conditions enforced from March 17, 2020, to May 11, 2020, in France, as well as the two less strict lockdowns from October 30, 2020 to December 15, 2020 and from April 3, 2021 to May 3, 2021. These sanitary measures forced drastic changes on population mobility and the dataset we explore in this work allows us to quantify their impact.

The dataset, collected by SFR, a major French mobile operator, temporally covers four different contexts, describing a *"usual"*, a *"radical lockdown"*, *"less strict lockdown"* and *"light lockdown and school holiday"* urban life. From a human mobility study perspective, this represents a rare opportunity, as a radical change in population mobility habits of an important EU capital can not be planed for research purposes. Note that the literature on human mobility focused on very big planned events, such as the Olympic Games or the Football World Cup, which induce flow increase in some particular areas of the city (e.g. airports, touristic spots, stadiums, and sports centers), while the large part of the local population pursues a "normal" urban life, with maybe an adaptation of some transit zones. On the other hand, when crisis situations are studied, it is usually too late to retrieve "normal" mobility information about the area of interest. In our case, the entire population is impacted and the dataset also covers the period immediately before the lockdown, allowing for an original case study.

We adapt one of the frameworks presented earlier in Chapter 2, for automatic land

use inference from mobile phone data [15], we classify each region in Paris as either *residential*, *activity*, or *other* area. This allows us to observe how the usage of any given area changes once the lockdowns are established. We define an original metric, based on the so-called signature of the base stations, which allows us to analyze the population mobility in a given geographical area. This work was partially published in [113].

This chapter is organized as follows. First, we present the dataset we use in this chapter. Then we study the Paris land use. Finally, we study the epidemic importance of zones based on the mobility peaks present in their signatures.

4.2 Data

For business (e.g., billing, network management purposes) and legal reasons, mobile operators log events generated during mobile communications. Practically, for each voice call, text message, and mobile data traffic session, as well as in some user mobility situations (known as location area updates), the operator collects a series of network signalling events [24]. These events are associated with individual mobile devices and are time-stamped and geo-referenced. In their most basic form, the locations of the geo-referenced events correspond to the cell towers where the signalling event was generated, which makes mobile network data an obvious proxy for the trajectories of the mobile network subscribers.

In this chapter, we use a dataset built from call detail records (CDR) [114] and network signaling data (NSD) [8], provided by the French telecom operator SFR. This dataset is already pre-processed by the operator, representing an aggregation of mobile network data of all the mobile phone subscribers of SFR and covering 20% of the entire population of the studied area. Note that we look for a relative importance comparison among zones, which can be perfectly obtained with this available 20% of the entire population. A more complete population coverage would require data availability from the four French cellular operators.

The original network data collected by the operator contain sensitive information regarding the users. Therefore, to avoid subscriber identification and to ensure a proper level of privacy, the provided dataset consists of user counters, aggregated both in time and space. We note that the European privacy regulations consider as personal data any individual trajectory data, such as those used in [91], even when the users' identifiers are pseudonymized.

4.2.1 INSEE IRIS

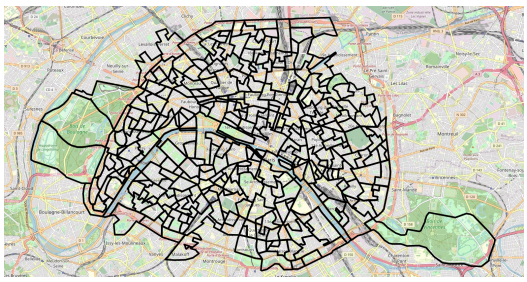
More formally, we observe a large metropolitan area $\mathcal{A}$ in the French region Ile de France, which basically comprises the Paris large metropolitan area, divided into a number of geographical zones $a \in \mathcal{A}$. For privacy reasons, the mobile operator cells, which represent the finest granularity achievable by the operator, were aggregated based on the official units for statistical information (also known as IRIS zones), defined by the French National Institute of Statistics and Economic Studies (INSEE). The IRIS zones must respect geographic and demographic criteria and have borders that are clearly identifiable and stable in the long term[1]. IRIS zones offer the most sophisticated tool to date to describe the internal structure of more than 1,900 municipalities in France with at least 5,000 inhabitants. Even this level of granularity

[1] `https://www.insee.fr/en/metadonnees/definition/c1523`

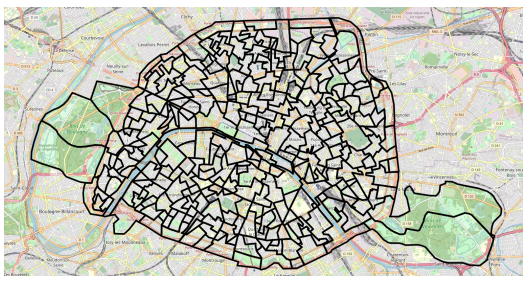
is too fine in some cases, as some IRIS contain very few active users who could be identified this way. Therefore, some IRIS are grouped together by the operator to ensure a proper level of anonymity.

4.2.2 SFR zones

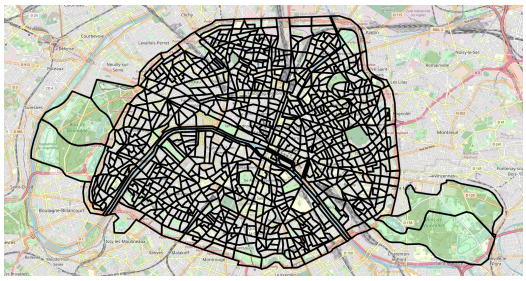
Practically, we have two datasets, the first describes user presence in a geographical area (also called attendance) for the periods before the first lockdown and during the first lockdown. The second describes the attendance during the second and the third lockdowns. The datasets cover the city of Paris (i.e., the French department with code 75), for a total area of 93.76 $\texttt{km}^2$ for the first dataset and 105.44 $\texttt{km}^2$ for the second dataset. The first dataset does not cover some zones in northeast, east and small part of south of Paris. There are 992 (only 911 are used in the first dataset) INSEE IRIS zones in the city of Paris, which are further aggregated into 326 zones in the first SFR dataset and 301 zones in the second. To achieve this level of aggregation, INSEE IRIS can be merged into groups with size from 2 to 16 INSEE IRIS for the first dataset (an average merging of 2.9 INSEE IRIS) and from 2 to 11 INSEE IRIS for the second dataset (an average merging of 3.3 INSEE IRIS). The higher average merging INSEE IRIS for the second dataset explains the fact that we have fewer zone despite a larger study area compared to the first dataset. The difference in the zones given to us is due to a slight modification of the mobile network configuration in Paris. For example, a new configuration of antenna orientation or power settings has an effect on the mapping of cells to the INSEE IRIS.



(a) SFR zones 1st dataset.



(b) SFR zones 2nd dataset.



(c) INSEE IRIS.

Figure 4.1: Comparison of SFR zones with INSEE IRIS

Fig.4.1 shows the division provided by SFR denoted hereafter as *SFR-IRIS-1st* and *SFR-IRIS-2nd* (which we will also call "zones" thereafter) and the original INSEE IRIS fragmentation of Paris. In the central metropolitan area, the SFR-IRIS-1st and SFR-IRIS-2nd zones are smaller, due to their high user attendance, when compared with zones covering parks or more humanly sparse zones.

4.3 Paris land use

An urban area is not used by its population in a uniform way. Depending on the settlements, arrangements, and inputs the population brings to a given area, its land use is different. Typically, the most basic division of an urban area is *residential areas* and *activity areas*. *Activity* areas can be further divided into *office areas*, *leisure areas*, *educational areas*, etc. Mobile phone data have successfully been used in the past as a proxy for urban land use [66, 58]. Therefore, in this section, we exploit the available dataset to characterize the land use before and during the first lockdown period, trying to assess the impact on land use of these early sanitary measures, and the impact of less strict sanitary measures of the second and the third lockdown. It is important to note that other methods (e.g. official surveys) could be used to characterize the land use during a normal period (i.e. before the lockdown). However, we are not aware of any other method that would allow us to characterize the land use in an exceptional period, such as during the lockdowns. As a matter of fact, the municipality surveys imply significant costs and mobilize numerous people. For this reason, the Paris municipality only conducts mobility surveys once per decade, and this is in fact one of the highest frequencies at a global scale.

To detect *residential* and *activity* areas in a given urban region $\mathcal{A}$, we first need to classify its different subdivisions, trying to find recurrent and a typical patterns. To do this, we associate to each zone a in $\mathcal{A}$ a *signature*, defined as the time series of the user attendance p_a^t. Fig.4.2 shows an example of a signature. The x-axis represents

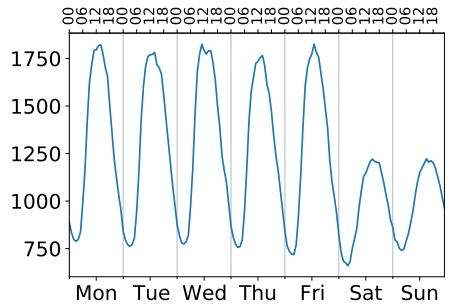


Figure 4.2: Signature example

the days of the week, with the hours evolution shown at the top of the figure. And the y-axis represents the user attendance.

By classifying the different signatures, we can detect the major types of land use in $\mathcal{A}$, mapped to the most significant classes obtained, as well as zones with a unique and specific land use, which remain outside the major classes obtained by the classification algorithm. This methodology is further detailed in [15], where it shows excellent performance in terms of accuracy when compared with ground truth information. Our methodology can be summarized through the following steps, further detailed below:

- We associate four signatures to each area a: one signature for the period before the lockdown and one signature for each of the three lockdown periods.
- We use the Pearson correlation coefficient to measure the similarity between any two signatures in the dataset.
- By using an unsupervised machine learning approach, we classify the signatures

in the dataset. Each class has *a characteristic signature*, obtained as the mean of all the signatures in that class.

- We label each class of signatures obtained in the previous step using information from municipality surveys produced before the lockdown.

4.3.1 Signature definition

Similarly to [15], we build a *typical week* signature. However, a particularity of our study is that each area a has four associated signatures, one for the typical week before the lockdown and one for the typical week during each of the lockdown periods. Practically, the typical week signatures we use represent the mean of the two weeks of data we have before and during the lockdowns, respectively. The signatures are normalized using a standard-score approach, in order to tackle the fact that the actual value of the user attendance p_t^a depends on multiple factors, such as the size of the area a or the smartphone adoption rate in the given area. This also allows comparing the signatures obtained before and during the three lockdowns, despite the significant difference in terms of user attendance. Practically, the normalization allows us to consider only the trend of each signature (i.e., the normalized typical median/mean week), which is a more relevant feature for signature classification.

4.3.2 Agglomerative hierarchical clustering

To detect the major types of land use in the studied area, we automatically classify the signatures by using an agglomerative hierarchical clustering (AHC) algorithm [115], which is a non-parametric and unsupervised method. As an unsupervised method, the classification is done directly using the test data, on which we have no prior knowledge of classes. More precisely, we use the average linkage method, also known as Unweighted Pair Group Method Arithmetic Mean (UPGMA) algorithm [116], which means that each newly formed class will be represented by the mean of its elements.

As well known in the literature, the UPGMA algorithm requires the use of a *stopping rule* to select the best level of aggregation in the clustering process to be used, i.e., the best clusters and their composition. Practically, the hierarchical algorithm builds all possibilities of clusters by fusing similar classes in an iterative manner. At the beginning, we start with each element in a separate class, then the two most similar elements or classes are grouped together, until only one single class is obtained. A dendrogram usually shows a graphical information of all clustering possibilities.

With respect to the stopping rule, [15] proposed a composed metric mixing seven different indices from the literature, as discussed in Chapter 2. However, these classical stopping indices were designed for totally different applications than ours, and it was difficult to find the intuition behind them. Therefore, we adapted the framework designed in [15], by using a more intuitive stopping rule, based on the Pearson correlation coefficient. Since our objective is to cluster together similar signatures, represented as time series, the Pearson correlation coefficient of two time series is a metric easy to understand and interpret. Therefore, at each step in the UPGMA algorithm, we group together the two clusters with the largest Pearson correlation coefficient and we stop the clustering once the correlation coefficient drops below a certain threshold. Since a correlation coefficient above 0.7 is usually considered to

indicate highly correlated variables, and after conducting a series of tests that demonstrate that the obtained clusters are well balanced (i.e. not one single highly dominant class), we decided to use a correlation coefficient threshold of 0.7 in the rest of this Chapter.

4.3.3 Labelling of zones

Using the UPGMA clustering algorithm, we obtain a total of 65 classes. Nonetheless, only a few of these classes contain the majority of elements. For instance, before the lockdown, three classes cover 81.5% of the total signatures, during the first lockdown, three classes contain 93.5% of the signatures, during the second lockdown, three classes contain 95.6% of the signature, and during the third lockdown, three classes contain 96% of the signature, the rest being small classes. The two most important classes are present before and during the three lockdowns. From these classes we obtain the type of a zone, as discussed below.

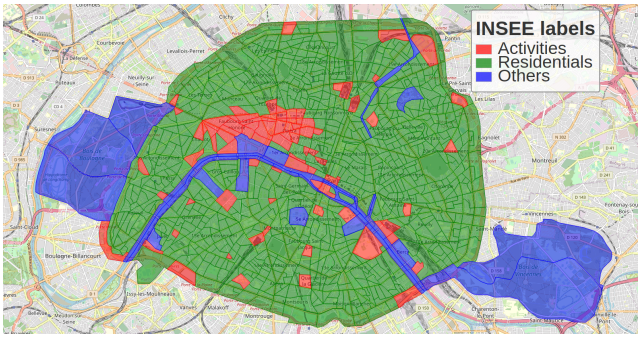


Figure 4.3: Map visualisation of INSEE class labels

Labelling method

To label the classes using known land use categories, we use the INSEE class labels as a reference. Based on census data, INSEE defines three different types of labels: activity, residential and diverse. IRIS zones labeled as residential have a population that is generally between 1,800 and 5,000. In terms of habitat types, they are homogeneous and their limits are based on the major cuts in the urban fabric (main roads, railways, waterways, etc.). Activity IRIS have more than 1,000 employees and have at least twice as many salaried jobs as the resident population. Finally, the diverse category covers large, specific areas that are sparsely inhabited and have a large surface area (leisure parks, port areas, forests, etc.).

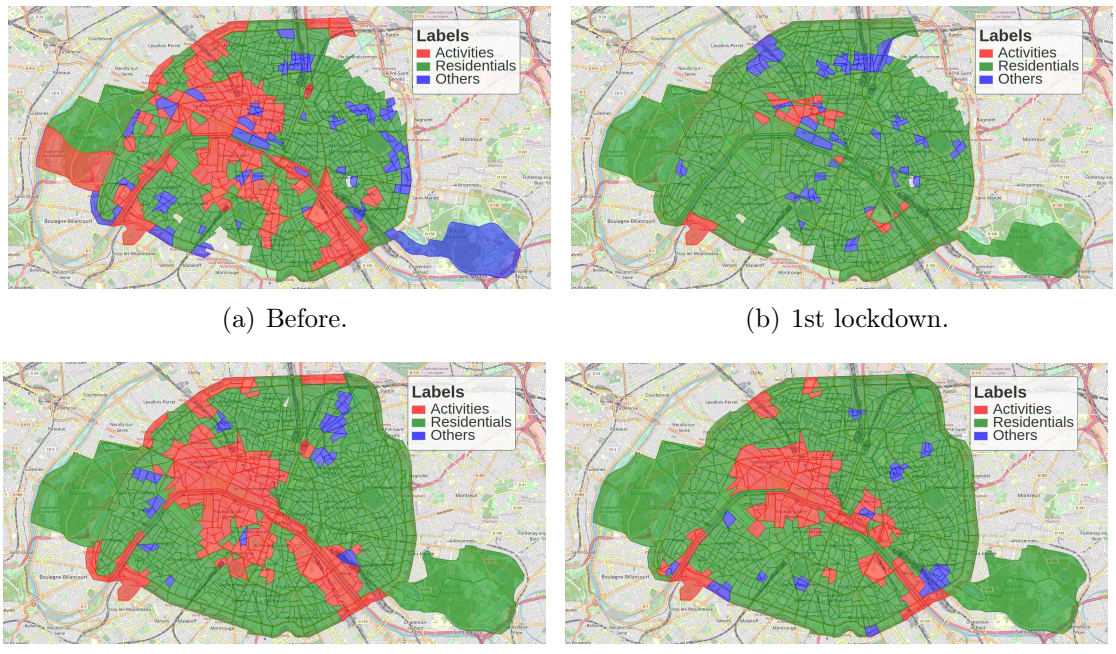
In Fig.4.3, we can see that *residential* IRIS in green are predominant in the INSEE labeling. This corresponds to about 87% of IRIS in Paris. *Activity* IRIS, in red, are less prevalent, about 9%. They are mainly in the west center and occupied by main train stations, while some other small IRIS are scattered all over Paris. Then, diverse IRISs (denoted as *others* in Fig.4.3), in blue, cover two large Parisian parks, the Seine River and other smaller parks.

In order to label mobile phone datasets in our study, we superimpose the classes obtained with the INSEE labels to those obtained by our clustering approach. A class will be labeled as *activity* or *residential* if its spatial occupation mainly corresponds to the corresponding INSEE label. For the classes covering very few elements, with an outlier behavior, we label them together as *others*. It is worth mentioning that

the methodology developed in [15], which we adopt in our study, has been shown to provide better land use information than official census data. We also recall that, for privacy compliance, some SFR-IRIS-1st and SFR-IRIS-2nd zones are the aggregation of several INSEE IRIS. Therefore, we are not looking for a perfect mapping between INSEE data and our automatically obtained classes. As a matter of fact, as discussed below, the *activity* and *others* classes are over-represented in our classification method with respect to INSEE data. However, to better compare the spatial distribution of different label along the different periods (before the 1st lockdown and the three lockdowns), in Fig.4.4 we use the correspondent INSEE IRIS for SFR-IRIS-1st and SFR-IRIS-2nd zones. To summarize, following this methodology on our dataset, we obtain three main labels, as shown in Fig.4.3, which are in line with the three classes defined by INSEE.

In more detail, we proceed as follows. First, all the classes with less than 5 elements (i.e., signatures) are grouped together as *others*. Next, using the classes identified before the lockdown as a reference, we label the *activity* and *residential* areas in Fig.4.4 using the INSEE labels as reference. The red cluster in Fig.4.4(a) covers 65% of the INSEE activity areas, therefore we label it as *activity*. It is important to note that the activity cluster detected by our method is larger than the one defined by INSEE (which is based on simple demographic statistics) and covers some zones denoted as residential or diverse by INSEE. This can be explained by the mixed land-usage of many zones in Paris, which combine housing and commercial and office activities. The other important class before the lockdown, depicted in green in Fig.4.4(a), represents for 73% zones considered as residential by the INSEE classification. We, therefore, label this class as *residential*.

Label distribution



(a) Before. (b) 1st lockdown.

(c) 2nd lockdown. (d) 3rd lockdown.

Figure 4.4: IRIS labels before and during the three lockdowns.

During the period before the first lockdown, we see that the *activities* zones are concentrated in the centre of Paris. This area has a high concentration of companies

and businesses. We also see in the southeast of Paris an activity area. These two areas are connected by almost continuous zones of activity.

For the periods during the three lockdowns, shown in Fig.4.4(b), 4.4(c) and 4.4(d), we apply the same labels previously defined. It is important to note that we do not have any ground truth information regarding the land use behavior during the three lockdowns. This underlines the interest in our methodology, where we cluster together the signatures before and during the three lockdowns, which allows us to directly observe the change in land use once the lockdowns are established.

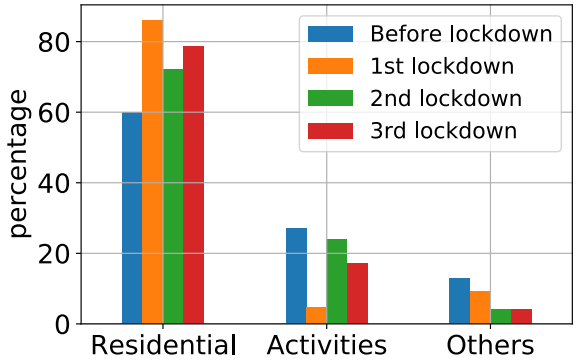


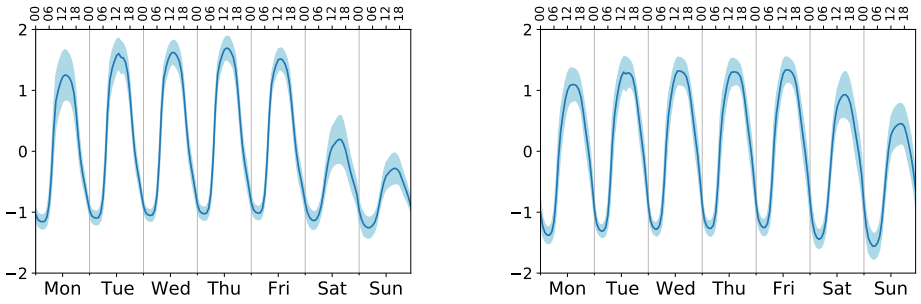
Figure 4.5: Evolution of the distribution of the different labels according to the study periods

In Fig.4.4(b), we can see that during the first lockdown, the majority of Paris is labeled *residential*, we have very few zones labeled *activity*. While for the two others lockdowns in Figure 4.4(c) and 4.4(d), the activity labeled areas are more important, quite comparable to the period before the first lockdown, and cover the same areas as in Fig.4.4(a). Indeed, these two lockdowns were less strict than the first one. However, we can see that, for the third lockdown, the labeled activity area is less important than for the second. This can be explained by the week of school holidays that falls in our study period. Fig.4.5 shows precisely this evolution of the different labels according to the different periods. It shows that the SFR-IRIS (or zones) labeled as *residential* have a significant increase between the period before the first lockdown and the first lockdown, from 59% of the study area to 86%. All of this is due mainly to the decrease in activity caused by the lockdown. Zones labeled as *activity* decreased from 27% to 4%. The percentage of zones labeled as *others* also decreased, some of them becoming *residential* during the first lockdown. During the second lockdown, we can see that the percentage of zones labeled as *residential* decreased to 72%. The zones labeled as *activity* increased to 24%. The percentage of zones labeled as *others* still decreased. For the third lockdown, the percentage of zones labeled as *residential* increased again to 78% and zones labeled as *activity* decreased to 17%. These allow a quantification of the expected impact of the strict first lockdown imposed by the French government on the population: most activities stopped and the population stayed at home, transforming the entire city of Paris into a big residential area. Then, we observe a restart of the activities, which we can see during the second lockdown which is less strict than the first one. For the third lockdown, we see a decrease in activity compared to the second lockdown. This can be explained by the presence of school holidays.

4.3.4 Characteristic signature and label evolution

Characteristic signature

Each label has an associated characteristic signature, defining the general trend of user attendance for the corresponding zones of that label. Fig.4.6 shows the main characteristic signatures in the dataset, namely those associated with the *activity* and *residential* labels. Fig.4.6(b) shows the characteristic signature for the residential class. User attendance is fairly the same all the week even though there is a small decrease on Sunday. The 24h pattern, with increased communication during the day and reduced communication at night is clearly visible in both characteristic signatures. However, when comparing the *activity* and *residential* signatures, we notice that the low level of the time series is reached later in *residential* areas (the drop happens when people go to sleep) than in *activity* areas (the drop happens when people leave the activity area).



(a) Characteristic signature - *Activity*. (b) Characteristic signature - *Residential*.

Figure 4.6: Characteristic signatures for *activity* and *residential* zones.

As shown in Fig.4.5, almost all SFR-IRIS labels turned into *residential* during the first lockdown. This is indeed the expected behavior following the sanitary measures taken by the government. However, some of the zones are still classified as *activity*, some join the others class, and we even notice a case where a former *residential* zone becomes *activity* area. In the following, we analyze the evolution of labels before and during the three lockdowns.

Label evolution

Table 4.1 shows the evolution of the different labels according to the different study periods. We have the percentage of each type of evolution. For example, the zones labeled *residential* during the four periods represent almost 51% of the zones present on the four periods. This is the dominant evolution. The second most important evolution of label is for the areas that remained labeled as *activity*, except during the first lockdown when they were labeled as *residential*. These areas represent about 7.3% of the areas present over the four study periods. Another important evolution is that the areas labeled as *activity* before the first lockdown, then labeled as *residential* during the first lockdown, then labeled *activity* again during the second lockdown, then during the third lockdown labeled as *residential* again('A-R-A-R'). This evolution represents approximately 5% of the zones present during the four periods. Another important evolution of the labels is the areas labeled *activity* before the

first lockdown and then for the other three periods is labeled *residential* ('A-R-R-R'). This evolution represents about 6.6% of the present areas during the four periods.

Label before	Label 1st	Label 2nd	Label 3rd	Percentage
Residential	Residential	Residential	Residential	50.94
			Others	3.09
			Activity	1.87
		Activity	Activity	1.43
			Residential	3.42
		Others	Residential	3.31
	Activity	Activity	Activity	0.11
	Others	Activity	Activity	0.11
		Residential	Activity	0.11
			Residential	2.43
Activity	Activity	Activity	Activity	2.21
		Residential	Residential	0.11
	Others	Activity	Activity	0.33
			Residential	0.33
		Residential	Residential	0.44
	Residential	Activity	Activity	7.28
			Residential	5.07
		Others	Residential	0.44
		Residential	Activity	1.10
			Others	0.11
			Residential	6.62
Others	Activity	Activity	Activity	0.22
	Others	Activity	Activity	0.88
		Residential	Others	0.77
			Residential	1.54
	Residential	Activity	Activity	0.33
			Residential	0.22
		Residential	Others	0.33
			Residential	4.85

Table 4.1: Percentage of each possible evolution of label from the period before the first lockdown to the third lockdown

As mentioned earlier, the second dataset covers areas that were not covered by the first dataset. These areas can be excluded from the comparisons. However, it is interesting to see that only about 8% of the areas change their label between the second and third lockdown, as shown in Table 4.2. They go from *others* to *residential*. However, this evolution represents only about 3.7% of the areas in common of the four periods.

It can also be interesting to have the proportion of each type of evolution by knowing the label of the zones before the first lockdown.

Label_2nd	Label_3rd	Percentage
Activity	Activity	1.18
Others	Residential	8.23
Residential	Residential	90.59

Table 4.2: Percentage of each possible evolution of label of zones that only exist in the second dataset

Residential. Table 4.3 shows the probability of each label evolution knowing that during the period before the first lockdown, an area is labeled *residential*. The probability of remaining labeled as *residential* during the three lockdowns is about 76%. We can also see that the probability that an area, during the first lockdown, is different from *residential* is approximately 4.12%. We will look later at what these different zones correspond to.

Label before	Label 1st	Label 2nd	Label 3rd	Percentage
Residential	Residential	Residential	Residential	76.24
			Others	4.62
			Activity	2.80
		Activity	Activity	2.14
			Residential	5.11
		Others	Residential	4.95
	Activity	Activity	Activity	0.17
	Others	Activity	Activity	0.17
		Residential	Activity	0.17
			Residential	3.63

Table 4.3: Probability of evolution of label knowing that before the first lockdown zones are labeled as *residential*

Activity. Table 4.4 shows the probability of each label evolution knowing that during the period before the first lockdown, an area is labeled *activity*.

We see that the probability that, during the first lockdown, the areas are labeled as *residential* is approximately 85.7%. The probability that the zones remain labeled *activity* during the 3 lockdowns is not negligible, it is about 9.1%.

Label before	Label 1st	Label 2nd	Label 3rd	Percentage
Activity	Activity	Activity	Activity	9.17
		Residential	Residential	0.46
	Others	Activity	Activity	1.38
			Residential	1.38
		Residential	Residential	1.83
	Residential	Activity	Activity	30.28
			Residential	21.10
		Others	Residential	1.83
		Residential	Activity	4.59
			Others	0.46
			Residential	27.52

Table 4.4: Probability of each evolution of label knwowing that before the first lockdown zones are labeled as *activity*

Others. Table.4.5 shows the probability of each label evolution knowing that during the period before the first lockdown, an area is labeled *others*.

Label before	Label 1st	Label 2nd	Label 3rd	Percentage
Others	Activity	Activity	Activity	2.41
	Others	Activity	Activity	9.64
		Residential	Others	8.43
			Residential	16.87
	Residential	Activity	Activity	3.61
			Residential	2.41
		Residential	Others	3.62
			Residential	53.01

Table 4.5: Probability of evolution of label knowing that before the first lockdown zones are labelled as *others*

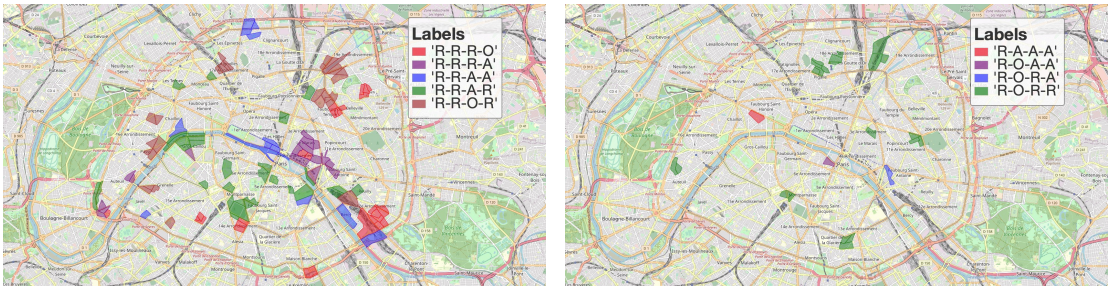
Being labeled as *residential* during the first lockdown is also is also the dominant behavior for areas labeled *others* before the first lockdown, with a probability of 62.6%. The probability that the area will continue to be labeled as *others* is also not negligible, at around 35%.

4.3.5 Signature and geographical analysis

In this section, we will analyze some specific patterns of label evolution, which can provide some insight on the impact the pandemic had on some geographical regions of Paris.

Evolution of residential areas. Fig.4.7 shows the geographical distribution of label evolutions when, during the period before the first lockdown, the areas are labeled *residential* and change label during at least one of the lockdowns.

Fig.4.7(a) shows the evolution of labels when, during the first lockdown, zones are also labeled as *residential*. We can see that the 'R-R-R-O' evolution has a higher area concentration in the southeast of Paris. This area is crossed by and adjacent



(a) Residential labelled zones during the first lockdown

(b) Activity or others labelled zone during the first lockdown

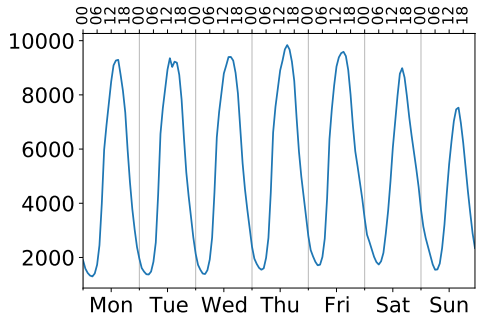
Figure 4.7: Residential labeled zones evolution during the three lockdonws

to railway lines near the Gare de Lyon and Bercy. For the 'R-R-R-A' evolution in purple, its highest concentration is in the centre of Paris in the Le Marais district and at the Bastille. For the 'R-R-A-A' evolution in blue, it covers the Hopital Bichat in the North of Paris. It also covers the quays of the Seine and a part of the commercial street Rivoli. For the evolution 'R-R-A-R', in green, covers some public gardens and parks. It also covers high schools. The opening of high schools in the second lockdown and the presence of school holidays in the third lockdown can explain this behavior. For the 'R-R-O-R' evolution in brown, there is a concentration of areas in the northeast of Paris in the Canal Saint Martin area.

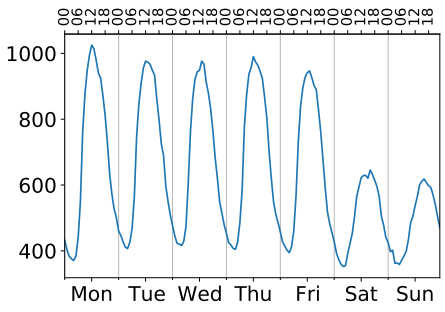
Fig.4.7(b) shows the evolution of labels when during the first lockdown the areas are not labeled as *residential*. We see that the 'R-O-R-R' evolution in green is the most important. The areas with this evolution cover partly small parks or near cemeteries. For the 'R-O-A-A' evolution, it corresponds to one area in the centre of Paris. It comprises many restaurants.

The evolutions 'R-A-A-A' in red, and 'R-O-R-A' in blue are also interesting, with their signatures represented in Fig.4.8 and in Fig.4.9. These two evolutions cover one area each. For the area with an 'R-A-A-A' evolution, this is the only area that was labeled *residential* before the first lockdown which become *activity* during the lockdowns. The zone with the evolution 'R-O-R-A' has an atypical signature during the first lockdown.

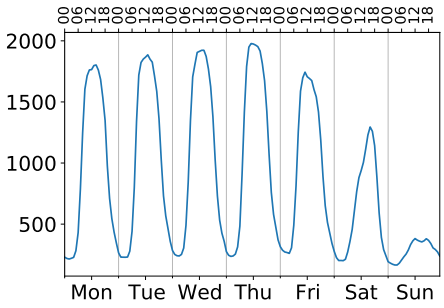
Before the lockdown, for 'R-A-A-A' evolution, in Fig.4.8(a), the signature is typical to the residential characteristic signature. During the first lockdown, for 'R-A-A-A' evolution, in Fig.4.8(b), we observe a typical signature of an activity IRIS, where there is a net decrease in user attendance during the weekend compared to weekdays. This zone corresponds to a part of Champs Elysees, and labeled *residential* before the lockdown because it is a famous touristic zone where there are many people during evenings and during the weekend. It is interesting to notice that this pattern changes during the lockdown, with much lower amplitude and an activity-like pattern, which is probably a result of some businesses and offices still active in the area during the lockdowns.



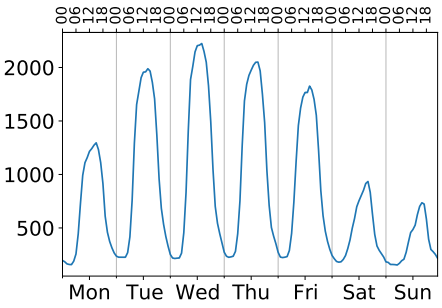
(a) Signature before the first lockdown of zone with 'R-A-A-A' evolution.



(b) Signature during the first lockdown of zone with 'R-A-A-A' evolution.

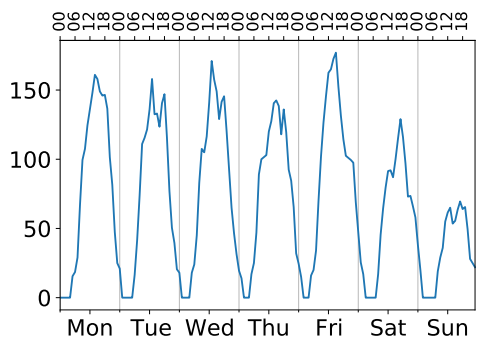


(c) Signature during the second lockdown of zone with 'R-A-A-A' evolution.

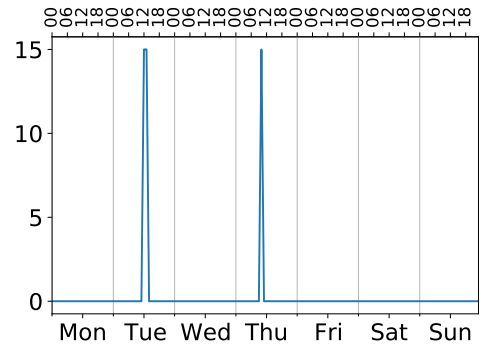


(d) Signature during the third lockdown of zone with 'R-A-A-A' evolution.

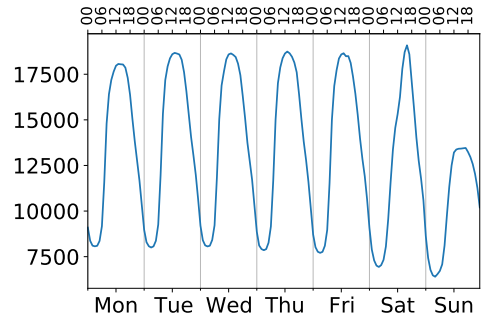
Figure 4.8: Signature before the lockdown and during the three lockdowns of 'R-A-A-A' evolution zones



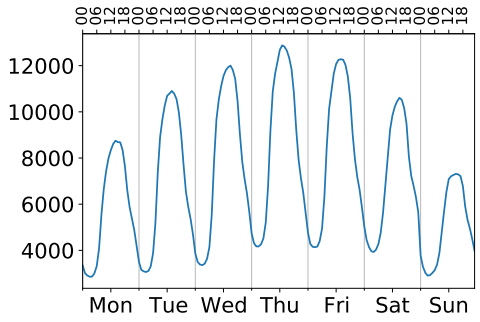
(a) Signature before the first lockdown of zone with 'R-O-R-A' evolution.



(b) Signature during the first lockdown of zone with 'R-O-R-A' evolution.



(c) Signature during the second lockdown of zone with 'R-O-R-A' evolution.



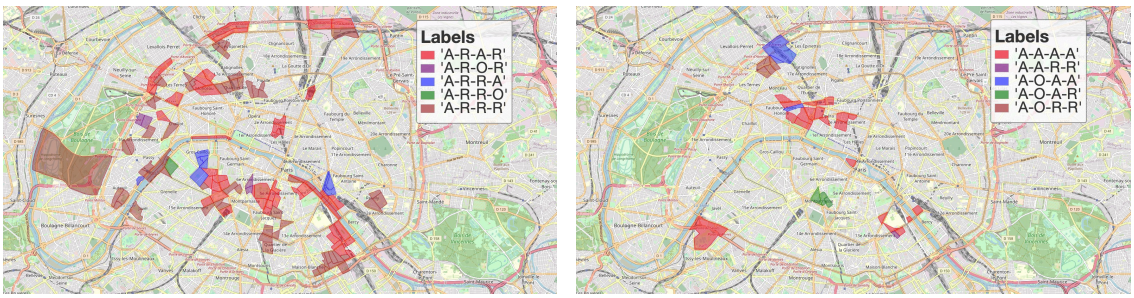
(d) Signature during the three lockdown of zone with 'R-O-R-A' evolution.

Figure 4.9: Signature before the lockdown and during the three lockdowns of 'R-O-R-A' evolution zones

Before the lockdown, for 'R-O-R-A, in Fig.4.9(a), the signature is similar to the residential characteristic signature in Fig.4.6(b). However, some particularities can be observed, namely the fact that the signature presents multiple peaks in some days, in early afternoon and in the evening. During the first lockdown, for 'R-O-R-A, in Fig.4.9(b), the signature shows that almost all presence disappears, with the exception of some sporadic peaks, one on Tuesday between 12 am and 1 pm and on Thursday at 10 am. This zone corresponds to an area that covers mainly the Opera Bastille. The fact that this area is labeled as residential before the lockdown is simply a side effect of evening events at the opera, that attract the public at late hours, a behavior associated with residential areas by our classification method. However, the signature during the lockdown, when the Opera is closed, clearly signals the fact that this is not a typical residential zone.

It is important to note that a number of touristic and cultural areas, such as Opera Bastille and Champs Elysees in our examples, are labeled as *residential* areas before the lockdown by our framework. As explained, this is due to increased weekend attendance and a late evening drop-in user presence, similar to residential areas. As shown in [15], with more fine-grained spatial and temporal data (a point every 15 minutes at a cell level in [15]), the signature-based approach can distinguish these touristic areas from residential ones.

Evolution of activity areas. Fig.4.10 shows the geographical distribution of label evolutions when, during the period before the first lockdown, the areas are labeled *activity*. We exclude the evolution 'A-R-A-A' where the zones are labeled as *residential* during the first lockdown, and become again labeled as *activity* during the two others lockdowns.



(a) Activity labelled zone before the first lockdown and residential labelled zones during the first lockdown

(b) Activity before the first lockdown and activity or others labelled zone during the first lockdown

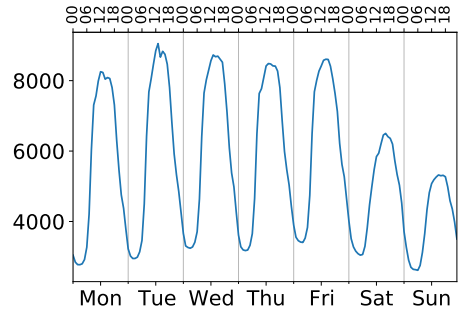
Figure 4.10: Activity labelled zones evolution during the three lockdowns

Fig.4.10(a) shows the evolution of labels when during the first lockdown the areas are labeled as *residential* and before the first lockdown they are labeled as *activity*. The two most important evolution are 'A-R-R-R' in brown and 'A-R-A-R' in red. The evolution 'A-R-R-R' covers areas containing hospitals, such as the Hopital Val de Grâce, the Hopital Saint Anne, but also the research center Institut Pasteur. It also covers green areas such as part of the Bois de Boulogne. The areas with this evolution cover a part of the railways to go to the Gare de l'Est and the Gare de Saint Lazarre, but also the road along the Seine on a the Western part. The 'A-R-A-R' evolution covers the Montparnasse and Est train stations. It covers universities such as Sorbonne Univerity Pierre and Marie Curie Campus or the faculty of pharmacy.

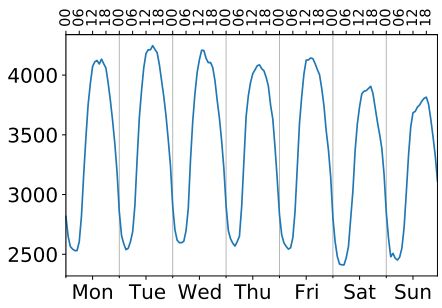
It also covers green spaces such as the Jardin de Luxembourg and sports complexes in the north of Paris. Moreover, this evolution covers a part of the east of the Seine and the François Mitterrand library in the southeast.

The evolution 'A-R-O-R' in purple, covers the University Catholic Institute of Paris but also many embassies in the west of Paris. The evolution 'A-R-R-A' covers the marina, the arsenal but also the UNESCO site and the military school.

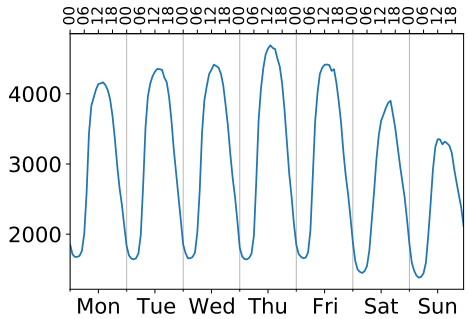
It is interesting to look at the only area that has an 'A-R-R-O' evolution with its signature in Fig.4.11.



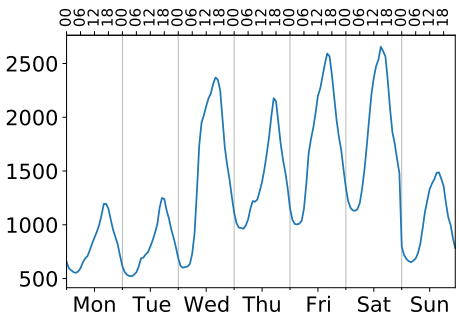
(a) Signature before the first lockdown of zone with 'A-R-R-O' evolution.



(b) Signature during the first lockdown of zone with 'A-R-R-O' evolution.



(c) Signature during the second lockdown of zone with 'A-R-R-O' evolution.



(d) Signature during the third lockdown of zone with 'A-R-R-O' evolution.

Figure 4.11: Signature before the first lockdown and during the three lockdowns of 'A-R-R-O' evolution zone

Fig.4.11(a) shows a typical signature of an area labeled *activity*. Indeed, there is a net decrease in user attendance during the weekend compared to weekdays, but also at the end of the day, during the week, there is a rapid decrease in attendance. During the third lockdown this area is labelled *others*, its signature is described by Fig.4.11(d). Compared to the period before the first lockdown, there is a general decrease in attendance. It went from a peak of almost 9000 users to just over 2500 users. During this third lockdown, the signature has a particular behavior. Monday and Tuesday are similar, with a single late peak around 6 pm and a lower attendance than the other days. Wednesday, Friday and Saturday also have similar behavior with a single peak between 3 pm and 5 pm, but with higher attendance. Wednesday is similar to Monday and Tuesday but with higher attendance. What we can see is that during the nights of Wednesday, Thursday and Friday the attendance remains quite high. On Sundays there is a decrease in attendance. This area contains the closest metro and RER station to the Eiffel Tower, the Bir-Hakein station. It is a place that serves as a place of passage for tourists, but also contains businesses.

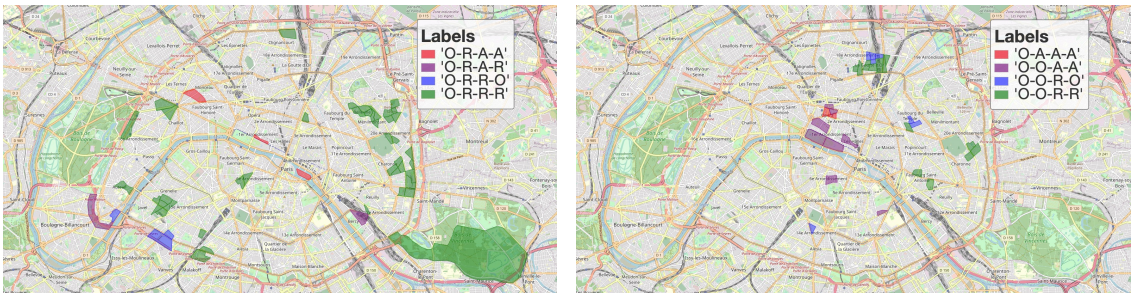
There are also evolutions where during the first lockdown, the zones do not become labelled as *residential*, as shown in Fig.4.10(b). For the 'A-A-A-A' evolution, in red, where the zones remain labelled *activity* throughout the study period, this concerns places of power like ministries, presidential palace, or embassies.

Only one area, in purple, has an 'A-A-R-R' evolution. This zone contains residential buildings but also a shopping centre.

The evolution 'A-O-A-A', in blue, covers two different areas. The first one covers a part of the railway for the Saint Lazare station, but also the court of Paris, a park and a shopping centre. The second is more in the centre of Paris and covers an area that contains many commerces and one of the Republican Guard barracks.

The evolution 'A-O-A-R', in green, covers an area that includes a university and primary, secondary and high schools. Only one groupe of IRIS in the northwest of Paris has an 'A-O-R-R' evolution in brown. This zone grouping includes primary, secondary and high schools, but also embassies and consulates.

Evolution of others areas. Fig.4.12 shows the geographical distribution of label evolutions when, during the period before the first lockdown, the areas are labeled *others*.



(a) Others labelled zone before the first lockdown and residential labelled zones during the first lockdown

(b) Others before the first lockdown and activity or others labelled zone during the first lockdown

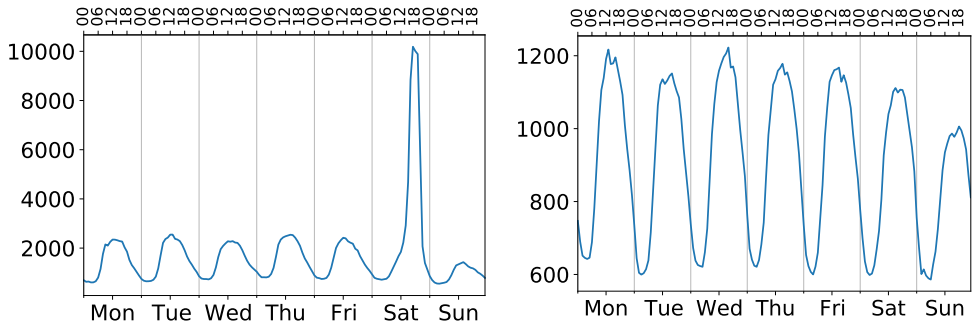
Figure 4.12: Others labelled zones evolution during the three lockdowns

Fig.4.12(a) shows the evolution of labels when, during the first lockdown, the areas are labelled as *residential*. The most important evolution is 'O-R-R-R' in green. A large proportion of the areas with this trend are on the outskirts of Paris, with a stronger concentration in the east. It includes for example part of the Bois de Vincennes. The areas with this development in the centre of Paris are areas with small shops. The closure of so-called "non-essential" shops can explain this evolution.

The evolution 'O-R-A-A', in red, includes three areas. The Ile Saint Louis, which is quite residential with small shops. The second zone is on the rue de Rivoli at the level of the Louvre Museum, which is an area frequented by tourists in normal times. The third zone contains an embassy and churches.

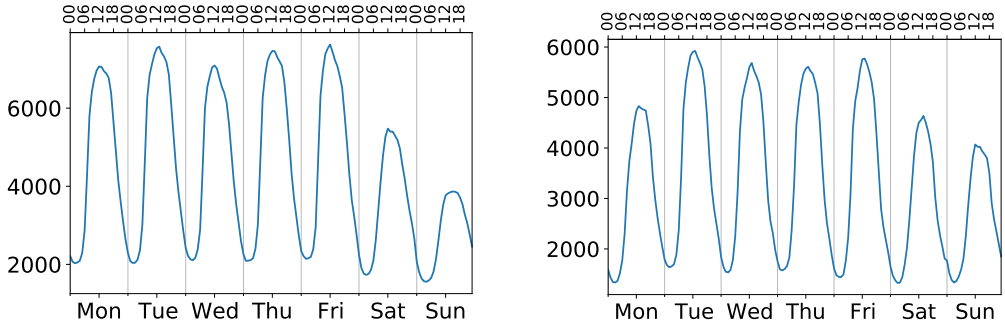
The 'O-R-R-O' evolution, in blue, has two zones. A zone in the southwest of Paris with a school and a high school. The second zone is a part of the Versailles exhibition park and an area with housing next to it. For the latter, the part with housing may have influenced the label during the lockdowns.

The 'O-R-A-R' evolution, in purple, covers two areas. One area covers the Bercy train station and the other others a sport complex that hosts the matches of the Paris Saint-Germain football team.



(a) Signature before the first lockdown of zone with 'O-R-A-R' evolution.

(b) Signature during the first lockdown of zone with 'O-R-A-R' evolution.



(c) Signature during the second lockdown of zone with 'O-R-A-R' evolution.

(d) Signature during the third lockdown of zone with 'O-R-A-R' evolution.

Figure 4.13: Signature before the first lockdown and during the three lockdowns of 'O-R-A-R' evolution zones

It is interesting to analyze the signature of the sport complex, in Fig.4.13. We can see, in Fig. 4.13(a), that on Saturday, before the lockdown, there is an outstanding peak in the signature in the late afternoon. This corresponds to a match day and results in a unique pattern and the *others* label. During the second lockdown, this area was typically active. During this second lockdown, there were football matches taking place in this area. However, the matches were played behind closed doors [117]. This may explain the fact that the area is labeled *activity* instead of *others* or even *residential*.

There are also evolutions where, during the first lockdown, the zones do not become labeled as *residential*, as shown in Fig.4.12(b). The evolution 'O-O-R-R', in green, is the most important. The areas with this evolution are quite concentrated in the north of Paris along or next to the railways that lead to the Gare du Nord. There are areas to the east of Paris with residences and schools and colleges. The area around the Saint Antoine hospital also had this evolution.

The second most important evolution is 'O-O-A-A', in purple. The areas with this evolution are mainly in the centre of Paris, with parks like the Jardin de Tuileries, the Bourse du Commerce, but also the famous Place Vendome with the Ministry of Justice. Some large Parisian stores also have this evolution.

The 'O-O-R-O' evolution, in blue, also covers the railways to the Gare du Nord, but focused on areas in the central east and south. The 'O-A-A-A' evolution also covers an area in central Paris. This area contains the Opera Garnier.

4.4 Metric for epidemic zone importance

Signatures can also be an indicator of how important is an area for the spread of a virus such as that responsible for Covid-19 disease. Indeed, the different peaks of attendance that a signature of an area presents can indicate a strong mixing of the population, if this value is largely higher than the average attendance of the area. Based on this assumption, we propose a metric to try to assess the importance of an area.

4.4.1 Peak to average ratio

The use of peaks in attendance signature is common in the detection of human mobility and presence. Therefore, the peaks can indicate some important information. We can suppose that peaks indicate population mixing in a zone. Indeed, peaks in human attendance may indicate how new people arrive in the zone, and how they leave. A large peak value compared to the mean attendance value in an area may indicate a higher risk of populations mixing (and spreading the virus during a pandemic).

Peak definition

Like in Chapter 3, we define a peak as a local maximum on a given signature, over a window with a predefined size w. In other words, we can define a peak as follows:

$$P_S = \{i | s_i \geq s_{i-k}, \forall k \in [-w, w]\} \tag{4.1}$$

In Eq.4.1, the set of peaks P of the signature S corresponds to the different time instant i when the signature value is greater than all other signature values in a window k preceding and following i.

Average peak to average signature ratio

With the peaks in each signature, we can compute a ratio based on peaks information. We propose to compute the average of all peaks in the signature over the mean of the whole signature. Formally, we have:

$$R_{Avg_z} = \frac{\frac{1}{n}\sum_{i=0}^{n} P_{z_i}}{\frac{1}{m}\sum_{i=0}^{m} s_{z_i}}, \forall P_z \in P_s, s \in S \tag{4.2}$$

In Eq.4.2, the average ratio R_{Avg} of a zone z is the mean of the n peaks P of the zone z over the mean attendance of signature s of the zone z. This ratio has the advantage of describing the general variation of attendance in a signature.

4.4.2 Geographical distribution

We propose to use the average peak to average (APA) ratio as an indicator of the importance of a zone for virus propagation.

Evolution of APA ratio

The analysis of the variation in the APA ratios over the four study periods may indicate the impact of the different measures taken by the French government to limit the spread of Covid-19, through the setting up of lockdowns.

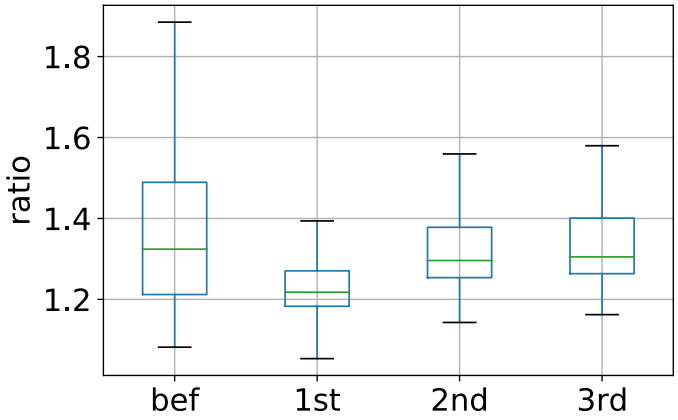


Figure 4.14: Boxplot showing the evolution of ratio during the four study periods

In Fig.4.14, we can see the boxplot distribution of the ratios according to the study period. We do not show the outliers to ease the readability of the results. Indeed, the first lockdown presents an outstanding outlier, with a ratio value of 56. The other outliers, regardless of the period, have values that do not deviate greatly from the upper bound and the lower bound of the boxplot. The value of these bounds are obtained with:

$$Bound_{up} = min(Value_{max}, Q3 + 1.5 \cdot (Q3 - Q1)) \tag{4.3}$$

$$Bound_{low} = max(Value_{min}, Q1 - 1.5 \cdot (Q3 - Q1)) \tag{4.4}$$

During the period before the first lockdown, we can see that the APA ratio values are quite dispersed compared to the other periods. We see this in the interquartile range, represented by the rectangle, or the values between the 1st and 3rd quartile. The interquartile range contains the ratios ranging from 1.21 to 1.48. The dispersion of the ratios before the first lockdown is also visible with the lower and upper bounds of the boxplot. Here we have a range of ratio values from 1.08 to 1.88.

During the first lockdown, if we do not consider the outlier value mentioned above, and discussed below, the dispersion of the ratios is less important compared to before the lockdown. Indeed, the interquartile range of the ratios goes from 1.18 to 1.27, and the lower and upper bounds of the boxplot go from 1.02 to 1.39. As we see from these values, during the first lockdown, we also have a decrease in the ratio values in addition to the decrease in their dispersion.

During the second and third lockdown, the dispersion of their ratios is quite similar. The values of the ratios contained in their interquartile range from 1.25 to 1.37 and from 1.26 to 1.40 respectively. And the values of their lower and upper bounds range from 1.14 to 1.6 and from 1.16 to 1.57. These values are more important than during the first lockdown, while still being tighter than the period before the first lockdown.

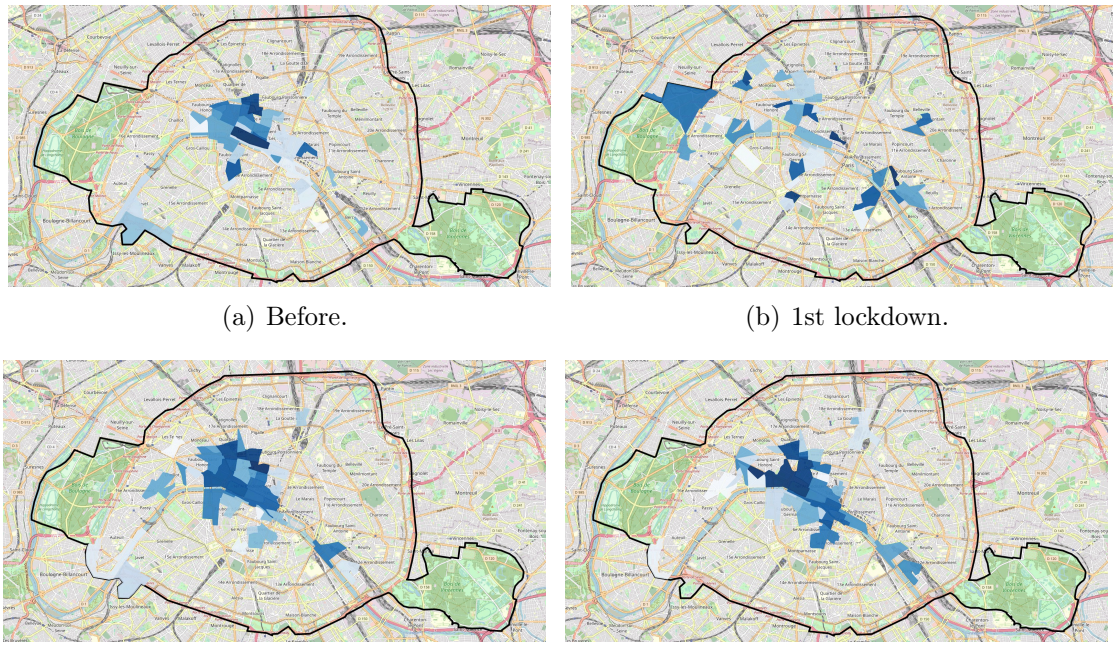
We can see that the median value of the ratios, represented by the green line inside the rectangles, is different according to the periods. The value of the median decreases during the first lockdown compared to the period before the lockdown. The median increases during the second lockdown, and during the third lockdown, approaching the median value of the period before the first lockdown.

In relation to the evolution of the ratios between the different periods we can assume that the period before the first lockdown presents heterogeneous areas, with

high population mixing. These areas have peak values that may be much higher than the average evolution of the signature. The implementation of movement restrictions during the lockdowns homogenised the ratios, by limiting their dispersion, and also decreased the ratios during lockdowns.

Top 10 APA ratio

It is interesting to analyze the spatial distribution of the zones with ratios in the top 10 percent of each period. These are zones with significant peaks, indicating a high degree of population mixing.



(a) Before. (b) 1st lockdown.

(c) 2nd lockdown. (d) 3rd lockdown.

Figure 4.15: Top 10 percent ratio for each periods.

In Fig.4.15, we can see the spatial distribution of the Top 10 percent zones with the most important ratios. The areas with dark blue colours have the highest ratios in the top 10 and the lighter the colours the lower the ratios of the corresponding areas.

We can see that, for the period before the first lockdown, in Fig.4.15(a), but also during the second lockdown, Fig.4.15(c), and the third lockdown, Fig.4.15(d), the areas with the highest ratios are concentrated in the centre of Paris. These zones are essentially labeled *activity*, according to the method we used earlier.

For the period before the first lockdown, the area with the Jardin de Tuileries has the highest ratio. This is a garden that is visited by about ten million people a year [118]. The areas in which the Parisian department stores are located are among the areas that have high ratios.

Contrary to the three other study periods, during the first lockdown, the areas with high ratios are scattered in Paris. The area with the highest ratio is the one described above, the Bastille opera, whose signature, during the first lockdown, is shown in Fig.4.9(b). Moreover, it is this area that has the outlier value discussed earlier. The value of this ratio can be explained by the presence of the two unique peaks in the signature, and except for the time of these peaks, the value of the

attendance is zero. Therefore, it had an average signature value of less than 1, and mean of the peaks are 15, which give a ratio of 56. We also see that some parks such as the Jardin des Plantes, the northern part of the Bois de Boulogne and the Jardin de Tuileries present high APA ratios. The train stations of Lyon, Montparnasse and Saint-Lazare also have important ratios.

During the second lockdown, the Paris centre has the highest ratio. The area with the highest ratio contains businesses such as the head office of a bank. The stations of Saint Lazare and Lyon also have high ratios. The Jardin des Tuileries still has a high ratio, and so does the quays of the Seine in the centre of Paris, but it is also joined by the Jardin de Luxembourg and some Parisian High school.

During the third lockdown, it is also the central areas of Paris that have the highest ratios. The area with the highest ratio is located on part of the Champs Elysées. The Jardin des Tuileries, the quays of the Seine in the centre of Paris, and the Jardin de Luxembourg with the Parisian High school also have high ratios. Along with the Gare de Lyon and Saint Lazare, the Gare du Nord is also in the top 10 areas with the highest ratios. The Bercy station also has a high ratio, however the area in which this station is located also contains the Bercy park.

This spatial distribution of the top 10 zones with the highest ratios shows us that the zones labeled as *activity* tend to have higher ratios. This can be explained by the fact that attendance in these areas increases during the day and decreases during the night. Concerning the Jardin des Tuileries, which is always part of the areas with a high ratio, this high ratio can be explained by the central place, which makes it a place of transit.

4.5 Conclusion

In this chapter, we used signature and unsupervised clustering methods to automatically detect land use. Using this method, we observed the evolution of different land uses during the three lockdowns that were imposed in France during the Covid-19 pandemic. The first lockdown was the strictest, and this was reflected in the land use. Most of Paris turned out to be labeled as *residential*, and it was this behavior that was expected by the authorities. The other two lockdowns were less strict, and the activity zones reappeared. However, we note that school holidays can also affect the land use, as in the third lockdown. The use of signatures allowed us to define metrics to evaluate the importance of each zone in the mixing of the population. These metrics indicate that areas that have been labeled *activity* by the unsupervised clustering method tend to have the most important metric values.

Chapter 5

eNodeB handover-based characterisation

5.1 Introduction

In previous chapters, we discussed about using mobile data as human mobility proxy and focused on the lockdown impact due to Covid-19 on the land use. Mobile data is also use for management and configuration of the network. It can help in understanding or discovering behaviors of the latter. In the case of user mobility, especially for micro mobility, which follows the movement of a user, it can be described in network terms by the handovers (HOs). HOs are performed when a user, in particular the UE, changes cells while being connected to the mobile network and being active on it (i.e communicating, call, data exchange). Therefore, this can give some idea of human mobility in a given area. For a better understanding of how HO works, the Appendix A describes the HO process in the different technologies. Intrinsically, HO is important for user Quality of Experience (QoE). Indeed, the fact that HOs occur while the user is communicating, a poor management or configuration of the network can lead to a degradation of the communication or even, in the most extreme cases, to the breaking of the connection. There are two types of HOs, incoming HOs (in_HO) and outgoing HOs (out_HO). For a given eNodeB, in_HO are HOs that end to this eNodeB, which represents the HOs destination. Out_HO correspond to HOs that have been initiated on a given eNodeB, which can be considered as the source eNodeB. Intuitively, an eNodeB treats these two types of HO differently. Indeed, their execution involves different network functions. We aim at profiling eNodeB with respect to in_HO and out_HO, investigating whether an eNodeB has a predominance in_HO or a predominance out_HO. Knowing these different profiles may allow the configuration of adequate network functions.

In this chapter, first, we discuss about the mobile data which allowed us to characterise HOs. Then we propose a methodology to characterise HOs depending on the profile of the eNodeB and the volume of the HOs performed.

5.2 Data

Our objective is to detect whether different eNodeBs have sufficiently different behaviors in terms of incoming and outgoing handovers and thus to show the existence of different handover profiles that an eNodeB can have. As mentioned before, we start from the observation that an incoming HO (in_HO) and an outgoing HO (out_HO)

do not consume the same number and type of resources. It would therefore be interesting to distinguish, if only for resource consumption, between eNodeBs with a more HO-emitting profile and/or with a more HO-receiving profile.

To show the existence of different eNodeB profiles, we used HO data from a French major operator, Orange. The data cover four geographical areas in France: Paris, Lyon, Poitiers, and Cognac. These four geographical areas represent the different urban areas that a country can have. Indeed, we have a metropolis of international importance, a large regional metropolis, a medium-sized city, and an area with a small city. The size and population density of a geographical area will reflect the number of eNodeBs covering it. Our study areas in Paris, Lyon, Poitiers, and Cognac are covered by 873, 490, 58 and 24 eNodeBs respectively. For all four geographical areas, we have access to HO data covering a one-month period from 16 March 2019 to 16 April 2019.

Concerning the HO data, we have access to the HO performed in 4G at the eNodeB level. This data represents, for each eNodeB, the total number of UEs that initiated a HO departing from it (outgoing HO) and the total number of UEs it received with an HO (incoming HO), aggregated over a 30 minute interval. A UE can perform more than one incoming HO and outgoing HO during this 30 minute interval. We do not have any information about the users, to guarantee the anonymity of the data.

5.3 Handovers profilling methods

To discover eNodeB profiles for each geographical area, we clustered per eNodeB handover data. We use, like in chapter 4, an agglomerative hierarchical clustering algorithm [115].

5.3.1 Typical week data

Prior to the clustering, we format the data. We based our format on a week basis. Indeed, in [66] they show that using week basis allows a clustering algorithm to find the difference in behavior that may occur on different days. Moreover, several studies [60, 119, 120] use week basis timeseries to validate their propositions. Week basis can also capture the vast majority of mobile traffic [121, 122]. We test two possibilities, either we use one typical week or all separate weeks, i.e. each eNodeB has four weeks to cluster. By comparing the results, the separate weeks and typical week approaches are quite similar. At least two weeks obtain the same label as the typical week for 65% to 100% of the cases, depending on the city. The use of a typical week can therefore be justified to allow all eNodeBs in the study area to have a profile, despite missing data during some brief time intervals. In addition, it helps to reduce the noise that individual weeks may present. The repetitiveness in human activities can also justify the use of the typical week, as from one week to another, in most cases the differences can be negligible. And finally, the use of the typical week can reduce the complexity of the clustering that is performed, since there are fewer elements to process.

In addition to the profiling of eNodeBs, we also conducted a study on the categorisation of eNodeBs according to the total volume of in_HO and out_HO that they have performed. This categorisation will enable us to classify eNodeBs according to the number of in_HO and out_HO. The resulting classes can then be a good indication of the amount of resource consumption of the eNodeBs, in addition to the

types of resources they consume, defined by their profiles. For this categorisation, we also use typical weeks for the same reasons mentioned above.

Typical week timeseries construction

We construct the typical week in three steps:

1. **day construction** which consists in building a set of in_HO and out_HO per eNodeB and per day and the definition of new metrics for eNodeB profiling.
2. **typical day construction** which consists in the definition of a typical day for a given eNodeB.
3. **typical week construction** for each eNodeB of the different geographical area is built from the concatenation of typical days.

Day construction: For each eNodeB, we have the total number of in_HO and the number of out_HO performed per each 30 minute interval. From this, we can create two sets containing the evolution of the number of in_HO and out_HO for a given day and for each day. A set will contain time ordered values of the HOs according to the time of day $\boldsymbol{T}$, i.e. from 00h to 23h30.

Formally, the set of days $\boldsymbol{D}$ is defined as all days of our dataset. $\boldsymbol{C}$ is the set of geographical areas represented in our data, $\mathbf{C} = \{Paris, Lyon, Poitiers, Cognac\}$. $A_c = \{a_c | c \in \boldsymbol{C}\}$ represents the set of eNodeB in each defined geographical area. Therefore, we define:

$$\mathcal{IN}_{a_c}(d,t) = vin((a_c,d),t) \forall c \in \boldsymbol{C}, a \in \boldsymbol{A}, d \in \boldsymbol{D}, t \in \boldsymbol{T} \tag{5.1}$$

$$\mathcal{OUT}_{a_c}(d,t) = vout(a_c,d),t) \forall c \in \boldsymbol{C}, a \in \boldsymbol{A}, d \in \boldsymbol{D}, t \in \boldsymbol{T} \tag{5.2}$$

In Eq.5.1 and Eq.5.2, $vin((a_c,d),t)$ and $vout(a_c,d),t)$ represent respectively the description of the total in_HO and out_HO performed within a city c at eNodeB a and at time t of the day d.

There is a need of a new representation of the evolution of in_HO and out_HO in order to extract the eNodeB profile. Indeed, by looking only on either $\mathcal{IN}_{a_c}(d,t)$ or $\mathcal{OUT}_{a_c}(d,t)$ we cannot decide if the eNodeB has more in_HO or out_HO profile. Therefore we define:

$$\mathcal{DIFF}_{a_c}(d,t) = \mathcal{IN}_{a_c}(d,t) - \mathcal{OUT}_{a_c}(d,t) \tag{5.3}$$

The handover difference (diff_HO), represented by $\mathcal{DIFF}_{a_c}(d,t)$ in Eq.5.3, is defined by the subtraction of $\mathcal{IN}_{a_c}(d,t)$ by $\mathcal{OUT}_{a_c}(d,t)$. This subtraction allows us to see which component of diff_HO is more important for an eNodeB. This is an indication for the future profile of the eNodeB.

Typical day construction: We can put the different days $\mathbf{D}$ we have into subsets $\mathbf{D}^\delta$ according to the days of the week, from Monday to Sunday, where $\boldsymbol{\delta} = \{MON, TUE, WED, THU, FRI, SAT, SUN\}$. We have $\mathbf{D}^\delta \subset \boldsymbol{D}$ where $\mathbf{D}^\delta$ contains all the days for a given day of the week. For example, $\mathbf{D}^{MON}$ contain all Mondays data. We can define the typical day td for in_HO, out_HO and diff_HO as follow:

$$\mathcal{IN}_{td_{a_c}}(\delta,t) = \mu_{\frac{1}{2}}(\{\mathcal{IN}_{a_c}(d,t) | d \in D^\delta\}) \tag{5.4}$$

$$\mathcal{OUT}_{td_{a_c}}(\delta, t) = \mu_{\frac{1}{2}}(\{\mathcal{OUT}_{a_c}(d, t) | d \in D^{\delta}\}) \tag{5.5}$$

$$\mathcal{DIFF}_{td_{a_c}}(\delta, t) = \mu_{\frac{1}{2}}(\{\mathcal{DIFF}_{a_c}(d, t) | d \in D^{\delta}\}) \tag{5.6}$$

$\mu_{\frac{1}{2}}(.)$ in Eq.(5.4, 5.5, 5.6) represents the median. Indeed, we use the median to construct the typical day. Therefore we have a typical median week. The use of the median is interesting since it allows to avoid outliers and exceptional values [15] that may occur on a given day. The other possibility was to use the average. However, the average can be affected by very important exceptional values. In our case, we will take the median HO value of a given day at a given time. For example, 8:00 am on a typical Monday of a given eNodeB is the median of the HO values at 8:00 am of all Mondays.

Hence, $\mathcal{IN}_{td_{a_c}}(\delta, t)$, $\mathcal{OUT}_{td_{a_c}}(\delta, t)$ and $\mathcal{DIFF}_{td_{a_c}}(\delta, t)$ are sets which contain each typical median day.

Typical week construction: tw is built by concatenation of all typical median days for a given eNodeB a_c :

$$\mathcal{IN}_{tw_{a_c}}(t) = \|_{\delta \in \boldsymbol{\delta}} \mathcal{IN}_{td_{a_c}}(\delta, t), \quad \forall c \in \mathbf{C}, a \in \mathbf{A}_c \tag{5.7}$$

$$\mathcal{OUT}_{tw_{a_c}}(t) = \|_{\delta \in \boldsymbol{\delta}} \mathcal{OUT}_{td_{a_c}}(\delta, t), \quad \forall c \in \mathbf{C}, a \in \mathbf{A}_c \tag{5.8}$$

$$\mathcal{DIFF}_{tw_{a_c}}(t) = \|_{\delta \in \boldsymbol{\delta}} \mathcal{DIFF}_{td_{a_c}}(\delta, t), \quad \forall c \in \mathbf{C}, a \in \mathbf{A}_c \tag{5.9}$$

In Eq.(5.7, 5.8, 5.9) $\|$ indicates the day-ordered concatenation of all typical median days. Thus, we use the typical median days, starting with typical median Monday and ending with typical median Sunday.

In order to cluster the typical weeks, we have to conduct additional processing. First, we normalise $\mathcal{DIFF}_{tw_{a_c}}$, the typical median week. Indeed, there can be a big difference between the values from one eNodeB to another. This difference can affect the clustering results since high values can induce large variations and the opposite for lower values. Normalization allows to avoid this, by considering only the evolution of the behavior in the typical week. To take into account only the median typical week behavior, we use standard score normalization, or Z-score, to normalise $\mathcal{DIFF}_{tw_{a_c}}$. Formally, for a geographical area c and eNodeB a_c:

$$\widehat{\mathcal{DIFF}}_{tw_{a_c}}(t) = \frac{\mathcal{DIFF}_{tw_{a_c}}(t) - \mu(\mathcal{DIFF}_{tw_{a_c}})}{\sigma(\mathcal{DIFF}_{tw_{a_c}})} \tag{5.10}$$

In Eq.5.10 $\mu(\mathcal{DIFF}_{tw_{a_c}})$ and $\sigma(\mathcal{DIFF}_{tw_{a_c}})$ denote respectively the mean and standard deviation of the typical median week $\mathcal{DIFF}_{tw_{a_c}}$.

To simplify the expression $\widehat{\mathcal{DIFF}}_{tw_{a_c}}(t)$, we note :

$$\mathcal{P}_{a_c}(t) = \widehat{\mathcal{DIFF}}_{tw_{a_c}}(t) \tag{5.11}$$

Moreover, for the categorisation of eNodeBs according to the volume of HO, the $\mathcal{IN}_{tw}$ and $\mathcal{OUT}_{tw}$ values have the same order of magnitude and also similar behaviors. Clustering them at the same time therefore seems coherent to have a similar basis for

their subsequent labelling. We have consequently created a new set comprising the two previous sets.

$$\mathcal{V}_c = \{\mathcal{IN}_{tw_c}, \mathcal{OUT}_{tw_c} \mid c \in \mathbf{C}\} \tag{5.12}$$

In Eq.5.12, $\mathcal{V}_c$ corresponds to the new set that we call Volume. It contains all typical median weeks, describing the in_HO and out_HO volume evolution for all eNodeB in the geographical area c. We use this new set for the clustering.

5.3.2 Agglomerative hierarchical clustering

Like in Chapter 4, we use Agglomerative Hierarchical Clustering (AHC) for the eNodeB profiling and categorisation. We aim to show the existence of profiles of eNodeBs, which are more producer and/or more receiver of HO. The profiles give us the major tendency of the evolution of the HO of the eNodeBs in the same cluster.These different profiles can later be useful for network management, especially in resource management, as shown in Chapter 6. The AHC allows us to obtain, in a simple way, these great tendencies described by the profiles. Despite its simplicity, the AHC is often used in the literature [90, 60], because it allows to have satisfactory clusters in terms of precision.

For the profiling of eNodeBs and the categorisation according to the HO volumes they handle, we use the AHC with two different configurations. Indeed, the nature of the inputs given to the AHC is different. For profiling we use normalized data, whereas for categorisation we use non-normalized data.

AHC eNodeB profiling

Two parameters can influence the results given by the AHC: the method for calculating the distance, or pairwise similarity, of each pair of elements, and the linkage method or inter-class distance.

Pairwise similarity is the method of calculating the difference between two elements. It allows to create a distance matrix which contains the distance between each pair of elements, i.e. all typical median weeks of all eNodeBs. It is used by the linkage method to agglomerate new elements to a cluster. Here, we use the Pearson correlation coefficient as distance.

The linkage is a method that allows to compute inter-clusters distance in order to choose the nearest and form a bigger cluster. Indeed, as the AHC progresses, at a given step a cluster can be composed of one or several elements, e.g. several typical median weeks of different eNodeBs. At the next step of the algorithm, the size of the cluster (i.e number of elements that comprise it) can increase. The possible new element of the cluster is the one that is at the closest distance to the cluster. Each element of the cluster has a different distance value to this element. These distances are represented in the distance matrix. Linkage is therefore a method which uses the distance matrix to find the closest element of the cluster. In the case of eNodeB profiling, we use the average method.

Pearson correlation coefficient pairwise similarity method for distance matrix: To compute the distance matrix, we use Pearson correlation coefficient as distance. The Pearson correlation coefficient distance, for two typical median week of

two eNodeB i and j is

$$\Delta_{c_{(i,j)}} = 1 - \frac{\sum\limits_{t\in\mathbf{t}} (\mathcal{P}_{i_c}(t) - \mu(\mathcal{P}_{i_c}))(\mathcal{P}_{j_c}(t) - \mu(\mathcal{P}_{j_c}))}{\sqrt{\sum\limits_{t\in\mathbf{t}}(\mathcal{P}_{i_c}(t) - \mu(\mathcal{P}_{i_c}))^2}.\sqrt{\sum\limits_{t\in\mathbf{t}}(\mathcal{P}_{j_c}(t) - \mu(\mathcal{P}_{j_c}))^2}} \quad \forall c \in \mathbf{C}, i, j \in \mathbf{A}_c \quad (5.13)$$

In Eq.5.13, we have Pearson correlation coefficient distance of two typical median weeks i and $j \in A_c$ of a given geographical area $c \in \mathbf{C}$. From Eq.5.13 we can define the distance matrix for the geographical area c:

$$\Delta_c = \begin{pmatrix} 0 & \Delta_{c_{(a,b)}} & \cdots & \Delta_{c_{(a,n)}} \\ \Delta_{c_{(b,a)}} & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_{c_{(n,a)}} & \cdots & \cdots & 0 \end{pmatrix} \quad (5.14)$$

The linkage method uses this distance matrix in Eq. (5.14) for building clusters.

In this case, the use of Pearson coefficient correlation as distance is coherent, since we aim to build cluster based on the behavior of the eNodeB. The Pearson correlation coefficient is a good indicator of the similarity and it has a stable performance [15].

Average linkage method: The average method is also called the UPGMA algorithm. It consits of averaging the distances between all the elements in a cluster and an outside element, in order to find the outside element the most similar (closest) to the cluster. Formally, the computation of the distance average can be described:

$$dist_c(u, v) = \frac{1}{|u|.|v|} \sum_{ij} \Delta_{c_{(u[i],v[j])}} \quad (5.15)$$

In Eq.5.15 we see that the distance between two clusters (elements) u and v corresponds to the sum distance of each element $i \in u$ and $j \in v$, divided by the product of their cardinalities, i.e. the size of cluster u and cluster v. The two clusters which have the smallest distance will fusion and this is repeated at each iteration until there is only one large cluster.

Using a linkage method, AHC outputs a whole family of solutions, represented by a dendrogram, with different possible number of clusters. The number of clusters will depend on a maximum inter-cluster distance, which needs to be defined.

AHC eNodeB HO volume categorisation

The pairwise similarity method used for the categorisation of the HO volume of eNodeBs is the euclidean distance. For the linkage method, we use the Ward algorithm [123].

Euclidean pairwise similarity: Euclidean distance can be expressed :

$$\Delta'_{c_{(i,j)}} = \sqrt{\sum_{t\in\mathbf{t}} (\mathcal{V}_{x_c} - \mathcal{V}_{y_c})^2}, \ \forall x_c, y_c \in \mathbf{A}_c \quad (5.16)$$

In Eq.5.16 we compute the Euclidean distance between two typical median week volumes x and y. As with the eNodeB profiling, this distance will be used to construct the distance matrix.

The use of Euclidean distance is justified here by the fact that it captures the variations that a typical median week volume may have in relation to another. This metric allows a distinction between an eNodeB with very high HO volumes and one with very low.

Ward linkage method: The variance minimization algorithm analyses the variance of a cluster and then deduces the distance between clusters by minimizing the variance withing each cluster. This can be expressed as follows [124] :

$$dist_c(u,v) = \sqrt{\frac{|v|+|s|}{Card_{Cl}}d(v,s)^2 + \frac{|v|+|z|}{Card_{Cl}}d(v,z)^2 + \frac{|v|}{Card_{Cl}}d(s,z)^2} \tag{5.17}$$

In Eq.5.17, u is the cluster composed by the union of clusters s and z. v is the cluster which we analyse to know if it also joins cluster u. $Card_{Cl}$ is the sum of the cardinality of $v + s + z$. Thus, we have $Card_{Cl} = |v| + |s| + |z|$. Regarding $d(v,s)^2$, $d(v,z)^2$ and $d(s,z)^2$, they correspond respectively to the variance of the distances of the set of elements that constitute the clusters v and s, v and z and s and z. Two clusters will merge when the value of their distance is the smallest and therefore the variance of the distances of the elements that compose them is minimal.

5.3.3 Cluster number

$\mathcal{P}_c$ cluster number. To choose the number of cluster we use, we conduct a visual analysis of different cluster numbers.

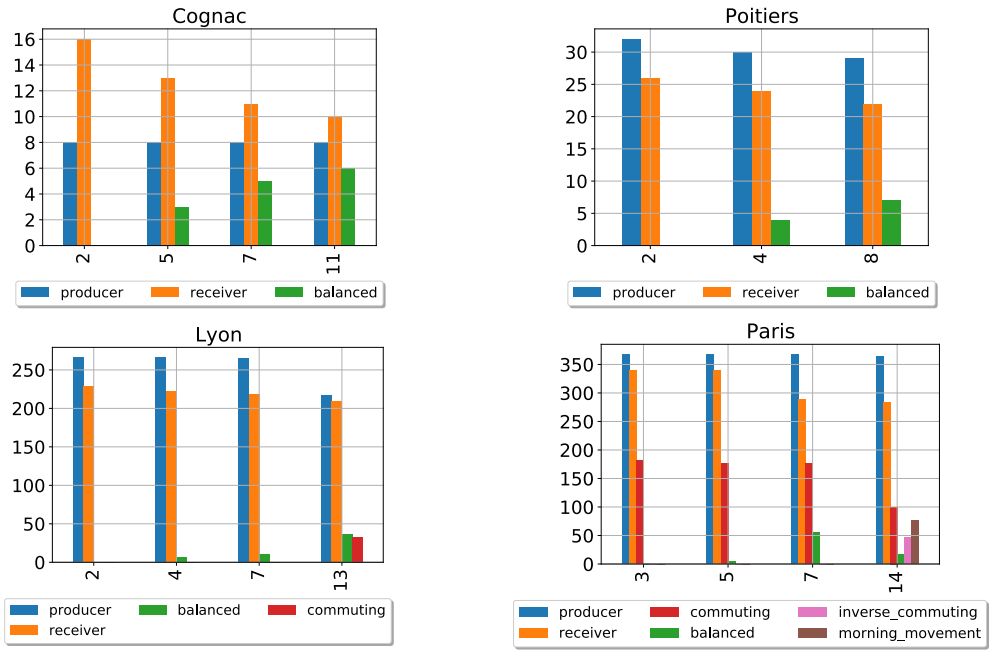


Figure 5.1: $\mathcal{P}$ aggregation level label analysis a) Cognac, b) Poitiers, c) Lyon and d) Paris

Looking at the different behaviors of $\mathcal{P}$ for each geographical area, we noticed visually that they have at least three different behaviors. By analyzing the results, after manual labeling, for a cluster number equal to 2, for the three geographical areas other than Paris, we find our assumption from the beginning, i.e. that there are eNodeBs that have rather producer profiles and others that have rather receiver

profiles. What we can see on Fig.5.1 is that there is a difference of about ten eNodeBs between the producer profiles and the receiver profiles. The x-axis indicates the number of clusters tested and the y-axis indicates the number of eNodeB in each profile that dominates. For Cognac, on the other hand, this difference proportion is more important given the small number of eNodeBs present in this study area. Also in Cognac, with a number of clusters equal to 2, contrary to Poitiers and Lyon, it is the receiver profile that contains the most eNodeB.

For our labeling, a profile or class can be composed of several clusters. For Cognac, we have chosen a number of clusters equal to 5. We chose this number because, for a higher number of clusters, we just have the appearance of some variant of the receiving profile. These profiles have some peculiarities but still have a receiver behavior tendency. For Poitiers, we have chosen 4 as the number of clusters. In this case, a larger number of clusters does not provide any additional information about particular areas. For Lyon, we chose 13 as the number of clusters. We chose this number because, a new interesting behavior appeared with a number of clusters of 13. As shown in Fig.5.1, a new profile, which we label as commuting, appears in Lyon at this point. For Paris, we chose a number of clusters equal to 14 to have the three basic profiles that we had in the three previous geographical areas. We chose 14 clusters, outputting two more different profiles, inverse commuting and morning movement which are quite different from the commuting profile. Morning movement profile presents a huge proportion of HO only in the morning.

$\mathcal{V}_c$ cluster number. Like $\mathcal{P}_c$, we conduct a visual analysis of different cluster numbers to choose the cluster number we use.

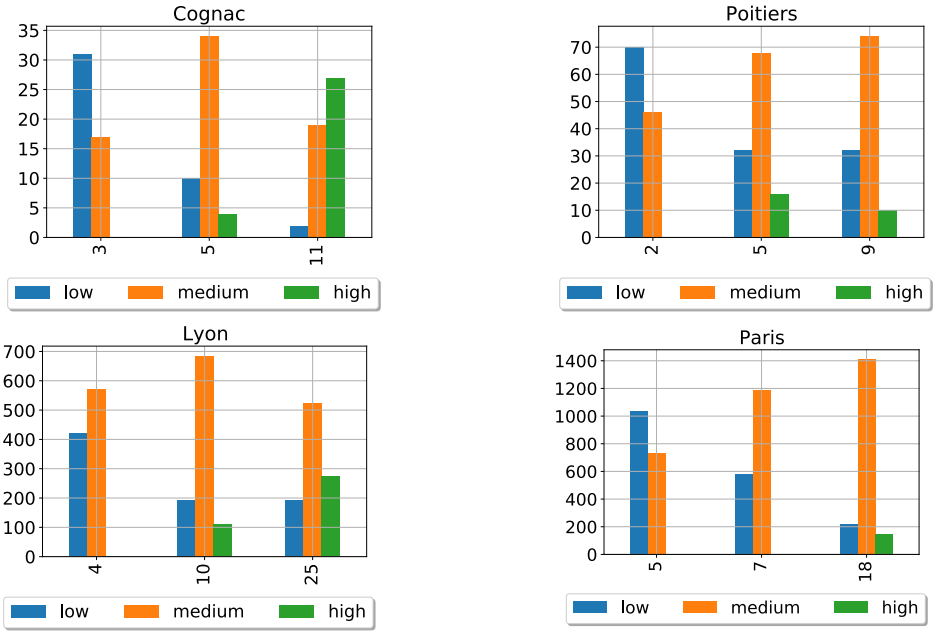


Figure 5.2: $\mathcal{V}$ aggregation level label analysis a) Cognac, b) Poitiers, c) Lyon and d) Paris

We test several aggregation levels for each geographical area. For each geographical area, we have three categories of eNodeBs: those with low HO volume, those with medium HO volume and those with high HO volume, as shown in Fig.5.2. The labelling methods of different categories are presented in the following section. For Cognac, we chose 5 as cluster number. We chose this value since we consider that

the number of eNodeB that have low or high volume values must be less important than the number of eNodeB that have medium volume values. For Poitiers, we chose a number of clusters equal to 9, because by manually checking the results, the repartition is more coherent than for 5 clusters. For Lyon, we chose 10 as the number of clusters for the same reason as for Poitiers. For Paris, we chose 7 as the number of clusters for the same reasons mentioned above. Although not visible in the figure, Paris has three categories, even if for the high volume category, there are much less eNodeBs than for the two others.

5.4 Results

Using the number of clusters defined previously, we now analyze the different profiles of eNodeBs and their categories.

5.4.1 eNodeB profile analysis

The labeling of clusters to obtain the eNodeB profiles is done by looking at the general trend of the different clusters. Different clusters can therefore be labeled with the same profile. Indeed two different clusters may have same trend, for example their elements have only positive values for diff_HO. In this case, the two clusters will be labeled with the same profile. Selecting a number of clusters equal to the number of defined profiles can prevent the detection of some actual profile, because a specific behavior can be hidden by the general trend of the cluster. Therefore, we chose a number of clusters higher than the excepted number of profiles. This allowed us to discover the existence of three basic types of profiles: producer, receiver, and balanced profiles. As a reminder, the profiles were obtained based on diff_HO, which is the difference between in_HO and out_HO.

Cognac

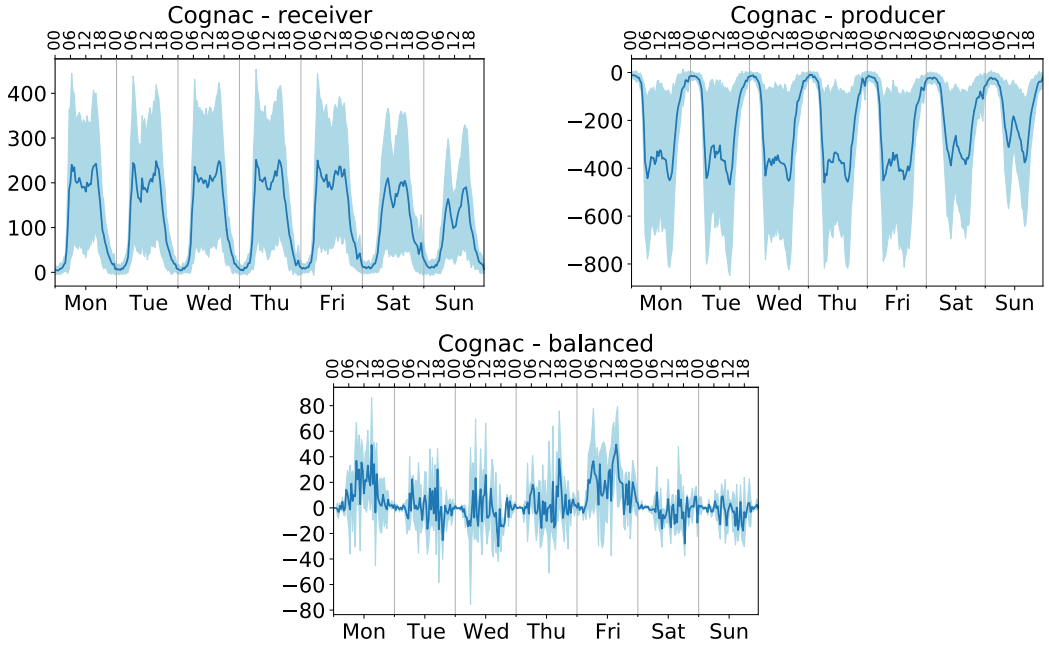


Figure 5.3: Profiles in Cognac: a) receiver, b) producer, c) balanced

For Cognac we have three different profiles, receiver, producer and balanced. In Fig.5.3, the y-axis represents the value of diff_HO. When this value is negative, at that moment, there are more in_HO than out_HO.

Receiver profile. In the case of the receiver profile, during the great majority of the time, the value of diff_HO is superior to 0. Therefore, we have a greater importance of in_HO compared to out_HO. This is the reason this profile corresponds to a receiver profile. Here, we see two main peaks during the day, one in the morning and one in the afternoon. The presence of these two peaks is often seen in the literature for the representation of mobile traffic. The appearance time of the morning peaks is different on weekdays and weekends. On weekdays, the morning peaks appears between 6:00 am and 7:00 am, while on weekend days, they appear between 9:00 am and 10:00 am. However, for the afternoon peaks, they appear between 3:00 pm and 4:00 pm, whether on weekdays or weekends. A difference that also exists between weekdays and weekends is the amplitude of the diff_HO value. On the curves of Fig.5.3, the continuous line in blue represents the average evolution of all the elements that form each profile, and the parts in lightblue represent the standard deviation around this average value. What we can see for the receiver profile is that, apart from a few exceptions during the night, the value of diff_HO is always greater or equal to 0. The maximum amplitude reached by the peaks is also higher for weekdays than for weekends.

Producer profile. Concerning the producer profile, the values of diff_HO are mostly negative. As for the receiver profile, the producer profile also presents two peaks, represented in reverse as valleys, during each day. These reverse peaks occur at the same time as the peaks for the receiver profile. Comparing the receiver and producer profiles, we can see that the amplitude reached by the producer profile is twice that of the receiver profile. The producer profile processes a larger number of out_HO than the receiver profile.

Balanced profile. For the balanced profile, the average value of diff_HO oscillates between positive and negative, yet we can see that on Monday and Friday this average value is mostly positive. However, when we consider the standard deviation, we have positive to negative oscillations every day. We can see that the value of diff_HO has a lower amplitude, it oscillates between -80 and 80 compared to the receiver and producer profile. The former has an average amplitude of 200 for the peaks during the week and the latter has an average amplitude of the peaks during the week is 400. We also notice that the amplitude of oscillations during the weekend is lower.

In Fig.5.4 we have the spatial distribution of each profile. These cells are obtained with a Voronoi projection, with base points the locations of the eNodeBs. Producer profiles are in blue, receiver profiles are in orange and balanced profiles are in green. The colored dots surrounding the cells correspond to the different locations of the edge eNodeBs. This is a choice of representation of edge points.

What we can see is that groups of producer cells are often adjacent to groups of receiver cells. There are two cells that are exceptions to this. However, one of these two cells is at the eastern edge of the Cognac study area, and both are adjacent to a balanced profile cell. For balanced profile cells, there are three in this study area. Two of them are surrounded by receiver and producer profile cells and the last one

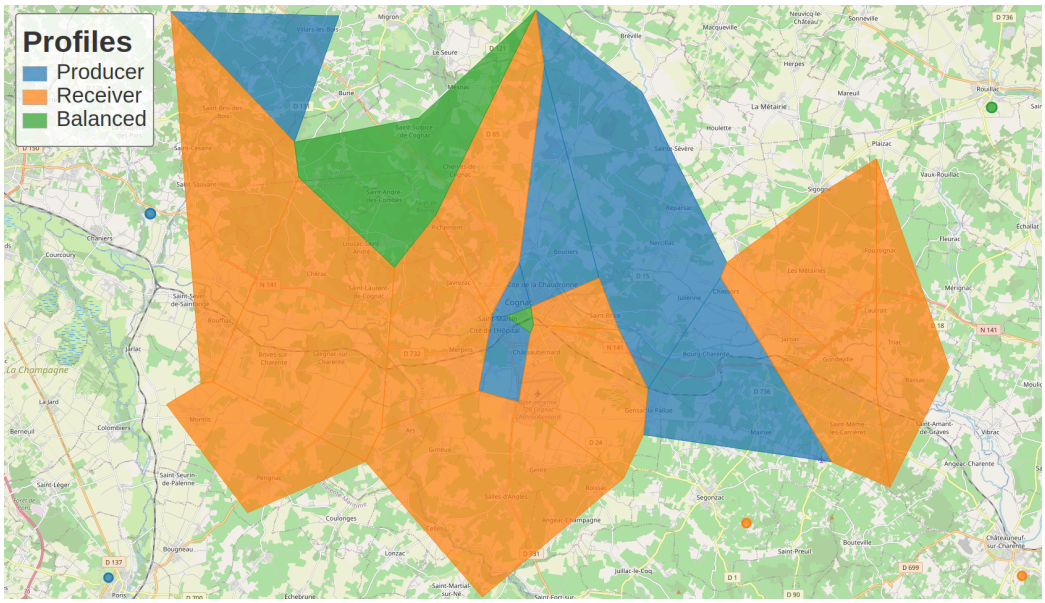


Figure 5.4: Cognac profile geographic distribution

is the green point in the Northeast at the edge of the study area. Of the two cells, one covers part of the Cognac train station, in the center of the area of study. The city centre of Cognac is a producer type. However, we see that the business and industrial zones to the east of Cognac, and the air base, located in south of Cognac train station, are of the receiving type.

Poitiers

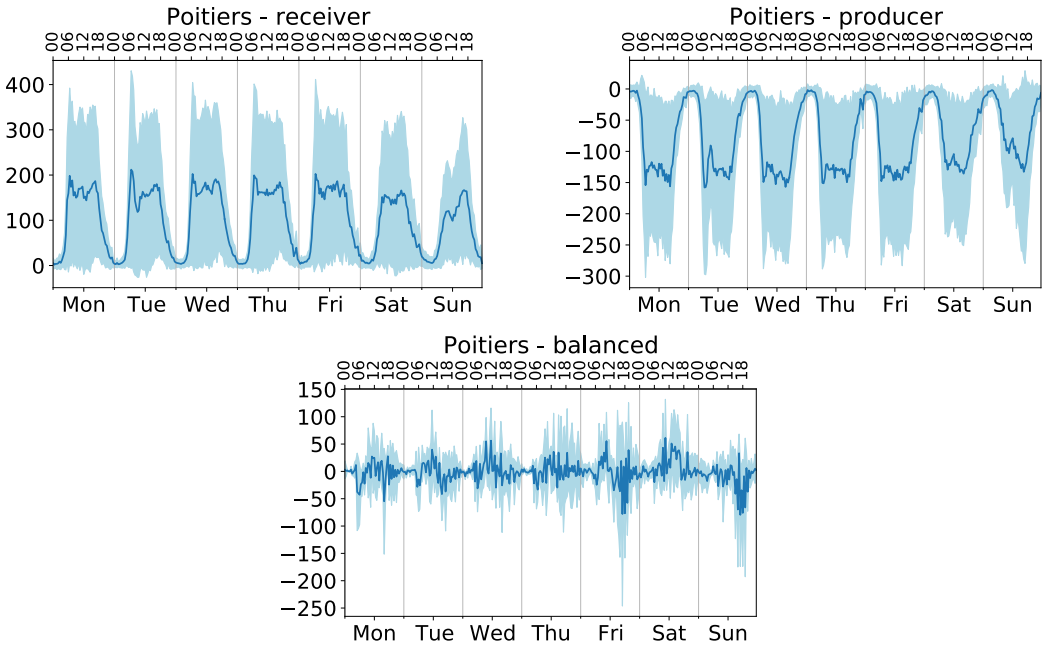


Figure 5.5: Profile in Poitiers: a) receiver, b) producer and c) balanced

Producer and receiver profile. For Poitiers, there are also three different eNodeB profiles in Fig.5.5. However, unlike Cognac, in Poitiers the average amplitude of the

receiver profile eNodeB is higher than the absolute value of the average amplitude of the producer profile. There are also two mains peaks in the day, for both producers and receivers. As for Cognac, the morning peaks for weekdays and weekends are out of sync while the afternoon peaks are synchronized. The timing of these peaks is like the timing of the peaks for Cognac. However, some weekdays present a third peak at mid-day, like Monday and Wednesday. Yet these are less important that the two main peaks in terms of amplitude.

Balance profiles. Concerning the balanced profile, we also have an oscillation of the diff_HO values between negative and positive. We can see that, on Saturday, the average value is mostly positive throughout the day. However, if we look at the standard deviation, we can see that there is still the presence of negative values.

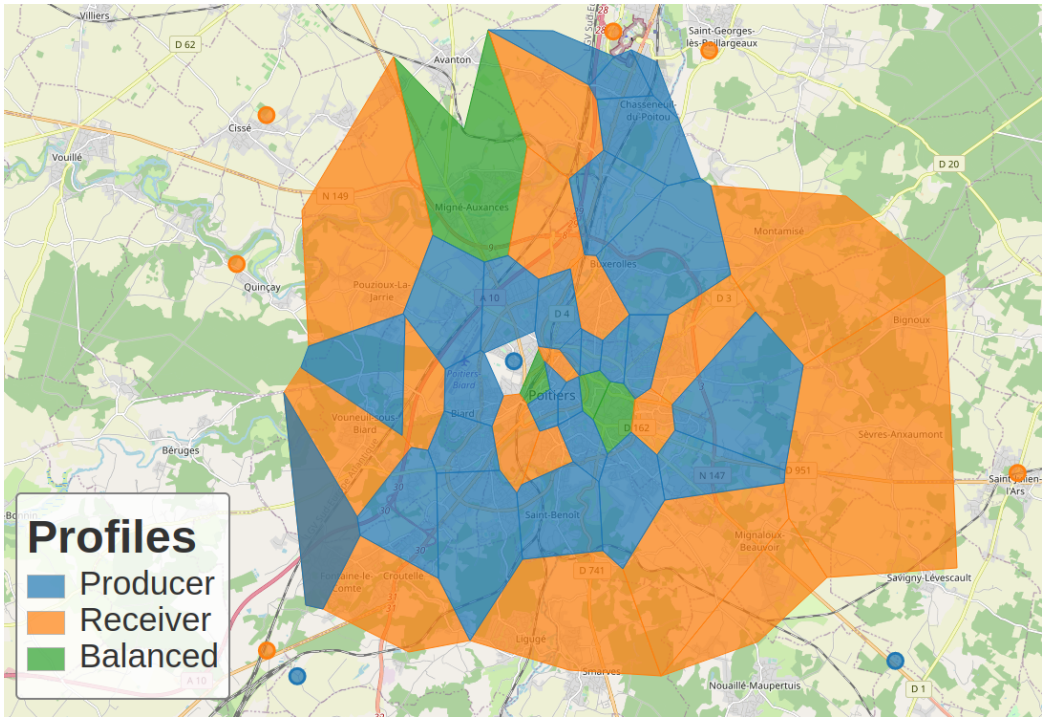


Figure 5.6: Poitiers profile geographic distribution

On Fig.5.6 we can see that in most cases, except in the south represented by the two blue points, the suburbs of Poitiers are receiving profile, and also cover the Futuroscope entertainment park. The city center is a bit more dominated by the producer profile. Just like for Cognac, the Poitiers train station, cell to the right of the blue dot in the middle of Poitiers, is covered by a balanced profile eNodeB. Like for Cognac, we can see that in the majority of cases, the producer profile eNodeB are adjacent to the receiver or balanced profile eNodeB. We can say that, for the geographical areas of Poitiers and Cognac, the city centers have eNodeBs with a producer profile and a balanced profile for the train station, and that the periphery has eNodeBs that are rather receivers.

Lyon

For Lyon, we have the three basic profiles, which also exist in Cognac and Poitiers, but we also have an additional profile called commuting in Fig.5.7.

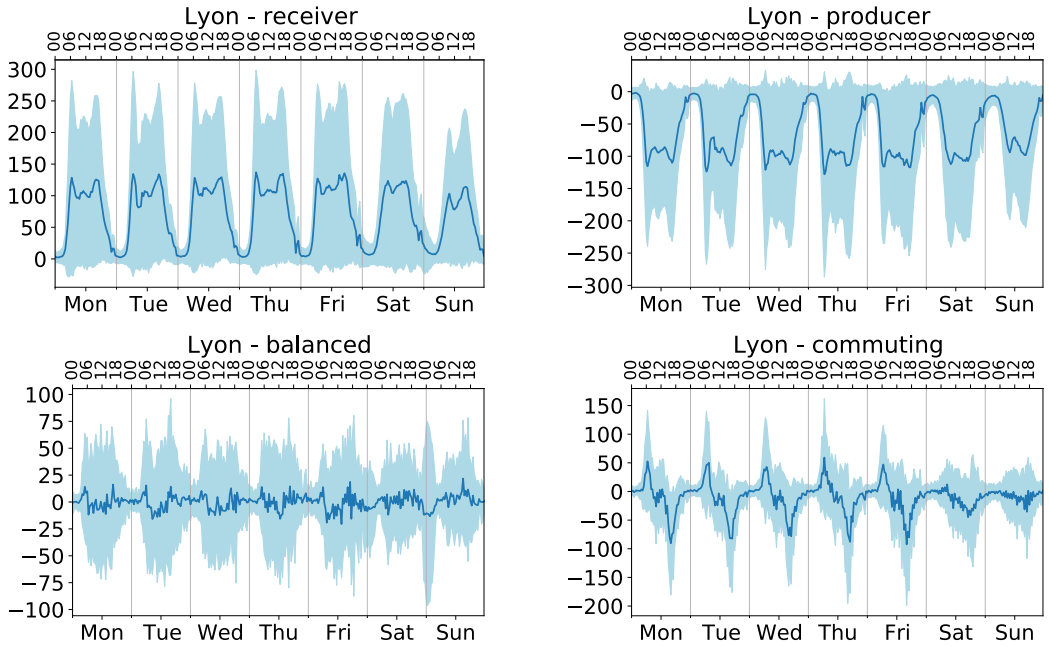


Figure 5.7: Profiles in Lyon: a) receiver, b) producer, c) balanced and d) commuting

Producer and receiver profile. The producer and receiver profiles also have the two daily peaks, but also a third less important peak at mid-day for only weekdays. The time appearance of the two peaks, both for weekdays and weekends, is similar to that of Cognac and Poitiers. Moreover, the average evolution of diff_HO are symmetrical with respect to zero for the producer and receiver profile, in other words, the variation of the absolute value of their average amplitude is similar. We can also see that the average amplitude in Saturday is equivalent to those of the weekday. On the other hand, Sunday presents a small fall of this average amplitude. When we look at the standard deviation, we see that the producer profiles have values that are constantly higher than 0 and for the receiver profile we have values that are constantly lower than zero.

Balanced Profile. For the balanced profile, there is also an oscillation between negative and positive values of diff_HO. We can see that the standard deviation is more important than for Cognac and Poitiers. Some eNodeBs with particular diff_HO behavior can greatly influence this variance. For example, we can see, at midnight from Saturday to Sunday, a large variation in the standard deviation. This variation is due to a single eNodeB in the northeast suburb of Lyon. The Voronoi cell of this eNodeB indicates that this eNodeB covers an industrial area but also a large parking lot with an access to the tramway to directly reach the city centre of Lyon.

Commuting profile. The commuting profile is a profile that is in the same family as the balanced profile oscillations between negative and positive values. However, a pattern clearly appears for this profile, which is not the one for balanced pattern. Indeed, for the days of the week, the value is positive during the morning, with a peak between 6:00 am and 7:00 am. This indicates a rather receiver profile. However, from 10:00 am onwards, we have negative values with a inverse peak between 4 pm and 5 pm. We have here a profile which is rather producer. Then, in the evening, we return to an oscillation of negative and positive values. During the weekend days, we

do not see the same patterns present during the weekdays. What we can see is that there is always a dip between 3 pm and 5 pm.

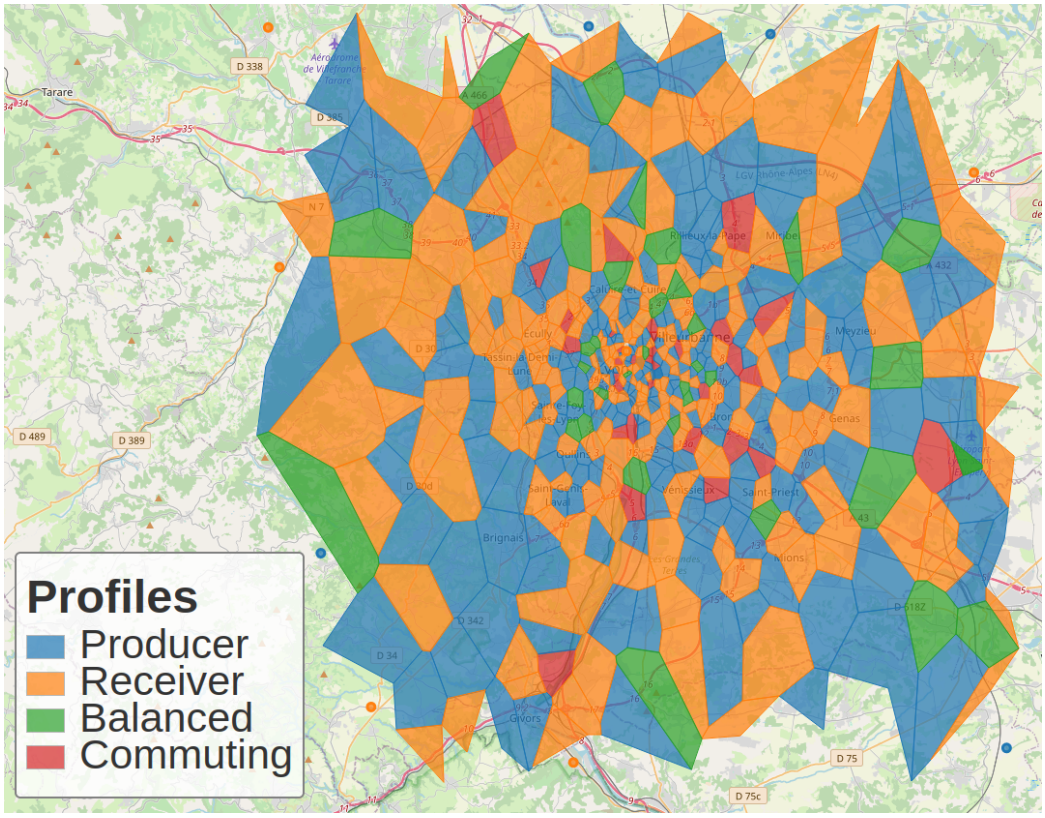


Figure 5.8: Profile in Lyon

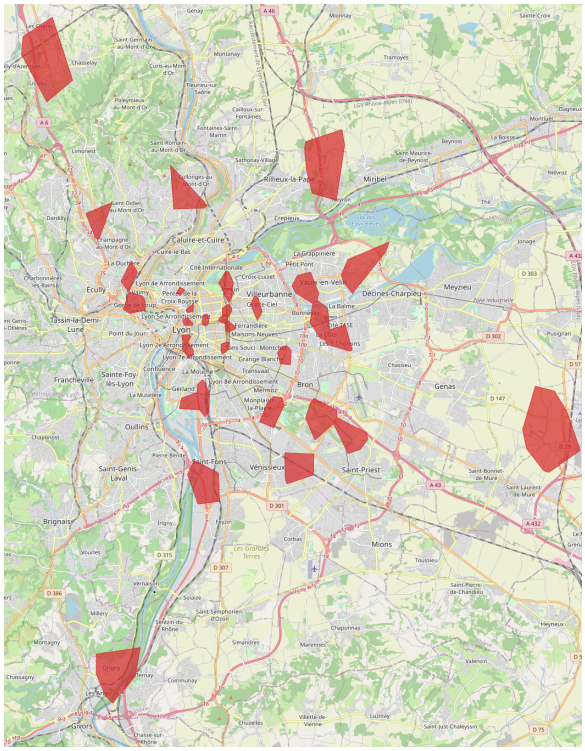


Figure 5.9: Lyon commuting profile geographic distribution

Concerning the spatial distribution of the profiles in Fig.5.8, we always have the Voronoi cells of the receiver profiles which are adjacent to the producer cells. If we

look at the cells with a commuting profile in Fig.5.9, they cover major traffic routes such as portions of highways or major railroads. This profile also covers part of Lyon's main train station, Lyon Part Dieu. It can be said that, to some extent, the commuting profile also integrates the balanced profile existing in smaller geographical areas, like Cognac or Poitiers e.g. the coverage of train stations.

Paris

Paris presents the three basic profiles present in all the different study areas, and the balanced profile is quite similar to balanced profile in Lyon. It also presents the commuting profile that appeared in Lyon. In addition, Paris presents two variants of the commuting profile, inverse commuting profile and morning movement profile, as shown in Fig.5.10.

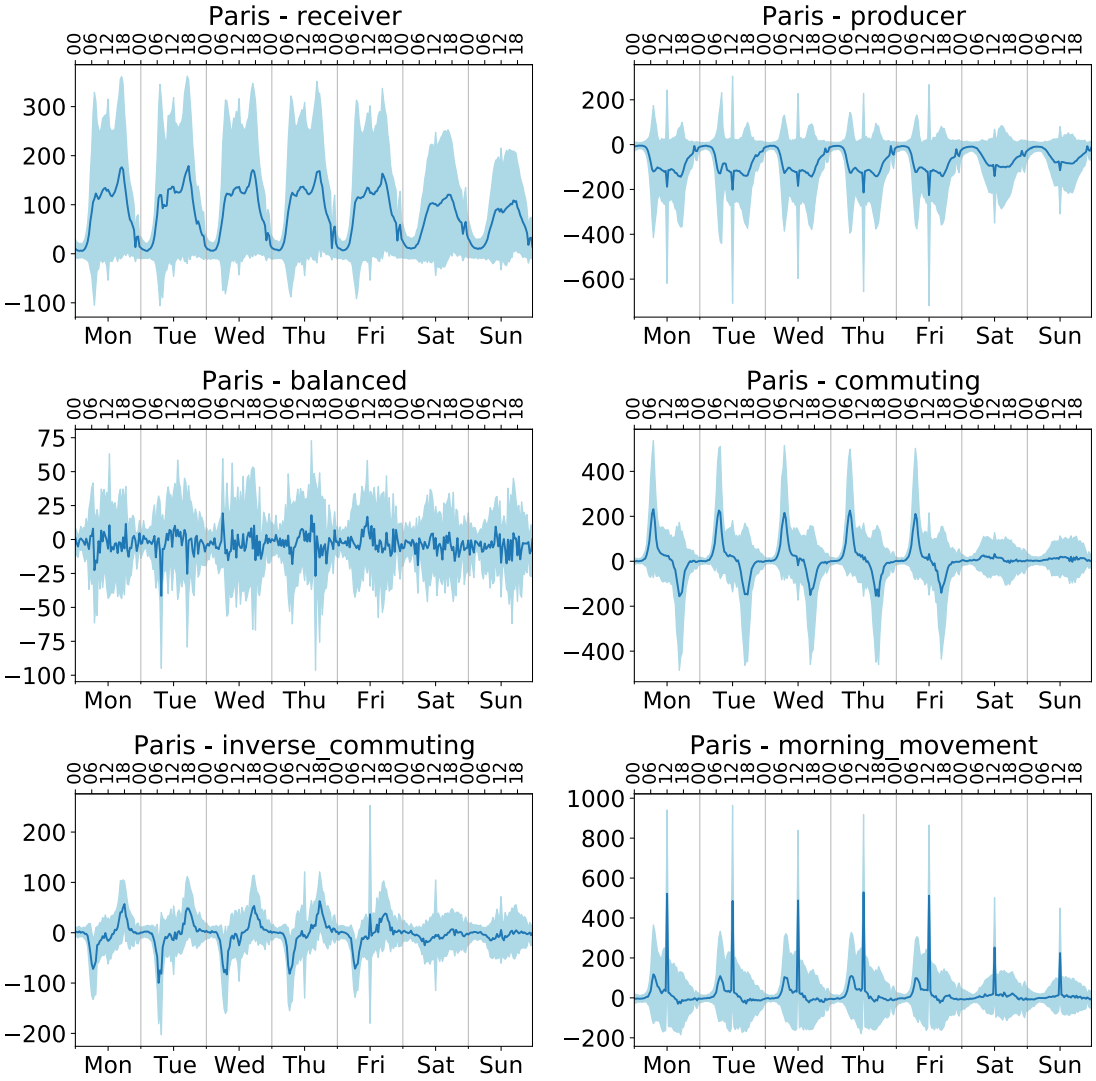


Figure 5.10: Profile in Paris: a) receiver, b) producer, c) balanced, d) commuting, e) inverse commuting and f) morning movement

Receiver profile. For the receiver profile, we have three peaks. We have seen the existence of these three peaks in Lyon and Poitiers but the mid-day peak was less important. Indeed, in Paris during the weekdays, the first morning peak, which takes

place between 6:00 and 7:00 am, has a smaller amplitude than the one which takes place between 10:00 am and 12:00 am. The afternoon peak occurs between 5:00 and 6:00 pm. We can see that the afternoon peaks during the weekdays occur later in Paris than in the other three study areas. For the weekend days, there are only two peaks. The morning peak occurs between 10:00 am and 12:00 pm and the afternoon peak between 4:00 pm and 5:00 pm. In relation to the existence and timing of the three peaks, we can assume that the second morning peak present in the weekdays corresponds to the only weekend morning peak. In this case, there is a real attenuation of the morning peak, until it disappears during the weekend. This assumption can also be made for Lyon and Poitiers. We can see that the average evolution, both on weekdays and on weekends, is always positive. Nevertheless, if we also look at the standard deviation, we have some values that are negative. These values are more important in the morning, at the time of the morning peak. This indicates a profile that is quite heterogeneous especially in the morning, for the days of the week.

Producer profile. The producer profile is quite different from the producer profile of the other study areas. Indeed, here we have three inverse peaks for the days of the week, however the central peak, between the morning peak and the afternoon peak, is the most important peak. This peak, quite clear, occurs at 12:00 pm. The time of appearance of the other peaks, morning, and afternoon is similar to the time of appearance of the latter with the receiver profile. For the weekend days, on the average evolution, we have only one peak, the one that occurs at 12:00 pm on weekdays. However, if we look at the values of the standard deviation, we can see the existence of the (inverse) peak in the afternoon, which takes place between 3:00 pm and 4:00 pm. We can see that the average evolution, whether it is during the weekdays or the weekends, is always negative. However, if we also look at the standard deviation, we have some values that are positive. These values are more important during the three peaks of the day. This indicates a profile that is quite heterogeneous, especially at the time of appearance of the peaks.

Commuting profile. Similar to the Lyon profile commuting, the Paris commuting profile has a peak in the morning and a reverse peak in the afternoon for weekdays. However, unlike Lyon, there is less noise between the peak and the reverse peak. Another similarity with Lyon is that, between the afternoon peak and the morning peak, during the evening, there is a continuous evolution around zero between 8:00 pm and 4:00 am. We have an almost perfect balance. The morning peak occurs between 7:00 pm and 7:30 am, a difference of 30 minutes to one hour with those of Lyon. For the reverse peak in the afternoon, this occurs, like in Lyon, between 4:00 and 5:00 pm. A difference with Lyon is that the evolution becomes negative only from 2:00 pm onwards, whereas for Lyon it is between 8:30 am and 10:00 am. Another difference with Lyon is also the amplitude of the peaks and the reversed peaks. Indeed, these peaks are more important for Paris. For the weekend days, there is no longer the presence of peaks and reversed peaks. The average evolution of diff_HO is almost continuous around 0. We have an almost perfectly balanced evolution even if there is a slight tilt towards positive values. Considering the standard deviation, we see that, during the weekend, we have a balanced profile. This shows that the commuting profile is part of the same family as the balanced profile, but with a specific pattern during the week.

Inverse commuting profile. The inverse commuting profile also presents a peak and a reverse peak on weekdays. However, here the reverse peak occurs in the morning and the peak occurs in the afternoon. Therefore, we have a profile that is more of a producer in the morning and a receiver in the afternoon, in contrast to the commuting profile. The reverse peak in the morning occurs between 6:30 and 7:00 am. The afternoon peak occurs between 4:30 and 5:30 pm. We can see from the evolution that the diff_HO values become positive between 12:00 pm and 2:00 pm. This shows even more the inverted side of the profile compared to the commuting profile. Compared to the amplitude of the commuting profile peaks in Paris, the inverse commuting profile is less important, both for the reverse peak and the peak. Concerning the weekend days, we have similar evolution to the commuting profile. We can say that inverse commuting is part of the balanced profile family, with another specific pattern for the weekdays.

Morning movement profile. For the morning movement profile, it presents two peaks for weekdays and one peak for weekend days, one peak in the morning and one at 12:00 pm. The 12:00 pm peak is also present during the weekend. The morning peak occurs between 7:00 am and 7:30 am, this peak has a low amplitude compared to the 12:00 pm peak. From 4:00 pm during the week, we have an average evolution of diff_HO almost continuous until the morning peak. If we take the average evolution of diff_HO, we have in the majority of the cases positive values, therefore we have rather a receiver profile. However, if we consider the standard deviation, we can see that the profile also presents some negative values. Concerning the weekends, we have an evolution similar to the inverse commuting and commuting profiles, except for the presence of the peak at 12:00 pm.

Concerning the spatial distribution of the profiles in Fig.5.11, we always have the Voronoi cells of the receiving profiles which are adjacent to the producing cells. We focus on the spatial distribution of the different commuting profiles in Fig.5.12.

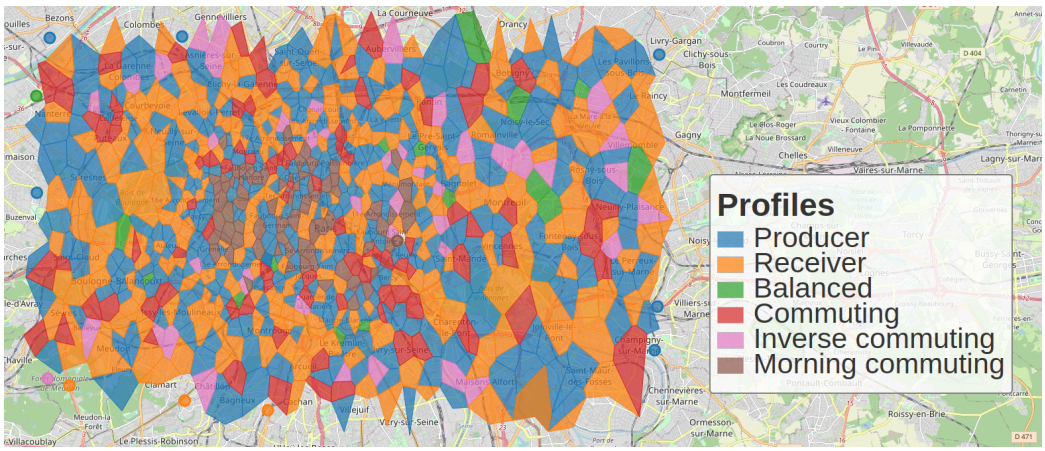


Figure 5.11: Profile in Paris

In Fig.5.12, the city of Paris is delimited by the black continuous line. The commuting profile, in red, covers the major communication axes, some freeway portions, gates for the entrance to Paris. For example, the Porte de Bercy with the A4 highway interchange in the south-east of Paris is one of the communication axes with the highest occupancy rate [125]. It also covers very busy areas such as the first French business district, Paris-La Défense [126] or the very busy roundabout of the Place de

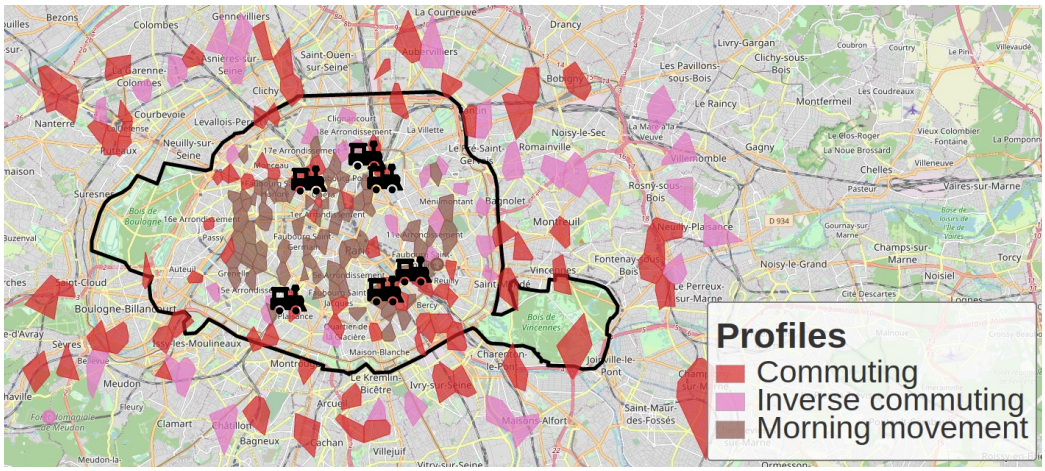


Figure 5.12: Paris commuting, inverse commuting and morning movement profile geographic distribution

l'Etoile [127]. It also covers railway lines. In addition to these major communication axes, the commuting profile also covers the main Parisian train stations, represented by the little train icons in Fig.5.12. These stations are the arrival points of trains coming from all over France and from the international, but also of trains from the Paris suburbs and the Ile-de-France. They are also served by several metro lines and sometimes the RER which also connects Paris to the suburbs.

The inverse commuting profile also covers major communication axes, such as the gateways to the East of Paris like the Porte de Bagnolet or the Porte de Vincennes. It also covers a portion of the A3 highway at the entrance of Paris. It also covers portions of railway lines and metro station in Paris.

The morning movement profile exists only inside Paris intramuros. It covers certain green spaces or parks such as the Champs de Mars or the Luxembourg garden. It covers certain places of activity such as the embassies or the presidential palace l'Élysée, but also a part of the national assembly and the town hall of Paris.

Taking the Voronoi cells of the three types of profile commuting, we cover 29% of the metro station entrances. However, metro stations can have several entrances, which might not all be covered by the same Voronoi cell. These profiles also cover the five busiest metro stations in Paris.

5.4.2 eNodeB categorisation analysis

In the following, we discuss the spatial distribution of the different categories of eNodeBs according to the volumes of HO they process. We have three categories: low, medium, and high. The definition of these three categories was done in an arbitrary way for each cluster. It is also relative to the geographical area. Each cluster can be defined by an interval. The lower bound of the interval corresponds to the average HO volume (in_HO or out_HO) of the eNodeB with the lowest value in the cluster. The upper bound corresponds to the average HO volume of the eNodeB with the highest value of the cluster. From these intervals, we defined the different categories. The low category is defined by the cluster L having the lowest lower bound. All the clusters having, at least, their lower bound included in the interval of L are categorized as low. For the high category, it will be defined by the cluster H having the highest upper bound. All the clusters having at least their upper bound included in the interval of H are categorized as high. The rest of the clusters are

categorized as medium.

In the following, we equate in_HO with out_HO. Indeed, the categories of eNodeBs for their in_HO and out_HO are almost always the same. For Cognac, two eNodeB have different categories between in_HO and out_HO, this represents about 8% difference. The categories that change are adjacent, i.e. from medium to low or from low to medium. For Poitiers, there are also two eNodeBs that have different categories between the in_HO and out_HO. This represents approximately a 3% difference. Here again the categories that change are adjacent. For Lyon, like Cognac and Poitiers, there are only two eNodeBs that have different categories between in_HO and out_HO. We therefore have a lower percentage of differences, at 0.4%, because the number of eNodeBs in Lyon is greater. And for Paris, there are three eNodeBs which have different categories between the in_HO and the out_HO, which represents 0.3% difference.

Cognac

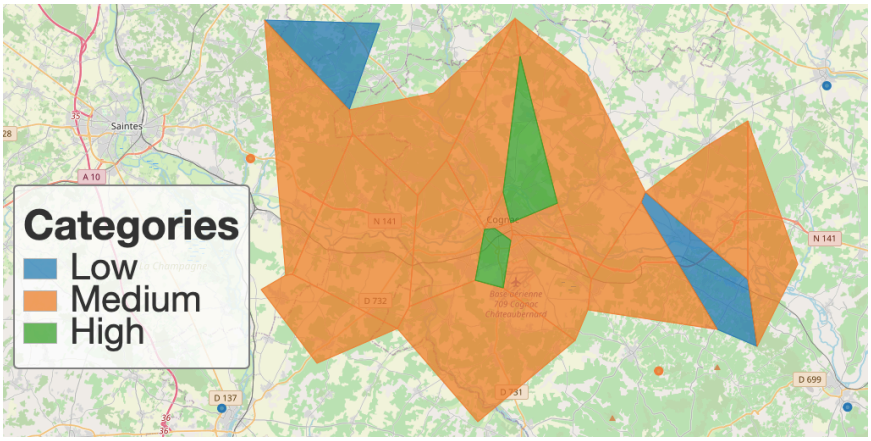


Figure 5.13: Cognac categories

As expected, Cognac is dominated by the medium category, in orange in Fig.5.13. The high category, in green, covers part of the city centre of Cognac, and part of the railroad leading to the train station. The lower category, in blue, covers the outlying suburbs of Cognac, and part of the railroad and a departmental road.

Poitiers

Like Cognac, Poitiers is dominated by the medium category in Fig.5.14. The high category covers a part of the city center, and mostly the northern part of Poitiers. On this northern part, it covers the high speed train line and the A10 highway. For the low category, it covers a part of the city center, but also surrounding villages, departmental roads and other railway lines and a southern part of the A10 freeway.

Lyon

Like Cognac and Poitiers, for Lyon the medium category, in orange in Fig.5.15, is the most important category. The high category is interesting in the case of Lyon. Indeed, it covers the eastern part of the study area. This part corresponds to the main communication routes, freeway to the Alps, freeway serving the international airport of Lyon, Saint Exupéry, and the high speed train line serving the airport.

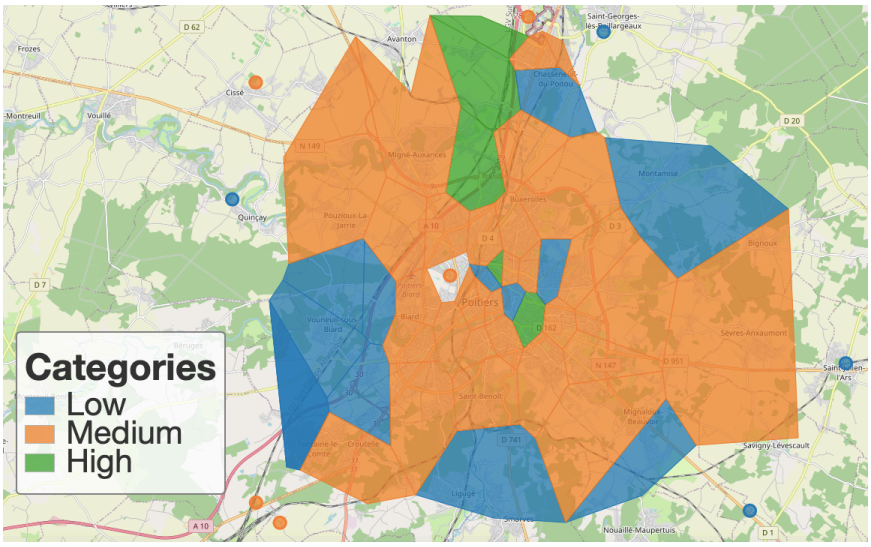


Figure 5.14: Poitiers categories

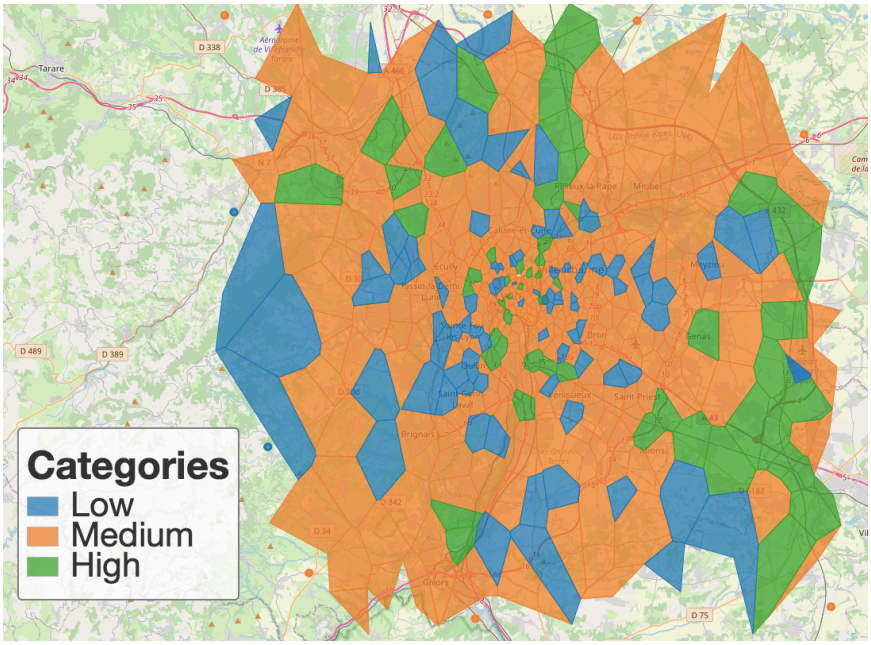


Figure 5.15: Lyon categories

This category also covers the northern part of the study area. In this part, it covers the high-speed train line to or from Lyon's main train station, Part-Dieu. It is in this part that there is also the connection with the high speed train line coming from the airport. The high category also covers, areas in the center of Lyon, railroads in the city of Lyon, but also highways, part of the A7, part of the A6. The low category covers some areas of the city of Lyon and the surrounding cities, but also areas further away from the city of Lyon, covering departmental roads and villages. It also covers a part of the A89 freeway.

Paris

The medium category is the most important in all the study areas, we see it also in the case of Paris in Fig.5.16. The high category is not very important, it covers three zones, two of which are gates of Paris. It covers the Porte de Bercy with the A4

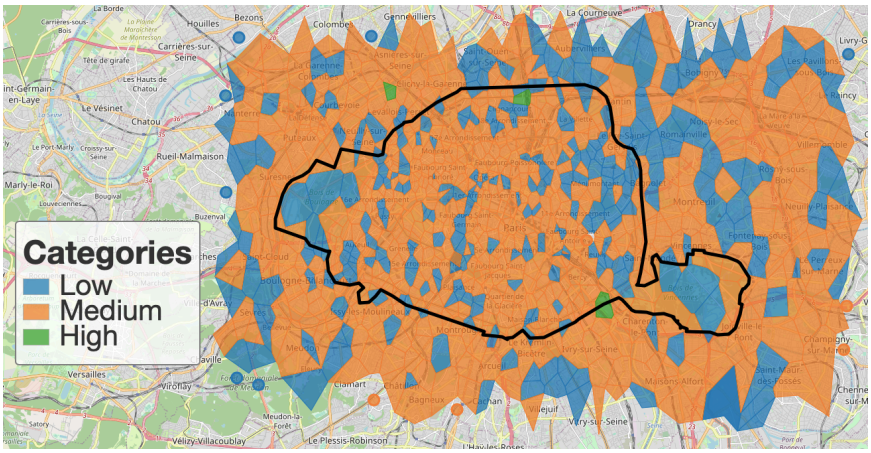


Figure 5.16: Paris categories

freeway interchange. The low category covers some green spaces and parks, like the Bois de Boulogne, the Bois de Vincennes, the park of Buttes Chaumont, the garden of Luxembourg or even the cemetery of Pére Lachaise. However, it also covers portions of freeways, railroads, and major train stations.

5.4.3 Link between categories and profiles

	Cognac			Poitiers			Lyon			Paris		
	L	M	H	L	M	H	L	M	H	L	M	H
Producer	0.25	0.5	0.25	0.2	0.74	0.06	0.22	0.66	0.12	0.36	0.64	0
Receiver	0.15	0.85	0	0.42	0.5	0.08	0.2	0.7	0.1	0.33	0.66	0.01
Balanced	0.34	0.66	0	0.25	0.5	0.25	0.17	0.69	0.14	0.56	0.44	0
Commuting	0	0	0	0	0	0	0.06	0.79	0.15	0.28	0.71	0.01
I commuting	0	0	0	0	0	0	0	0	0	0.4	0.6	0
M commuting	0	0	0	0	0	0	0	0	0	0.09	0.91	0

Table 5.1: Proportion of profiles in each category

Table 5.1 shows the importance of the medium category. Indeed, we see that, for any profile except the balanced profile for Paris, they belong to at least 50% to the medium category. The second most important category is the low category, which is even the most important for the balanced profile of Paris. The high category has the fewest representatives. However, for some profiles, it is sometimes at least as important as the low category, for example the commuting profile in Lyon, the producer profile in Cognac or the balanced profile in Poitiers.

Table 5.2 shows us the importance of the producer and receiver profiles. The most important profile for each category is always either a receiver or a producer. We can see for Lyon that the distribution of balanced and commuting profiles among the categories are in the same orders of magnitude. For Paris, the commuting profile has a greater presence among the different categories. For example, for the high category, only the receiving and commuting profile are present. There is also a relative importance of the morning movement profile especially for the medium category.

			Producer	Receiver	Balanced	Commuting	I commuting	M commuting
Cognac	In	L	0.4	0.4	0.2	0	0	0
		M	0.24	0.65	0.11	0	0	0
		H	1	0	0	0	0	0
	Out	L	0.2	0.6	0.2	0	0	0
		M	0.29	0.59	0.12	0	0	0
		H	1	0	0	0	0	0
Poitiers	In	L	0.35	0.59	0.06	0	0	0
		M	0.61	0.33	0.06	0	0	0
		H	0.4	0.4	0.2	0	0	0
	Out	L	0.27	0.66	0.07	0	0	0
		M	0.63	0.32	0.05	0	0	0
		H	0.4	0.4	0.2	0	0	0
Lyon	In	L	0.49	0.43	0.06	0.02	0	0
		M	0.42	0.43	0.07	0.08	0	0
		H	0.46	0.36	0.09	0.09	0	0
	Out	L	0.48	0.44	0.06	0.02	0	0
		M	0.42	0.43	0.07	0.08	0	0
		H	0.46	0.36	0.09	0.09	0	0
Paris	In	L	0.46	0.32	0.03	0.1	0.07	0.02
		M	0.39	0.32	0.01	0.12	0.05	0.11
		H	0	0.67	0	0.33	0	0
	Out	L	0.45	0.33	0.03	0.1	0.07	0.02
		M	0.39	0.31	0.01	0.12	0.05	0.12
		H	0	0.67	0	0.33	0	0

Table 5.2: Proportion of categories in each profile

5.5 Conclusion

To conclude, we have shown the existence of different profiles describing the behavior of eNodeB in terms of HO. There are three basic types of profiles. Cities with low or medium density have only these three basic profiles. When the population density of the city increases, other profiles appear. For a large regional metropolis, there is an additional profile in addition to the three basic profiles. This profile represents the main communication axes. For a global metropolis, in addition to the three basic profiles and the one that appeared in the regional metropolis, we have the appearance of two other profiles that are derivatives of the commuting profile. We also categorized the eNodeBs according to the volume of HO they process. By considering both HO volume (categories) and variation (profiles), we distinguished multiple classes of eNodeBs and analyzed their placement on the maps of four urban regions of different sizes. Beyond obvious sociological interest, this can also be used for network management purposes, as detailed in Chapter 6.

Chapter 6

Slicing approach for mobility management

6.1 Introduction

The traditional approach for a mobile network is denoted as "one size fits all" [128]. This approach consists in dimensioning the network in a static way, in order to satisfy the needs of each offered service. Indeed, the various services do not have the same needs, and thus the same quantity of resources necessary to their operation. The network is sized so that it can meet the demands of each service at peak times, i.e. when the maximum demand is recorded. The network is therefore oversized to be able to meet the needs of each service, at their peak use. This over-sizing leads to an over-reservation of resources that remain not used. Indeed, the network reserves the number of resources needed for the peak traffic, which implies that during periods of lower traffic, these reserved resources are wasted.

This over-dimensioning and undifferentiated treatment are also present for the management of mobility. Indeed, the number of HO evolving during the day, the network is dimensioned in such a way to satisfy the peak of HO. Moreover, some users are static and therefore do not need the same operations as mobile users. However, in the current system, static users are controlled in the same way as mobile users. To better adapt the resources to be used, the concept of network slicing is interesting. This concept makes it possible to adapt the resources on the data plane according to the type of service. We can thus ask the following question: can control plane slicing also bring benefits? Indeed, mobility management is realized on the control plane. We can also ask, what is the gain obtained by creating a slice for static users, different from the slice for mobile users? Also, what is the gain if mobility management resources are dynamically allocated?

To answer these questions, we will first look at the network slicing principle. Second, we will present the data that will allow us to have an indication of the number of static users. Then we will present the gain in terms of resources that can be saved by the implementation of a slice for static users. Finally, we will present the resource saving realized during a dynamic reconfiguration for a slice dedicated to mobile users.

6.2 Network slicing

Slices are logical networks over a single physical network. Each logical network can be configured to meet the specific needs of the services it accomodates. Slices are implemented in the 5G network to enable efficient and flexible use of resources.

6.2.1 Service definition

There is a diversity of services that can pass through the 5G network. Among these, the main are [129, 130]:

- Massive Machine Type Communications (mMTC) privileges the density of terminals that can be attached to the network, but also their energy efficiency. Internet of Things (IoT) services can be included in this service area.
- Enhanced Mobile Broadband (eMBB), which focuses on user throughput and therefore sacrifices can be made in terms of latency, density of terminals that can be managed, or reliability. Streaming services belong to this type of service.
- Ultra-Realiable Low Latency Communication (URLLC) is concerned with reliability, latency and mobility. Intelligent transportation or even remote surgery can fit into this domain.

These services being heterogeneous, they do not have the same needs and constraints. They have heterogeneous QoS requirements and heterogeneous Key Performance Indicators (KPIs). Network slicing is a solution that allows to provide each service with a QoS and to satisfy their Service Level Agreement (SLA). The SLA, here, describes the slice and the constraints imposed on it.

6.2.2 Slice isolation

Generally, in a 5G network we have the coexistence of several slices, i.e. virtual networks, one per service, sharing a single physical network. The isolation of the resources used between these various slices is consequently important [100]. The isolation makes it possible to guarantee the performance, as each slice must fulfill the conditions to satisfy the service. It does not consider the possible problems on the other slices. Isolation also plays a role in security and privacy. Attacks on a given slice must not affect the other slices. It must also facilitate the management of the slices, as each slice must be able to be monitored and managed independently.

6.2.3 Network slicing specification

The implementation of the network slicing requires several elements [131]:

Network slice template

Network slice template (NST) is a representation of the virtual network functions used by a service and the resources needed to meet all the requirements of the various functions. In other words, the NST is a list of virtual network functions and the different hardware resources needed to implement these virtual network functions. It is practically the specification of the slice.

Network slice instance

Network slice instance (NSI) is an instance created from NST. It is the effective implementation of the various network functions required for a network service (NS). These NS are virtual network function (VNF) chains and are deployed by Network Function Virtualization (NFV). In the case of a physical network function (PNF), which is the physical equivalent of VNF, the network functions and services are performed by dedicated hardware. The NS are defined by the number of VNFs and their order in the chain. This chain is allocated on Network Function Virtualisation Infrastructure (NFVI) [132]. The NFVI is the physical infrastructure on which the network function virtualization is deployed [128]. The NFVI is therefore characterized by the hardware resources that compose it.
A network slice allows the management and monitoring of the NSI. SLA compliance must be guaranteed for each service. The QoS that is provided must also be monitored.

6.2.4 New components

The management of slices requires new components in the mobile network architecture.

Management system

The management system is constituted of by two components. The Operation Support System (OSS) manages, allocates, configures, and supervises the resources that are used by the different VNFs. The Business Support System (BSS) adapts the deployment of the service to the user requests.

Management and orchestration

The NFV management and orchestration (MANO) component is composed of the orchestrator, VFN managers, and virtualized infrastructure manager. It handles the management tasks on the VNFs, their deployment and their supervision. It also manages the life cycle of the physical and virtual resources that are on the virtualization infrastructure. MANO deploys and supervises the virtual machines and the resources used, in addition to the resources available.

6.2.5 Mobility management in sliced architectures

Network slicing introduces complexity to the realization of HO. Indeed, in addition to the various HO execution scenarios, cf. Appendix A, network slicing introduces the inter slice HO [133]. This inter slice HO consists of changing the assigned slice while the user is still trafficking. This type of HO does not necessarily imply user mobility.

The management of mobility according to the services provided by the different slices can also be complicated [134], since the different services have different constraints.

For mobility management, related work evaluates the mobility that users may have by profiling their movement speed [135] using deep learning techniques. Knowing this, they can assign network functions that may be required for different travel speeds.

Our work is complementary to this idea, while remaining more general. Indeed, we want to show the advantages that can present the implementation of a slice at

the level of the control plane. However, we also want to show the benefits of having different slices for static and mobile users, without proposing methods for detecting user mobility and slice implementation.

6.3 Data

To be able to evaluate the gains that the implementation of slices on the control plane can have for static users, it is necessary to have an idea of the proportion of static users seen by the network. By using the HO data presented in Chapter 5.2 together with new data on the presence of users, we can deduce the static users. The user presence data is as follows.

We have the same geographical area as in Chapter 5, namely Cognac, Poitiers, Lyon, and Paris. The data covers the same 1-month period, from 16 March to 16 April 2019. We have 4G user presence data in line with the HO data. This presence data describes the number of users in the eNodeB cells who are connected. The network counts the number of International Mobile Subscriber Identity (IMSIs) present at the eNodeB in one minute. IMSI is a unique international number that identifies a subscriber. However, we do not have the IMSI of the users, just the number of different IMSIs that are connected. We therefore do not have the possibility to identify the subscribers. The user presence data have a granularity of one minute, which is different from HO data, aggregated over 30 minutes intervals.

6.4 Static users statistics

To deduce the proportion of mobile users, and thus the one of static users, in the presence data, we compute the ratio of users who make an HO. However, as HO data and presence data have different temporal granularities, it requires a pre-processing step to calculate the ratio. This pre-processing aims to put the two types of data on the same temporal scale.

In the following we are working on the typical days base, created by the same method seen in Chapter 5.

6.4.1 Temporal scaling pre-processing

The granularity of the HO data is 30 minutes, and that of the presence data is one minute. We mentioned in Chapter 5 that, during the 30 minutes period, a user can make several HOs at an eNodeB. The solution we have chosen to put the two types of data on the same time scale is to compute an average per minute of the HO values. This allows us to compare the HO values with the values of the users present while reducing the probability that a user is counted multiple times. To do this, we simply divide the HO values by 30. Afterward, we keep the user presence values only every 30 minutes to have a similar granularity to the HO data.

6.4.2 Proportion of mobile users

Taking the median typical day of each study area, we can show the evolution of the proportion of users making an HO among all users present, which represents the ratio

defined below. Formally we have:

$$S(t) = \frac{(in\_HO[t] + out\_HO[t])/2}{U[t]} \tag{6.1}$$

We have the average of the in_HO and out_HO over the number of users present $U[t]$. We take the average of the in_HO and out_HO to avoid counting the same HO twice in two different eNodeBs. Indeed, an in_HO on an eNodeB is an out_HO on another.

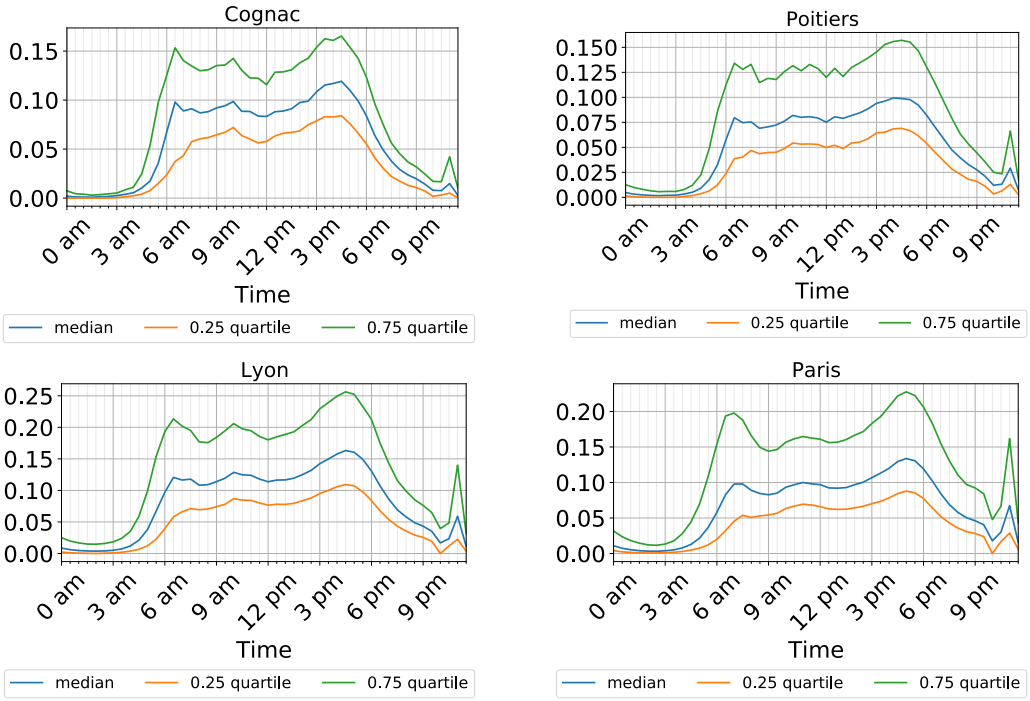


Figure 6.1: HO proportion evolution zone a) Cognac, b) Poitiers, c) Lyon and d) Paris

On the different curves of Fig.6.1, we observed the evolution of the proportion of HO in the present users along a typical day on the four study areas. We have three daily representations per geographical area. The green curve represents a typical daily 0.75 quartile. The blue curve is a typical median day, and the orange curve is a typical 0.25 quartile day.

We can see that, for all the different typical days in the different geographical areas, there are two main peaks in the day, one in the morning and one in the afternoon. The morning peak occurs at 6.30 am for all geographical areas, except for Paris, where the peak occurs at 7:00 am. The afternoon peak occurs at 4.30 pm, although the Paris peak occurs later, at 5:00 pm. There is also a peak at 11:00 pm in all geographical areas. As we seen in Chapter 4 user presence data also have peaks which also exist in the data we use in this Chapter. However, these peaks do not occur at the same period as HO peaks. For example, on a typical day in Paris, Lyon, and Poitiers, the number of present users increases throughout the morning until it reaches a peak between noon and 2pm. For Cognac typical day, there are two peaks in the user presence number, one at 8.30 am and the other at 3 pm.

If we take the typical 0.75 quartile day, the proportion of HO at peak ranges from 0.15 to 0.25 depending on the geographical area. If we take the typical median day, the proportion of HO is lower, ranging from 0.10 to 0.16. We can therefore say that,

in the majority of the time, users are static (do not make HO). Processing static users differently from mobile users, for example for signal quality measurement reports to the eNodeB, is coherent given the large proportion of static users.

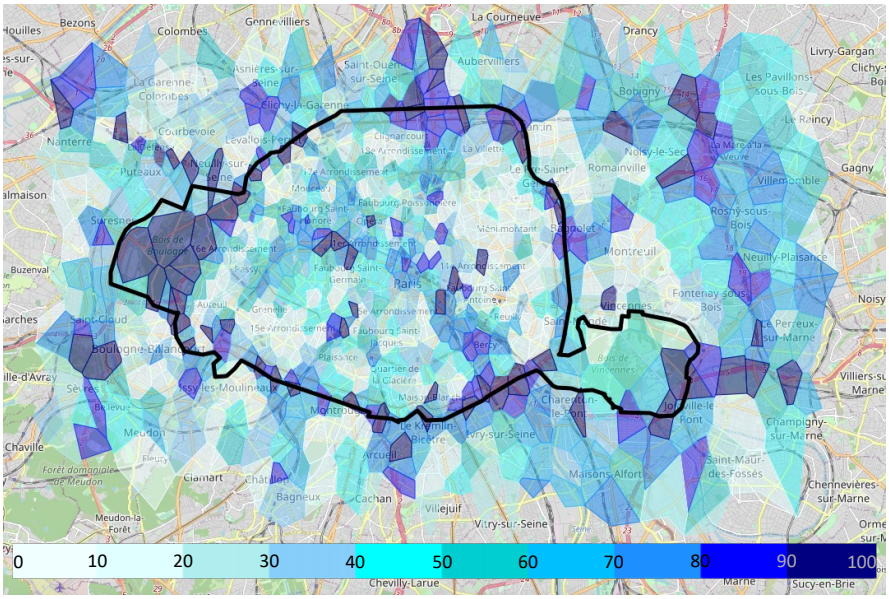


Figure 6.2: Paris spatial distribution of HO proportion

Looking at the spatial distribution, we can see that the highest proportions of HOs are on major roads. This can be seen in the example of Paris in Fig.6.2. Here we have averaged the proportion of HO for each eNodeB. Then we classify the different proportions in different intervals. The lowest 10% are classified in the interval [0-10], and so on, until the highest 10% in the interval [90-100]. Each interval is represented by a colour, ranging from white for the interval [0-10] to dark blue for the interval [90-100]. Each eNodeB is represented by a Voronoi cell approximating the area covered by the eNodeB.

We can see that the eNodeBs that cover the main communication routes are among the 30% of eNodeBs with the highest proportion of HOs among the users present. An example of this is the ring road around the city of Paris, but also the road between Paris and the business district La Défense. We can see that in the Bois de Boulogne, the western part of the city of Paris, we have eNodeBs with a high proportion of HO, whereas in the Bois de Vincenes the proportion of HO is lower. This can be explained by the presence of a national road that passes in the middle of the Bois de Boulogne.

This higher proportion of HO on major roads is also found in the four other geographical areas.

6.5 Static users slice

The cost of static users to the network depends on their number. These costs are induced by the preliminary phase at the initiation of the HO, described in Appendix A. Indeed, each user, whether static or mobile, is continuously executing this preliminary phase, conducting channel quality measurements and transmitting these reports periodically to the network (with the same period for all users). By setting up a static slice, we can potentially reduce the number of network functions required, since we no longer consider static and mobile users in the same way. The UE will therefore report fewer signal quality measurements and therefore conduct fewer computations.

In addition, there is the existence of the phenomenon of HO ping pong [93]. This phenomenon corresponds to the fact that a user, even a static one, can still trigger an HO to a cell and then shortly after do an HO again to return to the initial cell. A static slice can therefore be interesting to reduce the number of HO.

The cost reduction for the network can be achieved by reducing the number of operations, and therefore the reduction of measurement reports. Indeed, the measurement reports generate communication costs, if only between the eNodeB and the UE, but also computation costs. The eNodeB must process the measurements reported by the UE, and then calculate whether it should trigger an HO. Intuitively, a static user may not perform an HO while maintaining communication. Our objective in this work is not to detect if a user should be considered static or mobile. We consider such a mechanism already exists in the network. We do not focus either on the exact calculation of the computation and communication gain this approach gives for a particular user. In fact, mechanisms where the period of measurement reports is adapted to the user mobility already exist in the literature [136, 137, 138]. Our objective is to estimate, at the level of a geographical region, the impact of a static slice. We can calculate, artificially, the cost reduction, given by the implementation of a static slice, by considering a reduction of the number of static users managed by the network since this approach decreases the number of reported measurements per user.

Formally we can model this reduction as:

$$Cost_s = \alpha \cdot N_s \tag{6.2}$$

In Eq.6.2 we have the cost represented by the static users which is expressed by the number of static users N_s weighted by α, a coefficient. This coefficient represents the reduction in the number of operations, artificially approximated by the reduction in the number of static users. An $\alpha = 1$ corresponds to the current situation, where static users are treated as mobile users. An α less than one corresponds to the reduction in the number of measurements performed by static users in a static slice.

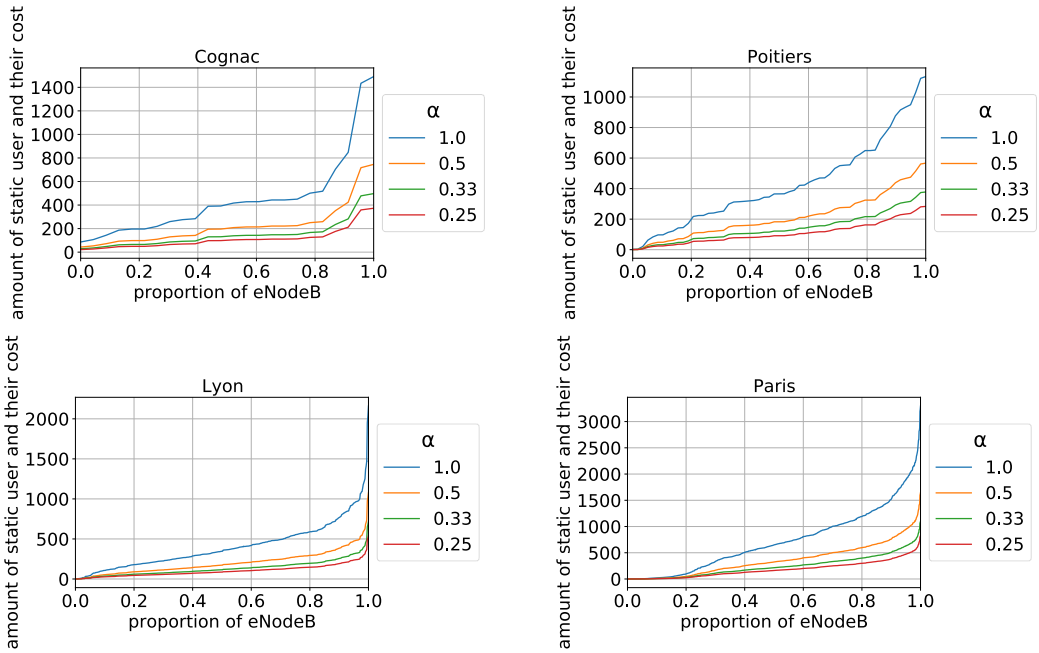


Figure 6.3: Cost reduction by α coefficient a) Cognac, b) Poitiers, c) Lyon and d) Paris

In Fig.6.3 we can see the distribution of the average cost reduction as a function of α, depending on the eNodeBs present in the different geographical areas. If we take an α coefficient equal to one, we have the ordered distribution of the current costs per eNodeB. For Cognac, we have a slight increase on the first 75%, we go from about 81 static users, for the eNodeB with the lowest number of static users, to about 450 static users. Then we notice a sharp increase, to go to approximately 1490 static users for the eNodeB having the biggest number of static users. Here we can see, that for Cognac, a small number of eNodeBs concentrates large number of static users.

For Poitiers, we have a linear increase for the first 80%, then we have another linear increase but with a more important slope. We can see that the average concentration of static users is lower for Poitiers than for Cognac for eNodeBs with a large number of static users.

For Lyon and Paris, we have an exponential distribution. Therefore, we have a high concentration of static users on few eNodeB. For Paris, and to a lesser extent for Lyon, we have eNodeBs with a small average number of static users.

The gain obtained for different values of α is straightforward to compute with our simple model. We simply indicate it in Fig.6.3 to give numerical intuition for this gain and its distribution.

6.6 Mobile users slice

The current system, set up by the operators, is static and designed to meet HO needs during the peak of the day. Therefore, there is no reconfiguration during the day. There is also no differentiation made between in_HO and out_HO.

We imagine thus a mobile slice which allows the reconfiguration of the resources during the day with an objective of economy of these resources. Indeed, the amount of reserved resources is no longer based on the peak of the day.

Concerning the types of HO, we distinguish three types of possible scenarios while using a dynamic reconfiguration of resources. The first scenario is to make no distinction between in_HO and out_HO, as is assumed to be done today. The second scenario is to consider the difference in resources that can be used by the in_HO and the out_HO. And the third scenario, it is the use of the profiles defined in Chapter 5 for the assignment of the resources for in_HO and out_HO.

6.6.1 Dynamic reconfiguration

Dynamic reconfiguration of resources can be done periodically every t time. This gives us reconfiguration windows with several window sizes. Thus, we considered reconfiguration window sizes of 0.5, 1, 2, 3, 4, 6, 8 or 12 hours. The resources reserved in each window need to cover the maximum number of resources consumed during the reconfiguration window. Thus, the larger the reconfiguration window size, the closer the number of resources reserved is to the current system and fewer resources can be saved. Without loss of generality, we consider that, in the current system and in the proposed sliced approach, the peak demand for a given period can be predicted with good accuracy.

In Fig.6.4 we can see an example of different reconfiguration windows: 0.5h in blue, 3h in orange and 12h in green. Each curve represents the proportion of the reconfiguration cost compared to the current system. For example, the green curve from 00:00 am to 12:00 pm consumes the equivalent of 0.8 proportion of the current

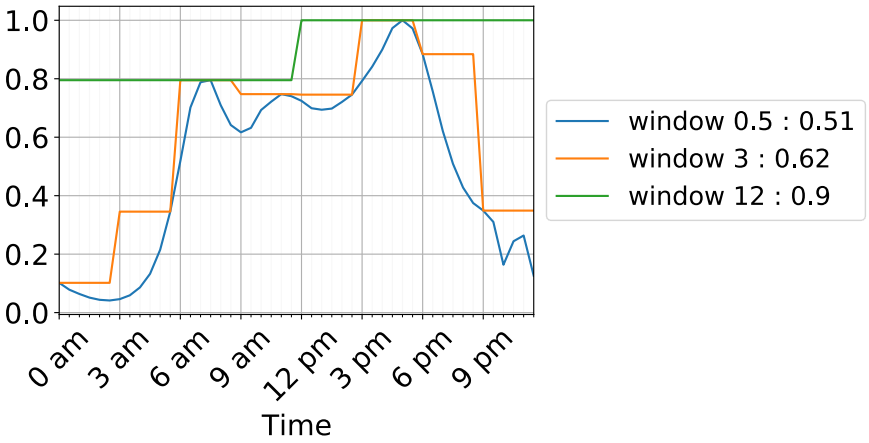


Figure 6.4: HO cost for different reconfiguration windows. Average values for 24h are provided in the legend

system; in other words, we only reserve 80% of the resources we would normally reserve.

Looking at the blue curve, for the reconfiguration window of 0.5h, we can see that we do not have the step function that we have on the reconfiguration windows of 3h and 12h, and all the other reconfiguration windows. This is due to the fact that our HO data have a granularity of 30 minutes. It is thus this window value which is used as base for the resources reservations for the other windows of configuration. The 30 minute reconfiguration window represents an average cost ratio of 0.51 compared to the current system.

For the 3h reconfiguration window, in orange the value of each step corresponds to the maximum values that the blue curve has within the three-hour window. With a reconfiguration window of 3 hours, we only reserve the maximum resources for 3 hours, contrary to the current situation where this number of resources are reserved all day. This allows to reserve only 0.62 of the resources that we would normally reserve.

For the reconfiguration window of 12h, in green, we make only two reconfigurations of the reserved resources for one day. With this reconfiguration window, we reserve at least for twelve hours the same quantity of resources that we would normally reserve. However, we can still have a reduction of the reserved resources and therefore a reduction of the costs. In this example, we have an average cost ratio of 0.9 compared to the current system.

6.6.2 No distinction between handover types

The first scenario that can be made of the in_HO and out_HO when dynamically reconfiguring the resources to be reserved is to consider them equivalent. In this case, to be able to define the quantity of the resources to be reserved, we sum the in_HO and the out_HO. We suppose that this scenario is the one which is currently implemented in the operator networks, without the reconfiguration during the day, because it does not make intervene any other analysis, therefore less complex to implement.

On Figure 6.5, we can see the proportion of the reconfiguration cost, compared to the current system, according to the size of the reconfiguration window. We show the average evolution of the cost on the four study areas: Cognac in blue, Poitiers in

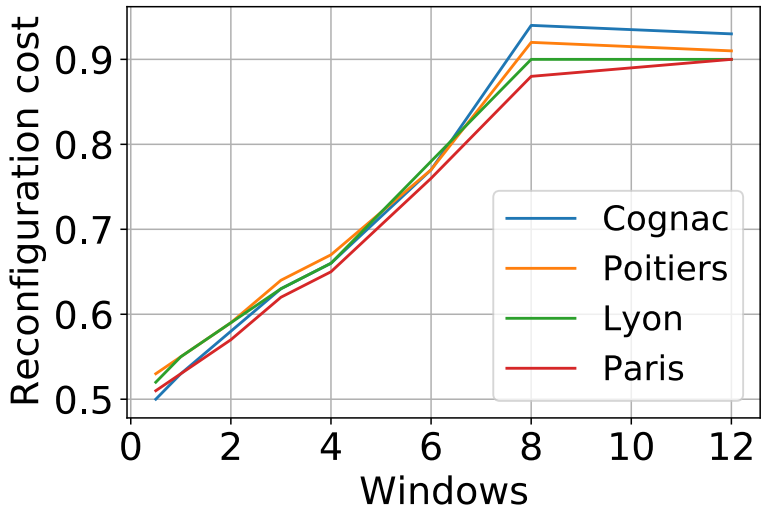


Figure 6.5: Reconfiguration proportion cost using different reconfiguration window sizes.

orange, Lyon in green and Paris in red.

For the four study areas, we have a linear increase of the costs for the reconfiguration windows ranging from 30 minutes to 8 hours. We notice that although these 4 regions have different sizes and mobility behaviors, they show an almost identical trend.

We can see that the four study areas, for a window size of 30 minutes to 6h and 12h, have similar cost proportions, within a 0.03 interval. For the 12h reconfiguration window, we see that the proportion of cost for Cognac is higher than for the other, but it remains within the same 0.03 interval. For the 8 hour reconfiguration window, the difference in proportion between the different study areas is greater, it is 0.06 for Cognac and Paris.

We can see that the higher the reconfiguration window size, the higher the proportion of the cost. However, we have a stagnation for the high window sizes.

6.6.3 Handover difference consideration

Our intuition in Chapter 5 is that the in_HOs and out_HOs do not consume the same number of resources and therefore do not have the same costs. Indeed, by looking at the procedures of HO in 4G in Appendix A, we see that the source eNodeB and the target eNodeB do not have the same interactions with the network, nor the same computation requirements. For example, the source eNodeB will communicate with the UE to get the signal quality measurements, then make the calculation to initiate or not the HO.

To differentiate the resources needed for in_HOs and out_HOs, we have defined two types of costs, computation costs and signalling costs.

Computation cost. This represents the set of computation operations that the eNodeB needs to perform, whether it is for the decision to command or not an HO, or for the admission control, i.e. checking if the eNodeB has the necessary resources to host a new UE through an HO. Looking at the HO procedure we see that, theoretically, an out_HO does more computation than an in_HO. For example, for the out_HO, the source eNodeB must check if the measurements reported by the UE

require the triggering of an HO and then calculate what will be the target eNodeB in case of HO, whereas the target eNodeB, in HO preparation, just checks if it has enough resources to manage a new UE. We can approximate the cost compared to the number of realised HO. We can formally describe the computation cost by:

$$Cost_{c_{out}} = \beta \cdot Cost_{c_{in}} \tag{6.3}$$

In Eq.6.3 we consider that the computation cost for the out_HO is equal to a coefficient β multiplied by the cost of in_HO. In the case of 4G, we estimate that the coefficient β is equal to 1.2. The computation cost of an out_HO is thus 1.2 times the cost of an in_HO.

Signalling cost. This corresponds to all the network communications that the eNodeB makes to request the possibility of making an HO, to warn of the decision of an HO, or to exchange with the UE for the establishment of a new link. The HO procedure allows us to see that, theoretically, an in_HO makes more network communications than an out_HO. We can formally describe the cost as follows:

$$Cost_{si_{in}} = \gamma \cdot Cost_{si_{out}} \tag{6.4}$$

In Eq.6.4 we considered that the cost of signaling for one in_HO is equal to the coefficient γ multiplied by the cost of out_HO. In the case of 4G, we estimate that γ is equal to 2.5.

The coefficient β and γ can differ according to technologies.

We now consider this second scenario, where in_HO and out_HO have different costs. We can reserve the right number of resources for these two types of HO. This solution requires the exact knowledge of the number of in_HO and the number of out_HO at each time, which can introduce complexity. However, we believe that this is an optimal scenario, since the reserved resources represent exactly what is used at a given time.

We will compare the results of this scenario with the other scenarios in the remainder of this chapter.

6.6.4 eNodeB profiles for resource saving

The third scenario is the use of the profiles defined in Chapter 5. To recall, P is the set of eNodeB which is labeled as a defined profile. The use of the profiles of the eNodeB is interesting because it makes it possible to consider the difference of the number of resources which are used by the in_HO and the out_HO. Nevertheless, we do not need to know the exact number of in_HO and out_HO, which reduces the complexity of the prediction. Indeed, each profile is defined by a proportion of in_HO and out_HO. In addition, we have a third proportion which we call uncertain_HO. For this proportion of uncertain_HO, we assign to it the highest costs for the computation and for the signaling, in other words, we assign to it the cost of computation of out_HO and the cost of signaling of in_HO. This cost assignment strategy is the same as the one applied for the first scenario.

Proportion computation

For the calculation of the proportion of in_HO, out_HO and uncertain_HO for each profile, we use the total number of HO. From this, for each profile, we make

a statistical study on the distribution of in_HO and out_HO. Each profile is thus characterized by the maximum and minimum proportion of in_HO and out_HO among the eNodeB in that profile. For in_HO, we are sure that the proportion of in_HO is at least as great as the minimal proportion of in_HO in the sum of the HO. For out_HO, we are sure that it is at least as great as 1− *maximum proportion of in_HO* of the profile. We have thus the value of uncertain_HO, as the subtraction between the maximal proportion of in_HO and its minimal proportion. Formally, we can write these values as follow:

$$in\_HO_p(t) = min(\{P_p^{in}(t)\}_w) \tag{6.5}$$

$$out\_HO_p(t) = 1 - max(\{P_p^{in}(t)\}_w) \tag{6.6}$$

$$uncertain\_HO_p(t) = max(\{P_p^{in}(t)\}_w) - min(\{P_p^{in}(t)\}_w) \tag{6.7}$$

In Eq.(6.5, 6.6, 6.7), $P_p^{in}(t)$ represents, for a profile p, the set of proportion P of in_HO in depending of the time t and the reconfiguration window size w.

Proportion representation

We use the profiles that exist in all study areas, which are producer, receiver and balanced. However, we have additional commuting-related profiles for Lyon and Paris. These profiles are treated in this chapter as balanced profiles since they have, as discussed in Chapter 5, similarities.

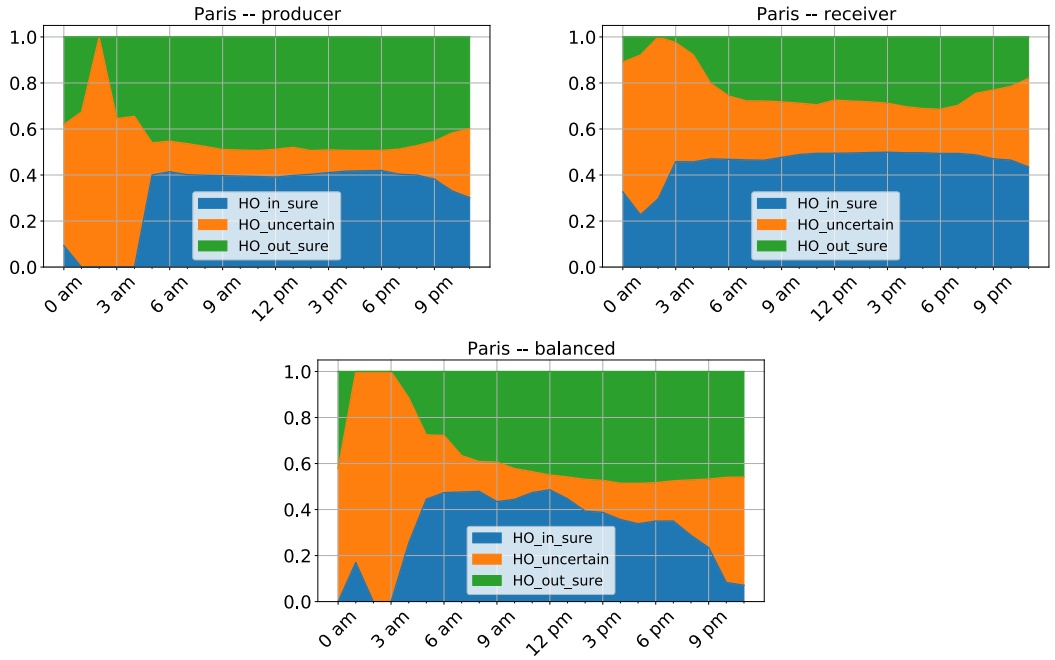


Figure 6.6: In_HO, out_HO and uncertain_HO proportion per profile example for Paris

We can graphically represent Eq.(6.5, 6.6, 6.7). Taking the example of Paris, we can see in Fig.6.6 these different proportions calculated on the basis of a 1 hour configuration window. We have three colors, green, orange, and blue, which represent respectively the proportion of out_HO, the proportion of uncertain_HO and in_HO.

In this example, for the producer profile, we can see that, from 5:00 am onwards, the proportion of out_HO is more important than the proportion in_HO. This shows the importance of the out_HO for the producer profile. For the proportion of uncertain_HO, we see that it is very important during the night. This can be explained by the impact that small variations can have when there are not many HO.

For the receiver profile, we see that, contrary to the producer profile, the in_HO are more important than the out_HO. However, for the receiver profile, the proportion of uncertain_HO during the night is also important.

For the balanced profile, we see that the importance of the proportions of in_HO and out_HO changes during the day. The proportion of in_HO is more important than the proportion of out_HO from 5:00 am to 1:00 pm and the order of importance is reversed afterward. This behavior is explained by the integration of the commuting profile in the balanced profile. In a normal balanced profile, the proportion of in_HO and out_HO is roughly equivalent throughout the day. The proportion of uncertain_HO is also important during the night.

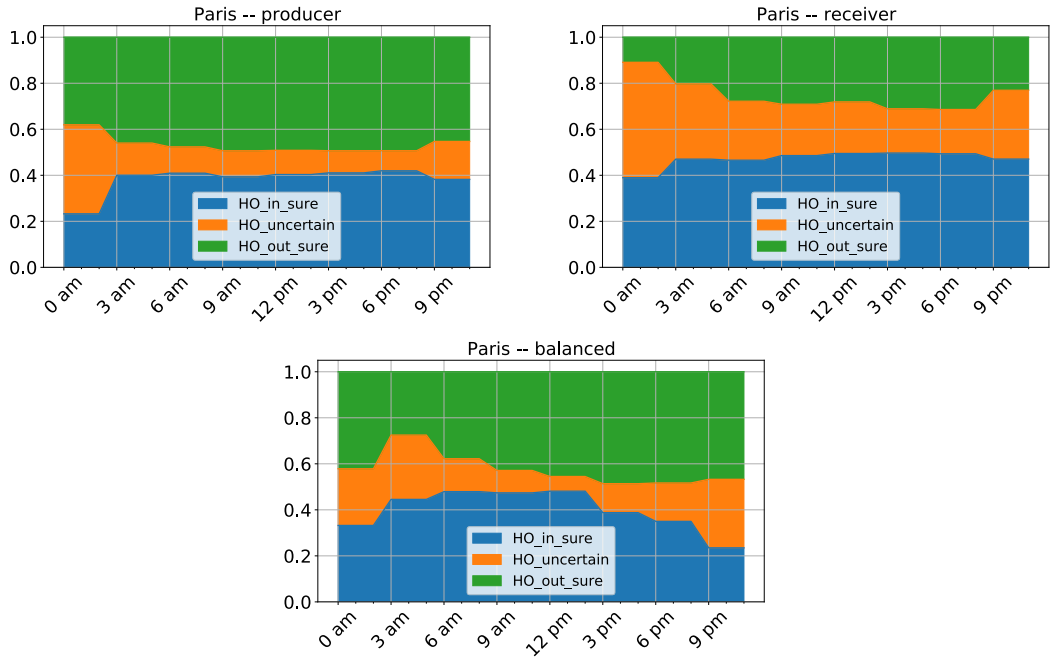


Figure 6.7: In_HO, out_HO and uncertain_HO proportion per profile example for Paris with reconfiguration window of 3h

It is also interesting to show the proportion of in_HO, out_HO and uncertain_HO with different reconfiguration windows. Fig.6.7 shows us the graphical representation of these different proportions for a reconfiguration window size of 3 hours. What we can see is that the importance of uncertain_HO, for all profiles, is less important during the night than with a window size of 1 hour. However, we still have the same tendency in the shape of distribution of the different proportions. For the producer profile, we also keep the importance of out_HO compared to in_HO and the opposite for the receiver profile. We can say that the bigger the size of the reconfiguration windows, the less the proportion of uncertain_HO is important during the night. This is due to the low value of HO during the night compared to the day.

6.7 Comparison of the different scenarios

6.7.1 Gain

We can represent the gains obtained by the different scenarios with respect to the current system. Formally, this gain can be represented by:

$$G^{w}_{k,n} = 1 - \frac{cost\_scenario^{w}_{n}}{current\_cost} \tag{6.8}$$

Eq.6.8 represents the gain G for $k \in [computation, signalling]$ and for the different scenario n. This gain is based on the ratio of the $cost\_scenario$, of the scenario n with a reconfiguration window size w, over the current system cost. We can represent these gains by a cumulative density function (CDF).

We present the gains as a function of different reconfiguration window sizes by taking the example of Paris in Fig.6.8. Indeed, all study areas have comparable gain behaviors.

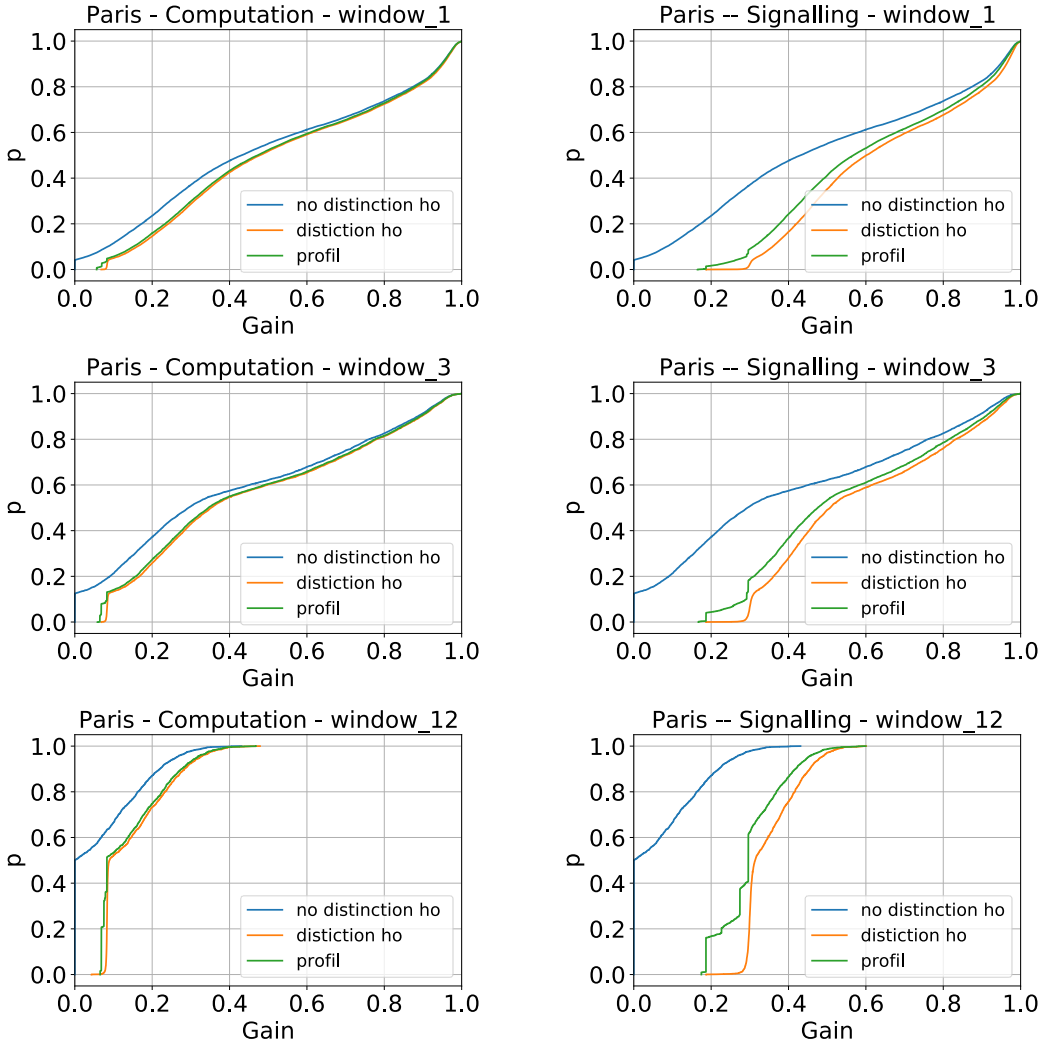


Figure 6.8: Computational and signalling gain compare to current system for Paris example with different configuration window size, 1h, 3h and 12h

In Fig.6.8, we can see that the second scenario, in orange, the one where we distinguish between in_HO and out_HO, is the one with the best gain, even if the

gain with the use of profiles is very close, especially for the computation cost. The scenario that has the least gain is the first one, when we do not distinguish between in_HO and out_HO. We can explain this by the fact that scenario 2 has a total knowledge of the network, so we consider that it is almost ideal. We also notice that the use of profiles, in green, produces gains that are very similar to the second scenario.

With a reconfiguration window size of 1 hour, for the computation, we see that we have a gain of at least 55% for 50% of the cases, and this for any scenario. And for the signalling gain, for the second and third scenario, for 50% of the cases, we have at least a gain of 60%. However, with the first scenario, for 50% of the cases, we have only a gain of 40%.

For a reconfiguration window of 3 hours, we see that there is a general drop in gains. For the computation, for 50% of the cases, we have at least 30% of gains for the first scenario and 35% of gains for the two other scenarios. Concerning signalling, for 50% of the cases, we have at least a 30% gain for the first scenario and a gain of at least 45% and 50% for the third and second scenario respectively.

For a reconfiguration window of 12 hours, the gain decreases further. The maximum gain for the computation is 50% for all scenarios, and for 50% of the cases for the first scenario there is no gain, while for the two other scenarios the gain is only 10%. For signalling, as for computation, the first scenario has no gain for 50% of the cases and the maximum gain is 44%. For the other two scenarios, 50% of the cases represent a gain of at least 30% and a maximum gain of 60%.

We also see that the larger the configuration windows the lower the gain. All other study areas have similar behavior.

6.7.2 Cost

Another way to compare the costs of the different scenarios with the current system is to sum up the costs of the HOs realised by all the eNodeBs in the different geographical areas. For the graphical representation, we use a 3h configuration window with value of β and γ respectively 1.2 and 2.5.

In Figures 6.9, 6.10, 6.11 and 6.12, the y-axis represents the sum of the costs, which can be approximated here to the reserved resource, for all of the HOs performed over the given geographical area as a function of the time of day on the x-axis

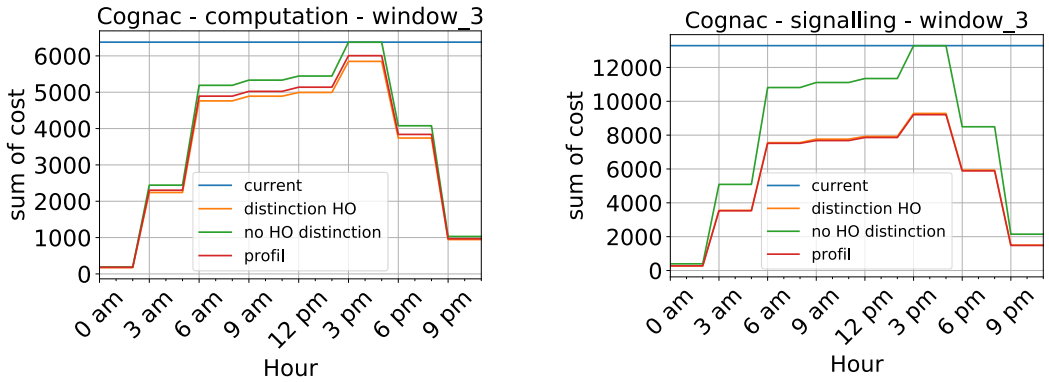


Figure 6.9: Computational and signalling total cost comparison of the current system to the three scenarios in Cognac

In Fig.6.9, we can see the total costs for the computation and the signalling for Cognac. We can see that, in both cases, the peak costs for the three scenarios are

reached between 3:00 pm and 6:00 pm. During this peak, the costs of the first scenario are similar to the costs of the current system, and the two others scenarios have a slightly lower cost. We can see here another representation of what we have seen previously in Fig.6.8 on the Paris example. Indeed, for the computation costs, the costs of the three scenarios are very close, the small difference is during the day between 6 am and 6 pm. For the cost of signalling, the difference in cost between the first scenario and the two other is significantly bigger. This difference is between 3 am and 8 pm. We can also see that the signalling cost is twice as high as the calculation cost. These costs are dependent on the coefficients β and γ.

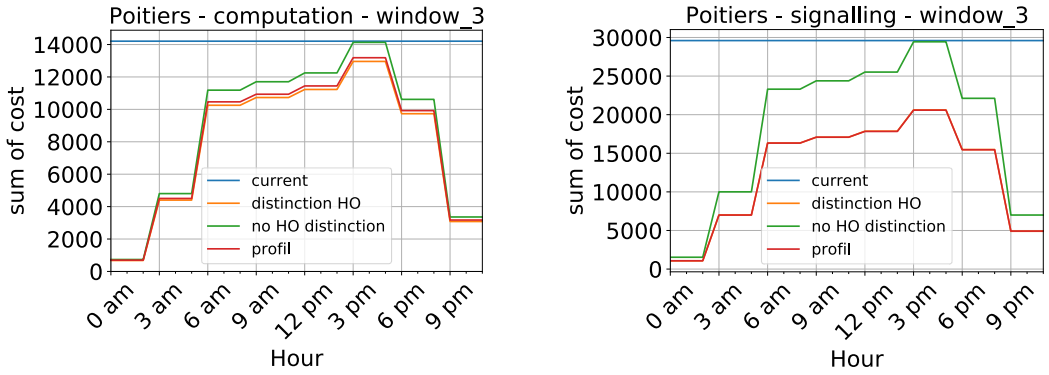


Figure 6.10: Computational and signalling total cost comparison of the current system to the three scenarios in Poitiers

For the case of Poitiers in Fig.6.10, we can make the same remarks as for Cognac. However here we can see that the sum of the costs, whether for computation or signalling, is greater than those of Cognac. This can be explained by the larger number of eNodeBs. On another hand, like Cognac, the costs of the second and third scenarios are very close to each other. This similarity is greater for the cost of signalling than for the cost of calculation.

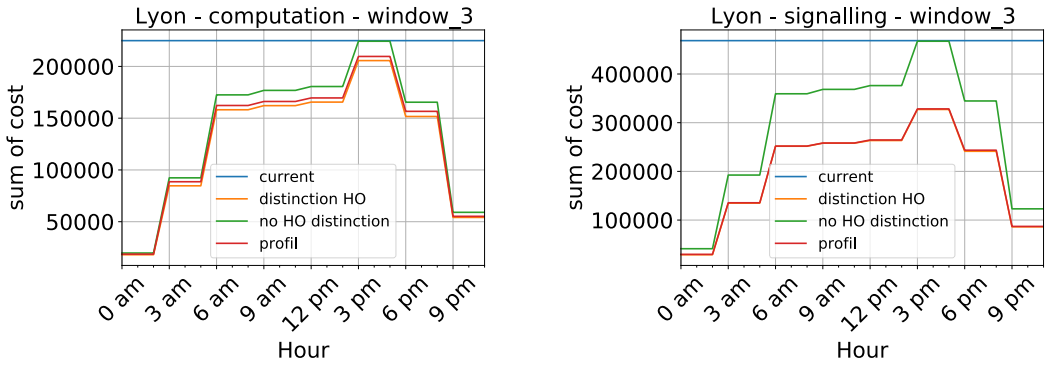


Figure 6.11: Computational and signalling total cost comparison of the current system to the three scenarios in Lyon

For Lyon in Fig.6.11, we can make the same remarks as for Cognac and for Poitiers. Here, the cost is higher than in the two other cities.

For the case of Paris, in Fig.6.12, we can also make the same remarks as for the previous cities and with costs that are more important than the other cities. However, during the peak of the costs, between 3 pm and 6 pm, we can see that the first scenario has a cost that is quite different from the cost of the current system contrary to the three other cities. Indeed, in all three cities, there is also a difference between the

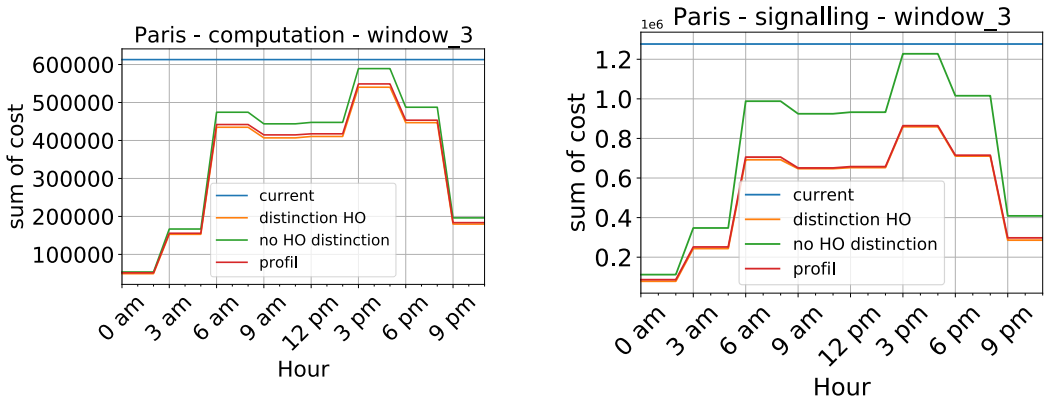


Figure 6.12: Computational and signalling total cost comparison of the current system to the three scenarios in Paris

peak cost for the first scenario and the current system. However, this difference is small that they can be confused on the graphs. The larger difference that exists for the Paris case can be explained by a larger number of eNodeBs.

6.8 Conclusion

In this chapter, we have seen that considering the distinction between mobile and static users through different network slice can save resources. Concerning the possible network slice for mobile users, the implementation of a dynamic reconfiguration of resources throughout the day with reconfiguration windows can allow resource saving compared to what is currently done. This resource saving can be more important if we have a precise knowledge of the real number of in_HO and out_HO performed by the eNodeB. However, this solution can be complex. The use of eNodeB profiles can therefore be interesting, each profile being defined by the proportion of in_HO and out_HO that it processes. This solution has the advantage, in our case, of having comparable resource savings with perfect network knowledge.

Chapter 7

Conclusion

The use of mobile data has become interesting both for describing human mobility and for improving the design and management of mobile networks. In this thesis, we show examples of how mobile data can be used in both cases.

7.1 Summary and discussion

Each of our contributions uses and analyses mobile phone data to describe some phenomenon, to evaluate the accuracy of the data and to evaluate the possible benefits that a new architectural paradigm could bring to the network.

First, we assessed the accuracy of aggregated mobile phone data by cross validation with sensor data collected by motion detection sensors. We defined several metrics, correlation between the two data sources, peak synchronization and start and end of the days synchronization, to compare the two data sources, represented by their time series. The analysis of the results is conducted on a spatiotemporal scale. Our results indicate that there is a strong correlation between the two datasets. When considering all peaks in the data, synchronization is pretty good between the two sources of data, even though some days present lower synchronization. The synchronization of the start of the day, when human activity begins, is also pretty good. These results validate the realism of mobile phone data. However, there are also some differences between the two data sources. We noticed that there is generally a delay of 30 min between the two time series, to reach the best results in terms of correlation. Moreover, there is also a weaker synchronization of the two most important peaks between the two time series. Finally, the end of day synchronization is very low, especially on Sunday and in summer.

This is a first step towards assessing the quality of mobile phone data. However, we compared aggregated mobile data and motion detection sensor data covering a commercial industrial area in a small city suburb. This area being quite characteristic, the observations we made may be also specific to this area.

Second, we studied the effects of different sanitary measures with the restriction of mobility, due to the emergence of Covid-19, on land use. We used agglomerative hierarchical clustering and the concept of signature for the detection of land use. This allowed us to study the evolution of the latter, in normal time, without restriction and with the different levels of restriction represented by the three different lockdowns. During the first lockdown, the majority of the city of Paris is labeled as *residential*.

Indeed, this lockdown has the strictest mobility restriction measures put into place, thus, the fact that most of the area became *residential* was expected. During the other two lockdowns, land use is similar to that of the study period before all the restrictions. However, during the third lockdown, we observe a "residentialisation" of the different areas, albeit at a negligible level compared to the first lockdown. The presence of school holidays during this period may explain this. We also used the mean-peak-to-average ratio in the signatures to try to assess the possible importance of the different zones in the propagation of the Covid-19 virus. The spatial distribution of the ratios shows that the areas labeled as *activity* have the highest values and represent areas where populations mix.

While the methodology proposed in Chapter 4 is used to obtain quantitative results regarding the impact of the Covid-19 pandemic on urban mobility, the approach we describe can be applied to any contextual scenario with an impact on human mobility, which brings the generality of our contributions. A possible application of the proposed technique is the provision of energy-efficient and cost-effective network infrastructures adapted to the modified population mobility patterns during these special situations. Indeed, telecommunications operators planned and deployed their infrastructure with certain attendance and mobility patterns in mind, therefore, understanding how these patterns change in special situations is a first step in the proposal of flexible and adaptive approaches. As a final potential impact, our results are a valid indication of activity loss, and consequently, of the imposed socio-economic impact activity-labeled areas suffered during the lockdown. The hypothesis of using peak-to-average ratio to assess the importance of areas in the epidemic spread needs to be further validated. Indeed, we need to compare high value areas with ground truths to prove our intuition behind this metric.

Then, we showed the existence of different profiles of eNodeBs in terms of handovers. After studying geographical areas with different urbanization levels, we obtain three types of profiles. The first one, *Producer*, has a higher proportion of outgoing handover compared to incoming handover. The second profile, *Receiver*, has an inverse proportion to the producer. The last one, *Balanced*, oscillates between producer and receiver. We also found different categories of eNodeBs according to the volumes of handover they handle. To define these profiles and categories, we also used agglomerative hierarchical clustering. This technique, despite its simplicity, requires, as we have seen in Chapter 5, a lot of manual verification to check the coherence of the results.

Finally, we have shown the possible advantage of implementing the slicing concept for mobility management. The implementation of a slice dedicated to static users can decrease the number of operations that the network has to do. Indeed, by knowing that a user is static, operations related to mobility, such as measuring the quality of the radio signal, can be performed less frequently. The reduction of operations will allow the reduction of the number of resources that are reserved. A specific slice can also be considered for mobile users. A dynamic reconfiguration of the resources reserved by this slice can reduce the resources to be reserved by up to 50% compared to what is currently done.

In Chapter 6, we showed a kind of proof of concept on the advantages brought by the implementation of a slice for static users and a slice for mobile users. For the reconfiguration of the resources for the mobile slice, we used the profiles defined in Chapter 5. These profiles, as indicated earlier, were obtained with an AHC. Our

objective with this clustering algorithm is to follow the overall tendency in the evolution of the HO. However, to get finer results, indicating particular behaviors, e.g. *producer* at certain times and *receiver* at others, a lot of manual analysis would be needed if using AHC. To facilitate the detection of these particular profiles, more powerful algorithms are needed. The handover resource reconfiguration can also be achieved by making predictions, in contrast to our approach which is based on statistical characteristics of profiles. We have also assumed that, in the current system, the reconfiguration of resources is done once a day. However, this reconfiguration is at the discretion of the operator, thus, the time between different resource reconfigurations can vary. If the period between two configurations set up by the operator is longer than 24h, the implementation of slices with dynamic reconfiguration can reduce costs by reserving even fewer resources.

7.2 Perspectives

Following the discussions in this manuscript, several perspectives can be considered.

Comparison of user presence information from mobile phone and sensor data

The good correlation value between the mobile data and the motion detection sensor data was obtained on an area that is representative of a specific behavior. In order to generalize our observations, we should conduct similar studies on different areas. These different areas must have different land uses. Indeed, the behavior of an area that is more activity-oriented may differ from areas that are more residential. Similar observations could generalize the observations we have described in this manuscript.

Covid-19 impact on Parisian mobility behavior

We used the peaks in the signatures to define a ratio that could estimate the importance of an area in the epidemic spread. We will be able to check the relevance of the proposed metric by finding data indicating where people are infected.

Another perspective is to use the evolution of the ratios of the different areas as another way to define land use. Indeed, if we look at the evolution of the peak-to-average ratio according to different periods, for different areas, we see different patterns. Therefore, it would be interesting to see the distribution of zones with similar evolutions of the metric, and then to check if there is a correlation between these distributions and the type of land use in those zones.

Handover characterization and slicing opportunity for mobility management

We have defined different handover profiles that an eNodeB can have, and we also defined several categories of eNodeBs in terms of handover volume. However, for our study on the benefits of mobility-oriented slicing, we have only used the type of profiles that an eNodeB can have. Nevertheless, in a future perspective, considering the category of the eNodeB, may allow to have pre-determined quantities of resources needed. This may reduce the complexity of resource allocation as knowing the exact number of handovers will no longer be a requirement. However, this may be less

effective in reducing costs in relation to reserved resources compared to using profiles alone. Moreover, to define in_HO and out_HO cost, we conduct a theoretical study of the handover procedure. To validate these costs and to have a more accurate model, we should use measurements conducted during real handovers.

Publications

International Journal with peer review

- H. C. Fanticelli, S. Rabenjamina, A. C. Viana, R. Stanica, L. S. D. Oliveira, A. Ziviani, "Data-driven Mobility Analysis and Modeling: Typical and Confined Life of a Metropolitan Population, "in *ACM Transactions on Spatial Algorithms and Systems (TSAS)*", vol. 8, no. 23, pp. 1-33, Sept. 2022.

International conference with peer review

- S. Rabenjamina, R. Stanica, O. Iova, and H. Rivano, "Comparison of User Presence Information from Mobile Phone and Sensor Data,"*25th ACM Conference on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWIM 2022)*", Montreal, Canada, 2022, pp. 223-231.

/

Appendix A

Handover mechanism

A.1 Introduction

The mobile network defines two types of mobility: macro mobility and micro mobility. Macro mobility allows the access to a service over the entire territory covered by the network operator. For instance, the macro mobility allows a French user to travel across France and benefit from the service at his new location. Also, roaming allows to benefit from services despite not using the user's own operator's network, but a third-party operator. This happens most of the time when travelling abroad. For micro mobility, its management allows the continuity of the service while the user is moving, thus changing the cell to which he was previously attached. Therefore, HOs represent a component of user micro mobility. The management of HOs aims to achieve a satisfying Quality of service (QoS) and QoE.

In this chapter, we will first describe the handover execution, then we discuss about how it is implemented in the different current mobile architectures.

A.2 Handover execution

There are three main modes of execution of HO: *(i)* hard HO *(ii)* seamless HO and *(iii)* soft HO. However, not all mobile network technologies implement all these different execution modes.

A.2.1 Hard handover

Hard HO is one execution method for a HO. We can see in Fig.A.1 three different steps: (1) before the execution of HO, (2) during the execution of HO and (3) after the execution of HO.

Before the execution of the HO, the mobile phone is linked to the transceiver (TRX1) by a radio link, and to the core network (CN) through TRX1. A transceiver is a part of the radio access network (RAN). The latter is usually composed of antennas, or base stations, and base station controllers.

During the execution of the HO, the user's mobile phone is moving, and a new cell is more adequate. For this phase, in hard HO, first the mobile is still linked to the previous TRX1. However, there is suspension of normal operations except for the radio resource management layer. Then, the disconnection of the signaling link occurs. This deactivates the previous allocated channels established through TRX1 and releases them. Afterward, it activates new channels with TRX2 and their

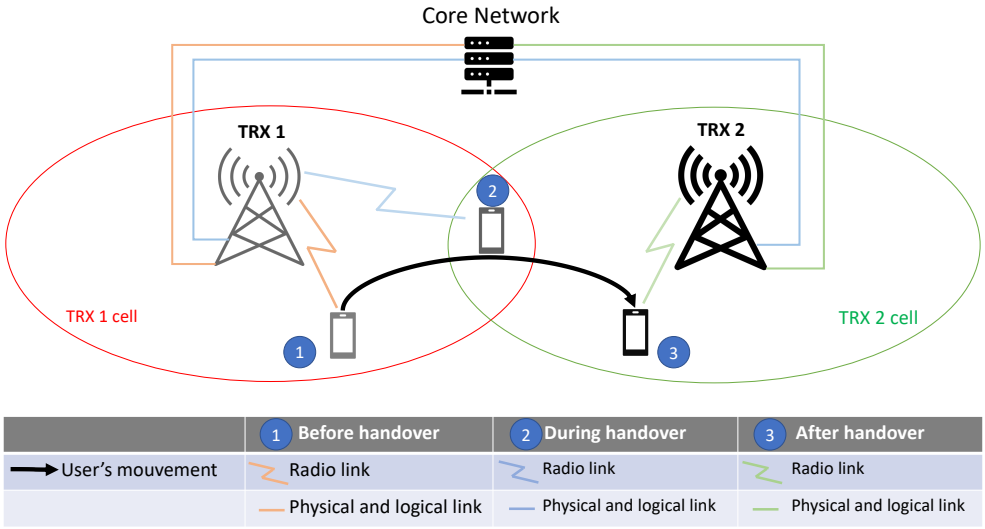


Figure A.1: Hard HO execution

connection if necessary. Finally, it triggers the establishment of a data link connection on the channels. In other words, with a hard HO execution, while the mobile is still linked to TRX1, information about the latter will first be sent from the core network to TRX2. Then the mobile phone disconnects its link to TRX1 and connects to TRX2, hence there is only one radio channel at a time. Therefore, there is a cut-off in the connection between UE and RAN. If this cut-off takes longer than a few milliseconds it may interrupt the user's communication, thus affecting the QoS and QoE.

After the execution of the HO, all information for the mobile is received through TRX2.

A.2.2 Seamless handover

Similar to hard HO, we show the three steps for seamless HO in Fig.A.2

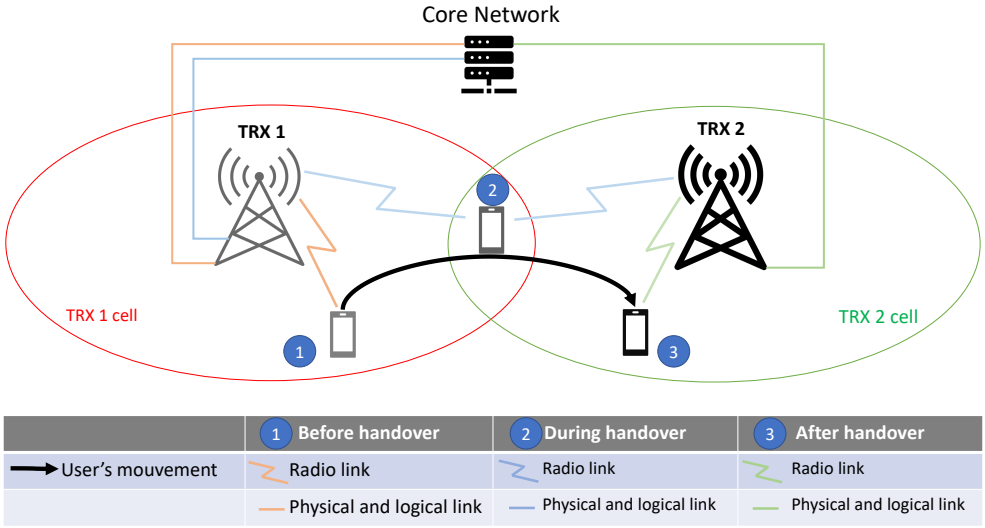


Figure A.2: Seamless HO execution

For seamless HO, the phases during and after the execution of the HO are similar

to those in hard HO. Hence, we just discuss about the process during the execution of the HO.

During the execution of the HO, the mobile is still linked to TRX1, and the CN establishes a link to TRX2. The process of disconnection to TRX1 is occurring at the same time to the connection to TRX2. The release of channels via TRX1 therefore occurs at the same time as the establishment of channels via TRX2. In other words, in seamless HO, the network releases resources during the request for new resources in a new cell. This minimises the probability of outage and therefore maintains the QoS; however it consumes more resources.

A.2.3 Soft handover

We also show the three steps in the execution of soft HO in Fig. A.3. Once again, we focus on the process during the execution of HO.

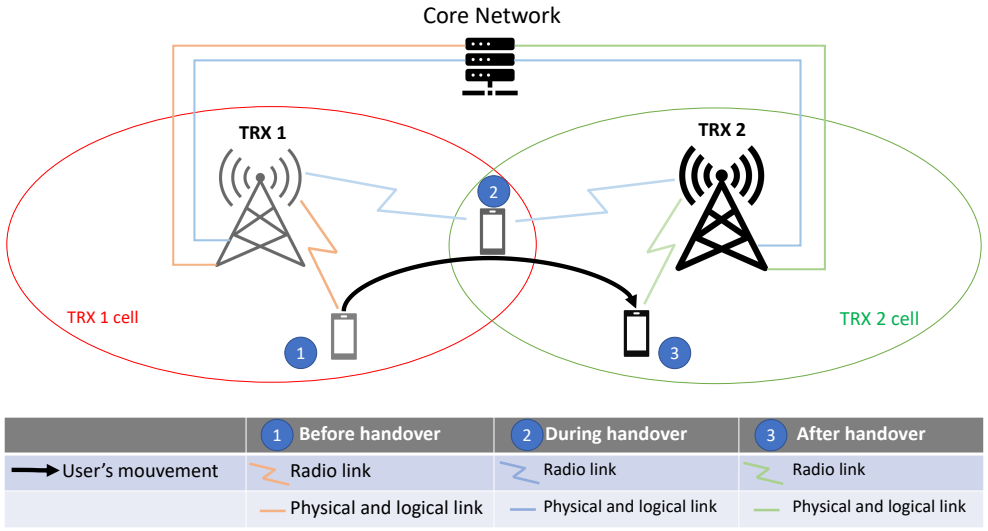


Figure A.3: Soft HO execution

During the execution of the HO, while linked to TRX1 and the CN, the mobile phone establishes a link to TRX2 and the CN via TRX2. The disconnection from TRX1 and the CN via TRX1 occurs only when the connection to TRX2 and the CN via TRX2 is established. The release of channels via TRX1 therefore occurs after the establishment of channels via TRX2. In other words, resources are released after obtaining new resources. The two links and two streams (from CN to UE) are active simultaneously for a short period of time. Thus, there is no disconnection, as the traffic is transmitted in both cells simultaneously. This will provide a good QoS, yet the load on the network and the radio interface increases.

A.2.4 Handover execution scenarios

There are two main HO execution types scenarios, intra cell HO and inter cell HO. Intra cell HO is mainly due to degradation of the radio link between the UE and the antenna. This degradation is caused by interference, but not by the distance of the antenna from the user. Inter cell HO is often due to the mobility of the user. The physical measurements reported by the UE show that there is a better cell. However, inter cell HO can also be used for traffic balancing.

A.3 GSM handover

GSM has an architecture that allows voice communication only. Its evolution, GPRS or EDGE, i.e. 2G+, instead allow data exchange in addition to voice. Hence, in GSM, HO occurs only when the user is calling.

A.3.1 GSM architecture

The GSM protocols are based on the Integrated Services Digital Network (ISDN), which uses the Signaling System Number 7 (SS7) for the signalisation.

To better understand HOs scenario, we present GSM architecture in Fig.A.4

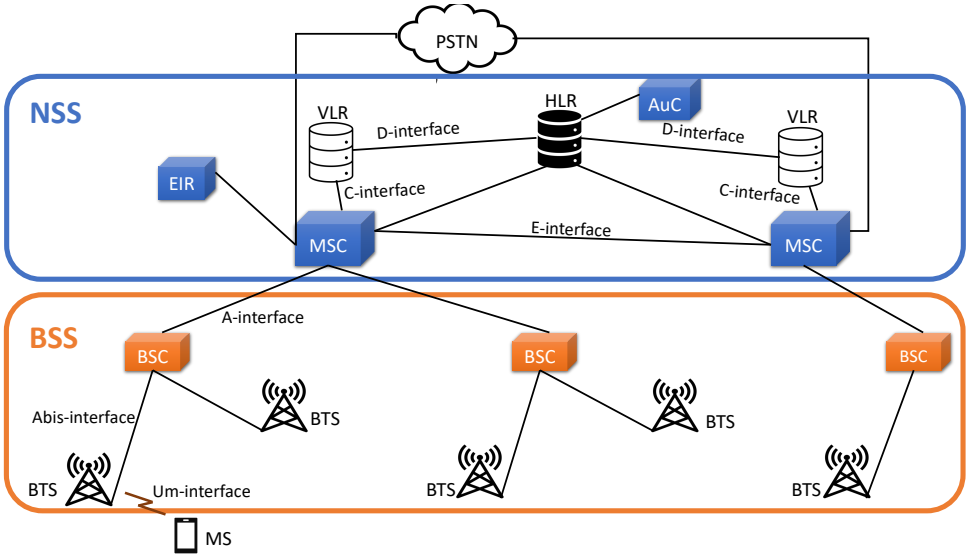


Figure A.4: GSM architecture

GSM architecture can be divided into two parts, the Network Subsystem (NSS) and the Base Station Subsystem (BSS). These two parts can be approximate to the CN and the RAN, respectively. These two parts have different components. NSS is composed from:

- Mobile Switch Center or **MSC** is responsible for the connection with the public switched telephone network (PSTN) and the other public land mobile network (PLMN). It also controls calls, signalisation and billing. MSC manages the update of the localisation and the authentication of the access of the subscriber in GSM network. Moreover, MSC manages interMSC HO.
- Home Location Register or **HLR** is the central data base of the PLMN. It contains subscriber data. HLR handles requests from Gateway MSCs and VLRs. From the operator side, it deals with adding, deleting and activating subscribers and services. From the subscriber side, it processes the activation of additional services.
- Visitor Location Register or **VLR** contains subscriber information on a given geographical area. It also controls authentication and encryption procedures, call set-up time, and the location updates.
- Equipment Identity Register or **EIR** contains mobile terminal identity and their permissions.

- Authentication Center or **AuC** is a data base which contains an individual key per subscriber (Ki), the same as in the subscriber SIM card. It implements encryption algorithms, e.g. A3 and A8.

The BSS is composed by:

- Base Station Controller or **BSC** is responsible for the establishment, release and maintenance of all connections from cells that are connected to it. It is responsible of the management of radio resources, and HO.
- Base Transceiver Station or **BTS** is the interface element that connects the mobile devices with the network. It also manages physical layer with encoding, modulation, power control.

In GSM, user's mobile phone is called mobile station (MS). All components of the network are interconnected with different interfaces, for instance BSC and BTS are connected via Abis interface and MSC and BSC via A interface. MSC and BSC are components which are involved in the HOs.

A.3.2 Handover scenarios

In GSM there are different scenarios of HO: intra BSC, intra MSC and inter MSC.

Intra BSC HO occur when an active MS changes cells. In this case, the HO destination cell has its BTS managed by the same BSC as the BTS of the HO source cell. In this scenario, the BSC can control the HO.

For intra MSC HO scenario, the HO destination cell BTS is not managed by the same BSC that manages the BTS of the HO source cell. However, the BTS are managed by the same MSC. In this scenario BSCs cannot control the HO, as no direct signaling exists between the two BSCs. In this case the source BSC requests the MSC to initiate the HO to the HO destination cell.

In inter MSC HO scenario, the HO destination cell and the HO source cell are managed by different BSC and different MSC. In this scenario the source MSC asks the MSC that controls the destination BSC to establish a link.

Despite these different scenarios, the procedure for initiating the HO is the same.

A.3.3 Handover procedure

We can divide the HO procedure into three phases: *(i)* radio link evaluation, *(ii)* HO triggering and *(iii)* execution procedure.

Radio link evaluation: During a call, the MS will make measurements to continuously assess the quality of the current radio link with the BTS, but also the signal quality of the other adjacent cells. The quality of this link will be evaluated with the received signal strength indication (RSSI). The RSSI represents the total signal power received. These measurements are sent periodically to the BSC, each 480 ms. In case the quality of the current link is worse than a given threshold, the BSC informs the MSC. The MSC identifies the target cell and/or channel while considering the traffic load.

HO triggering: The decision whether to perform a HO is based on the comparison of the quality of the current link with other possible links. In the case of GSM, a HO is triggered if the current signal strength plus a certain threshold plus a hysteresis margin is lower than the target signal strength: $RSSI_{source} + threshold + hysteris < RSSI_{destination}$.

Execution procedure: In GSM, when HO is initiated, a hard HO is performed. Therefore, there is a cut-off MS call.

A.4 UMTS handover

The UMTS core network is divided into two domains: circuit switched domain and packet switched domain (Fig.A.5). The circuit switched domain is the same as the GSM core network. The packet switched domain allow UMTS, unlike GSM, to be able to exchange data, for example with the Internet.

A.4.1 UMTS architecture

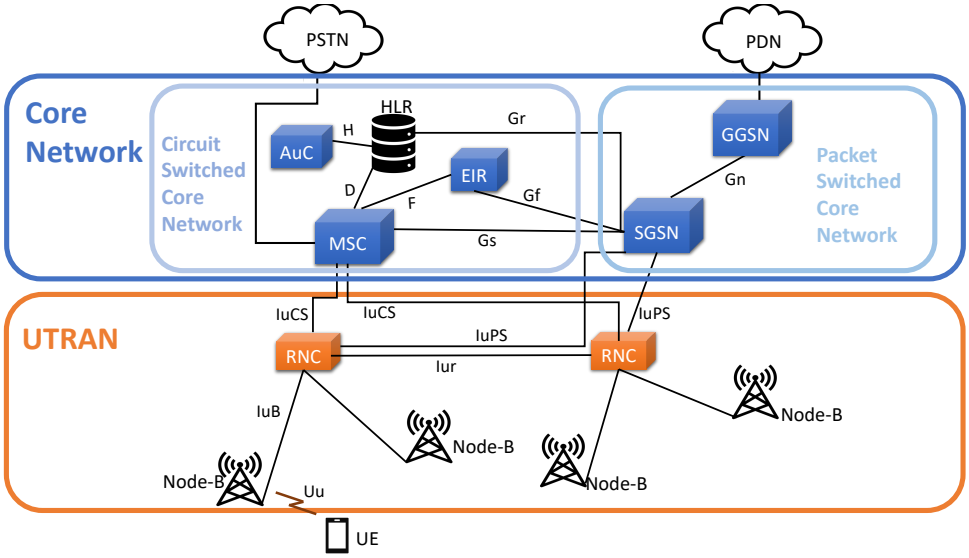


Figure A.5: UMTS architecture

Like GSM, UMTS architecture is divided into two parts, the CN and the RAN, called UMTS Terrestrial RAN (UTRAN). In the following, we will define only the components of the packet switched domain, since, as mentioned earlier, the circuit switched domain is the same as in GSM. However, the packet switch domain will share some components with the circuit switched domain. These shared components are: AuC, HLR and EIR, with the same role as in GSM. The UMTS packet switched core network is composed of:

- Serving GPRS Support Node or **SGSN**, which manages the packet connection of all UEs in its routing area, but also the billing and mobility management (other than HO).
- Gateway GPRS Support Node or **GGSN** is the last router in the operator network. It processes protocols in UMTS but also from external networks. In other words, it translates the protocols.

The UTRAN is composed of:

- Radio Network Controller, **RNC**, which controls and handles radio resources through the Radio Ressources Control (RRC) protocol. It also controls HO. RNC handles Node-B load and congestion control, but also the access control. It is also in charge of the scheduling of the data transmission. RNC performs everything that the BSC does, however, an RNC can be connected directly with another RNC on the Iur interface. Therefore, a direct dialogue between different RNCs is possible without going through the CN.
- **Node-B** is an evolution of the GSM BTS, meaning that it controls physical layer parameters, like coding and modulation. It manages the error control, the power control on the UE side, and receives the feed-back for its own power control. In addition to what the BTS did in GSM, the Node-B also accomplishs some tasks that were done by the BSC.

In UMTS, the HO is controlled by the RNC, when the UE is in the RRC connected state.

A.4.2 Handover procedure

Like in GSM, we can divide the HO into three phases: *(i)* radio link evaluation, *(ii)* HO triggering and *(iii)* execution procedure.

Radio link evaluation: During a call or exchange of data, the signal quality in the uplink and downlink is continually assessed. The uplink is assessed by the UTRAN and the downlink by the UE. The feedback of the signal measurements from the UE can be done in two ways: *(i)* as in GSM, periodically from 0.25 s to 64 s, but also *(ii)* triggered, when certain conditions defined by the RAN are met. Unlike GSM, UMTS uses other criteria in addition to RSSI to assess signal quality and HO decisions. These other criteria also consider the environment surrounding the UE. The UE will also measure the quality of signals received from other neighbouring nodes, which can be considered as noise. These other criteria are: received signal code power (RSCP) and energy per chip over noise (EcNo). The RSCP represents the received power level of the pilot frequency of a base station. In UMTS, several Node-B can transmit on the same frequency, with specific codes. The UE computes the power level of the Node-B, after demultiplexing of the codes. Hence, for example, it can indicate that the mobile is at UMTS cell-edge and no neighbouring UMTS cell is available. The EcNo, which is equal to $RSCP/RSSI$, represents a signal over noise ratio, where RSSI is the total noise (all transmissions take place on the same frequency). The higher this value is, the better the signal quality.

HO triggering: By considering all these criteria, the RNC can choose the best cell to which the UE can execute a HO. The criteria for the new cell must bring an improvement over the current cell.

Execution procedure: UMTS can perform hard and soft HO. Before changing the cell, resources must be reserved on the Iub interface (nodeB $\leftrightarrow$ RNC) and, if necessary, also on the Iur interface (RNC $\leftrightarrow$ RNC). In the case of a hard HO, this prior reservation of resources will result in a low cut-off time of about 100 ms on average.

A.5 LTE handover

Unlike UMTS, the LTE architecture is designed to be all-IP. Therefore, in LTE, the HO occurs when the UE is exchanging data while moving.

A.5.1 EPS architecture

The LTE design allows to simplify the network architecture. The whole architecture is called Evolved Packet System (EPS). The CN is called Evolved Packet Core (EPC) and the RAN is called evolved UTRAN (E-UTRAN), as shown in Fig.A.6.

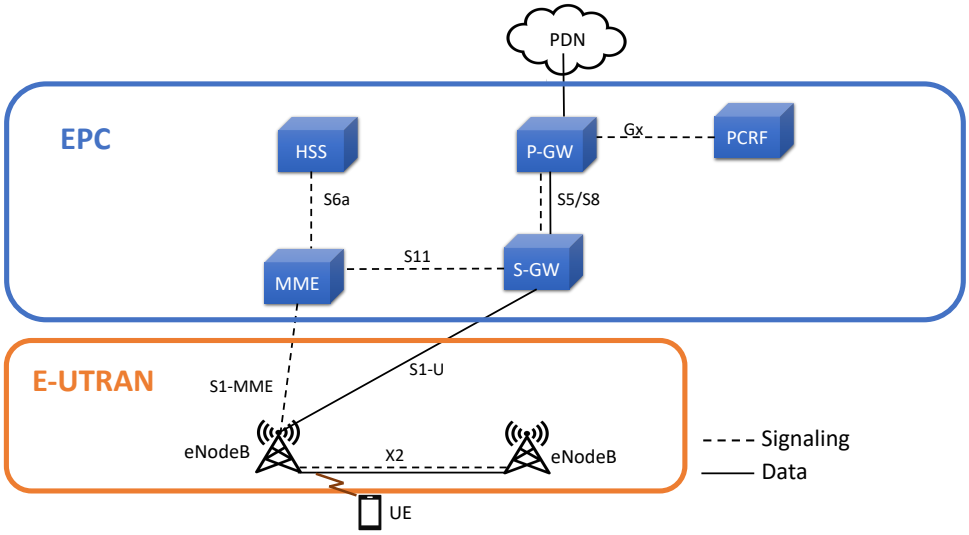


Figure A.6: EPS architecture

The EPC comprises several entities:

- mobility management entity or **MME** manages the network and handles all the signaling between the EPC and the UE. It has several roles, including managing authentication, updating the tracking area for mobility management, selecting the S-GW and P-GW that will serve the UE. Moreover, it also manages the bearer (establishment, maintenance and release of data bearers).
- serving gateway or **S-GW** acts as a router between the P-GW and the eNodeB over which all user IP packets pass. The S-GW buffers packet download for the UE in idle state, until the MME successfully pages the UE and re-establishes the bearer. It also serves as an anchor for the data bearer during a HO.
- packet data network gateway or **P-GW** is an IP anchor point for the data bearer. The P-GW allocates IP addresses to UE. It enforces QoS policy, like priority and throughput requirements. Furthermore, P-GW is the network gateway to external PDNs, e.g. application server or the Internet.
- home subscriber server or **HSS** is a database containing the subscription information of UEs, such as their QoS profile and access restrictions. It also contains the identity of the MME serving the UE. In addition, it integrates the authentication centre containing the security keys of the subscribers necessary for their authentication. In analogy with the previous generations, the HSS can be considered as a fusion of HLR and AuC.

- policy control and charging function or **PCRF** sets up policy contol decision making. For instance, it sets rules on control flow-based charging functionalities which are implemented by the P-GW.

The E-UTRAN comprises a single equipment, the evolved NodeB (eNodeB). The eNodeB is an autonomous entity that integrates functionalities previously provided by a centralized BS controller. It manages radio resources (allocation, bearer, mobility scheduling). It also performs IP packet compression (PDCP) and security functions. In addition to these functionalities, it also provide services of the previous generation of BS. Therefore, it deals with the air interface, modulation/demodulation, encoding/decoding, multiplexing/demultiplexing. The connection of the eNodeB with the EPC in the S1 interface is provided by a high-speed backhaul. Two eNodeBs can be connected to each other via the X2 interface, which can be used for HO management and for interference management.

A.5.2 Handover procedure

In the following we will describe each step in the LTE HO procedure. The triggering of the HO is preceded by a preliminary phase with the transmission of measurements from the UE. Then the HO is performed in three phases: *(i)* HO preparation, *(ii)* HO execution and *(iii)* HO completion.

Preliminary phase: In this phase, during data exchange, the UE will continuously assess the quality of the uplink and downlink. Like in UMTS, signal quality is also assessed by several metrics: RSSI, Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ). The RSRP allows the comparison of cells using the same carrier frequency. It assesses the environment in which the UE operates. The RSRQ is equal to $RSRP/RSSI$, therefore the better this value is, the better the signal is compared to interference generate by other cells. The feedback of these measurements to the eNodeB can be transmitted periodically or triggered. The activation of the transmission of the measurements by the UE to the eNodeB is done when the power received from the eNodeB is lower than a threshold. After the transmission of these measurements, the eNodeB analyses the power levels indicated by the UE for the current eNodeB and the neighbouring eNodeBs.

HO preparation: Once the HO decision is made, the eNodeB prepares to execute the HO by reserving the resource at the target or destination eNodeB.

HO execution: In this phase, the eNodeB sends the order to the UE to perform the HO. In this phase, the packets will be rerouted, the tunnels and connections will be modified, the radio connection with the new eNodeB will be established.

HO completion: During this phase, the UE is attached to the destination eNodeB and all data intended for it is sent directly to the destination eNodeB.

There are two different types of execution of a HO in LTE: X2 HO and S1 HO. X2 HO use the X2 interface to perform HO and S1 HO needs more interaction with the MME.

A.5.3 X2 handover

Two eNodeBs, source and target, linked by the X2 interface, can perform an X2 HO when a UE is changing cell, moving from the cell covered by the source eNodeB to the target eNodeB. Fig.A.7 shows the X2 HO procedure.

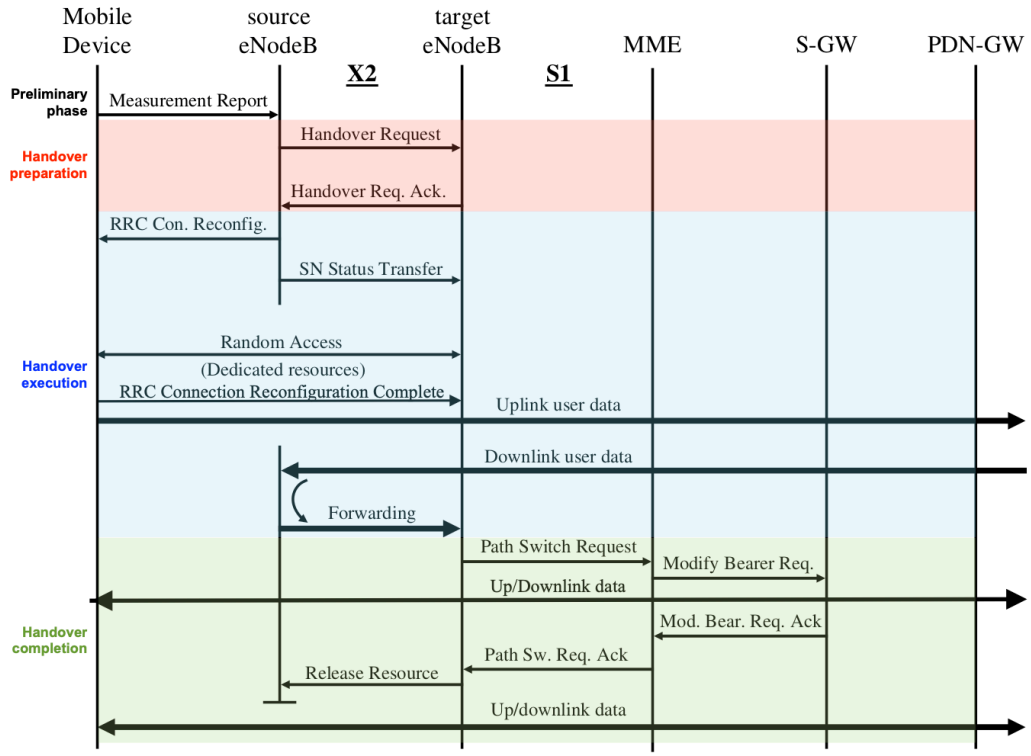


Figure A.7: X2 HO basic procedure

In the preliminary phase, the UE reports measurements, to the eNodeB, on signal quality. Then the eNodeB decides or not to perform HO. If it decides to perform a HO, it turns to the HO preparation phase.

HO preparation: When it is decided to perform a HO, the source eNodeB chooses the target eNobeB according to the measurements reported by the UE. The source eNodeB will then send a HO request to the target eNodeB. This message is *X2AP HO Request*.

When the HO request is received, the target eNodeB performs an admission control, which consists of verifying whether it is capable of receiving this new UE, for example by checking whether its cell is not saturated. If it can admit it, the target eNodeB will choose a preamble which will be used by the UE in the random access phase (RACH). This preamble is sent in the acknowledgement message with a message *X2AP HO Request Acknowledge*. The latter will indicate to the source eNodeB that the target eNodeB can receive the UE. In case the target eNodeB is not able to receive a new UE, the admission control leads to a refusal, and a negative acknowledgement with a message *X2AP HO Preparation Failure* is sent back to the source eNodeB. After receiving this message, the source eNodeB can try to detect if there is another target eNodeB or keep the UE on the old cell as long as possible. In some cases, the source eNodeB will not succeed so the connection will be interrupted

HO execution: When the source eNodeB receives the *X2AP HO Request Ac-*

knowledgement message, it can instruct the UE to change cell, by sending a *RRC Connection Reconfiguration* message. This message contains information that will allow the UE to connect to the target eNodeB, such as the new Radio Network Temporary ID (RNTI) of the UE, the identifier of the target eNodeB and the preamble previously provided by the target eNodeB.

At the same time, the source eNodeB sends a *SN Status Transfer* message to the target eNodeB. This message indicates the sequence number of the last valid uplink data block. As the eNodeB no longer accepts uplink traffic from the UE, this message will allow the target eNodeB to request uplink retransmission if it detects that data blocks are missing. This provides continuity in data transmission.

The UE will establish a link with the target eNodeB using the random access procedure. To do this, the UE sends the random access preamble it has been allocated to the target eNodeB on the Physical Random Access Channel. The target eNodeB sends a random access response to the UE. The use of the preamble given by the target eNodeB allows to use the reserved resources and therefore it is not necessary to perform a contention procedure. The random access procedure concludes with an acknowledgement message sent by the UE. This message is *RRC Connection Reconfiguration Complete*. From the reception of this message the radio connection with the target eNodeB is established. All the user's uplink data passes from the UE to the S-GW and P-GW via the target eNodeB. The downlink flow from the S-GW will still pass to the source eNodeB. Then, the source eNodeB will forward these flows to the target eNodeB which can then send them to the UE.

HO completion: The HO is now executed, however the MME is not yet aware of the HO. To avoid having the data forwarding phenomenon from the source eNodeB to the target eNodeB, and thus to correctly reroute the data to have a direct path, the target eNodeB must completely reroute the bearer tunnel on the downlink path, between S-GW and the target eNodeB. The target eNodeB sends a message *Path Switch Request* to the MME. The purpose of this message is to inform the MME of the route change and to request the S-GW to re-establish the bearer tunnel. For this, the MME sends a message *GTP-C Modify Bearer Request* to the S-GW. This message allows the exchange of the Tunnel Endpoint Identity (TEID) value on the eNodeB side, and thus to be able to re-establish the data tunnel between the target eNodeB and the S-GW. From this message, data transmission on the uplink and downlink channels is direct. However, the HO is not completely finished because the resources on the old eNodeB must be released.

For resource release, the S-GW will acknowledge the bearer modification request by sending a *GTP-C Modify Bearer Response* message to the MME. The MME then indicates to the target eNodeB that the re-establishment of the direct tunnels has been successful with a *STAP Path Switch Request Acknowledge* message. Finally, the target eNodeB indicates to the source eNodeB that it can completely release all its resources linked to the UE and thus confirm that the UE is completely managed by the new cell.

A.5.4 S1 handover

When two eNodeBs are not interconnected with a X2 interface, the HO is performed through an interaction with the MME via the S1 interface. From the UE point of view, there is no difference between S1 HO and X2 HO. Therefore, the preliminary phase is the same as X2 HO. Fig.A.8 shows the basic S1 HO procedure.

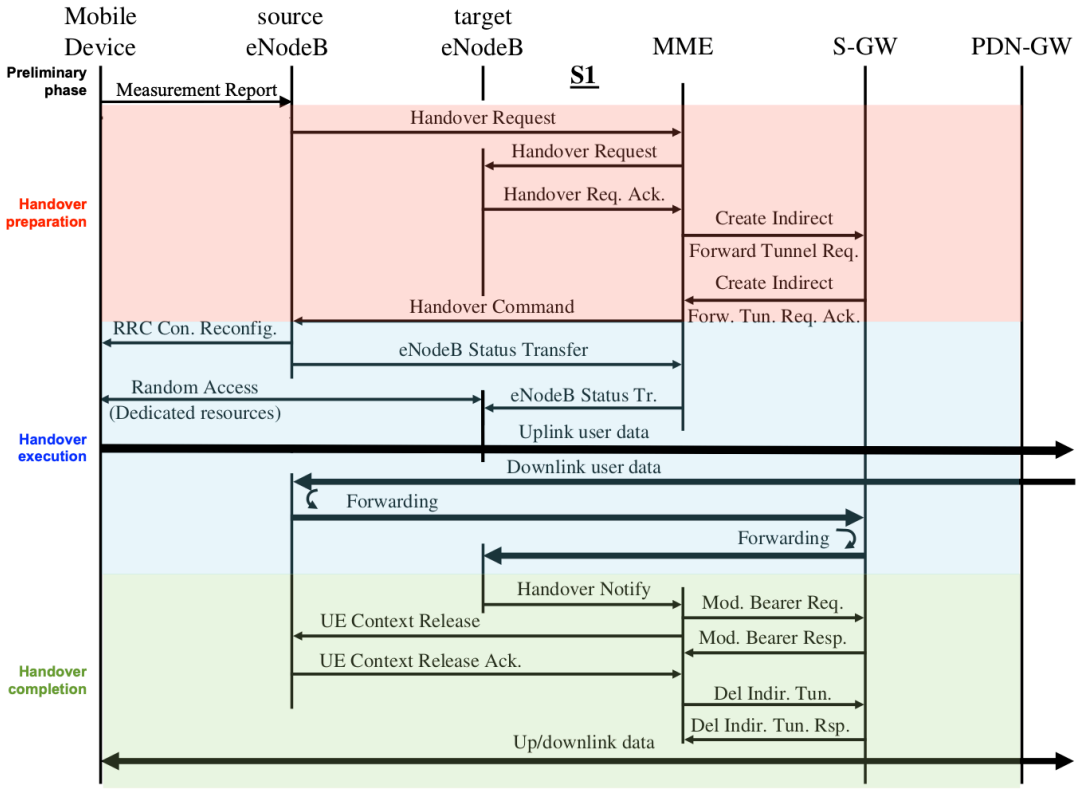


Figure A.8: S1 HO basic procedure

HO preparation: After choosing a target eNodeB, with which the source eNodeB does not have a direct link on the X2 interface, instead of sending a HO request directly to the target eNodeB, it sends it to the MME. The message will contain the information of the target eNodeB.

The MME will retransmit this request to the target eNodeB, which will perform an admission check. If it can handle the new UE, the target eNodeB will acknowledge this message positively with a *Handover Request Acknowledge*

To ensure that no packets are lost during the HO, the MME can request the S-GW to create a temporary indirect tunnel between the source eNodeB and the target eNodeB with the message *Create Indirect Forward Tunnel Request*. However, user data can also be forwarded directly with what is called direct forwarding. The S-GW confirms the creation of the temporary tunnel by acknowledging the MME request through a *Create Indirect Forward Tunnel Request Acknowledge* message. Once the temporary tunnel is created, the MME confirms the HO by sending a *Handover Command* message to the target eNodeB. This message contains information similar to the X2 *Handover Request Acknowledge* message.

HO execution: After receiving the *Handover Command*, the source eNodeB can instruct the UE to change cells, by sending it a *RRC Connection Reconfiguration* message, similar to X2 HO.

At the same time, the source eNodeB sends a *eNodeB Status Transfer* to the MME. The latter redirects the message to the target eNodeB. This message is similar to *SN Status Transfer* in X2 HO.

The UE will establish a link with the target eNodeB using the Random Access procedure, like in X2 HO. The forwarding process of user's downlink data is almost

similar to the X2. The difference is that, instead of forwarding the data directly from the source eNodeB to the target eNodeB via the X2 interface, the source eNodeB first forwards the data to the S-GW, which will then forward it to the target eNodeB.

HO completion: The HO is now executed. However, it is still necessary to modify the bearer data tunnel to have the data sent directly to the target eNodeB instead of being forwarded from the source eNodeB and release source eNodeB resources. The procedure for this is the same as for the X2 HO, with just a *Handover Notify* message from the eNodeB to the MME, instead of a *Path Switch Request* message.

Nevertheless, there is an additional step here. The network must delete the indirect forwarding tunnel on the S-GW created during the preparation phase. After this deletion, all resources that are no longer needed are removed and the user data flows directly to and from the target eNodeB.

Looking at the HO procedures in LTE, it is obvious that an incoming and an outgoing HO will not use the same number of resources. Differentiating eNodeBs with more incoming or outgoing HO profiles can therefore be interesting for resource management in the network.

A.6 5G handover

Like LTE, 5G is designed to be full IP. Therefore, HOs occur when a UE is exchanging data with the network while moving.

A.6.1 5G architecture

The 5G architecture, shown in Fig.A.9, is more complex than the LTE architecture, as it comprises more entities. However, each entity is virtualizable. The 5G network is divided into two parts, the CN represented by the 5G Core (5GC) and the RAN which is called New Generation RAN (NG-RAN).

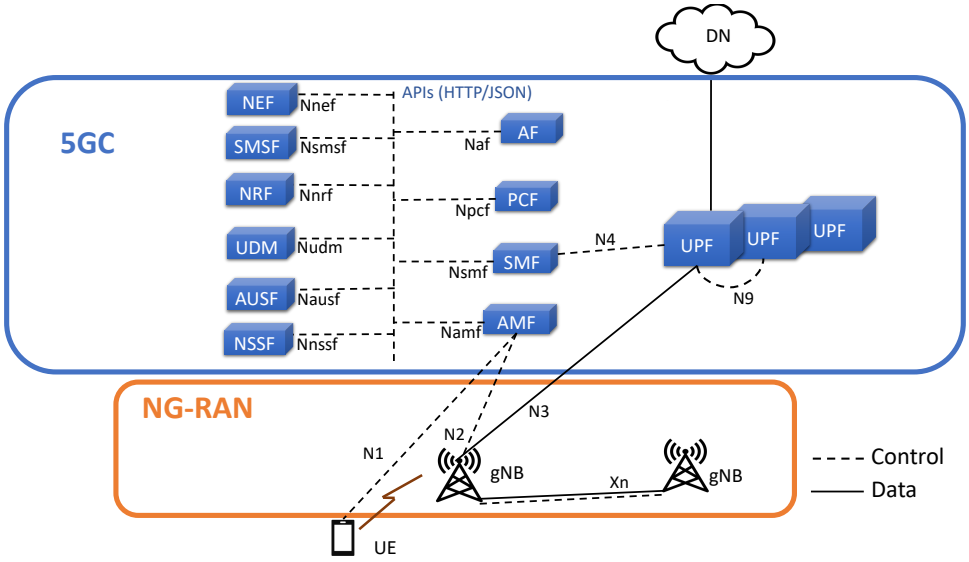


Figure A.9: 5G architecture

Designed to be service oriented, 5GC has more complex architecture than previous generations. Each entity or network function are virtual. The complexity in this

architecture can be explained in part by the virtualisation of network functions and the possibility of offering services from an external provider (other than the operator). Indeed, to have function instances that can be started dynamically or stopped, it is necessary to have well identified and well isolated functions. This is the reason the Control and User Plane Separation (CUPS) principle is used in 5G. CUPS makes it possible to separate everything that relates to the user plane from that which is in the control plane. And this functionality adds other network functions. However, this has the advantage of dynamism and flexibility. Indeed, one can for example add a new instance of function or remove some according to flows in the network, but also choose the location of these functions, according to the needs of the services which it implements.

5GC comprises several entities or network functions, the main ones are:

- Network exposure function (**NEF**) is the mandatory entry point for an external service provider, which will never dialogue directly with the functional entities of the network.
- Short Message Service Function (**SMSF**) is a service provided. In 5G, SMS can be transferred as signalling. The SMSF is therefore a sort of gateway between the external SMS server and the 5G network.
- Network repository function (**NRF**) is a directory service containing the list of virtual functions (available/used).
- Unified data management (**UDN**) and the Authentication server function (**AUSF**) can be assimilated to the HSS in LTE.
- Network slice selection function (**NSSF**) handles the selection of a slice considering the service-level agreement (SLA), and also provides the required QoE.
- Session management function (**SMF**) and Access and mobility function (**AMF**) can be assimilated to the MME in LTE. SMF also manages and supervises the PDN sessions in addition to the control plane.
- Policy Control Function (**PCF**) handles QoS, since it monitors and controls the traffic in the network for AMF and SMF.
- Application function (**AF**) provides the session related information to the PCF.
- User plane function (**UPF**) can be assimilated to S-GW and P-GW in LTE for user plane.

The NG-RAN, like in LTE, is comprised of a single entity, the gNodeB (gNB). However all control and computation functions can be virtual. Therefore, these functions can be relocated to a cloud. A gNB is composed of a physical part, which is the antenna (air interface), and two functions, which can be virtual, gNB-control unit (gNB-CU) and gNB-Distributed unit (gNB-DU). The gNB-CU provides mobility control, radio resource management (monitoring, allocation, etc) and session management. The gNB-DU provides the RLC, MAC and PHY layers functionality. gNB-CU and gNB-DU are interconnected with the F1 interface. One gNB-CU can be connected to several gNB-DU. And two gNB can be interconnected with the Xn interface. In addition to the features mentioned above, the gNB is also responsible for coding and modulation schemes, logical channels and SIBs, scheduling, encryption, data plane (SDAP, Service adaptation protocol), control plane (RRC), IP compression. It also manages QoS and network slicing.

A.6.2 Handover procedure

There are different procedures for HO in 5G, Xn and N2 [139], which occur when the UE moves in RRC connected state. These procedures are quite similar to X2 HO and S1 HO in LTE.

Xn HOs are only supported for intra-AMF mobility, therefore intra-NG-RAN. Indeed, like X2 HO, Xn HO can only be performed if two gNB are interconnected by interface Xn. However, unlike LTE, N2 HO in intra NG-RAN can sometimes be achieved, despite the presence of the Xn interface. N2 HO is mainly performed for inter NG-RAN and for inter AMF.

A.7 Conclusion

To conclude, we have seen in this chapter the different modes of execution for micro mobility control, especially the HO in the different generations of mobile networks. There are several types of HO. The hard HO is the HO type that is used in all mobile network generations. With the evolution of mobile network architecture, the HO scenario also evolved. However, LTE and 5G HO have similarities in their scenario and execution.

Bibliography

[1] V. Blondel, A. Decuyper, and G. Krings, "A survey of results on mobile phone datasets analysis," *EPJ Data Science*, vol. 4, no. 10, Aug 2015.

[2] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, 2010.

[3] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. International Conference on Knowledge Discovery and Data Mining*, 2011.

[4] D. D. C. Teixeira, A. Carneiro Viana, M. S. Alvim, and J. M. Almeida, "Deciphering Predictability Limits in Human Mobility," in *ACM SIGSPATIAL 2019 - 27th International Conference on Advances in Geographic Information Systems*, Chicago, United States, Nov. 2019.

[5] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, no. 7075, pp. 462–465, Jan. 2006.

[6] E. Mucelli Rezende Oliveira, A. Carneiro Viana, C. Sarraute, J. Brea, and I. Alvarez-Hamelin, "On the regularity of human mobility," *Pervasive and Mobile Computing*, vol. 33, pp. 73–90, Dec. 2016.

[7] E. Mucelli Rezende Oliveira, A. Carneiro Viana, K. Naveen, and C. Sarraute, "Mobile data traffic modeling: Revealing temporal facets," *Computer Networks*, vol. 112, pp. 176–193, Jan. 2017.

[8] M. Fekih, T. Bellemans, Z. Smoerda, P. Bonnel, A. Furno, and S. Galland, "A data-driven approach for origin-destination matrix construction from cellular network signalling data: A case study of lyon region $france$," *Transportation*, 2020.

[9] C. Sarraute, P. Blanc, and J. Burroni, "A study of age and gender seen through mobile phone usage patterns in mexico," in *ASONAM - IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, August 2014.

[10] R. Goel, R. Sharma, and A. Aasa, "Understanding gender segregation through call data records: An estonian case study," *PLoS ONE*, vol. 16, no. 3, March 2021.

[11] V. Traag, A. Browet, F. Calabrese, and F. Morlot, "Social event detection in massive mobile phone data using probabilistic location inference," in *SocialCom - IEEE 3rd International Conference on Social Computing*, October 2011.

[12] M. Berlingerio, F. Calabrese, G. D. Lorenzo, R. Nair, F. Pinelli, and L. Sbodio, "Allaboard: A system for exploring urban mobility and optimizing public transport using cellphone data"," in *Proc. ECML PKDD, Prague, Czechia*, Sep 2013.

[13] R. Wilson, E. Zu Erbach-Schoenberg, M. Albert, D. Power, S. Tudge, M. Gonzalez, S. Guthrie, H. Chamberlain, C. Brooks, C. Hughes, L. Pitonakova, C. Buckee, X. Lu, E. Wetter, A. Tatem, and L. Bengtsson, "Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 nepal earthquake," *PLoS currents*, vol. 8, Feb 2016, 26981327[pmid].

[14] M. Tizzoni, P. Bajardi, A. Decuyper, G. King, C. Schneide, V. Blondel, Z. Smoreda, M. Gonzalez, and V. Colizza, "On the use of human mobility proxies for the modeling of epidemics," *PLoS Computational Biology*, vol. 10, no. 7, Jul 2014.

[15] A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, and Z. Smoreda, "A tale of ten cities: Characterizing signatures of mobile traffic in urban areas," *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, pp. 2682–2696, 2016.

[16] S. Uppoor, C. Ziemlicki, S. Secci, and Z. Smoreda, "On mobile traffic distribution over cellular backhauling network nodes," in *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2016, pp. 726–731.

[17] M. Fiore, P. Katsikouli, E. Zavou, M. Cunche, F. Fessant, D. L. Hello, U. Aivodji, B. Olivier, T. Quertier, and R. Stanica, "Privacy in trajectory microdata publishing: A survey," *Transactions on Data Privacy*, vol. 13, Apr 2020.

[18] G. Ranjan, H. Zang, Z.-L. Zhang, and J. Bolot, "Are call detail records biased for sampling human mobility?" *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 16, no. 3, Jul 2012.

[19] Ericsson. Report and paper. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/

[20] W. Bank. Mobile cellular subscriptions. [Online]. Available: https://data.worldbank.org/indicator/IT.CEL.SETS

[21] A. U. Gawas, "An overview on evolution of mobile wireless communication networks: 1g-6g," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 3, no. 5, pp. 3130–3133, 2015.

[22] Thales-Group. 5g technology and networks (speed, use cases, rollout). [Online]. Available: https://www.thalesgroup.com/en/markets/digital-identity-and-security/mobile/inspired/5G

[23] M. Sauter, *From GSM to LTE-advanced Pro and 5G: An introduction to mobile networks and mobile broadband*. John Wiley & Sons, 2017.

[24] D. Naboulsi, M. Fiore, S. Ribot, and R. Stanica, "Large-scale mobile traffic analysis: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 124–161, 2015.

[25] A. Wesolowski, N. Eagle, A. Noor, R. Snow, and C. Buckee, "The impact of biases in mobile phone ownership on estimates of human mobility," *Journal of the Royal Society Interface*, vol. 10, no. 81, Feb 2013.

[26] L. Bengtsson, X. Lu, A. Thorson, R. Garfield, and J. Von Schreeb, "Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in haiti," *PLoS medicine*, vol. 8, no. 8, p. e1001083, 2011.

[27] A. Banchs, M. Fiore, A. Garcia-Saavedra, and M. Gramaglia, "Network intelligence in 6g: Challenges and opportunities," in *Proceedings of the 16th ACM Workshop on Mobility in the Evolving Internet Architecture*, 2021, pp. 7–12.

[28] M. Shafiq, L. Ji, A. Liu, J. Pang, S. Venkataraman, and J. Wang, "Estimating human trajectories and hotspots through mobile phone data," in *Proc. ACM SIGMETRICS Pittsburgh, USA*, Jun 2013.

[29] S. Hoteit, S. Secci, S. Sobolevsky, C. Ratti, and G. Pujolle, "Estimating human trajectories and hotspots through mobile phone data," *Computer Networks*, vol. 64, May 2014.

[30] S. Isaacman, R. Becker, R. Caceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky, "Ranges of human mobility in los angeles and new york," in *Proc. IEEE PerCom Seattle, USA*, Mar 2011.

[31] A. Ceselli, M. Fiore, A. Furno, S. Secci, and R. Stanica, "Prescriptive analytics for mec orchestration," in *Proc. IFIP Networking Zurich, Switzerland*, May 2018.

[32] B. Cici, A. Markopoulou, E. Frias-Martinezand, and N. Laoutaris, "Quantifying the potential of ride-sharing using call description records," in *Proc. ACM HotMobile Jekyll Island, USA*, Feb 2013.

[33] N. Caceres, L. Romero, and F. Benitez, "Exploring strengths and weaknesses of mobility inference from mobile phone data vs. travel surveys," *Transportmetrica A: Transport Science*, vol. 16, Feb 2020.

[34] Z. Qiu, J. Jin, P. Cheng, S. Secci, and R. Stanica, "State of the art and practice: Cellular probe technology applied in advanced traveler information systems," in *PProc. TRB 86th Annual Meeting, Washington, USA*, Jan 2007.

[35] R. W. Douglass, D. A. Meyer, M. Ram, D. Rideout, and D. Song, "High resolution population estimates from telecommunications data," *EPJ Data Science*, vol. 4, pp. 1–13, 2015.

[36] M. Lenormand, M. Picornell, O. G. Cantú-Ros, A. Tugores, T. Louail, R. Herranz, M. Barthelemy, E. Frias-Martinez, and J. J. Ramasco, "Cross-checking different sources of mobility information," *PloS one*, vol. 9, no. 8, p. e105184, 2014.

[37] F. Calabrese, G. D. Lorenzo, L. Liu, and C. Ratti, "Estimating origin-destination flows using mobile phone location data," *IEEE Pervasive Computing*, vol. 10, no. 4, Apr 2011.

[38] S. Bekhor, Y. Cohen, and C. Solomon, "Evaluating long-distance travel patterns in israel by tracking cellular phone positions," *Journal of Advanced Transportation*, vol. 47, no. 4, Jun 2013.

[39] C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, and M. C. González, "Unravelling daily human mobility motifs," *Journal of The Royal Society Interface*, vol. 10, no. 84, p. 20130246, 2013.

[40] D. Bachir, G. Khodabandelou, V. Gauthier, M. El Yacoubi, and J. Puchinger, "Inferring dynamic origin-destination flows by transport mode using mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 254–275, 2019.

[41] J. Ma, H. Li, F. Yuan, and T. Bauer, "Deriving operational origin-destination matrices from large scale mobile phone data," *International Journal of Transportation Science and Technology*, vol. 2, no. 3, Sep 2013.

[42] G. Chen, A. C. Viana, M. Fiore, and C. Sarraute, "Complete trajectory reconstruction from sparse mobile phone data," *EPJ Data Science*, vol. 8, no. 30, Oct 2019.

[43] L. Bonnetain, A. Furno, N.-E. E. Faouzi, M. Fiore, R. Stanica, Z. Smoreda, and C. Ziemlicki, "Transit: Fine-grained human mobility trajectory inference at scale with mobile network signaling data," *Transportation Research Part C: Emerging Technologies*, vol. 130, no. 103257, Sep 2021.

[44] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), May 2016.

[45] K. Grantz, H. Meredith, D. Cummings, J. Metcalf, B. Grenfell, J. Giles, S. Mehta, S. Solomon, A. Labrique, N. Kishore, C. Buckee, and A. Wesolowski, "The use of mobile phone data to inform analysis of covid-19 pandemic epidemiology," *Nature Communications*, vol. 11, no. 4961, Sep 2020.

[46] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, pp. 1–55, 2014.

[47] H. Huang, Y. Cheng, and R. Weibel, "Transport mode detection based on mobile phone network data: A systematic review," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 297–312, 2019.

[48] "Rapport sur le modele économique des transports collectifs," Jul 2021.

[49] M. Berlingerio, F. Calabrese, G. D. Lorenzo, R. Nair, F. Pinelli, and M. L. Sbodio, "Allaboard: a system for exploring urban mobility and optimizing public transport using cellphone data," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2013, pp. 663–666.

[50] M. Mamei, N. Bicocchi, M. Lippi, S. Mariani, and F. Zambonelli, "Evaluating origin–destination matrices obtained from cdr data," *Sensors*, vol. 19, no. 20, p. 4470, 2019.

[51] S. B. Elagib, A.-H. A. Hashim, and R. Olanrewaju, "Cdr analysis using big data technology," in *2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)*. IEEE, 2015, pp. 467–471.

[52] S. Manchanayake, D. Samarasinghe, L. Perera, H. Bandara, K. Kumaradasa, N. Premadasa, and A. Samarasinghe, "Potential upselling customer prediction through user behavior analysis based on cdr data," in *2019 14th Conference on Industrial and Information Systems (ICIIS)*. IEEE, 2019, pp. 46–51.

[53] Q. Lin and Y. Wan, "Mobile customer clustering based on call detail records for marketing campaigns," in *2009 International Conference on Management and Service Science*. IEEE, 2009, pp. 1–4.

[54] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and pois," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 186–194.

[55] X. Liu, J. He, Y. Yao, J. Zhang, H. Liang, H. Wang, and Y. Hong, "Classifying urban land use by integrating remote sensing and social media data," *International Journal of Geographical Information Science*, vol. 31, no. 8, pp. 1675–1696, 2017.

[56] M. Y. Santos and A. Moreira, "Automatic classification of location contexts with decision trees," 2006.

[57] S.-S. Wu, X. Qiu, E. L. Usery, and L. Wang, "Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use," *Annals of the Association of American Geographers*, vol. 99, no. 1, pp. 76–98, 2009.

[58] J. L. Toole, M. Ulm, M. C. González, and D. Bauer, "Inferring land use from mobile phone activity," in *Proceedings of the ACM SIGKDD international workshop on urban computing*, 2012, pp. 1–8.

[59] J. Reades, F. Calabrese, and C. Ratti, "Eigenplaces: analysing cities using the space–time structure of the mobile phone network," *Environment and Planning B: Planning and Design*, vol. 36, no. 5, pp. 824–836, 2009.

[60] A. Furno, M. Fiore, and R. Stanica, "Joint spatial and temporal classification of mobile traffic demands," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[61] H. Mao, Y.-Y. Ahn, B. Bhaduri, and G. Thakur, "Improving land use inference by factorizing mobile phone call activity matrix," *Journal of land use science*, vol. 12, no. 2-3, pp. 138–153, 2017.

[62] T. Pei, S. Sobolevsky, C. Ratti, S.-L. Shaw, T. Li, and C. Zhou, "A new insight into land use classification based on aggregated mobile phone data," *International Journal of Geographical Information Science*, vol. 28, no. 9, pp. 1988–2007, 2014.

[63] X. Zhan, S. V. Ukkusuri, and F. Zhu, "Inferring urban land use using large-scale social media check-in data," *Networks and Spatial Economics*, vol. 14, no. 3, pp. 647–667, 2014.

[64] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[65] F. Calabrese, J. Reades, and C. Ratti, "Eigenplaces: segmenting space through digital signatures," *IEEE Pervasive Computing*, vol. 9, no. 1, pp. 78–84, 2009.

[66] V. Soto and E. Frías-Martínez, "Automated land use identification using cell-phone records," in *Proceedings of the 3rd ACM international workshop on MobiArch*, 2011, pp. 17–22.

[67] H. Dong, M. Wu, X. Ding, L. Chu, L. Jia, Y. Qin, and X. Zhou, "Traffic zone division based on big data from mobile phone base stations," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 278–291, 2015.

[68] S. Grauwin, S. Sobolevsky, S. Moritz, I. Gódor, and C. Ratti, "Towards a comparative science of cities: Using mobile traffic records in new york, london, and hong kong," in *Computational approaches for urban environments*. Springer, 2015, pp. 363–387.

[69] B. Cici, M. Gjoka, A. Markopoulou, and C. T. Butts, "On the decomposition of cell phone activity patterns and their connection with urban ecology," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2015, pp. 317–326.

[70] A. Furno, D. Naboulsi, R. Stanica, and M. Fiore, "Mobile demand profiling for cellular cognitive networking," *IEEE Transactions on Mobile Computing*, vol. 16, no. 3, pp. 772–786, 2016.

[71] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, H. Janicke, and S. K. Jha, "A survey of covid-19 contact tracing apps," *IEEE Access*, vol. 8, pp. 134 577–134 601, July 2020.

[72] D. J. Leith and S. Farrell, "Coronavirus contact tracing: Evaluating the potential of using bluetooth received signal strength for proximity detection," *ACM SIGCOMM Computer Communication Review*, vol. 50, no. 4, October 2020.

[73] C. Castelluccia, N. Bielova, A. Boutet, M. Cunche, C. Lauradoux, D. Le Metayer, and V. Roca, "Robert: Robust and privacy-preserving proximity tracing," Inria, Technical Report, May 2020.

[74] A. Wesolowski, N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee, "Quantifying the impact of human mobility on malaria," *Science*, vol. 338, no. 6104, pp. 267–270, October 2012.

[75] S. Brdar, K. Gavric, D. Culibrk, and V. Crnojevic, "Unveiling spatial epidemiology of hiv with mobile phone data," *Scientific Reports*, vol. 6, January 2016.

[76] H. Yi, S. T. Ng, A. Farwin, A. P. Ting Low, C. M. Chang, and J. Lim, "Health equity considerations in covid-19: Geospatial network analysis of the covid-19 outbreak in the migrant population in singapore," *Journal of Travel Medicine*, vol. 28, no. 2, March 2021.

[77] S. Park, G. J. Choi, and H. Ko, "Information technology–based tracing strategy in response to covid-19 in south korea—privacy controversies," *JAMA*, vol. 323, no. 21, April 2020.

[78] C. J. Wang, C. Y. Ng, and R. H. Brook, "Response to covid-19 in taiwan: Big data analytics, new technology, and proactive testing," *JAMA*, vol. 323, no. 14, March 2020.

[79] N. Oliver, B. Lepri, H. Sterly, R. Lambiotte, S. Deletaille, M. De Nadai, E. Letouze, A. A. Salah, R. Benjamins, C. Cattuto, V. Colizza, N. de Cordes, S. P. Fraiberger, T. Koebe, S. Lehmann, J. Murillo, A. Pentland, P. N. Pham, F. Pivetta, J. Saramaki, S. S. V., M. Tizzoni, S. Verhulst, and P. Vinck, "Mobile phone data for informing public health actions across the covid-19 pandemic life cycle," *Science Advances*, vol. 6, no. 23, June 2020.

[80] Y. Hara and H. Yamaguchi, "Japanese travel behavior trends and change under covid-19 state-of-emergency declaration: Nationwide observation by mobile phone location data," *Transportation Research Interdisciplinary Perspectives*, vol. 9, March 2021.

[81] E. Willberg, O. Jarv, T. Vaisanen, and T. Toivonen, "Escaping from cities during the covid-19 crisis: Using mobile phone data to trace mobility in finland," *International Journal of Geo-Information*, vol. 10, no. 2, February 2021.

[82] P. S. Peixoto, D. Marcondes, C. Peixoto, and S. M. Oliva, "Modeling future spread of infections via mobile geolocation data and population dynamics. an application to covid-19 in brazil," *PLoS ONE*, vol. 15, no. 7, July 2020.

[83] K. Czech, A. Davy, and M. Wielekhowski, "Does the covid-19 pandemic change human mobility equally worldwide? cross-country cluster analysis," *Economies*, vol. 9, no. 4, November 2021.

[84] G. Pullano, L. Di Domenico, C. E. Sabbatini, E. Valdano, C. Turbelin, M. Debin, C. Guerrisi, C. Kengne-Kuetche, C. Souty, T. Hanslik, T. Blanchon, P.-Y. Boëlle, J. Figoni, S. Vaux, C. Campèse, S. Bernard-Stoecklin, and V. Colizza, "Underdetection of cases of covid-19 in france threatens epidemic control," *Nature*, vol. 590, no. 7844, pp. 134–139, Feb 2021.

[85] L. Di Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, and V. Colizza, "Impact of lockdown on covid-19 epidemic in île-de-france and possible exit strategies," *BMC Medicine*, vol. 18, no. 1, p. 240, Jul 2020.

[86] H. S. Badr and L. M. Gardner, "Limitations of using mobile phone data to model covid-19 transmission in the usa," *The Lancet*, vol. 21, no. 5, November 2020.

[87] G. Hadjidemetriou, M. Sasidharan, G. Kouyialis, and A. Parlikad, "The impact of government measures and human mobility trend on covid-19 related deaths in the uk," *Transportation Research Interdisciplinary Perspectives*, vol. 6, no. 100167, July 2020.

[88] T. Li, M. Zhang, Y. Li, E. Lagerspetz, S. Tarkoma, and P. Hui, "The impact of covid-19 on smartphone usage," *IEEE Internet of Things Journal*, April 2021.

[89] Y. Zhou, R. Xu, D. Hu, Y. Yue, Q. LI, and J. Xia, "Effects of human mobility restrictions on the spread of covid-19 in shenzhen, china: A modelling study using mobile phone data," *The Lancet Digital Health*, vol. 2, no. 8, pp. 417–424, August 2020.

[90] A. F. Zanella, O. E. Martínez-Durive, S. Mishra, Z. Smoreda, and M. Fiore, "Impact of later-stages covid-19 response measures on spatiotemporal mobile service usage," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 970–979.

[91] N. Ayan, N. Damasceno, S. Chaskar, P. de Sousa, A. Ramesh, A. Seetharam, and A. de A. Rocha, "Characterizing human mobility patterns during covid-19 using cellular network data," in *LCN - IEEE 46th Conference on Local Computer Networks*, October 2021.

[92] G. Romanillos, J. C. García-Palomares, B. Moya-Gómez, J. Gutiérrez, J. Torres, M. López, O. G. Cantú-Ros, and R. Herranz, "The city turned off: Urban dynamics during the covid-19 pandemic based on mobile phone data," *Applied Geography*, vol. 134, p. 102524, 2021.

[93] Y. Li, H. Deng, J. Li, C. Peng, and S. Lu, "Instability in distributed mobility management: Revisiting configuration management in 3g/4g mobile networks," in *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, 2016, pp. 261–272.

[94] A. Balachandran, V. Aggarwal, E. Halepovic, J. Pang, S. Seshan, S. Venkataraman, and H. Yan, "Modeling web quality-of-experience on cellular networks," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 213–224.

[95] M. Z. Shafiq, J. Erman, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Understanding the impact of network dynamics on mobile video user engagement," *ACM SIGMETRICS Performance Evaluation Review*, vol. 42, no. 1, pp. 367–379, 2014.

[96] Z. Ali, N. Baldo, J. Mangues-Bafalluy, and L. Giupponi, "Machine learning based handover management for improved qoe in lte," in *NOMS 2016-2016 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2016, pp. 794–798.

[97] U. S. Hashmi, A. Rudrapatna, Z. Zhao, M. Rozwadowski, J. Kang, R. Wuppalapati, and A. Imran, "Towards real-time user qoe assessment via machine learning on lte network data," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*. IEEE, 2019, pp. 1–7.

[98] Q. Duan and S. Wang, "Network cloudification enabling network-cloud/fog service unification: state of the art and challenges," in *2019 IEEE World Congress on Services (SERVICES)*, vol. 2642. IEEE, 2019, pp. 153–159.

[99] P. Rost, A. Banchs, I. Berberana, M. Breitbach, M. Doll, H. Droste, C. Mannweiler, M. A. Puente, K. Samdanis, and B. Sayadi, "Mobile network architecture evolution toward 5g," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 84–91, 2016.

[100] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5g with sdn/nfv: Concepts, architectures, and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 80–87, 2017.

[101] B. Blanco, J. O. Fajardo, I. Giannoulakis, E. Kafetzakis, S. Peng, J. Pérez-Romero, I. Trajkovska, P. S. Khodashenas, L. Goratti, M. Paolino *et al.*, "Technology pillars in the architecture of future 5g mobile networks: Nfv, mec and sdn," *Computer Standards & Interfaces*, vol. 54, pp. 216–228, 2017.

[102] D. Naboulsi, A. Mermouri, R. Stanica, H. Rivano, and M. Fiore, "On user mobility in dynamic cloud radio access networks," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 1583–1591.

[103] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri, "Towards realistic mobility models for mobile ad hoc networks," in *Proceedings of the 9th annual international conference on Mobile computing and networking*, 2003, pp. 217–229.

[104] M. Musolesi and C. Mascolo, "Designing mobility models based on social network theory," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 3, pp. 59–70, 2007.

[105] K. Xu, R. Singh, M. Fiore, M. K. Marina, H. Bilen, M. Usama, H. Benn, and C. Ziemlicki, "Spectragan: Spectrum based generation of city scale spatiotemporal mobile network traffic data," in *Proceedings of the 17th International Conference on emerging Networking EXperiments and Technologies*, 2021, pp. 243–258.

[106] Y. Wang, A. Yalcin, and C. VandeWeerd, "An entropy-based approach to the study of human mobility and behavior in private homes," *PLoS one*, vol. 15, no. 12, p. e0243503, 2020.

[107] Y. Benezeth, H. Laurent, B. Emile, and C. Rosenberger, "Towards a sensor for detecting human presence and characterizing activity," *Energy and Buildings*, vol. 43, no. 2-3, pp. 305–314, 2011.

[108] S. Rabenjamina, R. Stanica, O. Iova, and H. Rivano, "Comparison of User Presence Information from Mobile Phone and Sensor Data," in *MSWiM 2022 – 25th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Montreal, Canada, Oct. 2022.

[109] M. Dahan, A. A. Mbacké, O. Iova, and H. Rivano, "Challenges of designing smart lighting," in *EWSN 2020-International Conference on Embedded Wireless Systems and Networks*. ACM, 2020, pp. 1–6.

[110] S. Jiang, G. A. Fiore, Y. Yang, J. Ferreira Jr, E. Frazzoli, and M. C. González, "A review of urban computing for mobile phone traces: Current methods, challenges and opportunities," in *Proceedings of the 2nd ACM SIGKDD international workshop on Urban Computing*. ACM, 2013, p. 2.

[111] C. Iovan, A.-M. Olteanu-Raimond, T. Couronné, and Z. Smoreda, "Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies," in *Geographic Information Science at the Heart of Europe*. Springer, 2013, pp. 247–265.

[112] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, "Approaching the limit of predictability in human mobility," *Scientific Reports*, vol. 3, no. 1, p. 2923, Oct. 2013.

[113] H. Fanticelli, S. Rabenjamina, A. Carneiro Viana, R. Stanica, L. Santos de Oliveira, and A. Ziviani, "Data-driven Mobility Analysis and Modeling: Typical and Confined Life of a Metropolitan Population," *ACM Transactions on Spatial Algorithms and Systems*, vol. 8, no. 3, pp. 1–33, Sep. 2022.

[114] Y. Leo, A. Busson, C. Sarraute, and E. Fleury, "Call detail records to characterize usages and mobility events of phone users," *Computer Communications*, vol. 95, pp. 43–53, 2016.

[115] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: An overview," *Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2011.

[116] I. Gronau and S. Moran, "Optimal implementation of upgma and other common clustering algorithms," *Information Processing Letters*, vol. 104, no. 6, pp. 205–210, 2007.

[117] L. Figaro. Règlement, public, huis clos ... comment la ligue des champions vit avec le covid-19. [Online]. Available: https://www.lefigaro.fr/sports/football/ligue-des-champions/actualites/une-ligue-des-champions-1018522

[118] L. journal du dimanche. A paris, le jardin des tuileries se refait une beauté. [Online]. Available: https://www.lejdd.fr/Societe/a-paris-le-jardin-des-tuileries-se-refait-une-beaute-3994614

[119] C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, C. Ziemlicki, and Z. Smoreda, "Not all apps are created equal: Analysis of spatiotemporal heterogeneity in nationwide mobile service usage," in *Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies*, 2017, pp. 180–186.

[120] C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Perez, "How should i slice my network? a multi-service empirical evaluation of resource sharing efficiency," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, pp. 191–206.

[121] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling internet traffic dynamics of cellular devices," *ACM SIGMETRICS Performance Evaluation Review*, vol. 39, no. 1, pp. 265–276, 2011.

[122] R. Keralapura, A. Nucci, Z.-L. Zhang, and L. Gao, "Profiling users in a 3g network using hourglass co-clustering," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, 2010, pp. 341–352.

[123] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.

[124] scipy doc. scipy.cluster.hierarchy.linkage. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html#scipy.cluster.hierarchy.linkage

[125] LeFigaro. Circulation à paris : les points noirs de 2019. [Online]. Available: https://www.lefigaro.fr/fig-data/trafic-paris/

[126] Statista. Le top 10 des quartiers d'affaires. [Online]. Available: https://fr.statista.com/infographie/11938/le-top-10-des-quartiers-daffaires/

[127] L. Parisien. Paris : près de 110000 véhicules transitent chaque jour par le rond-point de la place de l'Étoile. [Online]. Available: https://www.leparisien.fr/paris-75/paris-pres-de-110000-vehicules-transitent-chaque-jour-par-le-rond-point-de-la-place-de-letoile-21-05-2022-PAGYJ2QTNJEC5AI5KYFVQ7ZDKU.php

[128] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega *et al.*, "Network slicing to enable scalability and flexibility in 5g mobile networks," *IEEE Communications magazine*, vol. 55, no. 5, pp. 72–79, 2017.

[129] ETSI, "Digital cellular telecommunications system (phase 2+) (gsm); universal mobile telecommunications system (umts); lte; 5g; release 15 description," European Telecommunications Standards Institute (ETSI), Technical Report (TR) 21.915, 10 2019, version 15.0.0. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/121900_121999/121915/15.00.00_60/tr_121915v150000p.pdf

[130] ——, "Digital cellular telecommunications system (phase 2+) (gsm) ; universal mobile telecommunications system(umts) ; lte ; 5g ; release 16 description;," European Telecommunications Standards Institute (ETSI), Technical Report (TR) 21.916, 07 2022, version 16.2.0. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/121900_121999/121916/16.02.00_60/tr_121916v160200p.pdf

[131] 3GPP, "Study on architecture for next generation system," 3GPP, Technical Report (TR) 23.799, 11 2016, version 14.2.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3008

[132] J. G. Herrera and J. F. Botero, "Resource allocation in nfv: A comprehensive survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 518–532, 2016.

[133] M. M. Sajjad, C. J. Bernardos, D. Jayalath, and Y.-C. Tian, "Inter-slice mobility management in 5g: Motivations, standard principles, challenges, and research directions," *IEEE Communications Standards Magazine*, vol. 6, no. 1, pp. 93–100, 2022.

[134] R. A. Addad, T. Taleb, H. Flinck, M. Bagaa, and D. Dutra, "Network slice mobility in next generation mobile systems: Challenges and potential solutions," *IEEE Network*, vol. 34, no. 1, pp. 84–93, 2020.

[135] I. Saffar, M. L. A. Morel, K. D. Singh, and C. Viho, "Deep learning based speed profiling for mobile users in 5g cellular networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–7.

[136] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5g mmwave mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2069–2084, 2017.

[137] M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Multi-connectivity in 5g mmwave cellular networks," in *2016 Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*. IEEE, 2016, pp. 1–7.

[138] ——, "Uplink-based framework for control plane applications in 5g mmwave cellular networks," *arXiv preprint arXiv:1610.04836*, 2016.

[139] ETSI, "5g;procedures for the 5g system," European Telecommunications Standards Institute (ETSI), Technical Specification (TS) 23.502, 05 2022, version 17.4.0. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123500_123599/123502/17.04.00_60/ts_123502v170400p.pdf

FOLIO ADMINISTRATIF

THESE DE L'INSA LYON, MEMBRE DE L'UNIVERSITE DE LYON

NOM : RABENJAMINA DATE de SOUTENANCE : 29/09/2023

Prénoms : Solohaja Ramahaimandamina Andriantoavina Mihael

TITRE : Human And Network Mobility Management Using Mobile Phone Data

NATURE : Doctorat Numéro d'ordre :2023ISAL0063

Ecole doctorale : InfoMaths

Spécialité : Informatique

RESUME :

Au cours de la dernière décennie, l'utilisation croissante des smartphones a entraîné une augmentation significative du volume de données échangées via les réseaux mobiles des opérateurs téléphoniques. Chaque nouvelle génération de réseau mobile génère davantage de données que la précédente. D'ici à 2027, on estime que 289 EB de données seront échangées par mois, dont 62% proviendront du réseau mobile 5G.
Cette disponibilité massive de données a ouvert de nouvelles perspectives de recherche, notamment dans l'étude de la mobilité. Les données mobiles permettent des études sur une population plus vaste et des zones géographiques étendues. Dans cette thèse, nous démontrons que les événements décrits dans les données mobiles peuvent être retrouvés dans d'autres sources de données. En comparant les données mobiles avec des capteurs de détection de présence humaine, nous constatons une corrélation satisfaisante. Cependant, certains événements, tels que la synchronisation des pics de présence ou la fin de l'activité en fin de journée, ont une similarité moindre.
Nous utilisons également les données mobiles pour étudier l'impact des confinements imposés par le gouvernement français sur l'utilisation du sol à Paris. Nos résultats montrent que le premier confinement a eu un impact radical sur les habitudes de déplacement et l'utilisation du sol, tandis que les deuxième et troisième confinements ont eu un impact moindre.
Enfin, nous exploitons ces données pour la reconfiguration du réseau mobile dans la gestion de la micro mobilité des utilisateurs, appelée handover. Les eNodeB, composants du réseau d'accès, peuvent avoir différents profils et catégories. En distinguant les utilisateurs mobiles des utilisateurs statiques, nous économisons des ressources en reconfigurant le réseau. La reconfiguration dynamique du réseau, en utilisant différents profils d'eNodeB, permet également d'économiser les ressources utilisées par les utilisateurs mobiles.

MOTS-CLÉS : Données mobile, Réseau mobile, Handover, gestion de la mobilité

Laboratoire (s) de recherche : CITI

Directeur de thèse : Hervé RIVANO

Président de jury :

Composition du jury :

Marco FIORE	Directeur de Recherche	IMDEA Networks	Rapporteur
Vania CONAN	Habilité à Diriger des Recherches	Thales	Rapporteur
Aline CARNEIRO VIANA	Directeur de Recherche	INRIA	Examinateur
Sahar HOTEIT	Maître de conférences	Université Paris Saclay	Examinateur
Stefano SECCI	Professeur des Universités	CNAM	Examinateur
Hervé RIVANO	Professeur des Universités	INSA-Lyon	Directeur de Thèse
Razvan STANICA	Maître de conférences HDR	INSA-Lyon	Co-directeur de Thèse