Compare Predictive Soil Mapping (PSM) products for the USA

D G Rossiter d.g.rossiter@cornell.edu

04-July-2021

Contents

Objective

This document explains the procedures to evaluate of maps of soil properties made by Predictive Soil Mapping (PSM) in the USA. It refers to a set of R Markdown scripts for the various steps of the procedure.

Data sources

There are several sources of soil property information for the USA:

- gNATSGO, entire CONUS;
- gSSURGO, by State;

- Global Soil Map v0.5, further abbreviated as *GSMv05*;
- Intermediate-scale gridded soil property and interpretation maps from averaged and aggregated SSURGO and STATSGO data, further abbreviated as *ISSR-800*;
- POLARIS Soil Properties;
- SoiLGrids250 from ISRIC further abbreviated as *SG2*;
- Soil Properties and Class 100m Grids USA, further abbreviated as *SPCG100USA*;
- LandGIS from the private company EnvirometriX.

SG2, POLARIS and LandGIS also give uncertainty estimates as the 5/95% quantiles, and gNATSGO/gSSURGO give low and high estimated values, as well as representative value.

Here is a brief explanation of each of these:

- *gNATSGO* is complete coverage of the best available soils information for all areas of the United States and Island Territories. It was created by combining data from the Soil Survey Geographic Database (SSURGO), State Soil Geographic Database (STATSGO2), and Raster Soil Survey Databases (RSS) offsite link image into a single seamless ESRI file geodatabase", here. Resolution is 30~m. Also available via Web Coverage Services (WCS).

  The map units often have multiple components and properties must be extracted by weighted aggregation from the relational database.

- *gSSURGO* is a set of statewide polygon maps with the associated SSURGO relational database that can be rasterized. ESRI file geodatabases; also available as 10~m rasterized coverages, here.

  The map units often have multiple components and properties must be extracted by weighted aggregation from the relational database.

- *Global Soil Map v0.5* is the USA's contribution to the GlobalSoilMap.net project, Gbased on Soil Survey Geographic Database (SSURGO), at nominal 100 m grid resolution.

  Partially explained in Hempel, J., Libohova, Z., Odgers, N., Thompson, J., Smith, C. A., & Lelyk, G. (2012). Versioning of GlobalSoilMap.net raster property maps for the North American Node. In Digital Soil Assessments and Beyond: Proceedings of the Fifth Global Workshop on Digital Soil Mapping 2012, Sydney, Australia (pp. 429–433). CRC Press. https://doi.org/10.13140/2.1.4670.2404.

  A promised 1.0+ is supposed to use a DSM approach from NCSS-KSSL point observations, i.e., a conventional DSM approach.

- *ISSR-800* "Intermediate-scale gridded soil property and interpretation maps from averaged and aggregated SSURGO and STATSGO data" at 800 m resolution.

- *POLARIS* is the result of harmonizing diverse SSURGO and STATSGO polygon data with the DSMART algorithm to produce a raster soil series map (30 m resolution) and then extracting property information from Official Series Descriptions.

  It is explained in: Chaney, N. W., Minasny, B., Herman, J. D., Nauman, T. W., Brungard, C. W., Morgan, C. L. S., McBratney, A. B., Wood, E. F., & Yimam, Y. (2019). POLARIS soil properties: 30-m probabilistic maps of soil properties over the contiguous United States. Water Resources Research, 55(4), 2916–2938. Scopus. https://doi.org/10.1029/2018WR022797

- *SoilGrids* is a system for global digital soil mapping that makes use of global soil profile information and covariate data to model the spatial distribution of soil properties across the globe. SoilGrids250 is a collections of soil property maps at six standard depths at 250 m grid resolution.

  SoilGrids250 filenames, procedures etc. are explained in a FAQ.

  The choice of the Goode Homolosine projection is explained in Moreira de Sousa, L., L. Poggio, and B. Kempen. 2019. Comparison of FOSS4G Supported Equal-Area Projections Using Discrete Distortion Indicatrices. ISPRS International Journal of Geo-Information 8(8): 351. https://doi.org/10.3390/ijgi8080351.

- *SPCG100US* is an adaption of the SoilGrids methodology for the USA, using the NCSS Characterization Database, the National Soil Information System (NASIS), and the Rapid Carbon Assessment (RaCA) point datasets and some covariates only available for the USA, giving a 100 m grid resolution.

  Explained in Ramcharan, A., T. Hengl, T. Nauman, C. Brungard, S. Waltman, et al. 2018. Soil property and class maps of the conterminous United States at 100-meter spatial resolution. Soil Science Society of America Journal 82(1): 186–201. https://doi.org/10.2136/sssaj2017.04.0122. This source was able to use 87 parent material classes four and drainage classes based on the representative soil components of gSSURGO map units.

- *LandGIS* follows the approach of SoilGrids V1 (ISRIC), with more points, additional covariates, and ensemble machine-learning methods.

Depth slices

- *gNATSGO*, *gSSURGO* gives values by genetic horizon, with no standard intervals, however they are served as WCS with any depth interval, by depth-weighted averaging over the requested depth slice.
- *GSMv05*, *SoilGrids250* and *POLARIS* predict as average values over the GlobalSoilMap.net specification depth slices: 0-5, 5-15, 15-30, 30-60, 60-100, and 100-200 cm. In fact SoilGrids250 (and POLARIS?) predict at the midpoint of these intervals but report the values as if they cover the interval.
- *SPCG100US* makes predictions for 7 standard soil depths (0, 5, 15, 30, 60, 100 and 200 cm). These are considered "point" predictions, i.e., the values for an infintesimal vertical extent at these depth points.
- *ISSR-800* can be produced for any depth slice, as a thickness-weighted average based on gSSURGO.
- *LandGIS* makes point predictions at 6 standard soil depths (0, 10 (not 5 or 15), 30, 60, 100 and 200 cm).

Download locations

See import scripts for details.

- gNATSGO, gSSURGO – accessd via Web Coverage Services (WCS)
- Global Soil Map v0.5 – limited access, US Government site; also here;

- SoilGrids250 web converage service;
- ISSR-800 Soil Properties;
- POLARIS soil properties;
- SPCG100US;
- LandGIS download via Zenodo.

Soil properties

Each download script explains which soil properties are included, how they are named, and their units of measurement.

Step 1: Import maps

Each product has an R Markdown file to import the product, in subdirectory `./scripts_importmaps`:

Regional maps:

- `gNATSGO_WCS_import.Rmd` – accessed via Web Coverage Services
- `GSM_USA_V05_import.Rmd`
- `ISSR_import.Rmd`
- `POLARIS_import.Rmd`

- `SoilGrids250_import.Rmd`
- `SPCG100USA_import.Rmd`
- `LandGIS_import.Rmd`

Local maps (30 m resolution):

- `gSSURGO_WCS_import.Rmd` – accessed via Web Coverage Services
- `POLARIS_import.Rmd`

To compare products, adjust each of these to the area of interest (AOI), property of interest, and the depth slice of interest, and import the product. To compare uncertainties also specify the quantile of interest, for those products that provide it. These should be adjusted in the YAML header.

The YAML headers include the default parameters and look like:

```
---
params:
   lrc_long: -76
   lrc_lat: 42
   size: 1
   voi.n: 4
   quantile.n: NA
   depth.n: 4
---
```

These can be overridden with the `params` argument in a call to `rmarkdown::render`.

The script `Import_all_regional.R` can be used to import one or more regional products, setting the parameters there:

```r
import.all <- function(lrc_long, lrc_lat,
                       size=1,
                       voi.n,
                       depth.n=1,
                       quantile.n=4,
                       which=1:6) {
  library(knitr)
  psm.list <- c(
    "gNATSGO_WCS",
    "GSM_USA_V05",
    "SoilGrids250",
    "POLARIS",
    "SPCG100USA",
    "LandGIS",
    "ISSR8")
  for (psm in psm.list[which]) {
    print(paste("Importing", psm))
    param.list <- list(lrc_long=lrc_long,
                       lrc_lat=lrc_lat,
                       size=size,
                       voi.n=voi.n,
                       depth.n=depth.n,
                       quantile.n=quantile.n)

    out.file.name <- paste0(lrc_long, '_', lrc_lat, '_',
                            voi.n, '_', quantile.n, '_',
                            depth.n, '.html')
```

```
    rmarkdown::render(paste0(psm, "_import.Rmd"),
                      params = param.list,
                      output_file = out.file.name,
                      envir = new.env(parent = globalenv())
                      )
  }
}
```

This function can be called as follows, for example:

```
import.all(lrc_long=-120,
           lrc_lat=38,
           size=1,
           voi.n=2,
           depth.n=2,
           quantile.n=4,
           which=c(3))
```

And of course the call to `import.all` can be put with a `for` statement to import several properties, depths, AOI at once.

The AOI is usually a $1 \times 1°$ tile in geographic coordinates, with integer corners. This is because POLARIS is provided in these tiles, and also because this is a large enough area to see the regional differences between products.

Two products provide uncertainty estimates: SoilGrids250 and POLARIS. These scripts also allow the choice of a quantile, as well as the mean prediction.

Step 2: Comparisons

A set of scripts in subdirectory `./scripts_compare` take the imported tiles and perform various comparisons.

Regional properties (250 m spatial resolution)

First is the direct comparison of the mapped property, in script `./scripts_compare/Compare_regional.Rmd`.

This script exports harmonized regional maps (same CRS, same resolution, same units, same area), which can then be compared for their spatial patterns by script `./scripts_compare/Compare_regional_patterns.Rmd`.

Regional uncertainty (250 m spatial resolution)

Two products provide prediction quantiles at 5, 50 and 95%, as well as the mean: SoilGrids250 and POLARIS. gNATSGO provides estimates of the low and high ends of the property range, with no probabilistic implication. These can be compared with script `./scripts_compare/Compare_regional_uncertainty.Rmd`. These quantiles and low/representative/high values must have been previously imported.

Regional pattern analysis (250 m spatial resolution)

Script `./scripts_compare/Compare_regional_patterns.Rmd` allows you to specify a sub-area of the full 1-degree tile to be classified by histogram equalization over all the products to compare, and then compares the patterns of the classified maps.

This script must follow script `Compare_regional.Rmd`, i.e., comparing the regional maps.

Local properties (30 m spatial resolution)

First is the direct comparison of the mapped property, in script `./scripts_compare/Compare_local.Rmd`.

This specifies an AOI within a $1 \times 1°$ tile, generally $0.15 \times 0.15°$, as an offset from the lower-right corner of the full tile.

This script exports harmonized regional maps (same CRS, same resolution, same units, same area), which can then be compared for their spatial patterns by script `./scripts_compare/Compare_local_patterns.Rmd`.

Local uncertainty (30 m spatial resolution)

Only POLARIS gives prediction quantiles at 5, 50 and 95%, as well as the mean. There is no other product with which to compare it.

Local pattern analysis

Script `./scripts_compare/Compare_local_patterns.Rmd` uses the sub-tile processed by `Compare_local.Rmd`. It classifies the selected propeerty by histogram equalization over all the products to compare, and then compares the patterns of the classified maps.

This script must follow script `Compare_local.Rmd`, i.e., comparing the local maps and selecting the sub-tile.

Ground overlays

Another way to qualitatively evaluate digital products is to produce ground overlays for Google Earth, for display within Google Earth and expert evaluation.

These overlays can be displayed along with SoilWeb Apps, including SoilWeb Earth which displays SSURGO (detailed) or NATSGO (generalized) soil map units in Google Earth.

There is one R Markdown file for each product, in subdirectory `scripts_ground_overlays`. These all use the native resolution, resampled to CRS EPSG:4326 if necessary, to match Google Earth.

Parameters

Each import script is parameterized via the YAML header, see this explanation.

In the script, if the parameters are defined either by calling `knitr::render` with parameters, or rendering directly in R Studio, these parameters take precedence over those defined directly in the script.

Here are the lists to which the parameters refer:

'# Property of interest {#voi}

Property names the different products. Note that all except GSM v0.5 and SoilGrids250 are missing one or more properties. Such products will be omitted from comparisons for those properties.

Properties that be compared are: clay, silt, sand weight concentrations; pH in 1:1 water; CEC; SOC; bulk density of the fine earth; coarse fragment volume proportion. See the import script for each product for a link to a description of the properties and their units of measure.

```
voi.list.gnatsgo <- c("claytotal_r", "silttotal_r", "sandtotal_r",
                   "ph1to1h2o_r", "ecec_r", "soc_r",
                   "dbthirdbar_lt2mm_r", "gravel_r")
voi.list.sg <- c("clay", "silt", "sand", "phh2o", "cec", "soc", "bdod", "cfvo")
voi.list.gsm <- c("claytotal_r_g_kg", "silttotal_r_g_kg", "sandtotal_r_g_kg",
                   "ph1to1h2o_r_ions_pHx10", "ecec_r_cmolc_kg", "soc_r_mr_g_gF",
                   "dbthirdbar_lt2mm_r_g_cm3", "gravel_r_vol_ratio_m3_m3")
```

```
voi.list.issr8 <- c("clay", "silt","sand","ph", "cec", "", "", "")
voi.list.polaris <- c("clay", "silt", "sand", "ph", "", "om", "bd", "")
voi.list.psu <- c("clay", "", "sand", "ph_h2o", "", "soc", "bd", "")
voi.list.landgis <- c("clay.wfraction_usda.3a1a1a",
                      "silt.wfraction_usda.3a1a1a",
                      "sand.wfraction_usda.3a1a1a",
                      "ph.h2o_usda.4c1a2a",
                      "",
                      "organic.carbon_usda.6a1c",
                      "bulkdens.fineearth_usda.4a1h",
                      "coarsefrag.vfraction_usda_3b1")
```

Quantile

Two products have quantiles as well as the mean prediction: SoilGrids and POLARIS.

```
quantile.list.sg <- c("Q0.05", "Q0.5", "Q0.95", "mean")
quantile.list.polaris <- c("p5", "p50", "p95", "mean")
```

Depth of interest

Depth slice names in the different products:

```
depth.list.sg <- c("0-5", "5-15", "15-30", "30-60", "60-100", "100-200")
# SPCGUSA100 predicts at points, these were averaged to GSM slices during import
# --- these have the SG names
# LandGIS predicts at points, these were averaged to GSM slices during import
# --- these have the SG names
depth.list.polaris <- gsub("-", "_", depth.list.sg)
depth.list.issr8 <- gsub("-", "", depth.list.sg)
depth.list.gsm <- c("000_005", "005_015", "015_030", "030_060", "060_100", "100_200")
```

Area of Interest (AOI)

The usual import is a $1 \times 1°$ tile, because that is how POLARIS data is served. Specify the *lower-right corner* using WGS84 geographic coordinates, and a tile size in degrees (by default 1°). These are then used compute the upper-right corner 1 degree west and north.

```
tile.lrc <- c(-77, 35)   # lower-right corner
tile.size <- 1
```