Use of Rapid Assessment of U.S. Soil Carbon Dataset to Calibrate a Surrogate Century Model

Yushu Xia[1], Hoyoung Kwon[2], and Michelle Wander[1]
[1] Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, U.S.A.
[2] Environmental & Production Technology Division, International Food Policy Research Institute, Washington, D.C., U.S.A.

Introduction

- National scale models that underpin estimates of soil health and productivity are typically run using models calibrated and validated at the point scale.
- Soil survey data that has wide spatial coverage and fine resolution provides an opportunity for more detailed model calibration. The quality of the data influences our ability to calibrate models to adequately represent soil and environmental conditions across the country.
- Use of a digital soil mapping approach may allow us to improve soil inventory quality by considering land use history and identifying explanatory factors used for up-scaling procedures.

Objectives

- Compare SOC stock estimates for cropland based on the widely used Unified North American Soil Map (UNASM) (Liu et al., 2013) with those based on the newer Rapid Assessment of U.S. Soil Carbon (RaCA) (Wills et al., 2013).
- Refine the RaCA dataset to include only lands that have been under long-term crop production and, compare SOC estimates using the dataset before and after the refinement.
- Use environmental covariates to scale up the refined RaCA dataset to county-level. Investigate the impacts of using county-level UNASM and RaCA databases for model calibration on SOC estimates.

Databases and Model of the Study

UNASM database

- UNASM database combines information from the Digital General Soil Map of the US (STATSGO2), Soil Landscape of Canada (SLCs) databases, and the Harmonized World Soil Database (HWSD v1.21).
- Variables, including SOC are reported for the topsoil (0-30 cm) and the subsoil layers (30-100 cm) after interpolation to a spatial resolution of 0.25 × 0.25 degrees. The measurements were taken over a period of more than 30 years (ca. 1970s to 2000s) and mixed methods including wet chemistry and dry combustion were used to quantify SOC.

RaCA database

- RaCA database compiled dry combustion measurements of SOC taken from both 0-30 cm and 0-100 cm across 6000 sites in the U.S. between 2010 and 2013 in a unified effort designed to assay stocks.

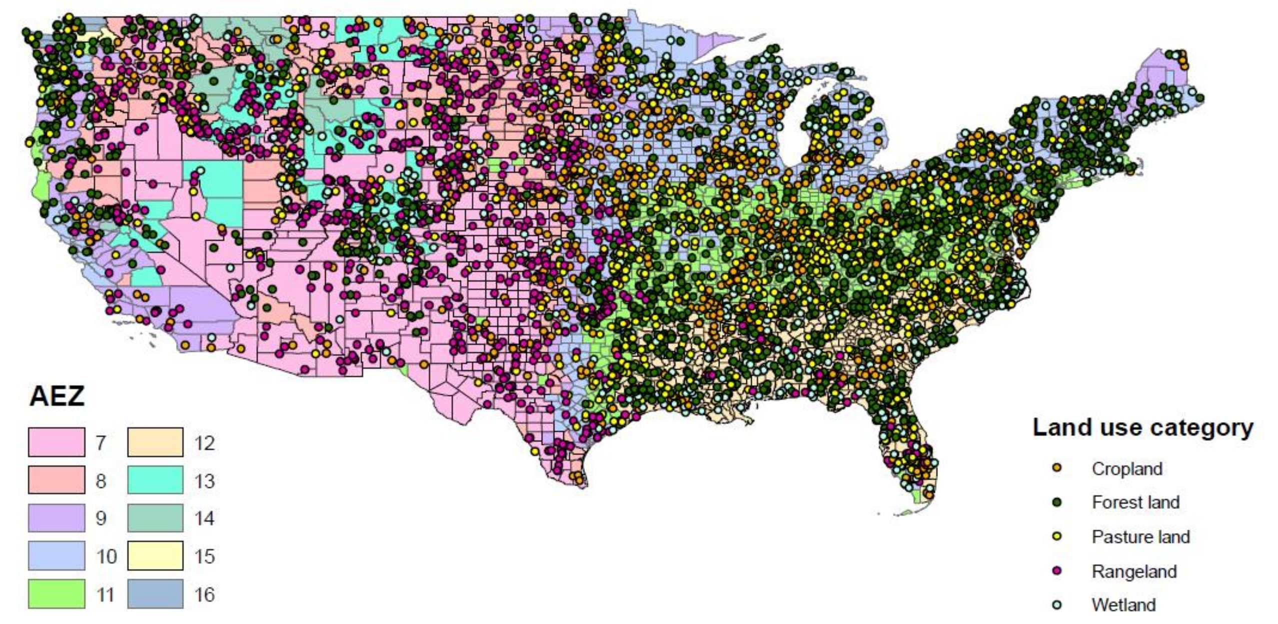

Fig. 1 RaCA sampling points across the conterminous U.S. grouped by land use category and Agro-ecosystem zone (AEZ).

Refined RaCA database

- 1000 RaCA measurement points labeled as 'cropland' were further refined to include only those locations with cultivation of main crops consisting of corn, soybean, small grains, or cover crops that had been cropped since 1992. Areas under cultivation were identified using USDA Cropland Data Layer (CDL), and areas managed as cropland since 1992 were identified using the National Land Cover Database (NLCD).

The surrogate Century (SCSOC) model

- SCSOC was developed by Kwon and Hudson (2010) using an inverse modeling approach to calibrate CENTURY's SOC sub-model within SAS for application to U.S. croplands (Kwon et al. 2017). It is decoupled from other sub-models and uses NASS observed yields to estimate residue return. Rate adjustments within the decoupled model currently reflect the influences of climate, dominant soil-type, and management factors described at the county-level.

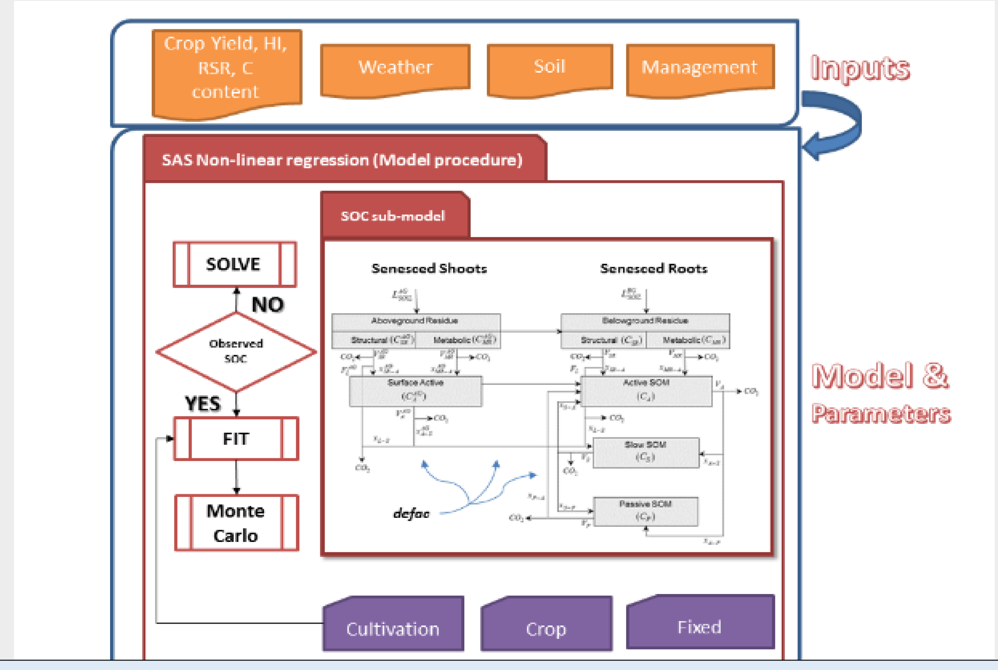

Fig. 2 SCSOC's structure, data inputs, and nonlinear regression (Model) procedure performed in SAS statistical software. Adapted from Kwon et al. (2017).

Scaling and Calibration

Scale up refined RaCA database

- Elevation, slope, soil bulk density (BD), soil pH, soil texture, temperature, and precipitation were extracted for refined RaCA data points, and for UNASM points gridded at 0.25 × 0.25 degrees across the U.S. Elevation and slope were extracted from U.S. Geological Survey, soil information was extracted from UNASM, and temperature and precipitation were extracted from the Climatic Research Unit Time Series map (Harris et al., 2013).
- Regression kriging method was used to estimate SOC of the gridded points with extracted environmental covariates. Then, point results within each county were averaged to get county average RaCA SOC estimations as the interpolated county-level RaCA database.

Model calibration and estimation

- SCSOC was initially 'FIT' in SAS against UNASM SOC observations to calibrate: (1) the temporally variable that reflects the influence of soil temperature and moisture on SOM decay, and (2) the initial SOC stocks and their distributions of SOM fractions at the start of the simulation. The fitted model was then used to estimate U.S. county-level SOC stocks that were averaged by AEZ for evaluation.
- U.S. county-level average SOC stocks were again estimated using SCSOC model calibrated with the interpolated RaCA database for the topsoil layer (0-30 cm) of cropped lands.

Results and Discussions

Compare UNASM with RaCA

Table 1. Summary of cropland topsoil (0-30 cm) SOC stocks measured from UNASM and RaCA dataset.

	No. of counties with records		Average		Min		Max		Standard deviation	
Database	*UNASM*	*RaCA*	*UNASM*	*RaCA*	*UNASM*	*RaCA*	*UNASM*	*RaCA*	*UNASM*	*RaCA*
AEZ			----------Mg ha^{-1}----------							
7	348	94	21.58	37.06	3.58	4.25	60.97	214.20	11.01	30.36
8	245	73	28.84	55.77	2.58	5.43	87.46	132.86	15.12	26.20
9	165	56	31.54	62.24	4.59	5.53	87.76	242.26	16.05	43.36
10	728	197	31.40	84.14	3.58	9.62	83.96	898.02	13.73	95.37
11	906	174	24.95	41.62	3.58	10.63	72.33	120.83	9.30	18.93
12	620	88	20.85	75.98	4.91	7.90	84.93	948.61	8.94	137.96
13	54	16	24.10	61.70	4.33	26.61	47.44	193.43	8.87	41.36
14	16	3	25.15	68.61	10.89	20.63	45.06	117.38	10.72	48.38
15	4	0	30.47	NA	10.32	NA	47.99	NA	15.63	NA
16	6	2	31.74	60.34	25.18	60.09	39.05	60.59	5.51	0.35
All	3092	703	25.93	60.96	2.58	4.25	87.76	948.61	12.26	75.64

- UNASM dataset has more observations and wider coverage across U.S. counties but RaCA database reports a wider data-range for variables within all AEZs. Lower variance in UNASM is due to larger sample numbers and possibly a smoothing effect of previous interpolation.
- For topsoil, average and median RaCA SOC values are 134% and 79.7% higher than UNASM values. Subsoil average and median values are 111% and 52.3% higher than UNASM values (data not shown).
- Both datasets have higher SOC values in AEZs 8, 9, and 10, and low values in AEZs 7 and 11.
- Evaluation of co-variates suggests the two data sets covered very similar resource ranges, with slightly but not significantly lower pH and higher precipitation and temperatures represented for areas covered by the UNASM dataset. Difference between the two inventories could stem from notable gains in SOC incurred during the past 20 years, differences in analysis methods used, lowering of UNASM estimates by interpolation, and/or a combination of these factors. We welcome your input on this.

Compare UNASM calibrated model outputs with RaCA

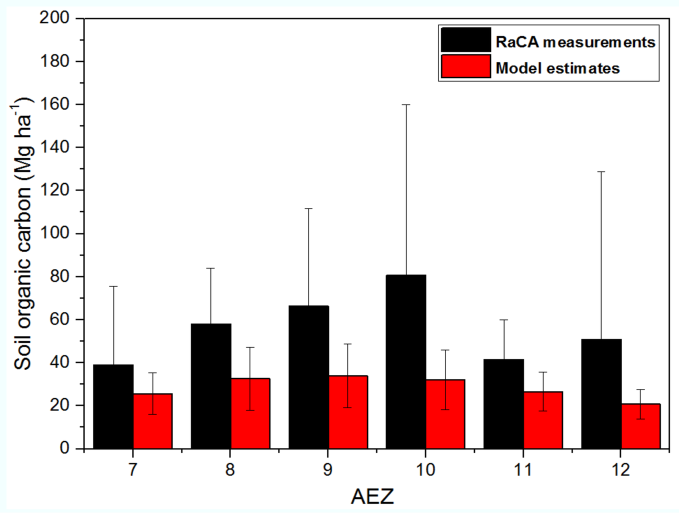

Fig. 3 Comparison of topsoil SOC stocks (0-30 cm) from RaCA measurements and SCSOC model outputs for croplands. Both values are from 2010 (n=515).

- Estimates of SOC stocks within AEZs produced with the $_{\text{UNASM-Cal}}$SCSOC model were much lower than RaCA inventory means. Better matching of inventory to modeling scenarios and land use histories may reduce differences.
- Model-based estimates had lower variance than measurements due to use of county-level climate and C input factors. Regression kriging with environmental covariates could improve spatial representation.

Compare UNASM calibrated model outputs with refined RaCA

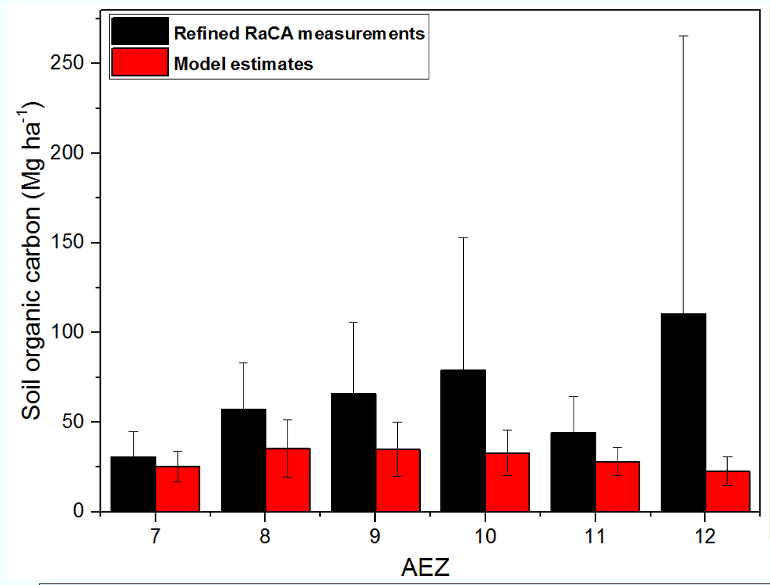

Fig. 4 Comparison of topsoil SOC stocks (0-30 cm) from refined RaCA measurements and SCSOC model outputs for croplands. Both values are from 2010 (n=251).

- Only 48.7% of the data points were retained after refinement.
- Differences between modeled and RaCA measured SOC were reduced in most AEZs after exclusion of non-cropped lands, but increased 8.2% and 191.9% for AEZs 11 and 12. Better accounting for land use did not significantly reduce differences between the modeled and inventory values.

Scale up refined RaCA dataset to county-level

- Significant predictors include elevation, precipitation, temperature, soil pH, and BD.
- Better prediction was obtained with lower SOC values. Observations with SOC > 200 Mg ha^{-1} were excluded as including them would increase RMSE by 128%. This indicates that modeling SOC with environmental covariates may have weak powers on extreme values from point measurements.
- The spatial distribution of refined RaCA data points is not homogeneous, which may affect the scaling up results.

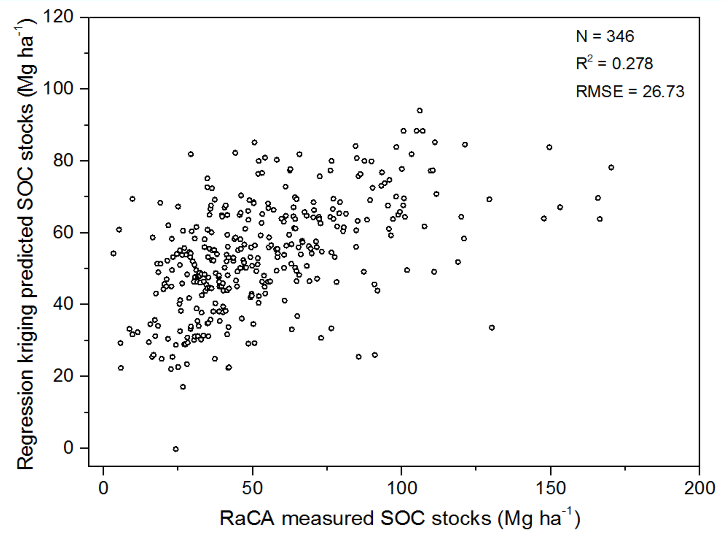

Fig. 5 Comparison of topsoil SOC stocks (0-30 cm) between refined RaCA measurements and results obtained with Regression Kriging method.

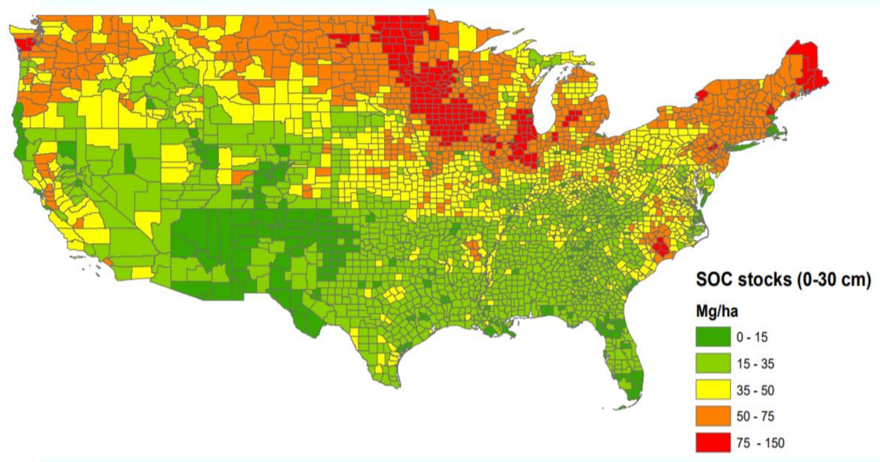

Fig. 6 Interpolated county-level RaCA topsoil SOC stocks (0-30 cm) for cropland using Regression Kriging method.

- SOC stocks are highest across the corn belts and in Northeast U.S. Spatial trends are parallel to estimates of SOC stocks made by NRCS (0-100 cm) using the RaCA dataset and ordinary kriging.
- Estimated county-level average and median SOC values (41.4 and 38.1 Mg ha^{-1}) are in between means of the original RaCA and UNASM databases (Table 1).

Model estimation of SOC after fitting interpolated RaCA dataset

- When SCSOC was fit to interpolated RaCA data, averaged county-level SOC stock (0-30 cm) estimates were 39.0±16.9 Mg ha^{-1} in 2010. This was 26.9% higher than the estimation made with $_{\text{UNASM-Cal}}$SCSOC. Averaged stock size estimated by $_{\text{RaCA-Cal}}$SCSOC for 1970-2000 was 38.7±17.6 Mg ha^{-1} and, the standard deviation of the prediction was reduced by 77.6% compared to the original RaCA data. This demonstrates the smoothing effects of the up-scaling process associated with gridded models. The standard deviation of the prediction was only 4.6 Mg ha^{-1} higher than that of the UNASM dataset, revealing the legacy of UNASM data interpolation.

Summary and Conclusions

- Comparison of two National SOC inventories suggest average stock values have doubled in the last 20 years. However, high variability in the RaCA dataset allows estimates to overlap the UNASM inventory.
- Model based estimates of SOC change made using a surrogate century model calibrated with the older UNASM database under-predict SOC stocks observed in the RaCA inventory. Model success was not improved by better accounting of land use history.
- Use of environmental co-variates (elevation, precipitation, temperature, and soil pH, and BD) to interpolate RaCA data was effective when SOC stocks were smaller than 200 Mg ha^{-1}.
- This effort demonstrated how upscaling through regression kriging and use of gridded data have a smoothing effect on SOC estimation that is retained on derived datasets and estimates.

Could SOM have increased this much since the mid-1980s or is this scaling/sampling uncertainty? How would you calibrate your model?

References

- Harris, I.P.D.J., Jones, P.D., Osborn, T.J. and Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. International Journal of Climatology. 34(3): 623-642.
- Kwon, H.Y. and Hudson, R.J., 2010. Quantifying management-driven changes in organic matter turnover in an agricultural soil: An inverse modeling approach using historical data and a surrogate CENTURY-type model. Soil Biology and Biochemistry. 42(12): 2241-2253.
- Kwon, H., Ugarte, C.M., Ogle, S.M., Williams, S.A. and Wander, M.M., 2017. Use of inverse modeling to evaluate CENTURY-predictions for soil carbon sequestration in US rain-fed corn production systems. PloS one. 12(2). e0172861.
- Liu, S., Y. Wei, W. Post, R. et al. Thornton. 2013. The unified North American soil map and its implication on the soil organic carbon stock in North America. Biogeosciences 10: 2915–2930.
- Wills, S., Seqeira, C., Loecke, T., et al. 2013. Rapid carbon assessment (RaCA) methodology: Sampling and Initial Summary. United States Department of Agriculture. Natural Resources Conservation Service.