**CFI MAP-2 Demand side study.**

Sampling design methodological note

**Creating spatial grids for stage 1 sampling**

**Lagos, Nigeria**

To create a representative sample of small business establishments (those with 10 or fewer employees, including the owner) at the city-level, CFI uses a geospatial sampling approach that divides an area of interest into blocks of size 150m by 150m to establish a sampling frame of blocks.

**Step 1: Defining the sampling boundary**

The first step in this process is to define the sampling boundary or perimeter. In Nigeria, we use sub-national administrative boundaries from the United Nations Humanitarian Data exchange[[1]](#footnote-1) and the sampling boundary is constructed using a combination of level-2 administrative units that cover the target city. In Nigeria, level-1 administrative units correspond to states and level-2 administrative units correspond to local government areas (LGAs). For this study, the Lagos sampling boundary is defined using the five LGAs the comprise the Lagos division in Lagos state, namely: Lagos Mainland, Lagos Island, Apapa, Surulere and Eti-Osa (Figure 1).

|  |
| --- |
| **Figure 1:** Level 1 and Level 2 administrative boundaries |
|  |
| Source: UN Humanitarian Data Exchange (HDX) |

Figures 2 and 3 display the study boundary overlaid over remote-sensing satellite data.

|  |
| --- |
| **Figure 2:** True-color day-time satellite composite |
| A map of a city  Description automatically generated |
| Source: Cloud-free mosaic of Sentinel-2 imagery via Sentinel-hub |

|  |
| --- |
| **Figure 3:** Night-time lights (NTL) |
| A pixelated image of a train  Description automatically generated |
| Source: NASA VIIRS (lack Marble) via Google Earth Engine |

**Step 2: Creating a grid of blocks (initial sampling grid) that covers the sampling boundary.**

Using a custom python script, the area contained within the sampling boundary is divided into a grid of blocks of size 150m by 150m. This initial grid covers areas that aren’t built-up, such as water bodies, shrubland or cropland, so an additional step is needed to eliminate those areas from the sampling grid. Since this additional step requires processing pixels over a large area using google earth engine, the initial sampling grid is split into sections of 2,000 blocks each to avoid running into processing limits (Figure 4).

|  |
| --- |
| **Figure 4:** Lagos, initial sampling grid (divided into sections) |
| A graph of a person's body  Description automatically generated |
| **Initial sampling grid dimensions**  East-west distance: 39.40 km  North-South distance 13.96 km  East-west N blocks: 94  North-south N blocks: 263  Total area of boundary polygon: 266.09 km2  Total area of block polygons: 277.53 km2  Total number of blocks in sampling grid: 12,371 |

**Step 3: Eliminating non-built blocks from the sampling grid.**

To remove blocks that do not contain built-structures, two datasets are examined: the 10m resolution Dynamic World land-cover dataset[[2]](#footnote-2) and the Open Buildings V3 dataset[[3]](#footnote-3) sourced through Google Earth Engine. For the Dynamic World land-cover dataset, a cloud-free composite covering the entire sampling boundary is first created using recent imagery (past 12-month data as of May 2024). The modal land-cover classification (the land cover class that has the highest estimated probability out of 9 possible classes[[4]](#footnote-4)) in the “label” band, for each pixel over the timeframe is computed (Figure 5). For each block, the percentage of pixels that are classified as “built” in the “label” band is computed. Only blocks with a composition of built pixels greater than 75% are retained for the final sampling grid. To classify blocks as built or not using the Open Buildings dataset, the total number of detected buildings (with a confidence estimate of greater than 0.75) is counted for each block in the grid. Only blocks with at least 1 high probability building are retained in the final sampling grid. For Lagos, we observed that there are a considerable number of blocks classified in the Dynamic World datasets as “Bare” that do contain buildings and roads (Figure 5B-5C).

|  |
| --- |
| **Figure 5A:** Land-use/land-cover (red = built pixels), Source: Dynamic World |
| A map of a city  Description automatically generated |
| **Figure 5B:** Building boundaries (white = building outlines), Source: Open buildings |
| A white light on a dark background  Description automatically generated with medium confidence |
| **Figure 5c:** Close up of area with poor correspondence of built-up classification and presence of buildings |
| A map with red spots  Description automatically generated  Buildings detected in areas that are not classified as “built” with a high probability probabilityp |

Figure 6 shows the resulting sampling grid using the two approaches. The Dynamic World dataset yields a sampling grid that, of the initial sampling grid, is 54% built-up, while the Open Buildings dataset yields a sampling grid that, of the initial sampling grid, is 75.03% built-up.

|  |
| --- |
| **Figure 6:** Built-up blocks |
|  |

Given the likely classification error evident in the dynamic world dataset for certain parts of the Lagos study area, we use the Open Buildings derived classification of “built” blocks to construct the final sampling grid (Figure 6).

|  |
| --- |
| **Figure 7:** Final sampling grid comprised of blocks with at least one detected “high confidence” building |
| A map of a city  Description automatically generated |
| Final sampling grid contains 9,282 blocks. |

**Step 4:** Randomly select initial sample of blocks from the sampling grid.

In the final step, an initial sample of 125 blocks is selected randomly for enumeration (Figure 7, Figure 8A, \*B).

|  |
| --- |
| **Figure 8:** 125 randomly selected blocks for enumeration |
| A map of a city  Description automatically generated |

|  |
| --- |
| **Figure 9A:** Sampled blocks in Lagos over Open Street Maps layer |
| A map of a city  Description automatically generated |
| **Figure 9B:** Detail |
| A map of a city  Description automatically generated |

1. https://data.humdata.org [↑](#footnote-ref-1)
2. https://developers.google.com/earth-engine/datasets/catalog/GOOGLE\_DYNAMICWORLD\_V1 [↑](#footnote-ref-2)
3. https://developers.google.com/earth-engine/datasets/catalog/GOOGLE\_Research\_open-buildings\_v3\_polygons [↑](#footnote-ref-3)
4. The 9 land-cover classes are: water, trees, grass, flooded vegetation, crops, shrub and scrub, built, bare, snow and ice. [↑](#footnote-ref-4)