



Visualising where commuting cyclists travel using crowdsourced data

David Philip McArthur*, Jinhyun Hong

Urban Big Data Centre, University of Glasgow, G12 8QQ, UK



ARTICLE INFO

Keywords:

Cycling
Crowdsourced data
Volunteered geographic data (VGI)
Strava

ABSTRACT

Encouraging more cycling is increasingly seen as an important way to create more sustainable cities and to improve public health. Understanding how cyclists travel and how to encourage cycling requires data; something which has traditionally been lacking. New sources of data are emerging which promise to reveal new insights. In this paper, we use data from the activity tracking app Strava to examine where people in Glasgow cycle and how new forms of data could be utilised to better understand cycling patterns. We propose a method for augmenting the data by comparing the observed link flows to the link flows which would have resulted if people took the shortest route. Comparing these flows gives some expected results, for example, that people like to cycle along the river, as well as some unexpected results, for example, that some routes with cycling infrastructure are avoided by cyclists. This study proposes a practical approach that planners can use for cycling plans with new/emerging cycling data.

1. Introduction

Encouraging people to cycle rather than travel by motorised transport is seen as a way of achieving more sustainable urban environments. The substantial benefits on offer have not been overlooked by policy makers. For instance, Transport Scotland (Transport Scotland, 2017) have set a vision of “10% of everyday journeys to be made by bike, by 2020”. They also note that cities will have to be the driver in achieving this ambitious aspiration.

One intervention which has been deployed in many cities is improving cycling infrastructure e.g., segregated cycle lanes, bicycle parking etc. Such measures have generally been shown to have a positive effect, although the evidence is somewhat mixed. For example, in a review Pucher et al. (2010) found statistically significant and positive effects in most of the aggregate level studies they looked at while noting that the evidence was more mixed for individual level studies. In another review, Ogilvie et al. (2004) found little effect of infrastructure changes. Evaluation, however, can be difficult due to a lack of data (Hankey et al., 2012; Heesch and Langdon, 2016).

One potential explanation is that the infrastructure changes were not well targeted to places where they would be of use to cyclists. One barrier to the effective targeting of infrastructure changes is the lack of good data on cycling to help planners understand overall cycling patterns. Currently, much of the data on cycling comes from either travel surveys or cycle counts. These methods tend to give sparse spatial and/or temporal coverage.

Travel surveys provide the opportunity to collect useful information about cyclists from representative samples, however they can be expensive to administer resulting in small sample sizes. Cycling typically has a low mode share (e.g., 1.6% for travel to work in Glasgow), hence many people must be surveyed in order to include a reasonable number of cyclists. There are also limitations to the sorts of data which can be captured using a survey. For instance, people responding to a survey will probably be able to recall their origins and destinations, however may be unable to give detailed route information.

Counting cyclists is another way to collect data. Such counts may be conducted manually or with a sensor, such as pneumatic tubes or inductive loops. Many sensors can only give accurate results if they are carefully installed and properly maintained (Norback and Janson, 2010; Romanillos et al., 2016). Ryus et al. (2014) note that in the US, the majority of cycle counts are conducted manually. This can reduce the number of locations surveyed, as well as creating a biased sample of locations.

New types of data have emerged in recent years which address some of the weaknesses of the traditional sources, allowing a better understanding of overall cycling patterns to be achieved. For example, having cyclists carry GPS receivers offers the chance to capture rich movement data, including data about origins, destinations and route choices. It can also be a way to deal with misreporting in surveys (Shen and Stopher, 2014). However, providing people with such devices and gathering the data can be expensive. This has severely limited the sample sizes in studies employing these methods.

* Corresponding author.

E-mail addresses: david.mcarthur@glasgow.ac.uk (D.P. McArthur), jinhyun.hong@glasgow.ac.uk (J. Hong).

Fortunately, the proliferation of smartphones means that many people already carry equipment capable of capturing their movements. The popularity of activity tracking apps also means that many cyclists are already collecting and publishing their own data. One of the most popular of these apps is Strava, which allows people to capture and share their running and cycling trips.

Such crowdsourced cycling data (sometimes referred to as volunteered geographic information or VGI) could be valuable in supporting decision making (Jestico et al., 2016). Crowdsourced mobility data has already been used by researchers to study topics such as routing and navigation (Hendawi et al., 2013; Prandi et al., 2014; Keler and J.D., 2016), disaster management (Haworth and Bruce, 2015; Granell and Ostermann, 2016), wheelchair routing (Zipf et al., 2016; Mobasheri, 2017), health (Griffin and Jiao, 2015) and cycling (Sun et al., 2017; Boss et al., 2018).

This paper will examine how data from the cycling app Strava can be augmented and used to gain a better understanding of how cyclists travel in a city. A typical research question asked by practitioners working with crowdsourced cycling data is where do cyclists cycle. This is relatively easy to see using tools such as Strava's global heatmap.¹ We extend this typical question to ask where do cyclists cycle when we control for travel demand between different origins and destinations. To address this question, we compare the routes taken by commuting cyclists with the route they would take if they minimised their travel distance in the city of Glasgow.

Through mapping the results, we highlight road links which cyclists avoid and those which they are drawn to. This provides valuable information about cyclists' choices. By combining these results with local knowledge and/or contextual information we can infer the relationship between cycling and other relevant factors such as the built environment and the public transport system. To demonstrate this process, we examine particular roads which are either heavily used or underused and discuss what factors may have led to this situation. In this way, our method can highlight anomalies in the road network which can then be subjected to further analysis; either quantitatively or qualitatively. We present qualitative discussions of the examples we select.

A key strength of our proposed methodology is that it can be easily replicated for cities around the world. Strava data is often available in places where other data may be lacking, for instance in the developing world (Musakwa and Selala, 2016). Part of the price for the abundance of data is that its representativeness is often unknown. Sizes and directions of bias may also change between different contexts. Still, the number of cyclists using Strava is likely to be much higher than the sample size which would be achievable in a conventional GPS survey; or indeed from other data sources. Having such rich data can provide extremely useful information for planners. There is emerging evidence that data from Strava corresponds well with urban cycling flows (Haworth, 2016).

The paper is structured as follows. Section 2 will review the literature concerning route choice, and in particular the studies using new and novel forms of data. This review will provide guidance about what we should expect to see in the data we create. Section 3 will outline the methodology we use to generate the enhanced Strava data. Section 4 will examine how the enhanced data can be visualised and analysed. Potential use cases are also presented. In the concluding section, we discuss some of the benefits and potential limitations of the data.

2. Related work

Encouraging more people to cycle requires an understanding of where cyclists like to cycle. This has been a topic of interest to researchers for decades. For a review of the early literature, see Hopkinson et al. (1989). It is well known that cyclists choose routes

differently from other road users (Ehrgott et al., 2012), hence special attention must be given to factors which are important to them.

As is to be expected with spatial interaction, cyclists tend to prefer shorter routes to longer ones (Dill and Gliebe, 2008; Fraser and Lock, 2011; Hood et al., 2011; Börjesson and Eliasson, 2012; Broach et al., 2012; Caulfield, 2014). This effect is more pronounced for utilitarian trips such as commuting trips (Sener et al., 2009). However using GPS data Dill and Gliebe (2008) found that their sample of 164 cyclists on average travelled 11% further than the shortest route for commuting trips. Clearly, other factors are also important when selecting a route.

While distance may be the dominant component in the impedance to cycling, there are other features which are important. For example, cyclists have been shown to prefer fewer junctions/red lights/stop signs (Sener et al., 2009; Menghini et al., 2010; Caulfield et al., 2012), fewer turns (Hood et al., 2011; Broach et al., 2012), and fewer steep inclines (Menghini et al., 2010; Fraser and Lock, 2011; Hood et al., 2011; Broach et al., 2012). These factors seem to increase the impedance values of routes and hence deter cyclists.

Environmental factors have received considerable attention in the literature. Cyclists are generally attracted to routes which are exposed to green space, or other attractive scenery (Fraser and Lock, 2011; Winters et al., 2011; Van Holle et al., 2014; Zhao and Li, 2017). The green space is experienced as an amenity which increases the utility of the journey. Some disamenities which have been shown to have a negative effect on cycling are noise, pollution and proximity to traffic (Dill and Gliebe, 2008; Sener et al., 2009; Fraser and Lock, 2011; Winters et al., 2011; Caulfield et al., 2012). Perhaps because of the negative externalities of traffic, cyclists usually prefer routes with low traffic volumes (Sener et al., 2009; Hood et al., 2011; Broach et al., 2012).

One popular intervention to promote cycling is the construction of cycling infrastructure. Such infrastructure may range from a demarcated lane to fully segregated infrastructure away from traffic. Cyclists have been shown to prefer routes with cycling infrastructure (Dill and Gliebe, 2008; Winters and Teschke, 2010; Fraser and Lock, 2011; Hood et al., 2011; Winters et al., 2011; Broach et al., 2012; Caulfield et al., 2012; Hankey et al., 2012; Caulfield, 2014). In general, the more segregated the infrastructure is from traffic, the more it is preferred by cyclists. For example, Wardman et al. (2007) found the presence of fully segregated infrastructure had the greatest impact on whether someone would choose to cycle, a finding supported by Gatersleben and Appleton (2007). The beneficial effect of segregated infrastructure seems to be particularly strong for female cyclists (Garrard et al., 2008).

The effect of infrastructure on attracting cyclists can be marked. For example, Tilahun et al. (2007) found that cyclists may be willing to extend their journey by up to 20 min if it means having access to good-quality cycling infrastructure. In a study of how cycling interventions might be appraised/evaluated, Börjesson and Eliasson (2012) found that cyclists put a higher value on travel time savings than people travelling by other modes of transport.

As noted by Song et al. (2017), while good cycling infrastructure may be necessary for encouraging non-cyclists to cycle, it is not likely to be sufficient. There are other factors which policy makers are able to influence which may have beneficial effects on cycling. Cyclists have been shown to be attracted to areas with mixed land-use (Frank and Pivo, 1994; Hankey et al., 2012; Zhao and Li, 2017). Population density (Saelens et al., 2003), accessibility to jobs and services (Frank and Pivo, 1994; Kockelman, 1997), and proximity to the central business district (CDB) have also been shown to encourage cycling (Hankey et al., 2012; Caulfield, 2014). Routes with slower moving traffic, which may be the result of lower speed limits, have also been shown to increase the attractiveness of a route (Sener et al., 2009; Hood et al., 2011; Caulfield et al., 2012; Caulfield, 2014). Parking policy can also play a role. Cyclists prefer streets without parking (Sener et al., 2009). Streets with side parking have been shown to be the least preferred (Tilahun et al., 2007).

In summary, several empirical studies have examined the important

¹ <https://www.strava.com/heatmap>.

determinants of the level of cycling as well as route choices. However, many studies utilised small samples that cover partial geographic areas. To better understand how cyclists travel in a city, larger, more detailed datasets are required.

3. Methodology

We take the City of Glasgow as our case study. Glasgow is Scotland's largest city and has expended significant effort in recent years to encourage more cycling. As part of these efforts, substantial investments have been made in cycling infrastructure. Much of this was connected to the city's hosting of the Commonwealth Games in 2014. With more investments currently being made, and more planned, it is important to gain an understanding of how cyclists travel in the city and how they might like to travel.

We utilise data collected from the Strava activity tracking app in Scotland in 2016. We consider only commuting trips since our discussion focusses on ways to reduce journey times. Such a discussion does not make sense for leisure trips; particularly if they are round trips. Commuting trips are identified in the Strava Metro data and can thus be easily extracted.

The app is primarily used by cyclists and runners to capture their activities. It is used around the world, which means that similar data to the data employed in this paper is available globally. Strava collects and aggregates the data before providing it to researchers² as a product called Strava Metro. The output of their aggregation comes in three main components: origin/destination flows; road link activity counts; and intersection information. The first two data sources are utilised in this study. The pertinent features of the data are briefly outlined below. The user guide provides further information about the structure of the data (Strava, 2015).

Cyclists can choose whether they wish to log a cycle trip using the app and upload it to Strava's server. For this reason, our dataset only covers trips made by Strava-using cyclists when they choose to log the trip. Users can also opt-out of having their data included in the Strava Metro datasets. If the trip is made available for inclusion in the Strava Metro data, then Strava processes the raw GPS trajectories into aggregate data.

The first Strava Metro product we use is the origin/destination table of trips. The output area where each trip started and ended is provided. Output areas are the smallest census geography used in the UK. In urban environments, the areas are small, making them useful for analysing details about trips in Glasgow. We retain only trips tagged as commuting trips and then aggregate the table to get an annual origin/destination matrix showing the flows between each pair of output areas in Scotland. We choose to work with data for the whole of Scotland to capture trips which may originate or terminate outside of Glasgow, but which pass through it.

The GPS points gathered by Strava are map-matched onto a road network before they are given to researchers as Strava Metro data. In this case, the points were map-matched to the OpenStreetMap (OSM) road network. Each link in the OSM data has the number of cyclists who used the link in each minute of each day. Once more, we retain only counts relating to commuters and then aggregate the data to get the total annual commuting count for each link in the network.

Some demographic information is given by Strava as part of Strava Metro about the sample, however this is only available for the whole sample and not the sample of commuters. In Table 1, we present the information for the full sample of users in and around Glasgow, stressing that it includes both the cyclists we are interested in along with the leisure cyclists.

Table 1

Summary of the Strava data for the area in and around Glasgow.

Age (years)	Male	Female		
< 20	847	207	Cyclists	15,292
25–34	2332	544	Trips	359,945
35–44	3316	456	Average distance (m)	23,296
45–54	2675	337	Median distance (m)	14,065
55–64	720	78	Average time (s)	4739
65–74	110	5	Median time (s)	2707
75–84	8	0	Average uploads	232
85–94	2	0		
Unknown	3075	580		
Total	13,085	2207		

Perhaps the most notable point is that the sample is dominated by males. This is to be expected as according to the National Travel Survey³ (NTS), 71% of cyclists are male. In our sample they make up 85%. This may be because the NTS covers all of England while we are focussing on an urban area in Scotland. Only a small number of cyclists (113) were captured in the corresponding Scottish Household Survey travel diaries. We therefore present data from the rest of the UK as a comparator. Observed differences may also be due to males being more likely to use the app than females. Further research is needed to understand the users of Strava and their behaviour.

The average trip length, at 23 km, seems somewhat high. There are a number of reasons why there may be outliers in the data. For example, some Strava users may continue to track a journey once they have left their bike and, for example, boarded a train. The most competitive, active cyclists are perhaps also more likely to use Strava than an average cyclist. The median figure looks more reasonable. Unfortunately, we do not have access to trip distances for the commuting trips alone.

In this state, the data already gives valuable information about how cyclists travel in the city. For example, it is possible to see which routes are most used. While it is valuable to see where people have cycled, this does not necessarily tell us much about which parts of the city people avoid and which they are attracted to. For instance, an area may not have many cyclists because people actively avoid it, or perhaps because it doesn't connect popular origins and destinations. It is not possible to see this in the raw data, however we suggest that it is possible to get some additional information by comparing the actual routes chosen with the routes which might have been chosen if the cyclists selected routes to minimise travel distance. Such information would be useful for planners developing cycling plans or when planning where to make investments in infrastructure.

We generate this additional information by utilising the origin/destination table and a simple traffic assignment model. The model takes an origin/destination table and loads it onto a transport network based on certain criteria. One of the simplest variants of the model is the All-or-Nothing shortest route. This method finds the shortest route between all origin and destination pairs, and then assigns the trips onto the relevant road links. We calculated such a measure using the origin/destination table provided by Strava and the OpenStreetMap road network. We removed motorways from the road network since cyclists cannot utilise these. We used popular transport planning software (TransCAD)⁴ to build the OpenStreetMap data into a routable network and to run the traffic assignment algorithm. Note that we are not trying to predict the link flows (as would usually be the aim of a traffic assignment model) but are constructing a counterfactual where all

³ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/577825/mode-use-2015-a-view-into-a-travel-week.pdf.

⁴ Although we use commercial software, open-source alternatives such as AequilibraE for Python could also be used. See <https://github.com/AequilibraE/aequilibrae>.

² At the time of writing, the Strava data used in this paper is available upon application from the Urban Big Data Centre at the University of Glasgow (www.ubdc.ac.uk).

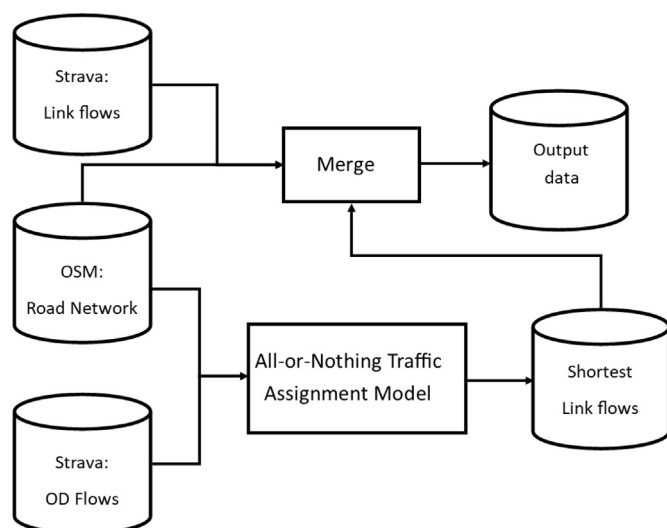


Fig. 1. Summary of data and methods.

cyclists took the shortest route. We use distance as our impedance measure here as an example. The analysis could also be run for different sorts of impedance measures such as a generalised travel cost. Turn penalties could also be incorporated into the distance measure to account for cyclists preferring routes with fewer turns (Broach et al., 2012). We use distance since this is an attribute which is available for any location for which Strava data are available.

After running the traffic assignment model, we now have a road network with two important attributes. Firstly, we know the total number of Strava cyclists who used each road link in the year. We also know how many people would have taken each link if all Strava cyclists had taken the shortest route. This information can be presented in different ways and can offer various insights about where cyclists cycle. For example, we can look at the correspondence between cycling infrastructure and cycling patterns, or actual versus expected traffic in different areas of the city. A diagrammatic summary of the data and methods is presented in Fig. 1.

We intend that planners would use maps of the data to identify anomalies and generate questions for further investigation. We select two examples to show how crowdsourced cycling data can be utilised to extract useful information about cycling patterns. We use Google Street View images and local knowledge to explain why selected roads/areas may be under or over-used. This can be seen as a quicker, less data-hungry approach to constructing a traditional route choice model.

4. Results and discussion

As outlined above, one of the benefits of using Strava data is that it shows which roads are popular with cyclists. In the top pane of Fig. 2, we see the city centre of Glasgow in the centre of the map with the observed annual volume of commuters displayed on the road links. In the bottom pane, we see the same map but showing the flows we would expect to see if people had taken the shortest route between their origin and destination. As expected, the maps show some similarities and it is relatively easy to see that both are consistent with the same set of journeys. For example, some of the major arterial routes are popular in both maps. However, it is easy to see examples of where substitutions have occurred e.g., where parallel links appear to be more used than we would expect from the shortest route.

In the study area presented above, a total distance of 718,798 km was travelled by commuters. If the shortest route had been selected, then a distance of 716,907 km would have been travelled. This represents a difference of 0.3%. This is substantially less than has been found in other studies. For example, Broach et al. (2012) found that for

commute trips, people travelled on average 11% more than the shortest path. This may be due to differing methodologies. For example, we have data on the origins and destinations of trips at the level of output area rather than the exact origin/destination. The differing geographic/cultural context as well as our substantially larger sample size may also explain some of the observed differences.

To add more context, we suggest presenting the data on the same map rather than in two separate panes. We have not added a basemap in the figures to increase their readability, however would suggest that planners use a basemap when exploring the data to help add context. We add an attribute to the road network which is the difference between the modelled flows and the observed flows. Fig. 3 shows disproportionately popular routes (where observed minus modelled flows ≥ 0) as solid lines and disproportionately unpopular routes (where observed flows minus modelled flows < 0) as broken lines. The thickness of the line corresponds to the size of the flows. The thinner the lines, the closer the modelled and observed flows are to one another. Classes are derived for both the popular and unpopular links using Jenks' natural breaks as implemented in ESRI's ArcMap software. The results are presented in Fig. 3.

One of the most dominant features of Fig. 3, which was also visible in the map of observed flows, is the popularity of the links along the north bank of the River Clyde (which runs through the centre of Glasgow). As well as offering some degree of scenic value, the northern route along the river has also been designated as a cycle route (which is shared with pedestrians). Displaying the data in this way allows us to have an indication of where these cyclists might otherwise have cycled. Notice for instance the disproportionately unpopular links running one block north of the route along the river; Argyle Street. If people were minimising distance cycled, we would expect that street to have more cyclists.

Argyle Street is soon due to be reconfigured as part of Glasgow City Council's so-called Enabling Infrastructure – Integrated Public Realm (EIIIPR) programme (better known locally as the Avenues programme). This is a £115 million investment in improving the public realm in the city centre. Among the aims are introducing segregated cycle lanes and increasing the space available to pedestrians and cyclists. It is hoped this will increase cycling in the city. Our analysis suggests that there is demand for using this street and that it could result in time savings for cyclists. Several of the other routes due to be improved as part of the avenues programme are also visibly underused according to our measure.

We now turn our attention to a qualitative discussion of some of the other interesting features highlighted by the data. Two examples of the sorts of analysis which could be done using the enhanced data are presented.

4.1. Example 1

Fig. 4 shows part of the study area containing Paisley Road West (highlighted). The traffic assignment model suggests that this should be a popular street for cycling; note that it is directly connected to a popular route. However, the observed flows are significantly lower. The images collected from Google Street View (Fig. 5) for this location give some hints as to why this may be the case.

A number of features identified in the literature review can be seen in Fig. 5. In part a) of the figure note that there are bus stops on both sides of the road. In the distance, traffic lights with a pedestrian crossing can also be seen. We would expect these factors to have a negative effect. On the positive side, there is mixed land-use and trees nearby. In part b) of the figure we can see that there is parking on both sides of the street. Note there is also a car waiting by the parked cars which is obstructing part of the road. The presence of shops nearby may be related to this.

There are other reasons why the road may be less popular. At the start of the portion of the road we are looking at there is a road with a

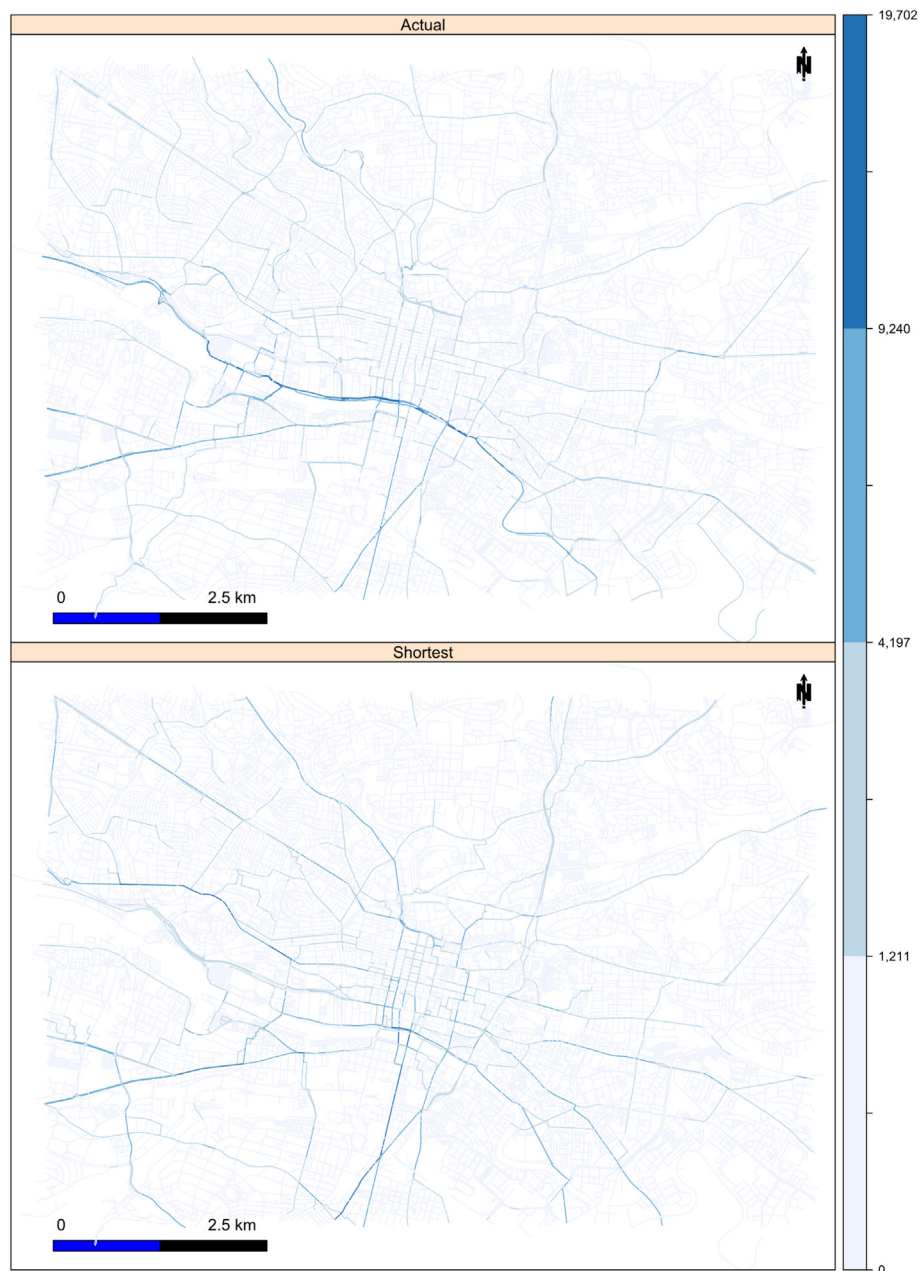


Fig. 2. Observed and modelled annual flows of cyclists in central Glasgow.

cycle lane (Whitefield Road) running north towards the river which is more popular than expected. It seems that cyclists choose to take this even though it may add extra distance to their journeys.

4.2. Example 2

Our previous example and the literature suggest that cycling infrastructure may be important in influencing where cyclists cycle. We examined roads where cycling infrastructure exists. One such example is Duke Street, highlighted in Fig. 6. While the route is used by cyclists, it is less popular than would be expected. Taking an image from Google Street View (Fig. 7) provides some clues as to why the route appears less popular than expected.

The cycling infrastructure on Duke Street consists of a shared cycle/bus lane which operates for only part of the day. Note in Fig. 7 that cars are parked in both lanes leaving little room for cyclists. These factors have been identified in the literature. We may also speculate that the

environment may be potentially problematic, for example there is a derelict property on the right side of the image. Data from the Scottish index of Multiple Deprivation (SIMD)⁵ shows this to be one of the most deprived areas of Scotland. It is also an area with a high crime rate. Some research has failed to find a link between cycling and crime (Hood et al., 2011), but Izadpanahi et al. (2017) note that further research is needed on this topic.

Importantly, this analysis does not mean that the road isn't used. Our data shows that this is a road with many cyclists. However, our augmented data shows that it is still not used as much as it should be expected given its location. A conventional approach to evaluation may consider the fact that the infrastructure is used as a sign of success, however our approach suggests otherwise.

To gain a more comprehensive view of the role of cycle infrastructure, we obtained information on the location of cycling

⁵ <http://www.gov.scot/Topics/Statistics/SIMD>.

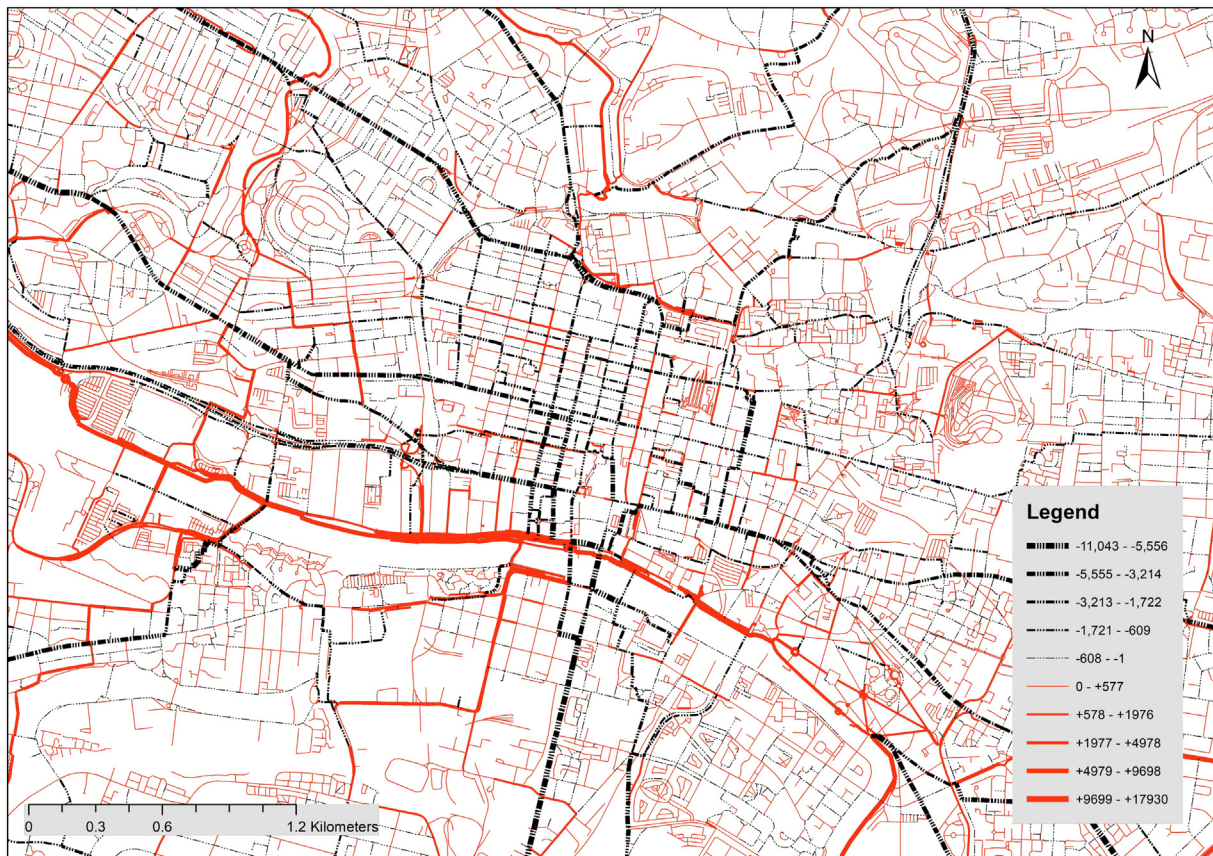


Fig. 3. Difference between observed flows and predicted flows.

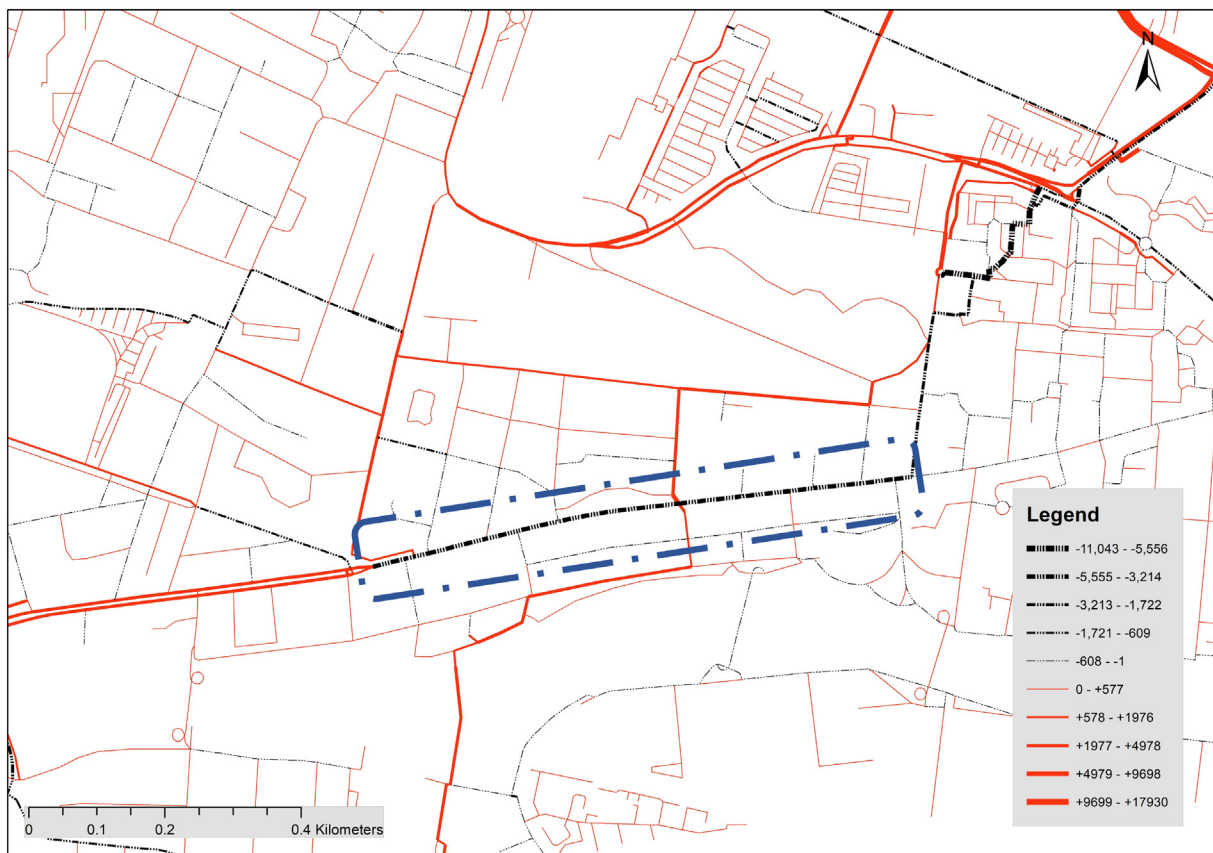


Fig. 4. Paisley Road West.



a)

b)

Fig. 5. Images of Paisley Road West from Google Street View.

infrastructure from Glasgow City Council. This was manually matched to the OpenStreetMap network. On links with cycling infrastructure, we found that in total, 8.2% more distance was travelled than was expected based on distance minimising behaviour. This suggests that people tend to be attracted to these links even if they are not links which help to minimise travel distance.

5. Conclusions and future work

New and novel forms of data present new challenges and opportunities to planners and policy makers. In this paper, we use data from the activity tracking app Strava to better understand how commuting

cyclists travel in Glasgow. The city has ambitious targets for cycling and is investing large amounts of money in cycling infrastructure. This makes it an interesting case and also increases the need for good quality evidence of where cycling occurs and where it might occur.

Many studies using Strava data to analyse patterns of cycling look only at observed flows. While this can reveal valuable insights, it doesn't necessarily tell us where people would like to cycle; only where they actually cycle. We suggest that one way to enhance the data is to compare the Strava flows to the flows which would be observed if cyclists took the shortest route. The approach allows us to compare the use of the network while controlling for the demand for travel between the origins and destinations in the area. We have illustrated different



Fig. 6. Duke Street is highlighted here. It has cycling infrastructure but appears to be less used than we would expect.



Fig. 7. Image of Duke Street from Google Street View.

ways in which this data could be visualised and analysed to support decision making.

The data and approach have numerous advantages. Data from the Strava app provide rich information about what cyclists do, where they cycle from and to, which roads they use, and how long they wait at junctions. The data is also available globally, which is particularly useful for countries where other data on cycling are unavailable. One key advantage of the method we propose is that it can be easily implemented anywhere Strava data is available. It does not require any additional data. The approach is also flexible. We use perhaps the simplest assignment method to illustrate the approach i.e., shortest path. However, any assignment method could be used with the results being compared to the observed flows. The observed data could also be used to calibrate a cycling route choice model.

There are potential limitations to the data. The cycling data is only available for users who use the Strava app and track their journey. This may introduce bias in terms of the cyclists we have data for and for the trips that they make. It could be the case that certain types of people are more likely to use the app than others; or that certain types of trips are more likely to be logged than others. Cyclists who use Strava may behave systematically differently to other cyclists. However, as the method works only using the Strava data, we are at least comparing like with like.

There may also be some issues introduced with the way the Strava data is produced. The process of attaching GPS points to road links (map matching) is subject to error. Some trips may be incorrectly assigned to a road link. This is more likely to occur when roads are dense. It is beyond the scope of this paper to enter into detailed evaluation of the quality of the data. However, there is growing literature on how to evaluate the quality of volunteered geographic information. For example, Mocnik et al. (2018) develop and present an ontology of data quality measures with examples of VGI. Senaratne et al. (2017) review various quality measures and indicators for VGI.

Overall, we think that the data and proposed method provides more advantages than disadvantages and can be a valuable tool for planners. It can highlight successes and failures in the current network, help appraise/evaluate interventions and raise new questions. Given the current lack of data on cycling in many cities, this represents a step forward.

Acknowledgements

The authors would like to acknowledge support from the Economic and Social Research Council-funded Urban Big Data Centre at the University of Glasgow (Grant ES/L011921/1). The data is available to others upon application to the centre. We would also like to thank four anonymous reviewers for constructive comments on previous versions of this paper.

References

- Börjesson, M., Eliasson, J., 2012. The value of time and external benefits in bicycle appraisal. *Transp. Res. A Policy Pract.* 46 (4), 673–683.
- Boss, D., Nelson, T., Winters, M., Ferster, C.J., 2018. Using crowdsourced data to monitor change in spatial patterns of bicycle ridership. *J. Transp. Health* 9, 226–233.
- Broach, J., Dill, J., Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transp. Res. A Policy Pract.* 46 (10), 1730–1740.
- Caulfield, B., 2014. Re-cycling a city – examining the growth of cycling in Dublin. *Transp. Res. A Policy Pract.* 61, 216–226.
- Caulfield, B., Brick, E., McCarthy, O.T., 2012. Determining bicycle infrastructure preferences – a case study of Dublin. *Transp. Res. Part D: Transp. Environ.* 17 (5), 413–417.
- Dill, J., Gliebe, J., 2008. Understanding and measuring bicycling behavior: a focus on travel time and route choice. *Urban studies and planning faculty publications and presentations*. Portland State University, Nohad A. In: Toulou School of Urban Studies and Planning.
- Ehrgott, M., Wang, J.Y.T., Raith, A., van Houtte, C., 2012. A bi-objective cyclist route choice model. *Transp. Res. A Policy Pract.* 46 (4), 652–663.
- Frank, L.D., Pivo, G., 1994. Impacts of mixed use and density on utilization of three modes of travel: Single-occupant vehicle, transit, and walking. *Transp. Res. Rec.* 1466, 44–52.
- Fraser, S.D.S., Lock, K., 2011. Cycling for transport and public health: a systematic review of the effect of the environment on cycling. *Eur. J. Pub. Health* 21 (6), 738–743.
- Garrard, J., Rode, G., Lo, S.K., 2008. Promoting transportation cycling for women: the role of bicycle infrastructure. *Prev. Med.* 46 (1), 55–59.
- Gatersleben, B., Appleton, K.M., 2007. Contemplating cycling to work: attitudes and perceptions in different stages of change. *Transp. Res. A Policy Pract.* 41 (4), 302–312.
- Granel, C., Ostermann, F.O., 2016. Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster management. *Comput. Environ. Urban. Syst.* 59, 231–243.
- Griffin, G.P., Jiao, J., 2015. Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *J. Transp. Health* 2 (2), 238–247.
- Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B., Xu, Z., 2012. Estimating use of non-motorized infrastructure: models of bicycle and pedestrian traffic in Minneapolis, MN. *Landscape Urban Plan.* 107 (3), 307–316.
- Haworth, J., 2016. Investigating the potential of activity tracking app data to estimate cycle flows in urban areas. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 41, 515.
- Haworth, B., Bruce, E., 2015. A review of volunteered geographic information for disaster management. *Geogr. Compass* 9 (5), 237–250.
- Heesch, K.C., Langdon, M., 2016. The usefulness of GPS bicycle tracking data for evaluating the impact of infrastructure change on cycling behaviour. *Health Promot. J. Aust.* 27 (3), 222–229.
- Hendawi, A.M., Sturm, E., Oliver, D., Shekhar, S., 2013. CrowdPath: a framework for next generation routing services using volunteered geographic information. In: *International Symposium on Spatial and Temporal Databases N. M. A. e. al. Berlin, Heidelberg*. Springer.
- Hood, J., Sall, E., Charlton, B., 2011. A GPS-based bicycle route choice model for San Francisco, California. *Transp. Lett.* 3 (1), 63–75.
- Hopkinson, P.G., Carsten, O.M., Tight, M.R., 1989. Review of literature on pedestrian and cyclist route choice criteria. In: *Working Paper 290*. University of Leeds, Institute for Transport Studies.
- Izadpanahi, P., Leao, S.Z., Lieske, S.N., Pettit, C., 2017. Factors motivating bicycling in Sydney: analysing crowdsourced data. In: *Proceedings of 33rd PLEA, International Conference on Passive and Low Energy Architecture, Volume III*, Transport.
- Jestico, B., Nelson, T., Winters, M., 2016. Mapping ridership using crowdsourced cycling data. *J. Transp. Geogr.* 52, 90–97.
- Keler, A., J.D., M., 2016. Safety-aware routing for motorised tourists based on open data and VGI. *J. Locat. Based Serv.* 10 (1), 64–77.
- Kockelman, K.M., 1997. Travel behavior as function of accessibility, land use mixing, and land use balance: evidence from San Francisco Bay area. *Transp. Res. Rec.* 1607, 116–125.
- Menghini, G., Carrasco, N., Schüssler, N., Axhausen, K.W., 2010. Route choice of cyclists in Zurich. *Transp. Res. A Policy Pract.* 44 (9), 754–765.
- Mobasheri, A., 2017. A rule-based spatial reasoning approach for OpenStreetMap data quality enrichment; case study of routing and navigation. *Sensors* 17 (11), 2498.
- Mocnik, F.-B., Mobasheri, A., Griesbaum, L., Eckle, M., Jacobs, C., Klöner, C., 2018. A grounding-based ontology of data quality measures. *J. Spat. Inf. Sci.* 2018 (16), 1–25.
- Musakwa, W., Selala, K.M., 2016. Mapping cycling patterns and trends using Strava Metro data in the city of Johannesburg, South Africa. *Data Brief* 9, 898–905.
- Norback, K., Janson, B., 2010. Automated bicycle counts. *Transp. Res. Rec.* 2190 (1), 11–18.
- Ogilvie, D., Egan, M., Petticrew, M., 2004. Promoting walking and cycling as an alternative to using cars: systematic review. *BMJ* 329 (7469), 1–5.
- Prandi, F., Soave, M., Devigili, F., De Amicis, R., Astyakopoulos, A., 2014. Collaboratively collected geodata to support routing service for disabled people. In: *Proceedings of the 11th International Symposium on Location-Based Services*.
- Pucher, J., Dill, J., Handy, S., 2010. Infrastructure, programs, and policies to increase bicycling: an international review. *Prev. Med.* 50, S106–S125.
- Romanillos, G., Austwick, M.Z., Ettema, D., De Kruijff, J., 2016. Big data and cycling. *Transp. Rev.* 36 (1), 114–133.

- Ryus, P., Laustsen, K.M., Proulx, F.R., Schneider, R.J., Hull, T., 2014. Methods and Technologies for Pedestrian and Bicycle Volume Data Collection: Phase 2. Transportation Research Board.
- Saelens, B.E., Sallis, J.F., Frank, L.D., 2003. Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Ann. Behav. Med.* 25 (2), 80–91.
- Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C., Haklay, M., 2017. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* 31 (1), 139–167.
- Sener, I.N., Eluru, N., Bhat, C.R., 2009. An analysis of bicycle route choice preferences in Texas, US. *Transportation* 36 (5), 511–539.
- Shen, L., Stopher, P.R., 2014. Review of GPS travel survey and GPS data-processing methods. *Transp. Rev.* 34 (3), 316–334.
- Song, Y., Preston, J., Ogilvie, D., 2017. New walking and cycling infrastructure and modal shift in the UK: a quasi-experimental panel study. *Transp. Res. A Policy Pract.* 95, 320–333.
- Strava, 2015. *Strava Metro: Comprehensive User Guide. Version 2.0 for 2015.*
- Sun, Y., Mobasheri, A., Hu, X., Wang, W., 2017. Investigating impacts of environmental factors on the cycling behavior of bicycle-sharing users. *Sustainability* 9 (6), 1060.
- Tilahun, N.Y., Levinson, D.M., Krizek, K.J., 2007. Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey. *Transp. Res. A Policy Pract.* 41 (4), 287–301.
- Transport Scotland, 2017. *Cycling Action Plan for Scotland 2017–2020: Cycling as a Form of Transport.*
- Van Holle, V., Van Cauwenberg, J., Deforche, B., Goubert, L., Maes, L., Nasar, J., Van de Weghe, N., Salmon, J., De Bourdeaudhuij, I., 2014. Environmental invitingness for transport-related cycling in middle-aged adults: a proof of concept study using photographs. *Transp. Res. A Policy Pract.* 69, 432–446.
- Wardman, M., Tight, M., Page, M., 2007. Factors influencing the propensity to cycle to work. *Transp. Res. A Policy Pract.* 41 (4), 339–350.
- Winters, M., Teschke, K., 2010. Route preferences among adults in the near market for bicycling: findings of the cycling in cities study. *Am. J. Health Promot.* 25 (1), 40–47.
- Winters, M., Davidson, G., Kao, D., Teschke, K., 2011. Motivators and deterrents of bicycling: comparing influences on decisions to ride. *Transportation* 38 (1), 153–168.
- Zhao, P., Li, S., 2017. Bicycle-metro integration in a growing city: the determinants of cycling as a transfer mode in metro station areas in Beijing. *Transp. Res. A Policy Pract.* 99, 46–60.
- Zipf, A., Mobasheri, A., Rousell, A., Hahmann, S., 2016. Crowdsourcing for individual needs—the case of routing and navigation for mobility-impaired persons. In: *European Handbook of Crowdsourced Geographic Information*, pp. 325.