ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XII-4/W1-2026
21st International 3D GeoInfo Conference 2026, 28–29 September 2026, Sofia, Bulgaria

Assessing and fixing LOD2 building models for Urban Energy Digital Twin

Oscar Roman[1], Giorgio Agugiaro[2], Ken Arroyo Ohori[2], Elisa Mariarosaria Farella[1], Fabio Remondino[1]

[1] 3DOM 3D Optical Metrology, Bruno Kessler Foundation, Trento, Italy - oroman@fbk.eu, elifarella@fbk.eu, remondino@fbk.eu

[2] 3D Geoinformation group, Department of Urbanism, Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, The Netherlands - g.agugiaro@tudelft.nl, k.ohori@tudelft.nl

Keywords: LOD2; Digital Twin; Graph-based repair; Energy Simulations; Smart cities

Abstract

Recent advances in 3D reconstruction have enabled the automated generation of city models from LiDAR point clouds and aerial imagery. Although these methods produce visually convincing models, they are primarily designed for visualization purposes and often lack the geometric and topological quality required for Urban Digital Twin applications, such as physics-based simulation, semantic analysis and predictive modelling. Automatically reconstructed LoD2 building models frequently exhibit defects including open boundaries, non-manifold configurations, inconsistent face orientations and degenerate geometries. This work presents a graph-guided framework for the topological repair of LoD2 building models. Mesh connectivity is represented as a graph, allowing boundary, adjacency and local geometric information to identify defective regions and guide robust repair operations. By combining topology-aware reasoning with geometry-constrained graph processing, the proposed method restores watertightness, improves 2-manifold consistency and preserves the structural integrity of complex building surfaces. Unlike conventional geometry-based post-processing techniques, the proposed approach explicitly targets topological correctness, producing high-quality building models suitable for simulation and analysis within Urban Digital Twins.

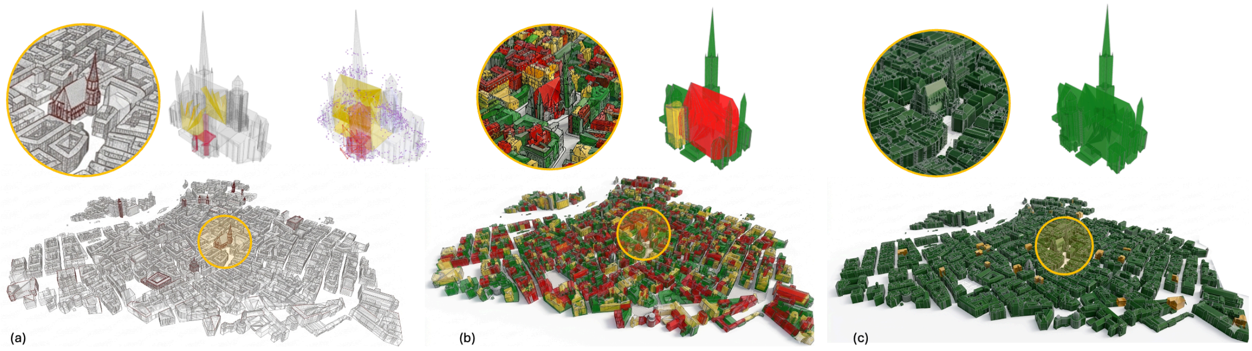

Figure 1. The proposed pipeline: analysis of building geometries in the input file and construction of the graph to analyse topological consistency (a); definition of the building status (open, non-manifold, watertight) (b); final result after applying repair strategies (c).

1. Introduction

Recent advances in 3D city modelling have significantly improved the automatic generation of building models from airborne LiDAR point clouds and aerial imagery (Wang et al., 2023; Amrullah et al., 2025; Jarząbek-Rychard et al., 2025). However, most reconstruction pipelines are optimized for visualization purposes, where geometric plausibility is sufficient and topological correctness is not enforced. Urban Digital Twins (UDTs) and in particular Urban Energy Digital Twins (UEDTs), impose fundamentally stricter requirements to serve as input to physics-based simulation engines such as EnergyPlus (U.S. Department of Energy, 2025) or computational fluid dynamics (CFD) solvers (Pađen et al., 2024): building geometries must be topologically valid, semantically classified and structurally well-formed. This gap is especially critical for LoD2 city models, which constitute the primary geometric basis for urban-scale energy analysis.

Despite appearing visually correct, LoD2 models are frequently affected by open boundary edges, non-manifold configurations, duplicate vertices, degenerate polygons and inconsistent face orientations, as well as missing or incorrect surface semantics (Biljecki et al., 2016). Such defects prevent the building envelope from being interpreted as a closed, orientable 2-manifold — a necessary precondition for interior – exterior discrimination. The consequences are operationally significant: without geometrically valid inputs, simulation tools that cannot robustly cope with invalid geometries may either reject the model outright or may resort to computationally expensive fallback strategies such as voxelization (Alkadri et al., 2020), at the cost of geometric fidelity and semantic granularity. In energy applications specifically, defective envelopes may propagate systematic errors into thermal load estimation, solar irradiation modelling and retrofit impact assessment. These defects are format-agnostic and pipeline-independent, manifesting across CityGML (Gröger et al., 2012; Arroyo Ohori et al., 2018), CityJSON (Ledoux et al., 2019) and SQL (Yao et al., 2018). Existing tools, such as val3dity (Ledoux, 2018), CityDoctor (Betz et al., 2025) and the FME GeometryValidator (FME Software), support well-established validation and partial repair but none integrates full topological correction.

1.1 Aim of the work

This paper proposes a methodology (Figure 1) for qualifying and repairing existing LoD2 building geometries (Figure 2) for urban-scale energy simulation. It aims to detect and correct topological defects, restore 2-manifold consistency and watertightness while minimizing geometric distortion, and generate semantically labelled envelopes compatible with energy tools (e.g., EnergyPlus) or CFD (Roman et al., 2025a) simulation workflows. The resulting models form closed thermal zones with

outward-oriented surfaces classified as roof, wall, floor, or virtual boundary, together with per-surface construction properties sufficient to produce valid simulation inputs.
Each CityGML Building or BuildingPart is converted into a face-adjacency graph using PyTorch Geometric (Fey and Lenssen, 2019) and processed through four stages: topological analysis, defect detection, progressive repair, and semantic assignment. The pipeline is fully automated and designed to scale to large urban datasets.

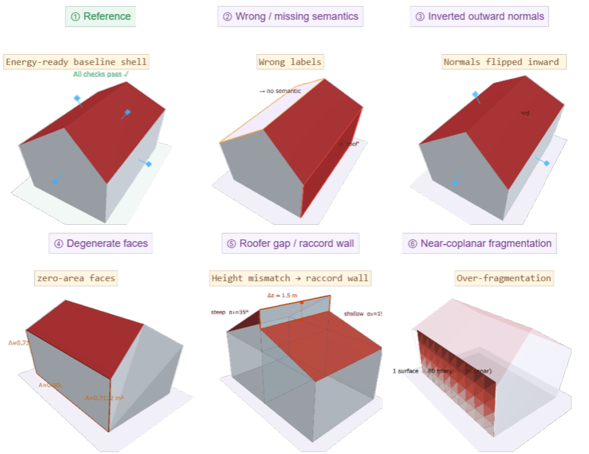

Figure 2. Some common geometric, topological and semantic errors in 3D geometries of building models.

2. Related works

The growing adoption of 3D city models raises a fundamental question of trust: are existing city models geometrically and semantically fit for the intended energy use?
Error severity is strongly application-dependent; defects that are negligible for visualization can substantially distort analytical results, including volume estimation, solar potential assessment and computational fluid dynamics simulations (Biljecki et al., 2016). Although existing validators support geometric and topological checking and repairing, e.g., CityDoctor (Betz et al., 2025), they are generally not designed for energy-oriented applications. Even so, they remain an essential step in model validation.

2.1 Requirements for Urban energy simulations

For energy-related simulations (Abbasabadi and Ashayeri, 2024; Longo et al., 2024), a building model must generally satisfy stricter conditions than those required for visualization (Agugiaro et al., 2018). At minimum and depending on the specific software requirements, the building envelope should nevertheless: (i) form a closed and watertight shell with no gaps or missing faces; (ii) be an orientable *2*-manifold with consistent outward-facing normals, ensuring an unambiguous interior–exterior distinction; (iii) be free of sliver polygons, degenerate faces, overlaps and self-intersections; and (iv) carry coherent surface semantics (e.g. for walls, roofs, etc.).
These requirements are strongly interdependent. Open shells prevent the derivation of enclosed volume and a thermally consistent boundary; non-manifold configurations undermine interior–exterior interpretation; incorrect normals invert the physical meaning of surfaces with respect to heat flux; and missing semantics hinder the assignment of thermal boundary conditions and construction properties.

Furthermore, formally valid geometry is not always sufficient: a model may pass geometric checks while remaining unsuitable for Urban Building Energy Modelling (UBEM) due to missing semantic differentiation or absence of adequate thermal zone attributes. This has been highlighted in previous work on geometric and semantic quality in urban 3D datasets (Biljecki et al., 2016) and remains central to the interoperability goals of energy-oriented urban modelling (Agugiaro and Padsala, 2025).

Feature to validate/check	**CityDoctor**	**val3dity**	**FME**	**our**
Purpose	Analysis + repair	Analysis	Analysis + repair	Analysis + repair
Input format	CityGML	CityGML, CityJSON, CityJSONSeq, JSON-FG, OBJ	CityGML, CityJSON (and many other formats)	CityJSON, mesh
Syntax/schema validation	✓	✓	✓	based on val3dity
Semantic validation	✓	✗	✓	partial
Geometric validation	✓	✓	✓	✓
Watertightness	✓	✓	✓	✓
Non-manifold detection	✓	✓	✓	✓
Shell orientation	✓	✓	✓	✓
Topological repair	partial	✗	partial	✓
Hole filling	✓	✗	✓	✓
Semantic surface classification	partial	✗	✗	partial
Energy-simulation validation	partial	✗	✗	✓

Table 1. Comparison of building model validation and repair features across tools with respect to the proposed solution.

2.2 Validation tools

Three tools are the most widely used in the literature:

- val3dity (Ledoux, 2018) is an open-source tool that performs rigorous geometric validation of 3D primitives according to ISO 19107[1], checking primitives in a hierarchical cascade, from `LinearRing` and `Polygon`, through `CompositeSurface` and `MultiSurface`, up to `Solid`, `CompositeSolid` and `MultiSolid`, where a lower-level failure stops all dependent checks.
- CityDoctor (Betz et al., 2025) is an open-source software that extends this with configurable validation plans covering schema, geometry and semantics, including application-specific checks for heating-demand simulation. It also provides a repair module to resolve topological and geometric issues.
- FME (FME Software) is a proprietary software that provides format-agnostic geometric checking - including non-manifold detection, shell closure and face orientation - within a broader *Extract, Transform, Load* (ETL) pipeline.

Previous studies applied these tools to further analyze building models. Biljecki et al. (2016) combined val3dity with Xerces[2] for XSD validation of CityGML files; Gao et al. (2024) proposed a robust method for fixing 3D building models via local remeshing; and Akın et al. (2025) introduced a graph-based approach for model validity checks. As summarized in Table 1, no single existing tool combines full geometric validation, semantic labelling and repair in one workflow like the proposed approach.

[1] https://www.iso.org/standard/66175.html

[2] https://xerces.apache.org/library.html

In the validation methodology presented in Section 4.1, only the open-source tools are considered.

2.3 Repairing methods

Repair approaches fall into two groups based on how much of the original geometry they alter.

- *Global reconstruction-based* methods rebuild entire surfaces from flawed input using volumetric or polyhedral representations. They handle severe corruption robustly but may distort sharp architectural features. Examples include ManifoldPlus (Huang et al., 2020), convex-hull shrink-wrapping for LoD2 buildings (Zhao et al., 2013) and unsigned distance function methods for point clouds (Hornung and Kobbelt, 2006).
- *Local topological repair* modifies only defective regions, preserving original geometry — critical for LoD2 buildings where roof and façade detail must be retained. Notable methods include topology-graph-based self-intersection resolution (Charton et al., 2021) and the hole-filling algorithm of Gao et al. (2024), which handles non-manifold edges and duplicate vertices without preconditions on normal orientation or semantic labels. Both strategies are typically combined with val3dity or CityDoctor to verify compliance.

3. Methodology

The proposed pipeline (Figure 3) processes LoD2 building geometries encoded in CityJSON. Each `Building` or `BuildingPart` — represented as a `Solid, CompositeSurface` or `MultiSurface` — is processed through three sequential phases: (i) detection of topological defects and energy-simulation compliance issues (Section 3.1), (ii) hierarchical repair escalating from local correction to global reconstruction (Section 3.2) and (iii) rule-based semantic labelling and thermal attribute derivation (Section 3.3).

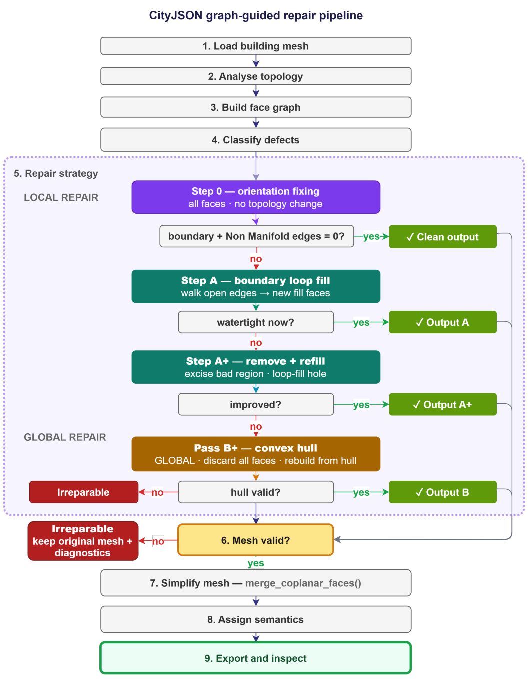

Figure 3. Complete workflow for the repairing algorithm.

Outputs are closed, orientable *2*-manifold solids with semantically labelled surfaces and per-surface thermal transmittance (U-value, W/m²K), suitable for multi-level Energy Digital Twin workflows at both solid-geometry and thematic-surface level (Roman et al., 2025b).

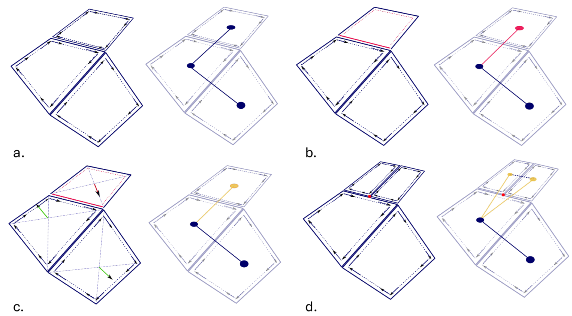

Figure 4. Graph topology: a. correct configuration; b. Detection of a hole in the surface; c. non-manifold geometry with incorrect normals; d. non-manifold geometry.

3.1 Inconsistency detection through graph analysis

Each building is represented as a face-adjacency graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, X, Z)$ constructed in *PyTorch Geometric* (Fey and Lenssen, 2019; *PyTorch v2.2.2*), where each node $v_i \in \mathcal{V}$ represents a mesh face f_i and each undirected edge $(v_i, v_j) \in \mathcal{E}$ connects two faces sharing a mesh boundary edge, encoding both geometry and connectivity so that defective regions (Figure 4) are identified as structurally connected subgraphs rather than isolated polygons. Node features are encoded as a 10-dimensional vector x_i:

$$x_i = \left[A_i,\ n_{\{x,i\}},\ n_{\{y,i\}},\ n_{\{z,i\}},\ h_i,\ d_i,\ p_i,\ b_i,\ m_i,\ \deg(i)\right]$$

where A_i is the face area; $(n_{x,i}, n_{y,i}, n_{z,i})$ the unit outward normal; h_i the centroid height normalized to the building footprint extent; $d_i \in 0{,}1$ a binary degeneracy flag for zero-area or collinear faces; p_i the planarity deviation as maximum vertex distance from the best-fit plane; b_i and m_i the counts of boundary and non-manifold edges incident to f_i; and $de\,g(i)$ the node valence. Each edge carries a 3-dimensional feature vector:

$$\mathbf{z}_{ij} = \left[\cos(\theta_{ij}),\ \beta_{ij},\ \mu_{ij}\right]$$

where $\cos(\theta_{ij})$ is the dihedral angle cosine between f_i and f_j; $\beta_{ij} \in \{0,1\}$ flags a boundary edge $(c(e) = 1)$; and $\mu_{ij} \in \{0,1\}$ flags a non-manifold edge $(c(e) > 2)$, with $c(e)$ denoting the edge incidence count. Topological consistency is assessed jointly via val3dity and edge-incidence analysis on $\mathcal{G}$.
Edges are classified as manifold $(c(e) = 2)$, boundary $(c(e) = 1$, signalling a hole or disconnected region) or non-manifold $(c(e) > 2$, violating the 2-manifold condition). Defining (Eqs. 1-2):

$$N_b = \mid e \in \mathcal{E} : c(e) = 1 \mid, \text{and} \qquad (1)$$

$$N_{nm} = \mid e \in \mathcal{E} : c(e) > 2 \mid \qquad (2)$$

a shell is watertight and topologically valid only when $N_b = 0$ and $N_{nm} = 0$. Figure 5 illustrates representative outcomes, where blue nodes denote valid faces, red nodes topologically inconsistent faces and yellow edges non-manifold connections.
However, topological validity (Figure 3) alone is not sufficient for energy simulation. In fact, a building is classified as energy-compliant if and only if:

i. $N_b = 0$ and $N_{nm} = 0$: the shell is watertight and *2*-manifold;
ii. all face normals are consistently outward-oriented, ensuring correct heat-flux direction;

https://doi.org/10.5194/isprs-annals-XII-4-W1-2026-293-2026

iii. no face is degenerate, self-intersecting or a sliver polygon ($with\ A_i > \varepsilon, \varepsilon = 10^{-4}$m²);
iv. every face carries a valid semantic label from {`RoofSurface`, `WallSurface`, `GroundSurface`};
v. every surface carries a valid thermal transmittance value U_i(W/m²K).

Conditions (i) – (iii) are evaluated directly from $\mathcal{G}$ (Figure 5); conditions (iv) – (v) are assessed after semantic assignment (Section 3.3).

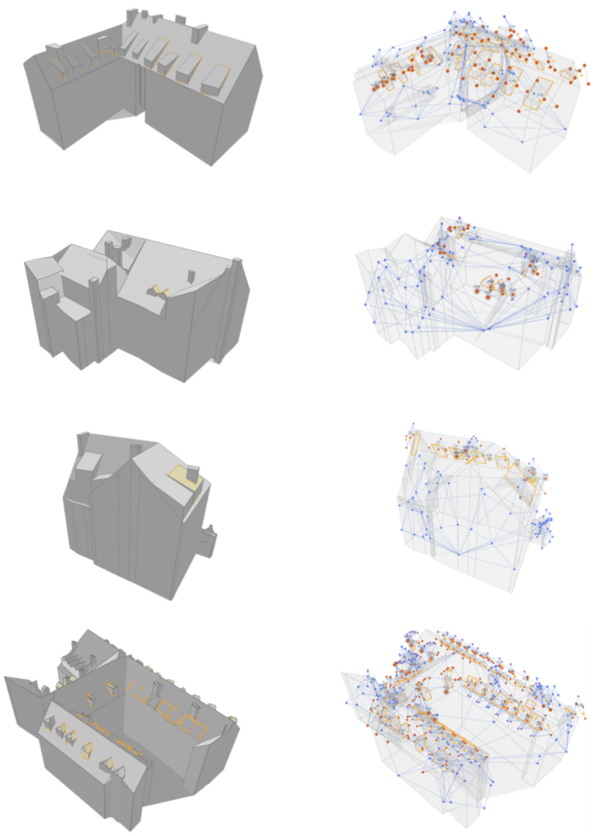

Figure 5. Topological inconsistencies in four buildings, shown in the face-based graph. Nodes represent $\mathcal{G}$ faces, lines represent $\mathcal{G}$ edges.

3.2 Repairing strategies

The repair strategy (Figure 3) follows an iterative process in which progressively more invasive operations are applied to defective buildings, moving from (i) local repairs, to (ii) extended local interventions and finally to (iii) global reconstruction strategies, where required.
Each stage is evaluated against the validity condition ($N_b = 0$, $N_{nm} = 0$); success at any stage terminates the process and advances the building to post-processing. The main steps involved are:

1. *Orientation fixing* (Figure 6). This is the first and least invasive intervention. Face normals are flipped to achieve a consistent outward mesh orientation without adding or removing any geometry. If this alone resolves all defects, the building is marked *clean* and it's already sent to post-processing.
2. *Boundary loop fill* addresses remaining holes by walking all open boundary edges to extract closed hole rings, which are then filled with new planar projected faces via constrained Delaunay triangulation (Blazek, 2005). The operation is strictly local, preserving all original geometry outside the filled regions. A watertight result at this stage is labelled *Output A*.
3. *Remove and refill* are applied when the region surrounding a hole is itself corrupted and cannot be filled directly. The defective neighborhood is recalled as a boundary constraint and the resulting gap is reconstructed locally. Though more invasive than the previous stage, it affects only the damaged region. Measurable reduction in N_b and N_{nm} produces *Output A+*.
4. *Alpha shape reconstruction* (Edelsbrunner and Mücke, 1994) is applied when local repair fails to produce a watertight shell. In such cases, the pipeline applies 3D alpha-shape reconstruction. The parameter α controls the reconstruction scale by defining the maximum circumradius R of the retained tetrahedra, with R ≤ α. Figure 7 illustrates how the method starts from the graph-detected defect and evaluates increasing α values to reconstruct a valid building shape. A candidate is valid when the numbers of boundary and non-manifold edges satisfy $N_b = 0$ and $N_{nm} = 0$.

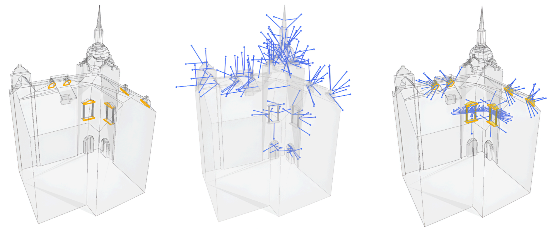

Figure 6. Example of fixed orientation of normals.

If several valid candidates exist, the one with the smallest deviation from the input geometry is selected; otherwise, the candidate that most reduces $N_b + N_{nm}$ is retained. If 3D reconstruction fails, the pipeline falls back to 2D alpha-shape footprint extrusion and, ultimately, a 3D *convex hull*.
Concretely, all steps involved in loop filling, remeshing and geometry restoration were carried out using a strictly constrained Delaunay triangulation approach (Blazek, 2005).

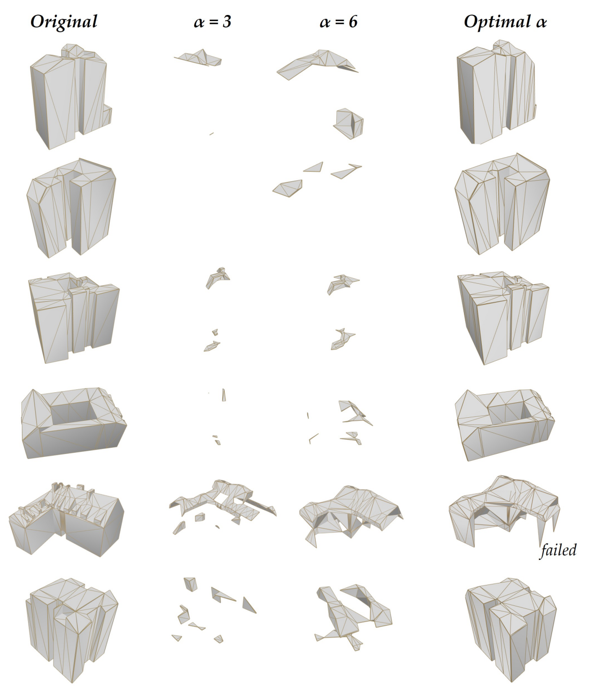

Figure 7. Effect of α on building reconstruction. The optimal α minimizes geometric deviation while ensuring a watertight manifold. Some buildings remain unresolved.

Likewise, the *alpha shape* (and *Convex Hull*) is strictly bounded by the outer vertices, ensuring that the input shape is strictly preserved.
Final failure marks the building as irreparable, retaining the original mesh alongside full diagnostics.

3.3 Semantic fixing and thermal values assignment

Remeshing can substantially alter both the number of faces and the triangulation pattern. As a result, semantic inconsistencies may arise, for example, when roof faces extend into walls or wall faces appear above the roof. To address this, we introduce a semantic checking and partial repairing module that also assigns labels when they are missing in the input dataset (Figure 8).
Face semantics are inferred from geometric rules based on centroid elevation (z_c), relative height and surface normal orientation (n_z). Specifically, a face is classified as `GroundSurface` if $z_c - z_{min} \leq 0.05$m, as `RoofSurface` if $n_z \geq 0.70$, or if $n_z \geq 0.30$ and $h_{rel} = \frac{(z_c - z_{min})}{(z_{max} - z_{min})} \geq 0.65$ and as `WallSurface` otherwise. For each repaired building, the workflow then derives thermal-envelope attributes, including the total envelope area (Eq. 3):

$$A_{env} = A_{roof} + A_{wall} + A_{ground}, \quad (3)$$

and the aggregate heat-loss coefficient *UA* (Eq. 4):

$$\text{UA} = \sum_i U_i A_i. \quad (4)$$

These outputs preserve repair-stage provenance and support direct comparison with val3dity results.

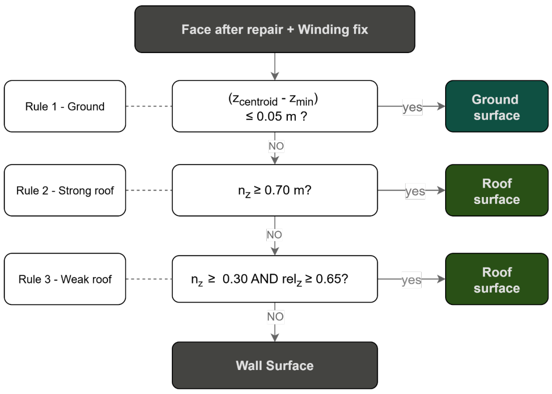

Figure 8. Assignment of semantics based on geometric rules.

3.4 Evaluation metrics

Geometric accuracy was evaluated by comparing the output models with the input buildings used as ground truth (GT). Two metrics were used: mean Intersection over Union (mIoU) and mean vertex position error on the outer envelope (mAP_v) (Eqs. 5 and 6):

$$\text{mIoU} = \frac{|I_{in} \cap O_{out}|}{|I_{in} \cup O_{out}|} \quad (5)$$

$$\text{mAP}_v = \frac{1}{|V|} \sum_{i=1}^{|V|} |v_i^{in} - v_i^{out}|_2 \quad (6)$$

Where I_{in} and O_{out} are the enveloped input buildings and output buildings, respectively, while $|V|$ is the number of corresponding vertices on the outer envelope and v_i^{in} and v_i^{out} are the positions of the i-th vertices in the input and output buildings.

Dataset	**LOD**	**# Buildings**	**Has semantics – [%] faces**	**AOI [km²]**
Berlin	LOD 2.2	8332	True/0.99	4.88
Delft	LOD 2.2	5456	True/0.97	2.42
Den Haag	LOD 2	1990	True/0.74	0.52
Hamburg	LOD 2.2	7248	True/0.99	1.14
Ingolstadt	LOD 3.0	379	True/0.55	0.09
Rotterdam	LOD 2.0	6708	True/0.97	4.21
Tallinn	LOD 2.0	4687	False/0.00	15.09
Wien	LOD 2.0	8752	True/0.99	2.40
Zurich	LOD 2.0	145865	True/1.00	151

Table 2. Summary of the datasets considered, with their source and associated tile identifiers or dataset size.

4. Results

The proposed algorithm was evaluated using a set of LoD2 city models drawn from the benchmark dataset provided by Wysocki et al. (2024)[3]. The datasets come from multiple sources and vary in their geometric characteristics. Their provenance and number of buildings and extension (area of interest, AOI) information is summarized in Table 2. The following sections report results for graph-based topological analysis, reconstruction metrics, mesh face reduction and visual assessment.

4.1 Approach validation

A controlled test set was assembled (146 buildings) using the Hamburg dataset to verify each repair stage independently, ensuring that the pipeline behaves correctly across the full range of defect types before application to large datasets.
Using the test buildings (Figure 9) the algorithm evaluated the buildings against a manually derived mesh ground truth (GT). As shown in Table 2, the dataset is provided by semantics, so after the repairing process, semantics have been restored, remapping faces as explained in Section 3.4, generating energy-compliant buildings.

	Geometric	**Energy compliant**	**mIoU**	**mPAv**	**Correct**
Input	25/146	24/146	-	-	0.17
Output	144/146	144/146	0.97	0.213	0.99

Table 3. Metrics for the controlled test.

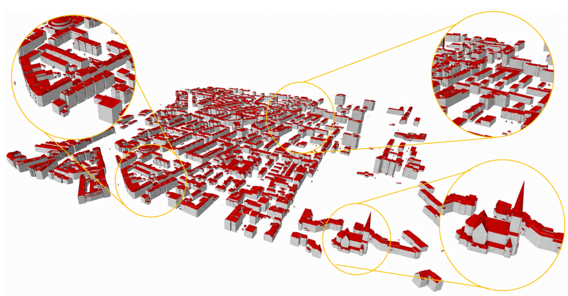

Figure 9. Controlled dataset for the workflow validation.

[3] https://github.com/OloOcki/awesome-citygml

4.2 Evaluation on all datasets - Analysis

Table 4 benchmarks CityDoctor and val3dity validators, and the energy-compliant analysis explained in Section 3.1.
The proposed method builds on val3dity for validation (column 'Geom' in Table 4), while column *Energy Compliant* extends the energy-compliance analysis beyond the metrics presented in Section 3.1.

Datasets	CityDoctor (geom.)	val3dity (geom)	ours Geom	ours Energy compliant
Berlin	0.98	0.97	0.98	0.46
Delft	0.98	0.98	0.98	0.00
Den Haag	0.62	0.62	0.62	0.88
Hamburg	0.58	0.94	0.94	0.31
Ingolstadt	0.40	0.70	0.70	0.05
Rotterdam	0.98	0.96	0.96	0.00
Tallinn	0.00	0.00	0.00	0.00
Wien	0.99	0.37	0.37	0.58
Zurich	0.98	0.98	0.98	0.36

Table 4. Analysis of the input datasets before repairing: valid buildings detected across the input datasets using different methods.

Concretely, the algorithm performs an in-depth assessment on both Building shell level (`Solid`) and also on `CompositeSurface` level on properties like (i) watertightness, (ii) surface tessellation quality based on mesh analysis, (iii) correct normal orientation of the building envelope at `Solid` and `MultiSurface` level, a (iv) the consistency of semantic and (v) thermal parameters for each component.

4.3 Evaluation on all datasets - Repairing

Results reported in Table 5 summarize the errors repaired at each iterative step of the workflow (local and global repairing steps).

Datasets	Invalid Buildings / Buildings	Local repairing	Global repairing	Errors left	Repaired buildings
Berlin	4499/ 8332	1516	817	**2166**	0.74
Delft	5456/5456	1410	3202	**844**	0.84
Den Haag	1990/1990	912	1078	**0**	1.00
Hamburg	5001/7248	4627	42	**332**	0.95
Ingolstadt	360/379	142	104	**114**	0.69
Rotterdam	6708/6708	3157	2103	**1448**	0.78
Tallinn	4687/4687	3942	323	**422**	0.91
Wien	3676/8752	2479	934	**263**	0.96
Zurich	93354/145865	89611	534	**3209**	0.97

Table 5. Performance of the proposed method, showing the percentage of buildings repaired at each stage and the final overall repair rate.

Furthermore, Figure 10 illustrates the distribution of repaired buildings across the successive repair and remeshing stages.
The results distinguish between local repair stages and global repair procedures (Alpha Shape and remeshing), as described in Section 3.2. For some datasets, including Zurich, Hamburg and Den Haag, the local repair loop is sufficient to correct all or almost all buildings. In several cases, topological inconsistencies are resolved solely through normal checking combined with vertex deduplication.
However, even if percentages are promising, the number of errors that remain in these datasets is still high. The results for the Zurich dataset, in particular, show that a substantial number of buildings cannot be successfully repaired.

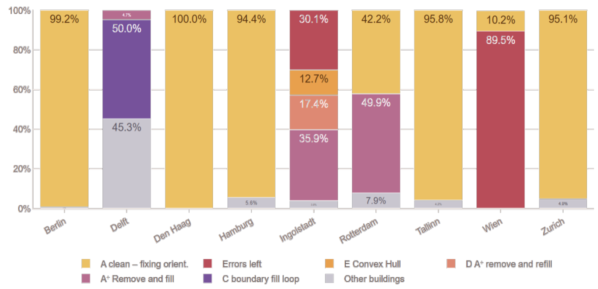

Figure 10. Percentage distribution of repair strategies applied to each dataset.

4.4 Topology Consistency and Energy Compliance

Table 6 reports energy-compliance rates before and after repair, together with geometric fidelity measured by *mIoU* and *mAPv* (Eqs. 5-6). *IoU* measures the overlap between the input and repaired building envelopes, whereas *mAPv* quantifies the average displacement of corresponding envelope vertices. Higher *mIoU* and lower *mAPv* therefore indicate better preservation of the original geometry. As shown in Figure 11, all datasets achieve high *mIoU* values, ranging from 0.932 to 0.989, indicating that the overall building envelopes are generally preserved.

Dataset	Energy Compliant [before/after]	mIoU	mAP_v
Berlin	0.46 / 0.78	0.932	0.154
Delft	0.00 / 1.00	0.941	0.966
Den Haag	0.88 / 0.92	0.983	0.491
Hamburg	0.31/ 0.96	0.967	0.272
Ingolstadt	0.05 / 0.70	0.989	0.457
Rotterdam	0.00 / 0.78	0.966	1.989
Tallinn	0.00 / 0.76	0.978	0.678
Wien	0.58 / 0.81	0.939	0.186
Zurich	0.36 / 0.97	0.941	0.674

Table 6. Energy compliance before and after repair, with geometric accuracy measured by *mIoU* and *mAPv*.

However, *mAPv* reveals greater variation in the magnitude of local geometric changes. Berlin and Wien show the lowest displacements, at 0.154 m and 0.186 m, respectively. Hamburg, Ingolstadt and Den Haag also exhibit relatively limited changes, with *mAPv* values between 0.272 m and 0.491 m.

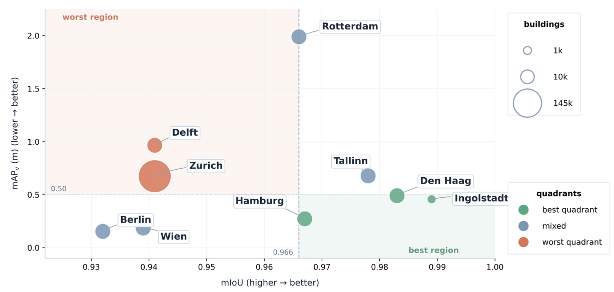

Figure 11. Per-dataset visualization of the results, expressed in terms of *IoU* and *mAPv*.

Tallinn and Zurich show larger, but still moderate, displacements of approximately 0.67 m.Delft and Rotterdam are the most challenging datasets, with *mAPv* values of 0.966 m and 1.989 m, respectively. These larger deviations reflect the higher proportion of corrupted geometries requiring global reconstruction.

The resulting envelopes satisfy all five energy-compliance conditions but deviate significantly from the original building form; simulation outputs for these buildings should be treated as approximate. Figures 12 and 14 show representative repair outcomes: input geometry (left), initial topological status - yellow for non-manifold, red for non-watertight (centre) - and repaired energy-compliant buildings in green (right).

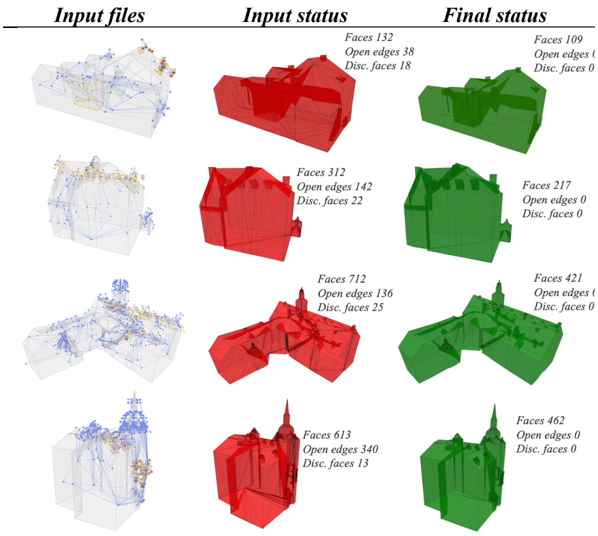

Figure 12. Building examples with initial inconsistences and final repaired and remeshed results.

4.5 RAM usage

Figure 13 shows the RAM consumption measured during execution of the complete repair pipeline, parallelized across all available CPU cores. Peak memory usage remained below 600 MB throughout the experiment, indicating a compact computational footprint. This efficiency is enabled by the lightweight face-adjacency graph, which supports defect detection, hole filling, self-intersection removal, vertex snapping and deduplication without requiring dense volumetric representations or memory-intensive global mesh structures.

4.6 Limitations

Tables 4–6 show that the pipeline substantially improves topological quality and energy compliance across all datasets. In Zurich, Hamburg and Den Haag, most buildings are resolved through local operations such as normal correction, boundary closure and vertex deduplication, although large datasets may still contain a significant number of unresolved cases. Wien, despite having comparatively valid geometries, also requires further processing to meet energy-simulation requirements.

Severely corrupted buildings may require global reconstruction, which can increase face and edge counts through surface subdivision. While *mIoU* remains high across all datasets, *mAPv* reveals greater variation in geometric change: Berlin and Wien show the lowest displacements, Hamburg, Ingolstadt and Den Haag remain relatively limited, and Tallinn and Zurich exhibit moderate deviations. Semantic classification is still constrained by rule-based methods and thermal properties such as U-values rely on simplified assumptions.

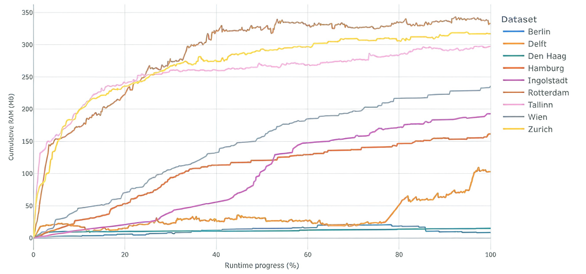

Figure 13. Normalized RAM usage over all the workflows.

5. Conclusions

This paper presented a graph-guided pipeline for qualifying and repairing LoD2 building geometries for urban-scale energy simulation. It combines a face-adjacency graph with hierarchical topology-aware repairs to restore *2*-manifold consistency and watertightness while preserving the original geometry whenever possible. A rule-based module then assigns surface semantics and thermal attributes.

Evaluation on nine real-world datasets comprising more than 190,000 buildings yielded post-repair energy-compliance rates of 70–100% and *mIoU* values of 0.932–0.989, indicating strong overall geometric preservation. Berlin and Wien showed the lowest vertex displacements, whereas Delft and Rotterdam exhibited the largest deviations because more buildings required global reconstruction. These cases also show that alpha-shape reconstruction may reduce architectural detail. The LoD3 Ingolstadt dataset further demonstrates the potential to extend the method beyond LoD2.

Future work should reduce reliance on global reconstruction, improve semantic classification for complex building forms and derive more realistic thermal attributes from material data or building typologies. Support for XML-encoded CityGML and multi-zone thermal model generation would further expand the pipeline's applicability.

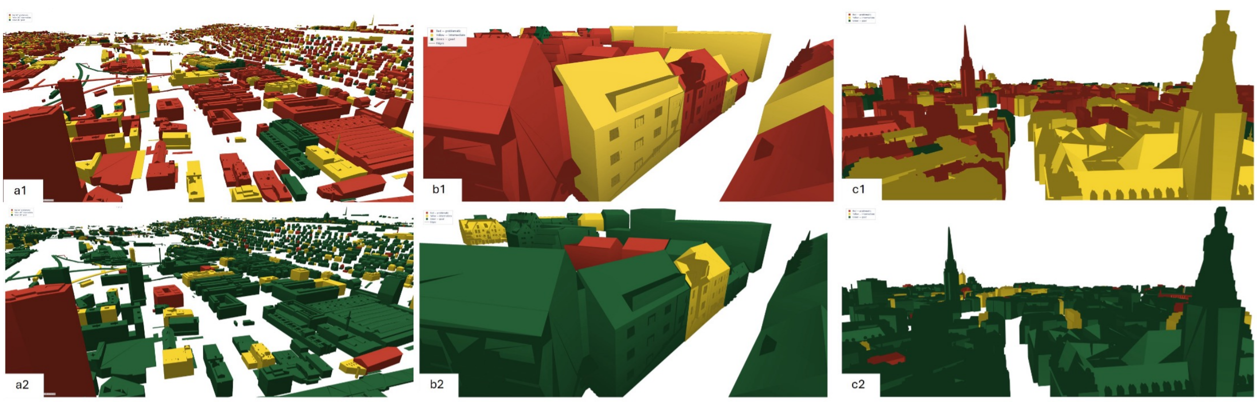

Figure 14. Initial and repaired geometries of three datasets: (a) Zurich, (b) Ingolstadt and (c) Hamburg. Panels a1–c1 show the initial states, while a2–c2 show the corresponding repaired states.

References

Abbasabadi, N., Ashayeri, M., 2024. Digital Twin for Citywide Energy Modeling and Management. *Artificial Intelligence in* Agugiaro, G., Benner, J., Cipriano, P., Nouvel, R., 2018. The Energy Application Domain Extension for CityGML: Enhancing interoperability for urban energy simulations. *Open Geospatial Data, Software and Standards,* 3(1), 2.

Agugiaro, G., Padsala, R., 2025. A proposal to update and enhance the CityGML energy application domain extension. *ISPRS Ann. Phot. Rem. Sens. Spat. Inf. Sci., X-4/W6-2025*, pp. 1-8.

Alkadri, M., De Luca, F., Turrin, M., Sariyildiz, S., 2020. A computational workflow for generating a voxel-based design approach based on subtractive shading envelopes and attribute information of point cloud data. *Remote Sens.* 12, 2561.

Akın, A.T., Usta, Z., Stoter, J., Arroyo Ohori, K., Cömert, Ç., 2025. Leveraging knowledge graphs and semantic web technologies for validating 3D city models. *International Journal of Geographical Information Science*, 1–27.

Amrullah, C., Panangian, D., Bittner, K., 2025. Polyroof: Precision roof polygonization in urban residential buildings with graph neural networks. Proc. JURSE, 1-4.

Arroyo Ohori, K., Biljecki, F., Kumar, K., Ledoux, H., Stoter, J., 2018. Modeling Cities and Landscapes in 3D with CityGML. *Building Information Modeling: Technology Foundations and Industry Practice.* Springer Int. Publishing, pp. 199-215.

Betz, M., Riegel, A., Coors, V., 2025. Storing quality validation results in CityGML with the QualityADE. *Int. Arch. Phot. Remote Sens. Spatial Inf. Sci., XLVIII-4/W15-2025*, 31–37.

Biljecki, F., Ledoux, H., Du, X., Stoter, J., Soon, K.H., Khoo, V.H.S., 2016. The most common geometric and semantic errors in CityGML datasets. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-2/W1*, pp. 13-22.

Blazek, J., 2005. Principles of grid generation. In: Blazek, J. (ed.), *Computational Fluid Dynamics: Principles and Applications*, 2nd ed. Elsevier Science, Oxford, pp. 373–413.

Charton, J., Baek, S., Kim, Y., 2021. Mesh repairing using topology graphs. *Journal of Computational Design and Engineering, 8(1)*, 251-267.

Edelsbrunner, H., Mücke, E.P., 1994. Three-dimensional alpha shapes. *ACM Transactions on Graphics, 13(1),* pp. 43-72.

Fey, M., Lenssen, J.E., 2019. Fast graph representation learning with PyTorch Geometric. *arXiv preprint arXiv:1903.02428.*

FME Software. Copyright (c) Safe Software Inc. Available at: https://www.safe.com (last access: Jul 2nd, 2026).

Gao, W., Peters, R., Ledoux, H., Stoter, J., 2024. Filling holes in LoD2 building models. *ISPRS Ann. Phot. Rem. Sens. Spat. Inf. Sci., X-4/W5-2024*, pp. 171-177.

Gröger, G., Kolbe, T.H., Nagel, C., Häfele, K.H., 2012. OGC City Geography Markup Language (CityGML) Encoding Standard, Version 2.0. *Open Geospatial Consortium, OGC Doc.* pp. 12-19

Hornung, A., Kobbelt, L., 2006. Robust and efficient photo consistency estimation for volumetric 3D reconstruction.*Proc. ECCV,* pp. 179-190.

Performance-Driven Design: Theories, Methods, and Tools, Wiley, 2024, pp. 251-276

Huang, J., Zhou, Y., Guibas, L., 2020. ManifoldPlus: A robust and scalable watertight manifold surface generation method for triangle soups. *arXiv preprint arXiv:2005.11621.*

Jarząbek-Rychard, M., Rigon, S., Boguslawski, P., Remondino, F., 2025. Automating 3D Building Modeling: A Comparative Study of Data- and Model-Driven Methods. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLVIII-G-2025, 701-707.

Ledoux, H., 2018. val3dity: Validation of 3D GIS primitives according to the international standards. *Open Geospatial Data, Software and Standards*, 3.

Ledoux, H., Arroyo Ohori, K., Kumar, K., et al., 2019. CityJSON: A compact and easy-to-use encoding of the CityGML data model. *In: ISPRS Geospatial Week.*

Longo, D., Turillazzi, B., Roversi, R., Lilla, S., Nucci, C.A., Piccinini, A., Costa, A., 2024. Urban digital twin and energy modeling: experiences and case study analyses. *International Journal of Architecture, Art and Design*, n.15, pp. 160-169.

Pađen, I., Peters, R., García-Sánchez, C., Ledoux, H., 2024. Automatic high-detailed building reconstruction workflow for urban microscale simulations. *Build Environ. 265*, 111978.

Queffélec, P.-L., Trouvé, N., Roussel, S., Wu, T., and Vallet, B., 2026. Point2WSS: Reconstructing LoD2 Buildings from Aerial LiDAR Data using Multimodal Learning and Weighted Straight Skeleton. *ISPRS Ann. Phot. Rem. Sens. Spat. Inf. Sci.,* XI-2-2026, 331-340.

Roman, O., Bassier, M., Ricciuti, S., Farella, E. M., Remondino, F., Viesi, D., 2025a: Digital Twins and CFD simulations for accurate sensor positioning, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-G-2025*, 1291–1298.

Roman, O., Bassier, M., Agugiaro, G., Arroyo Ohori, K., Farella, E. M., Remondino, F., 2025b. Scan-to-EDTs: Automated Generation of Energy Digital Twins from 3D Point Clouds. *Buildings, 15(22),* 4060.

U.S. Department of Energy's (DOE) Building Technologies Office (BTO), Energy Plus, 2025. Available at: https://energyplus.net/ (last access: Jul 9th, 2026).

Wang, R., Huang, S., Yang, H., 2023. Building3D: An urban scale dataset and benchmarks for learning roof structures from point clouds. Proc. *IEEE ICCV*.

Wysocki, O., Schwab, B., Beil, C., Holst, C., Kolbe, T. H., 2024. Reviewing Open Data Semantic 3D City Models to Develop Novel 3D Reconstruction Methods. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-4-2024*, pp. 493-500.

Yao, Z., Nagel, C., Kunde, F., Hudra, G., Willkomm, P., Donaubauer, A., Adolphi, T., Kolbe, T.H., 2018. 3DCityDB - a 3D geodatabase solution for the management, analysis, and visualization of semantic 3D city models based on CityGML. *Open Geospatial Data, Software and Standards,* 3, 5.

Zhao, J., Ledoux, H., Stoter, J., 2013. Automatic repair of CityGML LoD2 buildings using shrink-wrapping. *ISPRS Ann. Phot. Rem. Sens. Spat. Inf. Sci., II-2/W1*, pp. 309-317.