UTILITY OF GPS DATA FOR URBAN BICYCLE TRAFFIC PLANNING IN GERMANY: POTENTIALITY, LIMITATIONS AND PROSPECTS

STEFAN HUBER[1], SVEN LIßNER[1], ANGELA FRANCKE[2]

[1] Transport Ecology, Institute of Transport Planning and Road
Traffic at the Faculty of Transport and Traffic Science, Technical University Dresden, Germany

[2] Traffic and Transportation Psychology, Institute of Transport Planning and Road
Traffic at the Faculty of Transport and Traffic Science, Technical University Dresden, Germany

ABSTRACT

Planning bicycle infrastructure significantly depends on data that provide adequate information. Various studies indicate that GPS data, which have been collected via smartphone application by cyclists themselves, could provide that information. The article presents the results of a recently conducted study that evaluates the usability of such data for bicycle traffic planning in German cities. We used different methods (web-survey, focus group interview, data analysis) to investigate data needs of German municipal traffic planners and oppose it to the information deduced and computed from commercially available data provided by Strava Inc. The article reveals that the provided data are, in general, useful, but there are also serious limitations that must be considered.

Keywords: bicycle traffic, GPS data, traffic planning, smartphone.

1 INTRODUCTION

Since society struggles with negative effects of urban transport on people and environment (e.g. noise or pollutant emission), the bicycle offers high potential to achieve a more sustainable urban transport system [1]. The reputation of the bicycle as a clean, cheap and healthy means of transport finally led to increasing bicycle use in recent years in Germany [2]. In order to promote this development, urban areas need to implement favourable conditions for cyclists. A key factor for that is a well-developed infrastructure aligned to users' requirements [3]. Scarce financial situation of cities and municipalities, however, force traffic planners to prioritize in order to use the few financial resources in the most effective manner.

Therefore, reliable information are needed that serve as a robust basis for decision-making. However, data availability for bicycle planning is generally low and information is mostly based on extensive and costly data collection. Although there are commonly used data sources like automatic counting devices, manual counts or traffic surveys [4, 5] these data do not provide comprehensive movement patterns but local counts [6]. The increased distribution of GPS-enabled devices – especially smartphones – provides new possibilities to generate traffic data as more than 81% of all Germans used a smartphone in 2017 [7]. Furthermore, there are many fitness apps offering the opportunity to record routes or speeds when biking. In 2017, there were approximately ten million users across 195 countries using the Strava app and the number of users is constantly increasing [8]. App providers recognized the potential of the data and started selling anonymized and aggregate data.

Therefore, the major research question is: are the offered data reasonable for bicycle transport planning and do they comply with the requirements of transport planners of German municipalities? There are further sub-questions occurring in this regard: (a) what data are currently available and used for planning, (b) which data would be promising and which questions would planners try to answer using it, (c) could commercially available data fill this information gap and (d) what are major challenges for planners using these data?

The article proceeds as follows: in Section 2, we give an overview of the current state of research regarding the topic in general and specifically focussing on Germany. The methods used to answer the questions raised above are described in Section 3. In Section 4, the results of the study are presented. They are finally discussed in Section 5, which will close the article by drawing a conclusion and giving an outlook to further research possibilities.

2 CURRENT STATE OF RESEARCH

There has been quite intense research in recent years leading to various studies and articles related to the topic. A general overview of GPS technology, its development and utilization of GPS data for transportation research can be found in Ref. [9]. Reference [10] also discuss the implications of big data for analysis, transport modelling and planning in a more general way. Reference [11] further specified the utilization of such data by describing possibilities to use fitness data from private companies.

Strava has been one of the first companies to sell GPS-based data of cyclist as they started to provide data to local authorities or research institutions [13]. Since then, several studies used the data offered by Strava and other app providers in different areas (see Refs [12] and [13]).

Reference [6] compares the amount of cyclists using the Strava app with manual counts in Victoria, British Columbia (Canada). They incorporate the Strava data to predict cycling volumes at unknown locations and claim that the data allow a more detailed coverage of cycling volumes in the area. Reference [14] presents a spatial and temporal coverage of cycling patterns in Johannesburg (South Africa) derived from the Strava data of 2014. As there have been no data regarding bicycle traffic in Johannesburg, the study gives a first insight into the subject. In a further study, Ref. [15] developed a regression model in order to predict the number of cyclists counted at automatic counting points in Malmö (Sweden). The study shows that using the GPS data could potentially improve the prediction accuracy. Ref. [16] also uses the Strava data to predict ridership volumes. The study reveals how changes in cycling infrastructure could influence short-term cycling activities. The most recent studies by Refs [17] and [18] deal with similar aspects. Reference [17] use Strava data to monitor changes in spatial patterns of cyclists in Ottawa-Gatineau (Canada) and conclude that the Strava data are useful to observe changes in spatial and temporal distribution of cyclists. The study of Ref. [18] shows that the data generally match conventional data for Greater Sydney (Australia), but it also points out that there are still limitations because the data do not match for all locations.

The sections above show that there are several studies investigating the possibility to use crowdsourced data for bicycle transport planning in different areas. However, there is no research focussing on Germany and no study – neither in Germany nor in other parts of the world – going beyond pure data analysis. The found studies mostly investigate how to predict traffic volumes, but there is no assessment of what transport planners exactly need. Thus, the question remains if the data, offered by app providers, comply with the requirements of transport planners of German municipalities.

3 METHODS

We used different methods to shed light on the topic. These methods are presented in the following sections.

3.1 Survey

The survey is an appropriate method to investigate several issues and has widely been applied in transportation science because it possesses different advantages [19]. The web-survey, as

a special form of the survey, is very attractive for research as it is easy to conduct and cheap to implement. A comprehensive description of advantages and disadvantages can be found in Refs [20, 21].

Due to the advantages of web-based surveys, we conducted such a survey to investigate the topic and answer the research questions stated in Section 1. Since no survey mode is clearly superior to the others, we also ensured an analogue survey that could have been send to participants via post to minimize the disadvantages of the web-based survey (see Ref. [21]).

The survey initially has been sent to bicycle transport commissioners of all German cities with more than 100,000 inhabitants. However, the web-survey was further distributed by planning commissioner to other colleagues so that other cities (< 100,000 inhabitants) participated, as well. Within a first category, general questions have been addressed to traffic planning commissioners (e.g. city size, number of people responsible for planning, available financial and human resources for planning). The second category contained specific questions related to currently used data and experiences regarding the utilization of GPS-based data. The survey generated a return of $n = 61$. Within the sample, the majority of cities (31.1%) possess less than 100,000 inhabitants, 29.5% with 100,000–200,000 inhabitants, 24.6% of cities have 200,000–500,000 inhabitants and only few cities possess up to one million inhabitants or more (13.1%).

3.2 Focus group interviews

Conducting focus group interviews (FGIs) is a common method to investigate information, ideas, attitudes or perception of participants within a group where each group member possesses experience considering a given topic [22]. Originally used in marketing and communication research, the method found its way to social sciences decades ago and is being used for data collection by researchers across multiple disciplines, today [23]. Since FGIs are relatively easy to undertake, cost-efficient, moderate time consuming and very flexible [24], we chose this method to get in-depth knowledge based on the previously conducted survey.

Setting up the FGI, we complied with literature, which recommends small group discussion with selected group members (see Ref. [25]). We opted for non-random selection of participants in order to compile a homogeneous group and invited five transport planning commissioners of different German municipalities (Wolfsburg, Leipzig, Mainz, Dresden and Jena) to participate in the FGI. As recommended in literature (see Refs [22, 23–26]), a thoroughly prepared agenda and major questions helped to guide through the FGI. The major questions complied with the research questions raised in Section 1 and concerned data availability, data utilization and referring challenges as well as GPS-based data and question planners would try to answer using it.

3.3 GPS data analysis

In order to assess the utility of GPS-based data provided by private companies, we purchased a data set from Strava Inc. in 2015. We chose Dresden (Germany) as city for the case study. Within the city, which has about 550,000 inhabitants, more than 16% of the city population use the bike every day. This accounts for about 223,000 daily bicycle trips and results in a bicycle mode share of about 12% [27]. The Strava data of the study area contain around 3,200 users and 70,500 bicycle trips between June 2015 and June 2016. Due to data privacy, we did not receive GPS raw data. Instead, we sent the official road network file provided by the city

administration and a road network provided by Open Street Map to Strava. The company then carried out the map matching and all feasible calculation so that we finally received a road network file with supplementary information.

In order to check the values of traffic volume, we achieved linear regression of the Strava sample (explanatory variable) and the counting data from automatic counting points across the study area provided by the city administration (hourly traffic volumes for the period from January 2015 to June 2016). Furthermore, counting data from temporal manual counts (conducted from May to June 2016) were used, too.

As an important parameter for assessing the quality of the bicycle infrastructure, waiting times at intersections computed by Strava were validated using values from empiric measurements at nine intersections (measured during morning peak hour 7:45 to 10:00 AM).

To check the speed and speed distribution, the data of Strava (times cyclists spent on the particular links) were used to compute speeds, which then were opposed to empirical values. Therefore, speed measurements of 1,000 cyclists were carried out at three sites in the main cycling network and two locations in the secondary cycling network from 07:00 to 10:00 AM

Strava data also provide information about origination and terminating traffic and origin destination (OD) relations. Therefore, trip beginnings and ends are matched to zones (raster polygons of 1.0×1.0 km). Due to the fact that no detailed data for originating traffic are available for validation, we tested for correlation between trip beginnings and the number of inhabitants of the analysed zones using census data that were matched to the zones, too.

Further details referring to the analysis methods, supplementary data used for validation and variable computation can be found in [28].

4 RESULTS

In the following section, we present the study results. First, the currently used data and recent data needs of transport planners in German municipalities are presented. We subsequently show how purchased data could fill the gap between data availability and data needs and, thus, illustrate potentiality and limitation of the data.

4.1 Data availability and data utilization

Data availability and access to reliable data is of major importance to transport planners. However, referring to the survey and the FGI, data availability and utilization of proper data is a major challenge for municipal transport planners because available data mostly deal with motorized traffic. Most cities possess only few automatic counting points (mean = 3) that provide punctual bicycle transport volume year-round. However, more than half of the cities (50.8%) that participated in the survey do not have any automatic counting devices at all and, therefore, depend on short-term manual traffic counts. The survey results also reveal that short-term counts are rarely conducted so that manual counts do not provide a proper basis for broad planning. Nearly 25% of the cities never conduct short-term counts. More than 49% of the cities conduct a maximum of five counts per year, on average. This is quite alarming because the lack of proper data hampers planners in justifying projects to support bicycle transportation.

The survey and the FGI reveal that, in contrast to data availability, there are huge data needs. Broadly speaking, a bunch of data are needed for planning issues. Data needs and the issues planner would address are presented in Table 1.

Table 1: Data needs and issues to be addressed.

Data needs	Issues addressed
Traffic flow	Spatial distribution of cyclists, trip directions
Traffic volume	Number of cyclists using infrastructure
Surface quality	Comfort and safety
Safety	Points of danger and conflicting use
Dedication	Use of infrastructure that is not dedicated to bicycle use
Speed and waiting times	Quality of the bicycle transport system
Use of infrastructure	Differences in use of separated bike lanes, roadways etc.
Detour factor	Route choice and quality of network density
Daytime variation	Determining lane capacity
Trip purpose	Promoting bicycle traffic

Focussing on the use of GPS data, the survey reveals that about 77% of participants have not used GPS data, yet. However, the majority of planners assess that they are important for planning. More than 83% of planners that already used GPS data rate them important – so do more than 66% of planners that have not used such data, at all. However, GPS data remain unavailable for more than half of all planners although they would like to use it for planning and rate it 'helpful' for planning (70.5%).

There are several challenges that determine why planners are currently not using GPS data. A major reason for not using the commercially available data is the lack of human resources, as the situation of the investigated cites is very scarce in this manner. More than half of the cities (55.7%) employ only one full-time position or less for bicycle transport planning. Only 16.4% possess more than two. Furthermore, the lack of technical equipment is a major challenge for the majority of cities (58.7%). Concerns regarding data privacy protection are a challenging issue for most cities (53.6%), too. Additional costs are relevant as well and more than 43% of the cities face that obstacle: they stated a maximum budget of 2,600 Euro to purchase data for planning. For 27.5% of the cities, the budget is even less (1,000 Euro) and only few can afford more than 5,000 Euro (12.5%). Other major challenges refer to a lack of software needed for data processing (48.2%) and a lack of knowledge needed for data handling (32.9%).

The overall results of the survey and the FGI show that, on average, there is no or few data available for municipal traffic planning. In contrast, there are a lot of different questions and issues planners need to address. Facing scarce resources, the purchase and utilization of Strava data could be a solution for traffic planners to answer raised questions.

4.2 GPS data analysis

The analysis of the data provided by Strava reveals that several information, which are important to traffic planners (see Section 4.1), are directly deducible from the data or by further data processing. Referring to the results of the survey and the FGI, we present the results of the data analysis according to planners' data needs (see Table 1).

Traffic flow and traffic volume: The different data provided by Strava give an overview of the spatial distribution of cyclists within a city. Traffic flows and directional traffic volumes can easily be derived and illustrated using the data. There are two options to illustrate the data: (a) heat maps and (b) maps illustrating traffic volumes on links.

The heat map can be found online (see Ref. [29]) and offers a huge potential to planners because the access and utilization do not require special knowledge. It is an illustration of raw data (not matched to a map) and, thus, shows the spatial distribution of cyclists through illustrating the GPS trajectories. At first glance, heavily used links can be identified (see Fig. 1a). The missing projection on a certain GIS network (links) also helps planners to assess what ways are actually used by cyclist (e.g. shortcuts through inner courtyards). Therefore, heat maps can reveal important connections within the cycling network. However, heat maps do not provide any information about the number of cyclists on specific links because the visualized trips are not matched to the links of the network. Furthermore, an accumulation of GPS points, for example when slope is high and cyclist ride on low speed, leads to brighter glow effects of the map. The illustration then leads to the wrong suggestion that there is a higher amount of cyclists on the road section, which is actually not the case. Hence, a heat map does not provide any information on the direction of trips.

This particular gap can be overcome by using map matching procedures such as the ones reported by Ref. [30]. Strava matched the GPS trajectories of cyclists to specific links of the road network of the city of Dresden. The data are then aggregated to average traffic volumes on links. Due to low numbers of daily cyclists on network elements (only 0.5% of city inhabitants use the Strava app), average annual daily cyclists were used for aggregation. The data then provide information regarding the number of cyclists on a specific link for each direction and in a temporal resolution (daytime variation), too. Figure 1b illustrates the results of the map matching processed by Strava. It does not only reveal the distribution of cyclists in Dresden but illustrates the specific traffic volume on every link of the cities' road network.

The Strava data show a quite similar distribution of traffic volume on links as the data measured by counting stations. Therefore, the values for traffic volume are checked using a linear regression model. Linear regression is done on hourly basis for all counting stations (see Section 3). The regression model (equation: $f(x) = 0.5087x + 16.9719$) shows a relative
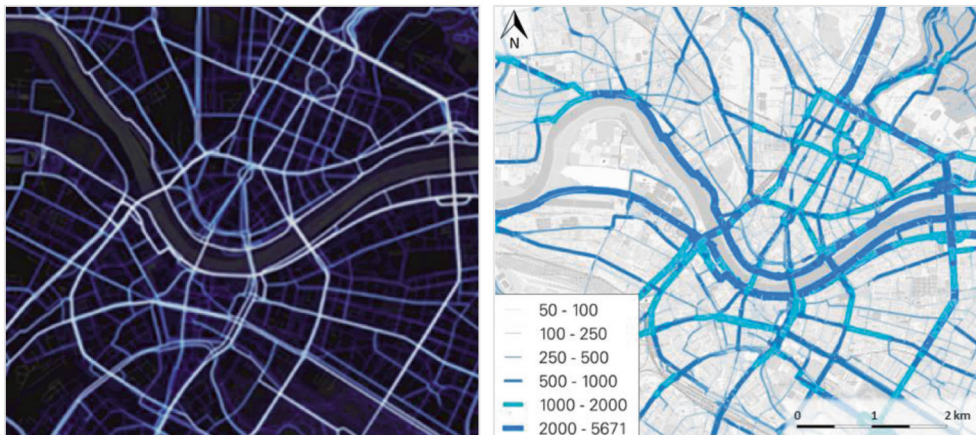

Figure 1: a) Heat map [29] and b) traffic volumes in the city of Dresden.

good correlation (0.87) and regression coefficient (0.754). This indicates a certain reliability and shows that Strava data offer high potential and could be used to illustrate spatial distribution of cyclists and traffic volumes on particular links.

However, using the data also brings serious limitations. First of all, the purchased data need to be visualized by the user since there is no visualization at all. Furthermore, users need to check and adjust the data because they contain duplicates and data from bicycle race events, which planners should not consider as it is preferably commute trips that should be used for transport planning. Specific knowledge is needed for plausibility checks and data visualization, which could be a challenging task for traffic planners in municipalities. Another limitation is the map matching executed by Strava as well as the map delivered to Strava for map matching. Some matched trips are hardly explainable and the map matching results sometimes seems to be counterintuitive.

Surface quality: As reported in Section 4.1, comfort and safety are major issues for planners and they associate it with the surface of bicycle infrastructure. However, the Strava data do not reveal any information about the surface quality as it is only data from the GPS sensor of smartphones that are collected. Nevertheless, there are by far more sensors integrated in smartphones that could deliver data to assess surface quality (e.g. data from acceleration sensor). However, these data are not processed so that the data provided by Strava do not allow any statement in that regard.

Safety: Safety issues, especially points of danger and conflicting use, are of major interest to planners. Strava data do not provide the information required by planners, which can be traced back to the functionality of the app and the motivation of app users. Even if single events (e.g. risky situations which cause different driving modes) are recorded by GPS sensors, they do not reveal any information about these single events. Accuracy and frequency of GPS point measurement is limited. Similar to the issue mentioned in the section above, other smartphone sensors could provide data to assess risky situations and differences in driving modes (e.g. data from gyroscope or acceleration sensor). However, these data are not collected.

Dedication: The use of infrastructure that is not dedicated to bicycle use is an important subject to planners. Although there is no direct information revealed by the data, differences can be detected by comparing the heat map (routes actually driven) and the map illustrating transport volumes (trips matched on links). Limitations mainly correspond to the limitations mentioned in the sub-section 'traffic flow and traffic volume'. Another limitation is the fact that planners are, up to now, not able to analyse potential ways where cyclists would like to ride but cannot do.

Speed and waiting times: Realized speed on different links in the network and waiting times at intersection can help to assess the quality of a bicycle transport system. Although Strava data do not provide speed on links, average travel times on links (provided by Strava) and link length (provided by the GIS network) can be used to calculate the average speed on particular links. Waiting times at intersection are processed and provided by Strava, directly (50% and 75% percentiles). At first glance, the provided data (waiting times) as well as the calculated data (average speed) may be used to assess a bicycle transport system. Results for the city of Dresden are illustrated in Figs 2a and 2b.

As described in Section 3, we cross-checked the data provided by Strava and compared speed values on links and waiting times at intersections with empirical values (see Section 3).

The expected speed differences between Strava app users (homogeneous sample of young and male people) and empirical values (heterogeneous sample of different cyclists) have
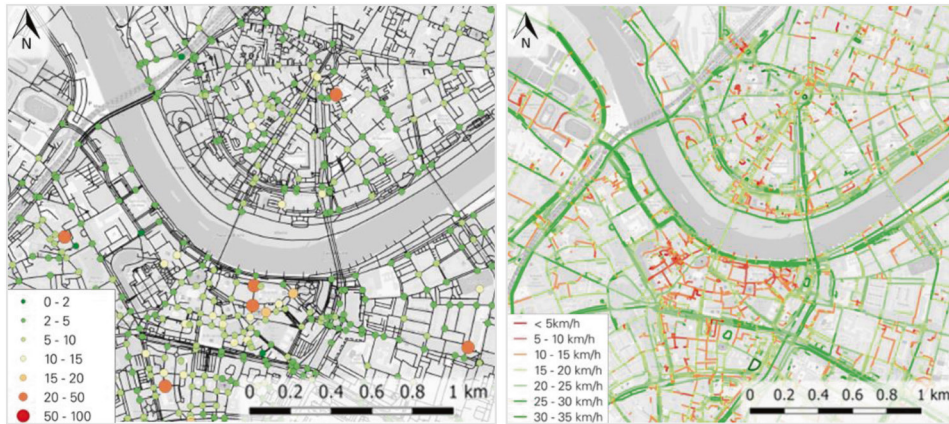
Figure 2: (a) Illustration of waiting times and (b) cyclists' speed on specific links in Dresden.

been affirmed. Average speeds differ significantly. The calculated speed based on the Strava sample shows a high overall mean (25.6 km/h), whereas the mean speed determined via measurements is obviously lower (19.75 km/h). Figure 3a illustrates the difference, which generally applies for all rides in the Strava sample (see Section 3.3). However, the distribution of speeds (shape of the graph) is similar. The overall average deviation, which is about 5.5 km/h, reveals the limitation of representativeness of the Strava sample.

Considering waiting times at intersections, Strava directly provides data. However, the empirical values of the comparison measurements indicate that Strava data do not represent actual waiting times. We found that the data actually represent crossing times. Strava obviously creates a buffer for each intersection and computes average waiting times on the basis of all GPS points within this buffer. This creates a high discrepancy in waiting times computed by Strava and the times investigated through measurements. Figure 3b illustrates an
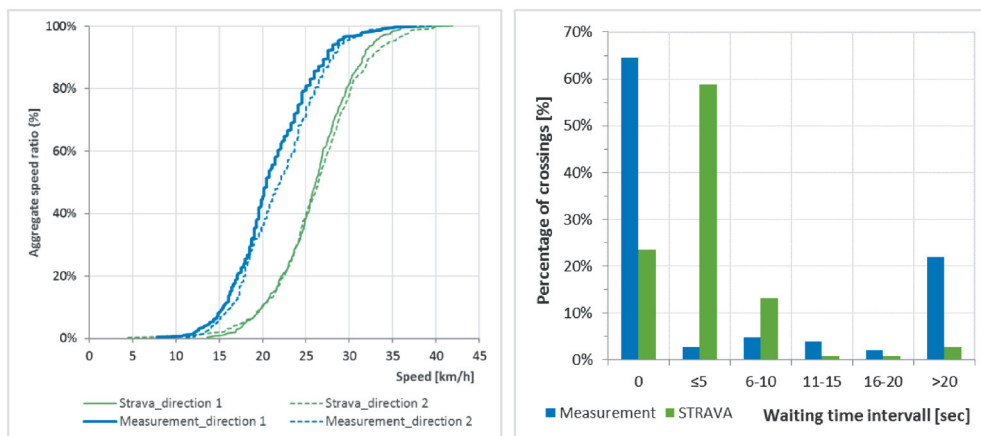

Figure 3: (a) Speed distribution of the Strava sample and measurements for the location 'Elberadweg' and (b) waiting times at the intersection Lennéplatz.

example. Compared to the field data, the number of passages without interference is underestimated by the Strava data and waiting times in the Strava data are much higher than waiting times measured empirically. Another limitation of the data is that waiting times are always matched to a single point, which reduces the actual complexity of intersections and, thus, limits explanatory power.

Use of infrastructure: The type of bicycle infrastructure actually used by cyclist is significant to planners in order to assess infrastructure measures. Differences in the utilization of separated bike lanes, roadways etc. are, therefore, an important information. However, the data provided by Strava cannot shed light on this utilization. The obvious reason is the accuracy of GPS system (maximum accuracy of ±10 m) and its systematic errors (e.g. atmospheric and multipath effects) [31]. GPS data are not suitable to detect which infrastructure is used by cyclists since they are not detailed enough considering the spatial scale.

Origin destination relations: Data considering originating traffic, terminating traffic and the relation between origins and destinations are significant for planning. These data provide information on spatial distribution of cyclists and trip directions (where trips start and end). Strava provides OD relations containing originating and terminating trips for each zone (polygons) and trips occurring on OD relations. Trips are not equally distributed across the city. Due to the fact that there are no OD data to check the data provided by Strava, we tested for correlation between originating trips in zones and its population (see Section 3) because population generally serves as a good proxy for originating traffic [32]. As expected, there is a correlation between trip origins and zonal population but correlation is not satisfactory ($r = 0.528$ and revised $r^2 = 0.268$). This correlation gives a first insight and serves as a simplified cross-check of the OD data. Quality as well as explanatory power of the method, obviously, needs to be discussed (see Section 5).

Detour factor: The detour factor is often used as a practical indicator to assess network density and it also provides information about network quality to a certain extent ('are cyclist choosing the most direct route?'). A low detour factor can indicate a dense route network. Although origins and destinations of trips are known, Strava does not provide a detour factor. Considering the provided data, it is, furthermore, not possible to calculate a simple detour factor because trip length of single trips and distance between origin and destination of the single trip are necessary for that calculation. Due to data privacy, no single trips are reported – nor are precise origins and destinations of trips. Thus, detour factor calculation remains infeasible.

Daytime variation and trip purpose: In order to determine the limits of bicycle infrastructure capacity or to further promote bicycle traffic, information regarding daytime variation of traffic and trip purpose is needed (see Section 4.1). Although Strava data do not provide any information about daytime variation, the aimed variable can be calculated using the data. The data contain timestamps of trips on the links of the network so that hourly traffic volumes can be computed. We calculated the number of cyclists on different links and furthermore processed the daily traffic variation. The comparison of calculated values with traffic count data (see Section 3) reveals that Strava data could be used to generate daytime traffic variation within the bicycle network. Figures 4a and 4b illustrate that, in general, the Strava graphs of daytime traffic variation show good fit to the graphs determined using empirical values from counting stations.

A serious limitation of the data is that the diurnal variation of trips can be derived on annual basis, only, because the data sample provided by Strava are too small. Thus, a seasonal variation of daytime trip variation cannot be identified with adequate level of reliability.
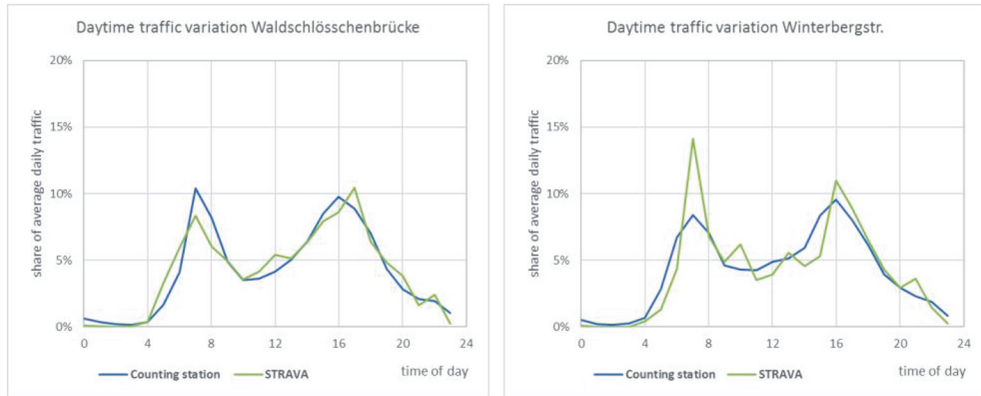
Figure 4: Daytime variation at location (a) and (b).

Due to the fact that app users record their route but trip purposes are not requested and recorded when cycling, reliable data regarding trip purpose cannot be determined from the Strava data. Nevertheless, Strava derives the trip purpose 'commute' from the raw data assuming that trips are 'commutes' when (a) the distance between origin O and destination D (O ≠ D) is longer than 1 km and (b) the OD relation is used regularly (no threshold reported).

As presented in this chapter, we can state that the data provided by Strava generally contain many information. It can be used directly to visualize and illustrate different issues (e.g. transport volume on links or waiting time at intersections) but to a certain extent and with certain constraints, only. Further information can be generated by using the data and further processing it (e.g. average speed on links or daytime variation of traffic demand). Nevertheless, there are a lot of issues that cannot be addressed using the Strava data because there are no information directly provided nor the aimed information can be computed using the data (e.g. use of infrastructure, detour factor or safety issues).

5 DISCUSSION AND CONCLUSION

The results presented in the previous chapter reveal the high potential but also the limitation of using the provided data. In order to evaluate the study results, we discuss the methods and results of the survey and the FGI in the following section. Subsequently, we debate the results of the data analysis and its shortcomings. We further point out the prospects of the used data and review the overall results of the study in the context of the current state of research.

At first glance, the return rate of the survey ($n = 61$) does not appear very high. However, the cities that participated in the survey nearly cover 25% of the German population. The data also show a good distribution over different city size (see Section 3) so that we assess the surveyed data representative for Germany.

Considering the FGI, we can state that it turned out to be a suitable method to get more in-depth knowledge – also because participants referred to different cities and city sizes that correspond to the survey, as well.

The results of the survey and the FGI partly confirm the presumptions we had before (see Section 4.1). They answer the raised research questions (see Section 1) and illustrate the currently available data as well as the data used by planners. Furthermore, the methods helped to investigate which data would be promising for planners and what questions they would try

to answer using it. Therefore, the results generated a proper basis for the analysis of the data provided by Strava. For the first time – to the best of our knowledge – requirements (data needs) of planners and the issues they would like to address using the data are included in such an investigation. This consideration allows a thorough evaluation of the benefit of the provided data, which is one major contribution of the article to the state of research.

We also pointed out major challenges for planners that would use the data. Nevertheless, further investigation is essential because answering this question is difficult when traffic planners have little experience in using such data.

The data analysis allowed assessing if commercially available data could fill the gap between information supply and demand. Section 4 clearly shows that data generally are useful to illustrate different aspects of bicycle traffic within the study area. Further computation of the data also allows deriving additional information. However, it also shows that there are certain limitations contrasting the potential of the provided data (see Section 4). Answering research question (c) is, therefore, not easy. A more general and distinct answer to this question is not feasible.

The data can help to address some issues, but they can fill the gap within certain limits, only. The analysis shows that the data could be used to illustrate traffic flows and traffic volume, speed on links and waiting times at intersections, originating and terminating traffic as well as spatial interaction (OD relation) and daytime variation of traffic. However, there are certain factors limiting the explanatory power. Quality (characteristics of app users) and the quantity (number of app users) of the Strava sample, for instance, are crucial because both directly influence the information provided (e.g. average speeds). Furthermore, the methods Strava used to compute further information are crucial (e.g. calculation of waiting times). There are also issues of planners that cannot be addressed, at all. There is, for example, no information regarding the surface quality, safety issues and utilization of different types of infrastructure or detour of cycling trips.

Nevertheless, the sections above also illustrate that there are many opportunities. One of the main prospects is the data itself. An enhanced data sample could enable the computation of further information to address issues that cannot be addressed by planners, yet. Adding supplementary data (see Section 4) is, therefore, a major challenge. Moreover, enhancing quality and quantity of the data sample is an important issue. Increasing the quantity of the sample as well as enforcing a more heterogeneous sample could result in more detailed and representative information.

Another major prospect is the improvement of data processing methods. First, considerable improvements should aim for map matching, trip segmentation and trip identification as well as trip purpose imputation. Map matching methods need to be more transparent and have to be improved to avoid multiple trip matching results that we found. A trip identification and segmentation model could ensure that the data contain actual bicycle trips (e.g. commutes), only.

Major improvements are also attainable considering the illustration of GPS trajectories within the heat map. The simple illustration of GPS points currently leads to the illusion of plenty of traffic where GPS point density is high. However, this is also the case where slope is high and cyclists ride on low speed. The prospect here is a (heat) map that illustrates processed GPS trajectories (e.g. through normalization).

Referring to dedication, data providers could offer further computation using the raw data. Providers could select, for instance, OD pairs that show high detour factors and compute the shortest paths between origins and destinations. Besides the ways cyclists ride (see Section 4), this method could illustrate where cyclists could ride but, actually, do not.

Similar data processing could provide information regarding detours of cyclists. Providers could use the individual GPS trajectories to calculate the ratio between trip lengths of actual trips and shortest route between origin and destination. They could then aggregate the results for each OD relation resulting in OD specific detour factors.

An important prospect aims at the calculation of waiting times at intersections. More detailed computation methods are needed to calculate realistic waiting times. A spatial resolution needs to be adapted, which goes beyond the simplified buffering and aggregation at intersections. Furthermore, waiting times should contain waiting times instead of crossing times (see Section 4.2).

Section 2 shows that there are several studies, which analysed similar data for different study areas. We found that there are similarities as well as differences between existing studies and our contribution.

Similar to the study of Ref. [14], the used data enable to illustrate spatial and temporal movement patterns of cyclists. Like Ref. [22], we state that the data are useful to observe changes in spatial and temporal distribution of cyclists across a city. The results of our study are also in line with the results of Refs [33 and 15] as we found correlation between Strava data, official counting results and the conducted measurements. Thus, we agree with Ref. [34] because our study also shows that crowdsourced data generally match the conventional counting data.

With respect to the limitations of the data, using it allows predicting the number of cyclists on links. However, we disagree with Ref. [16], which concludes that the data can be used to predict ridership volumes on different bicycling infrastructure. We argue that this is a more complex pattern, and methods of traffic demand modelling and route choice modelling have to be applied to investigate that subject.

Apart from international studies, our study gives a first insight into data utilization and validation for a German city. Furthermore, it is the first study that precisely contrasts data needs of planners and data potential. The study also points out the prospects of data computation and related utilization and is, therefore, unique in this manner.

Since several research has been done on Strava data, further research should focus on different data samples (e.g. from different providers) to investigate the differences and possible synergies of utilization. Further investigating the data and collecting additional ones, which comply with the requirements, will support traffic planners to address the most important issues. However, increasing their resources to handle the data or simplifying the access and utilization is a major issue in that regard. This could finally help planners to improve planning, increase bicycle mode share and, thus, reduce the negative effects of transport in cities significantly.

ACKNOWLEDGEMENTS

REFERENCES

[1] Kager, R. & Bertolini, L. & Te Brömmelstroet, M., Characterisation of and reflections on the synergy of bicycles and public transport. *Transportation Research Part A: Policy and Practice*, **85**, pp. 208–219, 2016. DOI: 10.1016/j.tra.2016.01.015.

[2] *Bicycle Traffic in Germany – Numbers, Data, Facts*. Federal Ministry of Transport and Digital Infrastructure, available at http://www.ziv-zweirad.de/uploads/media/rad-verkehr-in-zahlen.pdf. 2014, (accessed 23 March 2018).

[3] Birk, M. & Geller, R., Bridging the gaps: how quality and quantity of a connected bikeway network correlates with increasing bicycle use. *Transportation Research Board, 85th Annual Meeting*, Washington, 2006.

[4] Lißner, S., Francke, A., Chernyshova, O. & Becker, T., App-Daten für die Radverkehrsplanung – Eine explorative Datenanalyse von GPS-Daten im Radverkehr. *Internationales Verkehrswesen*, **69(1)**, pp. 48–52, 2017.

[5] Hyde-Wright, A., Graham, B. & Nordback, K., Counting bicyclists with pneumatic tube counters on shared roadways. *ITE Journal (Institute of Transportation Engineers)*, **84(2)**, pp. 32–37, ISSN 0162-8178, 2014.

[6] Jestico, B., Nelson, T. & Winters, M., Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*. **52**, pp. 90–97, 2016. DOI: 10.1016/j.jtrangeo.2016.03.006.

[7] Statista, Anteil der Smartphone-Nutzer in Deutschland in den Jahren 2012 bis 2017, available at https://de.statista.com/statistik/daten/studie/585883/umfrage/anteil-der-smartphone-nutzer-in-deutschland/, 2018 (accessed 23 March 2018).

[8] Thornhill, T., Globally cyclists clocked up 4.5 BILLION miles and Regent's Park is the UK's most popular bike route: GPS training app reveals how much exercise the world did in 2017. *Mail Online*, 2017.

[9] Romanillos, G., Austwick, M.Z., Ettema, D. & De Kruijf, J., Big data and cycling. *Transport Reviews*, **36(1)**, pp. 114–133, 2015. DOI: 10.1371/journal.pone.0074685.

[10] Milne, D. & Watling, D., Big data and understanding change in the context of planning transport systems, *Journal of Transport Geography*, Article in Press, 2018. DOI: 10.1016/j.jtrangeo.2017.11.004.

[11] Clarke, A. & Steele, R., How personal fitness data can be re-used by smart cities. *Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Adelaide, 2011.

[12] Wamsley, K., Optimal power-based cycling pacing strategies for Strava segments. Doctoral dissertation, University of Pennsylvania: Kutztown, 2014.

[13] Albergotti, R., Strava, popular with cyclists and runners, wants to sell its data to urban planners. *Wall Street Journal*, available at https://blogs.wsj.com/digits/2014/05/07/strava-popular-with-cyclists-and-runners-wants-to-sell-its-data-to-urban-planners/, 2014 (accessed 23 March 2018).

[14] Musakwa, W. & Selala, K.M., Mapping cycling patterns and trends using Strava Metro data in the city of Johannesburg, South Africa. *Data in Brief*, **9**, pp. 898–905, 2016. DOI: 10.1016/j.dib.2016.11.002.

[15] Holmgren, J., Aspegren, S. & Dahlström, J., Prediction of bicycle counter data using regression. *The 2nd Edition of the International Workshop on Data Mining on IoT Systems (DaMIS)*, 18–20 September, Lund, 2017.

[16] Heesch, K.C. & Langdon, M., The usefulness of GPS bicycle tracking data for evaluating the impact of infrastructure change on cycling behaviour. *Health Promotion Journal of Australia: Official Journal of Australian Association of Health Promotion Professionals*, **27 (3)**, pp. 222–229 2016. DOI: 10.1071/HE16032.

[17] Boss, D., Nelson, T., Winters, M. & Ferster, C.J., Using crowdsourced data to monitor change in spatial patterns of bicycle ridership. *Journal of Transport Health.* Article in Press, **9**, pp. 226–233, 2018. DOI: 10.1016/j.jth.2018.02.008.

[18] Conrow, L., Wentz, E., Nelson, T. & Pettit, C., Comparing spatial patterns of crowdsourced and conventional bicycling datasets. *Applied Geography*, **92**, 21–30, 2018. DOI: 10.1016/j.apgeog.2018.01.009.

[19] Richardson, A.J., Ampt, E.S. & Meyburg, A.H., *Survey Methods for Transport Planning*, available at http://www.no2hcf.co.uk/docs/Traffic_survey_form.pdf, 2006 (accessed 23 March 2018).

[20] Fleming, C.M. & Bowden, M., Web-based surveys as an alternative to traditional mail methods. *Journal of Environmental Management*, **90**, 284–292, 2009. DOI: 10.1016/j.jenvman.2007.09.011.

[21] Wright, K.B., Researching internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, **10(3)**, JCMC1034, 2005. DOI: 10.1111/j.1083-6101.2005.tb00259.x.

[22] Stage, F.K. & Manning, K., *Research in the College Context: Approaches and Methods*. Routledge, Abingdon-on-Thames, United Kingdom, 2015.

[23] Ledermann, L.C., Assessing educational effectiveness: the focus group interview as a technique for data collection. *Communication Education,* **39**, pp. 117–127, 2009. DOI: 10.1080/03634529009378794.

[24] Ausserer, K., Kaufmann, C. & Risser, R., The Focus Group Interview – A qualitative method to assess quality aspects of the traffic systems. *Proceeding of the 20th ICTCT Workshop*, Valencia, Spain, 2014.

[25] Williamson. K., Questionnaires, individual interviews and focus group interviews. *Research Methods (Second Edition) – Information, Systems, and Contexts*, Tilde Publishing, Prahran, Australia, 2013.

[26] USDOT, FHWA/FTA, *Public Involvement Techniques for Transportation Decision-Making*. Chapter 3, Section B, Designing programs to bring out community viewpoints and resolve differences, U.S.DOT, Washington DC, 2002.

[27] Ahrens, G.-A., *Sonderauswertung 'Mobilität in Städten SrV 2013'*, Technische Universität Dresden: Dresden, 2015.

[28] Francke, A., Lißner, S. & Becker, T., How can bicycle planning benefit from smartphone based data? A validation of Strava data. *International Cycling Conference*, 19–21 September 2017, Mannheim, Germany, 2017.

[29] Strava Inc., Global Heatmap. *Strava*, available at https://www.strava.com/heatmap#7.00/-120.90000/38.36000/blue/ride (accessed 23 March 2018).

[30] Bierlaire, M., Chen, J. & Newman, J., A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C*, **26**, pp. 78–98, 2013. DOI: 10.1016/j.trc.2012.08.001.

[31] Mansfeld, W., *Satellitenortung und Navigation – Grundlagen, Wirkungsweise und Anwendung globaler Satellitennavigationssysteme*, *Vieweg&Teubner | Springer, Wiesbaden*, 2010.

[32] Ortúzar, J.de D. & Willumsen, L.G., *Modelling Transport*, 4th ed., Wiley, New Jersey, 2011.

[33] McLafferty, I., Focus group interviews as a data collecting strategy, METHODOLOGICAL ISSUES IN NURSING RESEARCH. *Journal of Advanced Nursing*, **48**(**2**), pp. 187–194, 2004. DOI: 10.1080/13645579.1998.10846874.