

Multi-Level Sampling for Soil Analysis

Jiehua Chen
Africa Soil Information Service

July 23, 2014

1 Multilevel Sampling Procedure

The sampling procedure defined here combines stratified, clustered, and simple random sampling to optimally balance sampling and cost. Since remote sensing covariates data usually has a resolution of $100 \times 100 \text{ m}^2$, it is difficult to achieve reasonably high accuracy for predicting soil maps with a resolution finer than $100 \times 100 \text{ m}^2$. Therefore, the smallest resolution grid in the sampling procedure is $100 \times 100 \text{ m}^2$.

There are 3 parameters required as input:

1. latitude and longitude for the center of the sampling grid
2. M , the number of $1 \times 1 \text{ km}^2$ grids
3. N , the number of $100 \times 100 \text{ m}^2$ grids

Below is an outline of the sampling procedure:

1. Input the center location of $10 \times 10 \text{ km}^2$ grids, which are fully included in the ROI.
2. Divide each selected $10 \times 10 \text{ km}^2$ grid into four $5 \times 5 \text{ km}^2$ grids, and then further divide each $5 \times 5 \text{ km}^2$ grid into twenty-five $1 \times 1 \text{ km}^2$ grids.
3. Select M $1 \times 1 \text{ km}^2$ grids within each of the four $5 \times 5 \text{ km}^2$ grids.
4. Select N $100 \times 100 \text{ m}^2$ grids within each selected $1 \times 1 \text{ km}^2$ grid.

N , the number of $100 \times 100 \text{ m}^2$ grids sampled, can be chosen through power analysis, if prior knowledge of the variability of the interested soil properties is available. Since geostatistics models will be implemented in DSM, we require taking enough samples within a 1 km distance to enable the estimation of the spatial correlations at short distances. Therefore, we recommend N to be at least 4.

2 Advantages

Some advantages of this sampling procedure include:

1. Implementing stratified sampling at the $5 \times 5 \text{ km}^2$ level helps cover a wide enough area to capture spatial variability. Also, cluster sampling at the $1 \times 1 \text{ km}^2$ level helps reduce the transportation costs of field work.
2. The procedure can be easily automated and documented, simplifying statistical analysis and prediction work.
3. The procedure is flexible to customization, as both the sizes of grids and the allocations of samples at each level of the grids can be adjusted in accordance with variations in field scenarios and available information.