Use of machine learning to investigate soil carbon storage & dynamics

Umakant Mishra
Principal Member of Technical Staff
Livermore, CA
Aug 6, 2024

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

AI/ML IN SOIL PROPERTY PREDICTION: A BRIEF HISTORY

- 1982, Smith & Gregg – Role of extractable soil P in pasture growth
- 2001, Minasny & McBratney – Prediction of soil organic carbon content using artificial neural networks
- 2020, Mishra et al. – Ensemble Machine learning improves spatial prediction of SOC in data limited region
- 2022, Mishra et al. – ML to derive empirical relationships between environmental factors and SOC
- 2022, Gautam et al. – Climate change may release over 1.8 Pg of C from top soils in US by 2100
- 2024, Nyaupane et al. – Observed environmental controls on SOC can improve ESMs (under review)
- 2024, Salinas et al. – Reducing uncertainties in regional & global SOC stocks (under preparation)

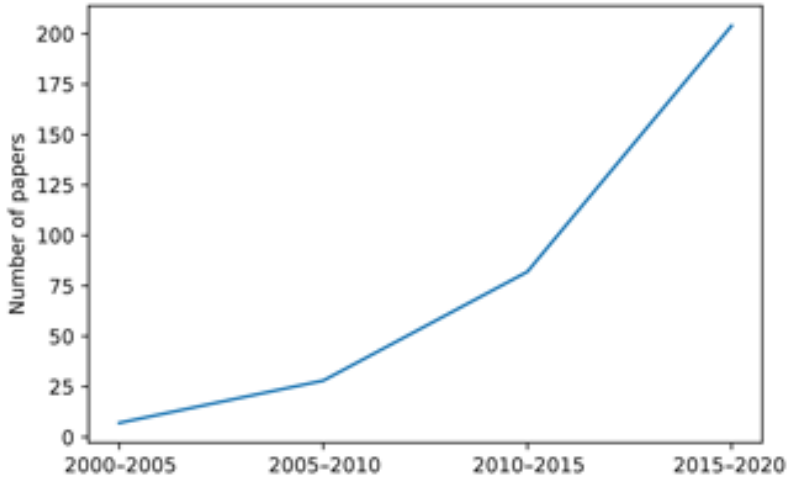

Figure: Number of publications using ML in soil carbon investigation (Padarian et al., 2020)

Why ML use has increased over time?

- Increased data availability
  - Remote & proximal sensing
  - Big & open datasets
- Algorithmic Advancements
  - Deep learning
  - Ensemble modeling
- Enhanced computational power
  - Graphics Processing Unit
  - Distributed learning

EMPIRICAL RELATIONSHIPS BETWEEN ENVIRONMENTAL FACTORS AND SOC PRODUCE COMPARABLE PREDICTION ACCURACY AS ML

Background & Objective

- Accurate representation of environmental controllers of SOC stocks in ESM land models could reduce uncertainties in future carbon-climate feedback projections.
- To derive observationally-based mathematical relationships describing environmental controls on SOC stocks.

Approach

Big data of soil and environmental factors were used in ML to generate nonlinear analytical expressions.

Results and Impacts

We developed ML based approach to derive analytical expressions for observationally-derived environmental controls on SOC stocks. Empirical relationships we derived can serve as benchmarks to evaluate environmental control representations of SOC stocks in Earth system land models, which could reduce uncertainty in predicting future carbon-climate feedbacks.

Mishra et al. 2022. Soil Sci. Soc. Am. J., doi:10.1002/saj2.20453

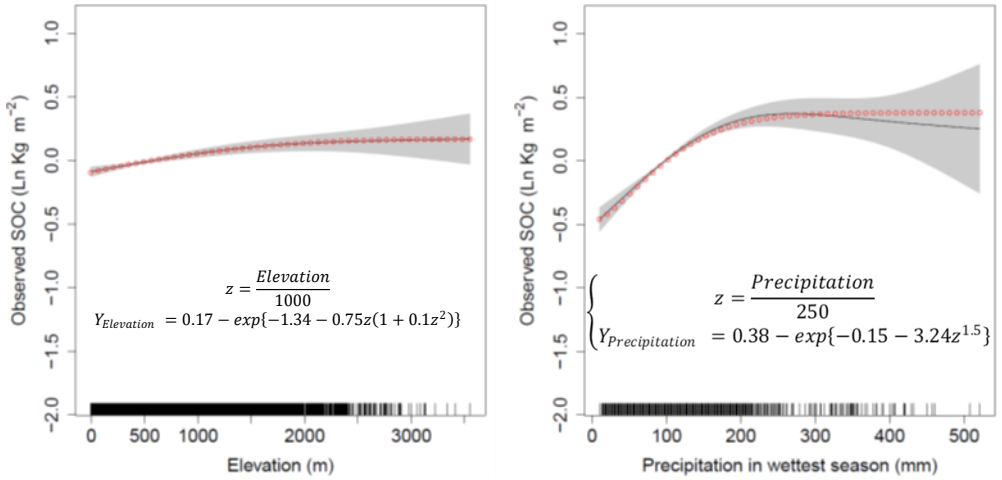

Figure: Deriving non-linear relationships that can be used to predict SOC stocks. The shade around the solid line indicates 95% confidence interval.

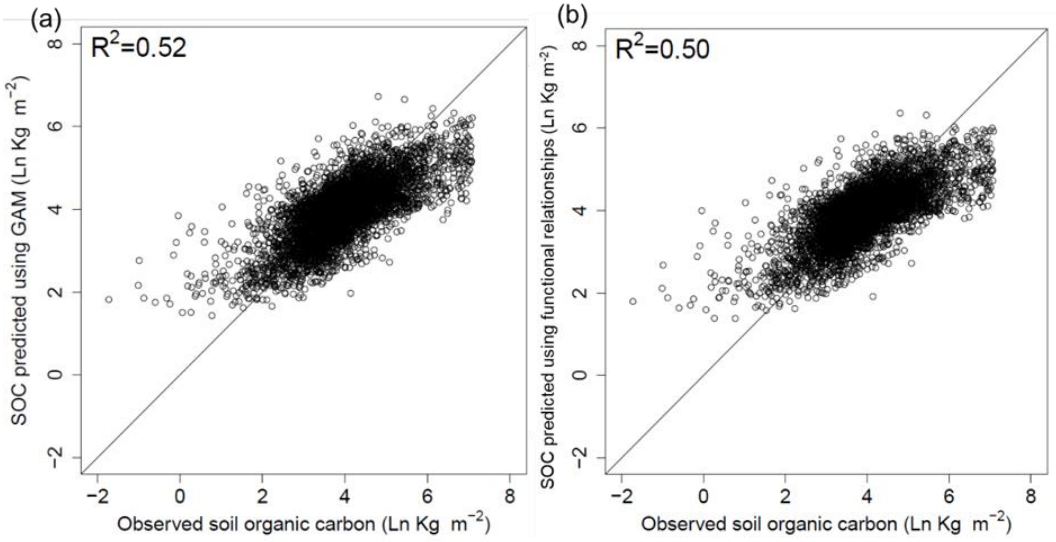

Figure: Comparison of predictions between (a) Machine Learning and (b) analytical model with 6 variables.

OBSERVATIONAL BENCHMARKS INFORM REPRESENTATION OF SOIL CARBON DYNAMICS IN LAND SURFACE MODELS

Background & Objective

- Representing SOC dynamics in ESMs is a key source of uncertainty in predicting carbon climate feedbacks.
- To benchmark CMIP6 ESM represented environmental controls on global SOC stocks.

Approach

Generate ML derived nonlinear analytical expressions of environmental controls on both field observations and model simulated environmental factors and SOC projections

Results and Impacts

Notable disparities exist in dominant environmental controllers and in the functional relationships between environmental factors and SOC stocks simulated by ESMs compared to observations. Our findings emphasize the need for benchmarking ESMs with observations to enhance our mechanistic understanding of the global soil carbon cycle.

Nyaupane et al. 2024. Biogeosciences (under review)

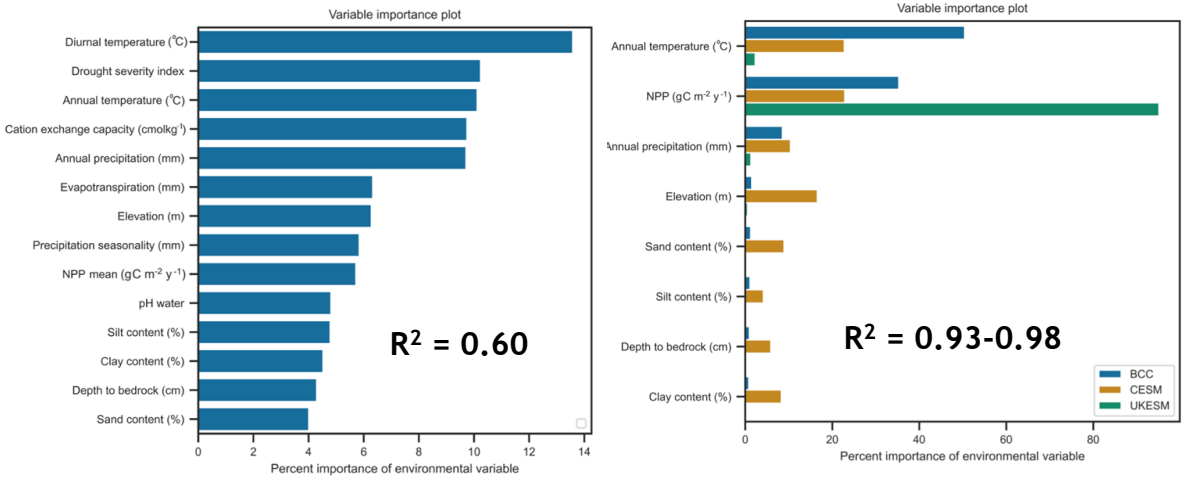

Figure: Importance of environmental factors in predicting global SOC stocks in observations (left) and three CMIP6 earth system models (right).

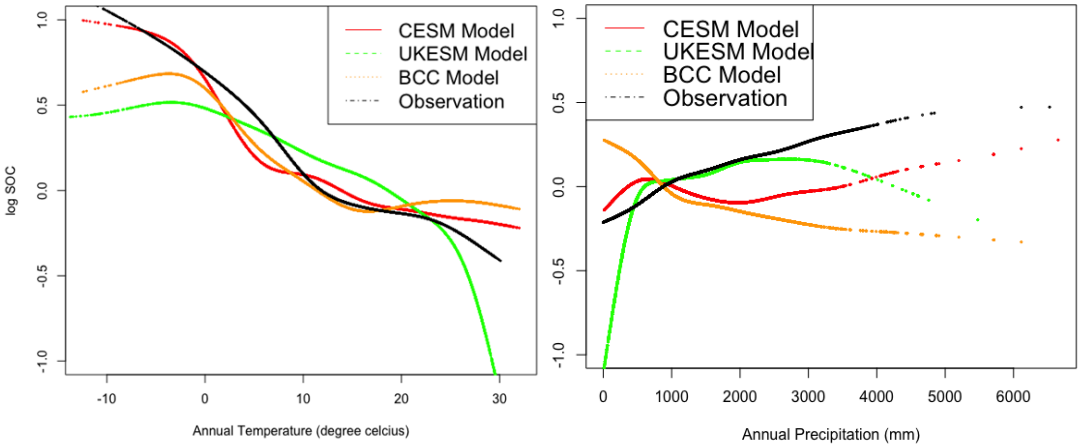

Figure: Predictive relationships between environmental factors and SOC stocks in observations (black line) and CMIP6 earth system models (different colors).

REDUCING UNCERTAINTIES IN GLOBAL SOIL CARBON ESTIMATES USING GENERATIVE AI BASED VERIFIABLE DIGITAL TWINS

Background & Objective

- Our recent study (Lin et al. 2023) reported large uncertainty in existing global SOC stock estimates:
  - 0-30 cm = 828 Pg C (577-1171)
  - 0-100 cm = 1873 Pg C (1086 – 2678)
- To reduce uncertainties in global SOC stock estimates by creating digital twins of soil profiles at required locations using generative AI.

Approach

Using global soil profile observations, data of soil types and environmental factors, and Generative AI

Results and Impacts

Preliminary outputs suggest promising results for digital twins of grassland soil profiles of Mollisol soil orders.

Our findings can be useful to constrain the existing uncertainties in global SOC stocks by proportionally creating digital twin of soil profiles for under sampled biomes and soil types.

Salinas et al. 2024. (under preparation)

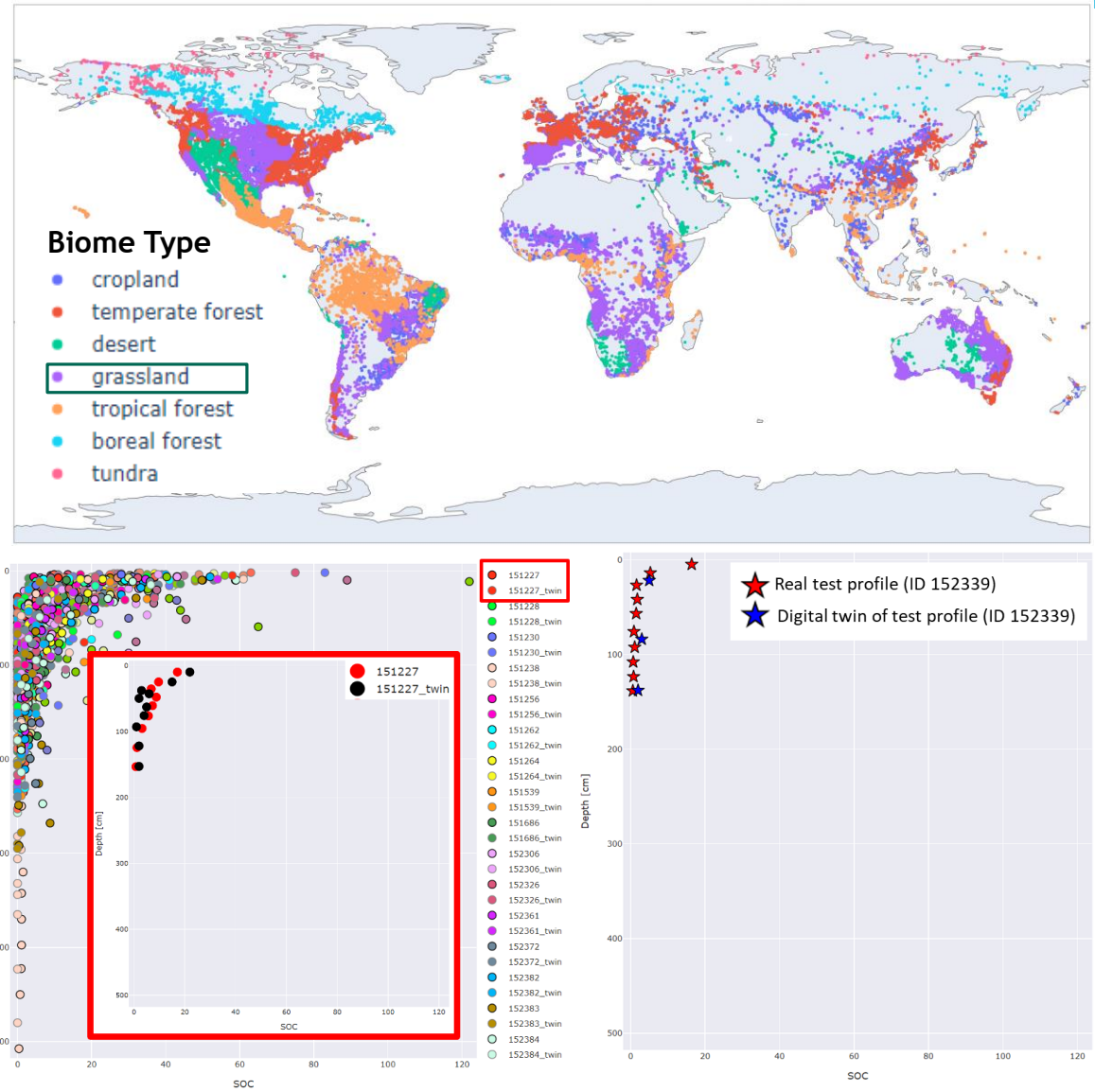

Figure: Digital twins of soil profiles of grassland and Mollisol soil order at specific temperature and precipitation range created using generative AI.

SUMMARY

- AI/ML has been used in soil carbon studies since last 23 years & it's use has increased exponentially after 2005-2010

- We have used ML to:
  - Improve spatial prediction of SOC stocks
  - Derive model benchmarks of environmental controls on SOC
  - Quantify and reduce uncertainty in SOC estimates

- AI/ML offers an opportunity to:
  - Extract patterns & enhance data-driven insights
  - Reduce uncertainties in SOC estimates
  - Improve study design & discovery processes
  - Incorporate new processes in ESMs

- The potential of AI/ML is yet to be realized in:
  - Developing hybrid models
  - Data augmentation
  - Reducing uncertainties

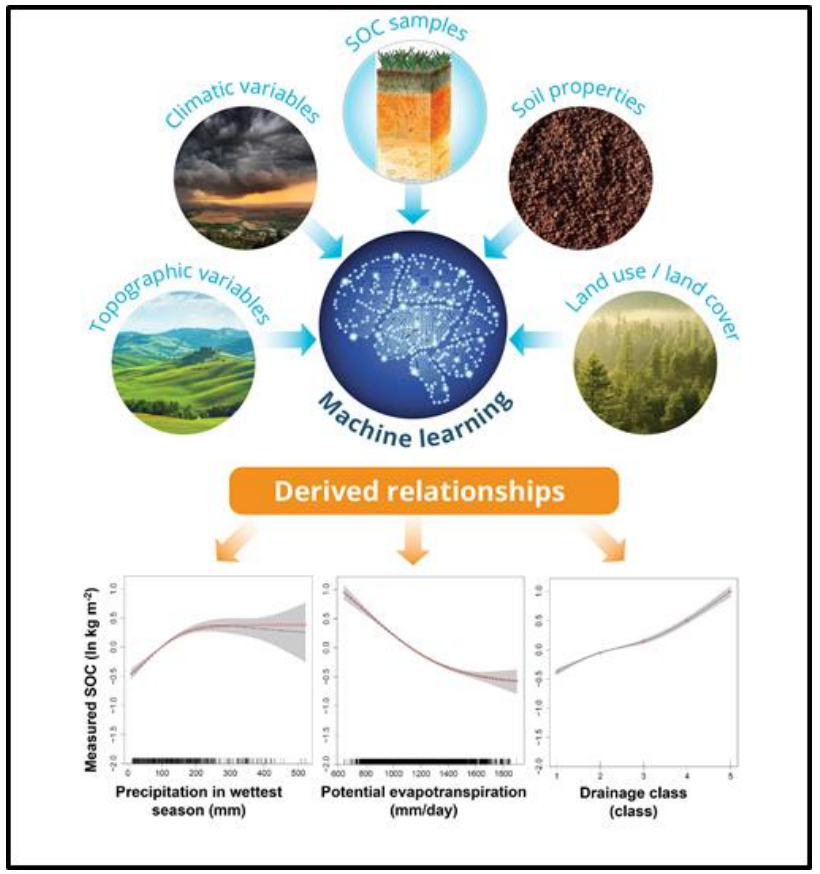

Figure: Deriving additive non-linear relationships that can be used to predict SOC stocks.

SOC dynamics is more complex than current model representations, and as the data sources are increasing rapidly, AI/ML can have a significant role in reducing the uncertainty in our understanding.