

Can Mobile Phone Data Improve the Measurement of International Tourism in France?

Guillaume Cousin* and Fabrice Hillaireau**

Abstract – Since July 2015, the *Banque de France* and the French Ministry for the Economy and Finance have been experimenting with the use of mobile phone data to estimate the number and overnight stays of foreign visitors in France. The purpose of the experiment is to assess the ability of such data to eventually replace, in part or in whole, the traffic data by mode of transport currently used to establish the representativeness of foreign visitor surveys (*Enquête auprès des visiteurs venant de l'étrangers* or EVE). Mobile phone data have yet to be incorporated into the method used to count tourists. However, estimates based on mobile phone data have a number of benefits in terms of the time required to obtain data, the level of temporal and geographical detail and short-term trend monitoring. This ongoing trial illustrates the difficulty of exploiting original Big Data and demonstrates the importance of drawing on traditional survey data to improve the quality of estimates.

JEL Classification : Z3, Z39

Keywords: Big Data, tourism statistics, mobile phone data

Reminder:

The opinions and analyses in this article are those of the author(s) and do not necessarily reflect their institution's or Insee's views.

* *Banque de France* (guillaume.cousin@banque-france.fr)

** *French Ministry for the Economy and Finance* (fabrice.hillaireau@finances.gouv.fr)

We sincerely thank François Mouriaux, Bertrand Collès, François-Pierre Gitton and Yann Wicky for their valuable contribution to the design and drafting of this article, as well as the Orange team with whom we conducted the experiment, in particular Sylvain Bourgeois, Jean-Michel Contet and Kahina Mokrani.

Received on 9 August 2017, accepted after revisions on 14 January 2019

Translated from the original version: "Les données de téléphonie mobile peuvent-elles améliorer la mesure du tourisme international en France ?"

Pour citer cet article : Cousin, G. & Hillaireau, F. (2018). Can Mobile Phone Data Improve the Measurement of International Tourism in France? *Economie et Statistique / Economics and Statistics*, 505-506, 89–107. <https://doi.org/10.24187/ecostat.2018.505d.1967>

Since July 2015, the Directorate-General of Statistics of the Banque de France and the Directorate-General for Enterprise (DGE) of the French Ministry for the Economy have been experimenting with the use of mobile phone data to estimate the number of foreign visitors in France and their overnight stays as part of a survey partnership aimed at developing the tourism satellite account and determining the balance of payments.¹

Counting foreign visitors and their overnight stays is necessary for the purpose of generating tourism statistics and estimating France's "travel services" trade balance. These statistics are of particular importance because of the weight of tourism in the French economy. In 2015, tourism consumption accounted for 7.27%² of GDP, 32% of which came from non-resident visitors. In the same year, France welcomed 84.5 million foreign tourists (DGE, 2016a), generating 52.6 billion euros of revenue from travel services recorded in the balance of payments.³

At present, visitor numbers are measured based on traffic data by mode of transport combined with counting sessions and surveys. Such sessions provide a detailed breakdown of border-crossing based on the country of origin of visitors, while traffic data by mode of transport serve as a basis for calculating extrapolation coefficients. Current estimates of visitor numbers require improvements which can, however, prove complicated when seeking to take into account rapidly-occurring changes such as the conversion of major airports into hubs – thereby multiplying the number of nationalities present on the same flight – and the difficulties involved in using road counts at borders in countries with multiple entry points (for example, France and Belgium share approximately three hundred border-crossing points). The purpose of experimenting with the use of mobile phone data is to evaluate the capacity of such data, in time, to replace traffic data by mode of transport for measuring foreign visitor numbers in France.

Mobile phone data have already been used to monitor tourism in Estonia and by a number of departmental and regional tourism committees, such as Bouches-du-Rhône Tourisme,⁴ to measure visitor numbers and flows. Therefore, mobile phone data appeared to represent a potential solution for overcoming the new limitations affecting, or likely to affect, current data sources.

This is because they are a rich source of information about the location and mobility of individuals. Gonzales *et al.* (2008) were among the first to use this data source to build a model of human mobility based on a sample of 100,000 mobile phones monitored over a 6-month period. This type of data has since been used to identify mobility patterns (Calabrese *et al.*, 2011, 2013), notably commuting (Aguilera *et al.*, 2014). Widhalm *et al.* (2015) sought to build a typology of urban activity patterns based on travel time, frequency and location. Many other uses have been explored in a wide range of areas (ONS, 2016).

Official statistical agencies have identified the potential benefits of Big Data for conducting population censuses (Vanhoof *et al.*, 2018; Givord *et al.*, this issue), but also for measuring tourism. Estonia has been a pioneer in this area, with an experiment reported in two main articles: Ahas *et al.* (2008), who found a strong correlation between mobile phone data and accommodation statistics, and Kroon (2012), who presented an experiment conducted by the Bank of Estonia involving the use of mobile phone data as a potential data source for estimating trade in travel services. Eurostat (2014) has since produced a feasibility study on such data for the purpose of monitoring tourism.

However, mobile phone data analysis is rarely used to measure tourism, and our experiment has the advantage of focusing on a relatively large country (twelve times larger than Estonia) welcoming a significant number of tourists (28 times more than Estonia). This paper also has the peculiarity of being written from the point of view of official statistics compilers and provides a different perspective to most other papers, aiming as it does to promote a regular operational use of Big Data for the development of statistical indicators. Given this background, it is important to test the quality of the indicators by comparing them

1. The partnership concerns the surveys known in French as SDT (Suivi de la Demande Touristique, or Tourism Demand Survey) and EVE (Enquête auprès des visiteurs étrangers, or Foreign Visitor Survey). The first survey involves collecting data on tourism demand among French nationals and is based on a representative sample of French households. The second data collection relates to tourism demand among non-residents visiting France and focuses on flows as evidenced by methods such as "border surveys". The partnership enables the integrated production of reference data for the official statistics for which each institution is responsible.

2. Cf. DGE, Compte satellite du tourisme (base 2010); Insee, Comptes nationaux (base 2010).

3. Cf. Webstat, Banque de France.

4. Cf. <https://www.myprovence.pro/bouches-du-rhone/projets-majeurs/projet-flux-vision-tourisme>.

to alternative data – in this instance, the reference survey on international tourism in France (EVE) combined with card payment data.

Mobile phone data have yet to be incorporated into the method used to estimate tourist arrivals, which has been limited to traffic data, counts and surveys. The test period highlighted many specificities in the use of Big Data: access to data, anonymisation and technical constraints, and quality of the data and of the indicators constructed on the basis of such data. It also meant that processing methods could be developed to render them more usable.

The Initial Requirement: Consolidation of the Current System Used to Estimate Foreign Tourist Arrivals

The Estimation of Foreign Tourist Arrivals is Based on Counts and a Survey Conducted at Border-Crossing Points

The current system used to estimate foreign tourist arrivals was developed based on the legacy of the border survey carried out between 1963 and 2001 with a view to bringing it into line with the wider context of the free movement of capital and the creation of the eurozone and of an area of free movement of persons (Schengen zone). The system is known in French as the *Enquête auprès des visiteurs venant de l'étranger* (EVE) and combines traffic data, counting sessions and a survey (Banque de France, 2015). The reason why monitoring international tourism presents such a challenge is that there is no sampling frame from which to carry out a traditional survey (such as is the case with “outgoing tourism”, for example⁵). The EVE system is thus based, first, on a traffic census at the country’s exit points (ports, airports, train stations offering international routes, road borders). Data on air, sea and rail passenger flows are collected from the various transporters and carriers, while road traffic flows are estimated by Cerema⁶ using fixed or mobile automata distributed along all the borders (more than one hundred and fifty counting points in total). The second stage involves detailing the total flow, i.e. breaking it down into resident and non-resident flows. This requires counting operations to be conducted by researchers located at different points throughout the country. In

airports, non-residents are counted in boarding lounges, which provides a basis for estimating the split between residents and non-residents based on a sample of flights and for extrapolating therefrom. In the case of travel by road, counting is conducted at border crossing points. The distribution of outgoing traffic by country or area of residence is then further specified using the responses to the EVE survey questionnaires. The EVE survey is thus based on a combination of traffic data (external data), specific counting operations (over a million vehicles counted at the borders, over 120,000 air passengers) and the survey itself, administered to over 80,000 visitors in both 2015 and 2016, the questionnaire being available in twelve languages.

The EVE Method Faces an Increased Need to Adjust the Traffic and Count Data

For counts (or tallies), the main challenge facing the EVE method in terms of statistical adjustment concerns road travel. For this particular mode of transport, the difficulties arise from both external traffic data, which are based on fewer measurement points, and counting and survey sessions, some of which are relatively unproductive. In order to obtain spontaneous responses at the end of their stay in France, travellers are interviewed in motorway rest areas near the border, where visitor numbers can vary unpredictably depending, for example, on the travel routes of coach operators. It follows that the extrapolated results relating to the distribution of outgoing visitor flows by geographical area can fluctuate erratically for some areas or countries or origin, meaning that specific corrective measures may need to be considered. Similarly, but to a lesser degree, splitting outgoing air traffic according to the country of origin of visitors needs to take into account the importance of transit in Paris airports and the role of major airports as hubs. The EVE survey is carried out with specific objectives in terms of questionnaires by area of origin. In the case of airports, the survey design is based on a sample of flights sampled in such a way as

5. French people travelling abroad. For these, a representative sample was obtained as part of the SDT (Suivi de la demande touristique, or Tourism Demand Monitoring) survey.

6. Cerema (Centre d'étude et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement) is an administrative public establishment under the joint authority of the French urban development and sustainable development ministries.

to collect questionnaires from visitors from a range of countries of origin. However, the link between flight destination and the country of origin of visitors is made tenuous by the fact that some passengers transit through an intermediate destination before returning to their country of origin. For example, Asian tourists are likely to fly from Paris to Frankfurt when leaving France, thereby complicating the targeting of the survey.

Mobile Telephony is a Potential Source for Counting Visitor Numbers

The decision to experiment with the use of mobile phone data was largely based on the potential of such data to address the issue of monitoring international tourism, as highlighted by several experiments. In France, some local authorities use such data to measure tourist arrivals. In Europe, Estonia has used mobile phone data as a main source for measuring incoming and outgoing international tourism since 2008-2010, while other countries have also shown an interest in using data of this kind, the potential of which has been highlighted in an in-depth study (Eurostat, 2014). However, the experiment conducted by the Banque de France and the DGE is unique by virtue of its scale, with the population of interest consisting of international tourists in France representing 85 million people a year. In addition, the territory where the count was conducted (metropolitan France) covers an area of 552 thousand square kilometres. By comparison, the number of tourist arrivals in Estonia is approximately three million a year in a country with a surface area twelve times smaller.

From a technical point of view, the use of mobile phone data is made possible by the fact that operators have access to the list of connections between cell towers and mobile phones, whether for the signals emitted passively or for mobile phone activity (calls, messages, reception of data via the Internet, etc.). The country of residence of the operator issuing the SIM⁷ card of mobile phones connecting to the French network is also known to operators based in France and provides a basis for building a mass database of the signals emitted by roaming mobile phones. Mobile phone data therefore include variables of interest for the production of tourism statistics.

Experimental Framework and Procedures

The Arrangement of a Service Contract Corresponds to the Financing of a Cooperative Research and Development Initiative

While experimenting with data obtained from mobile phones for the purpose of counting the number of foreign tourist arrivals in France evidently amounts to a Big Data approach, the context of the experiment cannot be said, however, to reflect an open data approach. This is because the data are held by the various operators with access to a mobile network within metropolitan France. To gain access to these data in the form of statistics, the Banque de France and the DGE issued an invitation to tender in the spring of 2015 and received two proposals, reflecting, on the one hand, the interest of operators in collaborating with official statistical agencies with a view to gauging a new field of application for their Big Data and, on the other, the need for public funding as part of an arrangement to share research and development costs and the costs specific to the provision of information within the context of the experiment. When interviewing the candidates, expectations in terms of the transparency of data collection and processing methods were a key discussion point. The experiment was based on a distinction between the detail of the aggregation and anonymisation algorithms, which relate to the protection of the service provider's intellectual property, the market shares of the operator selected for the different populations and territories examined being a matter of commercial confidentiality, and the determinant variables for assessing the statistical quality of the data and the adjustment choices, all of the methodological options having to be gauged by the Banque de France and the DGE. A more detailed explanation of the distinction is provided below.

At the end of the tender process, the contract was awarded to Orange Business Services. The experiment focuses on data dating from early July 2015 to late June 2017. The last available

7. A SIM (Subscriber Identity Module) card is a chip used in mobile phones to store information that is specific to a mobile network subscriber, in particular for GSM, UMTS and LTE networks. It also allows the data and applications of the user, of the user's operator or, in some cases, of third parties to be stored. A SIM card contains an IMSI number, made up of a mobile country code (MCC), a mobile network code (MNC) and a mobile subscription identification number (MSIN).

data point at the time of writing this paper was March 2017.

The selected bid has three characteristics: a pre-existing Big Data formatting module already available on the market, a balance between respect for intellectual property and methodological transparency, and a “phased” process governing the methodology used to develop the indicators.

The service provider had already developed a Big Data processing module adapted, in particular, to the needs of users looking for spatio-temporal data on visitor numbers (quantifying groupings of individuals in a given location, such as a cultural or sporting event). However, the proposed method had never been used for the purpose of observing international tourism over the entire national territory of France.

The service provider undertook to provide the two partners with a sufficient level of information to ensure that the method used would enable them to comply with the statistical quality standards established, in particular, for international and European institutions and be intelligible to the various audiences with an interest in tourism statistics. A knowledge and understanding of the methodology used serves to guarantee the independent ability to interpret results as well as the ability, where relevant, to make necessary revisions. At the same time, it was important to provide the service provider with an assurance of confidentiality in relation to the algorithm used to move from the basic datum of the signal transmitted by a SIM card to one or several towers to a raw datum representing a proxy of the anonymous physical person. The level of detail of the shareable methodological information therefore varies at stages 2, 3 or 4 and 5, as detailed below.

The mobile network operator’s pre-existing module does not record the detail of the movement of SIM cards prior to processing their movements by aggregating the data following the requirement of the study. The processing method validated by the *Commission nationale de l’Informatique et des libertés* (French National Commission on Informatics and Liberty, CNIL) requires that the behaviours studied be predefined. Predefined behaviours are the sole target of real-time incremental counts of connections to the provider’s networks, without personal data being stored.

The counting method includes five stages (Diagram):

- Stage 1: the counting criteria are defined in advance with the Banque de France and the DGE and correspond to the tourist behaviours of interest (arrival, overnight stay);

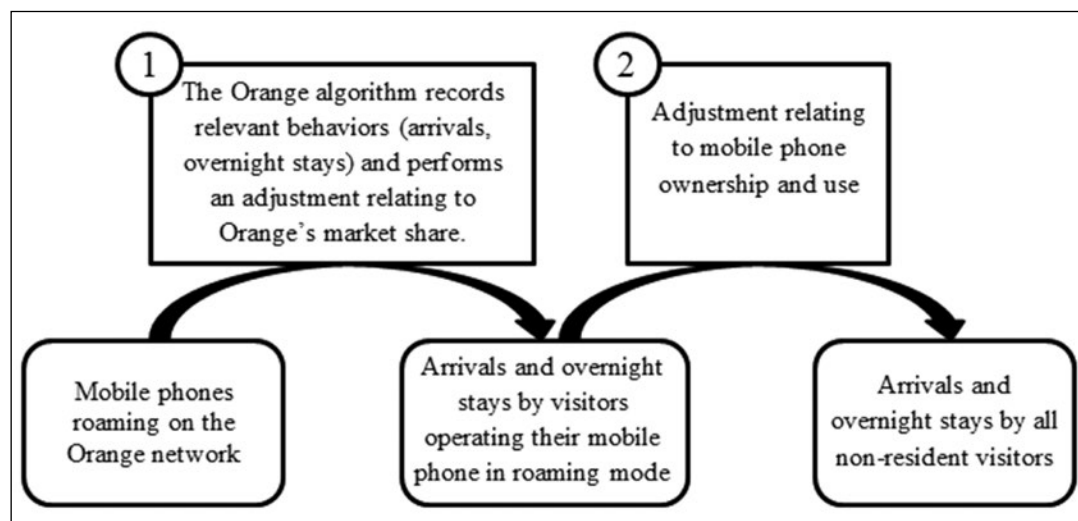
- Stage 2: mobile phone connection data are fed into an algorithm in real time. These involve signalling data, which include all of the communication between mobile phones and cell towers. The recorded signals are those emitted passively by mobile phones to connect to a cell tower based on their position as well as the data transiting via cell towers when mobile phones are being used (calls, SMS, use of mobile applications). These data originate solely from mobile phones equipped with SIM cards issued by a non-resident operator and roaming on the Orange network. They are not stored. The algorithm processes and anonymises the data as and when they are collected, rather like a meter. The algorithm constructs estimates of visitor numbers in terms of the number of mobile phones, by area of origin;

- Stage 3: the service provider carries out a “spatio-temporal” adjustment by aggregating the data on the different networks (“2G”, “3G”, “4G”) and correcting the effects associated with the constant changes being made to these networks (putting into service of new cell towers, temporary or long-term unavailability);

- Stage 4: the transition from connection data to an estimate of the number of foreign mobile phones present within metropolitan France requires an adjustment to the service provider’s market share of roaming customers, by country of origin and operator of origin in combination. Market share by country-operator is measured based on the distribution between the different operators of SMS sent from roaming mobiles, which is available in real time to the service provider;

- Stage 5: lastly, the number of visitors by area of origin is estimated based on a traditional statistical adjustment relating to mobile phone ownership and usage rates. Since it is carried out ex post on the results aggregated by country (or larger area) of assumed residence, this last adjustment lends itself to the execution of several scenarios.

Diagram
Simplified Methodology Used to Construct the Indicators



A Setup Stage is Required Prior to the Observation Period

The setup stage involves defining, in conjunction with the service provider, the different behaviours of interest in order that they can be measured. In the case of the present experiment, the definitions of tourist arrivals and overnight stays had to be translated into criteria relating to the presence of mobile phones on a network. An overnight stay was thus defined as the presence of a mobile phone between midnight and six o'clock in the morning. An arrival is counted when the first overnight stay is recorded following an absence the previous night. Such hypotheses allow mobile phone data to be interpreted in terms of behaviours. Some studies have used similar hypotheses to examine mobility patterns based, for example, on presence at home between midnight and eight o'clock (Akin & Sisiopiku, 2002). The initial criteria were gradually completed in order to correct a number of measurement deviations.

The setup stage is also an opportunity to define the territory of interest. This requires selecting the cell towers involved in the counts. A decision was made not to take into account the flows received by towers located on French soil but close to the borders, a factor deemed necessary when analysing the initial results generated from the operator's pre-existing module. The reach of cell towers being unaffected by administrative boundaries, it is important not to count foreign residents outside France. For

the data aggregated from the entire territory of metropolitan France, which are the primary target of the Banque de France and the DGE, we may therefore expect a slight underestimation of foreign tourist arrivals. Their numbers cannot be estimated since, over the experimental period, the overall noise does not allow bias measurements of such accuracy to be carried out. In the case of data broken down at a regional level, the problem of cell tower selection also arises at each administrative border, with the map of towers and their reach not being aligned with the map of regions and departments. This requires a specific focus on allocating cell towers on the basis of the spatial groupings sought.

The Series Obtained and their Usefulness

The Series Received

The operator provides the Banque de France and the DGE with estimates of the number of international tourist arrivals and overnight stays. These estimates are provided monthly within a theoretical period of one month. The indicators are provided at a daily frequency. Estimates are provided for 29 geographical areas of visitor origin. The data received are in CSV format. One file contains the arrivals while another records the overnight stays. Each file contains three columns: area of origin, date and number of overnight stays or arrivals for

the corresponding origin \times date intersection. The files received thus contain approximately 900 lines (29 geographical areas plus a total line times approximately 30 days). For a given day and country of origin, it is thus possible to determine the number of tourists who arrived and the number of tourists who spent the night in France on that day.

It is important to note that mobile phone data were initially chosen to calibrate the measurement of tourist flows travelling by road. The service provider was therefore required to include a distribution of tourist arrivals by border and mode of transport. The subtlety of the collection process and the operator's expertise were expected to identify the different means of transport, with rail travel being characterised, for example, by a high number of SIM cards operating at the same speed and over an identified route. In actual fact, the significant differences found between the data of the operator's pre-existing module and the contextual data (EVE survey) were such that the identification of the mode of transport was quickly abandoned. The approach ultimately adopted therefore focuses on the primary requirement to count the number of foreign visitors arrivals, aggregated for all modes of transport and borders.

While Initially Disappointing, the Quality of the Data Improved

Comparison of the indicators derived from mobile phone data and survey data provides a basis for assessing their reliability and examining potential sources of variance. Such comparisons have been carried out by several studies devoted to the analysis of mobility and the construction of origin-destination matrices. The results obtained using mobile phone data are, in some cases, close to the survey results, but are of a higher level (Calabrese *et al.*, 2013). In the field of mobility, a more recent study (Bonnet *et al.*, 2015) compared the results of the global transport survey with estimates derived from passive mobile phone data provided by Orange. The authors found a strong correlation between the two types of estimates and reached similar estimates in terms of the total number of trips in the Île-de-France region (9% difference). However, their study covered a short period (twelve days).

In the context of the experiment conducted by the Banque de France and the DGE, the first

deliveries of estimates in the third quarter of 2015 showed significant differences with the estimates of the EVE survey, the only available source for an estimate of tourist arrivals and overnight stays in metropolitan France. The indicator obtained from mobile phones pointed to over one hundred million tourist arrivals in the third quarter alone, whereas the approximate number of tourist arrivals is 85 million a year.

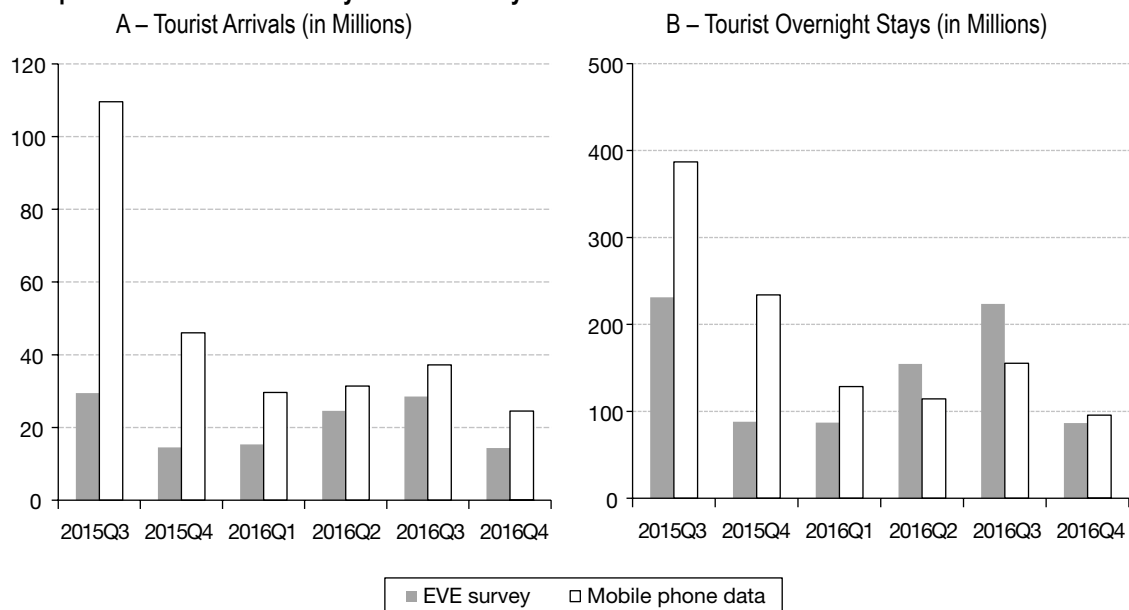
It soon became apparent that the indicators developed based on a single collection are unusable for various reasons set out below. Arrivals of tourists and day visitors (i.e. visitors who do not spend the night on French soil) in France were thus found to be of insufficient quality. Further work is therefore needed on "consolidated" indicators,⁸ such as the number of overnight stays, each overnight stay being defined as presence confirmed several times in the same location over a defined time frame. Overnight stays are therefore not counted in the case of arrivals previously defined by the measurement system. Rather, the number of arrivals is deduced from the number of observed overnight stays. This is entirely consistent with the letter and spirit of the international definition of a tourist: a tourist is a visitor whose visit includes at least one night in a territory which is not his or her habitual residence.⁹ Lastly, the goal of counting the number of one-day visitors (i.e. visits which do not include an overnight stay), made difficult by the exclusion of border areas where cell towers are likely to cover a portion of foreign territory, was soon abandoned, it being impossible to define it either directly or as a balance. The improvement work therefore focused mainly on an analysis of overnight stays.

To improve quality, various corrections were made throughout the experimental period; these are described below. The result was to bring the estimates of arrivals and overnight stays based on mobile phone data closer to trustworthy levels compared to the EVE survey estimates (Figure I). The difference between the estimates of the total number of overnight stays decreased from 67% in the first quarter delivered (third quarter 2015) to 10% for the final quarter currently available (fourth quarter 2016).

8. More generally, monitoring observations over time helps to overcome many of the defects of the measurement system, although the CNIL's requirements restrict such monitoring to 3 consecutive months.

9. See the website of the World Tourism Organization: <http://media.unwto.org/fr/content/comprendre-le-tourisme-glossaire-de-base>

Figure I
Comparison of Measurements by the EVE Survey and Mobile Phone Data



Coverage: Quarterly arrivals of non-resident tourists.
 Sources: Banque de France, DGE.

However, the quality of the estimates remains insufficient for three reasons. First, significant differences remain between the two sources in relation to countries or areas of origin. Some nearby areas appear to be overestimated whereas remote areas are underestimated. In the 4th quarter of 2016, for tourist overnight stays, the total difference of 10% between mobile phone data and the estimate obtained from the EVE survey arises from the compensation of very significant differences at country level. The estimates of overnight stays based on mobile phone data are 78% higher than those of the EVE survey for Germany, but approximately 80% lower for the United States, Canada and Brazil, for example. These differences stem in all likelihood from the values of the mobile phone usage rates used to adjust the data; for tourists originating from remote countries, the adjustment factors poorly reflect visitor behaviours. This limitation is inherent to mobile phone data: the quality of the estimates depends on knowledge of the operator's penetration rate by population segment. This limitation has been noted in several other studies, including when the population of interest is the resident population (Bonnet *et al.*, 2015). However, unlike the definitions of tourist behaviours, the adjustment factors relating to the use of mobile phones can be modified *a posteriori*. Therefore, the quality

of the data can be improved by furthering our understanding of behaviours.

Second, estimates of tourist arrivals are less robust than estimates of overnight stays. This is due to the problem of interrupted stays (see next section), which has a comparatively greater effect on arrivals than on overnight stays. Lastly, for some countries or areas of origin, there is a seasonality in the estimates obtained from phone data that differs significantly from the survey and which appears to be of limited credibility. For example, in the case of Spain, the EVE survey indicates that tourist overnight stays increase by 80% between the second and third quarters, reflecting the seasonality characteristic of the data from the professions in question (air traffic to tourist destinations, hotel occupancy, etc.), whereas the mobile phone data indicate a much lower increase of 13%. It follows that estimates obtained from mobile phone data are not yet of sufficient quality to replace the traffic data currently used.

The Data are Potentially Adapted to Short-Term Trend Monitoring

As noted above, mobile phone data have yet to be made more reliable at different levels. This is largely because of the adjustments

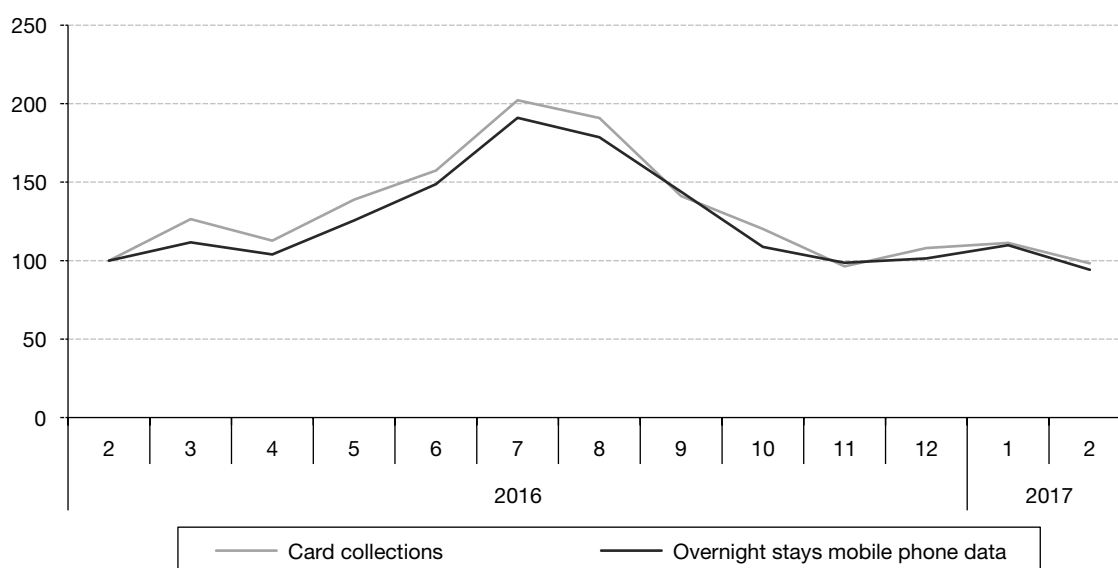
relating to mobile phone usage among visitors travelling from remote areas and artificially interrupted stays.

An analysis of these evolving data is of greater value, all the more so since such data are generally available before traditional survey data and provide a greater degree of chronological detail since daily estimates are available. The potential of mobile phone data for monitoring short-term trends can be seen by comparing them to the payment card data collected monthly by the Banque de France, which relate to cash payments and withdrawals made in France using a non-resident card and which are aggregated by country. Such spending is not an exact reflection of travel credits (use by residents of foreign cards, spending in currencies withdrawn in the country of origin or prepaid by simple bank transfer) but overlaps with it to a great extent, particularly in the case of specific nationals from particular countries among whom the use of card payments is the primary or even exclusive payment method. Payment card data also have the advantage of being available on a monthly basis with a detailed geographical breakdown, thereby allowing comparison with mobile phone data. Over the period February 2016 to February 2017,¹⁰ the revenue derived from payment cards and tourist overnight stays as measured

using mobile phone data were highly correlated: the correlation coefficient between the two series was 0.986 (Figure II). Furthermore, the high correlation between revenue derived from the collection of bank card data and tourist overnight stays estimated using mobile phone data was also found to be true for the different countries observed, albeit with some exceptions. Thus, while estimates of the number of overnight stays in levels show significant differences with the results of the EVE survey for certain countries (for example, the United States, Canada and Brazil), the correlation between overnight stays and payment card spending is high for all countries except Brazil (very limited use of payment cards), Morocco and Luxembourg (on account of the high proportion of cards issued in Luxembourg but used by residents of other countries). For the other countries, the correlation coefficient between overnight stays estimated based on mobile phone data and revenue obtained from card payments ranges between 0.66 and 0.97. Furthermore, some countries where the number of overnight stays is significantly underestimated using mobile phone data present fairly accurately estimated short-term trends

10. The decision to start in February 2016 is based on the fact that the last significant methodological change caused a break in series between January and February 2016.

Figure II
Tourist Overnight Stays Estimated Using Mobile Phone Data and Revenue from Payment Card Transactions
(Base 100 in February 2016)



Coverage: Expenditure in France using non-resident cards (excluding online) and overnight stays of non-resident tourists.
Sources: Banque de France, DGE.

does not use his or her mobile phone actively over that short period, she/he will not be represented in the numbers used to calculate market shares. The adjustment key is therefore underestimated and the extrapolation is carried out with an excessively high coefficient, implying an excessive estimate for the number of SIM cards of the relevant country over the considered period and territory. Because of this problem, it was necessary to make several changes to the measurement criteria. These changes are detailed in the next section.

Reception Interruption: A Cause of Underestimation of the Average Length of Stays and of Overestimation of Arrivals

The second measurement problem is in part related to the first and concerns artificial interruptions of stays. The counts available in the spring of 2016 indicate significantly excessive arrivals and overnight stays of an order of magnitude compatible with the contextual data, implying an excessively short average length of stay. To improve the measurement of the length of stays and arrival numbers, the Banque de France and the DGE requested an examination of the following intuition: stays may be artificially shortened by reception interruptions. Such interruptions can be caused by many factors: a mobile phone which has run out of battery or been deliberately turned off, travel to an area not covered by the Orange network, etc. The result is an automatic overestimation of the number of tourist arrivals: when reception resumes, if the interruption included an overnight stay, the SIM card is treated as a new arrival. The phenomenon also results in an underestimation of the number of overnight stays. For example, during a traditional one-week holiday stay on French soil, a SIM card signal may be detected regularly over three days, not be detected for two days and again be detected over two days prior to leaving the country. The operator's pre-existing system will therefore assume that two tourist arrivals have taken place, with the first arrival corresponding to a three-night stay and the second to a two-night stay.

To evaluate this hypothesis, a first test was performed in early 2016 over a limited area and period conducive to analysis: a hill station. The advantage of the chosen area is that we have a good understanding of tourist behaviours in such settings: foreign customers, a significant proportion of stays starting on Saturdays and

ending the following Saturday, and a clearly defined reception area on account of natural barriers. From this, it was found that the proportion of tourist stays concerned by artificial interruptions may be too high. This result is further supported by another finding at a national level over two observation periods: March 2016 and September-November 2016. For these periods, the number of arrivals was studied based on a requirement for absence prior to arrival. While this requirement is commonly set at one day, the aim of the observation was precisely to determine whether the arrivals recorded by the algorithm are genuine arrivals or artificial arrivals of individuals already present within the studied territory. Over the course of March, the application of a more stringent absence requirement (two days of absence prior to an arrival) resulted in the number of total arrivals decreasing by 13%. If the requirement is extended to six days of absence, the number of arrivals decreases by 37%.

The underestimation is therefore confirmed without it being possible to correct it on account of the inability to distinguish between artificially interrupted stays and genuine regular short stays, this distinction being necessary to establish the number of overnight stays. The application of a more stringent absence factor is not self-evident and would imply a departure from the statistical definition of tourism. Beyond the impact on the aggregates, which is significant if we exclude, for example, all arrivals prior to three days before the previous arrival, a more fundamental problem arises: should we rely on probabilistic reasoning, which would involve correcting an unsatisfactory measurement by adapting and altering definitions? A direct intervention on the source data, a physical identification of anomalies and correction prior to determining the aggregates appear preferable but are not feasible at present. While in terms of behaviour the identification of problematic cases appears unequivocal (sudden exit from the network, generally at a distance from a border, or over a route incompatible with an actual exit from the territory in question, and an equally sudden return), such an identification is not compatible with the operator's pre-existing measurement system. The question of interrupted stays is referred to in the study conducted by Estonia on international travel monitoring (Kroon, 2012). The authors mitigate the difficulty by adopting a number of hypotheses. They consider a mobile phone to

be present if the phone has remained inactive for less than 7 days and that the traveller has left if the period of inactivity is greater than 7 days. Such a solution is not ideal in the case of France, a major transit point.

The Country of Issue of the SIM Card and the Country of Residence May Differ

A third measurement difficulty concerns behaviours which weaken the hypothesis that the country where the SIM card was issued is the same as the country of residence of the owner of the mobile phone. Having been identified at the outset of the experiment by analogy with research conducted on French tourists, this difficulty has been in part corrected.

Converting the Number of SIM Cards into Tourist Arrivals is not Straightforward

Lastly, the final difficulty concerns the ultimate adjustment phase, which occurs *a posteriori*, independently of the operator's pre-existing system, in order to distribute arrivals and overnight stays according to the issuing area of the tourists. As noted in the section relating to data quality, the adjustment factors used to extrapolate tourist arrivals from the number of mobile phones are not always appropriate. The data relating to ownership rates in different countries come from GSM Alliance. However, there are significant limitations to using data on ownership rates among populations due, on the one hand, to the difference in representativeness between the total population of a country and the percentage of that population visiting France and, on the other, to the specific behaviours exhibited when travelling abroad. Ownership rates can vary according to the tariffs charged by operators in a given country and according to sociocultural factors (such as populations being more or less sensitive to security matters and with more or less intense connection habits).

In the absence of external data on ownership rates, the service provider combines the data relating to ownership rates with several sets of coefficients determined based on broad groups of countries defined according to their remoteness.

This adjustment problem, while it may not prevent a trend analysis¹¹ of indicators on a country-by-country basis, nevertheless complicates

the use of aggregate indicators. By positing that the trends in tourist arrivals for residents of country A are measured adequately and that the same applies to country B, the trends in overnight stays for residents of both countries taken together cannot be determined without drawing on external data relating to the respective contributions of both countries to tourist arrivals in the area considered. Thus, besides the numbers by country which are, in some cases, significantly underestimated, the evolving data are affected when considering a range of nationalities.

Corrections Made, Resulting Benefits and Limitations

The first correction made involved introducing a requirement for mobile phone loyalty to the operator's network in order to reduce the noise caused by sporadic connections by mobile phones connecting to the network only in the event of connection to their preferred network being lost. As indicated above, taking these phones into account results in an overestimation of the number of overnight stays and arrivals on account of the adjustment made by the operator in relation to its market share of roaming customers. In order to distinguish between regular and occasional users, the first criterion to be introduced concerns the total amount of time spent on the network, which must be higher than 9 hours over a 21-hour period. This first criterion was added for the delivery of the data from November 2015 and resulted in a decrease in the estimated number of overnight stays of around 30% (see Figure IV). It was further strengthened by adding a new loyalty requirement for the delivery of the data from February 2016. This requirement, applied systematically, requires that at least three network events be performed over the course of the 24-hour period prior to or following the recorded tourist overnight stay. Despite these successive improvements, sporadic mobile connections continued to introduce noise into the measurement. Current studies are tending to focus on selecting operators of the country of origin of visitors based on their loyalty to the network when roaming in France.

11. Over short periods at the very least: long-term analysis presupposes stability in mobile phone usage behaviour.

The second important measurement correction concerns French residents using a mobile phone with a foreign SIM card, which may be the case, for example, of cross-border workers (i.e. French residents working abroad) who have taken out a phone contract with a foreign company. Such behaviour tends to somewhat limit the validity of the hypothesis according to which the user's country of residence is the same as the country which issued the user's SIM card, resulting in an overestimation of the number of tourist overnight stays. The correction made drew on the significant contribution of the working group led by Tourisme & Territoires¹² in relation to the segmentation of the observed population. This is unavoidable for resident population data since the frequency and duration of trips make it possible to distribute individuals into different categories. The number of nights spent in a given territory means that someone carrying a mobile phone may be deemed to be a resident of that territory, regardless of the characteristics of his or her SIM card. Applied more simply to SIM cards with a foreign code, this method of segmentation serves to eliminate individuals who spend more than half of their nights in France over a two-month period. Therefore, the service provider added a non-residence requirement to the mobile phone data processing algorithm. Individuals who had spent more than one month on French soil over the

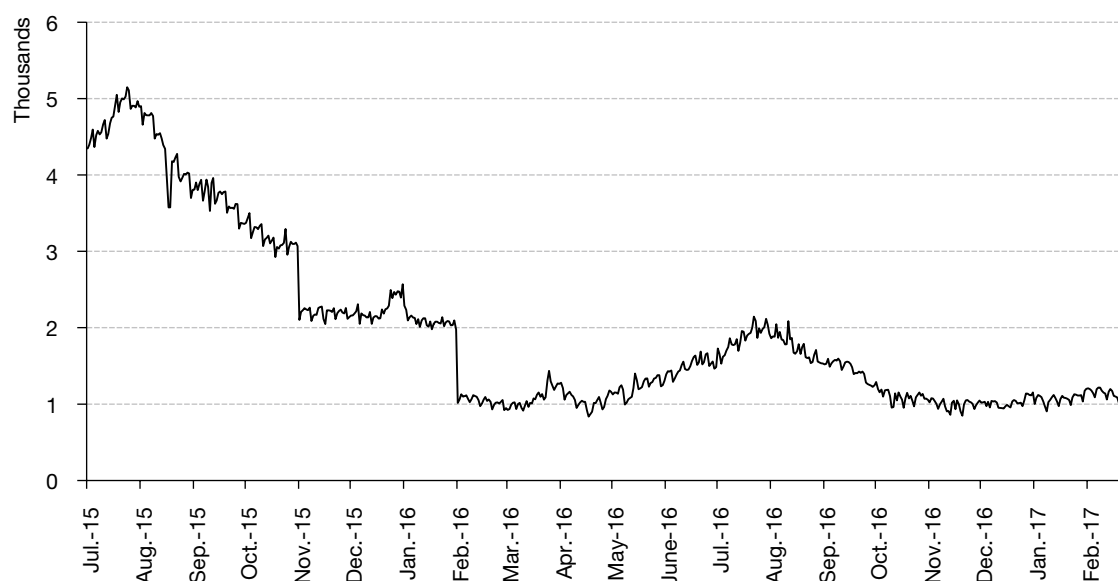
previous two months were treated as residents and were therefore excluded from the measurement of international tourism, thereby addressing the case of cross-border workers.

Because of the necessary learning process, the first data delivery to take account of this correction was the delivery made in February 2016. Combined with the increased requirement for loyalty to the mobile network, this correction resulted in a reduction of the total estimate of overnight stays of approximately 50%. The impact on overnight stays is significantly greater than on arrivals. The average length of a holiday stay in France is 6.8 days for all international customers (DGE, 2017), while residents spent almost all overnight stays in France. However, the correction remains imperfect since it implies excluding from the measurement tourists who stay in France for a period of more than one month, whereas the statistical definition includes stays of up to one year.

Insofar as the anonymity requirement prevents individual data relating to connections between mobile phones and cell towers from being stored, the alteration of the definitional

12. See <http://www.tourisme-territoires.net/zoom-sur-le-projet-flux-vision-tourisme/>

Figure IV
Tourist Overnight Stays Estimated Based on Mobile Phone Data



Coverage: Daily overnight stays of non-resident tourists travelling in France.
Sources: Banque de France, DGE.


criteria used to define relevant behaviours for monitoring tourism cannot be projected backwards. The effect of the corrections described is in evidence in Figure IV in the form of the breaks in the series observed between November 2015 and February 2016.


The Specific Problem of Behavioural Adjustment Requires an Exogenous Collection

Since an understanding of mobile phone usage behaviour among visitors was not enough to satisfactorily adjust the counting of SIM cards, the Banque de France and the DGE made the decision to incorporate a set of questions relating to the use of mobile phones into the EVE survey questionnaire. These questions (Box) were added in January 2017 and the first results will be available for analysis at the end of the year. It will then be possible to determine a coefficient for each of the main countries of residence and thereby to improve the adjustment by country of residence. However, while usage behaviours have been found to vary widely both temporally and spatially, the use of mobile phone data will prove more costly because of the requirement for a collection dedicated to adjustment. This will have an impact on the cost-benefit assessment to be carried out at the end of the experiment in order to decide whether or not to incorporate these data in the current production process.

Among the difficulties related to ownership and use rates are the family composition of tourist groups, which is likely to play a significant role: it is not simply a matter of estimating the number of individuals carrying mobile phones, the number of accompanying people also having to be measured. The issue of the impact of group composition is not specific to the measurement of foreign tourist traffic. Contextual data are available for French tourists: according to a study based on the permanent tourist demand monitoring (SDT) scheme carried out by the DGE and the Banque de France in the spring of 2015, while the mobile phone ownership rate is relatively consistent among residents aged 15 years and over, it varies very significantly up to the age of fifteen. The number of tourists aged under fifteen years accompanying a tourist aged over fifteen also depends heavily on the period (school holidays or term time), the type of accommodation (hotel/campsite/rental) and, therefore, the area. Inclusion of this impact will also be an important stage in improving the measurement of French tourist arrivals. The tourists-to-SIM cards ratio is inevitably higher in a family-friendly camping area at the height of the summer season than outside school holidays in an area dominated by work-related tourism and its representatives equipped with several mobile phones, or even with mobile phones equipped with several SIM cards; in such cases, the difficulty is to have suitable and sufficiently refined adjustment factors.

Box – Questions on Mobile Phone Usage, EVE 2017 Collection


 France is carrying out a trial to count the number of foreign visitors from the number of foreign mobile phones in the country. These three questions will help us to produce our statistics much quicker.

27 Thinking about you and those travelling with you, how many mobile phones did you have during your stay? *If you didn't have any, please put 0. One mobile with two SIM cards counts as 2.*  mobiles


28 During your stay, you mostly used your mobile(s):

With your usual mobile phone plan	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Only with Wi-Fi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
With a prepaid card bought in France	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

29 Still thinking about your stay, this/these devices were...

During the day: 

Switched on most of the time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Switched on from time to time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Switched off	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

At night: 

Switched on most of the time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Switched on from time to time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Switched off	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Thank you and have a good trip!

Future Developments and Uncertainties

The Abolition of Roaming Charges in the European Union

The abolition of roaming charges within the European Union has been in force since June 2017. Operators began to limit such charges to certain destinations a number of years ago. Consequently, tourist behaviours within the European Union are expected to become more standardised, the assumption being that everyone within the EU will eventually come to use their mobile phones as if they were in their country of habitual residence. This would help to improve the accuracy of estimates for European countries, although the transition phase between current habits and the anticipated standardisation may result in noise in the estimates, the expected increase in the use of roaming mobile phones abroad being likely to lead to an artificial increase in the number of visitors recorded.

However, this risk should be put into perspective since the system is not only based on actual mobile phone usage but also on passive connections to the network, which are not charged to users; therefore, the impact will not necessarily be significant. Furthermore, inasmuch as some operators already no longer charge roaming fees to their subscribers on internal destinations within the European Union, the transition phase has been underway for many months and its impact will be spread out over an extended period. The data collected among tourists as part of the EVE survey should enable such behavioural changes to be measured.

While European customers account for approximately 79% of foreign tourist arrivals in France (DGE, 2016b), customers from further afield represent a not insignificant and indeed growing proportion of total visitor numbers. For these customers, roaming charges are likely to remain high and to deter a large proportion of tourists from using their original SIM cards.

WiFi Connection

Specific connection practices may limit the scope of the measurement method based on mobile phone networks. Thus, some tourists, such as North American and Chinese travellers,

are already opting for WiFi hotspots and connection to an Internet network in order to use specific voice communication applications without having to connect to a mobile phone network. The use of such applications, popular among young travellers and technophiles, has so far eluded measurement. In addition, the availability of a WiFi connection is identified by tourist sites as a key pull factor and technical solutions are flourishing, for example in coastal areas. The dedicated questions added to the EVE survey in early 2017 should help to better gauge the scale of the phenomenon.

Supranational Contracts

Another development affecting the accuracy of the system (rather than its exhaustiveness) is the rise of supranational contracts, which is likely to weaken the link between the country of the detected SIM card and the country of residence of its user. This development may be further encouraged by the abolition of roaming charges within the European Union (see above), although the ban comes with some restrictions. These restrictions are designed to dissuade extreme uses (characterised by minority usage in the country of issuance of the contract). The difficulty of inferring the user's place of residence from the nationality of the SIM card is already established, as evidenced by the very high counts of overnight stays seemingly by visitors from Luxembourg in France: this finding – one of the first of the experiment carried out by the Banque de France and the DGE – has proved resistant to the various improvements made to the method. Therefore, the only explanation lies in the fact that Luxembourg companies sell contracts used by residents of other countries, which is no doubt connected to the number of cross-border workers in Luxembourg, who may reside in France but also in Belgium or Germany.

Assessment of the Experiment and Avenues for Further Improvement

The Type of Partnership Put in Place and the Research Approach Appear Suited to the Specific Features of These Big Data

The experiment suggests a number of key success factors for a partnership between a private company holding Big Data and the institutions responsible for producing official statistics. In the case of this ongoing

experiment, the research was based on a tried and tested data formatting module used to serve different needs to those set out here (analysis of events within a limited area – city, department, etc. – as opposed to a level-based assessment of visitor flows and length of stay across metropolitan France as a whole). The advantage is that datasets were available from the very outset of the partnership, which was conducive to an empirical approach and to comparison with the reference data held by the Banque de France and the DGE. The drawback lies in the low significance of the initial results in a context in which the upstream processing stage (phases 2 to 4) depended entirely on the supplier's expertise. This particular obstacle was overcome by putting in place a co-development approach. This requires that the parties commit proportionate and balanced resources, hence the importance of a partnership management structure that actively involves the experts and draws on the appropriate level of decision-making. In this context, the ability of the two parties to make adjustments rapidly (agile method) was found to be essential. For example, the Banque de France and the DGE decided to include a module in the 2017 survey questionnaire that allowed for variables to be collected on mobile phone usage behaviour which, if robust, will improve the potential for exploiting mobile phone data for statistical purposes.

The co-development approach and the agile method also appear to be adapted to the specific features of Big Data: the observation of very frequent events across the entire territory of metropolitan France requires greater computational capabilities than those required for current processing, hence the importance of high reactivity to adapt computational capabilities. Some definitional adjustments require a period of examination of variants, which is inherent to this type of experiment. Definitional stability, which is necessary to the construction of series and their comparison with existing sources, is not immediately achievable, the implication being that researchers need to be willing to interpret successive results that incorporate changes in method, which requires a high level of interaction between them. A test by sampling or by restricting the experiment to a limited territory would have served to mitigate this problem, but would not have achieved the goal of exhaustive measurement across the entire territory, which is one of the main contributions which these data are expected to make.

The Advances Achieved During the Experiment Open Up New Possibilities for Short-Term Trend Monitoring and the Regionalisation of Tourism Statistics

The setup of data processing provides results adapted to the short-term monitoring of tourist arrivals over the short-term and to the measurement of shocks in the case of one-off events (sporting event, festival, attack, comparison of populations in high season and low season). Excluding the setup period, the speed at which statistics can be generated from mobile telephone data (less than thirty days before the end of the observed month) is a definite advantage compared to survey-based collection methods while also being comparable to bank card data collection methods. Such uses require certain precautions, including the use of trend rather than volume indicators. As a second area of interest, the statistical processing derived from the experiment should provide, for each of the principal countries of residence, a satisfactory distribution of overnight stays across the thirteen metropolitan regions.

However, Long-Term Use Requires Other Improvements

To enable dissemination by different users, future research should enable significant improvements in two areas. The first such improvement concerns the reduction of the overall noise of the measurement. This relates to the very heart of the operator's pre-existing system, which served as a starting point for the experiment and implies changes to the basic algorithms. Bypassing insufficient quality by adapting the definitions of pre-defined behaviours cannot be deemed to be an acceptable solution. The second area of improvement relates to our understanding of mobile phone usage behaviour. The creation of an external database on mobile phone use rates among foreign visitors in France, segmented based on the principal countries of origin, is necessary. Given this, the cost of collecting high-quality external data represents one of the key factors in assessing the value of using mobile phone data. Since such data are intended to reduce our reliance upon or even replace the collection of external data relating to overall traffic by means

of transport,¹³ deploying a large-scale collection method to adjust the collected data to the data which they are designed to replace would hardly be appropriate.

In the short term, and in order to align with the adopted approach aimed at achieving visible progress according to relatively close milestones, the stakeholders in the experiment undertook to test a method designed to combine greater control of raw data quality and the preservation of basic algorithms: in order to limit the noise related to the sporadic connections and excessively short stays, it is necessary to measure a sporadicity rate for each of the foreign operators with a view to retaining the operators most loyal to the Orange network. One strength of this choice is that it will not be defined *a priori* on the basis of preferential roaming agreements, but will instead be measured on the ground. It will also be evolving, with the list of operators included in the counts having to be updated on a regular basis. The question of the representativeness of the different operators will arise, with the level of distinctiveness of customer profiles varying according to whether the operator is low cost, long-standing, targeted at technophiles, etc. While attractive in principle, this new version has yet to be evaluated.

Fulfilling the Initial Objectives of the Experiment Leads to Developing Processing that Stretches the Connection between Big Data and the Statistical Series Produced

At present, mobile phone data do not provide a basis for consolidating the data relating to traffic leaving the territory of metropolitan France. They cannot be substituted for traffic data and the EVE method therefore needs to be maintained in its current architecture.

The proposed solutions to improve the quality of estimates focus on sampling strategies. The selection of foreign operators with the fewest sporadic connections falls under this heading. Another possible solution is to monitor volunteers in order to achieve a better understanding of behaviours. Mobile phone operators are adept at monitoring samples of voluntary users and sell such studies more often than studies which focus on entire populations. Such studies have the benefit of avoiding some of the drawbacks faced when attempting to perform exhaustive measurements, such as the sheer volume of data and the inflexibility of anonymisation algorithms. In the case of tourism,

a method such as this would generate detailed results on mobility behaviours, including, in particular, travel frequency, duration and destination. For operators with access to a network in one or several border countries, monitoring on both sides of the border may be worth considering. Beyond the statistical dimensions (extrapolation of behaviours observed in a sample of volunteers to the entire population), adopting such a method requires a legal framework suited to the processing of data relating to natural persons.

In one sense, the idea of considering a sampling-based approach is a paradoxical outcome of the experiment, the initial motivation being to exploit a source of exhaustive data in a straightforward manner. To go in this direction presupposes evaluating the sustainability, and indeed the transparency, of such an approach, given the speed at which technologies and the associated behaviours tend to change. This could generate significant costs in maintaining the sampling frame, in a context in which official statistics are dependent on their users for clear information about their methods and any changes thereto.

To conclude that it is necessary to implement a strategy which involves the sampling-based processing of Big Data implies renouncing one of their assumed strengths: namely, the rapid generation, based on raw and exhaustive sources, of highly representative and easily interpretable results. In some sense, this amounts to altering them in order to transform them into traditional data or, put differently, data whose use goes hand-in-hand with acquisition and reprocessing costs.

* *
*

The experiment conducted by the Banque de France and the DGE therefore suggests viewing mobile phone data as an additional source of information and not as a source capable of replacing currently existing data collections. The same conclusion was reached by Eurostat

13. These data, and, in particular, the data generated by the Cerema survey for road travel, provide a point of reference for determining the survey design and serve to calibrate the statistical model ensuring the representativeness of the questionnaire data. It should be noted that visitor questionnaires remain necessary for understanding tourist behaviours: expenditure by type, type of accommodation and activities.

in its 2014 report on mobile phone data. The most pertinent uses of such data in the context of international tourism in France are short-term trend analysis and the regionalisation of data from the EVE survey. However, the monitoring of international tourism in France represents a very specific research context on account, first, of the size of the population of interest and its diversity (means of transport,

country of origin, mobile phone behaviour) and, second, of the specific characteristics of the territory (borders, surface area, transit and cross-border work phenomena). Using mobile phone data to produce tourism statistics in levels remains conceivable, subject to improving the algorithms and furthering our understanding of visitor behaviour pertaining to mobile phone usage. □

BIBLIOGRAPHY

Aguilera, V., Allio, S., Benezech, V., Combes, F. & Million, C. (2014). Using cell phone data to measure quality of service and passenger flows of Paris transit system. *Transportation Research Part C: Emerging Technologies*, 43(2), 198–211.
<http://dx.doi.org/10.1016/j.trc.2013.11.007>

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

Akin, D. & Sisiopiku, V. P. (2002). *Estimating Origin-Destination Matrices Using Location Information from Cellular Phones*. Puerto Rico, USA: Proc. NARSC RSAI.
https://s3.amazonaws.com/academia.edu.documents/7109318/PuertoRicoPaper_final.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1549286310&Signature=h7qw7eNszCA7KWGLEd4lvSaLzWw=&response-content-disposition=inline;filename=Estimating_origin_destination_matrices_u.pdf

Banque de France (2015). *Méthodologie – La balance des paiements et la position extérieure de la France*.
https://www.banque-france.fr/sites/default/files/media/2016/11/16/bdp-methodologie_072015.pdf

Bonnel, P., Hombourger, E., Olteanu-Raimond, A.-M. & Smoreda, Z. (2015). Passive Mobile Phone Dataset to Construct Origin-Destination Matrix: Potentials and Limitations. *Transportation Research Procedia*, 11, 381–398.
<https://doi.org/10.1016/j.trpro.2015.12.032>

Calabrese, M., Di Lorenzo, L. & Ratti, C. (2011). Estimating Origin-Destination Flows Using Mobile Phone Location Data. *IEEE Pervasive Computing*, 10(4), 36–44.
<http://dx.doi.org/10.1109/mprv.2011.41>

Calabrese, M., Di Lorenzo, G., Ferreira Jr., J. & Ratti, C. (2013). Understanding Individual Mobility Patterns From Urban Sensing Data: A Mobile Phone Trace Example. *Transportation Research Part C: Emerging Technologies*, 26, 301–313.
<https://doi.org/10.1016/j.trc.2012.09.009>

DGE (2016a). *Chiffres clés du tourisme*. Édition 2016.
https://www.entreprises.gouv.fr/files/files/directions_services/etudes-et-statistiques/stats-tourisme/chiffres-cles/2016-Chiffres-cles-tourisme-FR.pdf

DGE (2016b). *Le 4 pages de la DGE*, N° 60.
<https://www.entreprises.gouv.fr/etudes-et-statistiques/4-pages-60-touristes-etrangers-france-2015>

DGE (2017). *Le 4 pages de la DGE*, N° 71.
<https://www.entreprises.gouv.fr/etudes-et-statistiques/4-pages-71-touristes-etrangers-france-2016>

Eurostat (2014). Feasibility Study on the Use of Mobile Phone Positioning Data for Tourism Statistics. *Consolidated Report Eurostat Contract N° 30501.2012.001-2012.452*
<http://ec.europa.eu/eurostat/documents/747990/6225717/MP-Consolidated-report.pdf>

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

Kroon, J. (2012). Mobile Positioning as a Possible Data Source for International Travel Service Statistics. United Nations, Economic Commission for Europe, Geneva, Switzerland, 31 October-2 November 2012, *Seminar on New Frontiers for Statistical Data Collection*. <https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2012/mtg2/WP6.pdf>

ONS (2016). Statistical uses for mobile phone data: literature review. *Methodology working paper series* N° 8. <https://www.ons.gov.uk/methodology/methodological-publications/generalmethodology/onsworkingpaperseries/onsmethodologyworkingpaperseriesno8statisticalusesformobilephonedataliteraturereview>

Organisation mondiale du tourisme. *Comprendre le tourisme : glossaire de base*. <http://media.unwto.org/fr/content/comprendre-le-tourisme-glossaire-de-base>

Tourisme & Territoires. *Zoom sur le projet Flux Vision Tourisme*. <http://www.tourisme-territoires.net/zoom-sur-le-projet-flux-vision-tourisme/>

Widhalm, P., Yang, Y., Ulm, M. Athavale, S. & Gonzales, M. C. (2015). Discovering Urban Activity Patterns in Cell Phone Data. *Transportation*, 42(4), 597–623. <https://doi.org/10.1007/s11116-015-9598-x>
