ISRIC
World Soil Information

DSM 2025
Workshop: Evaluating Digital Soil Maps by their patterns

David G. Rossiter & Laura Poggio

ISRIC–World Soil Information

20-January-2025

Workshop objectives

1. Become familiar with the concept of **“spatial patterns”**
2. Become familiar with statistical methods to **describe** spatial patterns
3. Become familiar with concepts of **aggregation** and **segmentation**
4. Become familiar with **R packages** for pattern description, comparison and aggregation
5. *Optional* Become familiar with the `geopat2` Unix program for pattern-based segmentation
6. Consider what it means for a DSM product to represent **“true” spatial patterns** of soil properties and map units at different cartographic and categorical scales.

Workshop structure

1. Brief **presentation** of the problem and some approaches to addressing it
2. Tutorial **setup**: packages, directories, input maps
3. Exercise and discussion: Characterizing the **pattern of single maps: continuous**
4. Exercise and discussion: Characterizing the **pattern of single maps: classified**
5. Exercise and discussion: **Aggregation** by the `supercells` package
6. *Optional* Exercise and discussion: **Pattern-based segmentation** with the `geopat` Unix program
7. Discussion! the relation to **'reality'**

Preparation

Prior to the workshop:

1. Make sure R and R Markdown are installed and functioning properly.
2. Install the required packages; you can do this by `source`ing the script `InstallWorkshopPackages_2025.R`
3. Decide on a data source (see next).
   - If you decide to use your own data, prepare the GeoTIFF as explained next
   - If you decide to use the example SoilGrids tile, it will be provided at the workshop.
4. *Optional* For segmentation, Install GeoPAT from `https://github.com/Nowosad/geopat2`

Data sources

Option 1 **Preferred** Bring your own gridded map of soil properties, in an area where you are familiar with the “true” soil landscape pattern

- Pick an area with contrasting patterns of values, not necessarily of the range of values
- Size $\approx 480 \times 480$ pixels, covering about 1° x 1°
- Format as a multi-layer `GeoTIFF`.

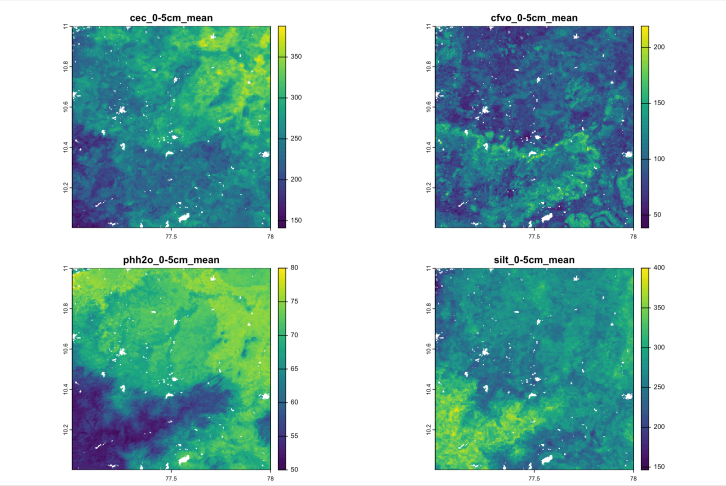
Option 2 Prepare a multi-layer 1° x 1°`GeoTIFF` of your area of interest, from SoilGrids v2.0 (see next slide)

Option 3 Use the prepared multi-layer 1° x 1°`GeoTIFF` of an area in Tamil Nadu, from SoilGrids v2.0 (supplied at the workshop)

Option 2: Preparing a `GeoTIFF` from SoilGrids

1. Load the R Markdown source `SoilGrids250_MakeRasterStack.Rmd` into **RStudio**
2. In Section “Directories” specify the **target directory** on your system for the GeoTIFF
3. Load `GetTiles.R` into RStudio
4. Modify the variables specifying lower-right corner latitude and longitude to specify your **area of interest**
5. **Select the properties and depth slices** which will make up your raster stack
   - one property, all depth slices; or all properties, one depth slice; or selected properties, all depth slices
6. **Select the appropriate code block and run it**. This will create the GeoTIFF.

GeoTIFF example



Tutorial document

Source `PatternAnalysisWorkshopTutorial_DSM2025.qmd`

- A Quarto[1] source notebook – similar to R Markdown[2]

Compiled HTML and PDF versions, using the Tamil Nadu example

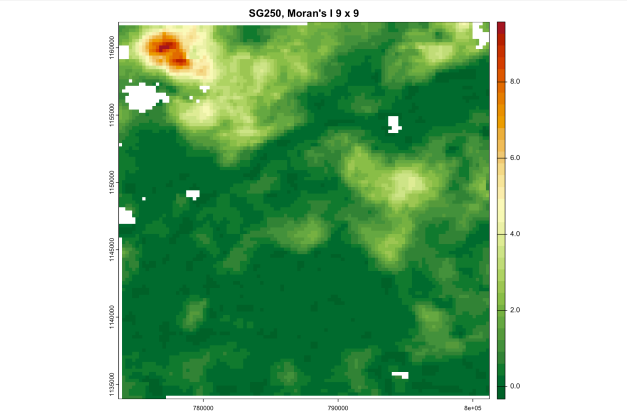
[1] `https://quarto.org/`

[2] `https://rmarkdown.rstudio.com/`

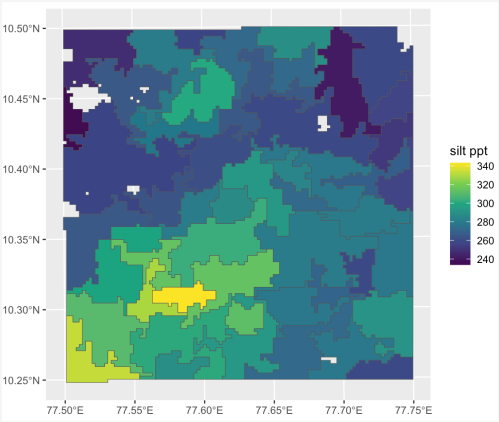
Working through the “Pattern Analysis” tutorial

1. Open the **tutorial source** in RStudio..
2. Open one of the **compiled documents** in a web browser (HTML) or PDF viewer (PDF), for reference.
3. Modify the source at heading `DSM product to evaluate` to specify the path and map files on your system.
4. Work through the source document by **running code chunks in sequence**, examining/interpreting the output.
    - Toolbar dropdown menu `Run`: either `Run Current Chunk` or `Run Next Chunk`
    - or, in the source code, click the “down” arrow at the right of the chunk header.
5. Feel free to **expand/adapt**!
6. Stopping points for **discussion** after each topic

Moving window local spatial autocorrelation example



Supercells example

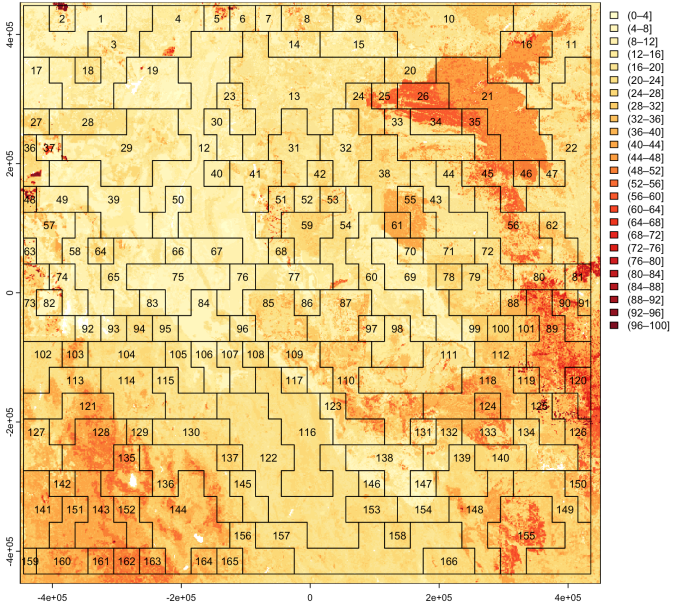


Discussion points (after “Patterns” tutorial)

- Which landscape-level metrics, seem most useful in characterizing the map? In determining if the map is “realistic”?
- How successful was aggregation in finding relatively homogeneous landscape units? Which parameters were optimal and why?
- How to communicate pattern analysis and supercells results to map users?
- Do we really know the “true” soil pattern at various scales?

Optional: Segmentation

- There will likely not be enough time for this, but just in case …and you can practice at home.
- At coarser scales, homogeneity of properties may not be possible or even desirable. This has led to the concept of **landscape segments**, defined by the **co-occurrence pattern of contrasting grid cells** of some pre-defined size.
- The **GeoPAT** suite is a set of **stand-alone Unix programs**
- These can be invoked in **sequence** to obtain a **segmentation** and an **evaluation of its quality**.
- We hope that the segments divide the soil landscape into units with homogeneous **internal patterns**.


Shift: 20000 m
(0–4]
(4–8]
(8–12]
(12–16]
(16–20]
(20–24]
(24–28]
(28–32]
(32–36]
(36–40]
(40–44]
(44–48]
(48–52]
(52–56]
(56–60]
(60–64]
(64–68]
(68–72]
(72–76]
(76–80]
(80–84]
(84–88]
(88–92]
(92–96]
(96–100]
4e+05
2e+05
0
-2e+05
-4e+05
-4e+05
-2e+05
0
2e+05
4e+05

Optional: Segmentation

1. The GeoPAT Unix executables must be installed on your system.
2. Load Quarto source `SegmentingSoilMaps.qmd` into RStudio.
3. Adjust the path to the GeoPAT program (line 113) for your system.
4. Adjust the path to the input files (line 126)
5. Render this file, or `Run All`. This will write a set of functions to `SegmentingSoilMaps_Functions.RData`
6. Download some area of the Global Soil Organic Carbon map, see script `PrepareGSOC.R` – adjust to your area.
7. Load Quarto document `SegmentingSoilMaps_CaseStudy_GSOC.qmd`.
8. Adjust the paths and areas and render or `Run All`

Discussion points (after “Segmentation”)

- Which grid size best segmented the landscape into groups of patterns?
- Do you recognized areas with known internal landscape patterns?
- How to communicate segmentation analysis results to map users?