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

Sampling design methodological note

**Creating spatial grids for stage 1 sampling**

**Addis Ababa, Ethiopia**

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 (area of interest)

The first step in this process is to define the sampling boundary or perimeter. In Ethiopia, we use sub-national administrative boundaries available from GADM and the sampling boundary is constructed using a combination 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 in Ethiopia |
| A map of ethiopia with green outline  Description automatically generated |
| Source: UN Humanitarian Data Exchange (HDX) |

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

|  |  |  |
| --- | --- | --- |
| **Figure 2:** Addis Ababa true-color day-time satellite composite | | **Figure 3:** Nighttime lights (NTL) |
| A satellite image of a city  Description automatically generated |  | |
| Source: Sentinel-hub, NASA VIIRS (Black Marble) | | |

**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:** Addis Ababa, initial sampling grid (divided into sections) |
|  |
| **Initial sampling grid dimensions**  East-west distance: 29.32 km  North-South distance: 29.43 km  East-west N blocks: 197  North-south N blocks: 196  Total area of boundary polygon: 553.02 km2  Total area of block polygons: 563.47 km2  Total number of blocks in sampling grid: 24,713 |

**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 data product and the Open Buildings data product. For the Dynamic World land-cover data, a composite covering the entire sampling boundary is first created using recent (past 3-month data), the most common highest probability land-cover classification for each pixel over the period is used (Figure 5A). For each block, the percentage of pixels that are classified as “built” is computed. Only blocks with a composition of built block pixels of greater than 75% are retained for in the final sampling grid. To classify blocks using the Open Buildings dataset, the total number of detected buildings (with a confidence estimate of greater than 0.85) is counted for each block in the grid (Figure 6, top panel). 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-up), Source: Dynamic World | **Figure 5B:** Building boundaries (white = building outlines), Source: Open buildings |
| A map of a red and green area  Description automatically generated |  |

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 64.17% built-up, while the Open Buildings dataset yields a sampling grid that, of the initial sampling grid, is 79.2% built-up.

|  |
| --- |
| **Figure 6:** Built-up blocks |
| **A red and blue map  Description automatically generated** |

In the case of Ethiopia, we use the Open Buildings derived classification of “built” blocks to construct the final sampling grid (Figure 6).

|  |
| --- |
| **Figure 6:** Final sampling grid comprised of blocks with at least one detected building |
| A map of a city  Description automatically generated |
| Final sampling grid contains 19,750 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 for enumeration (Figure 7, Figure 8).

|  |
| --- |
| **Figure 7:** 125 randomly selected blocks for enumeration |
| A map of a city  Description automatically generated |

|  |
| --- |
| **Figure 8A:** Sampled blocks in Lagos over Open Street Maps layer |
| A map of a city  Description automatically generated |
| **Figure 8B:** Detail |
| A map of a city  Description automatically generated |