Journal of Destination Marketing & Management 18 (2020) 100481

Contents lists available at ScienceDirect

Journal of Destination Marketing & Management

journal homepage: www.elsevier.com/locate/jdmm

Exploring new ways of visitor tracking using big data sources: Opportunities and limits of passive mobile data for tourism

Julian Reif[a,*], Dirk Schmücker[a,b]

[a] DITF – German Institute for Tourism Research at the West Coast University of Applied Sciences, Heide, Germany
[b] NIT Institute for Tourism Research in Northern Europe, Kiel, Germany

ARTICLE INFO

Keywords:
Big data
Passive mobile data
Tourist tracking
Spatio-temporal behaviour
Identification of tourists

ABSTRACT

Passive mobile data (PMD) are event data recorded by mobile network operators (MNOs) in the course of a consumer's use of mobile phones connected to public voice and data networks. Increasingly, MNOs provide such data for research and applications in tourism, anonymised according to national regulations and aggregated based on the technical and economic interests of the MNO. Alongside mobility research, it is evident that tourism research has been one of the early adopters of this data source. Possible applications of PMD in tourism research include the identification of tourists, the detection of temporal and spatial distribution patterns, and the analysis of spatial and temporal relations. However, a number of drawbacks have been identified. These include the results of anonymisation and aggregation procedures, and, most of all, the inability to identify tourist activities properly, as opposed to everyday or other non-tourist types of mobility. This paper analyses and aggregates the results of different research projects on different spatial levels in Germany in order to build a conceptual framework for the specific strengths and weaknesses of the use of PMD in tourism research. The study found that, at the current state of research, PMD can measure the mobility of people in space and time but are not suitable for correctly identifying tourists and distinguishing them from non-tourists. Destination management organisations (DMOs) that are working with PMD should be aware of these barriers and adapt their research questions accordingly. However, PMD can be a powerful instrument, particularly because of its high temporal and spatial granularity.

1. Introduction

The permanent generation, transmission and storage of digital data used in using mobile online devices opens up the possibility of going beyond standardised empirical surveys to potential new methods of observing tourist behaviour. People serve as sensors (Goodchild, 2007), leaving digital footprints that give researchers new ways of analysing their travel behaviour and producing new insights that were not possible using conventional market research methods. "A large part of the earth's population can now be used as a collection of data for (nearly) real-time, fine-grained spatial observations'"(Steenbruggen, Tranos, & Nijkamp, 2015, p. 336). The tourism industry can be considered one of the pioneers in the application of new big data sources (Demunter, 2017). The rise of new information and communication technologies and new big data sources promises to mitigate the shortcomings of traditional surveys and to reduce the participants' burden. Data from social media (e.g. Önder, Gunter, & Gindl, 2019), booking services (e.g. Batiste e Silva et al., 2018), destination cards (e.g. Zoltan & McKercher, 2015) and passive mobile data (PMD) (e.g. Ahas, Aasa, Roose, Mark, & Silm, 2008) are already being used to identify the spatio-temporal behaviour of tourists.

Passive mobile data are signal data that are generated during the operation of mobile networks of all kinds (GSM with GPRS/EDGE, UMTS/HSPA, LTE/LTE Advanced and 5G New Radio). These data can be recorded in the network without any activity on the user's side: they are generated automatically as soon as a mobile device, cell tower and the IT-backend of the mobile network operator (MNO) communicate. However, due to barriers in data access, this topic has received little consideration in the international research literature, so studies in this area are still scarce (Li, Xu, Tang, Wang, & Li, 2018; Shoval & Ahas, 2016).

In Germany, MNOs currently in the market provide tourism Destination management organisations (DMOs) with aggregated and anonymised data in their respective network for the tourism industry (a

* Corresponding author. DITF – German Institute for Tourism Research at the West Coast University of Applied Sciences, Heide, Germany.
E-mail addresses: reif@fh-westkueste.de, reif@uni-bonn.de (J. Reif), schmuecker@fh-westkueste.de (D. Schmücker).

https://doi.org/10.1016/j.jdmm.2020.100481
Received 29 November 2019; Received in revised form 26 August 2020; Accepted 28 August 2020
Available online 15 September 2020

fourth network operator is expected to enter the market in 2020 with the introduction of 5G New Radio). As the rising number of practical projects indicates, DMOs are interested in using those data to get new insights into visitor segments on which little information has been available so far (especially same-day visitors). Although MNOs offer PMD to the tourism industry, many challenges still have to be addressed, and necessary definitional work has to be done so that the data can be used effectively for (academic) tourism research and the practical purposes of a DMO respectively. This paper contributes to closing this gap by drawing on the author's experience in a number of destination management projects in Germany. More specifically, it analyses data from two research projects on different spatial levels (metropolitan and local) (Section 3.3) in order to build a conceptual framework for the specific strengths and weaknesses of the use of PMD in tourism research. The data and results focus on the use of PMD in Germany but can also be adapted for use in other European countries with similar data-protection rules.

In contrast to other papers that show applications of PMD under looser data protection rules (specifically in Estonia, see Ahas, Aasa, Mark, Pae, & Kull, 2007; Ahas et al., 2008; Kuusik, Tiru, Ahas, & Varblane, 2011; Nilbe, Ahas, & Silm, 2014; Raun, Shoval, & Tiru, 2020; Raun, Ahas, & Tiru, 2016; Saluveer et al., 2020; Tiru, Kuusik, Lamp, & Ahas, 2010), this paper shows the application of PMD in a stricter setting. Although an international or at least European perspective might be desirable, it has to be noted that both data-protection rules and data availability differ substantially between countries.

The following research questions were formulated and guided the research:

1. To what extent can PMD be used to discriminate tourists from non-tourists, according to international conventions?
2. How can PMD be used to identify tourist movement patterns?
3. What are the future implications for tourism research and practical implications for DMOs when working with PMD?

Beginning with a literature review and an overview of the current state of research (Section 2), a methodology for using passive mobile data, with a focus on describing current methods in the tourist identification process, is introduced in Section 3. Issues in identification and volume assessment are tackled in Section 4, which proposes three different approaches in detecting tourism activity out of PMD. Section 5 shows empirical evidence of PMD in depicting identification of visitor segments, inter-destination, and intra-destination movement patterns. Section 6 discusses the results and future implications for tourism research and for practical work with PMD by DMOs. The conclusions are presented in Section 7.

2. Literature review

2.1. A renaissance in tourist tracking

Due to new possibilities of tracking people digitally in space and time, the tracking of tourists is experiencing a renaissance. In a large body of literature (for a comprehensive list, see Shoval & Ahas, 2016), the spatio-temporal behaviour of tourists is receiving new attention, and the theoretical principles of the 1970s, such as that of time-geography (Hägerstrand, 1970) also seem to be *en vogue* again (Shoval, 2011). Researchers now have access to fine-grained data in space and time, allowing them to better comprehend the movement of people and to answer new research questions: "Tracking technologies are able to provide high-resolution spatial and temporal data that could potentially aid, augment, and advance research in various areas in the field of urban studies" (Shoval, 2008, p. 21). However, this is not only limited to the area of urban research.

From a DMO perspective, new digital tracking technologies offer numerous possibilities for answering questions within a sustainable tourism development framework. Questions about carrying capacity, allocation of hotels or transport possibilities in a destination and nudging tourists towards more sustainable alternatives can thus be answered more easily and reliably compared to traditional survey methods. Furthermore, DMOs can use this knowledge in the development of new tourism products and concentrate marketing budgets more effectively, which, in turn, can lead to an improvement in the onsite experience for tourists or in tourism acceptance (Edwards, Dickson, Griffin, & Hayllar, 2010; Edwards & Griffin, 2013; Shoval and Ahas, 2016, 2018).

From a methodological point of view, there are many possibilities for empirically assessing tourists in space and time. Analogue methods are direct observation (Keul & Kühberger, 1996) and the time-(-space)-budget method (Debbage, 1991), which are rarely used nowadays. Digital tracking methods, all of which have their specific advantages and disadvantages (Kellner & Egger, 2016), have become increasingly popular in recent times, especially in an urban context (Caldeira & Kastenholz, 2019). Altogether, GPS tracking (i.e. recording and analysing location signals received and recorded or redistributed by digital devices using satellite signals) seems to be the most commonly used technique to measure the spatio-temporal behaviour of tourists within the set of digital methods (Shoval & Ahas, 2016). Nevertheless, GPS tracking has its methodological challenges. The need for handing out the devices to participants at defined entry and exit points (Shoval & Isaacson, 2010) makes it time-consuming and thus expensive. Additionally, for app-based GPS tracking, willingness to participate is low (McKercher & Lau, 2009; Thimm & Seepold, 2016). To overcome these drawbacks, big data sources can be used to identify the spatio-temporal movement patterns of tourists.

2.2. Big data sources for tourist tracking

Big data are usually defined based on three main characteristics: volume, velocity and variety (Kitchin, 2013). These characteristics are also where the advantages of these new data sources over traditional market research methods in tourism lie. Big data are usually complete (high *volume),* they are available quickly (almost in real time, *velocity)* and come from a multitude of sources (structured, semi-structured and unstructured, *variety)*. In recent times, many more 'v-words' haven been used to describe big data sources, showing that the above-mentioned '3Vs' are not sufficient to define big data, as "there are multiple forms of big data" (Kitchin & McArdle, 2016, p. 8). Big data sources usually enable researchers and DMOs to analyse phenomena that traditional sample-based market research is not adequately able to detect. Kitchin and McArdle (2016, p. 8), sum up the differences of big data when they conclude: "Small data are slow and sampled. Big data are quick and n = all."

Frequently in tourism, big data can be described as data generated for some technical reason and then re-analysed for tourism research purposes (as is the case with PMD). Tourism research and DMOs can therefore often use big data as a secondary resource, as the data are not generated for specific touristic issues. Big data sources can therefore be classified for the purposes of tourism research into six domains with different characteristics in terms of the participation of the user in generating the data: mobile communication, sensors and wearable devices, cameras/lasers/satellites, business-process-generated data, websites and social media (Table 1).

Big data are of special interest for tourism (geography) in general and DMOs in particular if location information (usually geographical longitude and latitude) is attached (Bauder, 2019). An extra timestamp makes sure that the data source is well suited for research on the spatio-temporal behaviour of individuals. One of the great advantages of using big data is that facts no longer have to be asked about (e.g. a trip from A to B), as data traces can already be seen. However, one of the central requirements for the feasibility of big data for tourism research is distinguishing between tourist and non-tourist digital footprints. Tourist

Table 1
Big data sources with the potential for tourism and tourist tracking.

Domain	Category	Characteristic	Data type	Example
Mobile Communication	Device Data/Network	Passive	e. g. Passive Mobile Data	Raun et al. (2016)
			e. g. Wi-Fi	Bonné, Barzan, Quax, and Lamotte (2013)
			e. g. Passive GPS Data (Apps, ODK, SDK)	Brovelli, Minghini, and Zamboni (2016)
Sensors and Wearable Devices*	Device Data/Network	Active/ passive	e. g. Bluetooth	Versichele, Neutens, Delafontaine, and van de Weghe (2012)
			e. g. RFID/Beacons/NFC	Pesonen and Horster (2012)
			e. g. Physiological Sensors (Wristbands)	Shoval, Schvimer and Tami (2018)
Cameras/Lasers/Satellites	Network	Passive	e. g. Closed Circuit Television	Geng, Du, and Liang (2019)
			e. g. Satellite Images/Meterological Data	Guo (2016)
Business Process-generated Data	Network	Active/ passive	e. g. Financial Transactions	Romero Palop, Arias, Bodas-Sagi, and Lapaz (2019)
			e. g. Destination Cards	Zoltan and McKercher (2015)
			e. g. Booking Engines	Batista e Silva et al. (2018)
Websites	Network	Active/ passive	e. g. Open Data	Signorelli, Reis, and Biffignandi (2016)
			e. g. Searches in Search Engines	Bokelmann and Lessmann (2019)
			e. g. Clickstreams	Ward and Shafaghi (2013)
Social Media	User Generated Data/ Network	Active	e. g. Facebook, Twitter, Blogs	Önder et al. (2019)
			e. g. Photo Data (instagram, Flickr etc)	Salas-Olmedo et al. (2018)

Source: Authors, based on Bauder (2019); Demunter (2017); Li et al. (2018). Note: * Data has to be generated automatically and not collected intentionally.

digital footprints or data traces from tourist activities occur if a person can be considered a tourist, as defined by international conventions (United Nations, 2010, see Section 4.2). whether big data traces derive from tourist or non-tourist activity is not only of practical relevance with respect to an under- or over-estimation of tourist demand and – for example – the resulting economic effects. There is also the risk of neo-positivist attitudes: based on an evaluation of online photo data, Bauder (2019) argues that posting pictures is a highly individual and selective process and that the results obtained from the analysis of such data (e.g. tourist itineraries) are often generated without hypothesising about or reflecting a causal relationship. Although there are critical voices on the use of big data, raising questions about the epistemological change (Boyd & Crawford, 2012), it must be acknowledged that big data can have a strong impact on the production of knowledge about (tourism) geography (Singleton & Arribas-Bel, 2019): "Big data analytics can be seen as a new research paradigm, rather than a uniform method, that may utilise a diverse set of analytical tools to make inferences about reality using large data" (Xiang, Schwartz, Gerdes, & Uysal, 2015, p. 121).

2.3. Application of PMD in tourism research

Although tourism research in the field of PMD is still in its infancy (mostly due to restricted data access and privacy concerns), some researchers have been working in this field for more than 10 years: in Europe, this began with a research group from the University of Tartu (Estonia). One of the first attempts to analyse PMD was the work of Ahas et al. (2007), who showed seasonal regional patterns based on roaming data. Their dataset and factor analysis revealed typical time-space seasonality patterns, whereby the coastal areas in Estonia are visited mostly during summer, and the continental inland is visited during the winter season.

In their pioneering work, Ahas et al. (2008) introduced call detail records (CDR) as a new source for tourism research and detected diverging activity spaces for Latvians and Russians based on their first call in Estonia. In showing that correlations between conventional tourism statistics and PMD are higher in highly frequented areas and lower in regions with little tourist activity, the authors conclude that PMD can be used as a new approach for marketing analyses and to improve tourism infrastructure.

Furthermore, mobile data can be used to identify repeat visitors and to show destination loyalty (Tiru et al., 2010). Based on a dataset consisting of information about foreign visitors in Estonia who had visited the country in the past five years, the results demonstrated that repeat visitors made up to 30% of visitors, 64% of the number of visits and 70% of the total number of visiting days. Kuusik et al. (2011) investigated customer loyalty by using PMD; with mobile data, they were able to identify the duration, timing, density, seasonality and dynamics of visits and to distinguish repeat visitors.

The principle of distance decay (McKercher & Lew, 2003), which says that distance is a significant limiting factor that influences travel, can also be demonstrated using PMD: Nilbe et al. (2014) estimate distances travelled by event visitors using a passive mobile dataset on event visitors in Estonia in comparison to a group of regular visitors. Distance decay can be shown in both groups, but event visitors come to Estonia from shorter distances than regular visitors. Furthermore, Raun et al. (2016) use PMD to measure tourism destinations and demonstrate the application of big data in destination management. Using a set of data from foreign visitors in Estonia, they show that destinations can be differentiated by the geographical, temporal and compositional parameters of the visits.

The focus of the paper mentioned above is on the application of the data in tourism research and in analysing tourist behaviour. Statistical analyses, econometric models and case studies are common data processing techniques (Li et al., 2018). However, the question whether the data analysed in the above-mentioned body of research literature derives from touristic or non-touristic activities remains unanswered. For example, Raun et al. (2016) define tourism visitors (tourists) as all non-resident foreign visitors who use their mobile phones in Estonia and spend time in the country (without time limits). This means that data could also come from people who are passing through or come to Estonia for non-tourism reasons. In the case of international roamers, the probability of analysing actual tourist trips (as opposed to non-tourist trips) is quite high, but not given (see Section 4). However, Saluveer et al. (2020) propose a methodological framework for producing national tourism statistics from PMD. They use a negative definition of tourist signals when they classify all visitors who are not transit visitors, migrant workers or cross-border commuters as tourists. So far, there are only studies that focus on international roaming data, ignoring all the domestic tourist flows. Additionally, there are no scientific publications dealing with the specific situation in Germany.

In addition to the academic tourism literature, there is considerable debate about the usability of passive mobile data for tourism statistics. The key learnings from the 'Feasibility study on the use of mobile positioning data for tourism statistics (2012–2014)' are that mobile data are highly consistent with reference statistics and can be made available much more quickly than data from traditional sources (Ahas et al., 2014). They can be used as quick indicators, as a calibration source and potentially to strengthen current tourism demand surveys through mixed-mode data collection (e.g. number and duration of trips). Nevertheless, the heterogeneity of rules and regulations concerning access to mobile positioning data did not allow for useful application in the EU. There is a broad range of countries with more liberal regulations and stricter regulations (e.g. Germany). For example, Saluveer et al. (2020) were able to identify visitor groups based on the function (number of days spent and visits made each month) and duration of their visits to Estland (inbound tourism). This was only possible as the MNO provided them with data for a continuous three-year period with pseudonymous IDs "which are constant for each individual phone user for the whole period" (Saluveer et al., 2020, p. 5). These procedures cannot be adopted in Germany, as data protection rules are relatively strict in Germany. This was already one result of the European feasibility study (Ahas et al., 2014). The rules which the MNO have to follow are not codified in a law or regulation, but are rather decided on a case-by-case basis by the regulation office. Although in the meantime, with the introduction of the General Data Protection Regulation (GDPR), European data-protection rules were changed, the main restriction for the use of PMD in Germany, the 24-h cut-off rule, remained untouched (see Section 3.2).

It is worth mentioning that, beyond tourism geography, there are some attempts to use PMD to investigate other geographical questions like generational differences in spatial mobility (Masso, Silm, & Ahas, 2019), internal migration (Blumenstock, 2012), estimating literacy rates (Schmid, Bruckschen, Salvati, & Zbiranski, 2017), measuring ethnic segregation (Silm, Ahas, & Mooses, 2018), mapping changes of residence (Kamenjuk, Aasa, & Sellin, 2017) and tracking population movements after disasters (Bengtsson, Lu, Thorson, Garfield, & von Schreeb, 2011).

3. Methodology for using PMD

3.1. Passive mobile data

Passive mobile data (PMD) are event data recorded by mobile network operators (MNOs) in the course of the mobile device's use of public voice and data networks. The term PMD is used in this paper rather than the more generic 'location data' from mobile phones to emphasise that these data are obtained without any activity on the part of the user other than having the device switched on. In contrast to other means of obtaining location data (e.g. GPS signals, Wi-Fi signals), which require at least some sort of activity (i.e. making the GPS or Wi-Fi signal accessible), PMD are obtained by the unassisted activity of the network components (see also Table 1).

The location information is generated by the connection between the mobile device and a network antenna on the corresponding cell tower (for technical network details, see Sauter, 2017). Position accuracy depends significantly on the density of the cell towers. The distribution of cell towers is more concentrated in urban areas compared to rural areas (Shoval & Isaacson, 2010); the cell size can range from a few hundred square metres to several square kilometres. In contrast to active mobile data (AMD) where the location of the mobile phone is queried actively over a radio wave and for which the permission of the owner is required, e.g. in emergencies (Ahas et al., 2008), PMD is stored automatically during the use of the device. One can define two different methods for collecting passive mobile data: Call detail records (CDR) and signalling data (for details on the extraction of data for analysis, see Ahas et al., 2014). Against the backdrop of rapidly changing behaviour in the use of smartphones (e.g. the use of messaging services), CDRs, which are only generated for billing purposes, are becoming less important (Demunter, 2017). Empirical evidence from analysing CDRs shows that the event-triggered nature of those data produces a certain degree of bias in human mobility and that the results of the data have to be interpreted with caution (Zhao et al., 2016). However, this problem of temporal resolution is decreasing over time because the activity frequency of end devices is growing, and, thus, event signal density is increasing. MNOs collect and analyse all of the signalling data arising, including active cell changes (handovers) like calling (in- or outbound), sending/receiving SMS, usage of mobile internet and apps, switching the mobile device on and off and also passive events (e.g. automatic feedback from the mobile device to the cell).

3.2. Fundamental tourist identification processes in PMD

Ethical issues are of great concern when tracking tourists' tempo-spatial behaviour (Hardy et al., 2017). This is also the case in working with passive mobile data. Regulations in working with the data are strict, and only data from subscribers who have agreed to allow their data to be studied for statistical purposes (opt-in) can be used. Mobile signals are only analysed if there is a minimum of five signals per unit analysed. In Germany, MNO undergo a process of data anonymisation that has to be approved by the Federal Commissioner for Data Protection and Freedom of Information. A major result of this process is that the signals used have to be re-anonymised after 24 h. Against the backdrop of strict data protection regulations in Germany, identification of tourists in PMD datasets goes along the following basic lines (see Fig. 1):

- If the first signal from a mobile device and the last signal of the day are inside an area of interest (e.g. a tourism destination), the signal represents either a local inhabitant or an overnight tourist.
- If the first and last signals of the day are outside the area of interest, but inside this area during the day, the signal represents either a tourist day visitor or a commuter or some other group, e.g. day-to-day business visitors travelling for medical, administrative or other reasons.
- If the first signal of the day is inside the area of interest but does not reappear before the end of the day, the signal can represent an overnight tourist on their day of departure, a local inhabitant departing elsewhere for whatever reason or a late party guest (rehashing usually occurs at 3 a.m.). If the signal reappears within the same national network, the last signal of the day can be found for these cases. If the device travels abroad, it disappears, leading to the false impression that the last signal was inside the area of interest.
- In contrast, the first signal of the day outside the area of interest and the last signal of the day inside the area of interest may represent an overnight tourist on the day of arrival or a local inhabitant returning from a stay outside. Again, international roaming signals may simply appear, leading to a false impression.

The false identification of users of international roaming services can be mitigated, but not overcome, by looking specifically at traffic hubs such as stations and airports. The problem of correctly identifying tourists in such a setting is evident and will be addressed in more detail in Section 4 of this paper.

3.3. Data sources

In the course of analysing opportunities and limits in the use of PMD in tourism research, two datasets were used from two MNOs in Germany. The first dataset was provided by Motionlogic, a spin-off of Deutsche Telekom. The focus was on the analysis of same-day visitors to the city of Hamburg. For this reason, activity data from 40 million mobile devices in Germany were evaluated and extrapolated to the total German population based on Deutsche Telekom's local market share at the place of origin for domestic guests. Although the urban research area of Hamburg is densely covered by mobile phone antennae, smaller city districts had to be combined for analysis. The dataset enabled us to

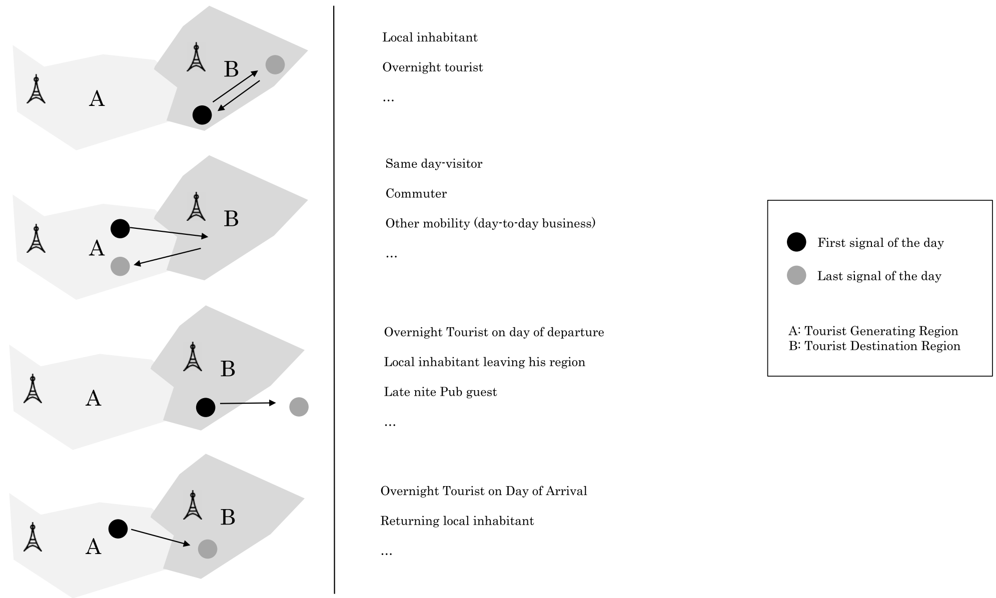


Fig. 1. Model of tourist and non-tourist-movement identification using PMD.
Source: Authors

analyse visitor flows to and in Hamburg on a monthly basis from May 2017 through April 2018, covering data from 401 source regions (based on administrative urban and county districts in Germany) and 98 destinations (city districts) in Hamburg. Additionally, analysis of intra-destination movement patterns for same-day visitors between the districts in Hamburg was possible on an annual basis. A domestic same-day visitor was defined as a person whose first and last mobile signals were recorded in the period between 3.00 a.m. and 2.59 a.m. the following day outside the city limits of Hamburg and had at least a 120 min stay in Hamburg.

The second dataset was supplied by Telefónica Next. Data were available for May and August 2018 on a daily basis for the seaside resorts St. Peter-Ording and Büsum, located on the North Sea coast of Germany. In order to restrict the area, the respective municipal boundaries were used. Extrapolation to the total German population based on geographically differentiated market shares and calibration was performed on the basis of socio-economic structural data such as differentiated census data. For this study, extrapolated and non-extrapolated data were used. Unlike in the first data set, Telefónica's data anonymisation process allowed for the identification of home and work locations of device users and, thus, different segments could be identified using the logical classification described in Section 4.2.1.

Both datasets underwent a process of data anonymisation that was approved by the Federal Commissioner for Data Protection and Freedom of Information before being handed over to the authors. Although both sample sizes cover the entire demand of one MNO, which is more or less a third of the German population, signals from other suppliers do not appear in the specific sample.

4. Issues in identification and volume assessment

One of the main challenges in working with passive mobile data for tourism purposes is the correct identification of tourists versus non-tourists within the vast number of anonymised electronic signals and based on the fundamental procedures described above. This section shows three ways to implement such identifications.

4.1. Validity and reliability

It needs to be noted that the detection of electronic signals from mobile networks captures devices, not users. These signals can be equivalent to a person, but this is not necessarily the case. For example, if one person carries more than one device (e.g. a smartphone and a tablet computer) or if devices are mainly used for machine-to-machine communication, interpreting a signal as equivalent to a person would clearly lead to invalid data. Thus, 'n = all' can lead to limitations in terms of data validity because not every recorded signal represents a human user (not to mention a tourist). Reciprocally, data from one MNO are restricted to the signals in their own network, while users of other networks or people without a switched-on mobile device or people who either do not have a device at all or do not have it with them are not captured. Therefore, data validity is also restricted because not all people are represented in the data. The same goes for devices outside the coverage of the network. In such cases, the network obviously cannot capture any signals from the device.

In terms of reliability, an outage of network or storage components may lead to incomplete data. It may also be the case that not all cell changes of a device are being captured by the network because one Mobile Switching Centre (MSC, Sauter, 2017) can cover a number of cells and handle the necessary handovers autonomously. It may therefore be correct that, in contrast to traditional sample data, big data derives from real user actions and not from surveys, which have a risk of being biased through information loss (Song & Liu, 2017). However, in the case of passive mobile data, it must be said that information loss is also a problem due to technical deficiencies. Finally, MNOs are commercial enterprises that are in competition with each other and do not want to disclose details about their algorithms. Researchers usually receive anonymised, standardised and extrapolated datasets and have no access to raw data, which makes it hard to assess the objectivity of the data-generating process.

4.2. Identifying tourists

Identifying tourists from electronic signals builds upon the formal definitions of tourism set forth by the United Nations (2010) and statistical bodies like Eurostat (2013). Using these definitions, tourism is the activity of visitors outside their usual environment, travelling to a primary destination and back, as long as they are not employed by a local entity and the whole trip does not take longer than a year.

Thus, tourism accounts for a large portion of mobility, but, of course, not all mobility is tourism. Specifically, the following forms of mobility must *not* be seen as tourism from the point of view of a destination (United Nations, 2010):

1. All forms of mobility which are in a person's usual environment, including shopping, sport and leisure, administrative and medical activities.
2. All forms of commuting, including trips to and from the workplace, school, university, etc, be they regular or irregular.
3. All forms of commercial mobility, including the transport of goods and deliveries, movements of agricultural and construction machinery, but also activities of taxi drivers, ship and aeroplane crews, train guards, bus drivers, etc.
4. All forms of mobility of inhabitants inside the destination: it is, however, not impossible that inhabitants may act as tourists in their workplace (depending on the size of the destination).

The challenge is to distinguish within the electronic signals between those that come from tourist activity and those that do not, based on the procedure described in Section 3.2. In practical terms and from the point of view of a destination, the most challenging distinctions are those between commuters and same-day visitors, on the one hand, and between inhabitants and overnight tourists, on the other.

Basically, there are three main ways of distinguishing tourists from non-tourists, in addition to simply recording the first and last signals of the day. The first and most promising is to identify regularities in the movement patterns of one device over the course of days, weeks, months or years, and thus distinguish regular movements (which are probably not tourism) from irregular or sporadic movements (which may be tourism). This approach, however, is not possible when data protection rules forbid long-term tracking. Data protection rules in Germany (Section 3.2) require rehashing every 24 h, and MNOs are only in the process of developing mechanisms to overcome this barrier for at least part of the signals in their network. For this case, besides the definition based on regularities in movement patterns, two different approaches are discussed, one building upon probabilities (Receiver Operator Characteristic – ROC) and one using classification approaches.

4.2.1. *Regularities in movement patterns based on home and work locations*

If data-protection rules allow, network operators can, within reasonable boundaries of reliability and validity, identify home and work locations for a device simply by analysing regularities in movement patterns. If the home and workplace are known, identifying tourists from the perspective of the destination is quite straightforward and simply follows the rules outlined in Table 2.

On closer examination, however, even this procedure has its pitfalls. A taxi driver bringing a passenger to some point inside the destination and returning to his or her origin will be counted as a tourist day-tripper (which may be true for the passenger, but not for the driver, see the second point in Section 4.2). Furthermore, irregular commuting will probably go by undetected, e.g. shift workers who work at different times of the day and whose movements are probably not very regular over the course of months or years. The same will probably be true for the identification of movements not outside the usual environment, e.g. medical treatments. However, one main problem can be solved with this approach, and that is to distinguish inhabitants from non-inhabitants (for empirical evidence using this approach, see Section 5.1).

4.2.2. *Probabilistic approaches*

If work and home locations are not known, probabilistic approaches can be used. These approaches use the true- and false-positive rates resulting from the application of a discriminant variable. Usually, a receiver operating characteristic (ROC) curve is used to find a point of discrimination.

A distance example can be used to illustrate the approach. The goal is to find the point of discrimination between commuters and same-day visitors. From official commuter statistics and from other market data, the percentage rates shown in Table 3 were derived. The distances are real distances in km between the place of living and the destination, as indicated by commuters (destination is the workplace) and day visitors (destination is for a day trip) in surveys.

In the example, 73% of the commuters did not travel more than 40 km, and 48% of day visitors (tourists) did not travel more than 40 km. If 40 km was to be used as the discriminant line and declare all signals *inside* the 40-km line to be commuters and all signals *outside* the line to be tourists, 52% of all tourists would correctly be classified – this is the true-positive rate (0.52). But 27% of commuters would be classified as tourists – this is the false-positive rate (0.27). Ideally, 100% of all visitors would be classified as true positive and 0% of all commuters as false positive. This point would be represented by the upper left corner in Fig. 2, and points closer to the ideal point are better discriminants than points further away from the ideal point. Furthermore, a good discriminant point would be one where the graph changes its direction from upwards to sideways, as can be seen in the idealised curve at 100 km. As can also be seen from Fig. 2, real data are relatively far away from the ideal point, and there is no clear advice as to which distance should be used for distinguishing between tourists and commuters.

The same approach could be used for other discriminant variables, e.g. the number of attraction points visited during a day. However, in practical terms, so far, no obvious parameter showed up to come close to the ideal point.

4.2.3. *Classification approaches*

If data protection rules allow, network operators can enrich the mobile data with contract data, e.g. the place of residence. This approach, however, has two main drawbacks in terms of data validity and reliability. First, the place of residence of the contract partner may not be identical to the real place of residence of the device user. Second, for some types of contracts, network operators may not have access to the place of residence from the contract database. This may be true for business contracts, contracts outside the country (international roaming) or customers who have prohibited the use of their contract data.

Table 2
Rules to classify different segments based on home and work location.

	Home Location in destination	Work Location in destination	First Signal of day in destination	Last Signal of day in destination	Signal during the day in destination
Inhabitant	YES	irrelevant	irrelevant	irrelevant	irrelevant
Commuter	NO	YES	irrelevant	irrelevant	irrelevant
Same-day visitor (tourist)	NO	NO	NO	NO	YES
Overnight tourist, on day of arrival	NO	NO	NO	YES	YES
Overnight tourist, on day of stay	NO	NO	YES	YES	YES
Overnight tourist, on day of departure	NO	NO	YES	NO	YES

Source: Authors

Table 3
False- and True-positive rates. Source: Authors. Data are smoothed real data for a major German city; reading example in the text is grey.

Distance class, km	Commuters, cumulated %	Day visitors, cumulated %	False-positive rate	True-positive rate
up to 30	37	42	0.63	0.58
up to 40	73	48	0.27	0.52
up to 50	81	52	0.19	0.48
up to 60	83	56	0.17	0.44
up to 70	88	59	0.12	0.41
up to 80	90	62	0.1	0.38
up to 90	92	64	0.08	0.36
up to 100	93	66	0.07	0.34
up to 110	95	68	0.05	0.32
up to 120	96	70	0.04	0.3
up to 130	97	72	0.03	0.28
up to 140	97	73	0.03	0.27
up to 150	99	75	0.01	0.25
up to 160	99	76	0.01	0.24
up to 170	100	77	0	0.23
up to 200	100	80	0	0.2
up to 300	100	89	0	0.11
up to 500	100	99	0	0.01
Over 500	100	100	0	0

Source: Institute for Employment Research, The Research Institute of the Federal Employment Agency and others

In such cases, where the place of residence is known for a portion of the signals, but not for all, classification approaches can help. In classification approaches, a portion of the data is used to identify movement patterns and to build a classification model. In the case described above, this would be the data where the place of residence is known and which can, therefore, be classified as tourist or non-tourist. The model derived from the classification algorithm is then applied to those data where the place of residence is unknown.

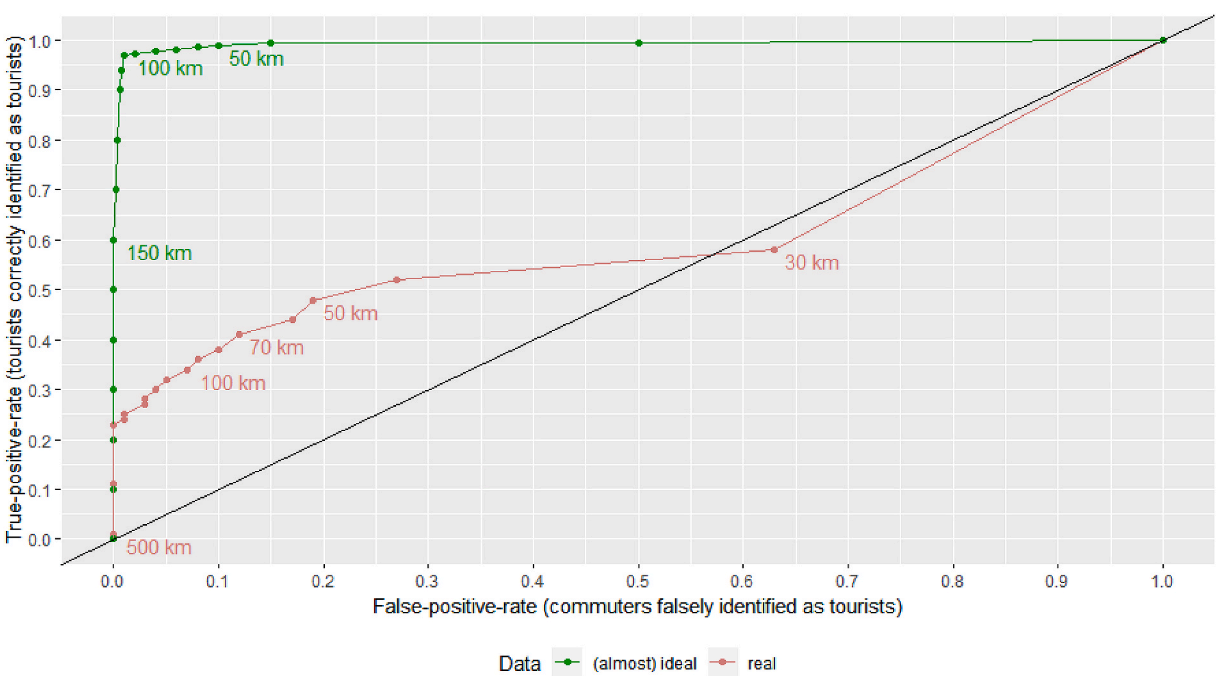


Fig. 2. Discriminating tourists and commuters by distance class. ROC curves for real and (almost) ideal data.
Source: Real data were drawn from the Institute for Employment Research, the Research Institute of the Federal Employment Agency and others and were smoothed; the ideal data are fictitious.

A number of model building ('eager') classifiers can be used for such an approach, e.g. Decision trees, Naïve Bayesian modelling, Artificial Neural Networks or Support Vector Machines (Beyerer, Richter, & Nagel, 2018; Witten, Frank, Hall, & Pal, 2017). To the best of the researchers' knowledge, no such research had been previously undertaken in the field of tourism. However, it can be expected that combining other (big data) sources can be a future point to validate PMD (see Section 6.2) and use them for classification.

4.3. Volume assessment

Closely related to the question of the identification of tourists is that of correctly assessing the volume of tourism and tourism flows, as a whole and in their various segments. Obviously, the correct identification of tourists is a prerequisite for a correct assessment of volume. However, identification is a necessary but not sufficient condition. Even if tourists are correctly identified in all their facets and segments, this would not automatically lead to a correct assessment of volume. The reason is that one network operator only sees signals in their own network, but not in other networks. Therefore, network operators tend to make a projection from their own market share. This can be done for a whole market (e.g. one country) or in a more granulated way, e.g. in states, regions, cities or even neighbourhoods. Mostly, the national market share is well known to the network operators, while their market shares in other countries and the effect this has on international roaming signals are less well known (Reif, 2019a).

One way to deal with this is to calibrate the data or put them through a plausibility check using external reference data. Data from accommodation statistics could be used, for example, to rectify the number of international overnight tourists by source market. Guest surveys covering day trips or overnight trips can also be used for calibration. However, the main value of using passive mobile data lies in its ability to produce new knowledge. Simply weighting the data so that they match other well-known sources is, therefore, counterproductive. Not calibrating the data may, however, lead to unexpected results that contradict well-established market knowledge. This again may lead to scepticism about or even rejection of information based on mobile data.

5. Identifying visitor segments and movement patterns using PMD

5.1. Visitor segments and seasonality

If the before-mentioned issues of quality criteria and identifying tourists were solved, PMD could be a powerful tool to monitor and analyse visitor flows for DMOs. At present, at least from a German perspective, tourist demand cannot be identified unambiguously. Nevertheless, on the basis of the present data, it can be shown that the strength of the data lies in showing finely resolved temporally and spatially movements between and within destinations as well as in showing the seasonality and structure of the tourist demand: questions that cannot be answered or easily financed with traditional market research instruments on this level of detail and granularity.

The plot in Fig. 3 shows inhabitants and different visitor segments in the two German seaside resorts of Büsum and St. Peter-Ording based on a definition using 'Home and Work Locations' as described in Section 4.2.1. One dot represents a day during the two months of May and August 2018. The colour of the dots indicates the day of the week (differentiated by Monday to Thursday, Fridays and weekends), the shape of the dot shows whether the specific date is a bank holiday. Tourist core segments can clearly be determined by larger variations in volume. However, specifically in May, the Pentecost bank holidays led to outliners in the data. As discussed in Section 4.2., this definition of visitor segments has its pitfalls. The numbers of commuters and inhabitants identified are, especially in St. Peter-Ording, significantly too small, compared to official reference statistics. Additionally, international roamers can be seen in the data, for example, the two big source markets of Denmark and Switzerland. However, at the moment, there is no way to extrapolate data from SIM signals with an international mobile country code (MMC), as the market share of the MNO in the source region is unknown and contractual relationships are unclear. Nevertheless, if data protection rules allow, this seems to be a great opportunity to identify tourists out of the data. However, the faulty volume assessment becomes even clearer when looking at the course of time in connection with reference statistics.

Using reference statistics from overnights stays from the local destination management organisation (TMS Büsum), measuring visitors'

Fig. 3. Using Home and Work Location to detect visitor segments in St. Peter-Ording and Büsum, Germany. Source: Authors; Data: Telefónica Next.

tax registrations without annual guest cards and without mobile home parking slots (blue line in Fig. 4), allows PMD to be validated. A correlation (Pearson correlation coefficient) of the mobile signals of the overnight visitors with the reference data shows that they are significantly correlated ($r = 0.779$; $p < .001$). Having information about anonymised raw data (red line) and extrapolated data (green line) (see Section 3.3), two things are learned that: (1) seasonal patterns can be emulated quite well with PMD (and this is relevant because not all tourist destinations have reference data of the same quality as the community here), and (2) the level of volume is clearly underestimated. The latter is remarkable, as one might expect that PMD cover a greater range of signals than the conventional methods used by DMOs, such as tax registration. At the moment, it can be seen that PMD are rather useful for describing the structure and seasonality of visitors in a destination, but are not an accurate tool for the determination of volumes.

5.2. Inter-destination movement patterns

Inter-regional tourist movement patterns assume circular spatial mobility as a constitutive element of tourism (Leiper, 1979). Diverging movement patterns of source-destination matrices have been analysed and described in a broad body of literature (e.g. Lau & McKercher, 2006; Oppermann, 1995). For the city of Hamburg, there are multiple options showing inter-destination tourist movements, some of which will be illustrated in this section, bearing in mind that data may also show a high rate of possible false positives.

Fig. 5 shows the inter-destination same-day visitor flows from the German federal districts (marked as the centroid of the respective district) to Hamburg during one year. Two things are learned from this explanatory visualisation: (1) in times of a tight network of flight connections, same-day visitors to Hamburg come from all over Germany, even from points of origin in the south, such as Munich or Frankfurt, and (2) the denser the source market, the bigger the visitor movements are, as indicated by the red arrows in Fig. 5. The problem with this is that the tourist area overlaps with the commuter area. Fig. 6 illustrates this problem.

As can be seen from Fig. 6, commuters and mobile signals show a strong correlation. As visualised on the left side, *up to* a distance of 125 km from the city of Hamburg, the official number of commuters from the city and regional districts has a strong influence on the mobile signals from same-day visitors. The adjusted R square is 0.964 ($p < .001$), and one can say that mobile signals measure commuters or same-day visitors (tourists) who behave like commuters (see section 4). On the right side, however, the adjusted R square is 0.568 ($p < .001$) indicating that more tourist activity is in the data, deriving from day visitors to the city of Hamburg with a distance of *more* than 125 km. Distance from the destination is clearly, then, one of the limiting factors of same-day visitors and commuters coming to Hamburg. Having in mind the first law of geography, that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236), a distance decay effect can be demonstrated: on an annual basis, 61% of commuters and 75% of same-day visitors measured by mobile signals came to Hamburg from a distance of up to 50 km. However, the distribution of distances (as percentages of the total number of commuters and mobile signals) can be modelled through power regression, following the function

$$y = b_0 \times t^{-b_1} \tag{1}$$

where b_0 is the constant and t the distance value raised to the power of b_1. Coefficient b_1, thus can be interpreted as a distance decay factor (Gao, Liu, Wang, & Ma, 2013; Taylor, 1971). The greater the value of b_1, the greater the influence of distance (Zhao et al., 2016). The results of

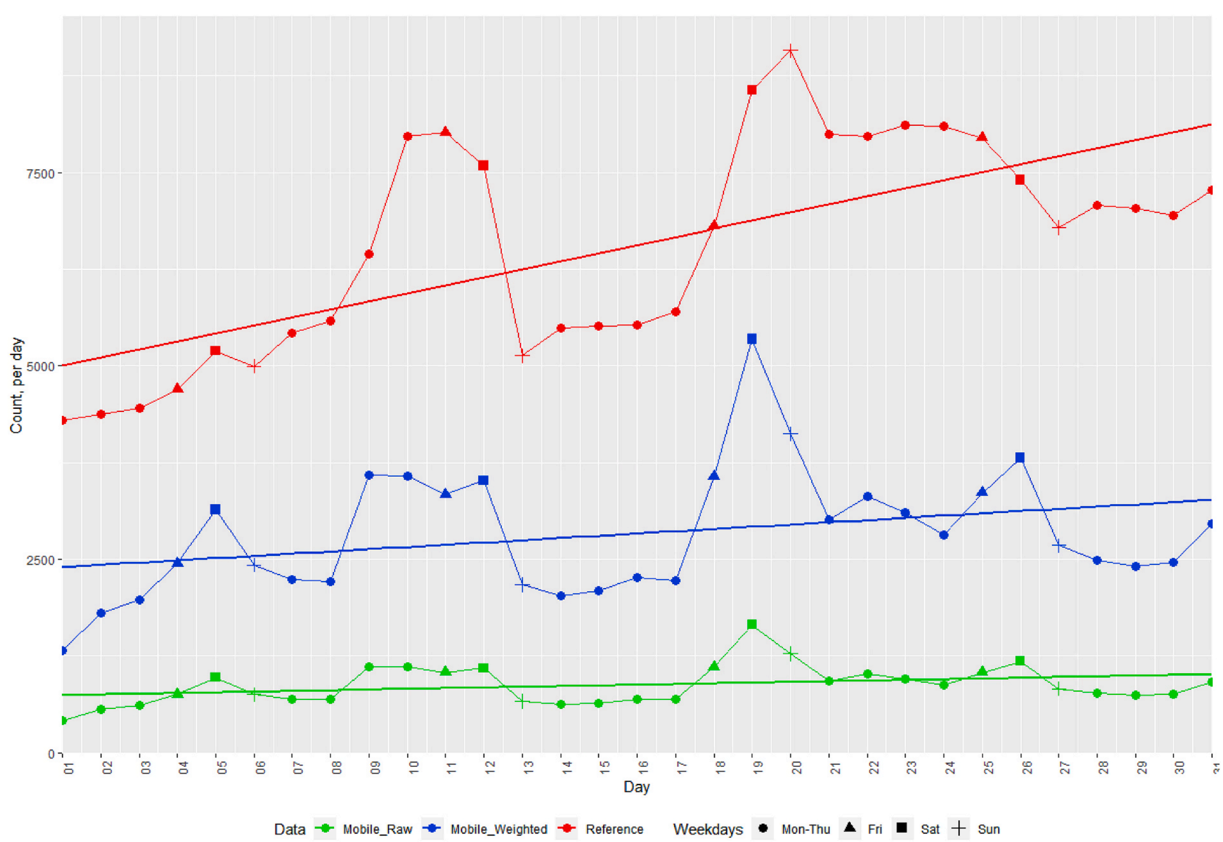


Fig. 4. Using reference statistics to validate PMD in Büsum, Germany. May 2018: Mobile Data (raw and weighted), Reference Data. Source: Authors; Data: Telefónica Next; TMS Büsum

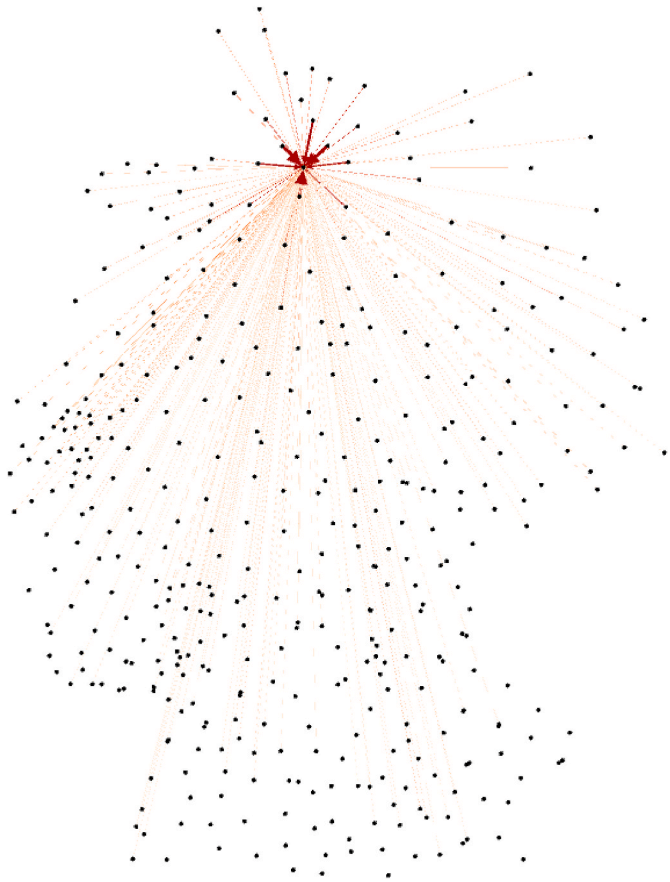


Fig. 5. Inter-destination visitor flows from German federal districts to Hamburg during one year.
Source: Authors. Data: Motionlogic

modelling both – commuter numbers and mobile signals – are shown in Table 4.

The distance decay factor can thus be estimated to be 1.707 for commuters and 3.859 for mobile signals. Note that the exponent is displayed as a positive number, but in fact is negative (see formula above), which leads to a falling curve which flattens out to the right, but can be transformed into a cumulative probability function (Fig. 7). The larger coefficient for PMD is in line with expectations, because, for example, same-day business visitors, who will accept longer distances compared to commuters, are included in the passive mobile dataset. The commuter values are in line with distance decay factor calculations in the literature, which range from 1.45 to 1.98 (Gao et al., 2013; González, Hidalgo, & Barabási, 2008; Zhao et al., 2016). As the function values represent percentages, these values can be transformed into cumulative percentages. However, there are some limitations in the available dataset. First, the data have a relatively coarse resolution for near distances. Second, there is a large number of objects for far distances with very small portions (or even zero) for many of them. Third, the distance from Hamburg is calculated based on the distance from the centroid of the respective source market, which leads to inaccurate distances due to the different sizes of the regions.

5.3. Intra-destination movement patterns

The manifold factors that influence intra-destination movement patterns can be summarised in terms of individual, trip-related and external, destination-related factors (Reif, 2019b). For example, focusing on the trip-related factors, one of the main factors affecting tourism patterns in a destination is the location of the hotel. Shoval, McKercher, Ng, and Birenboim (2011) show that distance decay is even relevant on the micro-level in a destination, as tourists spend most of their time near the hotel. In the case of temporal factors such as the duration of the visit, GPS tracking in urban areas shows that same-day visitors have a rather narrow activity space in urban centres (Reif, 2019b) due to the time available and to the tendency for the most relevant attractions to be in the inner districts of the city. However, a huge drawback of using PMD in comparison to active GPS tracking is that almost nothing is known from the data about the person who travels. Analysing intra-destination movement patterns has then consequentially to be discussed on a meta-level, analysing ways between Points of interests (POI), as well as between districts. Empirical evidence for ways between districts during the course of one year can be demonstrated using the first dataset (Section 3.3).

In the present case, only the movements of devices that had their first and last mobile signals outside Hamburg were examined. In a further step, only the movements that actually took place in Hamburg were looked at. The person must have spent at least 2 h in the respective district. It is important to stress that a person (device) can walk the same path several times during a day, and so it is counted several times accordingly. If someone does not move between the urban districts, the signal is not counted. Using methods from network analyses, Fig. 8 visualises the intra-destination movement patterns of same-day visitors between the districts of the city of Hamburg during one year. Interpreting the centre of the respective districts as nodes and the ways of the same-day visitors as edges, one can use the number of ways as a filter (edge weight) and visualise only those from 150,000 ways upwards to shed light on the paths that have been taken the most. The redder the arrows between the nodes, the more interaction takes place between the districts. Besides the inner districts of Hamburg's central business district (CBD), there are a lot of visitor flows between Fuhlsbüttel (Airport) and the district of Langenhorn (North) and between Bergedorf and Lohbrügge. A further indicator of the importance of a district in terms of visitor flows is the eigenvector centrality of the respective node (district). According to this method, a node is important, the more important its neighbouring nodes are, as eigenvector centrality weights adjacent nodes by their centrality (Ghajar-Khosravi & Chignell, 2017). The greener a node (district), the higher is the value of the eigenvector centrality. As can be seen in the CBD, the districts here obviously have the highest values, ranging on a scale for eigenvector centrality from 0 to 1 (e.g. Neustadt (1.0), Altstadt (0.97)). However, the districts of Wilhelmsburg (0.82) and Steinwerder (0.76) and other districts on the south bank of the Elbe river, districts where port and freight traffic usually takes place, also show high values of eigenvector centrality. Yet here there is another indicator of a high false-positive rate (see Section 4.2).

6. Discussion and implications

6.1. Opportunities and limits of PMD

In comparison to existing, traditional tourism databases, PMD have the advantage of potentially covering the whole tourism demand, including same-day visitors, visitors in holiday apartments and so on and make it possible – if data protection rules permit – to distinguish between different user groups.

Returning to the above-mentioned feasibility study (Ahas et al., 2014), it must be apparent from the available evidence that, at least in the current state of research based on the situation in Germany, PMD cannot be used as a calibration source. The opposite is the case; without external validation factors, the handling of the data should be treated with caution.

Finally, based on the actual findings of this paper and the literature (Ahas et al., 2014; Li et al., 2018), PMD have both opportunities and limitations for detecting the spatio-temporal behaviour of tourists (Table 5).

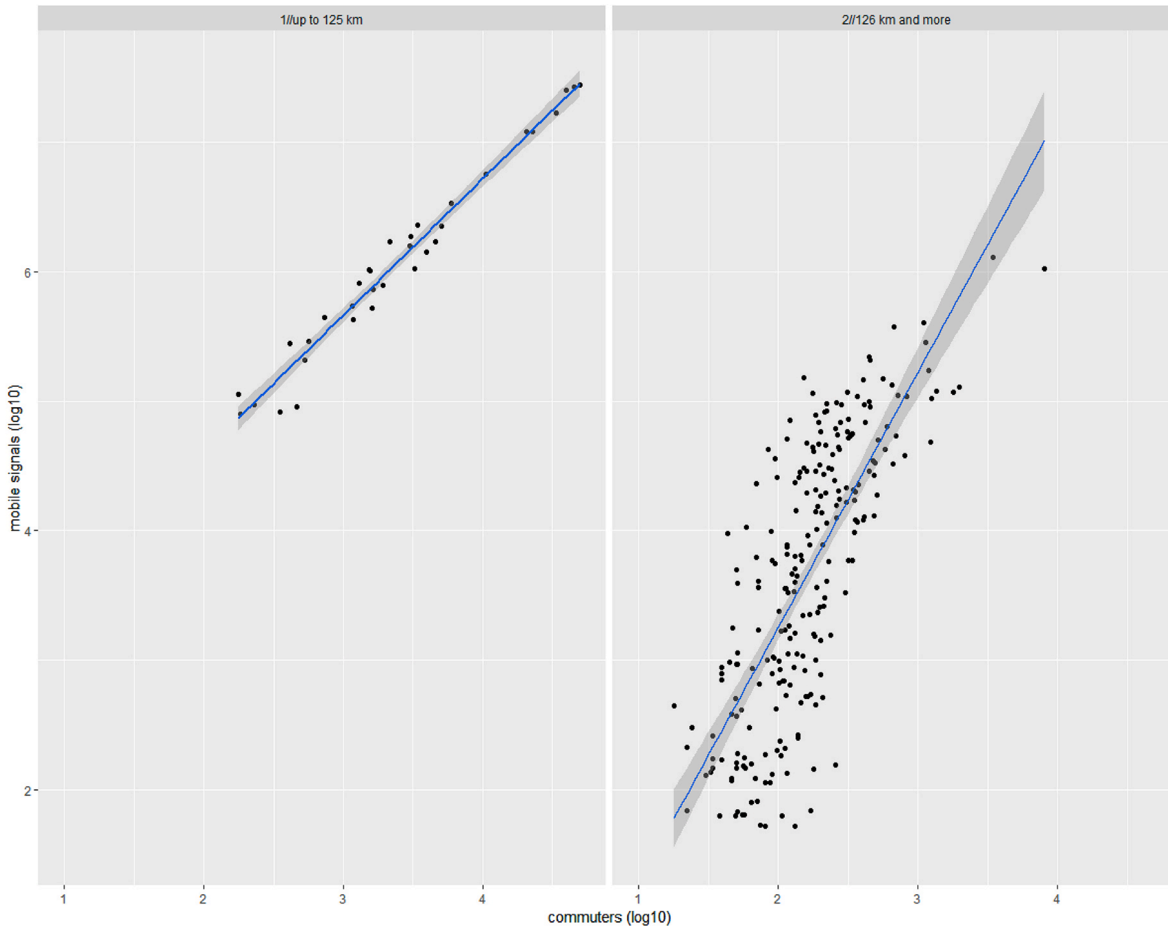


Fig. 6. Domestic mobile signals from same-day visitors versus commuters to Hamburg, Germany.
Source: Authors. Data: Motionlogic, Head of the Federal Employment Agency. Note: One dot represents one county district in Germany as a source market.

Table 4
Power Regression model and distance decay factor for official commuters and mobile signals.

Data	Model R^2	Model F	Model p	b_0 (constant)	b_1
Commuters	.533	445.9	<.001	691.46	1.707
Mobile signals	.681	584.0	<.001	11,358,708.16	3.859

Source: Authors. Note: Data points with zero percentage have been filtered

6.2. Implications for tourism research

Against the backdrop of the discussion of the idea that data-driven science could 'become the new paradigm of scientific method in an age of big data' (Kitchin, 2014, p. 6) and noting that, in realising this vision, the critical engagement of geographers is needed (Singleton &Arribas-Bel, 2019), this paper argues that data-driven tourism geography should contribute to the present research agenda for tourism geographies proposed by Bauder (2019):

(1) Discover big data sources that are either tourism-related or can be applied to tourism research
(2) Discover big data approaches
(3) Consider the ontologies of big data sources
(4) Develop research strategies beyond neo-positivist approaches
(5) Developing a data-driven tourism geography

This paper assessed PMD from a German perspective and described new ways of detecting tourist activities out of PMD. The present paper, therefore, can be seen as a contribution to points (1) and (2) of the proposed agenda.

However, most notably, the paper raises awareness of the importance of handling PMD with caution, as shown by the issues discussed in Section 4. At the moment, PMD cannot completely make tourism, as it is usually defined, identifiable (see Section 4.2). There are two ways out of this dilemma. One would be to redefine tourism to match the possibilities of the data source. Although this might seem to be an impermissible suggestion, fact more traditional data sources on tourism also do not cover the whole phenomenon, and defining tourism can depend on the point of view: an economic-practical understanding of tourism tends to prefer broader definitions, while an understanding of tourism as a socio-cultural construct tends to focus on tourists as leisure travellers (Gibson, 2016). This would lead to re-defining tourism as PMD signals (1) touching tourism points of attraction (e.g. a tourism-must-see point or a conference venue) or (2) moving at typical tourist places or (3) at typical tourist times. However, the actual applications for such a procedure would be relatively narrow and would assume that the epistemological subject is defined by the survey instrument; in a figurative sense, such as approach would resemble a drunken person who is looking for their lost car keys under the streetlamp because that is the only place they can see anything. This paper has showed that PMD measure mobility rather than tourism as a special form of mobility. The lack of detailed information on the person who travels does not allow PMD to define tourism on its own. Moreover, what about the intra-destination movement patterns of people who actually live in the destination and leave their usual environment or their regular activity space (Schönfelder & Axhausen, 2003)? Surveys show, for example, that intra-destination movements of

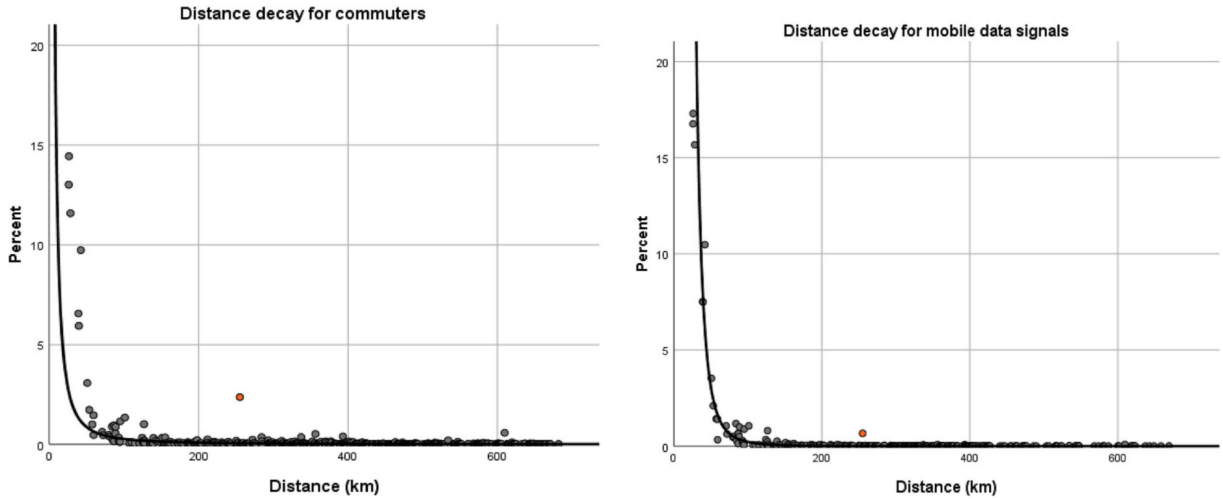


Fig. 7. Distance Decay function for commuters (left) and mobile data signals (right) into Hamburg, Germany.
Source: Authors. Data: Motionlogic, Federal Employment Agency. Note: Red dot: Berlin.

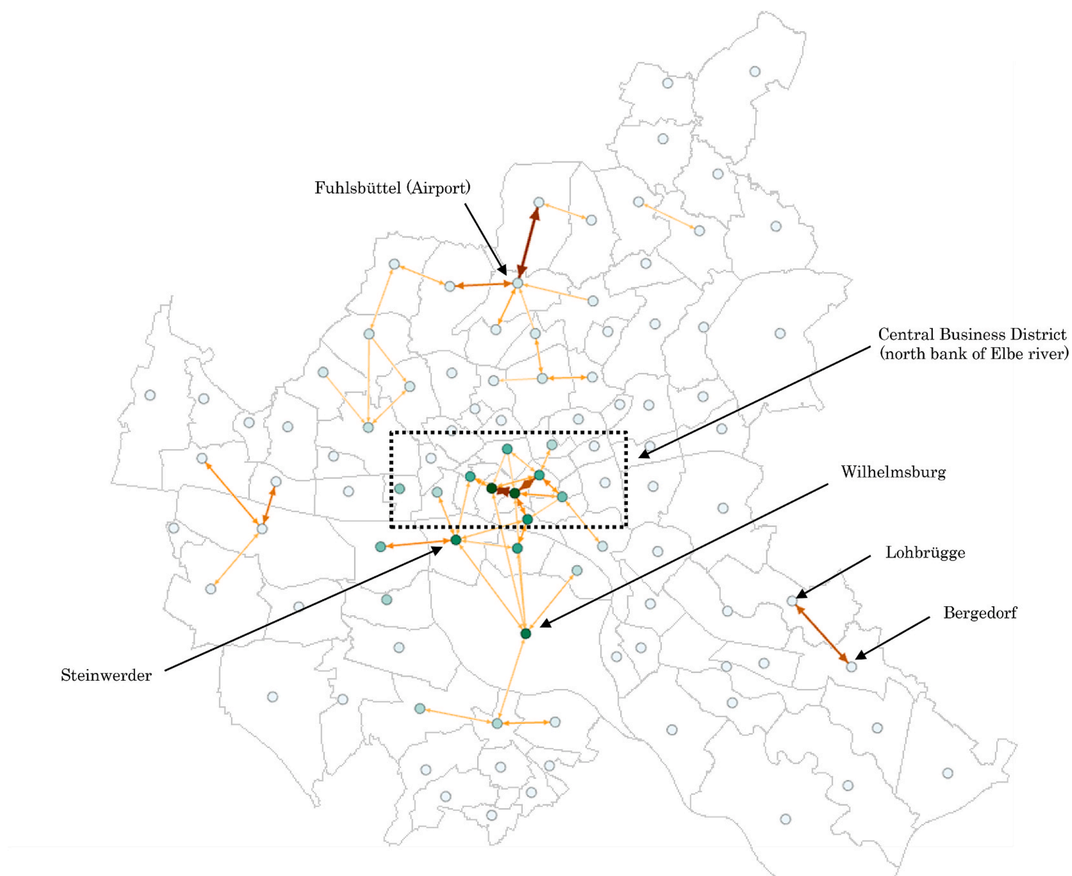


Fig. 8. Intra-Destination movement patterns of same-day visitors between the districts of the city of Hamburg during one year.
Source: Authors. Data: Motionlogic. Note: Only movements from 150,000 ways upwards are visualised in order to avoid confusion from too many arrows.

same-day visitors in inner-cities have the biggest market share of day tourism activities in Germany (Maschke, 2014). At the moment, PMD is not able to measure this kind of mobility.

However, if PMD are not able to determine tourist activity on their own, it may be possible to use other big data sources to validate the data. Passively generated GPS data can be used for classification, as movement patterns from *real tourists* are known (see Section 4.2.3).

For future research in the field of PMD, focusing on the following topics is suggested:

(1) Development of identification processes that can identify tourism based on the definition of the United Nations World Tourism Organization in spite of data protection rules.
(2) The use of other big data sources to classify mobile signals as touristic or non-touristic.

Table 5
Opportunities and Limitations of PMD for tourism.

	Opportunities	Limits
Quality Criteria	Complete data of an MNO (no samples)	Problems with reliability, validity and objectivity (e.g. high false positive rates)
People	Complete mobility of the end devices (incl. day tourism, VFR trips etc.)	Volume assessment issues (e.g. extrapolation, representativeness esp. with international roamers)
Time	Fine-grained data (almost real time) Longitudinal studies possible	24h-Re-Anonymisation
Space	Fine-grained data (city, districts, blocks, streets etc.)	Level of accuracy is dependent on the Global System of Mobile communication (GSM) in the destination No fine-grained movement patterns as with GPS tracking
Methodology	Cost-effective (depends on research focus) No burden on the side of tourist and no memory losses Less influence (from interviewers) through observation of behavior	MNOs company secrets of algorithms Little to no information on the individual or trip Cost-intensive (depends on research focus) Access to raw (anonymised) is not given Privacy concerns and Ethical issues
Epistomology	New perspectives and more accurate analysis for visitor behavior (e.g. seasonality, inter- and intra-destination patterns)	–
Ontology	PMD measures the mobility of devices	Problems in distinguishing tourist from non-tourist movements

Source: Authors.

(3) Application of statistics other than official accommodation statistics to validate PMD.
(4) Use of PMD to better understand the seasonality of tourism.
(5) Definition of key performance indicators (KPI) for PMD which help DMOs to monitor travel behaviour and to legitimise their marketing activities.
(6) Research on the international comparability of PMD datasets and, possibly, regulation rules would complement the views discussed in this paper and also in those cited in the literature review.

6.3. Practical implications for DMOs

It is nothing new that the tourism industry has a great need for information (Poon, 1988). Tourism marketers and DMOs depend on market research in order to satisfy the requirements of their different stakeholders and their marketing activities. Also, from a destination point of view, big data sources can be seen as a game-changer, having a considerable influence on tourism marketing activities (Stylos & Zwiegelaar, 2019). These impacts are obvious as traditional data analyses are not able to provide information on big volumes of data (e.g. web data) and possibly combine these data with the customer relationship tools from DMOs. However, to the best of the authors' knowledge, most of the projects on the application of PMD are experimental and result from the urge to try out new data sources to show that one is active in cutting-edge research. Without an adequate interpretation of this data (and usually the MNOs are not specialised in the field of tourism and the specific needs of the industry), DMOs are not able to translate the insights from PMD into concrete marketing decisions. Moreover, DMOs need some intermediaries to translate the technical language from the MNO to a branch-specific tourism language. These mediators can therefore help to validate the data with other market data. Furthermore, there is a great need to define KPIs out of the use of PMD, so that findings can be monitored over time and compared with other destinations to enhance the informative value of the data.

Additionally, with the present paper, DMOs get assistance in interpreting PMD. This can be of practical value with regard to the rising number of practical projects and use cases from DMOs. This paper recommends handling PMD in the context of reference frameworks (e.g. classical market research data or data from accommodation statistics). An isolated analysis of the PMD may lead to misinterpretations both of the volume of signals (absolute numbers) and the identification of tourism-specific target groups. A meta-network analysis of the use of big data in tourism comes to similar conclusions, as "[b]ig data cannot replace all data sources and industries should not disregard traditional observations or domain knowledge when making decisions" (Li & Law, 2020, p. 10).

PMD can, however, be used as a kind of early indicator. In combination with traditional market data, such as weather data, etc, forecast models of tourism demand can be calculated even more accurately. The same goes for calculating seasonality apart from a monthly based perspective using Gini coefficients. PMD can calculate this more precisely based on daily values, ignoring different holiday periods and bank holidays.

7. Conclusions

Referring to the research questions, the results of this research show that currently, there are a number of validity and reliability issues about PMD in the German market. Discussing three different approaches, the main barriers are seen in correctly identifying tourists and distinguishing them from non-tourists in the PMD.

Nevertheless, PMD can be used to identify inter- and intra-destination movement patterns, as the strength of the data lies in answering new research questions in showing finely resolved temporal and spatial data. Furthermore, analysing the seasonality and structure of tourist demand is a good way to apply PMD.

A research agenda is set out for tourism research in the future, focusing on identifying tourist signals out of PMD in spite of strict data protection rules. If target group identification is not reliable, however, this will inevitably lead to faulty volume assessments, both with respect to the number of tourists and tourism flows. This conclusion seems to be crucial for the usability of passive mobile data not only for tourism research but also for management.

The overwhelming technical potential lying in these data may sometimes lead to enthusiasm and the application of sophisticated analytical methods without always keeping the basics – data validity and reliability – in mind. As opposed to the argument of Demunter (2017), who sees access to the data as the main barrier, it can be perceived that, from a German perspective, the correct identification of tourism flows and its discrimination against other forms of mobility as the main challenge today.

Authors contribution

Julian Reif, Corresponding author, Study conception and design, Acquisition of data, Formal analysis, Visualisation and interpretation of data, Drafting of manuscript, Critical revision, Dirk Schmücker, Study conception and design, Acquisition of data, Formal analysis, Visualisation and interpretation of data, Drafting of manuscript, Critical revision.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

References

Ahas, R., Aasa, A., Mark, Ü., Pae, T., & Kull, A. (2007). Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management, 28*(3), 898–910.

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.

Ahas, R., Armoogum, J., Esko, S., Ilves, M., Karus, E., Madre, J. L., et al. (2014). *Feasibility study on the use of mobile positioning data for tourism statistics: Consolidated report*. Luxembourg: Publications Office.

Batista e Silva, F., Marín, H., Mario, A., Rosina, K., Ribeiro Barranco, R., Freire, S., et al. (2018). Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management, 68*, 101–115.

Bauder, M. (2019). Engage! A research agenda for Big Data in tourism geography. In D. Müller (Ed.), *A research agenda for tourism geographies* (pp. 149–158). Cheltenham: Edward Elgar Publishing.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & von Schreeb, J. (2011). Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Medicine, 8* (8), Article e1001083.

Beyerer, J., Richter, M., & Nagel, M. (2018). *Pattern recognition: Introduction, features, classifiers and principles*. Berlin, Boston: De Gruyter.

Blumenstock, J. E. (2012). Inferring patterns of internal migration from mobile phone call records: Evidence from Rwanda. *Information Technology for Development, 18*(2), 107–125.

Bokelmann, B., & Lessmann, S. (2019). Spurious patterns in Google Trends data: An analysis of the effects on tourism demand forecasting in Germany. *Tourism Management, 75*, 1–12.

Bonné, B., Barzan, A., Quax, P., & Lamotte, W. (2013). WiFiPi: Involuntary tracking of visitors at mass events. Madrid, Spain. In *2013 IEEE 14th international Symposium on "A world of wireless, mobile and multimedia networks" (WoWMoM) 4–7 june 2013* (pp. 1–6).

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662–679.

Brovelli, M. A., Minghini, M., & Zamboni, G. (2016). Public participation in GIS via mobile applications. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 306–315.

Caldeira, A. M., & Kastenholz, E. (2019). Spatiotemporal tourist behaviour in urban destinations: A framework of analysis. *Tourism Geographies, 16*(2), 1–29.

Debbage, K. G. (1991). Spatial behavior in a Bahamian resort. *Annals of Tourism Research, 18*(2), 251–268.

Demunter, C. (2017). *Tourism statistics: Early adopters of big data?*. Retrieved 25.11.19 from https://ec.europa.eu/eurostat/documents/3888793/8234206/KS-TC-17-004-EN-N.pdf/a691f7db-d0c8-4832-ae01-4c3e38067c54.

Edwards, D., Dickson, T., Griffin, T., & Hayllar, B. (2010). Tracking the urban visitor: Methods for examining tourists' spatial behaviour and visual representations. In G. Richards, & W. Munsters (Eds.), *Cultural tourism research methods* (pp. 104–114). Wallingford: CABI.

Edwards, D., & Griffin, T. (2013). Understanding tourists' spatial behavior: GPS tracking as an aid to sustainable destination management. *Journal of Sustainable Tourism, 21* (4), 580–595.

Eurostat. (2013). *Methodological manual for tourism statistics*. Luxembourg: Publications Office of the European Union. Version 2.1.

Gao, S., Liu, Y., Wang, Y., & Ma, X. (2013). Discovering spatial interaction communities from mobile phone data. *Transactions in GIS, 17*(3), 463–481.

Geng, Y., Du, J., & Liang, M. (2019). Abnormal event detection in tourism video based on salient spatio-temporal features and sparse combination learning. *World Wide Web, 22*(2), 689–715.

Ghajar-Khosravi, S., & Chignell, M. (2017). Pragmatics of network centrality. In L. Sloan, & A. Quan-Haase (Eds.), *The SAGE handbook of social media research methods* (pp. 309–327). Los Angeles, London: Sage Reference.

Gibson, H. J. (2016). Tourist. In J. Jafari, & H. Xiao (Eds.), *Encyclopedia of tourism* (pp. 957–959). Cham: Springer International Publishing.

González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *Geojournal, 69*(4), 211–221.

Guo, X. (2016). Application of meteorological big data. In *2016 16th international Symposium on Communications and information technologies (ISCIT)*. Qingdao, China.

Hägerstrand, T. (1970). What about people in regional science? *Papers - Regional Science Association, 24*, 7–21.

Hardy, A., Hyslop, S., Booth, K., Robards, B., Aryal, J., Gretzel, U., et al. (2017). Tracking tourists' travel with smartphone-based GPS technology: A methodological discussion. *Information Technology & Tourism, 17*(3), 255–274.

Kamenjuk, P., Aasa, A., & Sellin, J. (2017). Mapping changes of residence with passive mobile positioning data: The case of Estonia. *International Journal of Geographical Information Science, 31*(7), 1425–1447.

Kellner, L., & Egger, R. (2016). Tracking tourist spatial-temporal behavior in urban places: A methodological overview and GPS case study. In A. Inversini, & R. Schegg (Eds.), *Information and communication technologies in tourism* (pp. 481–494). Cham: Springer International Publishing.

Keul, A. G., & Kühberger, A. (1996). *Die Strasse der Ameisen. Beobachtungen und Interviews zum Salzburger Städtetourismus*. München.

Kitchin, R. (2013). Big data and human geography. *Dialogues in Human Geography, 3*(3), 262–267.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 1–12.

Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society, 3*(1), 1–10.

Kuusik, A., Tiru, M., Ahas, R., & Varblane, U. (2011). Innovation in destination marketing. *Baltic Journal of Management, 6*(3), 378–399.

Lau, G., & McKercher, B. (2006). Understanding tourist movement patterns in a destination: A GIS approach. *Tourism and Hospitality Research, 7*(1), 39–49.

Leiper, N. (1979). The framework of tourism: Towards a definition of tourism, tourist and the tourism industry. *Annals of Tourism Research, 6*(4), 390–407.

Li, X., & Law, R. (2020). Network analysis of big data research in tourism. *Tourism Management Perspectives, 33*, 100608.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301–323.

Maschke, J. (2014). *Sonderauswertung der Grundlagenstudie "Tagesreisen der Deutschen" für die Stadt und die übrige Metropolregion Hamburg*. München.

Masso, A., Silm, S., & Ahas, R. (2019). Generational differences in spatial mobility: A study with mobile phone data. *Population, Space and Place, 25*(2), 1–15.

McKercher, B., & Lau, G. (2009). Methodological considerations when mapping tourist movements in a destination. *Tourism Analysis, 14*(4), 443–455.

McKercher, B., & Lew, A. A. (2003). Distance decay and the impact of effective tourism exclusion zones on international travel flows. *Journal of Travel Research, 42*(2), 159–165.

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. *Journal of Urban Technology, 21*(2), 91–107.

Önder, I., Gunter, U., & Gindl, S. (2019). Utilizing Facebook statistics in tourism demand modeling and destination marketing. *Journal of Travel Research, 59*(2), 195–208.

Oppermann, M. (1995). A model of travel itineraries. *Journal of Travel Research, 33*(4), 57–61.

Pesonen, J., & Horster, E. (2012). Near field communication technology in tourism. *Tourism Management Perspectives, 4*, 11–18.

Poon, A. (1988). Tourism and information technologies. *Annals of Tourism Research, 15*, 531–549.

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

Raun, J., Shoval, N., & Tiru, M. (2020). Gateways for intra-national tourism flows: Measured using two types of tracking technologies. ahead of print *International Journal of Tourism Cities*. https://doi.org/10.1108/IJTC-08-2019-0123.

Reif, J. (2019a). Die Nutzung von Mobilfunkdaten in der Tourismusforschung—Das Beispiel Tagestourismus in Hamburg. In S. Groß, J. Peters, R. Roth, J. Schmude, & A. Zehrer (Eds.), *Wandel im Tourismus. Internationalität, Demografie und Digitalisierung* (pp. 31–50). Berlin: ESV.

Reif, J. (2019b). Touristische Aktionsräume und die Wahrnehmung von Crowding. Das Beispiel Tagestourismus in Hamburg. *Zeitschrift für Tourismuswissenschaft, 11*(2), 257–287.

Romero Palop, J. D., Arias, J. M., Bodas-Sagi, D. J., & Lapaz, H. V. (2019). Determining the usual environment of cardholders as a key factor to measure the evolution of domestic tourism. *Information Technology & Tourism, 21*(1), 23–43.

Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing big data sources. *Tourism Management, 66*, 13–25.

Saluveer, E., Raun, A., Tiru, M., Altin, L., Kroon, J., Snitsarenko, T., et al. (2020). Methodological framework for producing national tourism statistics from mobile positioning data. *Annals of Tourism Research, 81, 102895*.

Sauter, M. (2017). *From GSM to LTE-advanced pro and 5G: An introduction to mobile networks and mobile broadband* (3rd ed.). Hoboken: Wiley.

Schmid, T., Bruckschen, F., Salvati, N., & Zbiranski, T. (2017). Constructing sociodemographic indicators for national statistical institutes by using mobile phone data. Estimating literacy rates in Senegal. *Journal of the Royal Statistical Society A, 180*(4), 1163–1190.

Schönfelder, S., & Axhausen, K. W. (2003). Activity spaces: Measures of social exclusion? *Transport Policy, 10*(4), 273–286.

Shoval, N. (2008). Tracking technologies and urban analysis. *Cities, 25*, 21–28.

Shoval, N. (2011). Time geography and tourism. In J. Wilson (Ed.), *The Routledge handbook of tourism geographies* (pp. 174–180). London/New York, NY: Routledge.

Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: The first decade. *Tourism Geographies, 18*(5), 587–606.

Shoval, N., & Ahas, R. (2018). The potential of tracking technologies, smartphones and sensors for tourism management and planning of destinations. In C. P. Cooper, S. Volo, W. C. Gartner, & N. Scott (Eds.), *Applications of theories and concepts to tourism* (pp. 433–442). Los Angeles, CA: Sage Reference.

Shoval, N., & Isaacson, M. (2010). *Tourist mobility and advanced tracking technologies*. New York, Oxford: Routledge.

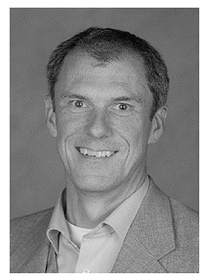
Shoval, N., McKercher, B., Ng, E., & Birenboim, A. (2011). Hotel location and tourist activity in cities. *Annals of Tourism Research, 38*(4), 1594–1612.

Shoval, N., Schvimer, Y., & Tamir, M. (2018). Real-time measurement of tourists' objective and subjective emotions in time and space. *Journal of Travel Research, 57* (1), 3–16.

Signorelli, S., Reis, F., & Biffignandi, S. (2016). *What attracts tourists while planning for a journey? An analysis of three cities utilising wikipedia page views*. Retrieved 25.11.19 from http://tsf2016venice.enit.it/images/articles/Papers_Forum/2.4_What%20attracts%20tourists%20while%20planning%20for%20a%20journey%20-%20an%20analysis%20of%20three%20cities%20through%20Wikipedia%20page%20views.pdf.

Silm, S., Ahas, R., & Mooses, V. (2018). Are younger age groups less segregated? Measuring ethnic segregation in activity spaces using mobile phone data. *Journal of Ethnic and Migration Studies, 44*(11), 1797–1817.
Singleton, A., & Arribas-Bel, D. (2019). Geographic data science. *Geographical Analysis, 1*(4). https://doi.org/10.1111/gean.12194.
Song, H., & Liu, H. (2017). Predicting tourism demand using big data. In Z. Xiang, & D. R. Fesenmaier (Eds.), *Analytics in smart tourism design* (pp. 13–30). Cham: Springer International Publishing.
Steenbruggen, J., Tranos, E., & Nijkamp, P. (2015). Data from mobile phone operators: A tool for smarter cities? *Telecommunications Policy, 39*(3–4), 335–346.
Stylos, N., & Zwiegelaar, J. (2019). Big data as a game changer: How does it shape business intelligence within a tourism and hospitality industry context? In M. Sigala, R. Rahimi, & M. Thelwall (Eds.), *Big data and innovation in tourism, travel, and hospitality* (pp. 163–181). Singapore: Springer Singapore.
Taylor, P. J. (1971). Distance transformation and distance decay functions. *Geographical Analysis, 3*(3), 221–238.
Thimm, T., & Seepold, R. (2016). Past, present and future of tourist tracking. *Journal of Tourism Futures, 2*(1), 43–55.
Tiru, M., Kuusik, A., Lamp, M. L., & Ahas, R. (2010). LBS in marketing and tourism management: Measuring destination loyalty with mobile positioning data. *Journal of Location Based Services, 4*(2), 120–140.
Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*, 234–240.
United Nations. (2010). *International recommendations for tourism statistics 2008*. New York, NY: United Nations. Studies in methods series M No. 83/Rev.1.
Versichele, M., Neutens, T., Delafontaine, M., & van de Weghe, N. (2012). The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent festivities. *Applied Geography, 32*(2), 208–220.
Ward, M., & Shafaghi, M. (2013). A literature framework analysis of online hotel accommodation process factors. In Z. Xiang, & I. Tussyadiah (Eds.), *Information and communication technologies in tourism 2014* (pp. 481–494). Cham: Springer International Publishing.
Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data mining: Practical machine learning tools and techniques* (4th ed.). Cambridge, MA: Morgan Kaufmann.
Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44*, 120–130.
Zhao, Z., Shaw, S. L., Xu, Y., Lu, F., Chen, J., & Yin, L. (2016). Understanding the bias of call detail records in human mobility research. *International Journal of Geographical Information Science, 30*(9), 1738–1762.
Zoltan, J., & McKercher, B. (2015). Analysing intra-destination movements and activity participation of tourists through destination card consumption. *Tourism Geographies, 17*(1), 19–35.

Julian Reif, graduated as geographer at the University of Bonn and has been project manager at the Institute for Management and Tourism (IMT) - now German Institute for Tourism Research (DITF) of the West Coast University of Applied Sciences since 2012. Julian has several years of experience in the management and implementation of primary and secondary market research projects for tourism demand. His main areas of work and research interests are: Tourism demand, digitalization in tourism and tourism impacts. He is currently doing his doctorate at the University of Bonn on the measurement of tourist spatio-temporal behavior.



Prof. Dr. Dirk Schmücker has been head of research at the NIT since 2007 and is Professor with focus on tourism demand and applied empirical tourism research at the West Coast University of Applied Sciences. He studied Applied Cultural Sciences and received his doctorate with a thesis on the information behavior of holidaymakers at the Faculty of Economics and Social Sciences of the University of Lueneburg. Prior to his current positions, he was managing partner of Project M GmbH. Dirk's work focuses on consumer research, in particular information and decision-making behavior, tourism impact studies and impact models, qualitative and quantitative market research methods as well as quantitative data analysis.