Framework for Spatial Predictive Modelling of Soil Properties Using Machine Learning

Name: Jeonghwan Choi

Affiliation: Junior Research Fellow in Geoinformatics, Department of Geography, University of Tartu, Estonia

Contact: jeonghwan.choi@ut.ee

@Lgeoinformatics

AGILE PhD School 2024
Date: November 27, 2024
Location: Universitat Jaume I, Castellón de la Plana, Spain

FutureScapes

UNIVERSITY OF TARTU Landscape Geoinformatics Lab

Estonian Research Council

Republic of Estonia Ministry of Education and Research

Water Smart Land

European Research Council Established by the European Commission

Co-funded by the European Union

Investing in your future

Problem Statement

- **Key Issue**: Mapping soil properties (e.g., Soil Organic Carbon, SOC) at a large scale is **expensive** and **logistically difficult.**
- **Why It Matters**: Accurate soil maps are needed for **sustainable agriculture**, **carbon sequestration**, and **climate change mitigation**.

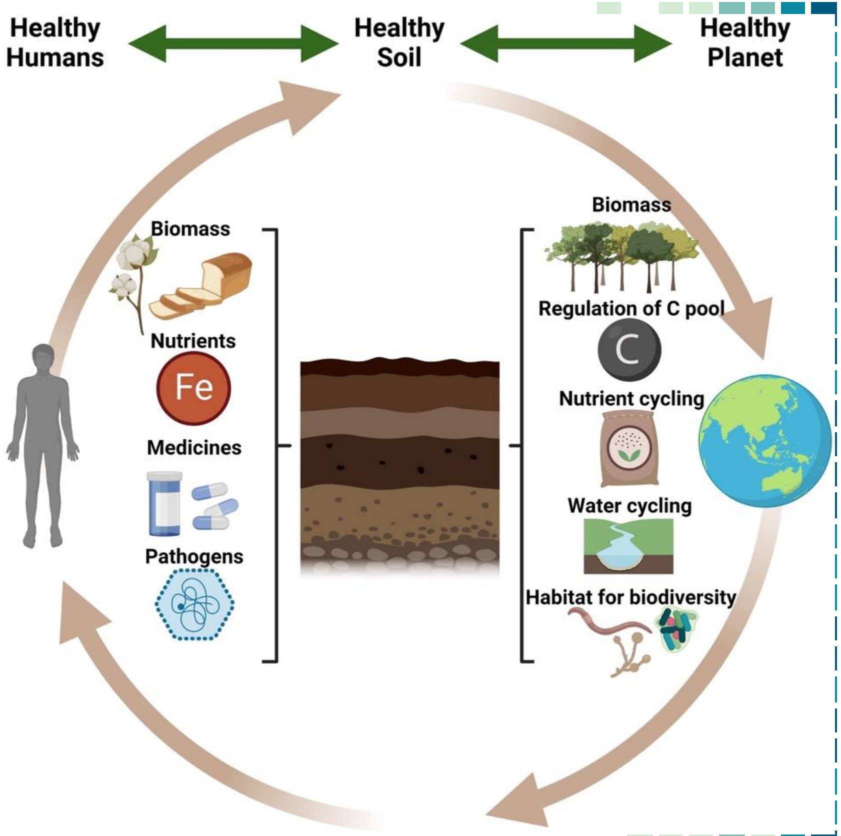


Figure 1: Healthy soil is inextricably tied to healthy people and a healthy planet (graphic based on Kopittke et al. 2023) / Credit: Michael Salzwedel.

Research Context & Motivation

- **Problem:** Traditional soil mapping methods are inefficient at large scales.
- **Goal:** Use **machine learning** to predict **soil properties** like SOC, while accounting for **spatial autocorrelation** to improve accuracy.

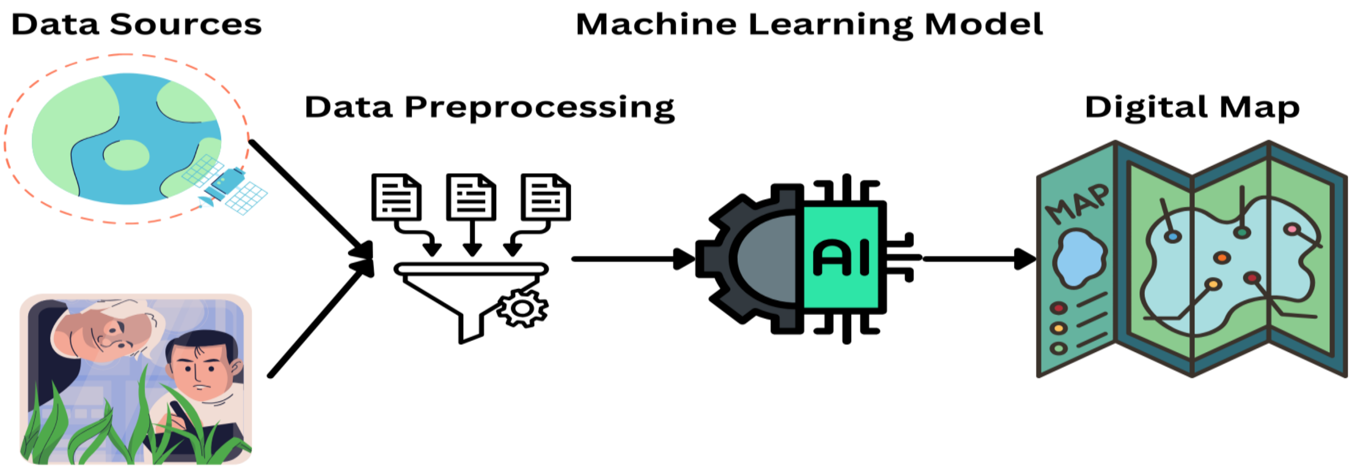


Figure 2: A simplified horizontal workflow diagram illustrating the progression from data sources (remote sensing and field samples) to machine learning methods, culminating in the creation of a digital map.

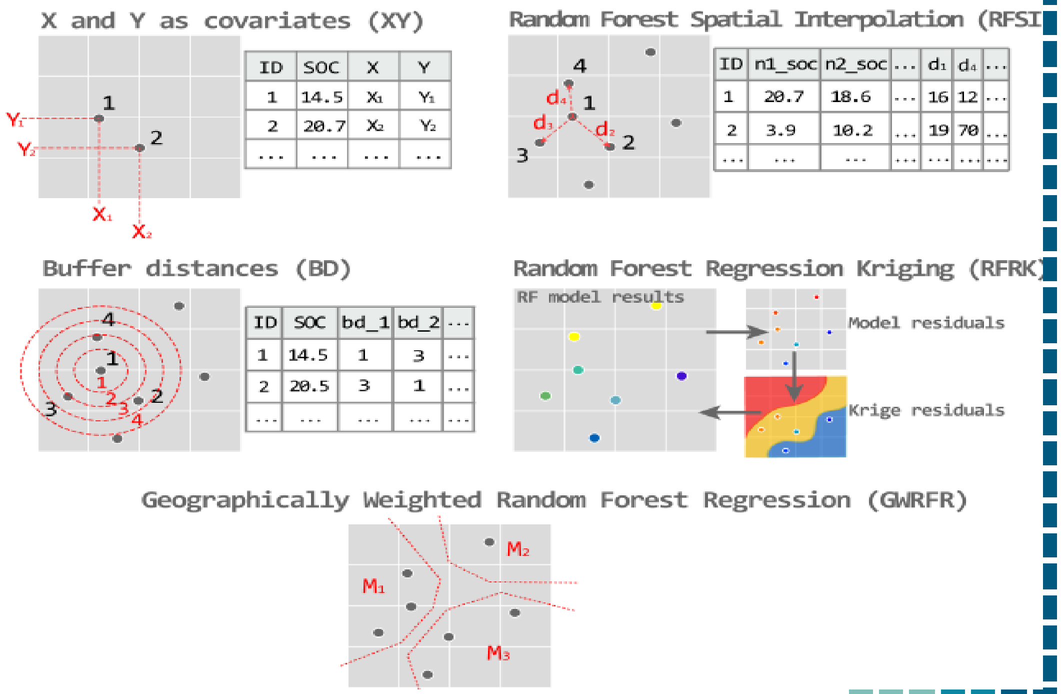
Methodology Overview

- Comparison of Baseline Random Forest and Spatial Autocorrelation-Adjusted Random Forest Models.

Table 1: Data Feature for Random Forest Construction)

Feature name	Description
clay	Clay fraction (top layer)
silt	Silt fraction (top layer)
sand	Sand fraction (top layer)
rock	Rock fraction (top layer)
twi	Terrain wetness index
tri	Terrain roughness index
slope	Slope
lsf	LS-factor
landuse	Land use classification
ndvi	Mean NDVI, July 2022
drained	Boolean value for drainage
soc	Soil organic carbon (% of mass)

Figure 3: Methods for incorporating spatial structure into random forest models



Spatial Autocorrelation (Completed Paper)

- **Findings:** Incorporating spatial autocorrelation improves **SOC prediction accuracy**, but the overall improvement is **small**.
- **Implication**: Validates the importance of spatial dependencies in soil prediction models.

Table 2: Model evaluation metrics (5-fold cross-validation)

Model	R^2	RMSE	MAE
Baseline vector	0.61	7.5	4.46
Baseline raster	0.6	7.5	4.39
XY	0.61	7.4	4.23
RFSI	0.63	7.29	4.31
BD	0.62	7.37	4.27
RFRK	0.61	7.47	4.39
GWRFR	0.6	7.49	4.57

Methods & Preliminary Findings

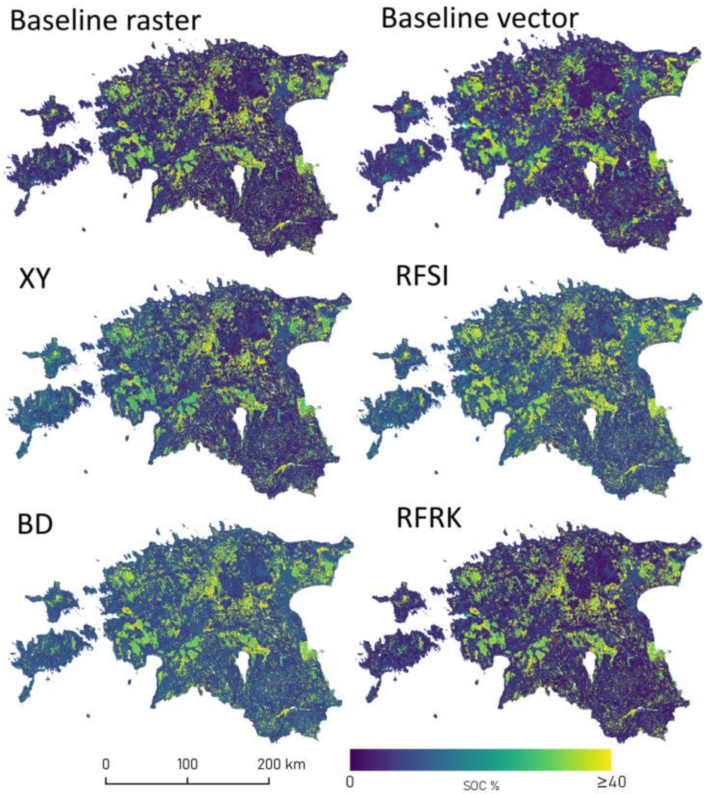


Figure 4: Comparison of predictions by spatial and non-spatial machine learning methods across Estonia.

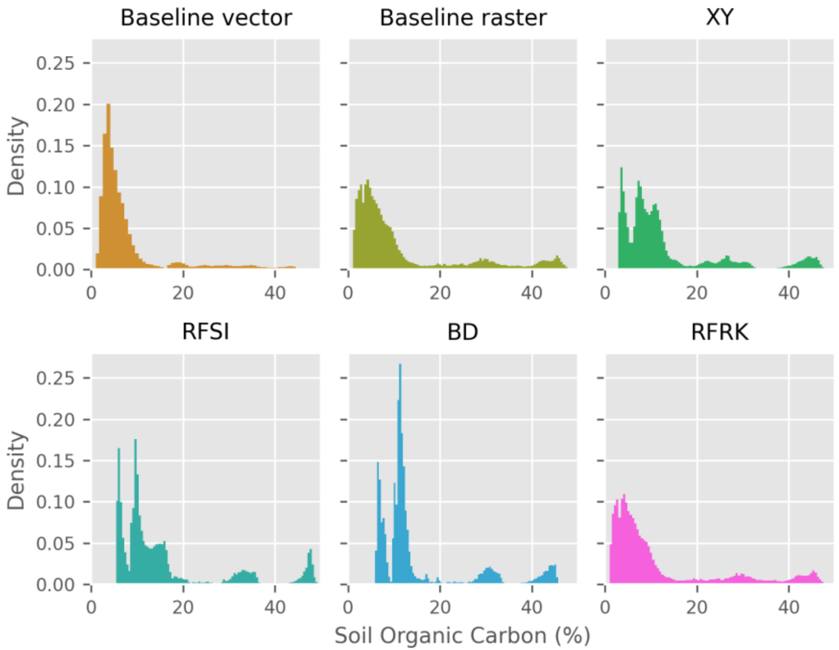


Figure 5: Distribution of SOC predictions across all of Estonia for each model.

Current Work – Sampling Design

- **Focus**: Improving **sampling design** to increase **data representativeness** and **model performance**.

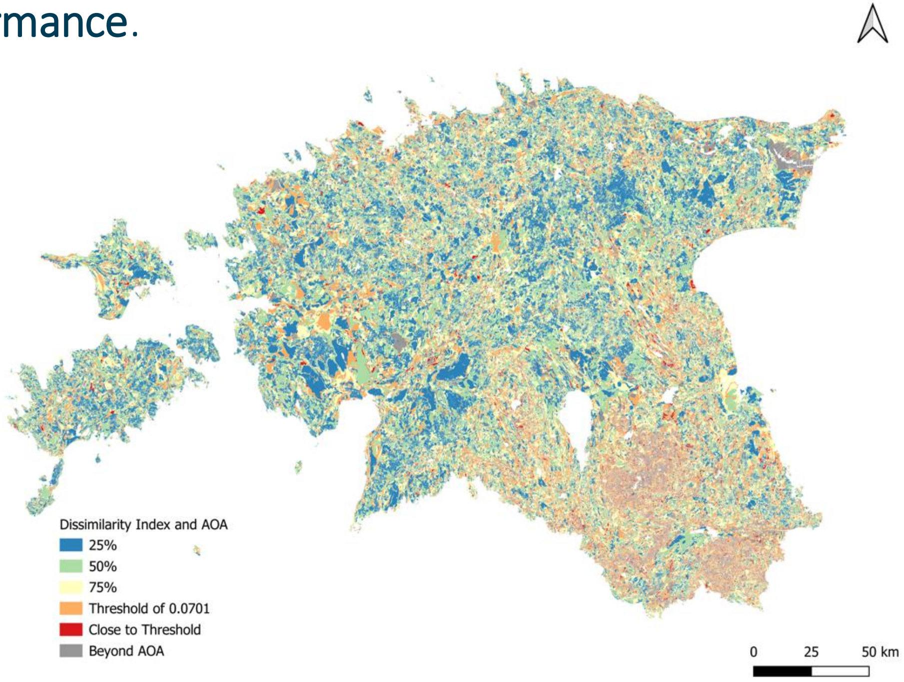


Figure 6: Dissimilarity Map of Training Samples and All Areas of Estonia.

Current Work – Sampling Design

- **Current Activities**: Visiting **potential sampling areas** across regions to collect data for SOC estimation.

Figure 7: Collecting soil samples from potential SOC sites for analysis.

Next Steps

- **Key Points:**
  - Analyse the collected samples and integrate them into machine learning models.
  - Investigate if there is any improvement in accuracy after adding additional selected samples
  - Build optimized field sampling method.

UNIVERSITY OF TARTU
1632
Landscape Geoinformatics Lab

Thank you!

FutureScapes

Estonian Research Council

Republic of Estonia
Ministry of Education and Research

Water Smart Land

erc
European Research Council
Established by the European Commission

Co-funded by the European Union

Investing in your future