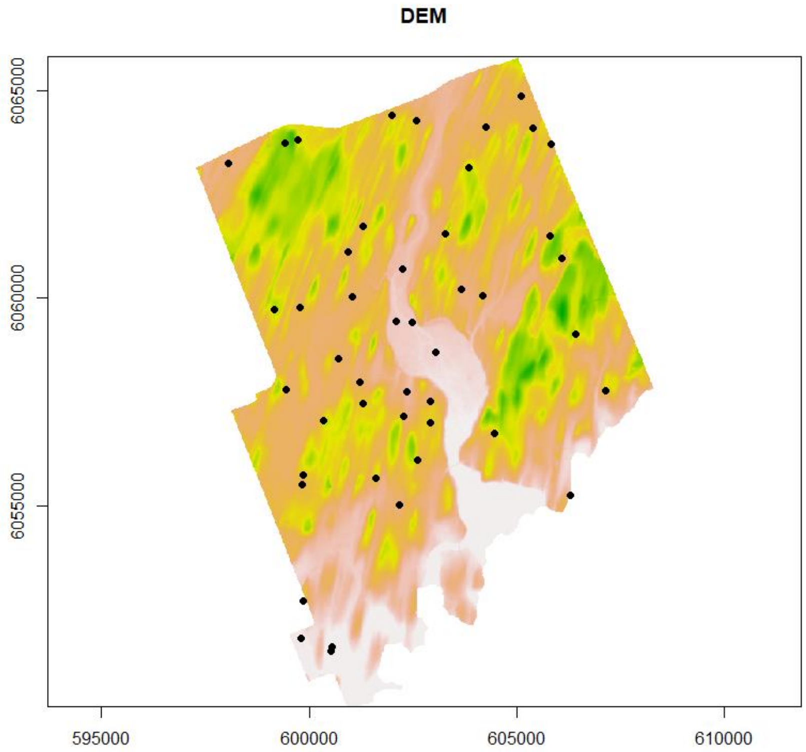
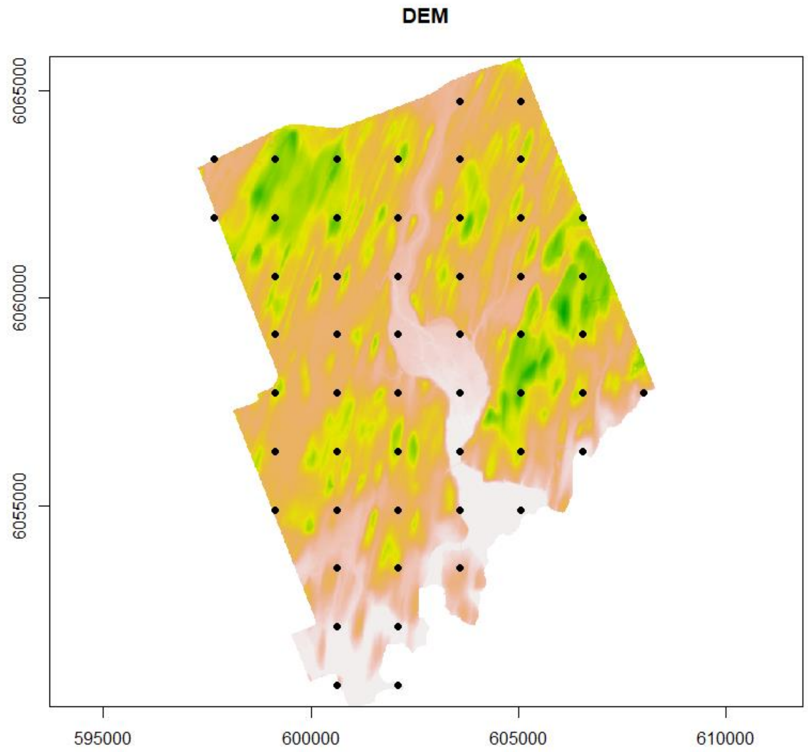




4. Sampling Design for Digital Soil Mapping

Canadian Digital Soil Mapping Workshop, 2020

CANADIAN SOCIETY OF SOIL SCIENCE
SOCIÉTÉ CANADIENNE DE LA SCIENCE DU SOL
Ontario
UNIVERSITY of GUELPH
DALHOUSIE UNIVERSITY
FACULTY OF AGRICULTURE

About Us

Brandon Heung

Dalhousie University

Brandon.Heung@dal.ca

Daniel Saurette

Ontario Ministry of Agri-Food and Rural Affairs

University of Guelph

Daniel.Saurette@ontario.ca

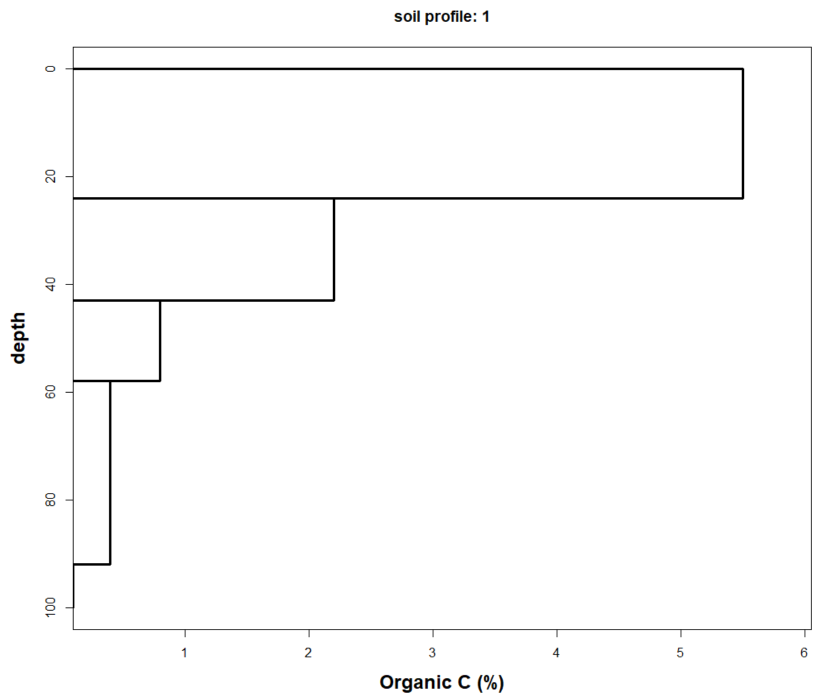
Outline

1. **Harmonizing soil profile data**

2. Sampling Design

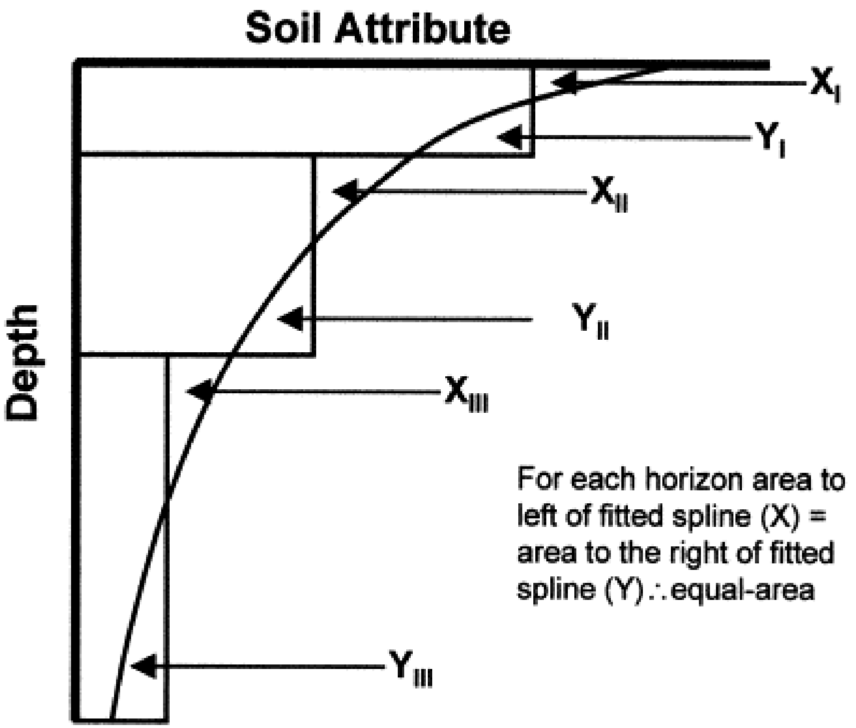
Depth Functions

- Soil sampling traditionally completed by pedogenic horizon (e.g., Ah, Ae, Bt, Ck)
- Bulk sample is collected from each horizon
- Analyses are completed and results represent the mean for that horizon
- In Digital Soil Mapping, we need the ability to predict for any depth interval
- 2 common approaches:
    - 3D predictions
    - stacked 2D predictions
- We will deal with stacked 2D approach



Depth Functions – 2D Stacking Approach

- To allow us to set prediction intervals, we need to convert the horizon data to a continuous function
- Most common approach is to use a mass preserving spline, or equal area quadratic spline function
- 2 parameters:
  - lam – controls smoothness
  - d – depth intervals
- User can specify as many depth intervals as wanted
- GlobalSoilMap.net specifications: 0-5, 5-15, 15-30, 30-60, 60-100 and 100-200 cm



Fun Fact:
One of the early pioneers of this approach, Dr. Ponce-Hernandez, is professor at Trent University!

Module 4.1: Depth Functions

- Hypothetical soil profile was sampled on a horizon basis.
- Particle size fractions, SOC, and pH were measured.
- Equal-area spline function will be used to convert the data into standard depth increments
  - 0-5 cm
  - 5-15 cm
  - 15-30 cm
  - 30-60 cm
  - 60-100 cm

Outline

1. Harmonizing soil profile data

2. Sampling Design

Sampling Design

- Sampling design is a critical first step in DSM, much like any experiment
- Two high-level types of sampling:
  - Probability sampling (design-based)
  - Non-probability sampling (model-based)

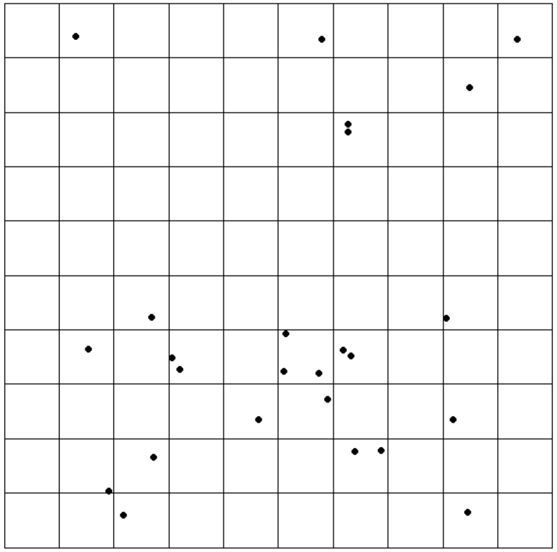
- Design-based Sampling must meet 2 requirements:
  - All units of the population have a probability of being selected
  - Selection probability of each unit is known

- Today we will review design-based sampling approaches:
  - Simple Random Sampling
  - Stratified Random Sampling – Equal
  - Stratified Random Sampling – Proportional
  - Grid Sampling
  - Conditioned Latin Hypercube Sampling

Simple Random Sampling

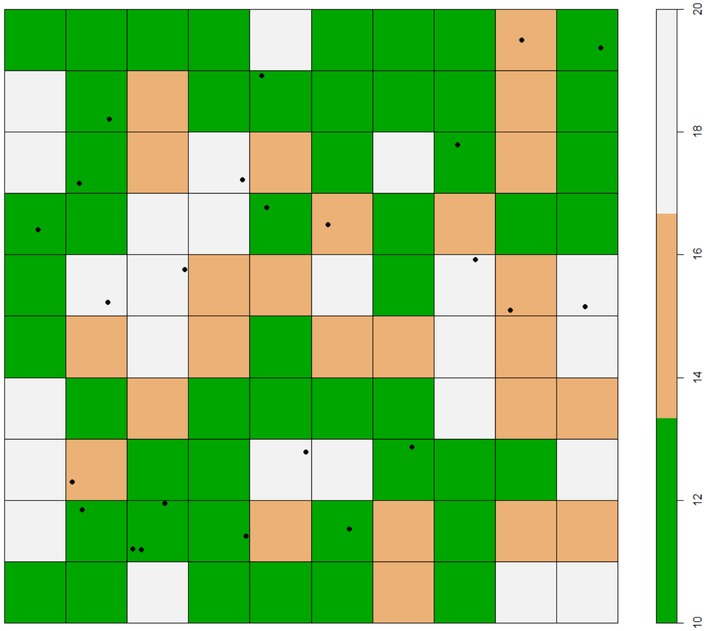
- All samples within the region have equal probability of being selected
- Advantages
  - Simple
  - Prior knowledge of the population not required

- Disadvantages
  - May not consider feature attributes (covariates)
  - Potential for clustered sampling
  - Efficiency may be reduced



Stratified Random Sampling

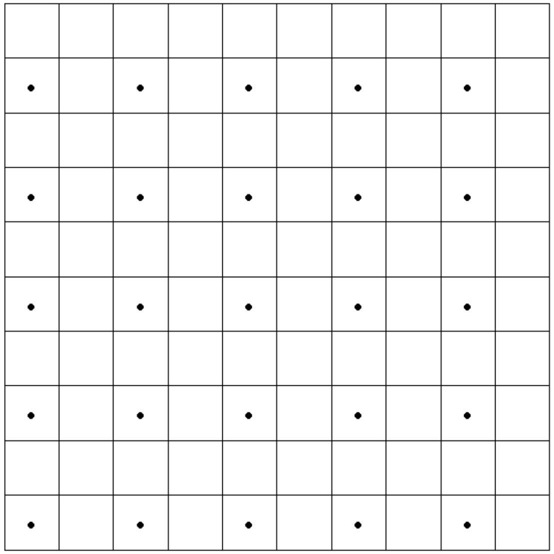
- Divide the population into strata (e.g. elevation)
- Sampling is random within the strata
- Equal sampling or proportional sampling by strata
- Advantages
  - Higher accuracy (if strata are meaningful)
- Disadvantages
  - May not consider feature attributes (covariates)
  - Strata knowledge may be flawed



Grid Sampling

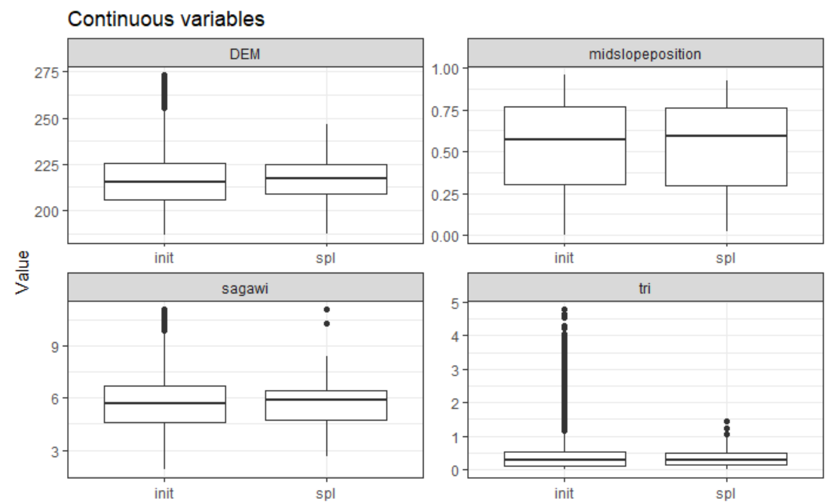
- Systematic sampling approach
- Must determine adequate grid spacing
- Advantages
    - Simple
    - Efficient
    - Low Cost

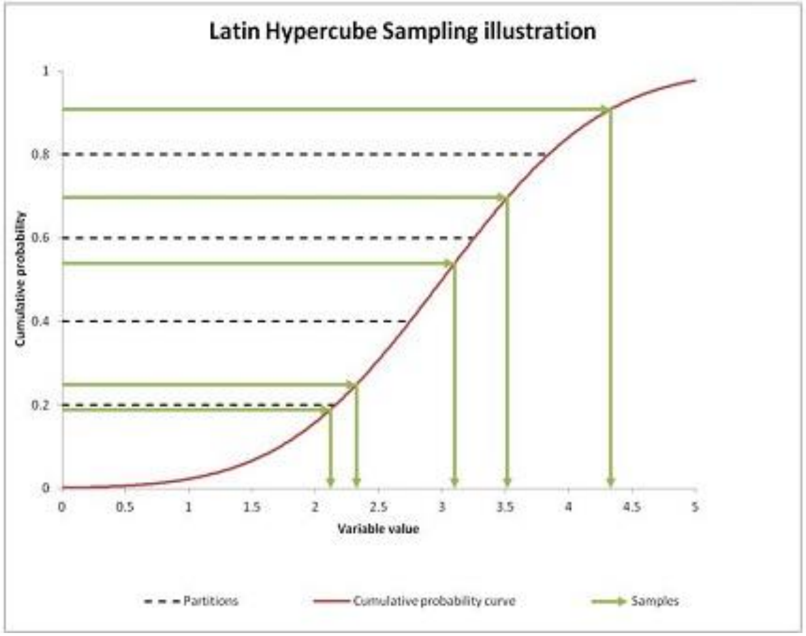
- Disadvantages
    - May not consider feature attributes (covariates)
    - Hard to determine appropriate grid size
    - Coarse grid could miss small-scale variability



Conditioned Latin Hypercube Sampling (cLHS)

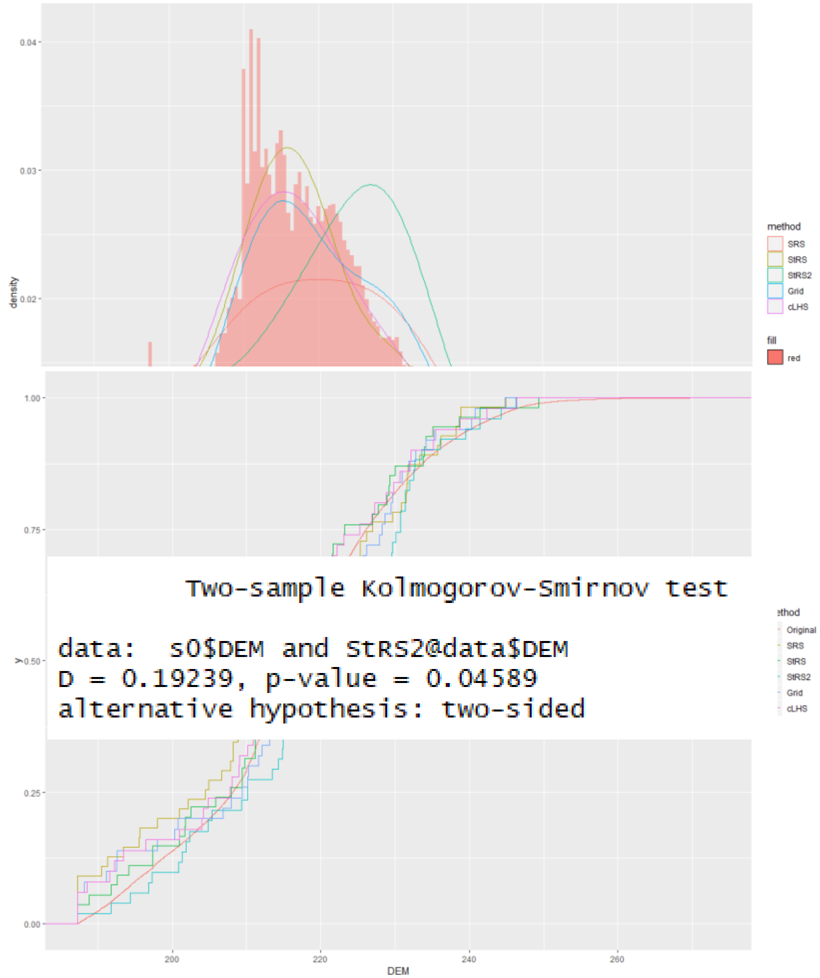
- Type of stratified sampling
- Full factorial experiment is not feasible with many covariates
- Covariates are each split into intervals based on their CDFs
- Selects one observation per partition of the CDF for each covariate
- Advantages
    - Closely resembles the distribution of the population of the covariates
    - Sample plan always optimized for feature space
- Disadvantages
    - Determining sample size
    - Access to sites can be difficult
    - No spatial component (e.g., potential for clustering and dispersion)





Assessing the Feature Space Coverage of a Sampling Plan

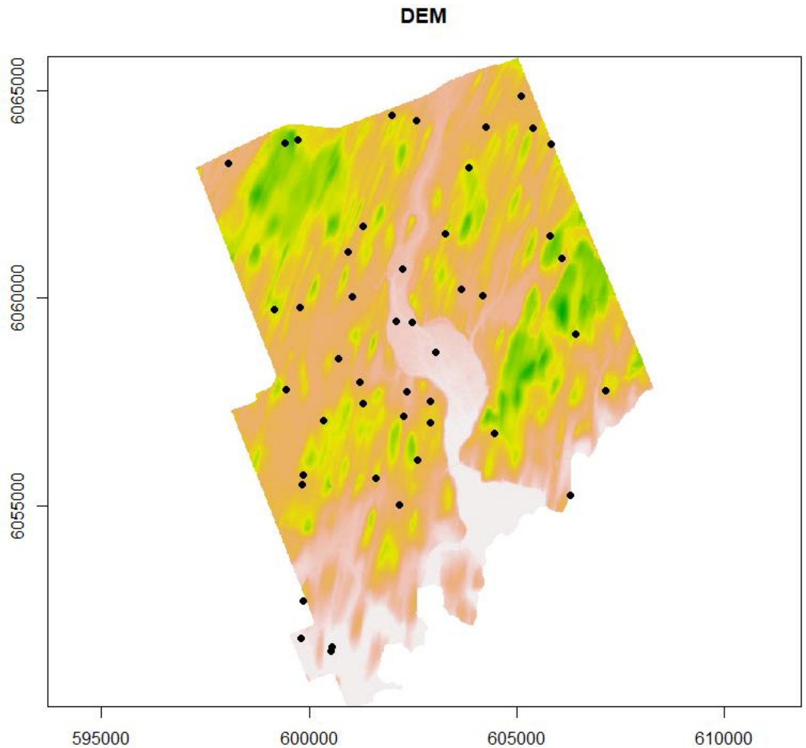
- Visual assessment comparing histograms
- Visual assessment comparing cumulative distribution functions
- Kolmogorov-Smirnov test
    - Non-parametric test of the equality of continuous distributions
    - Compares the CDFs
    - Null hypothesis is that both samples come from the same distribution
    - Reject at $p < 0.05$

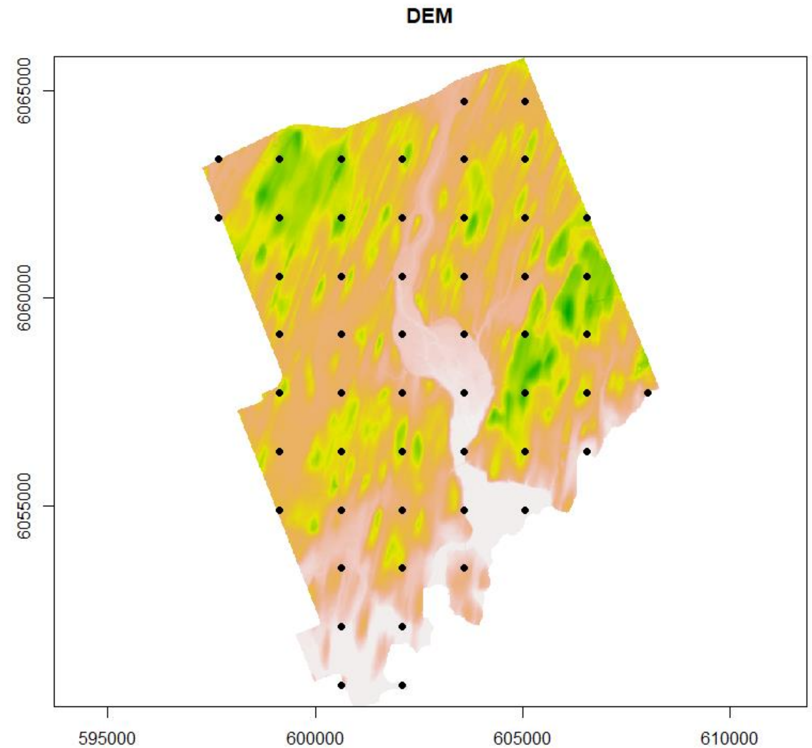


Module 4.2: Sampling Designs

- Generate sampling designs in R for Keene, ON:
    - Simple Random Sampling
    - Stratified Random Sample: Equal Sampling within Strata
    - Stratified Random Sampling: Proportional Sampling within Strata
    - Grid sampling
    - Conditioned Latin Hypercube Sampling

- Evaluating sampling designs





4. Sampling Design for Digital Soil Mapping

Canadian Digital Soil Mapping Workshop, 2020