Vis-NIR spectroscopy for multivariate classification of soil samples throughout Denmark into Danish Soil Classification System

Sayed Salman Tabatabai[1], Maria Knadel[1], and Mogens Greve[1]

[1]Aarhus University

November 22, 2022

Abstract

Different spectroscopy methods such as visible near-infrared (Vis-NIR) spectroscopy have proven to provide useful information on soil physical and chemical properties. The majority of previous studies have focused on multivariate regression methods such as partial least squares regression (PLSR) to predict soil characteristic features from soil spectra. The objective of these efforts was to provide precision data for agricultural and land management practices where the knowledge about the soil differences in each part of the field can be used for variable-rate irrigation, seeding, liming, fertilising and pesticide application. Since the currently available machines can only apply few (less than 10) variable classes of seed, fertiliser, etc., using soil classification seems to be a more appropriate option compared to regression. As a part of Danish National Soil Survey in 1980s, 2460 soil samples were collected from 789 soil profiles in 4 depths (0-30, 30-60, 60-100 and 100-200cm) throughout Denmark (some profiles did not have samples from all 4 depths) and tested for several soil characteristics including complete soil texture, organic carbon content and calcium carbonate in the lab. Based on these soil characteristics, all soil samples were classified into 8 soil types or 12 soil classes in the Danish Soil Classification System (JB system). Later the Vis-NIR spectra of samples were measured using a FOSS DS2500 spectrometer in the range of 350-2500nm. Partial least squares and support vector machines discriminant analyses (PLS-DA and SVM-DA, respectively) were used to calibrate and cross-validate classification models where soil Vis-NIR spectra were used to classify each sample from each depths into its corresponding JB soil type and soil class. The results show excellent classification accuracy and specificity (>80% and >90%, respectively) for samples from the same depths and when samples from all depths were combined. We found that the high false negative rate (low sensitivity) was mainly due to the models classifying samples in the neighbouring classes of the actual class (e.g. classifying a sample in JBC 2 as belonging to JBC 1 or JBC 3). This clearly indicates that calibrating the classification model on the uncertain hydrometer data (with 1.4-2% reproducibility error) was the main reason for classification of samples in the neighbourhood of the actual class. In conclusion, given the highly uncertain reference methods for soil classification, using Vis-NIR spectroscopy with PLS-DA provides a very rapid, inexpensive, and highly reliable method for soil texture classification on a national scale.

Vis-NIR spectroscopy for multivariate classification of soil samples throughout Denmark into Danish Soil Classification System

Salman Tabatabai*, Maria Knadel, Mogens H. Greve

Department of Agroecology, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark

* Corresponding and presenting author:
E: st@agro.au.dk || salman.tabatabai@gmail.com || T: +45 52793647

Extended Abstract:

Different spectroscopy methods such as visible near-infrared (Vis-NIR) spectroscopy have proven to provide useful information on soil physical and chemical properties. The majority of previous studies have focused on multivariate regression methods such as partial least squares regression (PLSR) to predict soil characteristic features from soil spectra. The objective of these efforts was to provide precision data for agricultural and land management practices where the knowledge about the soil differences in each part of the field can be used for variable-rate irrigation, seeding, liming, fertilising and pesticide application. Since the currently available machines can only apply few (less than 10) variable classes of seed, fertiliser, etc., using soil classification seems to be a more appropriate option compared to regression. As a part of Danish National Soil Survey in 1980s, 2460 soil samples were collected from 789 soil profiles in 4 depths (0-30, 30-60, 60-100 and 100-200cm) throughout Denmark (some profiles did not have samples from all 4 depths) and tested for several soil characteristics including complete soil texture, organic carbon content and calcium carbonate in the lab. Based on these soil characteristics, all soil samples were classified into 8 soil types or 12 soil classes in the Danish Soil Classification System (JB system). Later the Vis-NIR spectra of samples were measured using a FOSS DS2500 spectrometer in the range of 350-2500nm. Partial least squares and support vector machines discriminant analyses (PLS-DA and SVM-DA, respectively) were used to calibrate and cross-validate classification models where soil Vis-NIR spectra were used to classify each sample from each depths into its corresponding JB soil type and soil class. The results show excellent classification accuracy and specificity (>80% and >90%, respectively) for samples from the same depths and when samples from all depths were combined. We found that the high false negative rate (low sensitivity) was mainly due to the models classifying samples in the neighbouring classes of the actual class (e.g. classifying a sample in JBC 2 as belonging to JBC 1 or JBC 3). This clearly indicates that calibrating the classification model on the uncertain hydrometer data (with 1.4-2% reproducibility error) was the main reason for classification of samples in the neighbourhood of the actual class. In conclusion, given the highly uncertain reference methods for soil classification, using Vis-NIR spectroscopy with PLS-DA provides a very rapid, inexpensive, and highly reliable method for soil texture classification on a national scale.

A comprehensive study with more details is currently under review and will be published soon.

Vis-NIR spectroscopy for multivariate classification of soil samples throughout Denmark into Danish Soil Classification System

Salman Tabatabai*, Maria Knadel, Mogens H. Greve

Department of Agroecology, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark

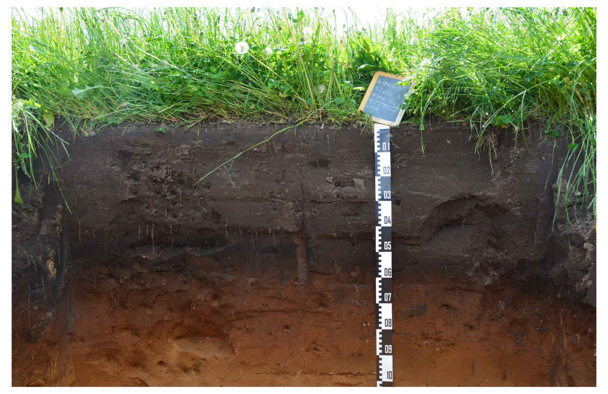
AARHUS UNIVERSITY

* Corresponding Author: Salman Tabatabai || Email: st@agro.au.dk || salman.tabatabai@gmail.com || Phone: +45 52793647

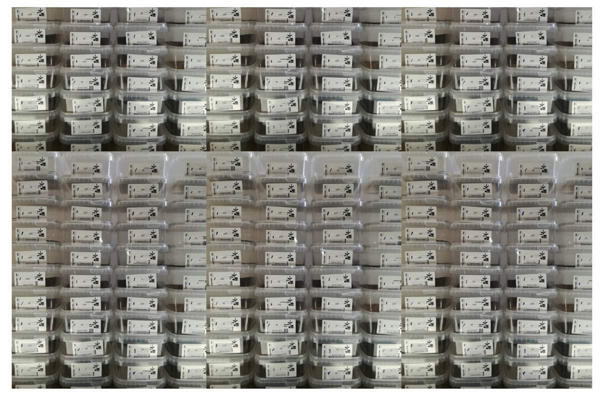
European Geosciences Union

Photography encouraged

The Problem

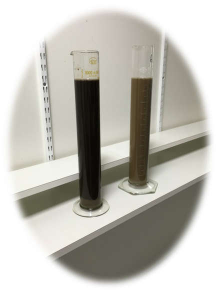


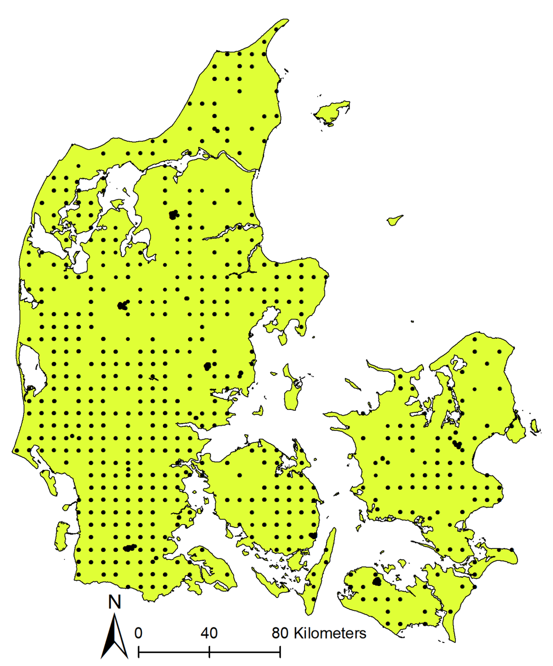
2/3 of Denmark under agriculture!



100's of soil profiles!

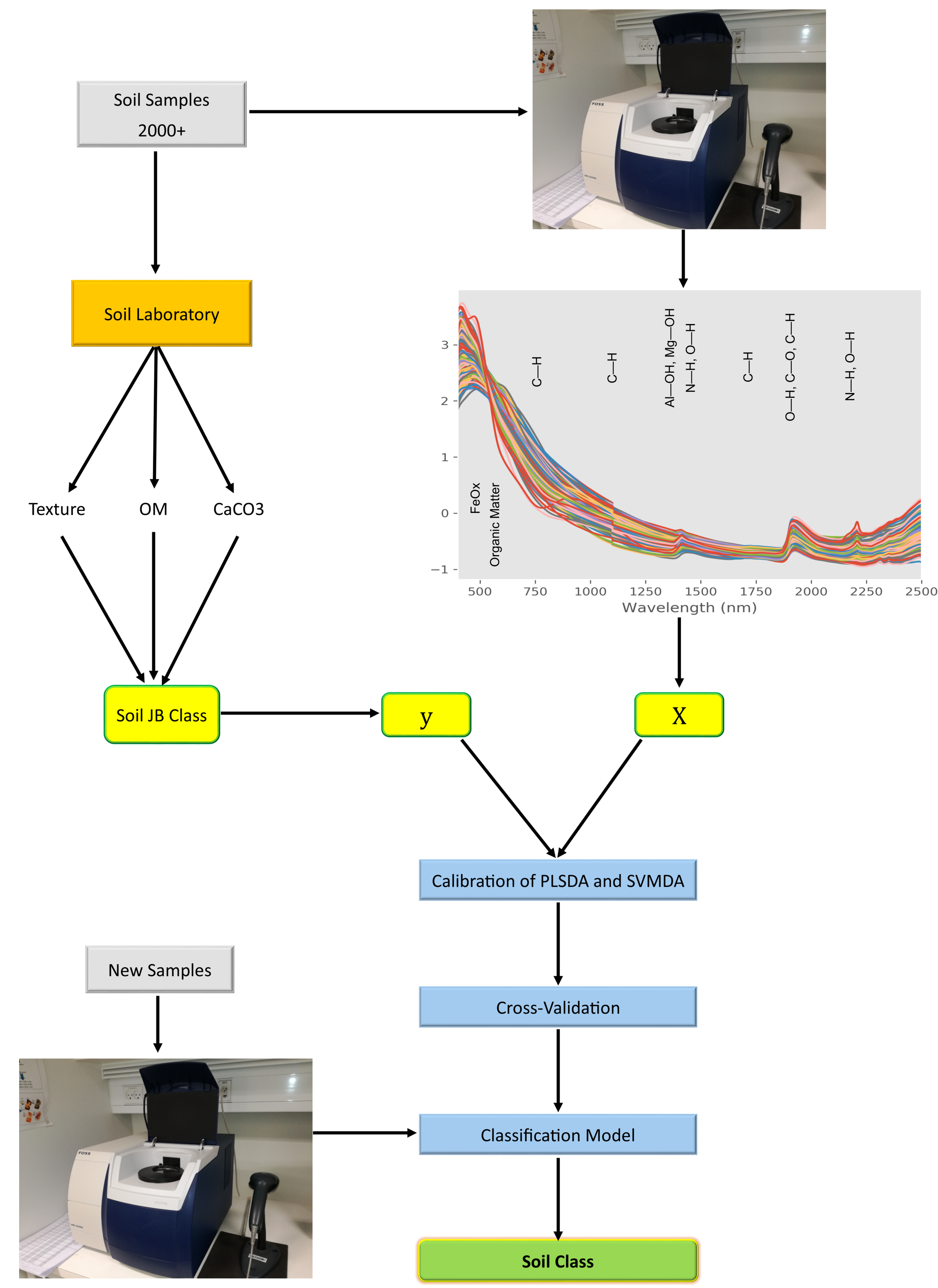
1000's of soil samples!





Soil Type	JBC	Percentage by Weight					
		Clay <2 µm	Silt 2-20 µm	Fine Sand 20-200 µm	Total Sand 20-2000 µm	Organic Matter 59.7% C	CaCO3
Coarse Sandy Soil	1	0-5	0-20	0-50	75-100	≤ 10	≤ 10
Fine Sandy Soil	2			50-100			
Clayey Sandy Soil	3	5-10	0-25	0-40	65-95		
	4			40-95			
Sandy Clayey Soil	5	10-15	0-30	0-40	56-90		
	6			40-90			
Clayey Soil	7	15-25	0-35		40-85		
Heavy Clayey Soil or Silty Soil	8	25-45	0-45		10-75		
	9	45-100	0-50		0-55		
	10	0-50	20-100		0-80		
Organic Soil	11					> 10	0-90
Calcareous Soil and Atypic Soil	12					≤ 10	> 10

Our Proposed Solution



What We did

What needs to be done from now on

Confusion Matrix

Actual Class / Predicted Class

Predicted \ Actual	1	2	3	4	5	6
1	46	28	13	7	1	0
2	24	33	7	10	2	0
3	10	7	35	13	3	1
4	4	15	18	39	5	1
5	0	0	7	9	58	14
6	0	1	4	6	13	58

For Class 1

- True Positive (TP)
- False Positive (FP)
- False Negative (FN)
- True Negative (TN)

Results

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Ratio of all true instances (TP and TN) to all samples

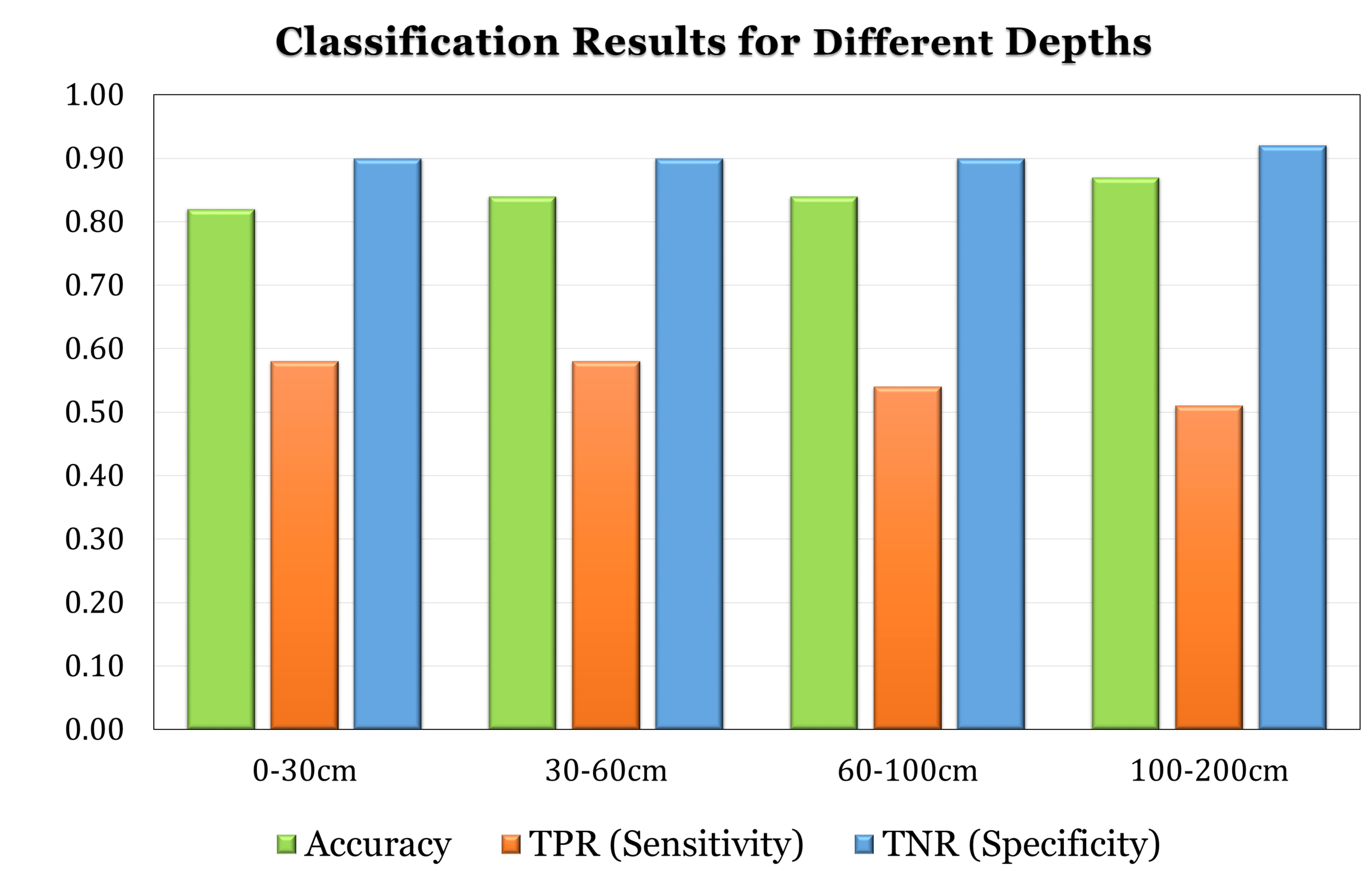
$$\text{TPR (True Positive Rate or Sensitivity)} = \frac{TP}{TP+FN}$$

Estimation of the model capability to avoid false negative (FN)

$$\text{TNR (True Negative Rate of Specificity)} = \frac{TN}{TN+FP}$$

Estimation of the model capability to avoid false positive (FP)

Calculated for each class



- Low Sensitivity (high FN) was due to some samples being classified in neighbouring classes (e.g. Class 2 classified as class 1 or class 3).
- Traditional texture determination (reference methods) are highly uncertain (reproducibility error of 1.4-2% for clay, minimum 1% for fine sand).
- Reproducibility error of traditional methods can change class of a soil sample 1 or 2 classes.