Proceedings of 7th Transport Research Arena TRA 2018, April 16-19, 2018, Vienna, Austria

Modeling Cyclists Traffic Volume – Can Bicycle Planning benefit from Smartphone based Data?

Sven Lißner*[a], Angela Francke [b], Thilo Becker [a]

[a]*TU Dresden, Chair of Transportation Ecology, 01062 Dresden, Germany*
[b]*TU Dresden, Chair of Transportation Psycology, 01062 Dresden, Germany*

Abstract

Good transportation planning requires reliable data. Nowadays many smartphone users record their routes and submit these GPS tracks to servers of smartphone application operators. These aggregate tracks are the base for a bunch of tools to close the gap in bicycle planning and evaluation. However, there is only few information about the app users. Therefore the question is if it is possible to derive predictions from the app data that are valid compared to field data. An analysis of field data collections in Dresden with a dataset collected by the smartphone app Strava with an overall of 3,200 cyclists and 70,500 rides was undertaken. The comparison focused on traffic volumes, speed and origin-destination matrices. Overall, the predicted values based on the Strava app sample were comparable to the permanent counting devices, especially in areas with higher traffic flow. Strava app data is with some limitations applicable for bicycle planning. Recommendations for the future use of Strava and similar data sources for bicycle planning and transportation research will be discussed.

Keywords: bicycle infrastructure planning; smartphone application; GPS tracks; big data

* Corresponding author. Tel.: +49 351 463 36929; fax: +49 351 463 37718
E-mail address: sven.lissner@tu-dresden.de

1. Introduction

1.1. A state of the art review of bicycle data

Data is the basis of all planning processes. For motorized private transport valuable data from different sources is already available. However, the classical data in bicycle traffic planning is based on the complex and active data collection by counting devices, traffic observations and traffic surveys. At the moment the best data is generated by permanent counting devices. However, a survey among German cycling coordinators (n = 61) which was conducted within this research revealed only low sensor availability, *M* = 3.05 devices per city, *SD* = 5.42, and 51% of cities without any devices.

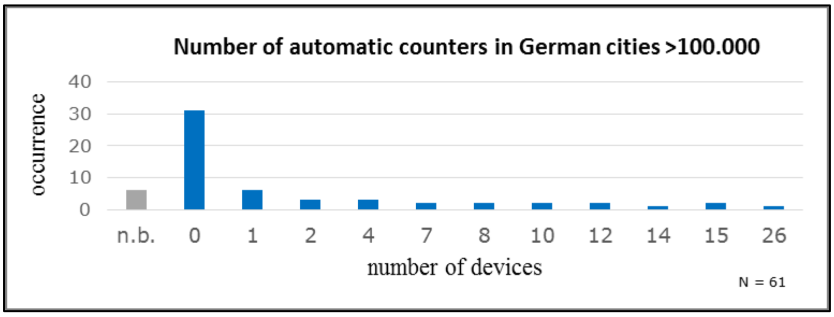

Figure 1: Number of automatic counting devices for cyclists in German cities, based on a survey amongst German cycling coordinators, n = 61

Therefore, these automatic counting devices can only deliver a snapshot picture of the actual traffic volumes and movement pattern in some cities. The analysis of cyclists numbers can be backed by manual counting approaches but they only give insight in a rather short period of time and are more expensive. The subnet and other parameters than the traffic volume, such as waiting time, route selection and speed, are not included in the data and, therefore, missing as information for bicycle planning.

1.2. Literature review and research question

The increasing use of devices that are GPS enabled and the growing amount of tracking apps make it possible to look at patterns of movement of road users in an aggregated manner and to derive a demand-oriented bicycle traffic planning. The number of users of smartphones increased in the last years. In 2016, 49 million people in Germany owned a smartphone (82 million inhabitants, Statista, 2017). A large number of these smartphones are equipped with a GPS sensor. In the last years, several smartphone applications for cycling have entered the market. They offer the possibility to monitor their mobility behavior or at least their rides by tracking their training routes, movement patterns and velocity via GPS and saving it. Users get an overview of their performance and they are able to compare their results with others. The number of users of these apps is also increasing and most of the app operators got also transferred the right of further data usage of the users. They can therefore offer the collected data in an anonymized form at the market. This leads to new exiting data sources for transportation research.
The utilization of GPS data in transportation planning and science is well documented and started in the late 1990s as a proof of concept (Lexington Area Travel Data collection Test, 1996) using data loggers fixed in cars. Newer approaches use wearable GPS loggers or smartphones (Broach, Dill, & Gliebe, 2012; Harvey & Krizek, 2007) but the study characteristic still remained. With the introduction of the iPhone in the year 2007, smartphones became available for a broad group of users. Since this break GPS data in transportation research can be divided in two subgroups: In the first subgroup user data is known due to small self-selected samples, like Krizek (2007) who had a sample size of 50 persons were users and trips are known. The second group contains GPS data provided by cycling apps or app data in general and usually has – due to privacy – a general lack of information about the users. Combined with a census survey it can lead to big representative samples like Schuessler and Axhausen (2008) or Menghini (2009) who had about 4,200 participants out of an even bigger study, which is not usually the case. Working with GPS data from smartphone apps includes the risk of having a strong bias due to the represented user group. The participants are no longer selected but choose to join the study but not on random basis. In most cases there is a bias towards middle aged male person with a slightly higher educational standard and income (Broach et al., 2012; Charlton, Sall, Schwartz, & Hood, 2011). In other cases it is towards a massive sportive background (Jestico, Nelson, & Winters, 2016). To analyze future datasets without knowing the raw data it has to be evaluated

how aggregated data can be interpreted. Therefore, the presented study fills the gap between available data from other sources and scientific approaches using aggregated big data.
The first difference occurs in the mixture of users: compared to punctual and permanent counting devices or representative surveys smartphone-based data collection is not likely to record all relevant user groups in cycling. It can be assumed that in the medium term, only a part of the cyclists' population is participating in the data collection process via smartphone app and this part can have a different behavior. Therefore, a projection of the actual traffic load from the smartphone-based sample is necessary for validation. Established projection methods, projecting for a whole working day or the average daily traffic volume from punctual traffic counts (e.g. 2 or 12h) such as HBS (2015) or Schiller (2011) cannot serve this purpose as the data structure is different. Smartphone-based data is recorded for longer periods (e.g. one year) but with a smaller sample size compared to the population. An alternative approach for projecting the traffic volume is therefore necessary. Strauss et al (2015) combined for example smartphone-based GPS and count data to estimate the Average Annual Daily Bicycle volume AADB (El Esaway, 2013) at the network of road segments and intersections in Montreal. They observed correlations between AADB predicted via GPS data and AADB from count data with R-squared values of 0.7 for signalized intersections, 0.58 at non-signalized intersections and between 0.48 and 0.76 for segments with and without bicycle infrastructure. Jestico et al (2016), carried out a similar approach, using only AM and PM peak hours to compare count data of 18 manual counting stations and 34 days with an R^2 of 0.58 over all peak periods which is quite weak considering the big time window. This leads to the question if the used sample is capable of modelling the population of cyclists in a specific city.
This article refers to this question: Is it possible to derive predictions from biased app data that are valid compared to field data? Within the research project the anonymized, purchasable GPS data of a (sport-oriented) app for smartphones by Strava Inc. is the basis of the validation process. General findings on the parameters traffic volume, origin-destination-matrices and trip length will be presented. Afterwards the results of an analysis of representativeness are discussed. The data is compared with empirical traffic data (counting devices and measurements) on the parameter traffic volume and with the number of inhabitants on the parameter origin-destination-matrices in the municipality of Dresden, Germany. Also, recommendations for the future use of Strava and similar data sources for bicycle planning and transportation research will be covered in the discussion.

2. Method

2.1. Description of samples and data collection

2.1.1. Study Area

The study area is the city of Dresden, Germany. Dresden has a population of about 553,000 inhabitants with an equal share of genders. The topography ranks as flat but includes some ascents. The ratio of cycling in mode split is 12%, and 16% of the inhabitants use the bicycle for every day transportation resulting in an average of 223,000 bicycle trips per workday (Ahrens, 2015).

2.1.2. Details of the Strava data

The used data for this study from Strava contains two data sets. The first contains 19,615 cyclists and 439,570 rides in Saxony, Brandenburg and Berlin. The second data set consists of 70,500 rides by 3,200 cyclists in Dresden. Men dominate the user sample (see Figure 2). This has mainly two reasons: Firstly, the average proportion of cyclists is not equally distributed in gender (Heesch, Sahlqvist, & Garrard, 2012). For the city of Dresden manual counting shows a distribution of about 60% male cyclists and only 40% female riders. Secondly, the fact of men being more affiliate to technic has to be considered. These two systematic biases are backed by a specific error when it comes to Strava data. Strava as a sports app is a highly competitive framework – this is also a fact, which relates much more to men (Gray, McHale, & Carré, 2016). The mentioned biases add up to a gender derivation of 80% male riders and only 20% female riders in the sample which is equal to the derivation in comparable studies (Jestico, Nelson, & Winters, 2016).

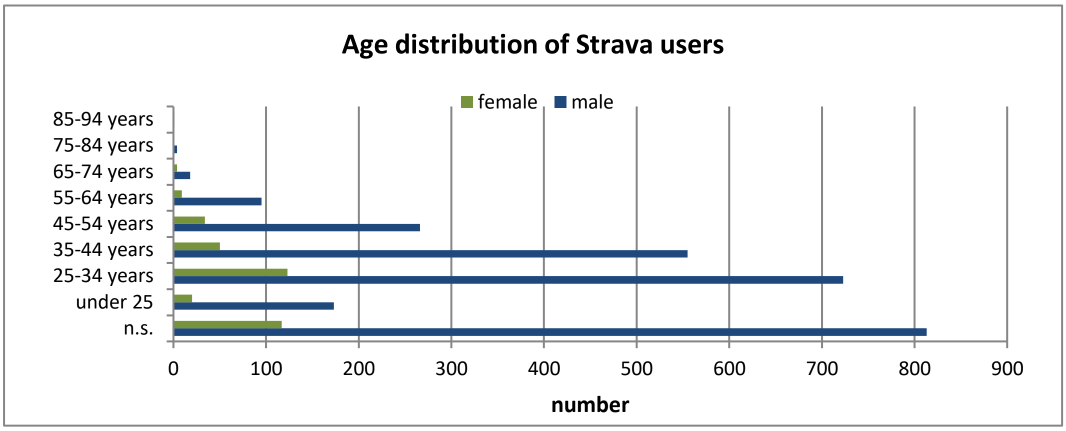

Figure 2: Distribution of age and gender of Strava users in the city of Dresden

Strava data is delivered in two different ways. There are pre-defined roll-ups for given timespans (year, time of day, etc.) where data is aggregated and matched to edges of a GIS network (OSM) and the core data which is minute-fine. These minute-fine data is also projected on road segments or nodes, so that routes, which can be assigned to users are not available. As a sports-oriented app, Strava offers the possibility to filter the data set for planners. The minute-fine data can be parted in commute trips and all trips. For commute trips, the data provider tries to eliminate all sportive rides by an algorithm, which flags the commute rides in two different ways: At first the starting point and the end of trip has to be at least 1km apart. Because daily rides, like the trip to work, are bi-directional it is supposed that users stop tracking after reaching the work place. Thus necessarily short trips less than 1km are defined as no commute trip, which leads to problems modelling the last mile problem. The second filter is the number of trips on a given relation. If a trip is realized several times with a certain start and ending point, the trip must be a part of daily mobility – at least that is the assumption. The third possibility is to flag the ride as commute which is done by the users themselves. This is not used by Strava due to doubts in reliability. Strava. Inc. states that the used algorithm detects at least 95% of the rides which are flagged as a commute. Other filters should be checked, too. In estimating the length of commute trips by counting the odd polygons passed, there is a significant amount of trips which show a higher trip length than usually found in commuting. Nevertheless, commute numbers are used for the upcoming validation and prediction of number of cyclists.

2.1.3. Field data

Two different kinds of classic counting data were compared to the GPS based data: data from six permanent counting devices in Dresden between 01/09/2015 and 31/05/2016 (400 to 2,000 cyclists per day, hourly volumes) and data from temporary manual counts in Dresden from 22/05/2015 to 12/06/2015, in the time period from 2pm to 6pm, Tuesday till Thursday (7,309 cyclists, hourly volumes). The temporary manual counting sights were selected with regard to the importance of traffic, parallel routes to the main roads and suspicion of high frequency of sport cyclists. The counting has been achieved with video technique. The comparison to Strava data was done for corresponding time-periods. The corresponding Strava cases for the time periods of the temporary manual counting were below 1%, as such too small for comparisons. Therefore, the entire Strava data set for the corresponding time frame is used as a reference. The compared data sets contain hourly values for the counting sites and average hourly values for Strava. For the comparison of origin-destination data and inhabitants of the city census data from 2011 was used. The European statistic network has been chosen as the basis polygon network. The number of inhabitants and starting rides could be linked to each cell.

2.2. Verification approach

Based on multivariate analysis the study tried to approximate a representative cyclist for the city of Dresden using empiric field data for speed, gender, trip length and allocation of trip end and start. The average Dresden cyclist, 60% male, riding with a speed of 19.75km/h, a trip length of 4km (Ahrens, 2015) is compared with AADB Strava data. Coherence is searched univariate for the parameters traffic volume, distribution of trip origins and trip length using data from both datasets at the same network links trying to see differences in cyclists' behavior (see Figure 3).

In addition to field measurements, which are described in the data section, a survey amongst Strava users (n = 182) shows a high percentage of users recording all bicycle trips with small influences on behavior on commute trips. Strava users cycle a bit more than before they activated the app (36%), but they mostly use the same routes. The main reason of Strava users to use the app is to have an overview of their overall trips (85%). The majority of them uses the app for each trip they take (Francke, Lißner, & Becker, 2017).

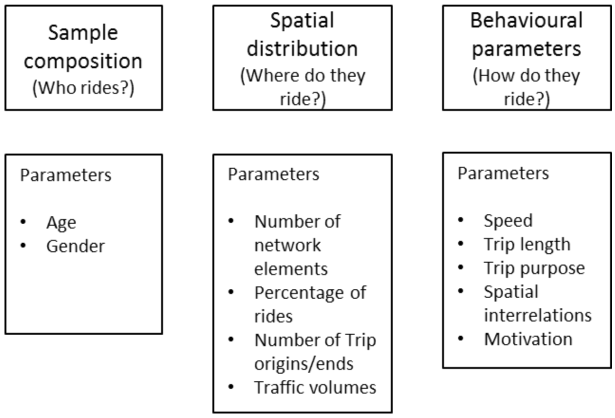

Figure 3: Overview of the methods and research questions

2.3. Data comparison methods

Traffic Volumes: Traffic volumes state the amount of cyclists per time unit. As in smartphone-based surveys only a part of cyclists participates in data collection the data needs to be projected to the total number of cyclists to draw conclusions for a certain area. The projection of the traffic volume is realized with a linear regression model. OSM is used as the road network. At first, a time and date variable is coded, then the cyclists of multiple edges are combined and duplicates are deleted with the criterion “more than one case in the same minute on two parallel edges”. This is necessary due to the fact that data in map matching was matched to several parallel edges whereas in reality the cyclists mostly took the cycle path at the counting site. Afterwards the data is aggregated to hours and value zero is added for hours without any cyclists recorded. Afterwards the hourly average Strava data volume are computed for the time of day when parallel empiric studies took place. In the mentioned empiric studies hourly volumes are counted for a distinct timespan (manual counting sites) or the whole year (automatic counters). For the automatic counting devices it has to be mentioned that the exact position of the examined site has to be taken into account. If there is a strong possibility of passing them without being count, it has to be thought of dropping this site from the built up regression or factoring model because of unrealistic numbers of cyclists compared to the GSSN data which is counted anyways.

Trip length: Whilst missing route data, trip length of commute data is estimated by using the number of passed polygons. In a first approach the polygon number was multiplied with the length of a side. This very basic estimation ignores the fact that some polygons are only touched a little by the original route. Therefore, some examples are computed by using the OSM navigator to reconstruct a possible route. Dividing estimated trip length $l_{e,i}$ by shortest route length $l_{r,i}$ an average factor a_i of 1.4 was derived.

$$a_i = \frac{l_{e,i}}{l_{r,i}} \tag{1}$$

Afterwards, the found data for commutes is compared with stated preference data from a survey (SrV 2013).

Origin-destination matrices: The data for the origin-destination matrices (OD-matrices) focuses on start point and destination of single user trips and includes an array of passed polygons between them. Because of data privacy only aggregated data of the users can be ordered from the providers of app data. Therefore on network level, e.g. for edges or nodes, only traffic volumes are available. Route choice data is not included in the original data set. Route choice can only be estimated by using OD-matrices. In the mentioned minute-fine data origins and destinations can be found together with the passed polygons for a specific trip. The OD data is aggregated in blocks based on the European statistic polygon network by the provider for privacy protection reasons. Polygons were built as regular squares with an edge length of 1,000m. The Strava data from the corresponding area was then projected onto this network. This procedure meets the requirements of privacy protection on the one hand and enables a good national comparison of the results on the other hand.

3. Results

3.1. Descriptive Statistics

An explorative data analysis looking at traffic volumes and origin-destination matrices is done for the area of Dresden. The general network coverage by the app users is evidently high – at least for the city of Dresden. Analysis shows a coverage of at least 90% of the GIS-Network with at least one ride in the timespan of one year. This number decreases to 68% if only network edges with at least ten rides are considered (see Tab. 1). The GIS network of the city is used for this analysis as the OSM network is too detailed to deliver meaningful results. The used network is less detailed in terms of parallel links and walkways or parking spaces.

Tab. 1: Distribution of rides on GIS-links

min. number of rides	total amount of GIS edges	percentage of GIS edges
0	1,703	11%
1	3,453	21%
10	5,180	32%
100	5,033	31%
1,000	794	5%

Figure 4 illustrates the coverage of the network with violet lines being closely distributed around the whole area of Dresden. 3,200 users and 70,000 rides were taken into account for the timespan of one year. About 50% of these trips were flagged as commute trips by the data provider.

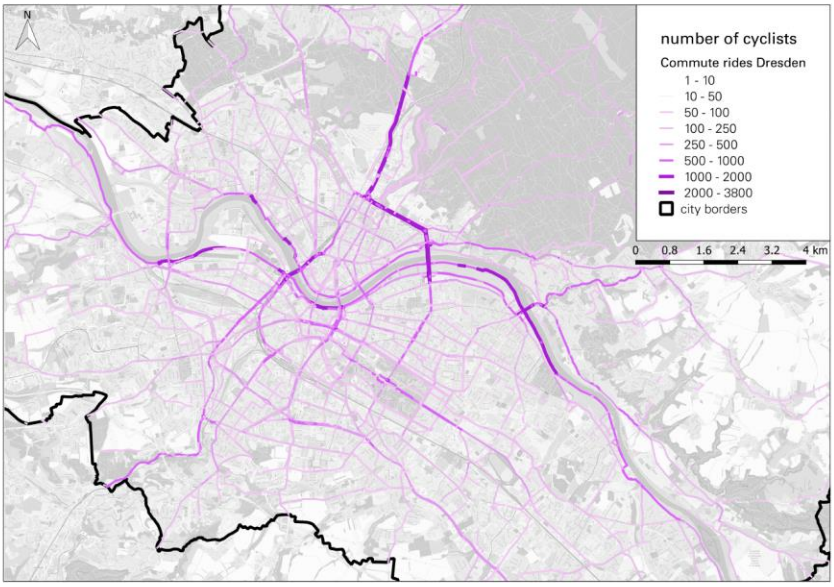

Figure 4: Cyclists volumes, city of Dresden, commute rides, 2016, based on Strava app data

The starts of all trips within the city limits of Dresden marked by Strava as 'commute' are shown in Figure 5. In order to capture incoming commuters in terms of location, traffic within the origin-destination matrix is examined which started or ended in the city of Dresden. Clearly visible is the strong focus on the city center, located in the middle of the map, with a high density of residential land use. North of the center emerges the city district 'Outer new city'. The predominantly young population there suggested a higher proportion of cyclists but also a higher proportion of app users. This thesis can be confirmed on the basis of originating traffic. Other strong sources in the north of Dresden, respectively in the north-central area (shown in purple) are congruent with the location of large employers.

In the city center and its' south there are other significant sources of bicycle traffic in (campus of the Technical University of Dresden and the nearby student accommodation). Equally striking is also the special contribution

made by the city forest in the north-east. Major sources of traffic generally tend to be more in the city center than in the urban environs. This seems to be reasonable due to the higher population within the central city district as well as the concentration of retail infrastructure and jobs.
Originating traffic cells for commuters are mainly found in the towns just outside Dresden. However, the absolute figures recorded for the commuters should generally be interpreted with caution, since figures beneath 200 trips could easily have been taken by one very active person on their own. The destination matrices of commuters show no significantly different distribution, so a bi-directional mobility pattern can be assumed. Origin-destination relationships enable, on the one hand, the identification of certain, possibly preferable, usage corridors. Secondly, they provide a way to identify particularly active cyclists as sole sources based on the absolute number of trips. This is not necessarily possible because data on cyclists are not available, and it is theoretically possible that several less active cyclists use the same route often. Compared with the urban population (> 550,000 inhabitants) the small sample of Strava users (approx. 3,000) results in a relatively low likelihood of having several less active users.

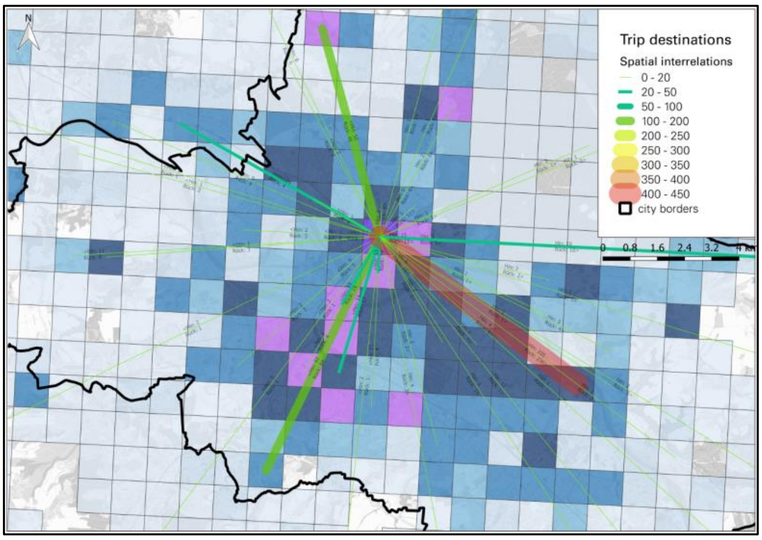

Figure 5: Trip origins and spatial interrelations for 'outer new city'.

Figure 5 shows all recorded origin-destination routes for a selected neighborhood in Dresden. The variety of different destinations and sources suggest a very heterogeneous mix of users and a variety of active users, while, secondly, the influence of one very active user is very clearly identified between the Laubegast part of south-east Dresden (arrow in red) and the outer part of the city's center. On this relation, about 450 trips were realized during the period of observation. This figure cannot be explained statistically by the number of inhabitants compared to other parts of the city. One possible explanation could be a very regular commuter, a regularly scheduled training session or an active gathering of cyclists. This problem area acts as an example for all procedures based on public participation. Very active users can lead to distortions. Looking at different timespans it is necessary is to normalize the user numbers or alternatively limiting the observation period.

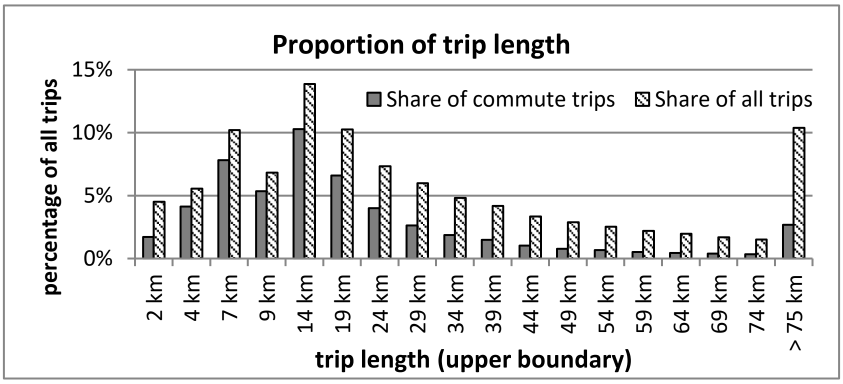

Figure 6: Proportion of trip length compared between trips for commuting and all sorts of trips
(Strava sample for Dresden, 06/2015-06/2016)

3.2. Analysis of representativeness – the length of trips characteristics

Figure 6 shows the distribution of total trip share for commute and all trips. Regarding the average length of commute trips, which is definitely beneath 10km, the Strava trip length shows a higher length in general. A peak can be observed at 14km, which refers to the more experienced Strava cyclists. Regarding average trip length a typical Strava commute is longer than the expected distance to commute. Traffic surveys state an average trip length of 5,8km per trip for Dresden for a workday (Ahrens, 2016). Moreover, the trip length distribution of flagged commutes is not significantly different from the distribution with sport rides. Only the second peak at more than 75km is not there. In this case again Strava data cannot be seen as representative in a scientific way which does not mean that the data cannot be used for planning purposes. Trip length can be an indicator for describing the experience level of a sample when it comes to cycling.

3.3. Validation process for traffic volumes

In order to validate cycling data traffic volumes can be seen as a very good indicator. In this section the projection of Strava data on basis of automatic counters in the city of Dresden is described. The daily traffic flow of the two data sets recorded by the permanent counting devices in Dresden and Strava is depicted in Figure 7 and 8. Differences in the daily traffic flow can be explained with the function of the respective road in the network. The traffic volume per day is between 400 and 2,000 cyclists in the permanent counting device data. The sample size in the Strava dataset is smaller. The sum of cyclists in the nine month period is between 500 and 5,500 cyclists per edge in a GIS network. Therefore, Strava data is depicted as ADB.

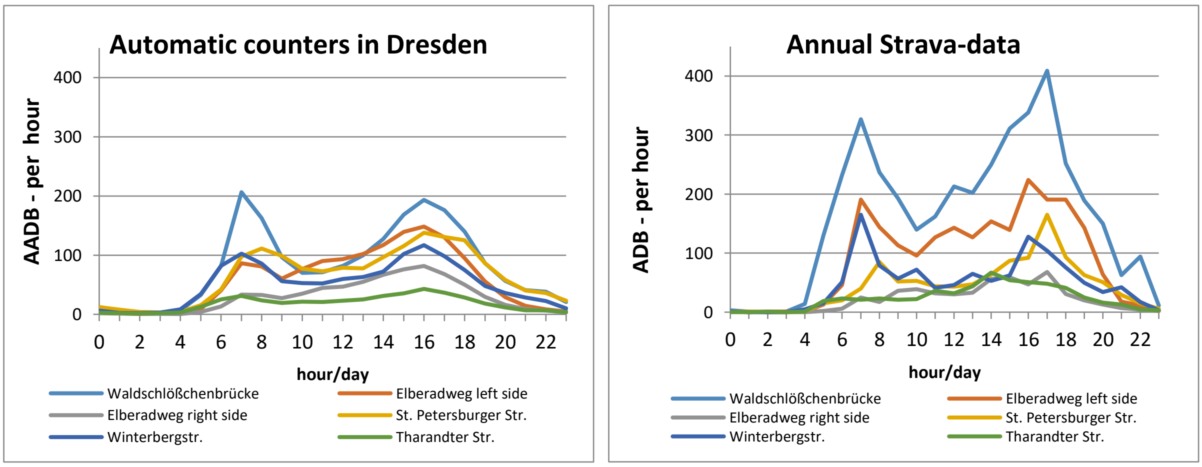

Figure 7: (a) Average traffic volume measured by the permanent counting devices during the period of investigation in Dresden (b) Strava traffic volumes during the period of investigation in Dresden

Strava commute data has a quite similar distribution over the total enquiry period as can be seen in Figure 8. The projection of the Strava data to the counting device data is done using a linear regression for every hour of the day and all counting devices and results in the projection, depicted in Figure 9. We use Strava data as explanatory variable (x) and derive the daily values, $f(x) = 0.5087*x+16.9719$.

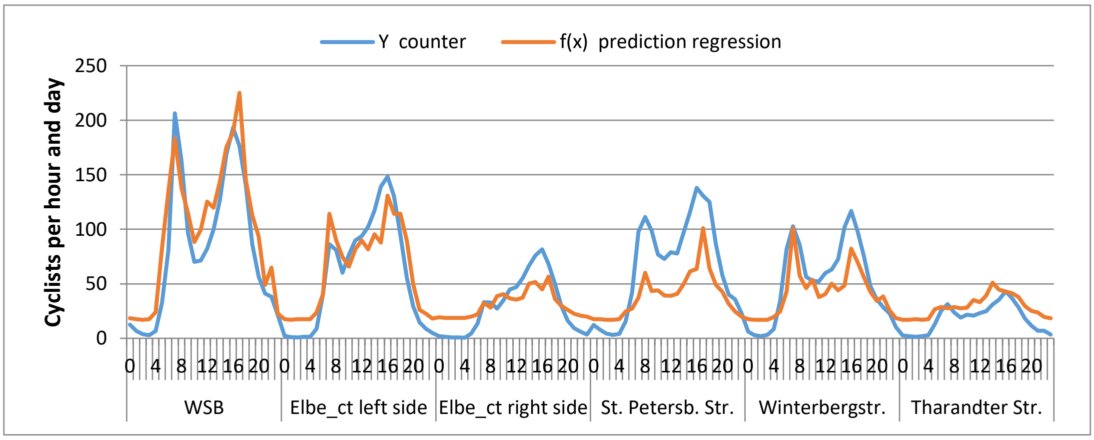

Figure 8: Projection for the permanent counting devices based on a linear regression

While evaluating the method it should be kept in mind that the linear regression model is using a constant value. This means that also roads without recorded Strava users have a minimal traffic volume of 17 cyclists per hour, which can be seen during night hours in Figure 8. This will not always hold in reality. Statistical parameters for forecasting reliability are very positive even without further explanatory variables, see Table 2.

Tab. 2: Statistical values assessing the quality of the prediction models

Statistical Value	Regression
Mean absolute error (MAE)	18.434
Relative error	0.364
Correlation	0.866
Coefficient of determination (r^2)	0.754

The high value for $R^2 = 0.75$ over all counting sites on daily basis shows that results in recent studies do not necessarily rely on Strava data being not feasible to predict model data. They are strongly influenced by the compared timespans or location in the road network (Jestico, Nelson, & Winters, 2016). It is highly recommended to use automatic counting sites for calibrating instead of short or medium term manual counting.

3.4. Comparison of origin-destination-matrices and population

The origin-destination matrices may also be influenced and distorted by a heterogeneous sample. Therefore, the regional distribution of the originating traffic is validated with the population of the corresponding originating traffic-cells, at first. The used statistical data is based on national census 2011. This approach is of course limited as originating traffic can, for instance, also be related to workplace. Further analysis is, therefore, recommended. With this limitation, the statistical analysis of the originating traffic data is only partly valid but delivers a first overview of the quality of the data. The statistical analysis results in a high positive correlation of both values, number of inhabitants and trip origins at basic level. The data set was analysed in the time period 01/2015 until 06/2016. We conclude that there is a close relation between these values. The results for the cities Dresden, Chemnitz, Berlin and Leipzig (see Table 3) show that this correlation (r) is differently distributed by location and stable at around 0.7. The correlation is however strongly influenced by areas with a low number of inhabitants and journeys. Using a filter for number of journeys > 100 leads to a correlation of 0.6.

Tab. 3: Correlation between trip origins and number of inhabitants

City	Correlation r	Coefficient of determination R^2	corrected r^2	covariance	Number of datasets
Dresden	0.689	0.474	0,472	583.818	394
Chemnitz	0.672	0.452	0,448	153.053	276
Leipzig	0.702	0.494	0,491	407.306	361
Berlin	0.678	0.460	0,458	1.110.467	1017
Overall	0.660	0,435	0,436	159.337	377048

It has to be kept in mind that there might be a systematic exclusion of members with lower income and lower educational standards from smartphone based data collection methods. This may lead to a further bias where areas with lower income and educational status are not equally represented in data. This is an issue which will clearly lead to problems in modelling approaches.

4. Discussion

Paying attention to the issues raised in the test of representativeness, GPS data especially generated by the Strava app is clearly useful for bicycle planning in municipalities. Assessments of how, when and where cyclists are moving in the network are possible as well as predictions on the traffic volume. The prediction of the traffic volume in Dresden shows that on the basis of hourly values of longer recording periods reliable predictions of the traffic volume are possible for the main road network. The originating traffic cells and the number of inhabitants also seem to be correlated. This finding is, however, not generalizable for all originating traffic cells. Some originating traffic cells with a high number of inhabitants and a rather low social background show a lower number of trips compared to the general findings. An influence by the homogeneous user group can be suspected here resulting in presumably social exclusion of inhabitants with a low-level education or low income. At this point further sociological studies are recommended. Special attention is also needed concerning very active app users, which can lead to distortion in data. It is furthermore recommended to achieve a normalization of the numbers of users on the supply side or to restrict the time period if applicable if more than one dataset is used. The lack of route data is a major concern when it comes to building models. It does not seem possible to estimate route choice decisions based on a biased aggregate user data. It also seems doubtable that Strava represents the population of cyclists in a given city due to motivational model and smartphone based approach.

5. Conclusion

In the long-term, the use of GPS data to be valuable for infrastructure planning. The active participation of cyclists can open up new possibilities in the communication process. In the data set, used in this analysis this possibility couldn't be fully exploited due to the low interactivity of the Strava app in this case. The strength of data clearly is related to its high potential regarding the visualization of cyclists and their trips in a municipality. The use of visualized data could become a strong tool in further political discussions. Ongoing research is required regarding the differences in cycling behaviour of different user groups to avoid social exclusion in bicycle planning. Due to the biased user group, Strava data is not feasible for modelling approaches.

Acknowledgements

The project is funded under the NRVP 2020 frame, Federal Ministry of Transport and Digital Infrastructure, Germany.

6. References

Ahrens, G.-A. (2015, April). *Sonderauswertung "Mobilität in Städten SrV 2013".* Retrieved 05 30, 2017, from https://tu-dresden.de/bu/verkehr/ivs/srv/ressourcen/dateien/2013/uebersichtsseite/SrV2013_Stadtgruppe_GrosseSrV-Vergleichsstaedte.pdf?lang=de

Broach, J., Dill, J., & Gliebe, J. (2012, 07 14). Where do cyclists ride? A route choice model developed with revealed preference Data. *Transportation Research Part A*, pp. 1730-1740.

Charlton, B., Sall, E., Schwartz, M., & Hood, J. (2011). Bicycle Route Choice Data Collection using GPS-Enabled Smartphones. *Transportation Research Board Annual Meeting*, (p. 10).

El Esaway, M. (2013). Development of Daily Adjustment Factors for Bicycle Traffic. *Journal of Transport Enginieering*, pp. 859-871.

Falkenberg, G., Blase, A., Bonfranchi, T., Draeger, W., Vortisch, P., Kautzsch, L., Zimmermann, A. (2003). *Bemessung von Radverkehrsanlagen unter verkehrstechnischen Gesichtspunkten.* Bergisch-Gladbach: Bundesanstalt für Straßenwesen (BASt).

Francke, A., Lißner, S., & Becker, T. (2017). Big-Data in Bicycle Traffic. Retrieved from Civitas.eu: http://civitas.eu/tool-inventory/guideline-big-data-im-radverkehr

Gray, P. B., McHale, T. S., & Carré, J. M. (2016, July 19th). A review of human male field studies of hormones and behavioral reproductive effort. *Hormones and Behavior*, pp. 52-67.

Griffin, G. P., & Jiao, J. (2015). Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *Journal of Transport an Health*, pp. 238-247.

Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Jornal of Transport Geografy*, pp. 90-97.

Romanillos, G., Zaltz-Austwick, M., Ettema, D., & De Kruijf, J. (2016). Big Data and Cycling. *Transport Reviews, 36:1*, pp. 114-133.

Schelewsky, M., Jonuschat, H., Bock, B., & Stephan, K. (2014). *Smartphones unterstützen die Mobilitätsforschung - Neue Einblicke in das Mobilitätsbverhalten durch Wege-Tracking.* Berlin: Springer Vieweg.

Schiller, C., Zimmermann, F., & Bohle, W. (2011). *Hochrechnungsmodell von Stichprobenzählungen für den Radverkehr.* Dresden.

Schüssler, N., & Axhausen, K. W. (2008). Identifying trips and activities and their characteristics from GPS raw data without further information. *Conference Paper ETH Zürich Research Collection*, (p. 29). Zürich.

Strauß, J., & Miranda-Moreno, L.(2017). Speed, travel time, and delay for intersections and road segments in the Montral network using cyclists Smartphone GPS dara. *Transportation Research Part D*, pp. 155-171.