Predicting Soil Properties from Hyperspectral Satellite Images

Frauke Albrecht[1,2], Caroline Arnold[1,2], Roshni Kamath[1,3], Kai Konen[1,4], Rıdvan Salih Kuzu[1,5]

[1]Helmholtz AI [2]German Climate Computing Centre DKRZ, [3]Jülich Supercomputing Centre, [3]Forschungszentrum Jülich GmbH, Germany, [4]Institute for Software Technology, German Aerospace Center (DLR), Cologne, Germany, [5]The Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Germany

INTRODUCTION

- Machine Learning Challenge provided by AI4EO[1]
- AI4EO aims to reduce the gap between Earth Observation and Artificial Intelligence

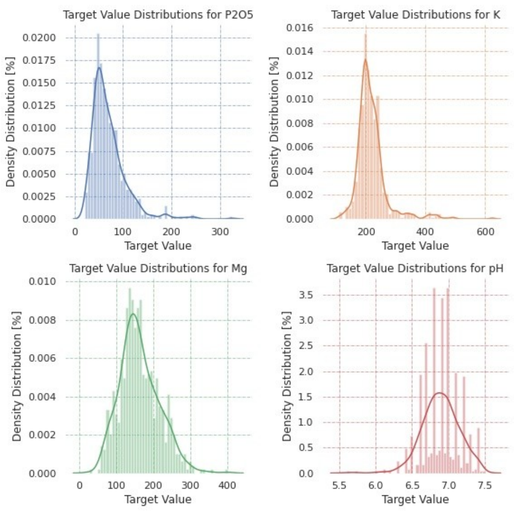


Fig.1: Distributions of target soil parameters

Objective:

- predict agricultural relevant soil parameters
- Computational lightweight solution

DATA

- Airborne hyperspectral images from an unspecified region in Poland
- Soil parameters: potassium (K), magnesium (Mg), phosphorus pentoxide (P2O5), pH value
- Patches according to the boundaries of the agricultural fields.
- 1732 training patches
- Each patch contains 150 hyperspectral bands

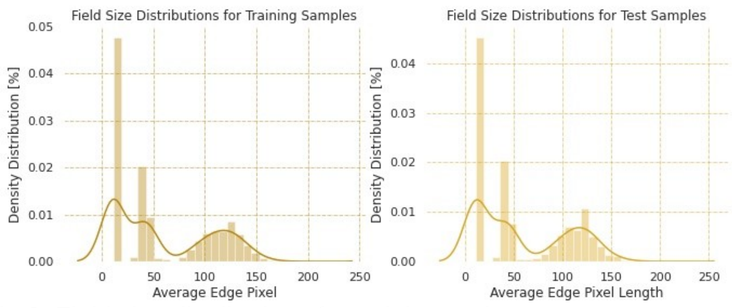


Fig.2: Field size distributions for training and test sets

FEATURE ENGINEERING:

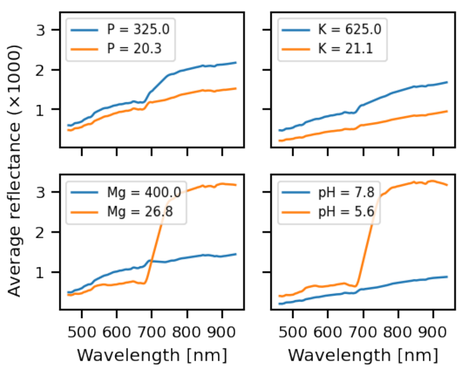
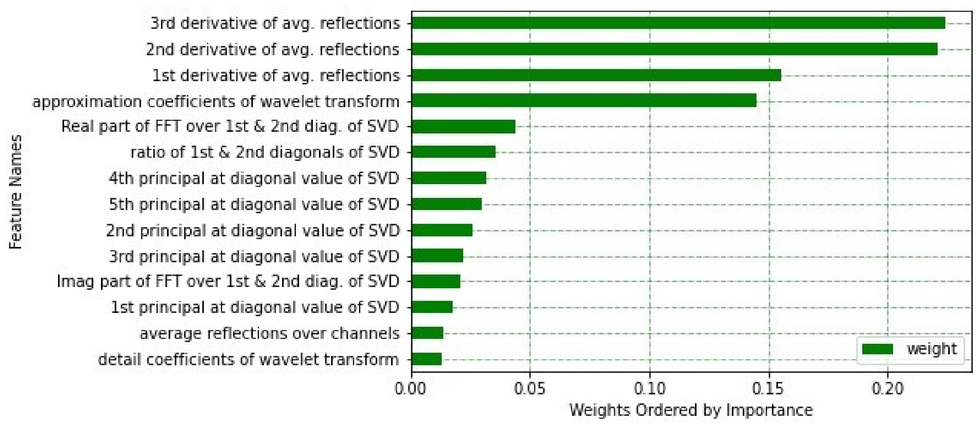


Fig. 3: Average reflectance for example patches

1. Average reflectance, its 1st, 2nd and 3rd order derivative
2. Discrete wavelet transforms of average reflectance: 1st, 2nd, 3rd, 4th level approximation and detail coefficients
3. First 5 diagonal values of the Singular Value Decomposition of each channel
4. The ratio of 1st, 2nd diagonals: σ1/σ2
5. Fast Fourier transform of σ1/σ2: real and imaginary parts

➔ For each field patch, a [1 × 2100] dimensional feature array is extracted For data augmentation, 1% random Gaussian noise is added to both input features and target values.

Fig. 4: Feature Importances



EVALUATION

The evaluation metric measures the improvement upon the baseline ($\mathrm{MSE}_{\mathrm{bl}}$) of predicting the average of each soil parameter.

$$\text{Score} = \frac{1}{4}\sum_{i=1}^{4}\frac{{\mathrm{MSE}_{\mathrm{algo}}}^{(i)}}{{\mathrm{MSE}_{\mathrm{bl}}}^{(i)}}, \quad \text{where:} \quad {\mathrm{MSE}_{\mathrm{algo}}}^{(i)} = \frac{1}{N}\sum_{j=1}^{N}(p_j^{(i)} - \hat{p}_j^{(i)})^2.$$

MODEL

- Best model so far:
  Hybrid model fusing Random Forest and k-Neighest Neighbors regressors
- Hyperparameter tuning using Optuna was included, but best results were achieved with the default settings
- 5-fold cross validation

Field Edge (pixel)	Model	P205	K	Mg	pH	Average
0-11	KNN	1.015	0.984	0.988	0.688	0.916
11+	RF	0.797	0.706	0.573	0.793	0.717
Entire Fields	**Hybrid**	**0.879**	**0.810**	**0.728**	**0.754**	**0.792**
Public Leaderboard Sore on Test Set					**0.79001**	

TEST SET PREDICTION

- 1154 test patches
- Current score on test set: 0.794

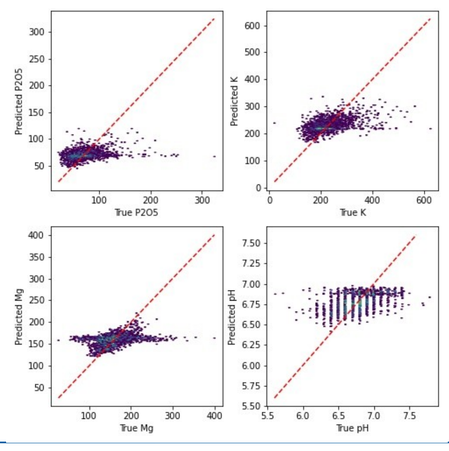


Fig.5: Ground-truth vs. predicted soil parameters

[1] https://platform.ai4eo.eu/seeing-beyond-the-visible