Prediction of certain soil parameters using reflectance spectroscopy and legacy soil data

Primary Supervisor: Dr. Ádám Csorba

Author: Mohamed Yassine Zaghdoudi

Institute of Environmental Sciences

Gödöllő 2024

Introduction

The soil science profession is dealing with an increase in demand for regional, continental, and international databases to monitor soil situations. Nevertheless, such information is lacking. Low-cost equipment for measuring soil characteristics across large areas is required. Soil spectroscopy is an efficient, low-cost, safe, and consistent analytical technique. As a result, we consider that the fundamental purpose of this research is to describe the current status of soil spectroscopy as well as its future application for soil monitoring. In this thesis, we will embark on an adventure to discover various aspects of soil spectroscopy. We will look at the basic ideas and procedures that underpin this technology. Moreover, we will investigate the methodology, data analysis, and practical applications made available by soil spectroscopy. These studies seek to address critical concerns about the ability to identify characteristics of soil, predict soil quality, and have an influence on a wide range of disciplines, from precision agriculture to environmental sustainability and mineral resource management. Overall, we will examine the landscape of soil spectroscopy, looking at its potential, limitations, and future prospects.

Materials and methods

This part of the study endeavored to find out whether the conducted VIS–NIR spectrum analysis information could potentially be used to predict soil characteristics. Therefore, to investigate the predictability of cation exchange capacity (CEC), pH, humus content, and clay content for 400 soil samples collected.

Pre-processing Methods:

Initially, the reflectance spectra are transformed into absorbance spectra to enhance spectral features. Signal processing involves noise removal, where a Savitzky-Golay filter is used, and differentiation to improve spectral characteristics. Scatter correction techniques like Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) are applied as well to reduce the impact of scattered light.

Calibration Sampling:

Kennard-Stone sampling (KSS) is used to select calibration samples evenly across the spectral distribution. The sample size is determined based on the mean squared Euclidean distance (msd) between the estimated probability density functions of subsets and the entire sample set.

Chemometric Analysis:

Partial Least Squares Regression (PLSR) with leave-one-out cross-validation is employed to calibrate spectral data with laboratory soil data. More specifically, PLSR selects orthogonal factors minimizing covariance between predictor and response variables, aiming to explain most variation. Models are evaluated based on root mean squared error of predictions (RMSEP) and cross-validated RMSE (RMSECV) to ensure model parsimony.

Results and discussion

After preprocessing and eliminating spectral outliers, 355 soil samples remained. Test-set validation was used to confirm the prediction models' stability with approximately 70 % calibration (around 255 samples) and 30 % validation samples (around 255 samples).

Characteristics of Soil Spectral Curves:

The reflectance spectra of 379 soil samples exhibited comparable overall shapes with distinct variations attributable to soil characteristics. Notably, the mean squared Euclidean distance (msd) values declined as the calibration set size increased, stabilizing after 100 samples. This observation informed the selection of 100 samples as the ideal size for calibrating the models.

Multivariate Calibration:

Utilizing Partial Least Squares Regression (PLSR), we achieved robust calibration models for soil properties. For instance, the optimal number of components for clay, humus, pH, and CEC were found to be 18, 16, 15, and 12, respectively.

Evaluation of Predictive Models:

The measured versus predicted plots revealed strong correlations between predicted and actual values for clay content and pH, with R^2 values of 0.91 and 0.85, respectively. However, humus content showed slightly lower predictive accuracy with an R^2 value of 0.72. CEC predictions exhibited moderate accuracy with an R^2 value of 0.61, indicating some variability in predictions.

Insights from Loading Plots:

Loading plots unveiled significant wavelengths influencing soil property predictions. For instance, clay content exhibited strong positive loadings between 350 nm and 500 nm, while humus and CEC showed prominent positive loadings around 1850 nm. These wavelengths are crucial for capturing information related to soil properties, enhancing the interpretability and predictive performance of the calibration models.