COMPARING THE CHARACTERISTICS OF BICYCLE TRIPS BETWEEN ENDOMONDO, GOOGLE MAPS, AND MAPQUEST

Angela Schirck-Matthews  
University of Florida, Geomatics Program  
3205 College Avenue  
Ft. Lauderdale, FL-33314  
Tel: 954-577-6392 Fax: 954-475-4125; E-mail: mathwiz961@ufl.edu

Hartwig H. Hochmair, Corresponding Author  
University of Florida, Geomatics Program  
3205 College Avenue  
Ft. Lauderdale, FL-33314  
Tel: 954-577-6392 Fax: 954-475-6317; E-mail: hhhochmair@ufl.edu

Dariia Strelnikova  
Carinthia University of Applied Sciences  
Department of Geoinformation and Environmental Technologies  
Europastraße 4  
A-9524 Villach  
Tel: +43-5-90500-2007 Fax: +43-5-90500-2210; E-mail: dariia.strelnikova@edu.fh-kaernten.ac.at

Levente Juhász  
University of Florida, Geomatics Program  
3205 College Avenue  
Ft. Lauderdale, FL-33314  
Tel: 954-577-6392 Fax: 954-475-4125; E-mail: levente.juhasz@ufl.edu

Word count: 6198 words text + 2 tables x 250 words (each) = 6698 words

Revised version submitted Nov 1, 2018
Numerous online trip planners, such as Google Maps and MapQuest, offer trip recommendations for cycling among other transportation modes. Whereas these platforms are widely used, little is known about the routing criteria embedded into their routing algorithms, and, more specifically, how recommended trips differ from routes travelled by cyclists. This study uses GPS tracking data from the Endomondo bicycle app to extract bicycle trips in Miami-Dade County, Florida. It compares Endomondo trip characteristics to the characteristics of bicycle trips obtained from Google Maps and MapQuest and the corresponding shortest path. Results highlight the attributes relating to road usage along a trip (e.g. percent of residential roads), trip geometry (e.g. number of turns) and surrounding land cover (e.g. percent of tree canopy) which significantly differ between the analyzed trips. These attributes should be considered in the refinement of routing algorithms to more closely resemble observed routes and hence to improve the usability of suggested trips from a cyclist’s point of view. Furthermore, the study estimates a multinomial logit model on observed Endomondo trips that provides insight into the routing behavior of Endomondo commute and sport cyclists, respectively.

Keywords: Sports app, Endomondo, route choice, cycling, decision making
INTRODUCTION

Selection of the optimal route between two network locations involves the evaluation of conflicting criteria, such as travel time, route complexity, safety, or scenery, which is true also for cycling trips (1, 2, 3, 4). Since a shortest and general single criterion optimal path is typically not the ideal trip, the computation of the best route is inherently a multi-criteria problem (5). Therefore, a thorough understanding of cyclist preferences with respect to the different criteria involved in the decision making process is important for the development of online trip planning systems for cyclists (6, 7, 8) and for the design of efficient cycling networks (9). Crowd-sourced information, such as geo-tagged shared images (e.g. Flickr), or GPS tracking data from bicycle apps (e.g. Strava, Endomondo) or route sharing Web sites (e.g. GPSies) provide big data for the elicitation of route selection behavior from individual cyclists (10, 11). These crowd-based data sources can supplement more traditional observation methods for cycling behavior such as phone interviews, intercept surveys, preference surveys or customized GPS tracking solutions. Compared to aggregate level studies (12, 13), disaggregate analysis frameworks have the advantage of better capturing the behavioral relationship between cyclist route preference and its determinants, such as on-street parking or roadway physical characteristics (14).

Although several commercial platforms provide routing directions for cyclists, most of these platforms do not reveal their routing algorithms and applied routing criteria. To quantify the match between observed cycling behavior and trips generated by such trip planners this study compares the characteristics between selected trips in Miami-Dade County logged on the Endomondo bicycle tracking app, and corresponding cycling trips suggested by Google and MapQuest. Strava, founded in 2009, and Endomondo, founded in 2007, are currently the most widely used bike tracking apps (15), with trip recordings on all five continents. As opposed to Strava, Endomondo facilitates access to GPS tracking data at the point level through a combination of HTTP requests (16). It is therefore particularly suited to extract observed route trips from cyclists. Despite this advantage it has so far received relatively little attention in the transportation community compared to other GPS tracking platforms, and not much is known about routing behavior of Endomondo users. The presented study explores therefore the following two research objectives:

- Compare the characteristics of observed Endomondo bicycle sport and commute trips with cycling trips recommended by Google and MapQuest and shortest paths;
- Analyze the route choice behavior of Endomondo sport and commuter cyclists using a multinomial logit (MNL) model.

RELATED WORK

Crowd-sourced GPS tracking data from bicycle apps (MapMyRide, MapMyFitness, Garmin Connect, Strava, Endomondo) are nowadays a common source for travel behavioral analysis studies. For example, the Strava Metro product matches uploaded GPS tracks to the underlying road network, such as OpenStreetMap (OSM) or NAVTEQ NAVSTREETS, provides segment-based trip and athlete counts at different temporal resolutions, activity counts by generalized origins and destinations, and the geometry of individual trips that are generalized to the centroids of analysis zones (census blocks) along the trips. Analysis tasks based on Strava data include the estimation of bicycle travel volume and exposure as well as the identification of factors associated with increased or decreased ridership (17, 18, 19, 20). Related studies found that land use diversity, presence of bike paths and bike lanes as well as bicycle parks, lower speed limits, and proximity
to water bodies (lakes, ocean) are associated with higher cycling volumes, whereas on street
parking discourages cycling.

GPS data from bicycle apps have also been used for route choice modeling, showing
preference for routes with bike facilities and a higher number of signals per mile, but aversion to
routes with many turns (specifically left turns) and one-way facilities, based on users of the Grid
Bikeshare app (11). Analysis of user data from GPSies showed that cyclists prefer to take longer
routes with less tumultuous traffic and that serve as social spaces (21). A review of revealed
preference methods for bicycle route choice is presented in (22), which includes GPS devices
(e.g., GPS loggers), customized smartphone apps, crowdsourcing data (recreational/sport apps,
Amazon Mechanical Turk), participant recalled routes, and accompanied journeys.

The presence of alternative routing criteria (safe, scenic, simple) in the trips suggested by
the popular routing platforms Google and MapQuest in addition to travel time for car routing
was analyzed in (23), showing that these platforms likely incorporate simplicity as an
optimization criterion in addition to travel time, and that no neighborhoods seem to be excluded
to make routes safer. Although previous work has studied the variety of routing criteria offered in
online bicycle routing tools (8), only little research has been conducted that evaluates the
outcome of routing requests for cyclists from major search engines. One such study (24)
proposes a bicycle routing procedure that adopts a user perspective for leisure activities. The
authors then compare the characteristics of computed routes to those generated by RouteYou,
Strava, Google Maps, and Brouter.

Transportation researchers acknowledge the different types of biases that come with GPS
tracking data from sport apps, including user self-selection bias towards younger, male cyclists
(17), being in-line with the fact that the cycling community itself tends to be biased towards
younger, male riders (20). A comprehensive comparison of user samples collected by bicycle
smartphone applications in North America and samples of cyclists from traditional travel surveys
showed furthermore that undersampled population segments include some minority groups and
lower-income groups (25). Other systematic effects include geographic biases, e.g. higher usage
around university campuses (26), and the consequence of mass events, such as bicycle
competitions, on the distribution of trip counts on route segments (27). The accuracy of Strava
count data has been assessed through comparison with community survey data (28) and manual
cycling counts (18, 29). Strava data has also been used to measure the effect of bicycle
infrastructure improvements on short-term changes in ridership (30, 31).

Compared to Strava, Endomondo trips are accessible as raw GPS point data. Whereas
Strava provides trip counts per segment at a specific temporal resolution (for example, in one-
minute intervals), map matching and trip counting procedures needs to be conducted in pre-
processing steps when handling Endomondo data. That is, although working with GPS raw data
requires additional data preparation for further analysis, the data processing method is
completely in the hands of the data analyst and thus highly transparent. Furthermore,
Endomondo data sets contain metadata about individual cyclists, such as gender or age, whereas
Strava, provides such information for the entire purchased sample, if at all, and only for selected
products. In Endomondo, metadata about individual trips include workout type, distance,
duration, and average and minimum speed. Only few studies are published that analyze cycling
patterns emerging from Endomondo data. One study identified frequently traveled network
portions for recreational trips in three European cities (32).
STUDY SETUP
Network and trip data
To achieve the two research objectives, four types of trips between selected origin-destination pairs were required, namely an observed trip in Endomondo, bicycle trips recommended by the Google and MapQuest applications, and the geometrically shortest route. To be able to obtain consistent trip characteristics for the four trip types and to run a route choice model, all these routes needed to be matched to the same network graph in ESRI’s ArcMap, which was based on OSM in this study. OSM data was downloaded from geofabrik (http://download.geofabrik.de/) in shapefile format, clipped to Miami-Dade County boundaries, planarized in ArcMap to allow turns at each intersection, and then used to build a network dataset that considers one-way restrictions and prohibits travel on motorways (with some exceptions). Footpaths were allowed for the computation of cycling routes, since many observed routes, especially in Endomondo and Google, use segments that are classified as footpaths in OSM.

Although automated map matching algorithms between GPS tracks and a routing graphs exist (33, 21), a manual mapping approach was chosen in this study to ensure that trips from all sources were correctly matched to the road network in ArcMap. It was also checked that the chosen Endomondo origin and destination points were available for routing in Google Maps and MapQuest, and that all segments of observed bicycle trips were present in the OSM network graph. Sometimes, Endomondo trips were found to traverse along sidewalks or follow off-road trails that were not part of the OSM network, which therefore needed to be digitized in the OSM network. This process of creating additional links was also reported necessary for other studies that analyzed bicycle travel behavior from smart phone applications (34), where cyclists went through parks, parking lots, and open fields and on campus sidewalks. Overall, this manual approach, together with the search for suitable Endomondo route trips resulted in a relatively small sample of 31 sport and 31 commute trips that was used for further analysis.

Endomondo
As a first step worldwide Endomondo cycling trips from January to March 2016 were downloaded and extracted, following the data retrieval procedure described in (16). The approach requests all workout IDs for a specified user within a chosen time interval, followed by GPS point downloads for these workouts. Returned trip points with their latitude and longitude values were then written into a PostgreSQL/PostGIS database. As an initial filter only trips with a mean speed between 2 km/h and 40 km/h, a length of over 500 m, and a duration of over 2 minutes were retained for further analysis. 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). For this analysis 31 sport and 31 transport (commute) trips in Miami-Dade County were used, where circuitous trips were excluded. Where necessary, trips were trimmed to a shorter portion of the original route that more closely reflected a directed travel from point A to B. Waypoints along these trips were then digitized in ArcMap, and routes were regenerated using the ArcGIS Network Analyst extension, based on a network dataset built from the planarized and extended OSM network. The 62 Endomondo trips were from 47 different users. Out of these 47 users, 38 users provided gender information in their user profiles; from these, 28 (73.7%) users were male, which is a lower proportion than for Strava with values above 80% (26, 20), but it is also slightly lower than the 77.9% male among Miami-Dade residents who commute by bicycle. Furthermore, 24 Endomondo users provided a date of birth in their profiles, which resulted in a median age of 46.5 years (mean = 45.8, min = 26.2, max = 66.3) at the time of travel.
Google, MapQuest, and shortest path

Google provides a bicycle routing option both on the Web based Google Maps interface as well as through an API. To obtain a Google route, the geographic coordinates of trip origin and destination of the corresponding Endomondo trip were entered in the online search fields of Google Maps. The application then returns one suggested route and usually some alternatives. The suggested route was then replicated in ArcMap through digitizing waypoints on the OSM network layer and running the Network Analyst. Google’s Directions API allows to download waypoints in JSON format which can be loaded into a GIS. These routes are, however, based on the road network that Google uses, and not on OSM. Therefore, this approach would require map matching as well.

MapQuest offers cycling directions only through an API\textsuperscript{1}, with the option to choose a default network or OSM as base data. For the presented study, the default network was chosen, so that, like with Google Maps, MapQuest’s proprietary road network could be evaluated as part of the chosen route. Using a HTML site that was customized with JavaScript code the MapQuest API was called with trip end point coordinates as input. The returned suggested cycling route was shown on a map and subsequently digitized on the OSM network layer. In addition, the geometrically shortest path between the same origin and destination point was determined in the Network Analyst under consideration of one-way restrictions.

FIGURE 1a shows for a chosen origin-destination pair on Miami Beach the trips based on the four sources. The Endomondo route to the left has most turns per km, follows primarily residential roads, and has a detour of 20.2\% compared to the shortest path. The Google based route to the right runs primarily on a footway (with bicycles allowed) along the beach and has the largest detour among all routes with 35.7\%. Compared to the first two routes the MapQuest route is simpler and more direct with a detour of only 10.6\% and follows mostly secondary roads. The shortest route (dashed) follows also mostly secondary roads and has one more turn than the MapQuest route. This example illustrates the different characteristics of bicycle trips generated from different sources. FIGURE 1b shows the same set of routes on top a 2-m resolution land cover grid which was used to measure the percentage of different land cover categories within a 15-m buffer around the identified routes.

\textsuperscript{1} https://developer.mapquest.com/documentation/directions-api/
GIS data layers
Determining the characteristics of routes required several GIS data layers. The percentage of road categories along a route was based on OSM highway classes, some of which were grouped to simplify road categories. For example, highways tagged “cycleway”, “path”, and “track” were grouped to class cycleway, and highways tagged “footway” or “pedestrian” were grouped to a class footway. Traffic signals were also taken from the OSM data set. During data preparation, OSM traffic signals, which are point features located at the end nodes of OSM street segments, were spatially aggregated to avoid double counting of traffic signals at major intersections. Traffic signals were then counted within a 15-m buffer around a trip polyline and reported as the number of traffic signals per km. Turns were considered as such if the turn angle between two connected route segments enclosed an angle of at least 40 degrees, and if the turn took place on a network junction with at least three adjacent road segments.

The 2-m resolution land cover map stems from an earlier tree canopy assessment study for Miami-Dade County (35) and was supplemented with GIS layers from the Miami-Dade County GIS Hub (e.g., location of bay water). Part of the land cover map with its land cover categories is visualized in FIGURE 1b. Furthermore, scenery related Flickr images were extracted using the Flickr API using keyword-based filters. Their density was computed within the 15-m buffer around each route. This was motivated by earlier studies showing that the density of shared images is higher along scenic routes than fastest routes (36, 10).

Analysis methodology
For both research objectives attributes of trips and their surroundings with a buffered area needed to be extracted from the different GIS layers, which was achieved through a customized C# script within ESRI’s ArcObjects framework. Computed attributes that were considered in this study were classified as belonging to the attribute groups road category, trip geometry, trip feature, and land cover, as follows:

FIGURE 1 Four bicycle trips between a trip origin and destination point in Miami Beach on top of OSM (a) and land cover (b) background layer.
- **Road category**: % primary road, % secondary road, % residential road, % cycleway, % footway
- **Trip geometry**: % detour; number of turns left, turns right, and turns per trip km
- **Trip feature**: number of traffic signals per trip km, number of scenic Flickr images per trip km
- **Land cover**: % tree canopy, % inland water (lakes, rivers), % bay or ocean, % roads and railroads

On-street bicycle facilities (bike lanes, sharrows) were not considered for two reasons. First, on-street bicycle facilities tend to be sparsely mapped in OSM in the analyzed county (37). Second, due to the small sample size of analyzed trips the statistical power of the regression model is limited. It would be further reduced by additional predictor variables. Therefore, the study focused on readily available OSM highway categories.

### Median comparison
For objective 1, medians of the above listed trip attributes were compared statistically between the set of Endomondo trips and trips from other sources (Google, MapQuest, shortest). For this purpose, a Wilcoxon-Pratt Signed Rank test was applied, which tests matched-pairs data for a common median. The analysis was conducted separately for sport and commute trips, based on Endomondo trip metadata. This analysis gives insight into how trip characteristics differ between the different sources, and how closely trips from trip planners resemble traveled trips in Endomondo. While oftentimes a paired T-test is used to compare two population means, this requires data to be normally distributed. However, this was not the case since the distributions of trip attributes were skewed to the right (see FIGURE 2).

### Discrete route choice model
Discrete route choice methods empirically model and analyze a decision maker’s preferences from a set of available alternatives, i.e., the choice set (38). The multinomial logit (MNL) model is a commonly used framework where the decision maker is given more than two alternatives to choose from. Random utility choice models, like the MNL model, assume that the utility of an alternative consists of a deterministic and stochastic (random) component. The MNL model is derived from the assumption that the error terms of the utility functions are independent and identically Gumbel distributed (39). The probability that a given individual n chooses alternative i within the choice set C_n is given by

\[ P(i|C_n) = \frac{\exp(V_{in})}{\sum_{j \in C_n} \exp(V_{jn})} \]  

The Independence from Irrelevant Alternatives (IIA) assumption that comes with the MNL model suggests that route alternatives should not be overlapping. There are several approaches to correct for the overlap in route alternatives, such as using the path-size logit (PSL) model, which adds an additional path-size factor PS to the deterministic component of the route utility if it is overlapping with other routes (39). We use the PS formulation provided in (40) which is calculated as

\[ PS_r = \ln \sum_{a \in r} \frac{L_r}{N_a} \]
where \( PS_r \) denotes the overlap factor for route \( r \), \( a \) is the index of a network link, \( l_a \) is the length of link \( a \) that is a part of route \( r \) and \( L_r \) is the length of route \( r \). \( N_a \) is the number of alternatives in the choice set which contain link \( a \). In this study the choice set consists of four trips (Endomondo, Google, MapQuest, shortest), and the Endomondo trip is the chosen alternative. Selected predictor variables were then used to estimate unrestricted multinomial logit choice models. The Null log-likelihood \( (L^0) \) is the same in all estimated models. It describes the value of the log-likelihood function when all parameters are zero, i.e., when the alternatives are assumed to have equal probability to be chosen. \( L^0 \) is computed as

\[
L^0 = \sum_{j=1}^{N} \ln \frac{1}{|R_j|} \tag{3}
\]

where \(|R_j|\) is the number of choice alternatives available to the individual decision maker in decision situation \( j \), and \( N \) is the total number of decision situations. The final log-likelihood \( (L^*) \) is the log-likelihood of the estimated model. The rho-square value is an informal goodness-of-fit index that measures the fraction of an initial log-likelihood value explained by the model and is computed as

\[
\rho^2 = 1 - \frac{L^*}{L^0} \tag{4}
\]

To check for potential multicollinearity in the tested models bivariate correlations were computed between all used predictor variables, which were low to moderate with a maximum absolute Pearson’s \( r \) of 0.65 between two variables. The NLOGIT 6 software package was used to estimate the MNL models.

**RESULTS**

**Comparison of trip characteristics between different sources**

With regards to objective 1, TABLE 1 shows the differences of median attribute values between Endomondo trips and the other three trip types, separated into sport trips (left half) and commute trips (right half). A positive value in a cell means that the median value for the corresponding attribute is higher for the Endomondo trip than for the other trip, whereas a negative value means the opposite. Numbers printed in boldface show significantly different medians as determined by the Wilcoxon-Pratt Signed Rank test.
TABLE 1 Differences of median attribute values between Endomondo (E) and Google (G), MapQuest (M), and shortest (S) trips.

<table>
<thead>
<tr>
<th>Road category</th>
<th>Median differences (Sport trips)</th>
<th>Median differences (Commute trips)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E-G</td>
<td>E-M</td>
</tr>
<tr>
<td>% Primary road</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>% Secondary road</td>
<td>-9.33</td>
<td>**</td>
</tr>
<tr>
<td>% Residential road</td>
<td>9.99</td>
<td>*</td>
</tr>
<tr>
<td>% Cycleway</td>
<td>9.06</td>
<td>9.06</td>
</tr>
<tr>
<td>% Footway</td>
<td>0.40</td>
<td>*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trip geometry</th>
<th>Median differences</th>
<th>Median differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Detour</td>
<td>9.57</td>
<td>12.39</td>
</tr>
<tr>
<td>Turns per km</td>
<td>0.29</td>
<td>0.52</td>
</tr>
<tr>
<td>Left turns per km</td>
<td>0.06</td>
<td>0.19</td>
</tr>
<tr>
<td>Right turns per km</td>
<td>0.28</td>
<td>0.40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trip features</th>
<th>Median differences</th>
<th>Median differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic signals per km</td>
<td>-0.18</td>
<td>-0.60</td>
</tr>
<tr>
<td>Scenic Flickr per km</td>
<td>-0.56</td>
<td>-0.78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Land cover</th>
<th>Median differences</th>
<th>Median differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Tree canopy</td>
<td>-0.43</td>
<td>2.16</td>
</tr>
<tr>
<td>% Inland water</td>
<td>0.23</td>
<td>0.24</td>
</tr>
<tr>
<td>% Bay or ocean</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>% Roads or railroads</td>
<td>-2.65</td>
<td>-3.33</td>
</tr>
</tbody>
</table>

The positive difference values across all trip types and data sources for the % detour variable indicate that Endomondo trips are significantly longer than Google and MapQuest suggested routes, and, as is to be expected, also shortest paths. This means that Endomondo cyclists are not focused on seeking short travel routes. The differences in median detour values are consistently higher for sport trips than for the commute trips, which indicates that Endomondo commuters are more concerned about trip distance than recreational cyclists. Routes chosen by Endomondo users have a significantly higher turn frequency than Google and MapQuest suggested routes (but not shortest paths), both for sport and commute trips. The probability density functions for turns per km depicted in FIGURE 2 for sport (a) and commute (b) trips provide a graphical depiction of this pattern, where peaks around smaller values are more pronounced for Google and MapQuest than for Endomondo and shortest paths. This supports earlier findings that Google and MapQuest tend to consider simplicity in their routing algorithms (23).

Another commonality shared between sport and commute trips is the higher share of

Sample size = 31

* p ≤ .05; ** p ≤ .01; *** p ≤ .001 (not adjusted for multiple testing)

(-) Mdn(E) = Mdn (G), but Mean(E) > Mean(G)

(+). Mdn(E) = Mdn (M), but Mean(E) > Mean(M)

(+). Mdn(E) = Mdn (M), but Mean(E) < Mean(M)
residential (local) roads in Endomondo trips compared to Google and MapQuest trips (compare FIGURE 2c for sport activities). This means that Endomondo users prefer local roads with little traffic and accept more turns along the trip as a trade-off.

![Graphs showing turns and residential/secondary road percentages across different data sources.]

**FIGURE 2** Probability density function of median values for turns per km (a, b), % residential road (c), and % secondary road (d) across different data sources.

For sport activities, the share of secondary roads (typically major urban streets with traffic lights, similar to minor arterials or collector roads) on Endomondo routes is significantly smaller than for all other sources (compare FIGURE 2d). Similarly, a higher percentage of cycleways and footpaths is observed for Endomondo sport trips than for Google and MapQuest trips. This shows that the two commercial routing applications do not match the cyclist’s travel behavior entirely in these aspects of observed route preferences. For commute trips, no such discrepancy is observed, meaning that suggested Google and MapQuest trips match observed cycling.
behavior along these characteristics. Footways identified on Endomondo trips often run on sidewalks, especially on major roads where travel on the traffic lanes can be considered unsafe. However, use of sidewalks is not explicitly suggested in Google and MapQuest trips, although bicycle riding on sidewalks is permitted in the State of Florida per Section 316.2065, Florida Statues, Bicycle Regulations\(^2\), provided that a cyclist yields the right-of-way to any pedestrian. Although cyclists show a preference for low-traffic residential roads, major roads can often not be avoided without excessive detour when there is a bottleneck in the transportation network without cycling infrastructure (e.g. a bridge). In such a case, however, sidewalk information for major roads in routing directions provided by the trip planning application could be useful for a cyclist to judge the safety of a recommended route.

Endomondo sport trips feature fewer traffic lights and more inland water land cover along trips than MapQuest trips, and Endomondo commute trips run more often near the bay or ocean than Google and MapQuest trips. These facts together suggest that (a) current trip planner applications could be improved by better integrating these land cover attributes in route search, and that (b) offering different cyclist profiles (e.g. sport vs. commuter cyclist) with modified weighting functions could lead to routes that better meet the needs of different cyclist groups.

The density of crowd-sourced imagery along the routes did not differ significantly between Endomondo and other trip sources, suggesting that this source is not necessary to replicate typical Endomondo trips. However, this kind of data source could be useful to compute scenic and popular routes (10, 41).

**Model results**

For objective 2, various route choice models were manually built in a step-wise approach and estimated, where the pathsize variable was included in all models since it accounts for the correlation between the alternatives in the choice set. Variables that were found to be significant in the median comparisons (TABLE 1) were first added to build more complex models. Using this stepwise approach, one final model for sport trips and one final model for commute trips was identified that maximized the goodness of fit value and that at the same time retained significant parameters \((p < 0.05)\), except for the PS overlap predictor (TABLE 2). With all tested intermediate models (as well as the final models shown), the arithmetic signs of coefficients, if statistically significant, were as expected. TABLE 2 also reports the best fitting models for sport and commute trips with a significant PS overlap predictor. However, these models contain fewer significant variables and have a lower goodness of fit value than the other final models.

Estimated models in TABLE 2 show that Endomondo sport cyclists prefer local roads, and that they avoid roads with heavier traffic and traffic lights. Local roads combine the advantage of low volume traffic areas and the absence of signalized intersections, which should be more prominently considered in routing applications for sport cyclists. The commute-based models show preferences of expected road types (residential, cycleway) as well as footways, where the preference of footways can be partially attributed to the use of sidewalks along major roads that are probably deemed unsafe for on-road travel. One of the models also shows a preference for routes with tree canopy. Hence urban design should consider planting trees where possible. Green and recreational space has been found to be positively associated with time spent on cycling in earlier studies (42).

\(^2\) https://www.flsenate.gov/Laws/Statutes/2012/316.2065
TABLE 2 Estimation results of unrestricted multinomial logit choice models for sport and commute trips.

<table>
<thead>
<tr>
<th></th>
<th>Reduced sport</th>
<th>Final sport</th>
<th>Reduced commute</th>
<th>Final commute</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Secondary road</td>
<td>Coeff.</td>
<td>-0.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>-2.40 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Residential road</td>
<td>Coeff.</td>
<td>0.041</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>2.28 *</td>
<td>1.97 *</td>
<td></td>
</tr>
<tr>
<td>% Cycleway</td>
<td>Coeff.</td>
<td></td>
<td>0.026</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td></td>
<td>1.99 *</td>
<td>2.57 *</td>
</tr>
<tr>
<td>% Footway</td>
<td>Coeff.</td>
<td></td>
<td></td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td></td>
<td></td>
<td>2.18 *</td>
</tr>
<tr>
<td>Traffic signals / km</td>
<td>Coeff.</td>
<td>-1.126</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>-2.41 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Tree canopy</td>
<td>Coeff.</td>
<td></td>
<td>0.300</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td></td>
<td>2.35 *</td>
<td></td>
</tr>
<tr>
<td>% Roads/railroads</td>
<td>Coeff.</td>
<td></td>
<td>0.173</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td></td>
<td>2.30 *</td>
<td></td>
</tr>
<tr>
<td>Overlap factor PS</td>
<td>Coeff.</td>
<td>2.026</td>
<td>1.404</td>
<td>1.789</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>2.19 *</td>
<td>1.43</td>
<td>2.49 *</td>
</tr>
<tr>
<td></td>
<td>Parameters</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Observations</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Final LL (L*)</td>
<td>-36.19</td>
<td>-32.81</td>
<td>-37.67</td>
</tr>
<tr>
<td></td>
<td>Null LL (L0)</td>
<td>-42.98</td>
<td>-42.98</td>
<td>-42.98</td>
</tr>
<tr>
<td>$\rho^2$</td>
<td></td>
<td>0.16</td>
<td>0.24</td>
<td>0.12</td>
</tr>
</tbody>
</table>

* $p \leq .05$

CONCLUSIONS

Through comparison of characteristics of Endomondo bicycle trips with bicycle trips suggested on Google and MapQuest this study revealed which aspects of cyclist route selection behavior are already well incorporated in those applications, and which ones could deserve closer attention, such as preferred routing through residential areas or considering the proximity to water bodies. Comparison of Google and MapQuest trips with Endomondo and shortest routes also revealed some apparent criteria that are already considered in those commercial applications, such as route complexity, and avoidance of overly long detours. This expands previous research on the characteristics of trip suggestions for motorized traffic in Google and MapQuest (23). Previous work has also identified which type of routing criteria users prefer to be offered in the user interface of bicycle trip planners (43). The current study expands this research by evaluating routes that were generated on two major search engines. For future work we plan to apply this research to other routing services, such as the MapQuest Open Directions API that is based on OSM. Replicating this study for other geographic areas, as well as integrating on-road bicycle facilities, and increasing trip sample size could also identify additional relevant criteria in bicycle travel behavior.
AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: Study conception and design: HH and AM; data collection: DS, AM, HH, and LJ; analysis and interpretation of results: AM and HH; draft manuscript preparation: HH and AM. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES


