

Research article

Mobile phone data and tourism statistics: a broken promise?

Laura Grassini^{1,*}and Gianni Dugheri²

¹ Department of Statistics, Computer Science, Applications, University of Florence, Italy

² Statistics Department of the Municipality of Florence, Florence, Italy

* **Correspondence:** Email: laura.grassini@unifi.it; Tel: +390552751500.

Abstract: Mobile phone data represents an original source of information about the movements of individuals across territories and, in particular, of visitors. The systematic use of this data for statistical purposes is expected to provide several advantages: timeliness, deeper geographical and time granularity, and reduction of the statistical burden of respondents. To check whether that expectation is well placed, this paper reports a bibliometric analysis and a subsequent literature review of the recent contributions to the use of mobile phone data in quantifying the volume of tourist flows. The main findings show that the systematic exploitation of mobile phone data for producing tourism statistics is still limited in terms of countries (Estonia, Indonesia) and domain (international flows). Furthermore, the basic definitions of visitors stated by the EU Regulation 692/2011 are rarely applied on mobile phone data, and the population of visitors/tourists is often derived as a residual group after having identified the other people movements (residents, commuters). Both the literature review and a brief case study of the Metropolitan City of Florence show the main weaknesses of mobile phone data, which include costs, privacy restrictions, statistical issues of representativeness, among others. Finally, it is clear that mobile phone data cannot completely substitute current surveys on tourism flows because they do not include some noteworthy information concerning a tourist's motivations and a trip's characteristics.

Keywords: big data; tracking technologies; Firenze; people movements; Biblioshiny; tourism statistics

JEL Codes: C8, C10, Z3

1. Introduction

In the last decade, the rapid advances and widespread use of tracking technologies (GPS, mobile phone positioning, georeferenced social data) have opened up new development potentials in the analysis of tourism with particular reference to the movement of visitors in a territory, monitoring of access to local attractors, connectivity between communication tools for enhancing

tourists' experiences (e.g., augmented reality), and support of Destination Management Organizations' activities.

As smartphones have become the preferred devices for accessing the Internet (Wang et al., 2012), they play an important role in providing information for tourism, especially about tourists' behavior. On the other hand, the increasing number of short trips, same-day visits, and non-rented accommodations (i.e., accommodation provided by family/friends, second homes) requires the production of tourism statistics to be adapted to the changes in tourist behavior (European Commission, 2011), thus communication devices are becoming a promising source even for large scale surveys (Daas et al., 2015; Eurostat, 2014b,c).

National Statistical Institutes (NSI) around the world are still experimenting with the use of mobile phone data (MPD) within the wider project of the exploitation of big data. Big data constitutes an information platform capable of feeding different forms of knowledge generation. The UNECE classification of big data identifies three major categories: social networks (blogs and comments, sms, etc.), traditional business systems (commercial and credit card transactions, etc.), Internet of things, the last one including also MPD for tracking purposes. Many sources of big data fall into the category of "administrative data". And just like administrative data, big data is analyzed according to survey frameworks that are different from those they were collected for. That is the real challenge, together with their integration with other forms of data sources such as official statistical data. For Europe, it is worth mentioning the Call launched by Eurostat in 2012 (Eurostat, 2014b), the UN Global Working Group on Big Data (UN, 2019), the analogous UNECE Working Group (UNECE, 2019), the ESSnet Big Data II project (Nurmi, 2018; Nurmi and Piela, 2019) *, and the Global Forum on tourism statistics in 2015 and 2016.

To date, several review studies on the use of MPD in tourism have been published: Kim and Law (2015), Law et al. (2018), Ortega-Fraile et al. (2018), Shoval and Ahas (2016), and Yuan et al. (2019). In those reviews, the perspective was that of marketers and consumers as the analyzes were mainly focused on practices in hospitality and tourism, and visitors' behavior. In all, the major findings proved that MPD have many effective applications in tourism practices, but there is still the question of whether they provide valuable support for the production of official tourism statistics.

The aim of this paper is to carry out a review of the experiences of the use of MPD in the perspective of quantifying tourist flows for the production of tourism statistics.

The paper is organized as follows. The next section summarizes the phenomena to be measured according to the EU Regulation 692/2011 (European Commission, 2011) and discusses the main strengths and weaknesses of MPD in tourism statistics. Moreover, section two provides an empirical evidence of the use of mobile phone data the quantification of tourists' flows in the Metropolitan City of Florence. Section three includes a bibliometric analysis of the documents retrieved from the Scopus database (years 2002–2021). The R package *bibliometrix* with the related Shiny App *Biblioshiny*, has been employed for that purpose (Aria and Cuccurullo, 2017). Moreover we carry out a literature review by analyzing a subset of retrieved documents, more focused on the issue of the production of tourism statistics. The major findings show that, after a great interest in the topic, limitations and costs of accessing data, the complexity in their management and preprocessing make the use of MPD still limited in terms of domain (only international tourism) and place (Estonia, Indonesia). However, MPD also is used for the compilation of the Balance of Payments (Saluveer et al., 2020). Concluding remarks are reported in the final section.

* https://ec.europa.eu/eurostat/cros/content/essnet-big-data-0_en.

2. Major phenomena in tourism and statistical measurement: current surveys and MPD

2.1. Major phenomena in tourism

In this section we briefly remind the fundamental concepts underlying the statistical measurement of tourist flows, emphasizing how such concepts are operationalized in current surveys and in the exploitation of MPD. The basic reference is the EU Regulation 692/2011 that, across EU member States, is the common framework for the systematic development, production and dissemination of statistics on tourism.

Article 2 of the cited Regulation establishes the measurement of the following tourism phenomena (European Commission, 2011; Eurostat, 2014a): *domestic tourism* (visits within a country by residents); *inbound tourism* (visits to a country by not residents); *outbound tourism* (visits abroad). The pivotal concepts involved in the identification of the above types of tourism is *usual environment* intended as the geographical area (though not necessarily a contiguous one), within which an individual conducts their regular life routines. From that, *visitors* identifies those who take a trip outside their usual environment; the term *tourists* indicates visitors with at least one overnight stay but a stay less than one year; *same-day visitors* are visitors with no overnight stay.

2.2. Current surveys

Two broad types of data sources are provided according to the cited Regulation: *demand-side* and *accommodation* data.

Demand-side data concern the domain of domestic and international (inbound, outbound) tourism (Table 1), and are recorded from the perspective of visitors through traditional household surveys (survey on residents, and the so called *border surveys*). Such surveys also gather a valuable set of information on individual and trip's characteristics, including the purpose of the visit. In addition, surveys on residents also measure the non-participation in tourism. With regard to non-sampling errors, the most important is likely the recall bias as a significant number of respondents may not accurately remember the number of visits, time and length of the trip, etc. Additional drawbacks are concerned with limitations in time and space granularity of the data (besides the issues of timeliness typical of sample surveys).

Accommodation data are recorded by official establishments (defined in Article 2 of the cited Regulation) and refer to domestic and inbound flows (Table 1). Such data report the number of customers (i.e., arrivals) and overnight stays classified by the features of establishments such as type (hotel, B&B, etc.), quality (number of stars), etc. Accommodation data fail to capture same-day visitors (as there must be at least one night spent) and unobserved tourism: that is, nights spent in unofficial accommodation (e.g., secondary houses, friends' and relatives' homes), and nights spent in official establishments deliberately unreported for fiscal reasons (De Cantis et al., 2015; Guizzardi and Bernini, 2012). In some EU countries, such as Italy, accommodation data are produced by a total survey. The minimal geographical scale is the municipality (LAU2); the georeferencing of accommodation establishments would improve data granularity, without resorting to big data (Othman et al., 2010). However, confidentiality issues do not allow this, nor is the release of tourist flows data for small municipalities allowed.

Table 1. Tourist flows and related statistical sources within brackets.

	To country A	Abroad
From Country A	Domestic flows (resident-household surveys and accommodation data)	Outgoing flows (resident-household and border surveys)
From Abroad	Inbound flows (border surveys and accommodation data)	—

Note: our elaboration of Figure 1–1.2 in Eurostat (2014a).

With regard to the identification of usual environment, no rule is applied in accommodation data as each customer is considered as a tourist. Conversely, in demand-side data, the Eurostat manual (Eurostat, 2014a) recommends, as a general rule, “to leave the interpretation of the usual environment to the subjective feeling of the respondent and to encourage a spontaneous reply”. Only in case of doubt, should the interviewer apply four objective criteria in a cascade manner (Antolini and Grassini, 2020; European Commission, 2011; Masiero, 2016): 1) the crossing of administrative borders (or the distance from the place of usual residence); 2) the duration of the visit; 3) the frequency of the visit; and 4) the purpose of the visit. For international trips, Eurostat (2014a) recommends a minimum threshold of 3 hours.

2.3. Mobile phone data

Several documents produced by NSIs can be referred to for the experimentation of MPD in tourism statistics. The main publications are some reports by Eurostat and the UN Global Working Group on Big Data for Official Statistics (Eurostat, 2014b,c; UN, 2019). In addition, we should also mention the ESSnet Big Data II project within the European Statistical System, under which a cooperation network between the Nordic EU countries was established[†].

MPD used for statistical purposes are generated by active or passive events. Active events consist of actions that are initiated by or targeted to a subscriber or mobile device (and generally reside in the billing system of the MNO: Mobile Network Operator). The billing domain stores call detail records (CDR) with also time attributes, and location information. Passive events consist of signaling or probing data that derive from a MNO’s activity of monitoring the network performance. They allow a major time granularity of information and do not depend on the behavior of the subscriber (who might have diurnal or daily differences in the use of mobile phone), but increase considerably the number of events, requiring a more capable technology for handling very huge datasets.

Tighter data can be produced with the direct involvement and agreement of the tourists, who are called to use special apps for tracking their movements. Those apps can produce continuous, fine grained data that can determine tourists’ movement and dispersal (Birenboim, 2016; Hardy et al., 2017). However, the use of such technologies is generally circumscribed to local attractions or involves a sample of subscribers in the case of larger territories. In some studies, enhanced mobile phones have been used to assess, with a cross check, the type of information derived from cell data (Raun et al., 2016).

In the followings, we refer to CDR data that are the most used information for statistical purposes. A crucial issue is the identification of the visitor (i.e. one-day visitor or tourist). That task requires the

[†] <https://norden.diva-portal.org/smash/get/diva2:1413891/FULLTEXT01.pdf>.

identification of the subscriber, with location and time attributes which in turn allow the recognition of subscriber's usual environment.

Regarding the identification of the subscriber, the identifier should be unique in the time span of the analysis. Major drawbacks are concerned with the facts that people may carry several devices with them (Dattilo et al., 2016). Moreover, any socio-demographic attribute of the subscribers—that we currently find in survey data—is usually collected in an external database to CDR, which the MNO uses for customer profiling.

About location, for domestic and inbound roaming data, the geographical reference is the coverage area of the network cell. The geographical reference in outbound roaming is limited to foreign countries and not smaller geographical regions (Eurostat, 2014c). However, due to the limiting information of the roaming service, any frequent trip abroad tends to expand the size of the usual environment to include the foreign country. Another case is when a foreign tourist buys a SIM of the country visited: they ceases to be a tourist. Of course, the location of CDR data is less accurate than the one measured with GPS. In particular, cell data cannot directly produce information at the LAU2 or lower level; thus, a data pre-processing which takes into account the distribution of cells in the municipal area is required. In this respect, there are significant statistical challenges not overcome yet; referring cell data to the territory, and inferring the reference population from the number of identified subscribers imply the use of specific statistical spatial models. For this purpose, the market share of the service provider is generally used. Thus, an additional issue stems from the increasing fragmentation of the market that would force NSIs to take agreements with several providers: in fact, in 2010, the market share of the top provider was 33% in Italy and Germany, 41% in France and 44% in Spain[‡].

The precision of the location is also associated with the size of the area: “[B]ecause visitors do not use a mobile phone at every place, the smaller the geographical level, the less representative the mobile data will be when it concerns tourism visits to those places” (Eurostat, 2014c). Another significant difficulty is the identification of the *main destination*, which is generally the place where the last event is located (Eurostat, 2014c), whilst this is directly indicated by the respondent in traditional surveys.

Together with location, time information is crucial for the identification of the usual environment, which is certainly the major concern in the use of MPD. A long-term unique continuous ID has to be used to identify the subscriber in order to detect a meaningful place (Eurostat, 2014c) or the so called *anchor point* (Ahas et al., 2010). This last study applied an observation length of 12 months, according to the Regulation principles. In another research (Seynaeve et al., 2016), the usual place of residence was roughly approximated by the place where the subscriber is most often observed at 4 a.m. over a given period of time. In practice, the municipality border criterion is difficult to apply as its success depends on the density of antennas in the territory.

By taking into account, subscriber ID, time and location, the identification of visitors/tourists is mostly made in a residual way, after having identified the other sub-populations present in the territory (residents, commuters, etc.).

Comparatively to accommodation statistics, MPD can identify same-day visitors, as well as those visitors who spend nights within their usual environment but are not able to distinguish whether they stay in a hotel or at home. There are studies aiming at quantifying sub-populations (visitors, etc.) by inferring human behavior from the buildings where subscribers stay (e.g., bank, public office, hotel);

[‡] Eurostat Telecommunication Statistics: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Telecommunication_statistics&direction=next&oldid=153514.

however, the feasibility of those approaches is questionable (Osaragi and Kudo, 2020). Comparatively to survey data, MPD fail to provide information on the characteristics of the tourist, of the trip, on the reasons for the trip. Another drawback of MPD (in comparison to both official data sources) is the impossibility of distinguishing the type of accommodation facility (hotel, bed and breakfast, etc.).

But, overall, the biggest challenge in the use of MPD is concerned with the restrictions due to data protection regulations that hardly allow access to the original MNO's data (even if anonymized). Regarding privacy restriction, some scholars argued that new forms of "data anonymization" should be explored to achieve a balance between the societal value of statistical data and the protection of privacy of individuals (de Montjoye et al., 2018).

Another noteworthy issue is the expensive cost related to data preprocessing. Basically, MNOs provide either event-based or aggregated data. For the purpose of producing statistics, aggregate data significantly reduces the options and lowers the quality (e.g., a longitudinal analysis cannot be carried out). However, this is the easiest option for obtaining data from MNOs.

At present, only two countries currently apply mobile phone technology for providing official statistical data: Estonia and Indonesia, who in 2017, also signed a Cooperation Contract for defining a "Quality Assurance Framework and Methodology Concept Development for Mobile Positioning Data." In Estonia, the private company Positium (originally a spin-off of University of Tartu) is developing methodology and technology for processing MPD for human mobility analyzes and statistical indicators. Data on international tourism are produced, that also serve in the estimation of the Balance of Payments. The great experience of Estonia, thanks to Ahas and collaborators' works, has provided a noteworthy wealth of knowledge even in the Eurostat project already mentioned.

In Italy, the use of MPD for producing official statistics is part of a wider project on the exploitation of new data sources (i.e. big data) for various domains: nowcasting of economic phenomena, small area estimates of unemployment, consumer price indexes, etc. In 2015, ISTAT (the Italian NSI) participated in the mentioned task force promoted by UNECE and launched a series of experiments, among which was the project *Persons and Places*, that should integrate the information on mobility collected through the Permanent census of population and housing (ISTAT, 2014).

2.4. A case study of MPD use: the Metropolitan City of Florence

In 2017, the Statistical Office of the Metropolitan City of Florence began an experiment in the use of MPD for quantifying the population present in its territory. A major concern was the presence of temporary visitors (i.e., same-day visitors and tourists), as the crowding in the UNESCO site of the city was becoming relevant. Vodafone released pre-processed data related to May-September of 2016 (five months). For privacy purposes and because Vodafone had already started an experiment of MPD preprocessing, released data were aggregated at the level of the census area (23 areas) with daily 6-hour slots (Morning 6.00–12.00 a.m., Afternoon 12.00–18.00 p.m., Evening 18.00–00.00 p.m., Nighttime 00.00–6.00 a.m.). The ACE average surface is 2.91 sqkm and the average population is 17,232 residents. Released data included two datasets of profiled subscribers: 1) unique subscribers and 2) statistical presences (each subscriber was weighted with time of stay). The profiled groups provided by Vodafone were as follows.

1. Residents: subscribers whose telephone "resides permanently" in the city in the survey month and in the previous and following month. The category includes registered residents who actually live

in the city and non-registered residents who usually live in the city. Conversely, it does not include registered residents who do not live in Florence.

2. Commuters: subscribers whose telephone is “resident” outside the municipality of Florence but who come regularly to Florence (at least 3 days a week and for 12 weeks, in the reference period of May-September 2016).
3. Italian visitors: subscribers whose phone is “resident” outside the municipality of Florence, who have occasionally come to Florence (not necessarily once) in the period.
4. Foreign visitors: all individuals whose cellular phone is “resident” outside Italy and who have occasionally come to Florence in the period.

By considering: 1) the number of Italian mobile phones managed by Vodafone (compared to the total), 2) the number of foreign mobile phones that are in roaming with Vodafone network (compared to the total), and 3) the numerical ratio between cell phones used and resident population (children, adults, elderly), Vodafone has developed an algorithm that makes it possible to report the MPD data to the population, but whose operation is unknown because it is protected by trade secret.

Examining the 2016 data, MPD patterns are broadly consistent with official accommodation data. Most of the foreigners were mainly in the historical center, with a great number of those visitors from the United States, followed by France and the United Kingdom. Most of the Italian tourists came from the regions of Tuscany, Lazio, Lombardy, and Emilia-Romagna.

Table 2 reports the monthly number of overnight stays of Italian and foreign tourists from accommodation data (Official Data: OD) and MPD. Specifically, MPD figures are the sum of the weighted presences of profiled groups 3 and 4, between 0:00 and 6:00 a.m. hours, assuming that whoever was present at that time was sleeping. The comparison involves monthly data, which is the referring period of released OD.

As expected, MPD are greater than OD figures. In particular, Italian MPD figures are more than double of OD ones, denoting the inclusion in MPD of non-official accommodations (e.g., visit relatives, friends, second homes, etc.). Moreover, commuters who go to Florence less than 3 times a week are also included. Conversely, foreign OD and MPD do not differ so much: traditional accommodation establishments may be more common among foreigners (i.e., minor use of non-official accommodation facilities), and with foreigners, there may be a minor under-reporting by firms for fiscal motivations. At the same time, there may be less phone use, as well, especially by people from distant countries (Saluveer et al., 2020).

Seasonal variation is less pronounced for foreigners. Moreover, MPD suggest that seasonality is greater than appears from OD (Figure 1). Both domestic and foreign series exhibit different pattern between May-Jun and Aug-Sep: in the first period, the concentration of overnight stays from MPD is greater than OD whereas the opposite occurs in Aug-Sep. The great discrepancies between domestic MPD and OD percentages raises the suspicion that MPD also include residents who, in August, leave the city for vacation.

Table 2. Metropolitan City of Florence: number of tourists' overnight stays from accommodation data (Official Data: OD) and MPD.

Time	Italian OD	Italian MPD	Italian MPD/OD	Foreign MPD	Foreign OD	Foreign MPD/OD	Total OD	Total MPD
May-16	182,435	537,664	2.95	714,189	864,838	1.21	896,624	1,402,502
Jun-16	186,383	587,400	3.15	686,642	838,890	1.22	873,025	1,426,290
Jul-16	145,003	415,617	2.87	811,144	889,266	1.10	956,147	1,304,883
Aug-16	143,196	294,779	2.06	762,087	833,590	1.09	905,283	1,128,369
Sep-16	164,211	454,350	2.77	728,235	795,750	1.09	892,446	1,250,100

Note: processing of Vodafone data by the Municipal Statistics Office of the Metropolitan City of Florence.

In conclusion, the MPD time pattern reflects the OD one but the comparison is not completely consistent as MPD processed by Vodafone do not follow the same rules of OD. Namely, the algorithm applied by Vodafone is not aimed at identifying visitors/tourists but rather to detect groups of people from their movements. In fact, the most important result of this analysis is the estimate of the number of those who usually live in Florence without being officially registered. This estimate is very difficult to obtain in other ways.

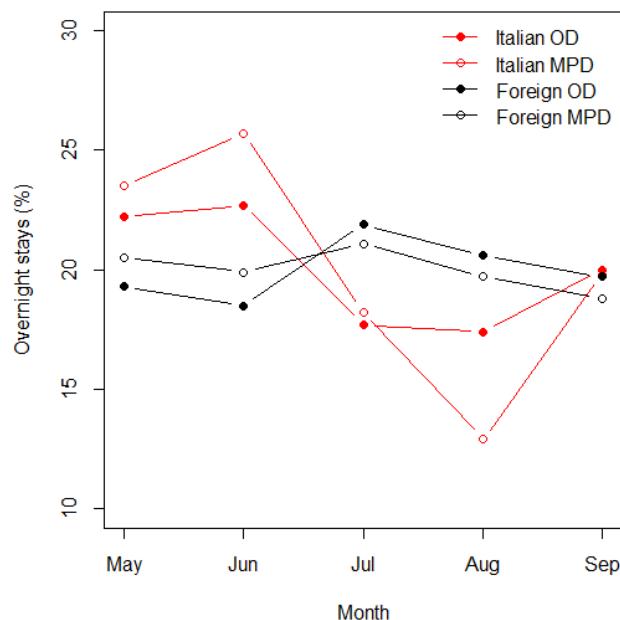


Figure 1. Metropolitan City of Florence. Monthly number of overnight stays (%). Note: our processing of 2016 MPD data.

3. Bibliometric analysis and literature review

The proliferation of contributions on the use of MPD in tourism has become evident, not only in scientific journals but also in reports and congress proceedings. We retrieved data from the Scopus database, an important supplier of scientific literature. The search was performed in August 2020 with

the following query applied to title field, keywords and abstract:

(“mobile positioning” OR “mobile tracking” OR “mobile phone” OR “smartphone”) AND “tourism”.

Through the adopted query we retrieved from Scopus a greater number of publications comparatively to Web of Science (WoS). As the merging of Scopus and WoS data is not completely feasible as some bibliometric statistics are not consistent with each other's, we used only Scopus data.

The 597 papers (corresponding to 1672 authors) retrieved were published between 2002–2021. Out of 597, 279 (46.7%) are journal articles, and 251 (42%) are conference papers, which shows how the topic is still in development with pilot and feasibility studies hosted in scientific meetings. As can be seen in Figure 2, there is a sharp upward trend in the yearly number of publications with a weak decrease in 2013–2015 and 2019, but perhaps, more recent data are not complete yet. The number of papers by the country of the corresponding author is reported in Table 3: China and Italy are at the top positions. The average number of authors per publication is 2.8, meaning that this kind of research is conducted by teams.

The retrieved papers have been published in 389 sources (journals and others), of which the most relevant are reported in Table 4. We argue that the main contributing sources are from ICT and environmental areas, but there is a significant presence of some top journals in tourism. There are 9919 citations, most of them in sources specialized in tourism (Table 5).

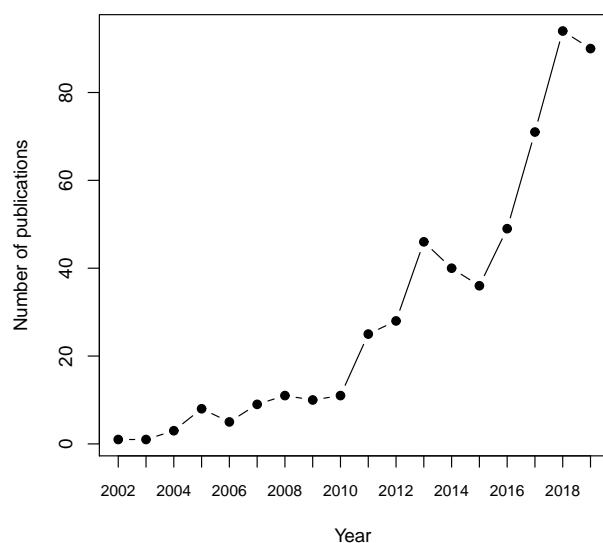


Figure 2. Number of retrieved papers by year. Note: our processing of retrieved Scopus data.

Table 3. Top corresponding authors' countries (total 1672 authors).

Country	Papers
CHINA	67
ITALY	38
INDONESIA	31
SPAIN	27
USA	26
UNITED KINGDOM	29
JAPAN	23
SOUTH KOREA	24

Note: our processing of retrieved Scopus data.

Table 4. Number of papers by source (sources with at least 5 retrieved articles).

Source	Number
Lecture Notes in Computer Science	26
ACM International Conference Proceeding Series	15
Tourism Management	10
Lecture Notes in Electrical Engineering	9
Advances in Intelligent Systems and Computing; Asia-pacifinc Journal of Tourism Research; Communications in Computer and Information Science	8
CEUR Workshop Proceedings; Current Issues in Tourism; IOP Conference Series: Earth and Environmental Science; Journal of Hospitality and Tourism Technology; Journal of Travel Research	7
Applied Mechanics and Materials; Journal of Physics: Conference Series; Sustainability (Switzerland)	6
Annals of Tourism Research; Information Technology and Tourism; International archives of the Photogrammetry Remote Sensig and Spatial Information Sciences; Journal of Travel and Tourism Marketing	5

Note: our processing of retrieved Scopus data.

Table 5. Top cited sources (over 50 citations).

Source	# citations
Tourism Management	605
Annals of Tourism Research	418
Journal of Travel Research	375
Computer in Human Behavior	130
Management Information Systems Quarterly	103
Current Issues in Tourism	99
Tourism Geographies	82
Journal of Travel and Tourism Marketing	70
International Journal of Tourism Research; Journal of Hospitality and Tourism Technology	69
International Journal of Hospitality Management	66
International Journal of Contemporary Hospitality Management	65
Journal of Business Research	64
Information Management; Journal of Marketing Research	55
Information Technology Tourism; Journal of Marketing	54

Note: our processing of retrieved Scopus data.

The results in Table 5 are consistent with the list of the most frequent keywords (i.e. counts greater than 10), which cover a total of 1477 occurrences. 758 occurrences (51.3%) are due to technical terms (phones, telephone sets, mobile communication systems, etc.), 359 (24.3%) to general terms used in

tourism studies (tourism industry, destinations, attractions, etc.), 286 (19.4%) are concerned with tourist and travel behavior (including 68 papers on “augmented” or “virtual reality”), and 25 (2%) to population statistics and surveys.

Due to these results, we applied an extensive reading of papers’ title and abstract in order to assess which of the retrieved publications could contribute most to the production of tourism statistics; although in some cases, these were circumscribed to local areas. We found 26 most significant papers, which are reported in Table 6 in chronological order starting from the most recent.

As we can see from Table 6, almost all studies employed CDR data, the exception being Kubo et al. (2020) who used aggregated data of passive events at the cell level. In some contributions, CDR data are accompanied by the enhancement of mobile apps that allow the production of GPS data but that require the direct involvement of the subscriber. For that reason, we identified such papers as “pilot studies” or “studies aiming at specific research questions” to be distinguished from “feasibility studies”, or from those that deal with larger scale experiments.

We can argue that in most cases the detection of tourists requires the simultaneous identification of different groups of people moving in a territory. A relatively easier task is the detection of international flows which appear to be the most frequent topic (13 out of 26).

A group of papers analyzed Estonia and constitutes the most developed works on MPD, as they involve Ahas and collaborators, who began their experiments on MPD in 2005 (Ahas and Mark, 2005). The papers by Ahas et al. (2007) and Ahas et al. (2008) include a systematic study on the use of MPD for producing tourism statistics. In both articles, authors considered roaming data of the foreign mobile phone call activities for 17 months (April 1, 2004 to August 31, 2005; call activities of 1.2 million subscribers from 96 countries). The papers include an in-depth description of the transformation from cell to municipality data, a discussion about the issues of data representativeness, which depends also on the proportion of phones and phone use across nationalities. In Ahas et al. (2007), MPD succeeded in describing different seasonality patterns related to different tourism products (e.g., mountain vs. coastal areas). They obtained a Pearson correlation coefficient between MPD and official data up to 0.99 in some regions. The work by Kuusik et al. (2011) applied the methodology developed by Ahas and collaborators for the identification of repeated visitors among foreign tourists. A repeat visit is detected if the subscriber visits Estonia several times. Nilbe et al. (2014) still analyzed inbound flows and distinguished regular and event visitors by applying the following rule: regular visitors are visitors who went to Estonia, but did not visit the events analyzed (i.e., no call activities occurred at any of the events considered). Raun and Ahas (2016) assessed CDR roaming data for counting the visitors in the Estonian counties, without disentangling sub-populations of movers. It is a study at large scale in terms of time (three years) and space (entire country). Authors found that the monthly number of nights spent in Estonia generated from MPD was strongly correlated (Pearson correlation coefficient=0.96, 36 monthly data), with the number of nights spent from official accommodation statistics. In another study, Sikder et al. (2016) faced the problem of identifying tourists. They applied several rules on CDR data by tracking people movements, and developed a 0-100 score measuring the likelihood of a subscriber to be a tourist. Finally, Saluveer et al. (2020) derived estimates of inbound and outbound tourists for Estonia by using roaming and domestic data which served for checking the periods spent abroad by Estonian residents. Four different groups were detected: 1) transit visitors, 2) migrant workers (including students), 3) cross-border commuters, and 4) tourists as the remaining group. According to Eurostat (2014a), the authors applied a period greater than 3 hours for inbound or outbound visits. Those authors

also compared time patterns of mobile phone estimates and current estimates from official data (even in this case the Pearson correlation coefficient between the number of nights spent at accommodation and MPD was high).

Other contributions analyzed Indonesia, Slovakia, Spain, France, Latvia, and South Korea.

Uluwiyah and Setadi (2017) aimed at quantifying international tourism in Indonesia. Usual environment was defined by the following rule: if a SIM was detected consecutively for 7 days or intermittently more than 20 days per month, the user was not considered a tourist. The authors discussed the representativeness of such data, and the estimation of foreign tourists was achieved also by using data from the Immigration office (i.e., integration between MPD and administrative data).

Table 6. Main features of the 26 papers retrieved from Scopus database.

Author (year)	Source	Data	Population groups	Territory/ scale	Aims/ findings
Xu et al. (2021)	Tour. Man.	CDR records	inbound flows	South Korea	Tourists dispersal
Arhipova et al. (2020)	Exp. Syst.	cell phone data aggregated at 15 min interval	detecting activities (and possibly tourism activity)	Latvia	Feasibility study
Hardy et al. (2020)	Ann. Tour. Res.	GPS and survey (app) data ($n < 500$) aggregated data at cell level (passive events)	inbound flows	Tasmania	Tourists dispersals
Kubo et al. (2020)	Tour. Man.	Roaming data	domestic, inbound	Japan (536 beaches)	Recreational value of beaches
Park et al. (2020)	Ann. Tour. res.	CDR records	inbound flows and outbound flows	South Korea	Detecting hot spots
Saluveer et al. (2020)	Ann. Tour. Rees.	Geograficky Casopis	inbound	Estonia	Feasibility study
Šveda et al. (2019)	Adv. in Transp. Police and Plann.	CDR dataset with internet usage (11 days)	detecting activities	Slovakia	Feasibility study
Wang et al. (2019)	ISPRS In. Jour. of Geo-Information	CDR data	Group movement pattern	Hainan (China)	Association between mobile usage and places visited
Zhu et al. (2019)	IEEE Conference Paper	CDR data	detect population groups (including tourists)	Hainan (China)	Feasibility study
Cousin and Hillaireau (2018)	Economie et Statistique (INSEE)	aggregated CDR data (29 provinces, 30 days)	inbound visitors	France	Feasibility study
García et al. (2018)	Boletin de Estadística e Invest. Oper.	CDR data and passive events	detect population groups (including tourists)	Spain (NUTS 2)	Feasibility study
Park et al. (2018)	Tour. Man.	CDR records	inbound flows	City (South Korea)	Planning non-stop flights
Rodríguez et al. (2018)	Sensors	GPS data (special App)	mobility behavior (sample)	Zeeland province (NL)	Answer research question
Wang et al. (2018)	Transp. Res. Part C	CDR data with internet usage	detecting activities	Shanghai	Mobile usage and places visited
Yun et al. (2018)	Asia Pacific Jour. Tour. Res.	GPS data	sample of involved visitors	Seoul	Answer research question
Hardy et al. (2017)	Inf. Tech. & Tour.	Mobile App - GPS data (472 tourists involved)	Tracking people movements	Tasmania (local area)	Pilot study

Continued on next page

Author (year)	Source	Data	Population groups	Territory/scale	Aims/findings
Uluwiyah and Setadi (2017)	IWBIS 2017	CDR data	inbound visitors	Indonesia	Feasibility study
Raun et al. (2016)	Tour. Man.	CDR sample (roaming) data	inbound flows	Estonia (county)	Feasibility study
Sikder et al. (2016)	IWBIS 2016	CDR data	Tracking people movements	Bangladesh	Feasibility study
Phithakkitnukoon et al. (2015)	Jour. Pervasive and Mobile Computing	GPS data (over 130,00 users)	Tracking people movements	Japan	Feasibility study
Nilbe et al. (2014)	Journ. Urb. Tech.	CDR roaming data (sample): 3 years	inbound flows: event and regular visitors	Estonia	Answer research question
Järv et al. (2012)	Plos One	CDR data	Domestic tourism through traffic flows	Estonia (Tallin)	Feasibility study
Kuusik et al. (2011)	Baltic Jour. of Management	CDR data	inbound flows	Estonia	Segmentation
Ahas et al. (2008)	Tour. Man.	roaming data	inbound flows	Municipality (Estonia)	Feasibility study
Ahas et al. (2007)	Tour. Man	roaming data	inbound flows	Municipality (Estonia)	Feasibility study

Šveda et al. (2019) defined an incoming tourist in Slovakia as a subscriber of a foreign SIM card with activity in the interval of 2–14 days during the calendar year (2016).

García et al. (2018) identified the group of tourists by distinguishing different people movements. Usual environment (i.e. residence) was assigned empirically, taking into account the different places where the cell phone had made overnight stays (between 0:00 and 8:00) in the previous six months. Inbound tourists seem to take more trips than in official surveys because multiple destinations of one trip are considered as different trips.

Even the paper by Cousin and Hillaireau (2018) aimed at developing a method for quantifying the volume of international flows, and extensively discussed the estimation of the number of foreign mobiles: it requires an adjustment taking into account the service provider's market share of roaming customers, by country of origin and operator in combination. In addition, another statistical adjustment is related to mobile phone usage. The authors concluded that estimates obtained from MPD were not yet of sufficient quality to replace the survey data currently used.

The paper by Arhipova et al. (2020) studied people movements for inferring economic features of Latvia's regions. Although not centered on tourism, this article is interesting, as due to privacy restrictions, it used aggregated data at 15-min intervals in the area of each cellular base station, for the period 2015–2018.

Finally, the work by Xu et al. (2021) included a large scale analysis of inbound flows also detecting their statistical distributions across territories. The number of inbound tourists to destinations follows a log-normal distribution, which indicates a noteworthy heterogeneity in tourism attractiveness of territories.

The remaining papers in Table 6 are pilot studies which used special mobile app with the direct involvement of the subscriber. These papers are useful not only for appreciating the potentiality of mobile data but also for knowing more about tourists' use of smartphone, which can affect data accuracy and reliability.

4. Conclusion

MPD and related tracking technologies are seen as promising sources for large-scale surveys due to the high pervasiveness of cellular phones within society.

The study carried out in this paper has attempted to highlight the recent contributions on the use of MPD for improving tourism statistics. Positive findings are that in many studies, and also in our empirical analysis, MPD show a variation over time consistent with and highly comparable to that of the respective official statistics (accommodation data). Another important fact is that standard procedures have been established for the use of MPD for producing statistics for international tourism, and estimating the Balance of Payments (i.e., Estonia, Indonesia). That experience has also shown how it is possible to share costs over different statistical domains.

However, most feasibility studies look at the correlation between the time patterns of the two official sources (demand-side and, more often, accommodation data), whereas the numerical figures are less discussed (Cousin and Hillaireau, 2018; Eurostat, 2014b; Saluveer et al., 2020). In addition, experiences are mostly concentrated on international tourism as it is relatively easier to intercept, thanks to roaming technology.

Rarely we have found the application of the standard definitions of a visitor, usual environment, etc., as DMOs tend to apply “multi-purpose” processing of data in order to gain efficiency, instead of personalizing the service that would be anyhow highly expensive (e.g. the case study of Florence).

MPD data are collected from the perspective of the tourist and a comparison with demand-side data would be more appropriate although feasible only at the regional/national level. Thus, we can argue that MPD can be used to potentially strengthen current tourism demand surveys through mixed-mode data collection (Eurostat, 2014b). That is, MPD can be seen as an alternative to the “bookkeeping system” or “diary”.

Anyway, major issues still remain, including 1) how NSIs may obtain a more stable and continuous access to the data of MNOs and organize MPD analysis in the wider process of the exploitation of big data to gain in efficiency, and 2) the privacy concerns. With the pandemic experience, people might accept a privacy-consensus use of MPD, as special apps (i.e., COVIDSafe in Australia, IMMUNI in Italy, among others) and statistical models are being used to understand the spread of COVID-19 and to predict and evaluate the impact of containment measures (European Commission, 2020).

After more than 15 years of work, the full exploitation of MPD is still far away, and some expectations have been disappointed. However, we cannot speak of “a broken promise”, and we close the paper citing Ahas et al. (2008): “... The world of ICT and society is developing so fast that today’s problems and limitations will easily be surpassed tomorrow or the day after tomorrow. This also pertains to the topic of data collection and cooperation with mobile operators and the standardisation of data and analysis methods until they reach the level of official statistics”.

A decisive step in the use of MPD could be the involvement of the NSIs because this would guarantee both the homogeneity of the definitions and the algorithms used. The involvement of the Institutes would also guarantee the achievement of substantial economies of scale and, consequently, a significant reduction in costs.

Conflict of interest

The authors declare no conflict of interest.

References

Ahas R, Aasa A, Mark Ü, et al. (2007) Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Manage* 28: 898–910.

Ahas R, Aasa A, Roose A, et al. (2008) Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Manage* 29: 469–486.

Ahas R, Mark Ü (2005) Location Based Services: New Challenges for Planning and Public Administration? *Futures* 37: 547–561.

Ahas R, Silm S, Järv O, et al. (2010) Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *J Urban Techol* 17: 3–27.

Antolini F, Grassini L (2020) Issues in Tourism Statistics: A Critical Review. *Soc Indic Res* 150: 1021–1042.

Arhipova I, Berzins G, Brekis E, et al. (2020) Mobile phone data statistics as a dynamic proxy indicator in assessing regional economic activity and human commuting patterns. *Expert Systems*, John Wiley & Sons. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/exsy.12530> (accessed August 2020).

Aria M, Cuccurullo C (2017) bibliometrix: An R-tool for comprehensive science mapping analysis. *J Informetr* 11: 959–975.

Birenboim A (2016) New Approaches to the Study of Tourist Experiences in Time and Space. *Tourism Geographies* 18: 9–17.

Chen Q, Hu Z, Su H, et al. (2018) Understanding Travel Patterns of Tourists from Mobile Phone Data: A Case Study in Hainan. 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), 45–51. Available from: <https://ieeexplore.ieee.org/document/8367096> (accessed August 2020).

Cousin G, Hillaireau F (2018) Can Mobile Phone Data Improve the Measurement of International Tourism in France? *Economie et Statistique* 505–506: 89–108.

Daas PJH, Puts MJ, Buelens B, et al. (2015) Big Data as a Source for Official Statistics. *J Off Stat* 3: 249–262.

Dattilo B, Radini R, Sabato M (2016) How many SIM cards in your luggage? A strategy to make mobile phone data usable in tourism statistics. 14th Global Forum on Tourism Statistics, Venice, Italy. Available from: <http://tsf2016venice.enit.it/programme> (accessed September 2020).

De Cantis S, Parroco AM, Ferrante M, et al. (2015) Unobserved tourism. *Ann Tourism Res* 50: 1–18.

de Montjoye YA, Gambs S, Blondel V, et al. (2018) On the privacy-conscious use of mobile phone data. *Sci Data* 5: 180286.

European Commission (2011) EU regulation No 692/2011 Of the European Parliament and of the Council of 6 July 2011. *Off J Eur Union* 192/17. Available from: <https://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:192:0017:0032:EN:PDF> (accessed January 2019).

European Commission (2020) *Joint European Roadmap towards lifting COVID-19 containment measures*. Available from: https://ec.europa.eu/info/sites/info/files/communication_-_a_european_roadmap_to_lifting_coronavirusContainment_measures_0.pdf (accessed September 2020).

Eurostat (2014a) Methodological manual for tourism statistics. Version 3.1. Available from: <https://ec.europa.eu/eurostat/documents/3859598/6454997/KS-GQ-14-013-EN-N.pdf> (accessed January 2019).

Eurostat (2014b) Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics. Consolidated Report. Eurostat, Luxembourg, ISBN: 978-92-79-39762-2.

Eurostat (2014c) Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics. Report 3a. Feasibility of Use: Methodological Issues. Eurostat. Available from: <https://ec.europa.eu/eurostat/documents/747990/6225717/MP-Task-3a-report.pdf/38716d6f-c2db-4ee3-92f3-0241bca55f94> (accessed September 2020).

García FC, Valverde MI, Mascuano JP, et al. (2018) Quality Implications of the use of Big Data in Tourism Statistics: Three exploratory examples. *Boletin de estadistica e Investigacion Operativa* 34: 138–148.

Guizzardi A, Bernini C (2012). Measuring underreporting in accommodation statistics: Evidence from Italy. *Curr Issues Tourism* 15: 597–602.

Hardy A, Birenboim A, Wells M (2020) Using geoinformatics to assess tourist dispersal at the state level. *Ann Tourism Res* 82: 1–17.

Hardy A, Hyslop S, Booth K, et al. (2017) Tracking tourists' travel with smartphone-based GPS technology: a methodological discussion. *Inf Technol Tourism* 17: 255–274.

ISTAT (2014) *Linee strategiche del Censimento permanente della popolazione e delle abitazioni. Metodi, tecniche e organizzazione*. ISTAT, Rome. Available from: https://www.istat.it/it/files/2014/11/Cens_Perm_pop.pdf (accessed December, 2020).

Järv O, Ahas R, Saluveer E, et al. (2012) Mobile Phones in a Traffic Flow: A Geographical Perspective to Evening Rush Hour. *PlosOne* 14. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0049171>.

Kim HH, Law R (2015) Smartphones in Tourism and Hospitality Marketing: A Literature Review. *J Travel Tourism Mark* 32: 692–711.

Kubo T, Uryu S, Yamano H, et al. (2020) Mobile phone network data reveal nationwide economic value of coastal tourism under climate change. *Tourism Manage* 77: 1–9.

Kuusik A, Tiru M, Ahas R, et al. (2011) Innovation in destination marketing: The use of passive mobile positioning for the segmentation of repeat visitors in Estonia. *Baltic J Managem* 6: 378–399.

Law R, Chan IIC, Wang L (2018) A comprehensive review of mobile technology use in hospitality and tourism. *J Hospitality Mark Manage* 27: 626–648.

Masiero L (2016) International tourism statistics: UNWTO benchmark and cross-country comparison of definitions and sampling issues. Statistics and TSA: Issue Paper Series. Available from: <http://cf.cdn.unwto.org/sites/all/files/docpdf/iptntourismstat.pdf> (access August 2020).

Nilbe K, Ahas R, Silm S (2014) Evaluating the Travel Distances of Events Visitors and Regular Visitors Using Mobile Positioning Data: The Case of Estonia. *J Urban Technol* 21: 91–107.

Nurmi O (2018) Improving the accuracy of outbound tourism statistics with mobile positioning data. 15th Global Forum on Tourism Statistics, Cusco, Peru. Available from: http://www.15th-tourism-stats-forum.com/pdf/Papers/S3/3_3_Improving_the_accuracy_of_outbound_tourism_statistics_with_mobile_positioning_data.pdf (accessed September 2020).

Nurmi O, Piela P (2019) The Use of Mobile Phone Data in Tourism Statistics, Proceedings of the 62nd ISI World Statistics Congress. Proceeding Special Topic Session, 4: 135–142. Available from: [https://2019.isiproceedings.org/Files/6.Special-Topic-Session\(STS\)-Volume-4.pdf](https://2019.isiproceedings.org/Files/6.Special-Topic-Session(STS)-Volume-4.pdf) (accessed September 2020).

Ortega-Fraile FJ, Ríos-Martín MÁ, Ceballos-Hernandez C (2018) Tourism and mobile devices: provenance of academic research through a bibliometric study. *J Tourism Anal* 25: 91–112.

Osaragi T, Kudo R (2020) Enhancing the Use of Population Statistics Derived from Mobile Phone Users by Considering Building-Use Dependent Purpose of Stay, In: Kyriakidis P, Hadjimitsis D, Skarlatos D, Mansourian A. (eds), *Geospatial Technologies for Local and Regional Development. AGILE 2019, Lecture Notes in Geoinformation and Cartography*, Springer, Cham.

Othman AG, Mohamed B, Bahauddin A, et al. (2010) A Geographic Information System Based Approach for Mapping Tourist Accommodations in the East Coast States of Malaysia. *World Appl Sci J* 10: 14–23.

Park SY, Pan B (2018) Identifying the next non-stop flying market with a big data approach. *Tourism Manage* 66: 411–421.

Park S, Xu Y, Jiang L, et al. (2020) structures of tourism destinations: A trajectory data mining approach leveraging mobile big data. *Ann Tourism Res* 84: 1–17.

Phithakkitnukoon S, Horanont T, Witayangkurn A, et al. (2015) Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan. *Pervasive Mobile Comput* 18: 18–39.

Raun J, Ahas R (2016) Defining usual environment with mobile tracking data. 14 Global Forum on Tourism Statistics, Venice, Italy. Available from: <http://tsf2016venice.enit.it/programme> (accessed September, 2020).

Raun J, Ahas R, Tiru M (2016) Measuring tourism destinations using mobile tracking data. *Tourism Manage* 57: 202–212.

Rodríguez J, Semanjski I, Gautama S, et al. (2018) Unsupervised hierarchical clustering approach for tourism market segmentation based on crowdsourced mobile phone data. *Sensors (Switzerland)* 18: 2972.

Saluveer E, Raun J, Tiru M, et al. (2020) Methodological framework for producing national tourism statistics from mobile positioning data. *Ann Tourism Res* 81: 1–13.

Seynaeve G, Demunter C, De Meersman F, et al. (2016) When mobile network operators and statistical offices meet— integrating mobile positioning data into the production process of tourism statistics. 14 Global Forum on Tourism Statistics, Venice, Italy. Available from: <http://tsf2016venice.enit.it/programme> (accessed September, 2020).

Shoval N, Ahas R (2016) The use of tracking technologies in tourism research: the first decade. *Tourism Geographies* 18: 587–606.

Sikder R, Uddin Md J, Halder S (2016) An efficient approach of identifying tourist by call detail record analysis, 2016 IWCI: 2016 International Workshop on Computational Intelligence, 136–141.

Šveda M, Krian F, Barkl P (2019) Utilizing mobile positioning data in tourism: who are the foreign visitors in Slovakia? When do they come and where they stay? *Geografický Časopis (Geographical Rev)* 71: 203–225.

Uluwiyah A, Setadi B (2017) Improving Data Quality through Big Data: case Study on Big-data Mobile Positioning Data in Indonesia Tourism Statistics. Proceedings—WBIS 2017: 2017 Internatinal Workshop on Big Data and Information Security. Available from: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8275101> (accessed August 2020).

UN (2019) Handbook on the Use of Mobile Phone Data for Official Statistics (Draft 2019). Global Working Group on Big Data for Official Statistic. Available from: <https://unstats.un.org/bigdata/taskteams/mobilephone/>.

UNECE (2019) New data sources for official statistics—access, use and new skills. Economic Commission for Europe, Conference of European Statisticians, Paris, ECE/CES/2019/41. Available from: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2019/ECE_CES_2019_41.pdf (accessed September 2020).

Wang D, Park S, Fesenmaier DR (2012) The role of smartphones in mediating the touristic experience. *J Travel Res* 51: 371–387.

Wang Y, de Almeida Correia GH, van Arem B (2019) Relationships between mobile phone usage and activity-travel behavior: A review of the literature and an example, *Advances in Transport Policy and Planning*, Ben-Elia, E. (ed.), Elsevier, 3: 81–105.

Wang Y, de Almeida Correia GH, van Arem B, et al. (2018) Understanding travellers' preferences for different types of trip destination based on mobile internet usage data. *Urban Plann Transp* 90: 247–259.

Xu Y, Li J, Belyi A, et al. (2021) Characterizing destination networks through mobility traces of international tourists — A case study using a nationwide mobile positioning dataset. *Aust Geogr* 82: 104195.

Yuan Y, Tseng YH, Ho CI (2019) Tourism information technology research trends: 1990–2016. *Tourism Rev* 74: 5–19.

Yun HJ, Kang DJ, Lee MJ (2018) Spatiotemporal distribution of urban walking tourists by season using GPS-based smartphone application. *Asia Pac J Tourism Res* 23: 1047–1061.

Zhu X, Su T, Yuan H, et al. (2019) Exploring group movement pattern through cellular data: A case study of tourists in Hainan. *ISPRS Int J Geo-Inf* 8: 74–83.



AIMS Press

© 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)