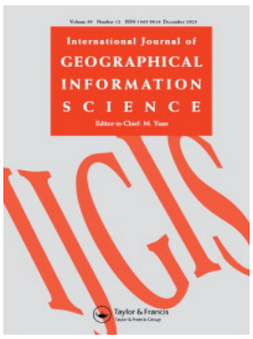


International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: www.tandfonline.com/journals/tgis20

Bidirectional enrichment of OpenStreetMap and CityGML 3.0 to facilitate cycling safety assessment

Weilian Li, Jannik Matijevic, Christof Beil, Lukas Arzoumanidis, Thomas H. Kolbe & Youness Dehbi

To cite this article: Weilian Li, Jannik Matijevic, Christof Beil, Lukas Arzoumanidis, Thomas H. Kolbe & Youness Dehbi (29 Dec 2025): Bidirectional enrichment of OpenStreetMap and CityGML 3.0 to facilitate cycling safety assessment, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2025.2595662

To link to this article: https://doi.org/10.1080/13658816.2025.2595662

Published online: 29 Dec 2025.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2025.2595662

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Bidirectional enrichment of OpenStreetMap and CityGML 3.0 to facilitate cycling safety assessment

Weilian Li[a,b], Jannik Matijevic[b], Christof Beil[c], Lukas Arzoumanidis[b], Thomas H. Kolbe[c] and Youness Dehbi[b]

[a]Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu, China; [b]Computational Methods Lab, HafenCity University Hamburg, Hamburg, Germany; [c]TUM School of Engineering and Design, Technical University of Munich, Munich, Germany

ABSTRACT

Safety concerns remain a barrier to the widespread adoption of cycling. Assessing cycling safety facilitates planning safer cycling routes, which helps to boost cycling confidence. However, existing research primarily concentrates on assessing cycling safety at regional or urban levels, with few studies assessing safety at the road segment level, often without considering detailed lane information. This article leverages the complementary strengths of OSM's rich semantic information on roads and CityGML with lane-level geometry to facilitate cycling safety assessment. Precisely, an informed map matching using Kernel Density Estimation (KDE) for bidirectional attribute transfer, cycling safety scores calculation, and CityGML enrichment with cycling safety are introduced in detail. OpenDRIVE data from the Test Track for Autonomous and Connected Driving (TAVF) in Hamburg, Germany, was converted to a CityGML 3.0-compliant structure using the *r:trån* tool and used together with the corresponding OSM data for experimental analysis. The experimental results show that integrating OSM and CityGML is conducive to improving cycling safety assessment at the road segment level. The assessment results are further embedded into bicycle-related semantics within CityGML 3.0 for the subsequent 3D representation of cycling safety, paving the way for safest path navigation and enhanced perception of cycling safety in 3D environments.

ARTICLE HISTORY
Received 10 April 2025
Accepted 22 November 2025

KEYWORDS
Cycling; OpenStreetMap; CityGML 3.0; safety assessment; Smart mobility

1. Introduction

Cycling has been globally promoted as a sustainable mode of transportation (Cao *et al*. 2020, Meng and Zheng 2023), offering benefits such as reducing air pollution (Biassoni *et al*. 2023), alleviating congestion (Ahmed *et al*. 2024), and improving urban livability (Li *et al*. 2024b, Ito *et al*. 2024). However, cyclists are vulnerable road users

CONTACT Youness Dehbi youness.dehbi@hcu-hamburg.de

with a high risk of injury and fatalities compared to motorists (Heeremans *et al*. 2022, Scarano *et al*. 2023, Panagiotaki *et al*. 2025), and the perceived safety risks in mixed traffic remains a significant barrier to the widespread adoption of cycling (Jacobsen 2015, Félix *et al*. 2019, Shoman *et al*. 2023).

In addition to investing in improved cycling infrastructure (Pucher and Buehler 2016), assessing the cycling safety of existing street networks can help cyclists plan the safest routes, particularly boosting the cycling confidence of older and novices (Schepers *et al*. 2017, Mora *et al*. 2021). However, to the authors' best knowledge, current research mainly focuses on developing comprehensive index for bikeability or safety from a macro perspective, with few studies addressing data-driven cycling safety assessment at a road-segment level (DiGioia *et al*. 2017, Götschi *et al*. 2018). Geospatial data plays an indispensable role in understanding and assessing cycling safety (Basiri *et al*. 2019). High-quality, detailed datasets are crucial for analyzing risk factors, identifying hazardous locations, and implementing targeted interventions to improve cycling infrastructure (Schepers *et al*. 2014).

OpenStreetMap(OSM)[1], with its extensive global coverage and detailed mapping of transportation networks, provides valuable data for cycling safety assessment (Vargas-Munoz *et al*. 2021, Herfort *et al*. 2023). The critical advantage of OSM-based assessments is the ease of data access, which enables large-scale cycling safety evaluations across different cities. However, given that OSM is collected through crowd-sourcing, the completeness and quality are heterogeneous across different geographical locations (Fan *et al*. 2016, Senaratne *et al*. 2017, Vargas-Munoz *et al*. 2021). Generally, OSM provides network-based parametric representations of street geometries (Beil *et al*. 2023). However, crucial information such as lane width, which is essential for the cycling safety assessment, is often incomplete in OSM (Li *et al*. 2024a).

CityGML[2] is an open standard released by the Open Geospatial Consortium (OGC) for storing and exchanging virtual 3D city models (Kolbe *et al*. 2005, Gröger and Plümer 2012, Biljecki *et al*. 2018). The newest CityGML version 3.0 can store lane-level geometry when such detail is supplied by the source data, in our case OpenDRIVE data of Hamburg's TAVF test field that we converted to CityGML with the *r:trån* tool (Schwab and Kolbe 2022, Schwab *et al*. 2020). The resulting dataset provides precise outlines of individual surfaces, opening new possibilities for comprehensive safety assessments of cycling routes (Beil *et al*. 2020, Kutzner *et al*. 2020, Kolbe *et al*. 2021). Moreover, integrating the results of cycling safety assessments into CityGML could further enrich bicycle-related semantics, paving the way for enhanced cycling safety perception in a 3D environment.

In this context, we aim to leverage the complementary strengths of OSM and CityGML to facilitate cycling safety assessment. On the one hand, detailed lane information can be retrieved from CityGML to augment the completeness of OSM for cycling safety assessment. On the other hand, cycling safety scores could enrich the semantic attributes of bicycle infrastructure within CityGML. The main contributions of this article are as follows:

1. The adoption of Kernel Density Estimation (KDE) to calculate emission probabilities in a Hidden Markov Model (HMM) for map matching. Rather than relying on

the commonly assumed Gaussian distribution for modeling spatial distances in probabilistic map matching, we employ KDE to capture the distribution of distances between OSM and CityGML, allowing for an informed and, hence, more accurate map matching. This non-parametric approach avoids the limitations of assuming normality and more accurately reflects the empirical characteristics of the data, particularly in complex urban environments such as road intersections. By incorporating this more accurate prior knowledge, KDE improves the reliability of the matching process and reduces mismatches that often occur under the Gaussian assumption, thereby enabling the accurate transfer of attributes between OSM and CityGML.

2. The bidirectional enrichment of OpenStreetMap and CityGML 3.0 to facilitate cycling safety assessment. On the one hand, the extracted additional indicators from CityGML, such as lane width, can be integrated with OSM to improve its data completeness, enabling a more comprehensive assessment of cycling safety. Conversely, the derived cycling safety scores can be inversely embedded into CityGML 3.0, enriching its bicycle-related semantic attributes for road infrastructure. Additionally, the development of a 3D visualization of cycling safety, offers an intuitive visual representation of safety conditions, which can serve as the basis for safest path navigation and perception.

The study workflow of this article is shown in Figure 1. Based on OSM and CityGML data, we first derive a list of indicators that contribute to cycling safety. Hidden Markov Model (HMM) is employed to match OSM and CityGML, enabling attributes

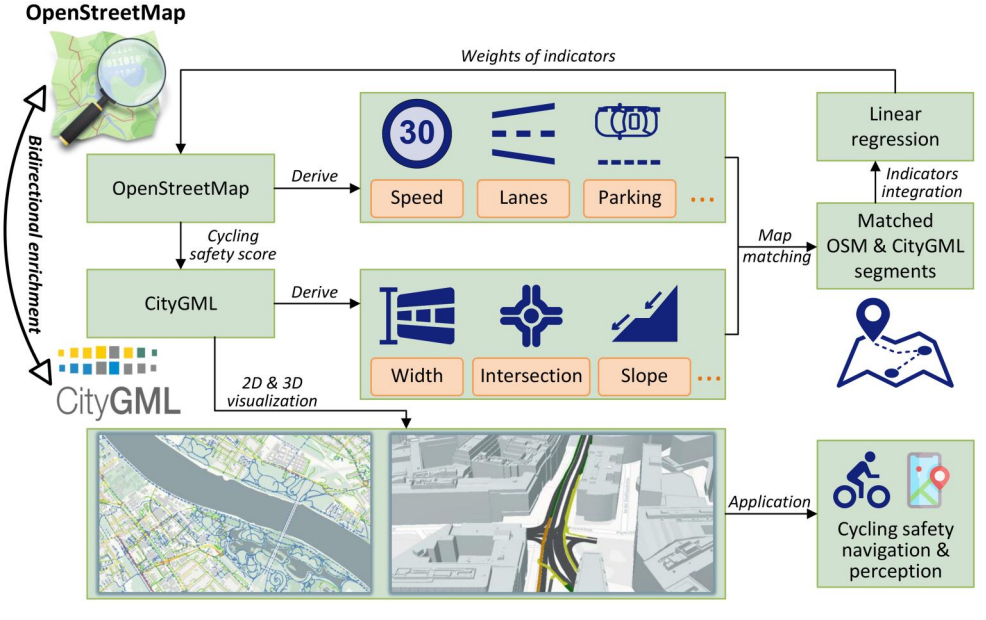


Figure 1. Overview of the study workflow. The process begins with deriving the cycling safety indicators from OSM and CityGML. Then, the extracted indicator attributes are integrated through map matching, and MLR is applied to determine the weights. Finally, the calculated cycling safety scores are embedded into CityGML 3.0 for 3D representation.

transfer between them. Once all necessary indicators are obtained, we use Multivariate Linear Regression (MLR) to induce the weights of these indicators by learning them from the observed data. The resulting weights facilitate the calculation of cycling safety scores, which can be further embedded into bicycle-related semantics in CityGML 3.0 for the subsequent 3D representation of cycling safety.

The remainder of this article is structured as follows: Section 2 reviews the related work of this study. Section 3 introduces the experimental data used in this study. Section 4 provides a detailed explanation of the methodology. Section 5 analyzes the experimental results. Section 6 summarizes the article and gives an outlook for future research.

2. Related work

Amid the pressures of urbanization and booming population, cities worldwide are increasingly promoting cycling as a mode of active mobility to strive for a more sustainable future (Götschi *et al*. 2018, Li *et al*. 2024b). This trend became especially evident after the COVID-19 pandemic, as people sought to avoid crowded spaces and engage in contactless outdoor activities (Buehler and Pucher 2021). However, car-centric urban development over the past decades has primarily considered vehicular traffic needs, often neglecting cycling and resulting in significant safety concerns (Meuleners *et al*. 2019, Scarano *et al*. 2023). Although many governments are working to improve cycling safety to alleviate the substantial health burden caused by bicycle crashes, the lack of safety for cyclists as vulnerable road users in mixed traffic remains a key challenge in promoting cycling (Wegman *et al*. 2012, Kamel *et al*. 2020, Heeremans *et al*. 2022).

Assessing cycling safety on existing roads is an effective way to encourage more people to ride with confidence (DiGioia *et al*. 2017). The most straightforward assessment approach involves large-scale public surveys to gather people's positive or negative opinions on cycling safety. For example, the German National Cyclists' Association (ADFC) conducts one of the largest surveys on the cycling climate worldwide[3], nearly 230,000 citizens participated in 2022, and the 11th survey results from 2024 will open soon. The survey's goal is to gather participants' ratings on cycling conditions, such as cycling safety, comfort, and infrastructure completeness, thus provide valuable insights that municipalities can use to make targeted improvements to the infrastructure. While this survey offers a valuable macro-level overview of cycling safety across cities, it is less useful for providing specific suggestions for individual cycling trips.

Similarly, statistical analysis of bicycle accidents can also be used to quantitatively assess cycling safety at regional or urban levels (Daraei *et al*. 2021). For example, Boss *et al*. (2018) used network kernel density estimation to monitor changes in the spatiotemporal distribution of cycling incidents in Vancouver, Canada. Alternatively, multiple-criteria decision-making (MCDM) is widely used to assess bikeability and can also be applied to cycling safety assessments (Lowry *et al*. 2012, Derek and Sikora 2019). The most striking feature of MCDM is that the weights of indicators are obtained directly from experts through a pairwise comparison of indicators. These weights are subsequently used to calculate the final score of cycling safety using analytic hierarchy process (Wang *et al*. 2024). However, it is challenging to judge which factor has a greater impact on cycling safety through pairwise comparisons,

an alternative approach is to use machine learning (e.g., Multivariate Linear Regression) that induces the weights of these indicators by learning the observed data on safety indicators (Li *et al*. 2024a).

Regardless of the data-driven assessment approach employed, geospatial data is essential for cycling safety analysis at the road segment level. OSM offers detailed mapping of transportation networks, providing valuable semantic road information for cycling safety assessments. In this context, Li *et al*. (2024a) propose a safety assessment method for cycling routes in urban environments based on OSM data. However, OSM data is often incomplete due to its crowdsourced nature (Feng *et al*. 2022), lacking crucial information such as lane width, which is vital for cycling safety assessment. The OpenDRIVE dataset, transformed into CityGML 3.0, offers highly detailed road geometry and rich semantic information, compensating for OSM's data completeness shortcomings (Kolbe *et al*. 2021). Leveraging the complementary strengths of OSM's rich road information and CityGML's highly detailed road geometry could enable a comprehensive cycling safety assessment, a task that, to the authors' knowledge, has not yet been explored in existing literature.

Integrating OSM and CityGML facilitates the comprehensive cycling safety assessment at the road segment level, and the resulting assessment contributes to enriching bicycle-related semantics within CityGML 3.0. However, this study must address two critical challenges: First, establishing geometric alignment between the two datasets is a prerequisite for mapping CityGML road semantics to OSM and embedding the cycling safety scores back into CityGML 3.0. Second, deriving the weights of safety indicators by learning from observed data incorporating new observations, e.g., lane width, and using them to calculate the cycling safety scores.

3. Data

As previously mentioned, this study utilizes two types of data: OSM and CityGML, and a snapshot of partial data is shown in Figure 2. The OSM street network for the study area, Hamburg, Germany, was obtained using OSMnx[4], a Python library that simplifies the retrieval and analysis of OSM data.

The process of retrieving CityGML data is shown in Figure 3. A UAV survey captured the original dataset, delivering road geometries accurate to 10 cm, which was then converted into an OpenDRIVE file using Trian3DBuilder software[5]. The resulting OpenDRIVE dataset was initially developed for the Test Track for Autonomous and Connected Driving (TAVF) in Hamburg, Germany. The CityGML data was derived from the OpenDRIVE dataset using the *r:trån* tool[6] (Schwab and Kolbe 2022), following the workflow described in Schwab *et al*. (2020).

The conversion process from the OpenDRIVE dataset to the CityGML 3.0 dataset starts by placing the data in a projected CRS (e.g., UTM zone 32 N). Each lane section is resolved into an explicit surface based on its width profile. Lanes designated as "Biking" are exported as CityGML Transportation features (e.g., TrafficArea) with an appropriate class/function indicating a cycle lane, while relevant source attributes (e.g., lane identifiers) are preserved in the feature. Lane predecessor/successor relations are validated and retained in the OpenDRIVE domain, yielding a geo-referenced, semantically rich surface model for subsequent analysis.

(a) Road network from OSM. (b) Lane information from CityGML.

Figure 2. An excerpt of the data used for this study. (a) Road network from OSM, including number of lanes, speed limit, etc. (b) Lane information from CityGML, including lane width, intersections, etc.

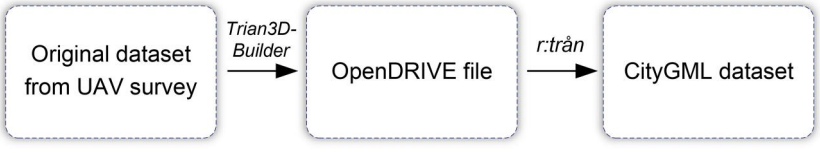


Figure 3. The retrieval process of CityGML dataset.

4. Methodology

4.1. Indicators extraction for cycling safety

In addition to extracting speed limits, the number of lanes, on-street parking, and lane types from OSM (Li *et al.* 2024a), this article seeks to obtain additional highly detailed lane information (e.g., width) from CityGML to augment the dataset for cycling safety assessment. For this purpose, the Feature Manipulation Engine (FME)[7], a visual workflow-based data integration platform, is used to retrieve and transform bicycle information from CityGML.

The retrieval process is shown in Figure 4. In the case of CityGML, a *UniqueIdentifier* Generator transformer assigns unique IDs to each traffic area, ensuring bidirectional enrichment between CityGML and OSM following the retrieval and matching processes. To align the 3D traffic areas from CityGML with the 2D OSM network, the *HullReplacer* transformer creates a concave hull around the filtered traffic areas, followed by the *CenterlineReplacer* transformer, which generates centerlines from these hulls. On the OSM side, the *AttributeFilter* transformer extracts lane types relevant to cycling safety, e.g. speed limit. Finally, the *NeighborFinder* transformer in FME is used to calculate the shortest distances between segments of the filtered OSM network graph and the centrelines of the cycling areas in the CityGML dataset. These distances are then manually inspected in descending order, inspecting only the largest values, until an appropriate maximum matching distance is determined, ensuring accurate

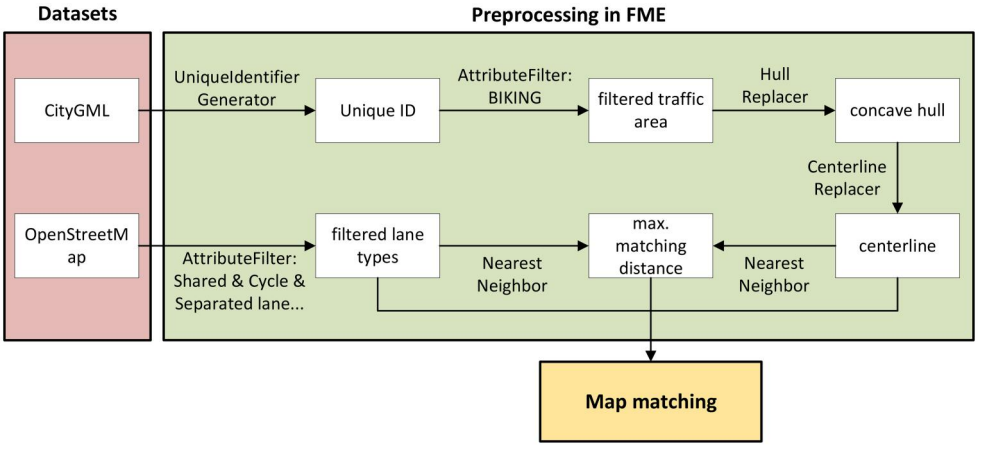


Figure 4. The retrieval process of indicators in FME. CityGML is used to extract the centerline and corresponding biking attributes, which are then utilized for the subsequent map matching with OSM. In addition, orthogonal distances between CityGML centrelines and nearby OSM candidates are sampled to build the residual set used to train the KDE emission model.

spatial alignment for the subsequent matching process. The same distance window ($d_{\max} = 10\,\text{m}$) is later used to constrain the candidate set for HMM emissions.

4.2. Mapping CityGML road semantics into OSM

4.2.1. OSM and CityGML matching

Figure 5 shows the matching workflow of OSM and CityGML. In this diagram, the OSM road segments and CityGML road centerlines serve as inputs, and the HMM algorithm is used to align them (Ghahramani 2001), enabling subsequent attribute transfer.

The detailed working process is as follows:

1. OSM and CityGML processing

   The T road centerlines are sampled into points with 1 m spacing, which are grouped into point sequences $O = \{o_1, \ldots, o_t, \ldots, o_T \mid 1 \le t \le T\}$, where each sequence $o_t = \{o_t^1, \ldots, o_t^n, \ldots, o_t^N \mid 1 \le n \le N\}$ represents the sampled points along the t^{th} road centerline. Here, $N = \mathcal{L}(o_t) + 1$, with $\mathcal{L}$ denoting the length of the t^{th} centerline. N could vary in different sequences.

   Similar to Lucks *et al.* (2021), we opted for a segment-wise matching to improve the computational efficiency. Thus, we create a buffer B_t with a radius of 10 m around each sequence o_t to extract the road segment e that intersects with the buffer, as shown in Equation 1.

$$E_t = \{e \in E \mid E \cap B_t \neq \emptyset\} \tag{1}$$

   where E is the original OSM road network, and $E_t = \{e_t^1, \ldots, e_t^m, \ldots, e_t^M\}$ denotes the set of road segments clipped from E that intersect with the buffer B_t.

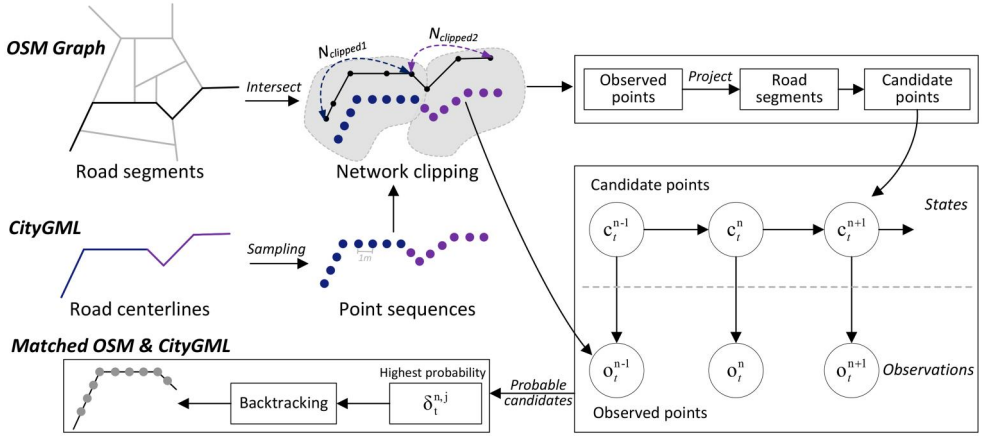


Figure 5. Matching workflow of OSM and CityGML. The road centerlines from CityGML are sampled into point sequences, which serve as the observed points. The orthogonal projections of these observed points onto OSM road segments are identified as the candidate points. Both the observed and candidate points are then input into the HMM map matching algorithm to determine the candidates with the highest probability.

2. Candidate preparation
   Given a set of road segments $E_t = \{e_t^1, ..., e_t^m, ..., e_t^M\}$ within the buffer B_t, each observed point o_t^n is projected onto the nearby road segments, resulting in candidate points $c_t^{n,s}$ where $s \geq 1$, representing the s^{th} candidate point for the n^{th} observed point within the t^{th} buffer, as shown in Figure 6. In our case, as the route directions of CityGML and OSM are roughly the same, there is only one candidate point in most cases, but multiple candidate points may occur at intersections.
3. HMM for map matching
   Taking the HMM-based matching process for OSM and CityGML within buffer B_t as an example, the process involves the following key components:

- Observations: A set of observed points $o_t = \{o_t^1, ..., o_t^n, ..., o_t^N | 1 \leq n \leq N\}$, where N represents the number of observed points within buffer B_t.

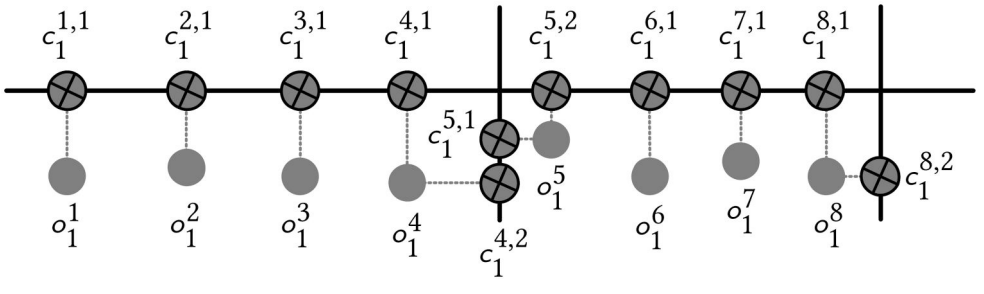


Figure 6. Example of candidate points for the observed points. Gray circles represent the observed points, while dark gray crosses indicate the orthographic projections of the observed points onto road segments, referred to as candidate points.

- States: A set of candidate points $c_t = \{c_t^{1,s}, ..., c_t^{n,s}, ..., c_t^{N,s} | 1 \leq n \leq N\}$, where n represents the index of observed points, and s indicates the possible candidate points, which are greater than or equal to 1. For example, o_1^4 has two candidate points $c_1^{4,1}$ and $c_1^{4,2}$.
- Transition probabilities: The transition probability p represents the likelihood of moving from the candidate points $c_t^{n-1,s'}$ to $c_t^{n,s}$ between the observed point sequence o_t^{n-1} and o_t^n. Here, s' denotes the candidate points of the previous position, and s refers to the candidate points of the current position. This probability accounts for the difference between the candidates along the road network and the observed movement, as shown in Equation 2.

$$p = \frac{1}{2\beta} e^{-\frac{|\,||c_t^{n-1,s'} - c_t^{n,s}||_{\text{route}} - ||o_t^{n-1} - o_t^n||_{\text{euclidean}}\,|}{\beta}} \tag{2}$$

  where $||c_t^{n-1,s'} - c_t^{n,s}||_{\text{route}}$ represents the distance along the network between candidate points $c_t^{n-1,s'}$ and $c_t^{n,s}$, $||o_t^{n-1}, o_t^n||_{\text{euclidean}}$ denotes the Euclidean distance between observed points o_t^{n-1} and o_t^n. The parameter β captures the variability in transition distances, which in our case is set to 1.
- Emission probability: Instead of a commonly assumed Gaussian distribution, we use a data-driven Kernel Density Estimation (KDE) distribution to calculate the emission probability of an observed point o_t^n given a candidate point $c_t^{n,s}$. Specifically, we compute the orthogonal distances between the observed points and the candidate points, e.g., $d(c_t^{n,s}, o_t^n) = ||c_t^{n,s} - o_t^n||_\perp$, to build a residual set $\mathcal{D} = \{d_i\}_{i=1}^I$. KDE is then applied to fit the distribution of all these real distances, as shown in Equation 3.

$$\hat{f}(d) = \frac{1}{Ih} \sum_{i=1}^{I} K\left(\frac{d - d_i}{h}\right), d \geq 0 \tag{3}$$

  where $\hat{f}(d)$ is the estimated probability density function for the distance d, d_i represents the individual distances, and I is the total number of distances in the residual set $\mathcal{D}$. The bandwidth h is determined by Silverman's rule (Silverman 1986), and K is the kernel function used for KDE, which, in our case, is the Gaussian kernel. Once the probability density function $\hat{f}(d)$ is created, the emission probability w of an observed point o_t^n given a candidate point $c_t^{n,s}$ can be obtained by evaluating the KDE at their orthogonal distance $d(c_t^{n,s}, o_t^n)$.

For the sake of replicability, the generic flow of the HMM-based matching process for OSM and CityGML is outlined in Algorithm 1.

Algorithm 1: The generic flow of OSM and CityGML matching

Input : a set of road segments $E_t = \{e_t^1, ..., e_t^m, ..., e_t^M\}$, a set of observed points $o_t = \{o_t^1, ..., o_t^n, ..., o_t^N\}$ where $1 \leq n \leq N$

Output : a set of candidate points c_t, transition probability p_t, emission probability w_t

initialize c_t, p_t, w_t as empty lists;
for $2 \leq n \leq N$ **do**
  $c_t^{n,s} \leftarrow \text{GetCandidates}(o_t^n, E_t)$;
  add $c_t^{n,s}$ to c_t;
  $p \leftarrow$ Calculate the transition probability from $c_t^{n-1,s'}$ to $c_t^{n,s}$;
  $w \leftarrow$ Calculate the emission probability from $c_t^{n,s}$ to o_t^n;
  add p to p_t;
  add w to w_t;
return c_t, p_t, w_t

(4) Most probable candidates finding

To determine the most probable path among candidate points corresponding to observed points, we use the Viterbi algorithm to identify a set of candidate points on road segments with the highest log-probability under the observed conditions (Viterbi 1967). The Viterbi algorithm is a dynamic programming algorithm for obtaining the maximum a posteriori probability estimate of the most likely sequence of hidden states from a sequence of observed events, as shown in Equation 4. Dynamic programming enables efficient solutions to complex global optimization problems, such as finding the most probable sequence of hidden states, by breaking the problem into overlapping subproblems and solving each subproblem only once, storing intermediate results for reuse.

$$\delta_t^{n,j} = \max_{1 \leq i \leq s'} \left[\ln \delta_t^{n-1,i} + \ln p_t^{i,j} + \ln w_t^{n,j} \right] \tag{4}$$

where $\delta_t^{n,j}$ represents the maximum log-probability of the most probable path leading to candidate point $\delta_t^{n,j}$. i denotes the index of the candidate points at the previous position s', while j indicates the index of the candidate points at the current position s. $p_t^{i,j}$ is the transition probability from the candidate point $c_t^{n,i}$ at position s' to candidate point $c_t^{n,j}$ at position s. $w_t^{n,j}$ is the emission probability of candidate point $c_t^{n,j}$ at position s given the observed point o_t^n.

The workflow for finding most probable candidate points and path reconstruction is summarized in Algorithm 2.

Algorithm 2: The workflow for most probable path reconstruction

Input : a set of candidate points $c_t = \{c_t^{1,s}, ..., c_t^{n,s}, ..., c_t^{N,s}\}$ where $1 \leq n \leq N$, transition probability p_t, emission probability w_t

Output : a set of matched points z_t^*

for $2 \leq n \leq N$ **do**
  for $1 \leq j \leq s$ **do**
  $\delta_t^{n,j} \leftarrow \max_{1 \leq i \leq s'}[\ln \delta_t^{n-1,i} + \ln p_t^{i,j} + \ln w_t^{n,j}]$;
  $a_t^{n,j} \leftarrow \operatorname{argmax}_{1 \leq i \leq s'}[\ln \delta_t^{n-1,i} + \ln p_t^{i,j} + \ln w_t^{n,j}]$;
  $z_t^N = \operatorname{argmax}_{1 \leq j \leq s}[\delta_t^{N,j}]$;
  add z_t^N to z_t^*;
for $n = N - 1$ ***down to*** *2* **do**
  $z_t^n = a_t^n z_t^{n+1}$;
  add z_t^n to z_t^*;
return z_t^*

4.2.2. Attribute transfer

Having acquired the matched path $z_t^* = \{z_t^{1*}, ..., z_t^{n*}, ..., z_t^{N*}\}$, where the asterisk $(*)$ indicates that these candidate points on road segments have been identified as the optimal matches during the map-matching process. The attribute transfer is performed by intersecting the matched paths z_t^* with the semantic attributes of CityGML and the road segments E_t.

4.3. Calculation of cycling safety scores

4.3.1. Computing indicator coefficients

With reference to the Level of Traffic Stress (LTS), which classifies the components of a road network based on the stress experienced by cyclists (Heeremans *et al.* 2022), we select a set of indicators contributing cycling safety, as shown in Table 1.

Table 1. List of the considered indicators in our study.

No.	Indicators	Description	Example values
1	Max speed	maximum speed allowed on a road segment	50km/h
2	Number of lanes	number of lanes on a road segment	3
3	Parking allowed	whether parking is allowed on a road segment	True
4	Incident to crossing	whether a road segment is incident to a crossing	False
5	Lane width	width of road segments	1.5m
6	Lane type	safety lane, cycle lane, separated lane, shared lane	cycle lane

Building upon the indicators (e.g., max speed) proposed by Li *et al.* (2024a), this study extracts additional indicators (e.g., lane width) for cycling safety from CityGML through map matching. We then introduce MLR to determine the weights of these indicators, paving the way for safety score calculations. Given L road segments as training examples, let $\varepsilon = \{e_1, ..., e_l, ..., e_L\}$, where each segment $e_l = \{e_l^1, ..., e_l^k, ..., e_l^K\}$ represents the normalized observation values of K indicators for the l^{th} segment $(1 \leq l \leq L)$. Each segment is labelled by an experienced cyclist

(averaging $>$ 300 km weekly) as training data. The labelling process incorporated expert knowledge and referred to the visual reference photos of the cycling environment, sourced from Mapillary. Specifically, the expert evaluated safety conditions by rating photos of each segment on a 5-point Likert scale (ranging from 1 to 5, where 5 indicates very safe and 1 indicates very unsafe), which can be denoted as $\hat{y} = \{\hat{y}_1, ..., \hat{y}_l, ..., \hat{y}_L | 1 \leq l \leq L\}$. Then, we can establish an MLR model for cycling safety, as shown in Equation 5. The least squares method is used to estimate the indicator coefficients $\hat{\theta}$.

$$\hat{y}_l = \sum_{k=1}^{K} \theta_k \ e_l^k \tag{5}$$

where the dependent variable $\hat{y}_l$ represents the cycling safety score of the l^{th} segment, e_l^k indicates the normalized observation value of the k^{th} indicator for the l^{th} segment, and $\theta = \{\theta_1, ..., \theta_k, ..., \theta_K | 1 \leq k \leq K\}$ denotes the coefficients of the indicators, where the sum of coefficients equals 1.

4.3.2. Calculating safety scores

By training an MLR model on a sample of L road segments, the coefficient for each indicator can be estimated, which will subsequently be used to calculate the cycling safety score for the entire road network. Therefore, given a street network, we represent it as a graph $G = (V, E)$, where $V = \{v_1, v_2, ...\}$ and $E = \{e_1, e_2, ...\}$. The proposed approach can determine edge weights $\omega : E \rightarrow \mathbb{R}^+ \geq 0$, which assigns a weight $\omega(e)$ to each edge $e \in E$. The edges in the graph are referred to as segments in the training examples. Equation 6 details the calculation of weights $\omega(e_l)$ for the l^{th} edge, where $\omega(e_l)$ represents the cycling safety score of the l^{th} road segment.

$$\omega(e_l) = \sum_{k=1}^{K} \hat{\theta}_k \cdot e_l^k \tag{6}$$

where e_l^k represents the k^{th} normalized observation value for the l^{th} edge, $\hat{\theta}_k$ denotes the estimated coefficient for the k^{th} indicator, K is the number of indicators.

4.4. Embedding safety scores into CityGML

CityGML offers two methods for augmenting the data model beyond the original standard's definitions (Kolbe 2009). The first is through generic objects and attributes, while the alternative is a built-in mechanism known as Application Domain Extension (ADE) for extending the data model (Biljecki *et al*. 2018). In our case, recording the unique IDs for the road centerlines allows us to associate the cycling safety scores with the corresponding traffic areas in the CityGML dataset and update the semantic information within the 3D model. With this enriched CityGML dataset, we could link the city objects with cycling safety data for interactive exploration in a 3D environment. Figure 7 shows a schematic diagram of the 3D representation of cycling safety within CityGML.

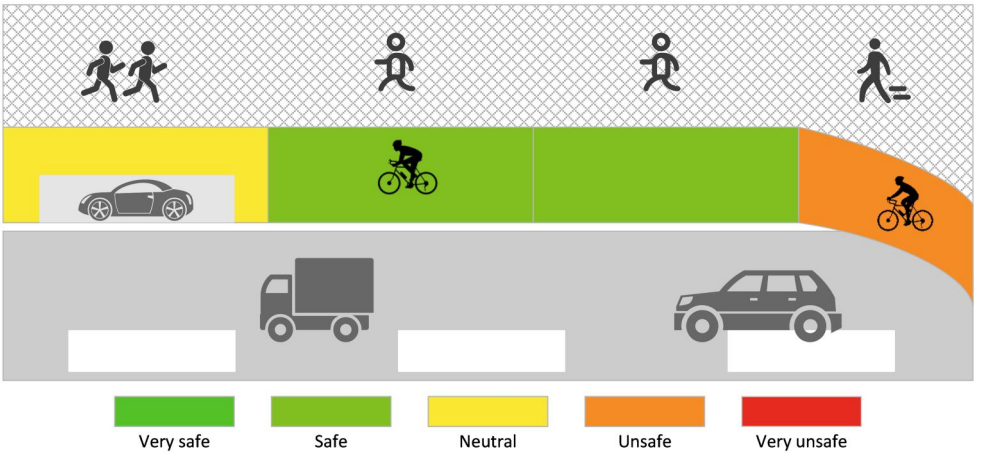


Figure 7. Schematic diagram of the 3D representation of cycling safety within CityGML. Different colors represent varying levels of cycling safety.

5. Results

5.1. Evaluation of matching between OSM and CityGML

To evaluate the matching between OSM and CityGML, the metrics defined in Equation 7, including Accuracy, Precision, Recall, and F1-score, are utilized to assess the performance.

$$\begin{cases} \text{Accuracy} = \dfrac{TP + TN}{TP + FP + FN + TN} \\ \text{Precision} = \dfrac{TP}{TP + FP} \\ \text{Recall} = \dfrac{TP}{TP + FN} \\ F_1\text{-score} = \dfrac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \end{cases} \tag{7}$$

TP (True Positives) indicates that the road segments that should be matched are correctly matched; FN (False Negatives) denotes that the road segments that should be matched are incorrectly matched; FP (False Positives) represents that the road segments that should not be matched are matched; TN (True Negatives) suggests that the road segments that should not be matched are unmatched.

Additionally, we evaluated the time cost of map matching with and without segmentation on a MacBook Pro M3 Pro with 18GB of RAM. The segmentation-based approach completed in 4.56 seconds, compared to 281.90 seconds without segmentation. This demonstrates that incorporating the concept of segmentation significantly improves the computational efficiency of the map matching process.

5.2. Comparison of matching performance across different approaches

Figure 8 illustrates the empirical distribution of the residual lateral distances between the OSM and CityGML centrelines, along with the corresponding KDE and Gaussian distributions. As shown, the KDE effectively captures the pronounced skew and the heavier upper tail of the data, whereas the Gaussian distribution places too much

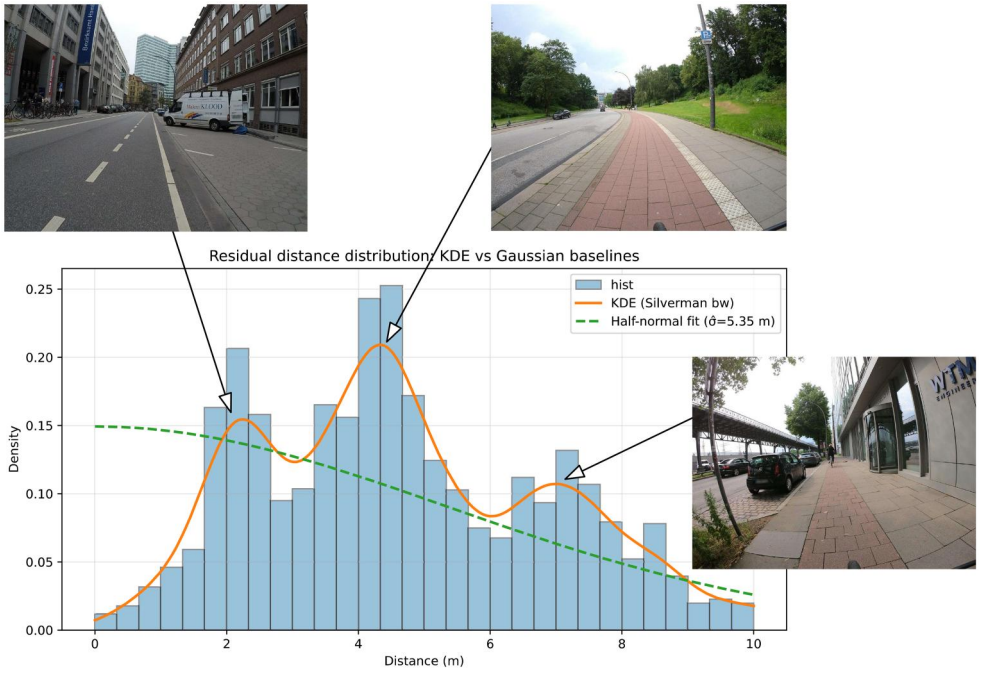


Figure 8. Residual distance distribution between the OSM and CityGML centrelines. The blue bars represent the empirical histogram of residual distances, the orange line fits the non-parametric KDE, and the green dashed line represents the Gaussian distribution. The pictures depict examples of cycling situations reflecting the corresponding distance peaks.

mass near $d = 0$ and underestimates the distance of different cycle paths. The KDE better reflects actual cycling conditions, which is why we use it to calculate the emission probability of observed points to candidate points, instead of the commonly assumed Gaussian distribution.

To compare the matching performance of a built-in matching method named Anchored Snapper in FME, HMM with Gaussian, and HMM with KDE, we manually labelled 154 road segments as ground truth. Of these, 114 segments are expected to be matched (positive cases) and 40 segments are not expected to be matched (negative cases). The results in Table 2 show that both the HMM with Gaussian and the HMM with KDE outperform FME's Anchored Snapper across all four metrics. Overall, using KDE instead of a Gaussian distribution in the HMM yields the highest scores, though the improvement over the Gaussian baseline is modest in our case. This is due to the limited spatial coverage of the ground truth, which consists of only two complex intersections where KDE can match observed points effectively, whereas the Gaussian distribution cannot.

Figure 9 shows an example of the matching result of FME and HMM with KDE at a street intersection. The FME also matches the CityGML with the road segments of

Table 2. Evaluation of matching performance between OSM and CityGML.

Approach	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
FME's Anchored Snapper	63.6	80.2	67.5	67.5
HMM with Gaussian	88.3	100	84.2	91.4
HMM with KDE	89.0	100	85.1	91.9

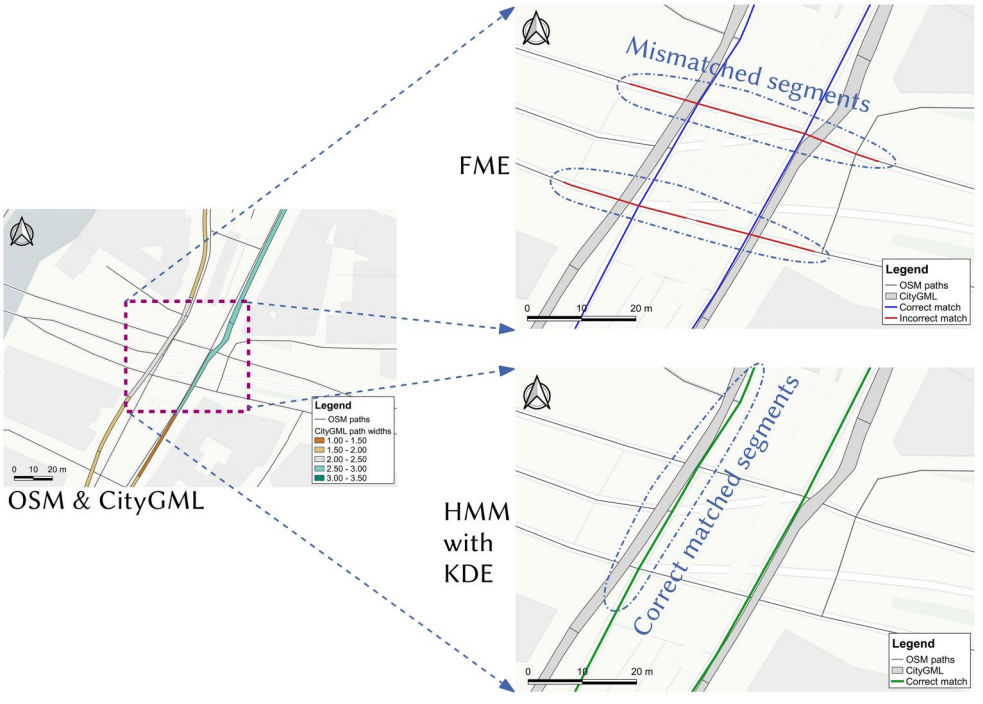
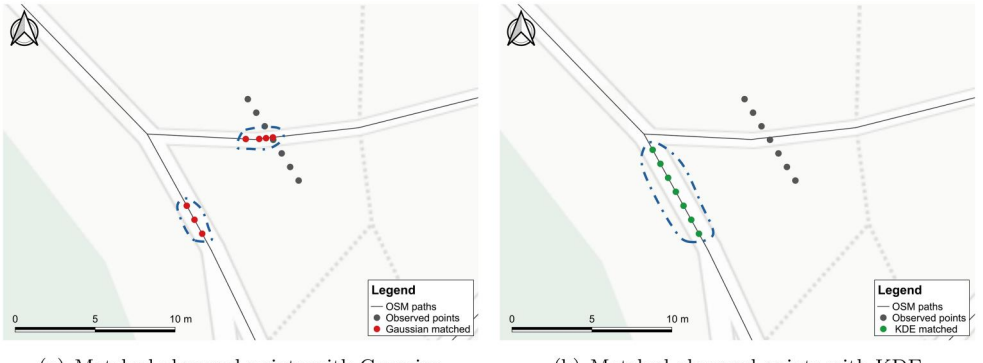


Figure 9. Comparison of map matching results between FME and HMM with KDE. The left subfigure shows the OSM road network and CityGML road lanes. The upper right subfigure shows the map matching results in FME, with incorrect matches highlighted by red lines, while the lower right sub-figure shows the correct matches from HMM with KDE.

OSM in other directions of the intersection, as indicated by the red road segments in the figure in contrast to the superior performance demonstrated by the HMM with KDE at the intersection.

Figure 10 illustrates a comparison between the Gaussian-based and KDE-based HMM matching approaches for a sample of observed points at a complex intersection.



(a) Matched observed points with Gaussian (b) Matched observed points with KDE

Figure 10. Comparison of Gaussian-based (left) and KDE-based (right) HMM matching for a sample of observed points at a complex intersection. The Gaussian model assigns most points to an incorrect road segment, while the KDE model correctly matches them to the corresponding segment.

In this example, the HMM with Gaussian (left) incorrectly assigns most observed points to the wrong road segment, which would lead to wrong results. In contrast, the HMM with KDE (right) correctly matches all observed points to the corresponding road segment, demonstrating its superior ability to better capture the true spatial relationship in this case.

5.3. Comparison of cycling safety scores

Figure 11 shows the correlation between cycling safety and lane width, indicating that lane width influences perceived or actual cycling safety. As lane width increases, the overall trend in cycling safety scores with lane width decreases, which means the wider the lane, the safer the cycling. It is important to note that the safety score does not decrease sharply, this is because cycling safety is influenced not only by lane width but also by various other factors, such as speed limits and the presence of parking.

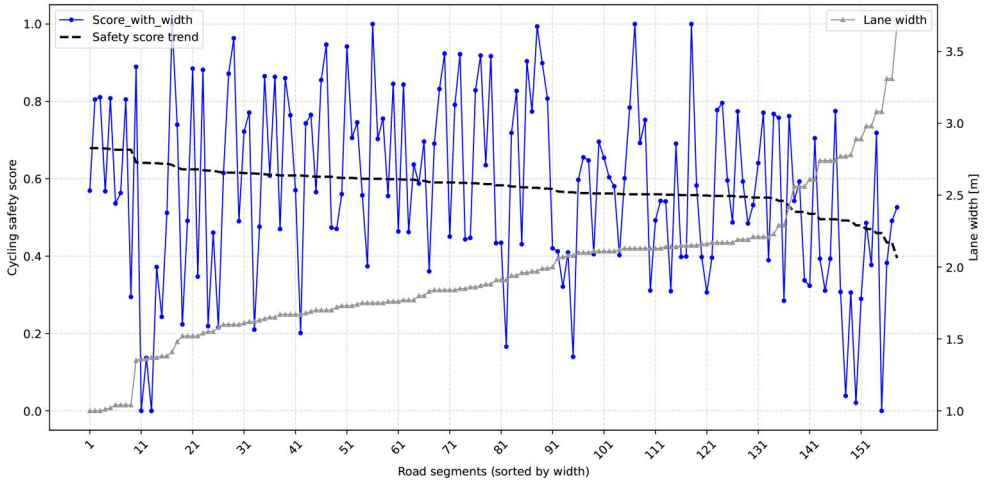


Figure 11. The correlation between cycling safety score and lane width. The gray line represents the lane width of road segments, and the X-axis is sorted in ascending order according to the width values. The blue line depicts the cycling safety scores with the consideration of lane width. The black line represents the trend in safety scores, illustrating the correlation between cycling safety score and lane width. Note that higher safety scores indicate lower safety.

Figure 12 compares the cycling safety scores before and after enrichment. It can be observed that when the lane width is below 2.0 m, the majority of cycling safety scores assessed with lane width consideration are higher than those without, indicating reduced safety. This observation aligns with the findings of the UK Department for Transport, whose *Local Transport Note 1/20* recommends a minimum width of 2.0 m for one-way protected cycle tracks to ensure sufficient space for overtaking and to accommodate non-standard bicycles (Department for Transport 2020).

Similarly, Schepers *et al*. (2014) report that Godthelp and Wouters (1978), based on experimental studies, recommended a minimum of 2.0 m for one-way cycle paths to allow safe lateral movement and overtaking. This 2.0 m benchmark is further echoed in Dutch guidance, where the CROW Design Manual for Bicycle Traffic identifies widths

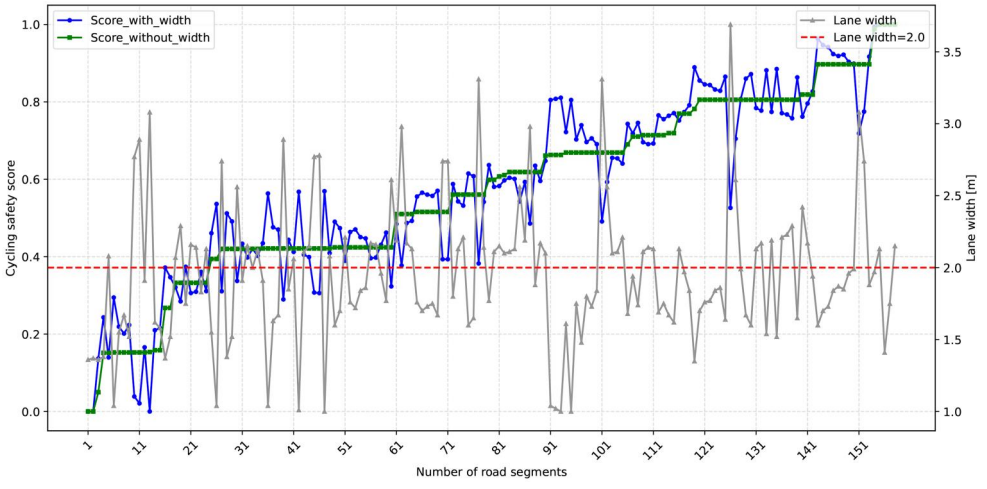


Figure 12. Comparison of cycling safety scores before and after enrichment. The gray line represents the lane width of road segments. The blue and green lines depict the cycling safety scores with and without the inclusion of lane width in the assessment, respectively. Higher safety scores indicate lower safety, and the red dotted line marks the lane width threshold for safe cycling.

of 2.0–2.5 m as the standard for one-way facilities on high-quality cycling routes (CROW 2017). Conversely, when lane width exceeds 2.0 m, the enriched safety scores are typically lower than the original scores, indicating a perceived or real improvement in cycling safety. This further supports the notion that widths above the 2.0 m threshold significantly enhance both comfort and safety for cyclists, especially in urban networks with varied rider speeds and vehicle types.

Additionally, we selected three road segments to visually explain the difference with the support of OSM and Mapillary[8], as shown in Figure 13. For example, the road conditions on edge 1 include a shared lane with a speed limit of 50 km/h, two lanes, and permitted roadside parking, resulting in a safety score of 0.663, indicating an

Figure 13. Cycling safety scores before and after enrichment are presented, with the three edges in the gray box representing scores that account for lane width. Higher scores indicate lower safety. The images on the right show the actual cycling environments of the three selected road segments, sourced from Mapillary.

unsafe level for cyclists. However, when 1.02 m road width extracted from CityGML is considered an indicator in the cycling safety assessment, the score increases to 0.808, reflecting an even more hazardous condition for cyclists due to the insufficient lane width. Studies have shown that the cycle lane should be at least 2.0 meters wide, with a recommended width of 2.5 meters, to ensure the stability of cyclists on the road (Lee *et al*. 2016, Yan *et al*. 2018).

5.4. Integrating cycling safety scores into CityGML

The left side of Figure 14 shows the CityGML dataset derived from an OpenDRIVE dataset, while the right side illustrates the corresponding hierarchical structure, highlighting key components related to transportation and cycling infrastructure in different colors. The core Transportation module includes the TrafficArea component, colored in grey, which can be further classified into Driving, Biking, and Sidewalk lanes. In our case, we integrate the attribute of cycling safety scores into the Biking component, thereby enriching the semantics of CityGML and paving the way for the subsequent representation of cycling safety in a 3D environment.

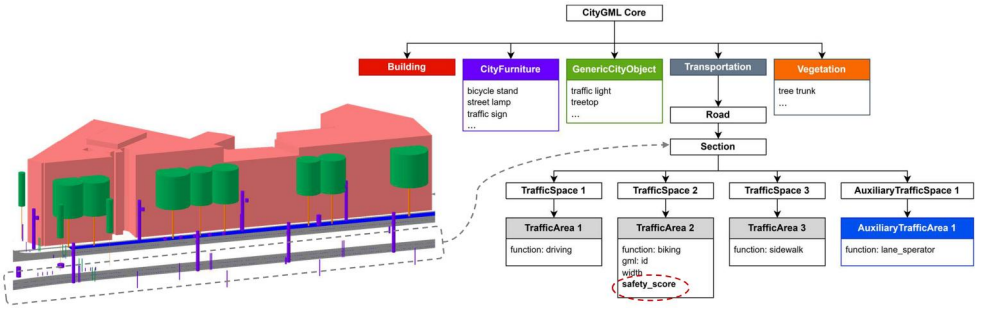


Figure 14. Integrating cycling safety scores into CityGML. The left subfigure shows the CityGML 3D models for the test track, and the right subfigure illustrates the corresponding hierarchy of CityGML core, where the cycling safety score can be embedded in the Biking class.

5.5. Three-dimensional representation of cycling safety scores

Figure 15 shows the result of cycling safety scores in a web-based Cesium visualization combined with semantic 3D city models. The cycle lanes and 3D city models are organized using lightweight 3D tiles, which ensure efficient streaming and rendering of large-scale 3D geospatial datasets (Li *et al*. 2021, 2024c). The left side of the figure shows the entire environment supported by the CityGML dataset, providing a comprehensive representation of the available road and cycling infrastructure. The right side presents a zoomed-in view of a selected road section, illustrating the cycling safety conditions with a traffic light-based color coding scheme. The green color represents a lower score than the yellow color, which means safer cycling. Three-dimensional representation facilitates intuitive cycling safety assessments by integrating enriched data attributes from both OSM and CityGML, it allows to perform further visibility analysis to be integrated as indicator.

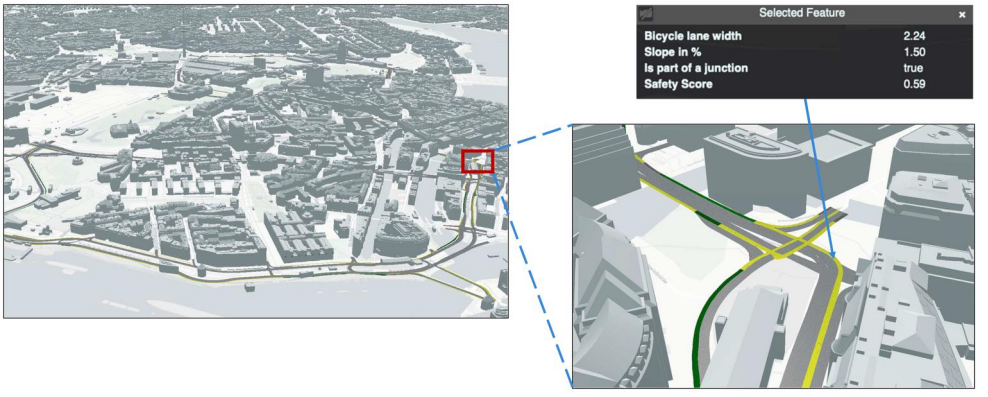


Figure 15. Three-dimensional representation of cycling safety scores. The 3D environment can serve as the basis for safest path navigation and perception.

6. Conclusion

Cycling is increasingly promoted as a sustainable transportation mode, but safety concerns remain a barrier to widespread adoption, particularly among older adults and novice cyclists. In this context, we proposed a novel approach that leverages the complementary strengths of OSM's rich semantic information on roads and CityGML's lane-level geometry to facilitate cycling safety assessment. First, we derived additional indicators of lanes (e.g., width) that contribute to cycling safety from CityGML, complementing the indicators available in OSM. Next, Hidden Markov Model (HMM) with Kernel Density Estimation (KDE) was employed to match the CityGML and OSM, enabling the bidirectional transfer of attributes. Then, we used Multivariate Linear Regression (MLR) to induce the weights of these indicators by learning from the observed values, and subsequently calculated the cycling safety scores for the whole road network. Finally, the cycling safety scores were embedded into bicycle-related semantics within CityGML for the subsequent 3D representation of cycling safety.

From a methodological innovation perspective, it is important to note that, unlike the commonly assumed Gaussian distribution for modeling spatial distances in map matching, we adopted KDE to capture the distribution of distances between OSM and CityGML, which is within an informed map matching to calculate the emission probabilities in HMM. This approach enhances the reliability of the matching process and reduces the mismatches that frequently occur under the Gaussian assumption, particularly in complex urban environments such as road intersections. The experimental results show that the matching F1-score of the HMM with KDE reached 0.919, outperforming both the built-in matching method of FME and the HMM with the Gaussian distribution. Additionally, instead of using a pairwise comparison to determine the weight of each indicator in the cycling safety assessment, the expert rated the selected track samples with reference to visual reference photos of the cycling environment. The rated scores were then used as training examples in MLR to derive the weights of the safety indicators. These weights are subsequently used to calculate the cycling safety score for the entire road network, which is embedded in CityGML 3.0 for 3D representation of cycling safety.

From a scientific contribution perspective, our study demonstrates that integrating OSM and CityGML is conducive to improving the comprehensive cycling safety assessment at the road segment level, and the assessment results contribute to the enrichment of bicycle-related semantics in CityGML. In particular, the findings of this study support the notion that widths above the 2.0 m threshold significantly enhance both comfort and safety for cyclists, which is aligned with the statement from the literature research, and the CityGML 3.0 with embedded safety score could serve as the basis for safest path navigation and perception in a 3D environment. Beyond its immediate applications, this work lays the foundation for incorporating advanced 3D-derived indicators for cycling safety, such as cyclist-driver visibility analysis. This could involve computing the sightline intersection between cyclists and drivers, accounting for the influence of buildings or parked vehicles.

Due to data quality issues in OSM, certain roads with two-way lanes are represented by a single geometric feature of roads in only one direction, which can impact the accuracy of cycling safety assessments. In the future, a key task will be to treat the entire OSM road network as a graph, using reasoning to complete the missing road segments and create a supergraph to better accommodate the dual nature of road features in CityGML. We will also explore the possibility of developing a standalone application dedicated to safest path navigation of cycling.

Notes

1. https://www.openstreetmap.org/.
2. https://www.ogc.org/publications/standard/citygml/.
3. https://fahrradklima-test.adfc.de/teilnahme.
4. https://osmnx.readthedocs.io/.
5. https://trian3dbuilder.de/.
6. https://www.asg.ed.tum.de/gis/software/rtron/.
7. https://fme.safe.com/.
8. https://www.mapillary.com.

Acknowledgements

The authors would like to express their gratitude to OpenStreetMap and OpenDRIVE communities for their invaluable contribution to open geospatial data, as well as to the editor and anonymous reviewers for their careful reading and constructive suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data and code availability statement

The data and codes that support the findings of this study can be found online at https://github.com/hcu-cml/bidirectionalEnrichment-OSM-CityGML-cycling.

Funding

This work was supported by the Federal Ministry of Transport (BMV) under the project Next Generation City Networking (NGCN) (Grant No. 19OI22008A), the Chinese Postdoctoral Science Foundation (Grant Nos. 2024T170742 and 2025M770237), and the Postdoctoral Fellowship Program of the China Postdoctoral Science Foundation (Grant No. GZC20232185),National Natural Science Foundation of China (42201446).

Notes on contributors

Weilian Li is an associate professor of the Faculty of Geosciences and Engineering at Southwest Jiaotong University. He completed his postdoctoral research in the Geoinformation Group at the University of Bonn in October 2023. Additionally, he spent one year as a senior research associate in the Computational Methods Lab at the HafenCity University Hamburg. He has a strong interest in topics related to smart cities, disaster management, and digital twins. He proposed the idea and contributed to this article's methodology, implementation, experiment, and writing.

Jannik Matijevic is currently pursuing his master's studies at HafenCity University Hamburg, where he also serves as a research assistant in the Computational Methods Lab. His research focuses on smart cities and sustainable mobility. He contributed to this article's methodology, implementation, experiment, and writing.

Christof Beil received his PhD degree at the Chair of Geoinformatics at the Technical University of Munich (TUM) in 2025. His research focuses on standardization and implementation of semantic 3D street space models and related applications. He contributed to this article's methodology, review, and proofreading.

Lukas Arzoumanidis received his Master of Science degree from the Institute of Geodesy and Geoinformation at the University of Bonn in 2023. He is currently pursuing a PhD at HafenCity University Hamburg. His doctoral research focuses on the automated information retrieval of historical documents and the deep generation of semantic 3D city models. His research interests include machine learning, with a particular emphasis on generative models and graph neural networks. He contributed to the methodology, review, and proofreading.

Thomas H. Kolbe is a full professor and the head of the Chair of Geoinformatics at the Technical University of Munich. He is the initiator and co-author of the international standard CityGML for semantic 3D city and landscape models, as well as the architect and co-author of the international standard IndoorGML. His key research areas include virtual 3D city models, urban digital twins, and 3D geoinformation systems. He proposed the idea, shaped the concept and contributed to this article's methodology and review.

Youness Dehbi is a full professor and the Chair of the Computational Methods Lab at HafenCity University Hamburg. He received his Ph.D. degree in the Institute of Geodesy and Geoinformation at the University of Bonn in 2016. His research focuses on urban reasoning, analytics, pattern recognition, and machine learning for spatio-temporal data. He proposed the idea, shaped the concept and contributed to this article's methodology, writing and review.

References

Ahmed, T., *et al.*, 2024. Bicycle infrastructure design principles in urban bikeability indices: A systematic review. *Sustainability*, 16 (6), 2545.

Basiri, A., *et al.*, 2019. Crowdsourced geospatial data quality: Challenges and future directions. *International Journal of Geographical Information Science*, 33 (8), 1588–1593.

Beil, C., *et al.*, 2020. Detailed streetspace modelling for multiple applications: Discussions on the proposed CityGML 3.0 transportation model. *ISPRS International Journal of Geo-Information*, 9 (10), 603.

Beil, C., *et al.*, 2023. Automatically evaluating the service quality of bicycle paths based on semantic 3D city models. In: *International 3D GeoInfo Conference*. Springer, 75–92.

Biassoni, F., *et al.*, 2023. Choosing the bicycle as a mode of transportation, the influence of infrastructure perception, travel satisfaction and pro-environmental attitude, the case of milan. *Sustainability*, 15 (16), 12117.

Biljecki, F., Kumar, K., and Nagel, C., 2018. CityGML application domain extension (ADE): overview of developments. *Open Geospatial Data, Software and Standards*, 3 (1), 17.

Boss, D., Nelson, T., and Winters, M., 2018. Monitoring city wide patterns of cycling safety. *Accident; Analysis and Prevention*, 111, 101–108.

Buehler, R., and Pucher, J., 2021. Covid-19 impacts on cycling, 2019–2020. *Transport Reviews*, 41 (4), 393–400.

Cao, M., *et al.*, 2020. Analysis of the spatiotemporal riding modes of dockless shared bicycles based on tensor decomposition. *International Journal of Geographical Information Science*, 34 (11), 2225–2242.

CROW CROW 2017., *Design manual for bicycle traffic*, Ede, Netherlands. *Dutch design standard recommending 2.0–2.5 m for one-way cycle paths*, Available from: https://crowplatform.com/

Daraei, S., Pelechrinis, K., and Quercia, D., 2021. A data-driven approach for assessing biking safety in cities. *EPJ Data Science*, 10 (1), 11.

Department for Transport 2020., *Cycle infrastructure design: Local transport note ltn 1/20*. UK Government. Table 5-2 recommends 2.0 m as a minimum for one-way cycle tracks, Available from: https://www.gov.uk/government/publications/cycle-infrastructure-design-ltn-120.

Derek, J., and Sikora, M., 2019. Bicycle route planning using multiple criteria GIS analysis. In: *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, 1–5.

DiGioia, J., *et al.*, 2017. Safety impacts of bicycle infrastructure: A critical review. *Journal of Safety Research*, 61, 105–119.

Fan, H., *et al.*, 2016. A polygon-based approach for matching OpenStreetMap road networks with regional transit authority data. *International Journal of Geographical Information Science*, 30 (4), 748–764.

Félix, R., Moura, F., and Clifton, K.J., 2019. Maturing urban cycling: Comparing barriers and motivators to bicycle of cyclists and non-cyclists in Lisbon, Portugal. *Journal of Transport & Health*, 15, 100628.

Feng, Y., Huang, X., and Sester, M., 2022. Extraction and analysis of natural disaster-related VGI from social media: review, opportunities and challenges. *International Journal of Geographical Information Science*, 36 (7), 1275–1316.

Ghahramani, Z., 2001. An introduction to hidden Markov models and Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15 (01), 9–42.

Godthelp, J., and Wouters, P., 1978. Koers houden door fietsers en bromfietsers. *Verkeerskunde*. In *Dutch. Recommends 2.0 m Minimum Width for Safe Bicycle Path Design*, 11, 537–543.

Götschi, T., *et al.*, 2018. Towards a comprehensive safety evaluation of cycling infrastructure including objective and subjective measures. *Journal of Transport & Health*, 8, 44–54.

Gröger, G., and Plümer, L., 2012. CityGML: Interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71, 12–33.

Heeremans, O., *et al.*, 2022. Group cycling safety behaviours: A systematic review. *Transportation Research Part F: Traffic Psychology and Behaviour*, 91, 26–44.

Herfort, B., *et al.*, 2023. A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. *Nature Communications*, 14 (1), 3985.

Ito, K., *et al.*, 2024. Translating street view imagery to correct perspectives to enhance bikeability and walkability studies. *International Journal of Geographical Information Science*, 38 (12), 2514–2544.

Jacobsen, P.L., 2015. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prevention: journal of the International Society for Child and Adolescent Injury Prevention*, 21 (4), 271–275.

Kamel, M.B., Sayed, T., and Bigazzi, A., 2020. A composite zonal index for biking attractiveness and safety. *Accident; Analysis and Prevention*, 137, 105439.

Kolbe, T., 2009. Representing and exchanging 3d city models with citygml. *In*: J. Lee and S. Zlatanova, eds. *3d geo-information sciences*. Berlin, Heidelberg: Springer, 15–31.

Kolbe, T.H., *et al.*, 2021. OGC City Geography Markup Language (CityGML) Version 3.0 Part 1: Conceptual Model Standard – International Standard. No. Doc. No. 20–010.

Kolbe, T.H., Gröger, G., and Plümer, L., 2005. CityGML: Interoperable access to 3D city models. In: *Geo-information for disaster management*. Berlin, Heidelberg: Springer, 883–899.

Kutzner, T., Chaturvedi, K., and Kolbe, T.H., 2020. CityGML 3.0: New functions open up new applications. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88 (1), 43–61.

Lee, C., *et al.*, 2016. Measurement of desirable minimum one-way bike lane width. *KSCE Journal of Civil Engineering*, 20 (2), 881–889.

Li, W., *et al.*, 2021. An augmented representation method of debris flow scenes to improve public perception. *International Journal of Geographical Information Science*, 35 (8), 1521–1544.

Li, W., *et al.*, 2024a. Safety assessment of cycling routes in urban environments. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W10-2024, 125–130.

Li, W., *et al.*, 2024b. Informed sampling and recommendation of cycling routes: leveraging crowd-sourced trajectories with weighted-latent dirichlet allocation. *International Journal of Geographical Information Science*, 38 (12), 2492–2513.

Li, W., *et al.*, 2024c. Visual attention-guided augmented representation of geographic scenes: a case of bridge stress visualization. *International Journal of Geographical Information Science*, 38 (3), 527–549.

Lowry, M.B., *et al.*, 2012. Using bicycle level of service to assess community-wide bikeability. *Journal of the Transportation Research Board*. Presented at the 91st Annual Meeting of the Transportation Research Board.

Lucks, L., *et al.*, 2021. Improving trajectory estimation using 3d city models and kinematic point clouds. *Transactions in GIS*, 25 (1), 238–260.

Meng, S., and Zheng, H., 2023. A personalized bikeability-based cycling route recommendation method with machine learning. *International Journal of Applied Earth Observation and Geoinformation*, 121, 103373.

Meuleners, L.B., *et al.*, 2019. Safer cycling and the urban road environment: A case control study. *Accident; Analysis and Prevention*, 129, 342–349.

Mora, R., Truffello, R., and Oyarzún, G., 2021. Equity and accessibility of cycling infrastructure: An analysis of santiago de chile. *Journal of Transport Geography*, 91, 102964.

Panagiotaki, E., *et al.*, 2025. The Oxford RobotCycle Project: A multimodal urban cycling dataset for assessing the safety of vulnerable road users. *IEEE Transactions on Field Robotics*, 2, 308–335.

Pucher, J., and Buehler, R., 2016. Safer cycling through improved infrastructure. *American Journal of Public Health*, 106 (12), 2089–2091.

Scarano, A., *et al.*, 2023. Systematic literature review of 10 years of cyclist safety research. *Accident; Analysis and Prevention*, 184, 106996.

Schepers, P., *et al.*, 2014. A conceptual framework for road safety and mobility applied to cycling safety. *Accident; Analysis and Prevention*, 62, 331–340.

Schepers, P., *et al.*, 2017. The dutch road to a high level of cycling safety. *Safety Science*, 92, 264–273.

Schwab, B., and Kolbe, T.H., 2022. Validation of parametric OpenDRIVE road space models. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W2-2022, 257–264.

Schwab, B., Beil, C., and Kolbe, T.H., 2020. Spatio-semantic road space modeling for vehicle–pedestrian simulation to test automated driving systems. *Sustainability*, 12 (9), 3799.

Senaratne, H., *et al.*, 2017. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31 (1), 139–167.

Shoman, M.M., *et al.*, 2023. Evaluation of cycling safety and comfort in bad weather and surface conditions using an instrumented bicycle. *IEEE Access*, 11, 15096–15108.

Silverman, B.W., 1986. *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Vargas-Munoz, J.E., *et al.*, 2021. Openstreetmap: Challenges and opportunities in machine learning and remote sensing. *IEEE Geoscience and Remote Sensing Magazine*, 9 (1), 184–199.

Viterbi, A., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13 (2), 260–269.

Wang, D., *et al.*, 2024. Evaluating the bikeability of urban streets using dockless shared bike trajectory data. *Sustainable Cities and Society*, 101, 105181.

Wegman, F., Zhang, F., and Dijkstra, A., 2012. How to make more cycling good for road safety? *Accident; Analysis and Prevention*, 44 (1), 19–29.

Yan, X., *et al.*, 2018. Recommended widths for separated bicycle lanes considering abreast riding and overtaking. *Sustainability*, 10 (9), 3127.