Journal of Urban Mobility 9 (2026) 100211

Contents lists available at ScienceDirect

Journal of Urban Mobility

journal homepage: www.elsevier.com/locate/urbmob

Leveraging Strava Metro Data to enhance urban cycling infrastructure development in Brussels

Sara Tori [a,1,*], Floriano Tori [b,1], Vincent Ginis [b,c,d]

[a] Mobilise Research Group, House of Sustainable Transitions, Vrije Universiteit Brussel, Pleinlaan 2, 1050, Brussels, Belgium
[b] Data Analytics Laboratory, Vrije Universiteit Brussel, Pleinlaan 2, 1050, Brussels, Belgium
[c] Applied Physics Group, Vrije Universiteit Brussel, Pleinlaan 2, 1050, Brussels, Belgium
[d] School of Engineering and Applied Sciences, Harvard University, Cambridge, USA

ARTICLE INFO

Keywords:
Cycling infrastructure
Missing links
Crowdsourced data
Strava metro
Neural networks

ABSTRACT

Cycling has seen a rising popularity over the past decade, booming even more since the COVID-19 pandemic. To accommodate this increase, appropriate infrastructure needs to be developed, requiring knowledge of traffic volumes, which can be challenging for active mobility. For cycling, one can overcome this through the use of automated or manual bike counters, but these counts are often only performed at strategic locations throughout cities. An additional interesting source of information can come from crowdsourced data which can fill the spatial gaps of counters. Our work explores how Strava Metro data can be used to estimate daily cyclist volumes in Brussels, Belgium, and guide new infrastructure development. We train a neural network to learn the relationship between Strava edge activity and automated counts, considering counter location within the region. We further augment the input Strava data with additional features, such as weather or day of the week. The trained model is then tasked to predict counts of a hypothetical counter on high-volume Strava edges without separated cycling infrastructure. This allows us to evaluate cycling volumes on these edges on equal footing with automated counter volumes. Our analysis identifies two categories of infrastructure gaps—(i) “missing links” where short discontinuities interrupt otherwise protected corridors, and (ii) “independent high-demand links” without separated cycling infrastructure. Our analysis and methodology offers a reproducible and robust approach to provide insights into cycling flows in cities, allowing us to identify locations for new cycling infrastructure.

1. Introduction

The current mobility system is one that is characterised by heavy car-dependence worldwide (Lewis & Grande del Valle, 2019; Mattioli et al., 2020; Urry, 2004). However, cities around the world are also increasingly prioritising cycling, in part due to the strong increase in cycling levels witnessed during and after the COVID-19 pandemic (Buehler & Pucher, 2024). Cycling, as a transportation mode, could help solve congestion and air pollution issues, and travelling actively further provides a number of health benefits (Fishman et al., 2015). However, for cycling to become a full-fledged alternative to individual motorized transport, traffic safety issues need to be addressed (Buehler & Pucher, 2024; DiGioia et al., 2017). When asked about their preferences, cyclists often express their preference for separated facilities (Broach et al., 2012; Winters & Teschke, 2010). Vulnerable road users specifically, such as elderly people, women, and those cycling with kids, place high values on separated road infrastructure (Hardinghaus & Weschke, 2022). In both stated and revealed preference environments, cyclists will take a non-direct, longer route if it allows them to be on separated infrastructure and not mixed with traffic (Broach et al., 2012; Hood et al., 2011; Krenn et al., 2014; Tilahun et al., 2007). To help further promote the uptake of cycling as a transportation mode, cities therefore need to implement bicycle-friendly policies and infrastructure, as a lack infrastructure is an important deterrent for people to cycle (McDonald, 2008). Identifying where protected cycling infrastructure should be installed to benefit cyclist is therefore a core planning challenge for policymakers.

Data-driven decision making is crucial in helping cities transition towards sustainability (Provost & Fawcett, 2013). Yet the implementation of bicycle-friendly policies often runs into the issue of a lack of

---

* Corresponding author.
E-mail address: sara.marie.tori@vub.be (S. Tori).
[#] These authors have contributed equally and share first authorship.

https://doi.org/10.1016/j.urbmob.2026.100211
Received 22 October 2025; Received in revised form 20 March 2026; Accepted 24 March 2026
Available online 4 April 2026
2667-0917/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

(high-quality) data on the subject (Broach et al., 2012; DiGioia et al., 2017). Bicycle route choice research has often been dominated by stated preference research, asking cyclists to rank or state their preference for different types of infrastructure (Broach et al., 2012). However, this does not provide policy makers with a complete picture of cycling needs. Often, real-time data across the city is missing, with cycling volumes usually being counted at limited fixed locations across the city (Roy et al., 2019), either by year-round counters, or by punctual manual counts. These, while a good first indication, fail to provide a full picture of cycling volumes in cities, since cyclists could cycle daily but never cross a counter on their route. Additionally, automated counters can be defect and can momentarily stop counting (El Esawey, 2018b), and they are expensive to install (Janež et al., 2022), although recent technological advancements have made the technology cheaper and more reliable. As a complement to the data derived from the counters, crowdsourced data is increasingly being employed, coming from apps such as Strava or other health apps (Lee & Sener, 2021; Nelson, Ferster et al., 2021, 2021). Crowdsourced data is user generated data (Xu & Nyerges, 2017), and can be an example of citizen science (Eitzel et al., 2017). Using this user generated data is less expensive than equipping research participants with GPS devices, providing an interesting data source (Broach et al., 2012).

For cycling specifically, Strava Metro data has the potential to be used for a variety of purposes (Lee & Sener, 2021). Strava Metro provides its users with anonymized and aggregated data on cycling and pedestrian activity (Strava, n.d.). The service launched in 2014, and since then, the dataset has been extensively used by researchers. Lee and Sener (2021) showed that Strava data correlates strongly with official counts, making it fit for research purposes, with Strava data representing around 7% of total bicycle volumes (Fischer et al., 2022). Lee and Sener (2021) documented five areas in which Strava Metro data have been applied to cycling research: i) to identify various travel patterns (Broach et al., 2012), ii) to estimate travel demand (Jestico et al., 2016), iii) to analyse route choice (McArthur & Hong, 2019; Orellana & Guerrero, 2019), iv) to evaluate infrastructure (Heesch et al., 2016; Hong et al., 2020), and v) to control for exposure in crash models (Sanders et al., 2017). Additionally it has also been used to deduce optimal counter location through ridership stratification (Brum-Bastos et al., 2019).

While popular in use, representativeness and sampling bias are known issues related to the use of Strava data (Venter et al., 2023). Contributors to Strava are more likely to be male (Jestico et al., 2016; Watkins et al., 2016), and belonging to an age group between 25 and 44 (Heesch & Langdon, 2016), with a focus also on fitness tracking rather than utilitarian usage of the app (Fischer et al., 2022). Moreover, the relationship between Strava counts and total cycling volumes varies spatially with socio-economic context and land use (Conrow et al., 2018; Livingston et al., 2021), meaning that a single global correction is insufficient for extrapolating from Strava data. These limitation related to data representativeness might however be addressed to some extent by fusing multiple datasets (Lee & Sener, 2021) or by using prediction that incorporate socio-economic factors (Dadashova et al., 2020).

Translating the spatially rich but biased crowdsourced Strava data into planning-relevant estimates of cycling demand remains an open challenge. Studies that fuse Strava with (manual) counter data have focused on improving volume predictions (Jean-Louis et al., 2024; Jestico et al., 2016), but do not explicitly connect these estimates to infrastructure decisions. Concurrently, topological approaches have been developed to identify gaps in cycling infrastructure from connectivity criteria alone instead of crowdsourced data to inform infrastructure planning. Such methods characterize levels of traffic stress (Furth et al., 2016) or prioritize missing links via betweenness centrality (Vybornova et al., 2023). However, these methods do not incorporate observed cycling volumes, meaning they cannot distinguish gaps that carry high demand from those that do not.

This paper makes two contributions to approach the challenges of fusing both Strava and topological approach to infrastructure planning. The first is extrapolating from Strava to total cycling volumes while taking the heterogeneity of cycling volumes across a city into account. For this, we train a neural network architecture that disentangles shared effects (such as temporal and meteorological effects) from location-specific scaling, allowing the model to accommodate heterogeneous cycling patterns across the city without the need for detailed spatial socio-economic covariates. The second is applying these predicted volumes to inform cycling infrastructure planning, grounding gap identification in measured activity rather than pure network topology. For this we first build a framework that identifies corridors with high cycling demand but an absence of separated infrastructure through Strava data. We subsequently integrate our neural network architecture to produce counter-equivalent estimates of cycling volumes on these corridors.

We apply our framework to Brussels, Belgium. Since the COVID-19 pandemic, cycling has known an important uptake in Brussels (de Séjournet et al., 2022; Pro Velo, 2025), yet the main barrier identified by these “new” cyclists in the city is a feeling that cycling in Brussels is too dangerous, paired with not feeling at ease cycling in traffic (Pro Velo, 2025). In order to foster the ongoing uptake in cycling and to encourage yet more people cycling, secure infrastructure therefore needs to be prioritised.

The remainder of this article is structured as follows: Section 2 provides an overview of the study area (Brussels), as well as the data and methodology employed in the research. Section 3 details our results, while Section 4 presents a discussion and conclusion of the results obtained, together with some limitations and recommendations for further research.

2. Materials and methods

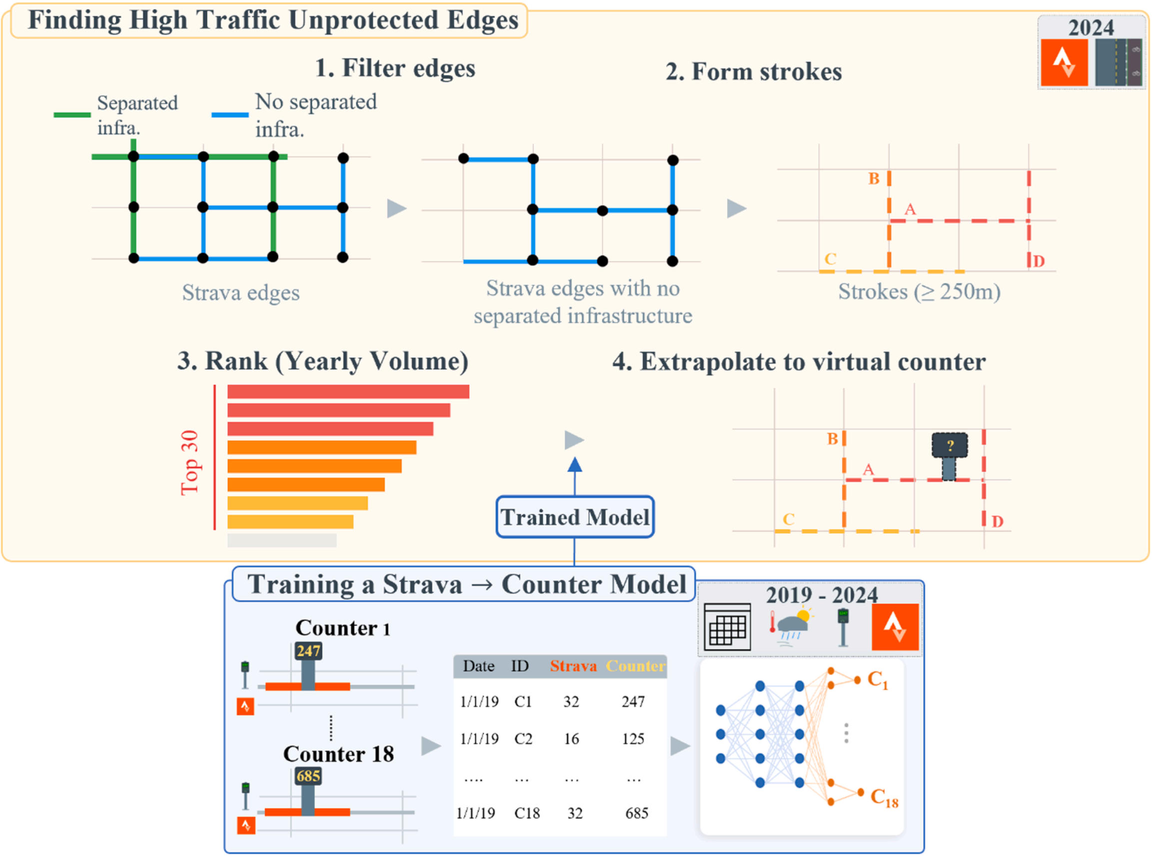
2.1. *Methodology and data collection*

The goal of our research was to provide a systematic framework to inform cycling infrastructure planning. Our research methodology is split into two stages (see Fig. 1). In the first stage, we identify high-volume cycling corridors in the Brussels-Capital Region that do not have separated cycling infrastructure. Infrastructure status is determined using the latest available infrastructure dataset, corresponding to the end-2024 network configuration. The resulting set therefore captures corridors with high cycling volumes during 2024 that still lacked separated cycling infrastructure at the end of 2024. The corridors highlight gaps in the current separated infrastructure network. In the second stage, we train a neural network to predict daily cycling volumes recorded by automated counters based on Strava counts. The training data for the model consists of matched automated-counter and Strava observations from 2019 to 2024. Once trained, we apply it to the previously identified high-volume corridors. The predictions made for these corridors are interpreted as cycling volumes a counter would have reported had one been present on these segments, enabling a direct comparison with volumes recorded by automated counters elsewhere in the city. Our analysis provides an empirical basis for prioritising the installation of separated cycling infrastructure on the identified corridors.

In this work we focus on weekdays, as cycling on weekdays is more strongly linked to utilitarian (commuting) activities than activities on Saturdays and Sundays (Miranda-Moreno et al., 2013). We focus on commuting as the goal is to improve infrastructure in such a way that cycling becomes an obvious alternative to the use of private cars. Additionally, commuting and non-commuting activities show different trip patterns (Lee & Sener, 2019).

2.1.1. *Automated counter data*

Automated counter data is available from the BCR and consists of 18 counters spread over different municipalities. The counters are located on bi- or monodirectional cycling lanes and vary in time of first activation. Only 8 were active on the starting date of January 1st 2019. For



Date	ID	Strava	Counter
1/1/19	C1	32	247
1/1/19	C2	16	125
....	...	...	...
1/1/19	C18	32	685

Fig. 1. Overview of the research methodology. The upper right corner of each panel denotes the data sources used for that process. The upper panel shows the corridor identification pipeline. Strava edges with separated infrastructure are filtered out (step 1) such that the remaining edges are aggregated into continuous strokes (step 2). The strokes of at least 250 m are retained and subsequently ranked by cumulative yearly volume (step 3). The Strava cycling volumes for the top 30 strokes are extrapolated to automated counter level volumes using the trained model (step 4). The lower panel shows the neural network training process. We use matched Strava and automated counter data from 2019 to 2024, augmented with weather and temporal features, to train the network. The architecture consists of shared layers and location-specific heads allowing the network to tailor its prediction based on the location of the edge.

each counter, the counter data, namely how many bikes have passed by, is updated every 15 min and is independent of the direction of travel. However, we note that automated counters do not solely register cyclists, as other forms of micromobility (e.g. e-scooters) are also counted. Cycling volumes in the BCR vary depending on the area of the city. In order to capture these differences, we use the counter data as a proxy for the cycling volume in a specific region. The cycling tendencies in a region are therefore associated with the cycling volumes measured by the automated counter. To create the regions associated with a counter we use Voronoi Tessellation and subdivided the BCR into 18 regions based on the location of the automated counters (see Fig. 2). The regions 1 and 7, as well as the region 9 and 13 originate from distinct counters on opposite side of the road. As can be seen on Fig. 2, counters are spread out over the BCR, with counters both located in the centre of the city (e.g. counters 18, 5, 7, 11), as well as counters located on the outskirts of the region to capture commuting cyclists originating from outside of the BCR (e.g. counters 16, 12, 3).

Data Cleaning The quarterly data of each counter was summed in order to obtain a daily count. Additionally, we removed Saturdays, Sundays and holidays from the dataset to reduce the influence of recreational cyclists to a minimum. Finally, days with a count of 0 after the initial activation of the counter were removed from the dataset as these were likely due to the counter not operating instead of reflecting real cyclist flows. In total we obtain 20,168 datapoints over all 18 counters. In Appendix A, Table 1, we highlight the per counter distribution of these datapoints.

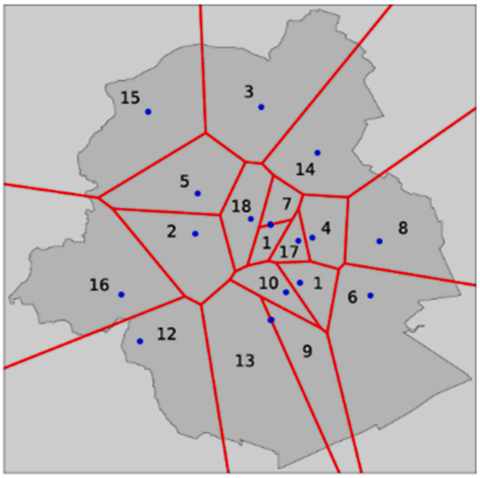


Fig. 2. Locations of the counters (blue dots) together with their Voronoi regions (red lines). Counters are both present in the centre region of the BCR, as well as on the outskirts of the region.

2.1.2. *Strava metro data*

Strava Metro data was provided in the form of daily directional counts of trips on Strava edges for the complete BCR. Strava edges do not represent complete streets, but directed network segments used within the Strava Metro data model to aggregate activity counts. A single street may therefore correspond to multiple Strava edges, for example because it is split at intersections, divided by direction of travel, or segmented according to the underlying reference network used by Strava (Strava Metro, 2025). While Strava also provides demographic information on the counts (age and gender) we did not take this information into account as the focus of our analysis lies on cycling flows in general. While the representativity of Strava data has increased during COVID-19 (Fischer et al., 2022), it can still remain limited (Lee & Sener, 2021; Venter et al., 2023) such that further categorisation of the data could lead to non-significant results. Counts on Strava edges is split into recreational vs. commuting (this is indicated by users when recording activities), and we considered both categories for our analysis as this distinction cannot be made for the automated counters. Similarly as for the counters we removed Saturdays, Sundays and holidays from the dataset. For each automated counter we associated Strava segments that represented the same location (in the sense that a Strava user being counted on the Strava edge would have been counted by the automated counter). The procedure used was automatic, based on distance, and we manually controlled the final selection. Further details can be found in Appendix B. For some counters, multiple segments could be identified to match with a counter location. To obtain the Strava volume of the counter we combined the counts from all the selected edges.

Data Cleaning The Strava edges associated with automated counters represent on average 6.7% of the cycling volumes of the counters, with a median percentage of 4.8%. This is in line with previous findings of 7% (Fischer et al., 2022). However, Strava daily counts and daily counts from the automated counters show a Pearson correlation coefficient between 0.66 and 0.88 (see Fig. 3). Looking at the correlation coefficient per counter for each year individually we find that in three instances ('CEK18': 2019, 'CB1143': 2020, 'CB1101': 2022) the correlation index is lower than 0. Closer inspection revealed that this is due to the counter malfunctioning during an extended period of time and we therefore removed these specific years for those datasets from the analysis. While two counter-Strava segment matches only reach a correlation of 0.60 (CEK049) and 0.53 (CVT387), we find that other counters obtain similar correlations as previously reported values. These values are 0.882 for daily correlation between Strava segments and manual counts in Glasgow (Livingston et al., 2021), and monthly count correlations of 0.76 to 0.96 in Ottawa-Gatineau (Boss et al., 2018), between automated counters and Strava segments. These findings highlight that, while Strava data does not represent absolute cycling volumes, it can provide information about cycling behaviour.

2.1.3. *Weather data*

Weather conditions are important determinants for cycling usage (El Esawey, 2018a; Hankey & Lindsey, 2016). We use opensource data from the Royal Meteorological Institute of Belgium[1] to include weather data into our analysis. We consider daily features such as "average temperature (°C)", "total precipitation quantity (mm)" and " average wind speed (in m/s) at 10 m altitude". This gives us a view of daily weather quality which can impact cycling decisions. The Royal Meteorological institute provides weather data of its 13 automated weather stations over all Belgium. For our purpose we select the weather station located in Uccle, Brussels. However, this weather station does not provide information on the average wind speed. To obtain this information we locate the three nearest weather stations from Brussels (which are towards the East, North and South-West of Brussels) and take the mean of the measurements reported in these weather stations to obtain an estimate of the average wind speed for Brussels.

2.1.4. *Bicycle infrastructure data*

Bicycle infrastructure in Brussels can be classified into two categories: separated and non-separated from motorised traffic. Using the BCR mobility dataset,[2] we obtain specific information on all bicycle infrastructure in the region. This dataset, provided as coordinate data, is not directly linked to specific Strava edges. The most recent update of the infrastructure data was in December 2024 and we use this version for all further analysis.

Data Cleaning To analyse areas in Brussels with high cycling traffic but lacking separated infrastructure, we assign an infrastructure label to Strava segments. This label indicates whether the road segment includes separated infrastructure. We attribute the separated label to any bicycle infrastructure denoted as "Tweerichtingsfietspad *(Two-way bike lane)*", "Eenrichtingsfietspad *(One-way bike lane)*", "Eenrichtings gedeeld voet/fietspad *(One-way shared foot/cycle path)*", "Tweerichtings gedeeld voet/fietspad *(Two-way shared foot/cycle path)*", 'Fiets-/voetgangersweg *(Bicycle/pedestrian road')*" or "Bosweg (*Forest path)*" within the dataset. These labels correspond to roads where cycling infrastructure is physically separated from motorised traffic in contrast with infrastructure designated by paint, which we exclude. As highlighted in Section 2.1.2, Strava edges do not correspond to complete streets, but rather to smaller directed segments. Consequently, infrastructure information which is provided at the street-level by the BCR dataset must be transferred to the corresponding set of Strava edges. This means that multiple Strava edges may inherit the same infrastructure label from a single street-level record. Because both the Strava data and the infrastructure dataset do not align perfectly at the segment level, a simple GPS-based match would have produced unreliable assignments. We therefore applied a more robust procedure described in Appendix C.

2.1.5. *Study area description- Brussels, Belgium*

Our research took place in Brussels, the capital city of Belgium (pop. 1.1 million (StatBel, 2022)). Cycling in the city has been on the rise over the past two decades (Henry et al., 2020), with the COVID-19 pandemic marking an acceleration point in cycling volumes (de Séjournet et al., 2022). Data from the latest travel behaviour survey indicates that cycling now accounts for 9% of the modal split in Brussels (Samyn & Lagrou, 2025). However, as highlighted in the introduction, this data originates from fixed counters (either manual or automated), but little is known about the volume and movements of cyclists outside of those fixed counters. This knowledge gap can therefore be bridged using crowdsourced data to complement it.

One of the contributing factors to the increase in cycling volumes in the city is the increase in cycling infrastructure (Vandenbulcke et al., 2011), which has evolved significantly since 2012 (see Fig. 4). In total in 2021, the Brussels Capital Region (BCR) counted 513 km of cycling infrastructure (L'Observatoire Good Move, 2021). In 2013, only 1,5% of the total road surface area of the city was equipped with cycling infrastructure (Brandeleer et al., 2016). During the COVID-19 pandemic, 38, 5 km of new, temporary cycling infrastructure was implemented, to provide Brussels citizens with more accessible public space (BXI, 2021). The majority of this infrastructure has been made permanent since then.

When looking at the map of current cycling infrastructure, it becomes clear that major axes in the city still need new (indicated in red in Figure) or improved (indicated in orange in Figure) infrastructure. In 2025, only 36% of the 'vélo PLUS' cycle network, which is a structural network for cycling throughout the city, had been realised, complemented by an additional 31% of completed cycling infrastructure that supports this structural network (L'Observatoire Good Move, 2021).

Yet as highlighted in the introduction, infrastructure that accommodates the needs of active mode users is an essential element of sustainable urban mobility transitions. Appropriate infrastructure contributes to safety feelings of cyclists, which is a major factor influencing cycling behaviour (Gössling & McRae, 2022). It is therefore crucial to further understand the routes that cyclists take, to have

---

[1] https://www.meteo.be/en/belgium

[2] https://data.mobility.brussels/mobigis/

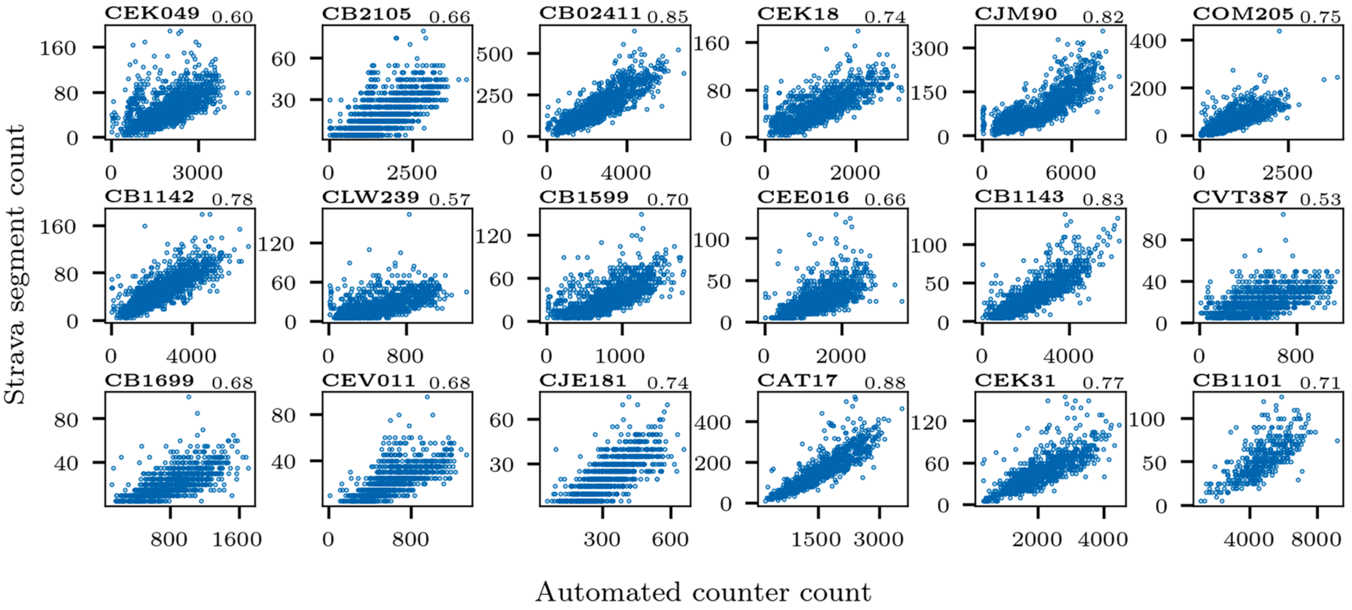


Fig. 3. Counter data and Strava data on the same location are well correlated with each other. Counts provided by the automated counters (x) vs the counts provided by the Strava segments (y). Each dot represents one day. In the right corner of each plot we report the Pearson correlation coefficient between both data sources.

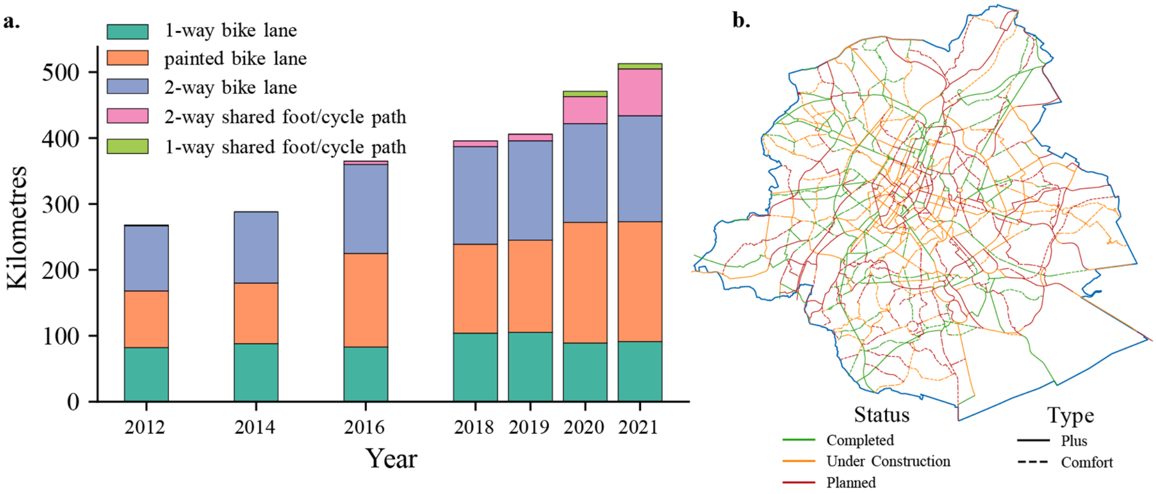


Fig. 4. a. Evolution of the kilometres of cycling infrastructure between 2012–2021 (reprinted from L'Observatoire Good Move, 2021). Dark green is a separated 1-way cycling lane, orange are painted (1-way) cycling lanes, mauve is a separated 2-way cycling lane, pink is a 2-way shared cycling and pedestrian lane and light green is 1-way shared cycling and pedestrian lane . For the two-way infrastructure the kilometres are counted doubled (as reported by the Brussels Mobility department). **b. Map of Brussels cycling infrastructure**. Red lines indicate bike lanes to be constructed, orange indicates bike lanes currently under construction and green are existing bike lanes (reprinted from L'Observatoire Good Move, 2021).

insights into which parts of the city are considered as safe to cycle on.

2.1.6. Machine learning for cycling volume predictions

Machine learning approaches have increasingly been applied to estimate bicycle volumes from crowdsourced data. Miah et al. (2023) provide a comparative study of modelling techniques for predicting bicycle flows where both shallow and deep neural networks were shown to produce higher accuracies in estimating daily bicycle volumes. Artificial neural networks, together with random forest methods, were also found to outperform traditional methods when predicting hourly cycling volumes based on crowdsourced data (Kwigizile et al., 2022). However, these studies typically train a single global model for all locations, meaning their accuracy gains most likely derive from flexible function.

A central challenge is that daily count magnitudes vary substantially across the 18 automated counters (see Section 2.1.2), reflecting the spatial heterogeneity of cycling activity across the city. Prior approaches address this spatial heterogeneity either through global models that apply a single coefficient set across all locations (Jean-Louis et al., 2024; Jestico et al., 2016), or through geographically weighted regression (GWR) (Dyason et al., 2025; Munira & Sener, 2020; Wei et al., 2021) which allows coefficients to vary spatially but requires continuous spatial covariates at all prediction locations. GWRs provide an improvement over traditional global regression models by enabling a more nuanced interpretation of the factors influencing cycling activity.

To accommodate this spatial heterogeneity while leveraging the improved capacity of neural networks, we draw from the field of multi-task learning, specifically through the use of shared representations (Caruana, 1993). We frame the task of predicting cycling flows across the entire BCR as 18 related but distinct tasks. We therefore implement a neural network architecture (see Fig. 5) with shared layers and

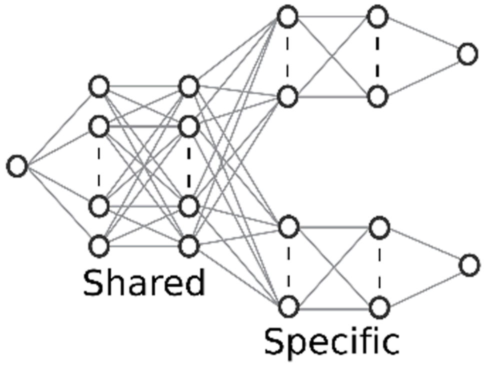


Fig. 5. Our network architecture is composed of shared layers and location specific layers to account for variations in location.

task-specific layers with output heads for each of the 18 automated counters. The shared layers learn globally transferable patterns, primarily meteorological and temporal effects, while each task-specific head learns the location-dependent scaling required to map from the shared representations to that counter's specific cycling volume and Strava penetration rate. This allows one network to predict cycling volumes over a region with very heterogeneous cycling patterns. Additionally, our multi-head neural network architecture addresses spatial heterogeneity without requiring enhanced spatial feature engineering required when constructing regression models. Unlike global models or GWR, which can in principle generate predictions at any location within the study area, our multi-head architecture is inherently bounded by the spatial region defined through its task-specific heads: the model can only predict cycling volumes within regions associated with one of the 18 counters it was trained on. However, the architecture can extends naturally as additional counters are deployed, requiring only limited further training for the new counter-specific head.

The shared layers of the model use features designed to capture the temporal and meteorological dynamics of cycling activity. The model input is therefore, next to Strava counts, the daily weather conditions and additional temporal information (the day of the week, the month and the year as well as if the given day is a public holiday). Specifically we provide the following features to the model:

- Strava Counts
  - We provide Strava counts of the day we want to predict and the counts of the previous 7days
  - Encoding: We normalise the values using min/max normalisation per counter location.
- Weather conditions: Average temperature (°C), Precipitation quantity (mm) and Wind speed (m/s)
  - Encoding: We normalise quantities with z-score normalisation.
- Temporal data: Weekday, Month and Year
  - Encoding: Year is normalised using min/max normalisation. Weekdays (Monday to Friday) are attributed a numerical value from 0 to 4 and subsequently encoded using sin and cos to conserve their cyclic nature. Months are similarly attributed to a numerical value between 1 and 12 and encoded with sin and cos.
  - We provide a binary value to indicate if the given day is a public holiday.

3. Results

3.1. Obtaining high-volume Strava strokes

The aim of our analyses was to see how the use of Strava Metro data as a complement to automated bike counters can help inform cycling infrastructure investments in the BCR. We therefore aimed to identify roads within the BCR which lack separated cycling infrastructure while currently accommodating a high cycling volume. As we possess infrastructure data from 2024 onwards we used Strava data for that year to assess cycling volumes. To identify these regions of high volumes and no separated infrastructure, we first removed all Strava edges with non-zero daily counts in 2024 which were attributed a separated infrastructure label as described in Section 2.1.4. Details about this matching are further described in Appendix C.

The edges that remain in the dataset correspond to Strava edges with no separated infrastructure. We then ranked these edges based on their yearly cumulative cycling counts over the entire span of the 2024 Strava dataset (which contains data up until 2024/09/30). From this ranking we then identified the most trafficked Strava edges. However, Strava edges can be short in length. The goal of our analysis was to focus on longer stretches of road which should accommodate new separated cycling infrastructure. To transform Strava edges to usable longer strokes we used the coins algorithm (Tripathy et al., 2021) from the momepy software package (v. 0.8.0) to calculate natural continuous traffic strokes from the most trafficked edges. The strokes formed by the coins algorithms are a combination of Strava edges. The edges are selected to represent a continuous direction.

From the strokes obtained, we only selected strokes with a length of 250 m or more. From these, we selected top 30 strokes of length longer than 250 m that have the largest flux of cyclists. Fig. 6a displays the geographical location of all 30 strokes within the Brussels Capital Region, colour coded with the total number of activities recorded on these strokes on Strava. The constructed strokes are composed of multiple Strava edges, with different aggregated cycling counts. To attribute a cycling volume to the strokes, we take the median value of the Strava edges that constitute the stroke. Fig. 6b shows the distribution of cycling counts within the identified strokes. For most strokes, except strokes with index 1 and 15, we find that the constituent Strava segments are consistent in count with each other.

From our Fig. 6a we, we can identify two important types of high-traffic strokes obtained from our analysis:

1. Missing Links: These are segments without separated infrastructure that interrupt otherwise continuous facilities. The identified strokes therefore connect two roads that already contain separated infrastructure. Examples of such strokes are strokes with ids 3, 85, or 106. The large volumes on these strokes could be partly due to cyclists choosing to use the network of separated infrastructure, instead of a shorter path with lesser quality infrastructure.
2. Independent High-Traffic Segments: These strokes are roads that do not connect two existing parts of the separated infrastructure network. Example of such strokes are stroke with ids 40, 53, 107, 99 or 45. These strokes often connect to or originate from separated infrastructure.

Missing links (strokes with ids: 3, 8, 29, 39, 52, 67, 77, 83, 86, 87, 91, 94, 108, 123) as can be seen in Fig. 6a have a total length of 9.70 km, while other identified strokes total 12.48 km. Proportional to the total existing separated infrastructure (379.80 km) this represents 2.55% and 3.29%. These proportions indicates that missing link infrastructure can represent a quick win in terms of separated infrastructure to be installed.

3.2. Estimated high-traffic counts

Due to the limited representativity of Strava data, we now aim to understand the cycling volumes recorded on the identified strokes within the context of cycling volumes in Brussels recorded by automated counters. We therefore aim to extrapolate, from the Strava counts on the strokes identified above, the counts a hypothetical counter would have registered had it been installed at that location (on separated infrastructure).

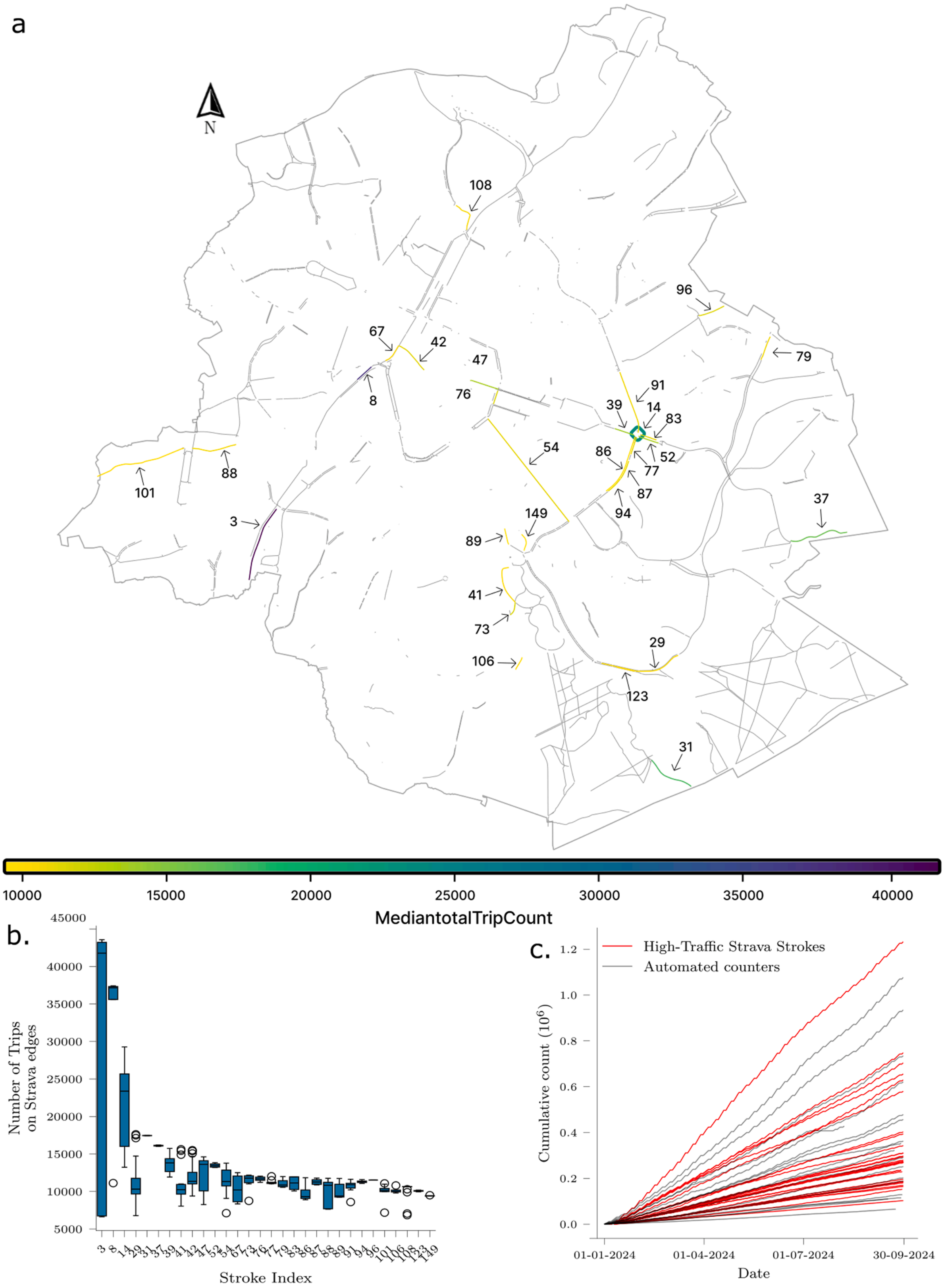


Fig. 6. Obtaining high-traffic Strava strokes and extrapolating their counts. The identified high-traffic are widespread over the BCR and could account for a similar volume of cyclists as measured by automated counters. **a.** The 30 identified high-traffic strokes with a length of >250 m. The color coding indicates the median total number of trips on the high-traffic stroke. We also show the North arrow for direction **b.** Strokes are composed of multiple Strava edges. For all but three strokes the spread in total counts over the constituent edges is fairly limited. **c.** When extrapolated using a neural network we find that the identified high traffic strokes (red) rival the automated counters (black) in cycling volumes during the year 2024.

3.2.1. Training a neural network

To train our network we use the matched Strava and counter data between 2019 and 2024 (see Section 2.1.2). The target of the network is to predict, given the trips recorded by Strava on the segment associated with the counter, the number reported by the automated counter. The data that we use consists of all the days with a valid cyclists count reported by the automated counters. We train the network on data between 2019 and 2024. We split the dataset temporally, using the first 90% as training data and the final 10% as test data. We make this 90–10 temporal split for each of the 18 counters. This means that not all counters receive the same absolute number of training and testing samples, as not all counters contain the same number of daily counts. We use the test set to evaluate how well our model predicts counter data based on Strava counts. Our final goal is to apply the trained model to the set of high-traffic strokes we identified in Section 3. We select the hyperparameters for our final model based on a hyperparameter sweep. The optimal set of parameters were a dropout of 0.004 and a learning rate of 0.004 with the AdamW (Loshchilov & Hutter, 2019) optimizer. The common part of the network architecture (the shared layers) consist of 3 layers with a width of 246, while the counter specific heads contain 3 layers of width 4.

Our final model obtains a Mean Absolute Error (MAE) of 515 cyclists (with a Mean Absolute Percentage Error of 24.66%) when aggregating over all counters. This performance varies across the different counters (with a minimum of 18.51% for region 4 and maximum of 36.09% in region 7). A full description of the error per counter can be found in Appendix D, Fig. 9.

3.2.2. Extrapolating high-traffic strokes

Once trained, we apply our network to the high-traffic strokes which do not have separated infrastructure. The predictions of the network aim to simulate the counts a hypothetical counter at this location would have provided. These counts therefore offer a more realistic portrayal of cycling volumes through these high-traffic strokes. To predict these counts, we apply our trained model on the edges that constitute the strokes, together again with weather and temporal features. Each edge is assigned to the counter region in which it falls, as determined by the Voronoi tessellation of the BCR (see Fig. 2). This approach allows us to make location specific predictions, but assumes that the relationship between Strava counts and total cycling volumes, as learned at the counter location, generalises to other edges within the same region. The network then produces predictions for that edge using the location-specific layers of the assigned region. After predicting the counts for each edge we aggregate, using a median, the daily prediction counts of each edge that constitute the high-traffic strokes This aggregated value then represents the predicted count for the stroke. In Fig. 6c we show how the volumes predicted for the high-traffic stroke (in red) compared to the volumes registered by the automated counters. We note that these volumes are comparable to volumes measured by the counters, indicating that real cycling volumes on the identified high-traffic strokes are of similar significance.

It is important to note that we do not estimate the global cycling volumes when extrapolating and do not aggregate the counts over all strokes. Cyclists may use multiple of the identified strokes in their travel journey (and will therefore be accounted for multiple times across different Strava edges), which could distort a global cycling volume prediction. Instead, our goal is to compare the cycling volumes of the identified strokes lacking infrastructure to volumes that are currently recorded by automated counters ensuring that the identified strokes are compared on equal footing.

4. Discussion and conclusion

In this paper, we developed a framework that leverages Strava Metro cycling data in order to identify segments in the BCR where separated cycling infrastructure is likely undersupplied. Because automated counters are globally sparse and manual counts are temporally sparse, Strava cycling data can provide rich spatial-temporal cycling volumes but, due to its limited number of users, does not reflect total volumes. In our study, we show that Strava data contains similar information on cycling behaviour and can provide more fine-grained insights into cycling volumes, making this an interesting data source for policy makers. Building on this relation, we use Strava data to identify corridors in the BCR where additional higher-quality, separated infrastructure would accommodate large cycling volumes. Our analysis translates the Strava counts on the high-traffic strokes into counter-equivalent daily volumes using a trained neural network. This offers a representative point of comparison for the cycling volumes on these segments.

We demonstrated that the combined use of official counter data and Strava Metro data allows for the implementation of evidence-based decision making, helping to provide a more comprehensive overview of cyclists in Brussels. Our results allowed us to concretely identify both "missing link" infrastructure, which can be seen as quick wins to connect existing separated infrastructure, as well as areas where currently no infrastructure exists but which are used by a high number of cyclists. By looking at patterns for existing cyclists, our research supports policy to encourage potential cyclists for whom the lack of safe infrastructure constitutes a barrier (Broach et al., 2012; Hardinghaus & Weschke, 2022). This type of results can therefore be used to increase evidence-based decision making, providing decision makers with cheaper and more reliable data when compared to automated counters (El Esawey, 2018a; Janež et al., 2022). This type of analysis can be used by decision-makers to prioritise cycling infrastructure investments in the BCR (distinguishing potential missing-link 'quick win' investments from independent high-demand corridors). Importantly, however, our analysis gives insight into directionality when it comes to cycling in the BCR: while we identify strokes that would benefit from separated cycling infrastructure, we do not claim that cycling infrastructure should be installed exactly on that specific street segment, as we did not look at feasibility in terms of urban planning. If it is not feasible for the identified strokes to accommodate separated cycling infrastructure, our results can indicate to decision-makers what flows are used by cyclists and, as such, infrastructure on a parallel trajectory would be equally satisfactory.

Importantly, for other cities, our analysis can be replicated using locally available data. This includes data from either fixed counters or manual counts, as well as Strava Metro data (accessible upon agreement with Strava). Our approach can be replicated using our code, available on Github. However, the neural network used in our analysis needs to be re-trained using local data.

Our work also includes some limitations. The first one is intrinsically linked to the lack of representativeness of Strava data, specifically considering lower socio-economic population groups (Venter et al., 2023). It therefore becomes important to use multiple data sources to correct for this, especially when the analyses are used to inform decision making. Linked to this is the substantial overrepresentation of commuters from outside of the BCR that we see in our dataset. This phenomenon reflects the intrinsic bias related to Strava data, namely the fact that it represents a subgroup of the cycling population. This is also visible in our Fig. 6, as a large proportion of the most populous Strava cycling routes are routes used by cyclists commuting from outside of the BCR. The counters with the highest observed daily volumes, such as CB1101 (see Fig. 3), are counters localised in the centre of Brussels and therefore register much more than solely (external Brussels) commuting cyclists. While this is a limitation of Strava data, we do accommodate this discrepancy when using a neural network to infer total daily volumes from Strava data, as we allow the model to adapt its prediction based on the region the edge is located, as is described in Section 3.2.1. A last limitation could also be linked to the route choice of cyclists: it could be that some cyclists take a certain route which is a bit longer but which has better infrastructure (Broach et al., 2012), as opposed to the shortest route which is worse off in terms of cycling infrastructure. This could bias the results and not show street segments that could benefit even more from separated cycling infrastructure.

For future work, we believe that including additional features into the model could be an interesting addition, such as topological information or traffic conditions. In our work, we used distance to associate regions of the BCR to automated counters, allowing us to account for difference in cycling flows throughout the city. It could however be interesting to associate edges based on cycling patterns for this. Additionally, our method is compatible with integrating other, manual counts, which could further increase the robustness of the model's predictions. It would furthermore be interesting to analyse the temporal evolution of cycling volumes before and after interventions that modify the available infrastructure. Finally, using results such as ours also requires clarity about the objective policy makers want to achieve. These objectives can be more than simply expanding cycling capacity, but could, for example, also focus explicitly on reducing accident risks. Integrating these trade-offs will contribute to a robust development of cycling infrastructure and allow the development of urban mobility's future trajectories.

Funding

Published with the support of the Universitaire Stichting of Belgium

Data availability

Code and public data are available on Github: https://github.com/FloTori/strava-for-cycling-infrastructure.

CRediT authorship contribution statement

Sara Tori: Writing – review & editing, Writing – original draft, Validation, Project administration, Conceptualization. **Floriano Tori:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Vincent Ginis:** Writing – review & editing, Validation, Supervision, Methodology.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Sara Tori reports financial support was provided by University Foundation Belgium. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Strava for providing access to Strava Metro. This paper includes aggregated and de-identified data from Strava Metro. The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.

F.T. would like to thank Brecht Verbeken and Andres Algaba for the stimulating discussions on the methodology.

Appendix

A. Data Overview

Table 1
Overview of number of useable days with data per counter.

Counter	ID	Number of Registered days	Counter	ID	Number of Registered days
CEK048	1	1235	CEE016	10	1337
CB2105	2	980	CB1143	11	1204
CB02411	3	1470	CVT387	12	680
CEK18	4	1230	CB1699	13	1072
CJM90	5	1497	CEV011	14	1065
COM205	6	1373	CJE181	15	920
CB1142	7	1380	CAT17	16	1018
CLW239	8	1146	CEK32	17	927
CB1599	9	1180	CB1101	18	454

B. Matching Strava edges to counters

We used the geographical information of the counters from Brussel Mobility which provide the longitude and latitude position of the counters. In order to find the relevant Strava segment for the counter we first identify the closest segment based on distance. For some counters this resulted in wrongly identified Strava edges, as the counter true location differed from the provided information. We therefore manually checked all counters and selected the true relevant edges. For some counters we also identified parallel edges which could also, given a margin on the gps detection of cyclists by Strava, have been counted by the counter. We therefore consider the summed counts of these parallel edges to obtain a total count.

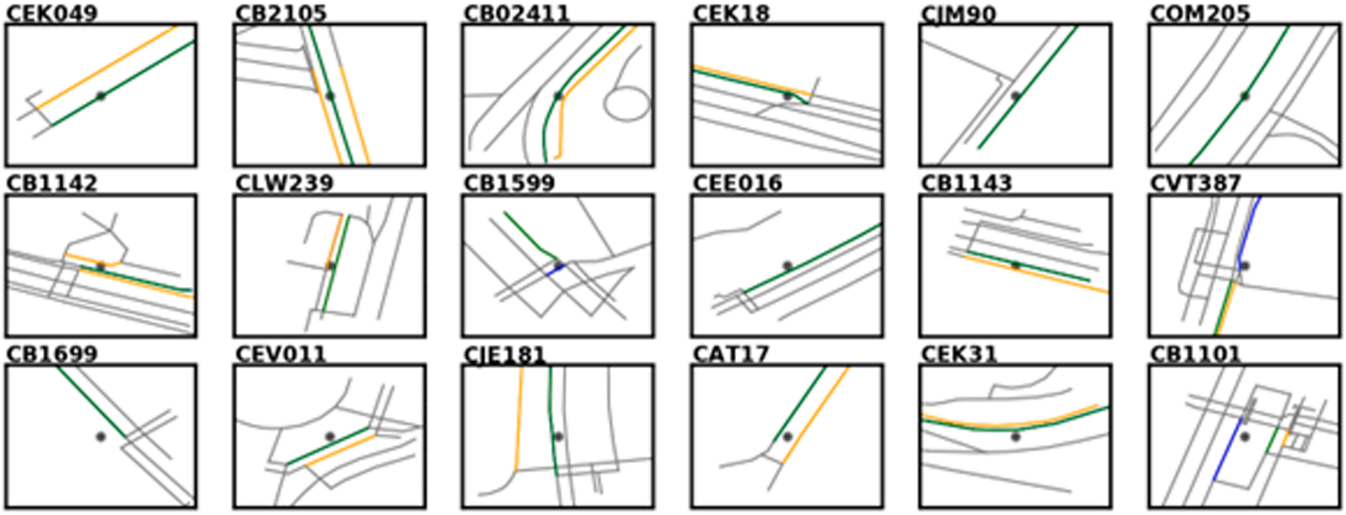


Fig. 7. Strava edges were selected using a combination of automatic and manual selection. The edges identified edges proximity are designated in blue. Edges ultimately selected as the main edges are designated in green and edges identified as potential parallels in orange. For most counter blue and green edges overlap, and we only display the green edge.

C. Automatically assigning infrastructure label to Strava segments

We use the dataset from Brussel Mobility which contains the coordinates of all cycling infrastructure in Brussels, both separated and non-separated from motorized traffic. We downsize this dataset by only considering the separated infrastructure. For our analysis we robustly identify the Strava edges aligned with the infrastructure datasets in order to exclude them from further analysis. Our identification approach identifies edges close to the infrastructure location within a certain tolerance. This means that we aim to select all edges that are on the same side of a street. We do this to accommodate possible fluctuations in the infrastructure gps information and possible fluctuations in the Strava dataset (e.g. a cyclists is identified via with a certain edge that does not directly overlap with the infrastructure even though the infrastructure was used).

Our approach is as follows: First we downsize the infrastructure dataset to only contain infrastructure separated from motorised traffic. Our goal is to identify, for each infrastructure piece, all edges within a certain buffer value. For each infrastructure segment we consider all Strava edges that lie within a distance b of the infrastructure. We then filter these candidates based on two criteria:

1. Edges must have an angle smaller than a certain value θ_{max} with the infrastructure segment. This avoid selecting edges that merely cross the infrastructure segment.
2. The overlap length between the edge and the infrastructure segment must be larger than a fraction ρ_{min} of the infrastructure piece. This filters edges that only overlap for a limited length (such as at the beginning or end of the infrastructure piece).

The values for the maximal angle θ_{max}, the minimal fraction ρ_{min} and the buffer size b are tunable parameters we optimise to obtain the optimal selection. We evaluate a selection of edges using an F1 score (based on precision and recall). To achieve this we first sample each infrastructure edge and Strava with points every 1 m If we consider a point (whether on the infrastructure or on the Strava edge) where d is the shortest distance to a point on the other dataset, we assign a weight $w_\sigma(d)$ to this point using a clipped gaussian kernel:

$$w_\sigma(d) = \begin{cases} 1 \quad d \leq \sigma \\ \exp\left(-\frac{(d-\sigma)^2}{2 \cdot \sigma^2}\right) \quad d > \sigma \end{cases}$$

Our chosen kernel has a flat band for distances smaller than σ and afterwards the weight drops off according to a gaussian weight. Precision is the mean over the weights of all the points on the selected edges (this computes how precise the selection is in covering the infrastructure), whereas recall is then the mean over the all points on the infrastructure segments (how well the selected set actually covers the infrastructure). We use a clipped gaussian kernel since we aim to identify parallel edges which could be associated to an infrastructure piece up to a certain distance away. Our kernel therefore weights all edges within σ with a factor of 1.

We find suitable parameters for our approach by sweeping over multiple possible values of the tunable parameters. For each combination (b, θ_{max}, ρ_{min}) we identify the set of edges and compute the F1 score with a sigma value of 4. The baseline set of edges is selected based on the highest obtained F1 score. Due to local variations in gps data of the infrastructure, the performance of a set of parameters can vary between areas. The Brussel Capital Region is therefore partitioned into a regular grid with areas of 1km^2 . For each square we compute the F1 score for the edges from the selected set and the infrastructure that lie within the square. In areas where the F1 score is below a threshold value of 0.95 we rerun the search algorithm with different (b, θ_{max},ρ_{min}) combination (in decreasing order of F1 score on the global set of edges) until 50 iterations are reached or until a local F1 score above the threshold value is obtained.

If we reach the maximal value of 50 iterations this indicates that the infrastructure and Strava edges are misaligned in that region, and that the $\sigma = 4$ value is too strict for the evaluation. We therefore repeat the process of local search with a higher σ value. We note that in this case the F1 values between regions evaluated with different σ value can not be compared. We continue this process until all local regions obtain (in their respective sigma value) a F1 score above the threshold value.

Finally, once we have obtained our final set of edges we perform a manual visual check to add or remove any edges that are not correct.

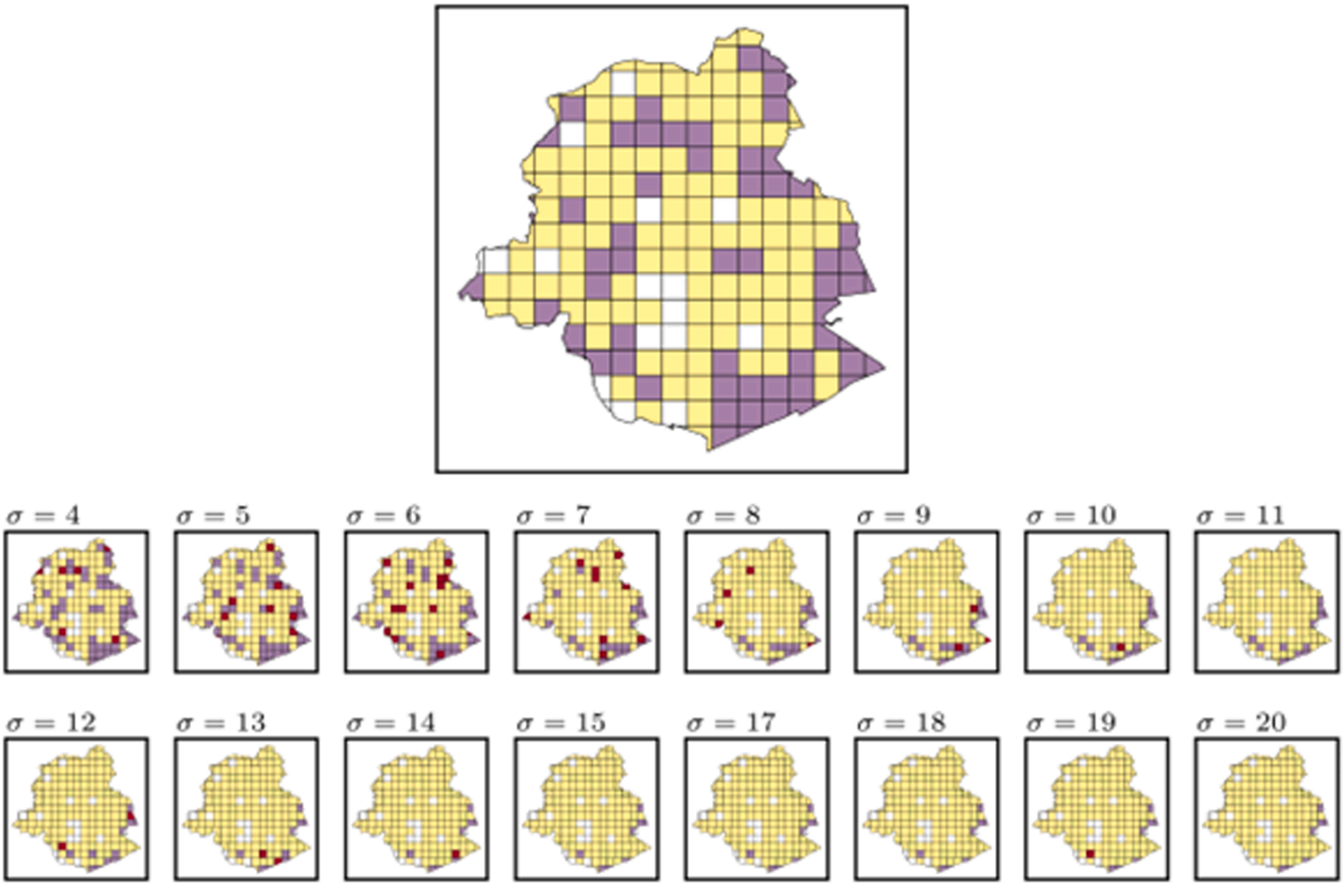


Fig. 8. Strava edges and infrastructure segments are matched in an iterative process. a. We first perform a global search over the entire BCR. We evaluate this selection locally within each square. Yellow squares indicate a region where the edge selection obtain a F1 score above a 0.95 threshold, purple squares fall below this threshold and white square contain no infrastructure. **b.** We iterate with different values of σ which loosens the evaluation. Red squares indicate a region that obtained a F1 score above the 0.95 threshold with the given σ.

D. Model analysis

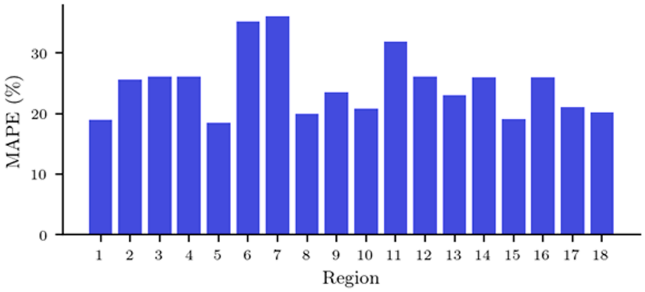


Fig. 9. Mean Absolute Percentage Error (MAPE) per region.

References

Boss, D., Nelson, T., Winters, M., & Ferster, C. J. (2018). Using crowdsourced data to monitor change in spatial patterns of bicycle ridership. *Journal of Transport & Health, 9*, 226–233. https://doi.org/10.1016/j.jth.2018.02.008

Brandeleer, C., Ermans, T., Hubert, M., Janssens, I., Lannoy, P., Loir, C., & Vanderstraeten, P. (2016). *Cahier 5 (a) de l'Observatoire de la Mobilité: Le partage de l'espace public en Région de Bruxelles-Capitale* (p. 89). https://data.mobility.brussels/home/media/filer_public/b3/27/b327a198-d758-4ae0-9fdb-c773f5009bfc/co5afr.pdf.

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice, 46*(10), 1730–1740. https://doi.org/10.1016/j.tra.2012.07.005

Brum-Bastos, V., Ferster, C. J., Nelson, T., & Winters, M. (2019). Where to put bike counters? Stratifying bicycling patterns in the City using crowdsourced data |. *Transport Findings*. https://doi.org/10.32866/10828

Buehler, R., & Pucher, J. (2024). COVID-19 and cycling: A review of the literature on changes in cycling levels and government policies from 2019 to 2022. *Transport Reviews, 44*(2), 299–344. https://doi.org/10.1080/01441647.2023.2205178

BXI. (2021). *Nouvelles pistes cyclables 'Covid': Plus de 38 kilomètres depuis avril 2020*. August 23. BX1 https://bx1.be/categories/news/nouvelles-pistes-cyclables-regionales-bilan/.

Caruana, R. (1993). Multitask learning: A knowledge-based source of inductive bias1. In *Proceedings of the Tenth International Conference on Machine Learning*.

Conrow, L., Wentz, E., Nelson, T., & Pettit, C. (2018). Comparing spatial patterns of crowdsourced and conventional bicycling datasets. *Applied Geography, 92*, 21–30. https://doi.org/10.1016/j.apgeog.2018.01.009

Dadashova, B., Griffin, G. P., Das, S., Turner, S., & Sherman, B. (2020). Estimation of average annual daily bicycle counts using crowdsourced Strava data. *Transportation Research Record: Journal of the Transportation Research Board, 2674*(11), 390–402. https://doi.org/10.1177/0361198120946016

de Séjournet, A., Macharis, C., Tori, S., & Vanhaverbeke, L. (2022). Evolution of urban mobility behaviour in Brussels as a result of the COVID-19 pandemic. *Regional Science Policy & Practice, 14*(S1), 107–121. https://doi.org/10.1111/rsp3.12525

DiGioia, J., Watkins, K. E., Xu, Y., Rodgers, M., & Guensler, R. (2017). Safety impacts of bicycle infrastructure: A critical review. *Journal of Safety Research, 61*, 105–119. https://doi.org/10.1016/j.jsr.2017.02.015

Dyason, D., Coetzee, C. E., & Kleyhnanhs, E. (2025). The role of spatial variability in developing cycling cities: Implications drawn from geographically weighted regressions. *Smart Cities, 8*. https://doi.org/10.3390/smartcities8040133

Eitzel, M., Cappadonna, J., Santos-Lang, C., Duerr, R., West, S. E., Virapongse, A., Kyba, C., Bowser, A., Cooper, C., Sforzi, A., Metcalfe, A., Harris, E., Thiel, M., Haklay, M., Ponciano, L., Roche, J., Ceccaroni, L., Shilling, F., Dorler, D., … Jiang, Q. (2017). Citizen science terminology matters: Exploring key terms. *Citizen Science: Theory and Practice*, 1–20.

El Esawey, M. (2018a). Daily bicycle traffic volume estimation: Comparison of historical average and count models. *Journal of Urban Planning and Development, 144*(2), Article 04018011. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000443

El Esawey, M. (2018b). Impact of data gaps on the accuracy of annual and monthly average daily bicycle volume calculation at permanent count stations. *Computers, Environment and Urban Systems, 70*, 125–137. https://doi.org/10.1016/j.compenvurbsys.2018.03.002

Fischer, J., Nelson, T., & Winters, M. (2022). Changes in the representativeness of Strava bicycling data during COVID-19. *Findings*. https://doi.org/10.32866/001c.33280

Fishman, E., Schepers, P., & Kamphuis, C. B. M (2015). Dutch cycling: Quantifying the health and related economic benefits. *American Journal of Public Health, 105*(8), e13–e15. https://doi.org/10.2105/AJPH.2015.302724

Furth, P. G., Mekuria, M. C., & Nixon, H. (2016). Network connectivity for low-stress bicycling -. *Transportation Research Record: Journal of the Transportation Research Board, 2587*(1). %252010.3141/2587-06.

Gössling, S., & McRae, S. (2022). Subjectively safe cycling infrastructure: New insights for urban designs. *Journal of Transport Geography, 101*, Article 103340. https://doi.org/10.1016/j.jtrangeo.2022.103340

Hankey, S., & Lindsey, G. (2016). Facility-demand models of peak period pedestrian and bicycle traffic: Comparison of fully specified and reduced-form models. *Transportation Research Record, 2586*(1), 48–58. https://doi.org/10.3141/2586-06

Hardinghaus, M., & Weschke, J. (2022). Attractive infrastructure for everyone? Different preferences for route characteristics among cyclists. *Transportation Research Part D: Transport and Environment, 111*, Article 103465. https://doi.org/10.1016/j.trd.2022.103465

Heesch, K. C., James, B., Washington, T. L., Zuniga, K., & Burke, M. (2016). Evaluation of the Veloway 1: A natural experiment of new bicycle infrastructure in Brisbane, Australia. *Journal of Transport & Health, 3*(3), 366–376. https://doi.org/10.1016/j.jth.2016.06.006

Heesch, K. C., & Langdon, M. (2016). The usefulness of GPS bicycle tracking data for evaluating the impact of infrastructure change on cycling behaviour. *Health Promotion Journal of Australia, 27*(3), 222–229. https://doi.org/10.1071/HE16032

Henry, A., Ermans, T., & de Smet d'Olbecke, F. (2020). Un retour de la bicyclette, aussi à Bruxelles ? *Brussels Studies. La revue scientifique pour les recherches sur Bruxelles /Het wetenschappelijk tijdschrift voor onderzoek over Brussel / The Journal of Research on Brussels*. https://doi.org/10.4000/brussels.4826

Hong, J., McArthur, D. P., & Livingston, M. (2020). The evaluation of large cycling infrastructure investments in Glasgow using crowdsourced cycle data. *Transportation, 47*(6), 2859–2872. https://doi.org/10.1007/s11116-019-09988-4

Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters, 3*(1), 63–75. https://doi.org/10.3328/TL.2011.03.01.63-75

Janež, M., Verovšek, Š., Zupančič, T., & Moškon, M. (2022). Citizen Science for Traffic monitoring: Investigating the potentials for complementing traffic counters with crowdsourced data. *Sustainability, 14*(2). https://doi.org/10.3390/su14020622. Article 2.

Jean-Louis, G., Eckhardt, M., Podschun, S., Mahnkopf, J., & Venohr, M. (2024). Estimating daily bicycle counts with Strava data in rural and urban locations. *Travel Behaviour and Society, 34*, Article 100694. https://doi.org/10.1016/j.tbs.2023.100694

Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography, 52*, 90–97. https://doi.org/10.1016/j.jtrangeo.2016.03.006

Krenn, P. J., Oja, P., & Titze, S. (2014). Route choices of transport bicyclists: A comparison of actually used and shortest routes. *The International Journal of Behavioral Nutrition and Physical Activity, 11*(1), 31. https://doi.org/10.1186/1479-5868-11-31

Kwigizile, V., Morgan Kwayu, K., & Oh, J.-S. (2022). Leveraging the spatial-temporal resolution of crowdsourced cycling data to improve the estimation of hourly bicycle volume. *Transportation Research Interdisciplinary Perspectives, 14*, Article 100596. https://doi.org/10.1016/j.trip.2022.100596

Lee, K., & Sener, I. N. (2019). Understanding potential exposure of bicyclists on roadways to traffic-related air pollution: Findings from El Paso, Texas, using Strava Metro data. *International Journal of Environmental Research and Public Health, 16*(3). https://doi.org/10.3390/ijerph16030371. Article 3.

Lee, K., & Sener, I. N. (2021). Strava Metro data for bicycle monitoring: A literature review. *Transport Reviews, 41*(1), 27–47. https://doi.org/10.1080/01441647.2020.1798558

Lewis, S., & Grande del Valle, E. (2019). San Francisco's neighborhoods and auto dependency. *Cities (London, England), 86*, 11–24. https://doi.org/10.1016/j.cities.2018.12.017

Livingston, M., McArthur, D., Hong, J., & English, K. (2021). Predicting cycling volumes using crowdsourced activity data. *Environment and Planning B: Urban Analytics and City Science, 48*(5), 1228–1244. https://doi.org/10.1177/2399808320925822

L'Observatoire Good Move. (2021). *Les aménagements cyclables et le matériel roulant à Bruxelles*. https://data.mobility.brussels/home/fr/observatoire/le-velo/les-amenagements-cyclables/.

Loshchilov, I., & Hutter, F. (2019). *Decoupled weight decay regularization* (arXiv:1711.05101). ArXiv. https://doi.org/10.48550/arXiv.1711.05101.

Mattioli, G., Roberts, C., Steinberger, J. K., & Brown, A. (2020). The political economy of car dependence: A systems of provision approach. *Energy Research & Social Science, 66*, Article 101486. https://doi.org/10.1016/j.erss.2020.101486

McArthur, D. P., & Hong, J. (2019). Visualising where commuting cyclists travel using crowdsourced data. *Journal of Transport Geography, 74*, 233–241. https://doi.org/10.1016/j.jtrangeo.2018.11.018

McDonald, N. C. (2008). Children's mode choice for the school trip: The role of distance and school location in walking to school. *Transportation, 35*(1), 23–35. https://doi.org/10.1007/s11116-007-9135-7

Miah, M. M., Hyun, K. K., Mattingly, S. P., & Khan, H. (2023). Estimation of daily bicycle traffic using machine and deep learning techniques. *Transportation, 50*(5), 1631–1684. https://doi.org/10.1007/s11116-022-10290-z

Miranda-Moreno, L. F., Nosal, T., Schneider, R. J., & Proulx, F. (2013). Classification of bicycle traffic patterns in five North American cities. *Journal of the Transportation Research Board, 239*(1). https://doi.org/10.3141/2339

Munira, S., & Sener, I. N. (2020). A geographically weighted regression model to examine the spatial variation of the socioeconomic and land-use factors associated with Strava bike activity in Austin, Texas. *Journal of Transport Geography, 88*, Article 102865. https://doi.org/10.1016/j.jtrangeo.2020.102865

Nelson, T., Ferster, C., Laberee, K., Fuller, D., & Winters, M. (2021). Crowdsourced data for bicycling research and practice. *Transport Reviews, 41*(1), 97–114. https://doi.org/10.1080/01441647.2020.1806943

Nelson, T., Roy, A., Ferster, C., Fischer, J., Brum-Bastos, V., Laberee, K., Yu, H., & Winters, M. (2021). Generalized model for mapping bicycle ridership with crowdsourced data. *Transportation Research Part C: Emerging Technologies, 125*, Article 102981. https://doi.org/10.1016/j.trc.2021.102981

Orellana, D., & Guerrero, M. L. (2019). Exploring the influence of road network structure on the spatial behaviour of cyclists using crowdsourced data. *Environment and Planning B, 46*(7), 1314–1330. https://doi.org/10.1177/2399808319863810

Pro Velo. (2025). *Les « nouveaux cyclistes »*. https://data.mobility.brussels/media/publications/reports/etudes/Les_nouveaux_cyclistes.pdf.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big Data and Data-driven decision making. *Big Data, 1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

Roy, A., Nelson, T. A., Fotheringham, A. S., & Winters, M. (2019). Correcting bias in crowdsourced data to map bicycle ridership of all bicyclists. *Urban Science, 3*(2). https://doi.org/10.3390/urbansci3020062. Article 2.

Samyn, W., & Lagrou, R. (2025). *Onderzoek Verplaatsingsgedrag (2023—2024)*. https://data.mobility.brussels//media/publications/reports/etudes/AnalyserapportOVG7.pdf.

Sanders, R. L., Frackelton, A., Gardner, S., Schneider, R., & Hintze, M. (2017). Ballpark method for estimating pedestrian and bicyclist exposure in Seattle, Washington: Potential option for resource-constrained cities in an age of big data. *Transportation Research Record, 2605*(1), 32–44. https://doi.org/10.3141/2605-03

StatBel. (2022). *Op 1 januari 2022 telde de wettelijke belgische bevolking 11.584.008 inwoners* | Statbel. https://statbel.fgov.be/nl/nieuws/op-1-januari-2022-telde-de-wettelijke-belgische-bevolking-11584008-inwoners-0.

Strava. (n.d.). *Strava | Metro*. Retrieved 4 July 2025, from https://metro.strava.com/.

Strava Metro. (2025). *Edge counts data export and download*. December 18. Strava https://stravametro.zendesk.com/hc/en-us/articles/360051202734-Edge-Counts-Data-Export-and-Download.

Tilahun, N. Y., Levinson, D. M., & Krizek, K. J. (2007). Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey. *Transportation Research Part A: Policy and Practice, 41*(4), 287–301. https://doi.org/10.1016/j.tra.2006.09.007

Tripathy, P., Rao, P., Balakrishnan, K., & Malladi, T. (2021). An open-source tool to extract natural continuity and hierarchy of urban street networks. *Environment and planning B: urban analytics and city science, 48*(8), 2188–2205.

Urry, J. (2004). The 'system' of automobility. *Theory, Culture & Society, 21*(4–5), 25–39. https://doi.org/10.1177/0263276404046059

Vandenbulcke, G., Dujardin, C., Thomas, I., Geus, B.de, Degraeuwe, B., Meeusen, R., & Panis, L. I (2011). Cycle commuting in Belgium: Spatial determinants and 're-cycling' strategies. *Transportation Research Part A: Policy and Practice, 45*(2), 118–137. https://doi.org/10.1016/j.tra.2010.11.004

Venter, Z. S., Gundersen, V., Scott, S. L., & Barton, D. N. (2023). Bias and precision of crowdsourced recreational activity data from Strava. *Landscape and Urban Planning, 232*, Article 104686. https://doi.org/10.1016/j.landurbplan.2023.104686

Vybornova, A., Cunha, T., Gühnemann, A., & Szell, M. (2023). Automated detection of missing links in bicycle networks. *Geographical Analysis, 55*(2), 239–267. https://doi.org/10.1111/gean.12324

Watkins, K., Ammanamanchi, R., LaMondia, J., & Dantec, C.L. (2016). *Comparison of smartphone-based cyclist GPS data sources*. https://www.semanticscholar.org/paper/

Comparison-of-Smartphone-based-Cyclist-GPS-Data-Watkins-Ammanamanchi/747441eeaf7fb3cc3b6c1ff78203a7b61cd6da39.

Wei, Z., Zhen, F., Mo, H., Wei, S., Peng, D., & Zhang, Y. (2021). Travel behaviours of sharing bicycles in the Central urban area based on geographically weighted regression: The case of Guangzhou, China. *Chinese Geographical Science, 31*(1), 54–69. https://doi.org/10.1007/s11769-020-1159-3

Winters, M., & Teschke, K. (2010). Route preferences among adults in the near market for bicycling: Findings of the cycling in cities study. *American Journal of Health Promotion, 25*(1), 40–47. https://doi.org/10.4278/ajhp.081006-QUAN-236. Scopus.

Xu, J., & Nyerges, T. L. (2017). A framework for user-generated geographic content acquisition in an age of crowdsourcing. *Cartography and Geographic Information Science, 44*(2), 98–112. https://doi.org/10.1080/15230406.2016.1163644