



Global Soil Nutrient and Nutrient Budgets maps (GSNmap) Phase I Technical Manual

Food and Agriculture Organization of the United Nations
Rome, 2022

Required citation:

FAO. 2022. *Global Soil Organic Carbon Map – GSOCmap v.1.6. Technical report*. Rome. <https://doi.org/10.4060/cb9015en>

The designations employed and the presentation of material in this information product do not imply the expression of any opinion whatsoever on the part of the Food and Agriculture Organization of the United Nations (FAO) concerning the legal or development status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. The mention of specific companies or products of manufacturers, whether or not these have been patented, does not imply that these have been endorsed or recommended by FAO in preference to others of a similar nature that are not mentioned.

The views expressed in this information product are those of the author(s) and do not necessarily reflect the views or policies of FAO.

ISBN 978-92-5-135899-3

© FAO, 2022



Some rights reserved. This work is made available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo/legalcode>).

Under the terms of this licence, this work may be copied, redistributed and adapted for non-commercial purposes, provided that the work is appropriately cited. In any use of this work, there should be no suggestion that FAO endorses any specific organization, products or services. The use of the FAO logo is not permitted. If the work is adapted, then it must be licensed under the same or equivalent Creative Commons licence. If a translation of this work is created, it must include the following disclaimer along with the required citation: "This translation was not created by the Food and Agriculture Organization of the United Nations (FAO). FAO is not responsible for the content or accuracy of this translation. The original [Language] edition shall be the authoritative edition."

Disputes arising under the licence that cannot be settled amicably will be resolved by mediation and arbitration as described in Article 8 of the licence except as otherwise provided herein. The applicable mediation rules will be the mediation rules of the World Intellectual Property Organization <http://www.wipo.int/amc/en/mediation/rules> and any arbitration will be conducted in accordance with the Arbitration Rules of the United Nations Commission on International Trade Law (UNCITRAL).

Third-party materials. Users wishing to reuse material from this work that is attributed to a third party, such as tables, figures or images, are responsible for determining whether permission is needed for that reuse and for obtaining permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

Sales, rights and licensing. FAO information products are available on the FAO website (www.fao.org/publications) and can be purchased through publications-sales@fao.org. Requests for commercial use should be submitted via: www.fao.org/contact-us/licence-request. Queries regarding rights and licensing should be submitted to: copyright@fao.org.

Cover design: ©FAO/Matteo Sala

Contents

Licence	xi
Abbreviations and acronyms	xii
Contributors and reviewers	xiii
Presentation and basics	xiv
What this Tutorial is not	xiv
The Training Dataset	xiv
Data Structure	xiv
How to Use This Manual	xiv
Manual Structure	xiv
Required downloads	xv
Step-by-step setup instructions	xv
0.1 What is R?	xv
0.1.1 Why R for Soil Science?	xv
0.2 Installing R and RStudio	xv
0.2.1 Installing R	xv
0.2.2 Installing RStudio	xv
0.2.3 Verifying Installation	xv
0.3 Understanding the RStudio Interface	xv
0.3.1 1. Console Pane (Bottom Left)	xv
0.3.2 2. Code Editor Pane (Top Left)	xv
0.3.3 3. Environment/History Pane (Top Right)	xv
0.3.4 4. Files/Plots/Packages/Help Pane (Bottom Right)	xvi
0.4 R Packages: Extending R's Capabilities	xvi
0.4.1 What are Packages?	xvi

0.4.2	Installing Packages	xvi
0.4.3	Loading Packages	xvi
0.4.4	Finding Help on Packages	xvi
0.5	R Basics: Objects and Data Types	xvi
0.5.1	Creating Objects with Assignment	xvi
0.5.2	Data Types in R	xvi
0.5.3	Checking Data Types	xvi
0.6	Data Structures in R	xvii
0.6.1	Vectors	xvii
0.6.2	Factors	xvii
0.6.3	Matrices	xvii
0.6.4	Data Frames	xvii
0.6.5	Lists	xvii
0.7	Operators in R	xvii
0.7.1	Arithmetic Operators	xvii
0.7.2	Comparison Operators	xvii
0.7.3	Logical Operators	xvii
0.8	Control Structures: Conditional Statements	xvii
0.8.1	If-Else Statements	xviii
0.8.2	Vectorized If-Else: <code>ifelse()</code>	xviii
0.8.3	Vectorized Cut: <code>cut()</code>	xviii
0.9	Control Structures: Loops	xviii
0.9.1	For Loops	xviii
0.9.2	While Loops	xviii
0.10	Functions in R	xviii
0.10.1	Using Built-in Functions	xviii
0.10.2	Creating Custom Functions	xviii
0.10.3	Function Arguments and Defaults	xviii
0.11	Data Manipulation with Base R	xviii
0.11.1	Subsetting and Filtering	xviii
0.11.2	Deleting Columns	xviii
0.11.3	Sorting Data	xviii
0.11.4	Aggregating Data	xviii
0.12	Data Manipulation with the Tidyverse	xviii
0.12.1	Loading the Tidyverse	xix
0.12.2	Tibbles: Modern Data Frames	xix
0.12.3	The Pipe Operator: <code>%>%</code>	xix
0.12.4	Data Manipulation with <code>{dplyr}</code>	xix
0.12.5	Renaming Columns with <code>rename()</code>	xix

0.12.6	Grouping and Summarizing Data by columns	xix
0.13	Combining Data Frames	xix
0.13.1	Understanding Joins	xix
0.13.2	Stacking Data with <code>bind_rows()</code>	xix
0.14	Missing Data	xix
0.14.1	Identifying Missing Data	xix
0.14.2	Handling Missing Data	xx
0.15	Data Reshaping with <code>{tidyr}</code> package: Pivoting	xx
0.16	Handling Below Detection Limit Data	xx
0.17	Working with Soil Data	xx
0.17.1	Setting Working Directory	xx
0.17.2	Load packages	xx
0.17.3	Importing Data from Files	xx
0.17.4	Exploring the Data	xx
0.18	References	xx

Data preparation for Digital Soil Mapping in R xxi

0.19	Introduction	xxi
0.20	Loading and exploring raw soil data	xxi
0.20.1	Basic data loading and structure	xxi
0.20.2	Data components and organization	xxi
0.21	Preparing site data for analysis	xxi
0.21.1	Adding a unique row identifier	xxi
0.21.2	Extracting and standardizing column names for sites . . .	xxi
0.21.3	Creating unique profile identifiers	xxi
0.21.4	Remove exact duplicate site records	xxi
0.22	Coordinate validation and correction	xxii
0.22.1	Check 1: Missing coordinates	xxii
0.22.2	Check 2: Valid coordinate ranges	xxii
0.23	Soil depth validation and correction	xxii
0.23.1	Check 1: Missing depth boundaries	xxii
0.23.2	Check 2: Negative depth values	xxii
0.23.3	Check 3: Zero-thickness intervals	xxii
0.23.4	Check 4: Invalid depth logic	xxii
0.23.5	Check 5: Profiles without a surface horizon (top > 0) . .	xxii
0.24	Preparing lab data: harmonization and validation	xxii
0.24.1	Prepare and standardize laboratory columns	xxii
0.24.2	Check 1: Check each property against feasible analytical thresholds	xxii

0.24.3	Check 2: Texture validation	xxii
0.24.4	Check 3: Correction of out-of-bounds laboratory values	xxiii
0.25	Resolving duplicated data in soil profiles	xxiii
0.25.1	Why duplicates occur	xxiii
0.25.2	Types of duplicate situations	xxiii
0.25.3	Check 1: Detect potential horizon duplicates within profiles	xxiii
0.25.4	Resolving Profile Duplicates	xxiii
0.25.5	Remove profiles not starting at the surface	xxiii
0.25.6	Result	xxiii
0.26	Harmonizing data for DSM	xxiii
0.26.1	Depth standardization	xxiii
0.26.2	Processing standardized data	xxiii
0.26.3	Exporting standardized data for Digital Soil Mapping	xxiii
0.27	Preparing data for spectroscopy analyses	xxiv
0.28	Summary of exported files	xxiv
	References	xxiv

Sampling design for soil surveys xxv

0.29	Introduction	xxv
0.30	Sampling methodologies for soil spatial survey	xxv
0.30.1	Selection of the sampling methodology	xxv
0.30.2	Inference Methods	xxv
0.30.3	Purpose of Sampling	xxv
0.30.4	Sampling Design Types	xxv
0.31	SoilFER Sampling Design	xxv
0.31.1	Soil Properties	xxv
0.31.2	Environmental Covariates	xxvi
0.31.3	Understanding the Methodology	xxvi
0.32	Tutorial Using R	xxvi
0.32.1	Setting Up the Environment	xxvi
0.32.2	Define Variables and Parameters	xxvi
0.32.3	Country ISO Code	xxvi
0.32.4	Custom Functions	xxvi
0.32.5	Load Country Boundaries and Legacy Data	xxvii
0.32.6	Load Environmental Covariates for PSUs	xxvii
0.32.7	Load Main Environmental Covariates	xxvii
0.33	Principal Component Analysis (PCA)	xxvii
0.34	Organize the Sampling Universe	xxvii
0.34.1	Load and Process Land Use Data	xxvii

0.34.2	Exclude Protected Areas	xxvii
0.34.3	Restrict to Accessible Slopes	xxvii
0.34.4	Aggregate to 100m Resolution	xxvii
0.34.5	Save Processed Land Use	xxviii
0.34.6	Filter Legacy Data	xxviii
0.35	Create Primary Sampling Units (PSU Grid)	xxviii
0.35.1	Generate 2km \times 2km Grid	xxviii
0.35.2	Trim to Country Boundaries	xxviii
0.35.3	Save PSU Grid	xxviii
0.36	Select PSUs with Sufficient Land Use Coverage	xxviii
0.36.1	Calculate Land Use Percentage	xxviii
0.36.2	Filter PSUs	xxviii
0.37	Rasterize PSUs for Covariate Space Coverage	xxviii
0.38	Compute Optimal Sample Size	xxviii
0.38.1	Run Optimization	xxviii
0.38.2	Extract Optimal Sample Size	xxviii
0.39	Covariate Space Coverage - Computing PSUs	xxviii
0.39.1	Prepare Function Parameters	xxviii
0.39.2	Perform K-means Clustering	xxix
0.39.3	Assign Clusters and Calculate Distances	xxix
0.39.4	Extract Selected PSUs	xxix
0.39.5	Visualize in Covariate Space	xxix
0.40	Compute SSUs and TSUs	xxix
0.40.1	Load High-Resolution Covariates	xxix
0.40.2	Process Each PSU	xxix
0.40.3	Combine SSUs and TSUs	xxix
0.41	Export Sampling Units	xxix
0.41.1	Create Cluster Raster	xxix
0.41.2	Join Cluster Info to TSUs	xxix
0.41.3	Create Site IDs	xxix
0.41.4	Export Shapefiles	xxix
0.42	Compute Alternative PSUs	xxix
0.43	Assign Country-Wide Unique IDs	xxix
0.44	Export Unified Datasets	xxix
0.45	Summary Statistics	xxx
0.46	Create Distribution Maps	xxx
0.47	GPS Device Preparation	xxx
0.47.1	Export to GPX Format	xxx
0.47.2	Load onto GPS Device	xxx

0.48	Sampling Procedure	xxx
0.49	Data Recording	xxx
0.50	Quality Control	xxx
	Appendix A: Software and Data Sources	xxx
0.50.1	Software Requirements	xxx
0.50.2	Data Sources	xxx
	Appendix B: Troubleshooting Common Issues	xxx
0.50.3	Memory Issues	xxx
0.50.4	CRS Misalignment	xxx
0.50.5	Empty Geometry	xxx
	Appendix C: Acronyms and Abbreviations	xxxix
TBD		xxxii
Soil Data Preparation		xxxiii
0.51	Introduction	xxxiii
0.52	Data cleaning	xxxiii
0.53	Soil Data Harmonization	xxxiii
0.54	Concepts on digital exchange of soil-related data (ISO 28258)	xxxiii
0.55	Plot data	xxxiii
0.56	Profile data	xxxiii
0.57	Element data	xxxiii
0.58	Specimen data	xxxiii
0.59	Definition of laboratory procedures	xxxiii
0.60	Data transformation	xxxiv
0.61	Organizing soil data according to the GloSIS database	xxxiv
0.62	Tools for harmonization of soil databases within GloSIS	xxxiv
Digital Soil Mapping and Modeling		xxxv
	Part I: Theoretical Foundations	xxxv
0.63	Fundamentals of Digital Soil Mapping	xxxv
0.63.1	Introduction to Digital Soil Mapping	xxxv
0.63.2	The SCORPAN Framework	xxxv
0.63.3	Statistical Theory for Predictive Soil Mapping	xxxv
0.63.4	Types of Soil Variables	xxxvi
0.63.5	Prediction Methods	xxxvi
0.63.6	Nonlinearity and Uncertainty	xxxvi
	Part II: Practical Application	xxxvi
0.64	Workflow Overview	xxxvi

0.65	Continuous Modelling (pH)	xxxvi
0.65.1	Nature of the Variable	xxxvi
0.65.2	Data and Covariates	xxxvi
0.65.3	Feature Selection (Boruta)	xxxvi
0.65.4	Model Training	xxxvi
0.65.5	Variable Importance	xxxvi
0.65.6	Validation Metrics	xxxvi
0.65.7	Prediction Maps	xxxvi
0.66	Nominal Classification (Clay_pH_Class)	xxxvii
0.66.1	Nature of the Variable	xxxvii
0.66.2	Feature Selection	xxxvii
0.66.3	Model Training	xxxvii
0.66.4	Validation	xxxvii
0.66.5	Prediction Maps	xxxvii
0.67	Ordinal Modelling (pH_Class)	xxxvii
0.67.1	Nature of the Variable	xxxvii
0.67.2	Ordinal Encoding	xxxvii
0.67.3	Feature Selection	xxxvii
0.67.4	Model Training	xxxvii
0.67.5	Validation	xxxvii
0.67.6	Prediction Maps	xxxvii
0.68	Area of Applicability (AOA)	xxxviii
0.68.1	Why AOA is Necessary	xxxviii
0.68.2	Dissimilarity Index (DI)	xxxviii
0.68.3	AOA Binary Mask	xxxviii
0.68.4	Local Point Density (LPD)	xxxviii
0.68.5	Predictions Masked to AOA	xxxviii
0.69	Summary	xxxviii
	References	xxxviii

Soil Data Sharing	xxxix
0.70 Data sharing and export formats	xxxix
0.71 Metadata	xxxix
0.72 Web Services	xxxix

References	xl
-------------------	-----------

Figures

Tables



Licence

The Technical Manual is made available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence

CC BY-NC-SA 3.0 IGO.

Abbreviations and acronyms

BD Bulk density

CEC Cation exchange capacity

CRAN Comprehensive R archive network

DSM Digital soil mapping

GEE Google Earth Engine

GSP Global Soil Partnership

INSII International Network of Soil Information Institutions

ITPS Intergovernmental Technical Panel on Soils

ME Mean error

MAE Mean absolute error

MEC Modelling efficiency coefficient

NDVI Normalized difference in vegetation index

QA/QC Quality assurance/quality check

RMSE Root mean squared error

SOC Soil organic carbon

SOM Soil organic matter

Contributors and reviewers

Presentation and basics

Placeholder

What this Tutorial is not

The Training Dataset

Data Structure

How to Use This Manual

Manual Structure

Module 1: Introduction to R and Preparation of Soil Data for Digital Soil Mapping

Module 2: Soil Sampling Design

Module 3: GloSIS Data Preparation

Module 4: Digital Soil Mapping

Module 5: Soil Spectroscopy

Module 6: Soil Data Sharing

Required downloads

Step-by-step setup instructions

Step 1: Download and extract the GitHub repository

Step 2: Download the Google Drive data

Step 3: Inspect the project folder structure

Step 4: Place Google Drive files into the correct module folders

0.1 What is R?

0.1.1 Why R for Soil Science?

0.2 Installing R and RStudio

0.2.1 Installing R

0.2.2 Installing RStudio

0.2.3 Verifying Installation

0.3 Understanding the RStudio Interface

0.3.1 1. Console Pane (Bottom Left)

0.3.2 2. Code Editor Pane (Top Left)

0.3.3 3. Environment/History Pane (Top Right)

0.3.4 4. Files/Plots/Packages/Help Pane (Bottom Right)

0.4 R Packages: Extending R's Capabilities

0.4.1 What are Packages?

0.4.2 Installing Packages

0.4.2.1 Install Packages from GitHub or Other Sources

0.4.2.2 Manual Installation of R Packages

0.4.3 Loading Packages

0.4.4 Finding Help on Packages

0.5 R Basics: Objects and Data Types

0.5.1 Creating Objects with Assignment

0.5.2 Data Types in R

0.5.2.1 1. Numeric (Numbers)

0.5.2.2 2. Integer (Whole Numbers)

0.5.2.3 3. Character (Text Strings)

0.5.2.4 4. Logical (TRUE/FALSE)

0.5.2.5 5. Dates

0.5.3 Checking Data Types

0.6 Data Structures in R

0.6.1 Vectors

0.6.1.1 Creating Vectors

0.6.1.2 Vector Operations

0.6.1.3 Accessing Vector Elements

0.6.2 Factors

0.6.3 Matrices

0.6.4 Data Frames

0.6.4.1 Creating Data Frames

0.6.4.2 Accessing Data Frame Elements

0.6.4.3 Adding Columns

0.6.5 Lists

0.7 Operators in R

0.7.1 Arithmetic Operators

0.7.2 Comparison Operators

0.7.3 Logical Operators

0.8 Control Structures: Conditional Statements

0.8.1 If-Else Statements

0.8.2 Vectorized If-Else: `ifelse()`

0.8.3 Vectorized Cut: `cut()`

0.9 Control Structures: Loops

0.9.1 For Loops

0.9.2 While Loops

0.10 Functions in R

0.10.1 Using Built-in Functions

0.10.2 Creating Custom Functions

0.10.3 Function Arguments and Defaults

0.11 Data Manipulation with Base R

0.11.1 Subsetting and Filtering

0.11.2 Deleting Columns

0.11.3 Sorting Data

0.11.4 Aggregating Data

0.12 Data Manipulation with the Tidyverse

- 0.12.1 Loading the Tidyverse
- 0.12.2 Tibbles: Modern Data Frames
- 0.12.3 The Pipe Operator: `%>%`
- 0.12.4 Data Manipulation with `{dplyr}`
 - 0.12.4.1 Selecting Columns
 - 0.12.4.2 Filtering Rows
- 0.12.5 Renaming Columns with `rename()`
 - 0.12.5.1 Creating/Modifying Columns with `mutate()`
 - 0.12.5.2 Arranging (Sorting) Data
- 0.12.6 Grouping and Summarizing Data by columns
- 0.13 Combining Data Frames
 - 0.13.1 Understanding Joins
 - 0.13.1.1 Join types
 - 0.13.1.2 Example datasets and join operations
 - 0.13.2 Stacking Data with `bind_rows()`
- 0.14 Missing Data
 - 0.14.1 Identifying Missing Data

0.14.2 Handling Missing Data

0.15 Data Reshaping with {tidyr} package: Pivoting

0.15.0.1 Wide vs Long Data Format

0.15.0.2 `pivot_longer()`: Wide to Long

0.15.0.3 `pivot_wider()`: Long to Wide

0.15.0.4 Advanced Pivoting: Multiple Measurements per Site (Replicates)

0.16 Handling Below Detection Limit Data

0.17 Working with Soil Data

0.17.1 Setting Working Directory

0.17.2 Load packages

0.17.3 Importing Data from Files

0.17.3.1 Reading CSV Files

0.17.3.2 Reading Excel Files

0.17.4 Exploring the Data

0.18 References

Data preparation for Digital Soil Mapping in R

Placeholder

0.19 Introduction

0.20 Loading and exploring raw soil data

0.20.1 Basic data loading and structure

0.20.2 Data components and organization

0.21 Preparing site data for analysis

0.21.1 Adding a unique row identifier

0.21.2 Extracting and standardizing column names for sites

0.21.3 Creating unique profile identifiers

0.21.4 Remove exact duplicate site records

0.22 Coordinate validation and correction

0.22.1 Check 1: Missing coordinates

0.22.2 Check 2: Valid coordinate ranges

0.23 Soil depth validation and correction

0.23.1 Check 1: Missing depth boundaries

0.23.2 Check 2: Negative depth values

0.23.3 Check 3: Zero-thickness intervals

0.23.4 Check 4: Invalid depth logic

0.23.5 Check 5: Profiles without a surface horizon (top > 0)

0.24 Preparing lab data: harmonization and validation

0.24.1 Prepare and standardize laboratory columns

0.24.2 Check 1: Check each property against feasible analytical thresholds

0.24.2.1 Load property thresholds

0.24.2.2 Identify out-of-bounds values

0.24.2.3 Reporting out-of-bounds for properties and audit trail

0.24.3 Check 2: Texture validation

0.24.4 Check 3: Correction of out-of-bounds laboratory values

0.24.4.1 Option 1: Targeted corrections (when error mechanism is known)

0.24.4.2 Option 2: Replace out-of-bounds values with NA

0.25 Resolving duplicated data in soil profiles

0.25.1 Why duplicates occur

0.25.2 Types of duplicate situations

0.25.3 Check 1: Detect potential horizon duplicates within profiles

0.25.4 Resolving Profile Duplicates

0.25.4.1 Check 1: Average duplicated horizons (same depth intervals)

0.25.4.2 Check 2: Resolve ProfID for multiple depth sequences

0.25.5 Remove profiles not starting at the surface

0.25.5.1 Check 1: Remove profiles with incomplete surface coverage

0.25.6 Result

0.26 Harmonizing data for DSM

0.26.1 Depth standardization

0.26.2 Processing standardized data

0.26.3 Exporting standardized data for Digital Soil Mapping

0.27 Preparing data for spectroscopy analyses

0.28 Summary of exported files

References

Sampling design for soil surveys

Placeholder

0.29 Introduction

0.30 Sampling methodologies for soil spatial survey

0.30.1 Selection of the sampling methodology

0.30.2 Inference Methods

0.30.3 Purpose of Sampling

0.30.4 Sampling Design Types

0.30.4.1 Example: Soils4Africa Sampling Design

0.31 SoilFER Sampling Design

0.31.1 Soil Properties

0.31.2 Environmental Covariates

0.31.3 Understanding the Methodology

0.31.3.1 PSU Selection Process

0.31.3.2 SSU and TSU Selection Process

0.31.3.3 Site Identification System

0.32 Tutorial Using R

0.32.1 Setting Up the Environment

0.32.1.1 Install Required Libraries

0.32.2 Define Variables and Parameters

0.32.3 Country ISO Code

0.32.3.1 Land Use Type

0.32.3.2 File Paths

0.32.3.3 Coordinate Reference System

0.32.3.4 Sample Size Calculation

0.32.3.5 Sampling Unit Definitions

0.32.3.6 Algorithm Parameters

0.32.4 Custom Functions

0.32.4.1 Covariate Space Coverage Function (CSIS)

- 0.32.4.2 TSU Generation Function
- 0.32.5 Load Country Boundaries and Legacy Data
 - 0.32.5.1 Visualize Boundaries and Legacy Data
- 0.32.6 Load Environmental Covariates for PSUs
 - 0.32.6.1 Optional Auxiliary Data
 - 0.32.6.2 Load and Process Auxiliary Data
- 0.32.7 Load Main Environmental Covariates
 - 0.32.7.1 Load and Process Soil Climate Data (Newhall)
 - 0.32.7.2 Convert Categorical Variables to Dummy Variables
 - 0.32.7.3 Merge All Covariate Layers
 - 0.32.7.4 Crop and Save Covariates
- 0.33 Principal Component Analysis (PCA)
- 0.34 Organize the Sampling Universe
 - 0.34.1 Load and Process Land Use Data
 - 0.34.2 Exclude Protected Areas
 - 0.34.3 Restrict to Accessible Slopes
 - 0.34.4 Aggregate to 100m Resolution

- 0.34.5 Save Processed Land Use
- 0.34.6 Filter Legacy Data
- 0.35 Create Primary Sampling Units (PSU Grid)**
 - 0.35.1 Generate 2km × 2km Grid
 - 0.35.2 Trim to Country Boundaries
 - 0.35.3 Save PSU Grid
- 0.36 Select PSUs with Sufficient Land Use Coverage**
 - 0.36.1 Calculate Land Use Percentage
 - 0.36.2 Filter PSUs
- 0.37 Rasterize PSUs for Covariate Space Coverage**
- 0.38 Compute Optimal Sample Size**
 - 0.38.1 Run Optimization
 - 0.38.2 Extract Optimal Sample Size
- 0.39 Covariate Space Coverage - Computing PSUs**
 - 0.39.1 Prepare Function Parameters

- 0.39.2 Perform K-means Clustering
- 0.39.3 Assign Clusters and Calculate Distances
- 0.39.4 Extract Selected PSUs
- 0.39.5 Visualize in Covariate Space
- 0.40 Compute SSUs and TSUs
 - 0.40.1 Load High-Resolution Covariates
 - 0.40.2 Process Each PSU
 - 0.40.3 Combine SSUs and TSUs
- 0.41 Export Sampling Units
 - 0.41.1 Create Cluster Raster
 - 0.41.2 Join Cluster Info to TSUs
 - 0.41.3 Create Site IDs
 - 0.41.4 Export Shapefiles
- 0.42 Compute Alternative PSUs
- 0.43 Assign Country-Wide Unique IDs
- 0.44 Export Unified Datasets

0.45 Summary Statistics

0.46 Create Distribution Maps

0.47 GPS Device Preparation

0.47.1 Export to GPX Format

0.47.2 Load onto GPS Device

0.48 Sampling Procedure

0.49 Data Recording

0.50 Quality Control

Appendix A: Software and Data Sources

0.50.1 Software Requirements

0.50.2 Data Sources

Appendix B: Troubleshooting Common Issues

0.50.3 Memory Issues

0.50.4 CRS Misalignment

0.50.5 Empty Geometry

Appendix C: Acronyms and Abbreviations

TBD

Soil Data Preparation

0.51 Introduction

0.52 Data cleaning

0.53 Soil Data Harmonization

0.54 Concepts on digital exchange of soil-related data (ISO 28258)

0.55 Plot data

0.56 Profile data

0.57 Element data

0.58 Specimen data

0.59 Definition of laboratory procedures

0.60 Data transformation

0.61 Organizing soil data according to the GloSIS database

0.62 Tools for harmonization of soil databases within GloSIS

Digital Soil Mapping and Modeling

Placeholder

Part I: Theoretical Foundations

0.63 Fundamentals of Digital Soil Mapping

0.63.1 Introduction to Digital Soil Mapping

0.63.2 The SCORPAN Framework

0.63.3 Statistical Theory for Predictive Soil Mapping

0.63.3.1 Mechanistic versus Empirical Approaches

0.63.3.2 Sources of Residual Variance

0.63.3.3 Covariates as Proxies of Soil-Forming Factors

0.63.3.4 The Universal Model of Soil Variation

0.63.3.5 Extending the Model: Space, Depth, and Time (3D+T)

0.63.4 Types of Soil Variables

0.63.5 Prediction Methods

0.63.6 Nonlinearity and Uncertainty

Part II: Practical Application

0.64 Workflow Overview

0.65 Continuous Modelling (pH)

0.65.1 Nature of the Variable

0.65.2 Data and Covariates

0.65.3 Feature Selection (Boruta)

0.65.4 Model Training

0.65.5 Variable Importance

0.65.6 Validation Metrics

0.65.7 Prediction Maps

0.65.7.1 Mean pH Map

0.65.7.2 Uncertainty Map (Standard Deviation)

0.65.7.3 Coefficient of Variation Map

0.66 Nominal Classification (Clay_pH_Class)

0.66.1 Nature of the Variable

0.66.2 Feature Selection

0.66.3 Model Training

0.66.4 Validation

0.66.5 Prediction Maps

0.66.5.1 Predicted Class Map

0.66.5.2 Maximum Probability Map

0.66.5.3 Uncertainty Map

0.66.5.4 Confident Predictions Map

0.67 Ordinal Modelling (pH_Class)

0.67.1 Nature of the Variable

0.67.2 Ordinal Encoding

0.67.3 Feature Selection

0.67.4 Model Training

0.67.5 Validation

0.67.6 Prediction Maps

0.67.6.1 Latent Values Map

0.67.6.2 Reclassified Ordinal Class Map

0.67.6.3 Standard Deviation Map

0.67.6.4 Class Instability Map

0.68 Area of Applicability (AOA)

0.68.1 Why AOA is Necessary

0.68.2 Dissimilarity Index (DI)

0.68.3 AOA Binary Mask

0.68.4 Local Point Density (LPD)

0.68.5 Predictions Masked to AOA

0.69 Summary

References

Soil Data Sharing

0.70 Data sharing and export formats

0.71 Metadata

0.72 Web Services

References



The Global Soil Partnership (GSP) is a globally recognized mechanism established in 2012. Our mission is to position soils in the Global Agenda through collective action. Our key objectives are to promote Sustainable Soil Management (SSM) and improve soil governance to guarantee healthy and productive soils, and support the provision of essential ecosystem services towards food security and improved nutrition, climate change adaptation and mitigation, and sustainable development.



Australian Government
**Department of Agriculture,
Water and the Environment**



**Ministry of Finance of the
Russian Federation**



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation



**Rural Development
Administration**

