A new benchmark on LoD 2 building reconstruction from aerial lidar and footprints

Florent Geniet[1], Emmanuel Séguin [1], Etienne Le Bihan [1], Liuyun Duan[2], Nicolas Girard [2], Ravi Peters[3], Bruno Vallet [1]

[1] Univ Gustave Eiffel, ENSG, IGN, LASTIG, F-94160 Saint-Mandé, France
(florent.geniet, emmanuel.seguin, etienne.le-bihan, bruno.vallet)@ign.fr
[2] LuxCarta - Zone de l'Argile – 460, Avenue de la Quiéra – Voie K – Bât 119 B, F-06370 Mouans-Sartoux, France
(lyduan,ngirard)@luxcarta.com
[3] 3DGI, Zoetermeer, The Netherlands, ravi.peters@3dgi.nl

Keywords: Building reconstruction, Digital twin, CityGML, LoD2, benchmark, lidar, building footprints

Abstract

The CityGML norm proposes specifications for the 3D representation of most urban objects, and in particular defines levels of details (LoD) for building models. The LoD2 corresponds to polyhedral roof structures representing the main roof slopes, but not superstructures as dormers and chimneys. Nowadays, several countries have a nationwide lidar program with densities well adapted to LoD2 building modeling and open topographic databases containing the building footprints, so automating LoD2 reconstruction from lidar and footprints allows the production of nationwide LoD2 models. This paper proposes a new benchmark to evaluate the quality, scalability and robustness of state of the art LoD2 building reconstruction from lidar and footprints. It is based on a subset of LiDAR HD data freely provided by IGN, the French mapping agency, along with building footprints derived from a high-quality, manually produced ground-truth dataset created by IGN. Results from various state-of-the-art algorithms are evaluated using the open source PyScoring tool that compares the results with the ground truth.

1. Introduction

1.1 Context

Digital twins are becoming a standard paradigm to create an interaction between real world objects and digital counterparts. In the domain of geographical information sciences, digital twins rely on a faithful 3D model of entire territories. An important layer of these 3D models are buildings, that can be represented using the CityGML specification. Producing such 3D buildings at large scale usually relies on a high density aerial lidar acquisition and building footprints (which may come from existing databases or be automatically produced themselves). The main stake to put up such large scale productions is having access to a 3D building reconstruction tool that is efficient, flexible and robust enough to cope with the important amount of data to process and variety of building shapes and sizes at country scale.

1.2 Objectives

This paper is interested in benchmarking tools that are able to

- automatically produce semantic 3D building models in the CityGML LoD2 specification, or something close enough to be easily converted. In practice, we used the CityJSON format that implements this specification (Ledoux et al., 2019).
- take as input classified aerial lidar and building footprints. Classification is important to filter out non building points projecting to building footprints (such as foliage).
- efficiently process very large amounts of data (region/country scale). In practice the problem is split by footprint, so the computing time should be reasonable even for very large areas.
- robustly cope with and adapt to the very large diversity of buildings that are present at such scales.

Benchmarking the large volume of produced data also requires automatically comparing the results with a ground-truth dataset using appropriate metrics.

1.3 Previous works

Because 3D building reconstruction is an important and difficult problem, many solutions have arisen over the years, and benchmarking efforts have already been done to evaluate objectively these approaches. The work most similar to ours is probably the ISPRS benchmark on urban object classification and 3D building reconstruction (Rottensteiner et al., 2012). The metrics that we propose are inspired by theirs (combining geometric and topological metrics). The main differences are that this work also focused on classification, and the input also covered airborne imagery while we focus on lidar.

More recently, a Multi-Purpose Benchmark was proposed for Urban 3D Reconstruction (Özdemir et al., 2019) that focuses on photogrammetric reconstruction, and another benchmark was also proposed to address the related problems of building point cloud completion (Gao et al., 2024). Finally, an evaluation of 3D building reconstruction from Low-Cost Terrestrial Photogrammetry was proposed in (Altman et al., 2017)

Newer benchmarks cover mainly learning-based approaches, which are reviewed in (Buyukdemircioglu et al., 2022), and a benchmark and dataset for learning roof structures from point clouds is proposed in (Wang et al., 2023). We chose not to address learning-based approaches because they do not yet seem robust enough to address the wide diversity of buildings on the country scale, and their results are extremely dependent on the data used for learning.

1.4 Contributions

- We selected existing methods that display such potential
- We selected test areas covering different building types and urban forms
- We collected lidar, building footprints and ground truth LoD2 reconstructions on these areas
- We proposed an extension of the open source PyScoring to compute 3D metrics (in addition to the 2D metrics already implemented).
- We produced LoD2 reconstructions with the selected methods on all test areas (possibly pre/post processing data when necessary)
- We quantitatively evaluated these reconstructions by computing intrinsic and extrinsic (by comparison with a ground truth) metrics using Pyscoring
- We analyzed the resulting metrics to give a clear overview of the pros and cons of each evaluated method.
- We put up a website [1] detailing the benchmark, its results and how to participate

2. Benchmark

2.1 Test Area Locations

For this benchmark, four test areas were selected around and inside Marseille, in France, as illustrated in Figure 1:

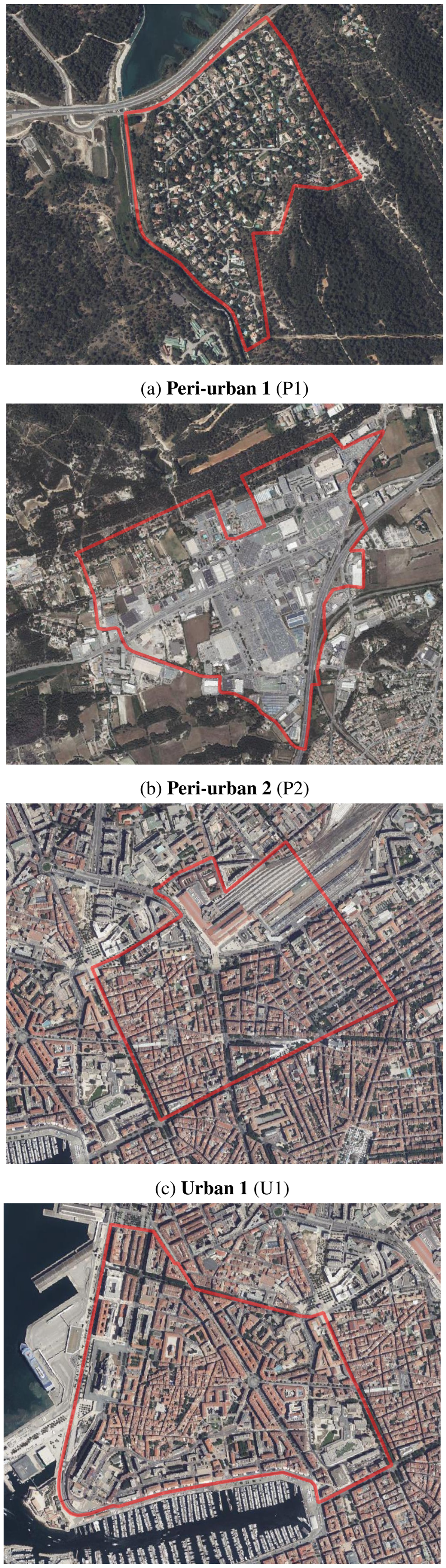
- **Peri-urban 1** (P1): single-family housing development, 316 footprints, 0.4 km². The abundant high vegetation, causing varying point density on the roofs makes this dataset challenging.
- **Peri-urban 2** (P2): commercial zone, 546 footprints, 1.5 km². This area will allow to assess the reconstruction of very large buildings.
- **Urban 1** (U1): densely homogeneous constructed area in the center of Marseille, 1735 footprints, 0.6 km². This area will allow to assess the reconstruction of strongly adjacent building clusters.
- **Urban 2** (U2): more diverse building types including remarkable buildings (eg : the Marseille Cathedral, large shopping malls), 1458 polygonal footprints, 0.8 km². This area will allow to assess the reconstruction of diverse and complex building clusters.

2.2 Datasets

In these four test areas, the following datasets have been used:

- **Classified LiDAR HD dataset** (Figure 2): The LiDAR HD product of IGN (the french national mapping agency) is a lidar point cloud with a density of at least 10 pulses/m² covering the whole french territory (except French Guiana). The point cloud is split in 1 km² tiles

[1] https://sites.google.com/view/lod2-building-benchmark



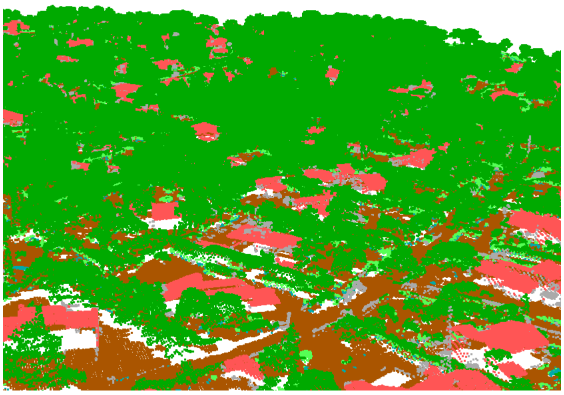
(a) **Peri-urban 1** (P1)

(b) **Peri-urban 2** (P2)

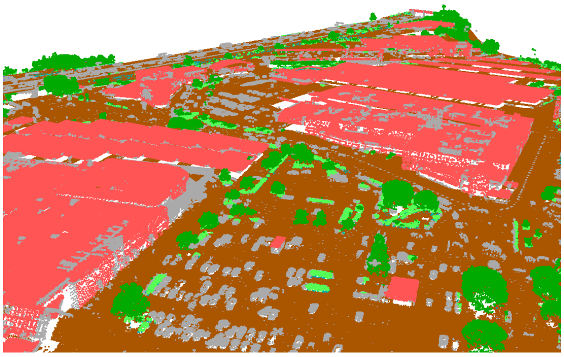
(c) **Urban 1** (U1)

(d) **Urban 2** (U2)

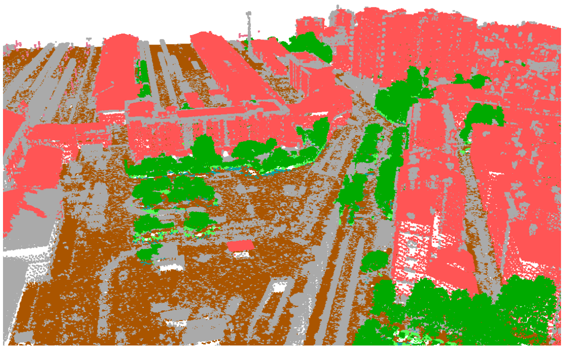
Figure 1. Orthophotos of the four test areas of the benchmark



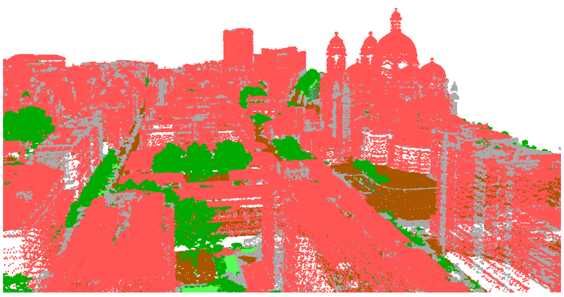
(a) **Peri-urban 1** (P1)



(b) **Peri-urban 2** (P2)



(c) **Urban 1** (U1)



(d) **Urban 2** (U2)

Figure 2. Semantized lidar scans of the four areas: red=buildings, brown=ground, dark green=high vegetation, light green=low vegetation, gray=clutter

that can be freely downloaded and comes with a point-wise classification produced by IGN. In this benchmark, we kept the points from classes 6 (building) and 67 (suspected building) to include all available potential building points.

- **3D LoD2 buildings ground-truth dataset** : These buildings have been produced manually by professional operators using IGN software GeoView® based on high-resolution aerial photography using stereo reconstruction. A complementary analysis was conducted to ensure the consistency of this product with the LiDAR HD point cloud dataset using CloudCompare tool Cloud/Mesh Distance. This high quality process took over 300 hours of operator hours.

- **Building footprints** : In order to avoid any discrepancies between the 3D models the choice was made to derive building footprints from the ground-truth dataset and use them as input for building reconstruction

This input data (lidar and footprints) is open and freely available to the scientific community and in particular to potential benchmark participants and can be downloaded from the benchmark website [2], but the ground truth will remain closed. In order to bootstrap the benchmark, we evaluate relevant state of the art methods in this paper.

2.3 Evaluated methods

We decided to evaluate a few methods that display the potential to fit the objectives listed above.

2.3.1 TerraScan/Vectorize Buildings The *Vectorize Buildings* tool of the TerraScan commercial software was used to reconstruct LoD2 buildings on the aforementioned datasets. The tool creates models representing each building. Each building has individual roof planes represented, and walls and floor extruded from roof edges to the specified offset from the ground level.

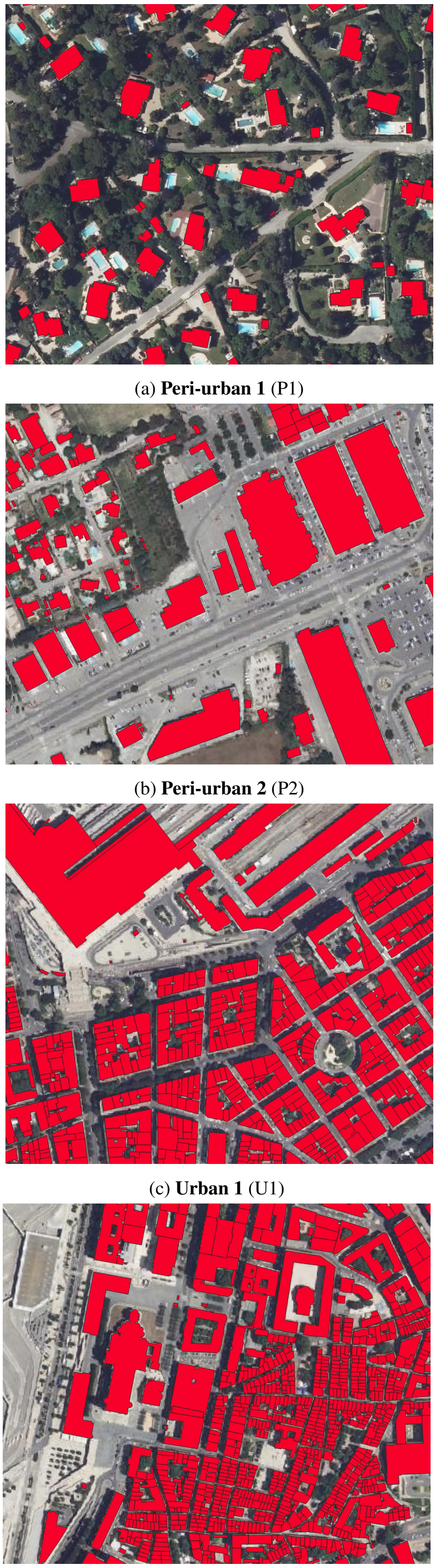
The following parameters were used for this current benchmark :

- Maximum gap : 1 meter
- Planarity tolerance : 0.100 meter
- Increase tolerance : 0.100 meter
- Minimum area : 15 m²
- Minimum detail : 3 m²

2.3.2 City3D is a fully automatic solution for the reconstruction of 3D buildings models developed by the 3D Geoinformation Research Group of the Delft University of Technology (Huang et al., 2022). The solution is based on the same approach as Polyfit (Nan and Wonka, 2017), an earlier automatic reconstruction solution that could not yet handle missing vertical planes in the point cloud. City3D takes large-scale aerial lidar point-clouds (that often miss these vertical planes) and 2D footprints as input, and returns a polygonal model.

As Polyfit, City3D is based on the technique of the 3D plane arrangement, it proceeds in three main steps:

[2] https://sites.google.com/view/lod2-building-benchmark



(a) **Peri-urban 1** (P1)

(b) **Peri-urban 2** (P2)

(c) **Urban 1** (U1)

(d) **Urban 2** (U2)

Figure 3. Input footprints for the benchmarked methods

1. The first step is to detect all the planar segments of the point cloud, and to compute all the planes to which the planar segments belong.
2. Then all the possible intersections between the planes are computed, and using these intersections, the planes are cut into several cells.
3. Finally, a binary linear programming formulation is used with an energy term that computes how much a cell is likely to belong to the building, this optimizes under a set of hard constraints which cells belong to the building an which cells do not, giving at the end a polygonal model.

The global energy term used in City3D is the weighted sum of four energy terms :

- The **data-fitting** term measures the fitting of the faces with the input point cloud.
- The **model complexity** term is used to control the compactness of the model and to avoid noise and outliers when input data are incomplete.
- The **point coverage** is used when there are missing data due to occlusion, in order to keep the faces resulting from missing areas as small as possible.
- The **roof preference** term is used when a building has several layers of roof, and gives the priority to the highest layer of roof.

In addition to these four energy terms, two hard constraints are added to enhance the topological correctness of the result:

- The result must have one single layer of roof.
- In cases where there is an ambiguity between two candidates faces nearly coplanar and close to each other, the priority is given to the face with the highest number of supporting points.

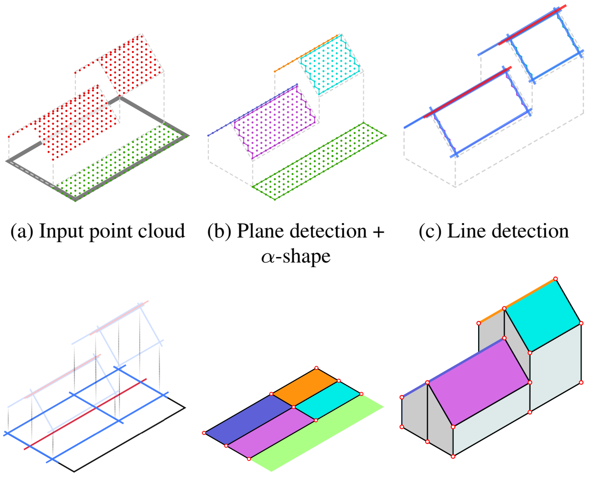
In practice, some post treatments had to be implemented to ensure that the resulting building models are watertight and well oriented.

2.3.3 Roofer (sometimes referred to as Geoflow) performs fully automatic building reconstruction from a point cloud containing at least the terrain and building classes and a building footprint polygon (see Figure 4a). It was developed at the Technical University of Delft and is used to create the nationwide Dutch 3DBAG dataset, containing over 10 million LoD 2 building models (Peters et al., 2022).

First, roofplanes are detected using a **region-growing algorithm** (see Figure 4b), and the roofplane boundaries are computed using the α**-shape algorithm** (see Figure 4b). Rooflines are then detected by intersecting the roofplanes and fitting lines to the α-shape boundaries (see Figure 4c).

Then the rooflines are clustered by similarity in orientation and proximity. A **representative line** is then selected for each cluster (see Figure 4d).

These lines, together with the footprint boundaries, are projected to create a 2D arrangement (a 2D polygonal partition)



(a) Input point cloud (b) Plane detection + α-shape (c) Line detection

(d) 2D projection + line regularization (e) Optimization 2D planar partition (f) Extrusion

Figure 4. Main steps in the roofer algorithm. Notice that the green terrain patch inside the footprint (a) is automatically detected (b) and clipped from the final result (f).

modeled using a DCEL data-structure. This arrangement is then optimized using a **graph-cut algorithm**. This labels the arrangement faces with the earlier detected roofplanes in such a way that an optimal balance is achieved between the compactness of the arrangement and the accuracy with respect to the input point cloud, using the energy term formulation proposed by (Zebedin et al., 2008). When adjacent faces are labeled with the same roofplane, they are merged. The result is a new arrangement typically with fewer and more regular faces, as illustrated in Figure 4e. The user can tune the *complexity-factor* parameter to steer the optimization balance either towards a more compact or a more geometrically accurate result (Peđen et al., 2024).

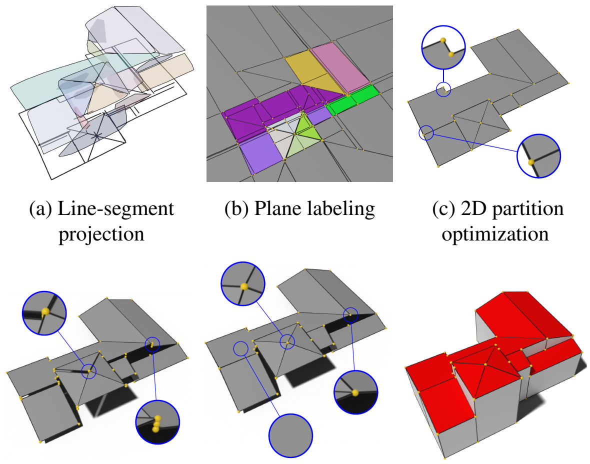
Finally, the 2D arrangement is extruded to create a **3D building model**, as shown in Figure 4f.

Like the other approaches in this benchmark, this approach is largely data-driven: its accuracy and robustness depend on the **quality and classification** of the input point cloud and the **correct alignment** with building footprints. Compared to City3D, Roofer is computationally more efficient, primarily because 1) it performs it's optimization on a 2D arrangement instead of a more complex 3D arrangement and 2) it uses graph-cut optimization instead of linear programming.

Roofer is open source and freely available via GitHub[3]. For this benchmark Roofer version 1.0.0 was used with the default parameters.

2.3.4 SimpliCity (Bauchet et al., 2024) is a tool developed by LuxCarta's R&D team for reconstructing LoD2 3D building models from LiDAR point clouds. A key advantage of our method is its approach to regularization while preserving the fidelity to the inputs. To avoid connectivity approximations and topological errors common in direct 3D primitive assembly, SimpliCity first simplifies and regularizes the building layout in a 2D polygonal partition before extruding it to a 3D polyhedron. This strategy ensures the planarity of the roof sections while preserving vertical discontinuities and horizontal roof edges.

[3] https://github.com/3DBAG/roofer



(a) Line-segment projection (b) Plane labeling (c) 2D partition optimization

(d) Cell extrusion (e) 3D mesh optimization (f) Extrusion

Figure 5. Main steps of the SimpliCity algorithm. (5a) Planar primitive contours and the building footprint are projected into a 2D plane. (5b) A kinetic simulation generates a dense polygonal partition. (5c) A global optimization procedure enhances the partition's regularity by enforcing geometric constraints. (5d, 5e, 5f) 2D partition is extruded, and a final 3D optimization ensures compactness and facet planarity.

The result is a scalable and efficient process that produces **lightweight, regularized, and compact** meshes that maintain fidelity to the input data. The generated models are watertight, 2-manifold, intersection-free, and conform to the CityGML LoD2 standard.

The core pipeline is demonstrated in Figure 5. First, line segments from the intersection of adjacent roof planes and their boundary are projected into the horizontal plane, together with the building's footprint line-segments (see Fig. 5a). A dense 2D polygonal partition is then built from these segments using a kinetic simulation (Fig. 5b) (Bauchet and Lafarge, 2018, Bauchet and Lafarge, 2020). Labels corresponding to detected roof planes or the ground are assigned to each partition cell. This assignment is formulated as an energy minimization problem to find the optimal label configuration by balancing three components: fidelity to the closest plane primitive, continuity between adjacent cells, and simplicity of edges. The partition is then further simplified and regularized via a global optimization that adjusts near-parallel and near-orthogonal edges (Fig. 5c). Each cell of the regularized 2D partition is then extruded according to its associated roof plane (Fig. 5d). Finally, another global optimization (Fig. 5e) fine-tunes the vertex z-coordinates to enforce strict facet planarity. This process merges nearly co-planar segments to enforce parallelism and orthogonality, yielding the final, clean 3D model (Fig. 5f).

In this benchmark, two SimpliCity results are evaluated :

1. Result A (SC-A) : Fully automated reconstruction with default parameters from the initial footprints.

2. Result B (SC-B) : Fully automated reconstruction with default parameters by taking an additional manual footprint enrichment. Important superstructure polygons are added to obtain better reconstruction regularity for small structures. SimpliCity leverages its kinetic partitioning scheme and regularization processes, giving it the flexibility to

incorporate supplementary geometric information for enhanced reconstruction quality.

2.4 Evaluation metrics

To evaluate the production from those different methods, we based our benchmark on an open-source Python tool named PyScoring[4], accessible in PyPI and as a GitHub repository. At first, PyScoring was a tool made to compare a dataset of 2D vectors (e.g. building footprints) with a reference (ground truth) by computing distance metrics between polygons or polylines. For this benchmark's purpose, we developed an extension of PyScoring to add geometric and topological 3D metrics.

2.4.1 Topological metrics The purpose of the topological metrics is to evaluate the segmentation of the roofs in our production dataset. With those metrics, we should be able to know if our method recovers the roof structure present in our ground truth dataset or if it is producing too much or not enough roof sections and if it is missing some roofs or making up some. The first task of computing topological metrics consists in pairing roof polygons from the ground truth reference (P_{GT}) and result (P_{res}) to be evaluated. This is done with an intersection over minimum area criterion:

$$\frac{|P_{GT} \cap P_{res}|}{min(|P_{GT}|, |P_{res}|)} > 0.5 \tag{1}$$

This criteria ensures no N to M matching but multiple polygons on one side can be matched to a single on the other, which we call **under segmentation** (when one P_{res} is matched with multiple P_{GT}) and **over segmentation** (when multiple P_{res} are matched with one P_{GT}). The topological metrics are simply the averages **Useg** and **Oseg** of these two numbers over all pairings defined by 1. The ideal algorithm producing the ground truth has **Useg** = **Oseg** =1.

2.4.2 Geometric metrics Geometric metrics aim at characterizing the geometric fidelity of the produced roof, mainly in terms of distance to the reference. In order for geometric metrics to be surfacic (the weight in the metric is proportional to the area), we rasterize the results with RasterIO at 0.2m ground resolution to compare then evaluate the precision of the Z coordinate in the roof sections of this raster grid. The proposed metrics are the mean of the error in z (Avg) and Root Mean Squared Error ($RMSE$) evaluated at each pixel of the raster.

Note that the geometric quality of the 2D subdivision of the footprint in polygonal roof segments is not evaluated directly. However, it participates to both topological metrics (as incorrect subdivision increase probability of over/under-segmentation) and to geometric metrics (as a wrong 2D partition will imply more errors in elevation). Moreover, only inner segmentation boundaries can differ between the methods as the outer boundary is an input of all evaluated methods.

3. Results

3.1 Qualitative results

Figures 6 and 7 show some results over a subset of the peri-urban area P1 with some relatively simple houses, and on a cathedral of U2 with very complex roof shapes, the two extremes of roof complexity. On the peri-urban area, the results are pretty similar between the evaluated methods, with City3D displaying some more complexity, in particular producing more vertical parts inside the footprint. On the cathedral, only Roofer and SimpliCity-B are able to recover the round towers correctly, and only Roofer recovers the small tower at the center of the central structure.

[4] https://github.com/IGNF/pyscoring

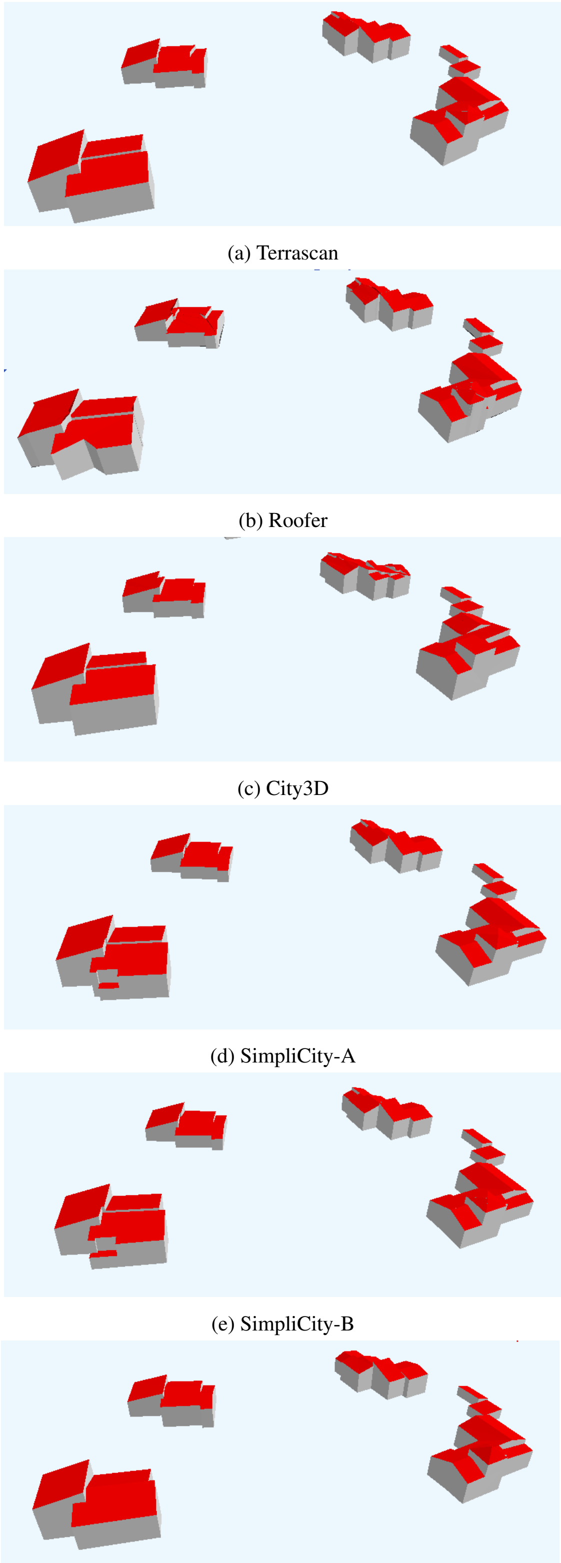
Area	Method	Intrinsics		Topology		Geometry	
		k GT	k R	USeg	OSeg	Avg	$RMSE$
P1	TScan	3.3	1.1	2.81	1.01	0.50	0.79
P2	TScan	3.7	2.9	1.64	1.39	0.15	0.70
U1	TScan	13.8	7.1	2.15	1.12	0.06	1.63
U2	TScan	20.4	7.5	2.96	1.10	0.85	3.55
All	Tscan	41.1	18.6	2.39	1.15	0.39	1.67
P1	Roofer	3.3	1.0	3.04	1.01	-0.48	0.73
P2	Roofer	3.7	1.7	2.12	1.11	-0.12	0.69
U1	Roofer	13.8	6.7	2.2	1.11	-0.00	1.60
U2	Roofer	20.4	6.8	3.2	1.09	0.53	2.94
All	Roofer	41.2	16.1	2.64	**1.08**	-0.28	**1.49**
P1	City3D	3.3	2.0	2.16	1.30	0.84	11.62
P2	City3D	3.7	x	x	x	x	x
U1	City3D	13.8	14.0	1.60	1.66	0.42	14.90
U2	City3D	20.4	10.9	1.98	1.46	0.58	2.43
All	City3D	41.2	29.0	**1.91**	1.47	0.61	9.65
P1	SC-A	3.3	1.2	2.58	1.01	-0.46	0.76
P2	SC-A	3.7	2.8	1.74	1.28	-0.09	0.95
U1	SC-A	13.8	7.5	2.07	1.13	0.03	1.57
U2	SC-A	20.4	6.2	3.51	1.06	-0.49	3.10
All	SC-A	41.2	17.8	2.48	1.12	**-0.25**	1.59
P1	SC-B	3.3	1.2	2.79	1.01	-0.47	0.74
P2	SC-B	3.7	2.7	1.73	1.37	-0.16	0.74
U1	SC-B	13.8	9.1	1.90	1.26	0.02	1.60
U2	SC-B	20.4	8.5	2.83	1.18	-0.51	2.88
All	SC-B	41.2	21.5	2.32	1.21	-0.28	**1.49**

Table 1. Quantitative evaluation of the compared approaches through topological and geometric metrics. k GT/R are the number of thousand polygons composing the roofs of the ground truth and evaluated result

3.2 Quantitative results and discussion

Table 1 provides the topological and geometric metrics for all evaluated methods and all areas, with a mean over all areas. We also give two intrinsic metrics (not requiring comparison with the ground truth): the number of roof structures in the ground truth and evaluated result. We see that City3D has the smallest under segmentation but at the cost of highest over segmentation, meaning it is the one that produces the finest grained structures. In practice, all methods aim at creating clean, regular, and simple roof structures. In doing so, they might merge two distinct but small roof elements from the ground truth into one, explaining the relatively higher under-segmentation (Useg) score. This is particularly visible with Roofer which has the smallest over segmentation but at the cost of highest under segmentation, showing that it generalizes the structures more. A limit of this benchmark is that the evaluated methods were probably tuned for a higher level of generalization than the high quality human produced ground truth where operators might have inferred details barely visible in the lidar (because they also had access to an orthoimage that helped interpretation). This explains the overall high under-segmentation (Useg) metric in most cases. However, because all evaluated methods have rather similar under-segmentation, we believe that this weighs the same on each.

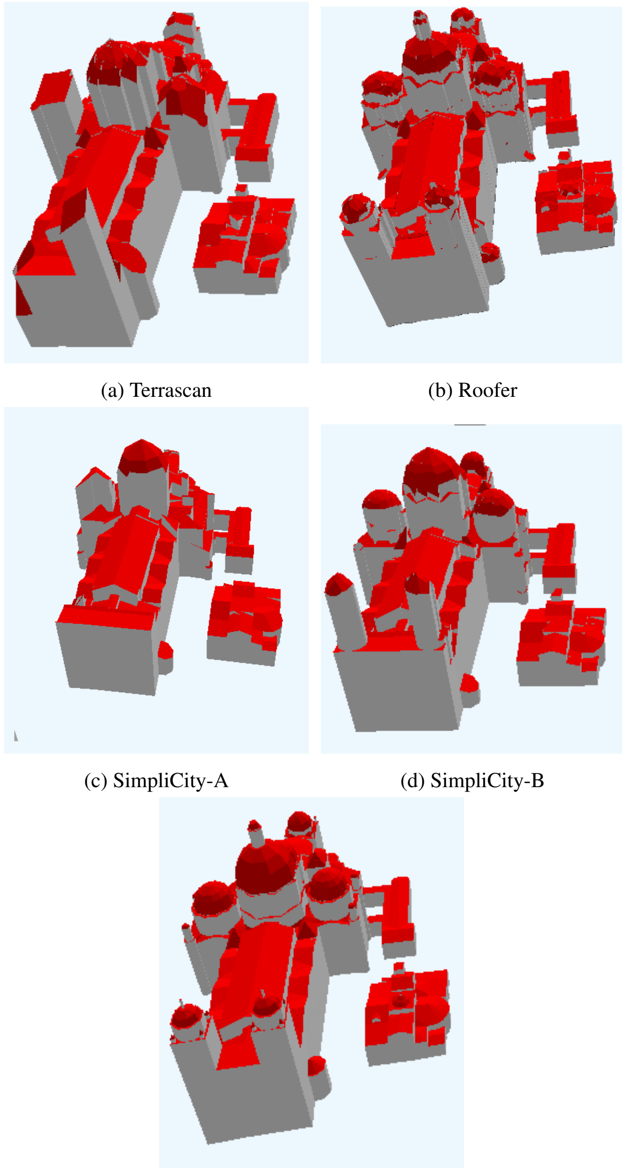
In terms of geometry, the results are quite similar (with a top



(a) Terrascan (b) Roofer (c) City3D (d) SimpliCity-A (e) SimpliCity-B (f) Ground truth

Figure 6. Results of the evaluated methods on a peri-urban area P1.

$RMSE$ score for Roofer and SimpliCity-B), apart from City3D which has a high error. The explanation is that in some difficult cases City3D produced a result with very high error that affects the mean significantly. It even sometimes failed completely on P2, which is why we do not provide metrics as removing the difficult cases on which it failed from the metrics computation would unfairly advantage it. This means that it can be considered less robust than the others, so probably less suited for a large scale production scenario. The metrics also does not measure regularity (face/edge horizontality/verticality, symmetry, parallelism, orthogonality), but when such regularity is present in the real world, operators have ensured that the ground truth has them, such that a method enforcing valid regularity priors should perform better than not.



(a) Terrascan (b) Roofer (c) SimpliCity-A (d) SimpliCity-B (e) Ground truth

Figure 7. Results of the evaluated methods on the cathedral of U2.

4. Conclusion and future works

4.1 Participating

This benchmark is open to future participation. Anyone developing a LoD2 building reconstruction method from aerial lidar can participate to the benchmark. All the input data over the 4 areas can be downloaded from the benchmark website [5], and results should simply be sent to lod2bench@ign.fr as 4 CityJSON files. The submitted results will then be evaluated using PyScoring and the resulting metrics will be added to a more complete version of Table 1 displayed on the website.

4.2 Conclusion

In this paper, we have presented a benchmark on LoD2 building 3D reconstruction from aerial lidar data and footprints, focusing on robust methods able to produce convincing results at city scale and cope with a large diversity of building types. Among the industrial tools and open source tools available, we have selected and evaluate 4 tools: TerraScan, Roofer, SimpliCity and City3D. While the first 3 give rather similar results with some minor differences analyzed in Section 3, City3D displays less robustness even if success cases usually give higher quality results. We believe that future methods should aim at enhancing reconstruction quality while being robust and general enough to cope with extremely diverse inputs.

4.3 Future works

In the future, we plan on keeping this benchmark open at least until the end of 2026 and to evaluate any result submitted to us through the simple procedure described on our web site. The results of this benchmark will be used in the context of the construction of a national digital twin over the whole French territory, which will be coordinated by IGN France. Other perspectives include benchmarking methods that will allow to handle the life-cycle of 3D city models:

1. building change detection methods such as (Huang et al., 2013)
2. qualification methods that aim at providing trustworthy quality measures of produced buildings or classification of detected errors, such as (Michelin et al., 2013)
3. interactive building model edition/correction such as (Geniet et al., 2024)

References

Altman, S., Xiao, W., Grayson, B., 2017. Evaluation of Low-Cost Terrestrial Photogrammetry for 3D Reconstruction of Complex Buildings. *ISPRS Annals IV-2/W4*, 199-201.

Bauchet, J.-P., Lafarge, F., 2018. KIPPI: KInetic Polygonal Partitioning of Images. *CVPR*, 3146-3154.

Bauchet, J.-P., Lafarge, F., 2020. Kinetic Shape Reconstruction. *ACM Trans. Graph.*, 39(5). https://doi.org/10.1145/3376918.

Bauchet, J.-P., Sulzer, R., Lafarge, F., Tarabalka, Y., 2024. Simplicity: Reconstructing buildings with simple regularized 3d models.

Buyukdemircioglu, M., Kocaman, S., Kada, M., 2022. Deep Learning for 3D Building Reconstruction: A Review. *ISPRS Archives, Volume XLIII-B2*, 359.

Gao, W., Peters, R., Stoter, J., 2024. Building-PCC: Building Point Cloud Completion Benchmarks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W5, 179.

[5] https://sites.google.com/view/lod2-building-benchmark

Geniet, F., Brédif, M., Vallet, B., 2024. Editing Watertight Manifold Polyhedra using Face Shifts with Automatic Topological Updates and Edge Flips. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W8-2024, 169–176. https://isprs-archives.copernicus.org/articles/XLVIII-2-W8-2024/169/2024/.

Huang, J., Stoter, J., Peters, R., Nan, L., 2022. City3D: Large-Scale Building Reconstruction from Airborne LiDAR Point Clouds. *Remote Sensing*, 14(9), 2254. doi.org/10.3390/rs14092254. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Huang, X., Zhang, L., Zhu, T., 2013. Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1), 105–115.

Ledoux, H., Ohori, K. A., Kumar, K., Dukai, B., Labetski, A., Vitalis, S., 2019. CityJSON: a compact and easy-to-use encoding of the CityGML data model. *Open Geospatial Data, Software and Standards*, 4(4).

Michelin, J., Tierny, J., Tupin, F., Mallet, C., Paparoditis, N., 2013. Quality evaluation of 3D city building models with automatic error diagnosis. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7/W2, 161–166. https://doi.org/10.5194/isprsarchives-XL-7-W2-161-2013.

Nan, L., Wonka, P., 2017. PolyFit: Polygonal Surface Reconstruction from Point Clouds. *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Venice, 2372–2380.

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

Peters, R., Dukai, B., Vitalis, S., van Liempt, J., Stoter, J., 2022. Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the Netherlands. *Photogrammetric Engineering and Remote Sensing*, 88(3), 165–170.

Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U., 2012. The ISPRS Benchmark on Urban Object Classification and 3D Building Reconstruction. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-3, 293–298.

Wang, R., Huang, S., Yang, H., 2023. Building3d: An urban-scale dataset and benchmarks for learning roof structures from point clouds. *International Conference on Computer Vision (ICCV)*. arXiv:2307.11914.

Zebedin, L., Bauer, J., Karner, K., Bischof, H., 2008. Fusion of feature-and area-based information for urban buildings modeling from aerial imagery. *Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part IV 10*, Springer, 873–886.

Özdemir, E., Toschi, I., Remondino, F., 2019. A multi-purpose benchmark for photogrammetric urban 3d reconstruction in a controlled environment. *ISPRS Archives XLII-1/W2*, 53.