Joschka Bischoff, Michal Maciejewski, Alexander Sohr

Analysis of Berlin's taxi services by exploring GPS traces

Analysis of Berlin’s taxi services  
by exploring GPS traces

Joschka Bischoff  
Technische Universität Berlin  
Department of Transport Systems Planning  
and Transport Telematics  
Berlin, Germany  
Email: bischoff@vsp.tu-berlin.de

Michal Maciejewski  
Poznan University of Technology  
Division of Transport Systems  
Poznan, Poland  
Technische Universität Berlin  
Department of Transport Systems Planning  
and Transport Telematics  
Berlin, Germany

Alexander Sohr  
German Aerospace Center (DLR)  
Institute for Transport technology  
Berlin, Germany

Abstract—With current on-board GPS devices a lot of data  
is being collected while operating taxis. This paper focuses on  
analysing travel behaviour and vehicle supply of the Berlin taxi  
market using floating car data (FCD) for one week each in 2013  
and 2014. The data suggests that there is generally a demand peak  
on workday mornings and a second peak over a longer time in  
the afternoon. On weekends, the demand peaks shift towards the  
night. On the supply side, drivers seem to adapt to the demand  
peaks very efficiently, with fewer taxis being available at times of  
low demand, such as during midday. A spatial analysis shows that  
most taxi trips take place within the city centre or from/to  
Tegel Airport, the city's largest single origin and destination.  
Drivers spend a large amount of their work time on waiting for  
customers and the taxi rank at Tegel Airport is the most popular  
one.

Keywords—Taxi Berlin, floating car data, FCD, taxi demand,  
taxi supply, Tegel Airport

I. INTRODUCTION

Although taxi services are one of the core elements of  
urban transport systems, the scientific focus on them had  
been limited until recent developments in information and  
communications technology (ICT). Nowadays taxis equipped  
with mobile devices may be monitored and managed remotely.  
This has opened up a broad range of new services, such as  
shared taxis, Uber and similar, or even autonomous taxis in  
the future. Due to the onboard devices a lot of data is being  
collected while operating taxis. These data make up a very  
extensive source of information on urban traffic that can be  
used, for example, to estimate time-dependent travel times.  
They can also provide an insight into the travel behaviour,  
which is the subject of this paper.

The paper deals specifically with taxis in Berlin. Being  
Germany’s capital and largest city, the city forms a hotspot  
for business, public services and tourists alike. In combination  
with a comparatively low percentage of private car ownership  
among its inhabitants, the city is an interesting location for  
taxi businesses.

II. RELATED WORK

There have been several studies on analysing taxi data all  
over the world, most of them focusing on the Asian market.  
In Europe, taxi trips in Lisbon were analysed to explore  
relationships between origins and destinations of taxi trips and  
predict taxi demand [1]. The results of such analyses may be  
used to improve the quality of taxi dispatch. In Hangzhou the  
trajectories of 5 500 taxis were analysed to determine taxi  
drivers’ best strategies to pick up passengers at a given time  
and location [2]. A study in Shenzhen explored taxidrivers’  
operation patterns with the focus on differences between the  
behaviour of top drivers and of ordinary ones [3]. For Beijing,  
a recommendation system for taxi drivers and passengers was  
implemented to help them in efficient mutual finding [4]. In  
Shanghai, spatio-temporal profitability maps were introduced  
to help taxi drivers with reducing cruising miles [5]. Similarly,  
prediction of taxi demand hotspots has been studied for Taipei  
[6] and for Hangzhou [7].

Taxi trajectories have been used for estimating urban link  
travel times, e.g. in Manhattan [8], and for time-dependent  
path finding. In Berlin, fleet travel times were reduced by  
measuring real-time travel times using floating car data from  
taxis [9]. Similarly, taxi trajectories were used to devised  
shortest path searching algorithms for Wuhan [10] and Beijing  
[11]. In Beijing, a parallel algorithm to process massive taxi  
traces for detecting traffic hot spots was proposed [12].

Taxi trajectories have been used also in many other areas.  
One approach is to use them to detect flawed urban planning  
[13], another one to determine land use by analysing taxi  
pick-ups and drop-offs [14]. Also spatial interaction models  
were calibrated using taxi trajectories [15].

III. DATA

In December 2012, around 8 000 taxis were licensed to  
operate in the city. Those vehicles were operated by around  
18 000 taxi drivers which were organised in 3 000 taxi
companies. Of these companies, 2,368 operate only one vehicle each, 206 companies have two vehicles and 427 three or more taxicabs [16]. Metered taxi fares are strictly regulated in Berlin. Market access for new operators comes only with technical entrance barriers, but, in contrast to many other countries, there is no limitation in the amount of available taxi licenses.

The data used in the study is based on historical trajectories of taxicabs managed by Taxi Berlin, Berlin’s largest taxi association dispatching roughly 70% of the city’s taxi fleet. Additionally a small fleet of about 30 taxis in Potsdam is connected to the system. The GPS position of each taxi is logged every 30–60 seconds depending of its current status. If the cab has a customer on board, the position is sent once in a minute, because it is not available for dispatching purposes and thus its position is less important for the taxi association. Consequently empty vehicles are tracked more frequently to warrant an allocation of the closest vehicle to new customers. Overall, around 2 million GPS positions are recorded per day. This way, a database of historical GPS traces arises. Each entry contains the following fields:

- temporary taxi identifier (ID) – changing several times a day to assure that taxi drivers’ privacy is not compromised (Taxi Berlin is an association and not the company hiring taxi drivers)
- coordinates – geographical location
- time – time the record is logged
- status – one of ten different statuses such as driving to a customer, driving with a customer, waiting for a customer, waiting at a taxi rank, etc.

To build trajectories from raw GPS positions, first some basic plausibility checks are performed. These include:

- the taxi’s location
- the entry’s uniqueness
- improbable location variations
- a validation of the currently driven speed

In the next step, based on the status information, the start and the end of the trajectories is determined. These trajectories are then map matched and travel times within the city can be derived, which is the main purpose of data collection both for commercial and research reasons [17].

Extraction of taxi demand and spatiotemporal supply distribution from the GPS traces results in the following data:

- hourly zone-based1 OD-matrices based on the vehicles’ statuses
- the amount of taxicabs per status and zone in five-minute intervals
- the duration and time stamp of each temporary taxi ID with the relevant vehicle status

There are, however, some limitations related to the properties of the trajectory database. First of all, due to the automatic anonymisation (IDs are not fixed over a day), it is not possible to provide full daily trajectories of individual taxi drivers. Secondly, data sets only include vehicles that are associated within Taxi Berlin. Consequently, it does not include taxi drivers which exclusively pick up passengers at taxi ranks or by roaming through streets. These drivers supposedly constitute for a large share of taxi supply at places with a high likelihood for on-the-spot trips, e.g. airports or night clubs. Last but not least, recent estimation suggest that up to 40% of all taxi trips in Berlin are black-market rides without proper fiscal records [18]. This may result in errors while analysing status data, as drivers will set their status to ‘available’ when, in fact, they have a customer on board. Hence, not all trips could be tracked correctly for this paper. Thus, the demand and amount of taxi trips may in fact be higher.

Two weekly data sets of GPS traces are used in this study. The first set dates from 15 to 21 April 2013, whereas the second includes data from 7 to 13 April 2014. The demand for taxi trips is depending on several outer impacts such as holidays, availability of transport alternatives and weather conditions. Overall, these outer impacts were rather similar in both weeks monitored in this study: Both weeks lie outside major German holiday periods, however, during the week in 2014, the federal states Niedersachsen and Bremen had school holidays. In neither week any large-scale event took place. Rapid public transit was running without any planned major interruptions during both time periods. For the week in 2013, the outside temperatures were overall higher than 2014. In neither week high rainfall had been measured.

IV. FINDINGS

A. Time-based taxi demand and supply

A detailed distribution of taxi supply (the number of taxis logged into the system) and demand (the number of requests submitted per hour) over a week are presented in Fig. 1 for 2013 and Fig. 2 for 2014. For the demand side, the following applies:

- Taxi demand during weekdays (Monday morning – Friday morning) follows a clear pattern with a major peak around 9 am and a smaller one during the afternoon.
- Friday is demand-wise the busiest day, with a somewhat smaller morning peak followed by a high demand during evening and night hours.
- The weekly peak-hour for taxi demand is Saturday night.
- Besides Friday and Saturday night, the demand for taxi services during weekends is low.

[1] For Berlin, zones are determined by LORs http://www.stadtentwicklung.berlin.de/planen/basisdaten_stadtentwicklung/lor/, for the surrounding areas according to municipalities.
As both figures show, the supply adapts surprisingly well to the demand. Peaks in demand are generally matched by a larger supply. For most of the time, the number of available taxis is higher than of hourly request submissions, which means that every taxi serves less than one request per hour. Saturday night in 2013 is the only exception to this with less taxis being in service than requests per hour.

Overall, the week in 2013 has a higher demand for taxi trips, with 198,445 trips in total versus 169,651 trips for the week in 2014. This is mainly due to a higher demand during Saturday night. However, the demand in 2014 is lower at most other times as well (see Fig. 3). On the other hand, the amount of active vehicles is higher during most times in 2014 (see Fig. 4). This goes in line with the trend of the steadily growing taxi fleet in Berlin over the last years[16].

C. Location-based taxi demand

Origins and destinations of taxi trips are generally spread all over the city. The majority commences and ends within the city centre (defined as the inner railway circle), as Figure 7 shows. The data suggest that there are more trips from the city centre to the outside than vice versa, though this could be due to the fact that many trips from Tegel Airport are made by taxis that are not affiliated with a radio-cab operator and most trips from Schönefeld airport are made by taxis which are registered in a surrounding parish rather than Berlin. Both types of taxis are not a part of the data set.

As to the distribution of trip distances (cf. Fig. 6), there are few trips of less than two kilometre beeline distance, whereas a trip distance between two and three kilometres is the most common. Longer distances are less and less likely – not even ten percent of all trips are ten kilometres or longer.
Trips between the city centre and Tegel Airport make up for most of the top OD relations and is hence also the most sought-after destination.

Other frequently used relations are almost exclusively within the city centre, with one notable exception (see below).

In general, the central zones are more likely to be destinations of taxi trips. Yet the picture is rather diverse with some outskirts also being a popular destination.

The city’s second airport, Schönefeld, is neither a destination nor an origin in the top 100 relations.

The notable exception is one relation are trips from southern Pankow to the northern part of the district, with a total of 106 trips. In the opposite direction only 44 trips have been measured. The relation seems unusual, but can be explained by construction works on the tram line M1 which connects the area during the week monitored. Especially in the northbound direction the offered bus replacement may not have provided the expected level of service, for example, due to lost connections. This may have lead to an increase in taxi demand.

Points of special interest are often a destination for taxi trips. These include train stations, the fair grounds and major event locations (the airport-related taxi traffic is analysed in Section IV-D). Trips to a selection of such locations are displayed in Figure 9. Of all train stations with regular long-distance connections, Hauptbahnhof (Central Station) has the biggest attraction for taxi traffic, with more than ten weekly trips from almost any zone in the city centre. Ostbahnhof (in the east) and Spandau (in the far west) stations attract taxi business to a much lesser degree and mainly on a district-wide level. Some zones notably have an equally high attraction to both Hauptbahnhof and Ostbahnhof. Finally, Südkreuz (in the south) and Gesundbrunnen (in the north) are almost exclusively destinations for short-range taxi trips. The fair ground has a
certain amount of trips from Tegel Airport and the western city centre around Kurfürstendamm. Since there was no bigger trade fair, this seems reasonable. For event locations, no clear pattern emerges. Taxi trips to zones near them are generally short-distance trips from within the city centre, with a notable exception around the Tempodrom. However the likelihood of travelling by taxi to a mass event is also depending on the actual event itself.

D. Airports

As already seen in the previous analyses, taxi trips to and from Tegel and Schönefeld airports constitute for a large amount of all taxi trips in Berlin. The following numbers are once more based on the week in April 2013. The amount of trips to Tegel is much higher (14,144 trips) due to its proximity to the city, the lack of a direct rail link and higher air passenger numbers. To Schönefeld, only 1,938 trips took place. On the other hand, with 7.48 km, the average beeline trip distance to Tegel is only half as long compared to Schönefeld (14.27 km). As Fig. 10 shows, origins for trips to the airports are similar in the city centre, but tend to differ in the outskirts. One possible interpretation would be that passengers departing in southern parts of the city would rather use a taxi to Schönefeld than to Tegel and vice versa. Thus the likelihood of taxi usages decreases with a growing distance from the airport.

E. Symmetries in taxi demand

For taxi dispatching, one of the more important questions is, whether taxi demand between different city areas is symmetric and whether trips in one direction occur at different times than in the other direction. In general, demand for taxi trips is rather symmetric between zones. The most notable exceptions may be explained by the aforementioned reduced share of Taxi Berlin in serving trips originating at both airports. Fig. 11 provides an overview. The asymmetry is best illustrated by taxi flows incoming to and outgoing from Tegel Airport.

F. Idle taxi distribution

Within Berlin, there are roughly 400 taxi ranks, which is where most taxi drivers gather. Figure 12 provides an overview of rank locations and the average number of idle taxis in each zone at each moment during the week in 2014. Tegel Airport attracts overall most vehicles. Drivers accept waiting times of up to several hours at this rank. One reason might be that the following ride has the chance of being especially rewarding. Furthermore, the general infrastructure for waiting taxi drivers is rather good here, so some drivers might prefer it over other ranks.
with biggest demand peaks being weekday mornings and weekend nights. These peak patterns are repeating both in 2013 and 2014. The top single origin and destination is clearly Tegel Airport, most trips however take place within the city centre. A correlation between taxi demand and interruptions in public transit is noticeable in at least one case.

ACKNOWLEDGMENT

The authors would like to thank the Einstein foundation for co-funding this paper and TAXI Berlin TZB GmbH for data provision.

REFERENCES


