



Modelling cyclists' route choice using Strava and OSMnx: A case study of the City of Glasgow



Mohammad Anwar Alattar^{a,*}, Caitlin Cottrill^b, Mark Beecroft^b

^a *Geography & Environment, School of Geosciences, University of Aberdeen, Aberdeen Scotland, UK*

^b *Centre for Transport Research, School of Engineering, University of Aberdeen, Aberdeen Scotland, UK*

ARTICLE INFO

Keywords:

Cycling
OSMnx
Strava
Spatial modelling
GIS

ABSTRACT

Previous research has demonstrated the influence of street layout on travel behaviour; however, little research has been undertaken to explore these connections using detailed and robust street network analysis or cycling data. In this study, we harness state-of-the-art datasets to model cyclists' route choice based on a case study of the City of Glasgow, Scotland. First, the social fitness network Strava was used to obtain datasets containing the number of cycling trips on each street intersection for the years 2017 and 2018. Second, we employed a Python toolkit to acquire and analyse the street networks. OSMnx was subsequently employed to quantify several commonly used centrality indices (degree, eigenvector, betweenness and closeness) to measure street layout. Due to the presence of spatial dependence, a spatial error model was used to model route choices. Model results demonstrate that: (1) cyclists' movement models were consistent for the years 2017 and 2018; (2) the presence of a spillover effect suggests that cyclists tend to cycle in proximity to each other; and (3) cyclists avoid streets with high degree centrality values and prefer streets with high eigenvector centrality, betweenness centrality and closeness centrality. These findings reveal cyclists' desired street layouts and can be taken into consideration for future interventions.

1. Introduction

Active travel (AT) comprises travel modes that incorporate physical activity for all or part of a journey (e.g., walking and cycling). AT is known to reduce many urbanization externalities, including traffic congestion, noise pollution, climate change, health inequality, and social exclusion, and is able to boost life quality by improving physical and mental well-being (Avila-Palencia et al., 2018; Grabow et al., 2012; Hamer and Chida, 2008; O'Dea, 2003; Rissel, 2009). In many places, cycling has the potential to substitute for the private car since a large share of trips made by car are <5 miles (Department for Transport, 2019; Hong et al., 2020).

In order to foster cycling, several studies have explored the role that street networks including streets, paths, bridges, and cycle routes exert on shaping travel behaviour, travel modes and more specifically, AT (Boeing, 2019b; Braçe, 2016; Emmanouilidis, 2013). A street network may comprise either planned or unplanned components through accretion (Hanson, 1989) and can be summarised by network centrality indices (explained in 2.2) and the shape movement of road users. Furthermore, street networks facilitate a range of socio-economic

activities, such as protests, public speeches, retail establishment locations and, eventually, the movement of people (Enström and Netzell, 2008; Mohamed et al., 2015). Street network centrality indices quantify the way in which each street segment relates to other streets within the network through various metrics (Emo et al., 2012). Natural movement theory, proposed by Hillier et al. (1993), states that street layout is a primary generator for an individual's movement. Numerous studies have been able to predict human movement through street layout. In particular, Hillier et al. (1993), Penn et al. (1998), and Hillier and Iida (2005) predicted 55–75% of human movement in London, while Read (1999) predicted between 60% and 70% in Amsterdam. Therefore, it is believed that cyclists' movement and street layout are also intertwined.

2. Background

Previous studies that have attempted to model cyclists' route choice have suffered from several limitations related to: (1) insufficiently detailed cycling data; (2) quantifying street network centralities; and

* Corresponding author.

E-mail address: m.alattar@hotmail.co.uk (M.A. Alattar).

(3) the usage of non-spatial models (Liu et al., 2016; McCahill and Garrick, 2008; Orellana and Guerrero, 2019; Raford et al., 2007).

Traditional cycling data sources often suffer from low spatiotemporal resolution, as seen in Raford et al. (2007), who deployed paper-based route collection; McCahill and Garrick (2008), who were limited to 16 intersection-level bicycle counts; and Liu et al. (2016), who used data counted manually by volunteers. Strava Metro can act as an alternative to traditional AT data sources, due to its spatial and temporal granularity. Strava¹ is a social fitness network and analytics platform that allows users to record, share and track their athletic activities (e.g. running, cycling and handcycling) via GPS-enabled devices such as smart phones and watches (Strava Metro, 2019). Strava offers a data service known as Strava Metro (hereinafter referred to as Strava) to help planners utilize such data. Numerous locations around the world are using Strava to make informed and effective AT interventions (Strava Metro, 2015). Data obtained from Strava is known as crowdsourced data, meaning it is generated by a series of individuals. Although there are other networks providing similar data (e.g. Endomondo²), Strava is the largest social fitness network for cyclists.

The second limitation can be overcome by employing OSMnx,³ a Python-based toolkit. In particular, OSMnx allows for easy data acquisition from OpenStreetMap⁴ (OSM), and the analysis and saving of street network data within any geographic information systems (GIS) environment. Inspired by Wikipedia, OSM is a collaborative web-mapping project that collects and disseminates data at the global scale. The data is openly available for public usage and is also known as open geospatial datasets. OSMnx extracts street networks from OSM, retains the topological features (i.e. curvatures) and accounts for street direction and non-planarity, namely it represents street networks using a three-dimensional model for grade separation, bridges, and tunnels (Boeing, 2017). Accordingly, it yields more precise centrality index calculations. Previous research has utilized Space Syntax, a well-known technique employed to analyse street layout; however, it is associated with topological errors and inconsistencies in representing street networks. Kim and Penn (2004) provide a detailed discussion on Space Syntax limitations. A review of open geospatial dataset and its application is provided in Section 2.3.

Cycling is inherently a spatial phenomenon; for example, cyclists will often avoid unpaved roads (Stinson and Bhat, 2004) and hills (Dill, 2009) (although fitness-oriented cyclists favour hills (Griffin and Jiao, 2015)). Accounting for this spatial dependence can provide more robust and insightful results. Thus, spatial regression models are believed to capture more insights than their non-spatial counterparts, with more model estimation power (Lee and Sener, 2020). Both Strava Metro and OSMnx are powered by OSM and thus these two datasets are compatible and facilitate spatial models. Accounting for the spatial variability has previously been deemed as challenging due to low spatio-temporal granularity in street network centrality and cyclists' datasets.

2.1. Study area

Glasgow is the largest and most populated city in Scotland. The majority of journeys (71%) are under 5 km; however, the cycle mode share for work and study is only 1.6% (Cycling Scotland, 2019). Glasgow has been selected as the area of interest in the current study (Fig. 1) due to: (a) the availability of rich data and literature; (b) the existence of AT communities; and (c) the significance of the city for Scotland, particularly the current challenges that can be mitigated through AT, and its ongoing endeavours to promote AT.

2.2. Aims, objectives and outline

This study aims to increase the understanding of cyclists' route choices within street configurations. Such an understanding will reveal favourable street characteristics that can be utilized in conjunction with other factors to foster cycling and consequently increase cycling ridership levels. The objective of this study is to model the number of cycling trips on each street intersection as a function of street network centrality indices (degree centrality [DC], eigenvector centrality [EC], betweenness centrality [BC], and closeness centrality [CC]).

The remainder of the paper is structured as follows: Section 2 reviews the key literature on AT data sources and street layout. Section 3 explains the data acquisition, preparation and analysis methods. The descriptive analysis and specifications of the empirical models used in the study are presented in Section 4. The results of the selected empirical models are interpreted and discussed in Section 5, followed by concluding remarks in Section 6.

3. Literature review

3.1. Traditional and emerging AT data sources

Interest in AT has led to the application of numerous data collection methods. AT data sources can be partitioned into: (1) traditional data sources, in which various conventional sensors are used to collect data - Table 1 summarizes traditional AT data advantages and disadvantages; and (2) emerging data sources that can be attributed to the proliferation of GPS-enabled devices such as smart phones and tablets (Lee and Sener, 2017).

Furthermore, emerging methods can be grouped in two categories, depending on the level of the subject (AT user) interaction required: The collection of passive data requires minimal levels of interaction between the AT user and the device (e.g. standalone GPS devices), while active data measurements (e.g. social fitness networks, in-house developed apps, and public participation geographic information systems) require more extensive levels of interaction (willingness to participate) (Lee and Sener, 2017). Many places around the world are developing in-house active data sources such as CycleTracks⁵ (San Francisco), CycleAtlanta⁶ (City of Atlanta), and CycleLane⁷ (Eugene, Oregon). For more in-depth reviews of various traditional and emerging data sources, readers are referred to Pritchard (2018) and Lee and Sener (2020). Strava Metro datasets are available commercially, with the price of the dataset depending on the number of users and type of data. In Queensland, Australia, the data service has been used to identify potential conflicts between road users as well as to prioritize cycling infrastructure and signage investment (Department of Transport and Main Roads, 2017). Moreover, Strava datasets are able to outperform data obtained from many traditional methods. For example, consider the collection of cycling data using a manual count procedure via a clicker; such a method is characterized by the requirement of pre-training for accurate results, and is also labour intensive, time-consuming, tiresome, and subject to vagaries of weather. Strava datasets overcome such limitations (Day et al., 2016; Jestico et al., 2016). Table 2 listed selected studies that have used Strava.

3.2. Route choice modelling

Crowdsourced data have previously been used in multiple cycling-related applications, such as identifying cycling patterns (Musakwa and Selala, 2016), infrastructure evaluation (Hong et al., 2018), and

¹ <https://www.strava.com/>.

² <https://www.endomondo.com>.

³ <https://github.com/gboeing/osmnx>.

⁴ <https://www.openstreetmap.org/>.

⁵ <https://www.sfcta.org/tools-data/tools/cycletracks>.

⁶ <http://cycleatlanta.org>.

⁷ <https://www.thempo.org/611/CYCLELANE---Bike-routes>.

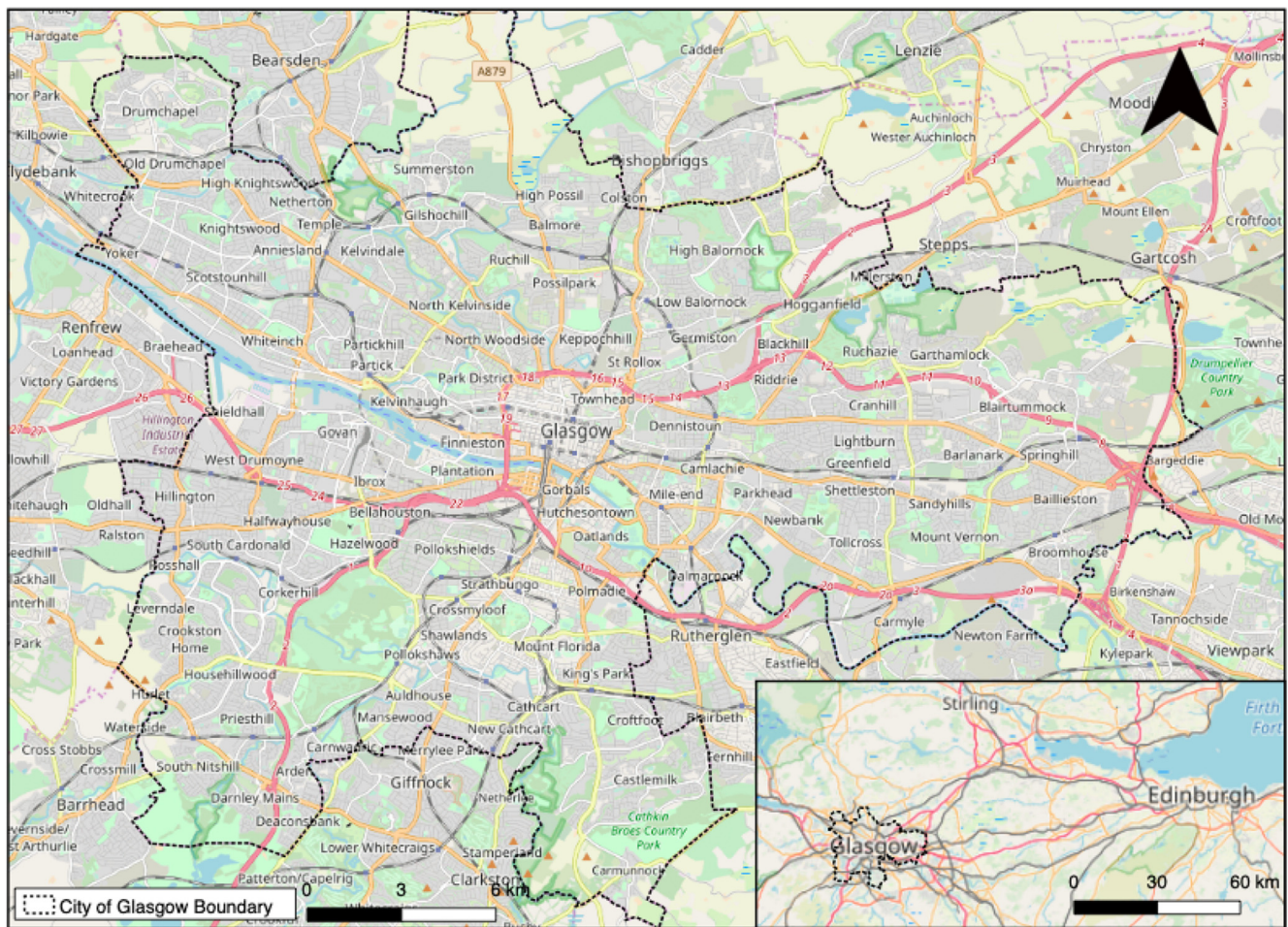


Fig. 1. Overview of area of interest. The delineated area shows the City of Glasgow. [Data: Basemap: OpenStreetMap, licensed under the Open Database License (ODbL)].

Table 1
Summary of traditional AT data sources.

Method	Advantages	Disadvantages
Traditional manual methods such as video recording, travel surveys, and handheld counters.	<ul style="list-style-type: none"> - Require a low level of technological sophistication (AMEC, 2011). - Collect additional information on AT users, such as the use of helmets, gender. - Can collect ground truth counts to validate other methods (Minge et al., 2017). 	<ul style="list-style-type: none"> - Low spatial and temporal quality, labour intensive, time-consuming, and tiresome (Day et al., 2016).
Traditional automated methods such as infrared sensors, magnetometers, and pressure pads.	<ul style="list-style-type: none"> - Can be implemented for lengthy periods and are less impacted by inclement weather conditions. 	<ul style="list-style-type: none"> - Cannot provide additional information about AT users. - Some sensors have further limitations related to their detection techniques^a (Lee and Senar, 2017).

^aFor example, magnetometers detect changes in magnetic fields within the approximation of the sensor created by ferrous metal objects, thus this sensor is not suitable for non-ferrous metal objects (e.g. carbon-fibre bicycles and pedestrians).

Table 2
Selected studies using Strava.

Study	Location	Remarks
Sun et al. (2017b)	Glasgow, Scotland	Estimated the average inhaled pollutant dosage (PM _{2.5}) of cyclists.
Sun and Mobasher (2017)	Glasgow, Scotland	Explored the association between cycling purposes (commuting and non-commuting) and air pollution exposure for both PM ₁₀ and PM _{2.5} .
Hong et al. (2019)	Glasgow, Scotland	Examined cycling volumes before and after the installation of four cycling infrastructure projects around the time of an international multi-sport event took a place
Venter et al. (2020)	Oslo, Norway	Examined the trends of outdoor recreational activities (cycling, walking, running and hiking) during COVID-19 partial lockdown.

air pollution exposure (Sun et al., 2017a, 2017b). The scope of this study is cyclists' route choice, which refers to investigating factors influencing the likelihood of choosing a given street segment along a cycling journey. Route choice models shed light on cyclists' preferences in order to aid transport planners in making routes more appealing for cyclists and identifying optimal locations to situate bicycle facilities (Lu et al., 2018). Table 3 reviews selected studies that integrated crowdsourced data into modelling route choices.

Table 3
Selected route choice studies.

Study	Location	Influencing factors
Sun et al. (2017a)	Glasgow, Scotland	Segment length, segment connectivity, traffic volume, residential land surroundings.
Lin and Fan (2020)	Charlotte, North Carolina	Segment length, number of through lanes, median household income, households in a census block, existence of bike routes, greenways, classified US routes, one-way routes.
Orellana and Guerrero (2019)	Cuenca, Ecuador	Primary roads, existence of bike routes, mixed land use, street slope, household density, living conditions index, number of intersections.
Fitch and Handy (2020)	University of California, Davis campus & San Francisco, California	Bike route, segregate bike route, segment length, route circuitry.
Lu, Scott and Dalumpines (2018)	Hamilton, Ontario	Bike facilitates, traffic volume, intersections.

3.3. Open geospatial data

In the context of ubiquitous ICT and the proliferation of open data, open geospatial data is becoming increasingly popular. Numerous platforms have been developed to collect geospatial data from users,

known as volunteered geographic information (VGI). Such platforms include OSM, Flickr, Twitter and Foursquare, with VGI platforms dedicated for cyclists including BikeMaps, BikeLaneUpRising and Safe-Lanes. Open geospatial data facilitate research transparency, reproducibility and scalability and many governmental bodies have adopted this concept. For example, [New York City's map portal \(http://maps.nyc.gov/doitt/nycitymap/\)](http://maps.nyc.gov/doitt/nycitymap/) provides a single access point for official geospatial datasets, including transportation, locations of airport, bicycle parking, and off-street parking (Mobasheri, 2020).

OSM has a noticeable presence in urban studies, as it provides updated data that is typically hard/expensive to obtain from other data sources. For example, OSM includes informal paths and passageways that are not included in official data. This presence can be witnessed in many studies. Graser et al. (2015) determined a close agreement when assessing the quality of OSM street networks for vehicle routing in Vienna by using official data as a reference, otherwise known as an extrinsic assessment. Mobasheri et al. (2017) evaluated the suitability of OSM for wheelchair users via both extrinsic and intrinsic assessments (i.e., the procedure of checking the completeness level of attributes such as sidewalk width and incline within OSM data itself) in various German cities, determining a high level of acceptability of the data. Similarly for cycling, Hochmair et al. (2015) assessed the completeness of OSM cycling data (lanes and trails) in selected US urbanized area using governmental datasets, satellite imagery and Google maps as a reference. In some locations, OSM data were found to be even more accurate than the reference data due to the greater number of mappers, though other regions were found to have some

Table 4
Definition and illustrations of centrality indices.

Index	Definition	Illustration [±]
Degree Centrality [DC]	Returns the fraction of streets that it is connected to (Hagberg et al., 2018; Newman, 2006).	
Eigenvector Centrality [EC]	Computes the extent of centrality of a given street based on the connectivity of the streets to which it has ties (Hagberg et al., 2018; Newman, 2006).	
Betweenness Centrality [BC]	Computes how often a given street appears on the shortest path between other streets in the network (Hagberg et al., 2018; Newman, 2006).	
Closeness Centrality [CC]	The reciprocal of the average shortest path from any given street segment to all other street segments in the network (Hagberg et al., 2018; Newman, 2006).	

± Hot (cold) colours represent high (low) values of the corresponding index.

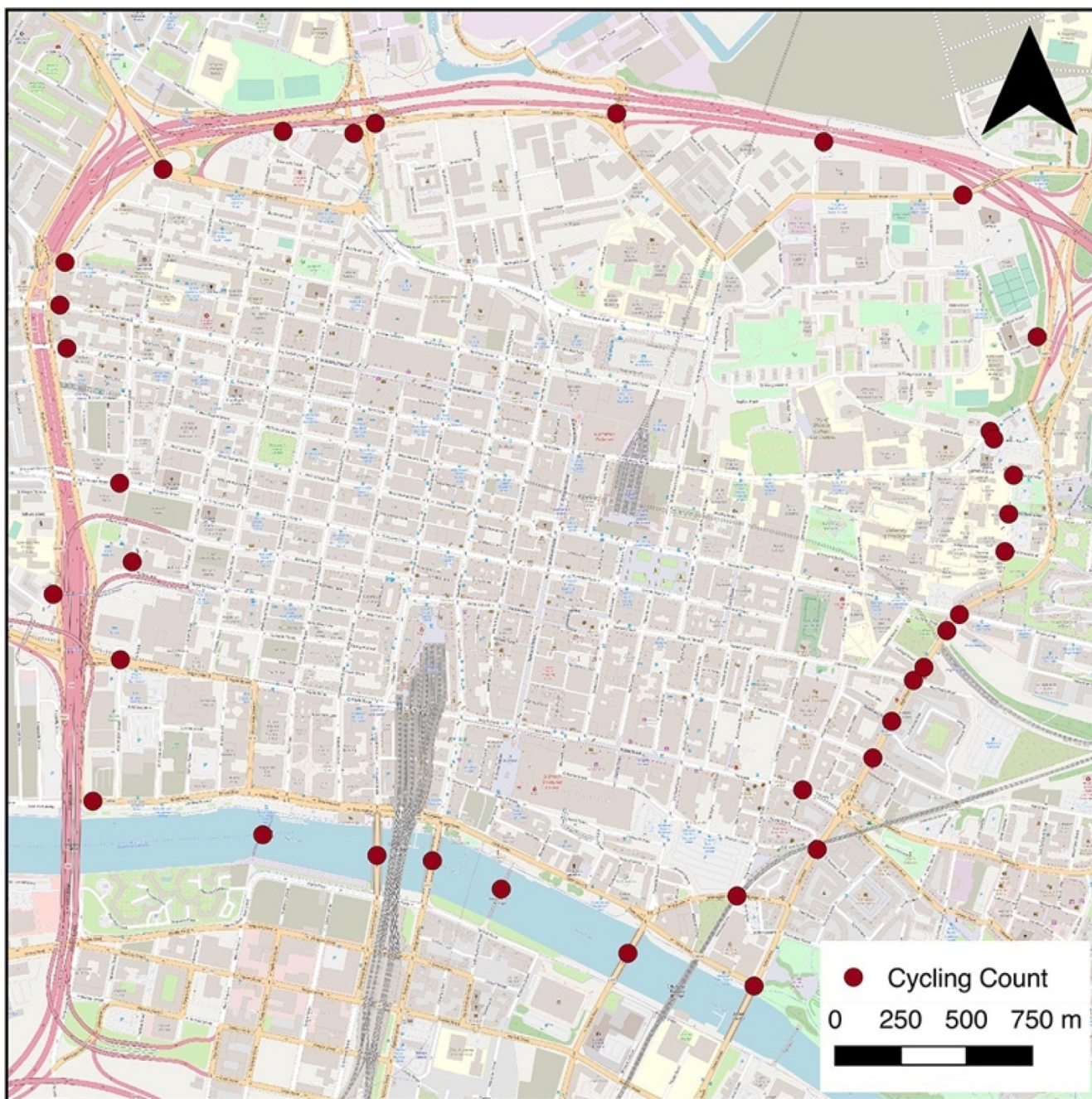


Fig. 2. Cordon cycle count locations (Glasgow City Council, 2018). [Basemap: OpenStreetMap, licensed under the Open Database License (ODbL)].

missing data. Rousell and Zipf (2017) prototyped a pedestrian-based navigation system with landmark extraction using only OSM data.

Using raw data from OSM to model street networks is cumbersome, as it does not provide network typology (connections and configurations). The development of the OSMnx toolkit (described in 4.2) has extended the concept of open geospatial data by allowing for downloading and modelling of street networks (i.e. walkable, bikeable, and drivable), in addition to building footprint, points of interest, and elevation data. Boeing (2020) made all street networks in the world openly available on the Harvard Dataverse and calculated various transportation-related indicators such as street grade and intersection counts. OSMnx also has applications in transportation studies. For example, Boeing (2019a) explored transportation efficiency using network circuitry (the ratio of street network distances to straight-line distances) for walkable and drivable networks in 40 US cities. Yen et al.

(2019) extended this investigation to include bikeable networks and applied it to Phnom Penh, Cambodia.

3.4. Street layout

Street layout has a primary role in AT mobility, safety, and wayfinding (Hillier et al., 1993). Rifaat et al. (2012) reported that loop and lollipop street patterns are associated more with pedestrian crash severity compared to gridiron street patterns. The impact of street angularity has been investigated by Dalton (2001), who concluded that people attempt to conserve street linearity in their journeys by avoiding unnecessary turns. In the current study, street layout is represented by four centrality indices that measure how each street segment relates to other streets within the network, namely the DC, EC, BC and CC indices, presented in Table 4.

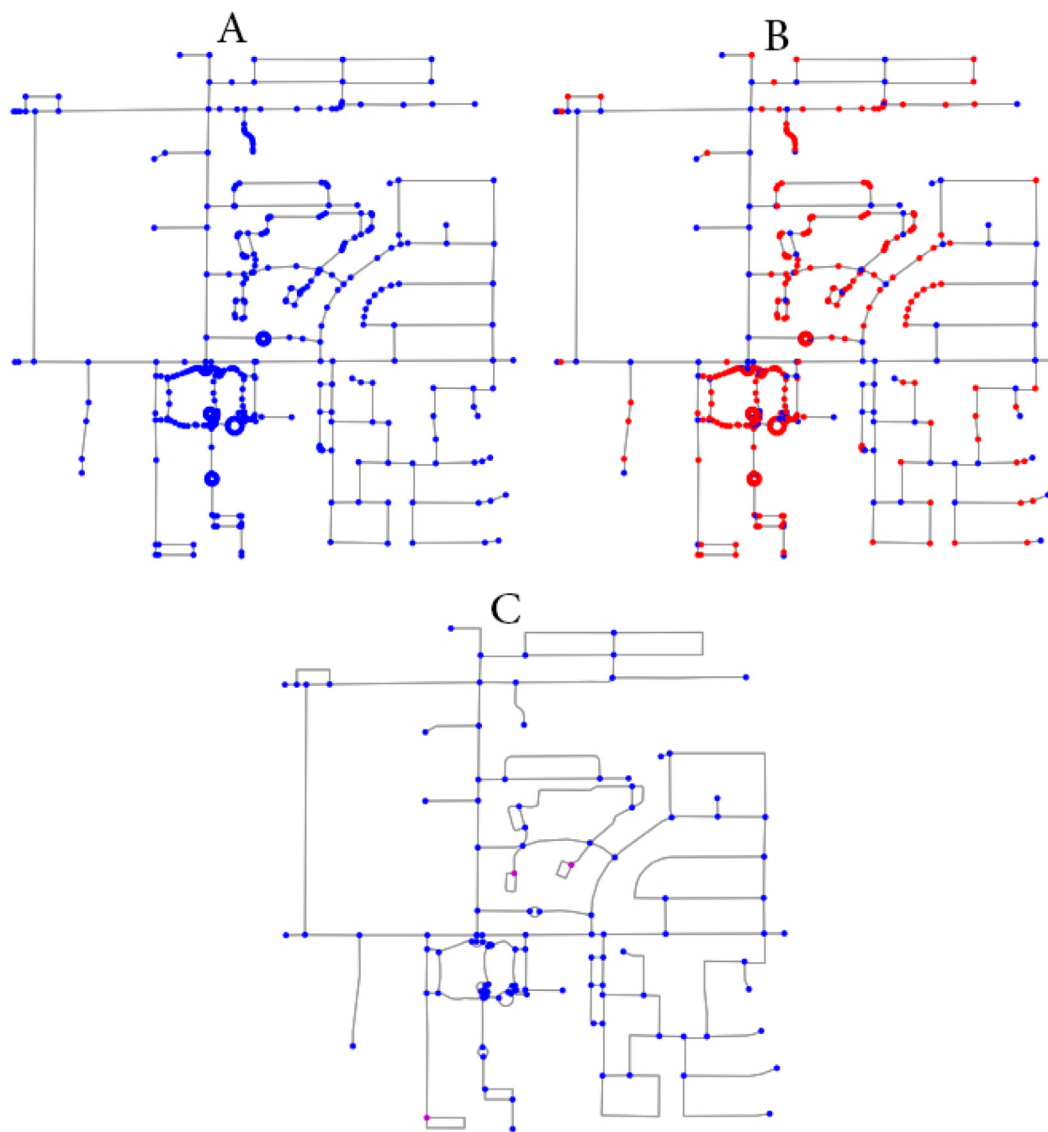


Fig. 3. OSMNx strict mode simplification (Boeing, 2017). (A) Original street network obtained from OSM. (B) Red coloured nodes are removed after simplification. (C) Strictly simplified network, with self-loop nodes coloured in magenta. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

DC refers to the number of street segments that are immediately connected to any given street. Well-connected streets offer maximal accessibility to other streets and alternative routes, while also maximising the number of intersections (Pucher and Buehler, 2012: 121). EC is a more advanced version of DC, which also includes the influence of a street in a street network. More specifically, a higher score is assigned to streets which are connected to multiple streets with high DC values (Newman, 2006).

BC measures the likelihood for any given street segment to be passed through as the shortest route from one street segment to all other street segments (Freeman, 1977; Hillier et al., 1987; Hillier and Iida, 2005). This index quantifies how well any given street functions as a unique bridge between other streets (Joss, 2016). BC is assumed to be associated with travel time minimisation, as it increases the likelihood that a given street segment serves as a shortest path within the street network, and thus acts as a trip option (McCahill and Garrick, 2008; Samson, 2017).

CC represents the number of turns that must be traversed in order to reach all street segments from any street segment of origin within a street network (Hillier, 2012). Compared to streets with a low CC value, those with high CC values can be reached with fewer turns, as they are closer to all other streets within the network (Koohsari et al., 2016). Kim and Sohn (2002) and Al-Shaheen (2012) found that streets with high CC often exhibit mixed land-use (residential, commercial, and work areas), a well-known variable that is correlated with AT (Frank et al., 2007; Kerr et al., 2007).

4. Methods

In this study, cyclists' route choices are modelled using street layout. Variables related to bike ridership were derived from Strava Metro, whereas OSMNx was used to calculate street network centrality indices. QGIS and GeoDa on macOS Catalina were used for data preparation, visualisation and analysis.

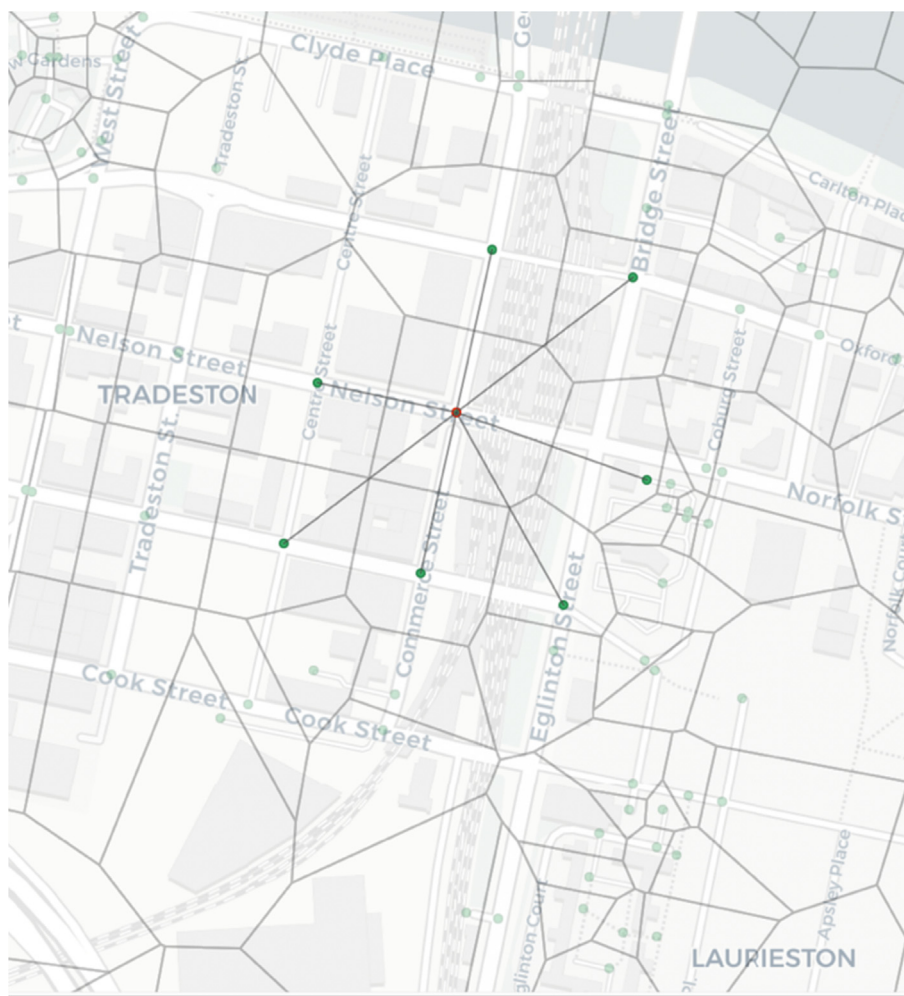


Fig. 4. Queen contiguity. Neighbour points are determined based on border sharing. Borders are delineated by Thiessen polygons, which assign an area of influence to the closet point member. The selected point shares the border with 7 neighbours.

4.1. Strava

Strava's 2017 and 2018 cyclist datasets for Glasgow were obtained from the University of Glasgow's Urban Big Data Centre.⁸ Each dataset has three products: Street, Origin/Destination, and Nodes. The Nodes product (point Shapefile) represents street intersections extracted from OSM and was used to model cyclists' movement. The Street product was employed for the purpose of visualization (shown in 5.1.1) and to extract temporal patterns. Core data files were layered to the shapefile to provide a summary for the entire year (known as yearly roll-up) for each point feature. The attribute ACTCNT, which denotes the count of trips through the intersection (hereafter CCT), was used to model cyclists' route choice. As for the temporal patterns, Edges – Hourly files were used to plot the Activity_Count variable, which denotes the number of trips on the street/trail segment at hourly and daily levels in the direction of the line digitization.

It should be noted that all counts must meet a minimum of 3 users before they are shown, and are rounded up to the nearest multiple of 5 (i.e. counts of 4 are shown as 5, counts of 8 are shown as 10 and so on). In addition to the aforementioned data, demographic information and other relevant data was acquired for the timeframe of the products.

In order to determine the extent to which the Strava cycling data represents actual cycle ridership in Glasgow, the 2017 and 2018 data-

sets were correlated with corresponding cordon cycle counts, which took place over two successive days between 6:00 am and 8:00 pm in September 2017 and 2018 at 36 locations (Fig. 2). This is a common practice to examine the representativeness of Strava (Conrow et al., 2018; Jestico et al., 2016; Roy et al., 2019). Note that the cordon cycle count was manually digitized using data obtained from Glasgow City Council (2018). Since the cycle count data reflects the number of cyclists going to and from the city, the Strava street product was used as it contains TACTCNT, which denotes the total count of trips on the street/trail segment, regardless of the direction of travel. Using linear regression analysis, the coefficient of determination (R^2) was found to be 0.69 and 0.67 for 2017 and 2018, respectively. This suggests that both Strava datasets adequately represent cycle ridership in Glasgow.

4.2. OSMNx

OSMNx, a Python-based toolkit, was used to analyse the street network as it overcomes the limitations of traditional methods (e.g. small sample size and an excessive network simplification). First, one can query if any street network exists on OSM using the location name or a predefined polygon. The toolkit automatically accounts for non-planarity, meaning that it represents street networks using a three-dimensional model for grade separation, bridges, and tunnels (Boeing, 2017).

⁸ <https://www.ubdc.ac.uk>.

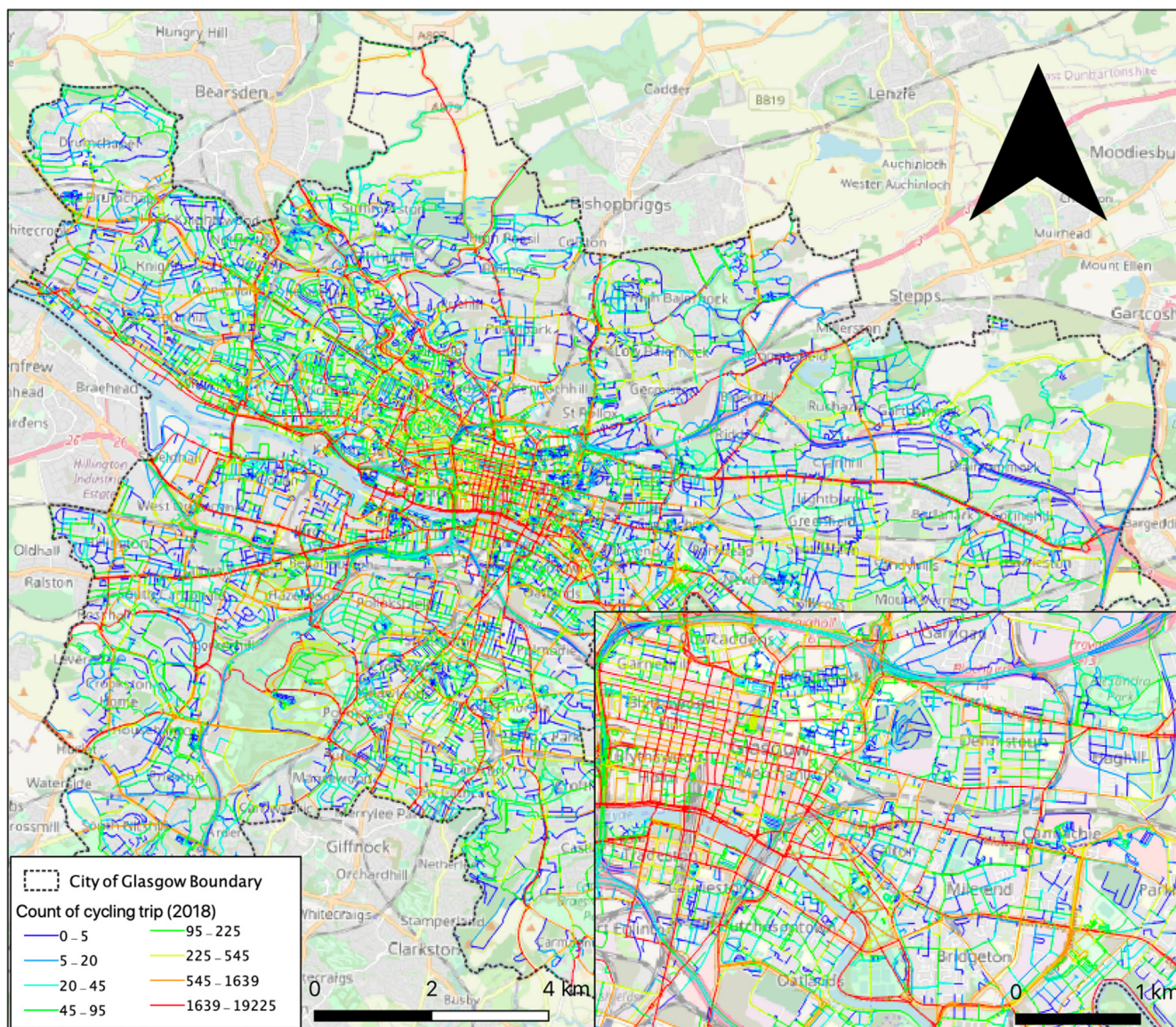


Fig. 5. CCT 2018 for Glasgow. Data obtained from the Strava Metro street product for 2018 [Basemap: OpenStreetMap, licensed under the Open Database License (ODbL)].

The toolkit then corrects and simplifies the street network, retaining all spatial geometry as the original OSM nodes are positioned on intersections, dead-ends, and along a single curved street segment. In the current study, the street network was simplified using the strict mode, which retains a node if it constitutes a dead-end, self-loop, and a true intersection of multiple street segments in which at least one of the streets continues through the intersection. Note that for the latter, the intersection of two streets, where both end at the intersection and thus create an elbow, does not constitute a node. Thus, as illustrated in Fig. 3, this procedure will eliminate excess street fragmentation, as is the case of Space Syntax, and yield better representative results (Boeing, 2017).

After acquiring the desired street network, OSMnx can calculate a wide range of indicators for both edges (street segments) and nodes (intersections). In this work, we focus on the centrality indices mentioned in Table 4. For the purpose of visualization, the street segments are illustrated in 5.1.2. The nodes were included in the cyclists’ movement model.

4.3. Data preparation

The cycling trip counts for both years (derived from the Strava node product) were merged with the centrality indices (derived from the OSMnx node product) using the QGIS NNJoin plugin (version 3.4.14-Madeira). This feature joins two shapefiles based on the nearest neighbour. The input and join layers were set as the centrality index and Strava layers, respectively. This results in two shapefiles with 35,561-point features corresponding to the attributes from the cycling trip count derived from Strava, while the centrality indices were derived from the urban network analysis. The nearest-neighbour distance tolerance was set to <10 m to allow points with less than a 10-m radius to be aligned, resulting in datasets with 31,374 point features.

GeoDa (version 1.14.0) was then deployed for the logarithmic transformation of all variables in order to reduce any skewness in the data. Following this, the weight matrix was created using queen contiguity to define neighbour points based on border sharing. The

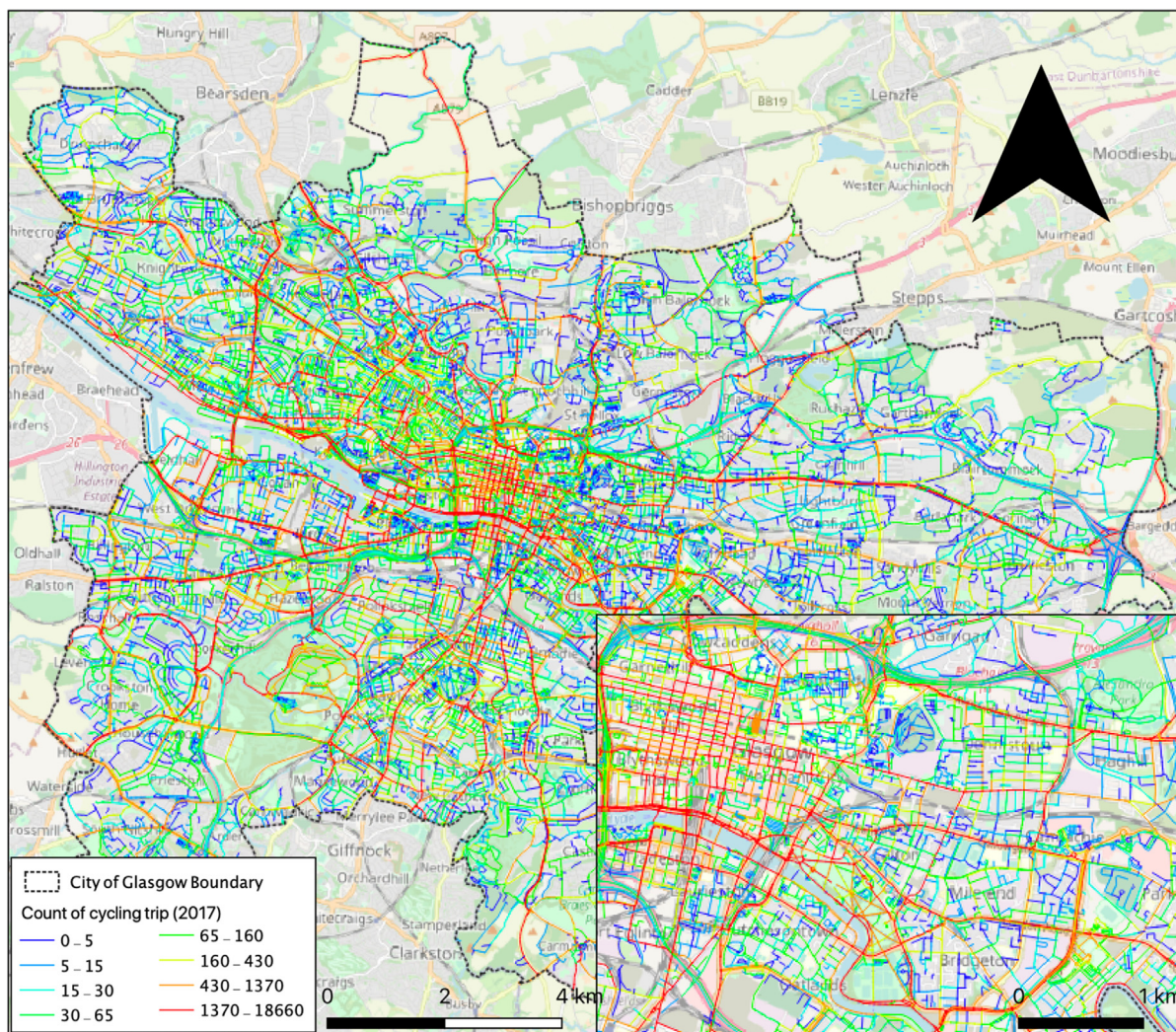


Fig. 6. CCT 2017 for Glasgow. Data obtained from the Strava Metro street product for 2017 [Basemap: OpenStreetMap, licensed under the Open Database License (ODbL)].

borders for each point were delineated via Thiessen polygons, whereby all locations are assigned to the closet member of the points to determine the area of influence of each point (Yamada, 2016). Fig. 4 depicts the queen contiguity for a given point.

5. Results

The results of both the descriptive and inferential analysis are reported in this section. The descriptive analysis generally consists of the spatial visualisation of the data obtained from Strava and OSMnx along with a descriptive summary of demographic information. The inferential analysis results report the output of the spatial regression model. The final section presents the procedures used to assess the adopted method.

5.1. Descriptive analysis

5.1.1. Strava

Figs. 5 and 6 map Glasgow Strava trip volumes for 2018 and 2017, respectively. Spatial cycling trends are similar for both years in the city centre. In particular, the distinctive gridiron street pattern in the city centre, as well as its arterial roads, allows for a high volume of cycling trips. The locator maps in both figures depict the cycle ridership in the

city centre for both years, clearly demonstrating the high number of cycling trips along the River Clyde and over the bridges across the River Clyde.

The temporal distribution of the cycling trips of each street segment in Glasgow is plotted in Fig. 7. The daily cycling trips plotted in Fig. 7A and B exhibit seasonal variability, where the number of trips plateaus in the non-winter seasons (around day 97 to 265). In 2018, the number of cycling trips peaks at day 252 (September 9th) and 220 (August 8th) in conjunction with major cycling events, namely, Pedal for Scotland and the Glasgow 2018 European Championships Cycling Time Trial, respectively. Similarly, for 2017, day 253 (September 10th) peaks in conjunction with Pedal for Scotland. The total hourly cycling trips plotted in Fig. 7C and D exhibited the same trends. The number of cycling trips peaks at around 8 AM and 5 PM, coinciding with commuting peaks, with a slight plateau at around 12 PM that may be attributed to recreational trips.

5.1.2. OSMnx

The analysis of the street layout characteristics in the city demonstrate the scattering of the DC values across the city (Fig. 8), with low DC values located on arterial roads and to the west of the city centre. Note that DC is affected not just by the intersections, but also by the street direction, which is evident in the western region of the city centre. In particular, an intersection of two bidirectional streets will

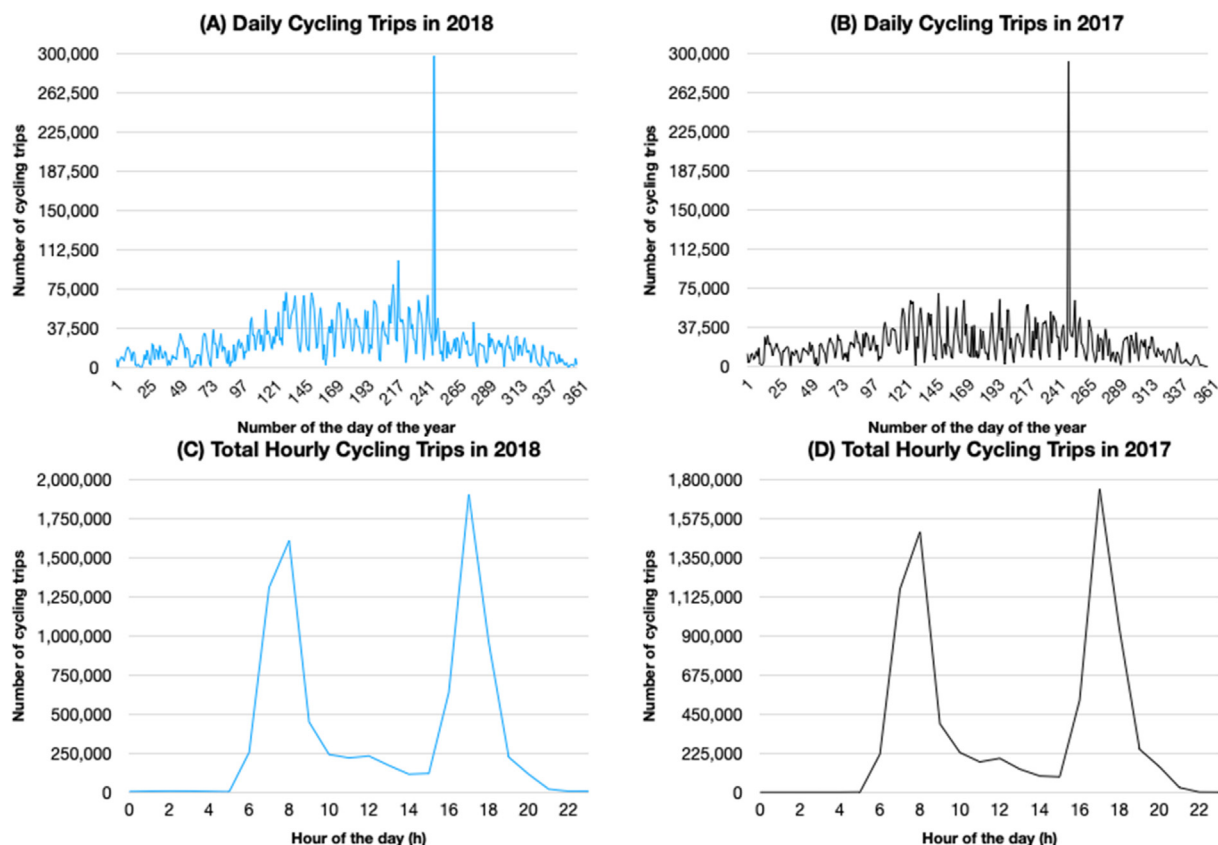


Fig. 7. Temporal distribution of CCT in Glasgow. (A) Daily cycling trips in 2018. (B) Daily cycling trips in 2017. (C) Total hourly cycling trips in 2018. (D) Total hourly cycling trips in 2017.

have a DC value of 8, while intersections of two one-way streets will have a DC value of 4. Thus, intersections of bidirectional and one-way streets will vary accordingly. However, the illustrated values are normalised. In comparison, as can be seen in Fig. 9, EC values exhibit an almost symmetric diffused pattern stemming from Glasgow Green, located in the east end of the city centre.

Fig. 10 presents the BC values for the City of Glasgow's street network. It can be seen that streets that function as bridges between sets of other streets (e.g. arterial roads, bridges and tunnels) exhibit high BC values. The lowest BC values correspond to streets located in the periphery of the city.

The CC captures the extent of the topological closeness between a street segment and all other street segments. It can be seen from Fig. 11 that the arterial road (A8) that passes through central Glasgow and its extensions towards the south west to the north east exhibits the highest CC values, and gradually decrease outwardly. This reflects the hierarchy and location of the roads, whereby arterial roads exhibit the highest values, and local roads (the lowest level in the hierarchy) located in the outskirts of the city hold the lowest values.

5.2. Inferential analysis

5.2.1. Spatial dependence

We adopted ordinary least squares (OLS) regression to obtain the residuals for CCT 2018 and 2017 and subsequently assessed the multicollinearity via the variance inflation factor (VIF). The VIF value was found to be < 10 (mean VIF = 1.06), implying the absence of multicollinearity between the explanatory variables (street network centralities). Spatial dependence analysis was performed to determine the appropriateness of spatial regression models. Spatial dependence occurs when a value at a given location is associated with those of

neighbouring locations. The presence of spatial dependence of the residual implies a violation of the basic OLS assumption of spatially independent residuals. Thus, spatial regression models can improve model estimations (Anselin, 2013).

Table 5 reports the diagnostics for spatial dependence. Both Moran's I and Lagrange Multiplier (LM) spatial dependence tests were found to be significant at the 1% level, indicating that the data was able to fit both the spatial error model and the spatial lag model, respectively. The spatial error model accounts for the spatial dependence of residuals (ϵ) in the neighbouring regions (spatial heteroskedasticity), while the spatial lag model accounts for the spatial dependence of the dependent variable (y) in the neighbouring regions (Anselin, 2013; Burkey, 2018). Local indicator of spatial autocorrelation (LISA) analysis was carried out with significance assessed by 9999 Monte Carlo permutations to identify spatial clusters (composed of High-High and Low-Low) and outliers (composed of High-Low and Low-High) from the OLS residuals for 2018 and 2017 (Figs. 12 and 13, respectively) (Anselin, 1995). High-High refers to features with high residual values that are surrounded by neighbouring features with high residual values; Low-Low refers to features with low residual values that are surrounded by neighbouring features with low residual values; High-Low refers to features with high residual values that are surrounded by neighbouring features with low residual values; and Low-High refers to features with low residual values that are surrounded by neighbouring features with high residual values.

5.2.2. Model selection criteria

Both the spatial lag and spatial error models are strong prediction models. Thus, in order to select the most appropriate spatial regression model, four selection criteria (reported in Table 6) were used to determine the optimal model goodness of fit (R^2), log likelihood (LogL),

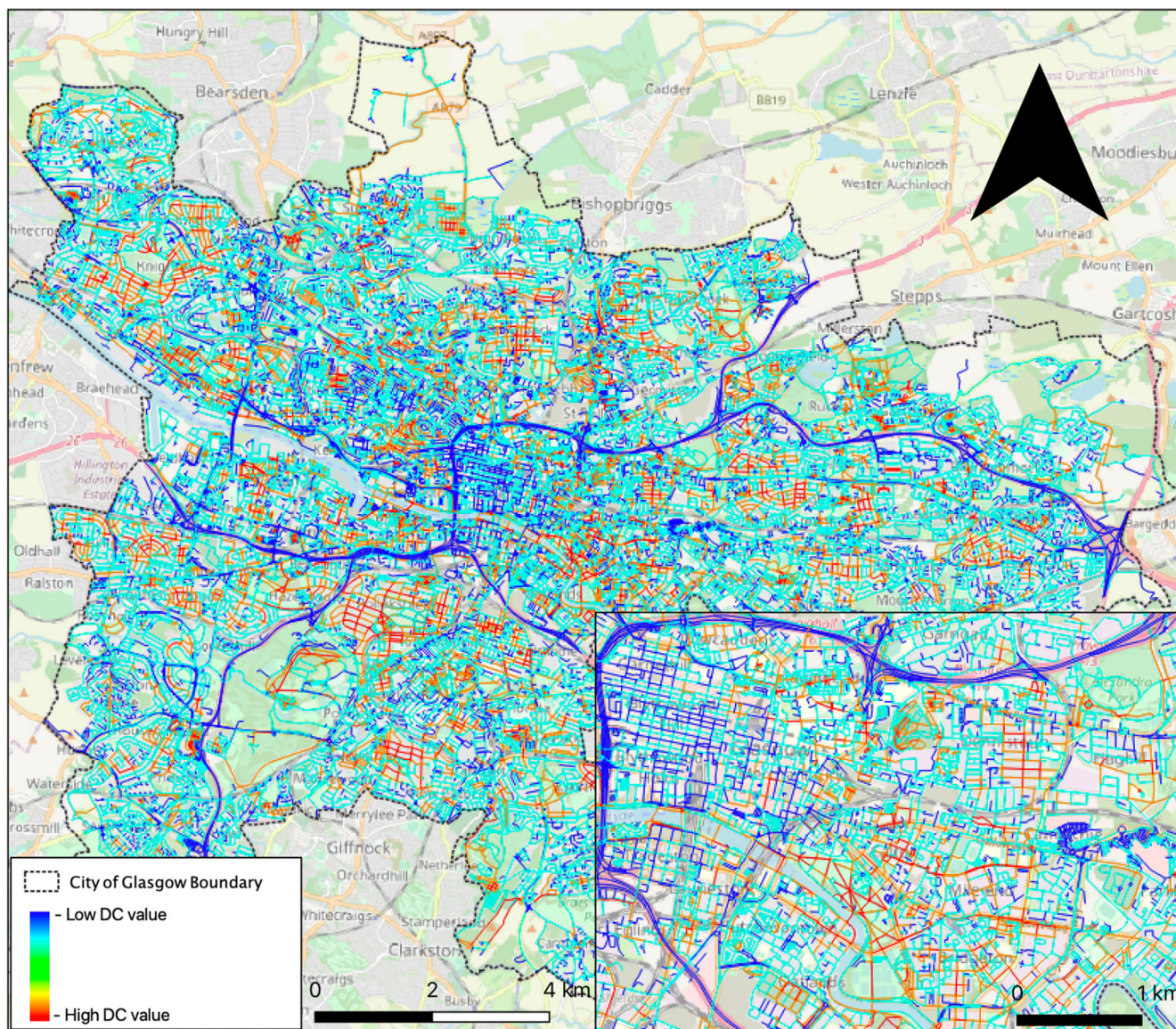


Fig. 8. Glasgow street network DC values. [Basemap: OpenStreetMap, licensed under the Open Database License (ODbL)].

Akaike info criterion (AIC), and Schwarz criterion (SC). Higher R^2 and LogL, and lower AIC and SC values imply a better regression fit (Chuai et al., 2012). The spatial error models for both years exhibit higher R^2 and LogL, and lower AIC and SC values (Table 6). This suggests the presence of statistical evidence of spatial heteroskedasticity, which occurs when the error variance is non-constant across observations (Brennan and Carroll, 1987: 443). Thus, the spatial error model fits the data more efficiently compared to the spatial lag model (see Table 7).

5.2.3. Spatial error model

Table 6 reports the results of the spatial error models for CCT for the years 2018 and 2017. All parameters were significant at the 1% level. In particular, the lag coefficient λ denotes the spillover effect of neighbouring intersections, such that an increase of 1% in the number of trips in an intersection leads to a 0.55% and 0.62% increase in adjacent intersections for the 2018 and 2017 models respectively. All centrality metrics were found to have a positive impact on the trip, with the exception of DC, which exhibited a negative impact.

In 2018, CC was observed to have the greatest coefficient of correlation a 1% increase in CC yielded a 1.15% increase in trips. This was followed by BC, where a 1% increase led to a 0.28% rise in trip numbers. EC demonstrated the smallest coefficient of correlation, implying that a large amount of the variation cannot be explained by the model, resulting in a dispersion of the prediction. In particular, a 1% increase in EC resulted in a 0.02% rise in the number of trips. DC was found to have a negative impact on the trip number, whereby a 1% rise in DC led to a 0.15% decrease in the number of trips. A similar trend was observed for the 2017 model.

5.3. Method assessment

Strava datasets for both years exhibited an adequate representation for cycling in Glasgow, with a linear correlation observed with cordon counts in the city centre. A visual inspection of the spatial distribution of cyclists also implies a plausible representation high CCT can be identified in the city centre and along the river, with low CCT values in highways. Furthermore, the CCT for both years were found to be

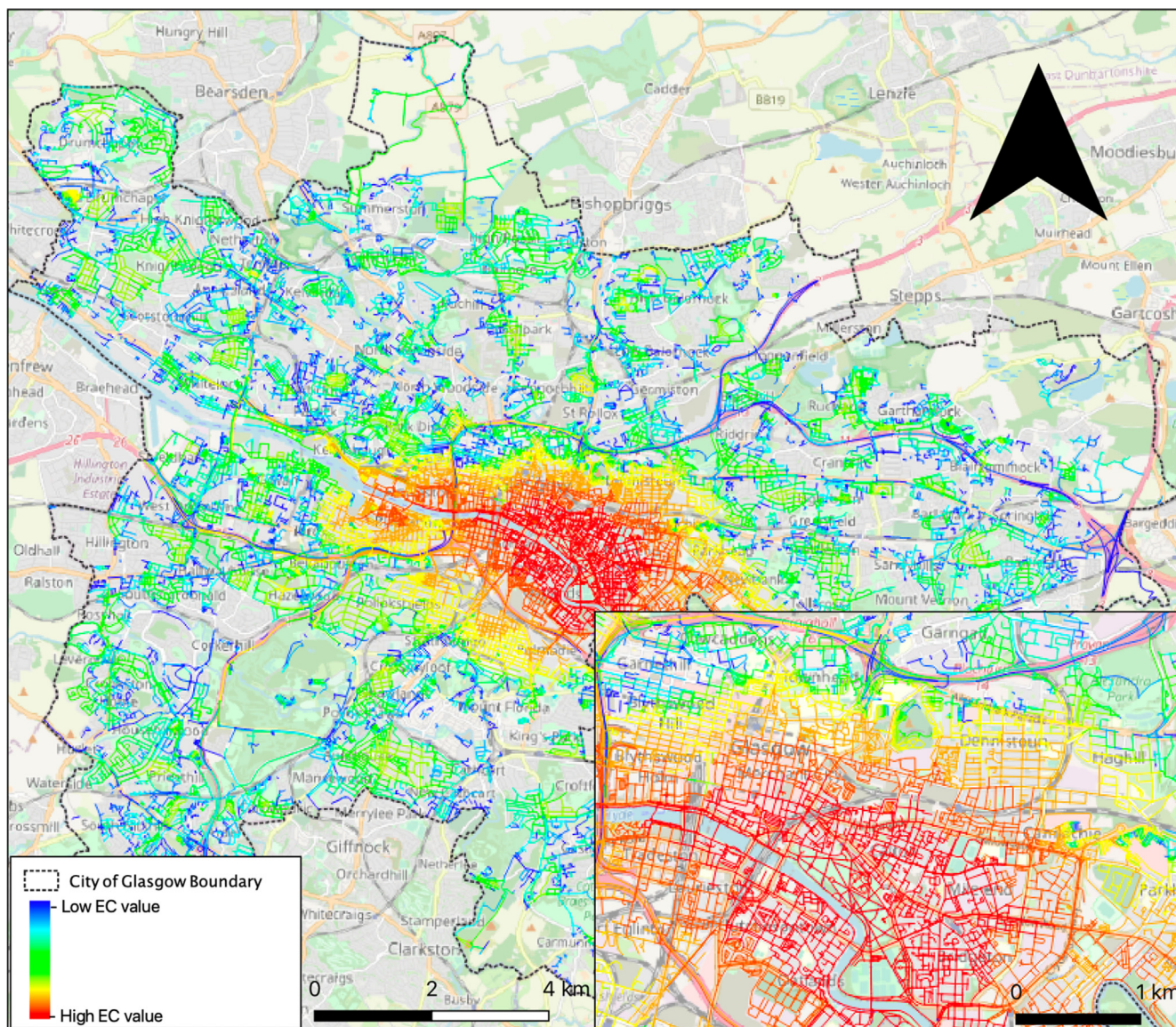


Fig. 9. Glasgow street network EC values. [Basemap: OpenStreetMap, licensed under the Open Database License (ODbL)].

sensitive to seasonality and both commuting and non-commuting peaks.

The adoption of OSM in the current study is demonstrated as suitable by the results. [Haklay \(2010\)](#) found a fairly accurate agreement between OSM and the national mapping agency for Great Britain, the Ordnance Survey, for London and England. Given the proliferation of GPS-enabled devices, this is believed to also be the case in Glasgow. Nonetheless, OSM is used as a background map for Strava, making it the optimal option for spatial matching.

The presence of spatial dependence and model selection criteria in [Table 6](#) are attributed to the nominate spatial error model and discarded OLS. Replicating the analysis for both years, the number of observations and similar coefficients trends introduces rigidity to the results.

6. Discussion

Results demonstrated the ability of the Strava datasets to model cyclists' movement. First, Strava was found to be strongly associated

with the cordon cycle counts. Second, total daily trips on all street segments were identified as sensitive to seasonal variability. Third, total hourly trips on all street segments clearly depicted commute and recreational patterns. Fourth, gender disparity reflects the prevalence of male cyclists. These trends are in accordance with findings from [McPherson \(2017\)](#), who reported similar temporal trends (both daily and hourly) in Glasgow's public cycle hire scheme, and [McPherson \(2017\)](#) and [Motherwell \(2018\)](#) who found gender disparity towards males.

The count of cycling trips (CCT) for the year 2018 was modelled and the results were validated by repeating the analysis for CCT 2017. A spatial error model was used to account for the spatial dependence in the residuals. The models' predictive powers were found to be moderate (0.42 and 0.53, respectively) with similar explanatory variable coefficients trends.

[Nordström and Manum \(2015\)](#) found that betweenness centrality (BC), a proxy for the likelihood of a given street to act as a shortest path between the origin and destinations, is insufficient for cyclist route predictions as less direct routes are used as a result of the poor

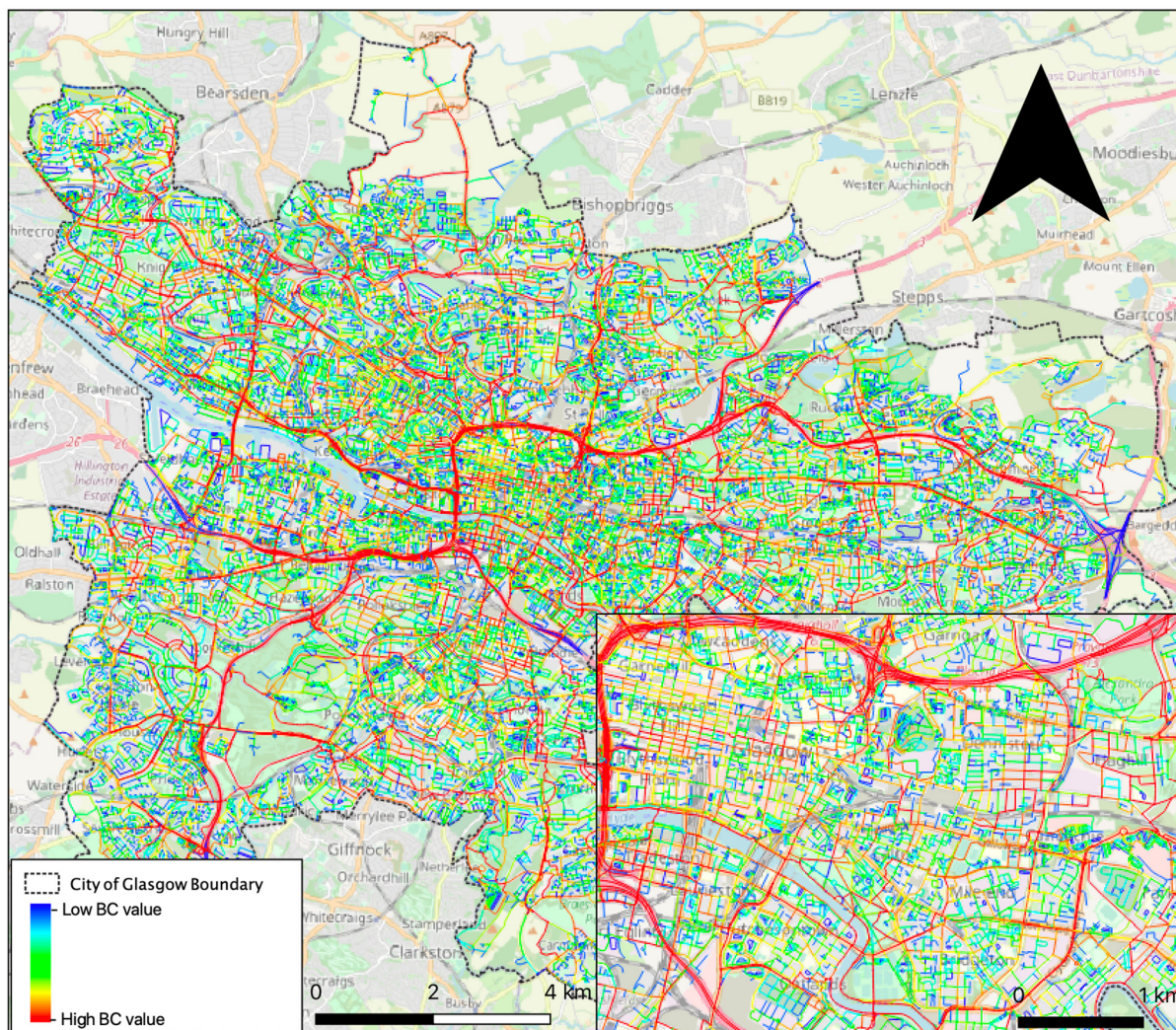


Fig. 10. Glasgow street network BC values. [Basemap: OpenStreetMap, licensed under the Open Database License (ODBL)].

safety of the shortest path routes. However, our findings suggest that cyclists in the City of Glasgow flow along the shortest-path, and this spillover effect may reinforce the safety-in-numbers theory, which states that ‘the greater the cyclists’ volume, the safer the path (Jacobsen, 2015)’.

Both basic and advanced measures of street connectivity were adopted in this study, namely, degree centrality (DC) and eigenvector centrality (EC), respectively. DC has a negative influence on trip counts, confirming that cyclists seem to prefer using less connected streets, and avoid intersections. Similar findings have been reported in Stinson and Bhat (2003), Menghini et al. (2010) and Snizek et al. (2013). Furthermore, previous research has indicated that intersections are perceived as potential areas for accidents due to: (i) the difficulty in anticipating how vehicles will move; (ii) noiseless and quiet vehicles which hinder localization; and (iii) the unpredictability of cyclists due to the lack of stop and indicator light usage (ECMT, 2000; Strauss et al., 2013). In addition, intersections can result in cyclists losing their momentum, which leads to a slow re-start as considerable energy is required to re-gain the momentum (McLean, 2017). As a countermeasure, the ‘Green Wave’ system, first introduced in Copenhagen, offers a seamless flow for cyclists via synchronized consecutive green traffic light signals (Pucher and Buehler, 2012).

Conceptually, EC is a more advanced variant of DC that measures the popularity of street segments. DC captures the immediate number of connections to other streets (which are likely to be intersections), whereas EC is able to capture the indirect number of connections and does not necessarily reflect intersections. By following the same reasoning of Hong et al. (2016), who developed a model of pedestrian exposure, EC estimates the extent to which a street segment is connected to other well-connected streets. Thus, cyclists are able to scatter to other desirable streets, such as those with less congestion and complex traffic situations. EC was observed to be high within Glasgow Green (see Fig. 10) and exhibited a gradual outward decreasing trend. The high number of cycling trips at this location may be a function of the proximity to the city centre, several landmarks, the River Clyde, and the presence of Glasgow’s most popular bike rental location operated by NextBike (McPherson, 2017). Glasgow Green, Scotland’s oldest public park, contains several landmarks, such as the largest terracotta fountain in the world (Doulton Fountain), McLennan Arch, a historical museum and a glasshouse, People Place’s and Winter Garden (Ashurst, 2008), which may also make it attractive to cyclists.

Closeness centrality (CC) values were found to be highest on major roads (e.g. the A8 and A77) and their surrounding areas, and gradually decreased towards the periphery of the city. These major roads accommodate cyclists with the presence of bridges, artwork and feature

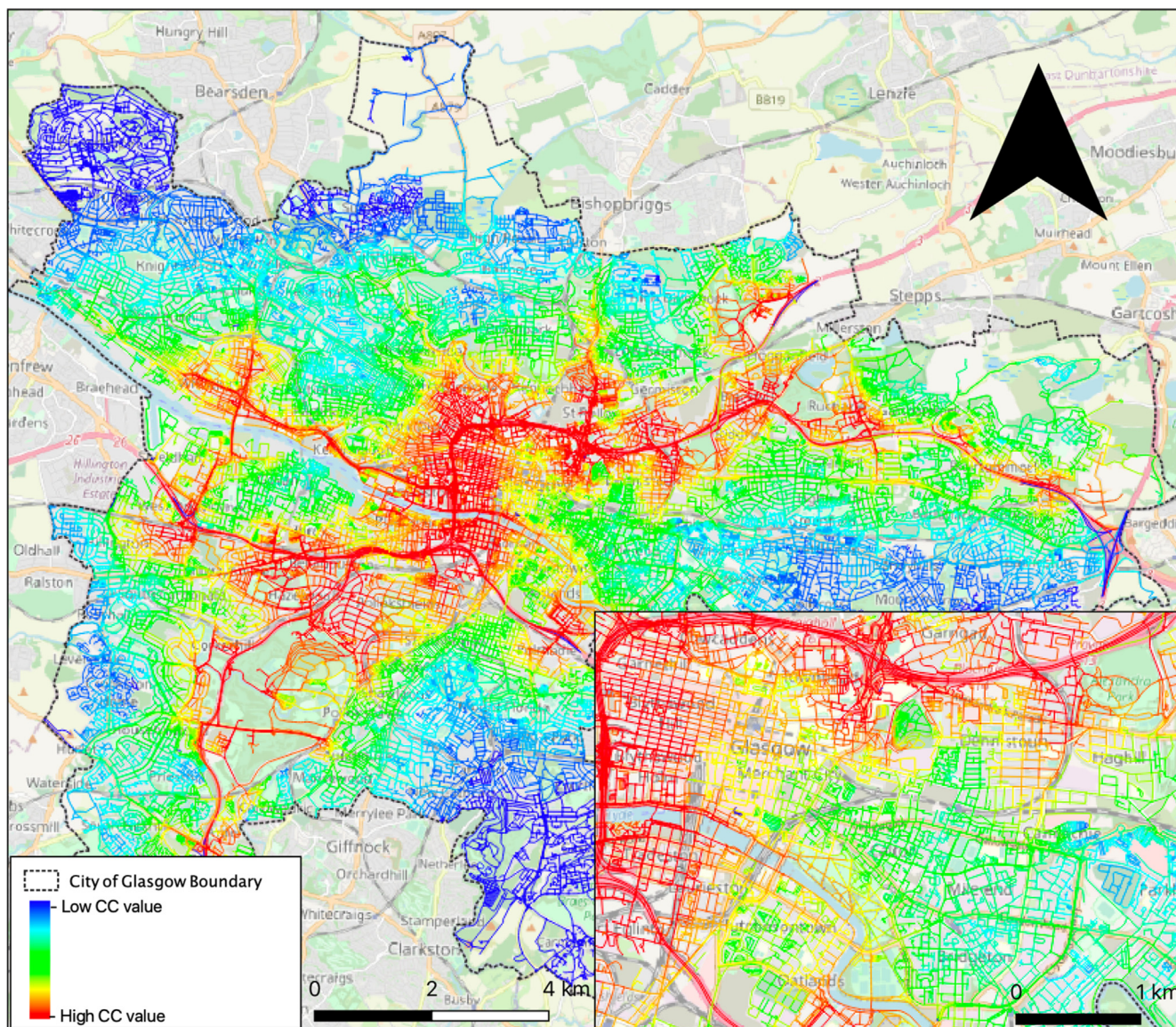


Fig. 11. Glasgow street network CC values. [Basemap: OpenStreetMap, licensed under the Open Database License (ODbL)].

Table 5
Spatial dependence diagnostics.

		Moran's I (Error)	LM (Lag)	Robust LM (Lag)	Lagrange Multiplier (Error)	Robust LM (Error)
CCT2018	MI/DF	0.44	1	1	1	1
	VALUE	64.00	3510.90	45.25	3961.25	495.60
	P-value	<0.001	<0.001	<0.001	<0.001	<0.001
CCT2017	MI/DF	0.50	1	1	1	1
	VALUE	75.51	5011.90	91.64	5691.90	771.64
	P-value	<0.001	<0.001	<0.001	<0.001	<0.001

lighting (Olsen et al., 2016). The correlation between CCT and CC indicates that cyclists are more likely to use accessible streets (i.e. streets that are topologically close to other streets within the network) rather than deep streets. In contrast, streets associated with low CC values are not easily accessible by many cyclists (Hillier and Sahbaz, 2008). This suggests that the infrastructure and bicycle facilities situated in the core of the city, including the city centre, will be more effective than those within the outskirts of the city, which is in agreement with McArthur and Hong (2019).

7. Conclusion

The current study proposes the use of a spatial model of CCT in the City of Glasgow to better understand cyclists' route choices and to identify optimal location for cycling facilities. The study overcomes limitations of previous research by: i) using a crowdsourced cycling dataset obtained from Strava; ii) deploying OSMnx to calculate street network centrality indices; and iii) accounting for the spatial dependence by using spatial errors models.

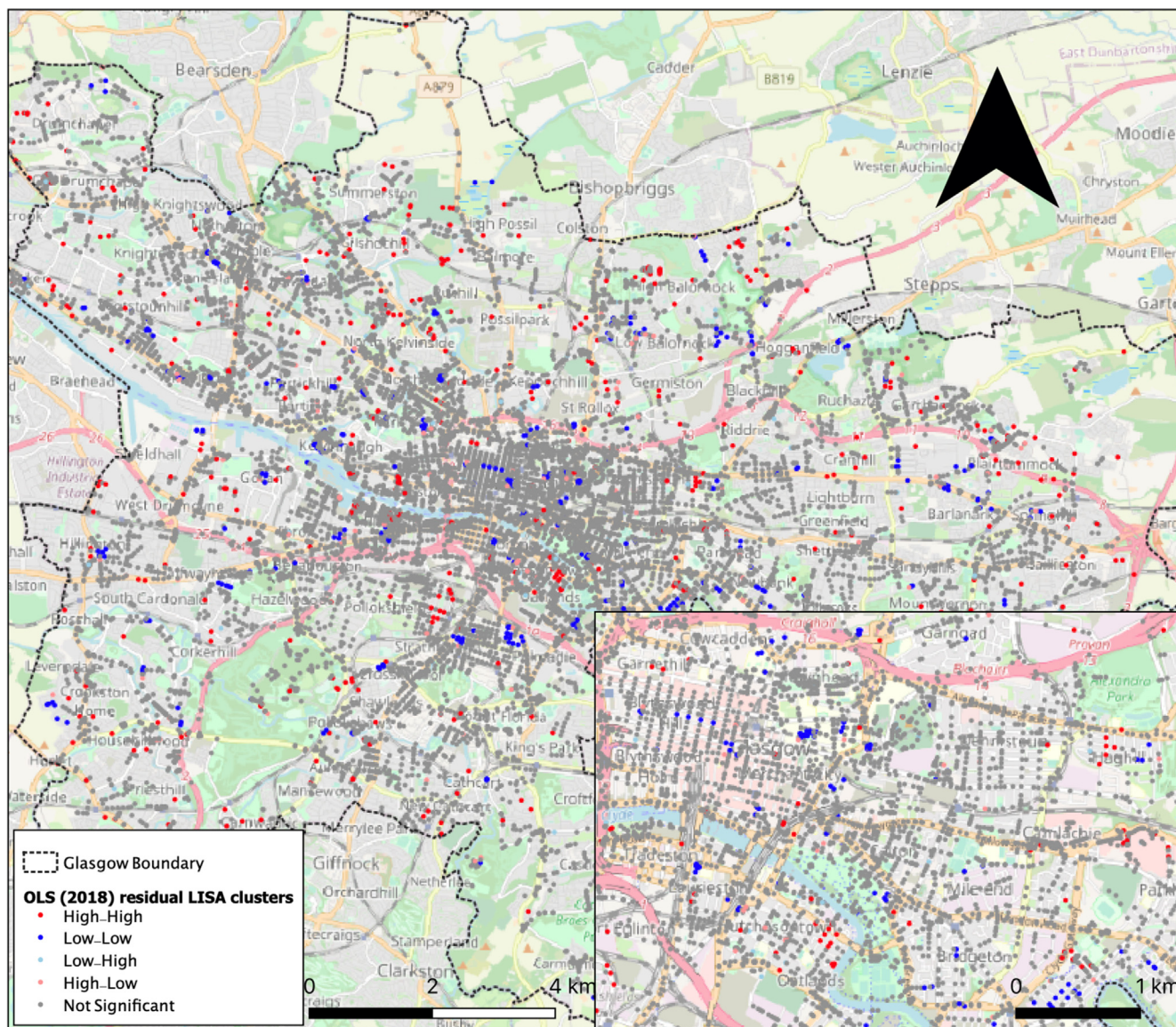


Fig. 12. OLS (2018) residual LISA clusters. [Basemap: OpenStreetMap, licensed under the Open Database License (ODbL)].

CCT was found to be significantly positively correlated with streets of high CC, EC, and BC values, and negatively correlated with those of DC values. This study can be applied to the planning of future street network configuration. For example, in order to increase ridership in grid pattern streets, Green Wave systems can be applied. Further, our results indicate that streets with high CC values should be prioritized for cyclist infrastructure since CC exhibits the highest coefficient of correlation with our proposed model.

8. Limitations

This study, however, is not without limitations. Strava datasets may be subjected to the self-selection bias that occurs when participants include themselves into a group, resulting in a nonprobabilistic sample. This may explain the overwhelming predominance of two segments, both age-wise (users aged between 25 and 54 years) and gender-wise (males), in the Strava dataset, supporting the

claim that females are under-represented, while fitness-focused users are over-represented (Hochmair et al., 2019; Sun, 2017). Our findings represent the demographic composition found in Strava datasets, which are to a large extent in line with the demographic composition seen in previous studies (McPherson, 2017; Motherwell, 2018).

8.1. Directions for future research

Future research might consider the following directions: (1) Extending the present study by incorporating different street network centrality indices, such as angular indices through the UrbanFormPy python package (Filomena et al., 2019); (2) modelling pedestrians' route choice using crowdsourced data, which is currently in its infancy stage (Griffin et al., 2014; Lee and Sener Ipek, 2020); or 3) considering additional mechanisms for gathering Strava-equivalent data from the under-represented groups identified above.

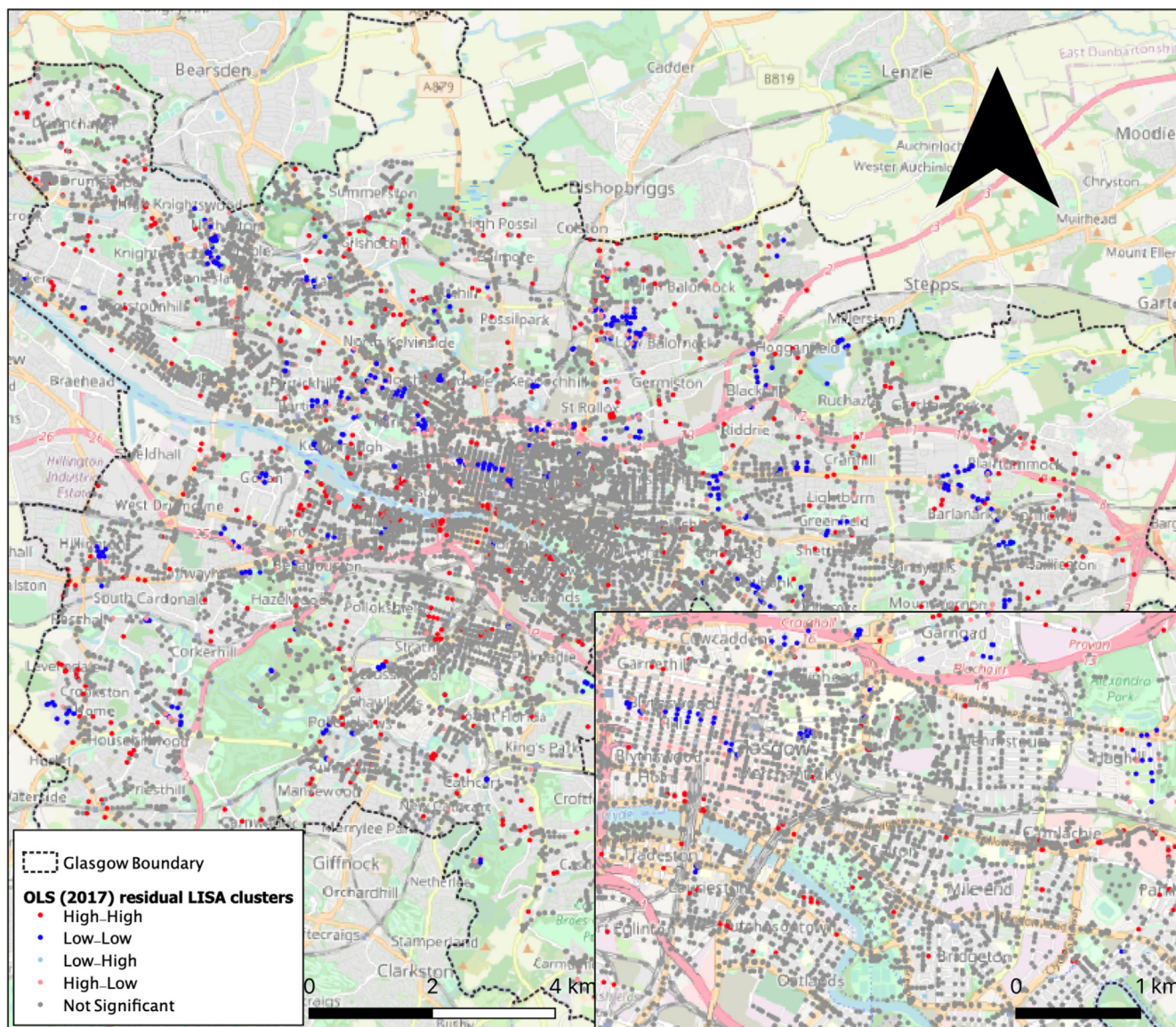


Fig. 13. OLS (2017) residual LISA clusters. [Basemap: OpenStreetMap, licensed under the Open Database License (ODbL)].

Table 6
Model selection criteria.

	Spatial Lag Model		Spatial Error Model	
	2017	2018	2017	2018
R ²	0.49	0.40	0.53	0.43
LogL	-13992	-11274.4	-13736.31	-11110.43
AIC	27996.1	22560.8	27482.6	22230.9
SC	28041.1	22605.4	27520.1	22,268

R² = Goodness of fit; LogL = Log likelihood; AIC = Akaike info criterion; SC = Schwarz criterion.

9. Data sources

9.1. Strava

9.1.1. Citation

Strava Inc. Economic and Social Research Council. Strava Metro data - Scotland, Glasgow, Manchester, Tyne and Wear [data collection]. University of Glasgow - Urban Big Data Centre.

9.1.2. Mandatory attribution

Data Licensed by Strava.

9.2. OSMnx

9.2.1. Citation

Boeing, G. 2017. "OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks." *Computers*,

Table 7
Spatial error model results.

Dependent variable: LogCCT 2018		R ² = 0.42	Number of observations: 12,359		
Variable	Coefficients	Standard error	z-value	P-Value	
CONSTANT	4.53	0.35	12.8	<0.001	
LogDC	-0.15	0.05	-2.61	<0.01	
LogBC	0.28	0.007	37.93	<0.001	
LogEC	0.02	0.002	7.25	<0.001	
LogCC	1.15	0.15	7.40	<0.001	
λ	0.55	0.007	71.11	<0.001	
Dependent Variable: LogCCT 2017		R ² = 0.53	Number of observations: 13,350		
Variable	Coefficients	Standard Error	z-value	P-Value	
CONSTANT	5.51	0.41	13.13	<0.001	
LogDC	-0.40	0.06	-6.08	<0.001	
LogBC	0.35	0.008	44.32	<0.001	
LogEC	0.02	0.003	8.38	<0.001	
LogCC	1.70	0.19	8.86	<0.001	
λ	0.62	0.006	91.33	<0.001	

Environment and Urban Systems 65, 126–139. doi: <https://doi.org/10.1016/j.compenvurbsys.2017.05.004>.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Al-Shaheen, Q.E., 2012. *The Role of Integrated and Diverse Retailing in the Production of Social Vitality: Reshaping Retailing in Kuwait City*. Heriot-Watt University.
- AMEC E&I and Sprinkle Consulting I, 2011. Pedestrian and Bicycle Data Collection. Available at: http://ahtd.info/yahoo_site_admin/assets/docs/Pedestrian_Bicycle_Data_Collection_FHWA_Dec.2011.14274024 (accessed 6 May 2019).
- Anselin, L., 1995. Local indicators of spatial association—LISA. *Geograph. Anal.* 27 (2), 93–115. Wiley Online Library.
- Anselin, L., 2013. *Spatial Econometrics: Methods and Models*. Springer Science & Business Media.
- Ashurst, N., 2008. The Investigation, Repair and Conservation of the Doulton Fountain, Glasgow Green.
- Avila-Palencia, I., Panis, L.I., Dons, E., et al., 2018. The effects of transport mode use on self-perceived health, mental health, and social contact measures: a cross-sectional and longitudinal study. *Environ. Int.* 120, 199–206. Elsevier.
- Boeing, G., 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput., Environ. Urban Syst.* 65, 126–139. <https://doi.org/10.1016/j.compenvurbsys.2017.05.004>.
- Boeing, G., 2019a. The morphology and circuitry of walkable and drivable street networks. In: *The Mathematics of Urban Morphology*. Springer, pp. 271–287.
- Boeing, G., 2019b. Urban spatial order: Street network orientation, configuration, and entropy. *Appl. Network Sci.* 4(1). Springer: 67.
- Boeing, G., 2020. Street Network Models and Indicators for Every Urban Area in the World. arXiv preprint arXiv:2009.09106.
- Braçe, O., 2016. Study of the effects of urban morphology on physical activity. *Revista de Estudios Andaluces* 33, 5–7.
- Brennan, M.J., Carroll, T.M., 1987. *Preface to Quantitative Economics & Econometrics*. South-Western Pub.
- Burkey, M.L., 2018. Spatial econometrics and GIS YouTube playlist. REGION 5 (3), R13–R18. <https://doi.org/10.18335/region.v5i3.254>.
- Chuai, X., Huang, X., Wang, W., et al., 2012. Spatial econometric analysis of carbon emissions from energy consumption in China. *J. Geogr. Sci.* 22 (4). Springer: 630–642.
- Conrow, L., Wentz, E., Nelson, T., et al., 2018. Comparing spatial patterns of crowdsourced and conventional bicycling datasets. *Appl. Geogr.* 92 (January). <https://doi.org/10.1016/j.apgeog.2018.01.009>. Elsevier Ltd: 21–30.
- Cycling Scotland, 2019. Annual Cycling Monitoring Report. Available at: <https://www.cycling.scot/mediaLibrary/other/english/6353.pdf> (accessed 4 June 2020).
- Dalton, R., 2001. The secret is to follow your nose: Route path selection and angularity.
- Day, C.M., Premachandra, H., Bullock, D.M., 2016. Rate of pedestrian signal phase actuation as a proxy measurement of pedestrian demand. *Transp. Res. Record*.
- Department for Transport, 2019. National Travel Survey: England 2018. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/823068/national-travel-survey-2018.pdf (accessed 12 August 2020).
- Department of Transport and Main Roads, 2017. Queensland State of Cycling Report. Available at: <https://cabinet.qld.gov.au/documents/2017/Jul/Cycling/Attachments/Report.pdf> (accessed 18 May 2019).
- Dill, J., 2009. Bicycling for transportation and health: the role of infrastructure. *J. Public Health Policy* 30 (1). Springer: S95–S110.
- ECMT, 2000. Safety in road traffic for vulnerable users. In: Paris, 2000. Available at: https://www.itf-oecd.org/sites/default/files/docs/00vulner_0.pdf.
- Emmanouilidis, T., 2013. Urban network analysis. centrality, sinuosity and shortcut detection. *Revue internationale de géomatique* 23 (3–4), 431–443. <https://doi.org/10.3166/riq.23.431-443>.
- Emo, B., Hoelscher, C., Wiener, J., et al., 2012. Wayfinding and spatial configuration: evidence from street corners. PUC.
- Enström, R., Netzell, O., 2008. Can space syntax help us in understanding the intraurban office rent pattern? Accessibility and rents in downtown Stockholm. *J. Real Estate Finance Econ.* 36 (3). Springer: 289–305.
- Filomena, G., Versteegen, J.A., Manley, E., 2019. A computational approach to ‘The Image of the City’. *Cities* 89, 14–25. <https://doi.org/10.1016/j.cities.2019.01.006>.
- Fitch, D.T., Handy, S.L., 2020. Road environments and bicyclist route choice: the cases of Davis and San Francisco, CA. *J. Transp. Geogr.* 85. <https://doi.org/10.1016/j.jtrangeo.2020.102705>.
- Frank, L., Kerr, J., Chapman, J., et al., 2007. Urban form relationships with walk trip frequency and distance among youth. *Am. J. Health Promotion* 21 (4, suppl). SAGE Publications Sage CA: Los Angeles, CA: 305–311.
- Freeman, L.C., 1977. A set of measures of centrality based on betweenness. *Sociometry*, 35–41. JSTOR.
- Glasgow City Council, 2018. City Centre cycle journeys more than double in less than a decade. Available at: <https://www.glasgow.gov.uk/article/23567/City-Centre-cycle-journeys-more-than-double-in-less-than-a-decade> (accessed 3 June 2019).
- Grabow, M.L., Spak, S.N., Holloway, T., et al., 2012. Air quality and exercise-related health benefits from reduced car travel in the Midwestern United States. *Environ. Health Perspect.* 120 (1), 68–76. DOI: 10.1289/ehp.1103440.
- Graser, A., Straub, M., Dragaschnig, M., 2015. Is OSM good enough for vehicle routing? A study comparing street networks in Vienna. In: *Progress in Location-Based Services 2014*. Springer, pp. 3–17.
- Griffin, G., Nordback, K., Götschi, T., et al., 2014. Monitoring bicyclist and pedestrian travel and behavior: current research and practice. *Transp. Res. Circ.* (E-C183).
- Griffin, G.P., Jiao, J., 2015. Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *J. Transp. Health* 2 (2), 238–247. <https://doi.org/10.1016/j.jth.2014.12.001>.
- Hagberg, A., Schult, D., Swart, P., 2018. NetworkX Reference Release 2.2. Tech. rep.
- Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plann. B: Plann. Design* 37(4). SAGE Publications Sage UK: London, England: 682–703.
- Hamer, M., Chida, Y., 2008. Active commuting and cardiovascular risk: a meta-analytic review. *Prev. Med.* 46 (1), 9–13. <https://doi.org/10.1016/j.ypmed.2007.03.006>.
- Hanson, J., 1989. Order and structure in urban design: the plans for the rebuilding of London after the Great Fire of 1666. *Ekistics*. JSTOR: 22–42.
- Hillier, B., 2012. The city as a socio-technical system: a spatial reformulation in the light of the levels problem and the parallel problem. In: *Digital Urban Modeling and Simulation*. Springer, pp. 24–48.
- Hillier, B., Iida, S., 2005. network effects and psychological effects: a theory of urban movement. In: *International Conference on Spatial Information Theory*, 2005, pp. 475–490.
- Hillier, B., Sahbaz, O., 2008. An evidence based approach to crime and urban design, or, can we have vitality, sustainability and security all at once. Bartlett School of Graduates Studies University College London.
- Hillier, B., Burdett, R., Peponis, J., et al., 1987. Creating life: or, does architecture determine anything? *Arch. Comportement/Arch. Behav.* 3 (3), 233–250.
- Hillier, B., Penn, A., Hanson, J., et al., 1993. Natural movement: or, configuration and attraction in urban pedestrian movement. *Environ. Plann. B: Plann. Design* 20(1). SAGE Publications Sage UK: London, England: 29–66.

- Hochmair, H.H., Zielstra, D., Neis, P., 2015. Assessing the completeness of bicycle trail and lane features in OpenStreetMap for the United States. *Trans. GIS* 19 (1). Wiley Online Library: 63–81.
- Hochmair, H.H., Bardin, E., Ahmouda, A., 2019. Estimating bicycle trip volume for Miami-Dade county from Strava tracking data. *J. Transp. Geogr.* 75. Elsevier: 58–69.
- Hong, J., Shankar, V.N., Venkataraman, N., 2016. A spatially autoregressive and heteroskedastic space-time pedestrian exposure modeling framework with spatial lags and endogenous network topologies. *Anal. Methods Accid. Res.* 10, 26–46. <https://doi.org/10.1016/j.amar.2016.05.001>.
- Hong, J., McArthur, D.P., Livingston, M., 2018. The Evaluation of Large Cycling Infrastructure Investments in the Glasgow Clyde Valley Planning Area Before, During, and After the Commonwealth Games: The Use of Crowdsourced Data (Strava).
- Hong, J., McArthur, D.P., Livingston, M., 2019. The evaluation of large cycling infrastructure investments in Glasgow using crowdsourced cycle data. *Transportation*. <https://doi.org/10.1007/s11116-019-09988-4>.
- Hong, J., McArthur, D.P., Stewart, J.L., 2020. Can providing safe cycling infrastructure encourage people to cycle more when it rains? The use of crowdsourced cycling data (Strava). *Transp. Res. A: Policy Pract.* 133. Elsevier: 109–121.
- Jacobsen, P.L., 2015. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prev.* 21 (4). BMJ Publishing Group Ltd: 271–275.
- Jestic, B., Nelson, T., Winters, M., 2016. Mapping ridership using crowdsourced cycling data. *J. Transp. Geogr.* 52, 90–97. <https://doi.org/10.1016/j.jtrangeo.2016.03.006>.
- Joss, C., 2016. Network Theory: An Overview. Available at: https://drive.google.com/file/d/0Bv0Dvld2_-lneXdUdGtZl81UjQ/view (accessed 4 October 2019).
- Kerr, J., Frank, L., Sallis, J.F., et al., 2007. Urban form correlates of pedestrian travel in youth: differences by gender, race-ethnicity and household attributes. *Transp. Res. D Transp. Environ.* 12(3), 177–182. DOI: 10.1016/j.trd.2007.01.006.
- Kim, H.-K., Sohn, D.W., 2002. An analysis of the relationship between land use density of office buildings and urban street configuration: case studies of two areas in Seoul by space syntax analysis. *Cities* 19 (6). Elsevier: 409–418.
- Kim, Y.O., Penn, A., 2004. Linking the spatial syntax of cognitive maps to the spatial syntax of the environment. *Environ. Behav.* 36(4). Sage Publications: 483–504.
- Koohsari, M.J., Owen, N., Cerin, E., et al., 2016. Walkability and walking for transport: characterizing the built environment using space syntax. *Int. J. Behav. Nutr. Phys. Act.* 13 (1), 121. <https://doi.org/10.1186/s12966-016-0448-9>.
- Lee, K., Sener, I.N., 2017. Emerging Data Mining for Pedestrian and Bicyclist Monitoring: A Literature Review Report.
- Lee, K., Sener Ipek, N., 2020. Emerging data for pedestrian and bicycle monitoring: sources and applications. *Transp. Res. Interdisc. Perspect.* Elsevier: 100095.
- Lee, K., Sener, Ipek Nese, 2020. Strava Metro data for bicycle monitoring: a literature review. *Transp. Res. Rev.* 1–21. DOI: 10.1080/01441647.2020.1798558.
- Lin, Z., Fan, W., “David”, 2020. Bicycle ridership using crowdsourced data: ordered probit model approach. *J. Transp. Eng. A: Syst.* 146 (8). American Society of Civil Engineers: 4020076.
- Liu, Z., Song, Z., Chen, A., et al., 2016. Exploring Bicycle Route Choice Behavior with Space Syntax Analysis. Report TRCLC 15-13.: 30p. Available at: https://wmich.edu/sites/default/files/attachments/u883/2017/TRCLC_RR_15_13_0.pdf%0Ahttps://ntl.bts.gov/lib/61000/61700/61773/TRCLC_RR_15_13_0.pdf%0Ahttps://trid.trb.org/view/1467671
- Lu, W., Scott, D.M., Dalumpines, R., 2018. Understanding bike share cyclist route choice using GPS data: comparing dominant routes and shortest paths. *J. Transp. Geogr.* 71, 172–181. <https://doi.org/10.1016/j.jtrangeo.2018.07.012>.
- McArthur, D.P., Hong, J., 2019. Visualising where commuting cyclists travel using crowdsourced data. *J. Transp. Geogr.* 74, 233–241. <https://doi.org/10.1016/j.jtrangeo.2018.11.018>.
- McCahill, C., Garrick, N.W., 2008. The applicability of space syntax to bicycle facility planning. *Transp. Res. Record* 2074(1). SAGE Publications Sage CA: Los Angeles, CA: 46–51.
- McLean, K., 2017. Protection Through Identity: Improving Cycling Through Design Informed by Biomimicry, Cognition, and Perception. *figshare*.
- McPherson, K., 2017. Glasgow's public cycle hire scheme. Available at: https://www.gcp.co.uk/assets/0000/6010/Next_bike_hire_data.pdf (accessed 12 April 2020).
- Menghini, G., Carrasco, N., Schüssler, N., et al., 2010. Route choice of cyclists in Zurich. *Transp. Res. A: Policy Pract.* 44 (9). Elsevier: 754–765.
- Minge, E., Falero, C., Lindsey, G., et al., 2017. Bicycle and pedestrian data collection manual. Minnesota Department of Transportation.
- Mobasheri, A., 2020. An introduction to open source geospatial science for urban studies. In: *Open Source Geospatial Science for Urban Studies*. Springer, pp. 1–8.
- Mobasheri, A., Sun, Y., Loos, L., et al., 2017. Are crowdsourced datasets suitable for specialized routing services? Case study of OpenStreetMap for routing of people with limited mobility. *Sustainability* 9(6). Multidisciplinary Digital Publishing Institute: 997.
- Mohamed, A.A., Van Nes, A., Salheen, M.A., 2015. Space and protest: a tale of two Egyptian squares. In: *SSS10: Proceedings of the 10th International Space Syntax Symposium*, London, UK, 13–17 July 2015, 2015.
- Motherwell, S., 2018. Are we Nearly there yet? Exploring Gender and Active Travel. Available at: <https://www.sustrans.org.uk/media/2879/2879.pdf> (accessed 31 May 2020).
- Musakwa, W., Selala, K.M., 2016. Mapping cycling patterns and trends using Strava Metro data in the city of Johannesburg, South Africa. *Data Brief* 9, 898–905. <https://doi.org/10.1016/j.dib.2016.11.002>.
- Newman, M.E.J., 2006. The mathematics of networks. Retrieved.
- Nordström, T., Manum, B., 2015. Measuring bikeability: space syntax based methods applied in planning for improved conditions for bicycling in Oslo. In: *Proceedings of the Tenth International Space Syntax Symposium*, London: University College London, paper, 2015.
- O'Dea, J.A., 2003. Why do kids eat healthful food? Perceived benefits of and barriers to healthful eating and physical activity among children and adolescents. *J. Am. Dietetic Assoc.* 103 (4). Elsevier: 497–501.
- Olsen, J.R., Mitchell, R., Oglivie, D., 2016. Effects of new motorway infrastructure on active travel in the local population: a retrospective repeat cross-sectional study in Glasgow, Scotland. *Int. J. Behav. Nutr. Phys. Act.* 13(1). BioMed Central: 77.
- Orellana, D., Guerrero, M.L., 2019. The influence of Space Syntax on cycling movement in Manta, Ecuador. In: *12th International Space Syntax Symposium*, 2019.
- Penn, A., Hillier, B., Banister, D., et al., 1998. Configurational modelling of urban movement networks. *Environ. Plann. B Plann. Design* 25(1). SAGE Publications Sage UK: London, England: 59–84.
- Pritchard, R., 2018. Revealed preference methods for studying bicycle route choice—a systematic review. *Int. J. Environ. Res. Public Health* 15(3). Multidisciplinary Digital Publishing Institute: 470.
- Pucher, J.R., Buehler, R., 2012. *City Cycling*. MIT Press Cambridge, MA.
- Raford, N., Chiaradia, A.J., Gil, J., 2007. Space Syntax: the role of urban form in cyclist route choice in Central London. *SafeTrec Res. Reports*. 18. Available at: <http://escholarship.org/uc/item/8qz8m4fz>
- Read, S., 1999. Space syntax and the Dutch city. *Environ. Plann. B Plann. Design* 26(2). SAGE Publications Sage UK: London, England: 251–264.
- Rifaat, S.M., Tay, R., de Barros, A., 2012. Urban street pattern and pedestrian traffic safety. *J. Urban Design* 17(3). Taylor & Francis: 337–352.
- Rissel, C.E., 2009. Active travel: a climate change mitigation strategy with co-benefits for health. *New South Wales Public Health Bull.* 20(2). CSIRO: 10–13.
- Rousell, A., Zipf, A., 2017. Towards a landmark-based pedestrian navigation service using OSM data. *ISPRS Int. J. Geo-Inf.* 6(3). Multidisciplinary Digital Publishing Institute: 64.
- Roy, A., Nelson, T.A., Fotheringham, A.S., et al., 2019. Correcting bias in crowdsourced data to map bicycle ridership of all bicyclists. *Urban Sci.* 3(2). Multidisciplinary Digital Publishing Institute: 62.
- Samson, N.P., 2017. The socio-spatial relations of the accessibility of Parks IN Chicago. In: *Proceedings of the 11th Space Syntax Symposium*, 2017, p. 14.
- Snizek, B., Nielsen, T.A.S., Skov-Petersen, H., 2013. Mapping bicyclists' experiences in Copenhagen. *J. Transp. Geogr.* 30. Elsevier: 227–233.
- Stinson, M.A., Bhat, C.R., 2003. Commuter bicyclist route choice: analysis using a stated preference survey. *Transp. Res. Record* 1828 (1). SAGE Publications Sage CA: Los Angeles, CA: 107–115.
- Stinson, M.A., Bhat, C.R., 2004. Frequency of bicycle commuting: internet-based survey analysis. *Transp. Res. Record* 1878(1). SAGE Publications Sage CA: Los Angeles, CA: 122–130.
- Strauss, J., Miranda-Moreno, L.F., Morency, P., 2013. Cyclist activity and injury risk analysis at signalized intersections: a Bayesian modelling approach. *Accid. Anal. Prev.* 59, 9–17. <https://doi.org/10.1016/j.aap.2013.04.037>.
- Strava Metro, 2015. Strava Metro Comprehensive User Guide Version 2.0. Available at: https://www.ubdc.ac.uk/media/1323/stravametro_200_user_guide_withoutpics.pdf (accessed 26 June 2019).
- Strava Metro, 2019. Big data for better cities. Available at: <https://metro.strava.com> (accessed 25 October 2019).
- Sun, Y., 2017. Exploring potential of crowdsourced geographic information in studies of active travel and health: Strava data and cycling behaviour.
- Sun, Y., Mobasheri, A., 2017. Utilizing crowdsourced data for studies of cycling and air pollution exposure: a case study using Strava data. *Int. J. Environ. Res. Public Health* 14(3), 274. DOI: 10.3390/ijerph14030274.
- Sun, Y., Du, Y., Wang, Y., et al., 2017. Examining associations of environmental characteristics with recreational cycling behaviour by street-level Strava data. *Int. J. Environ. Res. Public Health* 14(6). Multidisciplinary Digital Publishing Institute: 644.
- Sun, Y., Moshfeghi, Y., Liu, Z., 2017b. Exploiting crowdsourced geographic information and GIS for assessment of air pollution exposure during active travel. *J. Transp. Health* 6, 93–104. <https://doi.org/10.1016/j.jth.2017.06.004>.
- Venter, Z., Barton, D., Gundersen, V., et al., 2020. Urban nature in a time of crisis: recreational use of green space increases during the COVID-19 outbreak in Oslo, Norway. *Environmental Research Letters*. IOP Publishing.
- Yamada, I., 2016. Thiessen polygons. In: *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*. Wiley Online Library, pp. 1–6.
- Yen, Y., Zhao, P., Sohail, M.T., 2019. The morphology and circuitry of walkable, bikeable, and drivable street networks in Phnom Penh, Cambodia. *Environ. Plann. B Urban Anal. City Sci.* SAGE Publications Sage UK: London, England: 2399808319857726.