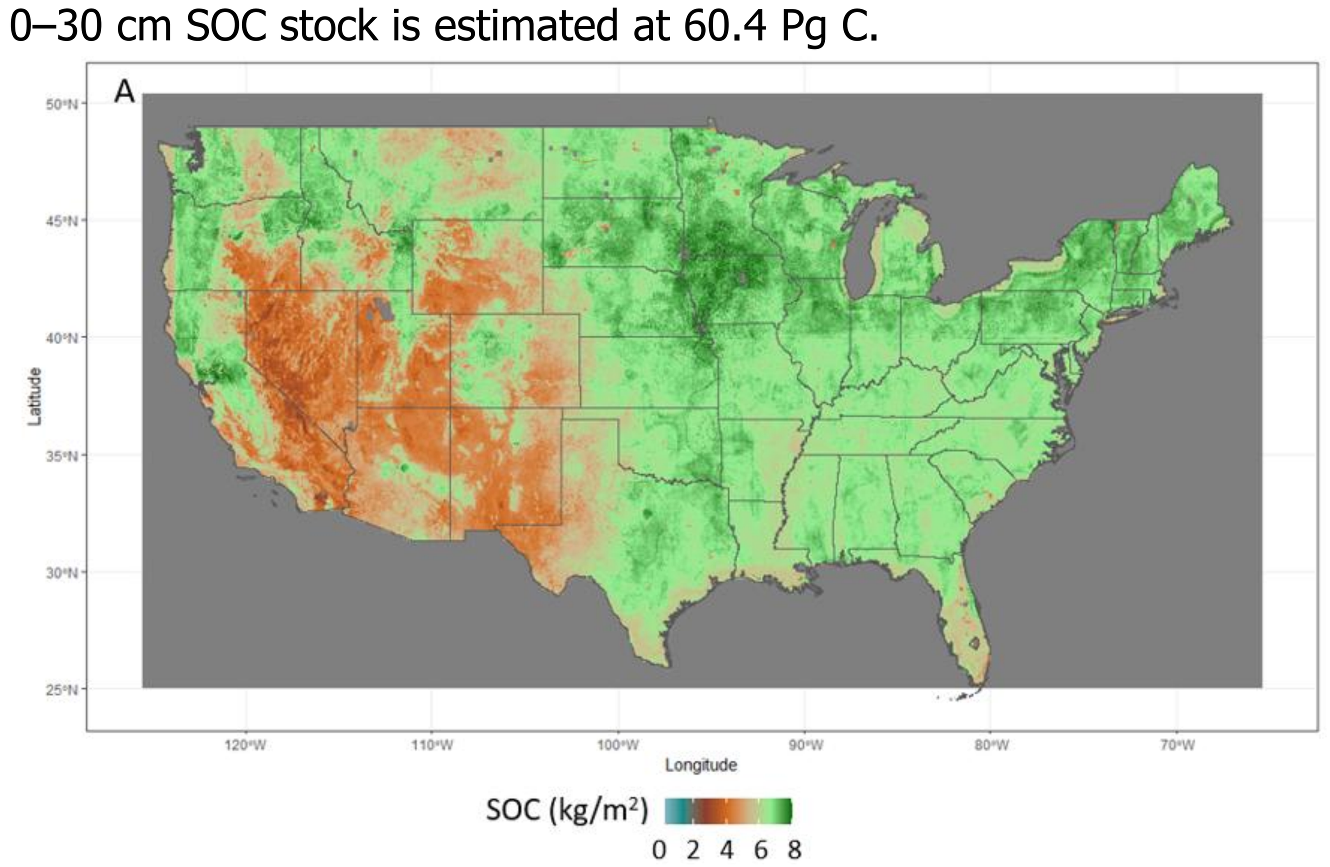
Project #104C
Methodology for Assessing Changes in Synergy with Soil Health

Motivation and Objectives

Reliable baseline estimates of U.S. soil health (SH) are crucial for supporting life cycle assessments (LCAs) and Sustainable Aviation Fuel (SAF).

Soil organic carbon (SOC) is a key indicator of soil health (Liptzin et al., 2022; Feeney et al., 2024; Maharjan et al., 2024). Improving SOC mapping supports soil quality and national-scale SH assessment. To establish a robust baseline for U.S. SOC, we harmonized multi-source SOC observations (ISCN, WOSIS, SoDaH, RACA, Sanderman) with co-located climate, land cover, topography, satellite-derived indices, and soil properties. We benchmarked multiple algorithms and selected Random Forest to produce a 30-m national SOC map for 1990–2022. The resulting 0–30 cm SOC stock is estimated at 60.4 Pg C.



Although previous efforts have produced meaningful results, there remain opportunities for improvement in several key areas.

- Duplicate and overlapping records: Several datasets contain overlaps and repeated entries, which may introduce statistical bias.
- Limited environmental variables: Previous studies primarily relied on a restricted set of predictors (e.g., annual temperature and precipitation, topography, vegetation indices such as NDVI), which may not fully capture the spatio-temporal variability of SOC.
- Modeling limitations: Traditional machine learning approaches have shown reasonable performance; however, they may be less effective in handling high-dimensional data and complex nonlinear relationships.

In our current and future work, we build on these aspects to update and iteratively improve SOC mapping as a basis for soil health assessment.

Methods and Materials

Re-examination of existing datasets:

- We compiled and harmonized multiple SOC datasets (ISCN, WOSIS, NCSS, SoDaH, RaCA), with NCSS included as a supplementary source.
- Particular attention was given to handling overlaps and duplicates to ensure that each field observation contributed only one valid entry.

Incorporation of additional environmental variables: To better represent the drivers of SOC formation, we introduced new soil and site attributes (e.g., pH, texture, drainage class) along with satellite-derived indices such as NIR, SWIR1, SWIR2, and EVI reflectance bands.

Summary

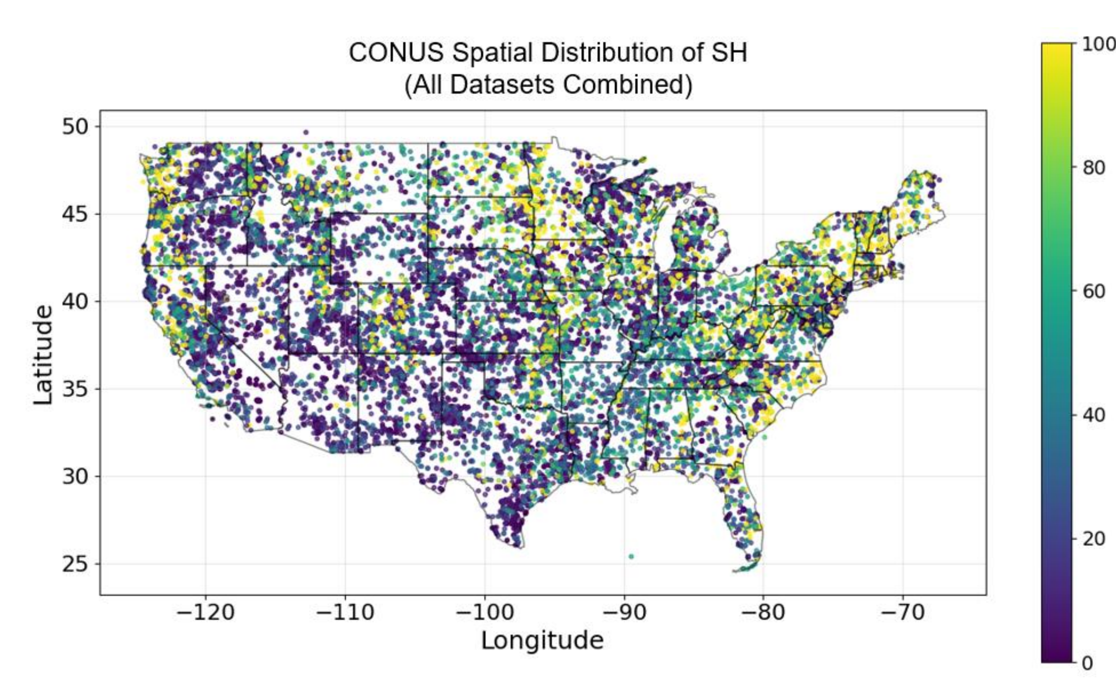
To ensure consistency across diverse data sources, we evaluated the characteristics of available SOC datasets. However, these datasets differ greatly in format, available fields, measurement units, and sampling depths. Some provide bulk density and carbon concentration, others only percentage values; sampling strategies vary from nationwide surveys to research-network compilations.
To ensure comparability, **all datasets were removed duplicates and standardized in this study**.
This included:

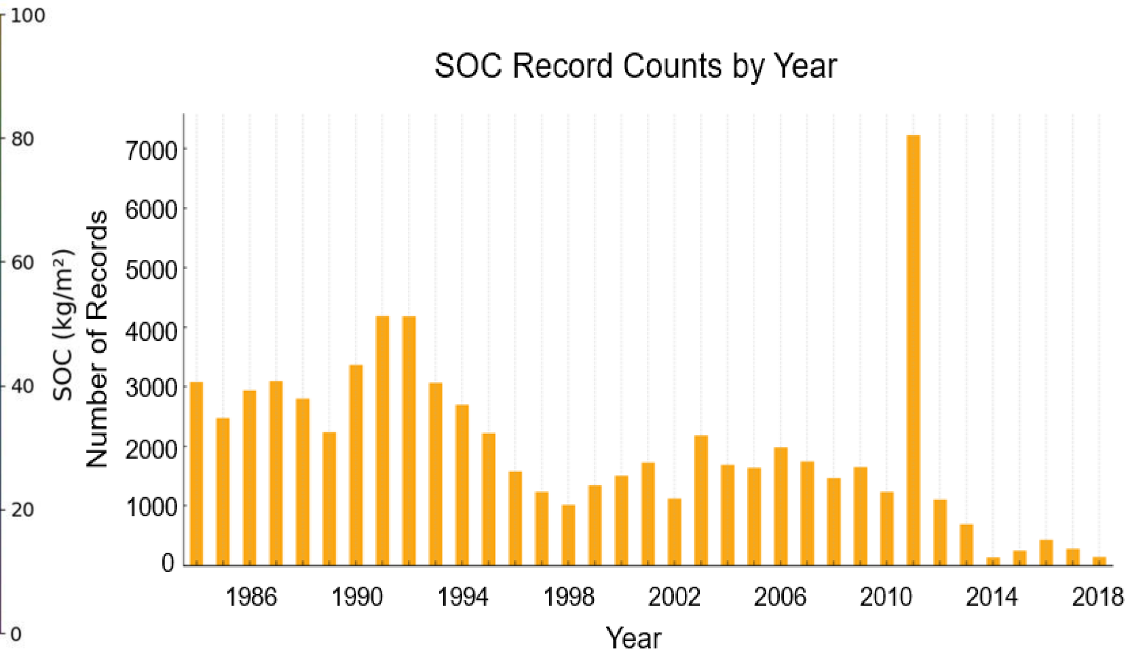
- Converting measurements to consistent SOC stock units (g C m^{-2} for 0–30 cm depth),
- Resolving overlaps and duplicate records across datasets,
- Standardizing metadata such as sampling location and year.

These harmonized datasets provide a consistent foundation for machine learning applications to improve SOC mapping and reduce uncertainties in national-scale soil health assessments.

Datasets	Type of Data	Sampling Strategy	Experimental / Analytical Methods	Spatial Coverage
RaCA (USDA NRCS -Rapid Carbon Assessment)	Field samples (nationwide campaign)	~6,000 randomly selected sites across USA	SOC by VNIR spectroscopy; bulk density from cores	USA
***NCSS** (USDA - National Cooperative Soil Survey)	Field samples + Lab measurements	County-scale field sampling, SSIR-42 protocols	Lab analyses (Kellogg Soil Survey Lab Manual)	USA
***ISCN** (International Soil Carbon Network)	Compiled database (researcher-contributed)	Contributed profiles; no uniform sampling	Contributor-supplied lab methods, harmonized by database	Global
***WOSIS** (ISRIC - World Soil Information Service)	Compiled database (global standardized)	Data from national archives	Standardized method codes; harmonization during import	Global
SoDaH (Soils Data Harmonization Database)	Research network compilation (LTER, NEON, NutNet, etc.)	Site-specific sampling; metadata documented	Unit conversions; standard depth pools	US-majority (with some global sites)
Sanderman et al. 2017	Global compilation prepared for modeling	No new sampling; based on ISRIC training set	Standardized SOC stocks from multiple sources for land-use change modeling	Global

* Removed duplicates and standardized by this study

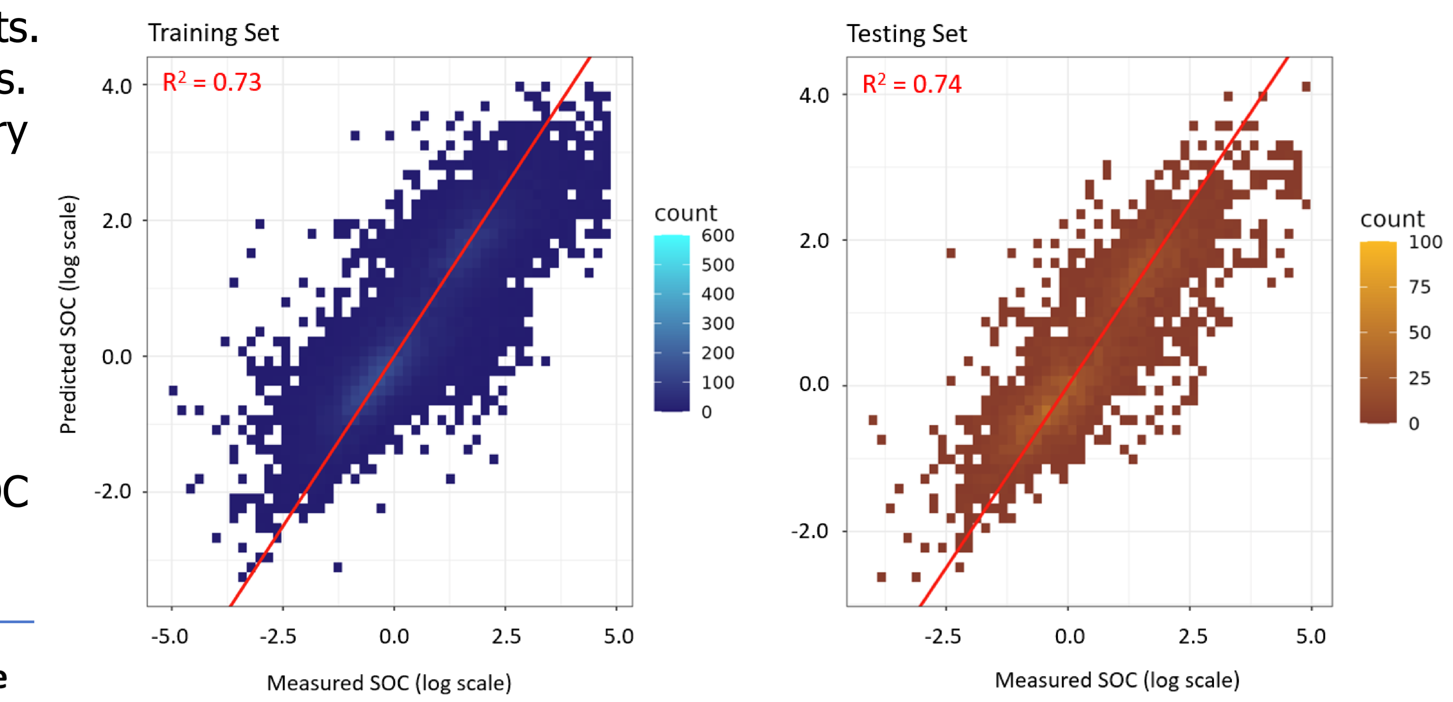




Lead investigators: Joshua Fu and T. Edward Yu, University of Tennessee
Project manager: Prem Lobo, FAA

October 15, 2025

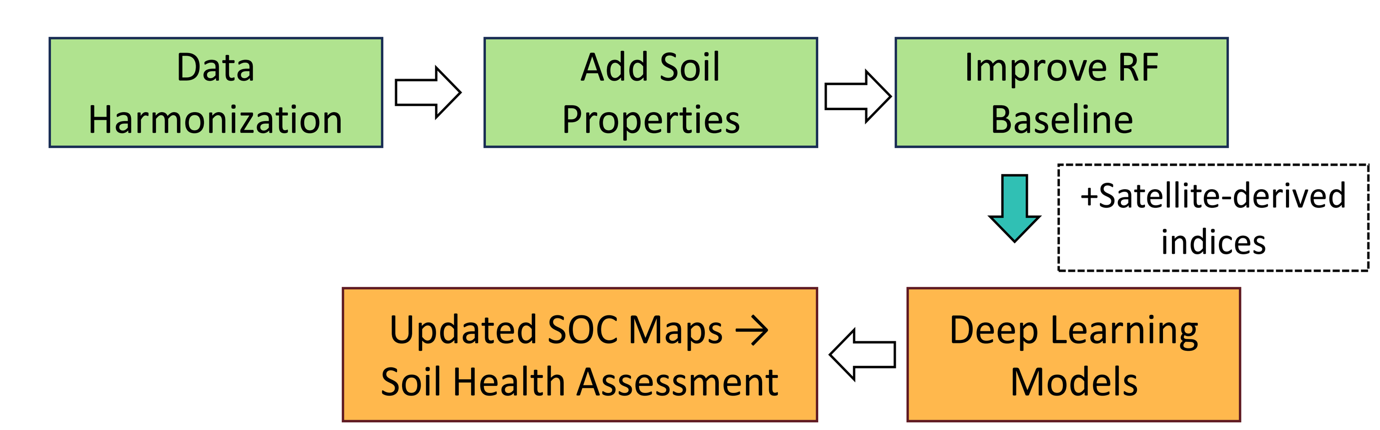
Results and Discussion



We harmonized and standardized multiple SOC datasets, strengthening their use as indicators of soil health, and incorporated additional soil properties (pH, soil texture, drainage class) to improve data consistency and predictive capacity. Using Random Forest models on the updated datasets, the validation R^2 increased to 0.74, a clear improvement compared to the previously reported value of 0.59.

This demonstrates the value of dataset cleaning and the inclusion of new soil attributes. Further improvements are anticipated by integrating satellite-derived indices and exploring deep learning methods to better capture complex relationships in SOC mapping.

Conclusion and Next Steps



Harmonizing SOC datasets and adding soil properties improved model performance (validation R^2 = 0.74 compared to 0.59 in previous work), emphasizing the importance of standardized data and enriched predictors.

Next, we will refine the modeling framework by assessing variable importance and reducing noise from less informative predictors. Satellite-derived indices will be added to further enhance spatial and temporal coverage. Building on these improvements, we aim to implement deep learning methods to better capture nonlinear and high-dimensional relationships, producing more robust and reliable SOC mapping products to support national soil health assessments..

Reference

1. Feeney, C. J., et al. 2024. Sci Total Environ 951:175642.
2. Liptzin, D., et al. 2022. Soil Biol Biochem 172:108708.
3. Maharjan, B., et al. 2024. Agrosyst Geosci Environ 7:e20504.
4. Sanderman, J., et al. 2017. Proc Natl Acad Sci USA 114:9575–9580.

This research was funded by the U.S. Federal Aviation Administration Office of Environment and Energy through ASCENT, the FAA Center of Excellence for Alternative Jet Fuels and the Environment, project 104C through FAA Award Number 13C-AJFE-UTENN-025 under the supervision of Prem Lobo. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the FAA.