**CFI MAP-2 Demand side study**

Sampling design methodological note

**Creating spatial grids for stage 1 sampling**

**Sao Paolo, Brazil**

To create a representative sample of small business establishments (those with fewer than 10 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. We use publicly available sub-national administrative boundaries, and the sampling boundary is constructed using a combination of one or more level-2 administrative units that cover the target city. In Brazil, level-1 administrative units correspond to states and level-2 administrative units correspond to municipalities and the boundary files are obtained from GADM[[1]](#footnote-1) using the GADM python API[[2]](#footnote-2). For this study, the sampling boundary is defined using the Sao Paolo municipality in Sao Paolo state (Figure 1).

|  |
| --- |
| **Figure 1:** Level 1 and Level 2 administrative boundaries |
| A map of a study area  Description automatically generated |
| Source: GADM |

Figures 2 and 3 display the study boundary overlaid over remote-sensing satellite data, illustrating that the administrative boundary, does not fully encompass the true extent of the Sao Paolo metropolitan (urban) area.

|  |  |  |
| --- | --- | --- |
| **Figure 2:** True-color day-time satellite composite | | **Figure 3:** Nighttime lights (NTL) |
|  | A map of the world  Description automatically generated | |
| Source: Cloud-free mosaic of Sentinel-2 imagery via Sentinel-Hub | 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, shrub 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:** Initial sampling grid (divided into sections) |
| A map of a region  Description automatically generated with medium confidence |
| **Initial sampling grid dimensions**  East-west distance: 46.74km  North-South distance 72.17km  East-west N blocks: 482  North-south N blocks: 312  Total area of boundary polygon: 1814.75 km2  Total area of block polygons: 1847.94 km2  Total number of blocks in sampling grid: 69,173 |

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

To remove blocks that do not contain built-structures, the 10m resolution Dynamic World land-cover dataset[[3]](#footnote-3) sourced through Google Earth Engine is used. A cloud-free composite land-cover dataset covering the entire sampling boundary is first created using recent imagery (past 4-month data as of March 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 5A). 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.

|  |
| --- |
| **Figure 5A:** Estimated land-cover, 10m resolution (red = built) |
|  |
| Source: Dynamic World, 10m resolution |

Figure 6 shows the two steps in the procedure described above. The Dynamic World dataset yields a sampling grid that, of the initial sampling grid, is 55% built-up. The final sampling grid contains 38,017 blocks, covering a total area of 1,015km2.

|  |
| --- |
| **Figure 6:** Identifying built-up blocks using land-cover estimates |
| A map of different colors  Description automatically generated |

|  |
| --- |
| **Figure 7:** Final sampling grid comprised of blocks with at least 75% of pixels classified as “built” in the Dynamic World dataset. |
|  |
| The final sampling grid contains 38.017 blocks, covering a total area of 1,015km2 |

**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 8, Figure 9A/B).

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

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

1. https://gadm.org [↑](#footnote-ref-1)
2. https://pypi.org/project/gadm/ [↑](#footnote-ref-2)
3. https://developers.google.com/earth-engine/datasets/catalog/GOOGLE\_DYNAMICWORLD\_V1 [↑](#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)