# A new machine-learning model to partition soil organic carbon into its centennially stable and active fractions based on Rock-Eval(r) thermal analysis

Marija Stojanova, Pierre Barré, Hugues Clivot, Lauric Cécillon, François Baudin, Thomas Kätterer, Bent T. Christensen, Claire Chenu, Ines Merbach, Adrián Andriulo, et al.

# PARTYsoc v3: A new machine-learning model to partition soil organic carbon into its centennially stable and active fractions based on Rock-Eval® thermal analysis


Marija Stojanova ♢ Pierre Barré ♢ Hugues Clivot ♠ Adrian Andriulo ♯ François Baudin ♣ Lauric Cécillon ∇ Claire Chenu ♡ Bent T Christensen △ Sabine Houot ♡ Thomas Kätterer ○ Ines Merbach ∀ Fabien Ferchaud †

♢Laboratoire de Géologie, École Normale Supérieure, CNRS, PSL University, IPSL, Paris, France ♠Université de Reims Champagne-Ardenne, INRAE, FARE, UMR A 614, Reims, France ♯ Instituto Nacional de Tecnología Agropecuaria, Buenos Aires ♣ ISTeP, UMR 7193 Sorbonne Université, CNRS, Paris ∇Research and innovation office, Ministry of agriculture and food sovereignty, Paris, France ♡ UMR ECOSYS, INRAE, AgroParisTech, Université Paris Saclay, Palaiseau, France △ Aarhus University, Department of Agroecology - Soil Fertility ○ Swedish University of Agricultural Sciences ∀ Helmholtz-Centre for Environmental Research † UMR Eco&Sols, Univ Montpellier, CIRAD, INRAE, IRD, InstitutAgro Montpellier, Montpellier

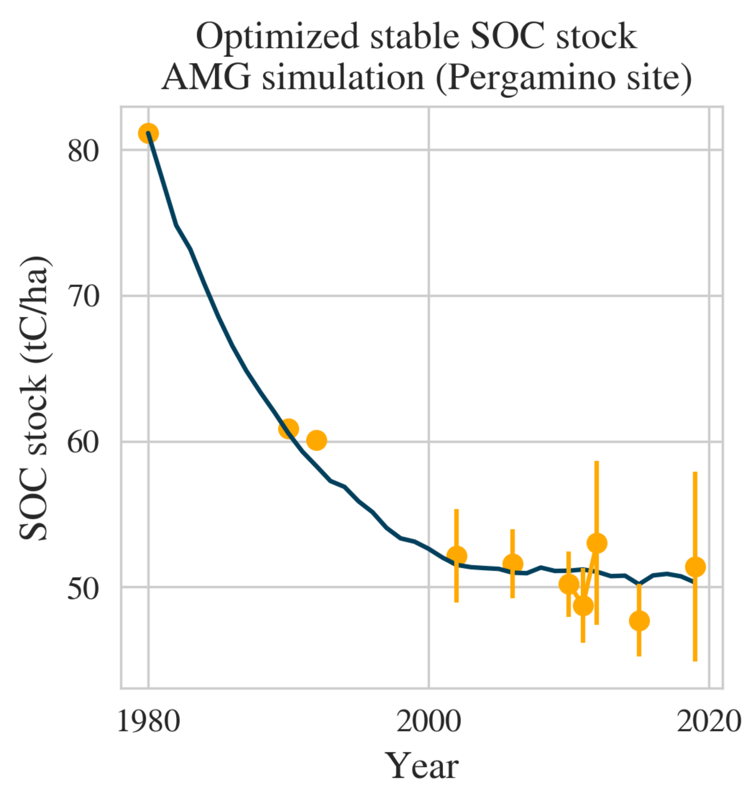

## Introduction

- Accurately simulating soil organic carbon (SOC) stock evolution is crucial for assessing the role of soil in climate regulation.
- SOC evolution is often modeled by dividing SOC into kinetic pools with contrasting residence times. **Initializing such kinetic pools remains a complex** task and a source of uncertainty in SOC simulations.
- In a previous study, Cécillon et al. [1] developed a machine-learning model (PARTYsoc v2) that uses Rock-Eval® thermal analysis results as input features to quantify the proportion of centennially stable and active SOC fractions.
- These proportions have been shown to be particularly effective for initializing the AMG model, enabling very accurate simulations of SOC stock evolutions for a dozen French agricultural long-term experiments [2].
- This present work builds **a new version of PARTYsoc**, validated on a larger data set, and extends the usefulness of the AMG model initialized with PARTYsoc to different parts of the world.

## Materials and methods

### Data

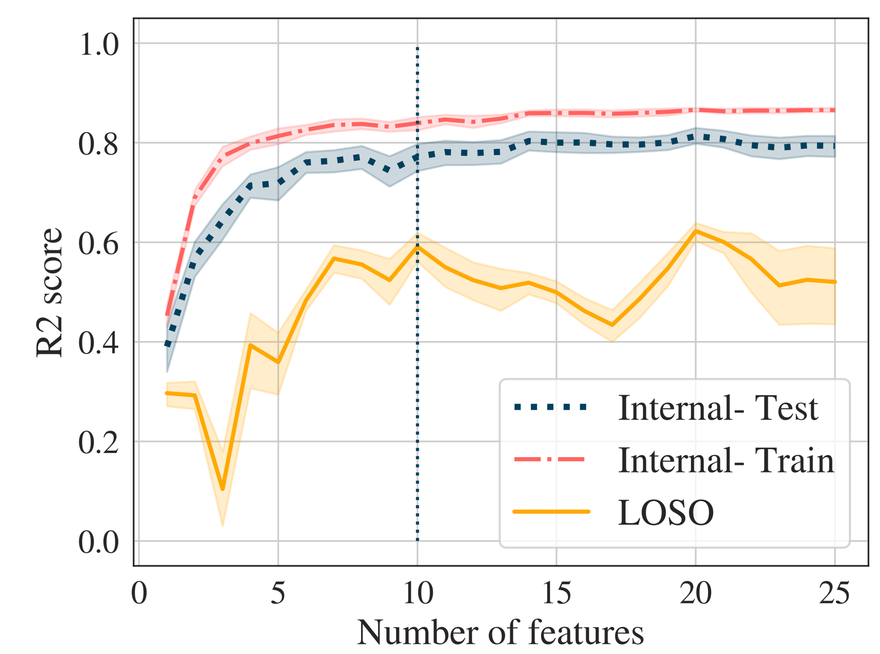
- Data collection of 18 sites with known crop yields and SOC stock evolutions and archived soil samples available for Rock-Eval® characterization.
- Sites are located in Europe and South America.
- The list of Rock-Eval® parameters [3] represents our potential **set of features**.
- The stable SOC stock size for each site is determined by optimization of the AMG model as in Clivot et al. (2019) [4].



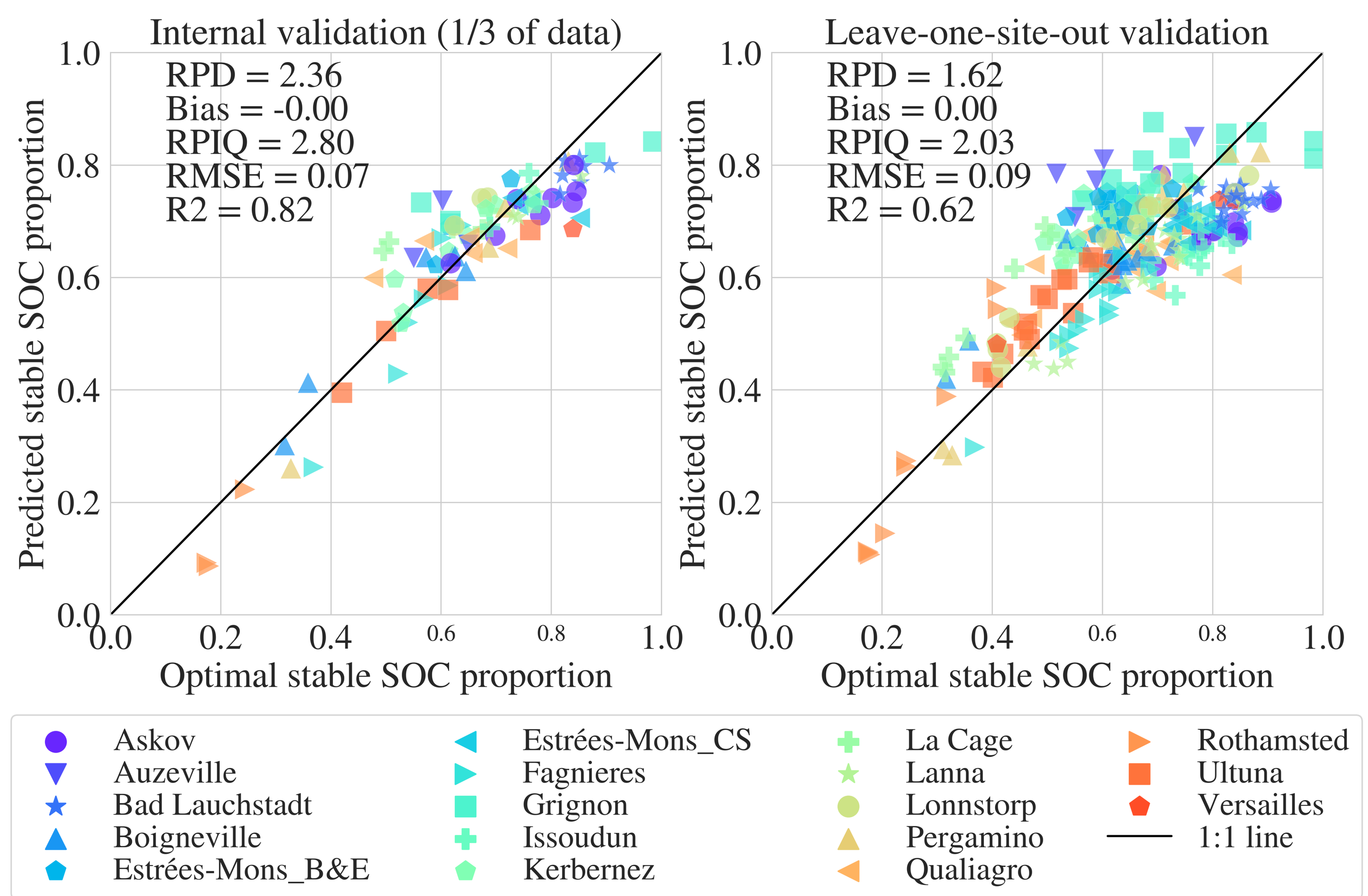

- Using the optimal stable SOC stock values, we calculate the stable SOC proportion of each sample from each site, representing our **target variable**.
- We reserve 1/3 of the data for testing, and use 2/3 for training.

### Feature engineering and model training

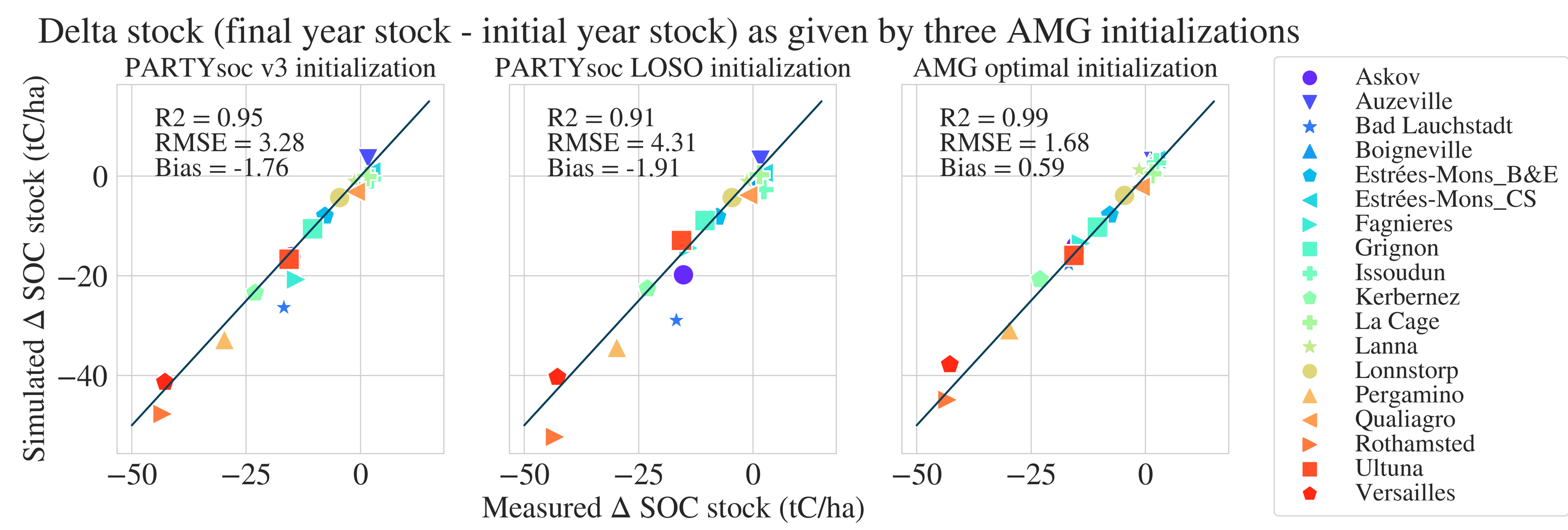
- We remove the highly correlated Rock-Eval® features using Spearman's correlation and a threshold of 0.9.
- We use AutoFeatureCorrelation and permutation importance to **rank the features according to their impact on the predicted target**.
- We run 30 repetitions of 5-fold cross-validation to find the **optimal number of features** for prediction on the test set and in Leave-One-Site-Out (LOSO) validation.
- The final model is a **combination of a Support Vector Machine (SVM) and a Beta regression**, trained using 10 Rock-Eval® features.




## Results




- On average, we obtain an $R^2 = 0.78$ and RMSE = 0.07 in internal validation, and $R^2 = 0.59$ and RMSE = 0.09 in LOSO validation.




- Using PARTYsoc v3 (LOSO) values to initialize AMG simulations offers comparable results to initializing with optimal AMG values.

## Conclusions and forthcoming research

- The proposed new PARTYsoc v3 builds upon and improves the original PARTYsoc v2 by incorporating a larger data set covering a **wider geographic area**, and a **and more parsimonious machine-learning model** (10 vs. 18 Rock-Eval® features in PARTYsoc v2).
- We are currently working on extending the data set as well as stabilizing the processes of feature selection and model parameterization.

## References


[1] L. Cécillon, F. Baudin, C. Chenu, B. T. Christensen, U. Franko, S. Houot, E. Kanari, T. Katterer, I. Merbach, F. Van Oort, *et al.*, "Partitioning soil organic carbon into its centennially stable and active fractions with machine-learning models based on Rock-Eval (R) thermal analysis," *Geoscientific Model Development*, 2021.

[2] E. Kanari, L. Cécillon, F. Baudin, H. Clivot, F. Ferchaud, S. Houot, F. Levavasseur, B. Mary, L. Soucémarianadin, C. Chenu, *et al.*, "A robust initialization method for accurate soil organic carbon simulations," *Biogeosciences*, vol. 19, no. 2, pp. 375–387, 2022.

[3] L. Pacini, T. Adatte, P. Barré, M. Boussafir, N. Bouton, L. Cécillon, V. Lamoureux-Var, D. Sebag, E. Verrecchia, A. Wattripont, *et al.*, "Reproducibility of Rock-Eval® thermal analysis for soil organic matter characterization," *Organic Geochemistry*, vol. 186, p. 104687, 2023.

[4] H. Clivot, J.-C. Mouny, A. Duparque, J.-L. Dinh, P. Denoroy, S. Houot, F. Vertès, R. Trochard, A. Bouthier, S. Sagot, *et al.*, "Modeling soil organic carbon evolution in long-term arable experiments with AMG model," *Environmental modelling & software*, 2019.


## Acknowledgements


We thank Florence Savignac for her help in conducting Rock-Eval measurements. This work would not have been possible without all the people who helped initiate and maintain these invaluable long-term trials. The analyses have been funded by the CNRS-INRAE "IMMORTAL" project.



Contacts: stojanova@geologie.ens.fr barre@geologie.ens.fr fabien.ferchaud@inrae.fr hugues.clivot@univ-reims.fr