TeMA

Journal of
Land Use, Mobility and Environment

The Special Issue collects six papers that use mobile phone data and spatial analysis techniques to study new urban critical features and social phenomena that arose with the Covid-19 pandemic. The applications of mobile phone data in the three study contexts investigated the potentialities of mobile phone data, as well as their limits. Compared to traditional methods of urban survey mobile phone data provide real-time maps of daily practices.

TeMA Journal offers papers with a unified approach to planning, mobility and environmental sustainability. With ANVUR resolution of April 2020, TeMA journal and the articles published from 2016 are included in the A category of scientific journals. From 2015, the articles published on TeMA are included in the Core Collection of Web of Science. It is included in Sparc Europe Seal of Open Access Journals, and the Directory of Open Access Journals.

Special Issue 2.2022

Mobile phone data for exploring spatio-temporal transformations in contemporary territories

print ISSN 1970-9889 e-ISSN 1970-9870
University of Naples Federico II

TeMA Journal of
Land Use, Mobility and Environment

Special Issue 2.2022

Mobile phone data for exploring spatio-temporal transformations in contemporary

Published by
Laboratory of Land Use Mobility and Environment
DICEA - Department of Civil, Architectural and Environmental Engineering
University of Naples "Federico II"

TeMA is realized by CAB - Center for Libraries at "Federico II" University of Naples using Open Journal System

Editor-in-chief: Rocco Papa
print ISSN 1970-9889 | online ISSN 1970-9870
Licence: Cancelleria del Tribunale di Napoli, n° 6 of 29/01/2008

Editorial correspondence
Laboratory of Land Use Mobility and Environment
DICEA - Department of Civil, Architectural and Environmental Engineering
University of Naples "Federico II"
Piazzale Tecchio, 80
80125 Naples
web: www.tema.unina.it
e-mail: redazione.tema@unina.it

TeMA. Journal of Land Use, Mobility and Environment offers researches, applications and contributions with a unified approach to planning and mobility and publishes original inter-disciplinary papers on the interaction of transport, land use and environment. Domains include: engineering, planning, modeling, behavior, economics, geography, regional science, sociology, architecture and design, network science and complex systems.

With ANVUR resolution of April 2020, TeMA Journal and the articles published from 2016 are included in A category of scientific journals. From 2015, the articles published on TeMA are included in the Core Collection of Web of Science. TeMA Journal has also received the *Sparc Europe Seal* for Open Access Journals released by *Scholarly Publishing and Academic Resources Coalition* (SPARC Europe) and the *Directory of Open Access Journals* (DOAJ). TeMA is published under a Creative Commons Attribution 4.0 License and is blind peer reviewed at least by two referees selected among high-profile scientists. TeMA has been published since 2007 and is indexed in the main bibliographical databases and it is present in the catalogues of hundreds of academic and research libraries worldwide.

TeMA

Journal of
Land Use, Mobility and Environment

Special Issue 2.2022

MOBILE PHONE DATA FOR EXPLORING SPATIO-TEMPORAL TRANSFORMATIONS IN CONTEMPORARY

Contents

TeMA
Journal of
Land Use, Mobility and Environment

TeMA Special Issue 2 (2022) 25-37
print ISSN 1970-9889, e-ISSN 1970-9870
DOI: 10.6092/1970-9870/8953
Received 4th February 2022, Accepted 14th November 2022, Available online 30th November 2022

A glimpse into mobile phone data: characteristics, organization, tools

Fabio Manfredini [a]*, Carmelo Di Rosa [b], Francesco Fagiani [c], Viviana Giavarini [d]

[a] Department of Architecture and Urban Studies
Politecnico di Milano, Milano, Italy
e-mail: fabio.manfredini@polimi.it
ORCID: https://orcid.org/0000-0002-6136-150X
* Corresponding author

[b] Department of Architecture and Urban Studies
Politecnico di Milano, Milano, Italy
e-mail: carmelo.dirosa@polimi.it
ORCID: https://orcid.org/0000-0002-6965-3806

[c] Department of Architecture and Urban Studies
Politecnico di Milano, Milano, Italy
e-mail: francesco.fagiani@polimi.it
ORCID: https://orcid.org/0000-0002-6688-5214

[d] Department of Architecture and Urban Studies
Politecnico di Milano, Milano, Italy
e-mail: viviana.giavarini@polimi.it
ORCID: https://orcid.org/0000-0002-8920-9999

Abstract

This paper aims to present the presence and mobility data provided by TIM, highlighting the acquisition methodology, the levels of spatial and temporal disaggregation, as well as the additional information related to age groups, gender, and classification of behaviours, which are directly supplied by TIM. The construction of a baseline based on mobile phone data for the comparison of temporal trends in the presence of people is also discussed.

At the same time, the supporting data obtained from traditional sources or ad hoc surveys will be presented to show how they can facilitate the interpretation of telephone data, its validation, and its use. Finally, a reference on the operational tools used for their processing and visualization will highlight the need to integrate skills, methodologies, and tools for the maximum exploitation of this wealth of information.

Keywords

Mobile phone data; Data manipulation; Presence.

How to cite item in APA format

Manfredini, F., Di Rosa, C., Fagiani, F., Giavarini, V. (2023). A glimpse into mobile phone data: characteristics, organization, tools. *Tema. Journal of Land Use, Mobility, and Environment*, 25-37. http://dx.doi.org/10.6092/1970-9870/8953

1. The generation of mobile phone data

In mobile phone research, the data on the users' activity acquired by the telephone network cells are provided as spatial and temporal data on the overall number of people and on their fluxes based on complex methodologies that the telephone operators define to attribute the data to comparable spatial areas.

The understanding of the general principles on the generation of telephone data together with the detailed description of the characteristics of the data is two fundamental conditions to use this data and to determine its potential in the context of urban studies. The intensity of mobile phone activity in a cell (i.e. the area covered around an antenna) is proportional to the presence of mobile phone users (Sevtsuk and Ratti, 2010; Reades et al., 2007; Ratti et al., 2006; Ahas and Mark, 2005). Most of the research in the last decade focused on the Erlang data provided at the time by mobile phone companies. Erlang[1] is the average number of concurrent contacts in a time unit and has been used to describe the correlation between mobile phone data and people's daily activities (Ratti et al., 2006; Sevtsuk and Ratti, 2010). In more recent years, mobile phone companies started to provide data about the overall number of users' presence and mobility. This made it possible to obtain a more directly usable figure as it is expressed as the number of active users instead of Erlang which is a measure not directly attributable to individuals but their activity. Moreover, TIM, like other telephone companies offering data analytics services, has developed a proprietary model for extracting, analysing, and providing data acquired from telephone cells, for small portions of territory such as 250x250 pixels or statistical aggregations such as municipalities or Census Areas (ACE). The model takes into consideration the land cover by assigning the probability of finding a person to the different land covers. According to this model, urban areas will therefore have a major weight than agricultural or forestry areas.

The value of mobile phone data for urban studies is widely recognized since the last decade (Steenbruggen et al., 2015; Blondel et al., 2015; Manfredini et al., 2013) mostly because of their spatial and temporal resolution which can be significantly higher than that of conventional data. In fact, while mobile phone data are available at a very fine temporal scale (hourly or sub hourly), conventional data have considerable limitations with respect to time dimension as they are updated on an annual or multi-year scale. Moreover, in order to intercept urban phenomena related to temporary populations and mobility issues, there is a need to acquire information that goes beyond administrative boundaries usually used in conventional data surveys. Mobile phone data as they are acquired from the activities of individual users, can be aggregated at different spatial scales (for istance square cells, customized areas, neighbourhoods) and can therefore describe complex urban phenomena very effectively and at the correct geographical scale.

Among the main applications of mobile phone data in urban studies, it is possible to cite:

- mobility behaviors of individuals at a global scale (Gonzalez et al., 2008);
- spatial structure of cities based on average distance between mobile phone users (Louail et al., 2014);
- emergency management (Gething et al., 2011; Dobra et al., 2015);
- mobility practices and travel behaviors (Pucci et al., 2015; Wang et al., 2018);
- land use classification and identification (Soto et al., 2011; Toole et al., 2012);
- temporary uses of urban spaces and tourism analysis (Manfredini et al., 2011; Ahas et al., 2008);
- improving quality of official statistics (Barcaroli et al., 2014; Vanhoof et al., 2018);

This rich but still partial list of mobile phone data applications for urban studies defines a possible urban agenda for improving the knowledge of urban phenomena that are difficult to intercept through conventional

[1] One erlang is the equivalent of one call in a specific channel for 3600 seconds in an hour. there are many ways in which a channel can carry a certain number of erlangs. For example, a traffic density of 3 erlangs can consist of three simultaneous calls, each lasting for an hour or of six calls, each of which are allocated 30 minutes or 180 calls, each of which occupy one minute.

data. The current research experience based on TIM data aims at investigating some emergent urban issues in the broad field of territorial fragilities in three different Italian contexts (Curci et al., 2022).

2. TIM mobile phone data

In recent years, TIM has developed a platform called DVI (Data Visual Insight), which makes it possible to view and download mobile phone data in text format (CSV) with a very high temporal resolution (available every 15 minutes for a while no longer than 2 years before the current date) on a spatial resolution of the ACE (Census Area).

Census Areas (ACE) are provided by ISTAT (National Statistics Italian Institution) and correspond to an aggregation of contiguous census tracts, the smallest Italian statistical unit consistent to the building block in urban areas. Census areas have been defined only for the Italian cities with an overall population of around 100.000 inhabitants. Every ACE has many inhabitants between 13,000 and 18,000, with some exceptions.

This data is collected through the network from the mobile phone activity of standard Sim cards[2] by TIM customers representing a market share of about 26% on a national basis (AGCOM, 2022[3]) at different levels of aggregation: Regions, Provinces, Municipalities, and ACEs.

Through the DVI platform, it is possible to access two different data categories at the ACE level: Presences and Mobility for the three active study areas selected, named by TIM "Scenario".

Case study	**Spatial data available**	**Research topic**
Milano	85 ACE	Near working during the pandemic in the Milano districts[4]
Piacenza	33 municipalities with the distinction of the main urban settlements	New home/work and tourism-related mobility practices in low-density mountain area during the Covid-19 pandemic[5]
Lecce	1 municipality with 9 coastal areas	Multiresidentiality in coastal areas marked by second homes and unauthorised construction[6]

Tab.1 Active TIM scenario available in the DVI platform

In two case study (Milano and Piacenza), new disaggregated areas have been selected to obtain data on the human presence for the analysis of specific urban phenomena. In particular, for the Piacenza case study, the main urban settlements of the area have been acquired in order to produce a detailed figure of the principal mobility patterns. For Lecce, some coastal areas have been isolated in order to analyse the seasonal patterns of presences in contexts characterized by a huge amount of second homes and unauthorised constructions.

2 Standard SIM cards are mainly used in consumer devices for human communication (voice, text or data).

3 AGCOM is the national Authority for Communications Guarantees and is the regulator and competition authority for the communication industries in Italy. AGCOM publishes quarterly the Communication Markets Monitoring System which provides, among other information, national data on mobile subscribers and mobile market shares

4 Mariotti, I., Giavarini, V., Rossi, F. & Akhavan, M. (2022). Exploring the "15-Minute City" and near working in Milan using mobile phone data. Tema. Journal of Land Use, Mobility and Environment

5 Curci, F., Kercuku, A., Zanfi, F. & Novak, C. (2022). Permanent and Seasonal Human Presence in the Coastal Settlements of Lecce. An Analysis Using Mobile Phone Tracking Data. Tema. Journal of Land Use, Mobility and Environment

6 Lanza, G., Pucci, P., Vendemmia, B. & Carboni, L. (2022). Impacts of the Covid 19 outbreak on two Apennine valleys. Remote-working and near-home tourism through mobile phone data. Tema. Journal of Land Use, Mobility and Environment

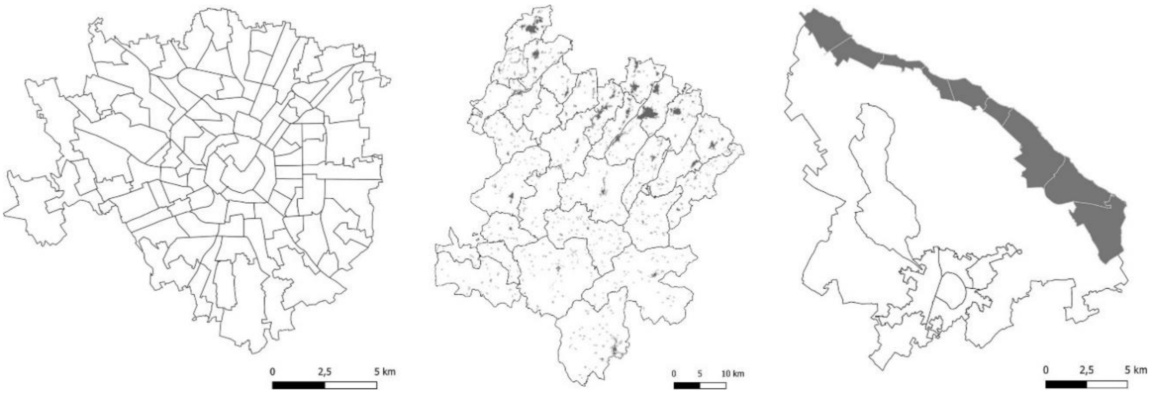

Fig.1 Maps of the three active scenarios (Milano - left, Piacenza - centre, Lecce - right). In grey colour are depicted the custom areas (urban settlements in Piacenza and coastal areas in Lecce)

2.1 Human Presence

The service provides presence data as an estimation of the number of people detected in the geographical area of interest (ACE basis), with socio-demographic details (gender, type of contract, age classes, nationality of the users, clusters category).

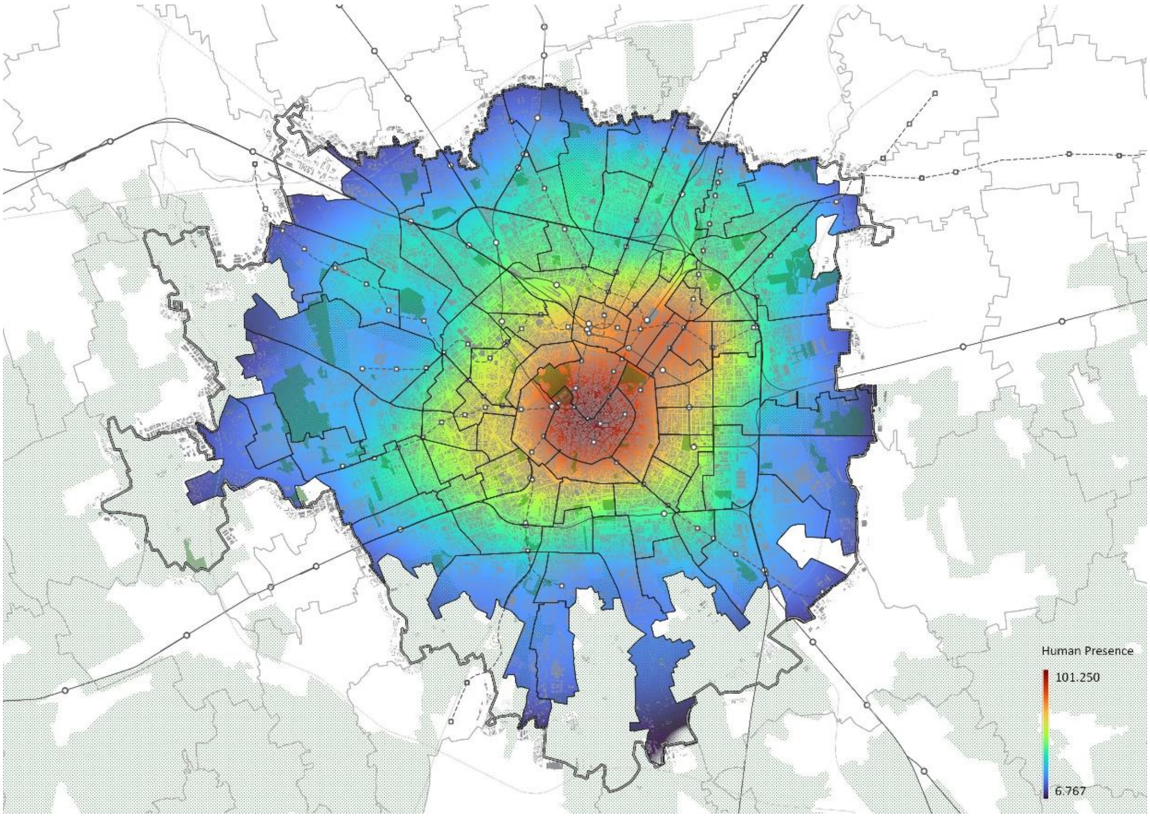

Fig.2 Presence map of Milan (date/time) available from DVI Insight

Fig.2 for example depicts the average number of overall human presence within a typical working day in April 2019 in Milano. The map shows through a heat map representation the amount of population calculated according to TIM market share, in every Census Area It is possible to observe the Milano districts which attract more population (the city centre, the area near the Milano Central station but also some suburban neighbourhoods in the south western side of the city where the overall number of presence is very high.

The type of contract can be business or customer and reflects the use of a mobile phone for personal or for business activity, according to the mobile plan subscriptions. Typically, a business contract is subscripted by companies for their employees who use corporate phones for job purposes.

The class ages (< 18 years old, 18-30 years old, 31-40 years old, 41-50 years old, 51-60 years old, and over 60 years old) together with the other variables, are directly extracted from the registry of TIM clients. The data for the youngest class is certainly underestimated compared to the real phenomenon because parents are used to purchasing a sim card and giving it to their children. Obviously, in these cases, the demographic data provides information on the owner of the sim and not on who is using it at that moment. For the other age groups, this issue is very small if not non-existent.

Field name	Field description
id	Unique ID counter
cluster	Description of TIM data
data_da	Beginning date and hour of the interval for which the data is provided
data_a	Ending date and hour of the interval for which the data is provided
numero_presenze	The number of people counted
layer_id	Spatial data code (Region + Province + Municipality + ACE)
layer_nome	Spatial data code
dettaglio(secondi)	Time interval (second)

Tab.1 Record layout for the human presences dataset

To define the presence data clusters category, it is necessary to define 2 types of TIM user ACE: Living ACE and Working ACE.

The Living ACE is derived from a set of events made by the user in the last 30 working days (Monday, Tuesday, Wednesday, and Thursday from 00:00 to 06:00 and from 22:00 to 24:00; Friday from 00:00 to 06:00). The 'residence' cell is calculated from the events during these periods. The idea beyond this Living ACE definition is that the residence is the ACE where a user spent most of the time in the last 30 days during the night.

The working ACE is derived from a set of events carried out by the user during the last 30 working days from 9:00 to 12:00 and from 14:00 to 18:00. The 'work' cell is calculated from the events during these periods. The idea beyond this ACE Work definition is that the work area is the ACE where a user spent most of the time in the last 30 days during the classical working hours.

These two definitions are directly provided by TIM and can lead to some misunderstandings concerning some categories of users who have different behaviours in terms of working hours or mobility practices.

Given the above definitions, it is possible to define the classes of travellers as follows:

- residents: counting of the Italian users of the TIM network, for each pixel of the map, who are (at the time selected in the timeline) in their ACE of residence;
- commuters: counting of the Italian users of TIM network, for each pixel of the map, who are (at the time selected in a timeline) in their ACE of work;
- intra-regional visitors: counting of the Italian users of TIM network, for each pixel of the map, who are (at the time selected in a timeline) outside their ACE of work or residence but in the region of residence;
- extra-regional visitors: counting of the Italian users of TIM network, for each pixel of the map, who are (at the time selected in a timeline) outside their ACE of work or residence and the region of residence;
- foreign visitors: counting of users with a foreign SIM card detected on the TIM network.

This classification can be used for the identification of particular patterns of the different categories of users related to the time variability of the presence in the different hours of the day for job, study, or leisure reasons, days of the week, and season for holidays and festivities.

For foreigners, the Presence dataset provides information on the number of users with a foreign SIM connected to the TIM network disaggregated by single nationality. This information therefore probably covers mostly tourists or temporary visitors and not immigrants living in Italy who have a TIM sim.

In research based on the use of mobile phone data, is worth noting that the technological constraints associated with how data is acquired and delivered are significant and greatly determine the development of the research itself.

This highlights a new role for researchers who must define research questions and methodologies that are strongly conditioned by the characteristics of the data source.

For all the three case studies, the human Presence dataset was available.

Fields	Categories
Gender	Male, Female
Type of contract	Business, Consumer
Class age	< 18 years, 18-30 years, 31-40 years, 41-50 years, 51-60 years, > 60 years
Traveller class	Intra-regional traveller, Extra-regional traveller, Residents, Commuters, Foreign visitors
Nationality	Italians, Foreigners (according to SIM)

Tab.2 Filters available in the human presences Scenarios

2.2 Mobility

The DVI platform also provides mobility data defined as the estimate of trips in the area of interest on an ACE basis, with details of where people come from or where they go.

The destination data is the analysis of movements from the territory of interest on an ACE basis. Displacement is considered to be detected when two consecutive events occur in the same cell.

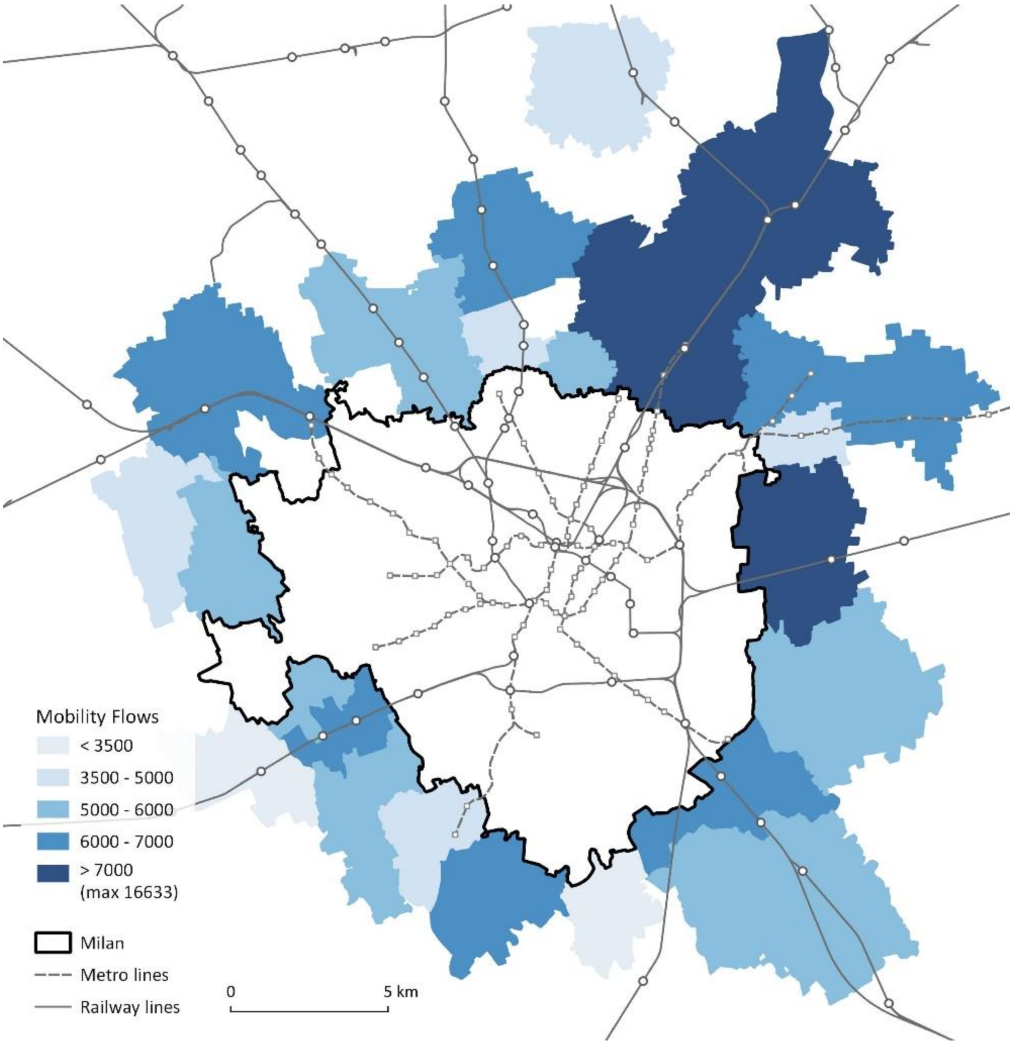

Fig.3 Mobility flow to Milano (3/4/2019)

In Fig.3 for example the overall amount of movement between some municipalities around Milano and the main city in a specific working day in April 2019 (3/4/2019) is depicted showing the mobility relationships at the metropolitan scale as detected from mobile phone data.

The origin data is the analysis of movements toward the territory of interest on an ACE basis, where the area of origin is based on the location of the users in the previous night. Provenances are a subset of Presences.

The tracking of mobile phone users' movements within the network is made possible by the passive location platform (PLS) that allows the estimation of the users' geographical position within an area whenever there is an interaction with the network. PLS keeps only the last available position and the instant when the update occurred.

The interactions with the network that are used by PLS define an event and can be:

- phone power on/off;
- return to coverage;
- call initiation (made and received);
- SMS (sent or received);
- location Update (periodic and non-periodic update app).

Mobility data is commonly represented by Origin-Destination data which is the most widely used tool to analyze and map mobility demand.

An Origin-Destination matrix represents the movements that affect the study area in a given period, subdivided by the zone of origin and destination. For each time interval, the trips between each pair of zones are available. The data on the movements of individual users are processed during the day of analysis and are therefore aggregated and provided in an anonymous form the following day.

The positions of the users are associated with the zones and, for each user, a movement is derived from the Origin Zone (which contains the cell used by the user in the time interval in which the origin is calculated) to the Destination Zone (which contains the cell used by the user in the time interval in which the destination is calculated).

Long-range movement is defined as a movement from location L1 to location L3 via Ln. In case the SIM does not generate events in the Ln location, only the movement from L1 to L3 will be recorded.

Further consideration of mobility data provided by TIM concerns the degree of complexity of the phenomena that can be detected.

If we take into account that mobility data is available at the ACE spatial aggregation both for Origin and Destination, we can easily conclude that the level of detail of the information is very high since it permits the detection of hourly Origin-Destination matrices at the ACE scale. This allows to analyse and map spatial relationships between a relatively small portion of territory with a temporal detail that cannot be achieved with other sources. It is possible to analyse mobility practices at night, at weekends, or at different times of the year. Mobility data is available only for the Milan scenario.

3. Elaboration of TIM data: from raw to usable data

The DVI Insight platform is useful for visualizing data but has proved to be ineffective for data extraction and elaboration. For this reason, we have obtained the raw data, which allows greater flexibility of use and ensures greater integration with the spatial analysis tools at our disposal. The complexity and the amount of raw data made it necessary to develop some preliminary elaborations to transform, clean, and finally, use the human Presences data. The following considerations and methodologies refer specifically to Milano which is the most complex and articulated case but can also be applied in general to other scenarios as they refer to the same types of data.

3.1 Pre-processing and preliminary elaborations

For the processing described below, in particular the preliminary cleaning and rearranging, given the amount of data, its organization, and type, Python programming language is used, and in particular the Pandas libraries (version 1.3.2) for the data frame management. Pandas is an open-source library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

Python is one of the most widely used languages for Data Analysis and Data Science.

The TIM data on human Presences in the municipality of Milan are also provided in CSV format, each file containing data for an entire week. Each file contains the following fields: cluster; date_from; date_to; attendance_number; layer_id; layer_name; detail (seconds).

The first operation performed by a script developed specifically for the project is to open the compressed CSV files and merge them into a single data frame containing all the human Presences data. In this way, the data for the entire period can be analyzed together:

- the year 2019: March, April, May | September, October, November;
- the year 2020: March, April, May, June, July, August, September, October, November, December.

The fields are also reorganized, keeping only those that are essential for further analysis and the result is as follows:

For the available period, the TIM data of Milan Presences consists of 18,971,793 records. The "cluster" field contains all TIM categories (Presences, Males, Females, Business, Consumer, Italians, Foreigners, <18 years, 18-30 years, 31-40 years, 41-50 years, 51-60 years, >60 years, Intra-regional Visitors, Extraregional Visitors, Commuters, Residents, Foreign Visitors) - (Total Attendances, Gender, Age Classes, Traveller Classes). If we consider only total human Presences, the number of records is reduced to 1,077,412. This number refers to the ACE of Milan with an hourly time interval between records. The available hours are 11,711 (almost 490 days analyzed).

Another preliminary operation to be performed is the conversion of the date-time format that TIM provides in UTC, into the local date and time format of Milan (Europe/Rome). The UTC format, which corresponds to Greenwich, differs by one or two hours from Milan (summertime and solar time) depending on the period of the year selected. To analyze presences with a time interval, it is therefore essential to convert the UTC format to the correct timezone (UTC+1). Another Pandas Python script has been used to perform this complex operation.

Once the correct time for Milan has been obtained, other information regarding the time range is added, which will be useful in the following steps of the analysis and which will facilitate further extractions and elaboration:

- day_name: Name of the day of the week;
- day_num: Number of the day of the week (Monday=0; Tuesday =1; Wednesday=2; Thursday=3; Friday=4; Saturday=5; Sunday=6);
- hour: Time of the day;
- date: Date without time reference in format yy:mm:hh;
- n_day: Progressive number identifying in ascending order the day within the period (first day = 0; second = 1; third = 2; etc);
- week: Number of the week within the year;
- n_week: Progressive number identifying in ascending order the week within the period.

This reorganization of the temporal aspect of the data greatly increases the amount of elaboration and complexity that can be carried out. For example, it is possible to compare the behaviours of the same day during the whole period or to calculate the average number of people per day or compare different weeks between them.

Finally, the presence of TIM users is recalibrated to obtain a reliable projection of the total presence of people (real number) according to the Tim market share in the municipality of Milan.
The final Milan database contains the estimated overall number of people on an hourly basis for the ACE for the whole period.

3.2 Baseline calculation

To facilitate the comparison between different periods we decided to build a reference trend for the presences in Milan, named baseline.
By baseline we mean the value we expect to find in a certain area, day, and time, in a 'standard' week, i.e. a week without any significant events that can be used to highlight specific behaviours or out-of-average trends during other periods. For the calculation of the baseline, we have used a methodology developed by Maas et colleagues for the calculation of presence data in Facebook 'disaster maps' (Maas et al., 2019).
The big difference is that Facebook calculates the baseline in the period before a critical event (for example a flooding event or an earthquake or the starting period of the pandemic). In our case, since we have data for the whole year, the baseline was calculated by selecting "standard" weeks, those with normal values (no major events, holidays, working week, open schools).
It is necessary to select the weeks that have the chosen characteristics for the baseline calculation.
By selecting standard weeks, it is possible to compare the attendance of any other period of the year against the baseline, at events, holidays, holiday periods, or periods where covid 19 measures were active to detect spatial and temporal patterns that diverge from the mean.
The values of the standard week (baseline) are calculated for each geographical area (ACE), for each time interval (hour) of each day of the week, and possibly for each TIM category.
Therefore, for a given location and time interval, the baseline dataset is composed of a set of counts from the same location over the same time interval on the same day of the week for multiple weeks preceding the crisis.
Once the period is selected, a dataset that contains only the data of the selected period is extracted. In our case, 6 weeks of 2019 are considered: the first three weeks of March, the second week of May, and the third and fourth weeks of October. i.e. working weeks with no major holidays, schools open.
The procedure for the construction of the baseline is the following:

- extraction of the baseline Dataset for each ACE, for each time interval, and possibly TIM category;
- calculation of the mean and the standard deviation for each Dataset;
- Winsorisation procedure by excluding extreme values beyond the determined thresholds: 'lower bound' and 'upper bound' (2.5 and 97.5 percentile) using a normal distribution (+ and - 2std from the mean = 95% values).

The winsorization is a process through which it is possible to eliminate extreme values by computing the mean and standard deviation of the pre-winsorization distribution, identifying the 2.5th and 97.5th percentiles of a Gaussian with that mean and standard deviation, and setting values outside those bounds to the lower and upper bound values if they are anomalously low or high, respectively.
The final baseline acts as a reference for the further analysis carried out to the Milano case as will be shown in the following paper.

4. Integration with conventional data

One of the findings of previous research based on the analysis of telephone traffic data is that integration with conventional data is needed to enhance both the capacity to interpret phenomena emerging from user-generated data and to build a better understanding of urban usages, in time and spaces (Pucci et al., 2015).

There is an ongoing debate over the potential of mobile phone data for official statistics (Daas et al., 2015; Struijs et al., 2014; Barcaroli et al. 2014). Some general issues are related to the nature of the data itself or the differences in terms of the dimension of the samples because conventional data refer to the universe of the population while mobile phone data refer to a subsample of the overall mobile phone subscriptions (i.e. active users). Mobile phone data are provided by private companies and in general, there are no common definitions about the format and the methodology used for data collection and there is no clarity on their costs. Moreover, some privacy issues must be taken into consideration according to national laws (De Montjoye et al., 2019).

For all these reasons we decided to integrate our mobile phone dataset with some other ancillary data acquired from conventional data sources and to provide researchers with some additional information useful for improving the interpretation of mobile phone data patterns both in the spatial and in the temporal dimension. Concerning the spatial dimension, we organized a database of variables and indicators on socio-demographic issues at the ACE level based on conventional or spatial data sources that can be easily integrated with mobile phone data. In particular, we covered the main socio-demographic issues (overall number of inhabitants, foreigners, class ages) from the Milano official registry and the main big attractors available through multiple spatial data sources available aggregated at the ACE level (distribution of schools, economic activities, university buildings, main residential areas, public parks, etc) that can help in determining the causes of variations in the presence of people during the different hours of the day.

Concerning the temporal dimension, we developed a calendar of events for Milano and official regulations about the different restrictions taken during the pandemic for the interpretation of mobile phone trends which are strongly influenced by these external factors. In the calendar is possible to distinguish weekend vs working days, holidays and festivities, big events occurring in the city, lockdown measures.

These indications pertain to institutional, recreational, and socio-cultural activities of particular relevance with differentiations for dates of national interest and those of local level, but also in general related to the calendar with the identification of holidays and weekdays.

In particular, since the period examined partially coincided with the emergence of the pandemic and subsequent containment measures, we identified and took into consideration the different dates when governmental limitations of activities and movements of the population have been active. These phenomena were some of the major factors of change in the use of the city and greatly impacted the behaviours of people in terms of daily, weekly and seasonal mobility.

Date	Activity
1 March 2020	Lockdown starts for some regions
4 March 2020	Schools and universities are closed
10 March 2020	Lockdown starts for Italy
21 March 2020	Prohibition of all travel
10 April 2020	Extension of lockdown until May 3rd
3 May 2020	End of Lockdown
4 May 2020	Reopening of activities
17 May 2020	Transfers between regions are permitted
12 June 2020	Reopening of the event without the public
15 June 2020	Reopening of cinemas and theatres, summer centres
29 July 2020	Extension of state of emergency

Tab.3 Extraction from the calendar of events database

Another information available concerns the school system calendar with the identification of the dates of the beginning and end of the school year and holidays for the various Italian regions. Attention was paid to some territorial differences, such as local festivities, to be able to monitor and adequately define the presences within the different territories.

Being Milano a large municipality with many events taking place there, it was necessary to make a selection of those with a significant impact in terms of the number of people involved. The dates of the Salone del Mobile fair, the Fashion Weeks, and the main football events at the San Siro stadium have been collected and identified.

All the information has been reported in a spreadsheet that can be easily consulted to provide consistent help for the present and future analyses based on mobile phone data.

The complexity of the telephone data requires the use of supporting information that can be integrated into the research design for an appropriate elaboration and understanding of the dynamics highlighted by it.

5. Tools and competences

Working with mobile phone data means processing a huge amount of data in the form of hundreds of tables with spatial and temporal information. To better exploit the potential of this source new tools are needed for data storage, elaboration, and visualization. In our workflow, we acquired competencies and developed integrated methodological frameworks based on Python scripting, Database Management System, Geographical Information, System, and visualization tools. Due to their complexity, an intersection between different competencies is needed to fully benefit from the opportunities offered by this kind of data. Computation and quantitative analytical skills are therefore required, together with the capacity to read temporal and spatial dynamics deriving from the analysis of big data.

References

AGCOM (2022) Osservatorio sulle comunicazioni n. 4/2021

Ahas, R., Aasa, A., Roose, A., Mark, Ü., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, *29* (3), 469-486. https://doi.org/10.1016/j.tourman.2007.05.014

Barcaroli G, De Francisci S, Scannapieco M, Summa D (2014) Dealing with big data for official statistics: IT issues*. Meeting on the management of statistical information systems*, Dublin, Ireland and Manila, Philippines, 14–16

Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ data science*, *4* (1), 10. https://doi.org/10.1140/epjds/s13688-015-0046-0

Curci, F., Lanza, G. & Manfredini, F. (2022). Mobile phone traffic data for territorial research. Opportunities and challenges for urban sensing and territorial fragilities analysis. *Tema. Journal of Land Use, Mobility and Environment.* http://10.6092/1970-9870/8892

Curci, F., Kercuku, A., Zanfi, F. & Novak, C. (2022). Permanent and Seasonal Human Presence in the Coastal Settlements of Lecce. An Analysis Using Mobile Phone Tracking Data. *Tema. Journal of Land Use, Mobility and Environment.* http://10.6092/1970-9870/8914

Daas, P. J., Puts, M. J., Buelens, B., & van den Hurk, P. A. (2015). Big data as a source for official statistics. *Journal of Official Statistics*, *31* (2), 249. http://dx.doi.org/10.1515/JOS-2015-0016

De Montjoye, Y. A., Gambs, S., Blondel, V., Canright, G., De Cordes, N., Deletaille, S., ... & Bengtsson, L. (2018). On the privacy-conscientious use of mobile phone data. *Scientific data*, *5* (1), 1-6. https://doi.org/10.1038/sdata.2018.286

Dobra, A., Williams, N. E., & Eagle, N. (2015). Spatiotemporal detection of unusual human population behavior using mobile phone data. *PloS one*, *10* (3), e0120449. https://doi.org/10.1371/journal.pone.0120449

Gething, P. W., & Tatem, A. J. (2011). Can mobile phone data improve emergency response to natural disasters?. *PLoS medicine*, *8* (8), e1001085. https://doi.org/10.1371/journal.pmed.1001085

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *nature*, *453* (7196), 779-782. https://doi.org/10.1038/nature06958

Lanza, G., Pucci, P., Vendemmia, B. & Carboni, L. (2022). Impacts of the Covid 19 outbreak on two Apennine valleys. Remote-working and near-home tourism through mobile phone data. *Tema. Journal of Land Use, Mobility and Environment.* http://10.6092/1970-9870/8915

Louail, T., Lenormand, M., Cantu Ros, O. G., Picornell, M., Herranz, R., Frias-Martinez, E., ... & Barthelemy, M. (2014). From mobile phone data to the spatial structure of cities. *Scientific reports*, *4* (1), 1-12. https://doi.org/10.1038/srep05276

Maas, P., Iyer, S., Gros, A., Park, W., McGorman, L., Nayak, C., & Dow, P. A. (2019, May). Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery. In KDD (Vol. 19, p. 3173). Shmelev, S.E. (2019). *Sustainable cities reimagined: multidimensional assessment and smart solutions*. New York: Routledge.

Manfredini, F., Pucci, P., & Tagliolato, P. (2013). Mobile phone network data: new sources for urban studies?. In *Geographic information analysis for sustainable development and economic planning: New technologies*, 115-128. IGI Global. https://doi.org/10.4018/978-1-4666-1924-1.ch008

Manfredini, F., Tagliolato, P., & Rosa, C. D. (2011, June). Monitoring temporary populations through cellular core network data. In *International Conference on Computational Science and Its Applications*, 151-161. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21887-3_12

Mariotti, I., Giavarini, V., Rossi, F. & Akhavan, M. (2022). Exploring the "15-Minute City" and near working in Milan using mobile phone data. *Tema. Journal of Land Use, Mobility and Environment.* http://10.6092/1970-9870/9309

Pucci, P., Manfredini, F., & Tagliolato, P. (2015). *Mapping urban practices through mobile phone data, 3* (4). Heidelberg, New York, Dordrecht, London: Springer. https://doi.org/10.1007/978-3-319-14833-5

Soto, V., & Frías-Martínez, E. (2011, June). Automated land use identification using cell-phone records. In *Proceedings of the 3rd ACM international workshop on MobiArch*, 17-22. https://doi.org/10.1145/2000172.2000179

Steenbruggen, J., Tranos, E., & Nijkamp, P. (2015). Data from mobile phone operators: A tool for smarter cities?. *Telecommunications Policy*, *39* (3-4), 335-346. https://doi.org/10.1016/j.telpol.2014.04.001

Struijs, P., Braaksma, B., & Daas, P. J. (2014). Official statistics and big data. *Big Data & Society*, *1* (1), 2053951714538417. https://doi.org/10.1177/2053951714538417

Toole, J. L., Ulm, M., González, M. C., & Bauer, D. (2012, August). Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD international workshop on urban computing*, 1-8. https://doi.org/10.1145/2346496.2346498

Vanhoof, M., Reis, F., Ploetz, T., & Smoreda, Z. (2018). Assessing the quality of home detection from mobile phone data for official statistics. *arXiv preprint arXiv:1809.07567*. https://doi.org/10.48550/arXiv.1809.07567

Wang, Z., He, S. Y., & Leung, Y. (2018). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, *11*, 141-155. https://doi.org/10.1016/j.tbs.2017.02.005

Image Sources

Fig.1: Authors' elaboration;

Fig.2: Authors' elaboration;

Fig.3: Authors' elaboration.

Table Sources

Tab.1: Authors' elaboration;

Tab.2: Authors' elaboration;

Tab.3: Authors' elaboration.

Author's profile

Fabio Manfredini

He is the responsible of the "Mapping and Urban Data Lab" (MAUD), Department of Architecture and Urban Studies, Politecnico di Milano. His main areas of expertise are methods and techniques of territorial and environmental analysis, geographical information systems, statistical and spatial analysis, mapping and data visualization. In the last years, he specialized in the use of novel data sources (mobile phone and social media data) for urban studies and for mobility mapping.

Carmelo Di Rosa

Senior technician of the "Mapping and Urban Data Lab" (MAUD), Department of Architecture and Urban Studies, Politecnico di Milano. His main areas of expertise are the design and the management of relational databases for the construction of mobility indicators for urban studies. VBA developer of user interfaces, expert in GIS and statistical analysis, he has collaborated in the publication of articles in national and international journals.

Francesco Fagiani

He is a research fellow at Politecnico di Milano, Department of Architecture and Urban Studies. He is graduated in Urban Planning and Policy Design, his primary expertise includes data analysis, with particular regard to spatial data and GIS. His recent work has focused on new data visualisation techniques, dynamic visualization and dashboards.

Viviana Giavarini

She has a degree in Architecture at the Politecnico di Milano. She works in the "Mapping and Urban Data Lab" (MAUD) of the Department of Architecture and Urban Studies, Politecnico di Milano where she develops data analysis and mapping activities to support researches aimed at analyzing urban and social transformations.