Comparison of Smartphone-based Cyclist GPS Data Sources

Corresponding Author: Kari Watkins, Assistant Professor, Civil & Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA 30332
Phone: (206) 250-4415, Fax: (404) 894-2278, Email: kari.watkins@ce.gatech.edu

Rohit Ammanamanchi, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA 30332
Phone: (704) 770-6960, Email: rammanamanchi3@gatech.edu

Jeff LaMondia, Assistant Professor, Department of Civil Engineering,
Auburn University, 238 Harbert Engineering Center, Auburn, AL 36849-5337
Phone: (334) 844-6284, Fax: (334) 844-6290, Email: jlamondia@auburn.edu

Christopher A. Le Dantec, Assistant Professor, Digital Media,
Georgia Institute of Technology, TSRB 316-A,85 5th Street NW, Atlanta, GA. 30308
Phone: (404) 385-7555, Fax: (404) 894-2278, Email: ledantec@gatech.edu

Submission Date: 8/1/15
Number of Words: 5357 + references (500 words)
Number of Tables/Figures: 5 (1250 words)
Total: 7107 words

Of note to reviewers: We have used color liberally in the graphics shown in this paper. We feel that the graphics are much better represented in color. However, we already have a plan of how to reformat the graphics for black and white in the event that this paper is (hopefully) chosen for publication.

ABSTRACT
It is important for planning agencies to have data on cyclist travel patterns, routes, volumes, and speeds, but access to such data is currently limited and often expensive to obtain. Many regions are looking toward the use of GPS data collected using smartphones to track cyclist trips, both via apps deployed by the agencies and, more recently, fitness-based apps providing anonymized user data by roadway segment. As regions begin to collect and purchase GPS-based data, there are many questions about potential uses in transportation planning, including the representativeness of the data. This paper provides a comparison of the data obtained from two smartphone-based apps, Cycle Atlanta and Strava, to begin to understand how GPS data can be used to map cyclist movements in an urban area. Analysis includes user demographic data and overall trip statistics, time-of-day, and geographic trips by segment comparisons. Differences between the two populations were found in terms of gender, age, percent commute trips, trip lengths, and preference for bike paths. Cycle Atlanta data was also compared to a set of manual bike counts and it was found that only about 3% of the cyclists counted had recorded their trip in Cycle Atlanta. The usage of GPS-based smartphone cycling app data is a promising new data source for transportation planning and design analysis, but should carefully take into account the likely bias from the self-selected users of such apps. These apps can supplement, but not replace large-scale count programs to establish system-wide cyclist volumes.

INTRODUCTION

Transportation has a central role in a healthy, sustainable society and economy. However, our current transportation system is associated with numerous societal problems, including congestion, pollution, energy consumption, equity issues, and health impacts. It is widely believed that bicycling as a mode of transportation could address many of these issues (1). The federal government (2) as well as many state and local transportation planning agencies have recently geared their policies towards promoting biking and walking in their long term visions. Despite this recent interest, literature shows that although 40% of the trips made in the U.S. are a bikeable distance of less than 3 miles, only 1.8% of such trips are bicycle trips (3). This low usage of bicycling as a mode has generally been attributed to safety issues with major safety perception factors including high speed limits, high traffic volumes, last mile disconnect in the network, and an absence of physically separated facilities for cyclists (4, 5).

Studies reveal that a substantial increase in the number of bicyclists can be achieved by providing facilities for safe riding (6) and therefore, it is important for planning agencies to know where the cyclists prefer to bike and their desire for dedicated facilities. However, data on cyclist travel patterns (routes, volumes, etc) is severely limited. Comprehensive count programs are critical to fill in the gaps in systemwide volumes over time (7), but these are expensive to conduct and cannot assess cyclist's individual behaviors, such as route choice. Therefore, many regions are looking toward the use of GPS data collected via smartphones to track cyclist trips. Initially this GPS data was recorded via apps developed by regional planning agencies, municipalities, and researchers. More recently, fitness-based apps have begun providing their anonymized data to regions via heat maps of cyclist trips and aggregated segment cyclist counts.

However, as regions and municipalities begin to collect and purchase this GPS-based data, there are many questions about potential uses in transportation planning, including the representativeness of the data. This paper provides a comparison of the data obtained from two smartphone-based apps, Cycle Atlanta and Strava, to begin to understand how GPS data can and cannot be used to map cyclist movements in an urban area. Additionally, the Cycle Atlanta data is compared to a set of manual bike counts conducted by the local business coalition, Midtown Alliance, to have a measure of what percentage of the cyclists in a region a good tracking program can capture.

LITERATURE REVIEW

Multiple studies have used cyclist GPS data over the past decade. The most prevalent use of cyclist tracking data is for inputs into travel demand models, such as length of route and route choice (8, 9, 10, 11, 12, 13, 14, 15, 16) as well as trip generation and distribution (17). These route choice studies often include a larger analysis on general cyclist travel behavior, including travel times (18) and the influence of infrastructure, such as bridges (19). These types of studies require only information about individual route choices from a randomly selected sample of cyclists. As such, characterizing cyclist movements system-wide has not been an issue for this work.

GPS tracking data have also been used to assess cyclist speed and acceleration as a tool for design and planning analysis (20, 21) or simulation (22), as well as categorizing bicycle environments (23). Cyclist GPS and high quality accelerometers have been used to assess pavement quality on cycling routes (24). More recently, multiple researchers have begun to use cyclist GPS data for safety analyses, including stopping behavior, speed, and wrong-way riding (25).

Many of the early GPS-based cyclist routing and speed studies relied on equipment mounted specifically to bicycles. However, beginning in 2009 with the Cycle Tracks app created by the San Francisco County Transportation Authority (26), the GPS in smartphones have been used to make the collection of data easier. Multiple regions have begun to collect data using Cycle Tracks such as Austin, Monterey, Raleigh, Fort Collins, Minneapolis, Seattle, Salt Lake City, Los Angeles, Toronto, and Lexington (KY) or rebranded and improved the original app, such as Lane County (OR), College Station, Charlottesville, Hampton Roads (VA), Atlanta, Montreal, Reno, and Philadelphia (27).

However, there are limitations to using smartphone data. Even high-end GPS equipment is not accurate enough to monitor cyclist position within a roadway to allow assessment of bike lane usage, sidewalk usage, etc (28). Perhaps most critically, despite the use of GPS data to assess cyclist exposure for injury risk analysis in at least one case (29), most researchers question the usage of GPS-based data for systemwide counts (7). Finally, there are potential equity issues in utilizing this data, such as the exclusion of individuals without access to smartphones, those who are unfamiliar with such apps, or those who are not interested in recording information. At this time, we are only beginning to understand which types of cyclists are being captured by this collection method.

It is also important to note that the use of Cycle Tracks or similar derivative smartphone apps requires the local agency deploying it to maintain a local server to collect the data, post-process the data for use, and upgrade the app periodically for the latest operating systems. Furthermore, these apps must be advertised to recruit cyclists and cyclists must agree to upload their trips at least initially and ideally continuously over time. Many cyclists already record their trips using smartphone apps to keep track of performance over time and compare routes and statistics with other cyclists. Therefore, in 2013, the Oregon Department of Transportation began a partnership with Strava to allow the use of data recorded by cyclists on their propriety app (30) and Strava subsequently began a program to display the data on their Global Heatmap and sell more detailed data to additional regions for use in planning efforts. Multiple regions have now purchased Strava Metro data, including Auckland, New Zealand, where the data have been used to understand locations cyclists avoid (31).

METHODOLOGY

In this paper, we will assess the differences between agency-sponsored smartphone app cyclist tracking data and fitness-based smartphone app cyclist tracking data, and provide an initial comparison to manual count data. The analysis uses three primary data sources: GPS data from the Cycle Atlanta app, data purchased from the Strava Metro program, and cyclist intersection counts from Midtown Alliance. Four analyses will be conducted to compare the datasets, including a comparison of user demographic data and overall trip statistics, a time-of-day comparison, a geographic comparison of trips by segment, and a comparison of the bike counts to the Cycle Atlanta app recorded trips.

Cycle Atlanta App

The first dataset uses the data collected through the smartphone app named Cycle Atlanta. In order to promote cycling in Atlanta, collaboration was set up between the Georgia Institute of Technology and the City of Atlanta's planning office to develop a smartphone app that would help in collecting data from bicyclists. The project was further facilitated by support from Atlanta Regional Commission who viewed the project as a means to foster "extensive public involvement by neighborhood residents, business owners, and the citywide cycling community" (32).

The smartphone app created for this initiative was named Cycle Atlanta, after the larger planning project for which the app was initiated, and was developed by an interdisciplinary team of researchers. The app was based off of San Francisco's CycleTracks (10), although Cycle Atlanta was substantially updated to make better use of current features available in iOS and Android as well as to include features that the City and local bicycle advocacy groups wanted in the app. The basic feature is trip recording, where the app uses the GPS of the phone to record the location of the user once per second. In addition to tracking cyclists' trips, the app also provides options to enter personal information, including age, email address, gender, ethnicity, home income, ZIP codes (home, work, and school), cycle frequency, rider type, and rider history. The app was launched in October 2012 and has been upgraded periodically.

Cycle Atlanta users were recruited over time via postcards handed out on the street and at cycling events, including the opening of the Atlanta Beltline Eastside Trail and the Atlanta Streets Alive ciclovia, as well as postings on social media sites for cycling groups and local agencies. About a dozen articles in popular media (Atlanta Journal and Constitution, the local NPR station, local TV networks, etc.) have also mentioned how to download the app. For this analysis, the Cycle Atlanta dataset included trips recorded that fall within a 5 mile radius of the intersection of Ponce de Leon Ave and Monroe Ave in Midtown Atlanta from October 10, 2012 to August 31, 2014. In this time period, Cycle Atlanta had a total of 1,541 users in the study area contribute 18,467 total trips, of which 11,082 were designated as commute trips (60%).

Strava App (Atlanta)

The second dataset includes anonymized data purchased from Strava for the same 5 mile radius of the intersection of Ponce de Leon Ave and Monroe Ave for August 1, 2013 to July 31, 2014. The data includes a total number of trips by time period of the day (very early morning, early morning, AM peak, midday, PM peak, early evening, and late evening) on segments on a roadway network the study team provided, as well as median speeds on each segment. In this time period, Strava had a total of 3,236 users in the study area contribute 51,408 total trips, of which 15,027 were designated as commute trips (29%).

Cycle Atlanta and Strava Comparison Analysis

Three analyses involved a comparison between Cycle Atlanta and Strava datasets. First, user demographic data collected within the apps was compared to understand how the gender and age of users differs. Second, the time-of-day that trips were recorded was compared for the two applications. The final comparison is the map of trips by segment generated from the two apps. In order to compare the number of trips on individual roadway segments between the two datasets, the data had to be normalized by the total number of trips recorded in each app, because the Strava data had a greater number of trips recorded. Therefore, a percent of variation between the Cycle Atlanta and Strava data was calculated based on the following formula.

Percent of variation of a given street segment:

$$V_i = \left(\frac{x_{S,i}}{N_S} - \frac{x_{CA,i}}{N_{CA}}\right) * 100\% \qquad (1)$$

Total Variation of all streets in a bounded area:

$$V = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_{S,i}}{N_S} - \frac{x_{CA,i}}{N_{CA}}\right| * 100\% \qquad (2)$$

Where:

V = Variation of route choices, in percent
n = total number of street segments
i = a specific street segment
$x_{\mathrm{CA,i}}$ = Cycle Atlanta count of cycling trips recorded on a given street segment
$x_{\mathrm{S,i}}$ = Strava trips recorded on a given street segment
N_{CA} = Total number of Cycle Atlanta trips recorded
N_{S} = Total number of Strava trips recorded

Midtown Atlanta Counts

In addition to the comparison between Cycle Atlanta and Strava data, a comparison between Cycle Atlanta data and cyclist counts conducted in Midtown Atlanta was made to estimate the percentage of trips that are being recorded by the app. A 6 month subset of Cycle Atlanta data from January 1, 2013 to June 30, 2013 was compared to manual counts of cyclists at 78 intersections taken by Midtown Alliance in March 2013. These counts include a total number of cyclists entering the intersection from all approaches during two time periods: AM Peak (7:30-9:30am) and PM Peak (4:00-6:00pm). Cycle Atlanta does not have enough data to compare the data of one day, therefore, it must be assumed that the manual counts can be extrapolated to represent the daily count of all weekdays over a six-month period. The Cycle Atlanta trips recorded over the weekdays of a continuous six-month period were aggregated for each intersection. The percentage of trips that are being recorded on Cycle Atlanta was estimated to be the ratio of the count of recorded trips through an intersection to the manual count of cyclists through that intersection multiplied by the number of weekdays in six months. A bootstrapping method was used to estimate the confidence interval around this percentage of trips. It is of note that bike counts were of total cyclists entering the intersection rather than screenline counts, therefore the Strava data was not in a format that allowed comparison.

RESULTS

The results for the four analyses are provided below.

Demographic Comparison

The demographic breakdown of Cycle Atlanta and Strava users who recorded trips during the time period is shown in Table 1. As shown in the table, both populations have a larger proportion of men than women, but the skew is greater for Strava (84%) than for Cycle Atlanta (76%). The Cycle Atlanta dataset is much younger than Strava, with 57% of the users being under 35 in Cycle Atlanta compared to only 37% in Strava. It is unknown how these gender and age breakdowns compare to the actual cycling population in Atlanta. An early analysis of Cycle Atlanta users compared to National Household Travel Survey data and Bike to Work Challenge data showed that Cycle Atlanta users tended to be younger and more likely male, but these datasets are known to have a greater representation of older people and include a very small portion of the cycling population as well (33).

In both applications, providing demographic data is optional. Only approximately 64% of Cycle Atlanta users report their demographic data in the case of both gender and age.

Interestingly, a much greater number of Strava users report their gender (95%) than their age (80%). This appears to be due to the positioning of data input prompts in the set-up of the application, with gender being the second prompt, but age requiring a user to go into their profile settings to specify. It is of note that Strava obtains demographic data from a greater percentage of users, but asks many fewer demographic questions than Cycle Atlanta. Additional demographic breakdowns of Cycle Atlanta data can be found in Misra et al (34).

TABLE 1. Demographic Breakdown of Cycle Atlanta and Strava (Atlanta) Users

	Cycle Atlanta		Strava (Atlanta)	
	Number in category	**% in category from respondents providing data**	**Number in category**	**% in category from respondents providing data**
Total Users	1,541		3,236	
Gender				
Male	741	76%	2,586	84%
Female	240	24%	482	16%
No Data	560		168	
Age				
Under 25	116	12%	171	7%
25 – 34	448	45%	775	30%
35 – 44	218	22%	800	31%
45 – 54	144	14%	600	23%
55 – 64	66	7%	205	8%
65+	9	1%	31	1%
No Data	540		654	
Total Recorded Trips	18,467		51,408	
“Commute” Trips	11,082	60%	15,027	29%
Avg. Trips per User	11.98 trips		15.89 trips	
Avg. Travel Distance	5.91 mi		20.28 mi	
Median Travel Distance	4.60 mi		16.58 mi	
Median Travel Time	26.5 min		83.6 min	
Median Speed	10.43 mph		11.90 mph	

Table 1 also shows several summary statistics regarding the datasets. The average number of trips recorded per user had fewer recorded in Cycle Atlanta (11.98 trips/user) than Strava (15.89 trips/user). With regard to commute trips, Cycle Atlanta had 60% versus 29% in Strava, although the pure number of commute trips was higher in Strava. Strava does not ask about trip purpose, but does give users an opportunity to designate a commute trip. Cycle Atlanta asks trip purpose, with options that include commute, school, work-related, exercise, social, shopping, errand, and other. Cycle Atlanta had substantially shorter average travel distances (5.91 mi) compared to Strava (20.28 mi); substantially shorter median travel distances (4.60 mi in Cycle Atlanta to 16.58 mi in Strava); and median travel times at 26.5 minutes in Cycle Atlanta compared to Strava’s 83.6 minutes. This results in a median speed, calculated from median distance over median time, of 10.43 mph in Cycle Atlanta compared to 11.90 mph in Strava. It is of note that a median travel time is reported rather than an average travel time, because a large percentage of users leave their phones recording after their trip is complete, extending the time elapsed beyond the travel time. This makes the average travel times difficult to use without adjustment.

Time-of-Day Comparison

The time-of-day that trips were recorded was available from both the Cycle Atlanta and Strava datasets for comparison. Strava's aggregated trips are provided by the following time periods:

- Very Early AM hours: 12am – 3:59am
- Early AM hours: 4am – 5:59am
- AM Peak Hours: 6am – 8:59am
- Mid-Day Hours: 9am – 2:59pm
- PM Peak Hours: 3pm – 5:59pm
- Early Evening Hours: 6pm – 7:59pm
- Late Evening Hours: 8pm – 11:59pm

The resulting breakdown in trips by time period for all trips, weekday trips, weekend trips, and weekday commute trips is shown in Figure 1.

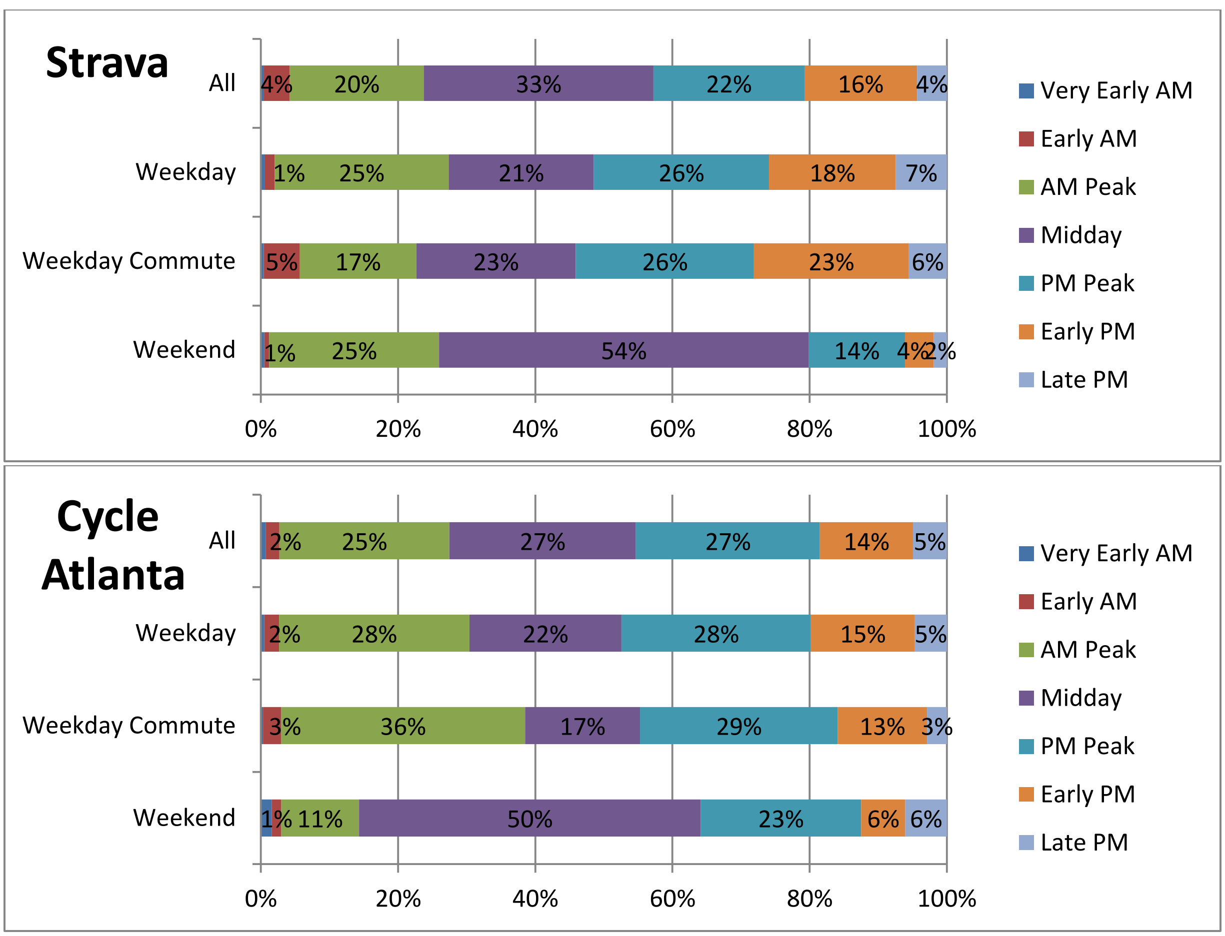

FIGURE 1. Time period of recorded Strava and Cycle Atlanta trips

As shown in Figure 1, very few trips are recorded overnight or in the early AM, with slightly more Strava trips recorded in these hours. Cycle Atlanta users record more trips in the weekday AM peak, especially commute trips. Strava has more trips recorded in the evening hours, including many trips designated as commute that occur after 8 pm. Many of the weekend trips occur midday in both apps, with Strava users having a similar proportion of trips in the AM peak, whereas weekend Cycle Atlanta users appear to travel later.

Roadway Segment Comparison

The total variation in trips from Cycle Atlanta to Strava is very small at 0.70% for the total trips and 1.433% for the commute trips. Interestingly, the variation is unexpectedly greater for commute trips, although both variations are small. The percent variations calculated for each roadway segment were mapped in ArcGIS to allow visual comparison between the two datasets. Figure 2 shows the percent variations for all data recorded by the app users by roadway segment.

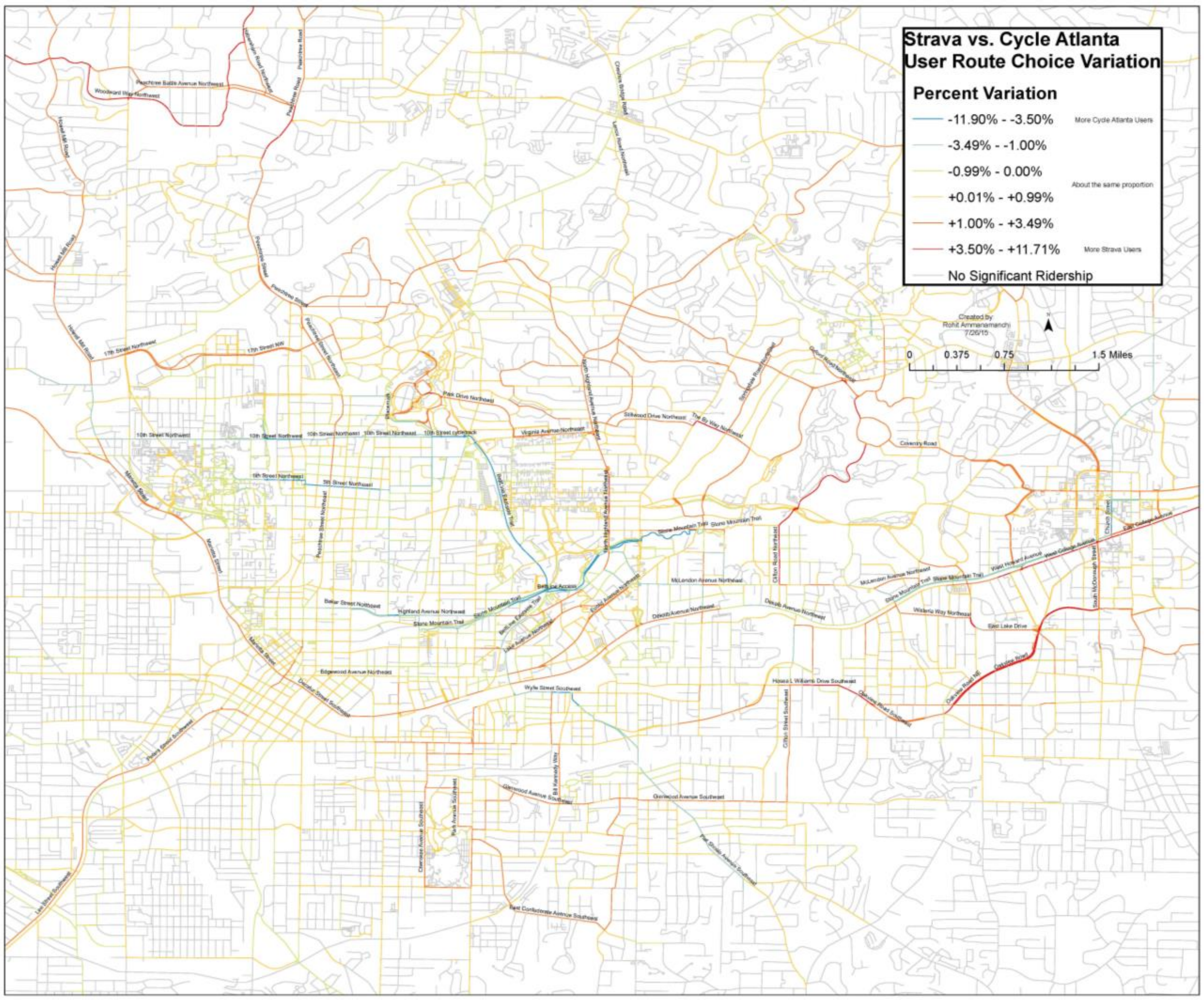

FIGURE 2. Comparison of Strava and Cycle Atlanta data on roadway segments – all data

As shown in Figure 2, Cycle Atlanta trips are concentrated around the Georgia Tech campus, where recruitment efforts and publicity were strongest, whereas Strava trips are spread over a much wider area, providing a greater overall representation. Both user groups show a strong preference for bike paths, but the preference is stronger in Cycle Atlanta, with a larger percentage of users on the Atlanta Beltline, the Freedom Park Trail, and the Stone Mountain Trail. This is particularly prevalent on the two parallel roads in the eastern portion of the graphic, West Howard Ave (greater proportion of Cycle Atlanta users) and West College Ave (greater proportion of Strava users). West Howard has an aging multiuse trail along the roadway that provides protection from motor vehicles, whereas West College requires cyclists to share the lane with motor vehicles, but the pavement is much better, allowing faster speeds.

Figure 3 shows the percent variations for trips designated as commute trips during the AM period of 6 am to noon by roadway segment. Again, Cycle Atlanta users show a preference for trails, including the pair discussed above. In addition, Cycle Atlanta users show a preference for streets within and near Piedmont Park, where motor vehicles are not allowed to travel, as well as less highly traveled roadways that connect trails.

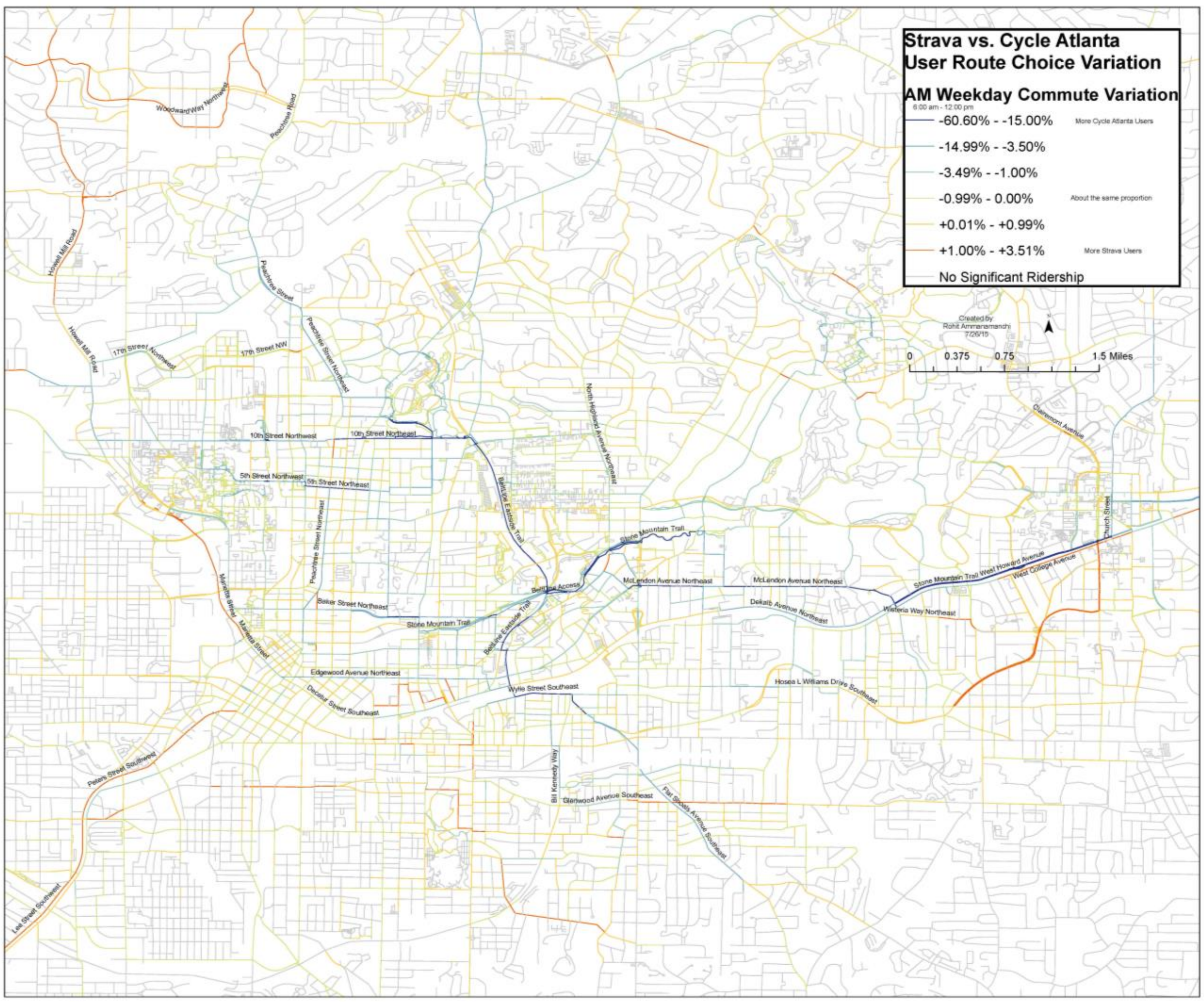

FIGURE 3. Comparison of Strava and Cycle Atlanta data on roadway segments – AM commute data

Midtown Atlanta Counts Comparison

In addition to comparing Strava and Cycle Atlanta data, an attempt was made to find data to compare Cycle Atlanta trips to actual cyclist trips on the network. As described in the methodology, counts were available from the Midtown Alliance for comparison between cyclist volumes at 78 intersections. In the AM peak, 3.88% ± 1.1% of the cyclists counted recorded their trip in Cycle Atlanta. In the PM peak, 2.45% ± 1.1% of the cyclists counted recorded their trip in Cycle Atlanta. The relatively early designation of PM peak (4 to 6 p.m.) compared to a typical campus and inner-city commute may contribute to the difference between the two peak hours. Figure 4 shows the ranges in these percentages by intersection, with some intersections having greater than 20% of the cyclists counted having recorded their trip.

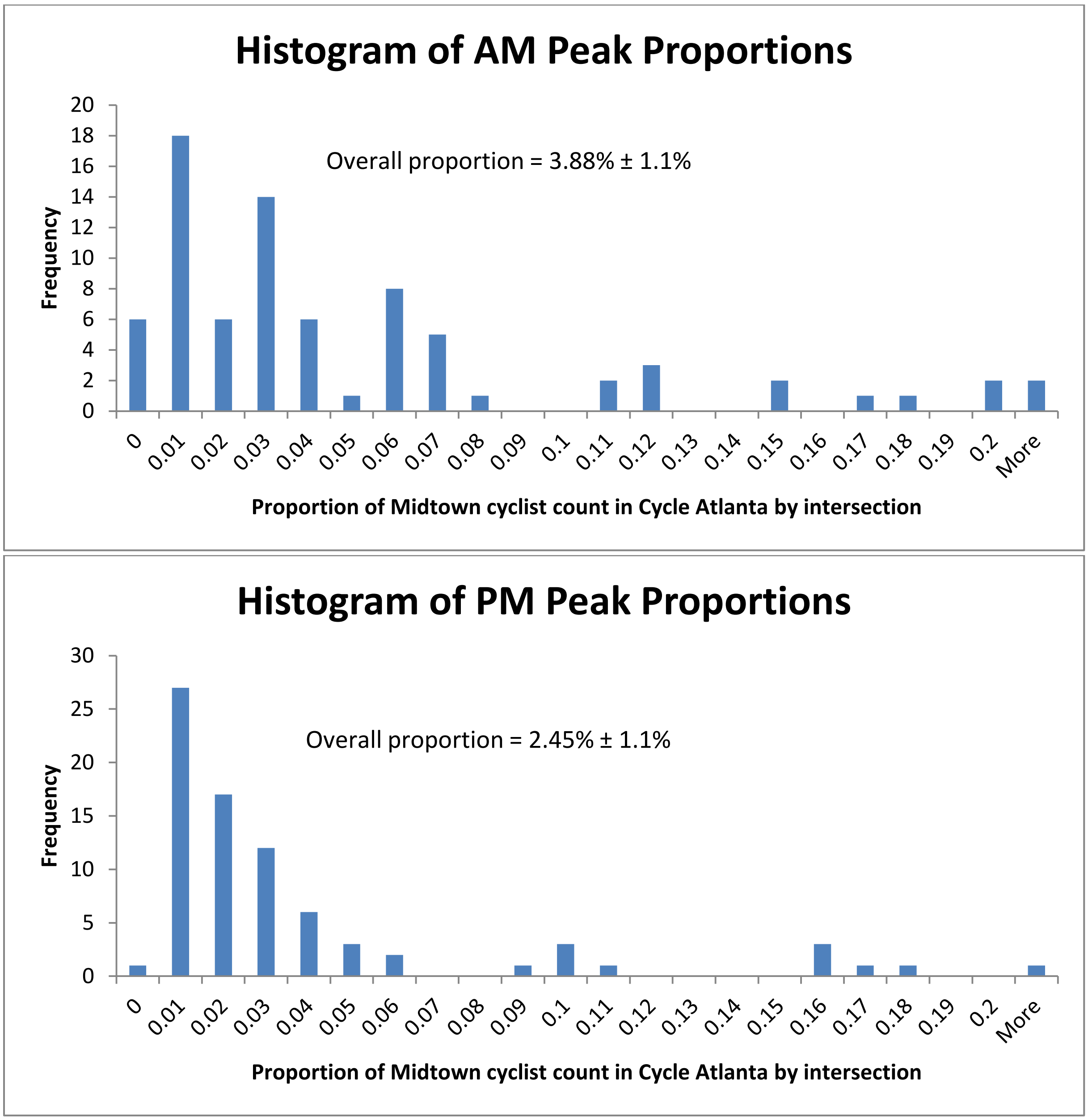

FIGURE 4. Proportions of Midtown cyclists in Cycle Atlanta data by intersection

DISCUSSION OF RESULTS

The question this paper is trying to address is if the population of cyclists recording trips in a fitness-based app such as Strava differs from an agency-monitored app such as Cycle Atlanta. In some cases the populations were similar, but noticeable differences between the two populations were also found.

In terms of demographics, a larger proportion of men record trips than women in both apps, but the skew is greater for Strava. The Cycle Atlanta dataset is much younger than Strava. Other demographic data could not be compared due to a lack of availability in the Strava data. It is unknown how these gender and age breakdowns compare to the actual cycling population in Atlanta, because there is limited data about the cycling population available. Regions such as Atlanta that wish to use such data sources need to conduct travel surveys in which cyclists are over-sampled to obtain demographics for comparison to these data sets. In addition, smartphone apps, including Strava, must at least periodically ask more demographic questions for a comparison to be conducted.

Regarding trips rather than users, the Cycle Atlanta dataset contains twice the percentage of commute trips than the Strava dataset, although the total number of commute trips was smaller. Furthermore, trips recorded in Cycle Atlanta were only about 30% of the length of Strava trips in terms of average and median travel distances and median travel time. The median speed was comparable, but noticeably slower in Cycle Atlanta at 10.43 mph compared to 11.90 mph in Strava. Although the majority of the trips occurred during similar time-periods, more Strava trips occurred in late and very early hours when athletes are more likely to travel, even for trips that users labeled as commute.

Perhaps the more critical question that must be asked is if the differences in population matter in the use of the data for transportation planning and design purposes. Assessment of the trips by segment found that Cycle Atlanta trips are concentrated around the Georgia Tech campus, where recruitment efforts and publicity were strongest, whereas Strava trips are spread over a wider area. Critical to planning efforts, Cycle Atlanta users show a strong preference for bike paths, cycle tracks, and low speed roads connecting the network between bike paths. In both apps, users recording trips are self-selected. In the case of Strava, users likely include a larger portion of athletes who are concerned about performance and may take more direct yet faster routes. In the case of Cycle Atlanta, users likely include a larger portion of cycling activists who are sufficiently interested in the project to be willing to share data and invest time without any personal gain.

For the purpose of travel demand models, information such length of route, route choice, route travel times, and the influence of infrastructure are critical to the development of models and understanding of cyclist choices. Such analysis regarding individual routing and trip choices can be obtained from a randomly selected sample of cyclists. GPS-data from apps such as Cycle Tracks and Cycle Atlanta are ideal for this purpose as long as the recruitment efforts ensure the cyclists are somewhat representative of the larger population.

Proprietary apps such as Strava must aggregate their data to protect the privacy of users and therefore individual decisions cannot be monitored. This includes some travel demand analysis such as route choice and some safety and planning analysis, including acceleration. Both existing fitness apps, such as Strava, and deployed agency-sponsored apps can be used to obtain data on link travel time and average speed to compare one link to another, although the level of fitness may come into play using raw numbers for links. Similarly, a gap analysis to understand roadways that are avoided by cyclists can be conducted using data from both types of apps, assuming the sample is sufficiently large. However, the preference for dedicated cyclist infrastructure over speed shown in Cycle Atlanta should be taken into account. Origin-destination data can be obtained from both apps, as well as other behavioral data such as time at nodes to analyze stopping behavior and directional totals for wrong-way riding.

For many analyses, the primary data requirement is volumes of cyclists. To date, cyclist volumes cannot be obtained from crowdsourced apps. These apps include a very small sample of self-selected cyclists. Therefore, large-scale count programs geared directly toward obtaining cyclist volumes are still desperately needed and should be pursued by regional planning agencies and state departments of transportation. This data can be used, among other things, to verify cyclist data obtained from smartphone apps to understand if they represent the cycling population and to develop factors to allow greater use of smartphone app-based data in the future.

In the case of Cycle Atlanta, only 3.88% ± 1.1% (AM) and 2.45% ± 1.1% (PM) of the cyclists counted passing through intersections in Midtown Atlanta recorded their trip in Cycle Atlanta. Even within the primary area of recruitment, there was a wide range in the proportion of cyclists recording trips who passed through each intersection. Therefore, any use of crowdsourced apps to understand cyclist volumes must be considered carefully. In particular, such apps are often used for a distribution of trips into an area to produce a heat map of where cyclists are traveling from. This assumes the app data is a representative sample of the entire population. Although some researchers are beginning to develop methods for such uses of the data, most researchers question the usage of GPS-based data for systemwide counts.

CONCLUSION AND FUTURE RESEARCH

The usage of GPS-based smartphone cycling app data for transportation planning and design analysis should carefully take into account the likely bias from the self-selected users of such apps. In this study, differences were found in the user population and recorded trips between the fitness-based smartphone app Strava and the agency-deployed app Cycle Atlanta. In addition, the percentage of the total cycling population recording trips in Cycle Atlanta was found to be only approximately 3% with substantial variation by intersection. Therefore, data from both types of GPS-based apps should be compared to other local data sources and weighted appropriately. Future research should include methods for matching the data for such weighting processes.

As long as the user population is taken into account, Strava data provides an opportunity for agencies to obtain data without deploying their own app. With Cycle Tracks and Cycle Atlanta, the local agency deploying the app must maintain a local server to collect the data, post-process the data for use, and upgrade the app periodically for the latest operating systems. Both Strava and Cycle Atlanta data can be used for gap analysis to understand roads avoided by cyclists, origins and destinations of cyclists, some cyclist behaviors, and broad changes in patterns due to new infrastructure. Individual routing data from an app such as Cycle Tracks or Cycle Atlanta is required for route choice models and other cyclist behaviors, particularly to segment data by user characteristics.

A major issue often encountered in the utilization of such data is the equity of obtaining data from a smartphone that may not be usable by the entire population. Smartphone ownership is now believed to be approximately even across gender and race, with limited influence on ownership based on income (35). Smartphone ownership differs substantially by age, but the skew toward the younger population is similar to the skew of the cycling population. The percentage of smartphone users with a current data plan is lower than smartphone owners;

however some smartphone apps, such as Strava, can be run without a continuous data connection, allowing upload when wifi is again connected.

Finally, in the case of both Strava and Cycle Atlanta, the data produced is only as good as the recruitment method and retention of users over time. Cycle Atlanta showed substantial numbers of trips recorded where recruitment efforts were high, however the larger reach was small and tapered off over time. Current self-selected Strava users may be skewed toward spandex-clad cyclists, but additional users can be recruited to record their trips to provide greater representation in a region. In personal communication, Strava indicated they are willing to work with agencies to cap fees at the pre-recruitment level when agencies advertise the app to recruit additional users.

As mentioned, both app-based data sources can only be used by keeping the likely biases in mind and assessing the impact of those biases on the particular analysis being conducted. Using both types of app data in combination with counts can give a robust picture of cycling in a region. However with all three data sources, critical populations that should be the target of our designs are missing, namely the casual cyclist and the potential cyclist who is not yet riding, but would like to be. Methods to assess the infrastructure desires of these future cyclists are a necessary component of increasing cycling mode share. In the meantime, smartphone-based apps can quickly show how new infrastructure is being used by the larger cycling population to justify infrastructure investments.

ACKNOWELDGEMENTS

The authors would like to thank the many users of crowdsourced GPS-apps for providing their data for planning efforts; Strava for their review of this work, providing additional data and answers, and running the Cycle Atlanta data through their process for direct comparison; and Midtown Alliance for their cyclist count data. Thanks also go to multiple students involved in the Cycle Atlanta project, especially Aditi Misra. Funding for this project has been provided by the Southeastern Transportation Research, Innovation, Development and Education Center and the Georgia Department of Transportation with initial funding to set up Cycle Atlanta provided by the City of Atlanta, GVU Center, and the Institute for People and Technology. However, opinions, omissions, and errors are the responsibility of the authors.

REFERENCES

1. Sallis, J.F., Frank L.D., Saelens BE, Kraft M.K. Active transportation and physical activity: Opportunities for collaboration on transportation and public health research. Transportation Research Part A 28(4): 249-268, 2004.
2. Federal Highway Administration. United States Department of Transportation Policy Statement on Bicycle and Pedestrian Accommodation Regulations and Recommendations, http://www.fhwa.dot.gov/environment/bicycle_pedestrian/overview/policy_accom.cfm
3. Pucher, J., Buehler, R., Merom, D., & Bauman, A.Walking and Cycling in the United States, 2001-2009: Evidence from the National Household Travel Surveys. American Journal of Public Health, Vol. 101(S1), pp. S310-S317, 2011.
4. Dill, J. and Carr T. Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them – Another Look. 2003.
5. Buehler, R., Pucher, J. "Cycling to Work in 90 Large American Cities: New Evidence on the Role of Bike Paths and Lanes," Transportation, Vol. 39 (2), 409-432. 2012.

6. Pucher, J. and R. Buehler. Making Cycling Irresistible: Lessons from the Netherlands, Denmark and Germany. Transport Reviews: A Transnational Transdisciplinary Journal. 28(4): 495-528, 2007.
7. Ryus, P., Ferguson, E., Laustsen, K., Schneider, R., Proulx, F., Hull, T., Miranda-Moreno, L. Guidebook on Pedestrian and Bicycle Volume Data Collection. NCHRP Report, Issue 797, Transportation Research Board, 2014
8. Harvey, Francis; Krizek, Kevin J; Collins, Reuben. Using GPS Data to Assess Bicycle Commuter Route Choice. Transportation Research Board 87th Annual Meeting, Transportation Research Board, 2008
9. Menghini, G; Carrasco, N; Schussler, N; Axhausen, K W. Route Choice of Cyclists in Zurich. Transportation Research Part A: Policy and Practice, Volume 44, Issue 9, 2010
10. Hood, J., Charleston, B., Sall, E. A GPS-based bicycle route choice model for San Francisco, California. Transportation Letters: the International Journal of Transportation Research 3: 63-75, 2011.
11. Broach, J., Dill, J., Gliebe, J. Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transportation Research Part A: Policy and Practice, Volume 46, Issue 10, 2012, pp 1730-1740
12. Casello, J. and Usyukov, V. Modeling Cyclists' Route Choice Based on GPS Data. Transportation Research Record: Journal of the Transportation Research Board, Issue 2430, 2014, pp 155–161
13. Jackson, S., Miranda-Moreno, L., Rothfels, C., Roy, Y. Adaptation and Implementation of a System for Collecting and Analyzing Cyclist Route Data Using Smartphones. Transportation Research Board 93rd Annual Meeting, Transportation Research Board, 2014.
14. Krenn, P., Oja, P., Titze, S. Route Choices of Transport Bicyclists: A Comparison of Actually Used and Shortest Routes. International Journal of Behavioral Nutrition and Physical Activity, Volume 11, Issue 1, 2014
15. Shen, Q., Chen, P., Schmiedeskamp, P., Bassok, A., Childress, S. Bicycle Route Choice: GPS Data Collection and Travel Model Development. Pacific Northwest Transportation Consortium, 2014
16. Dalton, A., Jones, A., Panter, J., Ogilvie, D. Are GIS-Modelled Routes a Useful Proxy for the Actual Routes Followed by Commuters? Journal of Transport & Health, Volume 2, Issue 2, 2015, pp 219-229
17. Casello, J., Nour, A., Rewa, K., Hill, J. Analysis of Stated-Preference and GPS Data for Bicycle Travel Forecasting. Transportation Research Board 90th Annual Meeting, Transportation Research Board, 2011
18. Dill, J. and Gliebe, J., Understanding and Measuring Bicycling Behavior: A Focus on Travel Time and Route Choice. Portland State University; Oregon Transportation Research and Education Consortium; Research and Innovative Technology Administration, 2008
19. Melson, C., Duthie, J., Boyles, S. Influence of bridge facility attributes on bicycle travel behavior. Transportation Letters: The International Journal of Transportation Research, Volume 6, Issue 1, 2014
20. El-Geneidy, A., Krizek, K., Iacono, M. Predicting Bicycle Travel Speeds Along Different Facilities Using GPS Data: A Proof-of-Concept Model. Transportation Research Board 86th Annual Meeting, Transportation Research Board, 2007

21. Parkin, J. and Rotheram, J. Design speeds and acceleration characteristics of bicycle traffic for use in planning, design and appraisal. Transport Policy, Volume 17, Issue 5, 2010, pp 335-341
22. Henriques, N. and Bento, C. Integration of GPS Traces and Digital Elevation Maps for Improving Bicycle Traffic Simulation Behavior. Transportation Research Board 92nd Annual Meeting, Transportation Research Board, 2013
23. Joo, S., Oh, C., Jeong, E., Lee, G. Categorizing bicycling environments using GPS-based public bicycle speed data. Transportation Research Part C: Emerging Technologies, Volume 56, 2015
24. Bíl, M., Andrášik, R., Kubecek, J. How comfortable are your cycling tracks? A new method for objective bicycle vibration measurement. Transportation Research Part C: Emerging Technologies, Volume 56, 2015, pp 415-425
25. Langford, B., Chen, J., and Cherry, C. Risky riding: Naturalistic methods comparing safety behavior from conventional bicycle riders and electric bike riders. Accident Analysis and Prevention, Vol 82, 2015.
26. Charlton, B., Hood, J., Sall, E., Schwartz, M. Bicycle Route Choice Data Collection Using GPS-Enabled Smartphones. Transportation Research Board 90th Annual Meeting, Transportation Research Board, 2011
27. San Francisco County Transportation Authority. Cycle Tracks for iPhone and Android, http://www.sfcta.org/modeling-and-travel-forecasting/cycletracks-iphone-and-android, accessed 7/15/15.
28. Lindsey, G., Hankey, S., Wang, X., Chen, J., Gorjestani, A. Feasibility of Using GPS to Track Bicycle Lane Positioning. University of Minnesota Intelligent Transportation Systems Institute, 2013
29. Strauss, J., Miranda-Moreno, L., Morency, P. Mapping Cyclist Activity and Injury Risk in a Network Combining Smartphone GPS Data and Bicycle Counts. Transportation Research Board 94th Annual Meeting, Transportation Research Board, 2015
30. Davies, A. Strava's Cycling App is Helping Cities Build Better Bike Lanes, Wired, 2014, http://www.wired.com/2014/06/strava-sells-cycling-data/
31. Norman, G; Kesha, N. Using smartphones for cycle planning. Institution of Professional Engineers New Zealand (IPENZ) Transportation Conference, 2015, Christchurch, New Zealand, 2015
32. The City of Atlanta. Cycle Atlanta: Phase 1.0, 2011. http://documents.atlantaregional.com/lci/2012Applications/Innov_Atlanta_CycleAtlanta.pdf
33. Poznanski A. Analyzing demographic and geographic characteristics of "Cycle Atlanta" smartphone application users. MS thesis, Georgia Tech, 2013.
34. Misra, A., Watkins, K., Le Dantec, C. Socio-demographic Influence on Rider Type Self Classification with respect to Bicycling. Transportation Research Board 94th Annual Meeting, Transportation Research Board, 2015.
35. Smith, A. and Page, D. U.S. Smartphone Use in 2015, Pew Research Center, 2015, http://www.pewinternet.org/files/2015/03/PI_Smartphones_0401151.pdf