

# Seven hundred thousand roofs: the built footprint of Gabon as detected from space: where the roofs are, how large, how sure, and what a confidence threshold removes

Hodologic Research

Hodologic, <https://hodologic.fr>

Preprint · 4 October 2026

## Abstract

Google Open Buildings v3 is the first inventory of buildings that covers the whole of Gabon. A model reads satellite imagery and draws a contour around every roof it recognises, with its ground area and a confidence score. This paper describes what that dataset contains for Gabon, so that whoever uses it knows what they are counting. We cut it to the country, 738 615 footprints kept out of 9 480 086 read, 29 refused and accounted for, and answer four questions: where the roofs are, how large they are, how sure the model is of them, and what a confidence threshold removes.

*The roofs are concentrated to an extreme degree.* They occupy a twentieth of the territory, and within it 10 % of the occupied cells hold 89 % of the roofs. The Libreville agglomeration alone holds 55 %; the ten largest built patches hold 85 %; fifteen towns hold 75 %.

*The roofs are small.* The median footprint is 71 m<sup>2</sup>; 14 % are under 20 m<sup>2</sup>, a size at which the model may be seeing a shed or a shadow rather than a dwelling. All the roofs together cover 76 km<sup>2</sup>.

*Confidence is lowest where roofs are isolated and where they are packed.* Nationally, 29 % of footprints score below 0.75. In the scattered rural cells, 54 % do. In the densest urban blocks, 36 % do in Libreville and 44 % in Port-Gentil. Everywhere in between, 27 to 29 %.

*A confidence threshold therefore removes the hamlets and the informal quarters first, not noise at random.* A cut at 0.75 keeps 71 % of the country's roofs but 46 % of the scattered ones; a cut at 0.80 keeps 54 %. Whoever applies a threshold should publish, next to it, what it removes in those two fabrics.

**Keywords:** building footprints; Open Buildings; settlement; confidence; Gabon.

## 1 Introduction

### 1.1 A building inventory for the whole country

Gabon had no inventory of its buildings that covered the whole country with one method. Google Open Buildings [1] provides one: a deep-learning model detects footprints on high-resolution satellite imagery across Africa, and its third release covers Gabon entirely, with a contour, a ground area and a confidence score for each footprint.

A detected inventory is not a surveyed one. It is an estimate with a score, and the score is the only thing that separates a roof from a shadow. This paper describes the dataset for Gabon as a user of it must know it: what it contains, where its footprints are and how they cluster, how large they are, what the confidence score is correlated with, and what a threshold on it does to the picture of the country.

### 1.2 Questions and contributions

We ask (Q1) where the detected roofs are and how concentrated the built footprint is; (Q2) how large they are and what the size distribution says about what is being detected; (Q3) what the confidence score

is correlated with, and what a threshold removes and where.

The paper contributes (i) a documented cut of Open Buildings v3 to Gabon, with every refused footprint accounted for (Section 3.1); (ii) a description of the built footprint at three grains, cells, patches and towns, which shows how far the built footprint is concentrated (Sections 4.1, 4.2, 4.4); and (iii) a measurement of the confidence score against local density at two scales, country and city block, which shows that confidence is lowest at both extremes (Sections 4.6, 4.7), with the consequence for any threshold stated class by class (Section 4.8).

## 2 Data

Google Open Buildings v3 [1, 2] distributes footprints as CSV files by S2 cell, each row carrying a WKT polygon, its area in square metres and a confidence score, the model's probability that the polygon is a building. Four cells cover Gabon (107, 109, 1a7, 1a9), 2.37 GB and 9 480 086 footprints, most of them in neighbouring countries. The imagery date is not published per building. The data are licensed CC BY 4.0. The nine provinces come from the administrative layer used by the GIA platform; a footprint belongs to the province that contains its centroid. The file used, its date and SHA-256 checksum are given in Appendix A.

## 3 Method

### 3.1 Cutting the dataset to the country

The four tiles are read line by line. A footprint is kept if its centroid falls inside a province; its contour is taken from the WKT and rounded to six decimals (eleven centimetres), which shortens the file by a third without moving a wall. Every line read is counted as kept, outside the country, or refused with a reason, and the producer refuses to write if the three do not sum to the lines read: 9 480 086 read, 8 741 442 outside Gabon, 738 615 kept, 29 refused (28 multi-ring geometries, 1 malformed). The refusals matter more than their number: a multi-ring geometry read as a simple polygon would produce a wrong contour without any signal. The cut was repeated independently by two people with two methods; the nine provincial counts coincide, and the difference of 29 between the two totals is exactly the refusals.

### 3.2 Cells, classes and patches

Footprints are counted in cells of  $0.01^\circ$  (about 1.1 km) by their centroid, taken as the mean of the contour's vertices; the cell is the floor of the coordinates divided by 0.01, and density bins include their lower bound (a block of exactly 60 footprints is in the densest bin). Cells are grouped into five settlement classes by their count: *scattered* (1 to 9 footprints per cell), *hamlet* (10 to 49), *village* (50 to 199), *town* (200 to 999) and *city* (1 000 and more). Occupied cells that touch, including diagonally, form *built patches*; a patch is the smallest unit one could call a settlement at this grain. Towns are also described within a window of  $\pm 0.1^\circ$  (about 22 km) around the place point of fifteen main towns.

### 3.3 Confidence against density

To see what the confidence score is correlated with, we measured it against local density at two scales: nationally, in the 1.1 km cells, binned by their count; and in five cities (Libreville, Port-Gentil, Franceville, Oyem, Lambaréné), in cells of  $0.001^\circ$  (about 110 m, one hectare), binned the same way. For each bin we report the mean confidence, the share of footprints below 0.75 and the median footprint area.

**Table 1:** Footprints by settlement class of their 1.1 km cell.

| Class (footprints per cell) | Cells | Footprints | Share  | Built area           | Mean footprint     | Mean conf. | Below 0.75 |
|-----------------------------|-------|------------|--------|----------------------|--------------------|------------|------------|
| Scattered (1–9)             | 4 994 | 11 631     | 1.6 %  | 0.6 km <sup>2</sup>  | 55 m <sup>2</sup>  | 0.755      | 54 %       |
| Hamlet (10–49)              | 1 332 | 31 002     | 4.2 %  | 2.6 km <sup>2</sup>  | 84 m <sup>2</sup>  | 0.800      | 29 %       |
| Village (50–199)            | 706   | 67 238     | 9.1 %  | 6.3 km <sup>2</sup>  | 94 m <sup>2</sup>  | 0.807      | 27 %       |
| Town (200–999)              | 349   | 164 344    | 22.3 % | 18.2 km <sup>2</sup> | 111 m <sup>2</sup> | 0.811      | 27 %       |
| City (≥ 1 000)              | 215   | 464 400    | 62.9 % | 48.6 km <sup>2</sup> | 105 m <sup>2</sup> | 0.803      | 29 %       |

### 3.4 Reproducibility

The cut is the platform’s `extraire_bati_detecte.py`. The statistics are three Python scripts reading the cut file, the charts a fourth; the maps are rendered with Leaflet on Esri basemaps, World Light Gray Canvas for the country and World Imagery for the footprint windows. Everything runs in a few minutes on a laptop.

## 4 Results

### 4.1 Where the roofs are

The 738 615 footprints occupy 7 596 cells of 1.1 km, a twentieth of the country’s area (Figure 1). Within that twentieth the concentration is extreme: 1 % of the occupied cells hold 34 % of the footprints, 5 % hold 78 %, 10 % hold 89 %. Outside the cities the picture is the one the national roads draw: strings of villages along the RN1, RN2 and RN3, the forestry towns of the Haut-Ogooué and the Ogooué-Lolo, and little else. Large parts of the interior hold no detected roof at all, which is a statement about the country and about what the model can see under forest, in a proportion the data cannot separate.

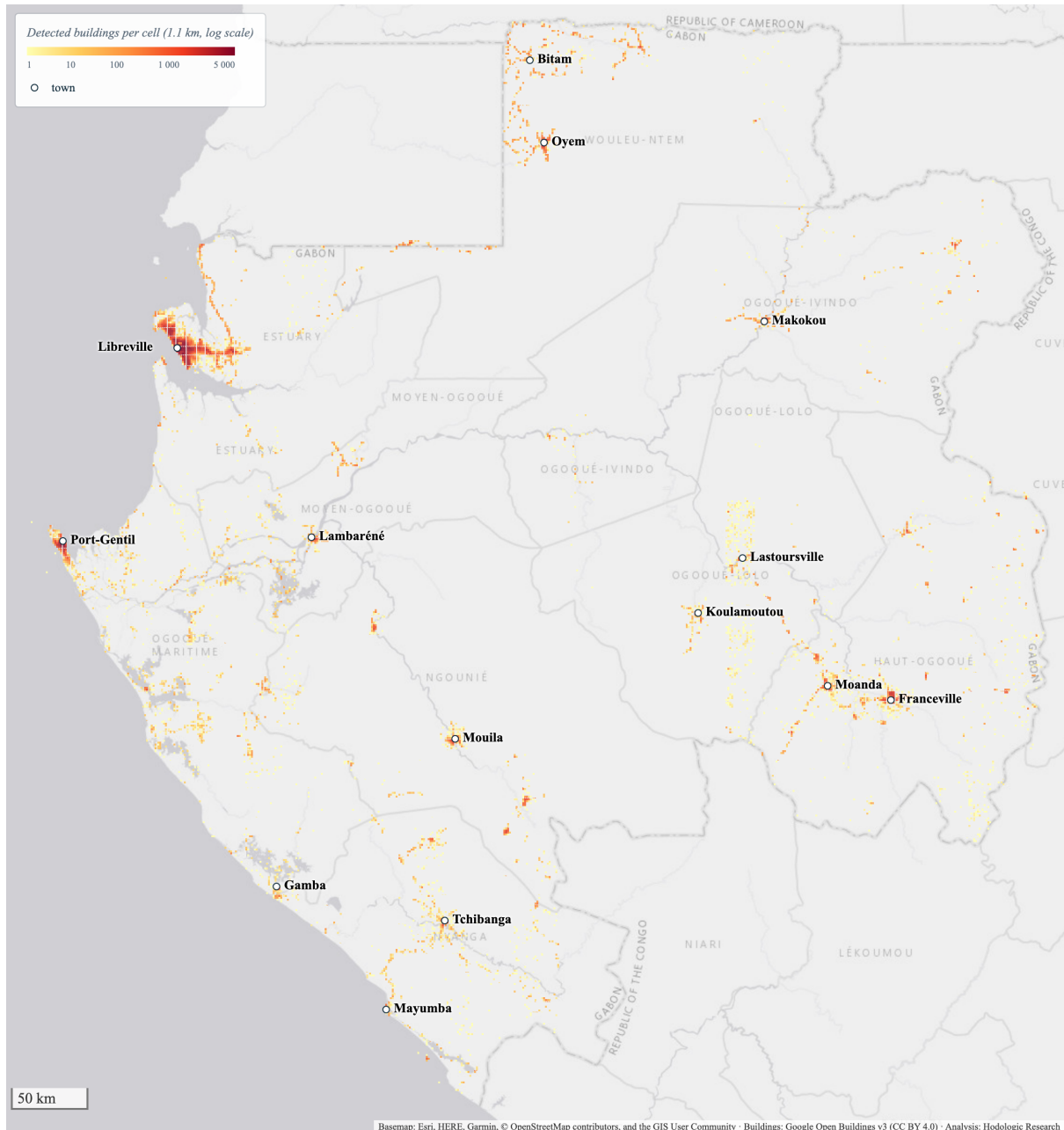
Grouped into patches of touching cells, the country has 1 954 built patches. The largest, the Libreville agglomeration from Cap Estérias to Ntoun and Owendo, spans 771 cells and holds 405 005 footprints, 55 % of the national stock, on 44 km<sup>2</sup> of roofs. The second is Port-Gentil (82 624 footprints), then the Moanda–Franceville basin and the towns of the Woleu-Ntem. The ten largest patches hold 85 % of all footprints and the hundred largest 97 %; at the other end, 1 209 patches are a single cell and 1 689 hold fewer than fifty footprints. The settlement structure of the country, as seen from space, is one metropolis, a handful of towns, and a long tail of hamlets.

### 4.2 Five settlement classes

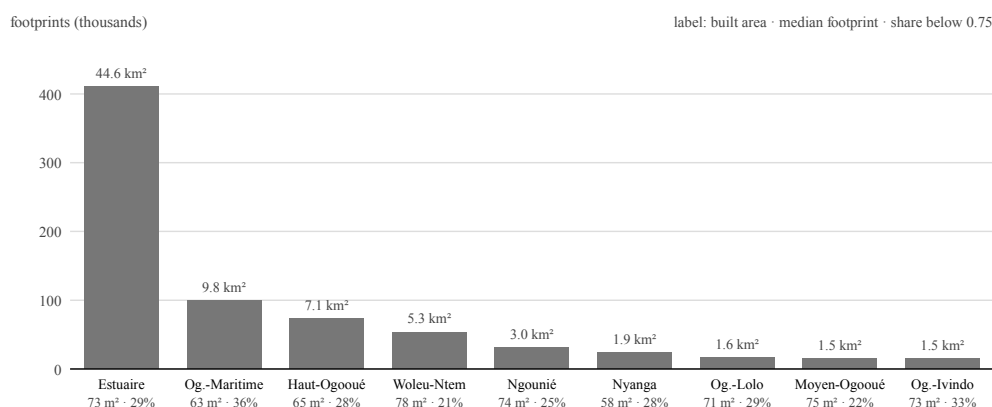
Two thirds of the cells are in the scattered class and hold 1.6 % of the footprints; 215 city cells hold 63 % (Table 1). The classes differ in more than count. The mean footprint grows with density up to the town class (55 m<sup>2</sup> scattered, 111 m<sup>2</sup> in towns) and eases slightly in cities, where small roofs pack. Confidence is flat across hamlets, villages, towns and cities (0.80 to 0.81, with 27 to 29 % below 0.75) and collapses in the scattered class, where the mean is 0.755 and more than half the footprints fall below 0.75. A lone roof in a clearing is the hardest thing for the model to be sure of, and the scattered class is where a rural accessibility analysis looks for the people farthest from everything.

### 4.3 Provinces

The Estuaire holds 410 540 footprints, 56 % of the country, on 44.6 km<sup>2</sup> of roofs (Figure 2). The Ogooué-Maritime (98 939) and the Haut-Ogooué (73 446) follow, then the Woleu-Ntem (53 787) and the Ngounié (31 294); the four remaining provinces hold between 15 000 and 24 000 each. Median footprints range



**Figure 1:** Detected footprints per 1.1 km cell, log scale. The strings of villages along the national roads are visible, as are the three agglomerations that dominate the count: Libreville and its estuary, Port-Gentil, and the Moanda–Franceville basin. Basemap: Esri World Light Gray Canvas.



**Figure 2:** Detected footprints by province; above each bar the built area, below it the median footprint and the share below 0.75.

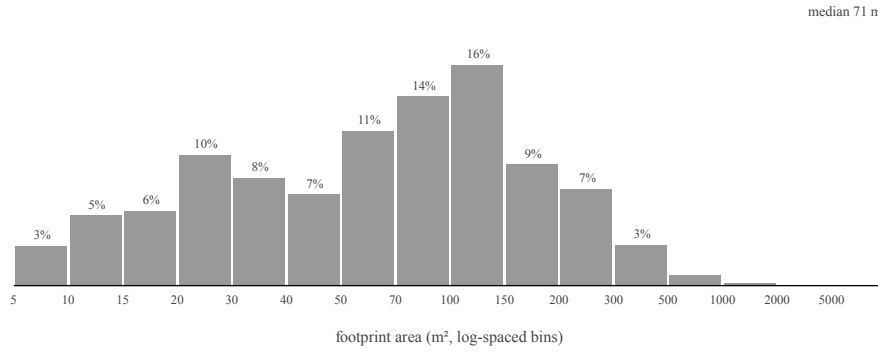
**Table 2:** Footprints within 11 km of the place point of fifteen towns ( $0.2^\circ \times 0.2^\circ$  windows, which may overlap for neighbouring towns). Together the fifteen windows hold 75 % of the national stock.

| Town          | Footprints | Share of country | Built area           | Median footprint  | Below 0.75 |
|---------------|------------|------------------|----------------------|-------------------|------------|
| Libreville    | 308 800    | 41.8 %           | 33.1 km <sup>2</sup> | 74 m <sup>2</sup> | 29 %       |
| Port-Gentil   | 81 467     | 11.0 %           | 8.5 km <sup>2</sup>  | 69 m <sup>2</sup> | 36 %       |
| Franceville   | 34 055     | 4.6 %            | 3.2 km <sup>2</sup>  | 65 m <sup>2</sup> | 28 %       |
| Oyem          | 27 496     | 3.7 %            | 2.8 km <sup>2</sup>  | 79 m <sup>2</sup> | 22 %       |
| Moanda        | 19 201     | 2.6 %            | 2.0 km <sup>2</sup>  | 76 m <sup>2</sup> | 27 %       |
| Mouila        | 18 670     | 2.5 %            | 1.8 km <sup>2</sup>  | 78 m <sup>2</sup> | 24 %       |
| Tchibanga     | 14 172     | 1.9 %            | 1.2 km <sup>2</sup>  | 70 m <sup>2</sup> | 23 %       |
| Lambaréné     | 11 953     | 1.6 %            | 1.3 km <sup>2</sup>  | 80 m <sup>2</sup> | 19 %       |
| Bitam         | 8 652      | 1.2 %            | 0.9 km <sup>2</sup>  | 81 m <sup>2</sup> | 19 %       |
| Makokou       | 8 228      | 1.1 %            | 0.8 km <sup>2</sup>  | 85 m <sup>2</sup> | 35 %       |
| Koulamoutou   | 7 331      | 1.0 %            | 0.7 km <sup>2</sup>  | 80 m <sup>2</sup> | 24 %       |
| Gamba         | 5 740      | 0.8 %            | 0.6 km <sup>2</sup>  | 70 m <sup>2</sup> | 29 %       |
| Lastoursville | 5 456      | 0.7 %            | 0.6 km <sup>2</sup>  | 77 m <sup>2</sup> | 28 %       |
| Ndendé        | 3 142      | 0.4 %            | 0.3 km <sup>2</sup>  | 78 m <sup>2</sup> | 24 %       |
| Mayumba       | 1 478      | 0.2 %            | 0.1 km <sup>2</sup>  | 69 m <sup>2</sup> | 32 %       |

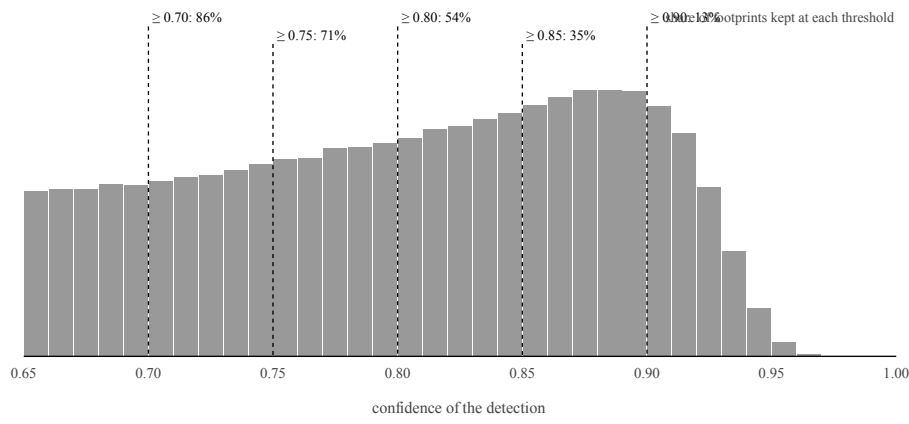
from 58 m<sup>2</sup> in the Nyanga to 78 m<sup>2</sup> in the Woleu-Ntem, and the share below 0.75 from 21 % in the Woleu-Ntem to 36 % in the Ogooué-Maritime, the province of Port-Gentil. Provincial medians of confidence are close (0.789 to 0.833); the score does not vary by province so much as by the kind of fabric a province is made of.

#### 4.4 Fifteen towns

Libreville's window alone holds 42 % of the country's footprints and Port-Gentil's 11 %; the two together, 53 % (Table 2). The table also separates two kinds of towns. The coastal and mining cities (Port-Gentil, Franceville, Makokou, Mayumba) have small median footprints and a high share of low-confidence detections; the towns of the north and the Ngounié (Oyem, Bitam, Lambaréné, Mouila) have larger footprints and the surest detections, 19 to 24 % below 0.75. The difference is one of fabric: compact informal quarters on the one hand, spaced village-type plots on the other.



**Figure 3:** Distribution of footprint area, log-spaced bins.



**Figure 4:** Distribution of the confidence score of the 738 615 footprints, and the share kept at each threshold.

#### 4.5 How large the roofs are

The median footprint is 70.8 m<sup>2</sup>; the first quartile is 31.8 m<sup>2</sup> and the third 132 m<sup>2</sup> (Figure 3). The distribution has two modes, one around 20–30 m<sup>2</sup> and one around 100–150 m<sup>2</sup>, and 13.7 % of footprints are under 20 m<sup>2</sup>, a size at which a model is as likely to be seeing a shed, a kiosk or a shadow as a dwelling. At the other end, 11 % exceed 200 m<sup>2</sup> and 1 % exceed 500 m<sup>2</sup>: warehouses, schools, markets, and the occasional merged block of touching roofs. Together the 738 615 footprints cover 76 km<sup>2</sup> of ground, 0.03 % of the country.

#### 4.6 The confidence score

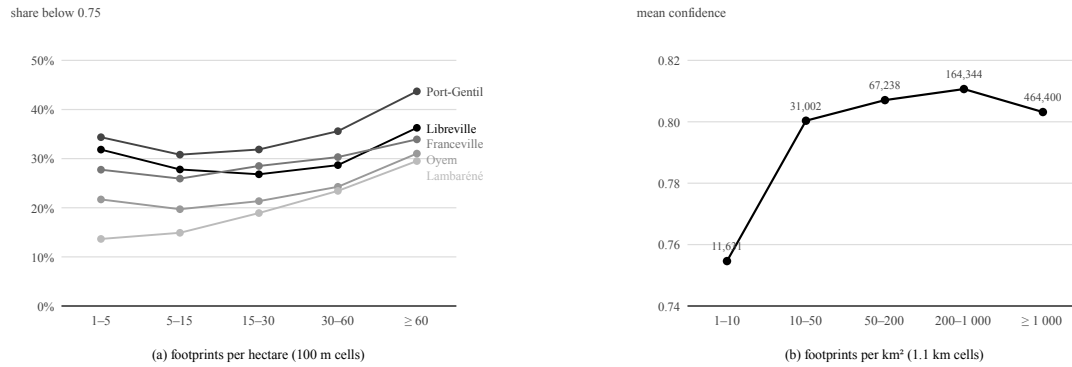
The score runs from 0.65, the floor of the distributed data, to 0.987, with a median of 0.811 and a long left shoulder (Figure 4). A threshold removes a large share of the set quickly: 0.70 keeps 86 % of footprints, 0.75 keeps 71 %, 0.80 keeps 54 %, 0.85 keeps 35 %, 0.90 keeps 13 %. What matters is not how much a threshold removes but what, and the next two sections show that it does not remove at random.

#### 4.7 Confidence is lowest at both ends of density

Two measurements at two scales give two different pictures, and both are true (Figure 6). Nationally, the least confident detections are the isolated ones: in the scattered class the mean confidence is 0.755 and 54 % of footprints fall below 0.75, against 0.80 to 0.81 and 27 to 29 % everywhere denser (Table 1).



**Figure 5:** Detected footprints outlined on satellite imagery and coloured by confidence, windows of about 1.1 km: (top) Libreville, Nombakélé-Akébé; (bottom, left to right) Port-Gentil, dense quarters; Franceville; a village in the Nyanga. In the packed quarters the low-confidence class dominates; in the spaced village plots nearly every roof is detected with high confidence. Imagery: Esri World Imagery (Esri, Maxar, Earthstar Geographics, and the GIS User Community).



**Figure 6:** (a) Share of footprints below 0.75 against block density (footprints per hectare) in five cities. (b) Mean confidence against cell density (footprints per 1.1 km cell) for the whole country; the labels give the number of footprints in each bin.

Inside cities the relation reverses. In each of the five cities measured, the share of footprints below 0.75 is lowest in blocks of 5 to 30 footprints per hectare, the ordinary residential fabric, and rises in the densest blocks: 36 % in Libreville and 44 % in Port-Gentil above 60 footprints per hectare, against 27 to 32 % in the middle range. The median footprint shrinks at the same time, from 77–82 m<sup>2</sup> to 62 m<sup>2</sup>. The densest blocks are the informal quarters, where small roofs touch and the model sees edges less clearly (Figure 5).

#### 4.8 What a threshold removes

The consequence for any threshold is direct, and it can be stated by class. A cut at 0.75 keeps 71 % of the national stock; it keeps 46 % of the scattered class, 71 to 73 % of hamlets, villages, towns and cities taken as wholes, and about 60 % of the densest urban blocks. A cut at 0.80 keeps 54 % nationally and far less than that in the two fabrics where confidence is lowest. The threshold therefore removes the hamlets and the informal quarters first. A threshold is therefore a decision about which settlements remain on the map, and it should be published as such, with the share it removes in the scattered class and in the densest blocks next to the share it removes overall.

### 5 What the data do not say

**Detected is not verified.** A footprint at 0.68 is the model's guess, and the confidence score is the only information that qualifies it. Every count in this paper carries its threshold and its bias; a total without them is not a measurement.

**The imagery has no date.** Google does not publish the acquisition date per building. A footprint may be a roof of 2018 or of 2023, and a settlement built or abandoned since will not be reflected. The imagery under Figure 5 is Esri's, not the imagery the model read; where an outline and the picture disagree, the cause may be a difference of date rather than an error of detection.

**A roof is not a household.** A hangar, a school and a house are one footprint each; 14 % of footprints are under 20 m<sup>2</sup> and may not be dwellings at all. Building density is an index of settlement, not a census.

**The forest is a blind spot.** Under canopy the satellite does not pass, and the empty interior of Figure 1 is partly a statement about what cannot be seen. The low confidence of the scattered class is the visible edge of that limit.

## 6 Use

For a producer of territorial indicators, three uses follow.

- *One inventory for the whole country.* The detected set covers the nine provinces with one method and one score, the second and third cities as well as the scattered rural settlement. Used with its confidence, it is the first uniform picture of where the roofs of Gabon are.
- *A threshold is a product decision.* Any consumer of the data should publish the threshold used and the share it removes in the scattered class and in the densest blocks. At 0.75, those shares are 54 % and 36 to 44 %.
- *A layer that changes with the zoom.* At the country scale a 10 m house is a hundredth of a pixel; the data are shown as density (Figure 1). Below about 1.5 km of map width, the contours themselves on the imagery, coloured by confidence, where each outline can be checked against the roof it claims (Figure 5). The two forms serve two uses, and a map product should offer both.

## 7 Relation to the GIA platform

GIA (<https://gabonassets.org>), operated by Hodologic, cut the detected set in September 2026 and uses it since 15 September 2026, with the method documented in its design notes. The figures in this paper are Hodologic Research’s own computation from the cut file; every figure that the platform serves is the platform’s to publish.

## Acknowledgements

Building data: Google Open Buildings v3 (CC BY 4.0). Basemap and imagery tiles are the intellectual property of Esri and its licensors, used under Esri’s terms for static maps in academic publications. The authors thank the GIA team for the design notes that document the cut and the switch.

## References

- [1] W. Sirko, S. Kashubin, M. Ritter, A. Annkah, Y. S. E. Bouchareb, Y. Dauphin, D. Keyzers, M. Neumann, M. Cissé, J. Quinn. Continental-scale building detection from high resolution satellite imagery. arXiv:2107.12283, 2021.
- [2] Google Research. Open Buildings v3, polygons dataset, CC BY 4.0. <https://sites.research.google/open-buildings/>

## A Input and scripts

| File                 | Date       | Size          | SHA-256         | Features |
|----------------------|------------|---------------|-----------------|----------|
| bati_detecte.geojson | 2026-09-15 | 205 361 283 B | 45527a6a...c7be | 738 615  |

`stats.py`, `stats2.py` and `blocks.py` compute the distributions, classes, patches, town windows and the confidence-against-density tables; `charts.py` draws the figures; the maps are rendered with Leaflet on Esri basemaps, World Light Gray Canvas for the country and World Imagery for the footprint windows. Scripts and outputs are available from the authors.