Food and Agriculture Organization of the United Nations

GLOBAL SOIL ORGANIC CARBON Map

TECHNICAL REPORT

GSOCmap v.1.6

itps
INTERGOVERNMENTAL TECHNICAL PANEL ON SOILS

GLOBAL SOIL PARTNERSHIP

Country Guidelines on Digital Soil Mapping

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

ISBN 978-92-5-135899-3

Cover design: ©FAO/Matteo Sala

Contents

Figures

Tables

Foreword

Abbreviations and acronyms

BD Bulk Density

CO_2 Carbon dioxide

CRF Coarse fragments

DM Dry matter

DSM Digital soil mapping

GAUL Global Administrative Unit Layers

GHG Greenhouse gas

GSOCmap Global Soil Organic Carbon Map

GSOCseq Global Soil Organic Carbon Sequestration Potential Map

GSP Global Soil Partnership

HWSD Harmonized World Soil Database

ISCN International Soil Carbon Network

INSII International Network of Soil Information Institutions

IPBES Intergovernmental Platform on Biodiversity and Ecosystem Services

IPCC Intergovernmental Panel on Climate Change

IPR Intellectual Property Rights

ITPS Intergovernmental Technical Panel on Soils

LDN Land Degradation Neutrality

NDVI Normalized difference in vegetation index

NPP Net Primary Production

P4WG Pillar 4 Working Group

QA/QC Quality Assurance/Quality Check

RMSE Root mean square error

SDF Soil Data Facility

SDG Sustainable Development Goals

SISLAC Latin America and the Caribbean's Soil Information System

SOC Soil organic carbon

SOM Soil organic matter

SPADE/M Soil Profile Analytical Database of Europe of Measured Parameters

SWRS Status of World's Soil Resources

UNCCD United Nations Convention to Combat Desertification

WFS Web Feature Service

WoSIS World Soil Information Service

Contributors

Prepared by:
Global Soil Partnership Secretariat
Guillermo Federico Olmedo
Kostiantyn Viatkin
Isabel Luotto
Moritz Mainka
Yusuf Yigini
Ronald Vargas
Mario Guevara Santamaría

Second Intergovernmental Technical Panel on Soils
Luca Montanarella - European Commission, Joint Research Centre (*Chair*); Saéb AbdelHaleem Khresat - Jordan; Isaurinda Dos Santos Baptista Costa - Cape Verde; Sopon Chomchan – Thailand; Juan Antonio Comerma - Venezuela; Talal Darwish - Lebanon; Gunay Erpul - Turkey; Fernando Garcia Préchac - Uruguay; Siosiua Halavatau - Tonga; Oneyda Hernandez Lara - Cuba; Rainer Horn - Germany; Amanullah Khan - Pakistan; Pavel Krasilnikov - Russia; Bhanooduth Lalljee - Mauritius; Botle Mapeshoane - Lesotho; Neil McKenzie - Australia; Maria de Lourdes Mendonca Santos -Brazil; Ahmad S. Muhaimeed - Iraq; Nsalambi V. Nkongolo - Democratic Republic of the Congo; Brajendra Parmar - India; Daniel John Pennock - Canada; Gary Pierzynski - USA; Peter de Ruiter - The Netherlands; Miguel Taboada - Argentina; Kazuyuki Yagi – Japan; Martin Yemefack - Cameroon; Gan Lin Zhang – China

Third Intergovernmental Technical Panel on Soils
Rosa Poch - Spain (*Chair*); Nsalambi V. Nkongolo - Democratic Republic

of the Congo; Matshwene Edwin Moshia III - South Africa; Lydia Mumbi Chabala - Zambia; Générose Nziguhueba - Burundi; Edmond Hien - Burkina Faso; Ashok K. Patra - India; Jin Ke – China; Chencho Norbu – Bhutan; Jun Murase - Japan; Mohammad Jamal Khan - Pakistan; Costanza Calzolari - Italy; Ellen R. Graber - Israel; Peter de Ruiter - The Netherlands ; Alexey Sorokin - Russian Federation; Adalberto Benavides Mendoza - Mexico; Martha Marina Bolanos Benavides - Colombia; Fernando Garcia Préchac - Uruguay; Lucía Helena Cunha dos Anjos - Brazil; Samuel Francke Campana - Chile; Kutaiba M. Hassan - Iraq; Rafla Sahli Epse Attia - Tunisia; Saéb AbdelHaleem Khresat - Jordan; Gary Pierzynski - USA; David Allen Lobb - Canada; Siosiua Halavatau - Tonga; Megan Balks - New Zealand

Chapter 1

Background

1.1 The importance of soil organic carbon

1.2 Objectives of soil organic carbon mapping

This is a citation (Xie, 2015).

1.3 Data policy

1.3.1 Data sharing principles

The GSP Data Policy has been endorsed by partners of the Global Soil Partnership during the 5th GSP Plenary Assembly in June 2017 (GSP & FAO, 2017) in order to promote and govern soil data sharing for data products including GSOCmap contributions, and considering harmonization and interoperability requirements.
The GSP data policy aims to ensure that:

- every existing ownership right to shared soil data are respected;

- the specific level of access and the conditions for data sharing are clearly specified;
- the ownership of each dataset and web service are properly acknowledged and well-referenced; and
- the data owners are protected from any liability arising from the use of their original and/or derived data.

It is recommended that data owners comply with the following open data principles:

1. Accessibility: the data shall be divulged through the Internet (web services).
2. Availability: the data is presented in a convenient, platform-independent, and in line with standard formats (e.g. web feature service WFS).
3. License: the formal concession of the usage and access rights over the data shared.
4. Cost: data shall be shared free of cost, or at no more than a reasonable reproduction cost, preferably by downloading it from the Internet.
5. Re-use and Redistribution: data must be provided and licensed under terms that permit its reuse and redistribution, including intermixing with other datasets.
6. Global benefit: any user must be able to access, use and redistribute data of the Global Soil Information System. However, inherited restrictions by national data policies shall be accepted.
7. Metadata: data describing the products of the Global Soil Information System will by default be open for access.

The data shared by the countries shall contain the relevant soil information representative for the area portrayed. The shared datasets contain the best available information for a given area and topic, however, they are subject to potential restrictions based on the institutions' or countries' data policy.
The data shared by the countries should be quality controlled which means that the data have been technically evaluated to ensure data integrity, correctness, and completeness; errors and omissions are identified and, if possible, addressed.

1.3.2 Ownership, data rights and citation

In the case of original data, the rightful data owner keeps full ownership of it. All intellectual property rights (IPR) and copyrights pertaining to the data owner remain intact and are respected by the soil data facility (SDF) host. All data providers must communicate to the SDI host their IPR and data use policies. Thus, the ownership of all data made available through the GSP soil portal need to be clearly specified. This is an important prerequisite to allow this data to be accessible through the soil SDF.

In the case of derived data, the deriving institution becomes the rightful owner. However, all original data must be accredited and correctly cited. According to the Pillar 4 Implementation Plan, each global-level derived GSP data product will be quality-assured by the Pillar 4 Working Group. This includes agreements about the correct citation.

The data owner shall ensure that the data shared can be used and interpreted by the authorized users in general; this includes providing the proper citations, as well as providing information over the ownership of such data for acknowledgement purposes. Users shall acknowledge the source of data provided through the Global Soil Information System.

All providers of original data (data owners) are responsible to define and clarify the IPR and licensing. Any user of this data, such as the SDF host, has to respect the national data policies and/or licensing involved with the retrieval of the respective web services. In the case of data provided to the central repository, a bilateral agreement/license may be required (between the national data owner and SDF host), depending on and in conformity with national rules.

More information about the data policy can be accessed at GSP & FAO (2017).

Chapter 2

Setting-up the software environment

Y. Yigini

This cookbook focuses on SOC modeling using open source digital mapping tools. The instructions in this chapter will guide user through installing and manually configuring the software to be used for DSM procedures for Microsoft Windows desktop platform. Instructions for the other platforms (e.g. Linux Flavours, MacOS) can be found through free online resources.

2.1 Use of R, RStudio and R Packages

R is a language and environment for statistical computing. It provides a wide variety of statistical (e.g. linear modeling, statistical tests, time-series, classification, clustering, etc.) and graphical methods, and is highly extensible.

2.1.1 Obtaining and installing R

Installation files and instructions can be downloaded from the Comprehensive R Archive Network (CRAN).

1. Go to the following link https://cloud.r-project.org/index.html to download and install **R**.
2. Pick an installation file for your platform.

2.1.2 Obtaining and installing RStudio

Beginners will find it very hard to start using **R** because it has no Graphical User Interface (GUI). There are some GUIs which offer some of the functionality of **R**. **RStudio** makes **R** easier to use. It includes a code editor, debugging and visualization tools. Similar steps need to be followed to install **RStudio**.

1. Go to https://www.rstudio.com/products/rstudio/download/ to download and install **RStudio**'s open source edition.
2. On the download page, *RStudio Desktop, Open Source License* option should be selected.
3. Pick an installation file for your platform.

2.1.3 Getting started with R

- **R** manuals: http://cran.r-project.org/manuals.html
- Contributed documentation: http://cran.r-project.org/other-docs.html
- Quick-**R**: http://www.statmethods.net/index.html
- Stackoverflow **R** community: https://stackoverflow.com/questions/tagged/r

2.2 R packages

When you download **R**, you get the basic **R** system which implements the **R** language. **R** becomes more useful with the large collection of packages that extend the basic functionality of it. **R** packages are developed by the **R** community.

2.2.1 Finding R packages

The primary source for **R** packages is CRAN's official website, where currently about 12,000 available packages are listed. For spatial applications, various packages are available. You can obtain information about the available packages directly on CRAN with the `available.packages()` function. The function returns a matrix of details corresponding to packages currently available at one or more repositories. An easier way to browse the list of packages is using the *Task Views* link, which groups together packages related to a given topic.

2.2.2 Most used R packages for digital soil mapping

As was previously mentioned, **R** is extensible through packages. **R** packages are collections of **R** functions, data, documentation and compiled code easy to share with others. In the following Subsections, we are going to present the most used packages related to digital soil property mapping.

2.2.2.1 Soil science and pedometrics

aqp: Algorithms for quantitative pedology. A collection of algorithms related to modeling of soil resources, soil classification, soil profile aggregation, and visualization.

GSIF: Global Soil Information Facility (GSIF). Tools, functions and sample datasets for digital soil mapping. GSIF tools (standards and functions) and sample datasets for global soil mapping.

ithir: A collection of functions and algorithms specific to pedometrics. The package was developed by Brendan Malone at the University of Sydney.

soiltexture: The *Soil Texture Wizard* is a set of **R** functions designed to produce texture triangles (also called texture plots, texture diagrams, texture ternary plots), classify and transform soil textures data. These functions virtually allow to plot any soil texture triangle (classification) into any triangle geometry (isosceles, right-angled triangles, etc.). The set of functions is expected to be useful to people using soil textures data from different soil texture classification or different particle size systems. Many (> 15) texture triangles from all around the world are predefined in the package. A simple text-based GUI is provided: `soiltexture_gui()`.

2.2.2.2 Spatial analysis

raster: Reading, writing, manipulating, analyzing and modeling of gridded spatial data. The package implements basic and high-level functions, processing of very large files is supported.

rgdal: Provides bindings to Frank Warmerdam's Geospatial Data Abstraction Library (GDAL).

RSAGA: The package provides access to geocomputing and terrain analysis functions of SAGA GIS from within **R** by running the command line version of System for Automated Geoscientific Analyses (SAGA).

sf: The package provides an improvement on **sp** and **rgdal** for spatial datasets. It is using the newly Simple Feratures (SF) standard which is widely implemented in spatial databases (PostGIS, ESRI ArcGIS), and forms the vector data basis for libraries such as GDAL and web standards such as GeoJSON (http://geojson.org/).

sp: The package provides classes and methods for spatial data. The classes document where the spatial location information resides, for 2D or 3D data.

2.2.2.3 Modelling

automap: This package performs an automatic interpolation by automatically estimating the variogram and then calling **gstat**.

caret: Extensive range of functions for training and plotting classification and regression models.

Cubist: Regression modeling using rules with added instance-based corrections. Cubist models were developed by Ross Quinlan.

C5.0: C5.0 decision trees and rule-based models for pattern recognition. Another model structure developed by Ross Quinlan.

gam: Functions for fitting and working with generalized additive models.

gstat: Variogram modeling with simple, ordinary and universal point or block (co)kriging, sequential Gaussian or indicator (co)simulation. The package includes variogram and variogram map plotting utility functions.

nnet: Software for feed-forward neural networks with a single hidden layer, and for multinomial log-linear models.

2.2.2.4 Mapping and plotting

Both **raster** and **sp** have handy functions for plotting spatial data. The following packages can be used as functional extentions.

ggplot2: Besides using the base plotting functionality, this is another useful plotting package.

leaflet: Create and customize interactive maps using the Leaflet JavaScript library and the **htmlwidgets** package. These maps can be used directly from the **R** console, from **RStudio**, in **Shiny** apps and **RMarkdown** documents.

plotKML: Writes sp-class, spacetime-class, raster-class and similar spatial and spatiotemporal objects to KML following some basic cartographic rules.

2.2.3 Packages used in this cookbook

The authors of this cookbook used a number of different **R** packages. All required packages used in the cookbook can be installed using the following code and the `install.packages()` function when starting a new SOC mapping project. Alternatively, the code for the installation of the needed packages is included at the beginning of each Chapter.

```
# Install all required R packages used in the cookbook
install.packages(c("aqp", "automap", "car", "caret", "e1071",
                   "GSIF", "htmlwidgets", "leaflet", "mapview",
                   "Metrics", "openair", "plotKML", "psych",
                   "quantregForest", "randomForest", "raster",
                   "rasterVis", "reshape", "rgdal", "RSQLite",
                   "snow", "soiltexture", "sf", "sp"))
```

2.3 R and spatial data

R has a large and growing number of spatial data packages. We recommend taking a quick browse on **R**'s official website to see the spatial packages available: http://cran.r-project.org/web/views/Spatial.html.

2.3.1 Reading shapefiles

The Environmental Systems Research Institute (ESRI) provides a shapefile format SHP which is widely used for storing vector-based spatial data (i.e., points, lines, polygons). This example demonstrates use of **raster** package that provides functions for reading and/or writing shapefiles.

```
library(raster)
# Load the soil map from a shapefile *.shp file
soilmap <- shapefile("MK_soilmap_simple.shp")
```

We may want to use these data in other GIS environments such as ArcGIS, QGIS, SAGA GIS, etc. This means we need to export the `SpatialPointsDataFrame` to an appropriate spatial data format such as a shapefile `*.shp`.

```
# For example, we can select the soil units classified as
# fluvisols according to WRB
Fluvisols <- soilmap[soilmap$WRB == "Fluvisol",]

# Save this as a new shapefile
shapefile(Fluvisols, filename = 'results/fluvisols.shp',
          overwrite = TRUE)
```

2.3.2 Coordinate reference systems in R

We need to define the Coordinate Reference System (CRS) to be able to perform any sort of spatial analysis in **R**. To clearly tell **R** this information we define the CRS which describes a reference system in a way understood by the PROJ.4 projection library. Find more information at this link: http://trac.osgeo.org/proj/.

An interface to the PROJ.4 library is available in the **rgdal** package. An alternative to using PROJ.4 character strings, we can use the corresponding yet simpler EPSG (European Petroleum Survey Group) code. **rgdal** also recognizes these codes. If you are unsure of the PROJ.4 or EPSG code for the spatial data that you have but know the CRS, you should consult http://spatialreference.org/ for assistance.

```
# Print the CRS for the object soilmap
soilmap@proj4string
```

The following example shows how a spatial object from a (comma-separated values) CSV file (`*.csv`) can be created. We can use the `coordinates()` function from the **sp** package to define which columns in the data frame refer to actual spatial coordinates. Here, the coordinates are listed in columns X and Y.

```
# Load the table with the soil observations site information
dat_sites <- read.csv(file = "data/site-level.csv")

# Convert from table to spatial points object
coordinates(dat_sites) <- ~ X + Y

# Check the coordinate system
dat_sites@proj4string

# As the CRS is not defined, we can assign the correct CRS as we
# have information about it. In this case, it should be EPSG:4326
dat_sites@proj4string <- CRS("+init=epsg:4326")

# Check the CRS again
dat_sites@proj4string
```

2.3.3 Working with rasters

Most of the functions for handling raster data are available in the **raster** package. There are functions for reading and writing raster files from and to different formats. In DSM, we mostly work with data in table format and then rasterize this data so that we can produce a continuous map. For doing this in **R** environment, we will load raster data in a data frame. This data is a digital elevation model (DEM) provided by the International Soil Reference and Information Centre (ISRIC) for FYROM.

```
# For handling raster data, we load raster package
library(raster)
```

```
# Load DEM from the raster *.tif files
DEM <- raster("covs/DEMENV5.tif")
```

We may want to export this raster to a suitable format to work in a standard GIS environment. See the help file for writing a raster `?writeRaster` to get information regarding the supported grid types that data can be exported into. Here, we will export our raster to ESRI ASCII (American Standard Code for Information Interchange), as it is a common and universal raster format.

We may also want to export our DEM to KML (Keyhole Markup Language) file (`*.kml`) using the `KML()` function. `KML()` is a handy function from the **raster** package for exporting grids to KML format. Note that we need to re-project the data to the World Geodetic System (WGS) WGS84 geographic. The raster re-projection is performed using the `projectRaster()` function. Look at the help file `?projectRaster` for more information.

2.4 Other DSM software and tools

- **GRASS GIS**: Available at https://grass.osgeo.org/ (free and open source).
- **SAGA GIS**: Available at https://sourceforge.net/projects/saga-gis/files/ (free and open source).
- **QGIS**: Available at http://www.qgis.org/en/site/forusers/download.html (free and open source).

Chapter 3

Product specifications

3.1 Generic target specification

3.2 Metadata specifications

Chapter 4

Data collection and processing

4.1 Different scenarios of country driven action

4.1.1 Delivery of the maps produced by the countries

4.1.1.1 Data submission form

4.1.2 Joint efforts

4.1.3 GSP gapfilling

4.1.3.1 Spatial modeling using publicly available data

4.1.3.2 Using publicly available SOC stock maps

4.2 Data processing and compilation of the GSOCmap

Chapter 5

Blocks

5.1 Equations

Here is an equation.

$$f\left(k\right) = \binom{n}{k} p^k \left(1 - p\right)^{n-k} \tag{5.1}$$

You may refer to using `\@ref(eq:binom)`, like see Equation (7.1).

5.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref(thm:tri)`, for example, check out this smart theorem 7.1.

Theorem 5.1. *For a right triangle, if c denotes the* length *of the hypotenuse and a and b denote the lengths of the* ***other*** *two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html.

5.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html

Chapter 6

Blocks

6.1 Equations

Here is an equation.

$$f\left(k\right) = \binom{n}{k} p^k \left(1 - p\right)^{n-k} \tag{6.1}$$

You may refer to using `\@ref(eq:binom)`, like see Equation (7.1).

6.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref(thm:tri)`, for example, check out this smart theorem 7.1.

Theorem 6.1. *For a right triangle, if c denotes the* length *of the hypotenuse and a and b denote the lengths of the* ***other*** *two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html.

6.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html

Chapter 7

Blocks

7.1 Equations

Here is an equation.

$$f\left(k\right) = \binom{n}{k} p^k \left(1 - p\right)^{n-k} \tag{7.1}$$

You may refer to using `\@ref(eq:binom)`, like see Equation (7.1).

7.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref(thm:tri)`, for example, check out this smart theorem 7.1.

Theorem 7.1. *For a right triangle, if c denotes the* length *of the hypotenuse and a and b denote the lengths of the* ***other*** *two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html.

7.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html

References

Xie, Y. 2015. *Dynamic documents with R and knitr*. 2nd edition. Boca Raton, Florida, Chapman; Hall/CRC. (also available at http://yihui.org/knitr/).

Thanks to the financial support of

European Commission

Ministry of Economic Affairs of the Netherlands

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Ministry of Finance of the Russian Federation

ifa
International Fertilizer Industry Association

ISBN 978-92-5-135899-3

9 789251 358993

CB9015EN/1/03.22