How well does Predictive Soil Mapping represent soil geography? An investigation from the USA Supplementary material

David G. Rossiter^{1, 2}, Laura Poggio¹, Dylan Beaudette³, and Zamir Libohova⁴ ¹ISRIC-World Soil Information, Postbus 353, Wageningen 6700 AJ, NL ²Section of Soil & Crop Sciences, New York State College of Agriculture and Life Sciences, 233 Emerson Hall, Cornell University, Ithaca NY 14853 USA ³USDA-NRCS Soil Survey Division, 19777 Greenely Rd., Sonora, CA, 95370 USA ⁴USDA-NRCS-National Soil Survey Center, 100 Centennial Mall North, Room 152, Lincoln, NE 68508-3866 USA

Correspondence: David G. Rossiter (david.rossiter@isric.org)

Copyright statement. Creative Commons Attribution 4.0 License

1 Introduction

This is a supplement to the journal article "How well does Predictive Soil Mapping represent soil geography? An investigation from the USA". The Predictive Soil Mapping (PSM) products discussed are POLARIS soil properties (Chaney et al., 2019)

5 (abbreviation PSP), SoilGrids v2.0 (abbreviation SG2) (Poggio et al., 2021), and Soil Properties and Class 100m Grids of the United States (abbreviation SPCG) (Ramcharan et al., 2018). The reference product is gSSURGO.

2 Data sources

2.1 gSSURGO

gSSURGO is delivered as gridded coverages at 30 m horizontal resolution on an Albers Equal Area projection covering the 10 CONUS, with standard parallels at 29.5° and 45.5° N and the central meridian at -96° E on the NAD83 datum, which uses the GRS80 ellipsoid. gNATSGO is a 90 m resolution generalization of this. Access to gSSURGO is via a Web Coverage Service (WCS) interface provided in the soilDB R package (Beaudette et al., 2021b). Each $1 \times 1^{\circ}$ tile requires about 63 Mb for the 30 m gridded map and 610 Kb for the associated data tables. Aggregation to the chosen Global Soil Map depth slice is by weighted averaging over the horizons or layers in the slice using the slab function of the aqp "Algorithms for Quantitative

15 Pedology" R package (Beaudette et al., 2013, 2021a). These are then combined over the map unit components by weighted averaging, using the estimated component proportions for the map unit which the grid cell represents. The resulting tile is about 12 Mb.

The transfer from unrectified photos to topographic base has not always been flawless, and in addition polygons may have been mis-drawn on the original survey. Fig. 1 shows an example of such discrepancies on a map of Tompkins County, NY,

20

design scale 1:24 000. SMU Ab (alluvial soils) includes the lower hillside and does not cover much of the alluvial plain. SMU Ro (rock outcrop) includes the alluvial area at the junction of two streams. SMU MaB (Mardin channery silt loam, 2–8% slopes) does not properly delimit this slope class. These map units are indeed the correct ones for the soil landscape, but for a given 30 or 250 m pixel the mapped soil may be incorrect.



Figure 1. SoilWeb view of SSURGO map units. Point inquiry at $-76.4861^{\circ}\text{E}, 42.2986^{\circ}\text{N}$. Dimensions 660×720 m

Since gNATSGO and gSSURGO are compiled from diverse field surveys over many years, in some areas there are artefacts of that survey process. The left panel of Fig. 2 shows a particularly egregious example: not only is the E-W State boundary between Missouri (N) and Arkansas (S) obvious, but also the survey of the eastmost Missouri county in this figure (Ripley) shows an abrupt, artificial soil boundary with Oregon county, to its west. Contrast this with the predictions by SG2 in the right panel of Fig. 2, which because it is based on globally-consistent covariates does not show such artefacts. PSP (center panel of Fig. 2) partly solves this problem by its harmonization algorithm.

30 2.2 SoilGrids v2.0

SG2 (SoilGrids team, 2021b) predicts at 250 m resolution for the standard depth slices specified by GlobalSoilMap on an equalarea Interrupted Goode Homolosine (IGH) projection on the WGS84 datum (Moreira de Sousa et al., 2019). User-specified tiles for a property, depth slice, and quantile (5%, 50%, 95%, mean) are downloaded from the SoilGrids data store (SoilGrids team, 2021a) using the Web Distributed Authoring and Versioning (WebDAV) protocol, as a Virtual Raster Table (VRT) from

35 the SG2, as explained by de Sousa (2020). This is then converted to the tile using the gdal_translate function of the rgdal R package (Bivand et al., 2021).



Figure 2. Maps of clay concentration % in the 30-60 cm depth slice, Missouri-Arkansas border; lower-right corner near Pocahontas AR. (left) gNATSGO; (center) PSP; (right) SG2

2.3 Soil Properties and Class 100m Grids of the United States

SPCG predicts at 100 m resolution for seven point depths (0, 5, 15, 30, 60, 100 and 200 cm) in the same projection as gSSURGO. Each property and depth is a separate grid covering the CONUS, about 370 Mb. These are downloaded from the ScholarSphere (PennState University Library) data repository (Ramcharan et al., 2017).

2.4 POLARIS

40

45

POLARIS is supplied as $1 \times 1^{\circ}$ tiles, one for each property, Global Soil Map depth slice, and quantile (5%, 50%, 95%, mean). The grid is 1 arc-second of longitude and latitude resolution, i.e., 0.0002777778° on the WGS84 datum, equivalent to ≈ 32 m latitude, and proportionally smaller longitude depending on latitude. These are downloaded from the POLARIS repository (Chaney, 2019) and need no preprocessing. Each tile is 8–35 Mb, depending on the complexity of the spatial pattern.

3 Matching

The four maps are made compatible for regional comparison by (1) resampling to 250 m to match SoilGrids v2.0, using the resample function of the terra R package (Hijmans et al., 2021; Hijmans, 2021); (2) masking areas not predicted by one or more products (e.g., urban areas, water); (3) trimming to the minimum area covered by all the products.

50 For local comparison, both gSSURGO and POLARIS are already in nominal 30 m resolution, but in different coordinate reference systems. To make them compatible, gSSURGO is resampled into the POLARIS grid, and then both are resampled into the UTM projection covering the study area.

4 Environmental covariates and geographic scope

The three PSM products (SG2, PSP, and SPCG) use a large number of gridded GIS coverages as the basis of their predictive models, as explained in the respective papers.

SG2 uses the following global covariates, i.e., available over the entire area to be predicted.

- Climate: temperature, precipitation, snowfall, cloud cover, solar radiation, wind speed.
- Ecology: bioclimatic zones and ecophysiographic regions.
- Geology: soil and sedimentary thickness; rock types.
- 60 Land Use/Cover: from sources such as ESA, NASA and USGS.
 - Elevation and terrain morphology: including numerous morphology indexes and land form classes.
 - Vegetation Indexes: NDVI, EVI and NPP. Raw bands from Landsat and Modis products
 - Hydrography: global water table, inundation and glaciers extent and surface water change.

PSP uses a similar set, and also coarse-resolution (≈ 2 km) estimates of U, Th, and K γ-ray decay products. SPCG uses
a similar set, but also parent material classes (87) and drainage classes (4) covering the CONUS. These are derived from the Official Series Descriptions of the dominant component of the SSURGO polygon at the grid cell centre.

5 Example area and soil property

The example area is in central New York State, bounding box (-77 – -76° E), (42–43° N); the subtile for pattern evaluation is (-76.8 – -76.6° E), (42.2–42.4° N) The regional geomorphology is described by Bloom (2018). The underlying bedrock
ro is a sedimentary sequence from Ordovician (north) to upper Devonian (south), with a wide variety of sedimentary facies. A north-south transect of the bedrock geology map (New York State Geological Survey, 1970) is shown in Fig. 3. This shows a chronological and topographic sequence from Upper Silurian (N) through Upper Devonian (S) sedimentary rocks, notably the Onondaga limestone (green "Don") and Tully limestone (crosshatched red, "Dt"), which were spread southward as glacial till.

The entire area has been glaciated, but the portion north of about 42° 15' (Valley Heads terminal moraines) somewhat more

- 75 recently than the southern portion. Many glacial features are present: ground moraine, deep glacial troughs with proglacial lake sediments, beach lines, outwash valley trains, kame terraces and hanging deltas. Soil reaction in the northern half is largely controlled by limestone spread by the glacier from outcrops of the Onondaga and Tully limestones, decreasing to the south. A fragment of the surficial geology map (New York State Geological Survey, 1986) is shown in Fig. 4. This shows the strongly-expressed glacial features ground moraine (pink), proglacial lakes (brown), organic swamps (green), bedrock or very
- 80 thin soil cover (red), till moraine (purple), kame moraines (yellow), and outwash sand and gravel (yellow). These features are well-known to the field soil surveyors.



Figure 3. Bedrock geology of Central New York State, transect from $N-43^{\circ} N$ (left) – 42° , centred on -76° 30' E. Orientation N (left) to S (right). Source: (New York State Geological Survey, 1970)



Figure 4. Surficial geology of Central New York State near Moravia NY. Legend: ground moraine (pink; if stippled shallow over bedrock), proglacial lakes (brown), organic swamps (dark green), bedrock or very thin soil cover (red), till moraine (purple), kame moraines (orange), lacustrine sand (light green), outwash sand and gravel (yellow). Source: (New York State Geological Survey, 1986)

6 Regional spatial patterns

6.1 Uncertainty

85

Fig. 5 shows the differences between the IQR of the PSM products and the low-high range from gNATSGO. Note that each figure has its own stretch. Both PSM products almost everywhere have substantially wider ranges than gNATSGO, however the pattern of differences is not similar. For example, the difference with SG250 is much larger in the north of the study area, whereas PSP has the larger differences in the southern hills.



Figure 5. IQR/range 5/95% vs. low/high, POLARIS-gNATSGO (left), SG250-gNATSGO (right), pHx10, 0-5 cm

7 Local spatial patterns

A second example of POLARIS disaggregation is shown in Fig. 6. This is the predicted silt concentration within \approx 3 ha field

- 90 and some adjacent woodland, all within the Chenango series alluvial fan, south of the intersection of Robinson Hollow Road and NY State Route 79 in Tioga County. Values range from 37% (darkest red) to 55% (lightest red), a range with significant management implications. These values come from the constituents listed for this map unit. The named series Chenango is assigned 75% of the area, with a surface soil of 39.7% silt concentration. The other five inclusions have different predicted silt concentrations. However, in this field, there seems to be no justification to map any of these inclusions. For example, the Tioga
- series inclusion (5%) is found on higher positions of flood plains, and the Middlebury series inclusion (5%) is found in recent alluvium, but this field is all on the alluvial fan terrace. This disaggregation is clearly not based on land use, and there is no terrain or parent material differentiation in this almost flat field. There would be no basis for differential management of each $\approx 700 m^2$ grid cell, as is implied by the fine resolution and strong differentiation within the field.



Figure 6. Ground overlay from PSP in the Chenango gravelly loam map unit, silt % 0-5 cm. Centre of image $-76^{\circ}17'03$ "E, $42^{\circ}22'40$ "N

Copyright statement. Creative Commons Attribution 4.0 License

100 *Author contributions.* DGR conceptualized the approach, did most of the writing, wrote the R Markdown documents and performed the example case study. LP provided PSM expertise and detailed knowledge of SG2. DB and LZ provided USA-specific expertise, in particular about the NRCS and its products and services. All authors collaborated on the motivation, methods and conclusions.

Competing interests. There are no competing interests.

References

120

- 105 Beaudette, D., Roudier, P., and Brown, A.: Algorithms for Quantitative Pedology. R Package Version 1.29, ncss-tech, https://github.com/ ncss-tech/aqp, 2021a.
 - Beaudette, D., Skovlin, J., Roecker, S., and Brown, A.: soilDB: Soil Database Interface. R Package Version 2.6.3., ncss-tech, https://github. com/ncss-tech/soilDB, 2021b.

Beaudette, D. E., Roudier, P., and O'Geen, A. T.: Algorithms for Quantitative Pedology: A Toolkit for Soil Scientists, Computers & Geo-

110 sciences, 52, 258–268, https://doi.org/10.1016/j.cageo.2012.10.020, 2013.

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., Baston, D., Rouault, E., Warmerdam, F., Ooms, J., and Rundel, C.: Rgdal: Bindings for the 'Geospatial' Data Abstraction Library, https://CRAN.R-project.org/package=rgdal, 2021.

Bloom, A. L.: Gorges History: Landscapes and Geology of the Finger Lakes Region, Paleontological Research Institution, Ithaca, New York, 2018.

- 115 Chaney, N.: Index of /POLARIS/PROPERTIES/v1.0, http://hydrology.cee.duke.edu/POLARIS/PROPERTIES/v1.0/, 2019.
 - Chaney, N., Minasny, B., Herman, J., Nauman, T., Brungard, C., Morgan, C., McBratney, A., Wood, E., and Yimam, Y.: POLARIS Soil Properties: 30-m Probabilistic Maps of Soil Properties over the Contiguous United States, Water Resources Research, 55, 2916–2938, https://doi.org/10.1029/2018WR022797, 2019.

- Hijmans, R.: The Terra Package R Spatial, https://rspatial.org/terra/pkg/index.html, 2021.
- Hijmans, R. J., Bivand, R., Forner, K., Ooms, J., and Pebesma, E.: Terra: Spatial Data Analysis, https://CRAN.R-project.org/package=terra, 2021.
- Moreira de Sousa, L., Poggio, L., and Kempen, B.: Comparison of FOSS4G Supported Equal-Area Projections Using Discrete Distortion Indicatrices, ISPRS International Journal of Geo-Information, 8, 351, https://doi.org/10.3390/ijgi8080351, 2019.
 - New York State Geological Survey: Geologic Map of New York, New York State Geological Survey, Albany, NY, http://www.nysm.nysed. gov/research-collections/geology/gis, 1970.

New York State Geological Survey: Surficial Geologic Map of New York, New York State Geological Survey, Albany, NY, http://www. nysm.nysed.gov/research-collections/geology/gis, 1986.

- 130 Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.
 - Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., and Thompson, J.: SoilProperties and Class 100m Grids United States, https://doi.org/10.18113/S1KW2H, 2017.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., and Thompson, J.: Soil Property and Class Maps

135 of the Conterminous United States at 100-Meter Spatial Resolution, Soil Science Society of America Journal, 82, 186–201, https://doi.org/10.2136/sssaj2017.04.0122, 2018.

SoilGrids team: SoilGrids, https://files.isric.org/soilgrids/latest/data/, 2021a. SoilGrids team: FAO SoilGrids, https://www.isric.org/explore/soilgrids/faq-soilgrids, 2021b.

de Sousa, L. M.: WebDAV: Direct Access to the Maps in VRT Format, https://git.wur.nl/isric/soilgrids/soilgrids.notebooks/-/blob/master/ markdown/webdav_from_R.md, 2020.