Cyclists do better. Analyzing urban cycling operating speeds and accessibility

Article in International Journal of Sustainable Transportation · March 2019

DOI: 10.1080/15568318.2019.1575493

2 authors:

Gustavo Romanillos
Complutense University of Madrid

Javier Gutiérrez
Complutense University of Madrid

Cyclists do better. Analysing urban cycling operating speeds and accessibility

Gustavo Romanillos*[1] and Javier Gutiérrez[1] (2019)

[1] tGIS Transport, Infrastructure and territory Research group, Universidad Complutense de Madrid.

* Corresponding author. Email: gustavro@ucm.es.

*This is a preprint version. Paper originally published in Journal of Urban Planning and Development on 1 March 2019. Available at: https://www.tandfonline.com/doi/full/10.1080/15568318.2019.1575493

Cite as: Gustavo Romanillos & Javier Gutiérrez (2019) Cyclists do better. Analyzing urban cycling operating speeds and accessibility, International Journal of Sustainable Transportation, DOI: 10.1080/15568318.2019.1575493.

Abstract: Given the growing interest in promoting more sustainable urban transport, over the past years, many researchers have analysed cycling mobility from different perspectives. However, some important aspects remain underexplored. One of them is the study of cycling speeds. The first goal of this research is to analyse the impact of a wide range of factor on cyclists' speeds. Based on the examination of thousands of GPS routes, the investigation conducts diverse OLS regressions in order to analyse cyclists' speed according to the diverse local factors that affect cyclists along their journey, such as the slope, the existence –and type– of bike infrastructures, the average traffic speed or the density of traffic lights or intersections. Cycling speed is also analysed according to cyclists' gender or age, the purpose of the journey, or even the weather conditions. The research includes the analysis of regular cyclists' trips, as well as the analysis of bike-messengers' routes.

The results obtained shed light on the influence of these factors on cyclists' speed by quantifying their specific impact, and diverse models predict cyclists' travel times in the current scenario but also in future ones that may correspond to the implementation of new infrastructure or policies. In addition, the models allowed us to pursue the second goal of this study: to conduct a comparative analysis of accessibility for different transport modes, and then evaluate the competitiveness between them. The results evidence that cycling is not only a sustainable transport mode, but the most competitive for small-medium distances.

Keywords: cycling, bike mobility, speed, accessibility, transport competitiveness

1. Introduction and background

In the context of an increasing interest in promoting a more sustainable urban transport, many researchers have analysed cyclist's behaviour in order to obtain the understanding necessary to foster effective policies and build useful cycle infrastructures. With different purposes, these analysis had traditionally been based on household or specific group surveys carried out through Stated Preference techniques (Kroes and Sheldon, 1988; Ortúzar, Iacobelli and Valeze, 2000), Revealed Preference methods that, for instance, asked cyclists to design their routes on a map (Ben-aiuva & Morikawa, 1990), or mixed techniques that combined both methodologies (Yang and Mesbah, 2013). However, over the last ten years, the emergence of new location devices and the consequent availability of geolocated data have led to a growing number of studies (Romanillos et al., 2016) that have shed light on different aspects of cycling mobility that were underexplored.

Especially significant was the research based on the analysis of Global Positioning System (GPS) data collected through different initiatives and, more recently, the studies based on the thousands of GPS routes collected by some app companies widely used by cyclists, Strava perhaps being the most remarkable one. However, although the volume of data that these companies make available surpasses the data collected by any of the previous research initiatives by far, the data have an important limitation: in order to preserve the anonymity of users, the data are always aggregated, so the single GPS tracks are not available and there is not even any associated information about cyclists (such as age or gender) or the trip (such as the purpose of the journey). The impact of these factors was exceptionally studied by some research initiatives (Hudson *et al.*, 2012) focusing on the analysis of their own data samples, yet in any case, they rely on the analysis of the data aggregated by route. Furthermore, most of these studies are essentially focussed on the development of route-choice models (Romanillos *et al.*, 2016) rather than on a detailed study of cycling patterns according to socio-demographic profiles or to different characteristics of the network. Thus, there are some important cycling aspects or dynamics that have not been properly explored.

One of the important cycling dynamics that remains underexplored is the estimation of cycling operating speed according to different local factors (slope, traffic, types of roads or bike lane, etc.), the different characteristics of cyclists (gender and age) or other variables, such as the purpose of the journey. The analysis of cyclists' operating speeds according to these (and other) factors is essential for the study of key aspects of cycling mobility, such as the estimation of travel times — already studied by Salonen & Toivonen (2013) — and the potentially derived accessibility analyses, the study of cycling competitiveness in relation to other transport modes — already explored by Ellison & Greaves (2011); Börjesson & Eliasson (2012) and Witlox & Tindermans (2004) — and for the correct development and calibration of cycling route-choice models, as well as for the evaluation of cycling risks and the planning and design of bike infrastructures and policies.

However, with some exceptions mostly focussed on Chinese case studies, cycling speeds have yet to be studied in detail. Liu, Shen and Ren (1993) analysed cyclists' free-flow speed in Beijing and reported an average of 14kph. (Wei, Huang and Wang, 1997) studied the different free-flow speeds on regular roads and in segregated bike lanes, resulting in 13.9kph and 18.2, respectively. Cherry (2007) conducted an exhaustive research on electric bike mobility in Chinese cities, reporting a free-flow speed of 18.2kph, significantly higher than the 13.0khp that corresponded to classic bikes. Just one year later, Lin, He, Tan, & He (2008) went one step further and analysed the operating speeds of 552 e-bicycle riders and 232 bicycle riders in the city of Kunming, China, according to age and gender. Though their results were interesting and evidenced different speeds, the followed methodology presented some remarkable limitations. The study was only conducted on bicycle-exclusive lanes and in straight and flat-road sections. The selected road was then video-recorded and the operating speed was measured by manually registering the time that it took cyclists to ride a measured distance. Apart from this, cyclists were classified into three different age groups (under 25, 25-50 and over 50 years old), based on the researcher's decision. Similar methodologies of vision-based analysis of cyclists' speed, relatively improved by automated video analysis techniques, have been applied (Kassim *et al.*, 2012), but with a limited sample in terms of locations. More recently, similar results have been found in Hangzhou, China, by (Jin *et al.*, 2015) when estimating cycleway capacities, and by Xu et al. (2015), who developed several models with the similar aim of predicting cycling flow speed according to the cycleway width, the type of bike and different characteristics of cyclists (age and gender). Finally, other studies have also analysed cyclists' crossing speeds at signalized intersections and traffic lights (Guo *et al.*, 2014), a fundamental measure when evaluating crash risk and its severity, which has also been highlighted and studied by Vlakveld et al. (2014) and Xu et al. (2015). In summary, as far as we know, cycling speeds have not been studied considering a wide range of factors at the same time, so the knowledge that we have about it is always narrowed to a very limited sample and the influence of specific and local aspects.

This research has two main goals. The first is to perform a detailed analysis on cyclists' operating speeds, according to a wide range of factors. The second is to apply the results of this analysis to estimate cycling travel times and conduct a comparative analysis on accessibility, evaluating competitiveness between different transport modes: cycling, walking, private car and public transport.

With this aim, initially, in relation to the first goal, we have studied the thousands of trips collected through the *Madrid Cycle Track* research initiative (Romanillos and Zaltz Austwick, 2015) by exploring their correspondent GPS dataset at a track-point level. Cyclists' operating speed can be then analysed according to important local factors that affect cyclists along the different sections of their journey, such as the slope, the type of road, the existence –and type- of bike infrastructures, the motor traffic speed or the distance to a signalized intersection or a non-signalized one. Since the initiative collected volunteers' information as well, cyclists' speed is also analysed according to age and gender, in addition to the purpose of the journey. The research includes the analysis of regular cyclists' trips, as well as the analysis of bike messengers' routes. Bike messengers companies are emerging in many major cities, playing a growing role in the delivery of packages and mail (Hong, Wei and Wei, 2006; Kidder, 2008; Fleming, 2012). Although cycle couriering has been considered a competitive transport mode for the delivery of time-sensitive materials in the core of many metropolitan areas (Kidder, 2011), as far as we know, bike messengers operating speeds have not been analysed in detail yet. For this research, bike messenger routes were collected from four different companies through the same initiative. Following a similar methodology, their different operating speeds are analysed according to the different types of bikes (regular, cargo, e-cargo bikes and e-tricycles), and then compared to the results obtained for regular cyclists. By conducting diverse OLS regressions to analyse cyclists' speed considering all these factors, it was possible to estimate cyclists' travel times for the whole street network (and not only in the street-network arcs where we have cyclists' records) and, furthermore, it would also be possible to predict cyclists' travel times and accessibility in future scenarios.

Finally, regarding the second goal, a series of isochrones from a central point in Madrid were calculated, according to the estimated cycling speed, an average walking speed, private car speeds in Madrid (considering a high-resolution network and accurate traffic speeds obtained from the *TomTom*® database) and public transport travel times according to the General Transit Feed Specification (GTFS) data obtained from the Regional Transport Consortium of Madrid. Different maps illustrate the diverse isochrones and a table summarizes the area covered by each transport mode, revealing that cycling is the most competitive one for trips under 20 minutes.

The article is structured as follows: after the introduction, Section 2 and 3 describe the data and the methodology, respectively, Section 4 shows the main results, and Section 5 presents the conclusions and final remarks.

2. Data

2.1. Cyclists' data collection: The Madrid Cycle Track initiative

This research is based on the analysis of the cyclists' trips collected through the *Madrid Cycle Track* initiative (www.huellaciclistademadrid.es). Since the data gathered have already been described in detail by Romanillos & Zaltz Austwick (2015), only the basic figures will be introduced here. The initiative collected 6,022 cycle routes uploaded by 328 volunteers, taking into account regular cyclists and bike-messengers, resulting in 48,122 km of cycling tracks obtained from 3,970 journeys (Figure 1). Participants were not selected by researchers. Regular cyclists responded to dissemination of the initiative through different channels, essentially cycling association social networks. In this sense, it cannot be guaranteed that the sample is completely random, since cyclists connected to these channels could be more active and experienced than the average cyclist. The bike messenger sample are all different cyclists working at the companies.

When uploading the routes, the *Madrid Cycle Track* online platform allowed volunteers to provide extra information about their routes (such as the purpose of the journey) or about themselves (such as age and gender). Regarding the purpose of the journey, 42.19 % of the routes corresponds to commuting, 23.28% to leisure, 10.76% to study, 8.55% to shopping, 7.76% to sport and 7.47% to errands. In terms of gender, the proportion of males and females in this sample was 72%-28%, respectively. These figures, reflecting the gender imbalance, match other existing local surveys (DOYMO, 2011; Monzon de Cáceres, Rondinella and Muñoz López, 2011). This study does not include routes travelled with e-bikes, since the sample obtained was very poor. In addition, four different bike messenger companies participated in the initiative, providing 2,052 routes and 10,777 cycled kilometres. The total number of bike messengers participating in the initiative was 23.

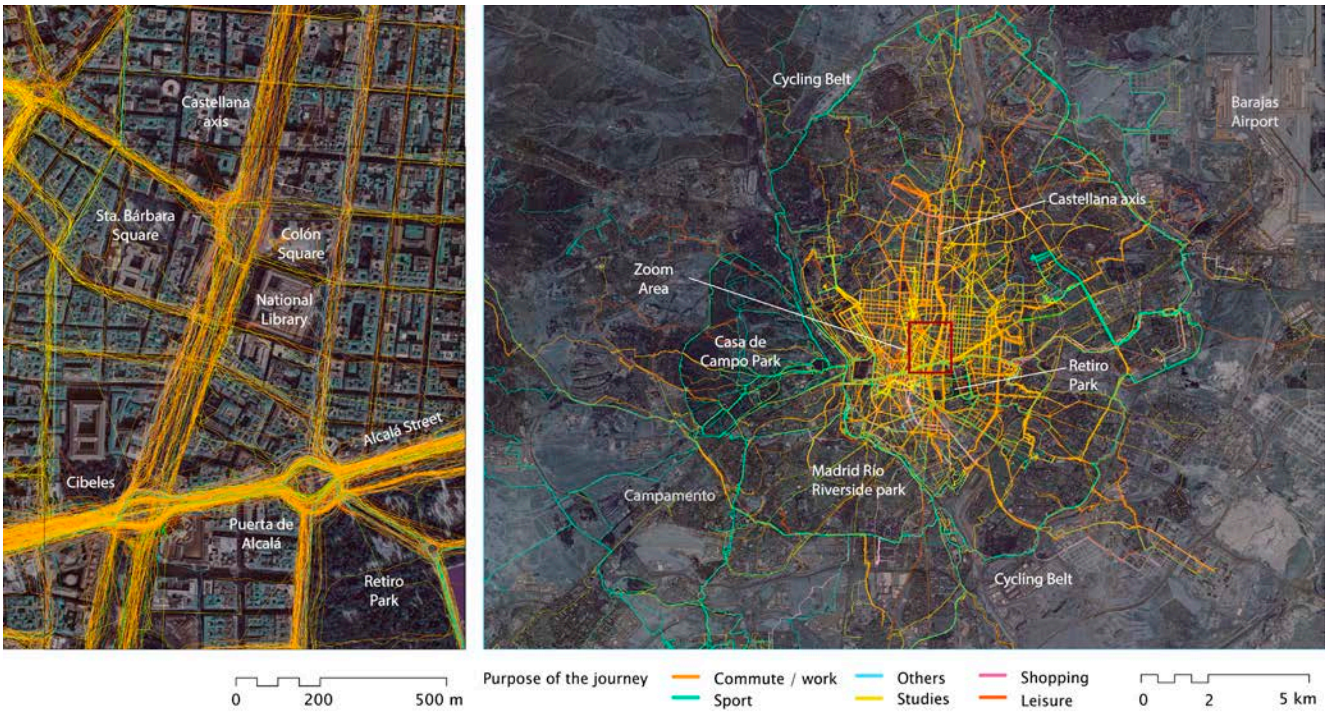


Figure 1: GPS routes collected through the Madrid Cycle Track initiative.

While the paper published by Romanillos & Zaltz Austwick (2015) focussed on visualising the global cycle track of the city of Madrid, this study explores the GPS dataset collected at a track-point level. The GPS app registered cyclists' location every two seconds, providing records on latitude, longitude and local time. By generating the GPS track-lines from the points, it was possible to calculate cyclists' speed every two seconds. Before going into the process of GPS data cleaning (described later in the Data preprocessing sub-section), the simple visualization of these GPS track lines according to cyclists' speed (Figure 2) evidences the impact that different factors have on it. For instance, the figure illustrates several cyclists' tracks along a steep street (Avda-, showing the speed asymmetry between those going up (from left to right) and those going down (from right to left), as well as the impact of street junctions, where cyclists' speeds are significantly reduced. The goal of this study is to explore, at the level of spatial accuracy that GPS records bring, the impact of these and other factors. In this sense, the Madrid Cycle Track initiative collected information about other variables that have a potential impact on cycling speeds, such as age and gender (regarding cyclists) or the purpose of the journey (regarding routes).

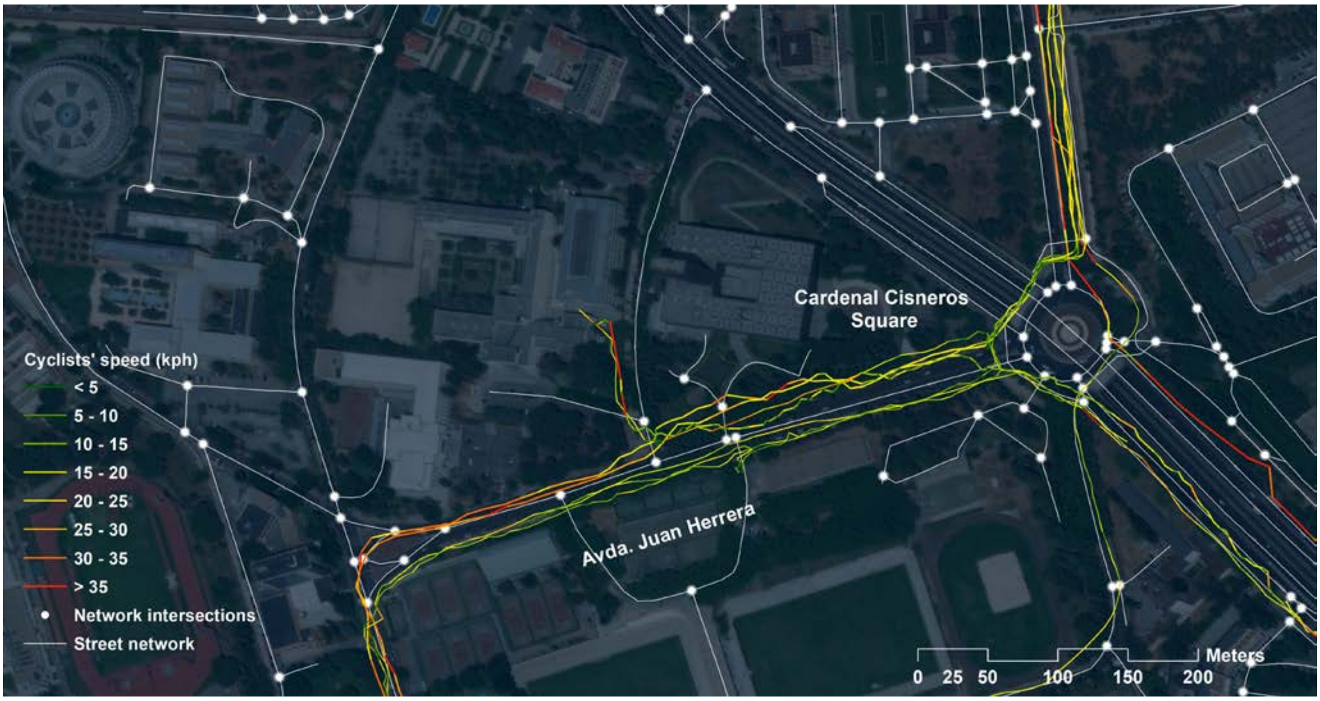


Figure 2. GPS track lines represented according to cycling speed.

2.1. Other data sources

In order to improve the analysis by including other variables that potentially affects cyclists' operating speeds, the routes collected were fed with data from different sources. All the datasets were integrated in a Geographic Information Systems (GIS) environment.

First, the GPS track lines were map-matched to a detailed street network based on the March 2013 version of *TomTom*® for the Spanish road network. The *TomTom*® network is actually the most accurate street network found in Madrid, contemplating not only roads but also pedestrian streets and bike infrastructure, and includes over 160,000 street-segments for the metropolitan area of Madrid that cover the collected cycling routes. In addition, TomTom® databases provided relevant information, such as the direction of traffic, or information on variables that could influence cyclists' operating speed, such as the maximum motor traffic speed or the real average traffic speed per road segment and according to different time frames (average real speed on weekdays, during the weekend, on weekdays during rush hour, etc.).

After this, the *TomTom*® street network was edited and completed with relevant information obtained from other local data sources. Slopes were calculated for each street segment by calculating the elevation for each node of the street segments from a high resolution Digital Elevation Model (cell size = 5 meters) obtained from the National Geographic Institute of Spain (http://www.ign.es). Although the *TomTom*® street network was supposed to contemplate the existing bike infrastructure, it was not complete. Because of this, the network was edited and the bike infrastructure updated, including all the eight different kinds of bike infrastructure considered in the Madrid Cycling Master Plan, listed in Section 3.2.1. This information was downloaded from the Madrid Open Data Portal (http://datos.madrid.es), where it was also possible to obtain the geolocated datasets of the existing traffic lights. When comparing accessibility between different transport modes, we considered the *TomTom*® speed profiles (every 5 minutes) in order to estimate the average speed at 8h.

Finally, in order to perform analysis of cycling competitiveness in relation to other transport modes, public transport (bus, train, tram and underground) travel times according to the General Transit Feed Specification (GTFS) data were obtained from the Regional Transport Consortium of Madrid. This data also provides travel times for every hour over the course of the day.

3. Methodology

3.1. Data preprocessing

3.1.1. Map matching process

The Madrid Cycle Track initiative collected the cyclists' routes through two different applications (*Map My Tracks* and *Garmin Connect*), which recorded cyclists' location every two seconds in GPX —or GPS exchange— format, providing records on latitude, longitude and local time. The GPX files were downloaded from the app and imported into a GIS environment, where GPS track-points were obtained and GPS track-lines were eventually generated from them. In order to analyse cyclists' routes according to the different variables that we were considering, the GPS track-line obtained must be matched to our street network. This is commonly known as the map-matching process, and it has been tackled by researchers following different procedures (Schuessler and Axhausen, 2009). For this study, based on the map-matching algorithm created by Dalumpines & Scott (2011), a new version was developed, improving an aspect that was relevant for the purpose of our research.

The commonly known map-matching process has been tackled by researchers following different procedures, described and classified by Schuessler & Axhausen (2009), who also implemented the advanced map-matching algorithm used by Hood et al. (2011) when analyzing the cycle tracks collected in San Francisco. The results obtained revealed the complexity of the problem: only 1,454 out of the 2,282 original traces were matched to the network, and still some of them could contain the significant errors common to the map-matching geometric procedures, that even led to some researchers to check each route and manually correct the errors (Snizek, Sick Nielsen and Skov-Petersen, 2013). Much better map-matching results are obtained when combining geometric and topological procedures, since they "consider the connectivity of the network in assessing the feasibility of a route" (Hudson *et al.*, 2012). Between these hybrid approaches, the algorithm developed by (Dalumpines and Scott, 2011) was especially interesting for our research, because of its easy integration into a GIS environment (using ArcGIS's Network Analyst tools) and because of the results obtained by (Larsen *et al.*, 2013) in the map-matching process applied to the routes collected in Texas, with 88% success.

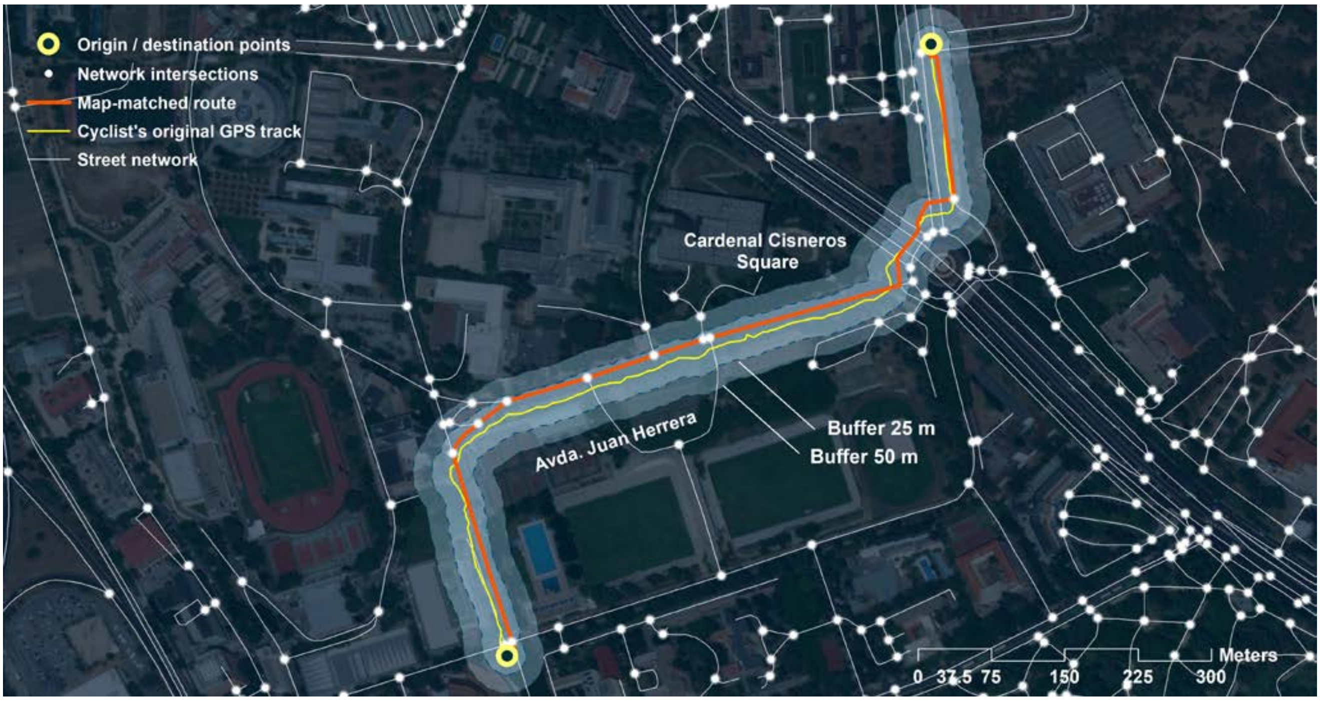


Figure 3. Illustration of the Map-matching process. Route map-matched within the 25-m distance buffer.

For this study, based on the map-matching algorithm created by Dalumpines & Scott (2011), a new version was developed, improving an aspect relevant to the purpose of our research. The procedure basically creates a buffer around the GPS track-line that constrains the estimation of the shortest path between the origin and the destination, by using Dijkstra's algorithm. In this process, the definition of the buffer distances determines the results: a buffer too small may prevent the matching of many routes, while a buffer too wide may lead to inaccurate or incorrect routes. After a sensibility analysis, Dalumpines and Scott concluded that with buffer distances below 25 m, the shortest-path algorithm did not find any routes, and with buffer distances above 60 m, inaccurate routes were the result. Our improvement in this sense was to develop a model that did not fix a specific buffer distance, but rather a dynamic buffer that ranges from a lower and an upper limit that we established from 25 to 250 m. The algorithm starts by attempting to find the route within the buffer that corresponds to the minimum distance and, if it does not succeed, the process is repeated by incrementing the buffer distance by 25 m, till the route is eventually matched. Each route will have a field that informs on the buffer distance at which it was generated, so that we can obtain an idea of how accurate the matched route is. This way, the output routes are as accurate as possible and, in our case, it was possible to map-match almost all the GPS tracks (96%). Of course, some of them are matched with buffer distances over 60 and even 100 m, which certainly could be considered unacceptable, but here comes the following appreciation that we considered relevant. There are GPS routes that are quite accurate and merely contain exceptional errors in some specific segments (probably due to a temporally poor GPS signal). As a result, these routes often cannot be matched considering low buffer distance. However, for the purpose of this study, we did not need entire matched routes, but rather accurate map-matched route segments. Our algorithm allowed us to select a larger amount of route segments perfectly matched out of routes that otherwise could have been discarded, since it informs on the distance between the original GPS track and the route for all the route segments.

3.1.2. *Selection of route segments*

The resulting map-matched routes, network arcs that consist of different network segments, were then split into single-route network-segments, considering all the vertices of the street network. Through a “spatial join” operation, we eventually merged, for any route segment, the data related to all properties of the street network (such as type of street or bike infrastructure, motor traffic speed, etc.) with the information from the original cyclists’ route. A selection of these data-enriched route segments was the basis for our analysis. The selection process is explained next.

In addition to the aforementioned buffer distance value, other criteria were considered in order to establish what could be considered a quality control for the analysis of cycling speeds in each of the route segments: general matched route buffer distance in map-matching process < 150 m, maximum track lines distance to map-matched route (could be considered a relative buffer distance, measuring the actual accuracy of the segment) < 10 m and average speed < 60 kph. (segments with higher speed were mainly errors). Additionally, we considered four criteria that basically select the route segments that are long enough to avoid errors when estimating speed (since the GPS collects points every 2 seconds, the speed estimated in a very short segment —5 seconds— could be easily over-or underestimated, if we considered 4 seconds or 6 seconds respectively): minimum network arc length > 20 meters, minimum time duration of track line ≥ 8 seconds, minimum number of track-points attached ≥ 4 and a ratio between the map-matched route arc length and the one of the original GPS track-line ranging from 0.75 to 1.25.

As result of this filter, the sample obtained for the analysis, including both casual cyclists and bike messenger routes, consisted of 227,284 out of the 361,610 total route segments (62.8%). In terms of the street network considered for the analysis (and taking into account that some of these route segments are overlapped, in other words, they are different records on the same street segment), the number of street-network segments with route information included in our analysis was 35,464 out of the 117,858 total segments (30.0%) existing for the study area.

3.2. *Estimation of operational speeds through Ordinary Least Squares regressions*

The Ordinary Least Square (OLS) regression is one of the most common techniques when exploring the relationship between a certain variable that we want to explain —the dependent variable— and the variables we believe influence this dependent variable —the explanatory variables —. In this section, the factors that may influence cyclists’ speed are identified and the potential explanatory variables are selected accordingly. Then, the different OLS regressions to be performed for diverse purposes or potential applications are defined.

The use of OLS regressions requires several underlying assumptions. One of them is that the sample of observations is random. This means that observations should be independent. The problem is that this is not actually the case with our observations, since a cyclist’s speed along a route segment may be influenced by his/her speed in the previous route segment. Although the average length of route segments is 84.28 m and, in consequence, the influence in some street segments (the longest ones) is probably not remarkable, this influence could be statistically significant for the rest of the segments. If this is the case, the observations would have a temporal autocorrelation or, more specifically, a serial correlation, since speed at one segment would be affected by the speed in the previous segment. This is a particular and simple case of *OLS with Time Series Data*, and applying a simple Finite Distributed Lag (FDL) model would allow us to correct the conventional OLS model (Wooldridge, 2015). FDL models allow us to include one or more variables that affect the dependent variable with a lag. Considering (Y_t) as the dependent variable at the time period (*t*), (β) as the coefficients, (ε) the residuals and (X) the explanatory variables, an example of an equation or an FDL of order one, for a single time period before (*t*) would be: $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \ldots \ \beta_n X_n + \varepsilon$. In consequence, considering the possible existing temporal autocorrelation in our observations, we have adopted an FDL model and have included, as a potential explanatory variable, cyclists’ speed in the segment before the one analysed.

3.2.1. Identification of factors affecting cyclists' speed and selection of explanatory variables

Cyclists' operating speed may be affected and determined by a wide range of factors. Different studies have analysed the impact of some of them, such as slope (Monzón de Cáceres *et al.*, 2008), the existence of segregated bike lines (Wei, Huang and Wang, 1997), the effect of electric assistance in the case of e-bikes Cherry (2007), the influence of age and gender Lin, He, Tan, & He (2008), the impact of signalized intersections and red lights (Guo *et al.*, 2014) or even the weather (Helbich, Böcker and Dijst, 2014). However, as we observed in the introduction section, most of these studies presented some limitations, in terms of either the methodology followed or the sample on which the analysis is based. In addition, some important factors are missing and, in any case, the study of a wide range of factors simultaneously had not yet been carried out. The information collected through the Madrid Cycle Track initiative provided a good opportunity to conduct a more complete analysis including the factors listed on Table 1 as possible explanatory variables. For the different analyses conducted, routes without information on any of these variables were not included.

Table 1: List of possible explanatory variables, description and expected sign

Explanatory Variable		Description	Expected sign
1.	Slope	Slope in percent rise, estimated by calculating the elevation for each node of the GPS route segments from a high resolution Digital Elevation Model (cell size = 5 meters).	Negative
2.	Intersections / km	Calculated as the ratio of number of intersections per route segment (km).	Negative
3.	Traffic lights / km	Considered as the ratio of number of traffic lights per route segment (km).	Negative
4.	Age	Age reported by the volunteers (all of them over 18 years old).	Negative
5.	Female	Gender, reported by the volunteers, introduced as a dummy, considering male by default (because around ¾ of the sample are males)	Negative. Males are expected to cycle faster.
6.	Purpose of the journey	Classified as commuting, sport, leisure, study, shopping and errands.	Diverse signs according to categories.
7.	Type of road regarding bike infrastructure	According to the Madrid Cycling Master Plan classification: Bike lane on the sidewalk, segregated bike lane on the sidewalk, non-segregated bike lane, segregated bike lane in parks/countryside with adapted surface, segregated bike lane in parks/countryside without adapted surface, segregated bike lane in parks or in the countryside with a slightly adapted surface, lane shared with cars (no infrastructure) and "lane with cycling preference and speed reduction".	Diverse signs according to categories.
8.	Total trip time (min)	Total trip time in minutes, obtained from the GPS records.	Negative
9.	Total trip distance (m)	Total trip distance in meters, calculated from the map-matched route length, not from the GPS track-lines, which may lead to under-or overestimations depending on the amount of noise in the GPS point records.	Negative
10.	Net increase in altitude (m)	Net increase in altitude (meters), obtained from the map-matched route difference in origin and destination elevation data.	Negative
11.	Net increase in altitude (m)	Accumulated increase in altitude (meters), calculated from the map-matched route accumulated difference in the route segments elevation.	Negative
12.	Maximum traffic speed (kph)	Maximum traffic speed (kilometres per hour) per street segment, according to TomTom® database.	Positive
13.	Real average traffic speed (kph)	Real average traffic speed (kilometres per hour) per street segment, for the weekdays-mornings time frame according to TomTom® database.	Positive
14.	Type of bike	Only in the case of bike messengers' routes, classified as normal bicycle, cargo bicycle and cargo tricycle	Diverse signs according to categories.
15.	Weather	As in the case of bike messengers' routes (provided by the Garmin Connect platform linked to the messengers' GPS devices), reported as sunny, cloudy or rainy	Negative for cloudy and rainy values.

3.2.2. Definition of the different OLS regressions to perform

Three different OLS regressions with diverse purposes or applications have been defined. In all of them, the categorical values of some variables (gender, purpose of the journey, type of road in terms of bike infrastructure and type of bike and weather) have been introduced as dummies, following the usual methodology for this purpose (Suits, 1957; Hardy, 1993).

The first OLS regression estimates an average cycling speed for each street-network segment according to its properties or conditions, and it is useful for creating general cycling isochrones or accessibility analyses. It considers the following six explanatory variables: slope, intersections, traffic lights, type of road according to bike infrastructure, maximum traffic speed and real average traffic speed. The model does not include any information related to the cyclists or their specific route and, in consequence, it does not consider the potential temporal autocorrelation of cyclists' speeds.

The second model assigns an average cycling speed to each route segment according to the street-network and other properties related to the trip or the cyclist and, considering more information, provides a more accurate speed estimation, which is useful for router apps when estimating travel times. The model is based on the analysis of 15 explanatory variables, the ones introduced in the previous model, as well as age, gender, purpose of the journey, total trip time, total trip distance, net increase altitude, accumulated increase altitude, maximum traffic speed and average traffic speed. As it was explained in the introduction of Section 3.2, cyclists' average speed along one route segment could be influenced by their speed when they arrive at said segment. For this reason, this model also includes, as a new variable, cyclists' speed in the previous route segment (kph), applying a simple Finite Distributed Lag (FDL) as previously described.

Finally, the third model focuses on bike messengers' routes, and estimates an average cycling speed to each route segment according to the street-network and other properties related to the trip, the conditions (weather) or the type of bicycle. As with the previous model, this one provides for more accurate travel times based on more specific information. The model is based on the analysis of 14 explanatory variables, the ones introduced in the first model, as well as type of bicycle, weather conditions and —again— total trip time, total trip distance, net increase altitude, accumulated increase altitude, maximum traffic speed and average traffic speed. In addition, and for the same reasons considered regarding the second model, we have also considered bike messengers' speed in the previous route segment (kph).

Because traffic speed variables were important but not present in all the network segments, the three models have two sub-models, the one applied to the roads with motor traffic and the one applied to streets without motor traffic.

The different OLS regressions performed are based on the analysis of a specific set of cyclists' route arcs (model observations), since they study cyclists' speed according to different variables, and each observation must contain data on all of these variables. This is the reason why more specific samples were selected for the different analyses. The tables in the Results section inform about the number of observations on which each model is eventually based.

3.3. Estimation of cyclists' travel times

As stated in the introduction, the analysis of cyclists' operating speeds is crucial for the study of essential aspects of cycling mobility. One of these aspects is the estimation of cyclists' travel times, which at the same time will be the base for subsequent analyses of accessibility and competitiveness. The previous models allow us to predict cyclists' average speed according to different variables at the level of street arc. In this section, we applied the models in order to estimate cyclists' travel times for an entire route (usually made of hundreds of street-network segments), and we calculated the correlation between real and estimated travel times by the first model (which considers the street-network properties) and the second one (which, in addition, includes other variables related to the trip or the cyclist). It is important to highlight that the models were applied to a number of routes that were not considered when performing the models (control routes). The number of control routes was defined as the 10% of the sample considered when performing the OLS models that correspond to casual cyclists (first and second models). This sample consisted of 2,290 routes, so we applied the first and the second models to 229 control routes.

3.4. Estimation of cyclists' accessibility and comparison to other transport modes

As an application case for the obtained results, we decided to perform a comparative analysis of accessibility, in order to find out whether cycling is not only a sustainable transport mode, but also a competitive one. With this aim, the areas covered by the isochrones that correspond to a range of travel times (from 5 to 25 minutes, every 5 minutes) were calculated for the following transport modes: cycling, walking, private car and public transport (using an intermodal network including bus, underground, tram and train). In order to produce comparable results, the isochrones were calculated considering the same street network, applying the same tool —the *Service Area* tool contained in the *Network Analyst* extension of *ArcGIS*— and estimating, for all the transport modes, the average travel times for a weekday morning. With this tool, generated service areas and isochrones are not determined by the length of street network segments. The tool not only includes the entire street segments covered in the corresponding travel time but is able to interpolate the proportional part of the street segments to be included. Therefore, it estimates services areas and isochrones accurately. The specific methodologies followed when estimating the travel times for each transport mode are described next.

Firstly, cycling isochrones are calculated by considering the estimated cycling speed (and travel time) for each street-network segment estimated by the first OLS regression, which is considering the variables related to the properties or conditions of the street network (previously listed). When defining the *Service Area Analysis* settings, we allowed cycling in both directions along all cyclable streets, which is possible according to current legislation.

Secondly, to estimate walking isochrones, we looked at existing research studies focussed on determining an average walking speed. Given the fact that walking speed values are significantly different in different cities (Bettencourt *et al.*, 2007; Guo, Sayed and Zaki, 2016), we decided to adopt the average walking speed for the city of Madrid considered by (Ortega *et al.*, 2015) in their study on urban fragmentation of the Chamberí district in Madrid, which is 4kph. When defining the *ArcGIS Service Area Analysis* settings, walking in both directions along all walkable streets was allowed.

Public Transport isochrones were estimated according to the travel times and frequencies considered in the *General Transit Feed Specification* (GTFS) database, obtained from the Regional Transport Consortium of Madrid. By applying the *ArcGIS* tool *Add GTFS to a Network Dataset*, we were able to consider bus, underground, tram and train travel times for a working day at 8:00 am, and run the *Service Area* tool, as we did for the rest of transport modes.

Finally, car isochrones were calculated according to a "door-to-door" approach similar to the one considered by Salonen & Toivonen (2013). This approach estimates not only the travel time spent in the car, but also the average walking time that it usually takes people to arrive to the car, the average time spent looking for a parking place and then, the average walking time from the parking place to the final destination. Car speeds, travel times, as well as street directions, were taken from the *TomTom*® dataset (*TomTom*® Speed profiles), selecting average weekday morning speed values at 8h, which provides an accurate approach coming from the historical records. Although the average walking distances considered are difficult to estimate since they may greatly vary for each city, for this study, we have considered the 180 distance per walk established by (Kurri and Laakso, 2002), which, at a pace of 4.5 kph, leads to a 288-second travel time considering both walks. When it comes to the time spent looking for parking, we considered the average estimated in the case of Madrid by an study conducted in the city (Europa Press, 2006): 6.28 minutes per trip (376.8 sec/trip). In summary, for the global car travel time estimations we added 664 seconds per trip to the car travel times previously estimated. This is why the isochrones that correspond to 5 and 10 minutes (300 and 600 seconds, respectively) will be null under this consideration.

4. Results

4.1. OLS regressions results

4.1.1. First OLS regression results: Cyclists' speed according to street-network properties and conditions

The first OLS regression estimates an average cycling speed for each street segment according to the network properties. The results are shown in Table 2. Regarding the sub-model 1.1 (applied to roads with motor traffic), the R-squared value obtained is 0.431. All the explanatory variables have the expected coefficient signs (according to what was previously hypothesized in Table 1) and with respect to the elasticities between these variables and cyclist's speed, the model offers relevant information. The variables with highest (negative) impact on speed according to the *standardised coefficients* (StdCoef) obtained, are *Street intersections/km* and *Slope*. In addition, the model is very sensitive to the *Real Average Traffic Speed*, so cyclists seem to feel forced to speed up when riding next to cars circulating faster than on other streets (for example, an increase of 20kph in traffic speed leads to an increase of 3kph in cyclists' speed on average). Other variables have a significant impact, such as *Traffic Lights/km,* or different bike infrastructure that increase cyclists' speed (such as *Segregated* or *Non-segregated bike-lanes*) or reduce it (such as Bike-lane on sidewalks).

The R-squared value obtained might be considered moderate, but some general remarks on the OLS regressions results when predicting speeds at specific street segments must be taken into account. The dependent variable that we are modelling is cyclists' speed at a specific route segment, and the R-squared 'modest' values obtained can be explained by factors of a different nature. First, there are aspects related to changing conditions on the street segments over time that dramatically affect cyclists' speed. Speed at a specific road segment will be completely different whether cyclists arrive to a green traffic light or to a red one, or whether they find traffic or not at non-signalized intersections. Then, there are other factors related to the particular conditions of the cyclists, personal variables related to physical condition or health, as well as to the type of bike they are riding (road bikes, foldable bikes, mountain bikes, etc.). In actuality, similar OLS R-squared results were obtained (R^2=0.414) when analysing walking speeds (Bohannon 1997), especially when considering the potential variables with impact on such a complex functional activity (Fritz & Lusardi 2009), which, therefore, is less predictable than other transport modes. In any case, the first type of these factors affects the estimation of cyclists' speed at a particular route segment, but not when considering an entire route, which is usually made of hundreds of segments. At the end of the trip, the number of red or green traffic lights found will probably be balanced, as well as the number of intersections with or without traffic, so when analysing the estimated vs. real trip travel times (next section), the global performance of the model will be much better.

The analysis of Variance Inflation Factor (VIF) values of the independent variables revealed that there is no multicollinearity between them. Koenker BP tests revealed the existence of heteroscedasticity (see Koenker BP test values in the corresponding OLS result tables), and the analysis of the scatterplot graphs of standardised residuals in relation to predicted dependent variable values showed that this heteroscedasticity is essentially due to the existence of greater errors when predicting highest speed values. This was expected, given the problem of predicting cyclists' speeds along route segments with traffic lights. Street segments that, according to the values of different explanatory variables (for instance, with negative slope values), should bear high cyclist speeds, can present low values when cyclists find red traffic lights along their way. Since heteroscedasticity was found, we relied on Robust Probabilities values (Rousseeuw and Leroy, 2005) to determine coefficient significance in the models, overcoming this limitation. Robust Probabilities values are shown in the corresponding OLS result tables.

Spatial autocorrelation of errors was revealed through Moran's Index. However, the analysis of the residuals on a map showed that many of them were concentrated around intersections with traffic lights, for the reasons previously discussed, and especially on streets with high slope values. The use of Geographically Weighted Regressions, frequently applied to improve the results of OLS regressions with spatial autocorrelation of errors, did not improve the results. This fact can be explained because, in street segments with a significant slope, residuals of different sign can be found when predicting speed for cyclists going up and down at the same location.

Regarding the sub-model 1.2 (applied to roads without motor traffic), the R-squared value obtained is 0.380. Again, all the explanatory variables have the expected coefficient signs (Table 1) and with regard to the elasticities, the model offers relevant information and some variations in comparison with the previous submodel. The variable with the highest impact on speed, again according to the *standardised coefficients* (StdCoef) obtained, is *Slope* rather than *Street intersections/km*, what makes sense considering that probably the presence of traffic is what has the greatest impact when arriving to intersections. This is the reason why the impact of *Traffic lights/km* is also notably reduced. The impact of slope is relatively lower, probably because traffic pushes cyclists to go faster, also on steep streets. In terms of infrastructure, some remarkable differences were found. Cyclists increase their speed significantly in the different kinds of bike lanes, and even bike lanes on sidewalks, with a negative impact on traffic roads, have a positive impact here. Since, as in the previous submodel, Koenker BP tests revealed the existence of heteroscedasticity, we relied on Robust Probabilities values to determine coefficient significance in the models. Spatial autocorrelation of errors was revealed through Moran's Index once again, for the reasons previously described. The analysis of the VIF values of the independent variables proved that there is no multicollinearity between them.

4.1.2. Second OLS regression results: Cyclists' operating speed according to street-network properties and conditions and other aspects related to the trip or the cyclist

The second model assigns an average cycling speed to each route segment according to the street-network and other properties related to the trip or the cyclist and, considering more information, provides a more accurate speed estimation (R-squared are higher than obtained in the previous OLS regressions). The results are shown in Table 3.

The sub-model 2.1 (applied to roads with motor traffic) showed an R-squared value obtained of 0.533. All the explanatory variables have the expected coefficient signs (Table 1) and, regarding the elasticities between these variables and cyclist's speed, the model provides important information. The model shows the significant different cycling speeds according to gender, females' speed being 1.75 kph lower that males'. According to the *standardised coefficients* (StdCoef) obtained, cyclists' speed in previous route segment shows a high elasticity. Although age has a negative coefficient, its impact is not so important. The purpose of the journey is also a key variable, with some particular purposes having a significant impact compared to the one considered by default in the models (working). Cyclists' speed when traveling for shopping, leisure or errands is lower and higher when sport is the purpose. The *Journey total duration* and the *Journey total elevation gain* have also a negative impact. The influence of the variables related to the street-network properties are similar to those obtained in the first model, with *Slope*, *Street Intersections/km*, *Traffic Lights/km* and *Real Average Traffic Speed* having the greatest impact, according to the *standardised coefficients* obtained.

With respect to sub-model 2.2 (applied to roads without motor traffic), the R-squared value obtained is 0.539. Again, all the explanatory variables have the expected coefficient signs, and regarding elasticities, the model reveals some interesting changes. The impact of gender is reduced significantly, so the presence of traffic affects females more than males. The influence of the different purpose of the journey are also reduced and even not any more significant in some cases (sport). The variables related to the street-network properties affects in a similar way as for the 1.2 sub-model.

In both sub-models, for the reasons already argued, spatial autocorrelation of errors was revealed through Moran's Index, and Koenker BP tests revealed the existence of heteroscedasticity. In consequence, we relied on Robust Probabilities values to determine coefficient significance of the models. The analysis of the VIF values of the independent variables revealed that there is no multicollinearity between them.

4.1.3. *Third OLS regression results: Bike messengers' operating speed according to street-network properties and conditions and other aspects*

The third model focuses on bike messengers' routes, and estimates an average cycling speed to each route segment according to the street-network and other properties related to the trip, the conditions (weather) or the type of bicycle. The data show that bike messengers' average speed is significantly higher than regular cyclists. The specific results are shown in Table 4 and analysed below.

The sub-model 3.1 (applied to roads with motor traffic) showed a R-squared value of 0.445. All the explanatory variables have the expected coefficient signs (Table 1), although the analysis of elasticities reveal some differences in the impact of some variables in relation to casual cyclists' results. According to the obtained standardized coefficients (StdCoef), cyclists' speed in the previous route segment shows high elasticity, as well as in the previous model. Although the factors that affect most messengers' speed are *Slope*, *Street Intersections/km* and *Real Average Traffic Speed*, the impact of *Traffic Lights/km* is significantly reduced in the case of bike messengers, compared to the values obtained for casual cyclists. Regarding the type of bike, there is no difference between casual bikes and bullit-bikes (which are electric-assisted), but the average speed of cargo-trikes is dramatically lower. When it comes to the weather, average speeds decrease on cloudy days, and especially on rainy days.

Regarding sub-model 3.2 (applied to roads without motor traffic), the R-squared value obtained was 0.565. All the explanatory variables have the expected coefficient signs (Table 1) and, with respect to elasticities, there are some differences compared to the previous model. The influence of bad weather is significantly lower and, in any case, not statistically significant (so what truly affects bike messengers is the combination of bad weather and motor traffic), the impact of intersections decreases by 50%, although the impact of traffic lights increases significantly, probably due to the fact that, even if messengers' speed were similar to the previous one at intersections, the reduction is higher simply because the average circulating speed is also higher. When it comes to the type of bike, bullit-bikes' average speed is clearly higher here. No data of cargo-trikes is available for the selection of routes that correspond to the selection of the variables of this model. The effect of cloudy and rainy days is slightly reduced.

The analysis of the VIF values of the independent variables revealed that there is no multicollinearity between them. However, and once again, both sub-models showed spatial autocorrelation of errors and heteroscedasticity. In consequence, we relied on Robust Probabilities values when determining coefficient significance of the models.

Table 2. First OLS regression results: Cyclists' speed according to street-network properties and conditions. (See Annex)

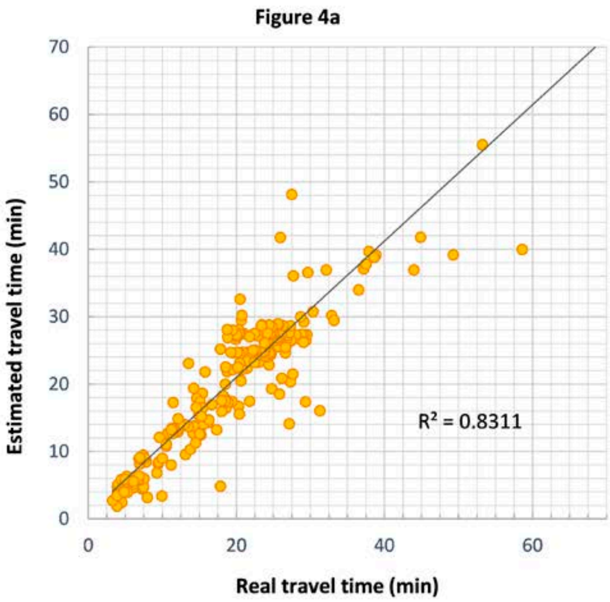
Table 3: Second OLS regression results: Cyclists' operating speed according to street-network properties and conditions and other aspects related to the trip or the cyclist. (See Annex)

Table 4: Third OLS regression results: Bike messengers' operating speed according to street-network properties and conditions and other aspects. (See Annex)

4.2. *Global performance of the model: Estimated vs. real travel times for an entire route*

Although the previous models showed moderate R-squared values when predicting cyclists' speed at a particular route arc (for the reasons already mentioned), the performance of these models when predicting cyclists' travel times for an entire route, which is usually made of hundreds of segments, is much better.

Figure 4a illustrates the correlation between real travel times and the time estimated by the first model (R^2=0.831), which only considers the street-network properties (estimated by applying the coefficients obtained from the first OLS regression), and Figure 4b shows the correlation obtained by the second model (R^2=0.887), which, in addition, includes others variables related to the trip or the cyclist (estimated by applying the coefficients obtained from the second OLS regression). Both correlations are obtained for a sample of 229 control routes (routes that were not considered when performing the models). The results obtained by the first model show how accurate the estimations of travel times may be when merely considering the street-network properties. The second model improves the estimation but, given the additional number of variables that it considers in comparison with the previous one, it also shows a clear limitation when predicting cyclists' operating speeds. This is not an unexpected result, since it seems reasonable to expect slightly different travel times even for the same trip completed by two different persons with similar characteristics and the same purpose of the journey.



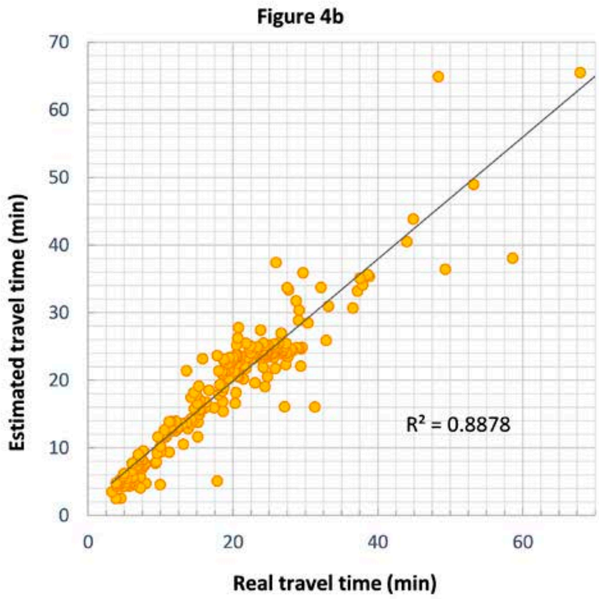


Figure 4: correlation between observed travel times and the time estimated by the first model (4a), which only considers the street-network properties, and the second model (4b), which, in addition, includes the others variables.

4.3. Application case: Estimated cyclists' accessibility and comparison to other transport modes

In this section, a comparative analysis of accessibility and competitiveness between different transport modes is performed, by calculating the isochrones that correspond to a range of travel times (from 5 to 25 minutes, every 5 minutes), for all the existing transport modes in Madrid: cycling, walking, private car and public transport (using an intermodal network, including bus, underground, tram and train). Cycling travel times were estimated by applying the first OLS model (according to street-network properties and conditions). Table 5 shows the area that corresponds to each isochrone (the urban area covered by each transport mode, according to the defined range of travel times), Figure 5 illustrates these relationships and Figure 6 illustrates the isochrones on four different maps, at the same scale. In the case of the estimation of car isochrones, due to the consideration of the "door-to-door" approach explained in Section 3.4, the areas that correspond to 5 and 10 minutes are null.

The results reveal that cycling is not only a sustainable transport mode, but is also the most competitive for small-medium distances, with Graph 2 clearly illustrating the better performance of cycling (in terms of the area covered) for trips under 21 minutes in length for the centre of Madrid. For trips over 21 minutes in length, the car begins to be more competitive, but only if we do not consider areas where private cars are forbidden, which is increasingly common in the central areas of many cities. For trips over 23 minutes in length, Public Transport begins to be more competitive, as well. In any case, long-distance trips often involve different transport modes, so the results should raise awareness about the suitability of promoting cycling as part of these multimodal trips.

Table 5: Area covered (Ha) according to travel time and transport mode

Journey duration (minutes)	5	10	15	20	25
Walking Area Covered	22.17	97.78	224.80	403.34	631.23
Cycling Area Covered	343.38	1,283.03	3,048.72	5,396.81	8,516.78
Car Area Covered	0.00	0.00	585.75	4,229.03	15,161.81
Public Transport Area Covered	106.84	741.97	2,061.75	4,635.63	9,017.16

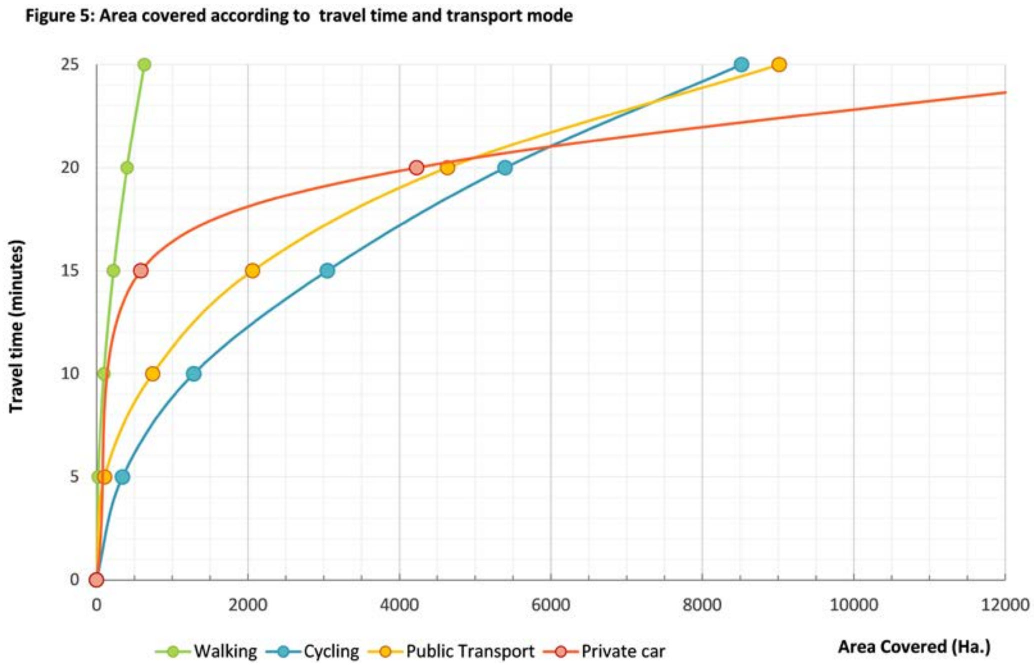


Figure 5: Area covered (Ha) according to travel time and transport mode

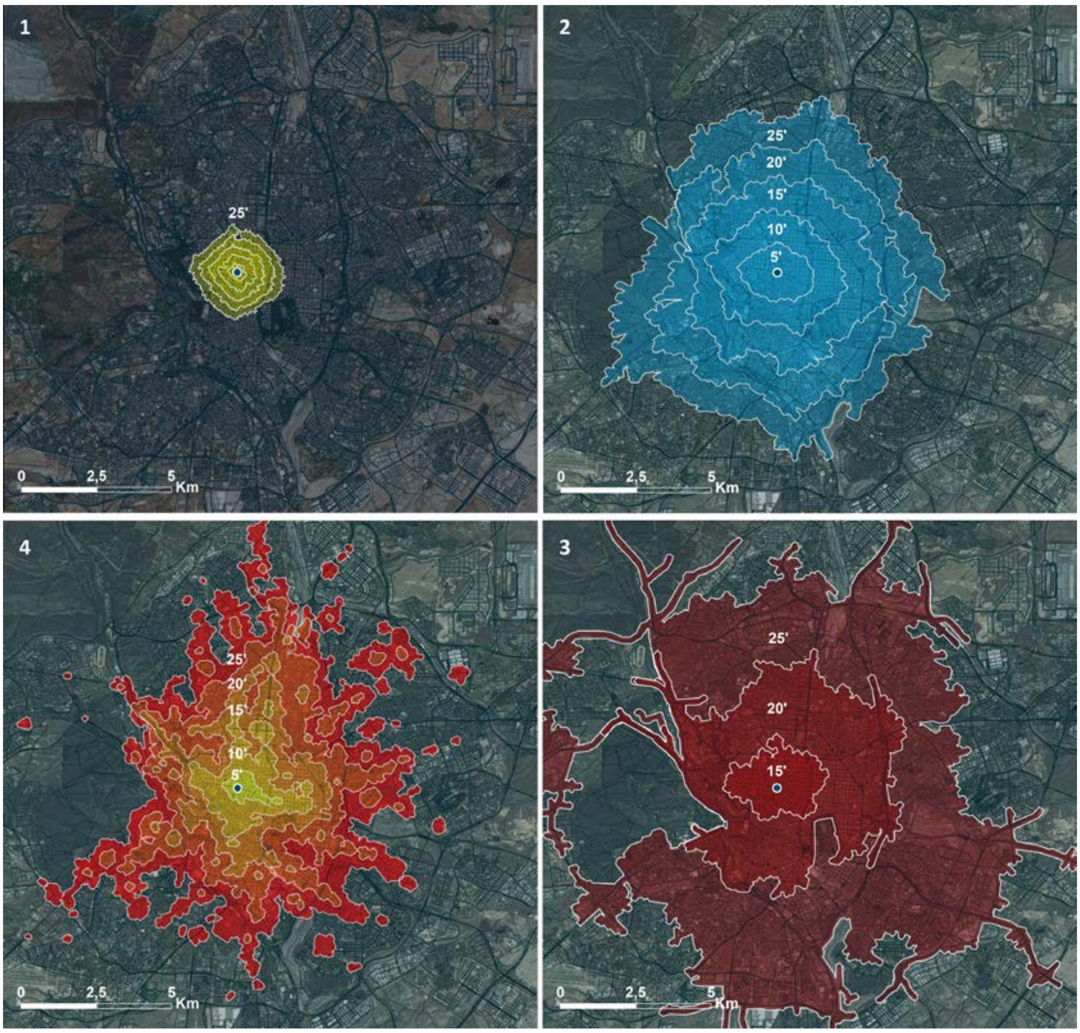


Figure 6: Isochrones according to transport mode: Walking (1), Cycling (2), Public Transport (3) and Private Car (4).

5. Conclusions

The famous quote by the American cyclist John Howard (1987), "*The bicycle is a curious vehicle. Its passenger is its engine,*" highlights an important fact for this study. Cycling mobility is extraordinarily complex because these human engines are affected by a wide range of factors that can have a different impact on each individual, since humans may behave differently under similar circumstances. This paper studies cyclists' operating speeds embracing this complexity, by analysing the impact of a wide variety of variables at the same time.

The results obtained support Howards' declaration, and present cycling as a particular transport mode, sensitive to many elements. The OLS models shed light on the influence of a wide range of factors on cyclists' speed, quantifying the specific impact of a wide range of variables of a different nature: from the street-network properties to other properties related to the trip (such as the purpose of the journey) or the cyclist (such as gender or age). Although the different models show modest R-squared values when predicting cyclists' speed at a particular route segment (for the reasons previously discussed), the models performed quite well when predicting cyclists' travel times for an entire route (usually made of hundreds of segments), showing a high correlation between estimated and real travel times for both the first model (R^2=0.831), which considers the street-network properties, and the second model (R^2=0.887), which, in addition, includes other variables related to the trip or the cyclist. Although this second model improves the travel time estimations, it also shows certain limitations when predicting cyclists' operating speeds, an expected result, since it seems reasonable to obtain slightly different travel times for the same trip completed by two different cyclists with similar characteristics and the same purpose for the journey.

Another relevant output is that the models also allow us to estimate cyclists' travel times for the entire street network and not only in the street-network arcs where we have cyclists' records (just 30% of the total) and, furthermore, they would also allow us to predict casual cyclists' as well as bike messengers' travel times and accessibility in future scenarios, given certain changes in the network, such as the execution of new infrastructure or implementation policies, such as slowing down traffic speed. In consequence, considering growing general cycling mobility and the particular emergence of bike messenger companies, the models can be considered as tools that may help decision makers when evaluating future scenarios.

Finally, this accurate estimation of cyclists' travel times also allows us to conduct the comparative analysis of accessibility, and evaluate competitiveness between different transport modes. The results show that cycling is the most competitive transport mode for small-medium distances (under 21 minutes in length for Madrid), which is a relevant finding not only for casual cycling mobility, but also for bike-driven parcel delivery service, since bike-messenger performance is even greater than casual cyclists'. Therefore, a consequent goal of this paper has become to raise awareness of cycling, not only an environmentally-sustainable transport mode, but also as the most efficient mode for an important range of distances.

Future research should cover one of the most important limitations of this study; it should expand the cyclists' routes sample and include the analysis of the GPS tracks coming from the Madrid Bike share system, whose electric bikes are being used by an increasing number of people (Romanillos *et al.*, 2018). E-bikes used by casual cyclists should be also included and analysed. In this case, it is expected that the different factors studied in this research affect cycling speeds and travel times in a different way (for instance, the impact of slope), so a comparative analysis would allow us to measure and model these differences. Another important limitation to highlight is that this study focuses on one single city. The impact of the variables analysed will probably be different in other cities. Future research would have to explore other cities and compare the results, so that stronger and more consistent findings will arise.

6. Annex. Tables.

Table 2. First OLS regression results: Cyclists' speed according to street-network properties and conditions

	Sub model 1.1 (roads with motor traffic)						Sub model 1.2 (roads without motor traffic)					
Independent variables	Coefficient	Std. Error	Robust_t	Robust_Prob	StdCoef	VIF	Coefficient	Std. Error	Robust_t	Robust_Prob	StdCoef	VIF
Intercept	13.896526	0.143329	99.5122	0,000000*	0.0000	-	14.946825	0.288699	49.6819	0,000000*	0.0000	-
Slope (percent rise)	-0.614379	0.011783	-48.8691	0,000000*	-0.3372	1.0392	-0.727377	0.024195	-27.9843	0,000000*	-0.4156	1.0254
Street Intersections / km	-0.110078	2.302086	-50.3755	0,000000*	-0.3329	1.2042	-0.058972	3.768646	-15.5466	0,000000*	-0.2333	1.1931
Real Average Traffic Speed (kph)	0.149815	0.003918	35.8919	0,000000*	0.2972	1.5016	-	-	-	-	-	-
Traffic Lights / km	-0.038570	3.241606	-11.9528	0,000000*	-0.0838	1.2327	-0.084582	6.742596	-12.2231	0,000000*	-0.1808	1.1146
Max. Traffic Speed (kph)	0.023694	0.003173	7.1872	0,000000*	0.0602	1.6161	-	-	-	-	-	-
Bike lane on the sidewalk *	-0.764160	0.111472	-7.3131	0,000000*	-0.0457	1.1032	1.413114	0.273998	4.96793	0,000001*	0.1719	5.9597
Non-segregated bike lane	1.087164	0.178602	6.4770	0,000000*	0.0389	1.0137	6.118493	0.918732	6.5874	0,000000*	0.0949	1.0899
Segregated bike lane in parks with a minimum adapted surface	2.577820	0.784055	2.0559	0,039795*	0.0209	1.0069	2.653539	0.283496	8.7532	0,000000*	0.2841	4.9449
Segregated bike lane in parks without adapted surface	-	-	-	-	-	-	2.232850	0.269059	7.8334	0,000000*	0.2831	6.2434
Segregated bike lane in parks with adapted surface.	-	-	-	-	-	-	4.766114	1.778681	2.4405	0,014707*	0.0370	1.0241
Adjusted R-Squared:	0.430794						0.38055					
Number of explanatory variables:	8						8					
Number of Observations:	14,144						3,325					
Joint F-Statistic: 1071.38	Prob(>chi-squared), (10) degrees of freedom: 0.000000*						Joint F-Statistic: 256.25	Prob(>chi-squared), (8) degrees of freedom: 0.000000*				
Koenker (BP) Statistic: 536.13	Prob(>chi-squared), (10) degrees of freedom: 0.000000*						Koenker (BP) Statistic: 94.90	Prob(>chi-squared), (8) degrees of freedom: 0.000000*				
Jarque-Bera Statistic: 1.24	Prob(>chi-squared), (2) degrees of freedom: 0,536509						Jarque-Bera Statistic: 29.78	Prob(>chi-squared), (2) degrees of freedom: 0.000000*				

* An asterisk next to a number indicates a statistically significant p-value ($p < 0{,}01$).

** Dummies by default: Working (Purpose-journey), Male (Gender), Normal bike (Type of bike), Sunny (Weather), No infrastructure (Bike inf.).

*** These models do not include any information related to the cyclists or their specific route and, in consequence, they do not consider the potential temporal autocorrelation of cyclists' speeds.

Table 3: Second OLS regression results: Cyclists' operating speed according to street-network properties and conditions and other aspects related to the trip or the cyclist

	Sub model 2.1 (roads with motor traffic)						Sub model 2.2 (roads without motor traffic)					
Independent variables	Coefficient	Std. Error	Robust_t	Robust_Prob	StdCoef	VIF	Coefficient	Std. Error	Robust_t	Robust_Prob	StdCoef	VIF
Intercept	9,707603	0,319172	29,4165	0.000000*	0,0000	-	9,553201	0,428530	17,9988	0.000000*	0,0000	-
Cyclists' speed in previous route segment	0,356995	0,007408	35,4937	0.000000*	0,4006	1,2874	0,455115	0,012741	27,5117	0.000000*	0,4787	1,1979
Slope (percent rise)	-0,510005	0,014269	-33,2682	0.000000*	-0,2787	1,1326	-0,526791	0,024218	-18,1907	0.000000*	-0,2802	1,1069
Real Average Traffic Speed (kph)	0,108916	0,004717	22,0976	0.000000*	0,2158	1,6263	-	-	-	-	-	-
Street Intersections / km	-0,050254	0,002664	-18,9623	0.000000*	-0,1500	1,1773	-0,034506	0,003579	-8,9267	0.000000*	-0,1284	1,1831
Traffic Lights / km	-0,050699	0,004204	-11,4088	0.000000*	-0,0998	1,2743	-0,083566	0,007261	-9,3096	0.000000*	-0,1479	1,1012
Female **	-1,755088	0,161629	-12,2507	0.000000*	-0,0874	- 1,2065	-1,244599	0,198741	-6,6378	0.000000*	-0,0818	1,1372
Leisure (Purpose of the journey)	-0,759388	0,104977	-7,1589	0.000000*	-0,0578	1,1879	-0,314167	0,146116	-2,1241	0,0337240	-0,0300	1,2996
Shopping (Purpose of the journey) **	-1,096129	0,159426	-7,1166	0.000000*	-0,0536	1,1303	-0,414898	0,155240	-2,7688	0,0056610	-0,0410	1,5693
Journey total duration (minutes)	-0,016023	0,002807	-5,8417	0.000000*	-0,0439	1,1006	-0,003529	0,002038	-1,4214	0,1553090	-0,0231	1,1818
Segregated bike lane in parks or countryside with adapted surface	4,938903	0,854686	5,4070	0.000000*	0,0426	1,0104	1,680420	0,283610	5,0366	0,0000010	0,1627	5,0277
Bike lane on the sidewalk **	-0,396956	0,130926	-3,4100	0,0006690	-0,0251	1,2792	1,178279	0,265099	3,6928	0,0002390	0,1305	5,7518
Study (Purpose of the journey)	-0,327553	0,136383	-2,4540	0,0141320	-0,0208	1,4019	-0,408424	0,210205	-1,8407	0,0657580	-0,0296	1,5516
Errands (Purpose of the journey)	-0,531909	0,180032	-2,8273	0,0047090	-0,0224	1,0740	-0,718439	0,284425	-2,9818	0,0028990	-0,0324	1,1000
Journey total Elevation gain	-0,001268	0,000547	-2,5111	0,0120410	-0,0176	1,0723	-0,000688	0,000516	-1,3572	0,1748430	-0,0168	1,0623
Non-segregated bike lane	0,431743	0,201437	2,2504	0,0244330	0,0160	1,0324	3,932708	0,917159	3,7636	0,0001820	0,0556	1,1231
Age	-0,010535	0,006951	-1,4840	0,1378480	-0,0138	1,5515	-0,030756	0,008358	-3,3729	0,0007700	-0,0553	1,5039
Segregated bike lane in parks or countryside without adapted surface	0,231522	0,243802	1,0048	0,3150150	0,0070	1,2792	1,288166	0,265763	3,9876	0,0000760	0,1505	6,4271
Sport (Purpose of the journey)	0,123793	0,167499	0,7406	0,4589670	0,0063	1,3616	0,148452	0,254433	0,6784	0,4975750	0,0076	1,1323
Max. Traffic Speed (kph)	0,001223	0,004083	0,2956	0,7675450	0,0032	2,0611	-	-	-	-	-	-
Lane with cycling preference and speed reduction	0,065944	1,276419	0,0606	0,9516300	0,0004	1,0027	-	-	-	-	-	-
Adjusted R-Squared:	0,5328						0,5393					
Number of explanatory variables:	20						17					
Number of observations:	8.702						3.074					
Joint F-Statistic: 476.29	Prob(>chi-squared), (20) degrees of freedom: 0.000000*						Joint F-Statistic: 212.59	Prob(>chi-squared), (17) degrees of freedom: 0.000000*				
Koenker (BP) Statistic: 563.04	Prob(>chi-squared), (20) degrees of freedom: 0.000000*						Koenker (BP) Statistic: 130.68	Prob(>chi-squared), (17) degrees of freedom: 0.000000*				
Jarque-Bera Statistic: 1513.92	Prob(>chi-squared), (2) degrees of freedom: 0.000000*						Jarque-Bera Statistic: 1773.16	Prob(>chi-squared), (2) degrees of freedom: 0.000000*				

* An asterisk next to a number indicates a statistically significant p-value (p < 0,01).

** Dummies by default: Working (Purpose-journey), Male (Gender), Normal bike (Type of bike), Sunny (Weather), No infrastructure (Bike inf.).

Table 4: Third OLS regression results: Bike messengers' operating speed according to street-network properties and conditions and other aspects

	Sub model 3.1 (roads with motor traffic)						Sub model 3.2 (roads without motor traffic)					
Independent variables	Coefficient	Std. Error	Robust_t	Robust_Prob	StdCoef	VIF	Coefficient	Std. Error	Robust_t	Robust_Prob	StdCoef	VIF
Intercept	11.387135	0.888138	10.1651	0.000000*	0.0000	-	15.027922	1.197565	11.1967	0.000000*	0.0000	-
Cyclists' speed in previous route segment (kph)	0.299205	0.017566	6.6461	0.000000*	0.3472	1.1588	0.411496	0.028968	12.8440	0.000000*	0.4520	1.2068
Slope (percent rise)	-0.634642	0.051655	-10.7053	0.000000*	-0.2424	1.0849	-0.669071	0.091649	-5.2918	0.000000*	-0.2234	1.1159
Street Intersections / km	-0.135268	0.019391	-6.9270	0.000000*	-0.1985	2.2584	-0.148271	0.025887	-5.3570	0.000000*	-0.2546	2.3554
Real Average Traffic Speed	0.121027	0.015397	6.7324	0.000000*	0.1899	1.6273	-	-	-	-	-	-
Cargo Trike (Type of bike) **	-4.482614	0.819780	-5.4293	0.000000*	-0.1066	1.0603	-	-	-	-	-	-
Journey accumulated elevation gain	-0.007725	0.001597	-5.0836	0.0000010	-0.0997	1.1846	-0.011104	0.002453	-4.4425	0.0000140	-0.1862	2.0166
Journey total distance (m)	-0.009856	0.003039	-2.6651	0.0077740	-0.0933	2.3074	-0.000290	0.002055	-0.1809	0.8564850	-0.0059	2.0941
Cloudy (weather conditions) **	-0.967724	0.280017	-3.4525	0.0005860	-0.0693	1.1196	-0.214245	0.295624	-0.7068	0.4800420	-0.0225	1.1490
Journey total duration (minutes)	-0.003042	0.001082	-2.0649	0.0390930	-0.0537	1.0152	-0.046832	0.006873	-7.5456	0.000000*	-0.2734	1.9187
Rain (weather conditions) **	-1.674932	0.883352	-1.9915	0.0465980	-0.0366	1.0401	-0.853137	1.375832	-0.7058	0.4806530	-0.0185	1.0614
Bullit Bike (Type of bike)**	1.044746	0.589898	1.7088	0.0877060	0.0343	1.0431	4.998283	1.180491	4.7901	0.0000030	0.1280	1.0897
Bike lane on the sidewalk**	-0.996514	0.755032	-1.3259	0.1850820	-0.0257	1.0573	-0.225321	0.856487	-0.2237	0.8230670	-0.0139	3.3457
Non-segregated bike lane**	2.425579	2.136889	1.1974	0.2313270	0.0216	1.0073	-	-	-	-	-	-
Max. Traffic Speed	-0.005507	0.013617	-0.4003	0.6889980	-0.0093	1.4781	-	-	-	-	-	-
No infrastructure but cycling preference and speed reduction **	-2.07011	5.219007	-5.0989	0.0000010	-0.0075	1.0046	-	-	-	-	-	-
Traffic Lights / km	-1.888438	19.153613	-0.0914	0.9271940	-0.0020	1.2032	-0.135903	0.045840	-2.3285	0.0202650	-0.1153	1.8021
Segregated bike lane in parks without adapted surface**	-	-	-	-	-	-	-0.538950	0.774797	-0.6006	0.5484030	-0.0418	4.3144
Segregated bike lane in parks with a minimum adapted surface**	-	-	-	-	-	-	0.885858	1.154429	0.7296	0.4659550	0.0296	1.7691
Adjusted R-Squared:	0.445163						0.564652					
Number of explanatory variables:	16						13					
Number of Observations:	1,548						520					
Joint F-Statistic: 78.57	Prob(>chi-squared), (16) degrees of freedom: 0.000000*						Joint F-Statistic: 52.78	Prob(>chi-squared), (13) degrees of freedom: 0.00000*				
Koenker (BP) Statistic: 221.44	Prob(>chi-squared), (16) degrees of freedom: 0.000000*						Koenker (BP) Statistic: 33.08	Prob(>chi-squared), (13) degrees of freedom: 0.00000*				
Jarque-Bera Statistic: 69.21	Prob(>chi-squared), (2) degrees of freedom: 0.000000*						Jarque-Bera Statistic: 134.20	Prob(>chi-squared), (2) degrees of freedom: 0.000000*				

* An asterisk next to a number indicates a statistically significant p-value (p < 0,01).

** Dummies by default: Working (Purpose-journey), Male (Gender), Normal bike (Type of bike), Sunny (Weather), No infrastructure (Bike inf.).

7. References

Bettencourt, L. M. a *et al.* (2007) 'Growth, innovation, scaling, and the pace of life in cities.', *Proceedings of the National Academy of Sciences of the United States of America*, 104(17), pp. 7301–7306. doi: 10.1073/pnas.0610172104.

Börjesson, M. and Eliasson, J. (2012) *The Benefits of Cycling: Viewing Cyclists as Travellers rather than Non-motorists*, *Cycling and Sustainability*. Emerald Group Publishing Ltd. doi: 10.1108/S2044-9941(2012)0000001012.

Cherry, C. R. (2007) Electric two-wheelers in China: analysis of environmental, safety, and mobility impacts. Citeseer.

Dalumpines, R. and Scott, D. M. (2011) 'GIS-based map-matching: development and demonstration of a postprocessing map-matching algorithm for transportation research', *Advancing geoinformation science for a changing world*. Springer, pp. 101–120.

DOYMO (2011) Estudio de seguimiento de impacto de la implantación del PDMC: Eje Prado - Recoletos - Castellana.

Ellison, R. and Greaves, S. (2011) 'Travel time competitiveness of cycling in Sydney, Australia', *Transportation Research Record: Journal of the Transportation Research Board*. Transportation Research Board of the National Academies, (2247), pp. 99–108.

Europa Press (2006) 'Los madrileños pierden cada día más de un cuarto de hora buscando aparcamiento', *El Mundo*, 22 May. Available at: http://www.elmundo.es/elmundo/2006/05/22/madrid/1148313020.html.

Fleming, S. (2012) 'Cycle Space: Architecture & Urban Design in the Age of the Bicycle'.

Guo, Y. *et al.* (2014) 'Red Light Running Behavior of Electric Bicycles at Signalized Intersections in China', *Transportation Research Record: Journal of the Transportation Research Board*, 2468, pp. 28–37. doi: 10.3141/2468-04.

Guo, Y., Sayed, T. and Zaki, M. H. (2016) 'Automated analysis of pedestrian walking behaviour at a signalised intersection in China', *IET Intelligent Transport Systems*. IET, 11(1), pp. 28–36.

Hardy, M. A. (1993) Regression with dummy variables. Sage.

Helbich, M., Böcker, L. and Dijst, M. (2014) 'Geographic heterogeneity in cycling under various weather conditions: Evidence from Greater Rotterdam', *Journal of Transport Geography*. Elsevier Ltd, 38, pp. 38–47. doi: 10.1016/j.jtrangeo.2014.05.009.

Hong, Z., Wei, Z. and Wei, L. (2006) 'Bicycle-Based Courier and Delivery Services in Beijing', *Transportation Research Board*, (80), pp. 45–51.

Hood, J., Sall, E. and Charlton, B. (2011) 'A GPS-based bicycle route choice model for San Francisco, California', *Transportation Letters The International Journal of Transportation Research*. J Ross Publishing, 3(1), pp. 63–75. doi: 10.3328/TL.2011.03.01.63-75.

Howard, J. (1987) *The Cyclist's Companion*. Rev Sub ed. Penguin Books.

Hudson, J. G. *et al.* (2012) *Using Smartphones to Collect Bicycle Travel Data in Texas*. Available at: http://utcm.tamu.edu/publications/final_reports/Hudson_11-35-69.pdf.

Jin, S. *et al.* (2015) 'Estimating cycleway capacity and bicycle equivalent unit for electric bicycles', *Transportation Research Part A: Policy and Practice*. Elsevier Ltd, 77, pp. 225–248. doi: 10.1016/j.tra.2015.04.013.

Kassim, A. *et al.* (2012) 'Vision-Based Analysis of Cyclists' Speed', in *Transportation Research Board 91st Annual Meeting*.

Kidder, J. L. (2008) 'Appropriating the city: space, theory, and bike messengers', *Theory and Society*, 38(3), pp. 307–328. doi: 10.1007/s11186-008-9079-8.

Kidder, J. L. (2011) *Urban flow: Bike messengers and the city*. First Edit. Ithaca, NY: ILR Press.

Kisters, C. *et al.* (2016) *Revisión y Actualización del Plan Director de Movilidad Ciclista de Madrid*. Madrid. Available at: http://www.madrid.es/portales/munimadrid/es/Inicio/Movilidad-y-transportes/Oficina-de-la-bici/Revision-y-actualizacion-del-Plan-Director-de-Movilidad-Ciclista/Plan-Director-de-Movilidad-Ciclista-2008-Revision-y-actualizacion?vgnextfmt=default&vgnextoid=04.

Kroes, E. P. and Sheldon, R. J. (1988) 'Stated Preference methods. An introduction', *Journal of Transport, Economics and Policy*, 22(1), pp. 11–25.

Kurri, J. and Laakso, J. M. (2002) 'Parking policy measures and their effects in the Helsinki metropolitan area', *The Helsinki Metropolitan Area Publication Series C*, 18.

Larsen, K. A. *et al.* (2013) 'Mapping Bicycle Route GPS Data to a Street Network', in *Bicycle Urbanism Symposium*. Seattle: University of Washington.

Lin, S. *et al.* (2008) 'Comparison Study on Operating Speeds of Electric Bicycles and Bicycles: Experience from Field Investigation in Kunming, China', *Transportation Research Record: Journal of the Transportation Research Board*, 1(2048), p. pp 52-59. doi: 10.3141/2048-07.

Liu, X., Shen, L. D. and Ren, F. (1993) Operational analysis of bicycle interchanges in Beijing, China.

Monzón de Cáceres, A. et al. (2008) Unibici Madrid. Estudio previo para la implantación de un sistema de préstamo de bicilcetas en la Ciudad Universitaria de Madrid. Madrid. Available at: https://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwi_4bOQ1-DVAhXEuhoKHTs3AD0QFggoMAA&url=http%3A%2F%2Fwww.transyt.upm.es%2Findex.php%2Fes%2Fcomponent%2Fphocadownload%2Fcategory%2F1-unibici%3Fdownload%3D2%3Aunibici&usg=AFQjCNGGwJbXfX.

Monzon de Cáceres, A., Rondinella, G. and Muñoz López, B. (2011) Estudio de evaluación de la demanda, impacto social y sobre la movilidad urbana correspondiente al eje de la red básica de vías ciclistas de las calles Atocha - Mayor - Alcalá.

Ortega, E. *et al.* (2015) 'Urban fragmentation map of the Chamberí district in Madrid', *Journal of Maps*. Taylor & Francis, 11(5), pp. 788–797. doi: 10.1080/17445647.2014.964341.

Ortúzar, J. D. D., Iacobelli, A. and Valeze, C. (2000) 'Estimating demand for a cycle-way network', *Transportation Research Part A: Policy and Practice*, 34(5), pp. 353–373. doi: 10.1016/S0965-8564(99)00040-3.

Romanillos, G. *et al.* (2016) 'Big data and cycling', *Transport Reviews*. Taylor & Francis, 36(1), pp. 114–133.

Romanillos, G. *et al.* (2018) 'The pulse of the cycling city: visualising Madrid bike share system GPS routes and cycling flow', *Journal of Maps*, 14(1). doi: 10.1080/17445647.2018.1438932.

Romanillos, G. and Zaltz Austwick, M. (2015) 'Madrid cycle track: visualizing the cyclable city', *Journal of Maps*. Taylor & Francis, 12(5), pp. 1–9.

Rousseeuw, P. J. and Leroy, A. M. (2005) *Robust regression and outlier detection*. John wiley & sons.

Salonen, M. and Toivonen, T. (2013) 'Modelling travel time in urban networks: Comparable measures for private car and public transport', *Journal of Transport Geography*. Elsevier Ltd, 31, pp. 143–153. doi: 10.1016/j.jtrangeo.2013.06.011.

Schuessler, N. and Axhausen, K. W. (2009) *Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT)*. Available at: http://www.baug.ethz.ch/ivt/ivt/vpl/publications/reports/ab568.pdf.

Snizek, B., Sick Nielsen, T. A. and Skov-Petersen, H. (2013) 'Mapping bicyclists' experiences in Copenhagen', *Journal of Transport Geography*, 30, pp. 227–233. doi: 10.1016/j.jtrangeo.2013.02.001.

Suits, D. B. (1957) 'Use of dummy variables in regression equations', *Journal of the American Statistical Association*. Taylor & Francis Group, 52(280), pp. 548–551.

Vlakveld, W. P. *et al.* (2014) 'Speed choice and mental workload of elderly cyclists on e-bikes in simple and complex traffic situations: A field experiment', *Accident Analysis and Prevention*. Elsevier Ltd, 74, pp. 97–106. doi: 10.1016/j.aap.2014.10.018.

Wei, H., Huang, J. and Wang, J. (1997) 'Models for estimating traffic capacity on urban bicycle lanes', in *76th Annual Meeting of the Transportation Research Board, Washington, DC*.

Witlox, F. and Tindermans, H. (2004) 'Evaluating bicycle-car transport mode competitiveness in an urban environment. An activity-based approach.', *World Transport Policy & Practice*, 10(4), pp. 32–42. Available at: http://www.eco-logica.co.uk/pdf/wtpp10.4.pdf#page=32.

Wooldridge, J. M. (2015) ‘Basic regression analysis with time series data’, in *Introductory Econometrics: A modern approach*. Nelson Education, pp. 1–17. Available at: http://www.getcited.org/pub/102998395.

Xu, C. *et al.* (2015) ‘Predicting Free Flow Speed and Crash Risk of Bicycle Traffic Flow Using Artificial Neural Network Models’, 2015.

Yang, C. and Mesbah, M. (2013) ‘Route Choice Behaviour of Cyclists by Stated Preference and Revealed Preference’, in *Australasian Transport Research Forum*. Brisbane, Australia. Available at: http://www.patrec.org/atrf.aspx Route.

8. Acknowledgements

The authors gratefully acknowledge funding from Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (TRA2015-65283-R and FPDI 2013/17001), and the Madrid Regional Government (SOCIALBIGDATA-CM, S2015/HUM-3427).