

# Seedling: A User-Friendly, Scalable, and Collaborative Digital Soil Mapping Workflow

Giulio Genova, Bas Kempen, Betony Colman, Diana Collazos, Maria Ruiperez Gonzalez, Laura Poggio

ISRIC – World Soil Information



# DSM workflow - Motivation

- Digital Soil Mapping (DSM) workflows have common features
- Modular structure to ease maintenance (functions and code library)
- Need-based tools and methods

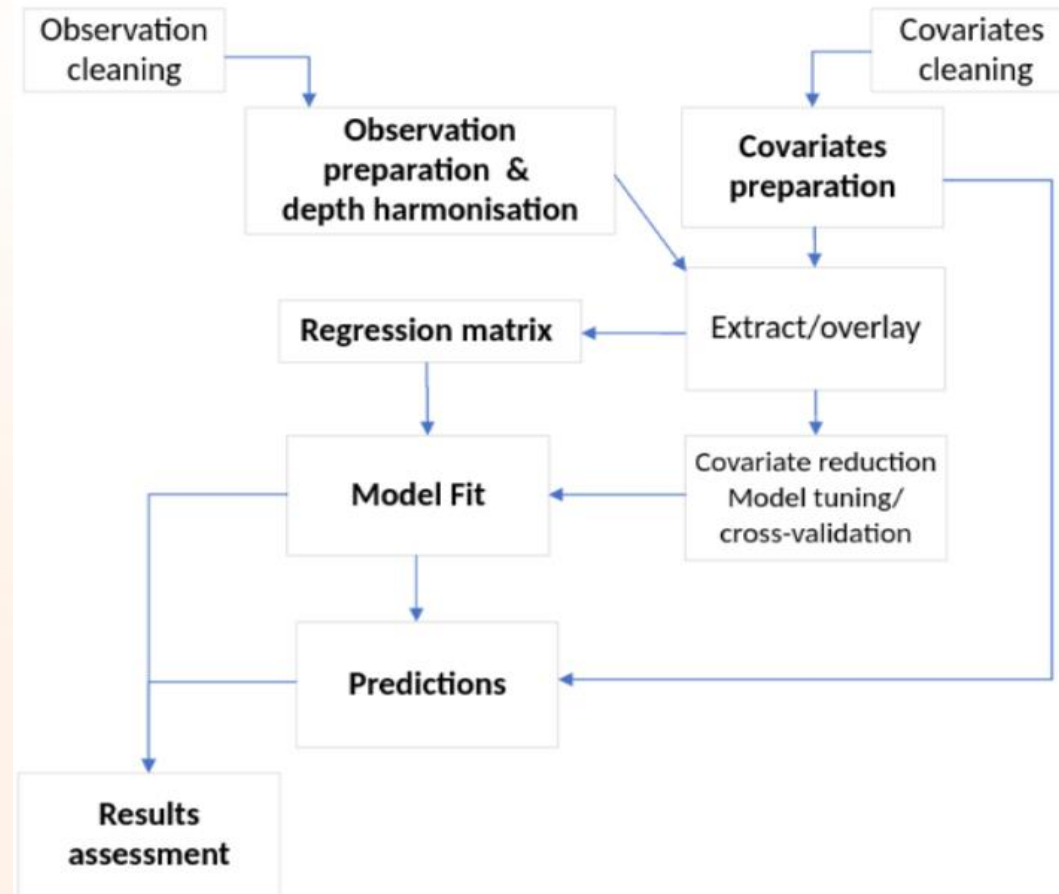


Operational tool that facilitates the process



# DSM workflow - Motivation

- DSM workflow Common features



# DSM workflow - Seedling

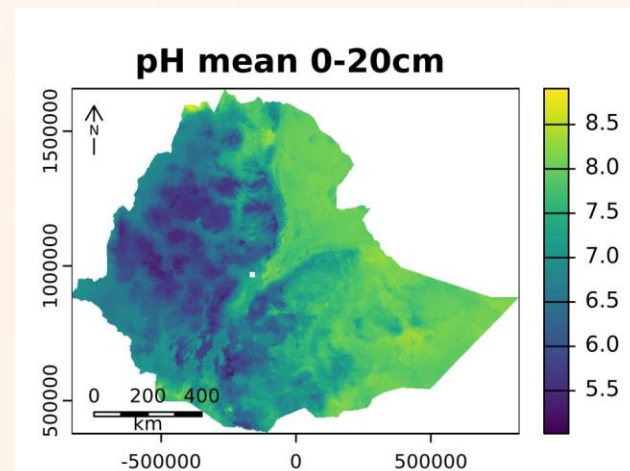
## Main characteristics

- Completely R based
- Easy to Use for people with limited DSM expertise
- Multiplatform (Windows, Linux, Mac)
- Running on single laptop/workstation (and on HPC, cloud-computing, etc.)



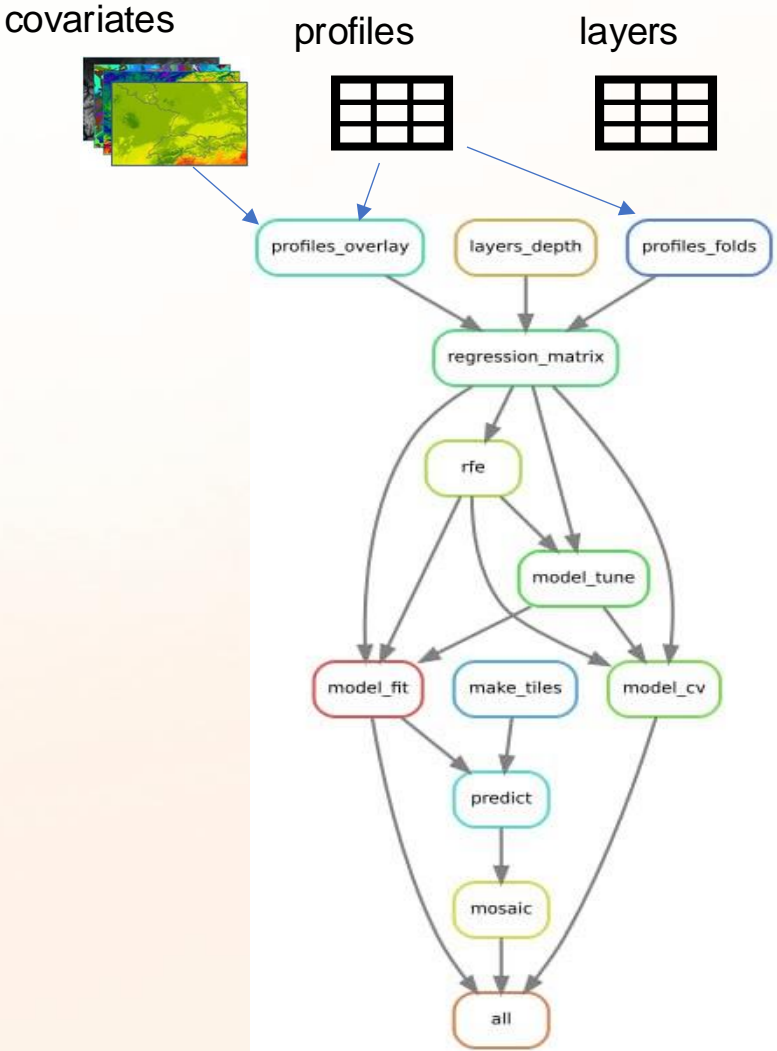
# Seedling - What is it

- What is it
  - A modular software stack to perform DSM. R package and configuration files
  - An operational tool
  - Based on SoilGrids workflow (but not limited to)
  - Starts with **standardised inputs**
  - Produces: Maps, Accuracy stats, Plots, and more
- What is it **not**
  - A data preparation tool
  - A training tool



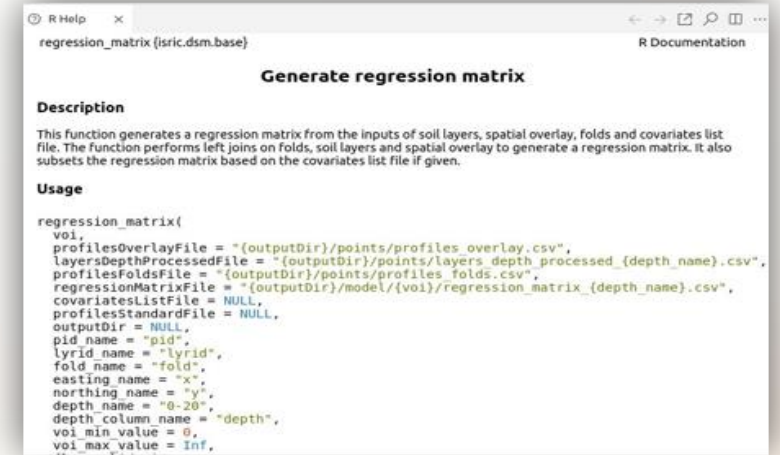
# Seedling - Main features

- What does it do
  - 2D and 3D modelling
  - Depth standardisation: weighted average of layers (for 2D)
  - Recursive feature elimination
  - Hyperparameter tuning
  - Accuracy (Cross validation) – Uncertainty maps
  - Parallel predictions with tiles
  
- Licence: GPL v3



# Seedling - Documentation

- R package documentation {isric.dsm.base}
- The online documentation contains:
  - How to install the workflow and its dependencies
  - How to setup and run the workflow
  - Additional information on modelling and design choices
  - The Licence



R Help x regression\_matrix (isric.dsm.base) R Documentation

### Generate regression matrix

**Description**

This function generates a regression matrix from the inputs of soil layers, spatial overlay, folds and covariates list file. The function performs left joins on folds, soil layers and spatial overlay to generate a regression matrix. It also subsets the regression matrix based on the covariates list file if given.

**Usage**

```
regression_matrix(
  vol,
  profilesOverlayFile = "{outputDir}/points/profiles_overlay.csv",
  layersDepthProcessedFile = "{outputDir}/points/layers_depth_processed_{depth_name}.csv",
  profilesFoldsFile = "{outputDir}/points/profiles_folds.csv",
  regressionMatrixFile = "{outputDir}/model/{vol}/regression_matrix_{depth_name}.csv",
  covariatesListFile = NULL,
  profilesStandardFile = NULL,
  outputDir = NULL,
  pid_name = "pid",
  lyrId_name = "lyrid",
  fold_name = "fold",
  easting_name = "x",
  northing_name = "y",
  depth_name = "0-20",
  depth_column_name = "depth",
  voi_min_value = 0,
  voi_max_value = Inf,
```



DSM Workflow Documentation

Table of contents  
Welcome

**DSM Workflow Documentation**

AUTHOR: ISRIC - World Soil Information  
PUBLISHED: 2023-06-16

**Welcome**

This is a Digital Soil Mapping workflow developed at ISRIC.

Starting from standardised inputs, it organizes data for modelling, performs modelling and predicts soil maps as final output.

It is a file-based workflow, meaning that each step has files as both inputs and outputs. Files are named and organised according to the main variables/options chosen by the user.

At this stage this workflow can only handle continuous properties.

Start with [Chapter 1](#) and follow the instructions to:

- Install ([Section 1.1](#)), Where we give installation instructions for the workflow and required dependencies
- Prepare the input data ([Section 1.2](#)), Where we explain how the input data should be prepared



# Seedling - Usage

To use the workflow, you will need two files

- Configuration file: where all the options are defined
- Workflow script: runs the steps of the workflow

```

# Input/Output files definition
outputDir: "./output/"
profilesStandardFile: "./input_example/points/geul_profiles_4326.csv"
layersStandardFile: "./input_example/points/geul_layers.csv"
covarsDir: "./input_example/covariates"
maskFile: "./input_example/other/geul_mask.tif"

# The variable of interest
voi: "pb"

voi_parameters:
  voi_min_value: 0
  voi_max_value: Inf
  transformation: "notransform"

# Dataset variables
pid_name: "pid"
lyrid_name: "lyrid"
easting_name: "x"
northing_name: "y"
top_layer_name: "top"
bottom_layer_name: "bottom"
depth_column_name: "depth"
fold_name: "fold"
stratum_name: "stratum"

# Model parameters
depth_name: "0-20"
depths_3D:
  - "0, 5, 15, 30, 60, 100, 200"
use_rfe: "dorfe"
do_tune: "notune" # "tune"
  
```

```

library(isric.dsm.base)
library(argparser)

config_location <- "config.yaml"

p <- arg_parser("Run the entire workflow")
p <- add_argument(
  parser = p, arg = "--config_file",
  help = "configuration file location", default = config_location
)
config <- yaml::read_yaml(parse_args(p)$config_file, eval.expr = TRUE)

# ## Prepare

profiles_overlay(
  profilesStandardFile = config$profilesStandardFile,
  covarsDir = config$covarsDir,
  profilesOverlayFile = config$profilesOverlayFile,
  covariatesListFile = config$covariatesListFile,
  outputDir = config$outputDir,
  easting_name = config$easting_name,
  northing_name = config$northing_name,
  crs_profiles = config$crs_profiles,
  pid_name = config$pid_name,
  covariate_pattern = config$covariate_pattern, overwrite = config$overwrite
)

layers_depth(
  layersStandardFile = config$layersStandardFile,
  layersDepthProcessedFile = config$layersDepthProcessedFile,
  outputDir = config$outputDir,
  pid_name = config$pid_name,
  lyrid_name = config$lyrid_name,
  
```





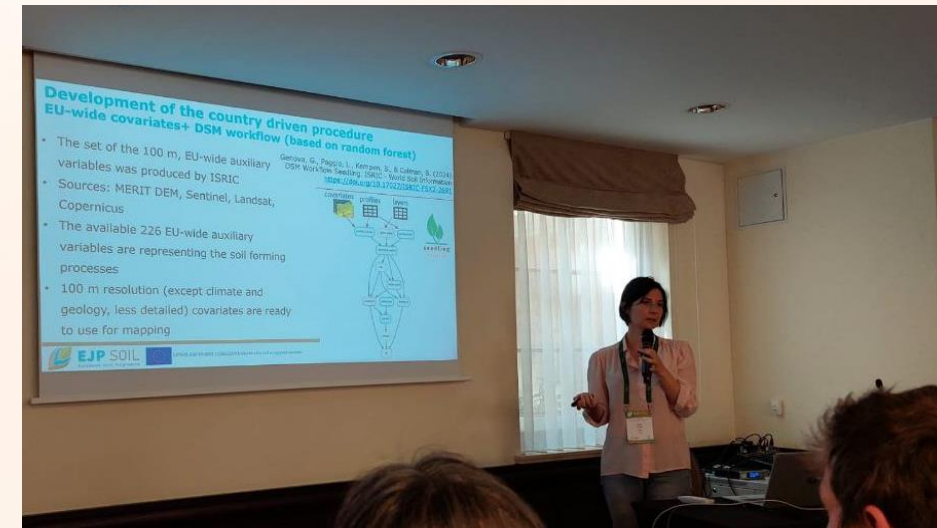
# Users and impact

Several users in different countries

- ISRIC's (internal projects/ production pipelines)
- DSM trainings. Rwanda, Kenya, Ethiopia, Nepal
- EJP soil. All national partners (Europe)
- PhD projects
- Other research groups non necessarily involved in projects
- Growing community?



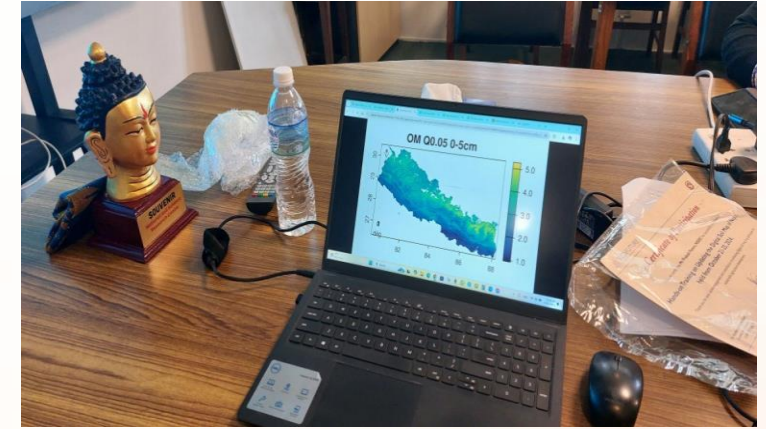
Rwanda – map comparison exercise



# Users and impact

Several users in different countries

- ISRIC's (internal projects/ production pipelines)
- DSM trainings. Rwanda, Kenya, Ethiopia, Nepal
- EJP soil. All national partners (Europe)
- PhD projects
- Other research groups non necessarily involved in projects
- Growing community?



# Conclusions

- Easy to use and scalable tool to lower the bar to enter DSM
- Evolving and improving
- Open source, community involvement?
- Reach out if you want to know more and get involved

<https://git.wur.nl/isric/dsm-general/dsm.workflows/seedling>





# *Thank You*

*giulio.genova@isric.org*

Event Digital Partner

**Event**   
A Unit of Krishi Junction

