Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Geo-Data-Driven HD Map Generation Workflow with Integrated Reference-Free Constraint-Based Verification

Ruidi He

ruidi.he@tu-clausthal.de

Clausthal University of Technology

Vaibhav Tiwari

Clausthal University of Technology

Mohanad Al-Ghobari

Clausthal University of Technology

Meng Zhang

Clausthal University of Technology

Andreas Rausch

Clausthal University of Technology

Research Article

Keywords: HD map generation, geo-data, driven mapping, representation-level verification, automated driving, constraint-based quality assessment, engineering workflow

Posted Date: June 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9727159/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Geo-Data-Driven HD Map Generation Workflow with Integrated Reference-Free Constraint-Based Verification

Ruidi He[1*], Vaibhav Tiwari[1], Mohanad Al-Ghobari[1], Meng Zhang[1*], Andreas Rausch[1*]

[1*]Institute for Software and Systems Engineering, Technical University of Clausthal, Clausthal-Zellerfeld, Germany.

*Corresponding author(s). E-mail(s): ruidi.he@tu-clausthal.de; meng.zhang@tu-clausthal.de; andreas.rausch@tu-clausthal.de;
Contributing authors: vaibhav.tiwari@tu-clausthal.de; mohanad.al-ghobari@tu-clausthal.de;

Abstract

High-definition (HD) maps are core artifacts for automated driving systems, but their generation commonly relies on sensor-intensive mobile mapping campaigns, while quality assessment often depends on high-precision reference data. These dependencies make HD map engineering costly and difficult to apply in settings where specialised measurement data or independently measured reference maps are unavailable. This paper presents an engineering-oriented geo-data-driven workflow for HD map generation with integrated representation-level verification. The workflow uses openly available geo-engineering datasets as the primary input source and transforms them into lane-level HD map representations of existing road environments through explicit intermediate representations and processing stages. To assess the generated representations without external reference maps, the workflow integrates executable constraint-based verification into the engineering process. Selected constraints are derived from specifications relevant to automated driving and road-design guidelines. They are evaluated directly on the generated lanelet-based representation to detect geometric, topological, and elevation-related inconsistencies. The workflow is evaluated using real-world shapefile-based road-network data from four cities in Lower Saxony, Germany, and controlled defect-injection scenarios. The real-world evaluation shows that the generated map representations satisfy the selected constraints

in the evaluated scenarios, while the defect-injection study demonstrates complete detection of the considered defect types without observed false positives. The results indicate that geo-data-driven HD map generation with integrated executable verification can provide a modular and inspectable complement to sensor-intensive mapping workflows under reduced sensing and reference-data availability.

Keywords: HD map generation, geo-data–driven mapping, representation-level verification, automated driving, constraint-based quality assessment, engineering workflow

1 Introduction

1.1 Motivation

High-definition (HD) maps are important artifacts in modern automated driving (AD) systems. They provide persistent prior knowledge about the geometry, topology and semantics of real-world road environments and complement onboard sensing, especially where local perception is limited by occlusion, range, weather, or computational constraints. As a result, HD maps are used across core modules such as localisation, perception, prediction, and motion planning. By providing a structured representation of road geometry, topology, and semantics, they help align these modules around a common model of the driving context. This makes HD maps safety-critical artifacts, because defects in map representations can propagate across modules and lead to system-level failures. For example, when a vehicle approaches a sharp curve after an uphill segment, the terrain may occlude the upcoming curvature from forward-facing cameras. In such cases, an HD map can support earlier speed adaptation through prior knowledge of the road geometry. However, inaccurate or inconsistent map information may instead lead to unsafe planning or control decisions. From a software engineering perspective, the generation of HD maps for real-world road environments can be viewed as an engineering workflow that incrementally transforms information about physical road infrastructure into progressively refined digital map representations through multiple processing stages. Conventional HD map generation approaches exhibit three engineering characteristics commonly.

First, conventional HD map generation approaches commonly rely on large-scale data collection using specialised high-end mapping sensor setups, such as 360 ° LiDAR, panoramic cameras, and high-precision GNSS/IMU units [1]. While such configurations provide rich and accurate mapping data, they require dedicated measurement campaigns and generate large raw-data volumes that involve substantial preprocessing, registration, fusion, and map-generation effort [1]. This sensing dependency motivates the investigation of alternative structured information sources for constructing initial HD map representations of existing road environments. Openly available geo-engineering data, such as road-network geometries, elevation data, and infrastructure-related attributes, describe existing physical road environments without requiring dedicated mapping campaigns. However, they are not HD maps: they

may lack lane-level detail, explicit topology, complete semantics, and consistency guarantees required by automated driving systems.

Second, conventional approaches are typically implemented as tightly integrated processing pipelines. In multi-sensor fusion and SLAM-based approaches, map generation is commonly formulated as a joint estimation problem in which LiDAR, IMU, camera, and positioning data are optimised within a unified framework [2, 3]. Although such approaches may contain intermediate data products, these representations are often pipeline-internal, tool-specific, or not exposed as independently verifiable engineering artifacts. This reduces transparency between processing stages, complicates localized diagnosis of transformation errors, and increases the risk that uncertainties propagate across subsequent stages. Similar challenges arise in many learning-based approaches, where intermediate processing is implicitly encoded in latent representations [4], limiting interpretability and independent control of intermediate results. For geo-data-driven HD map generation, this motivates a modular workflow with explicit intermediate representations and inspectable transformation stages.

Third, map quality assessment is commonly tied to external reference data or indirect evaluation signals. Since HD maps do not have a true ground truth beyond the physical world itself, existing approaches approximate references using high-precision sensor data such as dense LiDAR point clouds or carefully collected mapping datasets. However, when workflows reduce reliance on raw sensor data, representation-level consistency can no longer be directly assessed through comparison-based evaluation, increasing the risk that defects remain undetected. Together, these characteristics create a double dependency: map generation depends on specialized sensing campaigns, while quality assessment often depends on independently measured reference data. This work, therefore, investigates a modular and inspectable workflow for transforming open geo-engineering data into HD map representations of real-world road environments under reduced sensing dependency and integrates executable constraint-based verification for representation-level quality assessment.

1.2 Challenges

Given the above observations, two engineering challenges arise.

Challenge 1: Designing a modular map generation pipeline with reduced sensing dependency

The first challenge concerns how to generate HD map representations of existing real-world road environments without relying on specialized sensing campaigns as the primary data source. Addressing this challenge requires a modular and inspectable pipeline in which processing stages, intermediate representations, and transformation assumptions are made explicit.

Challenge 2: Establishing representation-level quality assessment without measured references

The second challenge concerns how to assess the consistency and structural quality of generated HD map representations when independently measured reference maps or high-precision ground-truth data are not available. Addressing this challenge requires practical geometric, topological, and semantic constraints that can be operationalized as executable verification checks directly on the generated map representation. Such

checks should support representation-level defect detection rather than relying primarily on comparison with measured reference data or indirect downstream performance signals [2, 4].

1.3 Research Questions

To address the challenges outlined above, this work investigates the following research questions.

- RQ1: How can HD map representations be generated from open geo-engineering data through a modular pipeline that reduces dependence on specialized sensing campaigns while preserving inspectable intermediate artifacts?
- RQ2: How can representation-level consistency of generated HD maps be assessed without relying on independently measured reference maps or high-precision reference data?
  - RQ2.1: How can practical geometric, topological, and semantic constraints be defined for representation-level assessment of generated HD map representations?
  - RQ2.2: How can these constraints be operationalized as executable verification checks within the generation workflow?

1.4 Paper Organization

The remainder of this paper is organized as follows. Section 2 reviews related work on HD map generation and quality control. Section 3 presents the proposed modular HD map generation pipeline together with its integrated constraint-based verification approach. Section 4 describes the implementation of the pipeline. Section 5 evaluates the pipeline using real-world geo-data and controlled defect scenarios. Section 6 discusses the implications and limitations of the approach, and Section 7 concludes the paper and outlines future work.

2 Related Work

2.1 HD Map Generation Workflows

HD map generation approaches can be broadly categorized into sensor-driven, learning-based, and geo-data-driven methods, depending on the primary data source and processing paradigm.

Sensor-driven HD map generation relies on large-scale data collection using mobile mapping systems equipped with LiDAR, cameras, and GNSS/IMU sensors. Such approaches typically involve multi-sensor fusion and tightly coupled processing stages [2, 3]. Although they can achieve high geometric precision, they require dedicated measurement campaigns, large raw-data processing effort, and specialized sensor configurations, which limit modularity and scalability, and make intermediate representation difficult to inspect independently [2].

Many recent approaches explore learning-based HD map generation from onboard perception data, often using Bird's-Eye View (BEV) representations [4, 5]. These methods support online inference in real-time and can reduce dependence on offline

mapping campaigns. However, their intermediate reasoning is often encoded in implicit latent representations, which limits interpretability, traceability, and explicit structural diagnostics. Their performance also depends heavily on large annotated datasets and may degrade under distribution shifts [4].

A third direction leverages structured geographic data, such as OpenStreetMap (OSM) or institutional GIS datasets, for HD map generation [6]. Such data provide broadly available information about road-network geometry, elevation, and infrastructure attributes, but they often lack lane-level detail, explicit topology, semantic completeness, and consistency guarantees required by AD systems [6]. Existing approaches address parts of this gap through structured transformation pipelines, lane-level map models, simulation-oriented conversion workflows, and OpenDRIVE-based road modelling [7–10]. However, these approaches primarily focus on conversion and modelling, while systematic representation-level verification is typically external, limited, or not explicitly integrated into the generation workflow.

Overall, existing HD map generation approaches involve a trade-off between measurement dependency, automation, and workflow transparency. Sensor-driven methods provide high-precision map data but require specialized sensing campaigns; learning-based methods support automated or online generation but rely on learned latent representations and annotated datasets; geo-data-driven methods offer a more modular and inspectable basis but require explicit transformation and consistency checks. This motivates generation approaches that treat intermediate map representations as inspectable engineering artifacts rather than only as final outputs of a conversion pipeline.

2.2 HD Map Quality Control and Verification

Ensuring representation-level consistency in HD maps is important since defects in geometry, topology, or semantics can affect downstream modules such as localization and motion planning. Existing HD map quality-control approaches differ primarily in how map quality is defined, what evidence is used, and whether diagnostics are available at the representation level. In the following, these approaches are broadly grouped into reference-based, learning-based, and constraint-based methods, with emphasis on their suitability for geo-data-driven methods without independently measured reference maps or high-precision reference data.

Reference-based approaches evaluate generated maps by comparing them with high-precision reference maps, dense point clouds, or annotated mapping datasets [11–13]. They provide quantitative evidence of agreement with a measured reference representation, but depend on costly data collection and inherit the assumptions and limitations of the reference data. Downstream performance-based evaluation can provide additional end-to-end evidence, but it is indirect and often provides limited diagnostic insight into representation-level defects.

Learning-based approaches estimate map quality, uncertainty, or anomalies using machine learning techniques [14, 15]. They can handle noisy and incomplete inputs and may capture complex defect patterns, but their diagnostics are often model-dependent and less interpretable. Moreover, they do not necessarily provide explicit guarantees that geometric, topological, or semantic requirements are satisfied.

Constraint-based approaches define representation-level consistency through consistency constraints over geometric, topological, and semantic properties of the map representation [16–18]. They provide interpretable diagnostics without requiring direct comparison with measured high-precision reference maps or reference data. However, existing approaches often apply such checks as standalone verification steps rather than integrating them into the map generation workflow. They also provide limited support for structured defect analysis across multiple constraint categories and transformation stages.

In summary, existing quality-control approaches provide complementary evidence but remain insufficient for geo-data-driven HD map generation workflows. Reference-based methods require independently measured reference data, performance-based methods provide limited diagnostic localization, and learning-based methods often lack explicit structural guarantees. Constraint-based verification is well-suited for interpretable representation-level diagnostics but is commonly treated as a separate checking activity rather than as an integrated part of the generation workflow. This motivates a unified engineering workflow that combines geo-data-driven map generation with executable, representation-level constraint verification.

3 Geo-data-Driven HD Map Generation Pipeline

This section describes the proposed workflow architecture. The workflow is organized as a sequence of processing components that transform geospatial input data for a user-defined region of interest into a lane-level map representation and subsequently evaluate this representation through executable consistency checks. The description focuses on the workflow structure, component responsibilities, exchanged representations, and the integration point between map generation and verification.

3.1 Initial Concept

The initial concept of the pipeline is derived from the two research questions addressed in this work. The first research question concerns the automated generation of lane-level HD maps from open geo-data, while the second concerns the verification of the generated maps with respect to geometric, topological, and semantic consistency constraints. This distinction motivates the separation of the workflow into a map generation part and a verification part. The generation part transforms open geo-data into a lane-level map representation, whereas the verification part evaluates the generated representation and reports detected inconsistencies.

Within the generation part, further decomposition is required because open geo-data are not directly suitable for HD map applications. They usually provide road geometries and attributes, but not complete lane-level structures, explicit lane boundaries, or verification-ready connectivity relations. Therefore, the system first retrieves and parses geo-data for a user-defined region of interest (ROI), then constructs a lane-level map representation, and finally converts it into a verification-compatible format. The ROI restricts processing to the relevant spatial area and avoids unnecessary extraction of large-scale geo-data. A frontend supports this workflow by allowing users to define the ROI, configure parameters, visualize the selected area, and initiate

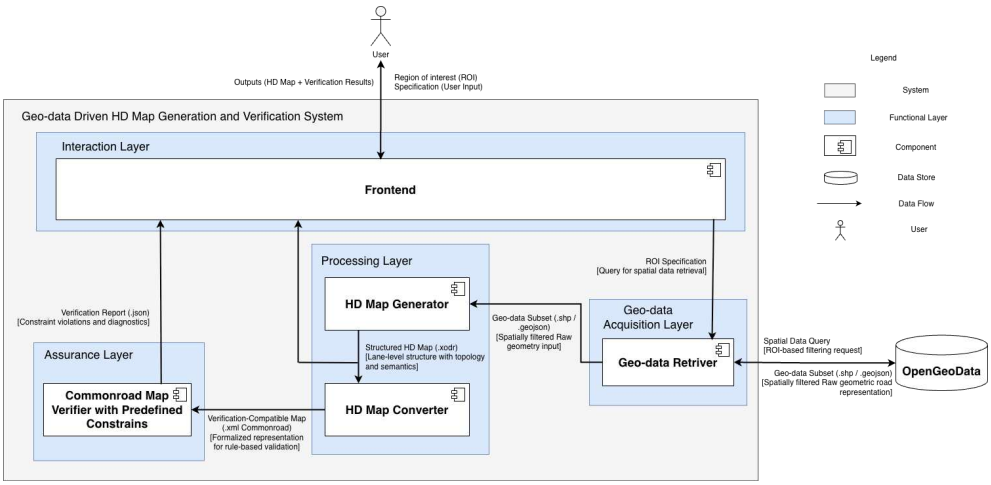


Fig. 1 System architecture of the geo-data-driven HD map generation and verification workflow.

the workflow execution. These considerations lead to a system architecture, which is described in the following subsection.

3.2 System Overview

The system is organized as a modular workflow that separates user interaction, geo-data transformation, map representation conversion, and constraint-based verification. Fig. 1 illustrates the component-based architecture of the proposed geo-data–driven HD map generation and verification pipeline. The workflow starts from a user-defined ROI, retrieves the corresponding geo-data from available geospatial datasets, normalizes and transforms the retrieved data into a lane-level map representation, converts this representation into a verification-compatible format, and finally evaluates executable constraints to produce diagnostic reports describing detected violations and affected map elements.

The workflow is structured into three conceptual layers. The interaction layer defines the spatial and execution context of the pipeline. The transformation layer incrementally constructs the generated map representation from geo-data through explicit processing stages. The verification layer evaluates the generated representation against executable constraints and produces diagnostic feedback without modifying the map. As indicated in Fig. 1, these conceptual layers are realised by five components with explicit interfaces: (1) the frontend component, (2) the geo-data retrieval and parsing component, (3) the HD map generator, (4) the HD map converter, and (5) the verification component.

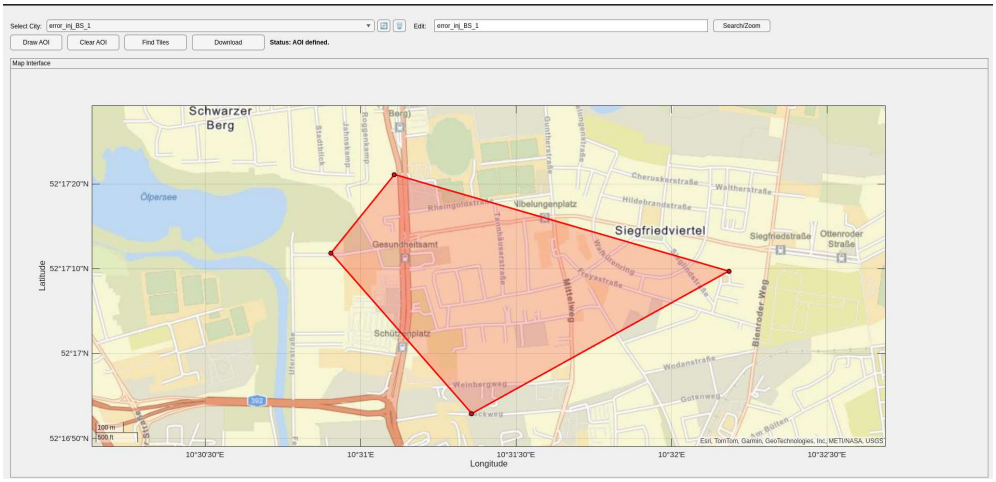


Fig. 2 Graphical selection of the region of interest (ROI) in the GUI. The target study area is defined interactively as a polygon over the basemap and is subsequently used for road-data clipping and DEM tile retrieval.

The frontend component implements the interaction layer and provides the graphical interface for workflow orchestration, configuration management, and execution control. It captures the user-defined ROI and relevant configuration parameters and aggregates them into a structured execution context. In the current implementation, the ROI is defined interactively on a map view by selecting or drawing a spatial bounding region around the target road network. Internally, the ROI is represented as a bounding box in latitude–longitude coordinates and stored together with execution parameters such as the selected input dataset, coordinate reference settings, elevation source, export options, and verification configuration. Fig. 2 illustrates an example ROI selection in the frontend, where the user marks the target area before starting the generation workflow. The selected ROI is converted into spatial boundaries that are used consistently for geo-data extraction and elevation sampling. By managing ROI selection, parameter handling, and initialisation of downstream processing stages, the frontend ensures reproducible and consistent pipeline execution.

The geo-data retrieval and parsing component constitutes the first stage of the transformation layer. Its purpose is to extract the road information required for a target study area from raw geospatial data and convert it into structured inputs for subsequent modelling. In the current implementation, the input geo-data are obtained from official open geospatial datasets published by the state of Lower Saxony, Germany, including Basis-DLM - Shapefile [19] and Digitales Geländemodell (DGM1) [20]. Basis-DLM provides vector-based representations of surface objects such as roads, water bodies, vegetation, and settlement areas, and is used here as the primary source of road centerlines and road attributes. As part of the ATKIS topographic data framework, it provides geometrically consistent and attribute-rich road data in the ETRS89 / UTM zone 32N reference system. DGM1 is a bare-earth digital terrain model derived from airborne laser scanning and provided in 1×1 km tiles with a raster resolution

of 1 m, making it suitable for fine-grained elevation sampling in lane-level HD map generation. Using the ROI boundary as a spatial constraint, the component filters road objects from the global road layer, referred to here as the VER1 layer, to produce a local road dataset. It further parses the key road attributes and organises them into semantically relevant descriptors, primarily including `OBJID`, `OBJART`, `OBJART_TXT`, `NAM`, `BEZ`, `FSZ`, `BRF`, `FAR`, `FKT`, `WDM`, and `ZUS`. Among these, `FSZ`, `BRF`, and `FAR` provide the core basis for recovering lane count, road width, and traffic direction semantics. In addition, the component performs coordinate reference system validation, projection information export, and caching of locally clipped results, thereby providing stable and standardised inputs for the downstream mapping stage.

The HD map generator converts the resulting local road centerlines and their key attributes into a lane-level RoadRunner HD Map. It first transforms the input centerlines into a local coordinate system in order to reduce the numerical scale effects of geodetic coordinates on subsequent geometric computation. It then clusters centerline endpoints to identify topological nodes in the road network and projects nodes located within road interiors back onto the centerlines to split the roads into segment units suitable for lane-level modelling. During semantic recovery, the system uses `FSZ` to infer the number of lanes, `BRF` to determine the lateral road width and lane expansion scale, and `FAR` to determine one-way or two-way traffic properties and assign lane travel directions accordingly. Based on these inferred semantics, the system automatically generates lane centerlines and lane boundaries and organizes them into *lane*, *lane boundary*, and *lane group* objects in the RoadRunner HD Map. For topology construction, it establishes predecessor and successor relations from the spatial clustering of lane endpoints and recovers basic traffic connectivity using direction constraints, distance constraints, and a greedy one-to-one matching strategy. In intersection areas, candidate connections are further classified according to the heading relationships between nearby incoming lanes and outgoing lanes, allowing straight, left-turn, and right-turn movements to be added and thereby enriching the topological semantics of the local road network. For three-dimensional map generation, DEM data are introduced after the two-dimensional lane geometry has been constructed. By selecting, mosaicking, and sampling elevation tiles covering the ROI, and by applying missing-value filling, smoothing, and grade constraints, elevation information is propagated to both lane centerlines and boundaries, producing an RRHD map with full 3D coordinates. Finally, the system outputs a local-coordinate `.rrhd` file and restores its corresponding global-coordinate version using the saved origin offset. The resulting map is then imported into RoadRunner for automatic scene construction and exported as `.rrscene`, `.fbx`, and `.xodr` files for downstream.

The HD map converter [21] transforms OpenDRIVE road network descriptions into CommonRoad scenario representations. The method converts roads and lane structures defined by reference lines and lateral parameters in OpenDRIVE into an explicit lanelet-based road network in CommonRoad. In OpenDRIVE, road geometry is defined by reference lines, while lane boundaries are implicitly described through lateral offsets, width functions, and lane sections. In CommonRoad, roads are represented as lanelets with explicit left boundaries, right boundaries, and centerlines, together with predecessor, successor, and lateral adjacency relations. The conversion therefore

extracts the reference lines, lane sections, width functions, elevation information, and lateral profile features from the OpenDRIVE file, reconstructs the inner and outer lane boundaries along the longitudinal path parameter, and discretely samples them to generate lanelet boundary point sequences. The centerline of each lanelet is computed from the corresponding boundary midpoints, and the connectivity information from the original road description is used to establish the lanelet topology. For three-dimensional road structures, longitudinal elevation and lateral profile information are combined with the recovered planar boundaries to preserve height-related geometry. The implementation is based on Python, uses `commonroad-io` to construct CommonRoad `Scenario`, `LaneletNetwork`, and `Lanelet` objects, and employs `pyproj` for coordinate transformation and georeferencing. The resulting road network is exported as a standard CommonRoad XML scenario file.

The converted CommonRoad map is verified through a rule-based process to assess road-structure consistency. The process takes the Lanelet network as input and evaluates predefined validity rules over lanelet boundary geometry, predecessor-successor relations, lateral adjacency relations, and road-element references. The rules are expressed in higher-order logic, with their syntax defined and parsed by ANTLR [22] to obtain executable expressions. CommonRoad road elements are then mapped to logical domains and evaluated through predicate and term functions. When a rule is violated, the corresponding lanelet or road-element identifiers are recorded as invalid states, providing explicit locations of structural inconsistencies. For large road networks, the verification can be applied to partitioned sub-networks and executed in parallel to improve efficiency.

To support rule extension, the implementation further includes an LLM-assisted rule generation script [23]. Given a natural-language rule requirement, the script constructs a prompt containing the current code context and invokes a large language model to generate modifications such as rule identifiers, formula mappings, and predicate functions. The generated outputs are integrated into the existing verification framework through structured patches or operations and checked by a basic smoke test before being used in the higher-order-logic verification workflow. Thus, the LLM is used to assist rule development and code generation, while the final map verification remains a deterministic higher-order-logic evaluation process. The overall workflow is implemented in Python, operates directly on `commonroad-io` map objects, and organises the results as rule-wise invalid states for subsequent analysis or text report generation.

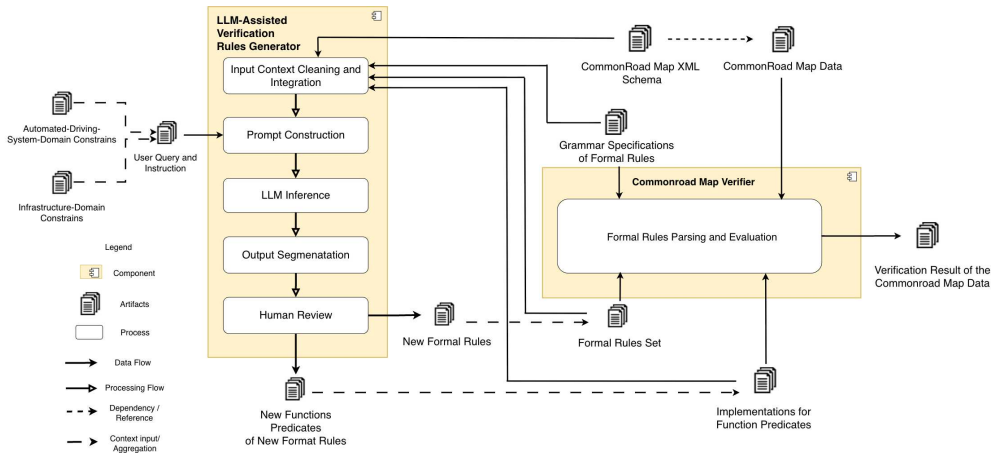


Fig. 3 LLM-assisted map verification workflow [23]. Constraint rules are constructed from heterogeneous sources with LLM assistance and then evaluated over the map representation by the verification component, producing interpretable diagnostics.

4 Evaluation

4.1 Evaluation Setup

The evaluation follows a sequential procedure in which selected regions of interest (ROIs) are used to generate HD maps, and the corresponding verification reports are manually inspected to assess whether the verification component identifies meaningful inconsistencies and provides interpretable diagnostic information. The evaluation comprises three complementary case studies: (1) verification of automated-driving domain constraints on real-world generated maps, (2) verification of road-design constraints, and (3) controlled defect injection. Across all studies, the verification pipeline processes the generated lanelet networks, evaluates the selected constraints, and produces reports containing detected violations and their spatial locations. The reported violations are then manually compared with the corresponding raw geo-data and generated map artifacts to assess whether they indicate actual inconsistencies. The evaluation is intentionally scoped as a feasibility-oriented engineering study rather than a comprehensive benchmarking campaign.

4.1.1 Case Study 1: Autonomous Driving Domain Constraints on Real-World Maps.

The first case study evaluates representation-level map constraints relevant to automated driving. For this purpose, 45 HD map samples are generated from real-world shapefile-based geospatial road-network data across four cities in Lower Saxony, Germany: Hannover, Braunschweig, Goslar, and Hildesheim. The selected regions of interest cover diverse urban road scenarios, including straight roads, curved roads, T-junctions, intersections, and merging lanes. The evaluated constraints address common

consistency requirements for automated-driving map representations [24], including elevation completeness, topological consistency, and geometric validity of lane boundaries.

Three constraint categories are evaluated:

- **Elevation Completeness:** Validates that all boundary points contain finite z-coordinates.
- **Successor Non-Self-Loop:** Ensures that a lanelet does not reference itself as a successor.
- **Polyline Validity:** Requires each boundary polyline to contain at least two geometrically distinct points.

4.1.2 Case Study 2: Road Design Constraints.

The second case study evaluates whether the map quality assessment in the pipeline can also be applied to constraints derived from road-design guidelines. For this purpose, a constraint is formulated based on the RASt 06 road-design guideline [25]. The evaluation focuses on built-up urban areas, since the selected requirement is applicable to this road context and can be assessed using the geometric information available in the generated lanelet networks. A total of 13 HD maps are generated across the same four cities used in Case Study 1, covering comparable urban road scenarios. Verification is performed using the same batch-processing pipeline as in the first case study. The evaluated constraint is defined as follows:

- **Minimum Turning Radius:** Validates that the turning radius of each lanelet remains above 10.0 m, as specified by RASt 06 for built-up urban roads.

4.1.3 Case Study 3: Controlled Defect Injection.

The first two case studies assess the quality of generated maps by applying selected constraints to real-world map samples and manually inspecting the reported violations. However, since naturally occurring defects are not systematically controlled, these studies cannot fully assess whether the verification rules detect known defect types reliably. Therefore, a third case study is designed to evaluate defect detection under controlled conditions. The third case study uses controlled defect injection to systematically assess detection behaviour and completeness. Unlike the previous case studies, where violations may occur naturally, known defects are intentionally introduced into a representative urban map from Braunschweig, Germany. Multiple modified map versions are generated from this baseline, each containing predefined defects. Three defect categories are considered, each with 10 injected defects:

- **Elevation Completeness Defect:** Non-finite z-coordinates (NaN) are introduced into selected lanelets to simulate incomplete elevation data.
- **Lane Width Defect:** The lateral distance between lane boundaries is artificially reduced to create unrealistically narrow lane sections.
- **Successor Self-Loop Defect:** Self-loop successor relations are introduced into selected lanelets to create invalid topological cycles.

4.2 Metrics

Case Studies 1 and 2 do not provide complete ground-truth annotations that distinguish all consistent and inconsistent map elements. Therefore, precision- and recall-based metrics cannot be computed reliably for these real-world studies, since the complete set of true violations and non-violations is unknown. Following prior work on constraint-based and ontology-driven validation [16, 17, 26], the evaluation instead reports detected violations per map and per constraint category, followed by manual inspection of the reported diagnostics. In contrast, Case Study 3 provides ground truth by map generation. Since the injected defects and their locations are known a priori, the controlled defect injection study enables quantitative assessment of detection behaviour, following metrics commonly applied in map change detection and anomaly detection [14, 27]. For this study, precision is used to measure how many reported violations correspond to injected defects:

$$\text{Precision} = \frac{N_{\text{correctly detected}}}{N_{\text{reported}}} \tag{1}$$

where N_{reported} denotes the number of reported violations and $N_{\text{correctly detected}}$ the number corresponding to injected defects.

Detection completeness is evaluated using the true positive rate (TPR):

$$\text{TPR} = \frac{N_{\text{correctly detected}}}{N_{\text{injected}}} \tag{2}$$

where N_{injected} denotes the total number of injected defects.

In addition, **false positives (FP)** are analyzed as reported violations that do not correspond to injected defects. This distinction allows the evaluation to assess both the correctness of reported diagnostics and the completeness of defect detection under controlled conditions.

4.3 Results

4.3.1 Case Study 1 and 2: Real-World Maps.

The evaluation demonstrates the feasibility of the proposed engineering-oriented workflow for constructing and assessing HD maps from real-world geo-data. As summarized in Table 1, no violations were detected across the 45 maps evaluated in Case Study 1 for elevation, polyline, and self-loop constraints, nor across the 13 urban road maps evaluated in Case Study 2 for the minimum turning radius constraint. These results indicate that the proposed workflow can generate structurally consistent HD maps across diverse urban scenarios using geo-data as the primary input source. In addition, the constraint-based verification component successfully detected controlled defects with high diagnostic precision.

These results indicate that the generated maps satisfy the selected representation-level and road-design constraints within the evaluated scenarios. They also demonstrate that the proposed workflow can process real-world shapefile-based geo-data, generate lanelet-based map representations, and execute constraint-based verification

Table 1 Verification results for the real-world map evaluation in Case Studies 1 and 2, showing the number of detected violations per evaluated constraint.

Case Study	Constraint	Maps	Violations Reported
Case Study 1	Elevation Completeness	45	0
	Successor Non-Self-Loop	45	0
	Polyline Validity	45	0
Case Study 2	Minimum Turning Radius	13	0

Table 2 Detection performance for the controlled defect injection experiments in Case Study 3. "Base." denotes violations detected in the original baseline map, "Inj." the number of injected defects, and "Det." the number of detected defects reported by the verifier.

Constraint	Base.	Inj.	Det.	TP	FP	Precision	TPR
Elevation	0	10	10	10	0	1.000	1.000
Lane Width	0	10	10	10	0	1.000	1.000
Self-Loop Connections	0	10	10	10	0	1.000	1.000

in a batch-processing pipeline. However, since no complete ground-truth annotations are available for these real-world maps, the results should be interpreted as evidence of feasibility and consistency with respect to the selected constraints, rather than as a comprehensive proof of map correctness.

4.3.2 Case Study 3: Defect Injection.

Table 2 summarizes the results of the controlled defect injection experiments. No violations were detected in the original baseline map before defect injection. After injecting defects, the verifier detected all 10 elevation completeness defects, all 10 lane-width defects, and all 10 successor self-loop defects. No false positives were reported for any of the evaluated defect categories. Consequently, precision and true positive rate (TPR) both reached 1.00 for all three defect categories. These results show that the implemented constraints can reliably detect the considered defect types under controlled conditions. In combination with the real-world evaluations in Case Studies 1 and 2, the defect injection study (Case Study 3) provides targeted evidence that the verification component can correctly and completely detect the selected defect types under controlled conditions.

5 Discussion

This work investigated how HD map generation and representation-level quality assessment can be integrated into a modular engineering workflow. The proposed approach

combines geo-data–driven map generation with executable constraint-based verification and organizes both activities within an inspectable workflow with explicit intermediate representations.

The workflow is not intended to achieve centimeter-level geometric accuracy comparable to dedicated sensor-intensive mapping systems. Since the generated maps depend on the completeness, resolution, and attribute quality of publicly available geodata, their absolute geometric accuracy is inherently bounded by the source data. The primary objective is therefore not precise physical alignment with the real-world environment but the generation of lane-level HD map representations satisfying selected structural constraints with integrated verification support.

The results indicate that structured geo-data can provide a practical basis for lane-level HD map generation without requiring tightly coupled sensor-processing pipelines or specialized mapping systems. In contrast to traditional SLAM-based approaches, the proposed workflow preserves explicit intermediate representations and separates data retrieval, map generation, representation conversion, and verification into distinct processing stages. This improves inspectability, supports incremental debugging, and allows individual components or constraints to be extended independently.

Map quality is assessed through executable consistency constraints evaluated directly on the generated map representations rather than through comparison against external reference maps. The evaluated constraints capture baseline geometric and topological consistency conditions derived from automated driving system specifications and infrastructure-oriented road-design rules. The real-world case studies show that the generated map representations satisfy the selected constraints in the evaluated scenarios, while the controlled defect injection study demonstrates that the implemented rules can detect the considered defect types with interpretable diagnostics under controlled conditions. These findings support the feasibility of integrating constraint-based verification into a geo-data–driven HD map generation workflow. Several limitations remain. First, the workflow depends on the completeness and consistency of geo-data attributes such as lane count, road direction, and road class. Missing or ambiguous attributes require heuristic assumptions, which may affect the correctness of the generated lane-level map representation. Second, the current constraint set primarily targets baseline geometric and topological consistency conditions and should be regarded as an initial feasibility-oriented rule set rather than a comprehensive map-quality model. Third, verification outcomes may be influenced by representation conversion effects, since constraints are evaluated on converted CommonRoad representations rather than directly on the originally generated OpenDRIVE maps. Finally, the evaluation is limited to selected urban scenarios in Lower Saxony, Germany, and primarily reflects European road-design assumptions, including those derived from RASt 06 [25]. Future work should extend the constraint set toward richer semantic and behavioral verification, including traffic-rule consistency, signal and intersection semantics, lane-transition legality, and route-level connectivity. Additional data sources, such as aerial imagery, point clouds, and onboard sensor observations may further improve semantic reconstruction and robustness. Broader evaluations across additional geographic regions, road structures, and defect categories are also required to assess the general applicability and robustness of the proposed workflow.

6 Conclusion

This work presented a geo-data–driven workflow for HD map generation with integrated representation-level quality assessment. The approach combines modular map generation with executable constraint-based verification, enabling direct and interpretable consistency analysis of generated map representations without relying on external reference maps or data.

The evaluation demonstrates the feasibility of generating lane-level HD map representations from publicly available geo-data and operationalizing geometric and topological consistency checks as executable verification rules. The real-world case studies show that the generated map representations satisfy the selected constraints in the evaluated scenarios, while the controlled defect injection study confirms that the implemented rules can detect the considered defect types with interpretable diagnostics. The results suggest that integrated verification can provide a practical and inspectable complement to traditional sensor-intensive mapping workflows, particularly for early-stage map generation, prototyping, and consistency-oriented assessment.

Additionally, the current approach remains limited by the quality and completeness of the underlying geo-data, representation conversion effects, and the restricted evaluation scope. Future work should extend the constraint set toward verification for richer semantics, incorporate additional data sources, and evaluate the workflow across broader geographic and road-design contexts.

References

[1] Tsushima, F., Kishimoto, N., Okada, Y., Che, W.: Creation of high definition map for autonomous driving. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **XLIII-B4-2020**, 415–420 (2020) https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-415-2020

[2] Asrat, K.T., Cho, H.-J.: A comprehensive survey on high-definition map generation and maintenance. ISPRS International Journal of Geo-Information **13**(7), 232 (2024)

[3] Bao, Z., Hossain, S., Lang, H., Lin, X.: A review of high-definition map creation methods for autonomous driving. Engineering Applications of Artificial Intelligence **122**, 106125 (2023)

[4] Kwag, J., Toth, C.: A review on end-to-end high-definition map generation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **48**, 187–194 (2024)

[5] Li, Q., Wang, Y., Wang, Y., Zhao, H.: Hdmapnet: An online HD map construction and evaluation framework. CoRR **abs/2107.06307** (2021) 2107.06307

[6] Elghazaly, G., Frank, R., Harvey, S., Safko, S.: High-definition maps: Comprehensive survey, challenges, and future perspectives. IEEE Open Journal of Intelligent

Transportation Systems **4**, 527–550 (2023)

[7] Bender, P., Ziegler, J., Stiller, C.: Lanelets: Efficient map representation for autonomous driving. In: 2014 IEEE Intelligent Vehicles Symposium Proceedings, pp. 420–425 (2014). https://doi.org/10.1109/IVS.2014.6856487

[8] Lopez, P.A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., Wießner, E.: Microscopic traffic simulation using sumo. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2575–2582 (2018). Ieee

[9] Dosovitskiy, A., Ros, G., Codevilla, F., López, A.M., Koltun, V.: CARLA: an open urban driving simulator. CoRR **abs/1711.03938** (2017) 1711.03938

[10] ASAM e.V.: OpenDRIVE Format Specification. Version 1.7 (2023). https://www.asam.net/standards/detail/opendrive/

[11] Kalenda, M.-N.B., Bonnifait, P., Mittet, M.-A.: Vector map quality metrics for contextual autonomous driving systems. In: 2025 IEEE International Conference on Vehicular Electronics and Safety (ICVES), pp. 367–373 (2025). https://doi.org/10.1109/ICVES65691.2025.11376526

[12] Chiang, K.-W., Srinara, S., Chiu, Y.-T., Tsai, S., Tsai, M.-L., Satirapod, C., El-Sheimy, N., Ai, M.: Creation and verification of high-definition point cloud maps for autonomous vehicle navigation. IEEE Internet of Things Journal **11**(23), 37582–37598 (2024) https://doi.org/10.1109/JIOT.2024.3435344

[13] Chiang, K.-W., Zeng, J.-C., Tsai, M.-L., Darweesh, H., Chen, P.-X., Wang, C.-K.: Bending the curve of hd maps production for autonomous vehicle applications in taiwan. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **15**, 8346–8359 (2022) https://doi.org/10.1109/JSTARS.2022.3204306

[14] Shaw, A.K., Jiang, K., Wen, T., Sah, C.K., Shi, Y., Yang, M., Yang, D., Lian, X.: Cleanmap: Distilling multimodal llms for confidence-driven crowdsourced hd map updates. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 3798–3807 (2025). https://doi.org/10.1109/CVPRW67362.2025.00365

[15] Pfeifle, M., Glander, K.-H., Szántó, M., Pfeifle, T.: Enhancing digital maps with ai-derived confidence information. In: Heintzel, A. (ed.) Automatisiertes Fahren 2024, pp. 105–117. Springer, Wiesbaden (2024)

[16] Maierhofer, S., Ballnath, Y., Althoff, M.: Map verification and repairing using formalized map specifications. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), pp. 1277–1284 (2023). https://doi.org/10.1109/ITSC57777.2023.10422044

[17] Qiu, H., Ayara, A., Glimm, B.: Ontology-based map data quality assurance. In: Eighteenth Extended Semantic Web Conference - Research Track (2021). https://openreview.net/forum?id=xNHyHnpLT5S

[18] Chiang, K.-W., Tsai, M.-L., Lin, S., Huang, Y.-E., Zeng, J.-C., Chang, Y.-F., Chen, J.-A., Huang, Y.-C., Yang, C.-S., Juang, J.-C., Wang, C.-K., Lin, C.-F., Lee, J., Darweesh, H., Li, P.-L.: Establishment of hd maps verification and validation procedure with opendrive and autoware (lanelet2) formats. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **X-1/W1-2023**, 621–627 (2023) https://doi.org/10.5194/isprs-annals-X-1-W1-2023-621-2023

[19] Landesamt für Geoinformation und Landesvermessung Niedersachsen (LGLN): Basis-DLM - Shapefile. OpenGeoData portal, accessed May 15, 2026 (2026). https://ni-lgln-opengeodata.hub.arcgis.com/documents/lgln-opengeodata::basis-dlm-shapefile/about

[20] Landesamt für Geoinformation und Landesvermessung Niedersachsen (LGLN): Digitales Geländemodell (DGM1). OpenGeoData portal, accessed May 15, 2026 (2026). https://ni-lgln-opengeodata.hub.arcgis.com/apps/lgln-opengeodata::digitales-gel%C3%A4ndemodell-dgm1/about

[21] Althoff, M., Urban, S., Koschi, M.: Automatic conversion of road networks from opendrive to lanelets. In: Proc. of the IEEE International Conference on Service Operations and Logistics, and Informatics (2018)

[22] ANTLR Project: ANTLR4. https://github.com/antlr/antlr4. GitHub repository, accessed May 15, 2026 (2026)

[23] He, R., Zhang, Y., Zhang, M., Rausch, A.: LLM-Assisted Tool for Joint Generation of Formulas and Functions in Rule-Based Verification of Map Transformations (2025). https://arxiv.org/abs/2511.01423

[24] Autoware Foundation: Autoware: Open-source software for autonomous driving. https://www.autoware.org/. Accessed: 2026-04 (2023)

[25] FGSV: Richtlinien Für die Anlage Von Stadtstraßen (RASt 06). Forschungsgesellschaft für Straßen- und Verkehrswesen, Köln, Germany (2006)

[26] De Meester, B., Heyvaert, P., Arndt, D., Dimou, A., Verborgh, R.: Rdf graph validation using rule-based reasoning. Semantic Web **12**, 117–142 (2020) https://doi.org/10.3233/SW-200384

[27] Zhang, P., Zhang, M., Liu, J.: Real-time hd map change detection for crowdsourcing update based on mid-to-high-end sensors. Sensors **21**(7) (2021) https://doi.org/10.3390/s21072477