Data-driven bicycle network analysis based on traditional counting methods and GPS traces from smartphone

Federico Rupi *, Cristian Poliziani, Joerg Schweizer

DICAM- University of Bologna, Viale Risorgimento 2, 40136, Italy; cristian.poliziani2@unibo.it (C.P.); joerg.schweizer@unibo.it (J.S.)
* Correspondence: federico.rupi@unibo.it; Tel.: +39-51-209-3345

Abstract: This research describes numerical methods to analyze the absolute transport demand of cyclists and then to quantify the road network weaknesses of a city with the aim to identify infrastructure improvements in favor of cyclists. The methods are based on a combination of bicycle counts and map-matched GPS traces and are demonstrated with the city of Bologna, Italy: the dataset is based on approximately 27,500 GPS traces from cyclists, recorded over a period of one month on a volunteer basis using a smartphone application. A first method estimates absolute, city-wide bicycle flows, by scaling map-matched bicycle flows of the entire network to manual and instrumental bicycle counts of the main bikeways of the city. As there is a good correlation between the two sources of flow data, the absolute bike-flows on the entire network have been correctly estimated.

A second method describes a novel link-deviation index, which quantifies for each network edge the total deviation generated for cyclists in terms of extra distances traveled with respect to the shortest possible route. The deviations are accepted by cyclists either to avoid unpleasant road attributes along the shortest route or to experience more favorable road attributes along the chosen route. The link deviation index indicates the planner which road links are contributing most to the total deviation of all cyclists – in this way, repelling and attracting road attributes for cyclists can be identified. This is why the deviation index is of practical help to prioritize bike infrastructure construction on individual road network links.

Keywords: GPS traces; cycling volumes; cyclists’ counts; cycling network; deviation index.

1. Introduction

Congestion of motorized traffic is one of the major problem for urban mobility, producing negative outcomes at economic, social and environmental level. In recent years, due to congestion, air pollution, climate change, energy scarcity and physical inactivity, an increasing importance has been attributed to sustainable transport modes, and in particular to cycling. Municipalities have drawn attention to these issues and started to implement different strategies to encourage a greater usage of bicycles on urban streets and to reduce car trips. In particular, many cities have decided to invest in the construction of quality bikeways with the intention to incentivize people to cycle even medium (and long) distance on a daily basis. Data on cycling volumes help to support this decision making; researchers have investigated the factors that influences ridership [1-15]. This data can be collected by the use of traditional manual or instrumental counts [3,11], which are characterized by some drawbacks. Traditional manual counts lack spatial detail and temporal coverage. Instrumental and permanent count stations do provide continuous data, but cover typically only a small number sections of the road network. More recently, the widespread use of smartphones and mobile applications for self-localization and navigation has increased the availability of observed cyclists’ data [8-10, 12-15]. This type of data provides detailed information about the origin/destination of the trips as well as the chosen routes.
This paper explains how to estimate the city-wide bicycle flows and how to identify weak points of the road network in terms of bicycle friendliness. Both methods are data driven, explicit and do not require the calibration of sophisticated models.

The methods are demonstrated for the city of Bologna, Italy, where bicycle flow-monitoring campaigns were conducted from 2009 to 2018. In order to show the importance of the relation between bikeway construction and bicycle flows, the measured flows are presented in function of bikeway meters per inhabitant. Cycling counts are compared with map-matched GPS traces recorded by a smartphone application to study the correlation between the two data sources. One of the main problems with GPS data is the representativeness of the data, because data collection is usually provided on a volunteer basis, which is not necessarily representative for the entire population [13]. Another problem is the level of detail of the network: in many cases, the success of identifying the correct network links from GPS points is limited if the bike network model is not sufficiently detailed [14, 15]. In the present work, care has been taken to obtain a realistic model of the bicycle network. The resulting good correlation between bicycle counts and map-matched GPS traces is exploited to reconstruct the absolute cyclists’ flows on all links of the analyzed road network, thus allowing to quantify the distribution of the ridership in Bologna.

The paper is organized as follows: section 2 describes the study area and the features of the bike network. Section 3 depicts the bicycle flows obtained by traditional (manual and instrumental) counting methods and by GPS data collected by smartphone application. Moreover, section 3 identifies a correlation between cycling counts and GPS data and describes the bicycle flow reconstruction method. In section 4 a deviation analysis is carried out. Concluding remarks and future research directions are presented in section 5.

2. Dataset description

2.1. Study area

Bologna is a northern Italian city with approximately 390,000 inhabitants [16]. The climate is convenient for cycling all year, with an annual average temperature slightly below 15°C and low rainfall (about 700 mm rain/year and 74 days of rain per year).

The home-to-work bicycle mode share was 8.2% in 2011 [17], which is relatively high compared with other medium to large Italian cities. Nevertheless, the car ownership equals 0.515 cars per inhabitant [16], which corresponds to 0.97 cars per household.

2.2. Bicycle network

The Municipality of Bologna has made substantial investments in bikeways during the past decade and to date the city offers 129 km bikeways of different types: exclusive access and mixed access with pedestrians or buses [16]. The bicycle network layout is composed of 13 main radial bicycle paths, connecting the suburbs to the city center, and of many other bikeways connecting the radial bike-paths. The bikeway meters per citizen increased by 45% starting with 0.228 m/citizen in the year 2009 and reaching 0.330 m/citizen in 2018 [16]. This is an almost linearly-increasing expansion of the cycling infrastructure.

3. Bicycle flow analysis

3.1. Cyclists’ flows from traditional counting methods

In the period 2009-2018, manual and instrumental counts of cyclists were carried out by DICAM-Transport of the University of Bologna [18]: the bicycle counts had been conducted from September to October of each year. In recent years, counting has also been performed in May with the aim to evaluate the difference in bicycle flows of different periods of the same year. Locations of bicycle counters had been selected adopting representative and targeted locations: the sites include different geographic areas of the city, different types of bikeways as well as “pinch points” (i.e. locations where cyclists must converge to cross a barrier) [11]. The 46 (bidirectional) road-sections monitored in 2018...
are showed in figure 1, highlighting the spatial distribution of measurement points. The monitored road-sections included the 13 main radial bicycle paths.

![Bologna map with marked road sections](image)

**Figure 1.** Road sections monitored in 2018.

Manual and instrumental counting was conducted at each road section from 8:30 a.m. to 10:30 a.m. on weekdays. The trips purpose during this time period is most likely “work” or “study” and trips with both do have a clear destination, thus excluding round-trips or random trips for recreation purposes. The total average flows have increased between 2009 and 2018 by approximately 75%, which is significantly greater than the increase in bikeway meters per inhabitant in the same period.

Fig. 2 shows the correlation between bikeway meters per inhabitant and the total average bicycle flows: each point represents one year from 2009 to 2018.

![Graph showing correlation](image)

**Figure 2.** Regression function between length of cycleways for inhabitant and bike flows.

As shown in Fig. 2, the total average bike flows are positively and highly correlated with the length of cycleways per inhabitant ($R^2 = 0.96$). In the city of Bologna, people use bicycles more often than in the past. Surely, such an increase in cycling is determined, like in other cities, by an integrated package of many different, complementary measures, including infrastructure provision, pro-bicycle...
programs, supportive land use planning and restrictions of car use [19]. However, today’s bicycle network of Bologna connects the most popular origins and destinations and, as a result, the expansion of the cycling network has been perceived as an increased level of safety. The increasing bicycle use is also related to an increasing bicycle use of females, growing from a share of below 30% in 2009 to a share of 44% in 2018 [18].

Using the regression function of figure 2, we can make the hypothesis that the addition of one bikeway centimeter per inhabitant generates an increase of about 100 cyclists per hour on the main sections of Bologna’s bicycle network. Based on the length increase of the bicycle network, the estimated bicycle mode share is currently around 10%, following the model proposed by Schweizer and Rupi [20].

3.2. Map matched cyclists’ volumes

A database with GPS traces has been obtained from a data collection initiative called the “European Cycling Challenge ECC” [21] which took place in May 2016. In particular, the city of Bologna participated in this initiative among other 51 cities from 18 European countries. In Bologna, 1123 participants, equal to 0.3% of the population, recorded the GPS traces of their bicycle trips during the month of May 2016 by means of a mobile phone application. The participation has been on a volunteer basis. The total distance travelled by all participating cyclists has been almost 200,000 km and the database contains over 7,998,000 GPS points, with 27,348 individual trips, covering the entire road network of Bologna.

In addition, information regarding the bike users has been provided, such as gender, age and profession: in particular, 40% of the sample were female, matching well with the share of females observed during the manual counts. Consequently, the sample of cyclists recording the GPS traces is representative for the gender of the counted cyclists.

The analysis focuses only on morning trips from 8:30 a.m. to 10:30 a.m. during work-days in order to obtain flow values comparable with manual and instrumental bicycle counts. In order to obtain bicycle flows on network links, the GPS data has been matched to the road-network, obtained by converting Open Street Map data into a SUMO network, as reported in Rupi and Schweizer [14]. The SUMO network has been manually corrected and enhanced, such that cyclists could potentially pass everywhere, including footpaths and the opposite direction of one-way roads (which is an “illegal” behavior in Italy). The final network contains 13,959 nodes and 38,324 links. In order to match the GPS points to network links with a high accuracy and to obtain a large number of correctly matched GPS traces, the entire map-matching analysis consists of 4 phases: (i) an initial filtering process, (ii) the actual map matching process itself, (iii) a post-filtering process and (iv) a final analysis of the matched routes. The employed map matching algorithm is based on a method proposed by Marchal et al. [22] and improved by Schweizer et al. [23]. Initially, many GPS traces could not be matched to the network due to missing links or missing access. Successively, the reasons for the failed matching of the trips have been analyzed in detail and missing network links or road access attributes have been added. Successively the mapmatching process has been repeated with a higher number of successfully matched trips.

After the map-matching process and a quality ensuring filtering process, 4,029 map-matched routes, collected from 842 users, have been used. These traces correspond to 91.6% of all traces recorded during the considered morning. It is worth mentioning that this is a significantly higher percentage than reported in other studies [14, 15]. Starting from these map-matched routes, the bicycle flows (as number of cyclists passing through each network link per hour) have been evaluated.

3.3. Estimated cyclists’ volumes

A linear regression between the manually and instrumentally counted cyclists and map matched GPS traces in the monitored road sections has been carried out. The map matched bicycle volumes have been multiplied by a coefficient c in order to minimize the difference between the measured flows and flows derived from GPS data.
The regression, shown in Fig. 3, is based on the flow-comparison at 23 monitored sections ($c = 0.91$).

![Figure 3](http://example.com/figure3.png)

**Figure 3.** Regression function between manually counted cyclists’ volumes and map matched cyclists’ volumes (May 2016).

The slope of the linear regression function is almost equal to one, highlighting that the average of map-matched cyclist-volumes are equal to the average of manually counted cyclist-volumes.

The relatively high degree of correlation between the measured flows and the flows from the map matched GPS traces is evident ($R^2 = 0.73$). This result is significantly better than the results obtained by other studies; i.e., Jestico et al. [10] have obtained an $R^2$ equal to 0.4 for the a.m. peak period. The reason for this difference is likely due to the more detailed representation of the Bologna road network, representing more precisely cyclists’ freedom to move on all links in both directions. Based on this correlation, one crowdsourced cyclist corresponds in average to 59 cyclists at a count station, which is consistent with previous findings in [10].

Although crowdsourced cyclists represent a small portion of all cyclists, the flows obtained from the map matched GPS data are consistent with the observed flows on the main sections of the Bologna cycle network. Given the good correlation between GPS dataset of year 2016 and manual/instrumental counts, the linear relation between both flow types has been used to determine the flows on all network links where GPS points have been detected. The resulting link flows in cyclists per hour per direction are shown in the Fig. 4. This map is particularly useful to quantify the spatial distribution of ridership and provide important cycling exposure data for safety studies. Starting from this map, it is possible to obtain the OD matrix of cyclists, the chosen routes and the bicycle flow on every link of the network - essential data for modelling the cyclists’ route choice behavior and for planning the bicycle network.
4. Deviation analysis

The deviation analysis aims at identifying the network links which are the most avoided by all cyclists who registered GPS traces. The analysis starts with the following basic assumption: given the choice of two routes with identical properties (same safety, pavement, environment, etc.), cyclists would always choose the shortest one. If this is true, the cyclist would only accept a longer route if it offers better properties (safer, quieter, etc.). From a different perspective, if certain road links are avoided by deviating on alternative links, then the avoided links are supposed to possess less attractive characteristics, even though these characteristics may be good in the absolute sense. In an ideal bicycle network, no cyclists should feel constraint to take a longer route due to some repelling characteristics of the shortest route or due to better characteristics of longer routes. The most “avoided links” of the city’s road network are therefore identified with the km of deviation caused to cyclists. The deviation index for each road link is calculated in the following way:

1. For all matched routes, calculate the shortest route by connecting the first and last edge of each matched route.

2. For each matched route, identify all non-overlapping sections where links deviate from the shortest route.

3. For each of these non-overlapping sections, calculate the partial deviation which is the difference between the length of the part of the chosen route and the length of the corresponding part of the shortest route; add the partial deviation to the deviation indexes of all links on the shortest route of this section, see also illustration of Fig. 5.

**Fig. 4.** Estimated bicycle flows in cyclists per hour during workday morning peak hours (from 8:30 to 10:30). Flows only on network links where GPS points have been detected.
Fig. 5 shows links 1, 2 and 3 which, despite they are part of the shortest route, are not chosen (solid line); whereas, links 4, 5 and 6 are part of the chosen route (dashed line). With the non-overlapping section between node A and B shown in Fig. 5, the chosen route is constituted also by links 4, 5 and 6, while the shortest route contains links 1, 2 and 3. If \( L_i \) is the length of generic link \( i \), then the partial deviation attributed to links 1, 2 and 3 equals to \( L_4 + L_5 + L_6 - (L_1 + L_2 + L_3) \). The deviation index of a road link is the sum of all partial deviations received from all non-overlapping sections of all matched trips. The deviation index for the central part of Bologna network is shown in Fig. 6.

![Figure 5. Illustration of the calculation of the deviation index for the non-overlapping route section between nodes A and B.](image)

![Figure 6. Deviation index determined for the central part of Bologna network.](image)

The highest deviation index can be seen on the main radial roads from and into the city center. As seen in Fig. 4, these are also roads with high bicycle flows. This means that many cyclists actually do use these radial roads but also many try to avoid them. Note that there are also roads in the city center with high bicycle flows, but generating almost no deviations. The deviation index quantifies the deviations generated by road links but does not identify the reasons for the deviations. However, it is evident that those radial roads with high bike flows and high deviations are characterized by an absence of reserved bike lanes, a high level of bus traffic, often on reserved bus lanes, and a high density of intersections. In contrast, roads where bicycle flows are high but deviations are low, are generally less trafficked, often provide bike lanes and are not part of major bus routes.
Analysing the road attributes of the chosen part and the shortest part of all non-overlapping sections of all trips, the causes for the deviations become clearer, see first 3 columns of Tab. 1. As expected, cyclists accept deviations in order to travel on roads with: 1.) a high share of reserved bikeways, 2.) a low priority (roads with one lane per direction and speed limits of 30km/h), 3.) a low intersection density and 4.) a low share of mixed access, such as bike&bus or bike&pedestrians. In average, the chosen route-parts are 20% longer with respect to the shortest route-parts.

The statistics of the road link attributes of the overlapping sections of each trip (i.e. all links where chosen and shortest routes coincide) are presented in the last column of Tab.1. It becomes evident that the values of the mixed road access share, the reserved bikeway share and the intersection density are in between the values of the shortest route (column 1) and the chosen route (column 2) of the non-overlapping sections. One could conclude that cyclists tend to deviate if road attribute values are below/above those of the overlapping sections. An exception is the low priority road share, where the overlapping sections show values even below the one of the shortest route.

<table>
<thead>
<tr>
<th>Non-overlapping sections</th>
<th>Overlapping sections</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shortest route</td>
</tr>
<tr>
<td>Total length [km]</td>
<td>8265</td>
</tr>
<tr>
<td>Mixed road access share</td>
<td>32.6%</td>
</tr>
<tr>
<td>Low priority road share</td>
<td>50.0%</td>
</tr>
<tr>
<td>Reserved bikeway share</td>
<td>16.4%</td>
</tr>
<tr>
<td>Intersection density [1/km]</td>
<td>18.5</td>
</tr>
</tbody>
</table>

However, the deviation index depends on the presence of route alternatives to the shortest route, and their respective road attributes: in case there are no feasible route alternatives to deviate a certain link, then the deviation index of the respective link is zero, even though attributes are unfavourable. In case the shortest route has favourable link attributes, but the alternative has even more favourable link attributes then the deviation index is high, despite the good conditions on the shortest route. The former case is the most severe as criticalities of unfavourable roads for cyclists without route alternatives remain undiscovered by the deviation analysis.

5. Conclusions

In this research, the cyclists’ flows obtained by traditional counting methods have been compared with GPS traces from smartphone at the same locations and during the same time period. Although crowdsourced cyclists represent often a small portion of all cyclists, they do represent well the ridership of Bologna in terms of cyclists’ volumes and gender distribution. This result emerges clearly by comparing traditional counting method with GPS traces, confirming their representativeness of the population. The correlation between cycling counts and GPS data collected by smartphones has been relatively high, with an $R^2$ value of 0.73. This correlation is significantly higher than the results obtained by other studies, most likely due to the more detailed representation of the Bologna network, including footpaths in parks and the possibility to cycle one-way roads in both directions. Due to this good correlation, it has been possible to estimate the absolute bicycle flows on all network links by an appropriate scaling of the map-matched flows. The cyclists’ routes are of great value for cycling infrastructure planning and the drafting of cycling policies. The proposed method, which combines bicycle counts at a few main road sections with area covering GPS traces, can readily be applied in other cities in order to reliably estimate the absolute bike flows of an entire urban area.

GPS data have been further used to determine the deviation index, which counts the total deviations that a road link causes to cyclists. The deviation index is useful to identify weak links of
the cycling network, but it does not identify the reason why certain road links are avoided. However, applying the deviation index to the Bologna road network, the highest deviation index have been seen on trafficked roads without physically protected bike lane. Also roads with reserved bus lanes, which are open for bicycles too, showed high deviation rates. Further analyses of chosen and shortest road sections have shown that cyclists are willing to make deviations when the alternative route provides a high share of reserved bikeways, a high share of low-priority lanes, a low intersection density and a low share of roads with mixed traffic (with buses and pedestrians). Obviously the deviation index does not reveal deviations if there are no alternatives to avoid a certain road link.

In future works, the representativeness of the results could be improved by statistically weighting the GPS traces according to different person attributes, such as occupation, gender or age. The deviation index of an individual link could be modelled by a linear combination of link attributes such as the presence of reserved bikeways, reserved bus lanes, vehicle flows-rates etc. Further research can combine an alternative routes analysis with the cyclists’ preferences revealed by the deviation index, in order to identify edges which prioritize intervention.

Author Contributions: Study conceptualization, Federico Rupi and Joerg Schweizer; Methodology, Federico Rupi and Joerg Schweizer; Programming and data retrieval, Federico Rupi, Joerg Schweizer and Cristian Poliziani; Formal GIS analysis, Joerg Schweizer and Cristian Poliziani; Writing—review and editing, Federico Rupi, Joerg Schweizer and Cristian Poliziani.

Funding: Open access funding provided by Department of Civil, Chemical, Environmental, and Materials Engineering (DICAM)—University of Bologna.

Acknowledgments: We are grateful to SRM Bologna srl for providing the GPS data of the European Cycling Challenge 2016.

Conflicts of Interest: The authors declare no conflict of interest.

References

5. Nelson, A. C., Allen, D. If you build them commuters will use them – association between bicycle facilities and bicycle commuting. Transportation Research Record 1997, 1578, 79-83
6. Dill, J., Carr, T. Bicycle Commuting facilities in major US cities: if you build them commuters will use them – another look. Transportation Research Record 2003, 1828, 116-123.


