Comparison of Cycling Path Characteristics and Park Popularity based on Data from GPS Fitness Tracker Apps

By Angela Schirck-Matthews in fulfillment of Marshall Plan Scholarship research assignment agreement

In conjunction with:

The Marshall Plan Foundation

Marshallplan-Jubiläumsstiftung
Austrian Marshall Plan Foundation
Fostering Transatlantic Excellence

University of Florida, Geomatics Program
3205 College Avenue
Ft. Lauderdale, FL-33314 USA
Tel: +1 954-577-6392 Fax: +1 954-475-4125
Home advisor: Hartwig H. Hochmair

UF | UNIVERSITY of FLORIDA

Carinthia University of Applied Sciences
Department of Geoinformation and Environmental Technologies
Europastrassa 4
9524 Villach, Carinthia Austria
Tel: +43 059-0500-7700
Local advisor: Gernot Paulus

KÄRNTEN
University of
Applied Sciences

Acknowledgements: Special thanks to Dariia Strelnikova for assistance with data extraction from several sources, to Levente Juhász for answering endless programming questions, to Alexandra Böösi in the international office at FH for her support and assistance during my stay in Villach, to my advisors Gernot Paulus and Hartwig Hochmair for their guidance and advice throughout this project and without whom this project would not have been possible, and to the Marshall Plan Foundation for financing this study abroad project.

Title page………………………………………………………………………………. 1

Table of contents…………………………………………………………………………. 2

List of abbreviations…………………………………………………………….……. 4

Abstract………………………………………………………………………………… 5

Keywords………………………………………………………………………………. 5

1. Introduction…………………………………………………………………………. 6

2. Related Research……………………………………………………………………… 7

2.1 VGI from smartphones in research……………………………………………….. 7

2.1.1 Route preferences……………………………………………………………….. 8

2.1.2 Spatial-temporal analysis……………………………………………….………… 8

2.1.3 Popularity of locations among cyclists…………………………………………… 8

2.1.4 Park Popularity for fitness activities…………………………………………….. 8

2.2 Quality assessment of VGI in research……………………………………………... 9

2.2.1 Manual cycling counts…………………………………………………………. 9

2.2.2 Travel surveys……………………………………………………………………... 9

2.2.3 Cycling volume………………………………………………………………… 10

2.2.4 Comparison of GPS data sources……………………………………………… 10

3. Data extraction……………………………………………………………………… 10

3.1 Cycling path characteristics (objectives 1 and 2)……………….……………………..10

3.1.1 Bikemap…………………………………………………………………………10

3.1.2 Endomondo……………………………………………………………………….. 10

3.1.3 MapMyFitness…………………………………………………………………. 11

3.1.4 Reference routes……………………………………………………………….. 11

3.1.5 Correcting GPS errors…………………………………………………………. 12

3.1.6 Trip characteristics…………………………………………………………….. 13

3.2 Park popularity (objectives 3 and 4)……………………………………………… 16

3.2.1 AllTrails…………………………………………………………………………... 16

3.2.2 MapMyFitness…………………………………………………………………. 17

3.2.3 Wikiloc…………………………………………………………………………. 17

3.2.4 Park characteristics…………………………………………………………….... 18

4. Data analysis……………………………………………………………………………… 21

4.1 Cycling path characteristics (objectives 1 and 2)…………………………………........ 21

4.2 Park popularity (objectives 3 and 4)……………………………………………........ 21

5. Results…………………………………………………………………………….…...... 22

5.1 Comparison of trip characteristics between South Florida and North Holland……... 22

5.2 Comparison of trip characteristics within South Florida and North Holland…….…... 25

5.2.1 Trips from each App compared to control………………………………………... 25

5.2.2 Trips from each app compared to each other……………………………….…..... 28

5.3 ***Characteristics of Florida State Parks that attract various fitness activities***………… 29

5.3.1 Hiking models………………………………………………………………….. 29

5.3.2 Cycling models………………………………………………………………… 32

5.3.3 Running models……………………………………………………………….... 34

5.3.4 Paddle-sport models…………………………………………………………… 37

5.4 Differences among apps used and physical activity tracked………………………….. 37

6. Discussion…………………………………………………………………………….…....38

7. Limitations and future work.......…………………………………………..……………… 41

7.1 Cycling path characteristics……………………………………………………………41

7.2 Park Popularity………………………………………………………………………… 41

8. Journal Submissions……………………………………………………………………... 42

9. References……………………………………………………………………………… 43

List of Abbreviations

AIC – Akaike Information Criterion

ACS – American Community Survey

API – Application Programming Interface

EC – European Commission

ECC – European Cycling Challenge

FDEP – Florida Department of Environmental Protection

FDEPGOD – Florida Department of Environmental Protection Geospatial Open Data

FDOT – Florida Department of Transportation

GPS – Global Positioning System

GPX – GPS Exchange

HTML – Hyper Text Markup Language

NBR – Negative Binomial Regression

OD – Origin-Destination

POI – Point of Interest

VIF – Variance Inflation Factor

VGI – Volunteer Geographic Information

Abstract

The benefits of physical activity are well documented, and parks play a key role in helping people stay physically fit. The Parks and Recreation department of the Florida Department of Environmental Protection initiated a program called *Florida Park Fit* aimed at attracting residents to State Parks for physical activity. Fitness tracker devices are a popular way for individuals to track their physical activity. Data from fitness tracker apps have become a prominent source for studying physical activity and cycling behavior. This type of crowd-sourced data provides larger datasets than were previously available and allows for the comparison of trip characteristics between geographic regions, and the study of temporal trends in bicycle ridership. The first portion of this research makes a comparison of trip characteristics among the three fitness tracker apps Bikemap, Endomondo, and MapMyFitness across two study regions, South Florida (Miami-Dade, Broward, and Palm Beach counties) and North Holland. The second portion uses Volunteer Geographic Information (VGI) from the fitness tracker apps AllTrails, MapMyFitness, and Wikiloc to assess which characteristics of Florida State Parks attract visitors for the four fitness activities, hiking, cycling, running, and paddle sports, across the various apps and to determine if different apps are used to track different activities. For the first portion, results show that cycling behavior observed in the three apps Bikemap, Endomondo, and MapMyFitness are similar relative to a set of control trips in each region (e.g. fewer primary roads than reference trips observed), but that there are some pronounced differences in trips recorded with the different apps between both regions. For example, Bikemap trips were significantly longer than Endomondo trips in North Holland, whereas the opposite was true in South Florida. This suggests that geographic region plays a role in how trip characteristics recorded on different apps compare to each other, demonstrating the presence of an additional aspect of geographic bias in crowd-sourced cycling data. For the second portion, in Florida State Parks, trail density for hiking, cycling, and running and number of canoe launches and percent water landcover for paddle sports were among the factors contributing to park visits for each activity. Presence of recreational facilities, restrooms, and the population of the surrounding area also were significant predictors. Comparisons among the three apps AllTrails, MapMyFitness, and Wikiloc found a dependency between the app used to track activity and the type of activities tracked.

Keywords
cycling, parks, fitness, fitness tracker apps, VGI, crowd-sourced data, Bikemap, Endomondo, MapMyFitness, AllTrails, Wikiloc

1. Introduction

Cycling plays an important role in current governmental planning efforts around the world to reduce individual motorized traffic and to increase the share of green transportation, which includes provision of bicycle infrastructure and introducing bicycle friendly policies (Faskunger, 2013). To be able to correctly forecast and facilitate the needs of cyclists, transportation planners rely on data describing cyclists' behavior and trip counts. Understanding spatio-temporal changes in cycling volume allows planners to assess the cost-effectiveness of added cycling facilities such as cycling lanes, paths, bridges, and parking areas (Dill & Carr, 2003; Gotschi, 2011). In addition, understanding cyclists' route-choice behavior aids in accurately facilitating cyclists' needs and bicycle network planning (Larsen & El-Geneidy, 2011).

Crowd-sourced spatial data, often called Volunteered Geographical Information (VGI) (Goodchild, 2007) from fitness apps is hosted on company Web servers and available with some restrictions, but nevertheless can provide researchers with rich datasets that were hitherto inaccessible. VGI is becoming a widely used source of data in research (Boss et al., 2018; Chen et al., 2011; McArthur & Hong, 2019; Norman & Pickering, 2017, 2019; Rupi et al., 2019), however, users of fitness apps represent only a small proportion of all cyclists, hikers, runners, and paddlers (e.g. kayak/canoe/stand-up paddle board) and the users of these apps show a demographic bias towards young and middle-aged males (Blanc et al., 2016; Griffin & Jiao, 2015). In addition to biased samples, different fitness tracker apps have differing target audiences. For example, Endomondo aims to attract sports enthusiasts of all kinds and recreational athletes between the ages of 15 and 50 (Facts About Endomondo, 2020), whereas MapMyFitness is focused more toward people trying to be more active who may not necessarily consider themselves athletes (About MapMyFitness, 2020). Endomondo also includes an option to select trip type (i.e. sport or transport) that is not available on all fitness tracker apps. Bikemap is different in that it is dedicated strictly to cycling enthusiasts with the goal of sharing cycling routes with other riders (About Bikemap, 2020). Wikiloc is targeted at more serious athletes who prefer rigorous activities (Norman & Pickering, 2019) and AllTrails if focused on making available a list of trails for hikers, mountain bikers, and runners using crowd-sourced data, e.g. trail reviews and photos (About AllTrails, 2020).

The state of Florida boasts the largest State Park system in the United States, with 175 parks, reserves, conservation areas, and trails, for a total of nearly 800,000 acres (FDEP - Parks and Recreation, 2020). In 2019 the Florida State Park system recorded over 29.4 million visits with an average of more than 2.4 million visits per month. Accessibility to parks can help prevent childhood and family obesity and lead to healthy lifestyle choices, such as using modes of active transportation (Blanck et al., 2012). Though State Parks are a popular destination for outdoor activities, not all State Parks are visited equally. Park popularity is affected by many factors, including size, facilities, and availability of trails and recreational activities (Neuvonen et al., 2010; Norman & Pickering, 2019). Understanding of factors that influence park visitation and popularity is essential for planners to determine allocation of resources, promotional activities, or facility maintenance (Buckley, 2009; Norman & Pickering, 2019). Using traditional means of obtaining up-to-date information about park visitation rates and activity patterns of visitors, such as surveys, interviews, or counters, are time consuming and costly (Di Minin et al., 2015). This led to the exploitation of VGI for this purpose (Wood et al., 2013), including social media data (Zhang & Zhou, 2018) and geo-tagged images (Heikinheimo et al., 2017).

Smartphones and wearable fitness devices, such as smart watches, are increasingly being used by individuals to keep track of their fitness activities and to challenge themselves to be more active by forming a social community of users. Besides the benefits to their users, selected GPS fitness tracker apps also provide researchers with crowd-sourced data about different types physical activities taking place in parks. Among the many fitness tracker apps available, Endomondo, MapMyFitness, Strava have the highest user numbers (Romanillos et al., 2016), followed by others such as AllTrails, Wikiloc, and Runkeeper. Data from GPS fitness tracker apps have been used to assess various aspects of visitor behavior in different parks and conservation areas of the United States and around the world (Neuvonen et al., 2010; Norman & Pickering, 2019; Santos et al., 2016; Schägner et al., 2016), although not for Florida. Only few studies analyze and compare user data from different GPS tracker apps, e.g. for cycling on urban road networks (Watkins et al., 2016) or for assessing the popularity of mountain biking, walking and running in national parks (Norman & Pickering, 2019).

As VGI becomes more prevalent in cycling research it is important to ascertain whether data obtained from different apps could introduce bias in the results. Only very limited research has been conducted so far to test for differences in trip characteristics among different fitness tracker apps (Watkins et al., 2016). With such differences present, using data from only one source in assessing travel behavior may be biased towards users of a particular app, which could be mitigated by the combination of data from various sources. Data from GPS fitness tracker apps, as well as surveys, and visitor counts have been used to assess various aspects of park visitation (Neuvonen et al., 2010; Norman & Pickering, 2019; Santos et al., 2016; Schägner et al., 2016), however there is lack of research for parks in the state of Florida. Therefore, this research has the following objectives.

(1) Analyze cycling trip data from Bikemap, Endomondo, and MapMyFitness for consistency of the observed travel patterns in the three apps across the two study regions (South Florida and North Holland).
(2) Identify differences in bicycle ridership patterns among the three apps Bikemap, Endomondo, and MapMyFitness within the two study regions.
(3) Analyze visit counts from AllTrails, MapMyFitness, and Wikiloc to assess which characteristics of Florida State Parks (e.g. proximity to major cities, supply of recreational trails, landcover) attract which activities (hiking, cycling, running, paddle sports) on the various platforms.
(4) Identify differences in activities tracked among the apps AllTrails, MapMyFitness, and Wikiloc in Florida State Parks.

2. Related Research

2.1 VGI from smartphones in research

Due to the availability of larger datasets through smartphone GPS tracking apps, researchers have begun to use these data to learn more about cycling behavior, including route choice and cycling volumes, in various cities and regions around the world (Boss et al., 2018; McArthur & Hong, 2019; Rupi et al., 2019).

2.1.1 Route preferences

In Glasgow, Scotland, the analysis of data from the Strava fitness app showed that commute cyclists frequented a cycling path along the river even though the distance traveled was much longer than the shortest route (McArthur & Hong, 2019). In addition, the researchers noted that cyclists were willing to ride up to 8.2% longer distances, compared to the shortest path, to stay on roads that had some sort of bicycling infrastructure, e.g. bike lanes or paths.

2.1.2 Spatial-temporal analysis

In Ottawa, Canada, crowd-sourced ridership data from Strava was used to evaluate the impact of cycling infrastructure change on spatial-temporal ridership patterns (Boss et al., 2018). The researchers used data separated in time by one year to map the differences in ridership as related to the changes in cycling infrastructure in the area. Generally, it was noted that there were significant differences in ridership based on infrastructure change, for example, the addition of a pedestrian/bicycle bridge showed a change in ridership over time with increased ridership not only on the bridge, but also on roads and paths surrounding the bridge. These changes also resulted in decreased ridership on less desirable network elements, e.g. a bridge shared with cars.

2.1.3 Popularity of locations among cyclists

Location based data from GPS tracking apps have been used to determine the popularity of locations among cyclists and other athletes. For example, to help tourists who need directions in unfamiliar areas, GPS trajectories for driving, hiking, cycling, etc. were used to build an algorithm for determining the most popular route between two locations (Chen et al., 2011). So, instead of the fastest route or shortest path given by other routing engines, cyclists, pedestrians, and tourists in general can find the route most popular among other tourists for their particular activity. VGI from GPSies (now part of AllTrails), MapMyFitness, and Wikiloc were used to assess park visitation in Queensland, Australia (Norman & Pickering, 2017). These data were found useful in predicting the relative popularity of trails within a park when compared to trail count data, but they were not useful in predicting actual counts. Differences among the apps were also noted.

2.1.4 Park Popularity for fitness activities

The benefits of physical activity are well documented, however approximately 25% of Americans remain sedentary (Brownson et al., 2001). Parks play a key role in keeping people physically active, and distance to parks, park size, and park features have been found to influence the likelihood of park usage (Zhang & Zhou, 2018). Additionally, access to parks and enjoyable scenery are factors found to promote physical activity (Brownson et al., 2001). Researchers in Beijing China, using social media check-in data from Weibo, found that neighborhood and community parks were the most popular (Zhang & Zhou, 2018). It was also noted that larger parks had more visits than small or medium sized parks. Key factors influencing park popularity were entrance fees, distance to bus stops and urban centers, and size. In Lisbon, Portugal data from GPSies and Wikiloc were used to assess mountain bike trail use and it was found that there was a difference between the trails offered by the park and where the cyclists actually road (Campelo & Nogueira Mendes, 2016) indicating the need for more trails and also policing of possible illegal trails (e.g. downhill/freeride). Norman & Pickering (2019) used VGI from

MapMyFitness, Wikiloc, and Strava to evaluate which factors influenced park popularity for cyclists, walkers, and runners. For parks that were close to urban areas, trail length was a factor that contributed to park popularity. However, for parks that were more remote, the distance between the park and urban area was an influencing factor. The researchers also found a significant difference in the rigor of activities between Wikiloc users and users of the other two apps, with users of Wikiloc preferring more strenuous activities. Data from GPSies was used to assess mountain bikers and runners in a major urban park in Lisbon, Portugal (Santos et al., 2016). Results indicate that mountain bikers and runners prefer unofficial trails, with only 51% and 64% respectively using official park trails. Use of parks for fitness activities engenders potential conflicts between pedestrians and cyclists (Santos et al., 2016). According to Santos et al., conflicts increase under certain trail conditions such as slope or terrain and in accordance with cyclist's speed or behavior, in addition to high potential overlap in trail usage. Factors contributing to park visitation in general include proximity to urban area, presence of water bodies, availability of trails, landcover diversity, and population in surrounding areas, (Norman & Pickering, 2017, 2019; Schägner et al., 2016). MapMyFitness, GPSies, and Wikiloc were compared for suitability to assess park visitation (Norman & Pickering, 2017). Because the three platforms differ, use of only one platform for assessing park visitation will not be indicative of all users or park types. MapMyFitness, as a fitness-oriented app, was determined to be most useful in comparing relative trail usage in urban parks, whereas Wikiloc was more useful in parks that were more remote.

2.2 Quality assessment of VGI in research

There has been some research into the suitability of using VGI for assessing cycling behavior, cycling volumes, and route choice**.** VGI have been compared to manual counts and cycling surveys to determine if the results obtained using VGI differ from those of data collected by other means (Blanc et al., 2016; Jestico et al., 2016).

2.2.1 Manual cycling counts

In Victoria BC, Canada, results from manual cycling counts were compared to VGI from Strava to determine the appropriateness of using VGI to measure spatial and temporal differences in ridership patterns (Jestico et al., 2016). The counts from Strava were matched to the timeframe and road segments of manual counts for comparison. The conclusion was, in part, that the data from Strava did well in modeling the cycling volume and a strong linear correlation was found between Strava volumes and the manual count. In Bologna, Italy researchers compared map-matched GPS traces obtained from participants in the European Cycling Challenge (ECC) with manual cycling volume counts (Rupi et al., 2019). The data from the ECC were selected to coincide with the times and places where the manual counts were taken. A linear regression model showed a strong correlation between the manually counted cycling volumes and those obtained from the ECC data, with $r^2 = 0.7264$.

2.2.2 Travel surveys

Data from travel surveys in seven different regions were obtained and commuters whose transportation mode was primarily cycling were considered in a comparison study by Blanc et al. (2016). A significant difference was noted in the demographics of smartphone app users and those of the travel survey sample. The researchers conclude that data obtained from travel surveys are more representative of the population than that obtained from smartphone apps. However, data from smartphone apps allow for larger sample sizes and could be useful for studies in areas where travel survey sample sizes are small.

2.2.3 Cycling volume

In Brisbane Australia Strava data were used to evaluate change in cycling volume after infrastructure changes (Heesch & Langdon, 2016). Traffic counts where no infrastructure changes occurred were used as a basis for comparison. Because the number of Strava users change over time and vary with location, it was determined that, although Strava data are useful in determining the impact of cycling infrastructure at a single location within a short timeframe, adjustment for differences in Strava use would be necessary to make inferences across locations or for longer time intervals.

2.2.4 Comparison of GPS data sources

Watkins et al. (2016) compared data from Strava with the city's own smartphone app, called *Cycle Atlanta*, in the city of Atlanta, GA. Some similarities and differences were reported. Both apps had predominately male cyclists, however Strava users were slightly older than those of Cycle Atlanta. More Cycle Atlanta users cycled for commute than Strava, and Strava users cycled longer distances than users of Cycle Atlanta.

3. Data extraction

3.1 Cycling path characteristics (objectives 1 and 2)

3.1.1 Bikemap

Since Bikemap does not provide an Application Programing Interface (API), a python script was written that searches through trip IDs sequentially on the Bikemap website and identifies trips located within Florida or North Holland, respectively. Since there were over five million trip IDs to search, copies of the program ran through different ID intervals until a sufficient sample size was obtained. The Florida trips were then parsed by city, where IDs of trips falling in the Tri-county study area (Palm Beach, Broward, and Miami-Dade counties) were retained for further processing. A similar approach was taken for North Holland. Next, trip data, including GPS points, were obtained for each trip ID in both regions by parsing the HTML Website of the searched trip. Only trips with a minimum length of 1 km and falling entirely inside the respective study areas were retained. The extracted GPS points were written to a file in GPS Exchange (GPX) format and then uploaded to a PostgreSQL database. This resulted in 596 trips from 232 unique users for the South Florida study area and 1114 trips from 560 unique users for North Holland.

3.1.2 Endomondo

Endomondo facilitates tracking of about 70 activity categories, including the four cycling activity types indoor, sport, transport (similar to commute) and mountain biking (off-road cycling). Endomondo was chosen as one of the sources for analysis since until recently it allowed access to a large volume of GPS tracking point data representing individual bicycle trips, as described in (Strelnikova, 2017). Currently, the default status of Endomondo activities is

"Share with Friends", which reduces the number of public activities that can be extracted. Therefore, for this study worldwide Endomondo cycling trips from January 2018 to February 2019 that were downloaded before the change in default status were used. The approach chosen returns all workout IDs for a specified user within a chosen time range, which is followed by GPS point downloads for these workouts. A minimum trip length of 1 km was used, which led to 841 trips from 298 unique users in the South Florida study area and 1182 trips from 362 unique users for North Holland. Out of the 298 South Florida users, 272 users provided gender information in their user profiles. From these, 206 (75.7%) users were male, which is slightly lower than the 77.9% of males among Miami-Dade residents who commute by bicycle. For North Holland, 343 out of 362 users provided gender information, of which 263 (76.7%) were male. Furthermore, 188 Endomondo users in South Florida provided a date of birth in their profiles, which resulted in a median age of 47 years (mean = 46.5, min = 22, max = 71) at the time of travel. For North Holland, the 252 Endomondo users who reported their birth date had a median age of 41 years (mean = 43.2, min = 19, max = 78) at the time of travel.

3.1.3 MapMyFitness

MapMyFitness data were extracted through the website's API using a python script. To obtain data from each of the two study areas, the API's "close_to_location" parameter was used to extract data within a given search radius of a latitude-longitude location. Since the maximum allowable search radius is 50 km, multiple locations were chosen so as to cover the entire study area. As with Bikemap and Endomondo, the search parameters also included a minimum trip length of 1 km. The extracted trip points were written directly to a PostgreSQL database and duplicate trip IDs (due to overlapping search areas) were eliminated from the results. Also, any trips not completely contained within the study area boundaries were eliminated. Using this approach, an initial set of 7474 MapMyFitness trips from 2255 unique users was obtained within the South Florida study area and 2833 trips from 604 unique users for North Holland. These numbers represent a random sample of all available trips from MapMyFitness in each study area.

In order to eliminate potential statistical bias of trip characteristics caused by highly active individual cyclists on any of the three platforms, a subset of trips was selected from the downloaded trips in such a way as to obtain at most two trips per user. This selection resulted in 314 trips from Bikemap, 429 trips from Endomondo, and 3189 trips from MapMyFitness for South Florida, and 775 trips from Bikemap, 537 trips from Endomondo, and 874 trips from MapMyFitness for North Holland (Table 1).

3.1.4 Reference routes

In order to control for differences in road supply and network structure when comparing Bikemap, Endomondo, and MapMyFitness trips within and between the two study areas, a set of reference routes (control) was prepared for each study area based on origin and destination (OD) points from the three apps that were at least one km apart to reflect more realistic bicycle trips. Applying this 1-km criterion, Bikemap had the smallest dataset for South Florida with 147 OD pairs. Therefore, to balance the contribution from each app, an equal number of OD pairs was randomly selected from the routes of the other two apps. This resulted in 441 OD pairs for South Florida. Reference routes were then computed as the shortest route on the OSM street network between each OD pair using ArcMap's Network Analyst. Though the shortest route is typically not the chosen one, it provides useful statistics about the type of road supply and surrounding features available in the study area and was therefore used as the reference dataset for South Florida. For North Holland Endomondo had the smallest trip set (196 OD pairs) that satisfied the 1-km criterion. Therefore, 196 OD pairs were also randomly selected from MapMyFitness and Bikemap, followed by the shortest path calculation, which resulted in a reference dataset of 588 routes for North Holland. Attributes for these reference routes were extracted in the same manner as for the observed routes from the three apps.

3.1.5 Correcting GPS errors

Each trip downloaded from Bikemap, Endomondo, or MapMyFitness, was examined for geometrical outliers due to possible GPS errors. Signal loss could lead to an erratic behavior as pictured in Figure 1a. Where possible, trips with the erratic behavior were fixed (*Figure 1b*). Other trips where tracking points were missing so that the traveled path could not be reconstructed, or trips that crossed waterbodies or ran on prohibited roads were eliminated.

a) Erratic behavior b) Corrected path

Figure 1 Erratic behavior with loss of GPS signal (a) and the corrected trip path (b).

Table 1 summarizes statistics related to trip download, processing, and analysis. Between about 2% and 6% of trips were removed due to reasons described above. For example, in South Florida 146 trips were eliminated from MapMyFitness for dropped GPS signals, and 32 trips were eliminated for having unclear paths or short repeated segments, e.g. traveling back and forth in front of a few houses.

Figure 2 depicts the trips that were retained for further analysis in each study area.

Table 1 Summary of cycling trips analyzed for trip characteristics.

	South Florida			North Holland		
	Bikemap	**Endomondo**	**MapMyFitness**	**Bikemap**	**Endomondo**	**MapMyFitness**
Trips extracted	596	841	7474	1114	1182	2833
Trips from max 2 per user	314	429	3189	775	537	874
Trips eliminated	12	12	178	9	7	48
Edits made	69	75	243	118	47	70
Trips analyzed	**302**	**417**	**3011**	**766**	**530**	**826**
Users of trips analyzed	232	298	2255	560	366	604

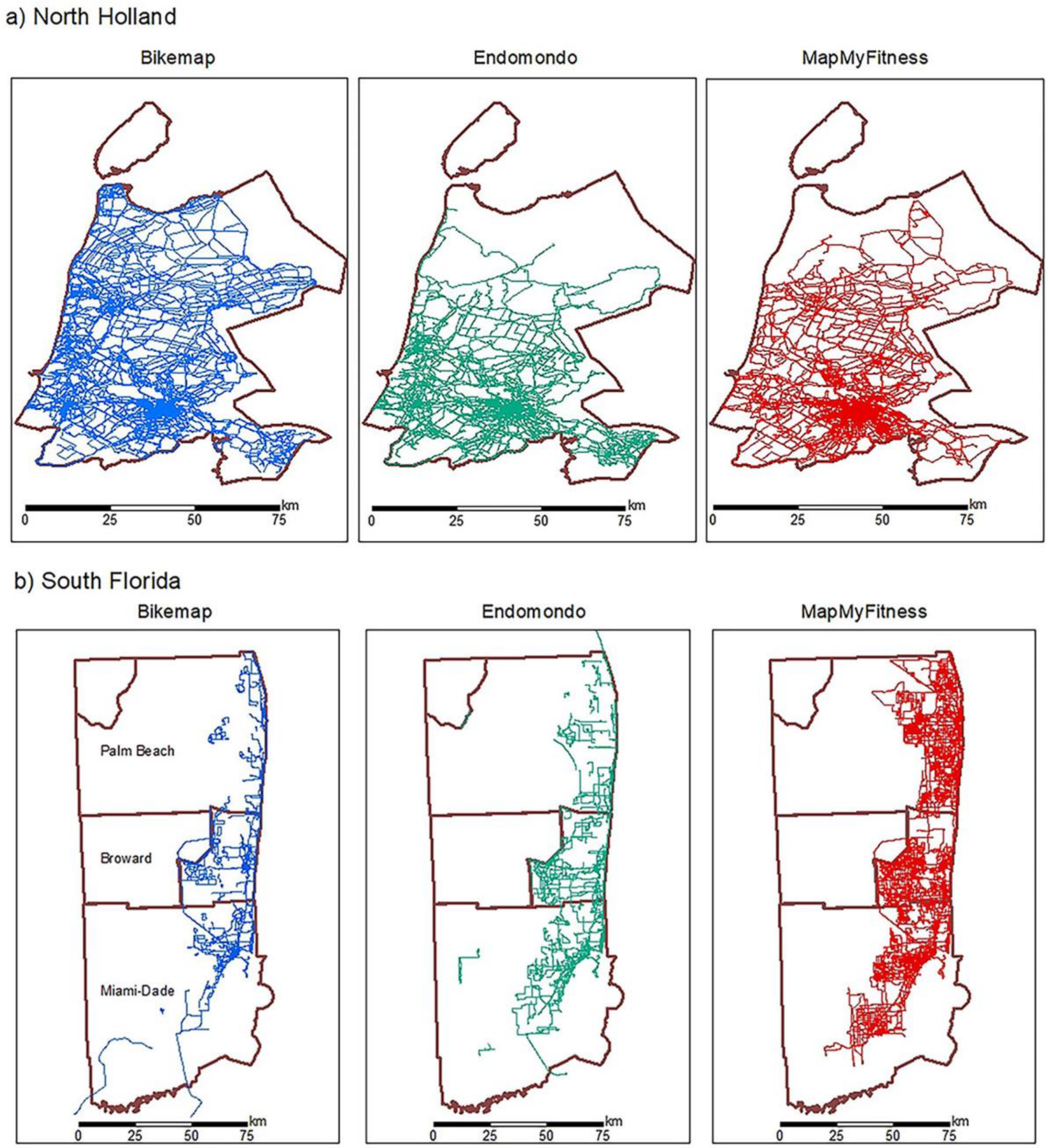


Figure 2 Trips analyzed from the various sources in the two study areas North Holland (a), and South Florida (b).

3.1.6 Trip characteristics

Customized python scripts were used to extract the trip characteristics from route geometries and their surrounding environment as described below. In addition to the trip characteristics described below, time of day, average speed, and trip duration were considered, but this information was not publicly available for Bikemap and MapMyFitness and therefore was not included.

Trip length. The trip length was taken from the geometric length of the trip polyline.

Shape ratio. To determine the circuitousness of a trip, both study areas were divided into 500 m × 500 m square grids. The shape ratio of a route was then calculated by dividing the number of grids the trip traversed by the length of the trip in km. A larger shape ratio indicates that the trip was more spread out, while a smaller ratio means the trip was more circuitous (Figure 3).

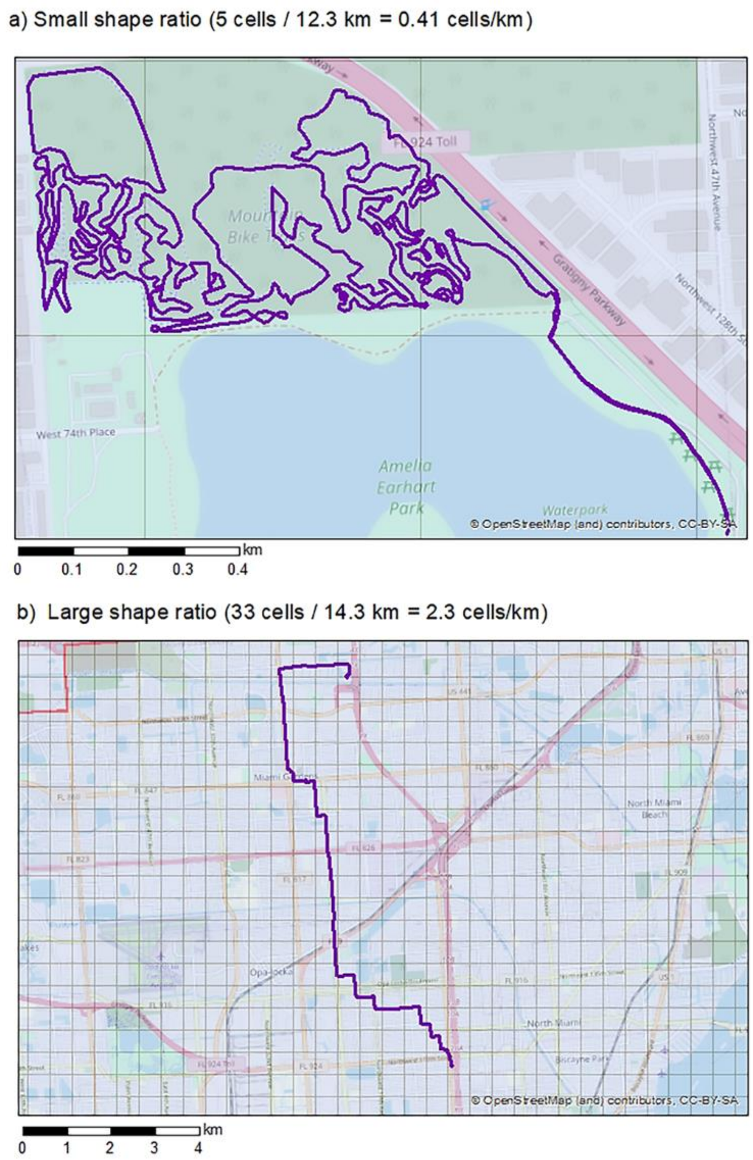


Figure 3 Difference between trips with a small shape ratio (a) and a large shape ratio (b).

Turns. To calculate the number of turns the angle at each vertex was determined. An angle of 40° or more (off a straight line) was considered a possible turn if it was within 30 m of a road network junction. The 30 m buffer was used to encompass the width of the road and to allow for GPS positioning errors along the route. To eliminate false turns, a possible turn was only counted if the previous vertex did not satisfy both criteria or if it was more than 30 m away from the current vertex. The number of turns, right turns, and left turns for a trip were divided by the trip length to determine the number of turns, right turns, and left turns per km.

Traffic signals. Traffic signal point data for each study area was downloaded from OSM and aggregated at intersections so that there was at most one traffic signal point at each network junction. The number of traffic signals located within a 30 m buffer around a trip polyline was divided by the trip length in km to determine the number of traffic signals per km.

Scenery. Flickr is a website where users upload and share geo-tagged pictures, primarily of interesting places. The density of images is higher along scenic routes (Alivand et al., 2015), and Flickr images can therefore also be used to compute interesting routes between two locations (Sun et al., 2015). To evaluate the scenery along the route, a 100 m buffer was placed around each trip and the number of Flickr images in the buffer was divided by the trip length in km to determine the number of Flickr images per km.

Population density. Global population density data from 2015 with nine arcsec (or approximately 270 m) resolution was obtained from the European Commission (see Table 2 for a list of all layer sources) as a raster layer and converted to polygons. Each trip path was intersected with the population layer and a weighted mean of population along the trip path was calculated.

Road type. Each trip path was split into 25 m long segments and each segment was snapped to the closest OSM network element. From this the percentage of the trip on each evaluated road type was calculated.

Land use. For South Florida, land use data was obtained from the Florida Department of Environmental Protection (FDEP) and for North Holland OSM was used for that purpose. It was not possible to use the same source for both regions as the OSM data did not sufficiently cover South Florida, and other sources of global land use lacked sufficient coverage for one of the two study areas. Only land use categories that were available on both sources (commercial, farm, forest, open land, industrial, recreational, residential) were compared. To calculate percent land use for each evaluated land use category, a 100 m buffer was placed around each trip polyline and intersected with the land use layer.

Table 2 List of sources for layer files used in this study.

Layer	Source	URL
State Park Features (park boundaries, park trails, park points of interest, park entrances, landcover)	Florida Department of Environmental Protection Geospatial Open Data Portal	https://geodata.dep.state.fl.us/
Major US cities	ESRI	https://www.arcgis.com/home/item.html?id=4e02a13f5ec6412bb56bd8d3dadd59dd
US airports	Natural Earth	https://www.naturalearthdata.com/downloads/10m-cultural-vectors/airports/
Florida cities	Official State of Florida Geographical Data Portal	http://geodata.myflorida.com/datasets/00a14c72b1034ab9b09dffbeb41e5304_8
Bays and oceans	U.S. Census Bureau TIGER/Line	https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-geodatabase-file.html
Census data	U.S. Census American Community Survey	https://www.census.gov/programs-surveys/acs/data.html
Population data	European Commission	https://ghsl.jrc.ec.europa.eu/ghs_pop2019.php

3.2 Park popularity (objectives 3 and 4)

Counts for park visitations for the activities hiking, cycling, running, and paddle sports were extracted from AllTrails, MapMyFitness, and Wikiloc using HTML parser or available APIs as described below.

3.2.1 AllTrails

AllTrails provides a search function that allows a user to enter a geographic term upon which matching trails are returned in a list together with activity recordings and trail reviews. The Website also provides a pre-set filter to return activity recordings for Florida State Parks (https://www.alltrails.com/us/florida/state-parks) or National Parks. When the user finds a satisfactory trail he/she can record his/her activity on the trail and link the recorded activity to the trail with the option to choose the activity undertaken (e.g. hiking, mountain biking, running) after the trail is completed. The user may also write a review of the trail and upload photos, both of which are displayed separately on the AllTrails website.

The Website does not offer an API, therefore activity counts were extracted using a customized HTML parser developed in Python. At the time of extraction there were 8040 activities recorded by users in Florida State Parks. Each record comes with one of 20 activity types from which a user can select when adding a new trail record. After the recordings of trails in State Parks were extracted, the tagged activities were filtered for the analyzed activities (hiking, cycling, running, paddling). For this purpose, activities tagged as backpacking, hiking, or walking were grouped in the hiking category. Activities tagged road-biking, mountain-biking, or tour-biking were grouped in the cycling category, and those labeled trail-running and paddle-sports were assigned to the running and paddling categories, respectively. Untagged activities or activities from categories not evaluated were excluded from further analysis. The GPS coordinates of the starting location of each recorded trip were extracted from the trip record link by parsing the link with the customized HTML parser and were then converted into a point feature class. Point features were subsequently intersected with the State Park polygons to ensure all recorded trips initiated within a State Park. Trips that did not initiate in a State Park were eliminated from analysis.

3.2.2 MapMyFitness

Unlike AllTrails, MapMyFitness trip recordings are not linked to a specific trail or location. Instead, MapMyFitness users choose their activity (e.g. cycling, running) prior to recording from an extensive list of available activities and then record their activity. Data regarding the activity (e.g. trip length, trip duration) as well as GPS points for the trip path are uploaded to the MapMyFitness website upon trip completion. MapMyFitness activity counts were extracted using the website's API through a Python script. The API's "close_to_location" function enables data extraction of activities within a given search radius around a location specified by geographic latitude and longitude. The function was used to extract the starting location (latitude-longitude) of all hiking, cycling, running, and paddling activities in State Parks. Hiking and walking activities were grouped in the hiking category and canoeing and rowing activities grouped in the paddling category. Activities labeled running and cycling went into their respective categories. The GPS coordinates for the starting location of each trip were intersected with the park boundaries to obtain activity counts per State Park.

3.2.3 Wikiloc

In Wikiloc users can pick their activity prior to recording from a list of 75 activities, which includes choices like off-roading and motorcycling as well as non-motorized activities (e.g. hiking, cycling, canoeing). The recorded trip information that is uploaded to the Wikiloc website includes GPS tracking points as well. Users can choose a trail uploaded by other users to follow and record the trail, but unlike AllTrails, the recording is not linked to the trail on the Wikiloc website. In order to extract activity counts from Wikiloc, a Python HTML parser program was used. The Florida trails were found by zooming to the state on a map on the Wikiloc website. Using that URL as the starting page, a Python program looped through the subsequent pages of search results to extract the URLs of all recorded Florida trips. The extracted URLs were then parsed and the GPS starting coordinates were obtained from the query response for all trips available in the state of Florida in the hiking/walking, mountain-biking/cycling, running, and kayaking-canoeing categories were downloaded. The GPS coordinates for the starting location of each trip were converted to a point feature class and intersected with the park boundaries in order to ensure all trips originated within a State Park.

Table 3 shows the total number of activities initially extracted for the different platforms, and the number of activities that satisfied the requirements and were subsequently used, i.e. trip starting point in a State Park and trip assigned to one of the four assessed activities. The high numbers of activities extracted relative to those finally used in analysis for MapMyFitness and Wikiloc is because for these platforms data extraction was achieved through geometric queries in which it was necessary to extract trips from wide areas outside of parks in order to ensure that all trips within the parks were counted.

Table 3 State Park visits extracted from each app and visits for the four activities hiking, cycling, running, and paddle sports that were in State Parks.

	AllTrails	MapMyFitness	Wikiloc
Visits extracted	8040	55,616	15,228
Hike visits	5285	1002	564
Cycle visits	463	1200	192
Run visits	61	1116	38
Paddle visits	50	22	215
Total visits analyzed	5859	3340	1009

3.2.4 Park characteristics

Analyzed parks. The State Park boundary shape file was downloaded from FDEP Geospatial Open Data (FDEPGOD) and added to an ArcGIS Pro map document. State Trails (not in a park) were deleted from the data as well as some parks where park entrance information was not available. In total 161 State Parks, Reserves, and Conservation Areas were retained for analysis (Figure 4).

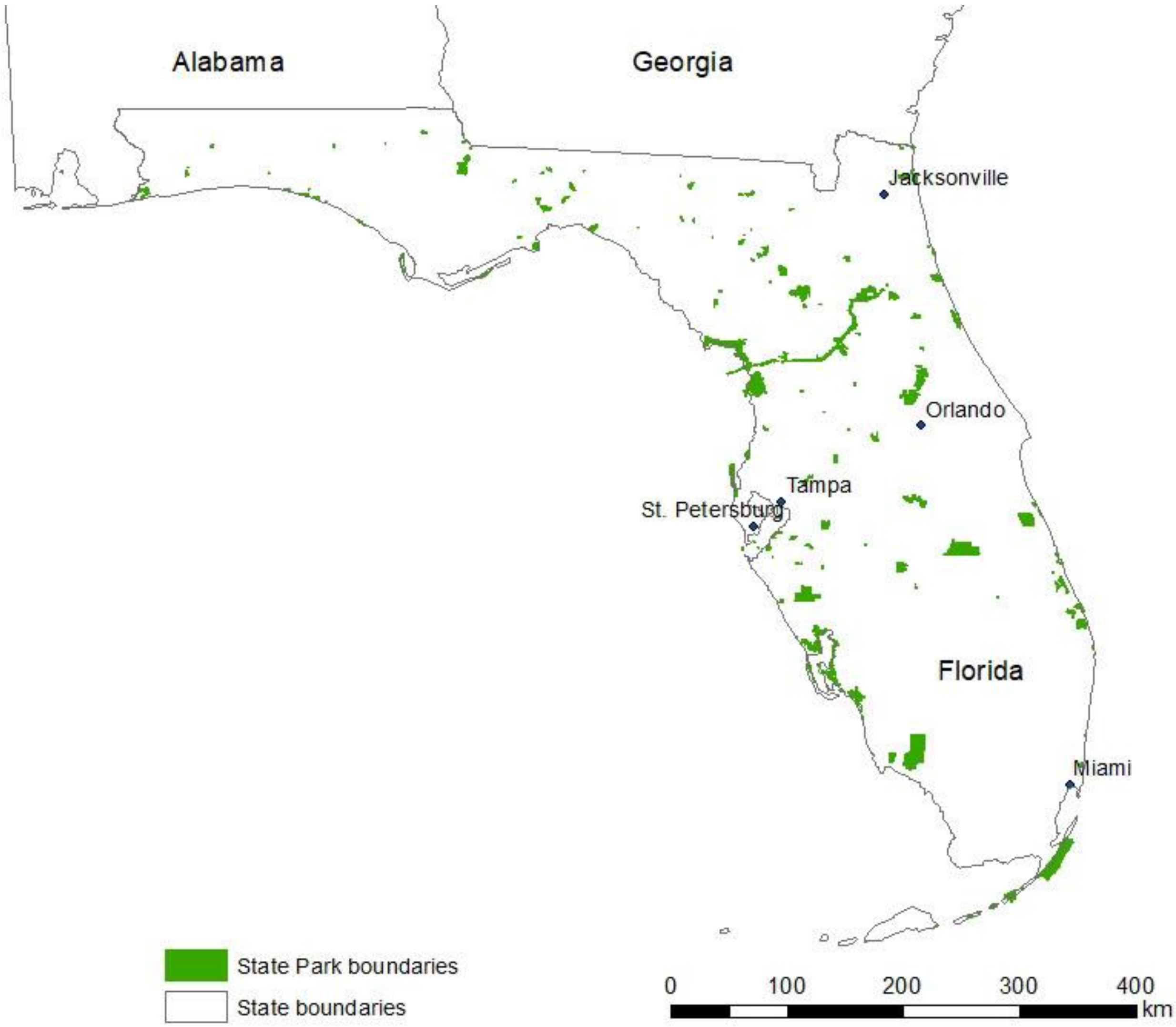


Figure 4 Boundaries of analyzed Florida State Parks.

Landcover, points of interest, and park features. Most sport activities in parks take place on trails or on areas of selected land cover types (e.g. rivers), and park facilities (POIs) provide the necessary infrastructure or comfort to conduct these sport activities (e.g. use of canoe/kayak launch areas for paddling). Therefore, we hypothesize that the percentage of certain landcover types, the density of trail types, and POI counts are positively correlated with the number of sport activities of some type. Statewide landcover data, which was obtained from the FDEP website as a polygon vector layer, was intersected with State Park boundaries to obtain percent landcover of various categories in each park. Out of the seven landcover categories available, landcover categories evaluated in this study were urban (consisting of mainly recreational facilities in parks), agricultural, rangeland, water, wetlands, and forest. Relevant POIs that were analyzed include number of restrooms, number of canoe launches, and number of campgrounds. Trail density was calculated for hiking trails, biking trails, combined hiking/biking trails, paddling trails, and nature trails by dividing the total length of trails (in km) for each category in the park by the area of the park in km^2. Park trails and water landcover features can be seen in Figure 5 for Rock Springs Run and Wekiva Springs State Parks as an example. Trails that including the "Equestrian" tag besides other tags were categorized by their other tag(s) for the analysis. For example, hiking/biking/equestrian was considered a combined hiking/biking trail. Information about entry fees charged was extracted from the State Parks boundary layer.

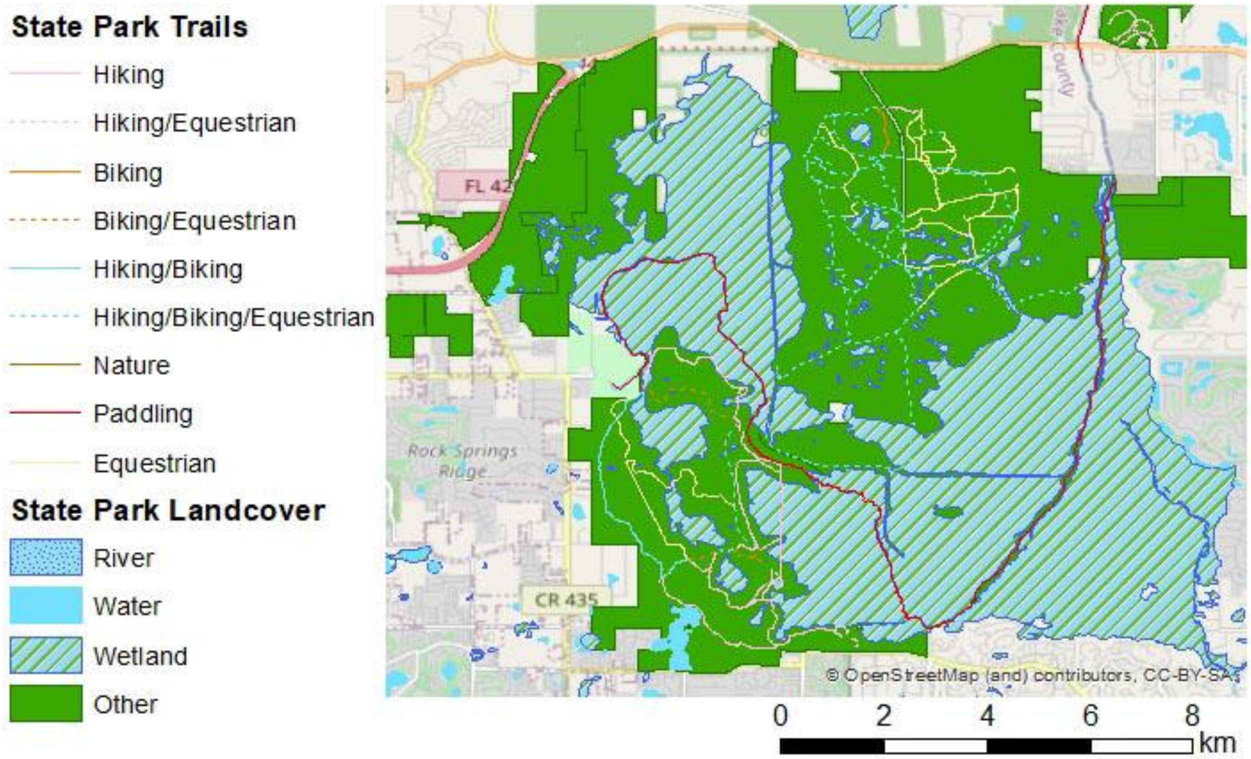


Figure 5 Trails and landcover in Rock Springs Run and Wekiva Springs State Parks

Drivetime to locations. A road network was created with HERE NAVSTREETS road data for the states of Florida, Alabama, and Georgia using ESRI's ArcGIS Network Analyst. Alabama and Georgia were included in the road network since many Florida State Parks are located near the Florida state border, and service areas around these parks reach into the neighboring states. Drivetime for road segments was computed as road segment length divided by average speed (taken from the speed category in NAVSTREETS). Furthermore, using a bay and ocean layer and a Florida cities layer, all Florida cities that touched a bay or ocean (coastal cities) were identified. Next, the ESRI Network Analyst was used to calculate the drivetime from each park's main entrance to the nearest large city (population greater than 10,000), the nearest international or regional airport, and the nearest coastal city, using the "Find closest facilities" function. All these attributes were used to consider potential park visitors in different ways. We hypothesize that a higher number of potential park visitors nearby is associated with higher sport activity numbers observed in a park. Furthermore, the number of other State Parks located within a 30-minute drive time from the centroid of each park was determined. A higher number of parks nearby means a cluster of recreational opportunities which increases visibility and is therefore expected to attract more sport activities in a park.

Demographics. Using the main entrance, a 30-minute drivetime service area was created for each park. Census block group geometries containing census data from the American Community Survey (ACS) were intersected with the service area for each park and percent White, percent Black, and percent Hispanic population were determined. Based on previous literature, it was hypothesized that parks in areas of higher percent of White population would be associated with more sport activities, but that the opposite was true for parks in areas with a higher percentage of Black or Hispanic population (Lawton & Weaver, 2008). Population data from the European Commission were intersected with the same 30-minute service areas to determine the total population nearby each park. Furthermore, median age and population of the nearest large city were also calculated for each park.

4. Data analysis

4.1 Cycling path characteristics (objectives 1 and 2)

Descriptive statistics for each app by study area were obtained and selected density plots were used to visualize the distributions of the trip variables. Since all variables tested were non-normally distributed (Figures 5 and 6), non-parametric tests were used to test if the two trip samples compared were likely derived from the same population. Trip characteristics were compared a) between the two study areas for each app (Table 4), and b) pairwise between the four sources (three apps and control) within each study region (Table 5 and Table 6), using a Mann-Whitney U test. This test is the non-parametric equivalent to the t-test but without the assumption of normality required by the t-test. Furthermore, various maps were used to visualize observed differences in trip characteristics between the apps and study regions, respectively.

4.2 Park popularity (objectives 3 and 4)

A correlation matrix was computed for the candidate predictor variables. The number of canoe rental locations and canoe/kayak launch areas were the only variables that showed significant correlation (Pearson $r = 0.77$, $p = 0.000$) and thus number of canoe rental locations was eliminated as a predictor variable. The remaining variables showed no significant linear correlation ($r < 0.52$ for all variable pairs). Twelve separate negative binomial regression (NBR) models, i.e. one for each of the dependent variables (hike, cycle, run, paddle for each platform) were built where the best fitting set of predictor variables was determined by the lowest Akaike information criterion (AIC). The variance inflation factor (VIF) for all 12 models was low (VIF < 3.2), indicating that multicollinearity among the predictor variables did not pose a problem. No significant spatial autocorrelation ($p > 0.05$) was present in residuals for 11 models, based on a Moran's I test that used an inverse distance weighting scheme. As an exception, the best fitting Wikiloc paddling model was the only one with significant spatial autocorrelation among residuals. Therefore, the model was modified by replacing one significant predictor variable (% urban landcover) by two other variables (% forest landcover and hiking trail density), which resulted in a higher AIC value but avoided spatial autocorrelation. The effect on the model outcome was modest since the arithmetic sign and general level of significance (i.e., $p < 0.01$ or $p < 0.05$) for all other predictors remained the same compared to the best-fitting model. In addition to regression models, to check whether the proportion of the recorded number of activities falling into the four different sport types differs among the apps, a chi-square test of independence was conducted, using activity count numbers from Table 3 in a contingency table.

5. Results

5.1 Comparison of trip characteristics between South Florida and North Holland

Whereas trip characteristics associated with a specific application, such as density of traffic signals along trips in Bikemap, could be directly compared between the two analyzed regions, differences in network structure between the two study areas (e.g. higher percentage of cycleways present among network segments in North Holland compared to South Florida) should be accounted for to better understand potential usage differences of a given app between both study regions. Therefore, the comparison between the two regions was not done directly, but rather relative to the difference in the above-mentioned reference routes (control) for the two regions.

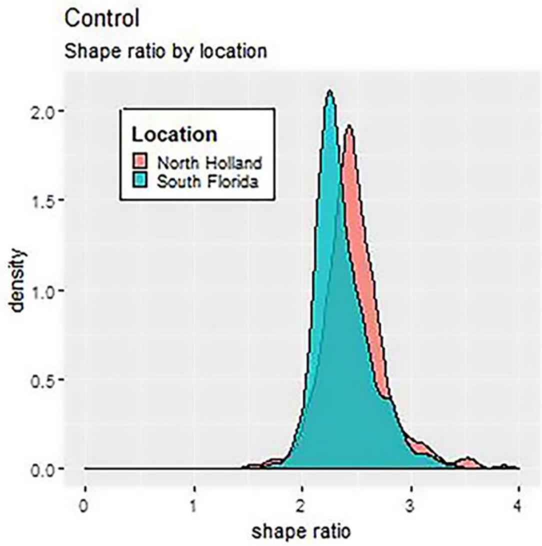
Bikemap users in North Holland undertake significantly longer trips than South Florida users (Table 4), where the difference is nearly five times what could be explained by differences in reference routes alone. As opposed to this, Endomondo trips are significantly shorter in North Holland than in South Florida, while MapMyFitness, though significantly longer in North Holland, do not differ much from reference routes with regard to trip length. In summary, the relative order of trip length between South Florida and North Holland varies between the apps.

Table 4 Median differences by location (North Holland - South Florida).

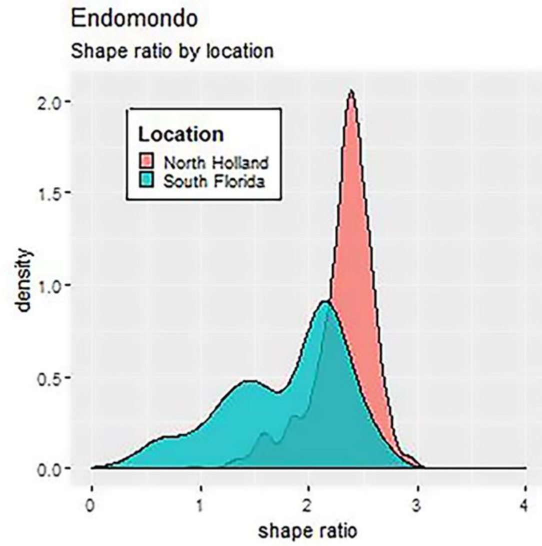
	Bikemap		Endomondo		MapMyFitness		Control	
Trip geometry								
Shape length	10249.09	***	-2340.48	**	2572.98	***	2168.20	***
Shape ratio	0.29	***	0.40	***	0.79	***	0.14	***
Turns per km	0.03		0.45	***	0.21	***	0.16	***
Turns right per km	-0.02		0.19	***	0.06	***	0.10	**
Turns left per km	0.05		0.22	***	0.13	***	0.10	**
Trip Characteristics								
Traffic signals per km	-0.10		1.08	***	1.52	***	2.35	***
Flickr images per km	1.60	***	-0.65	*	2.17	***	1.22	**
Population mean	-36.05	***	24.59	***	43.79	***	50.91	***
Road Category								
% Primary road	-3.64	***	-3.43	***	-0.54	***	-14.60	***
% Secondary road	-4.93	***	-4.30	***	2.44	***	5.57	
% Tertiary road	-4.22	*	1.36	**	0.96	**	1.06	
% Residential road	-12.16	***	-5.34	***	-14.05	***	-4.66	***
% Foot/pedestrian	1.12	***	2.04	***	5.20	***	0.16	***
% Cycleway	44.55	***	38.55	***	31.54	***	10.91	***
Land Use								
% Commercial	-15.50	***	-11.96	***	-6.19	***	-18.41	***
% Farm	0.86	***	0.39	***	0.05	***	0.00	***
% Forest	5.06	***	5.41	***	4.08	***	4.71	***
% Open land	20.25	***	17.03	***	19.70	***	14.59	***
% Industrial	1.12	***	1.47	***	1.01	***	0.88	***
% Recreation	-1.87	*	-2.75	***	-2.02	**	-0.15	
% Residential	-17.95	***	-5.29		-23.33	***	-3.53	

p < 0.05 *; p < 0.01 **; p < 0.001 ***

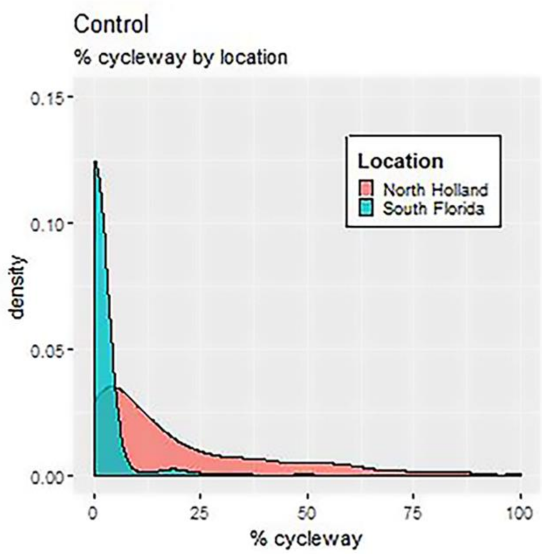
The shape ratio for all three apps is significantly larger in North Holland than South Florida. This indicates that North Holland cyclists' routes are less circuitous than those of South Florida cyclists (possibly due to many rural trips), where the difference goes beyond what can be expected based on network differences alone. The probability density plot of the shape ratio for control trips in North Holland and South Florida (Figure 6a) indicates that the shape ratio of trips in North Holland tends to be higher. For each of the apps this difference between the two regions is even more pronounced. An example is illustrated in the corresponding probability density plot for Endomondo (Figure 6b). Trips in Holland for all three apps run more frequently through farmland, forest, open land, and industrial land compared to South Florida, where in the case of farmland, open land, and industrial land this difference exceeds that of land use differences for references routes. South Florida cyclists ride in more residential areas than those of North Holland, and this difference is significant for users of Bikemap and MapMyFitness.



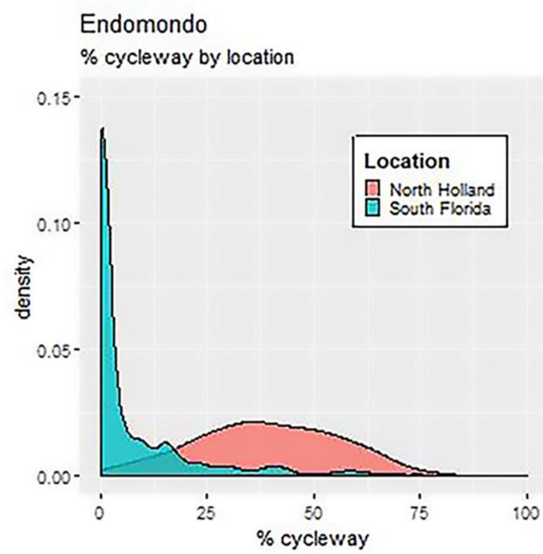
a) Shape ratio control



b) Shape ratio Endomondo



c) % cycleway control



d) % cycleway Endomondo

Figure 6 Shape ratio (a-b) and % cycleway (c-d) of South Florida and North Holland trips for control and Endomondo.

These differences in land use and shape ratio demonstrate that Florida cyclists tend to ride more circuitous paths in residential areas, whereas cyclists in Holland ride longer stretched out paths in more open areas away from residential locations. This is corroborated by the fact that the mean population density along cyclists' routes in North Holland is far less than should be, based on network differences alone, compared to South Florida (Table 4).

Although reference routes pass through more signals per km in Holland than South Florida, users of all three apps in Holland are making a greater effort to avoid traffic signals than South Florida users. In both regions, users of all apps choose routes with significantly fewer traffic signals per km than the control dataset would imply (Table 5 and Table 6, right half).

However, these differences in North Holland are between 1.8 (MapMyFitness) and 7 times (Bikemap) higher than in South Florida.

Some preferences for road type were found to hold across all three analyzed apps. For example, all three platforms indicate that South Florida cyclists use residential roads more frequently than cyclists in North Holland relative to what can be expected from difference in road supply alone. This suggests a strong preference for cyclists to use transportation network segments with little motorized traffic in South Florida, but a shortage of off-road cycleways or footpaths forces cyclists in that region to use residential roads instead. Also, control data suggest a much larger supply of primary roads for routing in South Florida (Table 4). However, trips logged by users of all three apps in South Florida include smaller proportions of primary roads relative to availability (with differences of over 10%). A possible explanation is that primary roads are considered less dangerous in North Holland where car drivers are more accustomed to cyclists. Another example of higher preference for a certain road type across all apps in one region compared to the other is use of footpaths and off-road cycling tracks. Reference routes in North Holland use about 0.2% more footpaths and 10.9% more cycleways than in South Florida (Figure 6c), but on top of this, routes observed in all three apps in North Holland utilize footpaths and cycleways much more frequently (e.g., plus 44.6%, 38.5%, 31.5% for cycleways). Compare also the distribution of % cycleway use between control and Endomondo in North Holland in Figure 6c and Figure 6d. Cyclists in North Holland appear to be more aware of these off-road cycling facilities, possibly because of their frequent rides and their familiarity with the area.

Whereas all three platforms demonstrate increased importance of certain trip characteristics in one region over the other, as demonstrated above, this is not the case for all trip characteristics analyzed. For example, the use of secondary and tertiary roads does not differ significantly between the two regions for the reference routes, however, it does differ significantly between regions among the different apps. For example, there are significantly fewer secondary roads in North Holland for two apps (Bikemap and Endomondo) but significantly more for MapMyFitness.

Overall, comparison of app-derived trip characteristics between both regions shows that there is an even higher use of network structure for the “extreme” ends of the network supply structure, i.e. primary roads (heavy traffic), foot paths and cycleways (no motorized traffic) in North Holland than in South Florida. As opposed to this, South Florida cyclists tend to use residential roads more often than cyclists in North Holland.

5.2 Comparison of trip characteristics within South Florida and North Holland

5.2.1 Trips from each App compared to control

To better understand cyclist preferences on the different apps within a study region, the characteristics of trips obtained from each app were, as a first step, compared to those obtained from the set of reference routes (control). There are some trip characteristics which differ between all apps and the control in the same direction, so that a general trend of route preferences can be inferred in a region. In other cases, the apps differ from the control in opposite directions indicating that no general conclusions can be formed. For this analysis, median differences in the right portion of Table 5 and Table 6 are more closely examined for South Florida and North Holland respectively, which show the difference between a trip characteristics of an app (Bikemap, Endomondo, or MapMyFitness) and that of the control. The main focus of this section is on the aspects which have not been discussed during the comparison of trips between the regions.

Table 5 Median differences South Florida for Bikemap (B), Endomondo (E), MapMyFitness (M), and control (C).

Differences:	**B - E**		**B - M**		**E - M**		**B - C**		**E - C**		**M - C**	
Trip geometry												
Shape length	-6247.37	***	-1149.00		5098.37	***	4975.24	***	11222.61	***	6124.24	***
Shape ratio	0.06	*	0.55	***	0.49	***	-0.29	***	-0.35	***	-0.83	***
Turns per km	0.71	***	0.45	***	-0.26	***	0.34	***	-0.37	***	-0.11	
Turns right per km	0.36	***	0.24	***	-0.12	***	0.22	***	-0.14	***	-0.03	
Turns left per km	0.31	***	0.22	***	-0.09	***	0.15	***	-0.15	***	-0.07	
Trip characteristics												
Traffic signals per km	0.14		0.67	***	0.52	***	-0.40	***	-0.54	***	-1.06	***
Flickr images per km	-1.50	***	1.63	***	3.12	***	0.14		1.64	***	-1.48	***
Population mean	31.58	***	32.64	***	1.06		1.48		-30.10	***	-31.16	***
Road Category												
% Primary road	-0.21		2.59	***	2.80	***	-12.25	***	-12.03	***	-14.83	***
% Secondary road	-0.55		5.53	***	6.08	***	1.34		1.89		-4.19	***
% Tertiary road	5.20	**	3.54	*	-1.66		6.66	**	1.46		3.12	
% Residential road	5.24	***	-2.57		-7.81	***	9.08	***	3.84	***	11.65	***
% Foot/pedestrian	-1.64	***	-0.01		1.63	***	0.91	***	2.55	***	0.92	***
% Cycleway	-0.31	***	0.00	*** (-)	0.31	***	0.00	*** (+)	0.31	***	0.00	*** (+)
Land Use												
% Commercial	3.33	***	8.90	***	5.57	***	-3.47	*	-6.80	***	-12.37	***
% Farm	0.00	** (-)	0.00	** (-)	0.00	(-)	0.00	(+)	0.00	* (+)	0.00	** (+)
% Forest	-0.42	***	0.00	(-)	0.42	***	0.00	(-)	0.42	***	0.00	(+)
% Open land	0.00	(-)	0.00	(-)	0.00	(-)	0.00	(+)	0.00	(+)	0.00	(+)
% Industrial	0.00	(+)	0.00	*** (+)	0.00	(+)	0.00	(+)	0.00	* (-)	0.00	*** (-)
% Recreation	-1.93	***	-1.81	***	0.12		1.55	**	3.48	***	3.36	***
% Residential	8.76	***	-7.52	***	-16.28	***	3.59	*	-5.17	**	11.12	***

$p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***

(-) Median difference was zero but mean difference was negative.

(+) Median difference was zero but mean difference was positive

Users of all apps across both regions demonstrated significantly greater use of residential roads, footways, and cycleways compared to control, and also show avoidance of primary roads and traffic signals. These results support earlier findings of preferences for riding on cycleways and roads with lower speed limits and less traffic (Bigazzi & Gehrke, 2018). Similarly, another study stated increased stress levels in cyclists on roads with higher speed limits or traffic volume, and, identified speed of traffic and lack of separation distance from traffic as the two main factors that are perceived as risk to cyclists (Christmas et al., 2010). In addition, all North

Holland apps showed significantly lower use of secondary roads than the control, but there is no clear pattern for South Florida. This could indicate that, possibly due to the scarcity of non-motorized route alternatives in South Florida, secondary roads are considered as somewhat acceptable (but not necessarily preferred) network links in that region.

Table 6 Median differences North Holland for Bikemap (B), Endomondo (E), MapMyFitness (M), and control (C).

Differences:	B - E		B - M		E - M		B - C		E - C		M - C	
Trip geometry												
Shape length	6342.20	***	6527.11	***	184.91		13056.13	***	6713.93	***	6529.02	***
Shape ratio	-0.05	***	0.04	*	0.09	***	-0.13	***	-0.08	***	-0.18	***
Turns per km	0.29	***	0.27	***	-0.02		0.21	***	-0.08		-0.06	
Turns right per km	0.16	***	0.17	***	0.01		0.10	***	-0.06		-0.06	
Turns left per km	0.14	***	0.14	***	0.01		0.10	***	-0.03		-0.04	
Trip characteristics												
Traffic signals per km	-1.03	***	-0.95	***	0.09		-2.85	***	-1.81	***	-1.90	***
Flickr images per km	0.75	***	1.05	***	0.30		0.52	**	-0.23		-0.53	
Population mean	-29.06	***	-47.21	***	-18.15	**	-85.48	***	-56.42	***	-38.27	***
Road Category												
% Primary road	-0.43	***	-0.51	***	-0.09		-1.29	***	-0.86	**	-0.78	***
% Secondary road	-1.18	***	-1.85	***	-0.66	*	-9.17	***	-7.98	***	-7.32	***
% Tertiary road	-0.37		-1.64	*	-1.27		1.38	*	1.75	**	3.02	***
% Residential road	-1.58	***	-0.69	**	0.90	*	1.58	***	3.16	***	2.27	***
% Foot/pedestrian	-2.56	***	-4.10	***	-1.54	***	1.88	***	4.43	***	5.97	***
% Cycleway	5.69	**	13.02	***	7.33	***	33.64	***	27.95	***	20.62	***
Land Use												
% Commercial	-0.21	**	-0.41	***	-0.21	*	-0.56	***	-0.35		-0.15	
% Farm	0.47	***	0.81	***	0.34	***	0.86	***	0.39	***	0.05	
% Forest	-0.77		0.98	***	1.75	***	0.35	***	1.13	***	-0.63	
% Open land	3.22	**	0.55		-2.67	**	5.66		2.44	**	5.11	***
% Industrial	-0.35	***	0.11		0.46	***	0.23		0.58	***	0.13	
% Recreation	-1.04	***	-1.66	***	-0.62		-0.17	***	0.87	***	1.49	***
% Residential	-3.90	***	-2.14	*	1.75		-10.83	***	-6.93	***	-8.68	***

$p < 0.05$ *; $p \leq 0.01$ **; $p < 0.001$ ***

Bikemap users are willing to accept more turns than the control in both study areas. As opposed to this, Endomondo trips reveal fewer turns than control (significant only for South Florida), whereas MapMyFitness users are in-line with the number of turns provided in control. Hence, the three apps reveal differences in trip behavior with regard to route complexity and no general trend for cyclists is evident.

Routes for all apps in North Holland and MapMyFitness and Endomondo in South Florida tend to pass through less densely populated areas, avoid commercial land, and have a higher proportion of open land compared to reference routes, revealing the general preference for natural environments along chosen routes. This is further supported by the finding that all apps (except for Bikemap in North Holland) reveal a higher share of recreational land use than expected compared to the reference routes.

In North Holland trips from all apps reveal less residential land use than expected, which could be explained by available route alternatives that pass through farmland, open land, and forest, most of which show a significantly higher percentage than expected. As opposed to this, in South Florida median differences for farmland, forest, and open land are almost all zero, part of the reason being that these land use types cover only a small portion of the area.

5.2.2 Trips from each app compared to each other

Whereas the previous analysis focused on identifying common patterns of trip characteristics revealed by the different apps relative to reference routes in the two study regions, the following discussion more closely examines notable differences in characteristics of trips among the apps themselves, within a region as well as between regions.

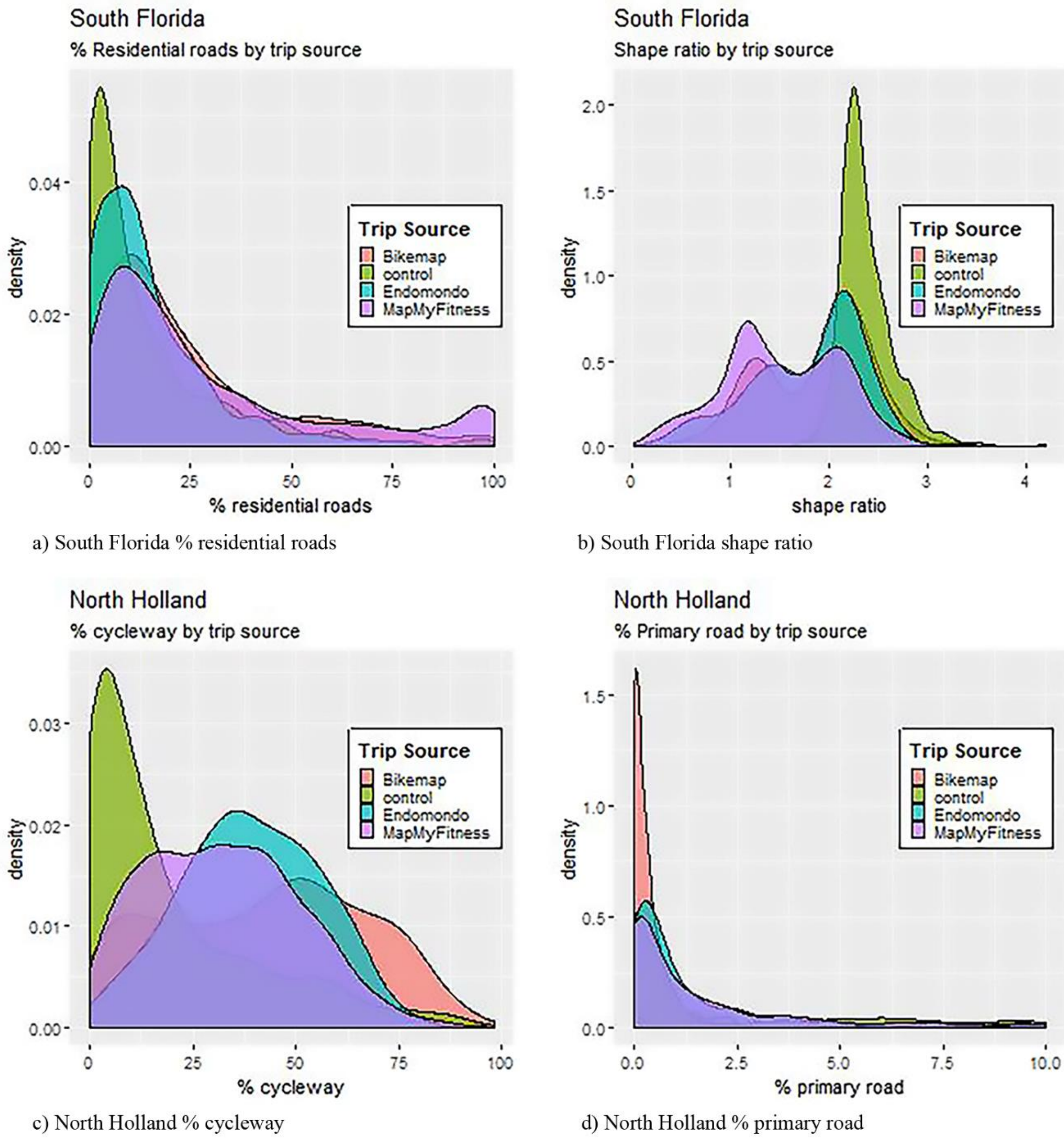
Endomondo users in South Florida travel significantly longer routes than users of both Bikemap and MapMyFitness (Table 5). This may be, in part, because Endomondo is more targeted toward athletes who undertake more serious exercise trips than those riding for leisure. However, in North Holland, Bikemap trips are significantly longer than those tracked on MapMyFitness and Endomondo (Table 6), whereas the latter two share similar lengths. This reveals inconsistent results between both regions as to which app tracks the longest trips.

In South Florida, Endomondo users use significantly more cycleways and foot paths than users of Bikemap and MapMyFitness, whereas routes from the latter two apps run more frequently on residential roads (Figure 7a). Therefore, users of different apps choose different types of road network segments in their search for low-traffic connections. As opposed to this, in North Holland, trips from Bikemap use the highest share of cycleways (Figure 7c), and MapMyFitness features most foot paths, whereas Endomondo trips show significantly more use of residential roads than the other two sources. Therefore, link types used to achieve the goal of low-traffic connections in the different apps are more or less flipped between the two regions, suggesting that the various apps are used differently in both regions with regard to low-volume network segments.

As for use of major roads, MapMyFitness trips feature a significantly smaller percentage of primary and secondary roads than the other two apps in South Florida, whereas the same is true of Bikemap in North Holland (see Figure 7d for primary roads in North Holland). Therefore, between the two analyzed regions, different apps track most trips that avoid high-volume traffic roads.

MapMyFitness trips have a smaller shape ratio (and thus more circuitous paths) than Bikemap and Endomondo trips in both analyzed regions (see Figure 7b for South Florida). This is in-line with the previous finding for South Florida where MapMyFitness trips tend to avoid major roads but rather follow residential roads, and for North Holland where MapMyFitness trips include more footpaths than trips from other sources.

Land use also plays a quite different role for routes extracted from the various apps when comparing both regions. For example, whereas in South Florida Bikemap trips have the highest and MapMyFitness trips the lowest percentage of commercial land use associated with their trips, respectively, the exact opposite is true for North Holland. However, other results, such as Bikemap trips passing through the fewest areas classified as recreational land use among all three apps is consistent across both regions.



a) South Florida % residential roads

b) South Florida shape ratio

c) North Holland % cycleway

d) North Holland % primary road

Figure 7 Examples of non-normal distributions from South Florida % residential roads (a) and shape ratio (b) and North Holland % cycleway (c) and % primary road (d).

5.3 Characteristics of Florida State Parks that attract various fitness activities

5.3.1 Hiking models

Hiking trail density played a significant role for users of all apps, where, as expected, a higher trail density increased the likelihood of park visitation for hiking (Table 7). For two apps

(AllTrails and MapMyFitness) the hike/bike combination trail density also had a positive impact, but bike trail density had a negative impact for AllTrails users. The latter might indicate that the presence of cyclists might be perceived as disturbance for hikers using this app. Availability of restrooms is an important factor for hikers amongst all apps. Parks with entrance fees experience higher hiking activity for AllTrails and MapMyFitness, pointing towards visitors' willingness to pay for certain services, such as better infrastructure or perceived safety (e.g. due to presence of park rangers at the entrance point) in State Parks (Zhang & Zhou, 2018). In terms of landcover, percent urban landcover was a positive predictor for MapMyFitness and Wikiloc users, and percent forest landcover was a positive predictor for AllTrails users. Higher population within 30 minutes of the park was significant for AllTrails and MapMyFitness, indicating that parks in more populated areas tend to have higher activity numbers, as expected. The population of the nearest large city decreased the likelihood of a visit for Wikiloc users, suggesting that users of this app tend to prefer more remote locations, which is in line with earlier findings (Norman & Pickering, 2017).

Table 7 Negative binomial models for AllTrails, MapMyFitness, and Wikiloc hiking activities; ln(area) used as offset.

		AllTrails Hike		MapMyFitness Hike		Wikiloc hike	
Intercept	Coeff	-5.94		-8.15		1.61	
	z	-4.16	***	-4.88	***	0.93	
Park facilities/characteristics							
Hiking trail density (km/km^2)	Coeff	0.42		0.38		0.29	
	z	3.62	***	3.99	***	2.31	*
Biking trail density (km/km^2)	Coeff	-0.78					
	z	-3.00	**				
Combo hike/bike trail density (km/km2)	Coeff	0.33		0.18			
	z	3.40	***	2.17	*		
No. of restrooms	Coeff	0.11		0.05		0.24	
	z	2.07	*	4.86	***	4.91	***
No. of campgrounds	Coeff			0.06			
	z			-3.50	***		
Fee charged (0 no 1 yes)	Coeff	1.30		0.33			
	z	3.46	***	3.19	**		
Landcover							
% Urban (recreational)	Coeff			0.01		0.05	
	z			4.67	***	4.48	***
% Forest	Coeff	0.01					
	z	2.12	*				
% Water	Coeff					-0.03	
	z					-2.48	*
Population statistics							
Population (30-min)	Coeff	0.42		0.12			
	z	3.64	***	3.18	**		
Population (nearest large city)	Coeff					-0.35	
	z					-2.10	*
Drive time							
Nearest airport	Coeff			0.01			
	z			3.85	***		
Moran's I (*p-value*)		-0.01 (*0.93*)		0.02 (*0.37*)		0.03 (*0.08*)	
Null log likelihood		-592.9		-419.61		-336.48	
Full log likelihood		-551.97		-369.11		-310.23	
McFadden's pseudo R^2		0.07		0.12		0.08	
N		5285		1002		564	

*p < 0.05; **p < 0.01; ***p < 0.001
Predictors not significant to any of the three apps not shown in table.

5.3.2 Cycling models

Biking trail or combined hike/bike trail density has a positive impact on reported cycling counts for MapMyFitness and Wikiloc but not for AllTrails (Table 8). This could be because AllTrails users are primarily hikers (compare ***Error! Reference source not found.***) and cycling might just be a secondary activity option for AllTrails park visitors without much focus on cycling trails. As with the hiking models, more restrooms increase cycling activities in State Parks and, along the same line, urban landcover is positively associated with cycling visits. As before, parks in more populated areas experience more cycling activities, which are operationalized either as higher population counts with the 30-minute service area (for AllTrails) or as shorter driving drive time to the nearest large city (for MapMyFitness). For Wikiloc, corresponding findings are mixed, since both population within 30 minutes but also longer distance to the nearest largest city are positively associated with cycling activity counts. The latter points towards preference for more remote locations of Wikiloc users. The cycling model for MapMyFitness is the only one with an ethnicity variable being significant, namely with a negative coefficient for % Black population within a 30-minute area, supporting earlier identified visitor patterns for non-urban parks (Lawton & Weaver, 2008).

Table 8 Negative binomial models for AllTrails, MapMyFitness, and Wikiloc cycling activities; ln(area) used as offset.

		AllTrails Cycle		MapMyFitness cycle		Wikiloc cycle	
Intercept	Coeff	-7.83		-2.21		-18.82	
	z	-4.93	***	-4.75	***	-4.07	***
Park facilities/characteristics							
Biking trail density (km/km^2)	Coeff			0.93			
	z			4.74	***		
Combo hike/bike trail density (km/km^2)	Coeff			0.35		0.62	
	z			3.71	***	3.78	***
No. of restrooms	Coeff	0.20		0.29		0.59	
	z	3.93	***	6.27	***	6.76	***
Landcover							
% Urban (recreational)	Coeff	0.06		0.05		0.08	
	z	5.75	***	5.61	***	5.25	***
% Barren	Coeff			0.05			
	z			2.59	**		
% Forest	Coeff	0.03					
	z	4.14	***				
Population statistics							
Population (30-min)	Coeff	0.28				0.61	
	z	2.15	*			1.98	*
% Black (30-min)	Coeff			-0.05			
	z			-2.91	**		
Median age (nearest large city)	Coeff					0.06	
	z					2.17	*
Attractions							
Nearby State Parks (30-min)	Coeff					0.23	
	z					2.04	*
Drive time							
Nearest large city	Coeff			-0.03		0.07	
	z			-3.55	***	2.75	**
Moran's I (*p-value*)		-0.01 (*0.87*)		-0.01 (*0.88*)		0.00 (*0.80*)	
Null log likelihood		-225.42		-373.47		-145.21	
Full log likelihood		-173.99		-315.34		-122.79	
McFadden's pseudo R^2		0.23		0.16		0.15	
N		463		1200		192	

*p < 0.05; **p < 0.01; ***p < 0.001
Predictors not significant to any of the three apps not shown in table.

5.3.3 Running models

Higher hiking and hike/bike combination trail density increases park visitations for running activities (Table 9). The number of available restrooms positively impacts visits while campgrounds have a negative impact. As with hiking, park entrance fees are associated with higher activity numbers. Surrounding population is positively associated with running counts for AllTrails, where at the same time higher percentage of Hispanic population is associated with lower running counts. Also, for MapMyFitness proximity to the nearest city increases running counts. The model for Wikiloc does not show a significant effect of nearby population on activity counts, pointing again towards a different type of recreationalists that use this platform compared to AllTrails and MapMyFitness.

Table 9 Negative binomial models for AllTrails, MapMyFitness, and Wikiloc running activities; ln(area) used as offset.

		AllTrails run		MapMyFitness run		Wikiloc run	
Intercept	Coeff	-14.36		-1.39		-6.76	
	z	-5.02	***	-4.10	***	-6.38	***
Park facilities/characteristics							
Hiking trail density (km/km^2)	Coeff	0.43				1.82	
	z	2.45	*			4.51	***
Biking trail density (km/km^2)	Coeff			0.45			
	z			2.49	*		
Combo hike/bike trail density (km/km^2)	Coeff	0.39					
	z	2.47	*				
No. of restrooms	Coeff			0.25		0.56	
	z			4.56	***	3.07	**
No. of campgrounds	Coeff			-0.28			
	z			-4.28	***		
Fee charged (0 no 1 yes)	Coeff	1.68		0.91			
	z	2.61	***	2.60	**		
Landcover							
% Urban (recreational)	Coeff			0.05			
	z			7.34	***		
% Barren	Coeff			0.05			
	z			2.91	**		
Population statistics							
Population (30-min)	Coeff	0.82					
	z	3.52	***				
% Hispanic (30-min)	Coeff	-0.10					
	z	-2.49	*				
Drive time							
Nearest large city	Coeff			-0.03			
	z			-3.23	**		
Moran's I (*p-value*)		-0.01 (*0.93*)		-0.00 (*0.95*)		-0.01 (*0.97*)	
Null log likelihood		-109.91		-422.65		-61.2	
Full log likelihood		-90.91		-382.37		-59.0	
McFadden's pseudo R^2		0.17		0.10		0.04	
N		61		1116		38	

•p < 0.1; *p < 0.05; **p < 0.01; ***p < 0.001

Predictors not significant to any of the three apps not shown in table.

5.3.4 Paddle-sport models

The number of canoe/kayak launch areas or water landcover are positive predictors for paddling activities on the different platforms, as can be expected (Table 10). Other park features play only a minor role for paddling activities. Population of surrounding area has a positive impact on paddling visits for all three platforms, as does the median age of the surrounding area. The latter might point towards more experienced recreationalists to undertake paddling activities, which require more skills and experience than other activities, such as hiking. As with running and hiking, the presence of an entry fee is associated with increased activities for paddling as well.

Table 10 Negative binomial models for AllTrails, MapMyFitness, and Wikiloc paddling activities; ln(area) used as offset.

		AllTrails paddle		MapMyFitness paddle		Wikiloc paddle	
Intercept	Coeff	-15.17		-21.07		-12.46	
	z	-4.89	***	-3.64	***	-4.93	***
Park facilities/characteristics							
Hiking trail density (km/km^2)	Coeff					82.99	
	z					2.57	*
No. of restrooms	Coeff					0.19	
	z					3.21	**
No. of canoe/kayak launch areas	Coeff	0.93				0.45	
	z	3.75	***			2.95	**
No. of campgrounds	Coeff					-0.24	
	z					-3.43	***
Fee charged (0 no 1 yes)	Coeff	3.54		2.27		1.99	
	z	3.26	**	2.02	*	4.11	***
Landcover							
% Forest	Coeff					-0.02	
	z					-2.55	*
% Water	Coeff			0.07			
	z			3.85	***		
Population statistics							
Population (30-min)	Coeff	0.55					
	z	2.44	*				
Population (nearest large city)	Coeff			0.94		0.55	
	z			2.12	*	2.70	**
Median age (nearest large city)	Coeff			0.11		0.06	
	z			2.91	**	3.88	***
Moran's I (*p-value*)		-0.02 (*0.33*)		-0.01 (*0.99*)		0.03 (*0.08*)	
Null log likelihood		-85.57		-65.02		-245.25	
Full log likelihood		-66.76		-61.48		-205.28	
McFadden's pseudo R^2		0.22		0.05		0.16	
N		50		22		291	

*p < 0.05; **p < 0.01; ***p < 0.001
Predictors not significant to any of the three apps not shown in table.

5.4 Differences among apps used and physical activity tracked

The chi-square test revealed a significant association between the app used and the type of sport recorded, $\chi^2(6, 10208) = 6094.2$, $p < 0.001$. More specifically, AllTrails users recorded more hiking activities than expected, while MapMyFitness users recorded more cycling and running activities than expected, and Wikiloc users recorded more paddling activities than expected.

*Figure 8***Error! Reference source not found.** shows the association plot for the relationship between app and sport categories. The signed height of each rectangle is the residual, given by $\frac{O-E}{\sqrt{E}}$, where O represents the observed value and E is the expected value under the assumption of independence. Therefore, a taller rectangle means a larger residual. The width of each rectangle equals $\sqrt{E}$ so that the corresponding area is the absolute difference in observed and expected values ($|O-E|$). For example, since the rectangle for Wikiloc paddle is tall, positive, and narrow, this means that the residual is large and positive, and that the observed value is large relative to the expected value.

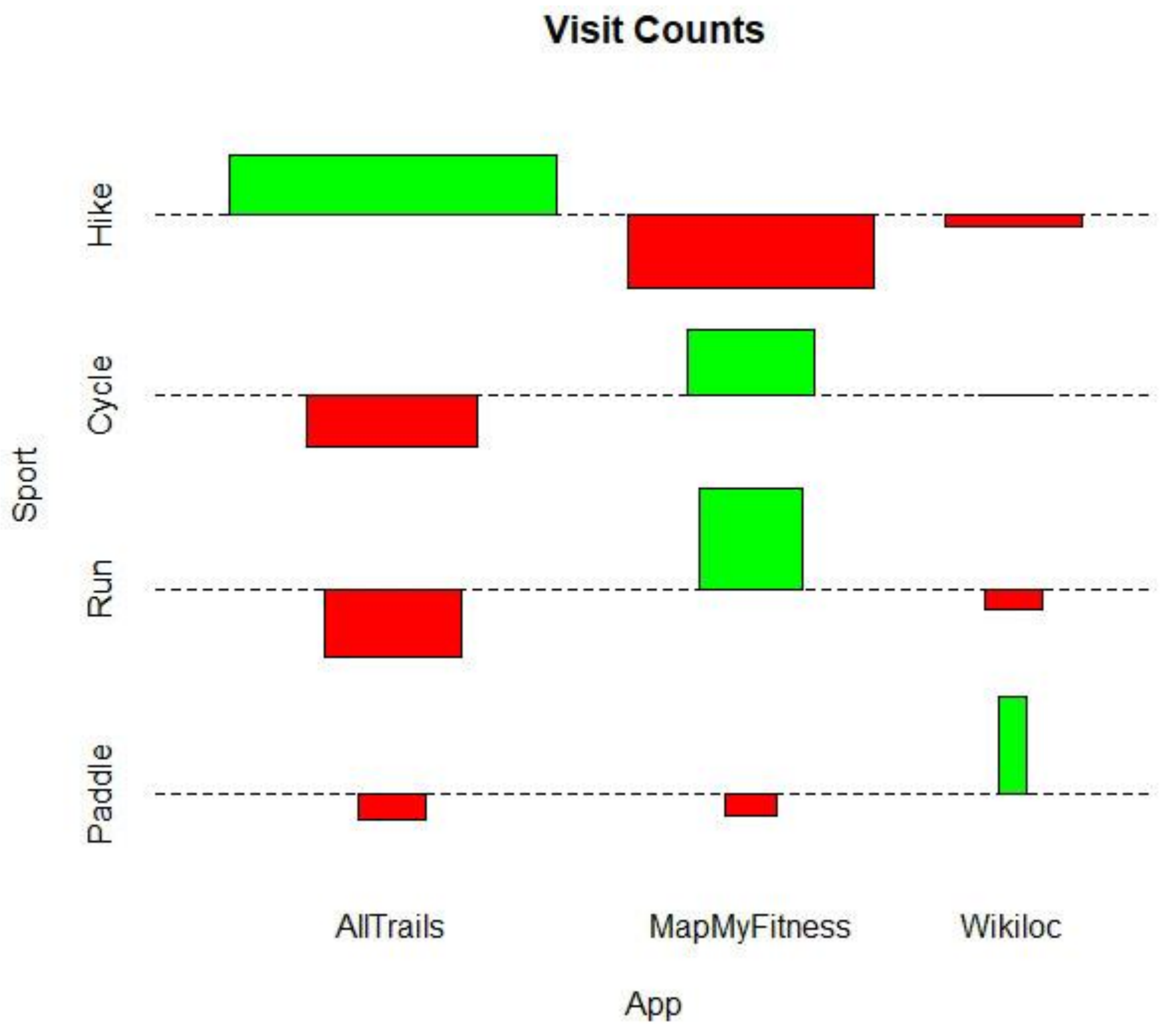


Figure 8 Association plot showing relationship between the app used and the sport recorded.

6. Discussion

The analyses conducted in this study reveal that significant differences exist between the characteristics of trips extracted from different fitness tracker apps for cyclists. Whereas comparison of app-based trip characteristics with those of a reference trip set (within each region) showed that these differences point by and large in the same direction (e.g. avoid traffic signals), differences become more apparent when comparing trip characteristics between trips from different apps directly (e.g. compare % cycleway along trips in Bikemap and Endomondo). Adding to this complexity is the fact that comparison results between apps fluctuate highly between analyzed regions. The observed median difference patterns point to the fact that use of data from only one app in assessing route characteristics could be biased towards the subset of cyclists who choose the app. Although the use of data from fitness tracker apps for measuring and modeling cyclist behavior has certain limitations and inaccuracies (Blanc et al., 2016; Griffin & Jiao, 2015; Rupi et al., 2019), this study is one of the first to provide valuable insights into comparative characteristics of trips logged by means of different apps. Furthermore, it is evident from the results that the route characteristics of users of an app identified in one geographic region will not necessarily transfer to a different region. Limited transferability of model results between regions has already been examined for other aspects of transportation research, such as tour-generation models (Nowrouzian & Srinivasan, 2012), but less so in the context of GPS tracker apps. Significant differences between the two study areas were made evident. For example, North Holland cyclists cycle on routes that are less circuitous than South Florida cyclists and they tend to travel through more farm, forest, and open land than South Florida cyclists as well. These differences go beyond what is explainable by network differences alone. Besides this, comparison results between apps may vary by region. For example, results show that Bikemap users in North Holland travel through less densely populated areas than users of the other two apps (Figure 9a), whereas in South Florida Bikemap users travel through more densely populated areas (Figure 9b) compared to MapMyFitness and Endomondo.

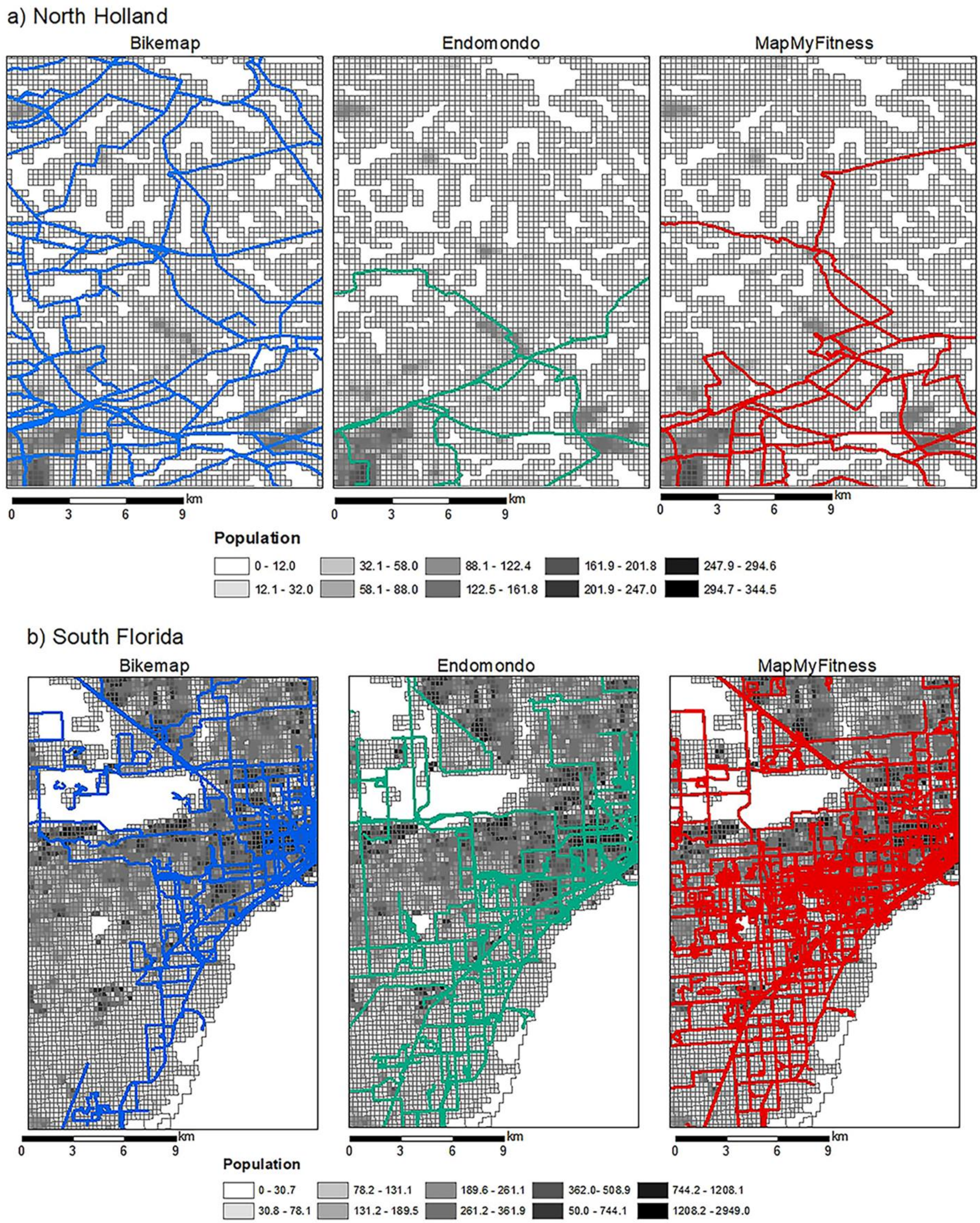


Figure 9 Bikemap cyclists travel primarily in less densely populated areas in North Holland (a) but primarily cycle in more densely populated areas in South Florida.

Endomondo allows for the extraction of personal information, including gender and age, however, no such information is publicly available for extraction from Bikemap or MapMyFitness. Therefore, comparison of ridership demographics and determining user selection bias between these apps was not possible.

This study also provides new insights into factors attracting visitors to Florida State Parks for physical activities, specifically hiking, cycling, running, and paddling. The models for all four sports identify some shared characteristics of increased park visitation for the assessed sports. For hikers, cyclists, and runners, the abundance of the appropriate trail is a positive predictor for park use for the specified activity. A similar result was found with regard to trail length in parks in Queensland Australia (Norman & Pickering, 2019). Trail density is not a contributing factor for paddling however, but, rather, the availability of canoe/kayak launch areas and water landcover were more often significant. Since Florida weather is conducive to year-round sporting activities such as those assessed, seasonality was not considered in the models. In areas with greater seasonal diversity, weather may contribute to lack of park visitation for the assessed activities during winter months and therefore the results may not be transferable to these regions. In fact, average daily maximum temperature was found to be a strong positive predictor for visitation of five national parks in Utah (Smith et al., 2018). In addition, climates with cold winters will attract visitors for winter sports like skiing or snowboarding in the cold season, rather than the sports considered in this research (Henderson, 2003). Visitors to parks for reasons other than physical activities, such as social gatherings, birding, reading, or family outings were not addressed, hence there may be differing factors influencing park choice for these activities.

7. Limitations and future work

7.1 Cycling path characteristics

The study demonstrated some geographic bias between the different apps, e.g. trips of one app running more frequently in areas of certain land use types than for other apps. In this study trip purpose was not considered in the analysis since only Endomondo asks users to specify their trip purpose. Consideration of trip purpose for analysis, where available in apps, could help to further refine the identification of commonalities and differences in trip characteristics from various apps and in varying regions, respectively. Another limitation of the study is the lack of information regarding trip duration on two apps (Bikemap and MapMyFitness), which made characteristics such as average speed incomparable between the apps. This information would have been useful in determining the athleticism of app users, as casual cyclists tend to ride at lower speeds than more serious cyclists. The analysis methods in this study do not apply to cycling volume or spatio-temporal changes of travel patterns. These can be considered aspects of future work.

7.1 Park popularity

Results showed that each app has its favorite sport type covered, suggesting that trip quantities derived from the different apps need to be read with caution since they will deviate from ground truth data both in absolute and even relative amounts. This did not pose a problem for this study which focused primarily on identifying factors associated with increased or decreased sport activities, but not on preference between sport types in parks. Also, fitness tracker apps tend to under-sample certain socio-economic groups. Whereas some studies already began to address the type of discrepancies between crowd-sourced park visitation numbers (Hausmann et al., 2018;

Tenkanen et al., 2017) or sports activity counts (Jestico et al., 2016) and alternative quantification methods (e.g. manual counts), this has not yet been conducted for the apps considered in this study or for State Parks, respectively, which can therefore be considered part of future work.

8. Journal submissions

The results of this research have been submitted to two journals and are currently under review. Objectives 1 and 2 results have been submitted to the International Journal of Sustainable Transportation with the title “Comparison of Cycling Path Characteristics in South Florida and North Holland among Three GPS Fitness Tracker Apps.” Objectives 3 and 4 results have been submitted to the Journal of Outdoor recreation and Tourism under the title “Identification of Structural, Environmental, and Socio-demographic Correlates of Outdoor Activities in Florida State Parks from Three Fitness Tracker Apps.”

9. References

About AllTrails. (2020). Retrieved July 5, 2020, from https://www.alltrails.com/about

About Bikemap. (2020) Retrieved May 8, 2020, from https://www.bikemap.net/en/about-bikemap/

About MapMyFitness | MapMyRun. (2020). *Under Armour*. Retrieved February 11, 2020, from https://blog.mapmyrun.com/about/about-mapmyfitness/

Bigazzi, A. Y., & Gehrke, S. R. (2018). Joint consideration of energy expenditure, air quality, and safety by cyclists. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 652–664. https://doi.org/10.1016/j.trf.2018.07.005

Blanc, B., Figliozzi, M., & Clifton, K. (2016). How Representative of Bicycling Populations Are Smartphone Application Surveys of Travel Behavior? *Transportation Research Record*, *2587*(1), 78–89. https://doi.org/10.3141/2587-10

Blanck, H. M., Allen, D., Bashir, Z., Gordon, N., Goodman, A., Merriam, D., & Rutt, C. (2012). Let’s Go to the Park Today: The Role of Parks in Obesity Prevention and Improving the Public’s Health. *Childhood Obesity*, *8*(5), 423–428. https://doi.org/10.1089/chi.2012.0085.blan

Boss, D., Nelson, T., Winters, M., & Ferster, C. J. (2018). Using crowdsourced data to monitor change in spatial patterns of bicycle ridership. *Journal of Transport & Health*, *9*, 226–233. https://doi.org/10.1016/j.jth.2018.02.008

Brownson, R. C., Baker, E. A., Housemann, R. A., Brennan, L. K., & Bacak, S. J. (2001). Environmental and Policy Determinants of Physical Activity in the United States. *American Journal of Public Health*, *91*(12), 1995–2003. https://doi.org/10.2105/AJPH.91.12.1995

Buckley, R. (2009). Parks and Tourism. *PLOS Biology*, *7*(6), e1000143. https://doi.org/10.1371/journal.pbio.1000143

Campelo, M. B., & Nogueira Mendes, R. M. (2016). Comparing webshare services to assess mountain bike use in protected areas. *Journal of Outdoor Recreation and Tourism*, *15*, 82–88. https://doi.org/10.1016/j.jort.2016.08.001

Chen, Z., Shen, H. T., & Zhou, X. (2011). Discovering popular routes from trajectories. *2011 IEEE 27th International Conference on Data Engineering*, 900–911. https://doi.org/10.1109/ICDE.2011.5767890

Christmas, S., Great Britain, & Department for Transport. (2010). *Cycling, safety and sharing the road: Qualitative research with cyclists and other road users*. Department for Transport.

Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, *3*(63). https://doi.org/10.3389/fenvs.2015.00063

Dill & Carr. (2003). Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them – Another Look. *Transportation Research Record: Journal of the Transportation Research Board*, *1828*, 116-123.

Facts About Endomondo. (2020). Retrieved February 11, 2020, from http://www.endomondo.com/files/press/Facts_EN.pdf

Faskunger, J. (2013). Promoting Active Living in Healthy Cities of Europe. *Journal of Urban Health*, *90*(1), 142–153. https://doi.org/10.1007/s11524-011-9645-7

FDEP - Parks and Recreation. (2020). Retrieved June 26, 2020, from https://floridadep.gov/parks

Goodchild, M. F. (2007). Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, *2*(2), 24–32. https://doi.org/10.2902/

Gotschi, T. (2011). Costs and Benefits of Bicycling Investments in Portland, Oregon. *Journal of Physical Activity and Health*, *8*(s1), S49–S58. https://doi.org/10.1123/jpah.8.s1.s49

Griffin, G. P., & Jiao, J. (2015). Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *Journal of Transport & Health*, *2*(2), 238–247. https://doi.org/10.1016/j.jth.2014.12.001

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Minin, E. D. (2018). Social Media Data Can Be Used to Understand Tourists' Preferences for Nature-Based Experiences in Protected Areas. *Conservation Letters*, *11*(1), e12343. https://doi.org/10.1111/conl.12343

Heesch, K. C., & Langdon, M. (2016). The usefulness of GPS bicycle tracking data for evaluating the impact of infrastructure change on cycling behaviour: GPS bicycle tracking data in evaluating cycling behaviour. *Health Promotion Journal of Australia*, *27*(3), 222–229. https://doi.org/10.1071/HE16032

Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey. *ISPRS International Journal of Geo-Information*, *6*(3), 85. https://doi.org/10.3390/ijgi6030085

Henderson, K. (2003). The Influence of Climate Variations on Visitor Activities in Pennsylvania's State Parks Introduction Pennsylvania Climatology. *Middle States Geographer*, *36*, 171-177.

Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*, *52*, 90–97. https://doi.org/10.1016/j.jtrangeo.2016.03.006

Larsen, J., & El-Geneidy, A. (2011). A travel behavior analysis of urban cycling facilities in Montréal, Canada. *Transportation Research Part D: Transport and Environment*, *16*(2), 172–177. https://doi.org/10.1016/j.trd.2010.07.011

Lawton, L. J., & Weaver, D. B. (n.d.). *Factors Associated with Non- visitation by Area to Congaree National Park, South Carolina*. 17.

McArthur, D. P., & Hong, J. (2019). Visualising where commuting cyclists travel using crowdsourced data. *Journal of Transport Geography*, *74*, 233–241. https://doi.org/10.1016/j.jtrangeo.2018.11.018

Neuvonen, M., Pouta, E., Puustinen, J., & Sievänen, T. (2010). Visits to national parks: Effects of park characteristics and spatial demand. *Journal for Nature Conservation*, *18*(3), 224–229. https://doi.org/10.1016/j.jnc.2009.10.003

Norman, P., & Pickering, C. M. (2017). Using volunteered geographic information to assess park visitation: Comparing three on-line platforms. *Applied Geography*, *89*, 163–172. https://doi.org/10.1016/j.apgeog.2017.11.001

Norman, P., & Pickering, C. M. (2019). Factors influencing park popularity for mountain bikers, walkers and runners as indicated by social media route data. *Journal of Environmental Management*, *249*(109413). https://doi.org/10.1016/j.jenvman.2019.109413

Nowrouzian, R., & Srinivasan, S. (2012). Empirical Analysis of Spatial Transferability of Tour-Generation Models. *Transportation Research Record*, *2302*(1), 14–22. https://doi.org/10.3141/2302-02

Romanillos, G., Zaltz Austwick, M., Ettema, D., & De Kruijf, J. (2016). Big Data and Cycling. *Transport Reviews*, *36*(1), 114–133. https://doi.org/10.1080/01441647.2015.1084067

Rupi, F., Poliziani, C., & Schweizer, J. (2019). Data-driven Bicycle Network Analysis Based on Traditional Counting Methods and GPS Traces from Smartphone. *ISPRS International Journal of Geo-Information*, *8*(8), 322. https://doi.org/10.3390/ijgi8080322

Santos, T., Nogueira Mendes, R., & Vasco, A. (2016). Recreational activities in urban parks: Spatial interactions among users. *Journal of Outdoor Recreation and Tourism*, *15*, 1–9. https://doi.org/10.1016/j.jort.2016.06.001

Schägner, J. P., Brander, L., Maes, J., Paracchini, M. L., & Hartje, V. (2016). Mapping recreational visits and values of European National Parks by combining statistical modelling and unit value transfer. *Journal for Nature Conservation*, *31*, 71–84. https://doi.org/10.1016/j.jnc.2016.03.001

Smith, J. W., Wilkins, E., Gayle, R., & Lamborn, C. C. (2018). Climate and visitation to Utah's 'Mighty 5' national parks. *Tourism Geographies*, *20*(2), 250–272. https://doi.org/10.1080/14616688.2018.1437767

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports*, *7*(1), 17615. https://doi.org/10.1038/s41598-017-18007-4

Watkins, K., Ammanamanchi, R., LaMondia, J., & Le Dantec, C. A. (2016). *Comparison of Smartphone-based Cyclist GPS Data Sources* (No. 16–5309). Article 16–5309. Transportation Research Board 95th Annual MeetingTransportation Research Board. https://trid.trb.org/view/1393960

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, *3*(1), 2976. https://doi.org/10.1038/srep02976

Zhang, S., & Zhou, W. (2018). Recreational visits to urban parks and factors affecting park visits: Evidence from geotagged social media data. *Landscape and Urban Planning*, *180*, 27–35. https://doi.org/10.1016/j.landurbplan.2018.08.004