

Machine learning in spatial predictive modelling: challenges and opportunities

Evelyn Uuemaa

Assoc. Prof. in Geoinformatics

Landscape Geoinformatics Lab, University of Tartu, Estonia



@Lgeoinformatics | @evelynuuemaa

evelyn.uuemaa@ut.ee

BalticGIT conference 8.-9. June 2023, Riga



Estonian
Research Council

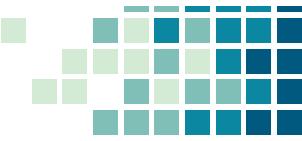
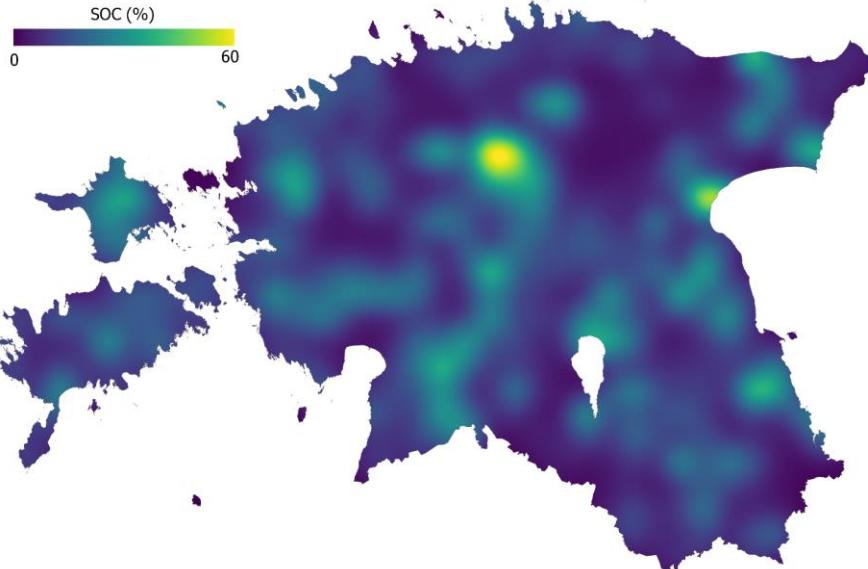
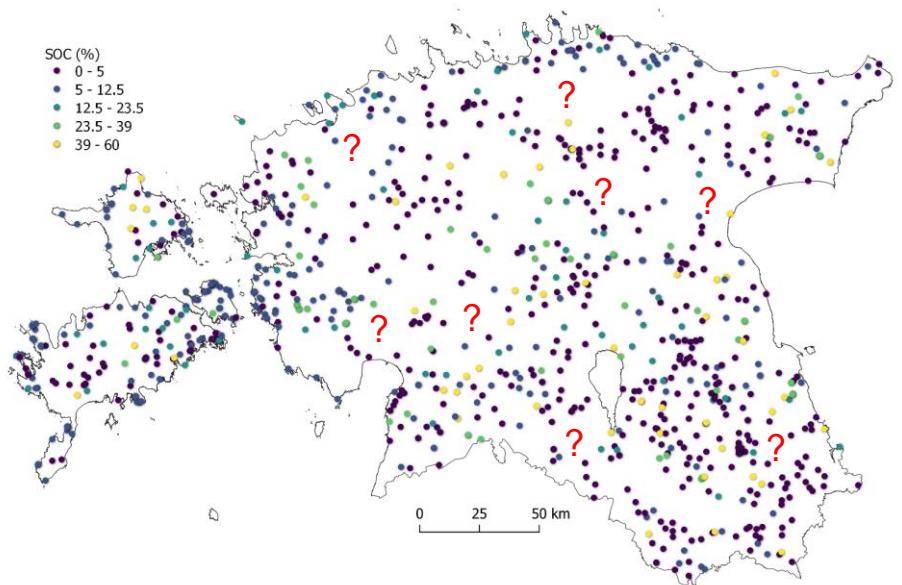
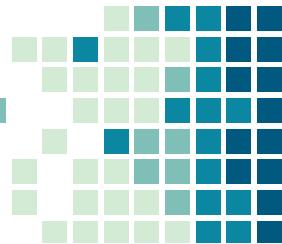


UNIVERSITY OF TARTU
Department of Geography



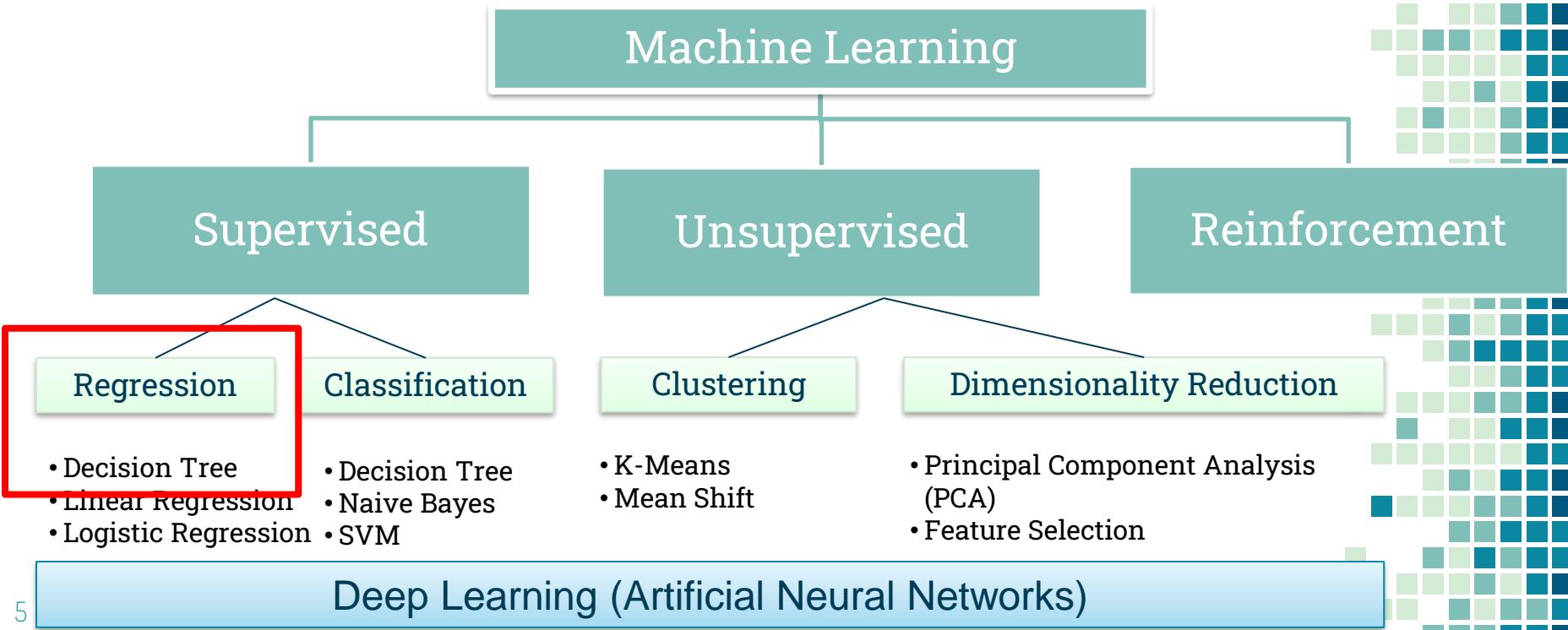
THE PROBLEM

We can't measure everything and everywhere



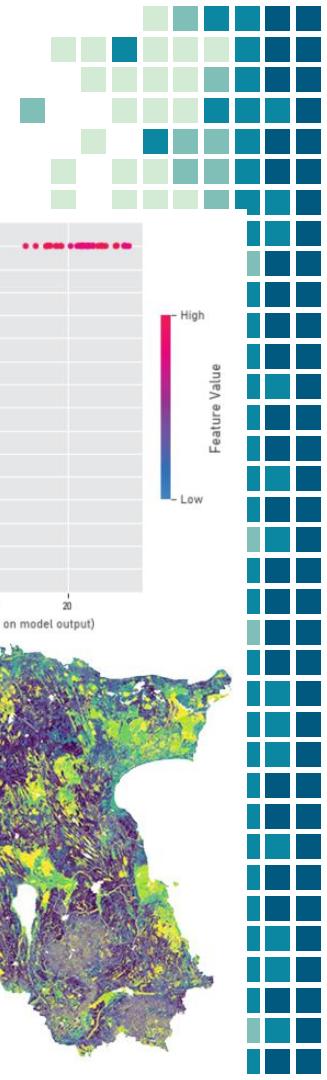
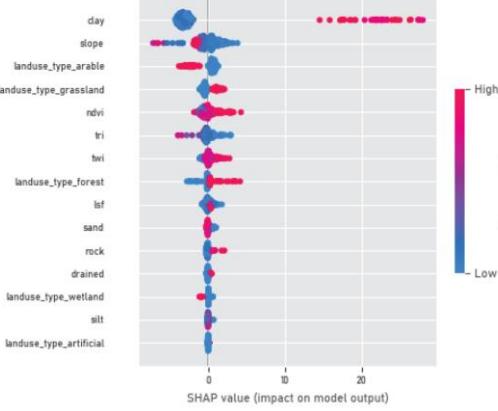
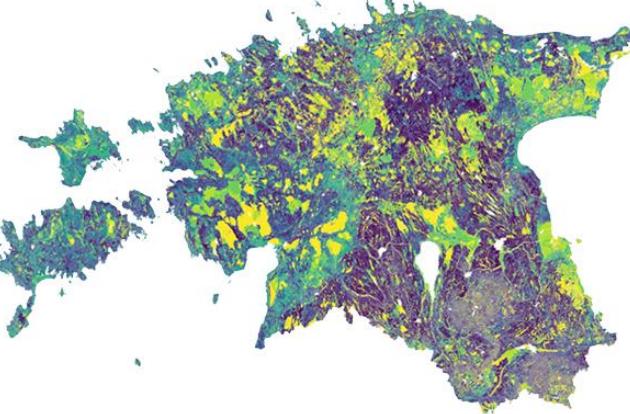
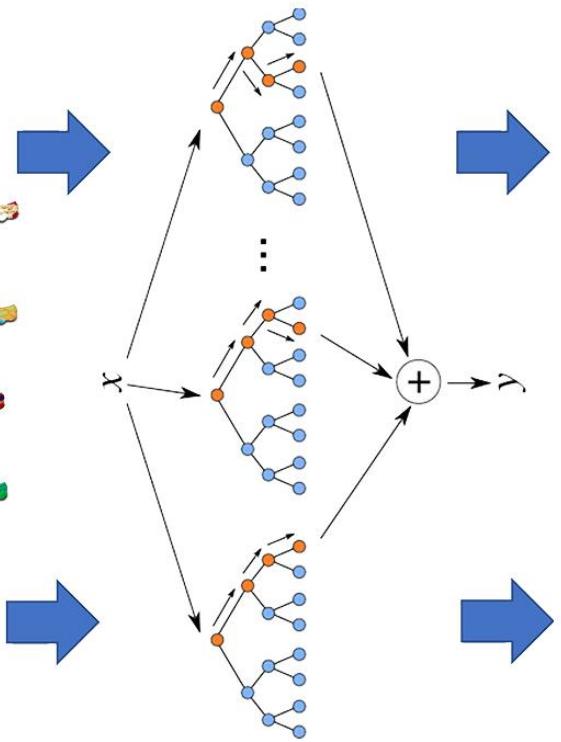
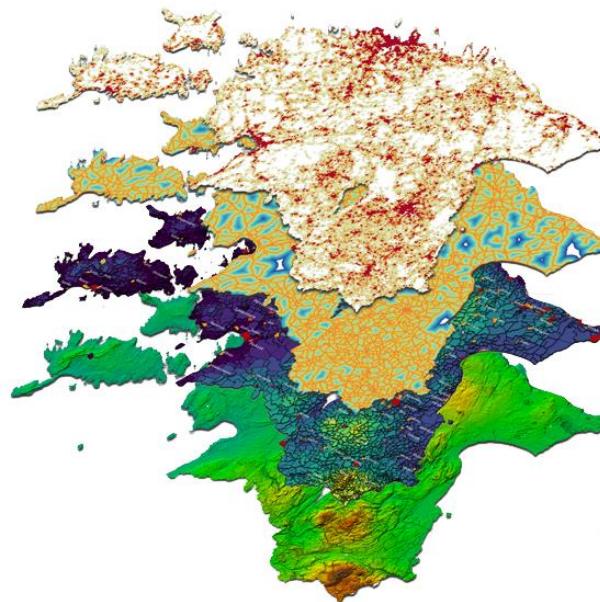
OPPORTUNITIES

Types of machine learning



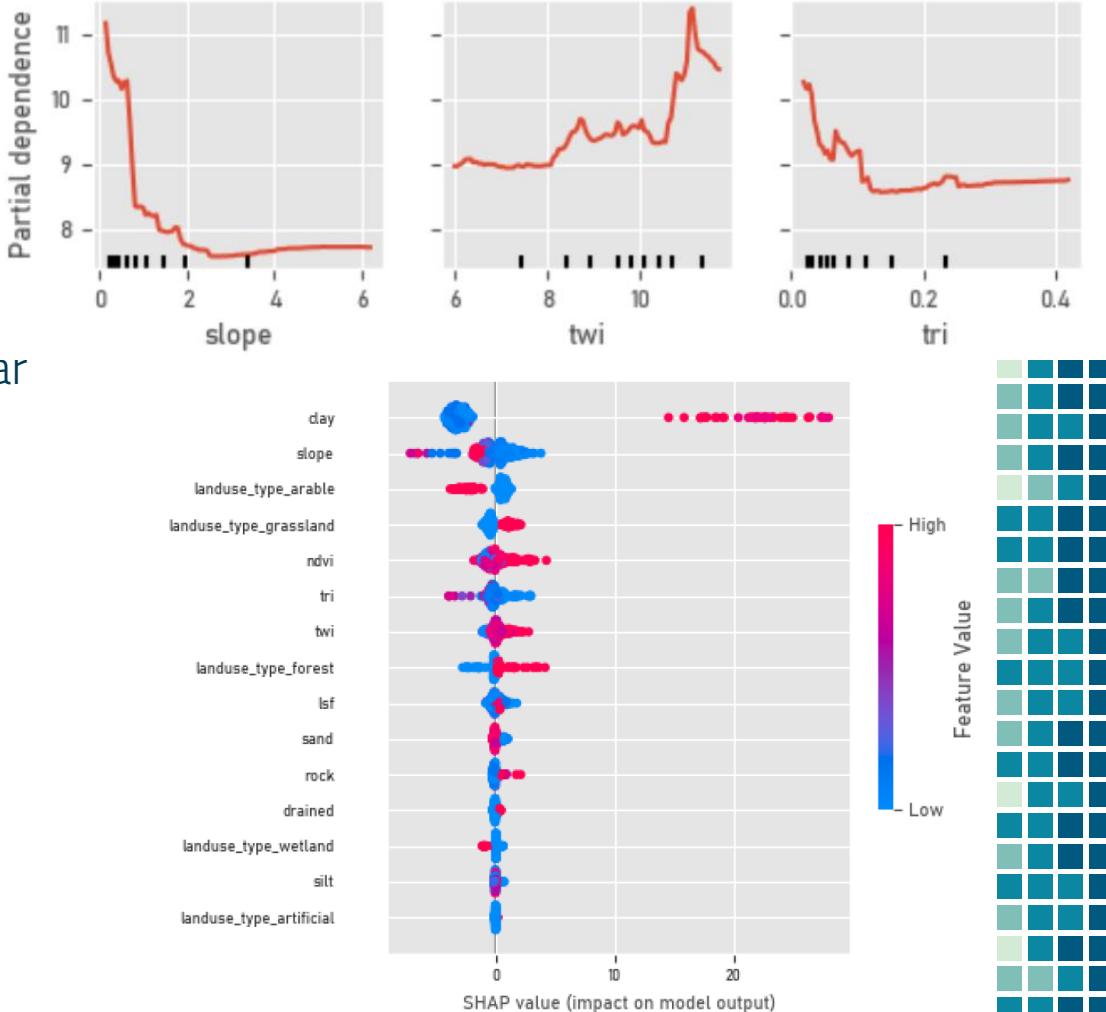
ML in spatial predictive modelling

Covariates aka predictors

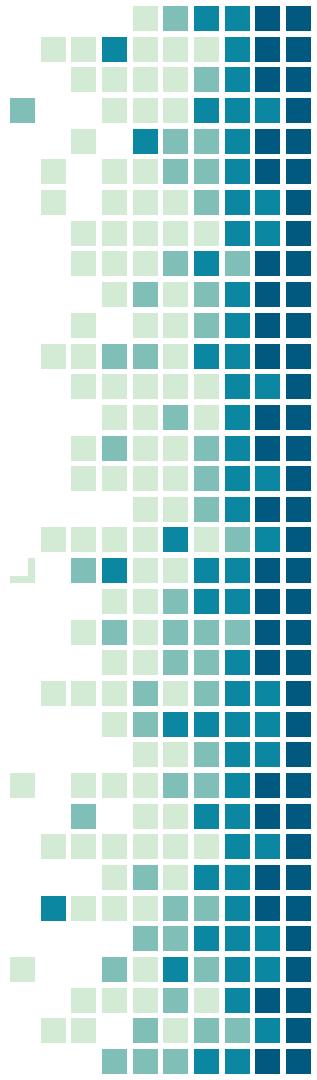


Advantages

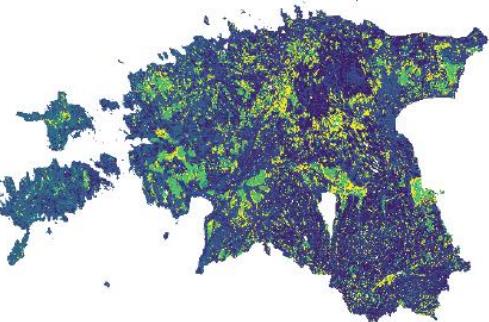
- No assumptions
- Can capture complex non-linear relationships
- Interpretability



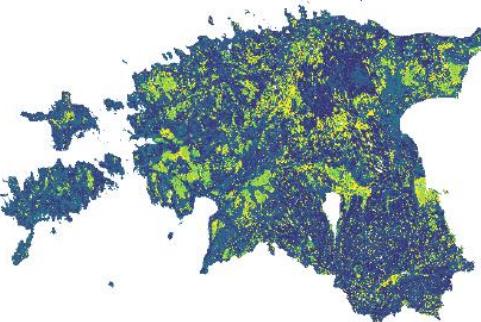
Machine learning is inherently not spatial
...but spatial aspect can be added



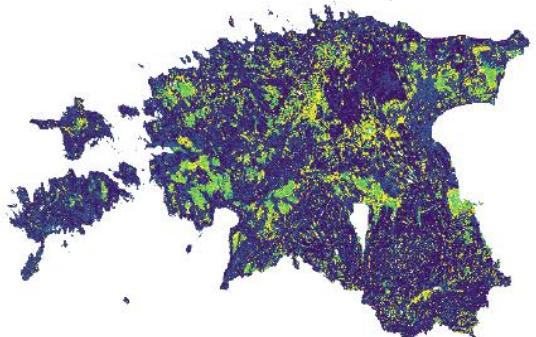
XY $R^2=0.62$



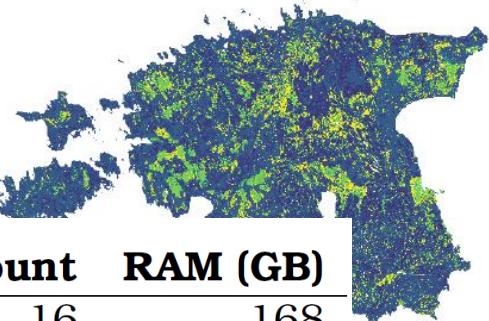
KNN $R^2=0.63$



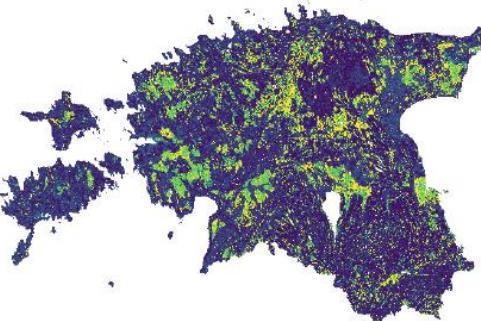
Baseline $R^2=0.59$



BD $R^2=0.62$



RFRK $R^2=0.61$

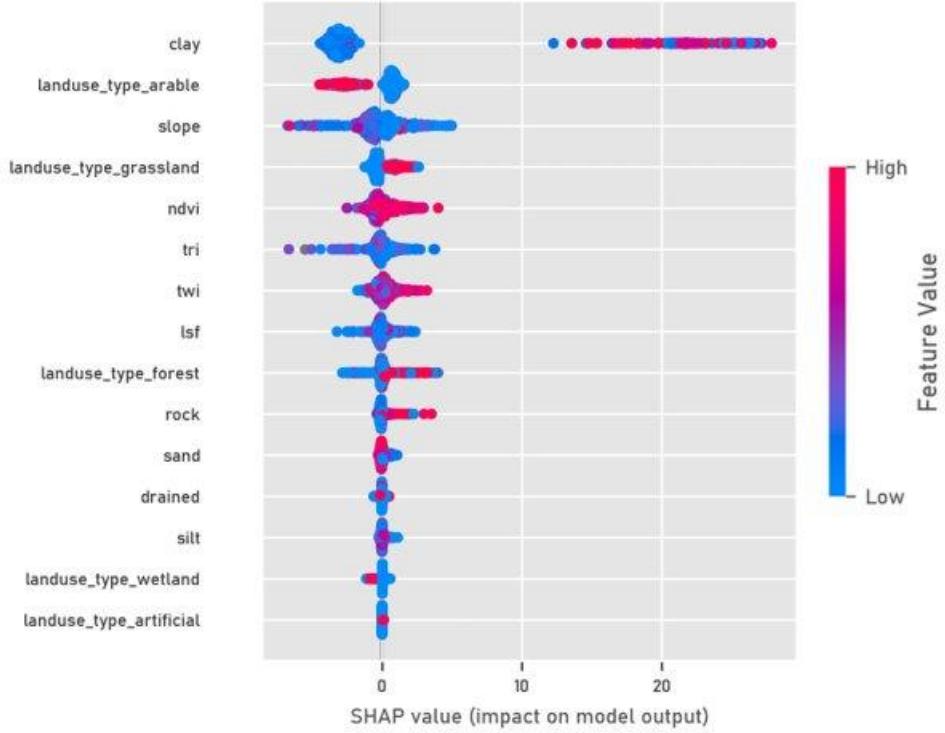


Model	Time	CPU count	RAM (GB)
Baseline raster	1h21m	16	168
XY	1h45m	16	148
KNN	1h43m	16	160
BD	4h13m	16	192
RFRK	1h51m	16	260

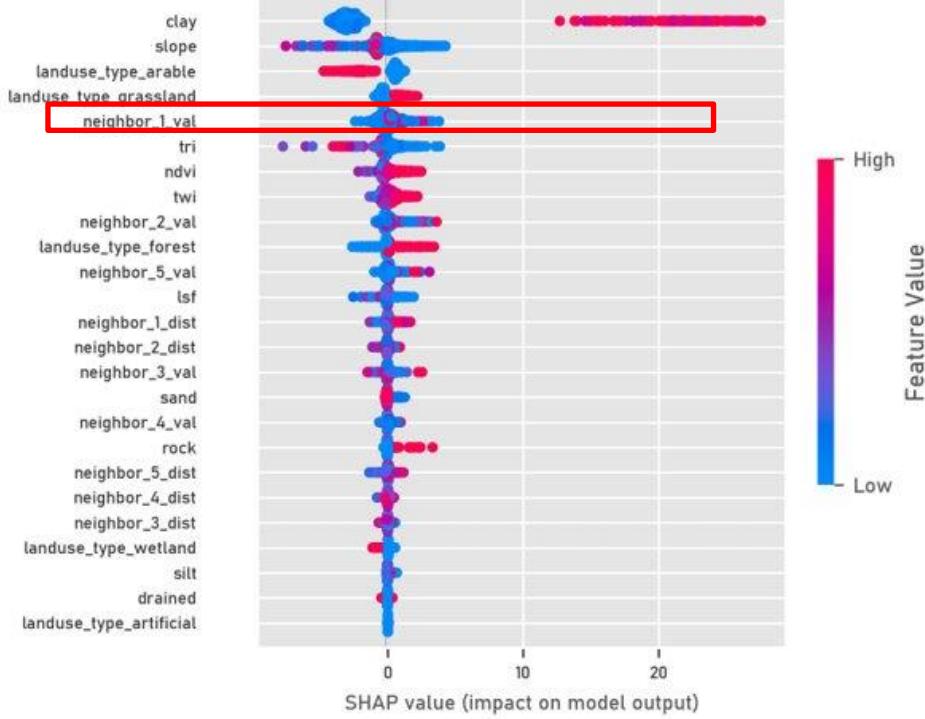


Harrison, 2023

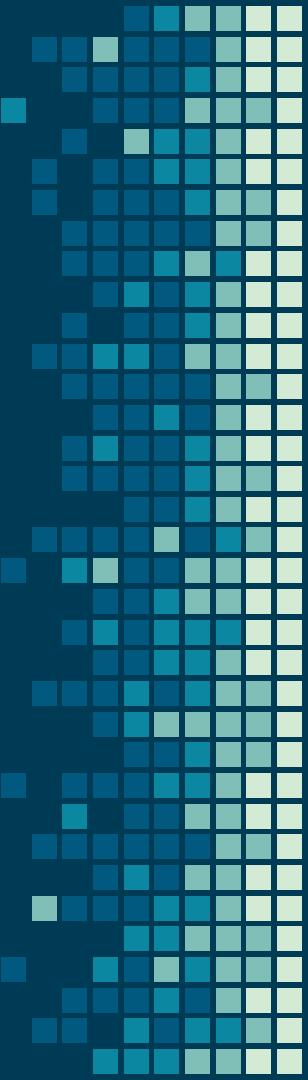
Baseline – no spatial awareness



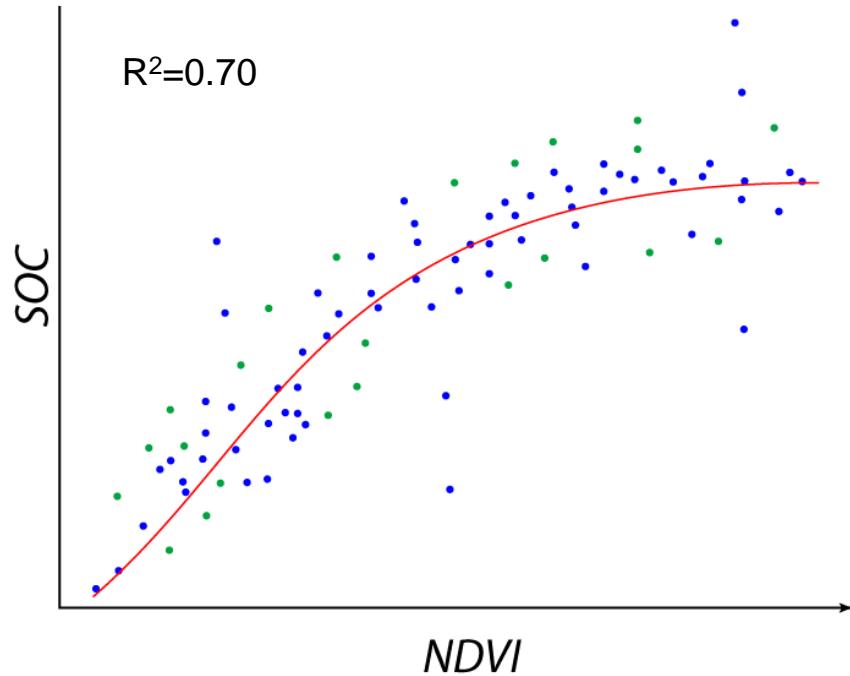
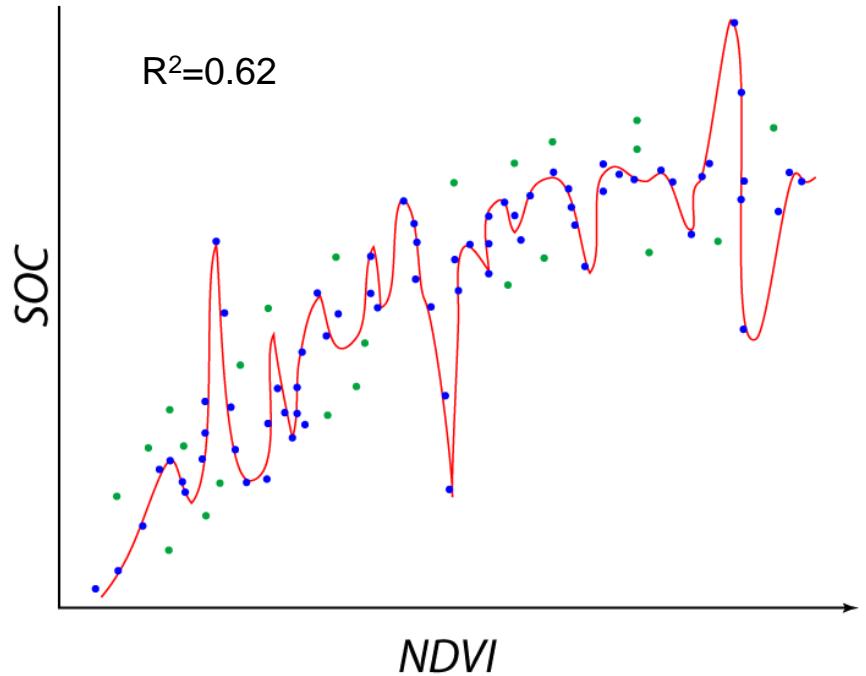
KNN



CHALLENGES



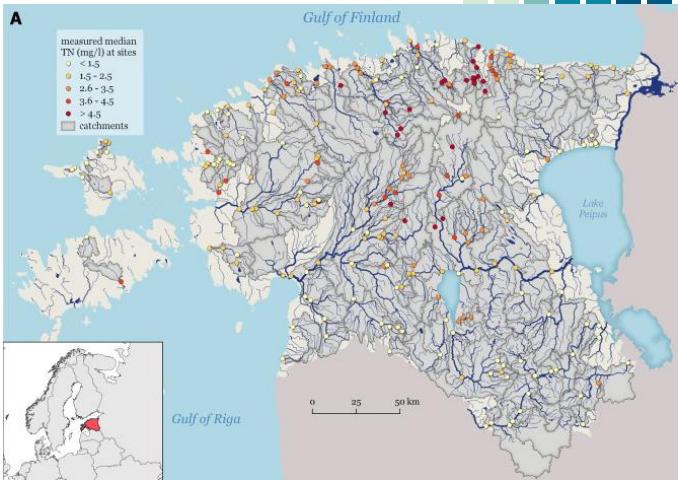
Overfitting



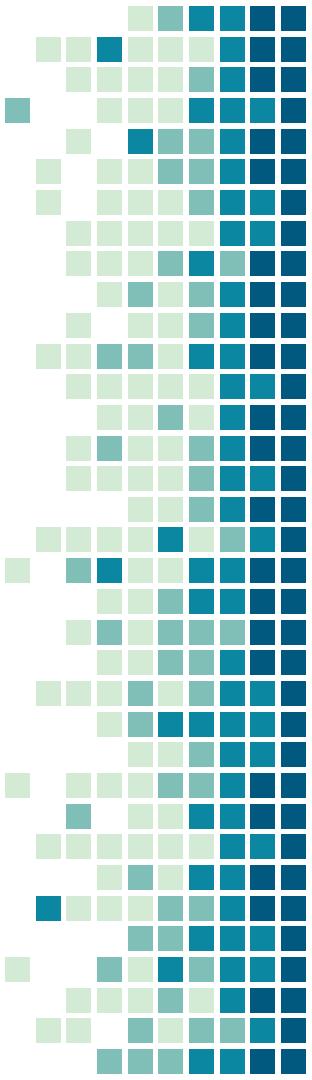
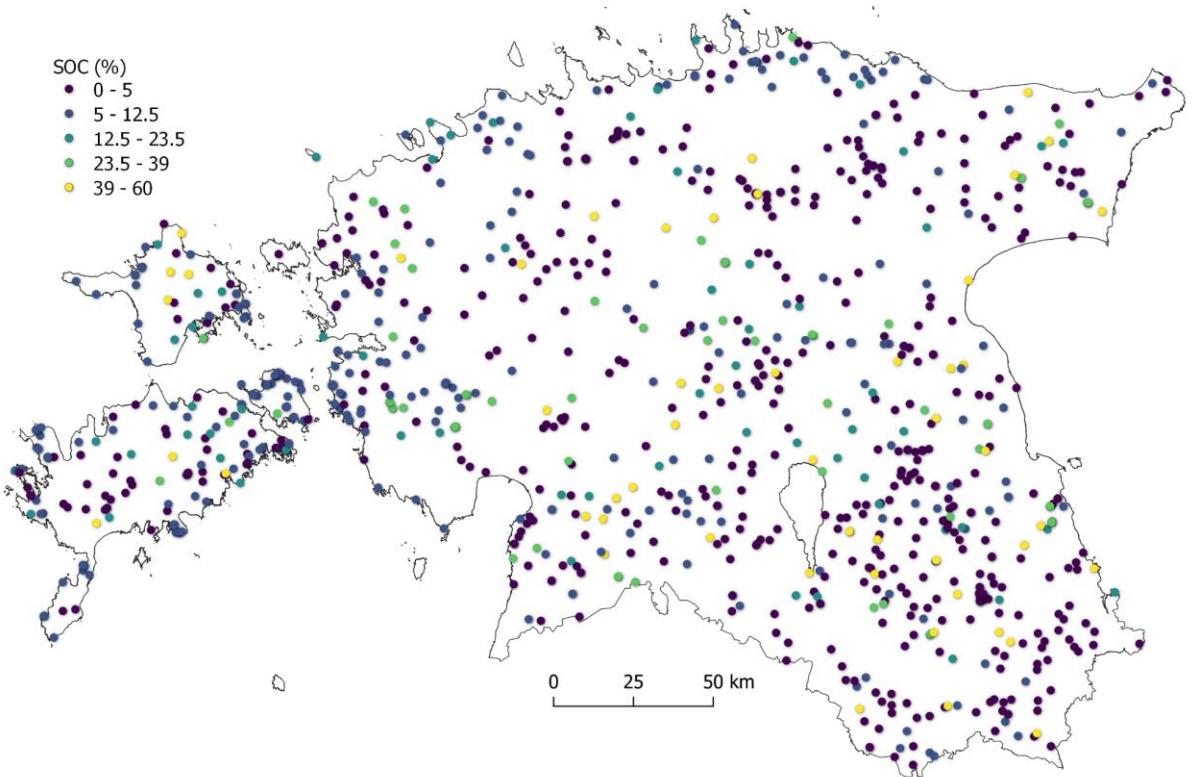
Multidimensionality

- Reducing multidimensionality and keeping only the meaningful covariates usually gives better results

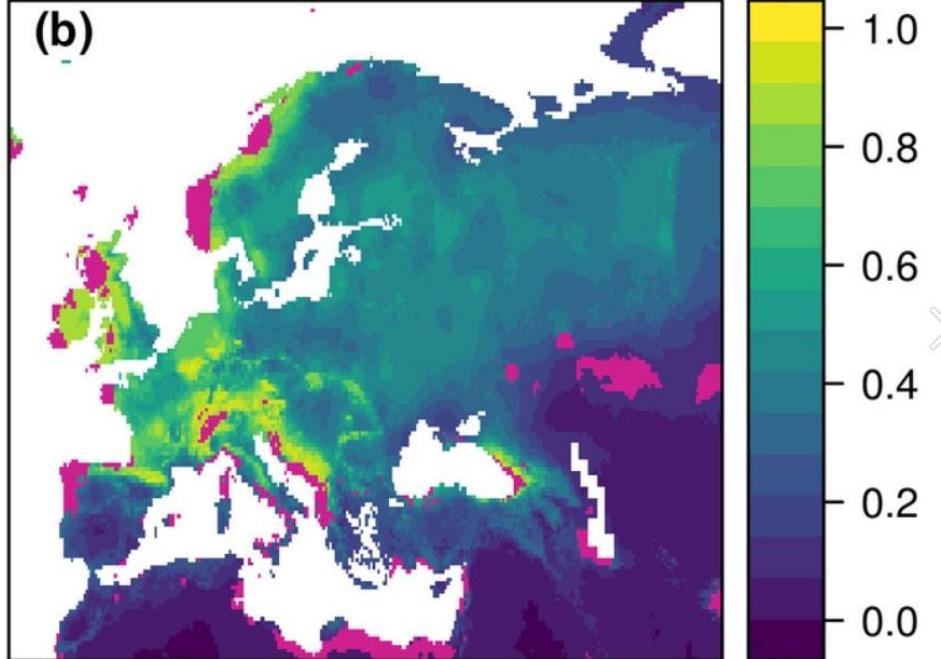
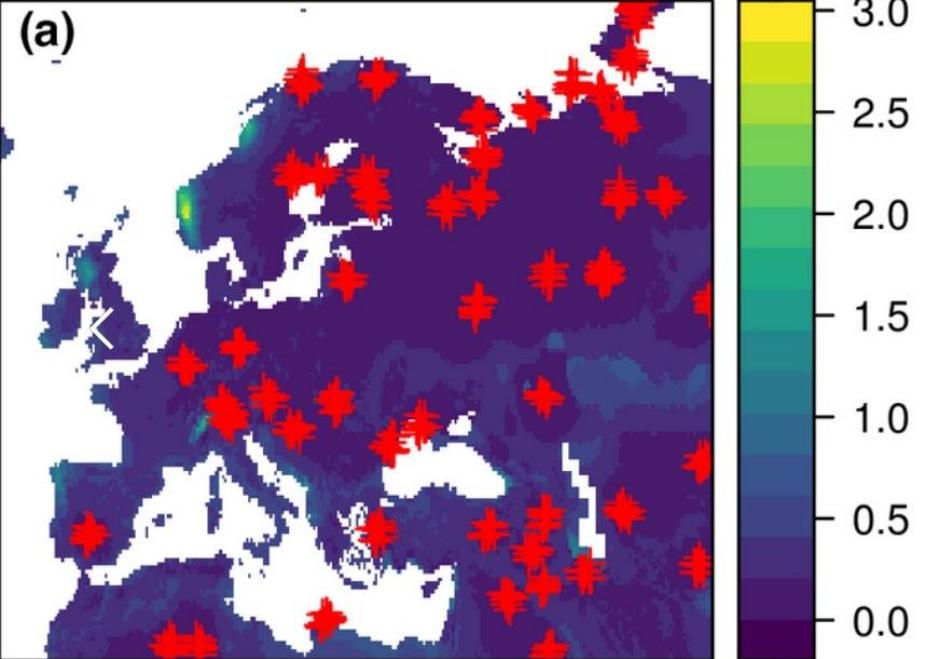
Attribute	TN_MODEL_V1	TN_MODEL_V2	TN_MODEL_V3	TN_MODEL_V4
n_features	62	55	47	38
r2_train	0.88	0.94	0.92	0.93
r2_test	0.79	0.83	0.82	0.83



Sampling (spatial) representativeness?



'Area of applicability' (AOA)



Carbage in, carbage out





Thank you!



evelyn.uuemaa@ut.ee

<https://landscape-geoinformatics.ut.ee/>

