Soils4Liberia

Sampling Design

Bas Kempen

13 February 2026

The primary objective of soil data collection in the Soils4Liberia project is to develop a new set digital (gridded) soil maps and derived information products with national coverage. To achieve this, a *spatial coverage sampling* design was selected to ensure good geographic representation across the area of interest. This approach is well suited to the project's objective (Brus, 2022; De Gruijter et al., 2006). This document describes the design of the sampling scheme.

Survey area

Although the digital soil maps will provide nationwide coverage, extensive areas of dense primary forest combined with a poor road network constrain the area that can be sampled within the project's timeframe and available resources. Hence, an *area of interest* (AoI) was delineated that excludes most of the primary forest as well as protected areas such as national parks. This allows the sampling campaign to focus on agricultural lands, plantations and forest margins (Figure 1).

The AoI covers 59,000 km^2, representing 61% of Liberia's total land area, and comprising most of Liberia's agricultural lands. A portion of the AoI is classified as '*difficult to access*' (grey areas in Figure 1). This area covers 16,675 km (28% of the AOI) and is not considered in the sampling design. The part classified as '*accessible*' is divided into areas: the 'breadbasket' agricultural lands and the 'other' agricultural lands (Figure 1).

The breadbasket lands comprise the counties of Nimba, Bong and Lofa, which include some of Liberia's most intensively cultivated agricultural regions. Because of their agricultural importance, these areas will be sampled at higher density in accordance with project requirements. The breadbasket lands cover 18,500 km^2 (32% of the AoI), and the other agricultural lands 23,725 km^2 (40% of the AoI). This means that the survey area covers 42,225 km^2, or 43,5% of Liberia's total land area.

Sampling design

The sampling design, based on *spatial coverage sampling*, is a nested design with three levels, each representing a sampling unit. Sampling locations are clustered at the third level to increase operational efficiency. Good spatial coverage is achieved by constructing *compact geographical strata* (CGS) using a clustering algorithm. The algorithm produces a set of geographic strata of approximately equal size that are as spatially compact as possible. Geographic stratification will lead to more precise predictions in case the variable of interest is spatially structured, which is the case for soil (Brus, 2022).

Primary sampling units

Because different sampling densities are required for the breadbasket lands and the other agricultural lands, the CGS were constructed done for each of these areas. The CGS serve as the *primary sampling units* (PSUs). Figure 2 shows the CSG covering the survey area.

There are 167 CGS within the breadbasket area, and 143 CGS in the other agricultural area. The average size of a CGS is 111 km^2 in the breadbasket area and 166 km^2 in the other agricultural area. This indicates that the sampling density in the breadbasket area is 1.5 times higher than in the other agricultural area.

Secondary sampling units

The *secondary sampling units* (SSUs) constitute the second level of the sampling design. Each SSU is defined as a 2 x 2 km area. The SSU is the centered on the centroid of the PSU such that the centroid of the CSG forms the center of the 2 x 2 km area.

Tertiary sampling units

Within the SSU, three sampling locations are selected for soil sampling. These locations constitute the third level of the sampling design and are referred to as the *tertiary sampling units* (TSUs). The first TSU is located at the SSU center (and thus at the center of the PSU), ensuring optimal spatial coverage. The other two TSUs are selected randomly within the SSU, subject to a minimum separation distance of 250 m between TSUs. Because all TSUs are situated within a 2 x 2 km area, these are within walking distance from each other.

Backup sampling units

Each PSU includes four backup SSUs in case the planned SSU at the PSU center cannot be sampled. These backup units are randomly selected within each PSU. Because the PSUs are relatively small compared to the size of an SSU, the five SSUs (one planned for sampling and four backups) within a PSU may overlap. Figure 3 shows the configuration of five SSUs within one PSU.

Each SSU includes three backup TSUs in case one or more of the planned TSUs cannot be sampled. In total, six TSUS are therefore selected within each SSU: three designated for sampling and three serving as backups. The backup TSUs are also selected randomly taking, subject to a minimum separation distance of 250 m. Figure 4 shows the configuration of six TSUs within one SSU.

Sample size and additional sampling locations

In total, the design comprises 310 PSUs. Within each PSU, three locations (TSUs) will be sampled, clustered within a 2 x 2 km area (SSU). This results in 930 planned locations.

Given the project requirement to sample at least 1000 locations, there is scope to include additional sites. It is therefore recommended to add an additional 90 locations organized into 30 clusters (each 2 x 2 km) containing three locations each. These locations can be selected purposively by the field campaign lead and subsequently digitized in a GIS.

Specific areas can be targeted for this purpose. These may include accessible agricultural areas outside the AoI, as well as areas that need to be better represented in the sample, such plantations and zones areas around mining sites. The sampling design would also benefit from the inclusion of additional sites within primary forest areas.

Sampling order

Each PSU contains five SSUs, and each SSU contains six TSUs. Both SSUs and TSUs are assigned sequential numbers (1-5 for SSUs and 1-6 for TSUs) that determine the sampling order.

In principle, SSU number 1 should be sampled. If it is not possible, SSU number 2 (the first backup) should be considered. If SSU number 2 cannot be sampled, SSU number 3 should be used, and so on in sequential order.

Similarly, TSUs numbered 1 to 3 are the primary sampling locations and must be sampled. If any of these cannot be sampled (e.g., because of denied access, the site being located on a road, etc.), TSU number 4 (the first backup) should be sampled, followed by numbers 5 and six in that specific order, if necessary.

Adhering to this sequence prevents preferential site selection, which could introduce selection bias and reduce the overall information content of the dataset. Such bias may ultimately compromise the quality of the soil maps derived from the data.

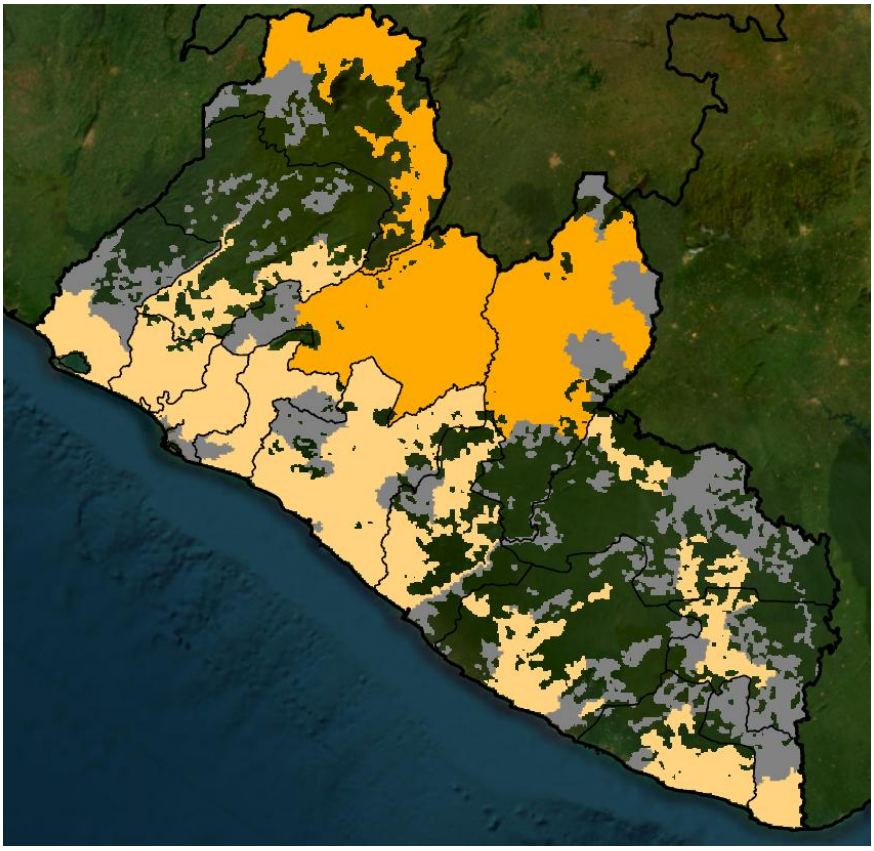

Figure 1. Area of interest (AoI) with three classes: accessible-breadbasket (dark yellow), accessible-other (yellow), accessible-bread basket, not accessible (grey). The AoI largely excludes the dense primary forests. The lines represent the county boundaries.

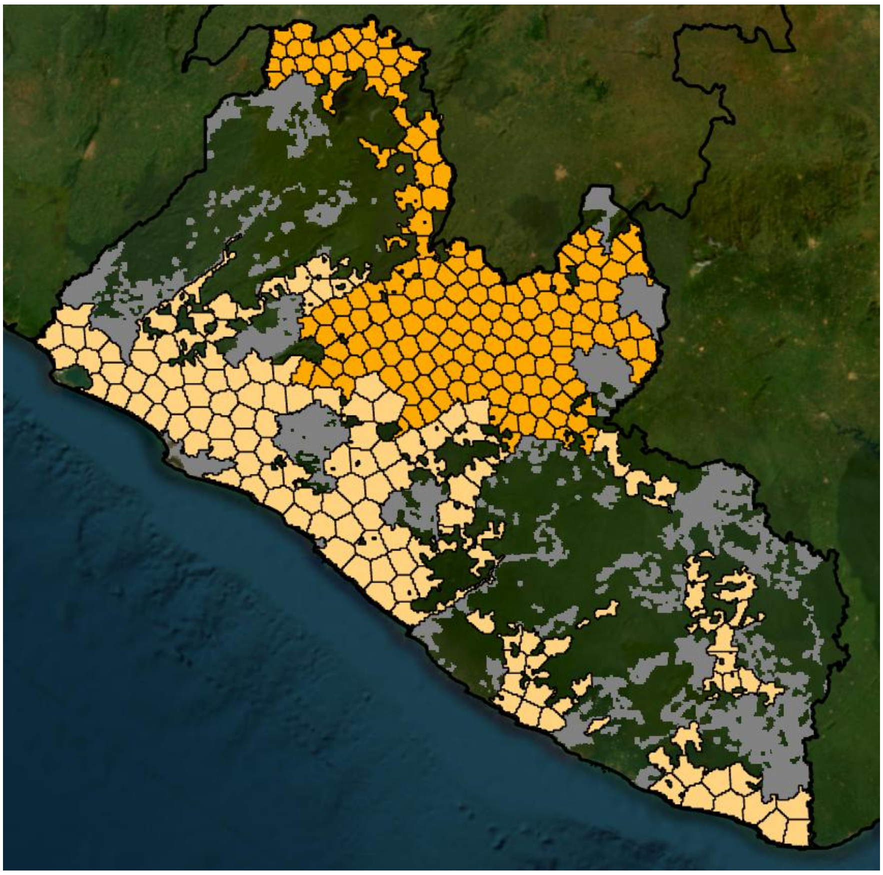

Figure 2. Compact geographic strata that serve as the primary sampling units of the sampling design.

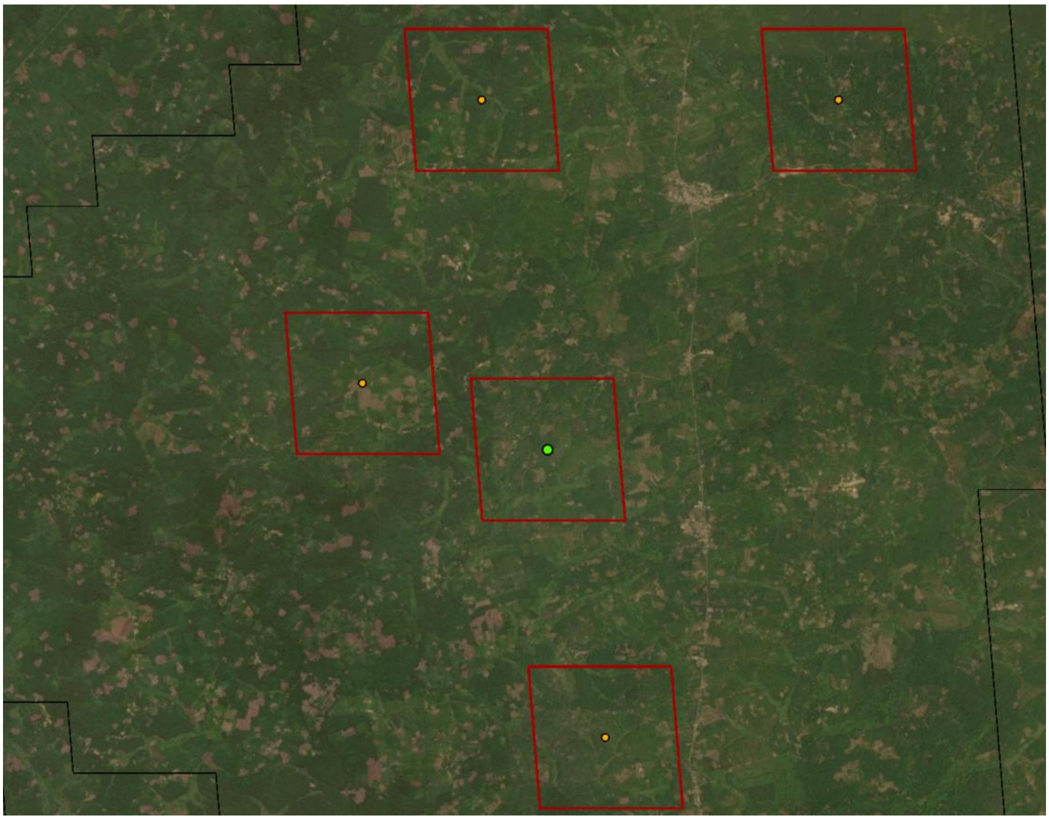

Figure 3. Location of five SSUs within one PSU (black outline). Each SSU is a 2 x 2 km area that is delinated by a square bounding box. The SSU with the large, green point is the planned SSU situated in the center of the PSU. The other four are backup SSUs (orange points designate the SSU centers).

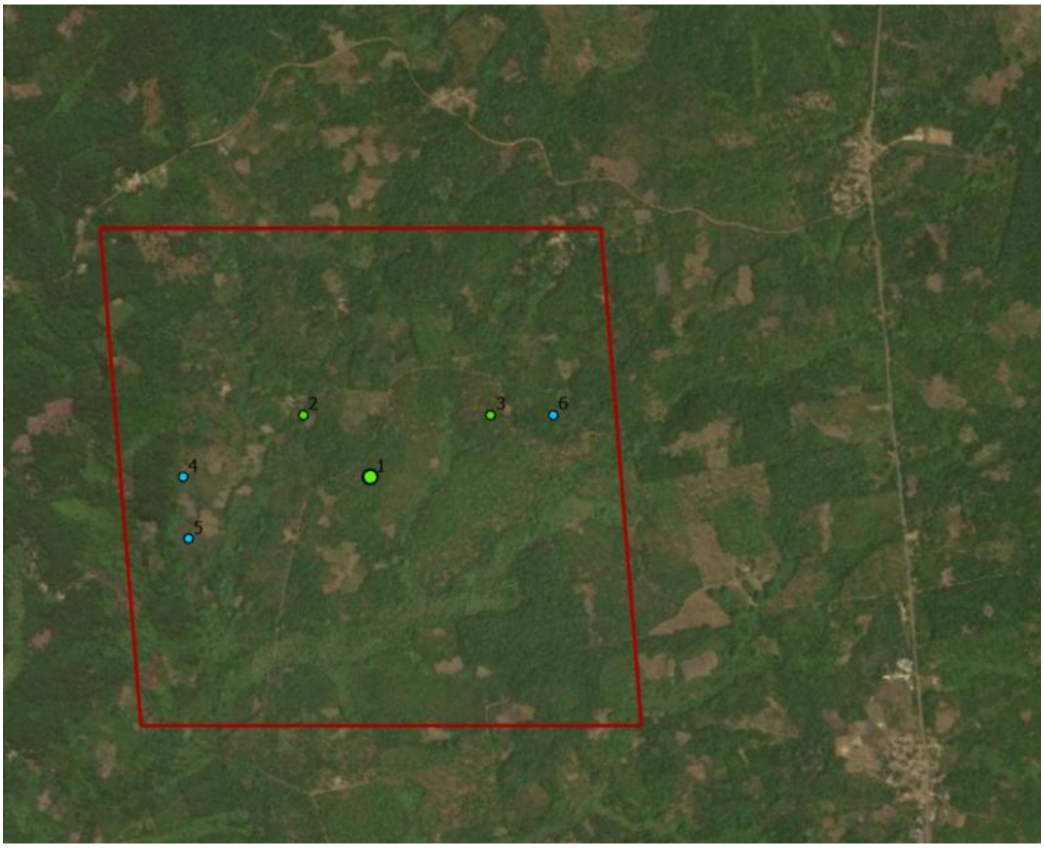

Figure 4. SSU (2 x 2 km area delinated by the squre bounding box) with six TSUs. The first three are planned for sampling (1-3), the others (4-6) are backup TSUs.

Sampling unit identifiers

The identifier of a sampling location is constructed by combining the two-letter country code, the (unique) PSU number and the SSU and TSU sampling order numbers, separated by dashes as follows:

LR<PSU>-<SSU>-<TSU>

For example, **LR001-1-2** refers PSU 001, SSU 1 (the first selected SSU within the PSU), and TSU 2 (the second selected TSU in the SSU).

PSUs in the 'breadbasket' agricultural land area are numbered 001 to 167, while PSUs in the other agricultural land area are numbered `501` to `643`.

Validation

Sampling units must to be validated before going into the field to assess whether they can be sampled. This process reduces the risk that field surveyors travel to a site to find out sampling is not feasible because of constraints such as for instance inaccessibility because of difficult terrain.

Validation takes place at the level of the SSU and is conducted using the Survey Data Management Tool (SDMT). Based on inspection of aerial imagery, the field survey lead must determine whether an SSU can reasonably be visited. Some SSUs are located close to roads, where accessibility is unlikely to pose a problem. Others may be located farther from the road network. For these SSUs, the field survey lead must assess whether the site can be reached with a reasonable amount of effort. Including SSUs that are farther from the road network is important to avoid selection bias.

If the field survey lead determines that a selected SSU cannot be visited, it may be rejected for sampling. The SDMT will then display the next (backup) SSU in the predefined sequence for validation.

Validation of the TSUs is typically carried out by the field surveyor based on on-site conditions. Surveyors are strongly encouraged to make a reasonable effort to reach planned TSUs that are located farther from the vehicle parking location. This is essential to avoid introducing selection bias and compromising the representativeness and integrity of the sample.

References

Brus, D.J., 2022. Spatial Sampling with R. CRC Press. Available at: https://dickbrus.github.io/SpatialSamplingwithR/

De Gruijter, J.J, Bierkens, M.F.P., Brus, D.J., Knotters, M., 2006. Sampling for Natural Resource Monitoring. Springer. Available at: https://link.springer.com/book/10.1007/3-540-33161-1