Transactions in GIS

Research Article Transactions in GIS, 2016, 20(6): 848–868

Conflation of OpenStreetMap and Mobile Sports Tracking Data for Automatic Bicycle Routing

Cecilia Bergman and Juha Oksanen

Department of Geoinformatics and Cartography, Finnish Geospatial Research Institute, National Land Survey of Finland

Abstract

This article investigates how workout trajectories from a mobile sports tracking application can be used to provide automatic route suggestions for bicyclists. We apply a Hidden Markov Model (HMM)-based method for matching cycling tracks to a "bicycle network" extracted from crowdsourced *OpenStreetMap* (OSM) data, and evaluate its effective differences in terms of optimal routing compared with a simple geometric point-to-curve method. OSM has quickly established itself as a popular resource for bicycle routing; however, its high-level of detail presents challenges for its applicability to popularity-based routing. We propose a solution where bikeways are prioritized in map-matching, achieving good performance; the HMM-based method matched correctly on average 94% of the route length. In addition, we show that the extremely biased nature of the trajectory dataset, which is typical of volunteered user-generated data, can be of high importance in terms of popularity-based routing. Most computed routes diverged depending on whether the number of users or number of tracks was used as an indicator of popularity, which may imply varying preferences among different types of cyclists. Revising the number of tracks by diversity of users to surmount local biases in the data had a more limited effect on routing.

1 Introduction

Smartphone-based mobility data has demonstrated its potential in the context of intelligent vehicle routing by providing valuable information at a scale and price unattainable by conventional monitoring methods, such as fixed sensors or non-consumer generated probe data (Batty 2013; Simontine 2012). We recognize the benefits of crowdsourcing for non-motorized traffic, for example, in the form of more comfortable paths for walking (Quercia et al. 2014) and cycling (Priedhorsky and Terveen 2008), and therefore investigate the problem of recommending popular cycling routes between two locations based on trajectory data from a mobile sports tracking application.

1.1 Route Planning Services for Cyclists

Cyclists often prefer to ride longer distances rather than the shortest way, based on both subjective and non-subjective factors (e.g. Dill and Gliebe 2008; Ehrgott et al. 2012; Menghini et al. 2010; Sener et al. 2009). Commuter cyclists, or more frequent riders in general, have proven to

Address for correspondence: Cecilia Bergman, Department of Geoinformatics and Cartography, Finnish Geospatial Research Institute, National Land Survey of Finland, Geodeetinrinne 2, FI-02430 Masala, Finland. E-mail: cecilia.bergman@iki.fi

Acknowledgements: The study is based upon work in the project SUPRA (Revolution of Location-Based Services: Embedded data refinement in Service Processes from Massive Geospatial Datasets) funded by Tekes (grant 40261/12), the Finnish Funding Agency for Technology and Innovation. We gratefully thank Sports Tracking Technologies Ltd. for providing us the workout tracking data, and Eimear Dunne for the language revision of the article. We would like to express our gratitude also to the three anonymous reviewers for their valuable comments.

 doi: 10.1111/tgis.12192

be more sensitive to distance than other utilitarian cyclists (Broach et al. 2012; Dill and Gliebe 2008), whereas recreational and sports riders who do not share the same time pressure are likely to place a higher value on other factors that make the route attractive. Traditionally, cyclists' route-finding has been a product of trial and error, and knowledge-sharing within the cycling community (Reddy et al. 2010). However, during the on-going renaissance of cycling as a mode of transportation as well as a popular sport and a recreational and tourist activity, a cycling option has been introduced in many route planners. Potential cyclists' lack of knowledge about routes that would be short, safe, simple and attractive at the same time (see Hochmair 2004) can hinder the transition from car to bike. Routing services are seen as a way to improve the attractiveness of cycling within the existing infrastructure and to increase its modal share as a sustainable means of transportation, which provides great benefits to the liveability of a city and its population's health (Dill et al. 2014; Su et al. 2010).

As cyclists' route choice decisions are influenced by many factors, which are often conflicting and context-dependent (e.g. Sener et al. 2009), existing route planners tend to offer the user multiple choices based on different optimization criteria, such as total elevation gain, vegetation, turns, and intersection with traffic. Considering that the best cycling route is hardly ever defined by any single route selection criterion but is rather a compromise between incommensurable objectives, it has been suggested that the interface of the route planner should support a trade-off between many preferred attributes (Hochmair 2004). Consequently, some routing services have aggregated different criteria into generally desirable objectives, such as "balanced" (*cyclestreets.net*), thereby easing the user's task. In addition to the limited information about the preferences of cyclists in different contexts (Lindsay et al. 2014), the availability of suitable attribute data of the required accuracy is one reason why many routing services are restricted to the local level. This applies to map data as well, as the cycling network is only partly consistent with the network of car traffic. Crowdsourcing has become a promising alternative for creating maps with a better coverage of cycling infrastructure (Hochmair et al. 2012), and can also transform routing services (Neis and Zielstra 2014; Shekhar et al. 2012). Cyclopath *(cyclopath.org)* is an example of a computational geowiki where cyclists can rate road segments according to their "bikeability", or add missing segments to the base map covering the area of Minnesota (Priedhorsky and Terveen 2008). Similarly, in Bikedistrict (*bikedistrict.org*), cyclists can rate segments in Milan on a three-stage scale from *"I like it! Do it more often"* to *"Don't ever take this road"*. A downside related to these implementations is that cyclists are required to recall the impression of their past experience on each block and manually add this to the service.

1.2 Mining Popular Routes from Crowdsourced Data

Mining meaningful collective information from crowdsourced data for automatic routing purposes has garnered increased interest in recent years. A number of studies have concentrated on recommending popular and attractive travel routes for tourists who need guidance in unfamiliar places. This work has been mainly based on other types of geo-referenced user-generated content than GPS trajectories, such as check-in records (e.g. Foursquare) and geo-tagged photos (e.g. Flickr) (e.g. Hao et al. 2010; Lu et al. 2010; Sun et al. 2015). If ordered by time, the aforementioned digital footprints can also be represented as sparse trajectories, i.e. sequences of time-stamped locations. Even with GPS trajectories, these methods have been more oriented towards finding interesting locations by extracting stay points and travel sequences between them than optimal routes in a road network (e.g. Yoon et al. 2012; Zheng et al. 2011).

Recording outdoor workouts with GPS-enabled handheld devices has become popular alongside the increasing number of location-based services which allow users to examine past exercise tracks and share them with friends or within web communities. The resulting GPS trace repositories of sports tracking applications provide a vast amount of local knowledge (Priedhorsky and Terveen 2008) in a corresponding way to the experience-based hidden intelligence of GPS-equipped commuters (Hendawi et al. 2013) or taxi drivers (Yuan et al. 2014), which can in turn be used to enhance route guidance services. To provide turn-by-turn route suggestions, we need either to mine a routable network from GPS trajectories (e.g. Chen et al. 2011) or, as in our case, to map trajectories to their equivalent road segments in a road network. This is a non-trivial task due to the noisiness of smartphone GPS data and the incompleteness and inaccuracy of road map data. The effect of matching accuracy has, however, been largely ignored in studies related to mining optimal routes from a massive amount of dense trajectory data (e.g. Chang et al. 2011; Hendawi et al. 2013). For an overview of map-matching methods, typically classified into geometric, topological and advanced algorithms, we refer the reader to Hashemi and Karimi (2014) or Quddus et al. (2007).

The nature of the crowd behind the data also requires more attention. Understanding how datasets are created and by whom is a prerequisite for successful geo-applications (Mullen et al. 2014). Analyses of human mobility based on crowdsourced trajectory datasets have been criticized for their biased sampling resulting from socio-demographic (e.g. age, gender, wealth) variation in, e.g. smartphone possession and usage (Yue et al. 2011), and for their focus on single-source empirical data (Zhang et al. 2013). It is also known that in projects which rely on volunteered user-generated content, contributions are typically very unevenly distributed (Adamic and Huberman 2002; Priedhorsky et al. 2010), which is known as "participation inequality" or the "90-9-1 rule" (Neis et al. 2013; Nielsen 2006). However, how this might affect routing, for example, has received little attention.

1.3 The Present Study

The tracking data provided by Sports Tracker (http://www.sports-tracker.com) has previously been used to investigate city dynamics (Ferrari and Mamei 2013) and to create heat maps which enable cyclists to compare potential routes by means of visual data mining (Oksanen et al. 2015). The main objective of our study was to extend the aforementioned work by providing the user with automatic popularity-based routing in a street network, by combining recorded workout trajectories with cycling-specific network data extracted from OpenStreetMap (OSM: *openstreetmap.org*). Previous studies where exact routes of cyclists have been inferred (Broach et al. 2012; Dill and Gliebe 2008; Menghini et al. 2010) have relied mainly on network data provided by local stakeholders, which has often been manually enhanced to include all relevant links. However, as a result of the high costs of vector data maintenance and the limited interest of authoritative and proprietary actors in mapping the infrastructure of non-motorized transport modes, many routing applications for cyclists which have emerged recently use the network data of OSM. The first national route planner for cyclists, OpenRouteService.org, was launched in April 2008, less than four years after the OSM project had been introduced (Schmitz et al. 2008). In addition to its good coverage of cycling paths, the OSM data is accessible to everyone and has a relatively homogeneous coding scheme for the whole world (Hochmair et al. 2012; Loidl et al. 2014).

We compare the effective differences between a modified Hidden Markov Model (HMM)-based algorithm, which calculates the most probable route using distance and topological data of the network, and a geometric point-to-curve method that considers only the distance

between each GPS track point and the road segments. Both matching methods are well-studied, but how differences in their accuracy affect popularity-based routing has not been investigated previously. After counting the number of tracks and users who have traversed each segment, we use the weighted networks to find routes with maximum popularity given a source location and a destination by Dijkstra's shortest path algorithm. Another objective of this study was to investigate whether alternative cost functions for calculating popularity indicators would affect routing due to the biased nature of tracking data. The compared cost functions are based on: (1) the number of users who have traversed each road segment; (2) the number of tracks on each segment; and (3) a value that combines the latter with Simpson's diversity index describing the distribution of tracks between different users.

Being sensitive to the variation in factors affecting route choices in different contexts and based on worldwide network data, our solution could in principle be implemented globally (given sufficient quality of OSM data and the availability of tracking data). Another tracking application, Strava (*strava.com*), provides a service with similar aspirations, but information related to its implementation does not cover issues related to the generation of the weighted network, calculation of popularity indicators, and their impact on routing (see Robb 2013). The remainder of the article is organized as follows. Section 2 presents the datasets and their pre-processing, followed by the methods used in the integration of map data and GPS trajectories resulting in a routable weighted graph. Section 3 presents the results, while the limitations of both the approach and the results are further discussed in Section 4. Conclusions are drawn in Section 5.

2 Data and Methods

2.1 Sports Tracking GPS Trajectories

The study area of southern Helsinki (Figure 1a) encompassed a total of 29,958 publically viewable cycling trajectories, of which approximately 80% had a valid timestamp. The data was collected by the mobile application Sports Tracker users in the period from January 2010 to June 2013. The following pre-processing steps were taken:

1. **Data reduction**. The original data was collected with a dense sampling interval of approximately one second. This may, however, vary during a workout, e.g. because of the poor availability of GPS signals in enclosed spaces or stopping at crossings (depending on the settings set by the application user, tracking can be paused automatically at low speed). On average, in 1.7e-03% of the cases, the difference between two consecutive track points was more than two seconds; in 2.5e-04% more than five seconds; and in 8.9e-07% more than one minute. The size of the dataset was reduced to consist of approximately equidistant points. Our empirical tests showed that points could be discarded as long as the distance from a given point was under the threshold of 30 m. In addition, trajectories of less than 20 points were discarded.
2. **Filtering.** In the beginning of each workout, the application user is asked to select a type of activity and the trajectory is labelled accordingly. In this study, the transport mode could therefore be taken as a given. Only tracks with valid timestamps were included in the dataset. From the remaining tracks all consecutive points whose speed exceeded 72 km/h (20 m/s) were discarded as noisy observations and, in addition, a few tracks (e.g. high-speed exercise sessions in the velodrome) were filtered out by their high average speed of over 40 km/h.

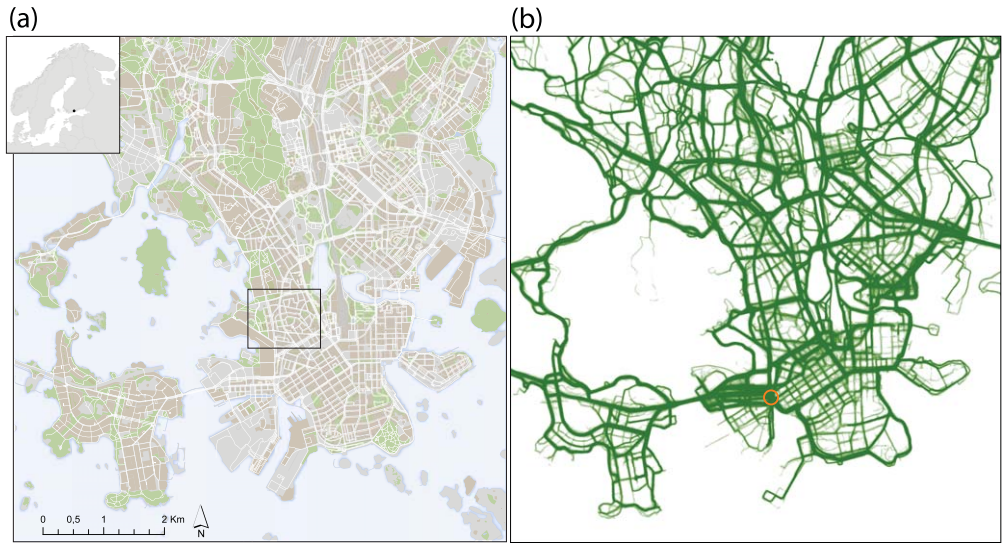

Figure 1 (a) Study area of southern Helsinki. The area inside the small rectangle is enlarged in Figure 3; and (b) Cycling trajectories within the study area drawn as lines with 5% opacity. The orange circle points the place of the square in Figure 12

3. **Slicing for long gaps.** If the distance between two consecutive track points exceeded the chosen distance threshold of 200 m, inferring the taken route between them was considered uncertain in an urban environment, and hence the trajectory was divided into multiple parts. Outages or "silent durations" may result from the unavailability of a GPS signal due to environmental conditions, or forgetting to turn on the tracking application after a break. Altogether almost 3,000 such outages were discovered in the filtered dataset.
4. **Interpolation for short gaps.** If the outage was more than 70 m but less than the slicing threshold (200 m), new points were interpolated along a straight line (to confirm the approximate equidistance we used the interval of 30-40 m) connecting two consecutive points following the method used by Thiagarajan et al. (2013). There were on average 1.3 short gaps in each track, i.e. altogether approximately 30,000.

The final dataset after all pre-processing steps encompassed a total of 23,290 tracks recorded by 1,994 cyclists (Figure 1b). The distribution of the tracks between users (Figure 2a) was extremely skewed; a few active users have shared hundreds of tracks each, while 65% of users have recorded at most 10 tracks. Less than 5% of users have recorded 50% of the tracks. The two-peaked diurnal distribution of the tracks (Figure 2b) indicates that the dataset included a high proportion of commuters. The median length of the tracks (including also the parts recorded outside the study area) was 13.5 km, whereas the mean length was almost 21 km. In addition, we divided the tracks into two sets based on the distance between their start and end points. When the distance exceeded a tenth of the length of the whole trajectory, the track was classified as a route between two different locations ("A-to-B tracks"). Otherwise, the track was classified as a circular route ("loop tracks"). Based on this simple division, we found out that only one-fourth of the tracks in the urban study area were circular workouts.

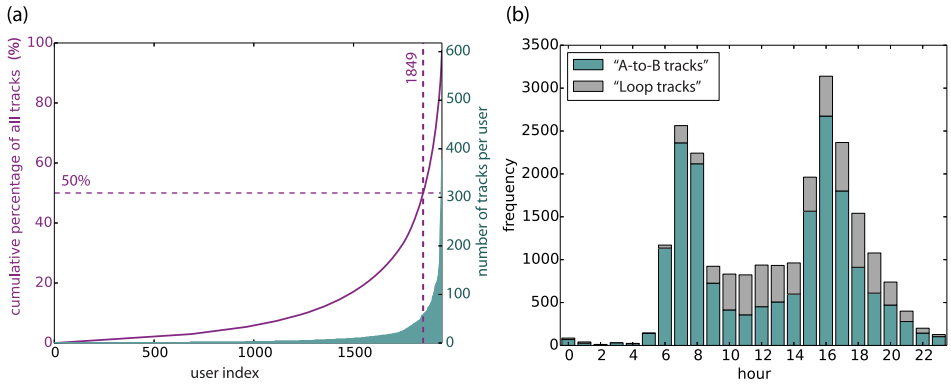

Figure 2 (a) The distribution of the recorded tracks between users. Users on the x-axis are ordered according to their number of tracks; and (b) The diurnal distribution of the tracks according to the starting time of the workout

2.2 *Extraction of "OSM Bicycle Network"*

To avoid routing along paths on which cycling is prohibited and to reduce the size of the dataset, only segments that are accessible to cyclists were included in the network. For instance, all motorways, footways, steps, and paths were excluded unless an additional tag was used indicating that cycling is permitted. The selection queries are listed in Table 1 (for a more comprehensive introduction of OSM coding conventions, see Hochmair et al. 2012). All extracted OSM features were line elements except for the pedestrian squares which are area features.

Table 1 Queries of road segments where cycling is allowed (in Finland) were conducted over the latest OSM export (August 28th 2014) provided by Geofabrik (http://download.geofabrik.de/). In OSM, highway tag is used to describe the role of all kinds of passages in the road network. The ones mentioned in the first column are included in the "OSM bicycle network" if the additional requirements are fulfilled

Highway type (key=highway)	Requirements related to additional tags of access permissions[a]	
Cycleway	No additional requirements	
(all)	Keep if	Bicycle=yes |designated |official |lane
Primary[b], secondary[b], tertiary[b], trunk[b], residential, living_street, unclassified, track	Drop if	Bicycle=no or (tunnel=yes and (access=no |private))[c]
Service	Drop if	Bicycle=no or access=private |no or service=parking_aisle

[a] Key=value; if many possible values are related to the same key, they are separated by |

[b] Also with _link-tags

[c] The usage of additional tags access=private and access=no was very heterogeneous, and hence they have not been excluded more comprehensively

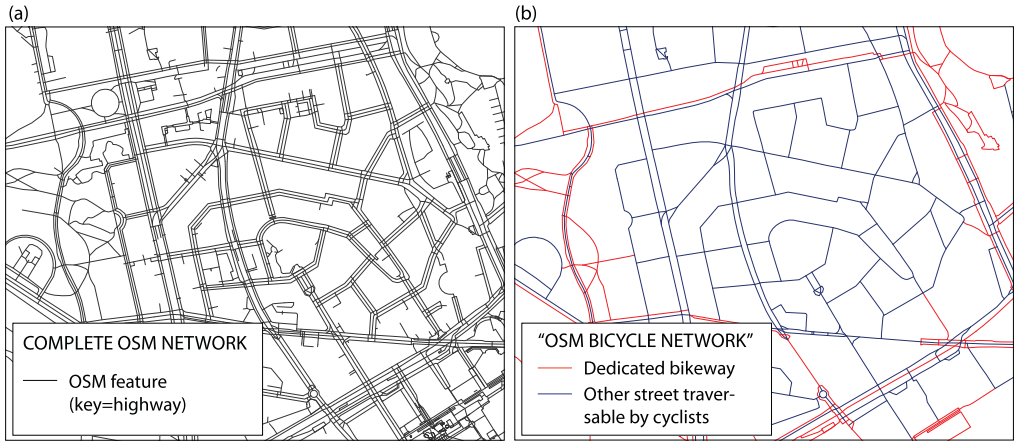

Figure 3 An extract of the network data showing: (a) all OSM features with key = highway; and (b) the extracted "OSM bicycle network", where dedicated bikeways are in red and other streets traversable by cyclists in blue

They were converted to single-line features to allow routing along their edges. Although this does not exactly reflect routes in real life, it preserved the network topology critical for our study. The resulting lines were split up at intersections and then grouped so that they formed solid links between intersections. Short dead ends of less than 200 m were excluded and, in addition, longer segments were split into shorter parts of equal length (at most 200 m; see Robb 2013). One-way restrictions were included according to the dedicated attributes. The final routable "bicycle network" (Figure 3b) included all passages where cycling is permitted, covering both dedicated bikeways (Table 2) and driveways that are traversable by cyclists. (Although cycling in driveways that have a parallel cycle track is in most cases prohibited by Finnish law (Road Traffic Decree 18§), cycling on them is popular and, more importantly, excluding them could have caused topological errors.)

2.3 *Map-Matching*

We compared two map-matching algorithms: a geometric point-to-curve method and an advanced HMM-based algorithm. The geometric method considers only distance when matching each point to the nearest segment, with a few exceptions. First, only road segments within a distance of 30 m of each point were considered as candidates. Track points that were far from

Table 2 Classification of bikeways in the study area based on OSM tagging

Highway type	Requirements related to additional tags[a]
Cycleway	No additional requirements
(all)	Cycleway[b]=lane |shared_busway |opposite_lane, bicycle=yes |designated |official

[a] Key=value; if many possible values are related to the same key, they are separated by |

[b] Including cycleway:left, cycleway:right, cycleway:both

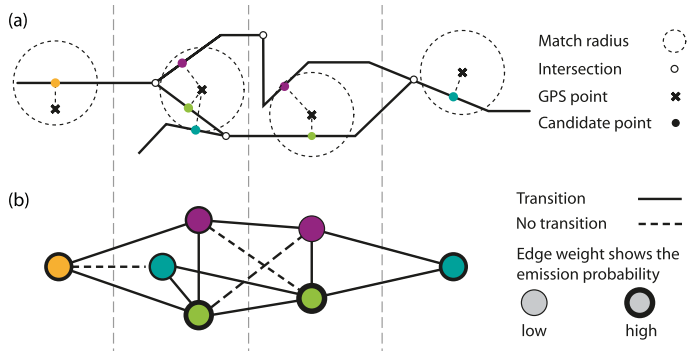

Figure 4 The idea of the HMM-based algorithm, illustrated as: (a) a schematic road network; and (b) a corresponding trellis diagram representing the Viterbi algorithm. The number of candidate segments is limited by the match radius. Based on the emission and transition probabilities, Viterbi iteratively calculates the most probable path in the network

paths where cycling is permitted could therefore be discarded. Second, if there was at least one bikeway (Table 2) available within 20 m, the track point was always matched to the nearest bikeway. This means that if there was a cycle track running parallel to a driveway, the aim was to match all cyclists who had been traversing either one of them to the bikeway, which we also want to return as a result of a routing request. It can also be noted that according to the observations by Lindsey et al. (2013) the accuracy of handheld GPS receivers does not support their usability in monitoring the use of bike lanes or other facilities.

HMM-based algorithms are in general the most accurate alternative for map-matching (Wei et al. 2013), and have therefore attracted much interest (e.g. Goh et al. 2012, Newson and Krumm 2009; Thiagarajan et al. 2013; Torre et al. 2009). The general idea of HMM is to find the most probable path through many possible states, where some state transitions are more likely than others. Individual road segments comprise the (hidden) states, whereas workout track points are the state measurements (visible observations) emitted by the states with a particular probability (Figure 4). Knowing the resulting observations, the probability of a state producing a certain observation (emission probability), and the probability of a state transitioning to another state (transition probability), the most likely sequence of states that could have resulted in the observed sequence of GPS points is calculated (Torre et al. 2009) using a dynamic programming technique called Viterbi decoding (see Forney 1973).

Emission probability was calculated for each observation-candidate state pair, based on the Euclidean distance between a point and a segment modelled as a zero-mean, normally distributed random variable, as suggested by Hummel:

$$Emission\ probability = \left(\sqrt{2\pi}\sigma\right)^{-1} \exp\left(-d(2\sigma)^{-1}\right), \tag{1}$$

where d is the Euclidean distance between the GPS point and road segment, and σ is the standard error of GPS positioning. To direct all trajectories to the cycle track(s) running parallel and next to a driveway, bikeways (Table 2) were favoured by multiplying their emission probability by three. In addition, transitioning to a bikeway was defined as two times more likely than to other road segments. We used a constant transition probability of 0.2 (bikeways 0.4) between all connected segments, which gave good results when used with a standard deviation $\sigma=4$.

The transition probability of staying on the same road segment is always one. The choice of parameter values is further discussed in Section 3.

The problem of missing connections was handled by discarding the track point if no route was found, i.e. if no road segments were available within the match radius, or the transition probabilities to all road segments within the radius were zero. As the HMM was not originally designed for large road networks, its computational efficiency was improved by calculating the emission probabilities only for road segments within 60 m of each track point. The threshold should be large enough to provide valid results while avoiding false routes, considering the road density of the network.

2.4 Cost Functions

After being processed via map-matching, trajectories could be represented as sequences of road segments. When calculating the popularity indicators assigned to segments, each traversed segment was counted at most once per trajectory, and the minimum number of users and tracks was set to five. After all, there is no reason to completely avoid non-traversed segments in routing.

The following impedance values were computed for each segment:

1. normalized inverse User Count

$$niUC = \left(\frac{u}{l}\right)^{-1} \tag{2}$$

2. normalized inverse Track Count

$$niTC = \left(\frac{t}{l}\right)^{-1} \tag{3}$$

3. normalized inverse Diversity-attentive track count

$$niDIV = \left(\frac{t\left(1-\sum p_i^2\right)}{l}\right)^{-1} \tag{4}$$

where u is the number of distinct users, t the number of tracks, and l the length of the segment. In the Simpson's diversity index $\left(1-\sum p_i^2\right)$, p_i is the ith user's proportion of tracks on the segment; in other words, the higher the diversity index, the more evenly tracks are distributed between users (e.g. McDonald and Dimmick 2003).

The inverse is taken because, although the goal is to maximize the resulting route's popularity, we use a minimization algorithm, which finds the route with the lowest cost. Normalizing the popularity indicator by the length of the segment makes segments comparable and allows us to avoid excessively long meandering routes. This is important considering that preferable routes are always compromises between distance and popularity. Because popularity is very unevenly distributed across the network, we also test how a logarithmic transformation of the popularity indicators will affect routing.

3 Results

3.1 Performance of the Map-Matching Methods

To measure the performance of the map-matching methods, 50 randomly selected workout trajectories were manually matched to the road network. Manual matching provided reasonable

Table 3 Performance of the map-matching algorithms: (a) with; and (b) without preference of bikeways

	TRUE POSITIVES		FALSE POSITIVES	
	HMM	Point-to-curve	HMM	Point-to-curve
(a) Mean hit ratio (with bike pref.)				
Number of segments of matched route	0.90	0.70	0.11	0.29
Length of matched route	0.94	0.91	0.09	0.31
(b) Mean hit ratio (without bike pref.)				
Number of segments of matched route	0.78	0.61	0.20	0.43
Length of matched route	0.85	0.84	0.16	0.48

"ground truth", as our aim was not to estimate the quality of the OSM network in the study area, and because the correct road corridor could be visually inferred largely without problems due to sufficient point density and accuracy. The reference data did not need to exactly correspond to the route taken if the network was missing a connection. Instead, tracks were matched as we aimed to match them automatically; e.g. in the case of parallel segments, the tracks were matched to bikeways, when possible. In cases where it was difficult to determine the correct side of the road, the track was mapped considering the cyclist's previous and next turns.

The method used in OSM for modelling intersections with multiple short segments was challenging for both matching methods. We therefore compared the fitted and manually matched data by determining both the number and length of segments which they had in common (Table 3). The HMM-based method achieved better results than the point-to-curve, especially with respect to the number of correctly matched segments (i.e. true positives). However, when comparing the length of the "correctly" matched route, the methods performed almost equally well, both catching over 90% of the length of the reference data. In other words, both methods do well with long segments, whereas the point-to-curve method misses short ones. The percentage of matched segments that have not been traversed according to the reference data (i.e. false positives) was clearly higher with the point-to-curve method, which can be seen as a higher popularity value of small intersecting roads (Figure 5).

Prioritizing bikeways significantly improved the performance of both matching methods (Table 3); e.g. with the HMM-based method, the fraction of the length of correctly matched routes increased from 0.85 to 0.94. When examining the effect of changing each parameter at a time while keeping all others constant, the sensitivity of the HMM-based method is less obvious when routes are considered for their entire length (Figures 6a-f). However, matching is sensitive to parameters in places with parallel ways, which can be seen in Figure 6g where the parameter combination used in this study resulted in a route matched along bikeways. The effect of σ is clear; values above four lead to a steady decrease in performance as less emphasis is placed on the distance between GPS points and road segments, whereas from two to four the ratio of false positives decreases faster than the ratio of true positives (Figures 6a and 6d). The other two parameters describe to what extent bikeways are favoured. Three-fold emission probability gives the highest ratio of true positives, and although the performance stays on a high level statistically with larger values (Figures 6b and 6e), visual inspection of matched

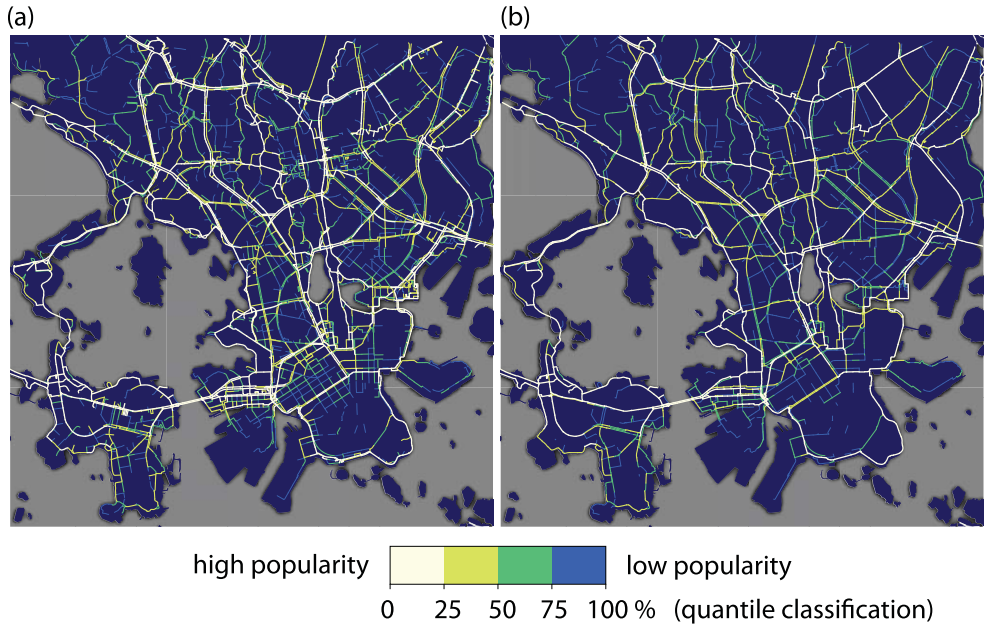

Figure 5 Popularity of the network measured as number of users based on: (a) the point-to-curve; and (b) HMM-based matching method. For clarity reasons, only segments with over 30 users are presented

routes confirmed that favouring bikeways too much will result in false matches. The effect of increasing the transition probability of bikeways increases until 0.6, but so does the ratio of false positives (Figures 6c and 6f). Obviously, searching for an appropriate combination of parameters that would conform to all diverse situations is difficult. Point-to-curve had two parameters: the matching radius (30 m), and the radius within which bikeways were favoured (20 m). Increasing the latter parameter enhanced the method's robustness to GPS errors, but at the same time increased the probability of matching points to bikeways along incorrect roads.

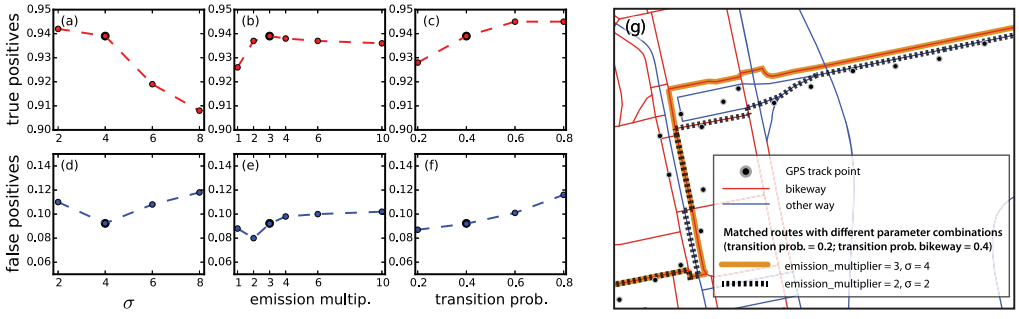

Figure 6 The HMM-based method's sensitivity to parameters measured by the change of the ratio of (a-c) true positives and (d-f) false positives. In each case, only the parameter in question is changed and all other parameters correspond to the values marked with a larger circle ($\sigma = 4$, emission multiplier = 3, transition probability of bikeways = 0.4, i.e., two-times the transition probability of other ways). (g) The parameter sensitivity of the HMM-based method shows in places with parallel ways

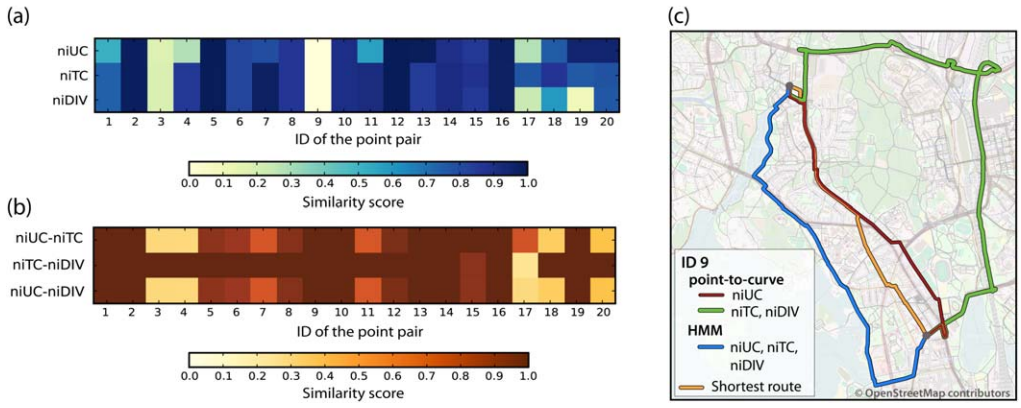

Figure 7 (a) Similarity of the routes computed with the point-to-curve-weighted network and HMM-weighted network, with different cost functions; (b) Similarity of the routes computed with the HMM-weighted network. For example, the values on the niUC-niTC axis represent similarity in length between the routes based on cost functions niUC and niTC; and (c) Route suggestions in case 9

During the manual matching process, many issues that were not correctly resolved by the HMM-based map-matching method could be identified. Most errors resulted from cycling on footpaths (e.g. along a one-way street where contraflow cycling is not legal in Finland); incompleteness of the OSM data; and changes in the road infrastructure during the study period. Furthermore, because most cycle tracks in Helsinki are two-way and often exist on both sides of the street, the algorithm can map the points to the cycle track on the wrong side due to inaccurate GPS measurements in urban canyons. Similarly, tracks can be matched to either one of the streets heading in the correct direction in places where a parallel residential street runs beside an arterial road. The algorithm was also unable to completely avoid matching to redundant side roads, and in a few places it tended to favour longer segments due to the transition probability. In addition, the selected threshold for reducing point density appeared to be too high in one complex intersection, where tracks were therefore matched to a parallel way.

3.2 Effects on Routing

For qualitative and quantitative analysis of the differences of map-matching performance and cost functions for routing, a set of 20 randomly generated origin-destination pairs was created such that the Euclidean distance between the origin and destination was at least 2 km. We calculated the share of the length where the routes follow exactly the same paths. The similarity score was calculated by dividing the length of the common route (l_{common}) with the average length of the routes computed with the network based on HMM (l_{hmm}) and point-to-curve (l_{ptc}) as follows:

$$l_{common}\left(\frac{1}{2}\left(l_{hmm}+l_{ptc}\right)\right)^{-1} \tag{5}$$

On average, the suggested routes followed identical paths 75% of their length irrespective of the map-matching method (Figure 7a). In the case where the routes diverged most clearly

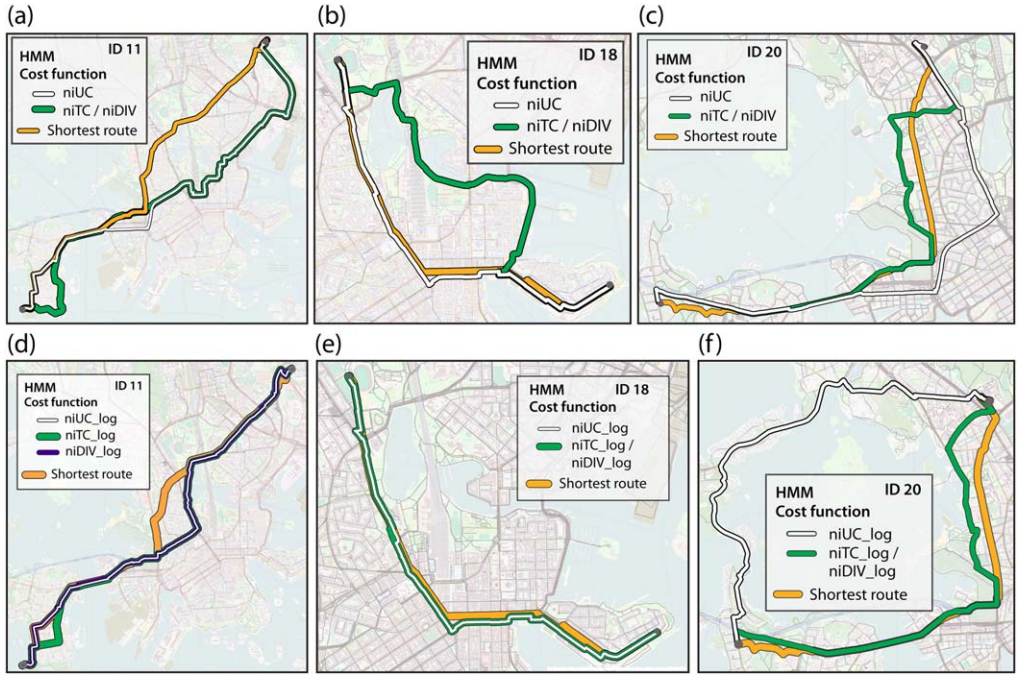

Figure 8 Routes in cases 11, 18 and 20 based on (a-c) cost functions presented in Section 2.4, and (d-f) cost functions where the number of tracks (t) and the number of users (u) are replaced by $\log(t)$ and $\log(u)$, respectively (© OpenStreetMap contributors)

(ID 9; Figure 7c), the point-to-curve-weighted network produced routes that had proportionally fewer short segments than the route computed using the HMM-weighted network. Based on this, the differences in routing might be explained by the point-to-curve method's inability to catch short segments, which could have a significant impact on routing despite their length. In general, the point-to-curve-weighted network routes were longer, but had fewer segments than the HMM-based routes.

Figure 7b presents the similarity of routes with different cost functions in a corresponding way. It is worth noting that routes based on the number of users diverge in most cases from routes based on the number of tracks. It appears that selecting the number of users as an indicator of popularity results in routes that are likely to be popular among utilitarian cyclists who often ride shorter distances and choose the most direct way to their destination (workplace, service, etc.), whereas the number of tracks returns routes also preferred by sports and other recreational cyclists (Figures 8a-c). Figure 9 shows separately the number of tracks between two locations ("A-to-B tracks") and the number of circular routes ("loop tracks"). While many paths are preferred by both groups, exercise riders are more clustered on routes along the coastline. The differences suggest that, unlike utilitarian cyclists riding mainly for purposes other than recreation, cyclists who record their workouts more frequently to follow their performance prefer to take the indirect, scenic, and quieter route option.

Irrespective of the map-matching method, the two cost functions that use the number of tracks (niTC, niDIV) give very similar results: in almost 90% of the cases, the routes follow identical paths. This was not surprising, considering that the few small Simpson's diversity

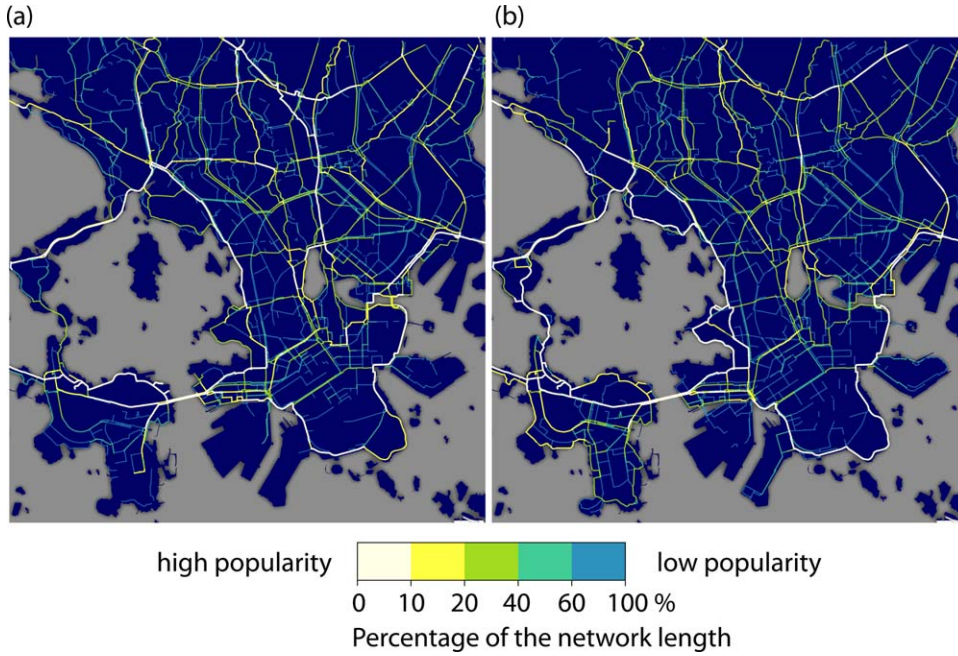

Figure 9 The most popular routes of: (a) cyclists riding from one location to another; and (b) cyclists riding loops which start and end at the same location. For clarity reasons, the presented network includes only the most popular one-third of the segments. Classification is based on length, i.e. the first class includes the most popular 10% of the network length, the second class the second most popular 10%, etc.

indices are related to segments outside the main cycling routes (Figure 10a). Figure 10b, which illustrates the spatial distribution of segments that have a small diversity index (i.e. tracks on them are unevenly distributed between users) and a rather large number of tracks, supports the two cases where multiplication by the diversity index impacts routing (Figures 11a and b). All routes computed using the HMM-weighted network with different cost functions are on average 20% longer than the shortest routes. With niUC, the difference is on average 19.4% (median 16.9%), and with niTC it is 20.9% (20.1%). Despite the inclusion of length in the cost function, the popularity-based routes occur primarily on the main cycling paths (Figures 11c and 11d). A logarithmic transformation of the number of tracks and users equalised the skewed distribution of popularity in the network, producing routes that were on average only 7% longer than the shortest paths (see Figures 8d-f, Figures 11c and d).

4 Discussion

4.1 Quality of the Road Map as a Limitation

Our approach is dependent on the completeness of the underlying network, and therefore has limited potential for worldwide implementation. The only available quality assessment concentrating on OSM cycling data, by Hochmair et al. (2012), has shown that the completeness of cycling facilities was already at a very good level in the selected metropolitan areas of the

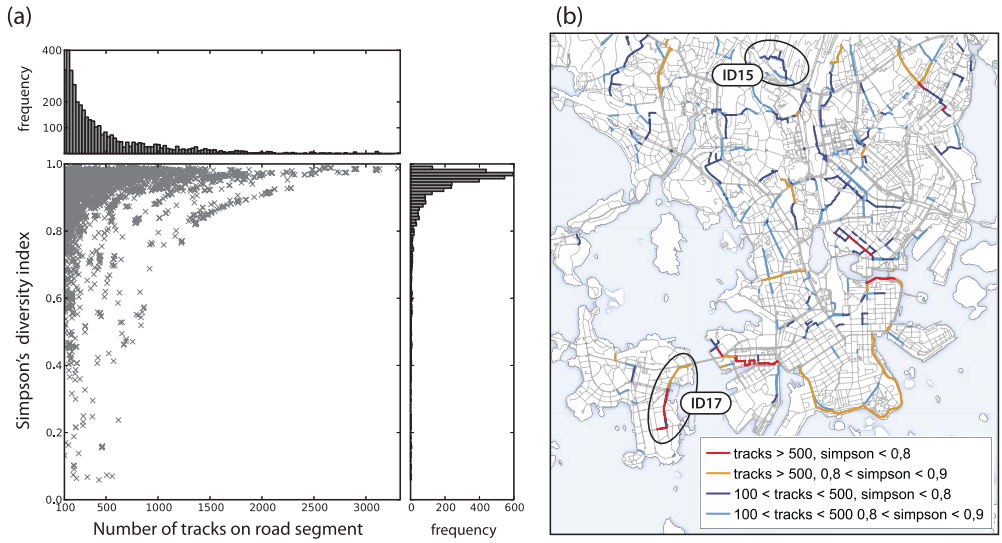

Figure 10 (a) The small Simpson's diversity indexes are primarily associated with segments outside the main cycling routes with many tracks. The illustration includes only segments with more than 100 tracks; and (b) The smallest Simpson's diversity indices that could have an effect on routing in the study area. Black ellipses highlight two such areas which seem to have affected routing in cases 15 and 17

US and Europe, outshining among others the cycling network of Google Maps. In addition, assessments of pedestrian-only paths (Zielstra and Hochmair 2010, 2011) have demonstrated similar results, emphasizing the potential of OSM in non-motorized routing. Several studies (Canavosio-Zuzelski et al. 2013; Girres and Touya 2010; Haklay 2010; Helbich et al. 2012) have suggested that the positional accuracy of OSM road features would be on average around 6 m or better. More attention has been paid to factors that are essential for routing applications, like topological correctness or the completeness of relevant attributes, such as turn or one-way restrictions (Graser et al. 2013; Neis et al. 2012).

It is well known, however, that OSM data is not of uniform standard, as it is up to the voluntary contributors to decide which areas and features are of interest to them and consequently mapped (Haklay 2010; Siebritz et al. 2012). Many studies have concluded that the OSM data is geographically heterogeneous and biased by remoteness (e.g. Girres and Touya 2010; Gröchenig et al. 2014; Helbich et al. 2012; Zielstra and Hochmair 2010; Zielstra and Zipf 2013) as well as by socioeconomic status (Haklay 2010), although exceptions may occur, for example, due to data imports from public domain datasets where agricultural areas may be more completely mapped than urban areas (Zielstra and Hochmair 2010; Zielstra et al. 2012). According to Ciepłuch et al. (2011), the lower the road class, the wider the gap between urban and rural areas. However, considering that the focus of our study was on finding popular routes that are used by many cyclists, we feel that it is safe to assume that the relevant paths are more completely mapped and annotated than small paths in general. An interesting option would be to use the HMM-based matching algorithm as a basis for automatic refinement of the OSM network using the tracked trajectories, as it can identify a situation where no connection is available (Torre et al. 2009; Wang et al. 2012).

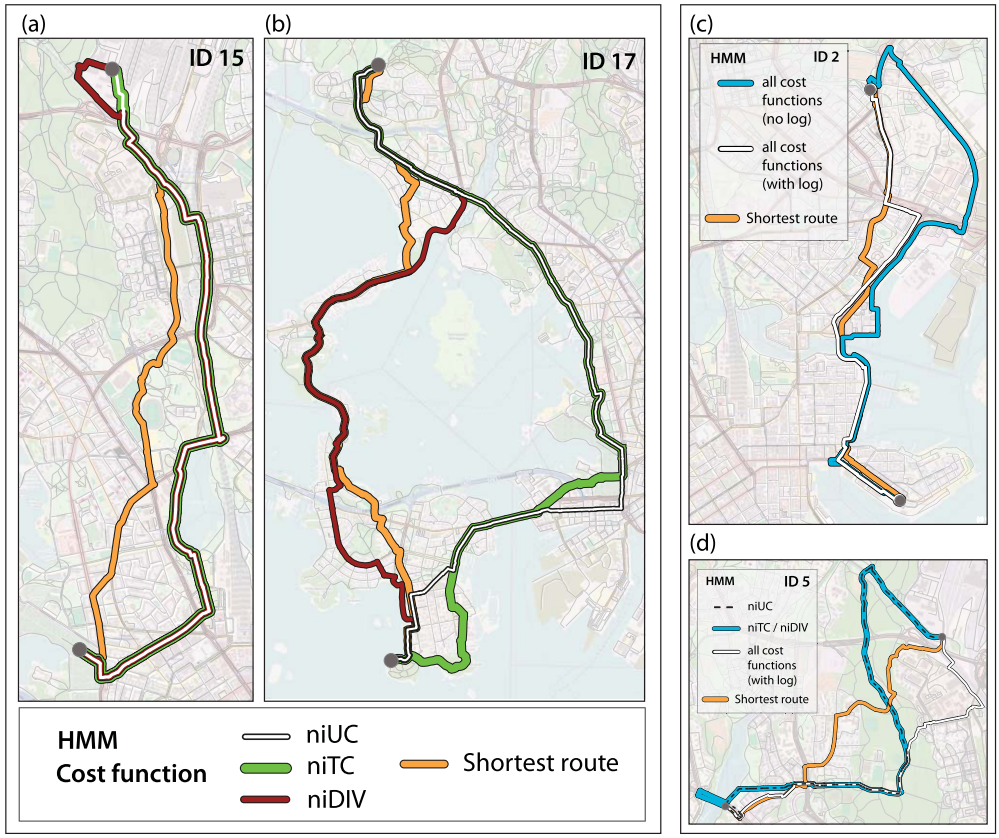

Figure 11 The Simpson's diversity index affected routing in two cases: (a) 15; and (b) 17. Routes in cases (c) 2, and (d) 5 before and after equalizing the popularity distribution (© OpenStreetMap contributors)

4.2 Richer Insight into the Data Through Multiple Popularity Indicators

As well as presenting a method for matching cycling tracks to the OSM network for routing purposes, this study provides insight into the biased nature of the dataset and how it can influence route suggestions. Although the generalizability of user-generated datasets can be questionable in many ways, we concentrated on the "participation inequality" related to the tracking activity of the application users.

The different types of route suggestions provided by the cost functions can indicate the distinctive preferences of cyclists with varying trip purposes, and thus show a pattern that corresponds to previous studies which claimed that the determinants of route choices differ between transportation and recreational cyclists (Broach et al. 2012; Dill and Gliebe 2008; Krizek et al. 2009; see also Heinen et al. 2010; Moudon et al. 2005; Stinson and Bhat 2005). Further insight into this division could be acquired, e.g. by looking at temporal differences between the route suggestions or at data related to land use and cycling facilities.

Zheng and Xie (2014) have argued that more experienced tourists should be given more weight when extracting attractive travel sequences. The diversity-attentive popularity indicator can therefore be compelling, because the inclusion of the number of tracks in the cost function

can be interpreted as giving more weight to the tracks of frequent cyclists. Furthermore, it can be questioned whether the effect of the diversity index would be greater in more peripheral areas where the number of cyclists is smaller than in the urbanized study area.

4.3 Towards Better Route Suggestions for Cyclists

Many cyclists might find the "popular route" suggestion to be a desirable alternative. Considering the four generalized principles of a good cycling route: short, safe, attractive and simple (Hochmair 2004), popularity can, in some way, provide all of them. It guarantees that the route is generally attractive and safe, which is an important aspect of route choice for all cyclists, but especially for those with less experience of cycling in the area, such as tourists. Results on how much longer cyclists are willing to ride compared with the shortest path are varying (e.g. Broach et al. 2012; Dill et al. 2008) and cannot be generalized due to their contextual nature and dependency on, for example, the availability of cycling facilities. More importantly, although the suggested travel distance could be limited by including length in the cost function, the routes were heavily inclined to follow the most popular cycling paths. Through equalizing the popularity in the network, we are more likely to avoid excessive detours, but simultaneously may compromise other desired factors. Because the method is based on local segment-level inferences, it is not possible to consider route-level factors, such as continuity of bicycle facilities, delays in difficult left-turns, or total travel time. In addition, although the effect of hills can be controversial, especially in the context of sports cycling, the direction of the trajectories ought to be considered. The Markov model-based routing algorithm by Chen et al. (2011) has computational efficiency comparable with the standard shortest path algorithms and could be considered as an enhancement.

Combining historical trajectories with the current network caused matching errors as the network is not static, but changing continuously. Because new connections affect the popularity of the old network, more recent tracks could be emphasized. Similarly, all links were seen as homogeneous, which may initially hide the popularity of new connections unless tracks are weighted according to their "age". Another topic that deserves attention in the future is the extraction of cycling events from the data.

4.4 Enhancing the HMM-Based Map-Matching Method

Overall, the HMM-based algorithm achieved good results despite the combination of two user-generated datasets. The method is robust regarding positional inaccuracies and therefore well suited for crowdsourced network data. However, it assumes that the network topology is correctly mapped. The main algorithmic problem was related to the preference of routes with fewer segments, which in some cases turned out to be a trade-off for prioritizing topological connectivity of the network by inclusion of the boundaries of cycleable squares. Favouring routes with fewer segments could potentially be avoided by determining the transition probability based on the difference between the great circle distance of the observations and the driving distance along the road network as suggested by Newson and Krumm (2009). This would, however, require using a computationally demanding shortest path algorithm (see also Wei et al. 2013). Our results support the need to find novel solutions to routing across polygon features which is a recognized shortcoming in existing pedestrian and bicycle route planners (Bauer et al. 2014). As the popularity of OSM-based routing services increases, it can well be that the routes of main cycling flows across squares will become more comprehensively mapped although there would be no visible, designated path (Figure 12). While reflecting the movement

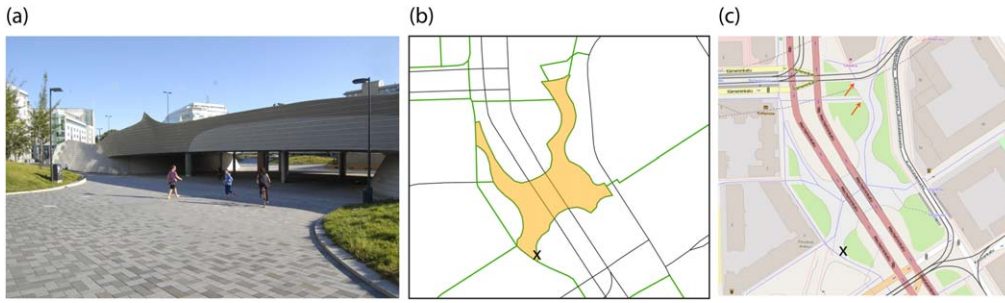

Figure 12 (a) A square at the end of an important cycle track 'Baana' in Helsinki; (b) as it was mapped in OSM during the data extraction in August 2014 (bikeways in green); and (c) in August 2015. During this period cycleways have been added in a skeleton-like fashion across the square although in real life there are no visible paths restricting the movement of cyclists. Two cycleways (highlighted with red arrows) are still missing connections across the square. The black 'x' shows where the photo (a) was taken. The location of the square is shown in Figure 1b

patterns in reality, this could in most cases be an adequate solution to popularity-based routing.

Systematic testing of all pre-processing parameters for the trajectory and network data in varying circumstances was beyond the scope of this study. However, the urban study area allowed us to recognize possible problems related to the chosen values, and a few points can be raised related to their effect on the performance of matching. First, the HMM-based method is sensitive to sampling interval, especially when combined to a road map represented at a high level of detail. Denser point interval would, especially in more rural areas, unduly increase the computation time, whereas in urban areas, sufficient density is difficult to determine beforehand. Rather, we could consider using a denser sampling only around intersections or a dynamic matching radius in HMM. Second, without the interpolation of new points, inferring routes would have been impossible if outages exceeded a certain distance. Obviously, the longer the slicing threshold, and thus the upper limit of interpolation, the more false routes could be inferred depending on the network's road density. Third, splitting road segments into shorter parts of at most 200 m allows more truthful popularity indicators along long roads, and does not, otherwise, affect the performance of map-matching, as long as the sampling density is higher.

5 Conclusions

OpenStreetMap is an attractive data repository for the cycling community. Crowdsourced workout data recorded with mobile applications, such as Sports Tracker, has also provoked a lot of interest due to the general lack of cycling data. Regardless of their incompleteness or problems related to generalizability, such volunteered geographic datasets have great potential value, for example in the implementation of new route-finding applications.

In this study we investigated the possibility of using sports tracking application data in providing bicycle-friendly routes based on their popularity. The advanced HMM-based algorithm, which was modified to be suitable for the detailed mapping of the OSM network, achieved better performance than the geometric method, and in many cases also resulted in partially different route suggestions. When compared to the manually matched set of 50 tracks, the

HMM-based algorithm was able to correctly match over 90% of the segments and approximately 94% of the route length. The biased nature of the tracking data also had an impact on route suggestions depending on the cost function used. Cost functions based on the number of users and tracks provided different route suggestions in most cases, possibly indicating the varying route choices of utilitarian and recreational cyclists. In a few cases, revising the number of tracks by Simpson's diversity index provided further alternative routes where cycling is regular, but the paths are not dominated by only a few bikers. Including the length of each segment in the cost functions produced routes that were on average 20% longer than the shortest paths. Due to the skewed popularity distribution of the segments, the routes were heavily inclined to follow the main cycling routes. Equalization of the distribution of popularity in the network significantly reduced the length of the route suggestions.

References

Adamic L A and Huberman B A 2002 Zipf's law and the Internet. *Glottometrics* 3: 143-50

Batty M 2013 Big data, smart cities and city planning. *Dialogues in Human Geography* 3: 274-79

Bauer C, Almer A, Landstätter S and Luley P M 2014 Optimierte Wegefindung für Fußgänger basierend auf vorhandenen OpenStreetMap-Daten. In Strobl J, Blaschke T, Griesebner G, and Zagel B (eds) *Angewandte Geoinformatik 2014, Beiträge zum 26, AGIT-Symposium Salzburg*. Berlin, Wichmann: 505-14

Broach J, Dill J, and Gliebe J 2012 Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A* 46: 1730-40

Canavosio-Zuzelski R, Agouris P, and Doucette P 2013 A photogrammetric approach for assessing positional accuracy of OpenStreetMap© roads. *ISPRS International Journal of Geo-Information* 2: 276-301

Chang K P, Wei L Y, Yeh M Y, and Peng W C 2011 Discovering personalized routes from trajectories. In *Proceedings of the Third ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, Chicago, Illinois: 33-40

Chen Z, Shen H T, and Zhou X 2011 Discovering popular routes from trajectories. In *Proceedings of the Twenty-seventh IEEE International Conference on Data Engineering*, Hannover, Germany: 900-11

Ciepłuch B, Mooney P, Jacob R, Zheng J, and Winstanely A 2011 Assessing the quality of open spatial data for mobile location-based services research and applications. *Archiwum Fotogrametrii, Kartografii i Teledetekcji* 22: 105-16

Dill J and Gliebe J 2008 *Understanding and Measuring Bicycling Behavior: A Focus on Travel Time and Route Choice*. Portland, OR, Oregon Transportation Research and Education Consortium No. RR 08-03

Dill J, Mohr C, and Ma L 2014 How can psychological theory help cities increase walking and bicycling? *Journal of the American Planning Association* 80: 36-51

Ehrgott M, Wang J Y, Raith A, and Van Houtte C 2012 A bi-objective cyclist route choice model. *Transportation Research Part A* 46: 652-63

Ferrari L and Mamei M 2013 Identifying and understanding urban sport areas using Nokia Sports Tracker. *Pervasive and Mobile Computing* 9: 616-28

Forney G D 1973 The Viterbi algorithm. *Proceedings of the IEEE* 61: 268-78

Girres J F and Touya G 2010 Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS* 14: 435-59

Graser A, Straub M, and Dragaschnig M 2013 Towards an open source analysis toolbox for street network comparison: Indicators, tools and results of a comparison of OSM and the official Austrian Reference Graph. *Transactions in GIS* 18: 510-26

Gröchenig S, Brunauer R, and Rehrl K 2014 Digging into the history of VGI data sets: Results from a worldwide study on OpenStreetMap mapping activity. *Journal of Location Based Services* 8: 198-210

Goh C Y, Dauwels J, Mitrovic N, Asif M T, Oran A, and Jaillet P 2012 Online map-matching based on hidden Markov model for real-time traffic sensing applications. In *Proceedings of the Fifteenth IEEE International Conference on Intelligent Transportation Systems*, Anchorage, Alaska: 776-81

Hakley M 2010 How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B* 37: 682-703

Hao Q, Cai R, Wang C, Xiao R, Yang J M, Pang Y, and Zhang L 2010 Equip tourists with knowledge mined from travelogues. In *Proceedings of the Nineteenth International Conference on World Wide Web*, Raleigh, North Carolina: 401-10

Hashemi M and Karimi H A 2014 A critical review of real-time map-matching algorithms: Current issues and future directions. *Computers, Environment and Urban Systems* 48: 153-65

Heinen E, van Wee B, and Maat K 2010 Commuting by bicycle: An overview of the literature. *Transport Reviews* 30: 59-96

Helbich M, Amelunxen C, and Neis P 2012 Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. In *Proceedings of the International GI_Forum 2012*, Salzburg, Austria

Hendavi A M, Sturm E, Oliver D, and Shekhar S 2013 CrowdPath: A framework for next generation routing services using volunteered geographic information. In Nascimento M A, Sellis T, Cheng R, Sander J, Zheng Y, Kriegel H-P, Renz M, and Sengstock C (eds) *Advances in Spatial and Temporal Databases*. Berlin, Springer Lecture Notes in Computer Science Vol. 8098: 456-61

Hochmair H H 2004 Decision support for bicycle route planning in urban environments. In *Proceedings of the Seventh AGILE Conference on Geographic Information Science*, Heraklion, Greece: 697-706

Hochmair H H, Zielstra D, and Neis P 2012 Assessing the completeness of bicycle trails and designated lane features in OpenStreetMap for the United States and Europe. In *Proceedings of the Ninety-second Annual Meeting of the Transportation Research Board*, Washington, DC

Lindsey G, Hankey S, Wang X, Gorjestani A, and Chen J 2013 *Feasibility of Using GPS to Track Bicycle Lane Positioning*. Minneapolis-St. Paul, MN, University of Minnesota Research Report No. CTS-13-16

Lindsey G, Nordback K, and Figliozzi M A 2014. Institutionalizing bicycle and pedestrian monitoring programs in three states: Progress and challenges. In *Proceedings of the Ninety-third Annual Meeting of the Transportation Research Board*, Washington, DC: 1-22

Loidl M, Krampe S, Zagel B, and Pucher G 2014 Aufbereitung von OpenStreetMap-Daten für GIS-Modellierungen und Analysen. In Strobl J, Blaschke T, Griesebner G, and Zagel B (eds) *Angewandte Geoinformatik 2014, Beiträge zum 26, AGIT-Symposium, Salzburg*. Berlin, Wichmann: 505-14

Lu X, Wang C, Yang J M, Pang Y, and Zhang L 2010 Photo2trip: Generating travel routes from geo-tagged photos for trip planning. In *Proceedings of the International Conference on Multimedia*, Florence, Italy: 143-52

McDonald D G and Dimmick J 2003 The conceptualization and measurement of diversity. *Communication Research* 30: 60-79

Menghini G, Carrasco N, Schüssler N, Axhausen K W 2010 Route choice of cyclists in Zurich. *Transportation Research Part A* 44: 754-65

Moudon A V, Lee C, Cheadle A D, Collier C W, Johnson D, Schmid T L, and Weather R D 2005 Cycling and the built environment: A U.S. perspective. *Transportation Research Part D* 10: 245-61

Mullen W F, Jackson S P, Croitoru A, Crooks A, Stefanidis A, and Agouris P 2014 Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features. *GeoJournal* 80: 587-605

Neis P, Zielstra D, and Zipf A 2012 The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007-2011. *Future Internet* 4: 1-21

Neis P, Zielstra D, and Zipf A 2013 Comparison of volunteered geographic information data contributions and community development for selected world regions. *Future Internet* 5: 282-300

Neis P and Zielstra D 2014 Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet* 6: 76-106

Newson P and Krumm J 2009 Hidden Markov map matching through noise and sparseness. In *Proceedings of the Seventeenth ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Seattle, Washington: 336-43

Nielsen J 2006 The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities. WWW document, http://www.useit.com/alertbox/participation_inequality.html

Oksanen J, Bergman C, Sainio J, and Westerholm J. 2015 Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data. *Journal of Transport Geography* 48: 135–144.

Priedhorsky R, Masli M, and Terveen L 2010 Eliciting and focusing geographic volunteer work. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, Savannah, Georgia: 61-70

Priedhorsky R and Terveen L 2008 The computational geowiki: What, why, and how. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, San Diego, California: 267-76

Quddus M A, Ochieng W Y, and Noland R B 2007 Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C* 15: 312-28

Quercia D, Schifanella R, and Aiello L M 2014 The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the Twenty-fifth ACM Conference on Hypertext and Social Media*, Toronto, Canada: 116-25

Reddy S, Shilton K, Denisov G, Cenizal C, Estrin D, and Srivastava M 2010 Biketastic: Sensing and mapping for better biking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, California: 1817-20

Robb D 2013 Introducing Routemaster, Strava Engineering blog. WWW document, http://engineering.strava.com/routemaster/

Schmitz S, Zipf A, and Neis P 2008 New applications based on collaborative geodata: The case of routing. In *Processing of the Twenty-eighth INCA International Congress on Collaborative Mapping and Space Technology*, Gandhinagar, Gujarat, India

Sener I N, Eluru N, and Bhat C R 2009 An analysis of bicycle route choice preferences in Texas, US. *Transportation* 36: 511-39

Shekhar S, Gunturi V, Evans M R, and Yang K 2012 Spatial big-data challenges intersecting mobility and cloud computing. In *Proceedings of the Eleventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, Scottsdale, Arizona: 1-6.

Siebritz L, Sithole G, and Zlatanova S 2012 Assessment of the homogeneity of volunteered geographic information in South Africa. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39: B4

Simontine T 2012 Android's Rise Helps Google Grow Its Traffic Surveillance System. WWW document, http://www.technologyreview.com/news/428732/androids-rise-helps-google-grow-its-traffic-surveillance-system/

Stinson M A and Bhat C R 2005 *A Comparison of the Route Preferences of Experienced and Inexperienced Bicycle Commuters*. Washington, DC, Transportation Research Board Paper No. 05-1434

Su J, Winters M, Nunes M, and Brauer M 2010 Designing a route planner to facilitate and promote cycling in Metro Vancouver, Canada. *Transportation Research Part A* 44: 495-505

Sun Y, Fan H, Bakillah M, and Zipf A 2015 Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems* 53: 110-22

Suvanto S and Oksanen J 2013 Privacy Aware Density Maps: Visualizing Popular Sports Routes in the Helsinki Region. WWW document, http://www.move-cost.info/highlights.php

Thiagarajan A, Ravindranath L, LaCurts K, Madden S, Balakrishnan H, Toledo S, and Eriksson J 2009 VTrack: Accurate, energy-aware road traffic delay estimation using mobile phones. In *Proceedings of the Seventh ACM Conference on Embedded Networked Sensor Systems*, Berkeley, California: 85-98

Torre F, Pitchford D, Brown P, and Terveen L 2012 Matching GPS traces to (possibly) incomplete map data: Bridging map building and map matching. In *Proceedings of the Twentieth International Conference on Advances in Geographic Information Systems*, Redondo Beach, California: 546-49

Wang Y, Liu X, Wei H, Forman G, Chen C, and Zhu Y 2013 Crowdatlas: Self-updating maps for cloud and personal use. In *Proceedings of the Eleventh Annual International Conference on Mobile Systems, Applications, and Services*, Taipei, Taiwan: 27-40

Wei H, Wang Y, Forman G, Zhu Y, and Guan H 2012 Fast Viterbi map matching with tunable weight functions. In *Proceedings of the Twenthieth International Conference on Advances in Geographic Information Systems*, Redondo Beach, California: 613-16

Yoon H, Zheng Y, Xie X, and Woo W 2011 Social itinerary recommendation from user-generated digital trails. *Personal and Ubiquitous Computing* 16: 469-84

Yue Y, Lan T, Yeh A G, and Li Q Q 2014 Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. *Travel Behaviour and Society* 1: 69-78

Yuan J, Zheng Y, Xie X, and Sun G 2013 T-Drive: Enhancing driving directions with taxi drivers' intelligence. *IEEE Transactions on Knowledge and Data Engineering* 25: 220-32

Zhang D, Huang J, Li Y, Zhang F, Xu, C, and He T 2014 Exploring human mobility with multi-source data at extremely large metropolitan scales. In *Proceedings of the Twentieth Annual International Conference on Mobile Computing and Networking*, Maui, Hawaii: 201-12

Zheng Y and Xie X 2011 Learning travel recommendations from user-generated GPS traces. *ACM Transactions on Intelligent Systems and Technology* 2(1): 2

Zheng V W, Zheng Y, Xie X, and Yang Q 2010 Collaborative location and activity recommendations with GPS history data. In *Proceedings of the Nineteenth International Conference on World Wide Web*, Raleigh, North Carolina: 1029-1038.

Zielstra D and Hochmair H H 2011 Comparative study of pedestrian accessibility to transit stations using free and proprietary network data. *Transportation Research Record* 2217: 145-52

Zielstra D and Hochmair H H 2012 Using free and proprietary data to compare shortest-path lengths for effective pedestrian routing in street networks. *Transportation Research Record* 2299: 41-47

Zielstra D, Hochmair H H, and Neis P 2013 Assessing the effect of data imports on the completeness of OpenStreetMap: A United States case study. *Transactions in GIS* 17: 315-34

Zielstra D and Zipf A 2010 A comparative study of proprietary geodata and volunteered geographic information for Germany. In *Proceedings of the Thirteenth AGILE International Conference on Geographic Information Science*, Guimarães, Portugal

