OVERVIEW OF THE SOURCES AND CHALLENGES OF MOBILE POSITIONING DATA FOR STATISTICS

ABSTRACT

The current paper gives an overview of the technical and methodological challenges of the use of location data from mobile devices for statistical purposes. Big Data is playing an important role in today's world of understanding the more and more social processes. Mobile data is considered as one of the most promising ICT (Information and Communication Technologies) data sources for measuring the mobility of people. Almost every person on the planet now has a mobile phone which he/she uses to communicate, access internet, conduct everyday banking and entertain. The digital footprint left by the users is very sensitive, but also highly valuable, as it provides new possibilities to measure and monitor the spatio-temporal activities of the population. For statistical purposes, such data provides new possibilities in terms of quality of the data as well as new opportunities. Statistics based on Big Data can be compiled automatically, in some cases almost in real time and requires less manual labour. Obviously the job of analysing and interpretation of the statistical indicators is left for statisticians and researchers, but the new concept of fast and expansive data collection should improve the quality of decision making process and results in public and private sectors.

In this paper, a technical overview of the sources for mobile positioning data is given with the main focus on the data stored by Mobile Network Operators. A short introduction of possible statistical applications explains some of the basic domains where the mobile positioning data can be and is used. As this is a new data source, there are a lot of challenges that have to be taken into account in order to be able to produce valuable and qualitative statistics. The main challenge is the access to the data where legal, administrative and commercial barriers have to be overcome. Obviously the most important question is the privacy protection aspect of accessing and processing the data. An overview of technological and methodological challenges is also provided in this paper as there are many general and domain-specific aspects that require attention.

Table of Contents

1 INTRODUCTION

This document provides an overview of technical and methodological aspects of accessing mobile positioning data from Mobile Network Operators (MNO) and other sources for location data of mobile devices. Mobile positioning data provides the possibility to analyse and understand spatio-temporal behaviour of humans in very different ways. However there is no single source for such data as the data is collected by many parties who usually are business corporations with no natural interest to provide the data about their users. As the location data of people can be considered as private data, the biggest problems are usually associated with getting access to the data and using legal ways to process the data. There are several regulations controlling what can and cannot be done with such data and overcoming these obstacles might require a lot of time and will to do so by all parties. The owners of the data are corporations who are business-oriented and do not easily provide the data for free, unless there are economic prospects for them in doing so.

MNOs are the main source for the mobile positioning data. Although the technical standards for MNOs are often the same, there are very different types and forms of location data that is stored. This makes it difficult to provide a specific technical description of what kind of data is available. Mobile apps developers also store the location data of their customers, but there is even more confusion as each developer has its own technical infrastructure which is developed for the needs of the specific app and not for the needs of statistical communities.

Even if accessing the data is the biggest challenge, processing the data is also not without its own pitfalls. The technology for processing large amounts of data can be very demanding. The methodological aspects are perhaps the most important questions for statisticians and there are both general and domain-specific methodological challenges to deal with.

Still, the possibilities of using mobile data can be vast as this data introduces new quality and often unprecedented new indicators in tourism, transportation, population statistics, and other domains, where the data can be used.

2 DESCRIPTION OF MOBILE NETWORK OPERATOR BASIC INFRASTRUCTURE

Almost all Mobile Network Operators (MNO) in the world, with very few exceptions,[1] use two main mobile technologies – GSM (Global System for Mobiles) and CDMA (Code Division Multiple Access). The market share of subscribers using CDMA worldwide is 15-25% (mostly North-American MNOs, some Asian MNOs) and 75-85% for GSM (the rest of the world) (EngineersGarage 2012).

The main difference between these systems lies in the core radio signalling technology, wherein CDMA uses spread-spectrum technology, which allows each subscriber to transmit over the entire frequency spectrum of the antenna all of the time, while GSM operates on the wedge spectrum wherein each subscriber is assigned a time slot which no other subscriber can have

[1] Mostly hybrid-system exceptions.

access[2]. The practical difference between CDMA and GSM is that within CDMA network, a mobile device is tied to a particular network, while within GSM, a SIM (Subscriber Identity Module) card is tied to a particular network. It is therefore easier to switch mobile devices within GSM networks by simply changing the SIM card. CDMA phones without GSM support cannot operate within GSM network and vice versa. This fact makes it more difficult to use roaming services while travelling in different countries.

It is expected that the two systems will eventually converge into a common global standard that has already started with the spread of LTE (Long Term Evolution) technology adopted by MNOs using both CDMA and GSM systems. However MNOs have expressed the opinion that the old generation networks will be upheld until at least until 2020 because of the number of subscribers using the old system phones (PCMag 2013).

With respect to the positioning data obtained, the CDMA and GSM systems will show some differences, however the more important differences are caused by the infrastructure of the specific MNOs. Regardless of the technological differences between specific systems, all network infrastructures have similar basic components (Figure 1).

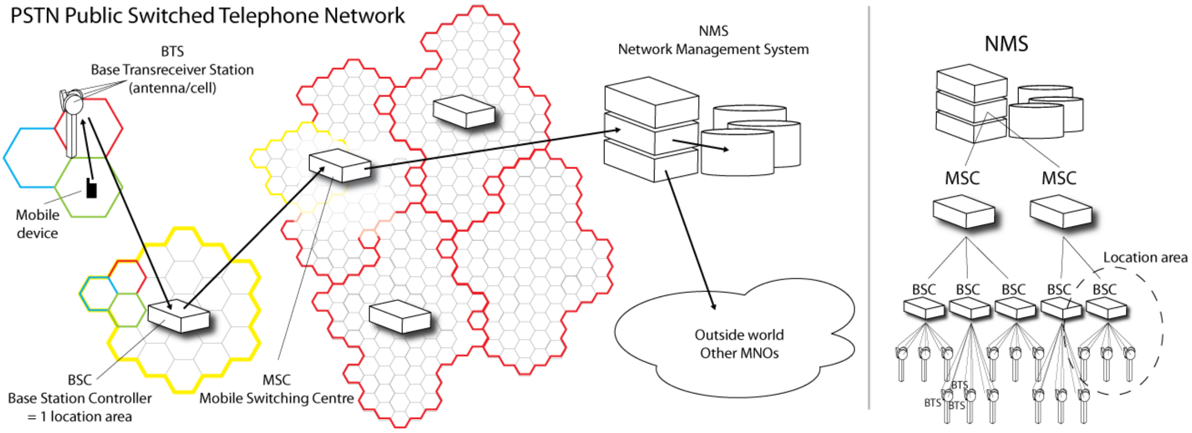

Figure 1. Simplified illustration and hierarchy of MNO's network components.

The smallest element in the network is the network antenna – BTS (Base Transreceiver Station) – to which the mobile devices are connected when they are subscribed to the network. The BTS is responsible for radio communications between the network and the mobile devices. It handles speech encoding, encryption, internet traffic and other radio signal specific tasks. There are different BTSs for different frequencies and protocols.

A number of BTSs in a proximate geographical area are controlled by a BSC (Base Station Controller). A set of antennae controlled by one BSC comprises a single Location Area (LA)[3]. If the phone is within a single Location Area, the BSC is responsible for handover procedures between one BTS to another. BSCs are controlled by MSCs (Mobile Switching Centres) which can be one or several depending on the size of the MNO[4]. MSC also accommodates the VLR (Visitor

[2] The CDMA network operates in the frequency spectrum of 850 MHz and 1900 MHz, the GSM network operates in the frequency spectrum of 850 MHz, 900 MHz, 1800 MHz, and 1900 MHz.
[3] The BSC handles allocation of radio channels, frequency, power, and signal measurements.
[4] The MSC handles call and messaging routing, handover between different Location Areas and MSCs, updating different registries within MSC and the central system.

Location Register) – the registry for holding the information about the Location Area and the BTS to which the mobile devices are connected.

MSCs report to the central system of the MNOs – NMS (Network Management System) where all administrative and central procedures, including data storage warehouses and network handling servers reside. Usually only data that is transmitted from MSCs to NMS is stored for different purposes and lower level data traffic is deleted. NMS accommodates different registries and databases that are important from the point of view of the location data e.g. OSS (Operation and Support Subsystem where the geographical data of the network is stored), data warehouses, billing databases, CRM (Customer Relationship Management), HLR (Home Location Register) and other components that are important from the aspect of location data. Components of MPS (Mobile Positioning System), one of the important elements for pinpointing subscribers' location for emergency and security services, also reside in NMS and MSCs.

Different MNOs may share some of the network equipment or network modules. For example it is not uncommon that MNOs share the antennae. There are also special types of MNOs (virtual MNOs) that do not possess any network infrastructure, but instead rent it from MNOs that have the network. In such case usually the data produced (e.g. CDRs) is transferred to both MNOs' central systems; however the virtual MNOs do not have the access to the some operational data.

CDMA and GSM are the main technological platforms in the mobile world. However there are also different standards and protocols including 2G, 3G, 4G, LTE, etc. Depending on these standards, the location data characteristics and availability may vary[5].

3 Sources and Forms of Location Data

Mobile positioning refers to the technical ability to locate mobile devices in time and space with certain accuracy using the mobile network infrastructure or other technology that includes location information of the mobile device. Mobile positioning data is the data collected about the location and movement of mobile devices. The location of the mobile devices can be obtained in real-time or historically. The owner of the located devices can be known or unknown. There are two main collection methods for obtaining the mobile positioning data:

a) Active positioning;
b) Passive positioning.

The difference between active and passive data collection methods is that for active positioning, a specific targeted request is made to locate the mobile device (request for location is made, location response is returned), while for passive positioning, historical data is collected and no active requests are sent. Active and passive data can both be monitored in real-time and historically, although in active positioning location requests have to be made regularly over a period of time to get the historical data.

3.1 Active Positioning Method

[5] The names and specifics of the components may vary in different infrastructure as this depends on the technical glossary of the specific technology vendor, size of the MNO and many other aspects. However most of the MNOs have a standard technical setup.

Active positioning method location request can be initiated from the mobile device itself (by its user) or from outside. In the latter case, the consent of the owner of the device is generally needed (for single location or for longer period) due to privacy reasons, except when the request is made by emergency services (e.g., E112 and E911 location enabling directives in EU and US) or security enforcement (mostly approved by court order), in which consent is neither sought nor needed.

Requesting active positioning usually requires authorisation and approval from the mobile device owner[6] unless it is done for requesting location of the mobile devices in case of emergency (e.g. E112 and E911 location enabling directives in EU and US). Another field where active positioning is used are so called “friend finder” or “find nearest” type of services where phone owner uses active positioning methods to find the location of friends/family members or to find places of interest in one’s proximity using the location of one’s mobile device. Location Based Services (LBS) including Location Based Advertisement (LBA), also use active positioning methods (Ericsson 2013). Active positioning refers to real-time or near-real time positioning of mobile devices. Moreover some services use periodical active pinging to store the location of the devices over a period of time (e.g. fleet management applications for tracking the location and routes of vehicles).

There are two basic methods for locating the phone using active positioning:

a) Using MNOs’ network infrastructure and MPS;
b) Using real-time location data from applications (“apps”) installed in the mobile devices.

3.1.1 Network Infrastructure-based Active Positioning

Mobile Positioning System is a common name for active positioning technical system employed by MNOs to introduce the ability to pinpoint the location of the mobile phones using network infrastructure. MPS uses different available methods to locate the mobile device: Cell Global Identity (CGI), trilateration of antennae, angle of received signal, and Timing Advance (TA - arrival time of radio signal from antenna to device and back). Most commonly MPS uses CGI+TA. As network cells can be relatively large (ranging between 0,002 km^2 to 1200 km^2), the exact location in the coverage area of the antenna is determined by timing advance and the angle of a radio signal. The practical precision of CGI+TA falls within the range of 50-2500 m. The precision is ultimately defined by the density of the network. If the mobile device has a GPS chip and Assisted-GPS (A-GPS) capability, then MPS is able to locate the phone with GPS accuracy (Figure 2).

[6] Upon the request for an active positioning of the mobile device, the device owner has to give the consent for the specific requester by replying to the location request SMS, app approval message or settings, or other appropriate means.

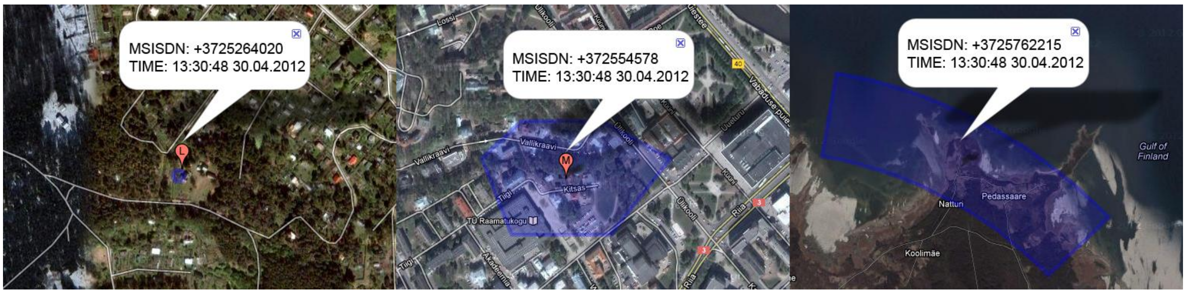

Figure 2. GPS precision (accuracy: ~10 m), 3G MPS (~200 m using trilateration) and 2G (~7 km using CGI+TA) positioning results from MPS.

Because MPS requests require network infrastructure resources, the number of devices that can be located simultaneously is limited. This can be improved by improving the technical recourses available for MPS, however this requires some investments in the infrastructure of the MNOs.

3.1.2 App-based Active Positioning

Active positioning also refers to the ability of installed application ("apps" from here onward) to use the location of the mobile device. This is used in map apps, navigation apps, travel guides, social media, location sharing, and other apps. Mobile apps use mostly three types of location methods:

a) GPS and A-GPS capability if available;
b) Wireless network (Wifi) location databases;
c) Network antenna-based location databases.

It must be noted that GPS does not work indoors and consumes mobile device's battery very fast, therefore other methods are often used to compensate or replace GPS[7].

Similarly to MPS, app-based location requests usually require permission from the phone owner and are mostly initiated by the phone owners themselves. However there are apps where locations obtained by the app are used also by outside parties – e.g. analysing travel behaviour of travel guide users, geographical marketing analysis based on the locations of the users of social media apps (e.g. Twitter, Facebook, Foursquare, etc.).

3.1.3 The Advantages and Weaknesses of the Active Positioning Method

The advantage of data obtained from this method is that the geographic information is generally very precise and accurate. Active positioning can be used by researchers for replacing or supplementing travel diaries in spatial behavioural analyses and for generating mobility statistics. Similar to surveys (in person, on-line or through telephone) active participation could be requested from the participants and sampling techniques could be used. The respondents could be selected and asked to participate by turning on the active positioning app or approve the periodical network-based location requests. The main weakness of this data source is the need to recruit the respondents resulting in a rather small sample size.

Advantages	Disadvantages
• Additional tool for spatial behaviour research and statistics;	• Small sample size because of the need to get consent from the subscribers;

[7] A-GPS is often used to fast-start the GPS as the initiation of GPS might take a long time before the actual coordinates can be calculated.

• Very small burden for respondents; • Relatively easy to set up and conduct positioning requests;	• Technological limits for the number of location requests via MNOs' network infrastructure;

3.2 Passive Positioning Method

Passive positioning relies on the location data that has been stored in the registries and databases of the MNOs or companies developing mobile apps. Usually this data is generated for other purposes than of collecting the location, therefore this data can be considered as residual data. Usually this residual data contains the information of activities of operational data of the mobile devices over time that can be linked to a location with specific geographical accuracy.

There are two basic sources of passive positioning data:

a) Data from MNOs' network infrastructure registries and databases;
b) Data from apps installed in the mobile devices.

3.2.1 Network Infrastructure-based Passive Positioning

In most cases, passive mobile positioning data is referred to the data originating from MNOs. MNOs store huge amounts of data for their business purposes including "location events" (e.g. network infrastructure quality monitoring (signalling quality), operational activity data, billing data, data for marketing and CRM, handover, Location Area updates, active positioning requests, etc.) that can be linked to geographical location. In most cases these events do not automatically have any location features but rather attributes that can be used to reference to geographic location. For example outgoing call initiation events include the antenna IDs to which the mobile devices were connected during the initiation of the call, which can be matched to an antennae registry that includes the geographical coordinates of the location and the coverage area of the antenna (Figure 3).

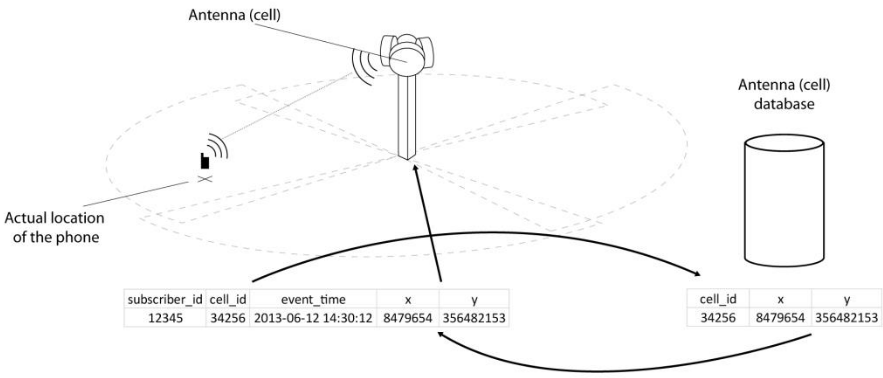

Figure 3. Illustration of the event coordinate mapping (Eurostat 2014).

Such call activity events are called Call Detail Records (CDR). CDRs include incoming and outgoing call activities, sending and receiving SMS and MMS messages. CDRs also include internet traffic between the mobile devices and the network which is often referred as DDR (Data Detail Record) or IPDR (Internet Protocol Data Records). MNOs can choose which CDRs they store depending on their business requirements (e.g. for billing purposes) and legal obligations (data retention directives requires MNOs to store all call activity information). CDRs are usually stored to internal data warehouses and billing management systems. The characteristics of the CDR data depend on which CDRs MNOs choose to store. For example if an MNO stores only outgoing call initiation CDRs, then the data might include an average of 2-3 CDRs per mobile device per day. If all abovementioned CDRs are stored, then this number might be between 200-300 CDRs per mobile device per day. Obviously the number of CDRs depends on the nature of the mobile device owner – more active phone subscribers leave more CDRs in the MNOs storage. Mobile devices with internet connection turned off or without the internet capability do not produce DDRs. Because many apps installed on smartphones actively use internet connection to exchange information even when the phone is in the pocket, a huge number of DDRs are stored for such devices.

CDRs can be stored in binary, XML (Extensible Markup Language), or plain text (CSV) format (Figure 4) in different registries and databases of MNOs. Usually CDRs include a lot of technical information about the activity recorded. However the most important elements of a single record are subscriber ID, time of the event, location reference of the event (antenna ID).


```
289345672907456;20130205120304;3725264020;3725544578;OPET-2;2G;SMS-RC;0452352561;24402
289345672907236;20130205133213;3725264020;3725544578;KESKL1;LTE;CLL-RC;29384567;24401
289345623406839;20130205204525;3725264020;3725544578;OPET-2;2G;CLL-OT;7892634590;24403
289345672903984;20130205230001;3725264020;3725544578;OPET-2;2G;IPDR;148631112;24402
```


Figure 4. Example of raw CDR file in CSV format.

Another source for location events are the Location Area updates that are done when the mobile device moves from one Location Area to another (change of BSC), and periodically by the network if the phone has been inactive for a period of time (usually 2 hours). Location Area updates are stored to HLR and the corresponding MSCs' VLR.

All abovementioned data sources are central and stored within the databases and registries of the NMS. There is a lot of data exchange within and below the MSC levels that are not transmitted to the central databases to reduce traffic throughput in the network and to reduce the data storage amounts. However this data includes different events that can be linked to the location of the subscriber and therefore is useful for generating statistics[8]. Obtaining such data often requires infrastructure investment (BTS probes, additional traffic to the NMS level, additional storage capacities) and is rather expensive.

Because of the different characteristics of the data (e.g. the temporal density and spatial accuracy), the applicability of the data might be different. For example, for the purposes of compiling national-level general statistics (not including the geographical distribution within the country), the reduced CDR data (2-3 location events per subscriber per day) might suffice as the more dense data does not improve the statistical indicators significantly and at the same time increases exponentially the need for processing resources. For high-accuracy traffic data, more dense and accurate data is required (e.g. Abis-level).

There are four main categories of data:

- Domestic data – data concerning the domestic subscribers within the country where MNO is situated;

[8] Such data includes operational high density information about radio signals and communication protocols between the BTS and BSC (so-called Abis-level traffic), inter-BSC handover, VLR registry updates, Erlang (antennae signal load), A-GPS protocols, MPS activity (active positioning), etc.

- Outbound roaming data – data concerning the domestic subscribers using the roaming service of another MNO. Usually the outbound roaming data refers to domestic subscribers using their mobile devices in foreign countries. Outbound data is received by the MNOs via Transferred Account Procedure (TAP) records from the roaming partner MNO where the domestic subscriber used the roaming service;
- Inbound roaming data – data concerning the roaming subscribers from another MNO. Usually the inbound roaming data refers to mobile devices being used domestically that originate from foreign countries;
- Antenna geographical reference data;
- Additional possible data concerning the socio-demographic profile and other characteristics of the subscribers that can be further categorized (e.g. age, gender, preferred language).

Domestic, outbound roaming, and inbound roaming data roughly correspond to the forms of tourism statistics (domestic, outbound, and inbound tourism). Although not without some reservations, the inbound roaming data represents the activity of foreigners, the outbound roaming data represents the travels of residents abroad and the domestic data represents the residents' activity within the country. An outbound roaming data from one country's MNOs is an inbound data of foreign countries' MNOs. See Annex 1 for domestic, outbound and inbound data examples.

The domestic and inbound roaming data are very similar as their origin and stored data are the same – for both forms, the events originate usually from within the MNO's own network. Outbound roaming data, as it originates from outside the MNO's network, is often limited in its characteristics to only include records for billing purposes (CDRs of call activities and sometimes internet usage) by partner MNOs (i.e. outbound data does not involve BSC-level data).

In order to transform inbound roaming and domestic data into location events, a geographical reference table should be used to assign coordinates for the events by using antennae location reference data (see Figure 3 and Annex 1). Outbound roaming data is generally stored with the reference to specific foreign country[9]. However in some cases, the antenna ID of the foreign MNO is also provided that makes it possible to obtain more precise location of the events within foreign countries.

MNOs might also hold additional data concerning for example monetary transactions that are made using mobile phones. For example in African countries, mobile phones are used to make transactions because the mobile network is often the only available mean for everyday banking. Near Field Communication (NFC) is also a new method to make payments using the mobile phones. Linking the mobile payment systems to credit cards is also in the development stage and the combination of mobile positioning methods with monetary transactions is a promising field with potential to generate new valuable statistics. The data for different transactions (sender, receiver, amount, etc.), if used via mobile phone paying system, is usually stored by the MNO. The applicability of such data is yet to be investigated.

3.2.2 App-based Passive Positioning

[9] Actually the reference is to the specific foreign MNO from what the foreign country can be deduced.

App-based passive positioning data is MNO-independent and is obtained and stored either to the mobile device or to an app provider's central databases by the apps that are installed on the mobile devices. There are thousands of apps for all types of devices (Android, IOS, Windows phone, etc.) that use and store location information. Mostly this location is used for app functionality (e.g. location sharing).

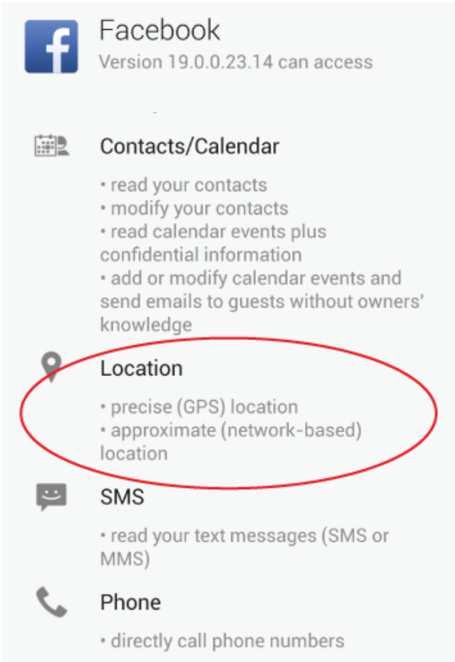

Figure 5. Example of location permissions required to be approved by the app user prior to installation of the Facebook app.

There are two main methods for apps to obtain the location of the device:

- GPS location;
- Wifi, network antennae or other network based locations (used in case GPS is not available).

The location of the data stored (e.g., an internal central database) and the specifics of the data stored are not always known and depend on a variety of reasons[10].

3.2.3 The Advantages and Weaknesses of the Passive Positioning Method

The main advantages of the passive positioning method is the cost-effectiveness of obtaining huge amounts of data involving all phone users (very large sample), and the different number of applications where such data can be used. The shortcomings lie in the problems of accessing the data including legal and business aspects. An important issue also involves the potential set-backs from the publicity, as the data source is highly sensitive in terms of private information which has caused much controversy in recent years.

Advantages	Disadvantages
• Very large sample size, representative data	• Difficult to access data (legislation and

[10] For GPS location, a simple set of coordinates and precision indication, or a full GPS-standard (NMEA 0183 or NMEA 2000) including might be stored (NMEA 2012). For network-based location, also an approximate location or more extensive collection of sighted antennae or Wifi hotspots with geographical reference can be stored.

source; • Data is quantitative, methodologically feasible; • Passive data collection (no burden on the respondents), high automation level of statistical production; • Can be used in very different statistical domains and produce new statistical indicators that are often previously unmeasurable; • Cost-effective compared to the data collection methods with same sample size (e.g. population census); • Possible to compile historical statistics and generate near-real time indicators; • Good coverage over time and space;	ownership of the data); • Privacy protection aspects and methodological aspects of processing highly sensitive data; • Possible bad publicity for providers (MNOs) and users (government) of the data; • Very few or almost no qualitative information about the sample; • Data quality aspects (geographic accuracy, density of the data, over- and under-coverage issues); • Processing of the data requires powerful computational resources; • Does not always correspond to official statistical indicators;

4 Potential Statistical Applications for Mobile Positioning Data

Mobile positioning data represents the model of the people's mobility. The knowledge about the mobility can be useful in a number of applications where such new data source can produce valuable additional, replacing or new statistical indicators. The data can be successfully used both in public as well as private sectors. Most examples here are based on passive mobile positioning data from MNOs, but other sources and methods can also be employed (e.g. app-based data for traffic density measurements).

4.1 Tourism

As the domestic, outbound roaming, and inbound roaming data corresponds very well to appropriate forms of tourism (domestic, outbound and inbound tourism), the data can be employed for assessing the tourism flows to, within, outside the country, and internationally. Although there are several over- and under-coverage issues and sources for bias (as with all data sources), the inbound roaming data represents the mobility of the foreign tourists within a country very well and has a very good correlation rate compared to existing data (Figure 6).

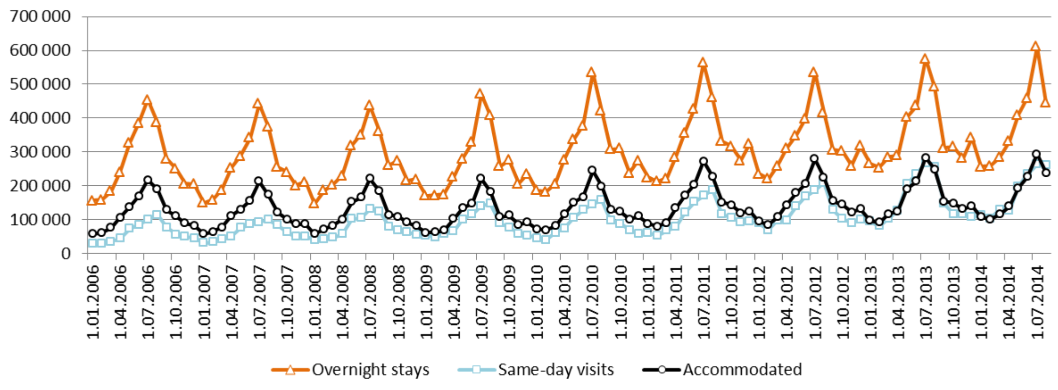

Figure 6. Coherence between overnight stays and same-day visits of inbound tourists from mobile positioning statistics (estimation based on the data of two MNOs), and official accommodation statistics (Statistics Estonia 2014).

Following statistical indicators can be compiled from mobile data in tourism domain:

- Number of trips (I, O, D[11]);
- Number of unique travellers (I, O, D);
- Duration of the visit in a destination country (I, O, D) / in a smaller sub-regions (I, D);
- Breakdown by the country of origin for foreign tourists (I);
- Breakdown by the home administrative unit within the country (O, D);
- Temporal breakdown: day/week/month (I, O, D);
- Overall duration of the trips in spent nights, hours, days present (O, D);
- Geographic accuracy: country (I, O, D), lower level administrative units (I, D);
- Trajectory of tourism trip (I – only inland, O – only country level, D);
- Repeating visits (I, O, D);
- Destination, secondary destinations, transits (I, D);
- Destination country, transit countries (O).

Within the tourism domain, the data can be used in official tourism statistics (domestic, inbound and outbound tourism) for national and sub-national levels; Balance of payments – travel item (inbound and outbound tourism); monitoring the tourism investment effects on large and small regional levels; tourism attractions centres and destination marketing (Figure 7); event visitors (domestic and inbound tourism); long-term and repeating tourists (place loyalty); international travels.

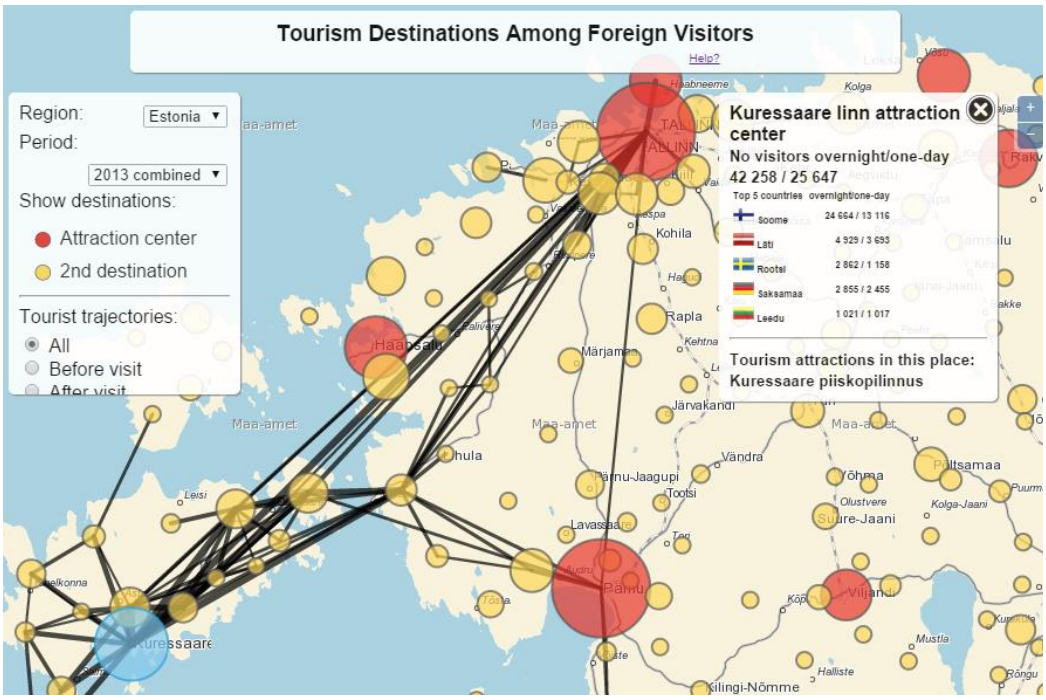

Figure 7. Application for monitoring tourism destinations for foreign tourists within Estonia (attraction centre = more than 10 000 yearly overnight stays) (Positium tourism demo, 2013).

4.2 Transportation Planning, Urban and Regional Planning, Commuters

One of the possibilities of the mobility data is measuring the regular everyday activities of the population. Compared to alternative methods, mobile data can provide more reliable, accurate,

[11] I – Inbound tourism, O – outbound tourism, D – domestic tourism;

timely, and less expensive alternative to traditional methods. By identifying the location of meaningful places (so-called anchor points), and everyday movement paths, one can assume the everyday travel patterns between home and work-time locations (workplace, school). The regular travel patterns combined provides the commuting model for the whole nation.

Following statistical indicators can be compiled from mobile data in transportation domain:

- Origin-destination matrices with hourly and daily travel numbers based on long-term average regular, or actual (for a specific date) data;
- Identification of everyday commuting patterns;
- Spatial accuracy up to 100 m^2 grid (depending on the available data accuracy);
- Breakdown based on the demography (depending on the data available from MNO);
- Average number of trips per person per day;
- Average or actual distance travelled;
- Average or actual travel times.

This data, combined with traditional databases results in a very powerful representation of everyday movements of the population and can be used in a variety of domains: catchment area of administrative centres (optimizing the public services network); Origin-Destination matrices for transportation planning; public transportation planning and optimisation (mobile data represents the demand side); everyday commuters (Figure 8); cross-border commuting; short-term and long-term traffic density data; input for national transportation model; geomarketing applications for businesses (including LBA); outdoor media poster location planning; planning new business locations.

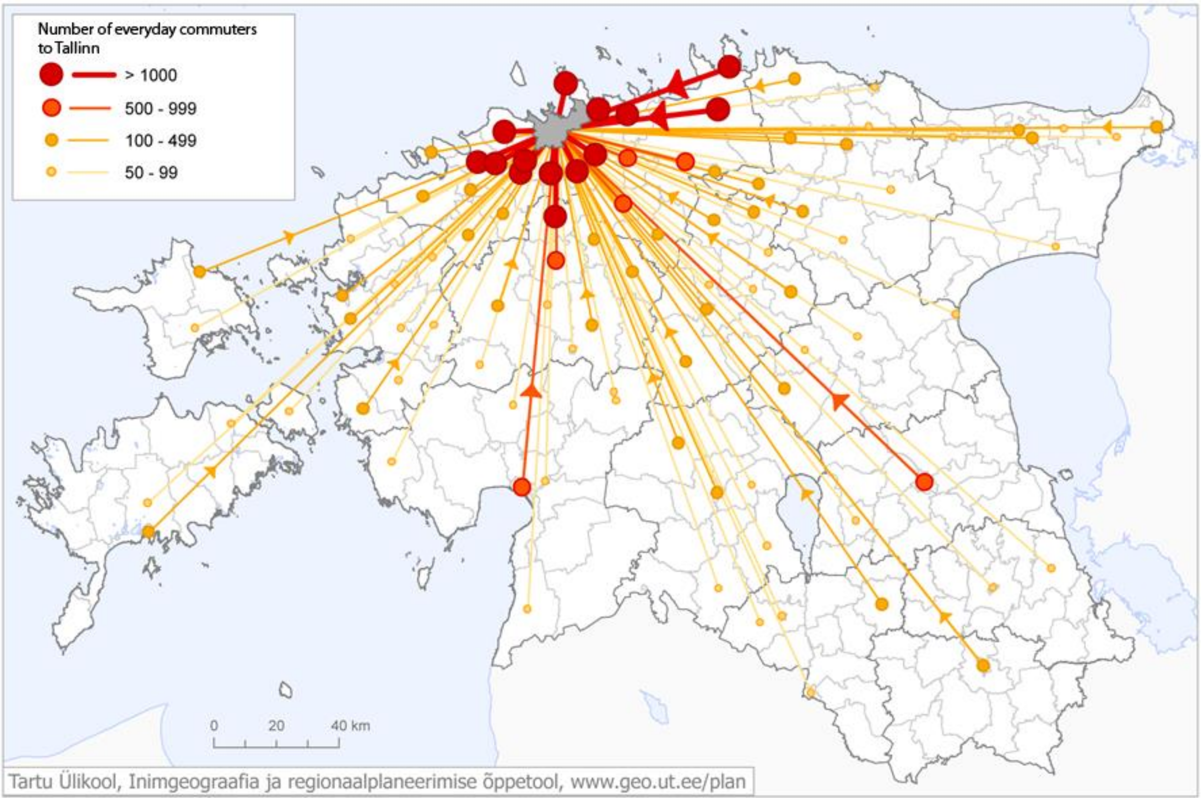

Figure 8. Example of everyday commuters to Tallinn, Estonian capital from other municipalities (University of Tartu, Laboratory of Mobility Studies, 2012).

4.3 POPULATION STATISTICS

Based on home, work-time and other types of anchors, regular and irregular movements, and tourism travels, monitoring some indicators for population statistics can be very fast and reliable. Mobile data could be used as of the intermediate sources for population statistics between population censuses, for assessing internal migration, for monitoring temporary population statistics, and other purposes. Based on the population statistics, the historical analysis, real-time monitoring and prediction of the people present in specific areas can be applied in risk assessment analysis, and for generating different emergency situation scenarios. The following statistical indicators can be generated from mobile positioning data:

- The number of residences geographically distributed according to available accuracy;
- The number of workplace, school, secondary home, and other regular locations;
- Internal migration based on the change of the residences within the country;
- Change of workplace over time;
- Cross-border migration based on the regular travels between different countries;
- Population grid statistics (1 km^2);
- Temporary population statistics
- Assessing temporary population (hourly, daily, weekly, monthly, etc.) (Figure 9);
- Real-time assessment for specific location during the large-scale event, gathering of people or actual emergency situations (e.g. what is the consistence of the crowd in specific location, how many people are affected by an earth-quake of hurricane) (Lu et. al. 2012);
- Risk assessment for law enforcement (planning the number of patrol units in the area based on the consistency of the temporary population).

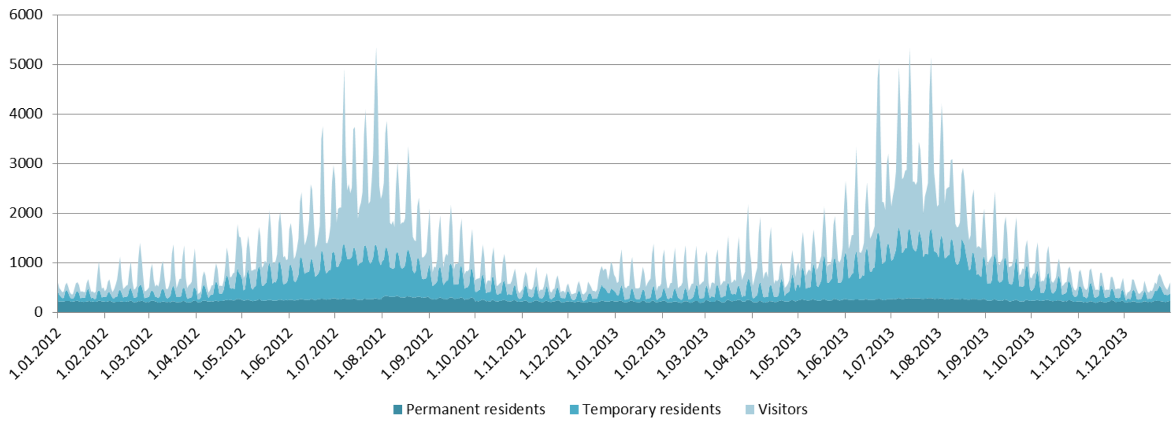

Figure 9. Daily temporary population statistics for a small rural municipality with many summer-hoses in Estonia. The chart represents the number of people present within the municipality during a 2 year period. The seasonality and weekly cycle can be clearly distinguished for second home residents (temporary residents) and visitors.

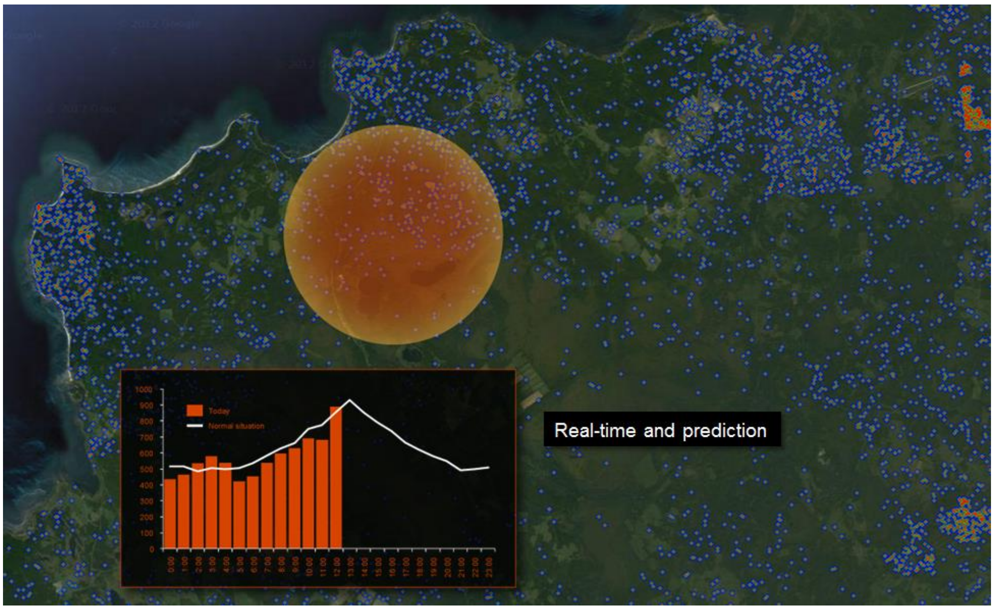

Figure 10. Illustrative example of the estimation of the number of people present in the forest-fire area and the prediction for near future based on historical movement patterns of local residents.

4.4 RESEARCH

There are a number of research areas where mobile positioning data can be and is used. All applications initially require research-based approach with applied theoretical framework of the domain. For example the development of the methodology for identifying anchor points, a theoretical background from human geography and behavioural analysis was applied with additional survey for measuring the accuracy of the anchor identification (Ahas, et. al. 2010). A number of private and public research organisations work with the mobile positioning data in very different domains:

- Tourism research, destination loyalty, travel behaviour;
- Co-presence and social networking analysis;
- Spatial segregation;
- Socio-economic effects of mobility of people;
- Mobility and climate change;
- Marketing research;
- Emergency and public safety research;
- Long-term migration research;
- Epidemiology (modelling the spread of infectious diseases) based on the co-presence algorithms;
- Psychology of the spatial mobility;

There are several disciplines that support the different research areas: computer sciences (how to process big data), and Geographic Information Systems (GIS).

5 CHALLENGES OF DIFFERENT DATA SOURCES

Three different subtypes of challenges can be identified while dealing with mobile positioning data from MNOs:

- Administrative aspects – accessibility to the data, legal questions, privacy protection, the cost of the data, business cases for MNOs;
- Technological – physical access to the data, data types and limitations, and processing the data;
- Methodological – the specific methods for processing the data, and generating the statistically reliable and usable results.

5.1 Administrative Challenges of Accessing the Data

The first and usually most important questions are concerning the accessibility of the data. The owners of the data are companies within whose systems the data is generated (MNOs). They are considered as data controllers and are responsible for any processing, storage, and external use of the data. Usually the privacy protection, electronic communication data, and telecommunication data legislations apply to these organisations. Additionally there might be legislation concerning the collection of statistical information (national statistics acts) that might apply.

For MNOs, the use of the data usually falls under the following categories:

- Use of the data for internal native technical of business purposes (network load, customer billing, etc.);
- Use of the data under legal obligation (providing access to the data for law enforcement based on court order, providing statistics for national statistics institutes, etc.);
- Use of the data for (business) purposes other than of native communication service (e.g. geomarketing purposes, LBA, big data revenue departments).

The first step for getting the access to the data is laying down the appropriate legislation that ensures the justification and legality of the data access and processing. For MNOs this means the legal obligation or legal entitlement to provide access to the data or to process the data.

In parallel with the legislation, the specific business model has to be developed – who is processing the data, under what legislation, who is responsible for processing and storing the data, who is responsible for protecting privacy, who is gaining the value from the resulting data, and what are the financial agreements between the players.

There might be a large number of interested parties and organisations involved in the process. On one part there are original data controllers (MNOs, app developers and other businesses who possess the data, also credit card companies). From the legislation point, there is data protection agencies and national telecommunication authorities who oversee that the data is used according to the law and acceptable privacy protection standards. There might be a number of technology providers who develop the technology to process the data. If MNOs are not directly providing the data, then there might be a data brokerage companies who process and aggregate the data, and deliver the results to the users. On the user side there is large number of public and private organisations who use the data in their own domain.

5.2 Technological Challenges of Processing the Data

There are a number of technological challenges to meet before and during the processing of the mobile data. These are connected to:

- Data source-specific questions – what data sources are available and does the data meet the requirements for expected results (Figure 11). The applicability of the data directly depends on the characteristics of the data, its spatial accuracy, temporal density and other aspects. For example for national monthly tourism statistics, simple billable CDRs might suffice, however for transportation model, the initial data has to be much accurate and more dense;
- Processing-specific questions – mostly concerning the technical ability and resources required to process large amounts of data within expected period of time. Required technical resources (parallel processing might be required) and maintenance costs are much higher if the system has to deliver near-real time results compared to historical analysis of the data over a longer periods of time (monthly);
- Scalability of processing performance as well as scalability of processing data for different domains. Some of the core processing of the data is similar for different domains, but the domain-specific results require individual processing and methodology. Several domain-specific processing requires additional resources and raises maintenance costs.

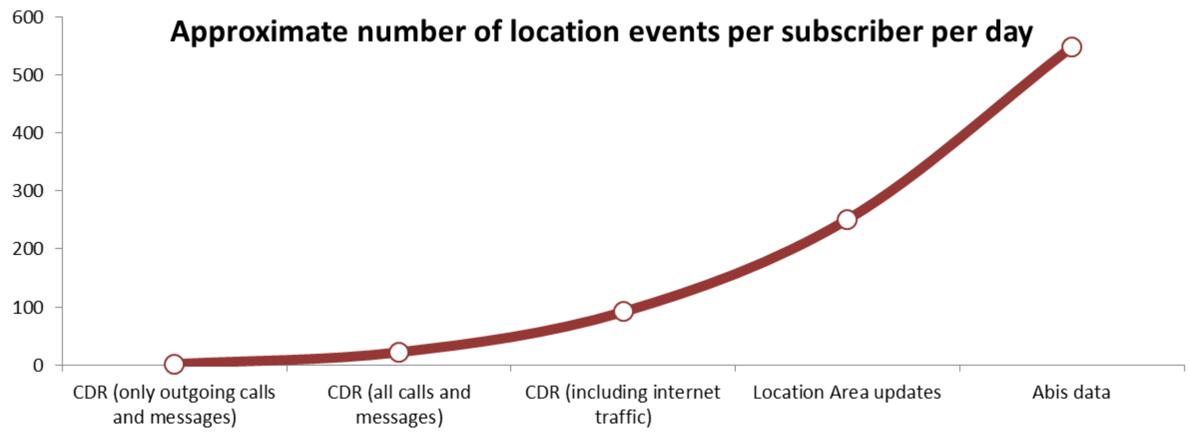

Figure 11. Approximate number of location events per subscriber per day depending on the data source used (cumulative number).

5.3 Methodological Challenges of Processing the Data

Methodological and technological challenges are tightly connected as the requirements for technical resources depend on the complexity of the methodology. There is no single specific methodology for processing the mobile positioning data. Different research groups have developed different methodological approaches depending on the domain of the research, type of the data, and the background of the research groups. However it is possible to develop one core methodology for processing raw data into mobility data model (representation of real life in a database) and implement different domain-specific methodologies on top of that. There are a number of very different methodological issues that need to be addressed; some of them are general and some specific to the domain where the data will be used.

5.3.1 Anonymization of Sensitive Data

Collection of location data of the subscribers or users of apps falls under the jurisdiction of data and privacy protection. It is a highly sensitive data as the personal information can be extracted for individuals that are included in the dataset. Therefore processing of such data requires either consent from the subscribers, legal entitlement (and/or obligation for MNOs), or transforming the data into anonymous data where individual data records cannot be re-identified and linked to a specific persons.

For both active and passive positioning mobile phone data, compilers of official statistics must take precautions to make the data anonymous and process the data according to the privacy protection rules. The legislations that usually govern the use of the data are (the examples are based on EU directives):

- Privacy protection legislation (e.g. Directive 1995/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data);
- Telecommunication data processing legislation (e.g. Directive 2002/58/EC (as amended by 2009/136/EC) of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector;
- Statistics Acts for different data sources and domains where data is used (e.g. Regulation No 692/2011 of the European Parliament and of the Council concerning European statistics on tourism).

Although legislation might provide the legal grounds for national statistical institutes to get access to and process the raw personalised data, it is often required to change the national legislation for this specific data source which might be a lengthy process. Alternative is to apply methods for processing the data anonymously. However there are no known algorithms for making the long term location data anonymous while preserving different aspects of the value of the data (e.g. longitudinal subscriber's data for calculating individual usual environment). Therefore different approaches might be used to solve the issue – e.g. data is processed within MNOs' infrastructure for aggregated statistics and initial sensitive data is not provided for the data processor (e.g. National Statistics Institutes – NSI).

5.3.2 Filtering out the Usable Data, Eliminating Causes for Bias

As any statistical data source, mobile positioning data includes incorrect values (outliers), missing periods of data, values that fall outside the spectre of interest, and other records that have to be identified, compensated, or possibly eliminated from the following processing flow as they might cause decrease of different quality aspects or biased results. The source for such errors in the data might be:

- So-called "machine-to-machine" devices – mobile devices that do not represent any human-related data (e.g. street-light communication modules) but are not eliminated from the database;
- Accidental roaming subscribers – roaming service users who in fact do not enter the county where they are using the roaming service thus "contaminating" the source of data for tourism statistics;
- Missing values – MNOs suffer occasional down-time when there is no data recorded;
- Incorrect data – a number of different possible sources for wrong values (e.g. wrong geographical coordinates for antennae, records with incorrect time values, duplicated records, etc.);

- Subscribers using several mobile devices – if it is possible to recognize, such records should be combined. If it is not possible, this should be handled during the estimation processing.

Some of such "wrongful" data might be eliminated, but in terms of overall data processing, this data might be useful for other purposes. Eliminating the problematic data might be done using some attributes from the records or some spatio-temporal algorithms might be used for identification. Unfortunately there might be cases where the wrongful data might not be recognised.

5.3.3 Domain-specific Methodologies

Depending on the domain the data is processed for, there are several methodological aspects that need to be taken into account. For example, within the tourism statistics domain, a definition of usual environment has to be defined and calculated for each subscriber, and activities within the usual environment have to be excluded from the frame. Within the short-term migration statistics (commuting), a different definition of usual environment and identification of home- and workplace has to be applied and non-regular travels have to be excluded. In most cases the definitions have to be extracted from the official definitions, however because in most cases, as the official definitions are based on specific data sources, the definitions are not directly applicable. For example the question of residency for tourists if often based on the subjective opinion of the traveller, however with anonymous passive mobile positioning, the subjective opinion cannot be asked. This limits the possibility to use unified definitions but often opens up new possibilities (e.g. defining the usual environment based on some quantitative measures – a 30 km buffer zone around the regular movement corridors of the subscriber).

There are also many problems with the data which have importance in specific domains. For example in the context of tourism statistics, tourists may acquire local pre-paid SIM cards in order to make cheaper calls while in the destination country. This generates a possible under-coverage of the tourists within the inbound and outbound roaming data.

5.3.4 Reference Data and Estimation

Because the mobile positioning data is basically a sample data where the sample size is the number of the subscribers from the data provider (i.e. all domestic subscribers, roaming service users of MNO), in order to provide statistical indicators for the general population (a number of tourists, commuters, etc.) it is necessary to conduct estimations for this specific population based on that sample. The data from MNOs includes a selection bias – the sample represents only the users of mobile devices – and is in many cases over- and under-covered (e.g. young children and senior people do not have mobile phones; some subscribers have several connected mobile devices, the subscriber base of MNOs is demographically skewed, technological differences of MNOs, cultural differences of phone usage, etc.). In order to be able to provide accurate estimates based on the mobile data, the reference data has to be used. A variety of data sources might be used as a reference data:

- Population census results (the number of residents, demographical distribution of the country, geographical distribution of residents, etc.);
- Customer profiling of MNOs (to assess the representativeness of the mobile data);
- Accommodation data, border statistics and survey, and other tourism statistics;

- Traffic loop counters;
- National and international transportation statistics;
- National transportation network;
- National land coverage maps;
- Administrative registries that include information about the residence, workplace and other locations;
- Etc.

There are some new indicators compiled based on the mobile data, that cannot be compared to any reference data because there is no other data source. Different quality dimensions also require assessing during the estimation process (validity of the data, accuracy, comparability, etc.).

If the data is provided by several MNOs, the quality of the data increases, the possible error margins decrease (because the sample size and the representativeness is better). However the important question within the methodology is how to process data from different MNOs. One option is to combine the initial data and process the data all together. Alternative is to process individual MNOs' data and combine the data during or after the estimation process. The latter might be inevitable if the processing of the data for some reasons is done within the infrastructure of MNOs.

6 CONCLUSION

Different Information and Communication Technologies' (ICT) data sources are becoming more and more important sources for statistical data. Big Data is a becoming an important source for information about the societies' processes and allows fast and previously unprecedented insights into many aspects of human mobility. Location data from Mobile Network Operators or phone apps is already used in a large scope of domains and is becoming more and more accessible to researchers, decision makers, and statisticians. Although not without many issues, the mobile positioning data is playing a pioneering role among the Big Data sources as the data is very good in describing the mobility of people.

GLOSSARY

Abbreviations used in the current document are mostly common and might not refer to official definition used in specific standard. Following abbreviations are used in the current document (alphabetical order):

Abis – level data traffic. Communication protocol for MSC-BSC-BTS communication. A potential source for location information, however often not stored because of the large amount of data traffic.

A-GPS – Assisted GPS, also network-assisted GPS. An auxiliary system for acquiring faster GPS location by using computational resources of the mobile network. A GPS location of the mobile device is acquired faster if the A-GPS system is available.

BSC – Base Station Controller. The name for the module that controls the activity of several BTSs in mobile networks.

BTS – Base Transreceiver Station. The name for antennae in mobile networks. Usually a BTS can be a directed 120 degree antenna (3 BTSs covers 360 degrees from a single tower) or round 360 degrees antenna.

CDMA – Code Division Multiple Access. One of the two main mobile telecommunication standards.

CDR – Call Detail Record. A basic record of communication activity between network and the mobile device. CDRs are incoming and outgoing calls, messages (SMS, MMS), internet data traffic. Only some of the CDRs may be stored within MNOs databases for limiting storage capacities.

CGI – Cell Global Identity. The global identification of the network antennae. In positioning data, the CGI refers to the method of locating the mobile device with the precision of antenna.

CRM – Customer Relationship Management. System for storing and processing customer information. CRM holds different additional data that can be used to profile the subscribers (e.g. socio-demographic data).

CSV – Comma Separated Values. One option for CDR data format.

DDR – Data Detail Record (also IPDR). A record of internet data traffic. DDR is a part of CDR.

Erlang – Antennae signal load measuring unit.

GIS – Geographic Information System. IT tools for processing geographical data.

GPS – Global Positioning System. A highly accurate US military-based satellite system for acquiring location in different parts of the planet.

GSM – Global System for Mobiles. One of the two main mobile telecommunication standards.

HLR – Home Location Register. The central database for details of each subscriber that is authorized to use the mobile network.

IPDR – Internet Protocol Data Record (also DDR). A record of internet data traffic. IPDR is a part of CDR.

LA – Location Area. The geographically limited area where BTSs are located which is controlled by a single BSC.

LBA – Location Based Advertisement. A part of LBS that concentrates on delivering advertisements that are based on the location of the mobile device to the subscriber.

LBS – Location Based Services. A common name for services that include location within the mobile network.

LTE – Long Term Evolution. 4 generation of mobile communication protocol. Also referred as 4G (with some reservations).

MNO – Mobile Network Operator.

MPS – Mobile Positioning System. System for actively locating the mobile phone from the network.

MSC – Mobile Switching Centre. The name for the module that controls the activity of several BSCs in mobile networks.

NFC – Near Field Communication. A short range communication protocol between a mobile device any device with NFC capability. A rather new technology meant for exchanging quick information and also for mobile payment applications (not implemented widely yet).

NMS – Network Management System. The name for the central system of a mobile network including all central databases and registries that are not distributed.

NSI – National Statistics Institute, National Statistics Office.

OSS – Operation and Support Subsystem. Managing system for supporting different network operations. The geographical data for antennae (BTS) is usually located within OSS.

SIM card – Subscriber Identity Module. SIM card is used as a subscriber identity in GSM standard.

TA – Timing Advance. A measure of the distance between the mobile device and closest antenna(e). Allows more accurate location in active and passive positioning methods (CGI+TA).

TAP – Transferred Account Procedures. A protocol for transmitting roaming data between MNOs.

VLR – Visitor Location Register. Registry keeping the record of the location of specific mobile device within the network. VRL is usually associated with specific MSC.

XML – Extensible Markup Language. One option for CDR data format.

7 REFERENCES

Ahas, R., Silm, S., Järv, O., Saluveer E., Tiru, M. 2010. Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones, Journal of Urban Technology, 17(1): 3-27.

EngineersGarage. 2012. Difference Between GSM and CDMA http://www.engineersgarage.com/contribution/difference-between-gsm-and-cdma

Ericcson. 2013. Location Based Services - a vital component in the Networked Society http://www.ericsson.com/news/130827-location-based-services-a-vital-component-in-the-networked-society_244129227_c

Eurostat, 2014. Feasibility study on the use of mobile positioning data for tourism statistics. http://epp.eurostat.ec.europa.eu/portal/page/portal/tourism/methodology/projects_and_studies

Lu, X., Bengtsson, L., Holme, P., 2012. Predictability of population displacement after the 2010 Haiti earthquake. PNAS 2012 109 (29) 11576-11581; published ahead of print June 18, 2012, doi:10.1073/pnas.1203882109.

NMEA – National Marine Electronics Association, NMEA 0183 standard description, 2012. http://www.nmea.org/content/nmea_standards/nmea_0183_v_410.asp

PCMag. Segan, S. 2013. CDMA vs. GSM: What's the Difference? PCMag. http://www.pcmag.com/article2/0%2c2817%2c2407897%2c00.asp

Positium tourism demo, 2013, http://demo.positium.ee/tourism/.

Saluveer, E., Silm, S., Ahas, R. 2011. Theoretical and methodological framework for measuring physical co-presence with mobile positioning databases. Gartner, G., Ortag, F. (Eds.) Advances in Location-Based Services. 8th International Symposium on Location-Based Services, Vienna 2011. Springer, 247-266

Statistics Estonia, 2014, http://www.stat.ee.

University of Tartu, Laboratory of Mobility Studies, 2012, http://www.geo.ut.ee/plan/en.html.

ANNEX 1. DATA EXAMPLES

Following examples include domestic, inbound roaming, outbound roaming, and antennae geographical reference data with different characteristics. The presented examples are formatted to pre-processing format.

Table 1. Example of domestic data for a single subscriber for 2 days within Estonia. This domestic data includes only outgoing CDRs (call activities and messaging).

msisdn	event_time	cell_id	msisdn	event_time	cell_id
372*******	2013-12-07 09:32:58	6468	372*******	2013-12-08 17:58:08	2756
372*******	2013-12-07 09:39:58	6487	372*******	2013-12-08 21:54:53	327
372*******	2013-12-07 10:32:45	7058	372*******	2013-12-09 08:11:30	7729
372*******	2013-12-07 10:57:30	7058	372*******	2013-12-09 13:13:57	6152
372*******	2013-12-07 11:28:08	6468	372*******	2013-12-09 13:17:46	6145
372*******	2013-12-07 11:32:28	7058	372*******	2013-12-09 14:48:38	4532
372*******	2013-12-07 11:57:14	4256	372*******	2013-12-09 14:50:05	4532
372*******	2013-12-07 12:50:37	7058	372*******	2013-12-09 14:52:33	4532
372*******	2013-12-07 12:52:14	6905	372*******	2013-12-09 14:57:17	4532
372*******	2013-12-07 13:14:59	6487	372*******	2013-12-09 16:06:37	5943
372*******	2013-12-07 13:21:59	7058	372*******	2013-12-14 00:21:54	3947
372*******	2013-12-07 14:08:03	6468	372*******	2013-12-14 01:04:03	3849
372*******	2013-12-07 15:19:19	7058	372*******	2013-12-14 09:56:43	3849
372*******	2013-12-07 16:23:59	7058	372*******	2013-12-14 10:36:06	3849
372*******	2013-12-07 16:47:02	6468	372*******	2013-12-14 10:36:07	3849
372*******	2013-12-07 16:49:50	2755	372*******	2013-12-14 10:36:09	3849
372*******	2013-12-07 20:35:59	7464	372*******	2013-12-14 11:19:39	3849
372*******	2013-12-07 21:29:55	7634	372*******	2013-12-14 13:37:41	4624
372*******	2013-12-07 21:54:35	7634	372*******	2013-12-14 16:16:48	1808
372*******	2013-12-08 00:38:45	7634	372*******	2013-12-14 20:01:14	1506
372*******	2013-12-08 00:41:59	7634	372*******	2013-12-14 20:03:21	3112
372*******	2013-12-08 03:23:09	7634	372*******	2013-12-14 23:57:45	3853
372*******	2013-12-08 03:23:40	7634	372*******	2013-12-14 23:58:02	3853
372*******	2013-12-08 03:51:32	4877	372*******	2013-12-15 10:48:24	4129
372*******	2013-12-08 09:41:47	6468	372*******	2013-12-15 18:25:13	7634
372*******	2013-12-08 12:28:43	6468	372*******	2013-12-15 18:34:42	7464
372*******	2013-12-08 17:07:21	7184	372*******	2013-12-15 19:53:39	7634
372*******	2013-12-08 17:57:05	2756	372*******	2013-12-15 21:42:59	7634

Table 2. Example of outbound roaming data for a single subscriber's 4 day trip from Estonia to Latvia (LV), Belgium (BE), Luxembourg (LU), and Netherlands (NL). This outbound roaming data includes all CDRs (call activities and internet usage records). The actual trip was a flight from Estonian capital Tallinn to Latvia (Riga) then a flight from Riga to Brussels following a train to Luxembourg, return to Brussels, transit to Amsterdam and flight back to Tallinn.

msisdn	iso_a2	event_time	msisdn	iso_a2	event_time
372*******	LV	2013-12-09 16:32:23	372*******	LU	2013-12-11 12:34:37
372*******	LV	2013-12-09 17:08:20	372*******	LU	2013-12-11 12:44:33
372*******	LV	2013-12-09 17:26:43	372*******	LU	2013-12-11 15:04:07
372*******	BE	2013-12-09 19:32:07	372*******	LU	2013-12-11 15:04:11
372*******	BE	2013-12-09 20:13:50	372*******	LU	2013-12-11 17:29:37
372*******	BE	2013-12-09 20:37:23	372*******	LU	2013-12-11 17:36:03
372*******	BE	2013-12-09 21:57:45	372*******	LU	2013-12-11 18:19:44
372*******	BE	2013-12-09 21:57:55	372*******	LU	2013-12-11 20:22:31
372*******	BE	2013-12-09 22:38:19	372*******	BE	2013-12-11 21:39:26
372*******	LU	2013-12-10 00:41:20	372*******	BE	2013-12-11 21:47:13
372*******	LU	2013-12-10 08:13:03	372*******	BE	2013-12-11 21:53:17
372*******	LU	2013-12-10 11:07:39	372*******	BE	2013-12-11 22:08:16
372*******	LU	2013-12-10 11:07:41	372*******	BE	2013-12-11 22:09:00
372*******	LU	2013-12-10 12:04:35	372*******	BE	2013-12-11 22:12:55
372*******	LU	2013-12-10 13:07:45	372*******	BE	2013-12-11 22:30:57
372*******	LU	2013-12-10 14:34:17	372*******	BE	2013-12-11 22:32:47
372*******	LU	2013-12-10 18:53:03	372*******	BE	2013-12-11 22:35:15
372*******	LU	2013-12-10 19:03:36	372*******	BE	2013-12-11 22:37:56
372*******	LU	2013-12-10 19:41:37	372*******	BE	2013-12-11 22:41:16
372*******	LU	2013-12-10 20:27:48	372*******	BE	2013-12-11 23:56:01
372*******	LU	2013-12-10 20:44:57	372*******	BE	2013-12-12 00:01:28
372*******	LU	2013-12-10 21:14:51	372*******	BE	2013-12-12 00:10:20
372*******	LU	2013-12-10 21:29:33	372*******	BE	2013-12-12 11:17:30

372*******	LU	2013-12-10 21:35:04	372*******	BE	2013-12-12 12:07:53
372*******	LU	2013-12-10 21:42:42	372*******	BE	2013-12-12 12:44:16
372*******	LU	2013-12-11 08:23:49	372*******	BE	2013-12-12 13:57:45
372*******	LU	2013-12-11 08:23:55	372*******	BE	2013-12-12 13:58:21
372*******	LU	2013-12-11 08:23:59	372*******	BE	2013-12-12 14:12:51
372*******	LU	2013-12-11 08:24:07	372*******	BE	2013-12-12 16:19:49
372*******	LU	2013-12-11 08:36:28	372*******	BE	2013-12-12 18:15:07
372*******	LU	2013-12-11 08:36:32	372*******	BE	2013-12-12 18:36:45
372*******	LU	2013-12-11 09:32:47	372*******	BE	2013-12-12 22:23:29
372*******	LU	2013-12-11 09:38:40	372*******	BE	2013-12-12 22:25:57
372*******	LU	2013-12-11 09:45:15	372*******	BE	2013-12-13 09:32:50
372*******	LU	2013-12-11 09:54:50	372*******	BE	2013-12-13 10:07:42
372*******	LU	2013-12-11 09:56:55	372*******	BE	2013-12-13 10:16:25
372*******	LU	2013-12-11 11:13:45	372*******	BE	2013-12-13 10:24:30
372*******	LU	2013-12-11 11:55:55	372*******	NL	2013-12-13 12:20:30
372*******	LU	2013-12-11 11:56:43	372*******	NL	2013-12-13 12:28:26
372*******	LU	2013-12-11 12:00:24	372*******	NL	2013-12-13 15:17:29
372*******	LU	2013-12-11 12:10:19	372*******	NL	2013-12-13 16:18:12
372*******	LU	2013-12-11 12:25:28	372*******	NL	2013-12-13 19:22:21

Table 3. Example of inbound roaming data for a Latvian single subscriber's traveling to Estonia for 4 days.

msisdn	event_time	cell_id	iso_a2
371*******	2013-12-08 16:10:55	7201	LV
371*******	2013-12-08 16:12:23	7201	LV
371*******	2013-12-08 16:14:25	7201	LV
371*******	2013-12-09 10:49:03	7201	LV
371*******	2013-12-09 15:32:50	7201	LV
371*******	2013-12-09 15:41:45	7201	LV
371*******	2013-12-09 15:45:25	7201	LV
371*******	2013-12-09 15:51:06	7201	LV
371*******	2013-12-09 17:31:42	101	LV
371*******	2013-12-10 10:23:43	6105	LV
371*******	2013-12-10 10:37:55	4562	LV
371*******	2013-12-10 10:52:06	4357	LV
371*******	2013-12-10 11:16:30	3770	LV
371*******	2013-12-10 12:32:43	4944	LV
371*******	2013-12-10 14:18:53	4944	LV
371*******	2013-12-10 17:35:49	4562	LV
371*******	2013-12-11 08:46:03	7201	LV
371*******	2013-12-11 08:47:21	7201	LV
371*******	2013-12-11 09:23:16	7201	LV
371*******	2013-12-11 12:23:31	7419	LV
371*******	2013-12-11 12:25:47	7560	LV
371*******	2013-12-11 12:27:28	3773	LV
371*******	2013-12-11 12:35:24	6886	LV
371*******	2013-12-12 14:51:35	4944	LV

Table 4. Example of antennae geographical reference table. The accuracy of this table depends on the size of the coverage area of the antenna.

cell_id	lat	lon	coverage_area
7201	59.22735	24.73454	MULTIPOLYGON
101	59.52334	25.37232	MULTIPOLYGON
6105	59.67211	24.96783	MULTIPOLYGON
4562	58.38213	24.40528	MULTIPOLYGON