Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics

Eurostat Contract No. 30501.2012.001-2012.452

Report 3b. Feasibility of Use: Coherence

14 April 2014

Feasibility study on the use of mobile positioning data for tourism statistics

Eurostat Contract 30501.2012.001-2012.452

Report 3b. Feasibility of Use: Coherence

Final report

14 April 2014

Authors of this report (in alphabetical order):

Rein Ahas (University of Tartu), **Jimmy Armoogum** (IFSTTAR), **Siim Esko** (Positium),
Maiki Ilves (Statistics Estonia), **Epp Karus** (Statistics Estonia), **Jean-Loup Madre**
(IFSTTAR), **Ossi Nurmi** (task coordinator, Statistics Finland), **Françoise Potier** (IFSTTAR),
Dirk Schmücker (NIT), **Ulf Sonntag** (NIT), **Margus Tiru** (project coordinator, Positium)

Table of Contents

Introduction

The tourism sector accounts for a significant part of the economy in many European countries. Given the sector's potential in terms of growth and employment, and also in terms of social and cultural integration, any appraisal of its competitiveness and position requires a good knowledge via official statistics of the volume of tourism, the characteristics of tourism trips, the profile of the visitor, tourism expenditure, and the benefits for the economies of the countries being visited. The adoption by the European Parliament and the Council of the European Union of Regulation 692/2011 concerning European statistics on tourism is a major step forward towards a harmonised system of tourism statistics for European countries. (Eurostat 2013b).

Two of the main sources of tourism statistics are accommodation statistics and statistics about participation in tourism, tourism trips and visitors. For accommodation statistics (regarding tourism supply), details are usually used from a census, a cut-off sample, or a probability sample survey taken in tourist accommodation establishments. The cut-off threshold may vary by country. Key variables in accommodation statistics are the number of persons being accommodated and the number of overnight stays by country of residence and by type of accommodation. Information about participation in tourism statistics (regarding tourism demand) is collected through surveys that cover households or individuals. Due to the high cost of such surveys, the sample is generally small and therefore estimates of detailed breakdowns may be of low reliability, and information extracted from small countries may not be disseminated. Other sources of tourism statistics are as follows:

- Structural business statistics (SBS) are a rich and comprehensive source of information on businesses in the EU. Data is available at the four-digit level of NACE and includes a wide range of economic indicators; however, most economic activities are not exclusively serving and relying upon tourists. Eurostat is currently exploring ways of integrating SBS and/or STS (short-term business statistics) data for certain so-called 'tourism characteristic activities' in the tourism statistics databases.
- In the Balance of Payments, travel is one of the main items within the services current account. It differs from other components in that it is demand-orientated, as consumers in the form of 'the traveller' move to the location of the service provider, 'the destination being visited'.
- The Community surveys on ICT usage by households and enterprises collect information on Europeans' usage of the internet when it comes to preparing or

booking travel and accommodation, and of the usage of ICT and e-business applications by enterprises in the tourist accommodation sector.

- Passenger transport plays an important role in tourism, and vice versa. While certain aspects of transport can be both tourism-related and non-tourism-related, other segments (such as, for instance, air transport) can be entirely linked to tourism activity.

- Border surveys are used in several countries to collect information about inbound tourism, the characteristics of visitors, and the tourism trips being taken. Border surveys provide information about the number of trips and/or visitors entering or leaving the country by crossing the border on land. The drawback is that border surveys are increasingly difficult and costly to conduct within the Schengen area due to the free movement of people across borders.

Although there is quite a large range of harmonised tourism statistics indicators available, some topics are still not covered or could be improved. For example: the total volume of inbound tourism, including the volume of accommodation below the threshold and accommodation without charge (non-rented accommodation), border crossing, regional tourism data, reliable detailed tourism demand data, same-day visits, efficient production, and the high quality of tourism statistics, etc.

1.1. Aims, Content and Structure of Report 3b

The focus of this report is on carrying out a quantitative and qualitative comparison of mobile-phone-based tourism statistics with reference statistics that include official tourism statistics and other available indicators that are related to tourism. The qualitative element of this report is to describe the reasons for deviations in these data sources. The report does not repeat the methodological issues described in Report 3a (Methodology) but rather quantifies the total impact of the methodology used to produce statistics on tourism flows that are based upon mobile positioning data.

The analysis is carried out by utilising the mobile phone based statistics from Estonia. These statistics are compared against the official tourism statistics made available by Eurostat and against other indicators that are related to tourism statistics.

The key questions addressed in this report include the following:

- How exhaustively do positioning-based statistics cover physical tourism flows as measured by the reference statistics?

- How well does positioning-based data support a breakdown into the various components of tourism: inbound, outbound, domestic tourism and/or same-day trips, rented accommodation and non-rented accommodation, personal and professional purpose trips?

- How does positioning-based data compare to existing statistical indicators?

- What are the reasons behind the deviations in existing statistical indicators? Are these deviations strengths or weaknesses of mobile positioning data?

- Does positioning-based data support previously unavailable levels of detail such as geography, time, nationality, etc.?

- Are there any previously unavailable metrics that can be compiled using positioning-based statistics?

1.2. Methods Used

The report focuses on carrying out a quantitative comparison of positioning-based data versus existing tourism statistics.

Mobile-positioning-based tourism statistics were provided by Positium, being based upon data taken from two Estonian mobile network operators. The data includes Estonian inbound, outbound and domestic tourism between 2008 and 2012. The data is presented in an in-depth format in Section 2. The project consortium's original plan was to carry out the tests using mobile positioning data taken from at least three consortium countries. As presented in Report 2, the consortium contacted 106 and discussed pilot data with nineteen MNOs. Due to the barriers described in Report 2, the consortium was unable to access suitable pilot data from other countries besides Estonia. Therefore, the tests presented in this report are based on Estonian mobile positioning data only.

The reference statistics were provided mainly by Statistics Finland and Statistics Estonia. Reference data that was available on the Eurostat website was also used where it concerned the outbound trips of Estonians to EU Member States. The reference statistics are described in Chapter 3.

The mirror statistics method is used to compare two metrics that describe the same scope of observation. An example of mirror statistics is the number of Finnish tourists in the Estonian positioning-based inbound statistics when compared to the number of Finnish tourists in the official Finnish outbound tourism statistics. These comparisons are visually plotted on appropriate graphs and are analysed using statistical indicators that measure both coverage and consistency.

1.3. Background of the Report

This report is part of a larger study which has the goal of assessing the feasibility of using mobile positioning data for the production of tourism statistics. There are six partners collaborating in the study. The entire study consists of five reports that address the main objective from various aspects. These reports are as follows:

a) Report 1 - Stock-taking contains an up-to-date description of the state of affairs of using mobile positioning data in research and applications in tourism statistics and related domains. Report 1 provides both a list and descriptions of the experiences of consortium members and other public and private institutions that have been involved in projects in which mobile positioning data has been accessed. The report provides references to existing problems and solutions in technology, methodology, regulations and other aspects of accessibility to the data that will eventually serve as important input for Report 2.

b) Report 2 - Feasibility of access provides an overview of the regulatory, financial, technical and other related topics that cover all aspects of data accessibility using, among other things, references to the existing project that are present in Report 1. Report 2 provides input for barriers as well as opportunities on this topic.

c) Report 3a - Feasibility of use, methodological issues: this report provides a methodology for the production of tourism statistics by using mobile positioning data. A detailed description of the production process is given. An evaluation of the quality of the methodology described is conducted.

d) Report 3b (this report) - Feasibility of use, coherence: Report 3b assesses the coherence of tourism statistics that have been acquired from various sources (mobile positioning data, accommodation statistics, household and individual surveys, transport statistics, etc.). Tourism statistics taken from several countries will be analysed. An evaluation of mirror statistics will be carried out and the possible usage of mobile positioning data to increase coherence will be analysed.

e) Report 4 - Opportunities and benefits: this report concentrates on the potential opportunities and benefits that can be brought to tourism statistics through the use of mobile positioning data. In Report 4, the consortium does not collect or research new data or information, but rather integrates the results from

previous reports into a structured and coherent assessment of the potential usage of mobile positioning data in the field of tourism.

This report covers the objectives of Report 3b. The current report is founded on the results from Reports 2 and 3a. This report serves as input to the subsequent Report 4.

2. Mobile Positioning Data

This chapter presents the Estonian mobile positioning data source that is available for the purpose of carrying out the coherence analysis. Mobile positioning data has been provided by Estonian mobile network operators (MNOs), and has been processed into relevant aggregates concerning tourism statistics. The methodology for carrying out this aggregation is described in Report 3a (Section 2).

Although Estonia is a small Member State within the EU, there are certain positive elements that can be useful when it comes to testing the coherence of tourism statistics that are based upon mobile positioning. Firstly, the main inbound and outbound tourism flows take place between Finland and Estonia, which are separated by the Gulf of Finland. Most tourists will use the ferry and will pass through the ports of Helsinki and Tallinn, which is something that is reflected in ferry passenger statistics. Secondly, Estonia has a border with a large non-Schengen country: Russia. The inbound flows of Russian tourists to Estonia will therefore be reflected in the Estonian Border Control statistics. Thirdly, Estonia has a land border with another EU member state - Latvia - making it possible to analyse tourism flows that take place across a land border.

The pilot data from Estonia contains inbound and outbound roaming and domestic call data between January 2008 and July 2013. Up to August 2011 the data is drawn from one MNO, and starting from September 2011 it is drawn from two MNOs. The initial data is based upon the Call Detail Records (CDR) that have been processed according to the methodology described in Report 3a, with these records having been aggregated and estimated to represent the general population of inbound, outbound and domestic tourism, excluding non-tourism travellers and trips so that the indicators that are compiled from the records are as fully comparable as possible to the official statistics.

2.1. Estonian Inbound Mobile Positioning Data

Inbound data is based upon the inbound roaming CDR of the Estonian MNOs. The data is presented both as indexes that are based upon raw indicators (before estimation) and the absolute numbers of estimated indicators which represents the general population on a country level and on the level of fifteen Estonian counties. From the perspective of inbound tourism, a trip is considered to be a single trip made by a foreign resident to the country of Estonia. A visit is considered to be a visit to a specific location (such as a county or municipality). A visit to the country can therefore include multiple county-level visits.

The methodology used to compile the aggregates is described in Report 3a Section 2.2.2.2. The data contains only trips that have been made by tourists. Residents and in-transit visitors have been excluded from the data, along with seamen and accidental border noise.

The inbound data contains the following variables:

- reference month;
- country of origin (iso_a2 code);
- the country of origin;
- whether it is an overnight or same-day trip - the distinction is made with a basis on the number of days that a subscriber is factually present at the location, based upon the network records (CDRs).
- total number of visits started during the reference month;
- total number of visits ended during the reference month;
- total number of nights spent there during the reference month;
- total number of days spent there during the reference month (a 6 hour visit = 1 day);
- average duration of stay during the reference month - the duration of the visit in days (a 6 hour visit = 0.25 days);
- the LAU level code for an Estonian county (for county-level data);
- name of Estonian county (for county-level data).

2.2. Estonian Outbound Mobile Positioning Data

Outbound data is based upon the outbound roaming data taken from domestic subscribers of Estonian MNOs. The data is presented both as indexes that are based upon the raw indicators (before estimation) and upon the absolute numbers of estimated indicators representing the general population that is involved in outbound tourism. The data is

presented at a foreign country level (showing the number of trips outside Estonia). From the perspective of outbound tourism, a trip is considered to be a single trip made by a local resident outside Estonia. A visit is considered to be a visit to a specific foreign country.

The methodology used to compile the aggregates is described in Report 3a Section 2.4.2.1. The data contains only trips that have been made by tourists. Foreign residents and transit visits to the same foreign countries have been excluded from the data, as has accidental border noise.

The pilot data contains the number of visits that have been made by Estonian residents to the top forty outbound destinations in terms of country. The data contains the following variables:

- reference month;
- destination country (iso_a2 code);
- the name of the destination country;
- whether it is an overnight or same-day trip (for the whole trip);
- whether it is an overnight or same-day visit (to a specific country);
- the total number of visits begun during the month;
- the total number of visits ended during the month;
- the total number of nights spent at the destination country during the month;
- the total number of days spent in the country (a 6 hour visit = 1 day);
- average duration of the visit to the country (a 6 hour visit = 0.25 days).

2.3. Estonian Domestic Mobile Positioning Data

Domestic data is based upon the CDRs of the domestic subscribers of Estonian MNOs within Estonia. The data is presented both as indexes that are based upon the raw indicators (before estimation) and the absolute number of estimated indicators that represent the general population on a country level and on the level of fifteen Estonian counties. From the perspective of domestic tourism, a trip is considered to be a single trip made by a local resident outside their usual environment, within the borders of the country of reference. A visit is considered to be a visit to a specific county.

The methodology used to identify the usual environment and to compile the statistical indicators is described in Report 3a Section 2.3.2.2. The data contains only trips that have been made by tourists. Transit visits have been excluded from the data, as has accidental border noise.

The data contains the following variables:

- reference month;
- destination county (LAU1 code) or country level;
- the total number of same-day trips;
- the total number of overnight trips;
- the total number of days in which an individual has been present in the county (a 6 hour visit = 1 day);
- the total number of spent nights in the county

Domestic data also includes information on the origin of the visitors (in terms of county); the number of visits to the destination by visitors; the duration of trips (outside the usual environment) and the duration of stays at the destination; plus travel routes, and several other indicators which were not used.

2.4. Estimation Example: From Raw to Estimated Data

Data used in the coherence analysis is aggregated and corrected using the methodology described in Report 3a. Figures 1-4 present an example of inbound data for one specific country with comparison to the main source for reference data - the number of monthly accommodated visitors from this specific foreign country. The original, uncorrected aggregates for the data from two MNOs (Figure 1) is compared to the accommodation statistics, using suggested candidates for six month average correction coefficients (Figure 2) that are based upon the mathematical discrepancy in the reference data (accommodation, surveys, border statistics, ferry traffic). Figure 3 and Figure 4 present the corrected data.
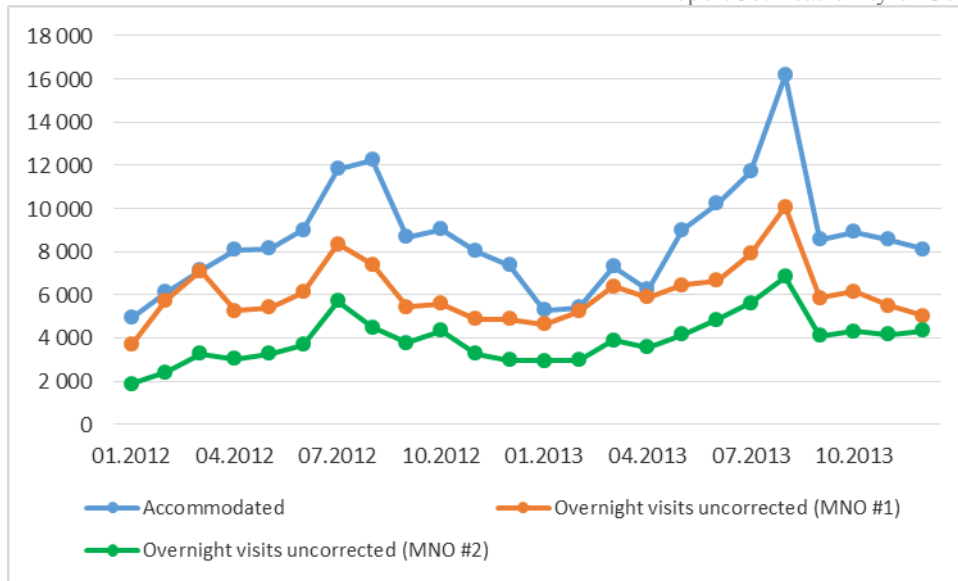
Figure 1. Uncorrected overnight visits when compared to accommodation statistics (Source: Positium, Statistics Estonia).
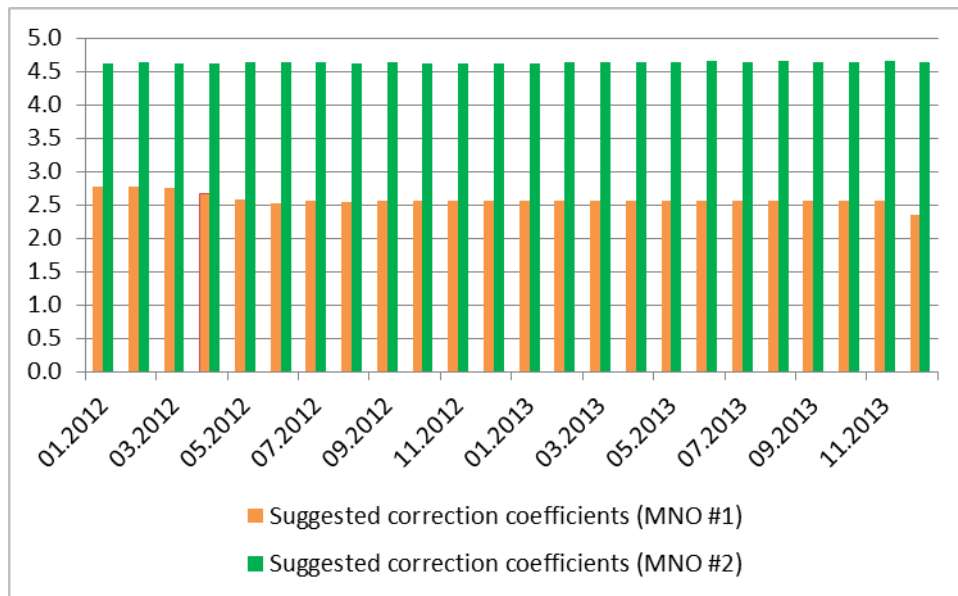

Figure 2. Correction coefficients used for each of two MNOs for one country
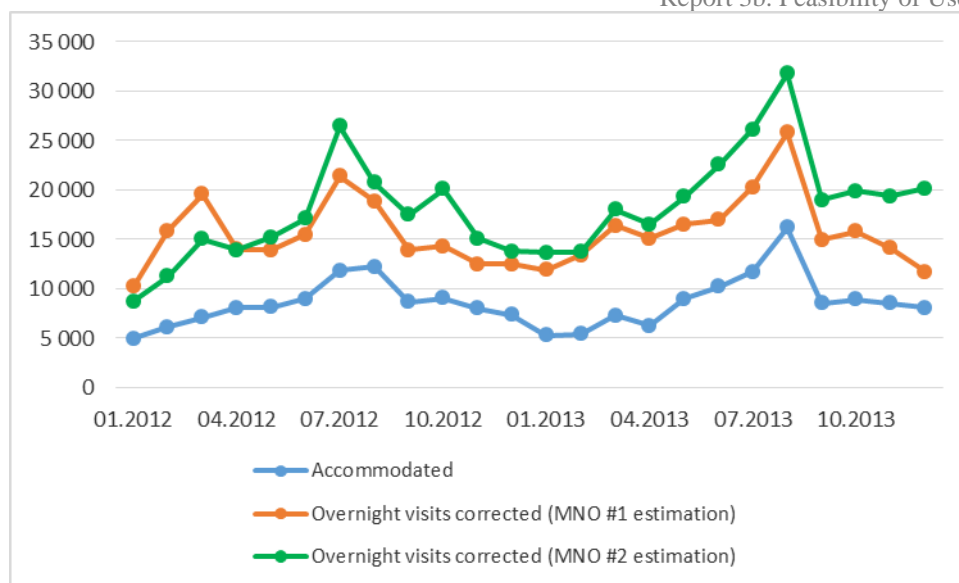
Figure 3. Corrected overnight visits when compared to accommodation statistics (Source: Positium, Statistics Estonia).
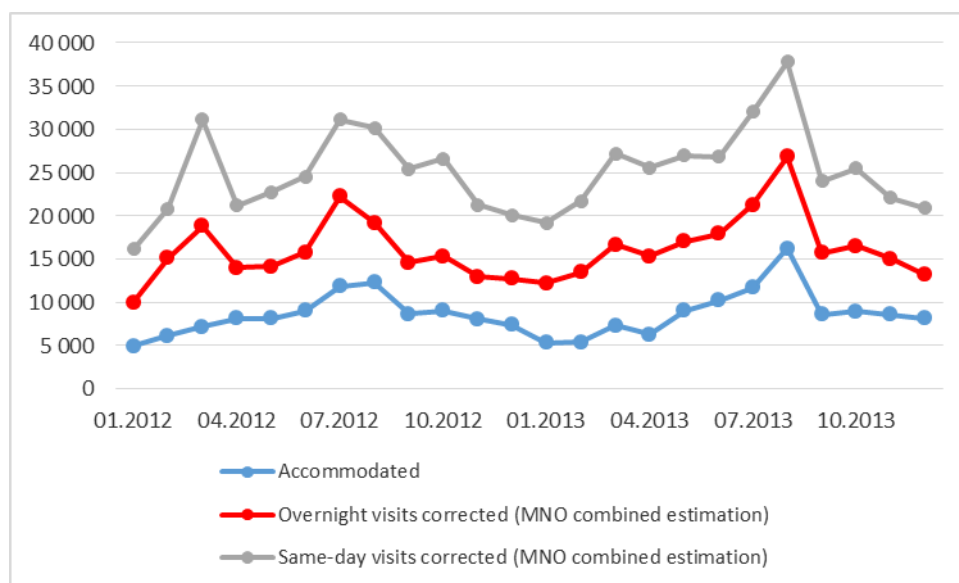
Figure 4. Corrected overnight and same-day visits when compared to accommodation statistics (Source: Positium, Statistics Estonia).

3. Reference Statistics

This chapter presents the statistical indicators that can be used as reference for analysing the coherence of mobile positioning data. The criterion for selecting this set of statistical indicators is that all of these indicators cover a certain part of the total tourism flows. The main reference statistics are the official tourism statistics that are compiled according to Regulation (EU) 692/2011.

In addition to the official tourism statistics, there are a number of other potentially relevant statistics concerning tourism flows. These other statistics include border interview surveys, travel surveys, passport control data, and passenger data. The availability and methodology of these statistics varies greatly depending on the country concerned.

3.1. Summary of Reference Statistics

Table 1 and Table 2 present the reference statistics in a structured format, using standard concepts of statistical quality. Mobile positioning data is also included in the tables so that a comparison can be made in terms of each quality dimension. Table 1 compares the quality dimensions of mobile positioning data to official tourism statistics that have been compiled using Regulation (EU) 692/2011 as a basis. Table 2 does the same for other potentially useful statistical data sources.

Table 1. Statistical quality dimensions in official tourism statistics and mobile positioning data.

	Mobile positioning data	Supply statistics (accommodation statistics)	Demand statistics
Target population	Outbound/domestic: the population of the reference country. Inbound: non-resident tourists	Accommodation establishments (all or above threshold)	Population over fifteen years
Frame	Data from mobile phone subscribers	Business/tourism register.	Population register or area frame
Source data	Administrative	Enterprise survey	Survey of individuals
Sampling design	Census/sample	Census/sample	Sample
Time units available	Day/week/month etc	Month	Month/Quarter/Year
Regional coverage	Any customised area	NUTS2	Country
Nationality breakdown	Possible	Possible	Residents only
Timeliness	1-2 weeks	5-8 weeks	7-8 weeks
Legal basis	None	Regulation 692/2011	Regulation 692/2011

Table 2. Statistical quality dimensions of tourism-related statistics data and mobile positioning data.

	Mobile positioning data	Border interview	Passenger statistics	Border control	Travel agencies	Payment card data
Target population	Outbound/domestic: the population of the reference country. Inbound: non-resident tourists	Inbound visitors to the reference country/outbound travellers upon departure or return	All passengers by transport mode (air/sea/rail, etc)	All passengers passing through border control.	All tourists using the services travel agencies in the reference country	Transactions of payment cards within the reference country
Frame	Data from mobile phone subscribers	Main border crossing locations	The transport authorities register.	The border authority register.	The travel agencies register.	The card issuers register.
Source data	Administrative	Survey of individuals	Administrative	Administrative	Administrative/Survey	Administrative
Sampling design	Census/sample	Sample	Census	Census	Census/sample	Census
Time units available	Day/week/month etc	Month/quarter	Day/week/month	Day/week/month	Quarter/year	Quarter/year
Regional coverage	Any customised area	Country	Country	Country	Country	Country
Nationality breakdown	Possible	Possible	Not possible	Possible	Not possible	Possible
Timeliness	1-2 weeks	10-20 weeks	1-2 weeks	1-2 weeks	12 weeks	7-8 weeks
Legal basis	None	None	EU Regulations 1365/2006, 91/2003, 437/2003, Directive 2009/42/EC	National	None	None

3.2. Official Tourism Statistics

Official tourism statistics that are compiled within the EU are done so based firstly on
Regulation (EU) 692/2011, which covers tourism statistics and national regulations. The
regulation is divided into two parts, these being supply and demand statistics. Supply statistics
concern internal tourism (specifically the capacity and occupancy of tourist accommodation
establishments), while demand statistics deal with national tourism (tourism trips made by
residents).

The major benefit of the official statistics is the harmonisation of outputs: each EU member state should produce the same minimum set of tourism statistics. The national statistical institutes also typically compile tourism statistics for the needs of national users which are often more extensive than those required by regulations.

The Member States transmit their statistics to Eurostat according to a predefined schedule in the regulation. After validation has been carried out, Eurostat makes the statistics available online. Eurostat also maintains a Methodological Manual (Eurostat 2013a) that contains guidelines, definitions and common standards for compiling tourism statistics.

3.2.1. Supply Statistics

Tourism supply statistics deal with capacity and the occupancy of accommodation establishments. Accommodation (either rented or non-rented) is a core tourism sector even if it is relevant for one section of the total number of visitors alone (i.e. tourists = overnight visitors). The economic importance of this sector can be seen from the results of Tourism Satellite Accounts (TSA) of some countries in which accommodation services accounted for between 15% and 20% of total internal tourism expenditure according to national TSAs (Eurostat 2013b). What's more, when looking at the physical flows, over 2.5 billion nights were spent at tourist accommodation establishments in the EU during the year 2012 (Eurostat Tourism Data).

Table 3. Quality dimensions of Supply Statistics.

Target population	Frame	Source data	Sampling design	Time units available	Regional coverage	Nationality breakdown	Timeliness	Legal basis
Accommodation establishments (in total or above threshold)	Business / tourism register	Enterprise survey	Census / sample	Month	NUTS2	Possible	5-8 weeks	Regulation 692/2011

Accommodation statistics play a key part in the system of tourism statistics in the EU and they have a long history in terms of data collection. The drawback involved in accommodation statistics is that, according to demand-side estimates, roughly only 50% of tourism trips (including overnight stays) are registered in accommodation statistics. The remaining 50% of trips are spent in less formal or smaller establishments (which fall outside the scope set by many Member States), in privately-owned dwellings (possible secondary residences), and with friends or relatives who offer accommodation for free.

Tourism accommodation establishments are classified and described in groups according to the NACE Rev. 2 classification as follows:

- 55.1 (hotels and similar accommodation),
- 55.2 (holiday and other short-stay accommodation) and
- 55.3 (camping grounds, recreational vehicle parks and caravan parks)

To reduce the burden on administrations and reporting units, the Regulation allows certain limitations in the scope:

1. For 'hotels and similar accommodation' and for 'holiday and other short-stay accommodation', the scope of observation shall at least include all tourist accommodation establishments which have ten or more bed places.
2. For 'camping grounds, recreational vehicle parks, and caravan parks', the scope of observation shall at least include all tourist accommodation establishments which have ten or more pitches
3. Member States that account for less than 1% of the total annual number of nights spent at tourist accommodation establishments in the European Union may further reduce the scope of observation to include at least all tourist accommodation establishments which have twenty or more bed places (or twenty or more pitches).

Where a limitation on the scope as described above in points 1), 2) or 3) is applied, an estimate is transmitted annually to Eurostat of the total number of nights spent during the reference year by residents and non-residents in the tourist accommodation establishments that have been excluded from the scope of the observation.

This study focuses on the relevant occupancy variables of accommodation statistics. The capacity variables (the number of establishments, the number of bed places and the number of bedrooms) are considered to be beyond the scope of mobile positioning data, as is the occupancy rates of rooms and bed places. The occupancy variables that will be used as reference statistics for mobile positioning data are as follows:

- **Number of nights spent**: A night spent (or overnight stay) is classed as being each night a guest or tourist (either resident or non-resident) actually spends in a tourist accommodation establishment (either sleeping there or merely staying there)
- **Arrivals:** An individual (in the form of a tourist) who arrives at a tourist accommodation establishment

17

The Member States report annually, six months after the end of the reference year, both on nights spent and on arrivals, using the following breakdowns:

1. **By type of accommodation** (NACE 55.1, 55.2, 55.3)
2. **By locality Type A** (densely populated area, intermediate area, thinly populated area)
3. **By locality Type B** (coastal or non-coastal area)
4. **By size class for establishment** (<25 bedrooms, 25-99 bedrooms, >99 bedrooms)
5. **By guest's country of residence**: EU Member States, Iceland, Norway, Switzerland, Russia, Turkey, Ukraine, South Africa, the USA, Canada, China, Japan, Republic of Korea, or Australia
6. **By guest's geographical region**: EU, EFTA, other Europe (excluding EU/EFTA), Africa, North America, South and Central America, Asia, Australia/Oceania/Other

In addition, the Member States report monthly data on arrivals and nights spent. Monthly data is available with a delay of between six to eight weeks and includes only resident/non-resident breakdowns broken down by NACE groups.

The full regulation can be found in Annex 2.

3.2.2. Demand Statistics

Tourism demand statistics deal with information on tourism trips, on the visitors who are making the trips, and on general participation in tourism. The scope of observation consists of residents of the country who are aged fifteen or over. Demand statistics therefore deal only with outbound and domestic tourism. The data is typically collected by means of household or individual surveys. The respondent is asked a series of questions concerning trips that ended last month or in the last quarter. In contrast, arrivals in supply statistics are registered at the beginning of their trip or visit.

Table 4. Quality dimensions of Demand Statistics.

Target population	Frame	Source data	Sampling design	Time units available	Regional coverage	Nationality breakdown	Timeliness	Legal basis
Population over 15 years	Population register	Survey of individuals	Sample	Quarter / Year	Country	Only residents	7-8 weeks	Regulation 692/2011

For the reference years 2012 and 2013, the regulation only requires the Member States to provide data on overnight trips. The data collection of outbound same-day trips will start from reference year 2014. Domestic same-day trips will be collected only every three years

starting with the reference year 2015, at which time data collection is voluntary. The first obligatory reference year for domestic same-day trips is 2018.

The data is transmitted to Eurostat both in table format and as a micro-data file. The micro-data file includes only overnight trips. The full description of statistics is transmitted according to the regulation; please refer to Annex 2.

The main mirror data for Estonian inbound mobile positioning data is the Finnish Demand Statistics. In 2012 the methodology for compiling the Finnish Demand Statistics was updated due to the regulation causing a slight break in time series. For the 2012 survey, the population was extended from people aged between 15 and 74 to those aged between 15 and 84, and more accurate weighting coefficients were also introduced, along with significant changes being made to the interview questionnaire. On account of the revision, the 2012 statistics are not fully comparable with previous years. The extension of the age group increased the numbers of trips by one to three per cent at the main level. The revision of the weighting coefficients increased the numbers of trips by another one to three per cent at the main level. It is difficult to define the effect of the changed questions to the results that have been obtained. Overall this methodological update resulted in an increase of between 2% to 5% in the number of trips when compared to previous years, depending on country. However, the increase seems to be even greater in the case of Estonia, as the number of trips to Estonia increased by nearly 12% (Statistics Finland 2013) in 2012 in the Finnish Demand Statistics, while ferry passengers between Finland and Estonia increased by only 3% in 2012 (Finnish Transport Agency 2013). In Estonian mobile positioning data the number of Finnish inbound trips is nearly the same for 2011 and 2012.

The reference to the regulation (EU 692/2011) is found in References chapter.

3.3. Border Interview

The border interview surveys have traditionally been the main means for compiling statistics on inbound flows. Recently, a large number of Member States, including Finland and Estonia, have ceased to produce these statistics due to budgetary reasons. It has also become increasingly difficult to conduct the surveys, especially for land borders within the Schengen area since vehicles simply drive across the border without stopping.

The border interviews are typically conducted at border crossing locations such as airports, ferry terminals and land border crossing sites. The border interview is targeted at non-residents and the interviews are conducted either upon arrival or departure. Through the

means of border interviews it is possible to obtain information on the purpose of a trip, the means of accommodation to be used, and especially on money spent on various commodities during the trip.

The border interview statistics used in this report are the Finnish Border Interview Statistics that are available up to the year 2012. Starting from 2013, Finland has also ceased to conduct the border interview.

3.4. Passenger Statistics

The geographical location of Finland and Estonia is ideal for using ferry passenger statistics as a reference for mobile positioning data. Finland and Estonia are separated by the Gulf of Finland, the distance between the capitals Helsinki and Tallinn being roughly 70km. The passenger flow between Estonia and Finland takes place mostly via the ferries that operate regularly between the two countries. Statistics on these passenger flows are compiled on a monthly basis by the Finnish Transport Agency and are based upon a census taken from a national register (PortNet).

The geography between Sweden and Estonia is similar, and regular ferry routes operate between Tallinn and Stockholm. Due to the greater distance between Tallinn and Stockholm, the ferry passenger statistics should be accompanied by air passenger statistics. In 2012, over 25% of passengers used a flight instead of a ferry for their passage between Sweden and Estonia. Statistics on both ferry and air passengers between Sweden and Estonia were provided by Statistics Estonia.

The main drawback in using passenger statistics is that, typically, there is no possibility of distinguishing one group from another in passenger statistics, since the nationalities of passengers are contained within the ship's manifest, which cannot be accessed for statistical purposes. Nevertheless, passenger statistics are useful for specific routes, such as the Finland-Estonia ferry where the majority of passengers are of Finnish or Estonian nationalities.

3.5. Border Control Statistics

Border control data contains inbound passenger flows that originate from outside the Schengen area. The data is collected from registers that are maintained by border authorities. The clear drawback is that border control provides data only for those passengers of non-Schengen countries who enter the Schengen area for the first time and who are required to present their passport at that point. In certain cases this is not a great drawback. For example,

most Russians travelling to Finland will enter the Schengen area by crossing the Finnish border. Therefore, data from the Finnish border control provides a very good estimate for the total number of Russians travelling to Finland. For other countries and nationalities the situation may be less ideal, as trips into the Schengen area may have started from another Schengen country.

Passport control data statistics typically contain the following information:

- reference time-period (e.g. month)
- type of flow: inbound or outbound
- name and location of the border-crossing site
- nationality of the passport-holder
- number of passport control checks (= number of passengers)

The Border Control statistics used in this report are those Estonian Border Control Statistics that cover Russians who cross the border into Estonia.

3.6. Travel Agencies

Certain travel agencies and related associations publish statistics on the number of trips that are sold by destination country. The drawback in terms of these statistics is that they only include package trips (including at least a flight and a hotel), with such trips being sold by certain travel agencies. These statistics may be useful as reference data for mobile positioning in cases involving certain typical holiday destinations where the majority of trips are package trips.

3.7. Payment Card Data

The travel item on the Balance of Payments often relies on payment card data that involves expenditure. Payment card data typically contains expenditure aggregates by card issuer country which typically is split into commodities that are consumed by tourists, such as car rentals, flights, accommodation, etc. Since travel expenditure is often correlated to the physical flows of tourists, payment card data may in certain cases act as reference statistics for mobile positioning.

4. Coherence Analysis

4.1. Analysis Framework

From the point of view of an individual country or member state, the total tourism flows can be broken down by using two dimensions as a basis: by domain and by type of accommodation used. The tourism domains are inbound, outbound and domestic tourism. Based upon the type of accommodation used, tourism trips can be classified into same-day trips and overnight trips with either rented or non-rented accommodation. A third dimension could be a geographical breakdown that is based upon the level of detail for the trip's destination, such as country, region or municipality.

Table 5 presents the framework that will be used for carrying out a series of analyses that compare mobile positioning data to domain-specific reference statistics. The comparisons will be made using Estonian mobile positioning data. The main country for mirror statistics is Finland, since from the Estonian perspective, the biggest tourism flows take place between Finland and Estonia and because the situation on the Gulf of Finland allows a good assessment of mirror statistics. This framework can also be used as a general 'test bench' once access is gained to mobile positioning data from other countries.

Table 5. Framework for comparing mobile positioning data and reference statistics.

Tourism Domain (of Estonia)	Mobile Positioning Data	Reference (Mirror) Statistics
Combined inbound tourism + outbound tourism		
Total trips	Indexed raw data (inbound+outbound)	Ferry passengers EE - FI Ferry+air passengers EE - SE
Inbound tourism (to Estonia)		
Total trips	Total trips	Demand Statistics (FI) Border Control (EE)
Overnight trips	Overnight trips Indexed raw data	Demand Statistics (FI) Supply Statistics (EE)
Same-day trips	Same-day trips	Demand Statistics (FI)
Nights spent on overnight trips	Overnight trips (nights spent)	Supply Statistics (EE)
Outbound tourism (from Estonia)		
Total trips	Total trips	Demand Statistics (EE) Border Interview (FI)
Overnight trips	Overnight trips	Demand Statistics (EE) Supply Statistics (EU)
Same-day trips	Same-day trips	Not available (begins 2014)

Domestic tourism (in Estonia)		Demand A
Total trips	Total trips	Demand Statistics (EE)
Overnight trips	Overnight trips	Demand Statistics (EE)
	Indexed raw data	Supply Statistics
Same-day trips	Same-day trips	Not available (begins 2018)

There are two types of mobile positioning data used in this framework: estimated data and raw indexed data. Estimated mobile positioning data is measured in terms of the number of trips or visits and has been fully processed according to the methodology presented in Report 3a. Raw indexed data (January 2012 = 100) is calculated from the raw number of subscribers with no 'estimation' corrections as described in Section 2.6 of Report 3a. This raw number of subscribers is indexed due to the sensitivity of the data.

The estimated mobile positioning data will be used when it is independent of the reference statistics. Independence here means that the reference statistics have not been used as an input into the process of estimating the data. The mobile positioning data that is used in the analyses in the forthcoming sections is by default estimated data, and exceptions to this are clearly indicated.

Indexed raw data will be used in addition to estimated data when the reference statistics have been used in the process of estimating the estimated data. For example, Estonian accommodation statistics have been used as an input when estimating the inbound mobile positioning data for Estonia, and this data is 'affected' by accommodation statistics. Therefore, comparing accommodation statistics to the estimated inbound mobile positioning might provide unrealistically accurate results.

4.2. Statistical Indicators for Coherence Analysis

There are two main aspects when it comes to analysing the coherence of mobile positioning data in measuring tourism flows. The first is the **coverage** of mobile positioning data in terms of completeness. The second is **consistency** over time, meaning how well mobile positioning data captures the trends and seasonal fluctuations that are associated with tourism flows. Separate statistical indicators were chosen to measure these two elements of coherence.

4.2.1. Coverage Indicator

The indicator that was chosen for measuring the **coverage** of mobile positioning is 'CC2. Asymmetry for mirror flows statistics - coefficient'. The 'ESS Guidelines for the

Implementation of the ESS Quality and Performance Indicators 2010' manual defines this indicator as being 'the absolute difference between inbound and outbound flows to and from a pair of countries divided by the average of these two values'. This indicator is calculated as follows:

$$CC2A_B = \frac{OF_{AB} - mIF_{AB}}{\frac{OF_{AB} + mIF_{AB}}{2}}$$

Where:

CC2 = Asymmetry for mirror flows statistics - coefficient

A = Country A

B = Country B

OF_{AB} = Outbound flow from country A to country B

mIF_{AB} = Mirror inbound flow

The analogy here is that the mobile positioning data is the primary tourism flow (= outbound flow) and the reference statistics are a mirror statistics source (= inbound flow). The value indicates the magnitude of potential under-coverage or over-coverage. If the value obtained is close to zero then the two data sources match perfectly. A negative value indicates potential under-coverage in mobile positioning data or over-coverage in the reference data. Similarly, a positive value indicates potential over-coverage in mobile positioning data or under-coverage in the reference statistics.

4.2.2. Consistency Indicator

The indicator that was chosen for measuring consistency over time is the Pearson product-moment correlation coefficient or, in short, Pearson's r. The indicator measures the linear correlation between the two variables X and Y. The value ranges from -1 to 1, where 1 is a total positive correlation, 0 is no correlation, and −1 is a total negative correlation. By definition, Pearson's r is calculated as the covariance of the two variables divided by the product of their standard deviations. The Pearson's r is calculated for a population (ρ) as follows:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Where cov is the covariance, σ_X is the standard deviation of X.

In the case of comparing reference statistics to mobile positioning data, the Pearson's r will indicate the correlation between the two statistics on tourism flows. The closer the value is to 1, the better a correlation there is with mobile positioning data to the actual tourism flows as indicated by the reference statistics.

4.3. Combined Inbound and Outbound Tourism

Combined inbound and outbound mobile positioning data is used as a reference for passenger statistics. When considering a single transport link such as, for example, the ferry route between Finland and Estonia, the passengers aboard such a vessel can be classified into three categories. The first group are those residents of Estonia who are returning from an outbound trip. The second group are those residents of other countries who will be inbound visitors to Estonia. The third group consists of transit passengers who will continue their trip to another country. All of these groups will be included in an aggregated number of passengers. Therefore, the correct reference for passenger statistics is to combine both the inbound and outbound mobile positioning data and presume that there is a certain degree of over-coverage in passenger statistics due to other nationalities and in-transit passengers.

The ferry passenger statistics have also been used as an auxiliary data source in the process of estimating Estonian mobile positioning data. Bearing this in mind, the analysis is made using both the estimated data, which has been influenced by the ferry statistics, and the raw mobile positioning data that is independent of ferry statistics.

4.3.1. Ferry Passengers between Estonia and Finland

Most of the passenger (and tourism) flows between Estonia and Finland travel via ferry between Helsinki and Tallinn. Passenger statistics for ferry passengers is based upon monthly statistics that are compiled by the Finnish Transport Agency, based upon their administrative data. The following charts compare the Estonian combined inbound and outbound mobile positioning data to the monthly ferry passengers crossing between Finland and Estonia between 2009 and 2012. Figure 5 is an index of raw mobile positioning data and ferry passengers where January 2012 = 100. Figure 6 provides the same comparisons on the number of inbound and outbound trips in mobile positioning data versus the absolute number of ferry passengers.
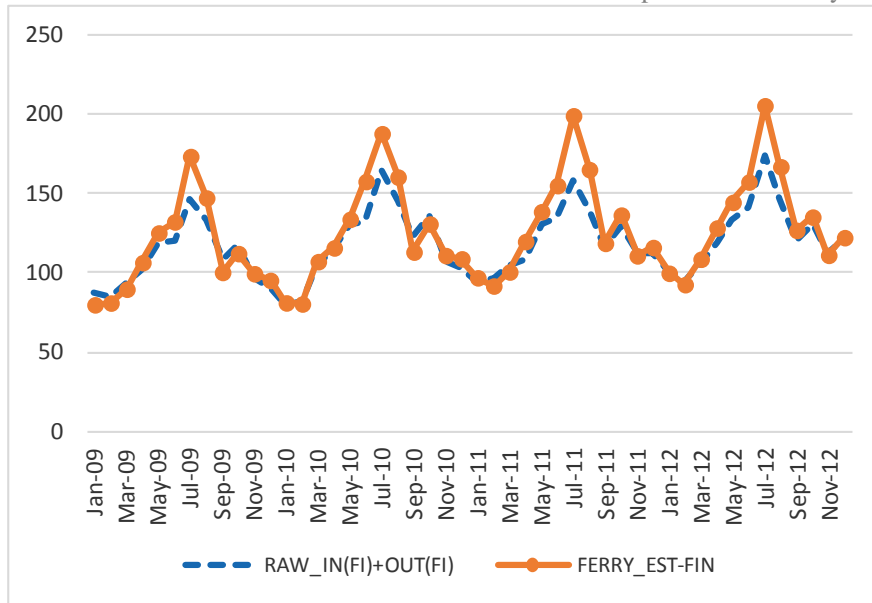
Figure 5. Monthly inbound (Finns to Estonia) + outbound (Estonians to Finland) raw mobile data index when compared to the ferry passenger index between Finland and Estonia; January 2012 = 100 (Source: Positium, Finnish Transport Agency).

Table 6. Coherence indicators: Ferry passengers between Estonia and Finland - raw data.

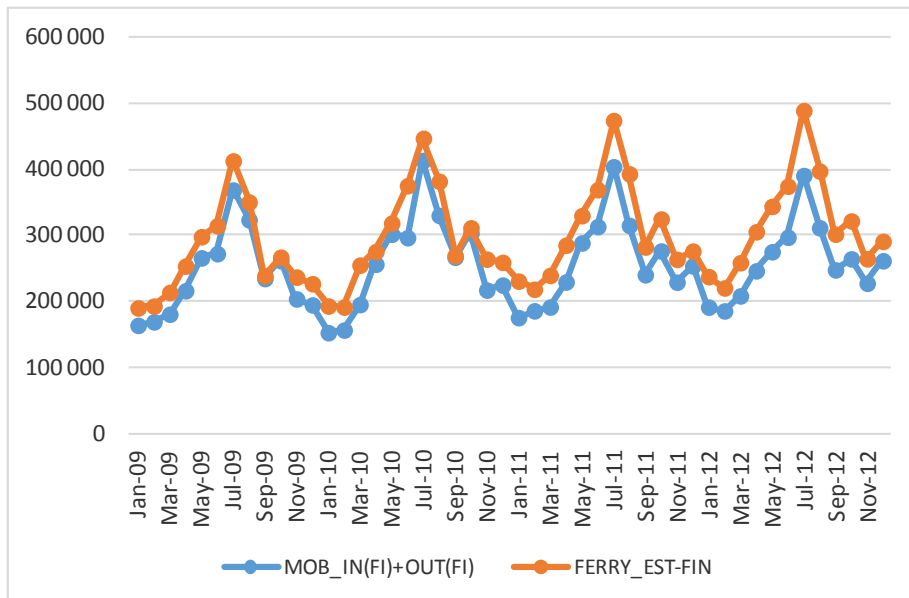| Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.96	N/A	N/A	N/A

Figure 6. Monthly trips inbound (Finns to Estonia) + outbound (Estonians to Finland) in mobile positioning data when compared to the number of ferry passengers sailing between Finland and Estonia (Source: Positium, Finnish Transport Agency).

Table 7. Coherence indicators: Ferry passengers between Estonia and Finland.

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.96	-0.16	-0.27	-0.01

Table 8. Coverage indicator (CC2) by monthly average (2009-2012)

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
CC2	-0.22	-0.17	-0.22	-0.17	-0.13	-0.19	-0.14	-0.17	-0.10	-0.10	-0.16	-0.12

In the case of Finland and Estonia, there is a near-perfect correlation (0.96) between ferry passengers and mobile positioning data. The correlation is as strong in the indexed raw data (

Table 6) as it is in the estimated mobile positioning data (Table 7). The estimation process does little to improve mobile positioning data in terms of consistency since the raw data is already nearly identical to ferry passengers in terms of seasonal basis and trend.

In terms of coverage, the number of trips in mobile data is always less than the number of passengers by an average of 16%. The reason for this under-coverage is that although Finns and Estonians constitute a majority of ferry passengers, passengers from other nationalities are also on board, as are transit passengers. Mobile positioning data, on the other hand, only includes outbound trips taken by Estonians and inbound trips taken by Finns. The under-coverage is therefore both expected and justified. There are some monthly differences in the CC2 with the least under-coverage in September-October (-0.10) and the most in January (-0.22), March (-0.22) and June (-0.19). This would indicate that the share of other nationalities (other than FI/EE) and transit passengers on board the ferry would be at its highest in January, March and June, and at its lowest in September and October. The absolute difference is naturally at its highest during the peak holiday months of June to August, at which time the highest number of tourists from other nationalities are also present.

4.3.2. Ferry and Air Passengers between Estonia and Sweden

Sweden is much farther away from Estonia than is Finland. The reference statistics used for Sweden originates in the total number of ferry and air passengers which is data that is provided by Statistics Estonia. The following graph (Figure 7) compares these passenger statistics to the quarterly mobile positioning inbound trips of Swedes to Estonia and the associated outbound trips of Estonians to Sweden between 2009 and 2012.
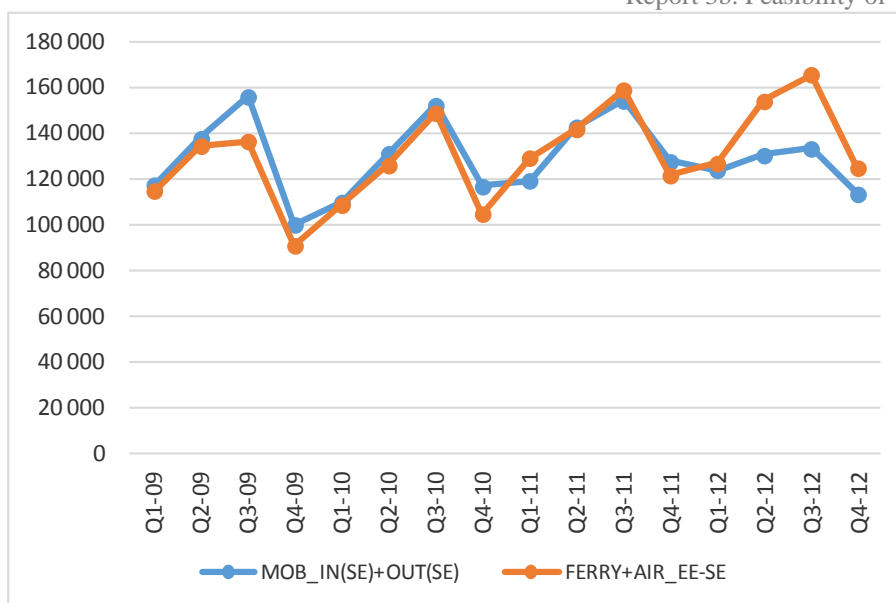
Figure 7. Monthly trips inbound (Swedes to Estonia) + outbound (Estonians to Sweden) in mobile positioning data when compared to the number of ferry and air passengers going from Estonia to Sweden (Source: Positium, Statistics Estonia).

Table 9. Coherence indicators: Ferry and air passengers between Estonia and Sweden.

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.77	-0.01	-0.22	0.13

Passenger statistics and mobile positioning data are relatively well correlated (Pearson's r = 0.77). The years 2009 to 2011 are clearly different from 2012. For mobile data, the difference between 2009-2011 and 2012 is the fact that the first three years are based upon data from one MNO while 2012 is based upon the data from two MNOs. In the case of Sweden, introducing another MNO clearly changes the trend. Since the basic presumption is that the number of trips in mobile positioning data should be less than the number of passengers, it is likely that for 2012 the mobile positioning becomes more accurate because the CC2 indicator is constantly negative during 2012, which is what it is expected to be. In fact, if Pearson's r is calculated separately for 2009-2011 and 2012, the value is roughly 0.9 for both. The interpretation is that there is a strong correlation when using data from either one or two MNOs, but introducing another MNO caused a break in the time series which, in this case, resulted in a consistency issue for the overall trend.

In conclusion, using passenger statistics as a reference yields encouraging results which show that mobile positioning data does indeed provide a good estimate for the total inbound and outbound tourism flows. Since this test is restricted to Sweden, Finland and Estonia, this conclusion applies to these countries and more testing and examples are needed in order that broader conclusions can be reached.

4.4. Inbound Tourism

This chapter focuses on the Estonian inbound tourism flows. The inbound mobile positioning data is compared to the following reference statistics:

1. Finnish Demand Statistics: outbound trips to Estonia;
2. Estonian Supply Statistics: arrivals and nights spent in accommodation establishments by non-residents;
3. Estonian Border Control statistics.

4.4.1. Total Inbound Trips: Demand Statistics

The main mirror statistics for total tourism trips in Estonian inbound mobile positioning data is the Finnish Demand statistics, which have been compiled according to EU Regulation 692/2011 on Tourism Statistics. Finns are by far the largest group of visitors to Estonia. In 2012 more than 40% of non-resident tourism trips to Estonia were made by Finns. The next largest group were Russians with a 15% share.

Figure 8 presents the number of total tourism trips to Estonia for each quarter during 2008-2012, both for inbound mobile positioning data and for Finnish Demand Statistics. The Finnish Demand Statistics include both same-day trips and overnight trips.
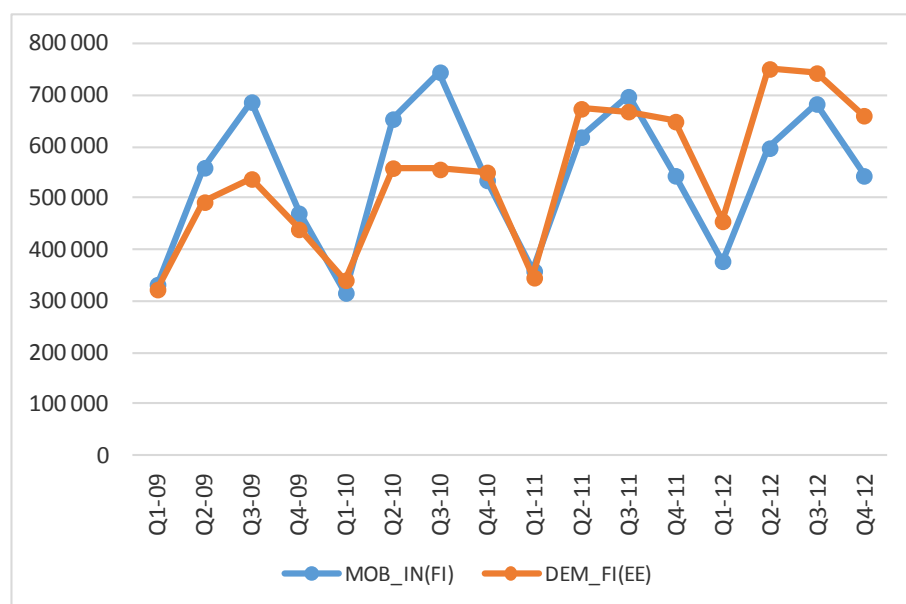

Figure 8. Quarterly inbound trips (Finns to Estonia) in mobile positioning data when compared to the total number of trips to Estonia in Finnish Demand Statistics (Source: Positium, Statistics Finland).

Table 10. Coherence indicators: Total inbound trips (Finns to Estonia).

Years	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
2009-2012	0.77	0.00	-0.23	0.29
2009-2011	0.83	0.05	-0.18	0.29
2012	0.95	-0.17	-0.23	-0.08

The impact of the methodological update in Finnish Demand Statistics, as described in Chapter 3.2.2, is clearly visible in Figure 8. In the years 2009-2011 the total number of trips in Demand Statistics are either less than or almost the same as in mobile positioning data. For 2012 the number of trips in Demand Statistics was always higher than in mobile positioning data. Due to the methodological change, the years 2009-2011 should be treated independently of 2012, which is something that is reflected in Table 10. The Pearson's r value is 0.83 for 2009-2011 but this improves to 0.95 for 2012.

Concerning the purpose of the trip – personal or business-related, there are limited possibilities to assess the trips based on the purpose. Because there is no possibility to ask the person, the assessment can be made based on the temporal or spatial characteristics of the mobile data. For example an assumption can be made that trips made on weekends (starting on Friday or Saturday and ending on Saturday or Sunday) are most probably personal trips. For remaining majority of the trips, assumption based on the purpose of the trip are almost impossible to make (see Figure 9.). According to Finnish Demand Statistics only between 9 to 13 per cent of all trips to Estonia are business-related each year.

Figure 9. The proportion of the purpose of the trip of Finns to Estonia from Finnish Demand Statistics and the share of the weekend trips based on the mobile positioning data (Source: Positium, Statistics Finland).

There are several findings that are based upon an interpretation of this analysis:

1. The methodological update improved the results of the Finnish Demand Survey where they concerned the total number of trips to Estonia - the correlation to mobile positioning data is higher.

2. The Estonian inbound mobile positioning data provides a consistent estimate on the number of Finnish inbound trips.

3. However, mobile positioning data appears to underestimate the number of Finnish inbound trips for 2012. The main reason for this underestimation could be the fact that some Finnish tourists don't use their mobile phones for calls or SMS during their trips. According to Eurobarometer 15% of Finnish travellers switch off their mobile phones and never use them during the travelling in other EU countries (Special Eurobarometer 414 2014).

4. The purpose of the visit (personal or business) can be presumed for only a part of the trips (e.g. weekend trips).

4.4.2. Total Inbound Trips: Border Control

The border control statistics are relevant as reference statistics when a large share of inbound tourism flows for a certain nationality enter the Schengen area by crossing the country's border. At the border, the tourists need to present their passport to the border authorities and they will be registered in the border control statistics for the inbound country. In Estonia, this is the case with Russian citizens entering Estonia by crossing the border between Estonia and Russia.

Figure 10 provides a comparison of monthly inbound trips made by Russians in mobile positioning data when compared to the number of border crossings made by Russian citizens to Estonia.
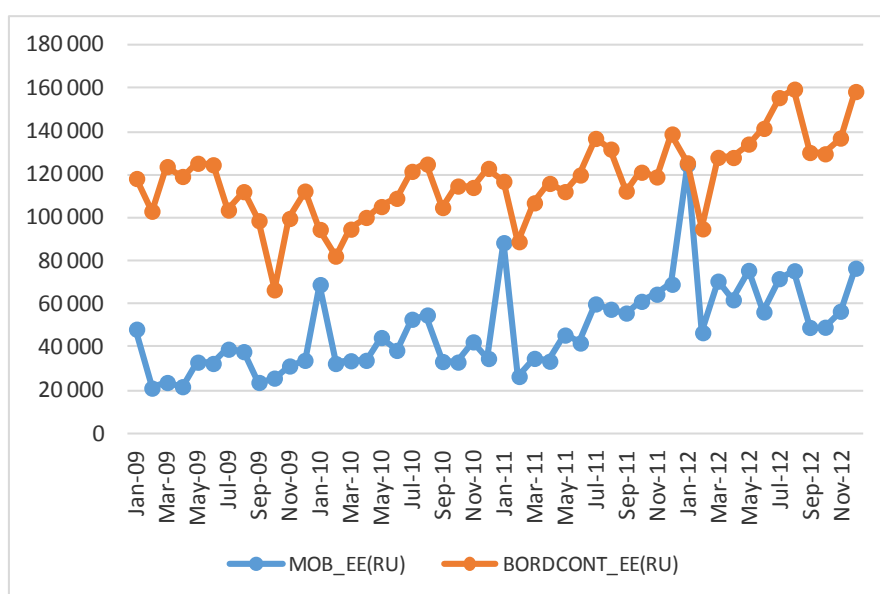

Figure 10. Monthly inbound trips (Russians to Estonia) in mobile positioning data when compared to the total number of border crossings made by Russians to Estonia (Source: Positium, Statistics Estonia).

Table 11. Coherence indicators: Border Control (Russians to Estonia).

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.55	-0.86	-1.38	-0.01

The correlation between border control and mobile positioning data is rather weak (0.55) in the case of Russian residents crossing the border into Estonia. Mobile positioning data also heavily underestimates the total number of Russian inbound trips to Estonia. The reason for such under-coverage can be found by looking beyond these statistics at the profile of the trips that are made by Russians. It is pretty well known that a high percentage of trips made by Russians to Estonia is made up of short shopping trips. During these trips, many Russian visitors will often not use the roaming services offered by Estonian MNOs at all. Either they will keep their phones offline or will remain within the coverage area of their home MNO such as, for example, on shopping trips to Narva, a city which is located just next to the Russian border.

This example illustrates the fact that very short trips to destinations which are close to the border may be seriously underestimated in mobile positioning data because these tourists might not use the roaming services of MNOs in the inbound country. To further validate this, Figure 11 breaks down the mobile positioning data for Russian visitors into same-day and overnight trips.
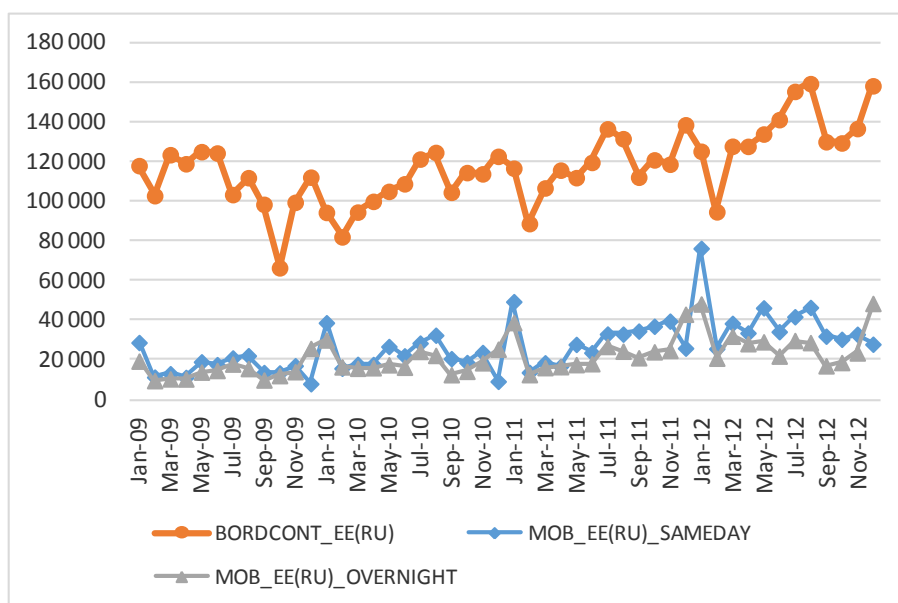

Figure 11. Monthly inbound trips (Russians to Estonia) in mobile positioning data when compared to the total number of border crossings made by Russians to Estonia (Source: Positium, Statistics Estonia).

Table 12. Coherence indicators: Border Control (Russians to Estonia).

Type of trip	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
All trips	0.55	-0.86	-1.38	-0.01
Same-day trips	0.48	-1.27	-1.73	-0.49
Overnight trips	0.56	-1.39	-1.69	-0.89

Based upon this analysis it is evident that a major proportion of Russian visitors in Border Control statistics are not captured in mobile positioning data. It's safe to assume that the majority of border crossings are same-day visits with some visitors even crossing the border several times a day. Even so, the correlation against mobile positioning data is actually worse (0.48) for same-day trips than it is for overnight trips (0.56).

Although at first this may seem to represent poor levels of performance for mobile positioning data, certainly not all short visits shown in Border Control data would qualify as same-day tourism trips. According to Regulation 692/2011, Article 2(1), those trips that qualify as same-day trips are visits which do not involve overnight stays, which are made outside one's usual environment, and which originate from one's normal place of residence. The methodological manual for tourism statistics (Eurostat 2013a) provides three further criteria for ruling out certain short trips from the tourism category in Section 1.3.1:

1. The **purpose** of the visit: it is useful to exclude from the tourism category those trips that are made (only) for the purpose of maintaining one's 'daily living' or as part of one's 'current life routine'.
2. The **duration** of the visit: a lower limit of three hours is recommended in order to allow for the exclusion of same-day trips that are most probably too short to include a 'tourist' element.
3. The **frequency** of the visit: a trip made repeatedly once a week is considered to be inside the usual environment and therefore is not included in the tourism category.

Border Control statistics do not rule out any trips since all passports are checked at the border. According to a Finnish study (Nurkka 2012), in 2012 a total of 35% of Russian respondents said that they come to Finland between one and four times per month and 6% at least once a week. Using a straightforward calculation it can be estimated that 25% of all border crossings are made by the most frequent Russian visitors who visit Finland every week and more than 50% by those who visit between one and four times a month. Additionally, 77% of visitors stated that shopping was the primary purpose of their visit. According to the

criteria laid out in the Methodological Manual (Eurostat 2013a), a large proportion of these visitors should therefore not be included in the tourism category.

Due to the reasons described above (no phone use/no need for roaming), many of these short trips are already being excluded from mobile positioning data, and mobile positioning data therefore has a higher share of 'real' tourism trips than Border Control statistics. Furthermore, in case inbound mobile positioning data has a non-changing subscriber ID, it is possible to exclude those subscribers who appear so frequently that they can be considered to be within their usual environment.

In conclusion, mobile positioning data is a far better source for measuring inbound tourism than is Border Control statistics. The under-coverage of short trips in mobile positioning data when combined with the possibility of filtering out frequent trips is a strength that is inherent in mobile positioning data. In contrast, simply too many visitors in the Border Control statistics are within their usual environment either due to frequency, duration or purpose of visit.

4.4.3. Inbound Overnight Trips: Demand Statistics

This section continues the analysis of Finnish Demand Statistics by restricting the scope to overnight trips only. In the Finnish Demand Survey, the respondent specifically reports trip duration in terms of nights spent. For overnight trips, at least one night spent was reported by the respondent.

The following graph (Figure 12) compares the quarterly inbound overnight trips of Finns to Estonia in mobile positioning data to the number of overnight trips to Estonia in the Finnish Demand Statistics during 2010-2012.
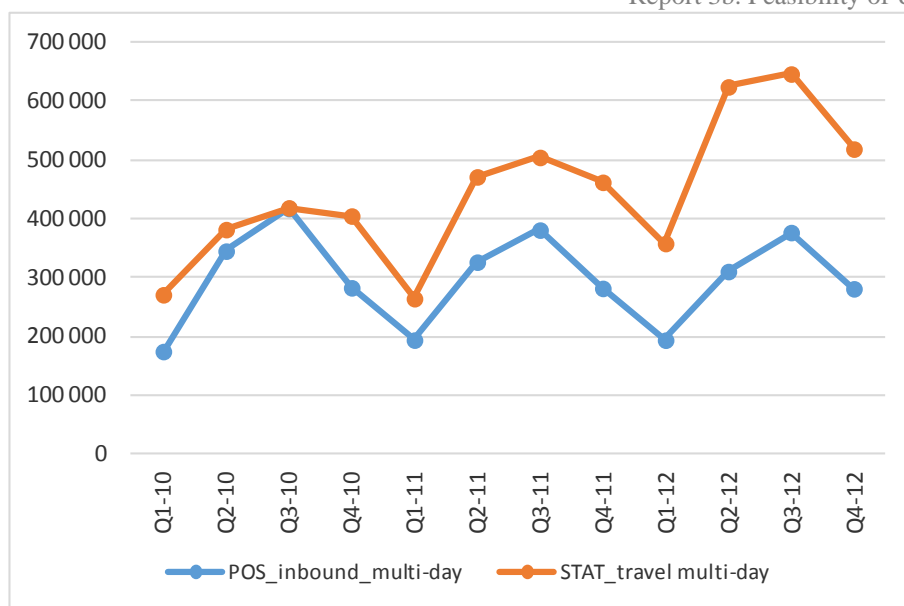
Figure 12. Quarterly inbound overnight trips (Finns to Estonia) in mobile positioning data when compared to the total number of overnight trips to Estonia in Finnish Demand Statistics (Source: Positium, Statistics Finland).

Table 13. Coherence indicators: Overnight inbound trips (Finns to Estonia).

Years	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
2010-2012	0.63	-0.39	-0.67	0.00
2010-2011	0.77	-0.29	-0.48	0.00
2012	0.96	-0.60	-0.67	-0.53

Due to the methodological update in Finnish Demand Statistics in 2012 the results are shown separately for 2012 and for 2010-2011. The correlation (0.96) for 2012 is as strong as it is for the total number of trips. However, the under-coverage issue is more severe for the total trips in Section 4.4.1. The number of overnight trips in Demand statistics for 2012 is more than double when compared to the number of overnight trips in mobile positioning data (CC2 average = -0.6).

The key for interpreting this under-coverage issue is to look at the methodology used for breaking down the total number of trips in mobile positioning data into same-day and overnight trips. According to the methodology presented in Report 3a, at least two activities (calls and SMS) are necessary at long enough time intervals for a trip to be classified as an overnight trip. If only one activity is registered for the visitor during the entire trip it will be considered to be a same-day trip in mobile positioning data, although in reality the individual concerned may have spent one or more nights during their trip without using their phone. This can easily lead to a situation in which many of the presumed same-day trips in mobile

positioning data are actually overnight trips in reality. The next analysis on same-day trips sheds further light on this assumption.

4.4.4. Inbound Same-Day Trips: Demand Statistics

This example focuses on the same-day trips made by Finns to Estonia in Finnish Demand Statistics. Figure 13 compares these trips to the Finnish same-day inbound trips in the Estonian inbound mobile positioning data.
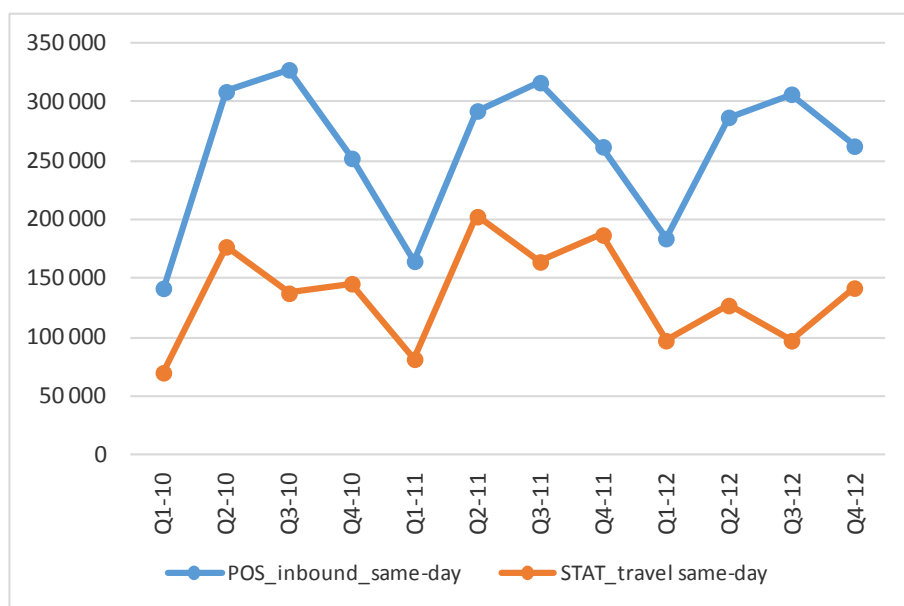

Figure 13. Quarterly inbound same-day trips (Finns to Estonia) in mobile positioning data when compared to the total number of same-day trips to Estonia in Finnish Demand Statistics (Source: Positium, Statistics Finland).

Table 14. Coherence indicators: Same-day inbound trips (Finns to Estonia).

Years	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
2010-2012	0.67	0.63	0.33	1.04
2010-2011	0.81	0.57	0.33	0.82
2012	0.25	0.76	0.60	1.04

The volume of same-day trips made by Finns to Estonia is nearly double in terms of mobile positioning data when compared to Finnish Demand Statistics. This over-coverage is systematic in each quarter. It's also interesting to observe that for the third quarter of each year, there is in fact a negative correlation. During the peak holiday season of July-September the number of same-day trips is very low according to Demand Statistics while mobile positioning shows a peak value. When looking at the respective Demand Statistics for overnight trips, it can be seen that they peak in the third quarter. This observation further supports the assumption that a part of overnight trips have 'leaked' into same-day trips in

mobile positioning data. This 'leakage' is clearly at its highest in the third quarter which includes the peak holiday months. An issue that is related to Demand Statistics is that same-day trips tend to be underreported due to recall bias. In other words, it's easier for the respondent to forget same-day visits than overnight visits. The recall bias can partly contribute to the differences between Demand Statistics and mobile positioning data. Nevertheless, the misclassification of overnight trips into same-day trips is more likely to be the primary reason for the over-coverage in mobile positioning data.

4.4.5. Inbound Nights Spent: Demand Statistics

The next example shows nights spent by Finns on overnight inbound trips to Estonia.
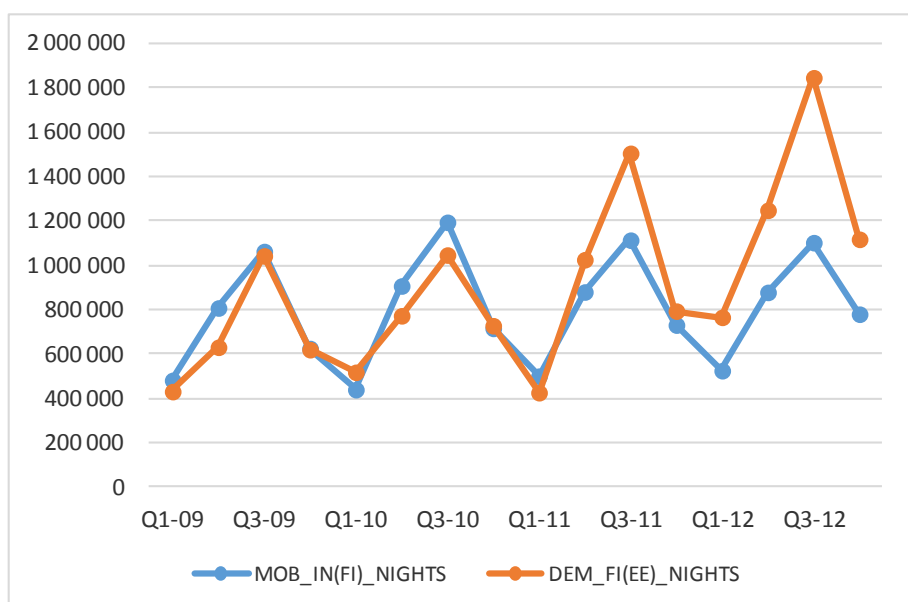
Figure 14. Nights spent on quarterly inbound overnight trips (Finns to Estonia) in mobile positioning data when compared to nights spent on overnight trips to Estonia in Finnish Demand Statistics (Source: Positium, Statistics Finland).

Table 15. Coherence indicators: Nights spent on overnight inbound trips (Finns to Estonia).

Years	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
2009-2012	0.79	-0.09	-0.50	0.25
2009-2011	0.87	0.01	-0.30	0.25
2012	0.98	-0.39	-0.50	-0.35

The results on nights spent are similar to the number of overnight trips presented in Section 4.4.3 where the reasons for the underestimation in mobile positioning data was explained. The magnitude of this underestimation is not as high for nights spent since the CC2 average for 2012 is -0.39 while it was -0.60 for the number of trips. Trips with several nights

spent will more likely be registered as an overnight trip in mobile positioning data than a trip with one night spent. The weight of these longer trips is higher for nights spent than in number of trips where the weight of each trip is one. This explains why the under-coverage is smaller (-0.39) for nights spent than for number of trips (-0.60). Still there is a significant under-coverage issue also for nights spent on overnight trips in the case of Finns to Estonia.

4.4.6. Inbound Overnight Trips: Supply Statistics - Finnish Arrivals

The focus will now be shifted to the other half of official tourism statistics: the supply statistics that are based upon the occupancy of Estonian accommodation establishments. This section focuses on Finns, the biggest group of inbound tourists to Estonia while the next section analyses all non-resident visitors to Estonia.

The following graphs present the monthly number of Finnish arrivals in Estonian accommodation establishments when compared to the number of Finnish overnight trips in mobile positioning data. Both absolute number of visitors (Figure 15) and an indexed raw data (Figure 16) (January 2012 = 100) are presented since the Estonian Supply Statistics are used in the process of estimating the data. The presumption is that visitors in supply statistics should always be less than overnight trips in mobile positioning data because a certain part of visitors stay in non-rented accommodation or rented accommodation below the threshold of official statistics.
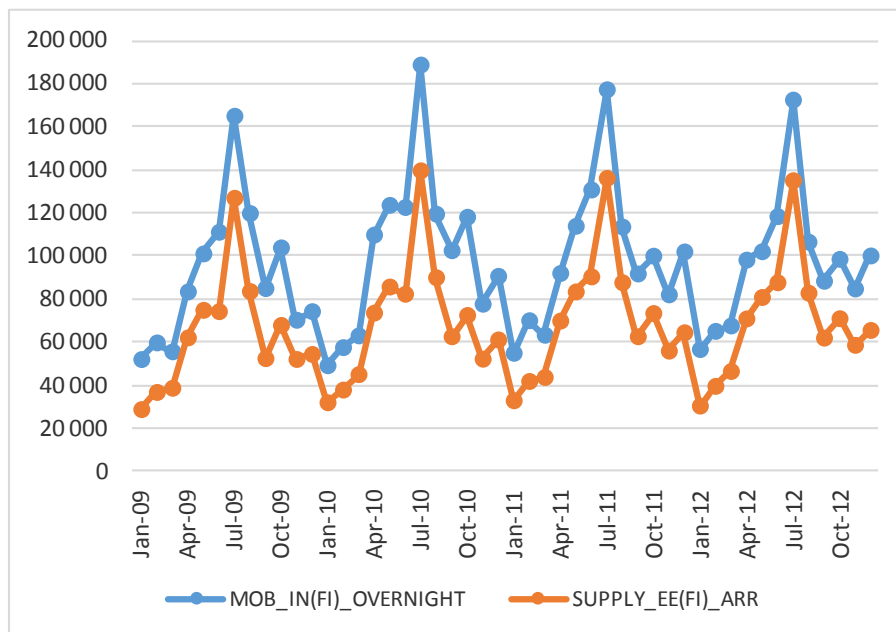

Figure 15. Monthly inbound overnight trips (Finns to Estonia) in mobile positioning data when compared to the number of Finnish arrivals in Estonian Supply Statistics (Source: Positium, Statistics Estonia).

Table 16. Coherence indicators: Supply Statistics arrivals (Finns to Estonia).

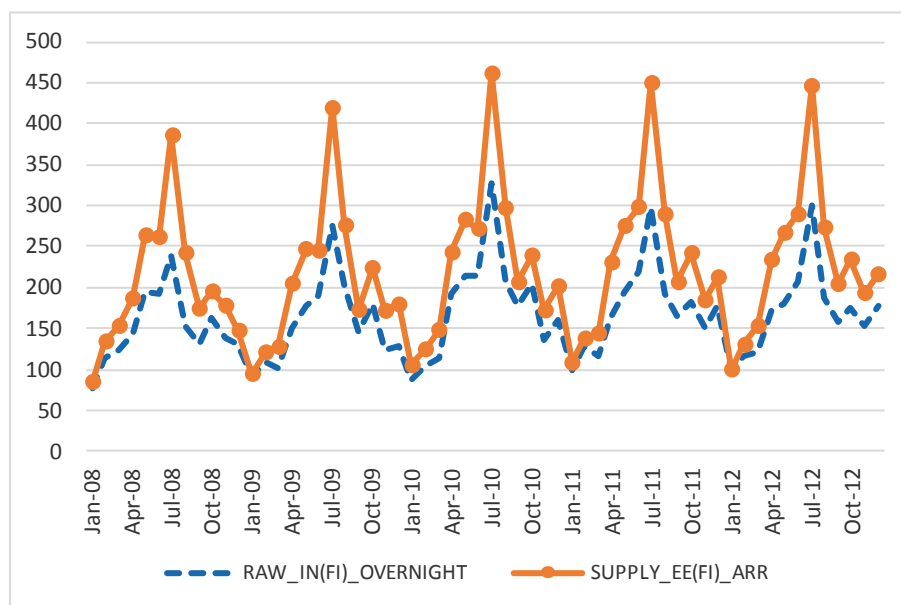| Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.98	0.37	0.27	0.48

Figure 16. Monthly inbound overnight trips (Finns to Estonia) raw mobile data index when compared to indexed Finnish arrivals in Estonian Supply Statistics; January 2012 = 100 (Source: Positium, Statistics Estonia).

Table 17. Coherence indicators: Supply Statistics arrivals (Finns to Estonia) - raw data.

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.98	N/A	N/A	N/A

Mobile positioning data provides a consistent estimate of the arrivals of Finnish tourists to Estonia. The correlation is almost perfect in both estimated and raw data. The over-coverage (0.39) in mobile positioning data is reasonable considering that many Finns stay in non-rented or non-registered accommodation which are excluded from Estonian Supply Statistics.

4.4.7. Inbound Overnight Trips: Supply Statistics - Nights Spent by Finns

The second main occupancy variable in Supply Statistics is nights spent. Nights spent in Supply Statistics measures the length of stay and similarly the duration of the trip can be measured in terms of nights in mobile positioning data. Figure 17 presents the nights spent by Finns in Estonia according to both Estonian Supply Statistics and inbound mobile positioning data.
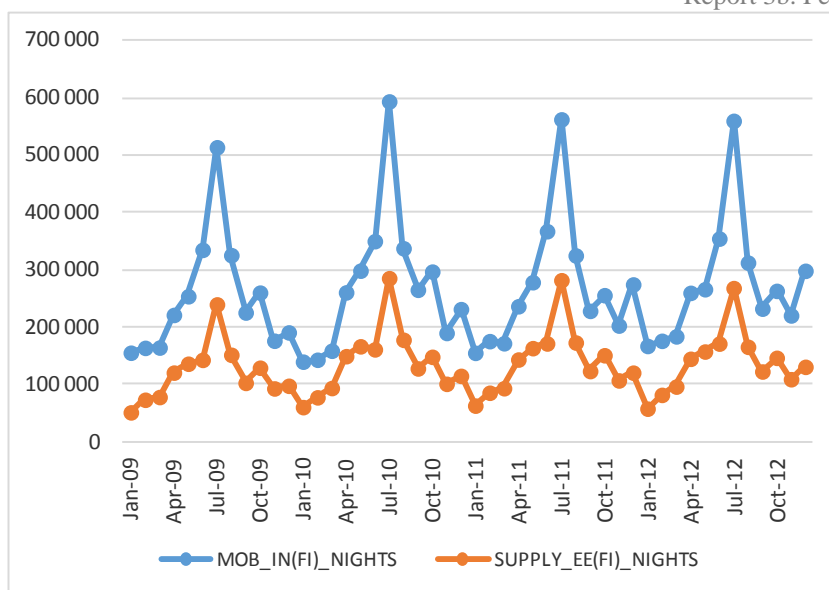
Figure 17. Monthly inbound nights (Finns to Estonia) spent in mobile positioning data when compared to the number of Finnish nights spent in Estonian Supply Statistics (Source: Positium, Statistics Estonia).

Table 18. Coherence indicators: Supply Statistics nights spent (Finns to Estonia)

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.97	0.67	0.49	1.01

Similarly to the number of Finnish arrivals, the correlation (0.97) is excellent for nights spent. However, the over-coverage (0.67) in mobile positioning data is even more substantial for nights spent than it is for the number of arrivals (0.37). Based upon Finnish Demand Statistics, a total of 85% of overnight trips to Estonia in 2012 were in fact short trips with between one to three nights spent. Roughly two thirds of the visitors on these trips used rented accommodation. Nevertheless, 15% of Finnish overnight trips to Estonia last more than three nights. The share of non-rented accommodation (such as time-share and holiday homes) increases as the duration of the trip increases, and consequently the nights spent on longer trips are covered less in Estonian Supply Statistics. This explains why the over-coverage is higher for nights spent than it is for arrivals. It can be considered as being a strength for mobile positioning data that longer trips are better covered in MDP than in they are in Supply Statistics.

4.4.8. Inbound Overnight Trips: Supply Statistics - Non-Resident Arrivals

This chapter provides the analysis for total inbound flows from the 27 EU countries to Estonia in terms of the number of trips. The comparisons are presented for the total EU 27 as well as the main countries of origin. The following graph (Figure 18) presents the total of

monthly EU-27 overnight inbound trips when compared to the number of EU-27 arrivals in Estonian Supply Statistics.
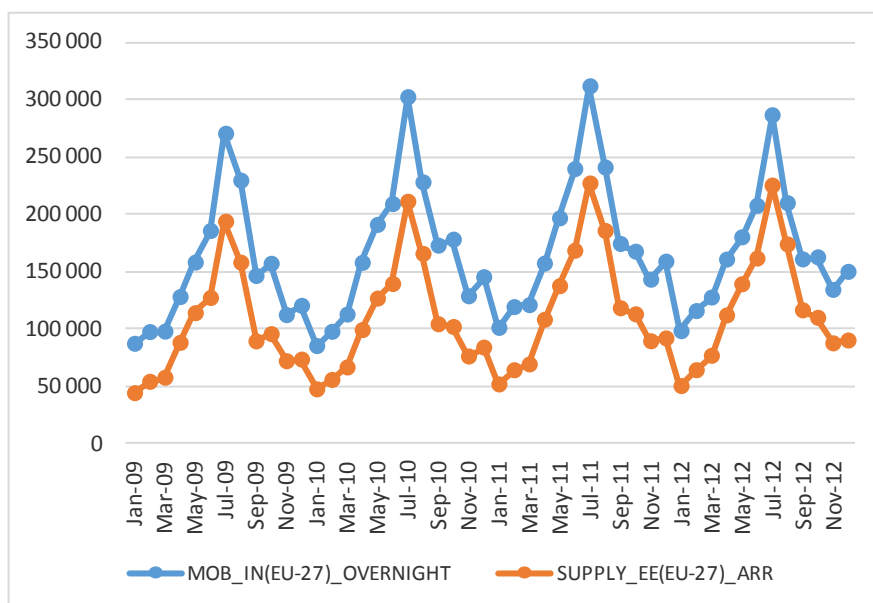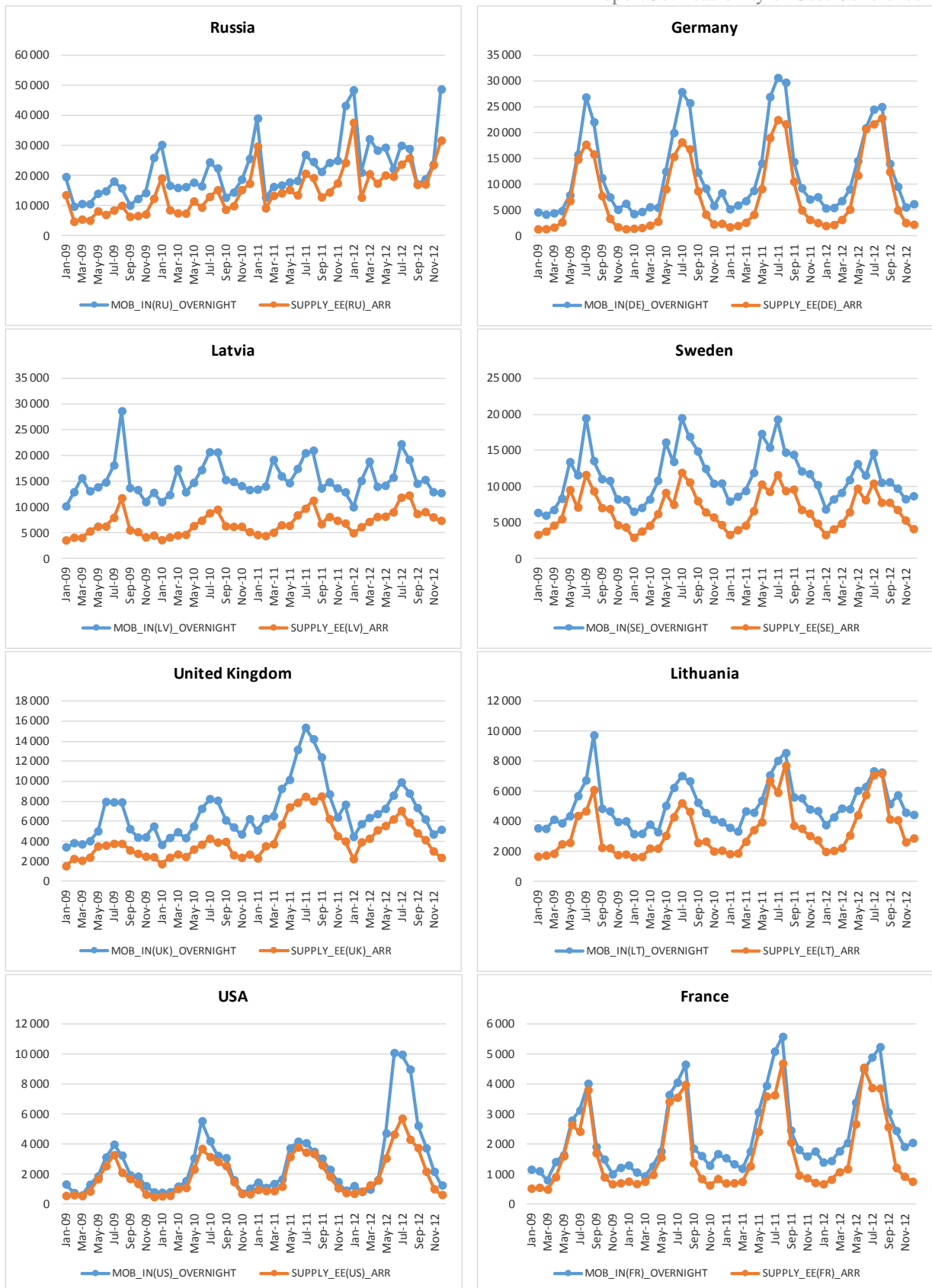

Figure 18. Monthly inbound EU-27 overnight trips in mobile positioning data when compared to the number of EU-27 arrivals in Estonian Supply Statistics (Source: Positium, Statistics Estonia).

Table 19. Coherence indicators: Supply Statistics arrivals (EU-27 to Estonia).

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.98	0.44	0.19	0.66

The consistency of overnight inbound trips for the total number of non-resident visitors is very good since the correlation is almost equal to one. The over-coverage in mobile positioning data is a little higher for EU 27 countries than it is for Finns alone. For every registered arrival in the Supply Statistics there were 0.44 other overnight trips in inbound mobile positioning data.

Figure 19 below breaks down the overnight inbound trips based upon main countries of origin.

Figure 19. Monthly inbound overnight trips in mobile positioning data when compared to the number arrivals by country of residence in Estonian Supply Statistics (Source: Positium, Statistics Estonia).

Table 20. Coherence indicators: Supply Statistics arrivals (non-residents to Estonia by country of residence).

Destination	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
Russia	0.92	0.42	0.01	0.75
Germany	0.97	0.62	0.01	1.31
Latvia	0.74	0.79	0.44	1.18
Sweden	0.94	0.53	0.30	0.83
United Kingdom	0.93	0.54	0.27	0.80
Lithuania	0.92	0.47	0.01	0.77
USA	0.93	0.31	-0.25	0.81
France	0.97	0.40	-0.01	0.93

Mobile positioning data provides a very consistent estimate on the number of overnight inbound trips on a country level. It should be remembered here that Estonian Supply Statistics have been used in the process of estimating the number of trips in mobile positioning data, so good results are expected. For each arrival in Supply Statistics there are between 0.31 to 0.79 additional overnight trips in mobile positioning data, depending on the country of origin.

While the general levels of consistency are good, there are certain differences depending on the country. This is especially the case for Latvia, a country neighbouring Estonia and clearly stands out with a seemingly-weak level of consistency (Pearson's r = 0.74). Visitors from neighbouring countries tend to make more same-day trips and overnight visits in non-rented or below threshold accommodation. When looking at Latvian visitors, there is a peak of Latvian overnight trips to Estonia in mobile positioning data between January and March while no such peak exists in Supply Statistics. One of the main reasons for this peak is ice fishing. It's generally known that in the first quarter of each year many Latvians travel to Estonia on ice fishing trips. Ice fishing is popular in Latvia, and Estonia provides a multitude of lakes, with Lake Peipus being the fifth largest lake in Europe. On these fishing trips the Latvians predominantly stay away from registered accommodation and therefore are not reflected in Supply Statistics. The example of the Latvians demonstrates that, while at first glance it may appear that mobile positioning data is not consistent, a closer examination in fact reveals a strength of mobile positioning data. In the case of Latvians travelling to Estonia, mobile positioning data has in fact captured a real tourism flow which is outside the scope of official tourism statistics.

4.4.9. Inbound Overnight Trips: Supply Statistics - Non-Resident Nights Spent

As a follow-up to the previous example, Figure 20 shows the inbound nights spent in Estonia by residents of EU 27 countries.
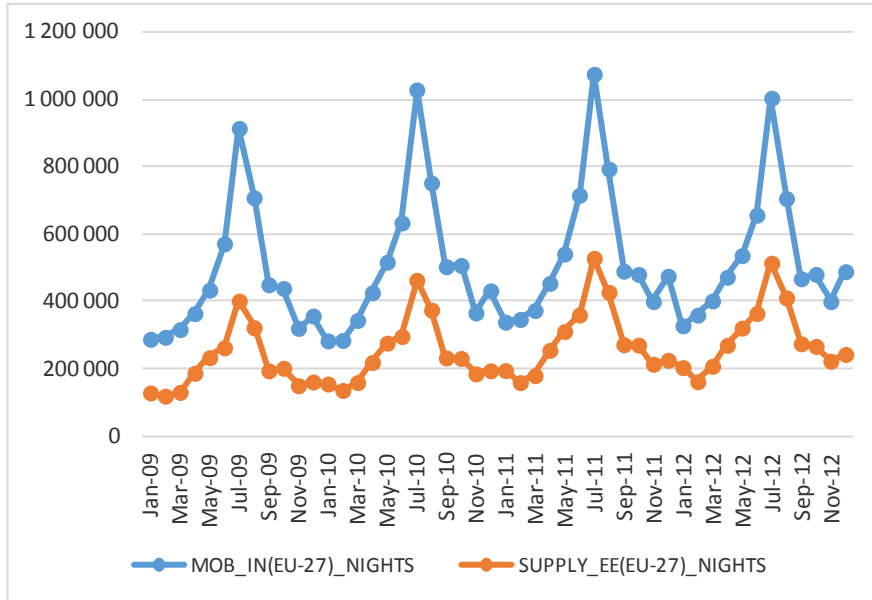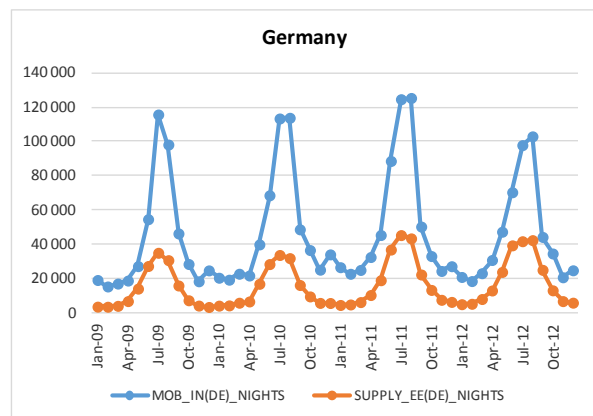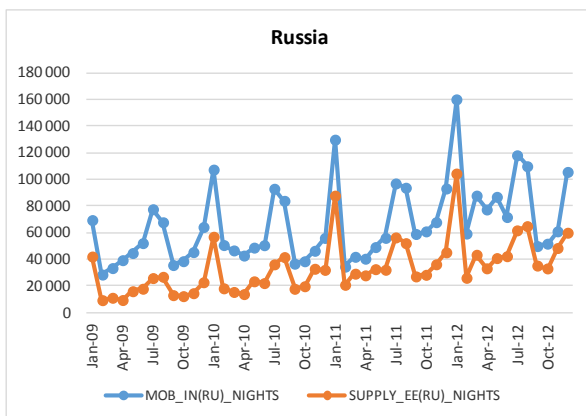

Figure 20. Monthly nights spent on inbound overnight trips in mobile positioning data when compared to the number of EU-27 nights spent in Estonian Supply Statistics (Source: Positium, Statistics Estonia).

Table 21. Coherence indicators: Supply Statistics nights spent (EU-27 to Estonia).

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.96	0.66	0.47	0.85

The results are very similar for EU 27 countries as they were for Finns alone. Over-coverage of nights spent (0.66) is greater than for arrivals (0.44). This kind of result is expected, since the share of non-rented or non-registered accommodation is typically higher on longer trips, as we have seen in the case of Finns to Estonia.
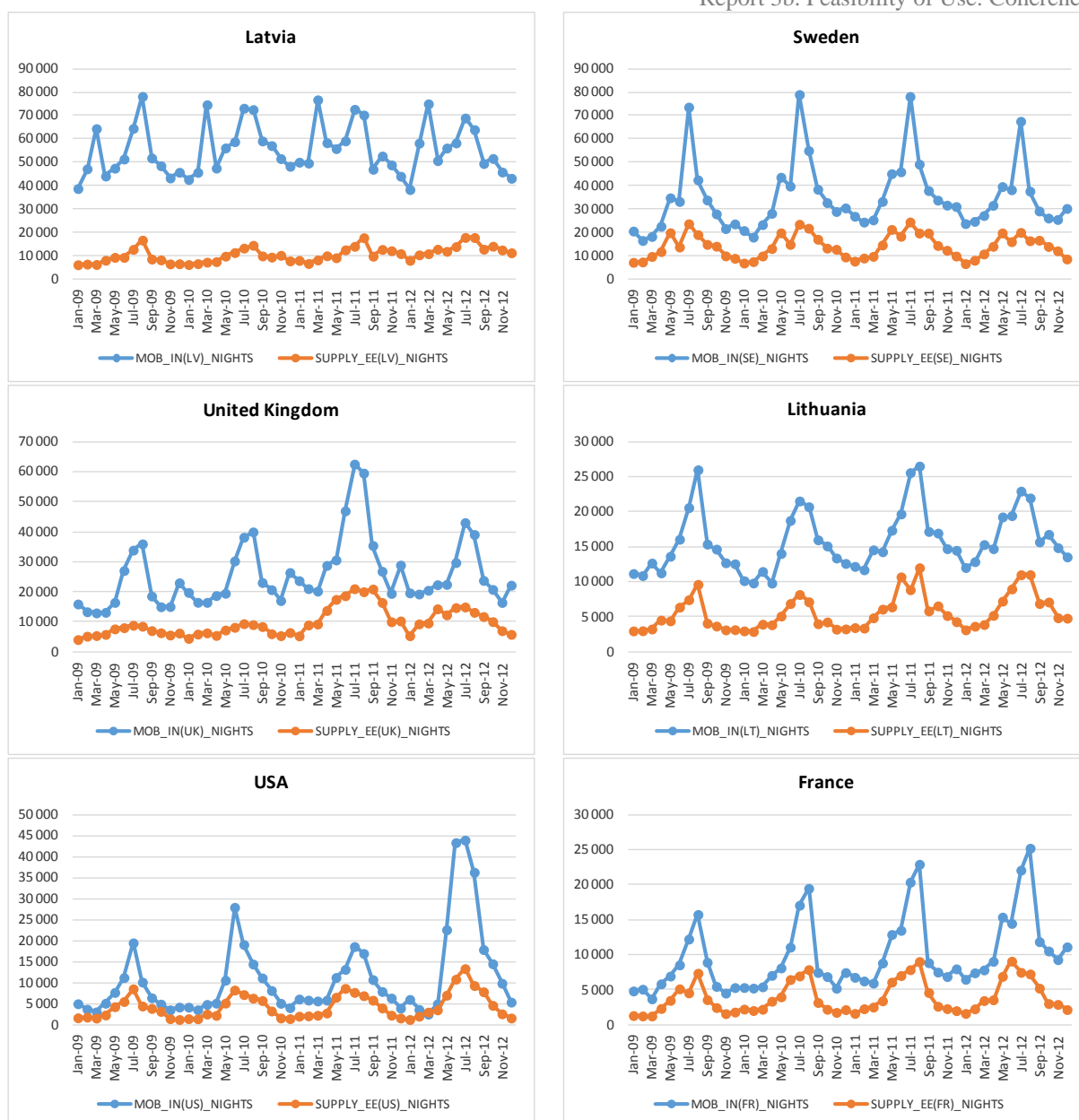
Figure 21. Monthly nights spent on inbound overnight trips by country in mobile positioning data when compared to the number of nights spent in Estonian Supply Statistics (Source: Positium, Statistics Estonia).

Table 22. Coherence indicators: Supply Statistics nights spent (non-residents to Estonia by country of residence).

Destination	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
Russia	0.93	0.70	0.23	1.23
Germany	0.94	1.05	0.56	1.54
Latvia	0.74	1.37	1.13	1.65
Sweden	0.86	0.85	0.55	1.14
United Kingdom	0.91	0.91	0.44	1.27
Lithuania	0.91	0.99	0.59	1.23

Destination	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
USA	0.91	0.81	-0.17	1.33
France	0.90	0.90	0.46	1.35

The results for nights spent are similar to those for the number of trips. The Pearson's r figures are slightly smaller for most countries and over-coverage in mobile positioning data is significantly higher. Additionally, the case of Latvian ice fishing trips is even more apparent when looking at nights spent between January and March each year. During those months the CC2 average for Latvians is 1.51 while for other months it's 1.33.

The conclusion for nights spent is that mobile positioning provides a consistent and comprehensive estimate for nights spent by non-residents. Supply Statistics are limited to those inbound tourists who stay at rented accommodation in above-threshold establishments. The strength of mobile positioning data is in the fact that it captures the total tourism flows. This is even more apparent for nights spent than it is for the number of trips.

4.4.10. Inbound Overnight Trips: Regional Supply Statistics

One of the main strengths of using mobile positioning data is the spatial dimension. For tourism statistics this enables the analysis of tourism flows by sub-regions within a country. The Estonian inbound mobile positioning data provides the breakdown of tourism trips by county of destination. Figure 22 compares the number of inbound trips to six main Estonian counties against the arrivals shown in Supply Statistics. These six counties make up more than 95% of arrivals in Supply Statistics.
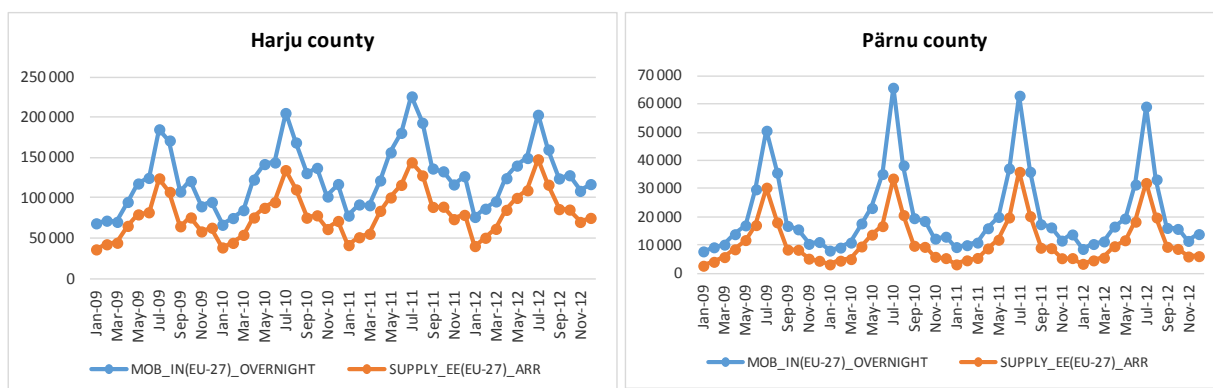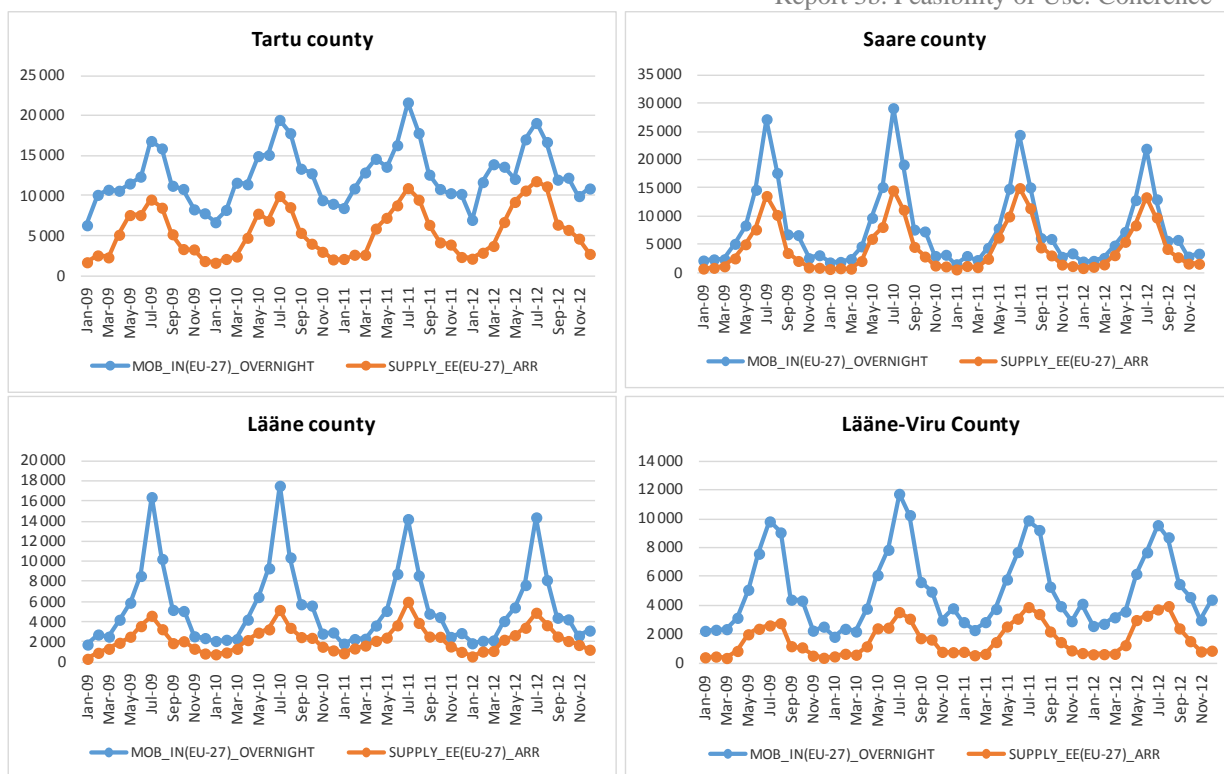
Figure 22. Monthly EU-27 inbound overnight trips by Estonian county in mobile positioning data when compared to the number of EU-27 arrivals in Estonian Supply Statistics (Source: Positium, Statistics Estonia).

Table 23. Coherence indicators: Supply Statistics arrivals by county (EU-27 countries).

County	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
Harju (Tallinn)	0.98	0.45	0.31	0.62
Pärnu	0.99	0.65	0.36	0.98
Tartu	0.88	0.86	0.27	1.32
Saare	0.97	0.66	0.23	1.13
Lääne	0.94	0.77	0.34	1.37
Lääne-Viru	0.94	1.08	0.70	1.49

The consistency of mobile positioning data is also at its best at the county level. The best Pearson's r values are achieved in the top two tourism counties of Harju (Tallinn) and Pärnu. The over-coverage of mobile positioning data varies depending on county.

4.4.11. Inbound Overnight Trips: Comparison of 2012 Reference Statistics

This chapter combines the demand and supply statistics available for 2012 - the first reference year of Regulation 692/2011 - concerning inbound overnight trips to Estonia. Based on the micro-data transmissions made by Member States, it was possible for Eurostat to

provide the number of inbound overnight trips and nights spent in Estonia from 23 Member States. At the time of writing this report, the data was available for all other Member States except Denmark, United Kingdom, Poland and Sweden. In demand statistics, 87% of inbound trips into Estonia with at least one overnight stay were made by Finns, 6% by Latvians, 2% by Lithuanians and remaining 6% by residents of other Member States for which data was available.

Figure 23 presents the comparison of 2012 inbound overnight trips in mobile positioning data compared to demand statistics of Member States and supply statistics of Estonia. The data is indexed so that number of trips in 2012 mobile positioning data is set to 100.
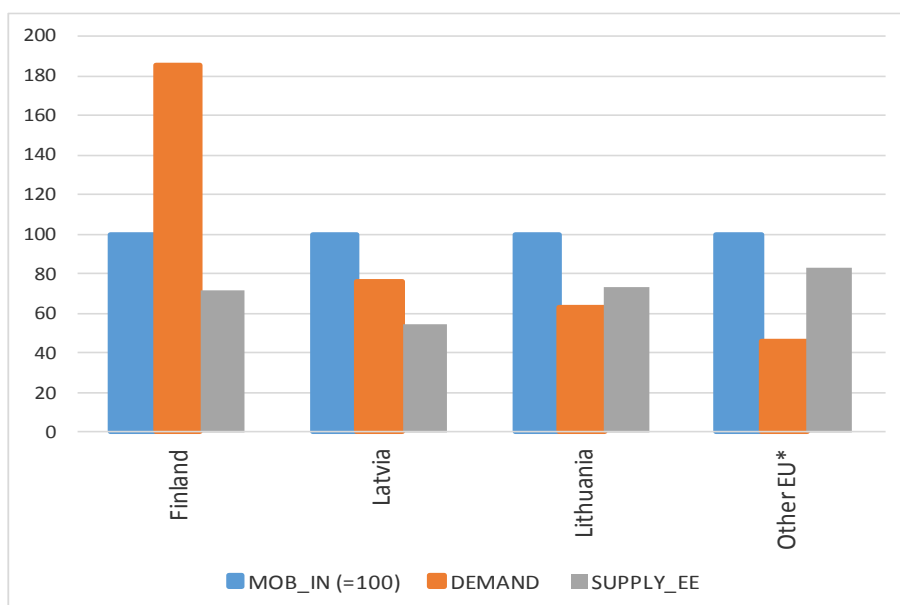

Figure 23. Inbound overnight trips to Estonia in 2012 in mobile positioning data (2012=100), demand statistics and supply statistics (Source: Positium, Eurostat, Statistics Estonia) *excluding Denmark, United Kingdom, Poland and Sweden.

The reasons for the deviations compared to Finnish demand statistics have been analysed in the previous chapters. For the other EU Member States the situation seems to be opposite: number of trips in mobile positioning exceeds the trips in demand statistics. For other Member States even the Estonian supply statistics exceed the trips in demand statistics. This is rather surprising, considering that demand statistics should include both rented and non-rented accommodation while supply statistics is concerned with rented accommodation only. Figure 24 presents the same analysis but based on nights spent on inbound trips to Estonia.
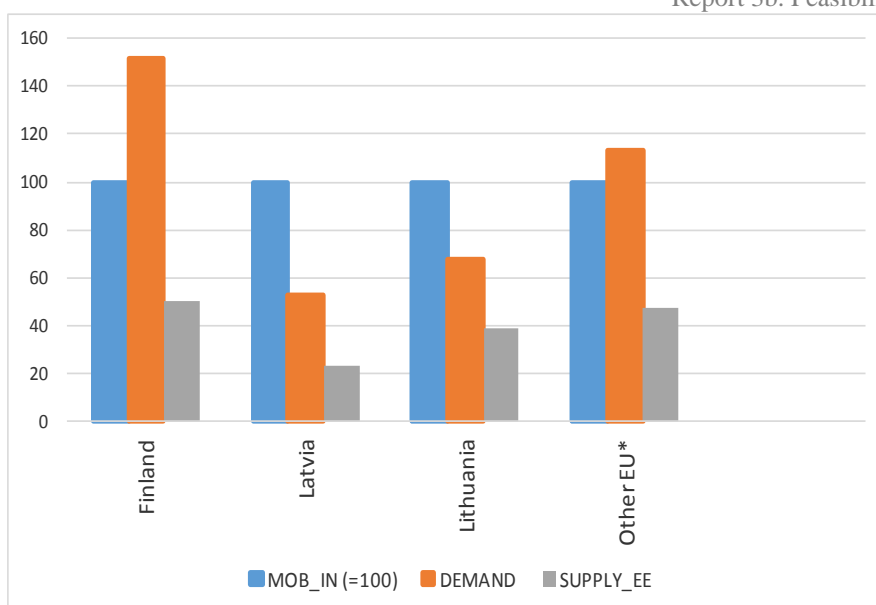
Figure 24. Nights spent on inbound overnight trips to Estonia in 2012 in mobile positioning data (2012=100), demand statistics and supply statistics (Source: Positium, Eurostat, Statistics Estonia) *excluding Denmark, United Kingdom, Poland and Sweden.

Again, the nights spent in Finnish demand statistics exceed mobile positioning due to reasons discussed in earlier chapters. For other EU member states the results are similar to Finland: nights spent according to demand statistics exceed both mobile positioning and supply statistics.

Due to the random variation involved especially with surveys used for demand statistics, it's difficult to make general conclusions that would apply to all countries. In general, supply statistics cover only rented accommodation and are exceeded by both demand statistics and mobile positioning data. Due to methodological differences in mobile positioning data and demand statistics, the results differ depending on country.

4.5. Outbound Tourism

This chapter provides an analysis of outbound mobile positioning data that is based upon the roaming of Estonian mobile subscribers in foreign cellular networks. The mirror reference statistics used include the following:

1. Estonian Demand Statistics: outbound trips made by Estonians
2. Finnish Supply Statistics: Estonians to Finland
3. EU-27 Supply Statistics: Estonians to EU-27 countries
4. Finnish Border Interview: survey of Estonians upon their departure from Finland

4.5.1. Total Outbound Trips: Demand Statistics (EU-27)

The Estonian Demand Statistics have been compiled by Statistics Estonia according to Council Directive 95/57/EC (which is valid until the end of 2011) and Regulation 692/2011. Figure 25 compares the quarterly outbound overnight trips to EU-27 countries in Estonian mobile positioning data to overnight trips in the Estonian Demand Statistics.
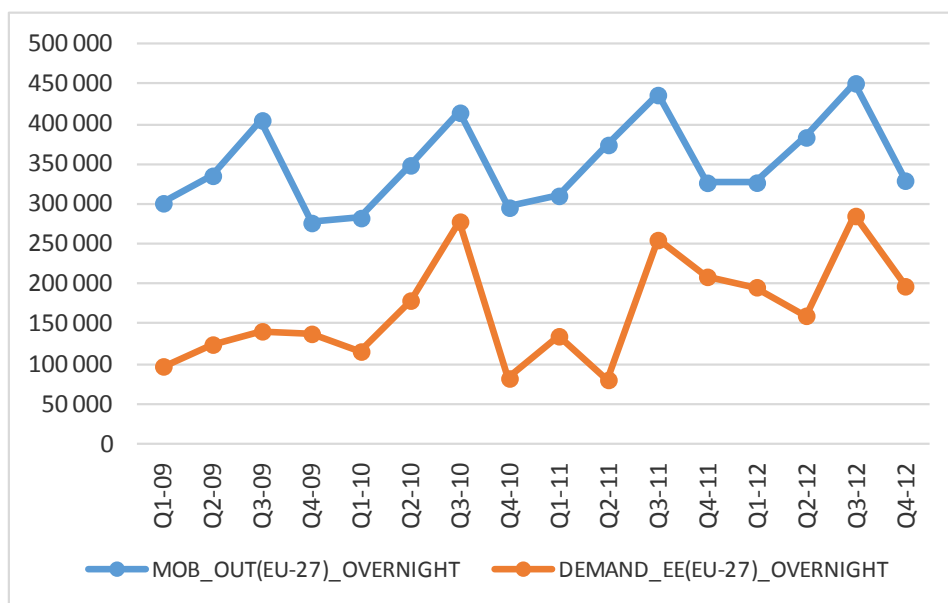

Figure 25. Quarterly outbound overnight trips (Estonians to EU-27) in mobile positioning data when compared to the number of overnight trips to EU-27 countries in Estonian Demand Statistics (Source: Positium, Statistics Estonia).

Demand statistics are obtained from the survey data, and the quarterly estimates that are presented in Figure 25 have coefficient of variation around 7%.

Table 24. Coherence indicators: Outbound overnight trips in Demand Statistics (Estonians to EU-27).

Years	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
2009-2012	0.67	0.74	0.39	1.29
2010-2011	0.63	0.80	0.39	1.29
2012	0.70	0.57	0.45	0.82

The consistency of mobile positioning data is rather weak when compared to Demand Statistics. In addition, it seems that the Estonian Demand Statistics constantly underestimates the number of outbound overnight trips. There is a slight improvement both in consistency and coverage for 2012, but overall the coherence levels are not very good. It's difficult to reach a conclusion based upon these reference statistics. Mobile positioning data seems to have a stable quarterly fluctuation and a slightly upward trend from year to year, while Demand Statistics seem to be affected by random variations.

There can be many reasons behind the differences that have been observed between the two estimates. First of all, the problem of recall bias in the survey is a known source of error. It is quite understandable that people may not remember all of their trips, which were completed over three months ago. It is believed that it influences more domestic trips but it probably leads also to an underestimation of the number of outbound trips.

In addition, the methodology change which was introduced at the beginning of 2010 in order to reduce other sources of measurement bias leads to incomparability over time in the demand statistics series. In addition, one needs to take into account the variability of the survey estimates, although the 95% confidence interval of the series on Figure 25 does not cover estimates that are based upon mobile positioning data.

In conclusion, it is possible to state that demand statistics and mobile positioning statistics provide different results, and the difference is not only due to the random nature of the demand statistics. It is likely that demand statistics are biased, but the size of the bias is unknown.

4.5.2. Total Outbound Trips: Finnish Border Interview

In the year 2012 Finland was the main destination for more than 25% of Estonian outbound trips. Statistics Finland conducted Border Interview Surveys up to the end of 2012. The interviews were carried out at main border crossing locations in Finland, such as Helsinki Airport and at the ferry terminal, through which most Estonian visitors arrive. Figure 26 displays the quarterly number of Estonian visitors (both same-day and overnight) to Finland in Estonian outbound mobile positioning data and in the Finnish Border Interview.
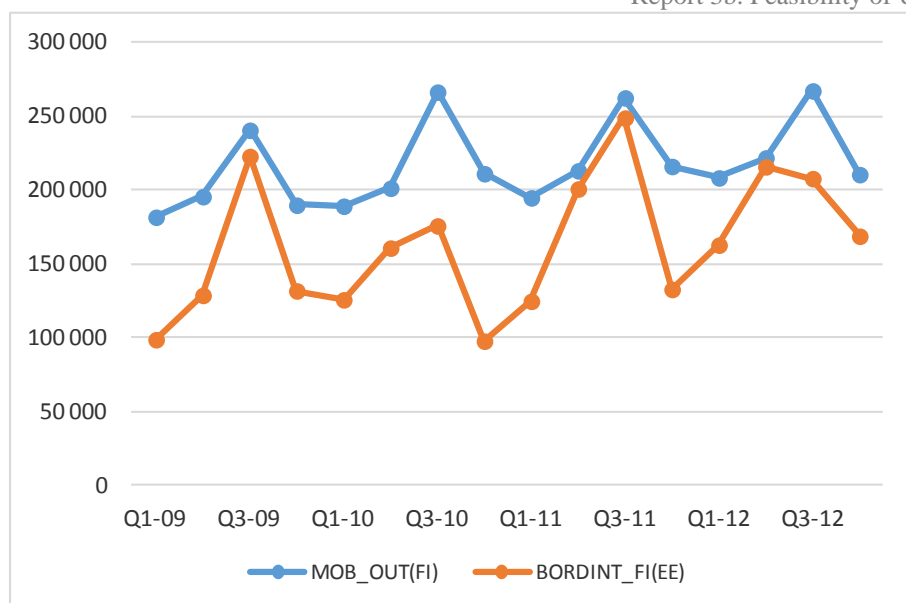
Figure 26. Quarterly outbound trips (Estonians to Finland) in mobile positioning data when compared to the number of Estonian visitors in the Finnish Border Interview Survey (Source: Positium, Statistics Finland).

Table 25. Coherence indicators: Finnish Border Interview (Estonians to Finland).

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.75	0.31	0.03	0.73

The consistency levels (0.75) are average, and both data sources have a slightly upward trend. The over-coverage in mobile positioning data varies from minor (0.03) to significant (0.73). Part of this variation is certainly due to the sampling error in the Finnish Border Interview. Regardless of the sampling error, there seems to be some real over-coverage in mobile positioning data as the number of visitors is higher each quarter (CC2 average = 0.31). There are at least two potential factors explaining this over-coverage.

Firstly, the Border Interview survey is carried out with persons who have a permanent residence outside of Finland. However, many Estonians commute to Finland and may report Finland as their permanent residence if, for example, they are in Finland during weekdays and return to Estonia only for weekends. In such cases, they will not appear in the Border Interview but may still have an Estonian mobile subscription which would appear in Estonian outbound mobile positioning data. The four year average CC2 value for the 'high-commuting' winter quarters one and four is 0.42, while for 'low-commuting' summer quarters two and three the average CC2 is only 0.19. This would indicate that commuting in fact plays a part in the over-coverage when compared to the Border Interview Survey, and the impact is bigger for quarters one and four where there is more commuting. The second factor that contributes to the over-coverage is that of Estonian transit air passengers. It's estimated that more than 50,000 Estonian air transit passengers travel via Helsinki Airport each year. Transit

passengers are not included in the scope of the Border Interview but they may appear in Estonian outbound mobile positioning data in case in which they use their phone in the transit area of Helsinki Airport.

4.5.3. Overnight Outbound Trips: Supply Statistics - Arrivals

Regulation (EU) 692/2011 obliges the Member States to separately report non-resident arrivals and nights that are spent by the residents of each member state. Arrivals for Estonians in EU countries were extracted from data tables that are maintained by Eurostat. The full time series for 2009-2012 are available for these eighteen EU countries: AT, BE, BG, CZ, DK, DE, ES, FR, FI, IT, LV, LT, LU, HU, PL, PT, SE, SK. Only these eighteen countries are included in Figure 27.
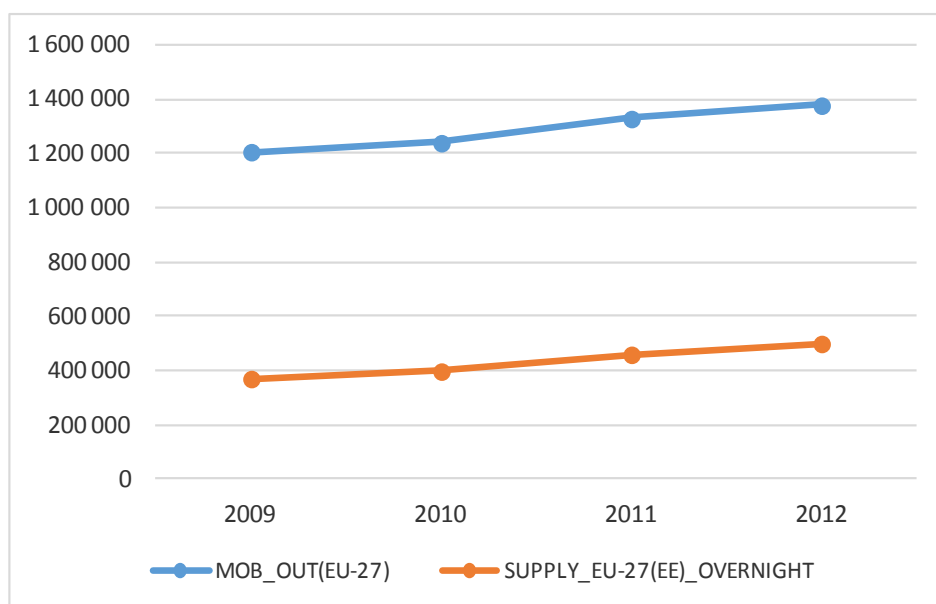

Figure 27. Annual outbound trips in mobile positioning data when compared to the number of arrivals in the Supply Statistics of eighteen EU countries (Source: Positium, Eurostat).

Table 26. Coherence indicators: Arrivals in Supply Statistics (18 EU countries)

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
1.00	1.00	0.94	1-06

The consistency of mobile positioning data against Supply Statistics is perfect (Pearson's r = 1.00). For every arrival in Supply Statistics there are nearly three trips in mobile positioning data. The difference is clearly more than in the inbound overnight trips to Estonia. This indicates that when Estonians go abroad the share of paid accommodation is less than for inbound trips to Estonia. Figure 28 breaks down the arrivals by main destinations of Estonian outbound trips.

Figure 28. Outbound overnight trips of Estonians to the main destinations in mobile positioning data when compared to arrivals for Estonians in Supply statistics (Source: Positium, Statistics Finland, Eurostat).

Table 27. Coherence indicators: Supply Statistics arrivals (for Estonians to main destinations).

Destination	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
Finland	0.89	1.41	1.22	1.58
Latvia	0.98	0.84	0.74	0.95
Sweden	-0.83	1.50	1.45	1.63
Lithuania	0.88	0.62	0.51	0.71
Germany	0.99	0.8	0.77	0.84
France	1.00	1.33	0.98	1.54
Poland	0.77	0.44	0.33	0.54
United Kingdom	0.86	0.55	-0.05	1.34

There are some real differences in over-coverage depending on destination country. Finland is the main outbound destination for Estonians but only a small share of the Estonian trips are reflected in Finnish Supply statistics. Similarly to the Finnish Border Interview, the issue that is related to commuting can also be seen here, as most commuters do not stay in accommodation establishments. The over-coverage is much less in Latvia and Lithuania, which are likely to receive less commuters from Estonia.

4.5.4. Same-Day Outbound Trips

According to Regulation (EU) 692/2011, the first reference year for same-day outbound trips is 2014. Due to lack of reference data, the assessment cannot be made for outbound same-day trips at this stage.

4.6. Domestic Tourism

In this chapter the coherence of Estonian mobile positioning data and official tourism statistics on domestic tourism are compared. Available data sources are: Estonian mobile positioning data about both same-day and overnight trips, official statistics on tourism demand (overnight domestic trips of Estonian residents from Tourism survey among individuals) and tourism supply (accommodation statistics collected from businesses). Official statistics on same-day domestic trips of Estonian residents is not available.

Domestic mobile positioning data cover trips of domestic subscribers outside their usual environment (but not abroad) that are not considered transit trips. Usual environment is calculated based upon county-level (administrative units, LAU 1). The usual environment is based upon following criteria: Usual environment of the subscriber comprises of the second

level administrative units (LAU 1, county) where a subscriber has been present at least once a week for at least ten weeks during the ninety day retrospective time window. Identified usual environments' time influence is extended for 90 days prior and following the identification day. Two types of estimates are available, raw estimates and estimates corrected for accommodation statistics. For raw estimates only indices are available (January 2012 = 100).

Estimates corrected for accommodation statistics are available with correct level values. Corrected domestic mobile positioning data is calculated for the whole of Estonia and fifteen counties. The results reflect the whole population.

4.6.1. Total Number of Trips

The total number of trips (overnight and same-day) for Estonia is available on the basis of mobile positioning data only.
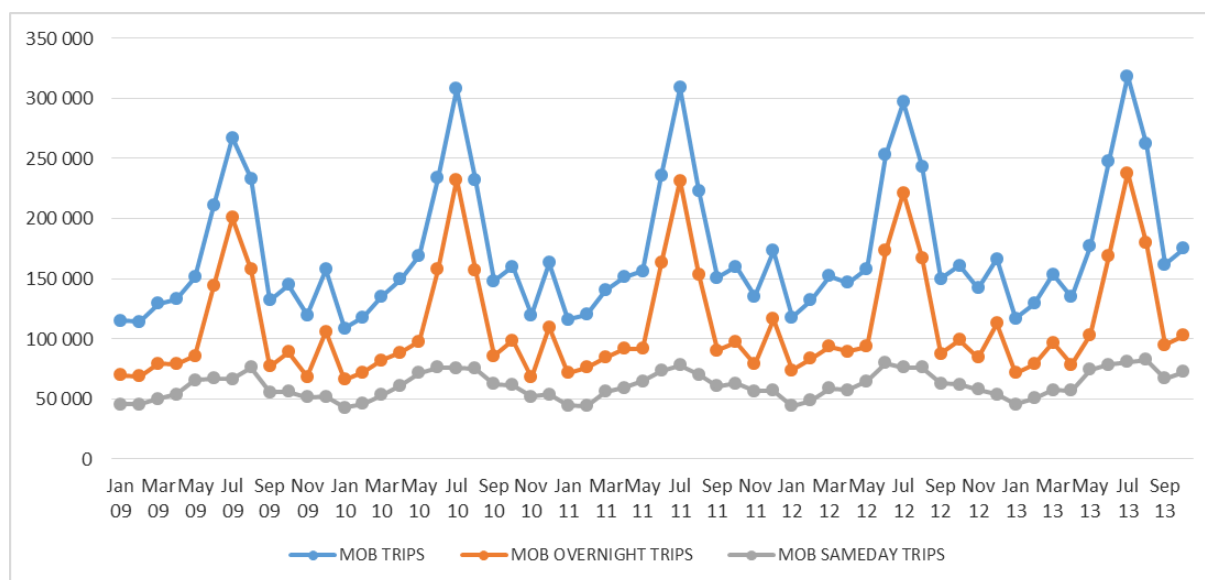

Figure 29. Domestic trips in mobile positioning data (Source: Positium).

Time series of the total number of domestic trips shows typical seasonality for Estonian residents. Most Estonian residents have holiday during summer months and trips tend to be longer then.
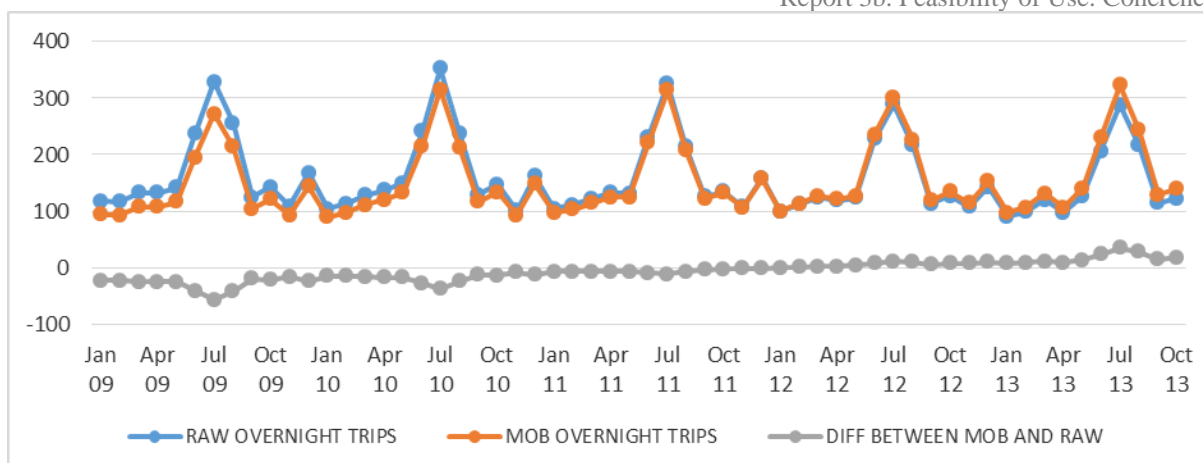
Figure 30. Monthly Estonian domestic overnight trips in mobile positioning data which is un-corrected when compared to data that has been corrected for accommodation statistics, January 2012=100 (Source: Positium).
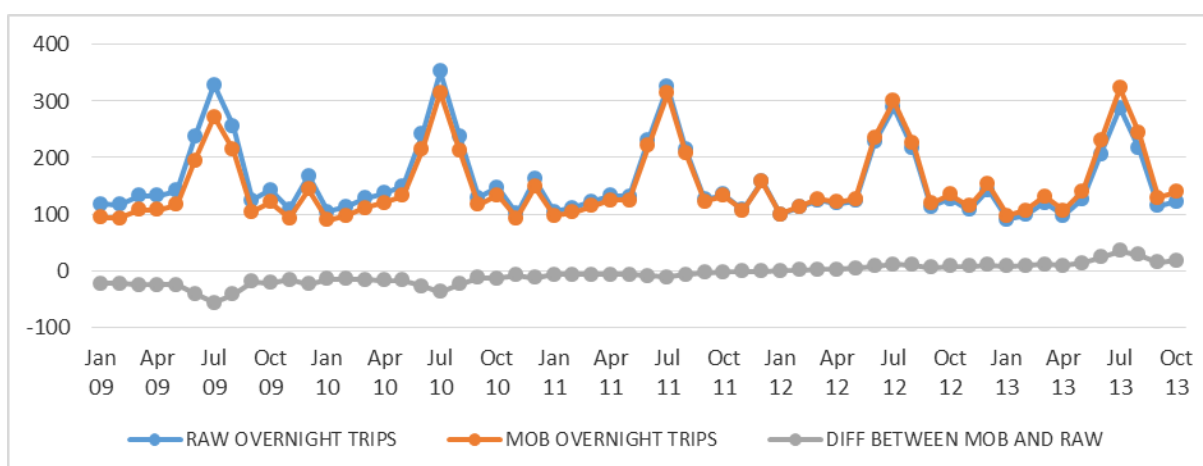

Figure 31. Monthly Estonian domestic nights spent in overnight trips in mobile positioning data that is un-corrected when compared to data that has been corrected for accommodation statistics, January 2012=100 (Source: Positium).

Figure 30 and Figure 31 show higher corrections of mobile positioning data for years 2009 and 2010 and for the summer months. As there is a very small difference between the corrected and uncorrected estimates, then in the following part of this section the corrected estimates and not indexes of uncorrected estimates are used.

4.6.2. Overnight Trips: Supply Statistics

For coherence analysis of overnight trips monthly domestic mobile positioning data (corrected absolute value, overnight) and accommodation statistics of residents are compared.
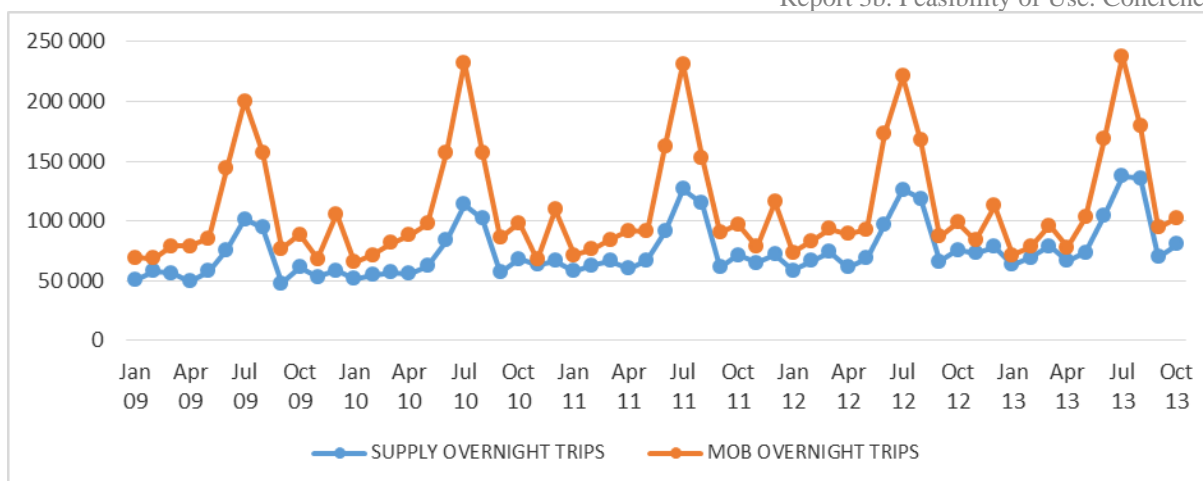
Figure 32. Monthly Estonian domestic corrected overnight trips in mobile positioning data when compared to accommodated domestic tourists in accommodation statistics (Source: Positium, Statistics Estonia).
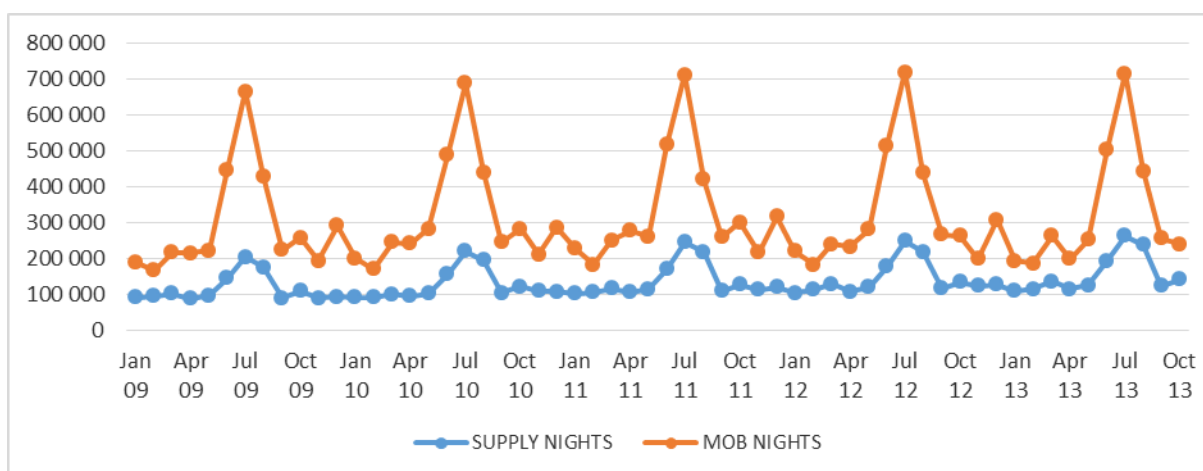

Figure 33. Monthly Estonian domestic corrected nights spent in mobile positioning data when compared to accommodated domestic tourists' nights spent in accommodation statistics (Source: Positium, Statistics Estonia).

Table 28. Coherence indicators: the number of overnight domestic trips and nights spent in domestic trip according to accommodation statistics and mobile positioning data (Jan. 2009-Oct. 2013)

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.98 (overnight trips)	N/A	N/A	N/A
0.91 (nights spent)	N/A	N/A	N/A

Difference between the mobile positioning and accommodation data is largest during the high season of tourism in summer and the difference is smaller when there is the low season of tourism (autumn-winter). The difference is explained by the fact that mobile positioning data cover all size rented and non-rented accommodation. At the same time, Estonian accommodation statistics include rented accommodation establishments with five and more beds. Small establishments work often seasonally and mostly during summer months. Using non-rented accommodation is also more common in summer when relatives and friends are visiting and free accommodation in camp sites is used. The bigger difference

can be noticed also for December data, which can be explained by the more frequent use of non-rented accommodation by the visiting relatives and friends during Christmas holidays.

All in all, the trends shown in the mobile positioning estimates coincide with the expectations and follow largely the trends in the accommodation statistics. The levels of the mobile positioning estimates are larger than expected but the accommodation statistics is not very good reference for the evaluation of the levels due to the different scope of observations.
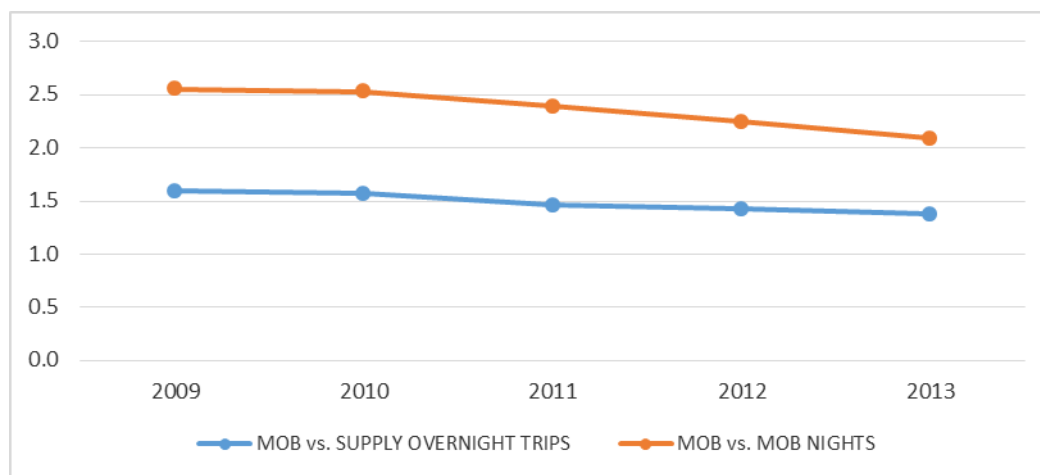

Figure 34. Annual ratio of mobile positioning and accommodation data (Source: Positium, Statistics Estonia).

Figure 34 shows that the annual ratio between mobile positioning and accommodation data is in decreasing trend indicating that the difference between the two sources is remaining largely the same or is slowly decreasing.

4.6.3. Overnight Trips: Demand Statistics

Following the coherence of domestic overnight trips in mobile positioning data and the tourism demand data for overnight domestic trips of Estonian residents is studied. Quarterly tourism demand statistics are available.
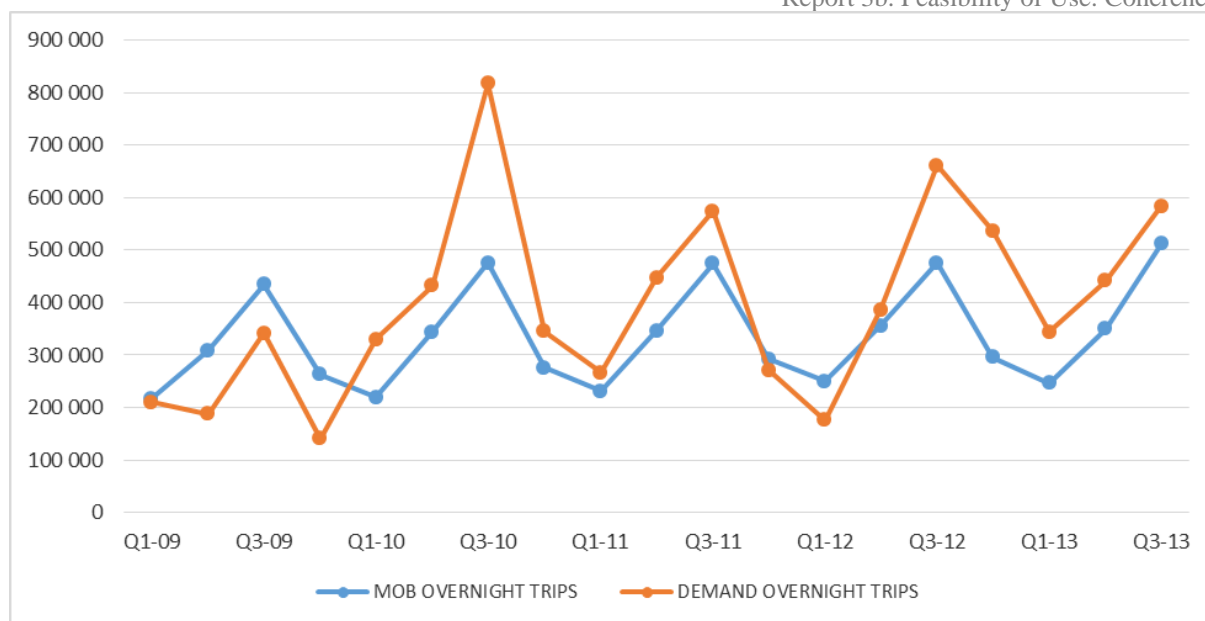
Figure 35. Quarterly Estonian domestic corrected overnight trips in mobile positioning data when compared to overnight trips made by Estonian residents in tourism demand statistics (Source: Positium, Statistics Estonia).

Table 29. Coherence indicators: the number of overnight domestic trips according to demand statistics and mobile positioning data (first quarter 2009 - third quarter 2013)

Pearson's r	CC2 average	CC2 minimum	CC2 maximum
0.79 (overnight trips)	N/A	N/A	N/A

Those estimates that are based upon the mobile positioning data in Figure 34 are based upon the conservative approach to usual environment i.e. usual environment used is at a county level.

The difference in the absolute value of overnight trips by mobile positioning data and tourism demand statistics is smaller. The mean of the absolute deviance between the two statistics is 105,000 overnight trips for the period 2009-2013.Taking into account that the estimates of domestic trips in the demand survey have coefficient of variation around 5.5% which leads to 95% confidence interval that does not include most of the mobile positioning estimates. However, it can be seen that the estimates of tourism demand statistics are very much influenced by methodology of survey. The quality of sample survey is also influenced by the sample size and response rate. The methodology of Estonian tourism demand survey changed in 2010 and again in the fourth quarter of 2012.

4.6.4. Overnight Trips: Rented and Non-rented Accommodation

According to the European 2012 tourism demand survey, the number of nights spent by accommodation type vary a great deal by different countries as shown in Figure 36. The

ratio of nights spent in rented accommodation differs almost five times between highest and lowest ratio. The highest ratio was 73%, which was reported in Austria, while the lowest ratio was 15%, which was found in Latvia.
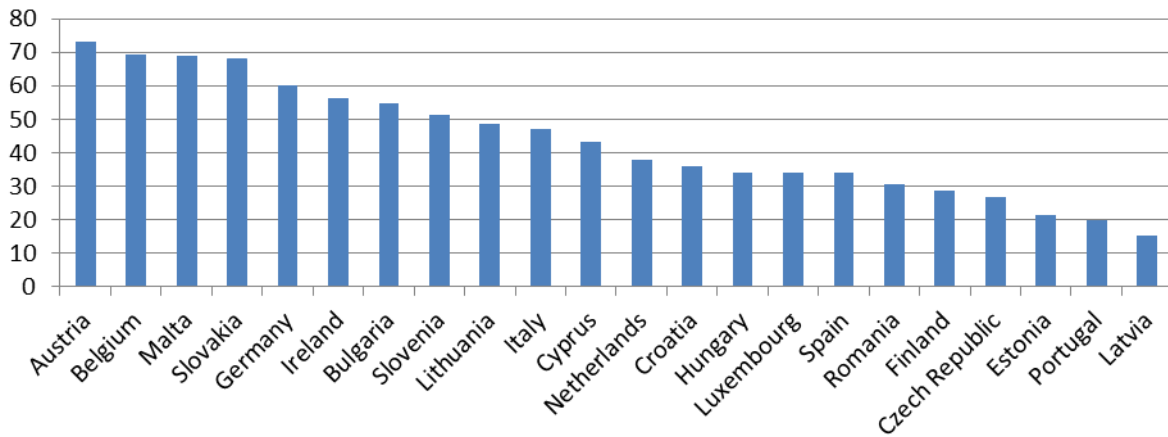

Figure 36. Annual ratio of nights spent by European residents in rented accommodation by country of residence in tourism demand statistics (Source: Eurostat).

Figure 37 shows that the estimate of tourism demand survey for nights spent in rented accommodation might be underestimated as it is smaller than the number of nights spent in accommodation statistics which excludes small accommodation establishments. Mobile positioning data contains nights spent in rented and non-rented accommodation and therefore, is accordingly higher than other two statistics. All sources show similar trends.
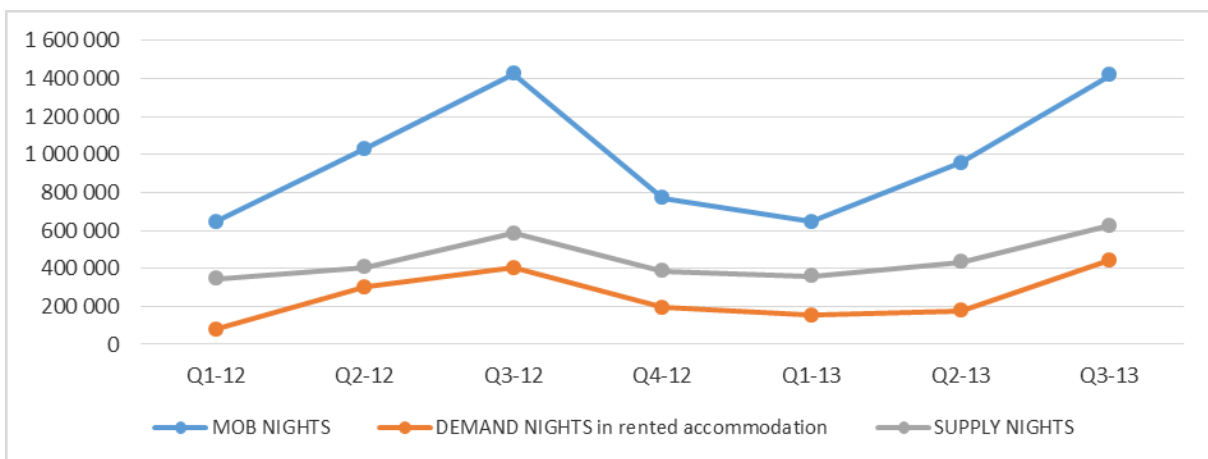

Figure 37. Quarterly domestic nights spent in mobile positioning data, in accommodation statistics and nights spent by Estonian residents in rented accommodation in the tourism demand survey (Source: Positium, Statistics Estonia).

Figure 38 compares the annual number of overnight trips in mobile positioning data and in tourism demand survey at a county level (based upon main destination). The data taken from six counties is presented.

Figure 38. Annual overnight trips in mobile positioning data and in the tourism demand survey at a county level (Source: Positium, Statistics Estonia).

Table 30. Coherence indicators: the number of overnight domestic trips to different counties according to demand side statistics and mobile positioning data (2009-2012)

Destination	Pearson's r	CC2 average	CC2 minimum	CC2 maximum
Estonia	0.82	N/A	N/A	N/A
Harju county	0.68	N/A	N/A	N/A
Ida-Viru county	-0.23	N/A	N/A	N/A
Lääne county	0.88	N/A	N/A	N/A
Lääne-Viru county	0.65	N/A	N/A	N/A
Saare county	0.79	N/A	N/A	N/A
Tartu county	0.50	N/A	N/A	N/A
Remaining Estonia	0.64	N/A	N/A	N/A

The county-level figures show that the data series for mobile positioning data and accommodation data is more stable and that for tourism demand statistics is more volatile. In the case of the tourism demand survey, one should keep in mind the fact that it tourism demand survey a sample survey and contains a sampling error. On the other hand, mobile positioning data may be influenced by the bias due to the errors mentioned in the Report 3a (see Section 3.2.).

4.6.5. Same-Day Trips

The data for domestic same-day trips is available only in mobile positioning data, in tourism demand statistics the data for same-day trips for Estonian residents is not available.
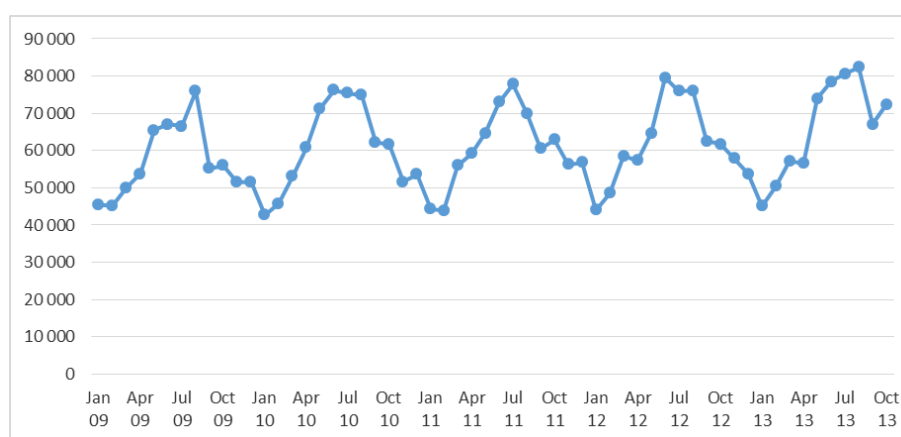

Figure 39. Domestic same-day trips in mobile positioning data (Source: Positium).
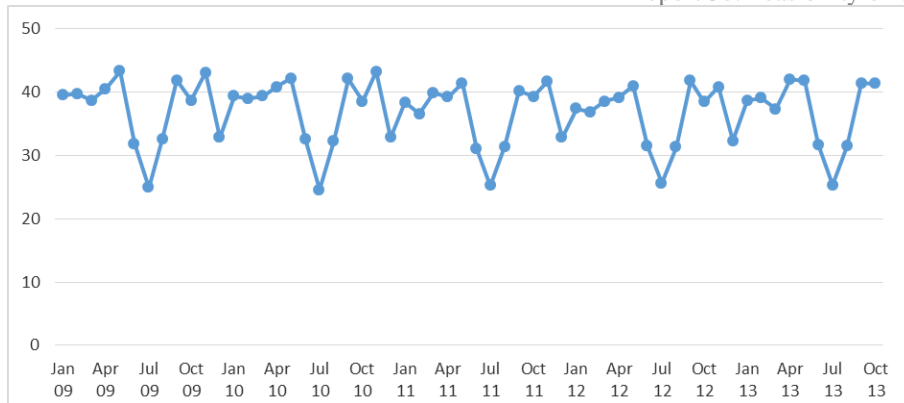
Figure 40. Share of same-day trips in domestic trips in mobile positioning data (Source: Positium).

The share of same-day trips is about 40% for low season and 25% for high season. Tourists have more overnight trips during summer period.

5. Conclusions

This report has analysed the coherence of mobile positioning data (mobile positioning data) in tourism statistics by comparing mobile positioning data from Estonian MNOs to selected reference statistics from Estonia, Finland and other EU countries. The analysis has been done both visually through graphs and statistically by using separate indicators for consistency and coverage of mobile positioning data against reference statistics. Figure 41 visualizes a sample of these comparisons for consistency and coverage, which provides an overview of mobile positioning data compared to various mirror statistics.
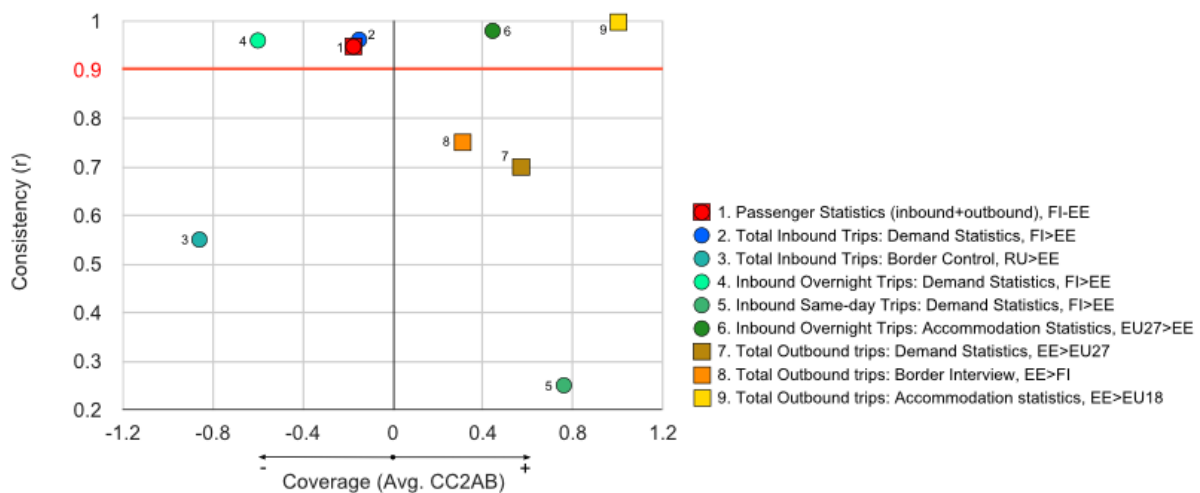

Figure 41. An overview of the two observed coherence indicators (consistency and coverage) compared against various mirror statistics (Demand, Supply, Border and Passenger Statistics).

This chapter presents the conclusions based upon the findings and observations from these analyses.

5.1. Total Inbound and Outbound Trips

Ferry passengers travelling between Finland and Estonia and the combined ferry and passenger numbers between Sweden and Estonia provided good testing corridors for analysing the total volume of inbounds and outbounds between Estonia and Finland/Sweden. The strength of these passenger statistics is that it is an administrative data source that includes a vast majority of the total number of tourism flows, but a drawback is that other nationalities and in-transit passengers are on board as well. Based upon the 2009-2012 monthly time series it can be concluded that MDP provides a very consistent estimate of the total number of inbound and outbound trips. Mobile positioning data underestimates the real passenger volume as reflected by the reference statistic due to the other nationalities and transit passengers on board. The absolute difference is greatest in the peak holiday months of June-August.

When compared to Demand Statistics, mobile positioning data also provides a consistent estimate of the total number of trips of Finns to Estonia when compared to Finnish Demand statistics. In 2012 the mobile positioning data slightly underestimates the number of trips when compared to Demand Statistics. The main potential reason for this underestimation is that some Finnish tourists to Estonia either do not have or do not use their mobile phone during the trip.

The consistency of mobile positioning data against Estonian Demand statistics is only moderate. The higher number of trips in mobile positioning data when compared to Demand Statistics implies that mobile positioning data contains many such Estonian outbound trips that do not qualify as tourism trips in the Demand Survey either due to frequency, purpose or duration of the trip. Further evidence for this is found by using the Finnish Border Interview Survey as reference data. Many Estonians commute to Finland making Finland part of their usual environment. The over-coverage in mobile positioning data is highest in quarters 1 and 4 that are outside the summer holiday season, which indicates that part of the outbound trips in mobile positioning data are made by these commuters. By definition the commuters should be excluded from tourism statistics, so further filtering should be carried out for outbound mobile positioning data that is based upon visit frequency. This should be possible in case the data contains a non-changing subscriber ID.

5.2. Same-day Trips versus Overnight Trips

Mobile positioning data is based upon passive positioning where the presence of the subscriber is triggered by calls and SMS messages. This provides both strengths and weaknesses for tourism statistics.

One of the strengths of mobile positioning data is the elimination of many short and frequent trips that include a border crossing as seen from the analysis using border control data. Majority of these same-day trips would qualify as being within the usual environment either due to frequency, duration or purpose of trip. Due to methodology used for mobile positioning data these trips are also excluded from the data as the visitor either doesn't use his phone or remains within the coverage of his home network.

Regarding overnight trips, the mobile positioning data provide a consistent estimate over time and the correlation to existing tourism statistics, both Demand and Supply, is mostly excellent. As seen from the example of Latvian visitors to Estonia, a weaker correlation to Supply Statistics may in fact indicate a strength of mobile positioning data as they capture a real tourism flow which is not represented in Supply Statistics that only includes rented accommodation in over-threshold establishments. While mobile positioning data cannot provide the split into rented and non-rented accommodation they are nevertheless a more comprehensive data source for overnight trips.

The inherent weakness of mobile positioning data is the potential misclassification of trips into same-day and overnight trips. At least two activities (calls and SMS) are necessary at long enough time intervals for a trip to be classified as an overnight trip. If only one activity is registered for the visitor during the entire trip it will be considered a same-day trip in mobile positioning data although in reality the visitor may have spent one or more nights during his trip without using the phone. The analysis using Finnish Demand Statistics provides quantitative evidence of this misclassification. The mobile positioning data clearly overestimates the amount of Finnish same-day inbound trips to Estonia and underestimate the amount of overnight trips.

Another potential weakness is the presence of cross-border commuters in mobile positioning data. According to definitions used in tourism statistics these commuters should be excluded from tourism statistics since they regularly commute to another country. Many Estonians commute to Finland regularly and might appear in the Estonian outbound mobile positioning data as seen in the analysis comparing mobile positioning data to data from the Finnish Border Interview. The exclusion of these commuters requires a non-changing

subscriber ID in mobile positioning data based upon which the frequent commuters could be excluded from the inbound and outbound mobile positioning data. As the raw data contains this kind of user ID, such filtering of frequent visitors could be carried out at the level of raw data.

5.3. Domestic Tourism and the Usual Environment

Domestic tourism differs from inbound and outbound tourism in the sense that the concept of usual environment critical for domestics trips. Most of the trips in inbound and outbound mobile positioning data are part of tourism with the exclusion of very short or very frequent trips. In domestic tourism the opposite is true: only a fraction of the total number of trips in mobile positioning data are relevant for tourism due to the concept of usual environment.

The trend in domestic overnight trips in mobile positioning data coincide largely with that in accommodation statistics. The difference, which varies depending on season, is largest during the high season of tourism in summer and smaller for the low season of tourism (autumn-winter). This is expected as the use of non-rented and small non-registered accommodation is highest during the summer months.

Compared to demand statistics, the difference in overnight trips is smaller than in accommodation statistics, as expected. Much of the differences of mobile positioning data to Estonian Demand statistics are attributable to the sample size, response rate and recent methodological changes in the survey.

All three data sources for domestic tourism (mobile positioning, accommodation statistics and Demand statistics) show similar trends, while the absolute number of trips is certainly highest in mobile positioning data.

At this stage it is not possible to assess the coherence of same-day trips due to lack of reference statistics. For overnight trips, the consistency of time series yields similar good results as for inbound and outbound tourism.

5.4. Overall Conclusions and Implications for Tourism Statistics

Based upon the requirements laid out in Regulation (EU) 692/2011 it is clear that mobile positioning data cannot fulfil these requirements. Although mobile positioning data provides good estimates for the number of trips, nights spent and destination, it does not

produce any additional data about those trips such as purpose of trip, type of accommodation or expenditure. For this information, additional data sources such as traditional surveys would still be needed.

The main strengths of mobile positioning data are:

1. Mostly excellent consistency over time for the number of trips and nights spent
2. Superior coverage for overnight trips when compared to Supply Statistics: covers also trips in non-rented or non-registered accommodation
3. Possibility to make breakdowns that are based upon region and nationality
4. Possibility to apply rules for usual environment, for example filtering out repeat visitors from inbound/outbound data using a non-changing subscriber id
5. Many short trips are excluded from mobile positioning data

The main weaknesses of mobile positioning data are:

1. Lack of additional information about the trip such as purpose of trip, expenditure, type of accommodation and means of transport used
2. The current methodology with its basis upon passive positioning cannot accurately break down trips into same-day and overnight trips
3. Over-coverage issues that are related to the usual environment due to purpose, duration or frequency of trips which can be partly solved by improving the usual environment algorithms
4. Under-coverage issues that are related to tourists that do not appear in mobile positioning data at all: some tourists either don't have or don't use their phone and therefore do not generate any mobile positioning data records

Based upon the analysis made in this report there are major opportunities for using mobile positioning data within tourism statistics. Although mobile positioning data might not always be accurate regarding the absolute level of trips or nights spent due to under- or over-coverage issues, on average mobile positioning data is very suitable for reflecting changes over time. Due to better timeliness when compared to official statistics, mobile positioning data would be particularly useful for providing quick indicators on the total number of tourism flows to complement the current Supply Statistics focusing on rented accommodation only. Mobile positioning data could also be used for producing more accurate regional tourism indicators. Mobile positioning data provides a source for calibrating the current surveys used for measuring tourism.

References

Legal documents

Regulation (EU) No 692/2011 of the European Parliament and of the Council of 6 July 2011 concerning European statistics on tourism and repealing Council Directive 95/57/EC: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:192:0017:0032:EN:PDF

Publications

Eurostat (2013a): Methodological Manual for tourism statistics. Version 2.1. Luxembourg (Cat. No: KS-GQ-13-007-EN-N): http://epp.eurostat.ec.europa.eu/portal/page/portal/tourism/methodology

Eurostat (2013b) Tourism Satellite Accounts (TSAs) in Europe - 2013 edition: http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-TC-13-006
Finnish Transport Agency (2013) Statistics on International Shipping 2012. Helsinki, 07/2013: http://www2.liikennevirasto.fi/julkaisut/pdf3/lti_2013-07_ulkomaan_meriliikennetilasto_2012_web.pdf

Nurkka, Pasi (2012) TAK Rajatutkimus 2012 esitysmateriaali: Rusgate.fi. Tutkimus- ja Analysointikeskus TAK Oy [Presentation]: http://www.rusgate.fi/sites/default/files/julkaisut/tak_rt_2012_imatra_lpr.pdf

Special Eurobarometer 414 (2014) e-Communications Household Survey and Telecom Single Market Survey Roaming Results. TNS Opinion & Social at the request of European Commission: https://ec.europa.eu/digital-agenda/en/news/e-communications-household-survey-and-telecom-single-market-survey-roaming-results-special

Statistics Finland (18 April 2013) Finnish travel abroad increased in 2012 [Press release]: http://tilastokeskus.fi/til/smat/2012/smat_2012_2013-04-18_tie_001_en.html

Websites

Eurostat Tourism Data Main Tables: http://epp.eurostat.ec.europa.eu/portal/page/portal/tourism/data/main_tables

Annex 1. Abbreviations and definitions

Abbreviation	Definition
AIR	Air passenger statistics
ARR	Arrivals in supply statistics
BORDCONT	Border control statistics
BORDINT	Border interview statistics
CC2	Coverage indicator: 'Asymmetry for mirror flows statistics - coefficient'
DEMAND	Official tourism demand statistics
DOM	Domestic
EE	Estonia
FERRY	Ferry passenger statistics
FI	Finland
IN	Inbound
MOB	Estimated mobile positioning data
mobile positioning data	Mobile positioning data
NIGHT	Nights spent in supply statistics
NON-RES	Non-resident tourist
OUT	Outbound
OVERNIGHT	Trips including overnight stay
r	Consistency indicator: Pearson's r
RAW	Indexed raw mobile positioning data (January 2012 = 100)
SAMEDAY	Same-day trips (no overnight stay)
SE	Sweden
SUPPLY	Official tourism supply (= accommodation) statistics