

Quantifying the accuracy of OpenStreetMap for longitudinal cycling-infrastructure change: A reproducible validation framework using Google Street View

Eugeni Vidal-Tortosa ^{a,*,} Víctor González-Parra ^{a,b,} Oriol Marquet ^{a,b,}

^a Department of Geography, Universitat Autònoma de Barcelona (UAB), Building B, UAB Campus, Carrer de la Fortuna s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain

^b Institute of Environmental Science and Technology (ICTA-UAB), Building ICTA-ICP, UAB Campus, Carrer de les Columnes s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain

ARTICLE INFO

Dataset link: <https://github.com/GEMOTT/osm-cycling-change-framework>

Keywords:

OpenStreetMap (OSM)
Cycling infrastructure
Urban data science
Change detection
Spatial validation
Reproducible workflows

ABSTRACT

Cities are rapidly expanding their cycling networks to promote sustainable mobility, yet robust longitudinal data needed to evaluate change are scarce. OpenStreetMap (OSM) provides a global, time-stamped record of urban infrastructure, but its reliability for detecting change remains uncertain. This study quantified the temporal accuracy of OSM for identifying cycling-infrastructure additions and removals, using Barcelona (2015–2023) as a case study within a reproducible open-source workflow with manual validation using street-level imagery.

We reconstructed baseline and follow-up cycling-infrastructure networks from dated OSM snapshots and applied geometric differencing to detect change. A stratified sampling design based on population density and centrality was used to validate detected changes against historical Google Street View (GSV) imagery, allowing estimation of precision, recall, and F1 scores. OSM-derived estimates indicate an 80.3% increase in cycling-infrastructure length, from 130.7 km to 235.7 km. For additions, validation shows high recall (0.96) and strong precision (0.82), indicating that OSM captures most network expansions. For removals, recall is high (1.00) but precision is low (0.20; $n = 5$), with wide uncertainty given the small sample. Apparent losses may reflect mapping edits rather than real infrastructure change.

By quantifying detection accuracy, this study provides a framework for calibrating OSM-derived measures of cycling infrastructure change. OSM snapshot differencing can serve as a scalable screening approach for identifying infrastructure additions, while apparent removals require cautious interpretation and independent verification to avoid mapping artefacts being mistaken for physical infrastructure loss. This framework supports more robust longitudinal use of OSM histories in planning and research.

1. Introduction

In recent years, many cities worldwide have expanded their cycling-infrastructure network in pursuit of cleaner, healthier and more equitable mobility (Szell et al., 2022; Buehler and Pucher, 2021). Reliable data on how these networks evolve are essential for evidence-based planning and robust longitudinal research (Hirsch et al., 2016). Such data enable policymakers and researchers to examine how infrastructure change relates to mode choice, safety and equity (Houde et al., 2018; Prince et al., 2025).

Yet longitudinal infrastructure data are often incomplete or inconsistent. Official inventories are rarely maintained systematically across years, and field audits, although precise, are resource-intensive and

difficult to replicate at scale (Winters et al., 2025; Dai et al., 2024). In the absence of consistent data, policymakers and researchers must rely on fragmented sources, increasing the risk of exposure misclassification and biased impact estimates (Batomen et al., 2026). As a result, comparable evidence on cycling-infrastructure change remains limited, constraining both informed policy development and robust research (Houde et al., 2018).

The growing availability of Volunteered Geographic Information (VGI) offers new opportunities to address these data gaps (Goodchild, 2007). Among such sources, OpenStreetMap (OSM) stands out for providing open, editable, and time-stamped spatial data on a wide range of urban features, making it a valuable resource for reconstructing

* Corresponding author.

E-mail address: eugeni.vidal@uab.cat (E. Vidal-Tortosa).

<https://doi.org/10.1016/j.jtrangeo.2026.104775>

Received 9 March 2026; Received in revised form 27 May 2026; Accepted 7 July 2026

Available online 21 July 2026

0966-6923/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

historical infrastructure networks and studying urban infrastructure change over time (Minghini and Frassinelli, 2019; Zhang and Pfoser, 2019).

Early comparisons with official datasets suggest that OSM can achieve good positional accuracy and substantial agreement in mapped features (Haklay, 2010; Girres and Touya, 2010). However, its use for longitudinal change analysis remains more uncertain. OSM data quality varies across space and time, and edits may reflect not only real-world changes but also evolving mapping practices, including delayed or retrospective updates (Barron et al., 2014; Senaratne et al., 2017; Minghini and Frassinelli, 2019). In addition, tagging practices and compliance can vary across local contexts, introducing further heterogeneity in how features are represented (Almendros-Jiménez and Becerra-Terón, 2018). This uncertainty in timing can influence analytical results when historical OSM snapshots are used as proxies for real-world change (Schmidl et al., 2021; Juhász, 2021).

Cycling infrastructure poses additional challenges because some facilities, particularly on-street cycle lanes, are represented as attributes of road segments rather than as independent geometries. This makes them less visually distinct, harder to map consistently, and more prone to omission or incorrect tagging (Hochmair et al., 2015; Ferster et al., 2023). Despite these challenges, empirical comparisons with reference datasets suggest that OSM can provide useful cycling-infrastructure inventories in many contexts, although agreement varies across cities and infrastructure types and depends on consistent tagging practices (Ferster et al., 2020). Evidence also suggests that OSM bicycle-infrastructure data quality varies spatially, particularly outside dense urban cores (Vierø et al., 2025).

Importantly, recent work has shown the value of OSM histories for measuring cycling-infrastructure network growth at scale, providing a scalable alternative when consistent historical official datasets are unavailable (Bres et al., 2023; Winters et al., 2025). At the same time, these studies highlight that observed changes may reflect not only real-world infrastructure change but also evolving mapping practices, delayed mapping, or revisions to existing OSM data.

For this reason, OSM-derived changes require independent validation. To assess whether changes detected from historical OSM snapshots reflect real on-street changes, an external source of evidence is needed. Street-level imagery, including Google Street View (GSV), has increasingly been used to audit and validate built-environment features remotely (Rundle et al., 2011; Askari et al., 2025; Hanibuchi et al., 2019). However, to our knowledge, no study has yet validated cycling-infrastructure change detected from historical OSM snapshots using GSV imagery.

Building on this gap, this study aims to quantify the accuracy of OSM for detecting cycling-infrastructure change. Using Barcelona over the 2015–2023 period as a case study, we implement a reproducible open-source workflow that combines longitudinal snapshot differencing with probability-based stratified validation against historical GSV imagery. The selected time window aligns with a broader research project examining infrastructure change across policy-relevant governance periods, although the present paper focuses exclusively on methodological validation. Detection performance is evaluated using precision, recall, and F1 scores. By integrating change detection and external validation within a transparent workflow with fully reproducible computational components and structured manual validation, the study provides a framework for calibrating OSM-based measures of cycling-infrastructure change.

2. Data and methods

2.1. Study setting

Barcelona is a large, dense Mediterranean city in northeastern Spain (around 1.7 million inhabitants within the municipal boundary) with an extensive and evolving cycling network. Between 2015 and 2023,

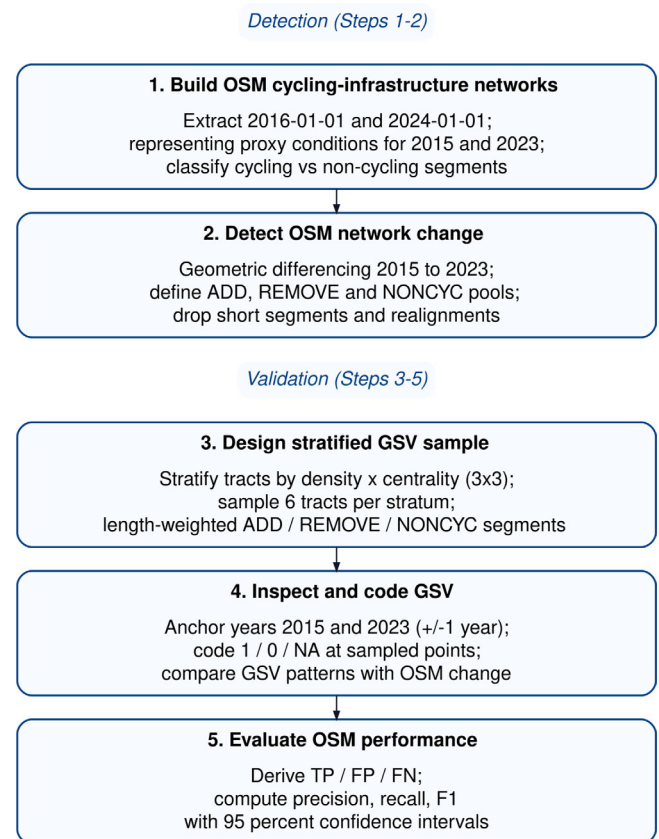


Fig. 1. Five-step workflow for OSM-based change detection and GSV-based validation. Steps 1–2 reconstruct cycling-infrastructure networks and detect change from historical OSM snapshots, while Steps 3–5 evaluate these detections using stratified manual validation with GSV imagery.

the city experienced substantial cycling-infrastructure expansion, generating measurable network change over time. These dynamics make Barcelona a suitable case for evaluating OSM-based network reconstruction, change detection, and validation procedures.

2.2. Data sources

We used three data sources. First, OpenStreetMap (OSM) extracts for Barcelona were retrieved for 1 January 2016 and 1 January 2024 using the *osmextract* R package (Gilardi and Lovelace, 2025). These Geofabrik snapshots represented baseline and follow-up cycling-infrastructure networks corresponding to the end of 2015 and 2023, respectively. Second, historical Google Street View (GSV) panoramas were used to validate a sample of OSM-detected changes. Third, census-tract boundaries (2015) and resident population counts (2022) from official statistical sources were used to support the stratified sampling design.

2.3. Analytical workflow

The analysis follows a five-step workflow (Fig. 1), implemented as a single pipeline with reproducible computational steps and structured manual validation. Steps 1–2 detect cycling-infrastructure change, while Steps 3–5 evaluate these detections using stratified GSV validation.

Step 1. Build OSM cycling-infrastructure networks

We constructed baseline and follow-up cycling-infrastructure networks from the dated OSM extracts described above, representing 2015 and 2023 conditions. From each extract, we derived a core cycling-infrastructure network. A segment was classified as cycling infrastructure if either (i) it was mapped as a dedicated cycleway (highway =

cycleway), or (ii) it was a street segment carrying explicit cycleway tags indicating a lane or track. In the second case, we searched cycleway-related keys (e.g., `cycleway`, `cycleway:left`, `cycleway:right`, `cycleway:both`, and other `cycleway:*` variants) and retained segments where any value matched `{lane, track, opposite_lane, opposite_track}` (OpenStreetMap contributors, 2026). Values were matched case-insensitively and, where multiple values were present, a segment was retained if any value matched the set above.

To maintain a conservative and visually interpretable definition of dedicated cycling infrastructure for subsequent GSV-based validation, we excluded `cycleway=shared_lane` tags. Although these tags may represent legitimate forms of cycling provision, they are often associated with more ambiguously defined facilities, such as advisory lane markings or shared-traffic treatments, which may not be consistently identifiable in historical GSV imagery. Including these tags would substantially broaden the operational definition of cycling infrastructure and could reduce consistency in the manual validation process. As a result, the extracted network should be interpreted as a conservative representation of clearly delineated cycling infrastructure rather than a comprehensive representation of all forms of cycling provision. This distinction may be particularly relevant in cities where cycling provision relies more heavily on shared or mixed-traffic facilities.

To reduce duplication from parallel OSM representations of the same facility (e.g., a dedicated cycle track mapped as a separate geometry while a cycle lane is also tagged on the adjacent carriageway), we applied a no-double-counting (NDC) rule. Specifically, on-road cycle-lane segments were excluded when they overlapped with nearby dedicated cycling infrastructure, defined as cases where at least 20 m of the lane segment and at least 50% of its length fell within a 15 m buffer around dedicated cycling geometries. The 15 m proximity threshold reflects typical spatial offsets between carriageway-centreline representations and adjacent segregated cycle tracks in dense urban settings, while remaining conservative enough to avoid conflating distinct facilities on opposite sides of wide corridors. The additional requirements of a minimum 20 m overlap and at least 50% segment inclusion ensure that exclusions reflect substantive parallel representation rather than incidental geometric proximity.

Step 2. Detect OSM network change

Cycling-infrastructure change between the baseline and follow-up networks was detected using geometric differencing in a projected CRS. First, each network was dissolved and buffered by 10 m to accommodate minor geometric misalignments and small positional edits. This tolerance absorbs small positional shifts introduced by re-digitisation or geometry refinement, while remaining conservative relative to typical urban carriageway widths. Additions were defined as segments present in the 2023 proxy network but absent from the buffered 2015 proxy network; removals were defined analogously as segments present in the 2015 proxy network but absent from the buffered 2023 proxy network. Resulting fragments shorter than 10 m were discarded to remove small artefacts arising from geometric differencing.

To avoid counting positional edits in OSM as real infrastructure change, we flagged likely realignments as cases where an added segment lay within 15 m of a corresponding removed segment, reflecting small shifts in mapped geometry rather than substantive infrastructure modification. These cases were excluded from the validation evaluation pool. The 15 m proximity threshold served a distinct purpose from the differencing buffer: it was intended to capture cases where added and removed fragments likely represent the same infrastructure feature that has been re-drawn, split, or slightly shifted in OSM rather than physically removed and rebuilt.

To define control locations and support the identification of false negatives, we constructed a non-cycling comparison pool (NONCYC) consisting of general street segments that did not meet the cycling-infrastructure tagging criteria. NONCYC candidates were restricted to

standard road classes (highway $\in$ {primary, secondary, tertiary, unclassified, residential and corresponding link classes, living_street, pedestrian}). To ensure a clear separation between cycling and non-cycling locations, we excluded any NONCYC candidate geometries located within 10 m of cycling infrastructure in either reference year (2015 proxy or 2023 proxy). Resulting fragments shorter than 10 m were discarded to remove small geometric artefacts.

The robustness of key geometric parameters used in this step (buffer distance, realignment threshold, and minimum segment length) was assessed through dedicated sensitivity analyses (see Section 3.3 and Supplement S1).

Step 3. Design stratified GSV sample

To evaluate the accuracy of OSM-detected cycling-infrastructure changes, we designed a stratified GSV validation across Barcelona's 1063 census tracts by population density and centrality (2015; Fig. 2 a). Population density was calculated as the 2022 resident population per square kilometre for each tract, using official census counts and tract area. Centrality was operationalised as a simple, reproducible proxy for inner vs outer areas, measured via Euclidean distance to Plaça Catalunya. Each variable was divided into terciles, and their combination produced nine density-centrality strata (D1_C1 to D3_C3). From each stratum, six tracts were randomly selected (total $n = 54$).

Within each sampled tract, we attempted to sample up to two OSM-detected additions, two removals, and one NONCYC control segment. Segments were sampled with probability proportional to their length, so longer segments were more likely to be selected than shorter fragments. Candidate segments were split at census-tract boundaries before sampling so that segment length (and therefore sampling probability) was calculated only within each tract.

Each sampled segment was converted to a single validation point located at its mid-length position (Fig. 2 b). Points were linked to the nearest available GSV panorama and exported to a structured coding workbook (Supplement S2). The same points were also provided in an interactive map with direct Street View links (Supplement S3).

Step 4. Inspect and code GSV

Each validation site was independently coded by two trained coders following a standardised protocol (Supplement S4). Coders assessed whether cycling infrastructure was present in a baseline and follow-up period corresponding to the OSM reference snapshots. When the nearest panorama did not provide a clear view of the sampled location, coders were permitted to adjust the Street View position slightly along the same street segment, provided that the original sampled point remained visible and spatially identifiable. For each site and period, coders recorded infrastructure presence/absence and the month of the imagery used.

Because GSV imagery availability varies across locations and years, coders prioritised imagery as close as possible to the reference dates. For the baseline period, coders attempted 2016 and 2015 first (using 2014 only as a fallback when required); for the follow-up period, coders attempted 2024 and 2023 first (using 2022 only as a fallback when required). When multiple candidate years were available, the observation closest in time to the reference date was selected to define the baseline and follow-up values.

Sites were coded as missing when available imagery did not allow a clear determination of cycling-infrastructure presence for the reference period (e.g., obscured views, roadworks, or ambiguous visual evidence). Missing observations were excluded from analyses of change, as these require clear identification of both baseline and follow-up conditions to determine whether a change occurred. Inter-coder agreement was assessed on usable sites, and any discrepancies were resolved through reconciliation to produce the reconciled validation dataset (Supplement S5).

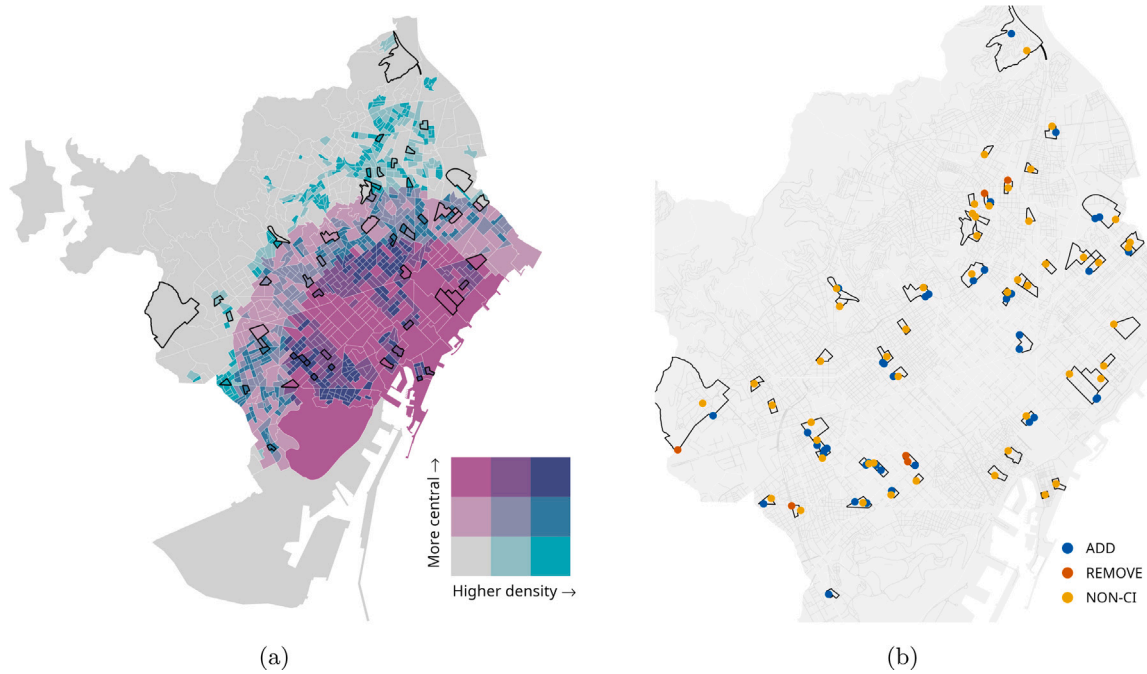


Fig. 2. Outputs used in the validation design: (a) Bivariate tract stratification; (b) Validation points by type.

Step 5. Evaluate OSM performance

Using the reconciled baseline and follow-up GSV observations, we evaluated whether OSM snapshot differencing correctly identified additions and removals of cycling infrastructure. For segments flagged as additions, a true positive (TP) corresponded to a 0 → 1 transition between baseline and follow-up, whereas a false positive (FP) corresponded to no change (0 → 0) or persistence of cycling infrastructure (1 → 1). For segments flagged as removals, a TP corresponded to a 1 → 0 transition, whereas an FP corresponded to no change (1 → 1) or no cycling infrastructure in either period (0 → 0). NONCYC control locations were used to identify false negatives (FN), defined as cycling-infrastructure changes visible in GSV but not detected by OSM differencing. Examples of TP additions and removals are shown in Fig. 3.

Validation performance was summarised using standard classification metrics. Precision, recall and the F1 score were computed as:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Pooled metrics were computed by first aggregating TP, FP and FN across addition and removal events, and then calculating precision, recall and the F1 score from these pooled counts. For precision and recall, 95% confidence intervals were computed using Wilson score intervals for the corresponding numerators and denominators, which provide stable coverage for proportions, particularly with small sample sizes or extreme values.

All data processing, change detection, sampling, and validation procedures were implemented in an open-source R workflow. Code availability is detailed in the Data availability section.

3. Results

3.1. Network change detected from OSM (2015–2023)

Between 2015 and 2023, the OSM-derived cycling-infrastructure network in Barcelona increased from 130.7 to 235.7 km, corresponding to a net growth of 105.0 km, an 80.3% increase (Fig. 4). Network lengths were calculated as the summed length of cycling-infrastructure

Table 1

Network totals and differencing-based change estimates: consistency check (2015–2023, Barcelona).

Metric	Value (km)
Total network length (2015)	130.7
Total network length (2023)	235.7
Net growth (2023–2015)	105.0
Additions	127.1
Removals	24.9
Additions - removals	102.2
Gap: (Additions - removals) - Net	-2.8

line geometries in OSM, without collapsing parallel or spatially separated facilities into a single corridor-level measure (e.g., cycle tracks on both sides of a street). Geometric differencing identified 127.1 km of added and 24.9 km of removed cycling infrastructure, yielding a net change of 102.2 km. The small discrepancy (−2.8 km) reflects minor geometric artefacts introduced by the buffering, dissolving, and fragment-filtering steps used in the differencing procedure, and indicates good internal consistency between aggregate network totals and change estimates (Table 1).

As an external consistency check, we compared OSM-derived cycling-infrastructure network length with official municipal data for Barcelona (Àrea Metropolitana de Barcelona, 2015; Ajuntament de Barcelona, 2023), one of the few cases where consistent longitudinal records are available. Both sources show a very similar increase over the study period (OSM: 130.7 to 235.7 km; official: 133.0 to 234.0 km), with only minimal differences in aggregate network length at either time point. This close correspondence provides additional support that OSM-derived estimates broadly reflect the aggregate growth trajectory.

OSM-detected additions were unevenly distributed across the city. Low-density strata (D1_C1, D1_C2 and D1_C3) accounted for 60.0% of all added length, with particularly large gains in low-density peripheral (D1_C1) and low-density central (D1_C3) areas (Table 2). Additions were also substantial in medium-density central tracts (D2_C3) and high-density central tracts (D3_C3), indicating that expansion was not confined to the urban periphery. In contrast, removals were limited in total length and were more unevenly distributed, with the largest shares observed in low-density central areas (D1_C3; 37.7%) and low-density

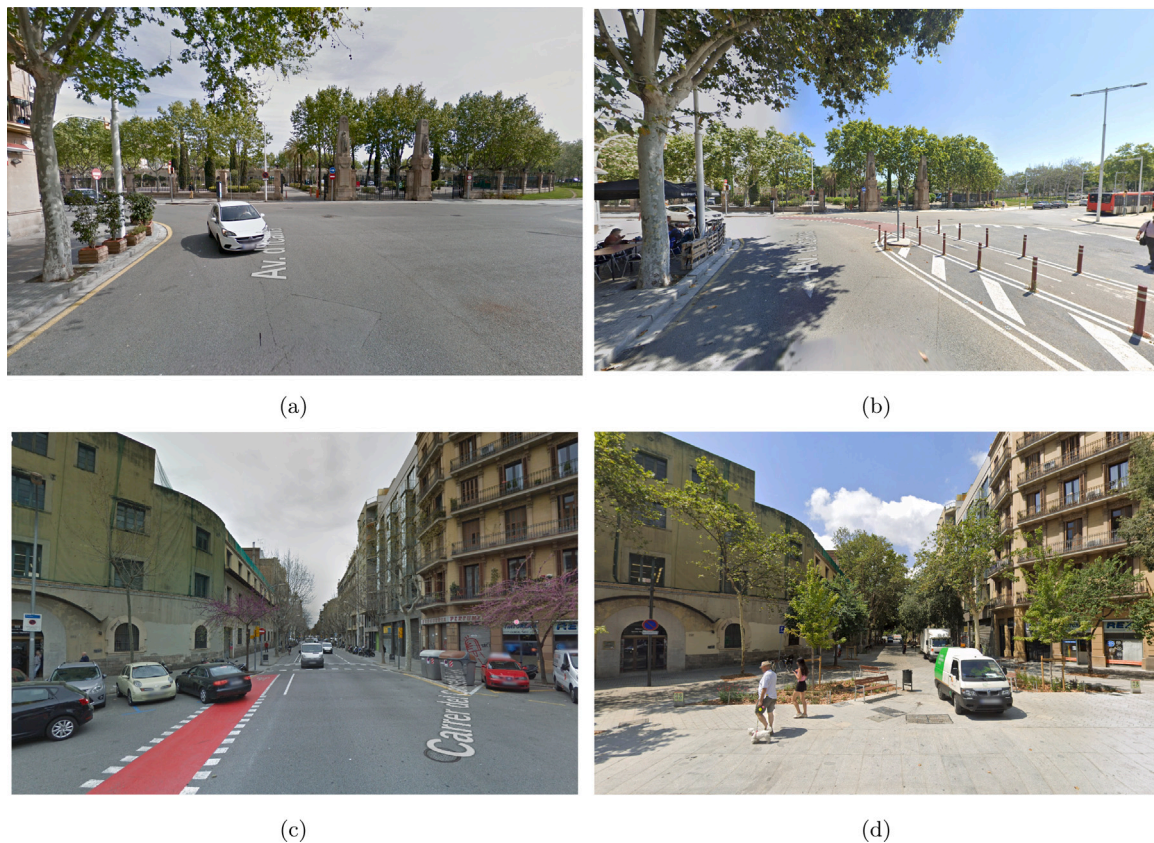


Fig. 3. Examples of true positive changes identified during GSV validation: (a–b) addition (0→ 1), from no cycling infrastructure in 2015 to a newly implemented segregated cycle track in 2023; (c–d) removal (1→ 0), from a cycle track in 2015 to its absence in 2023 following street redesign.

Table 2

OSM-estimated additions and removals (2015–2023) by density × centrality stratum.

Stratum	Definition	Added (km)	Removed (km)	Added (%)	Removed (%)
D1_C1	Low density, peripheral	31.0	7.1	24	29
D1_C2	Low density, intermediate	16.0	1.0	13	4
D1_C3	Low density, central	29.1	9.4	23	38
D2_C1	Medium density, peripheral	2.0	0.5	2	2
D2_C2	Medium density, intermediate	10.9	1.5	9	6
D2_C3	Medium density, central	15.1	2.8	12	11
D3_C1	High density, peripheral	1.4	0.1	1	0
D3_C2	High density, intermediate	9.7	1.2	8	5
D3_C3	High density, central	11.7	1.3	9	5
TOTAL	All strata combined	126.8	24.8	100	100

Note: Stratum-level totals are calculated from change geometries after assignment to census tracts. Consequently, minor differences from the overall network totals reported in Table 1 may occur due to the tract-assignment procedure.

peripheral areas (D1_C1; 28.8%), while all other strata each accounted for less than 12% of removed length.

3.2. Validation of OSM-detected changes using GSV

The evaluation pool consisted of 946 OSM-detected change segments (850 additions and 96 removals) after excluding short fragments and likely realignments. From this pool, we sampled 45 change segments for validation (39 additions and 6 removals), representing 4.8% of all eligible change candidates. Of these sampled change segments, 38 were usable for analysis (33 additions and 5 removals), as validation required GSV imagery for both reference periods. In addition, 48 NONCYC control sites were included to assess false negatives. Despite the reduction in usable change segments, validation points remained distributed across all density–centrality strata, consistent with

the stratified sampling design (Table 3). However, the realised validation sample was also constrained by the spatial distribution of eligible change segments and the availability of usable GSV imagery, and should therefore not be interpreted as a strictly representative sample of all detected changes.

Initial intercoder agreement on the binary presence/absence coding was 92% (79/86). The remaining 7 cases were resolved by consensus, and the reconciled baseline and follow-up values were used in the accuracy analysis.

Table 4 summarises the validation outcomes. CIs reflect sampling uncertainty in the GSV validation and are wider for classes with fewer usable sites. For additions, OSM shows high recall (0.96, 95% CI [0.82–0.99]) and good precision (0.82, 95% CI [0.66–0.91]), indicating that most real additions were detected, although some false positives remain. For removals, recall is 1.00 but with a very wide confidence

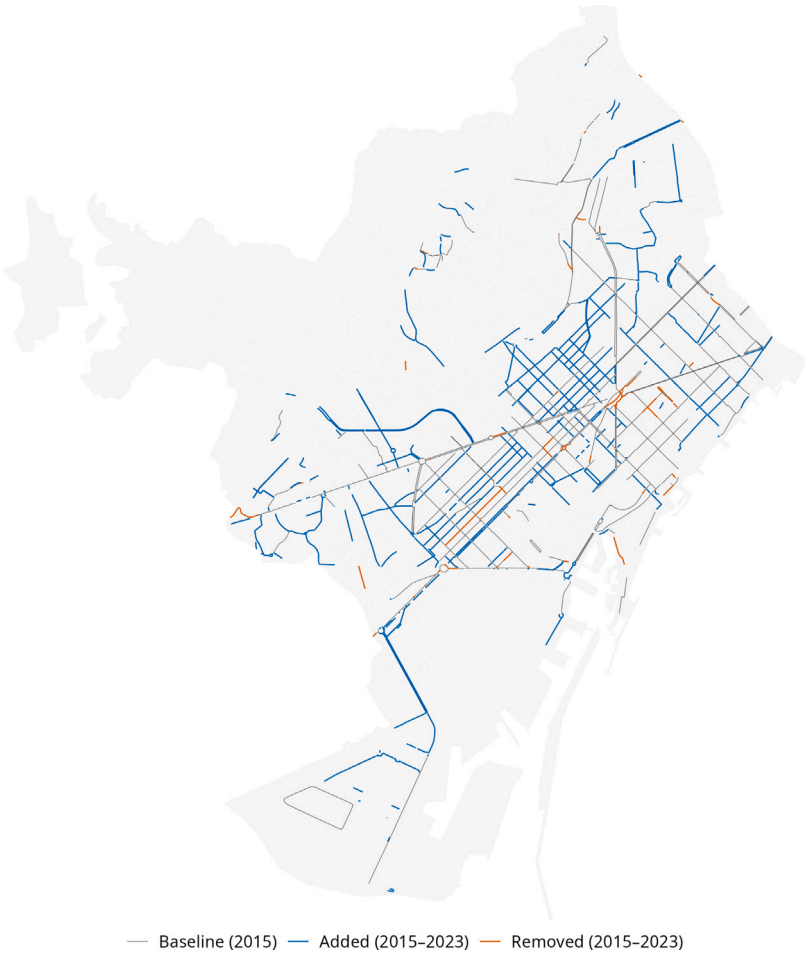


Fig. 4. OSM-detected cycling-infrastructure changes in Barcelona, 2015–2023.

Table 3
Validation points by class and stratum (usable points only).

Stratum	Definition	Add	Remove	NONCYC	Total
D1_C1	Low density, peripheral	5	1	6	12
D1_C2	Low density, intermediate	5	0	4	9
D1_C3	Low density, central	6	0	5	11
D2_C1	Medium density, peripheral	2	2	6	10
D2_C2	Medium density, intermediate	4	0	6	10
D2_C3	Medium density, central	3	1	5	9
D3_C1	High density, peripheral	1	0	6	7
D3_C2	High density, intermediate	3	1	5	9
D3_C3	High density, central	4	0	5	9
TOTAL	All strata combined	33	5	48	86

Table 4
Validation metrics for OSM-inferred cycling-infrastructure changes in Barcelona, 2015–2023.

Class	n (usable)	TP	FP	FN	Precision (95% CI)	Recall (95% CI)	F1
ADD	33	27	6	1	0.82 [0.66-0.91]	0.96 [0.82-0.99]	0.89
REMOVE	5	1	4	0	0.20 [0.04-0.62]	1.00 [0.21-1.00]	0.33
Pooled	38	28	10	1	0.74 [0.58-0.85]	0.97 [0.83-0.99]	0.84

Note: ADD and REMOVE refer to segments present only in 2023 or 2015, respectively; Pooled combines both change types. Only change events are included (n = 33 + 5 = 38); NONCYC sites are used to identify false negatives and are not counted as observations. Confidence intervals are Wilson score intervals.

interval due to the small sample size (n = 5), while precision is very poor (0.20, 95% CI [0.04–0.62]), suggesting that most OSM-detected removals do not correspond to true infrastructure loss. When pooling additions and removals, overall precision is 0.74, recall is 0.97, and the F1 score is 0.84, reflecting strong sensitivity to change but more limited specificity, particularly for removals.

3.3. Sensitivity to geometric parameters

Given the reliance on geometric parameters in Step 2, we assessed the robustness of change-detection results to alternative parameter choices. Sensitivity checks were conducted varying the differencing buffer (5 m, 10 m, 15 m), minimum segment length (5 m, 10 m,

20 m), and the realignment threshold (10 m, 15 m, 20 m). Net change estimates remained broadly stable across configurations, varying within approximately 8% of the baseline. Detected additions showed limited variation across parameter settings, whereas estimated removals were more sensitive, particularly to the realignment threshold. Despite this variability, removals remained substantially smaller than additions in all configurations, resulting in consistently positive net change. Full results are provided in Supplement S1.

4. Discussion

This study quantified the temporal accuracy of OSM-derived cycling-infrastructure change in Barcelona (2015–2023) by combining longitudinal snapshot differencing with stratified GSV validation. To date, the available literature has either validated OSM cycling infrastructure cross-sectionally against reference datasets (Ferster et al., 2020; Hochmair et al., 2015), used OSM snapshots to measure cycling-network growth at scale without externally validating whether detected changes reflect on-street infrastructure change (Winters et al., 2025), or applied street-level imagery to audit built-environment features independently of OSM (Rundle et al., 2011). By designing and implementing a transparent end-to-end workflow combining historical network reconstruction, geometric differencing, stratified spatial sampling, and independent visual verification, we produce reproducible computational outputs and validation-based accuracy estimates (precision, recall, and F1 score) for cycling-infrastructure change detection. Our workflow therefore provides a practical calibration strategy for planning and longitudinal research using OSM when consistent year-on-year inventories are unavailable.

Overall, OSM performed well in detecting cycling-infrastructure additions, with high recall and strong precision indicating that most real expansions were captured, although some false positives remain. For removals, recall was also high, but precision was very low, suggesting that most detected removals do not correspond to actual on-street cycling-infrastructure loss. Estimates for removals, however, are based on a very small number of validated cases ($n = 5$) and should therefore be interpreted with caution.

To understand the lower precision observed for OSM-detected removals relative to additions, it is important to consider how OSM records change over time. Because OSM objects can be split, merged, re-tagged, deleted, or recreated (Minghini and Frassinelli, 2019), structural edits between snapshots may produce apparent removals without corresponding physical change on the street. In practice, snapshot differencing may classify representation changes as removals when segments are re-drawn with new geometries, reclassified under different tags, or merged into adjacent features, even though the underlying infrastructure remains in place. Additions, in contrast, require the introduction of new geometries or tags and may therefore more often correspond to genuinely new on-street features. OSM-based differencing can therefore over-detect removals when edits reflect changes in representation rather than physical interventions. This is consistent with the nature of OSM as a continuously edited VGI resource, in which objects are routinely revised rather than fixed as in a static dataset (Goodchild, 2007).

From a practical perspective, for cities that lack consistently maintained inventories, OSM snapshot differencing can function as a screening tool for identifying likely infrastructure additions at scale, with targeted imagery checks providing an efficient quality-control layer. This matters because practitioners often need to monitor delivery (corridor completion, network coverage growth) and to allocate limited auditing resources; virtual audits have repeatedly been shown to reduce the burden of in-person observation while remaining usable for built-environment measurement (Rundle et al., 2011). Our results suggest that OSM-derived additions can be treated as generally informative, but apparent removals should not be interpreted as “rollback” without independent verification.

Beyond planning practice, these measurement characteristics also have implications for longitudinal research. The high recall and relatively strong precision observed for additions suggest that OSM histories can provide useful proxies for the timing and location of cycling-network expansion in settings where consistent historical inventories are unavailable. This is particularly relevant for studies evaluating infrastructure interventions over time. However, the substantially lower precision observed for removals indicates that inferred infrastructure loss from snapshot differencing should be interpreted cautiously. Over-detected removals may underestimate net network growth and introduce measurement error into longitudinal exposure measures. In this context, validation-derived precision estimates may help calibrate OSM-based change measures by providing an approximate correction for false positives. However, calibration of removals remains uncertain due to the small number of validated removal cases and should be interpreted cautiously.

At a broader conceptual level, this study contributes to efforts to assess the temporal validity of OSM and other forms of VGI. Longitudinal applications of OSM require attention to how edits, reclassification, and representation changes may be expressed as apparent cycling-infrastructure change. The approach presented here provides a practical way to distinguish true interventions from mapping artefacts, supporting more robust use of OSM histories in transport, health, and urban studies.

5. Generalisability and limitations

The workflow developed in this study is designed to be transferable to other cities where historical OSM snapshots, a suitable spatial sampling framework (e.g., census tracts or equivalent small areas) and sufficient street-level imagery are available. The computational components are fully reproducible, while the validation stage relies on structured manual interpretation of imagery following a standardised protocol.

Several limitations should be acknowledged. First, the overall validation sample was relatively small ($n = 86$ usable sites), particularly for removals ($n = 5$), resulting in wide confidence intervals for removal-specific metrics. This small removal sample reflects the limited number of eligible removal candidates in the OSM-detected change pool under our cycling-infrastructure definition and differencing rules, further reduced by realignment filtering and incomplete GSV coverage for both reference periods. Sensitivity checks expanding the number of sampled tracts and candidate removal segments yielded only marginal gains in usable removal sites, indicating that the primary constraint was the rarity and uneven spatial distribution of OSM-detected removals across the city rather than the sampling strategy itself. As a result, the reported removal metrics should be interpreted cautiously and treated as conditional on the sampled validation set rather than as directly generalisable estimates of detection accuracy.

Second, the construction of the NONCYC control pool may influence the interpretation of recall. NONCYC candidates were generated from street segments located outside the cycling-infrastructure buffers used in the differencing workflow. This design reduces ambiguity in the identification of non-cycling segments, but it may reduce the likelihood of identifying false negatives near already mapped cycling facilities, potentially leading to an overestimation of recall. As a result, recall estimates should be interpreted as conditional on the sampling framework rather than as a comprehensive measure of missed cycling-infrastructure changes across the full street network.

Third, the validation relies on GSV as an observational reference rather than direct ground truth, which introduces several related sources of uncertainty. Coverage is not uniform across space, and requiring imagery at both time points may introduce selection bias towards better-covered areas. Imagery dates also vary within each anchor year and across locations, occasionally creating temporal mismatches with

the OSM snapshots that may cause misclassification when infrastructure was built or removed close to the reference dates, a known limitation of GSV-based auditing approaches (Curtis et al., 2013). In practice, fallback imagery was used rarely for the baseline period (2014 imagery for 1 of 86 usable sites) but more frequently for the follow-up period (2022 imagery for 19 of 86 usable sites). The visibility of cycling infrastructure also varies by type — segregated tracks are generally easier to identify than painted or shared facilities — which can affect both coding consistency and inter-rater agreement, particularly for borderline or mixed designs. Previous studies and reviews of GSV-based auditing have similarly noted that image resolution, spatial and temporal coverage, and feature characteristics can affect audit reliability, with lower agreement often observed for small, temporally variable, obstructed, or visually ambiguous features (Rzotkiewicz et al., 2018; Hanibuchi et al., 2019; Rundle et al., 2011). Together, these factors mean that precision and recall should be interpreted as agreement conditional on observable locations rather than as a measure of absolute ground-truth accuracy.

Fourth, the validation stage becomes increasingly resource-intensive at larger scales because currently available GSV APIs do not provide systematic access to historical Street View imagery across multiple time points. As a result, longitudinal validation still relies heavily on manual review of historical panoramas through the Street View interface, and auditing effort increases with sample size.

Finally, Barcelona likely represents a relatively favourable OSM context, given its active mapping community, extensive cycling-infrastructure development, and relatively high mapped completeness (Herfort et al., 2023). Future research should therefore assess how this workflow performs in cities with different levels of OSM completeness and mapping activity.

6. Conclusions

This study quantified the accuracy of OSM for detecting longitudinal cycling-infrastructure change by implementing and validating a reproducible snapshot-differencing framework in Barcelona (2015–2023). Using probability-based stratified GSV validation, we estimated precision, recall, and F1 scores for OSM-detected additions and removals.

The results indicate that OSM-detected additions are generally reliable in this context, with high recall and good precision suggesting that most network expansions are captured. In contrast, apparent removals identified through snapshot differencing often do not correspond to on-street cycling-infrastructure loss; however, this evidence is based on a very small number of validated cases ($n = 5$) and is therefore subject to substantial uncertainty. These findings suggest that OSM histories can be informative for tracking network expansion, while inferred removals should be interpreted cautiously and, where possible, verified independently.

By producing quantitative accuracy estimates within a transparent workflow with reproducible computational components and structured manual validation, this study provides a practical basis for calibrating OSM-derived measures of cycling-infrastructure change and for accounting for measurement error in longitudinal evaluations. The transferability of the framework to other settings will depend on the availability of historical OSM snapshots, suitable spatial units to support stratified sampling, and sufficient street-level imagery.

More broadly, the findings support the use of OSM and other VGI sources for longitudinal urban change analysis, while underscoring the importance of validating temporal change signals before interpreting mapped edits as real-world infrastructure change.

CRedit authorship contribution statement

Eugeni Vidal-Tortosa: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Víctor**

González-Parra: Writing – review & editing, Investigation. **Oriol Marquet:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript the authors used ChatGPT (version 5.2) to support language editing and improve readability. All content was subsequently reviewed and edited by the authors, who take full responsibility for the final manuscript.

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.

Acknowledgements

This research was co-funded by the European Union (ERC, ATRAPA, 101117700). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. V. González-Parra was supported by the PhD grant AGAUR-FI (2025FI-100999) Joan Oró of the Department of Research and Universities of the Generalitat de Catalunya and the European Social Fund Plus.

We thank members of the GEMOTT research group for helpful comments and discussions, and the OpenStreetMap community for their contributions.

Supplements

The supplementary materials accompanying this article comprise the following files:

S1. Sensitivity analysis of change-detection parameters (XLSX)

Provides the results of sensitivity checks for key geometric parameters used in the OSM change-detection workflow, including the differencing buffer, realignment threshold, and minimum segment length.

The table reports the total length of detected additions, removals, and net change under alternative parameter settings, along with deviations from the baseline configuration.

The baseline scenario corresponds to the default parameter configuration used in the analysis (10 m differencing buffer, 15 m realignment threshold, and 10 m minimum segment length). Values are computed on the filtered evaluation dataset (after excluding short fragments and likely realignments) and therefore differ from the raw network-level change estimates reported in Table 1.

S2. Raw validation workbook (XLSX)

Contains the sampled segments and coding sheets completed independently by the two coders prior to reconciliation.

S3. Interactive validation points map (HTML)

Interactive HTML visualisation displaying validation points with direct Google Street View links. The map supports zooming, panning, layer toggling, and inspection of individual validation points. It can be

opened locally in any modern web browser (an internet connection is required to load basemap tiles).

S4. Google Street View validation protocol (PDF)

Details the standardised coding procedure, imagery selection rules, and reconciliation process.

S5. Reconciled validation results (XLSX)

Provides the reconciled validation dataset used to compute precision, recall, and F1 scores.

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jtrangeo.2026.104775>.

Data availability

The code and validation materials supporting this study are available in the supplementary files and at <https://github.com/GEMOTT/osm-cycling-change-framework>.

References

- Ajuntament de Barcelona, 2023. Carril bici dataset (2023 T4). <https://opendata-ajuntament.barcelona.cat/data/es/dataset/carril-bici>.
- Almendros-Jiménez, Jesús M., Becerra-Terón, Antonio, 2018. Analyzing the tagging quality of the Spanish OpenStreetMap. *ISPRS Int. J. Geo-Inf.* 7 (8), 323. <http://dx.doi.org/10.3390/ijgi7080323>.
- Àrea Metropolitana de Barcelona, 2015. Carril bici dataset (december 2015). Unpublished dataset.
- Askari, Sajad, Snyder, Devon, Li, Chu, Saugstad, Michael, Froehlich, Jon E., Eisenberg, Yochai, 2025. Validating pedestrian infrastructure data: How well do street-view imagery audits compare to government field data? *Urban Sci.* 9 (4), 130. <http://dx.doi.org/10.3390/urbansci9040130>.
- Barron, Christopher, Neis, Pascal, Zipf, Alexander, 2014. A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Trans. GIS* 18 (6), 877–895. <http://dx.doi.org/10.1111/tgis.12073>.
- Batomen, Brice, Wen, Richard, Lyeo, Joonsoo, et al., 2026. Pedaling towards safer streets: Evaluating the impact of cycling infrastructure on road safety. *Accid. Anal. Prev.* 225 (February), 108308. <http://dx.doi.org/10.1016/j.aap.2025.108308>.
- Bres, Raphaël, Peralta, Veronika, Le-Guilcher, Arnaud, Devogele, Thomas, Olteanu Raimond, Ana-Maria, De Runz, Cyril, 2023. Analysis of cycling network evolution in OpenStreetMap through a data quality prism. *AGILE: GIScience Ser.* 4 (June), 1–9. <http://dx.doi.org/10.5194/agile-giss-4-3-2023>.
- Buehler, Ralph, Pucher, John (Eds.), 2021. *Cycling for sustainable cities*. In: *Urban and Industrial Environments*. MIT Press.
- Curtis, Jacqueline W., Curtis, Andrew, Mapes, Jennifer, Szell, Andrea B., Cinderich, Adam, 2013. Using Google street view for systematic observation of the built environment: Analysis of spatio-temporal instability of imagery dates. *Int. J. Health Geogr.* 12 (1), 53. <http://dx.doi.org/10.1186/1476-072X-12-53>.
- Dai, Shaoqing, Li, Yuchen, Stein, Alfred, Yang, Shujuan, Jia, Peng, 2024. Street view imagery-based built environment auditing tools: A systematic review. *Int. J. Geogr. Inf. Sci.* 38 (6), 1136–1157. <http://dx.doi.org/10.1080/13658816.2024.2336034>.
- Ferster, Colin, Fischer, Jaimy, Manaugh, Kevin, Nelson, Trisalyn, Winters, Meghan, 2020. Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities. *Int. J. Sustain. Transp.* 14 (1), 64–73. <http://dx.doi.org/10.1080/15568318.2018.1519746>.
- Ferster, Colin, Nelson, Trisalyn, Manaugh, Kevin, Beairisto, Jeneva, Laberee, Karen, Winters, Meghan, 2023. Developing a national dataset of bicycle infrastructure for Canada using open data sources. *Environ. Plan. B: Urban Anal. City Sci.* 500 (9), 2543–2559. <http://dx.doi.org/10.1177/23998083231159905>.
- Gilardi, Andrea, Lovelace, Robin, 2025. Osmextract: Download and import OpenStreetMap data extracts. <https://docs.ropensci.org/osmextract/>.
- Girres, Jean-François, Touya, Guillaume, 2010. Quality assessment of the French OpenStreetMap dataset. *Trans. GIS* 14 (4), 435–459. <http://dx.doi.org/10.1111/j.1467-9671.2010.01203.x>.
- Goodchild, Michael F., 2007. Citizens as sensors: The world of volunteered geography. *GeoJournal* 69 (4), 211–221. <http://dx.doi.org/10.1007/s10708-007-9111-y>.
- Haklay, Mordechai, 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plan. B: Plan. Des.* 37 (4), 682–703. <http://dx.doi.org/10.1068/b35097>.
- Hanibuchi, Tomoya, Nakaya, Tomoki, Inoue, Shigeru, 2019. Virtual audits of streetscapes by crowdworkers. *Health & Place* 59 (September), 102203. <http://dx.doi.org/10.1016/j.healthplace.2019.102203>.
- Herfort, Benjamin, Lautenbach, Sven, Porto de Albuquerque, João, Anderson, Jennings, Zipf, Alexander, 2023. A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. *Nat. Commun.* 14 (1), 3985. <http://dx.doi.org/10.1038/s41467-023-39698-6>.
- Hirsch, Jana A., Meyer, Katie A., Peterson, Marc, et al., 2016. Obtaining longitudinal built environment data retrospectively across 25 years in four US cities. *Front. Public Health* 4 (April), 0. <http://dx.doi.org/10.3389/fpubh.2016.00065>.
- Hochmair, Hartwig H., Zielstra, Dennis, Neis, Pascal, 2015. Assessing the completeness of bicycle trail and lane features in OpenStreetMap for the united states. *Trans. GIS* 19 (1), 63–81. <http://dx.doi.org/10.1111/tgis.12081>.
- Houde, Maxime, Apparicio, Philippe, Séguin, Anne-Marie, 2018. A ride for whom: Has cycling network expansion reduced inequities in accessibility in montreal. *Can. J. Transp. Geogr.* 68 (April), 9–21. <http://dx.doi.org/10.1016/j.jtrangeo.2018.02.005>.
- Juhász, Levente, 2021. Towards understanding the temporal accuracy of OpenStreetMap: A quantitative experiment. <http://dx.doi.org/10.5281/ZENODO.5112236>, July.
- Minghini, Marco, Frassinelli, Francesco, 2019. OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date? *Open Geospatial Data Softw. StandArds* 4 (1), 9. <http://dx.doi.org/10.1186/s40965-019-0067-x>.
- OpenStreetMap contributors, 2026. Bicycle - OpenStreetMap wiki. <https://wiki.openstreetmap.org/wiki/Bicycle>. (Accessed 30 January 2026).
- Prince, Stephanie A., Thomas, Tyler, Apparicio, Philippe, et al., 2025. Cycling infrastructure as a determinant of cycling for recreation and transportation in montreal, canada: a natural experiment using the longitudinal national population health survey. *Int. J. Behav. Nutr. Phys. Act.* 22 (June), 71. <http://dx.doi.org/10.1186/s12966-025-01767-y>.
- Rundle, Andrew G., Bader, Michael D.M., Richards, Catherine A., Neckerman, Kathryn M., Teitler, Julien O., 2011. Using Google street view to audit neighborhood environments. *Am. J. Prev. Med.* 40 (1), 94–100. <http://dx.doi.org/10.1016/j.amepre.2010.09.034>.
- Rzotkiewicz, Amanda, Pearson, Amber L., Dougherty, Benjamin V., Shortridge, Ashton, Wilson, Nick, 2018. Systematic review of the use of Google street view in health research: Major themes, strengths, weaknesses and possibilities for future research. *Health & Place* 52, 240–246. <http://dx.doi.org/10.1016/j.healthplace.2018.07.001>.
- Schmidl, Martin, Navratil, Gerhard, Giannopoulos, Ioannis, 2021. An approach to assess the effect of currentness of spatial data on routing quality. *AGILE: GIScience Ser.* 2 (June), 1–12. <http://dx.doi.org/10.5194/agile-giss-2-13-2021>.
- Senaratne, Hansi, Mobasheri, Amin, Ali, Ahmed Loai, Capineri, Cristina, Haklay, Mordechai (Muki), 2017. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* 31 (1), 139–167. <http://dx.doi.org/10.1080/13658816.2016.1189556>.
- Szell, Michael, Mimar, Sayat, Perlman, Tyler, Ghoshal, Gourab, Sinatra, Roberta, 2022. Growing urban bicycle networks. *Sci. Rep.* 12 (1), 6765. <http://dx.doi.org/10.1038/s41598-022-10783-y>.
- Vierø, Ane Rahbek, Vybornova, Anastassia, Szell, Michael, 2025. How good is open bicycle network data? A countrywide case study of Denmark. *Geogr. Anal.* 57 (1), 52–87. <http://dx.doi.org/10.1111/gean.12400>.
- Winters, Meghan, Ferster, Colin, Laberee, Karen, 2025. Mapping change in cycling infrastructure across Canada: What, where, and for whom? *Can. J. Public Health* <http://dx.doi.org/10.17269/s41997-025-01139-w>, ahead of print, December.
- Zhang, Liming, Pfoser, Dieter, 2019. Using OpenStreetMap point-of-interest data to model urban change—A feasibility study. *PLoS ONE* 14 (2), e0212606. <http://dx.doi.org/10.1371/journal.pone.0212606>.