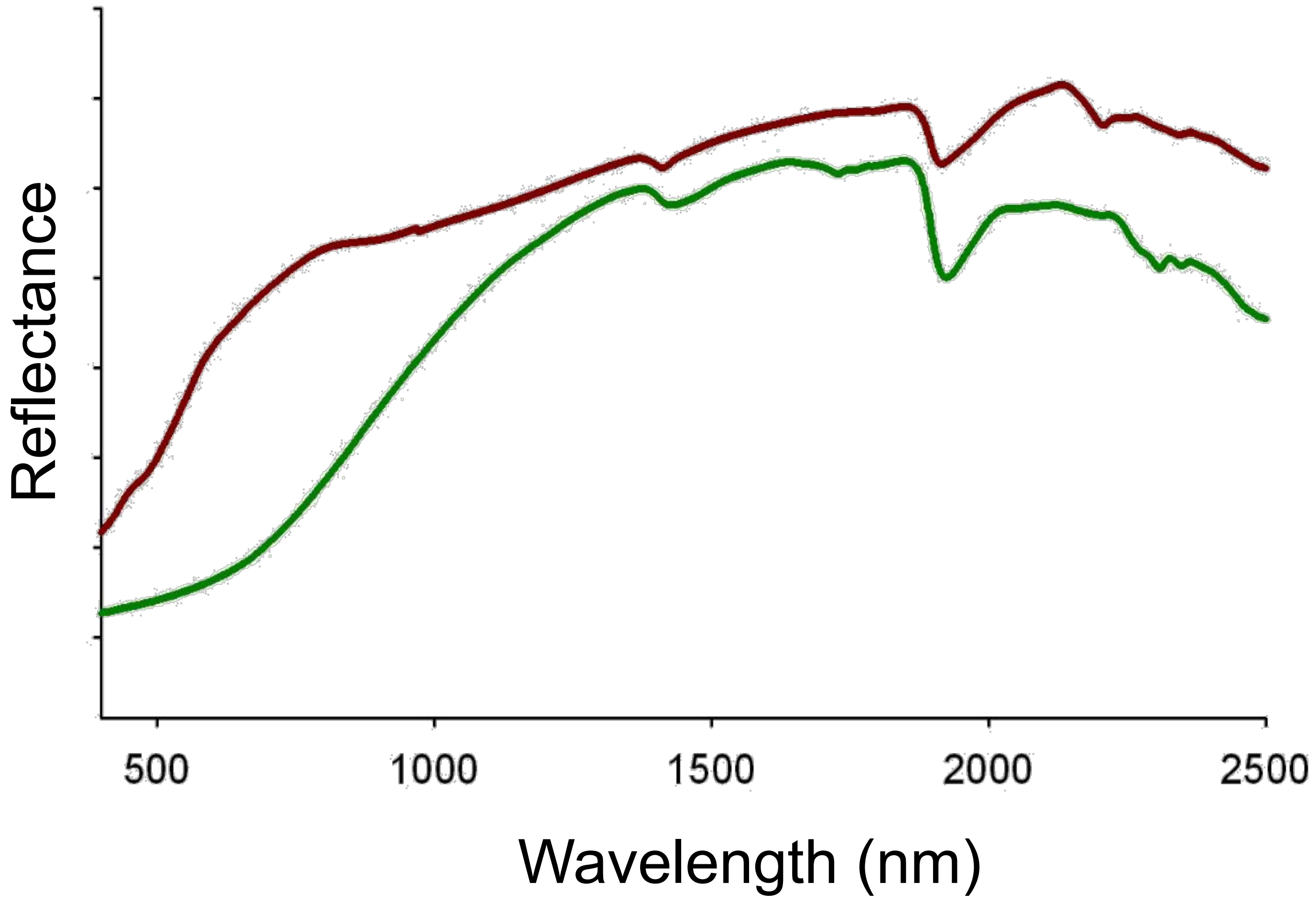
An introduction to soil spectroscopy

Bo Stenberg

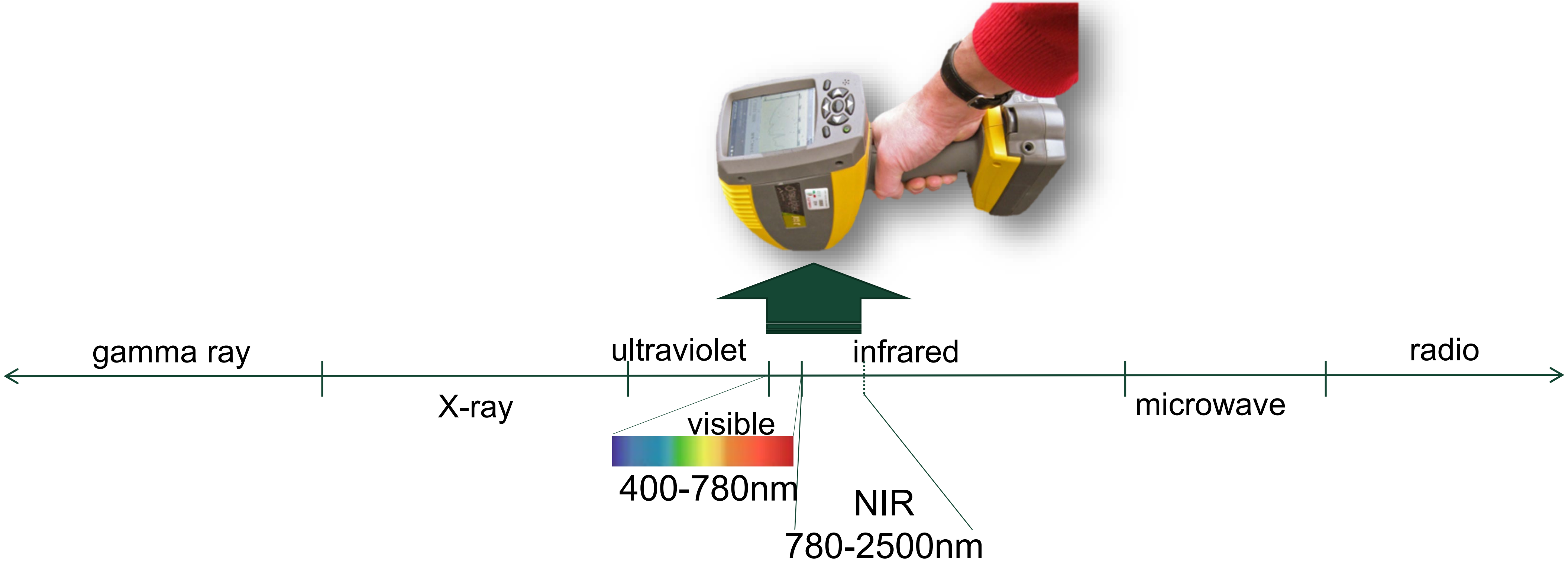
Soil and Environment
Swedish university of Agricultural Sciences

SCIENCE AND EDUCATION
FOR SUSTAINABLE LIFE

Soil Visible and near infrared (vis-NIR) Spectroscopy



Vis-NIR spectroscopy



gamma ray

X-ray

ultraviolet

visible

400-780nm

infrared

NIR

780-2500nm

microwave

radio

Frequency 28 000 – 4 000 cm^{-1}

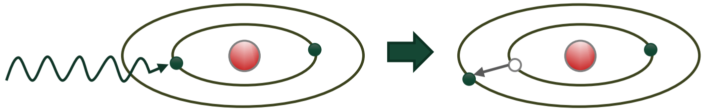
Source for confusion:
NIR: 780-1000 nm
SWIR: 1000-2500 nm

Advantages with Vis-NIR

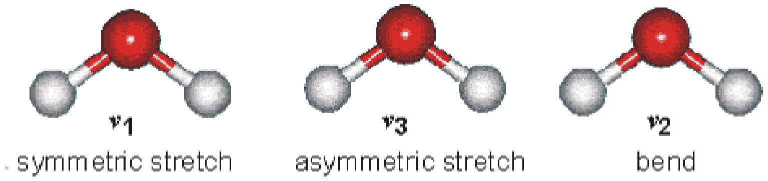
+ Rapid analysis in real-time
+ Minimum sample preparation
+ Non-destructive, Non-invasive
+ Flexible
+ Several parameters can be analyzed simultaneously
+ Rugged instruments for field analyses can be built

Visible and near infrared spectroscopy (vis-NIR)

- Absorption at different wavelengths hold information on the chemical composition of the material.
- In the visible region absorption is due to excitation of electrons.

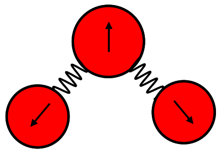


- With longer wavelengths, the absorptions is due to vibrations in chemical bonds within molecules, with the primary absorption in the MIR and overtones and combinations in the NIR region.

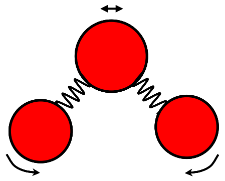

ν_1
symmetric stretch
ν_3
asymmetric stretch
ν_2
bend

B

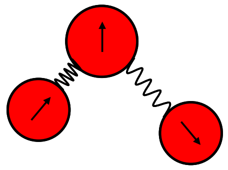
Example for liquid H_2O



Symmetric stretching
ν1



Bending
ν2



Asymmetric stretching
ν3

Absorption	Frequency	Wavelengh
Fundamentals	ν_1, ν_2, ν_3	$\lambda_{1,}\ \lambda_{2,}\ \lambda_3$
1st Overtone	$2\nu_1, 2\nu_2, 2\nu_3$	$½\,\lambda_{1,,}\ ½\,\lambda_{2,,}\ ½\,\lambda_3$
Combinations	$\nu_1 + \nu_2$ or $\nu_2 + \nu_3$	$1/(1/\lambda_1 + 1/\lambda_2)$ or $1/(1/\lambda_2 + 1/\lambda_3)$

	Fundamental	1st Overtone	Combinations
	(λ nm)	(λ nm)	(λ nm)
ν_1	3050	1525	2031 ($\nu_1 + \nu_2$)
ν_2	6080	3040	
$\nu_{3`}$	2870	1435	1950 ($\nu_2 + \nu_3$)

B

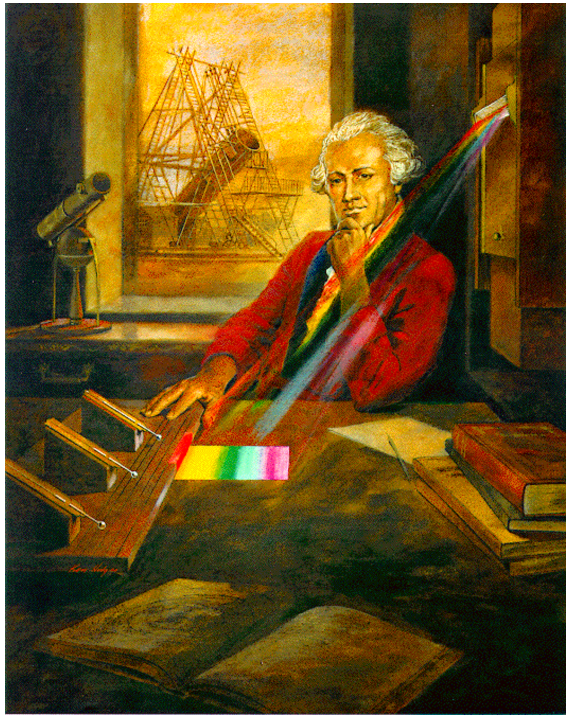
What kind of information can be found in the vis-NIR?

- Strong absorption by water (1400 & 1900 nm)
- Absorbtion by vibrations of molecular bonds (C-H, O-H, N-H and C=O), Combination bands and overtones from longer wavelengths

Successfully used for numerous applications:

- Agriculture – Forage and grain quality, soil mapping...
- Food industry – Process and quality control...
- Pharmaceutical industry – Process and quality control...
- Medicine – Blood status, body fat…
- Petrochemical industry – quality, polymers, plastics…
- Etc.

Pioneers for a wider use

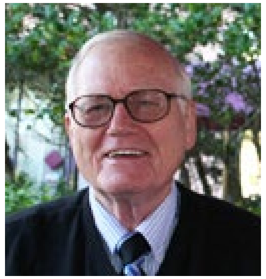


William Herschel discovered NIR radiation in 1800

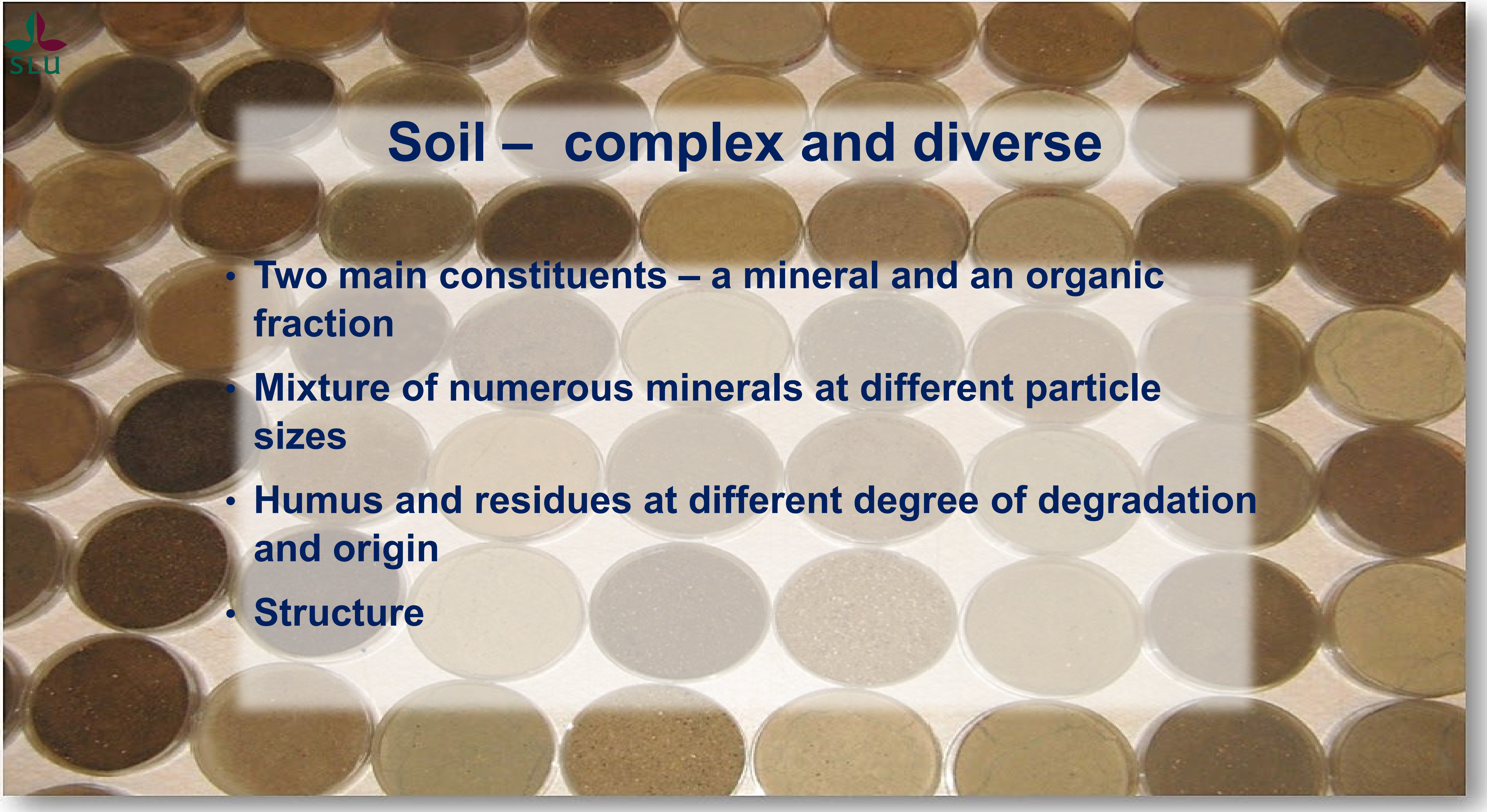
Karl Norris at USDA chose NIR for fast, quantitative determination of moisture, protein and oil in wheat in the 1950s. "Father of modern NIR spectroscopy"

He applied MLR for calibration in a publication in 1968 – the multivariate approach.

Phil Williams converted protein testing for Canada's export wheat to NIRS on the first commercial instrument (Dickey-John) in 1975-1976. Spread the technology over the world.



John S. Shenk installed in 1978 a portable instrument in a mobile van that brought the technology to the farm, hay-markets, etc. Former of ISI software

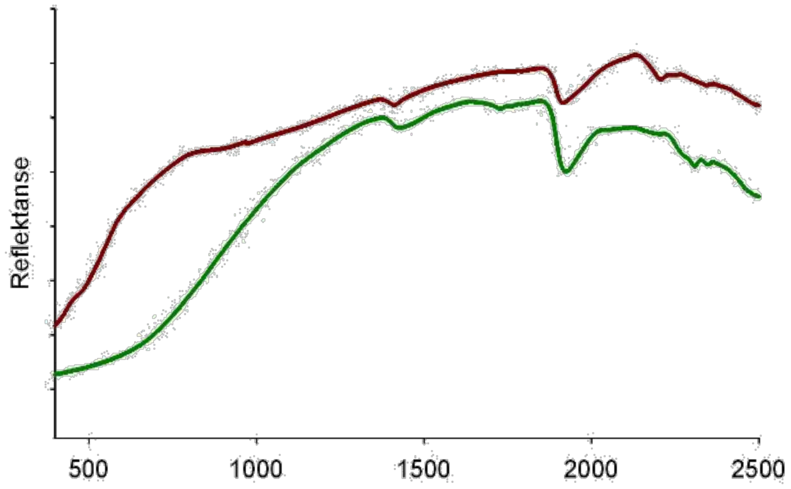


Soil – complex and diverse

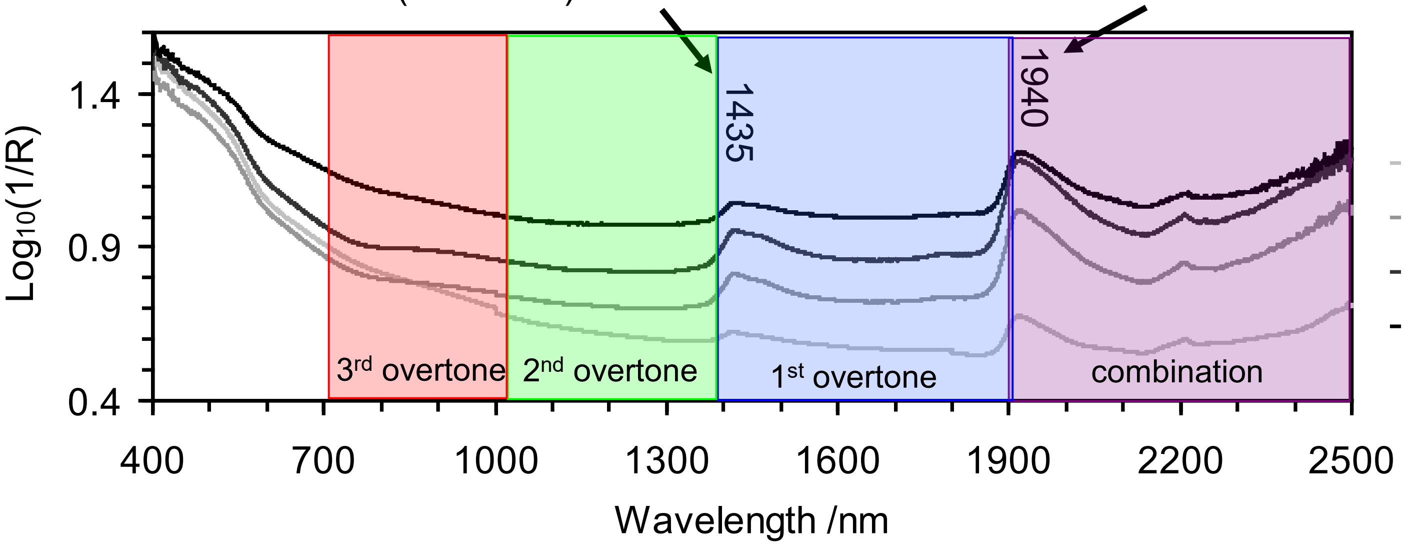
- Two main constituents – a mineral and an organic fraction
- Mixture of numerous minerals at different particle sizes
- Humus and residues at different degree of degradation and origin
- Structure

Main factors influencing soil spectra

Water
Clay minerals
Organic matter
Structure



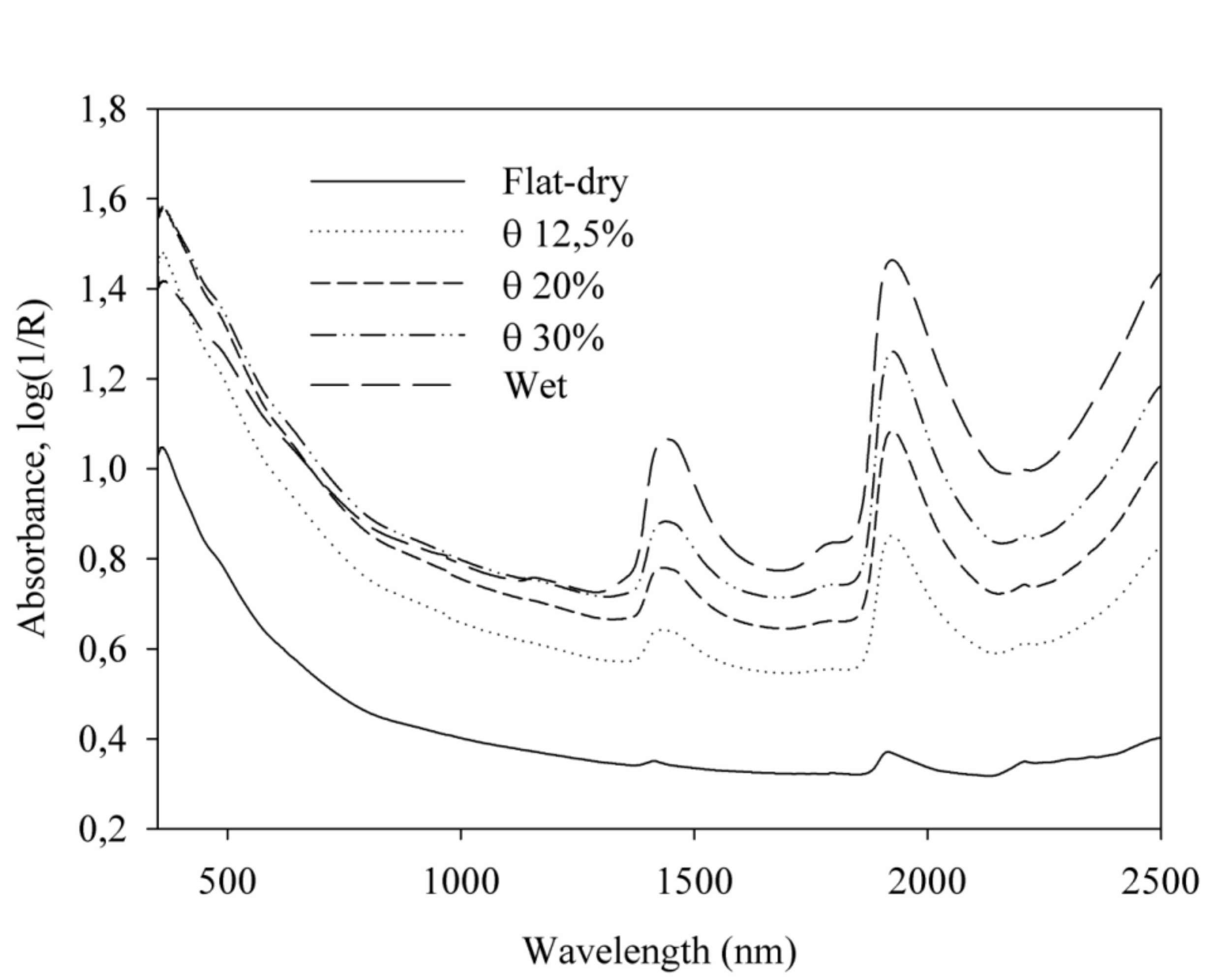
Main soil spectral features (Water)



1st overtone of OH stretch fundamental at 2870nm (3480cm^{-1})

H_2O bend & OH stretch combination of fundamentals at 6080nm & 2870nm (1645 cm^{-1} & 3480cm^{-1})

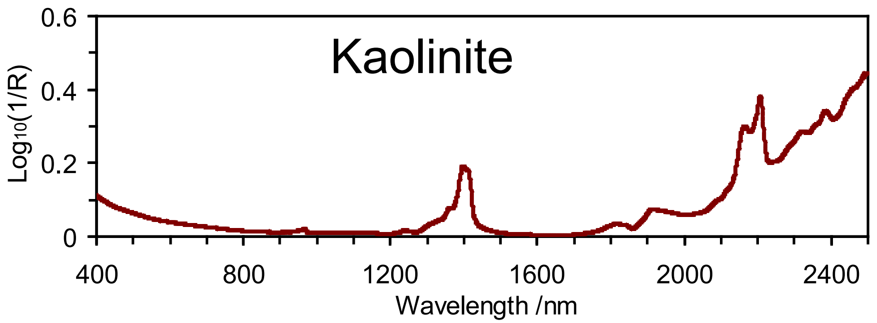
General influence of water on spectra

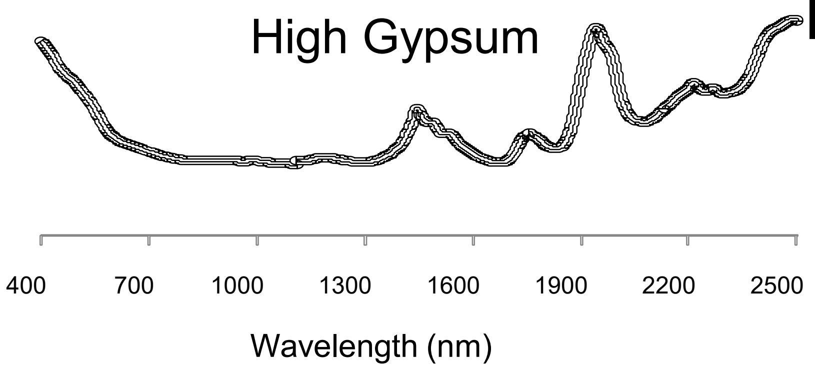


Clay fraction

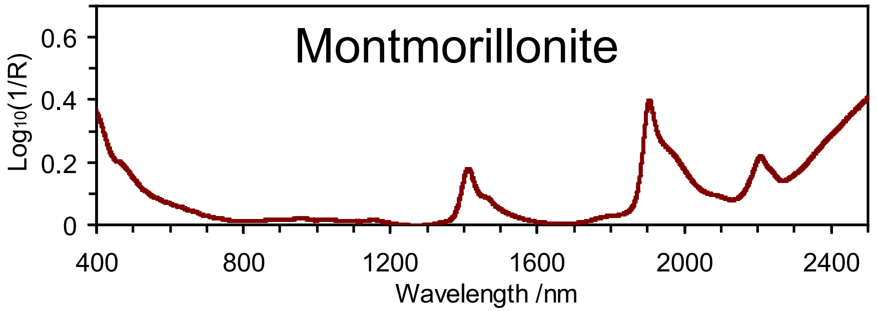
- **Size fraction <0.002 mm (2µ)**
- Secondary, layered clay minerals (illite, smectite, kaolinite, etc.)
- Primary minerals (quartz, feldspars, etc.)
- Carbonates
- Sesquioxides or metal oxides (gibbsite, goethite, hematite)

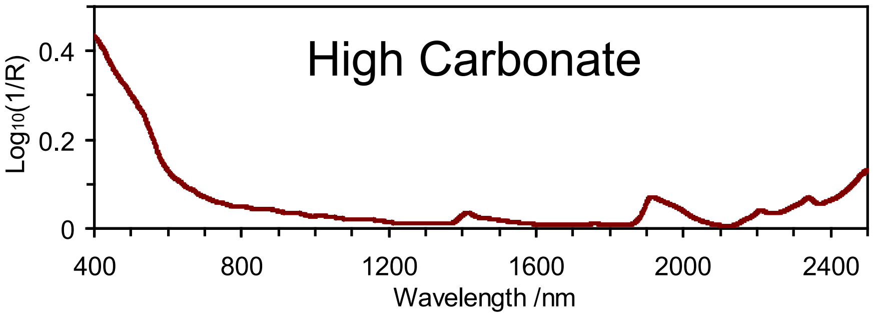
Vis-NIR clay mineral fingerprints

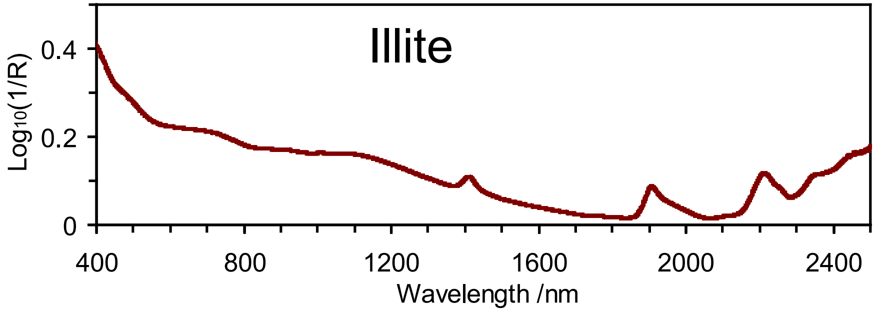


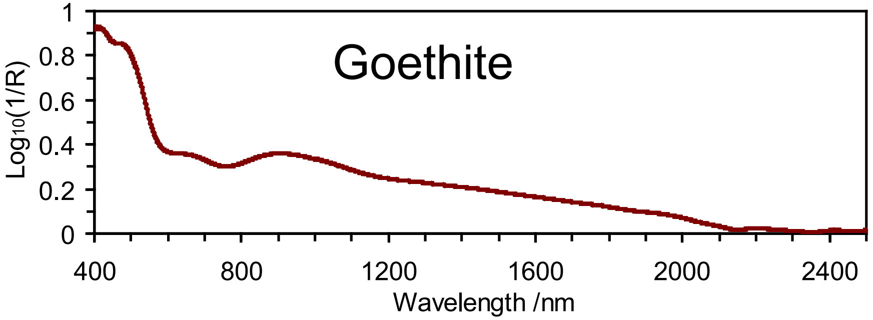


$CaSO_4\cdot 2H_2O$



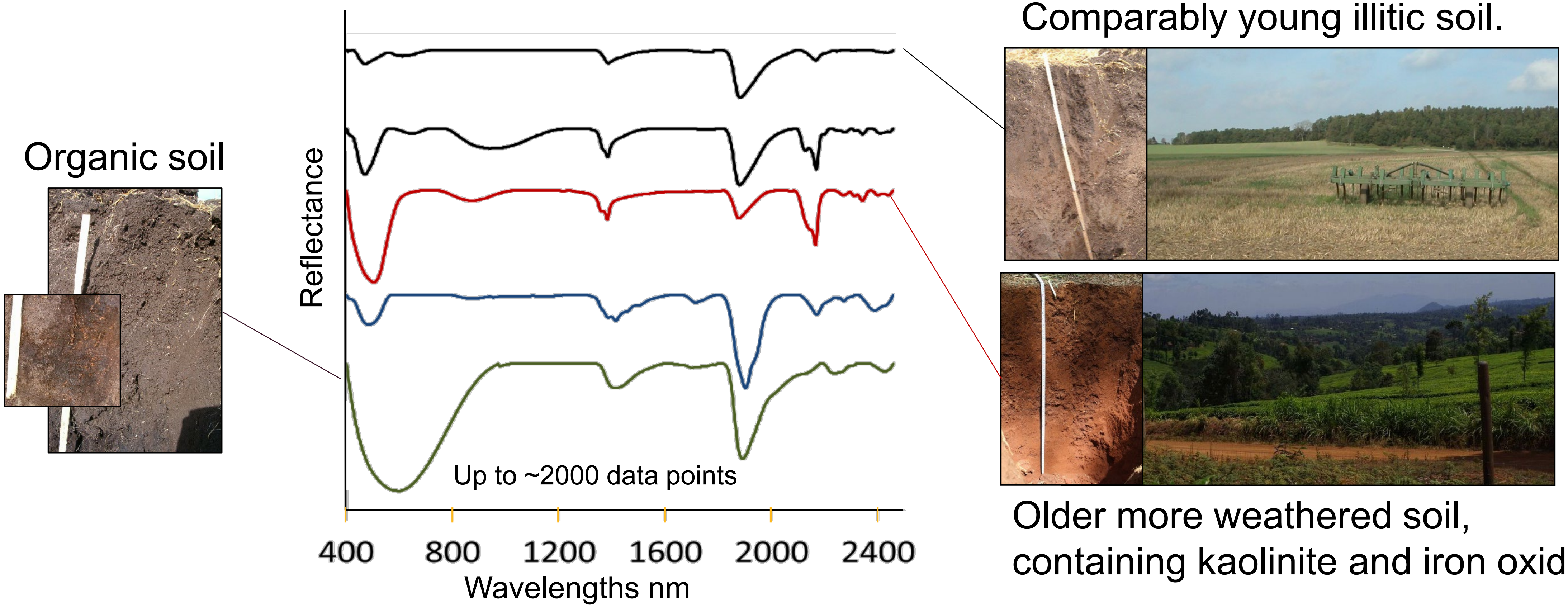






B

Natural soils – a mixture of everything

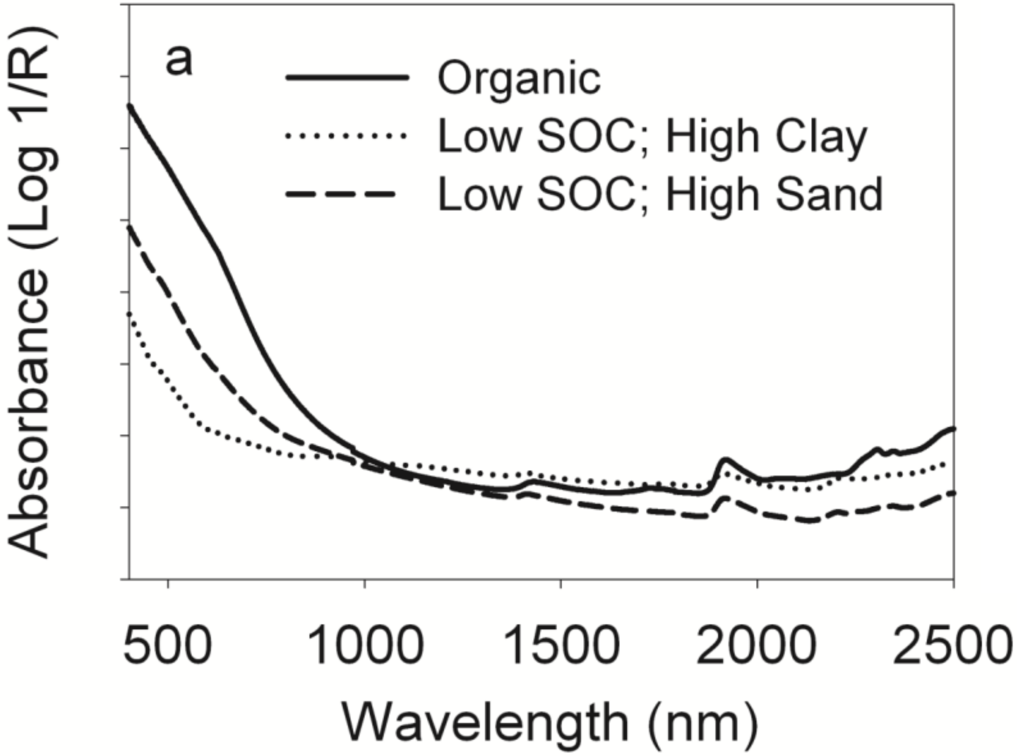


Organic soil

Comparably young illitic soil.

Older more weathered soil, containing kaolinite and iron oxides.

Effect of organic matter on soil spectra

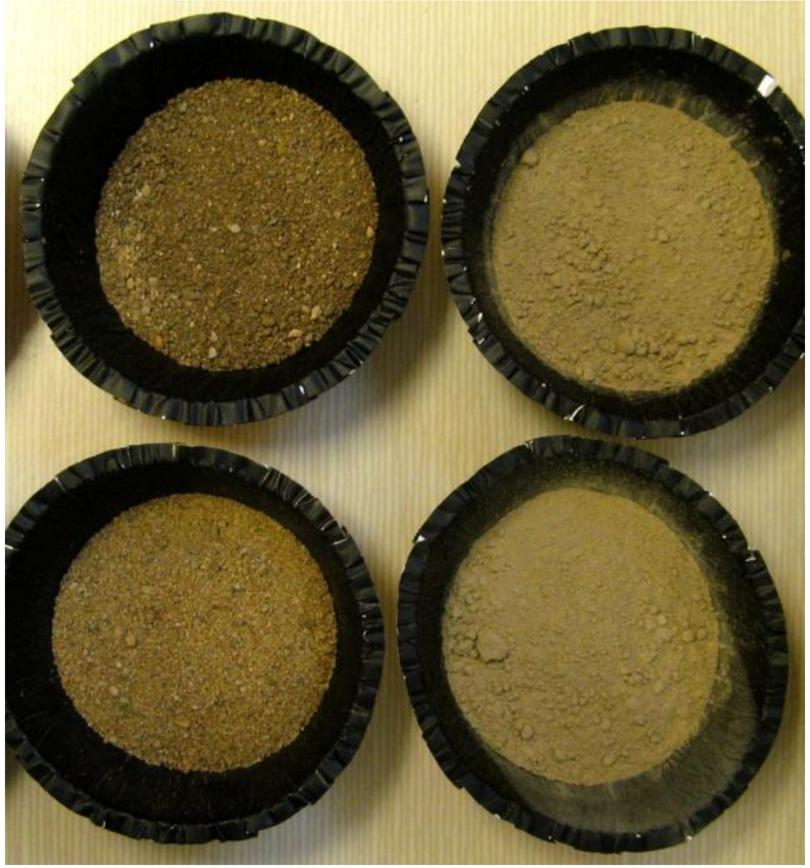


Interaction between texture and SOM

100% Sand

0% Sand (~40% clay)

~0.5% SOC



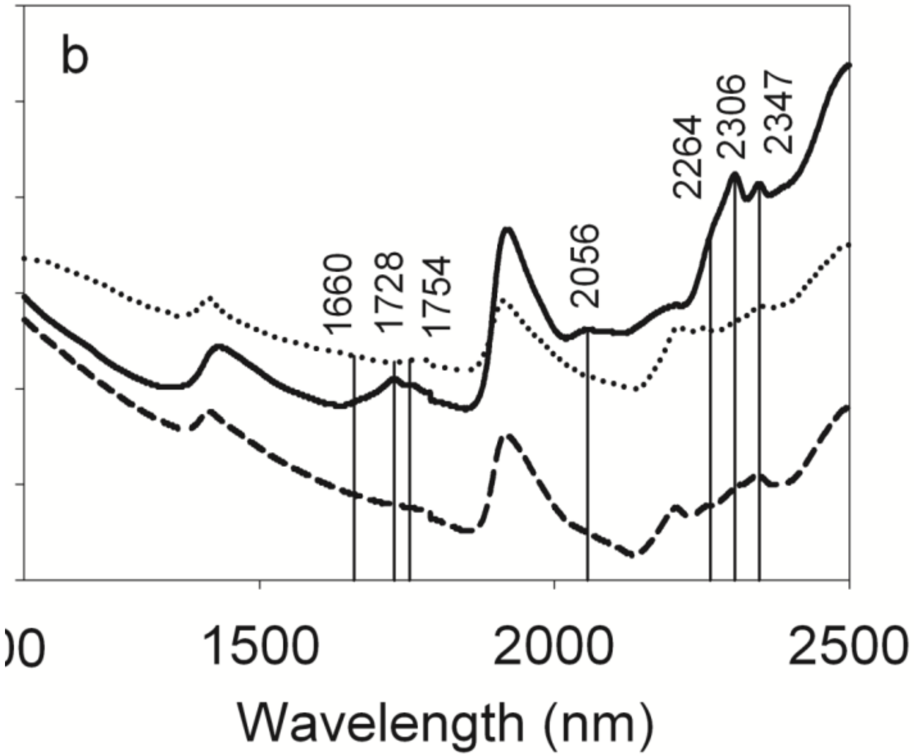
~2% SOC

0% SOC

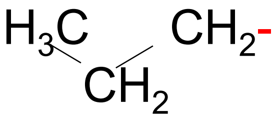
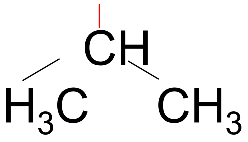
0% SOC

Common absorption bands of organic matter in soil

Alkyl fundamental at 3413 and 3509 nm in Mid-IR

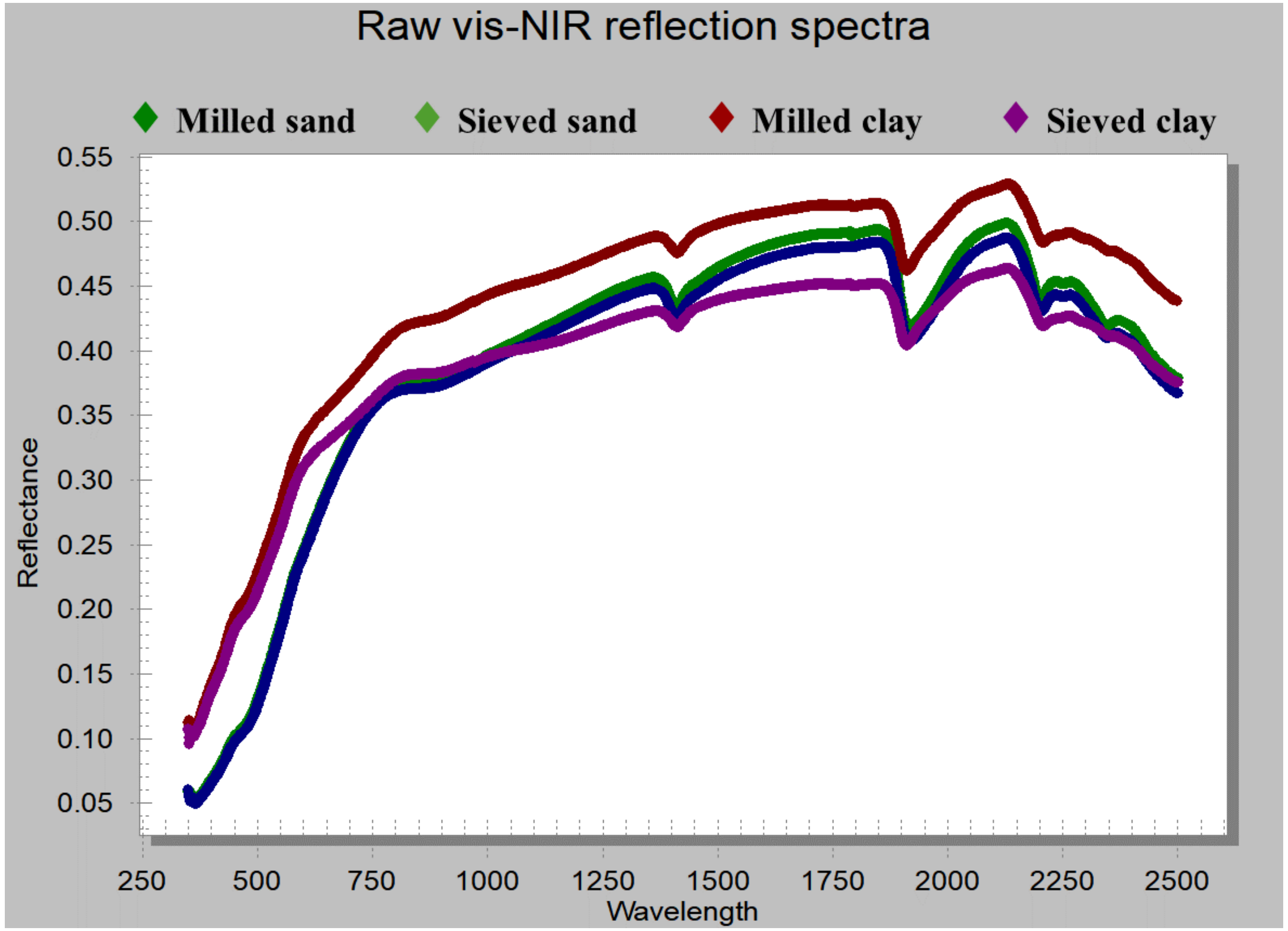


Example of alkyl groups

CH_3-	metyl group
CH_3-CH_2-	ethyl group
	propyl group
	isopropyl group

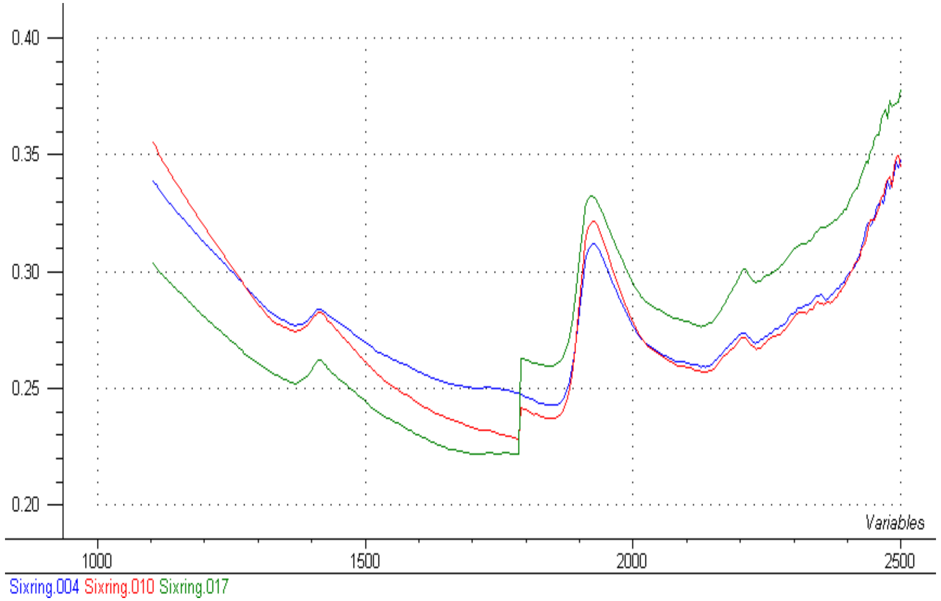
Effect of structure

- Sieving and grinding

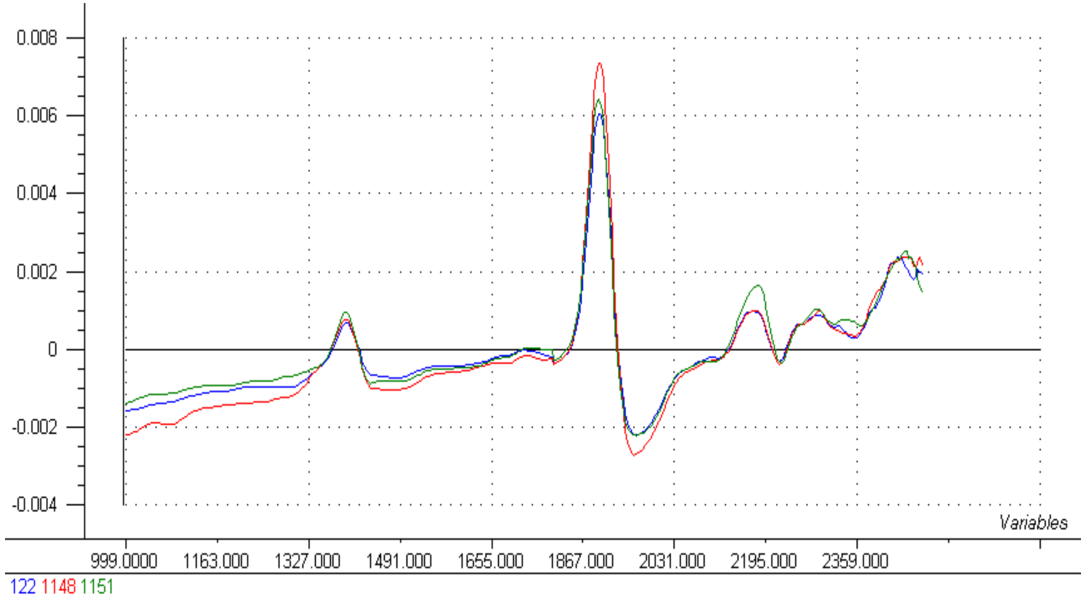


Scatter correction through transformation

Log 1/R



1st derivative



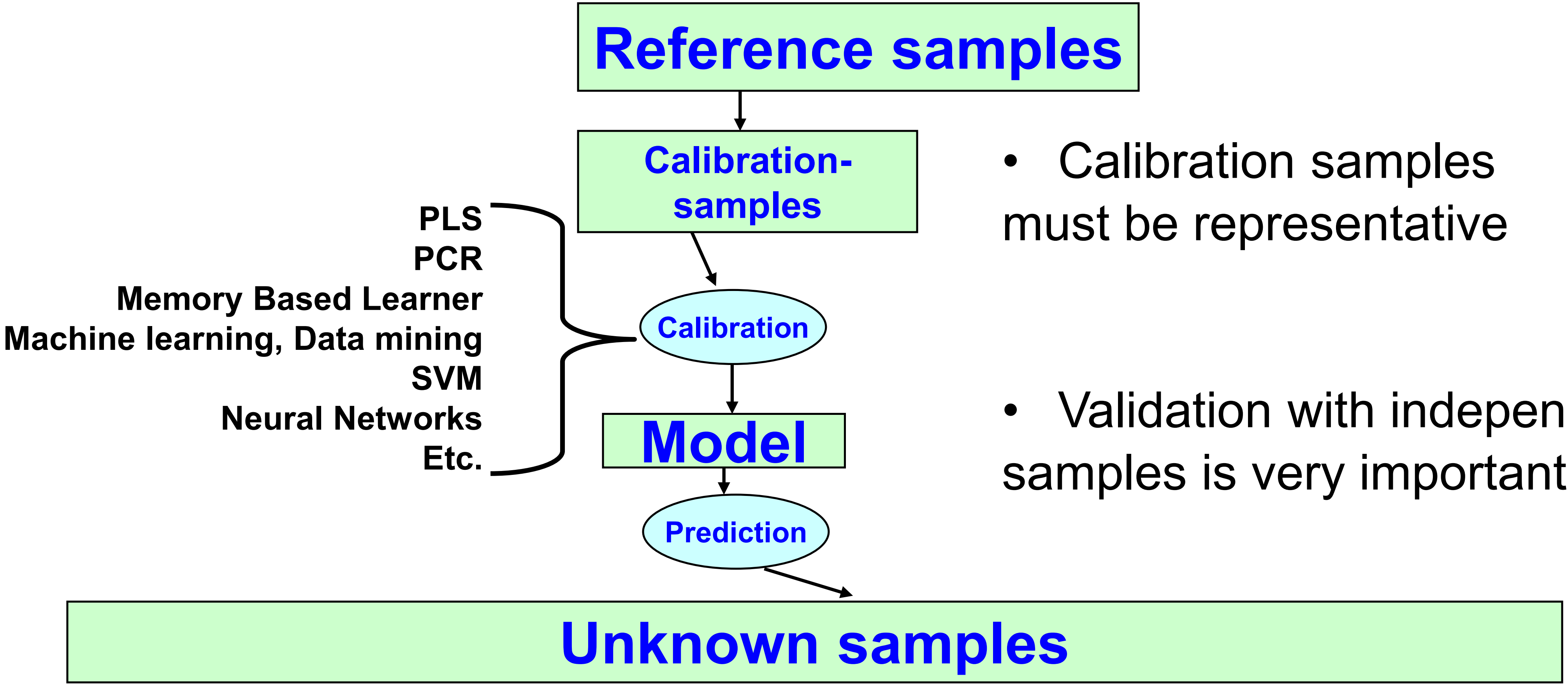
Plus and minus with diffuse reflectance

+ Rapid analysis in real-time
+ Minimum sample preparation
+ Non-destructive, Non-invasive
+ Flexible
+ Several parameters can be analyzed simultaneously
+ Rugged instruments for field analyses can be built

- Results can typically not be red directly, reference required
- Calibration is empirical

Principle for NIR predictions

- Based on empirical linear or non linear (data mining, machine learning etc.) calibrations



Reference samples

Calibration-samples

PLS
PCR
Memory Based Learner
Machine learning, Data mining
SVM
Neural Networks
Etc.

Calibration

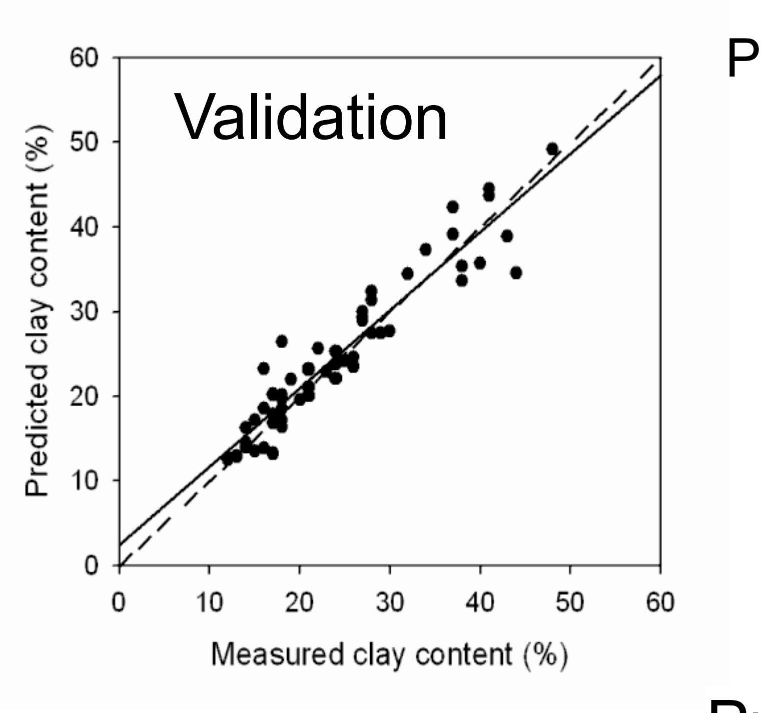
Model

Prediction

Unknown samples

- Calibration samples must be representative
- Validation with independent samples is very important

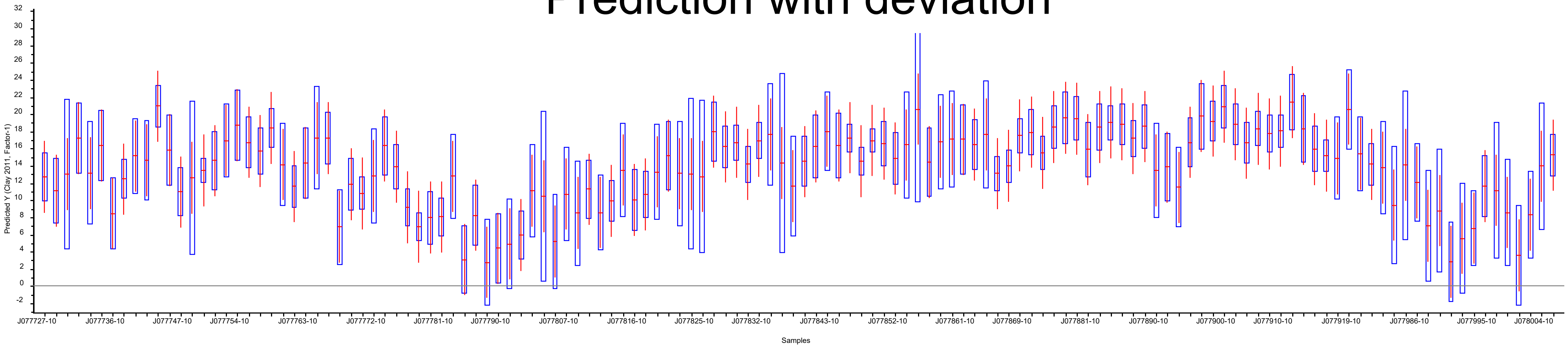
B



Prediction model relating :

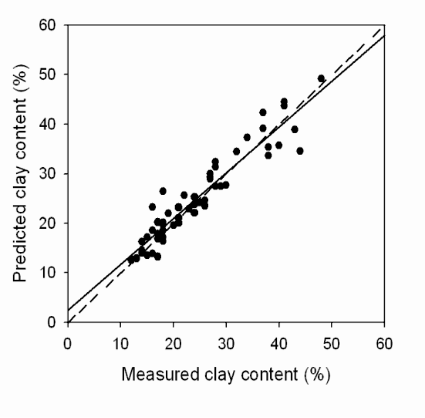
Y-space to X-space
Dependent to independent
Clay to spectra

Prediction with deviation



Statistical estimates of model performance

- Validation statistics



- R^2 = Explained variance by model

- RMSE = Root Mean Squared Error

$$RMSEP = \sqrt{\left[\left(\sum_{i=1}^{n} d_i^2\right)/n\right]}$$

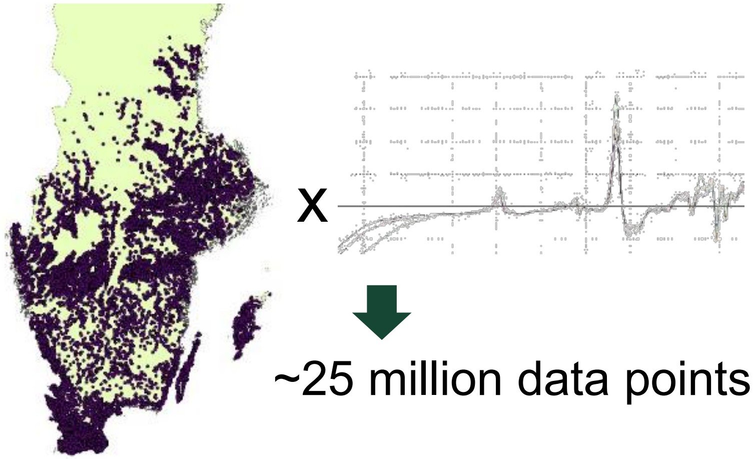
- RPD = Ratio of performance to deviation = $\frac{\text{Stdev}}{\text{RMSE}}$

What is good enough?

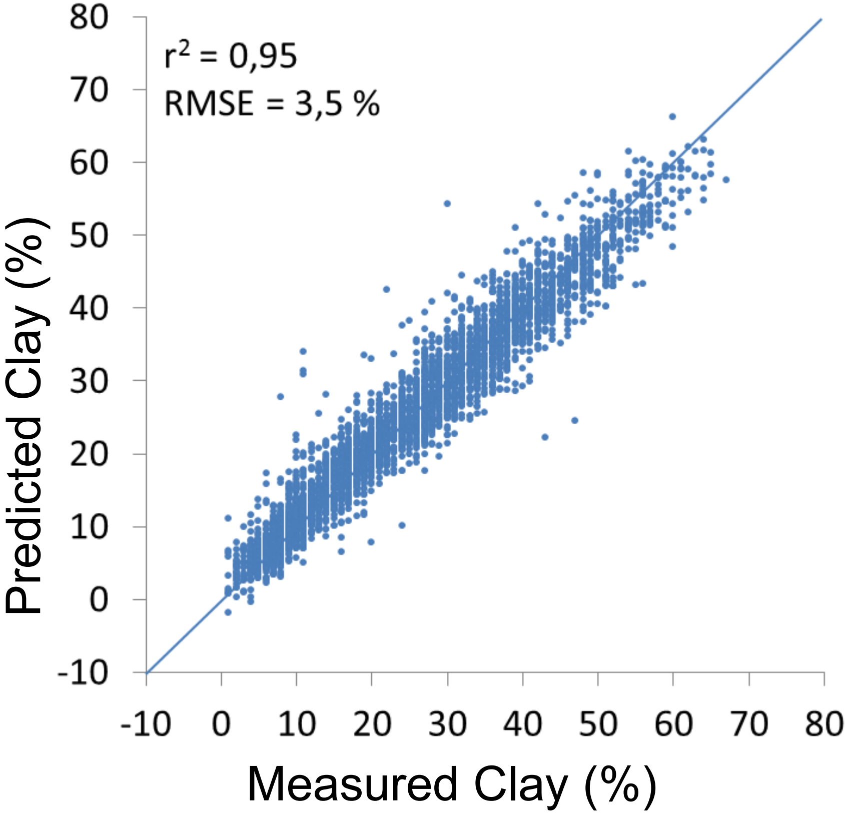
Large soil spectral libraries (SSL)

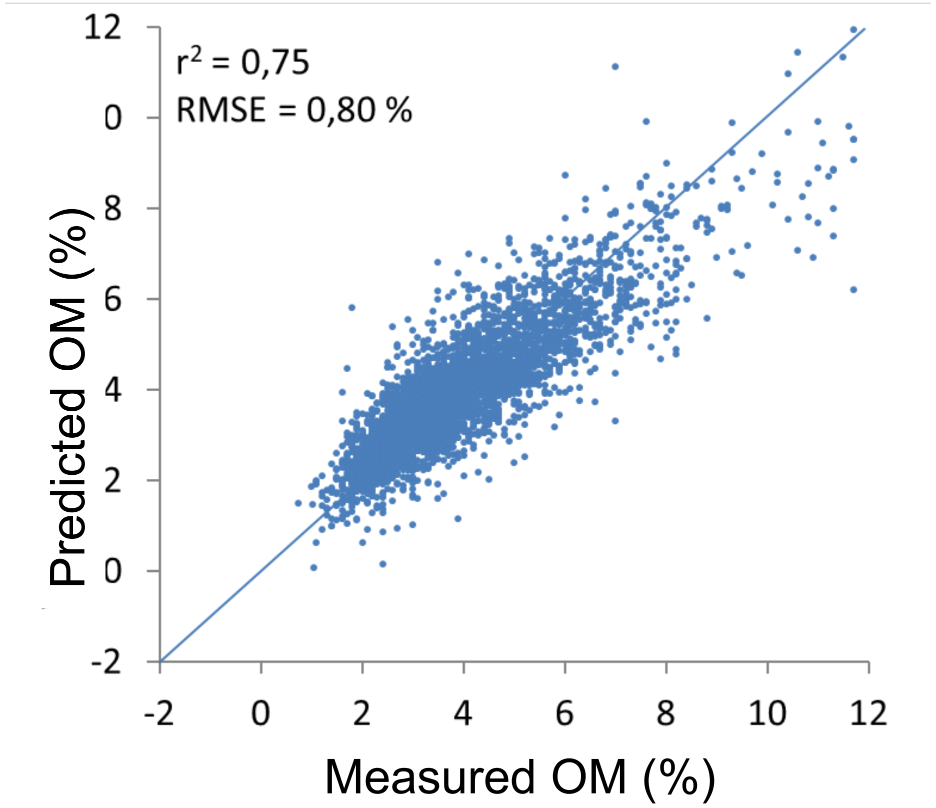
- Much effort to build large SSL’s and gain efficiency through general calibrations at the global/national/regional scale
- Many perform well when validated at their own scale

The Swedish national SSL ~12 000 samples Agricultural top soil


x

~25 million data points





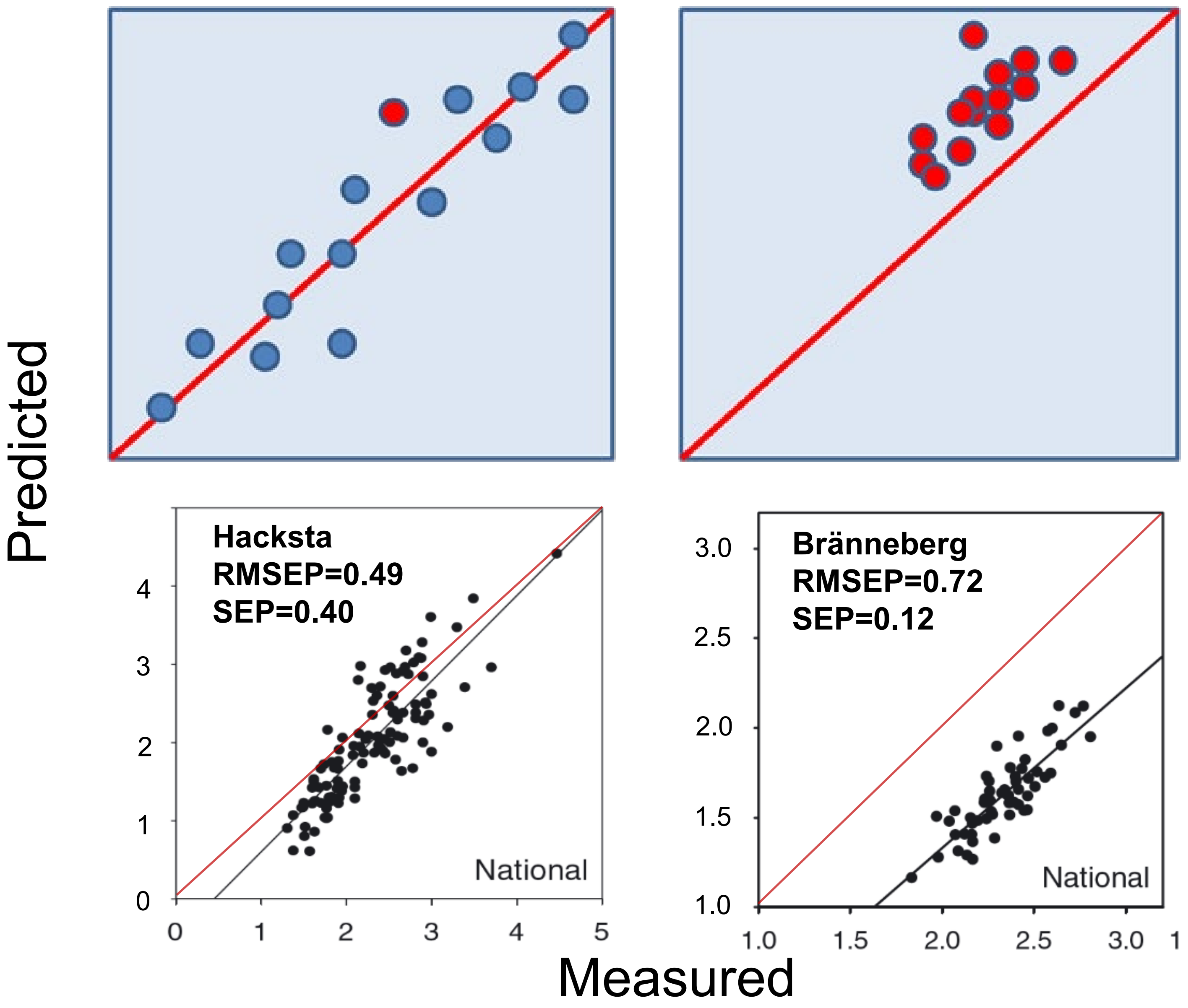
B

Large soil spectral libraries (SSL) for farm or field scald applications?

- Are large scale calibrations suitable to resolve small scale variations?
- Can this be better than local few-sample-calibrations (10-50 samples)

B

Large scale calibration for predictions at the farm scale



Stenberg / Minasney 2013, Pedometron

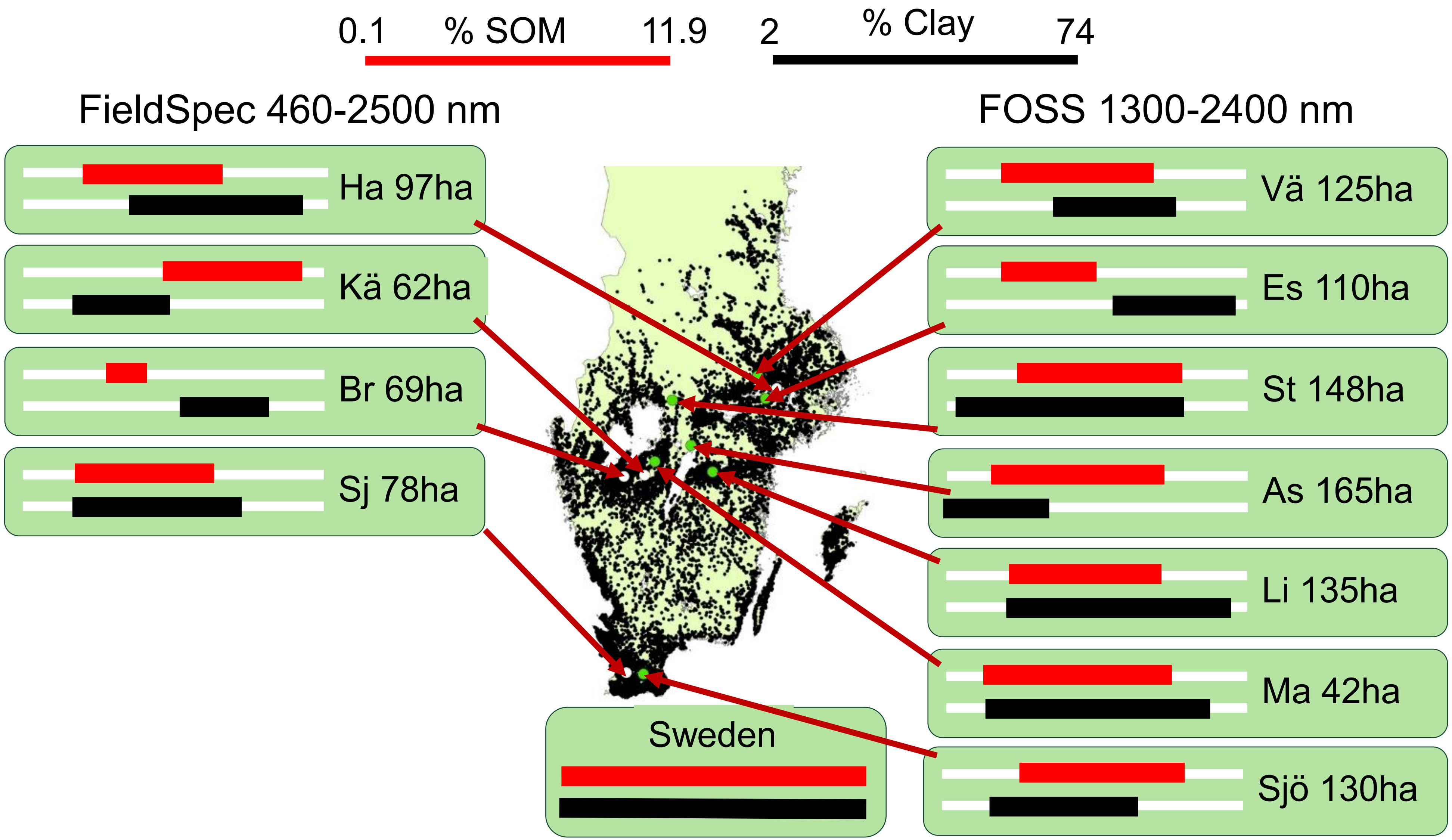
Wetterlind & Stenberg 2010, EJSS 61

B

Validation at the field scale

- We compared the National scale global PLS calibration with:
- The MBL (Memory Based Learner) PLS
- The National SSL spiked with 10-40 local samples
- PLS on 10-40 local samplas only

National and farm data sets



0.1 % SOM 11.9

2 % Clay 74

FieldSpec 460-2500 nm

FOSS 1300-2400 nm

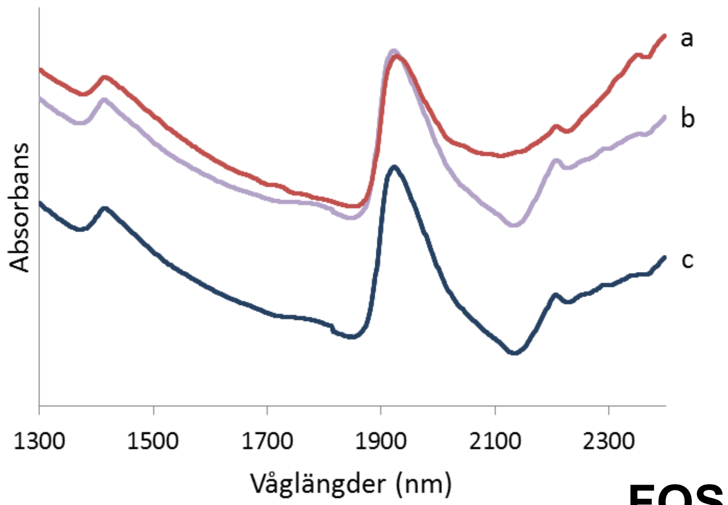
Ha 97ha
Kä 62ha
Br 69ha
Sj 78ha
Vä 125ha
Es 110ha
St 148ha
As 165ha
Li 135ha
Ma 42ha
Sjö 130ha
Sweden

J

Instrument transfer

Piecewise Direct Standardization

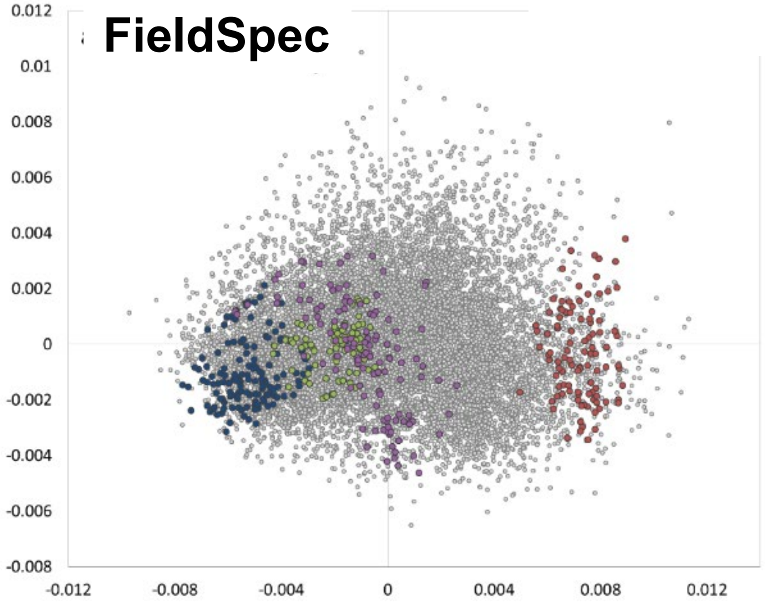
18 standard samples

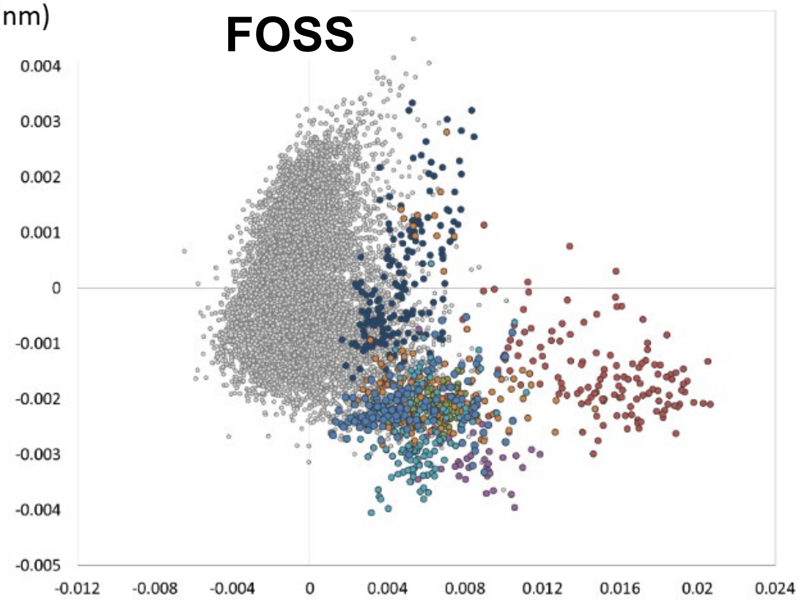


Original FOSS

Transformed FOSS

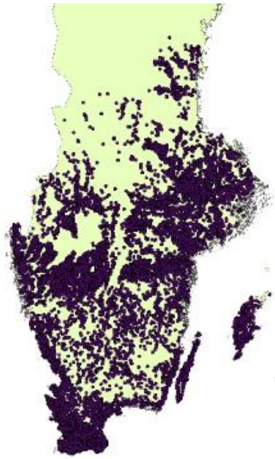
FieldSpec


FieldSpec

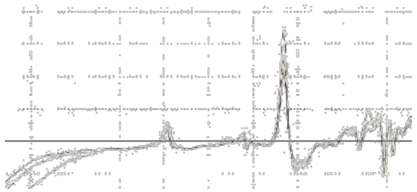

FOSS

J

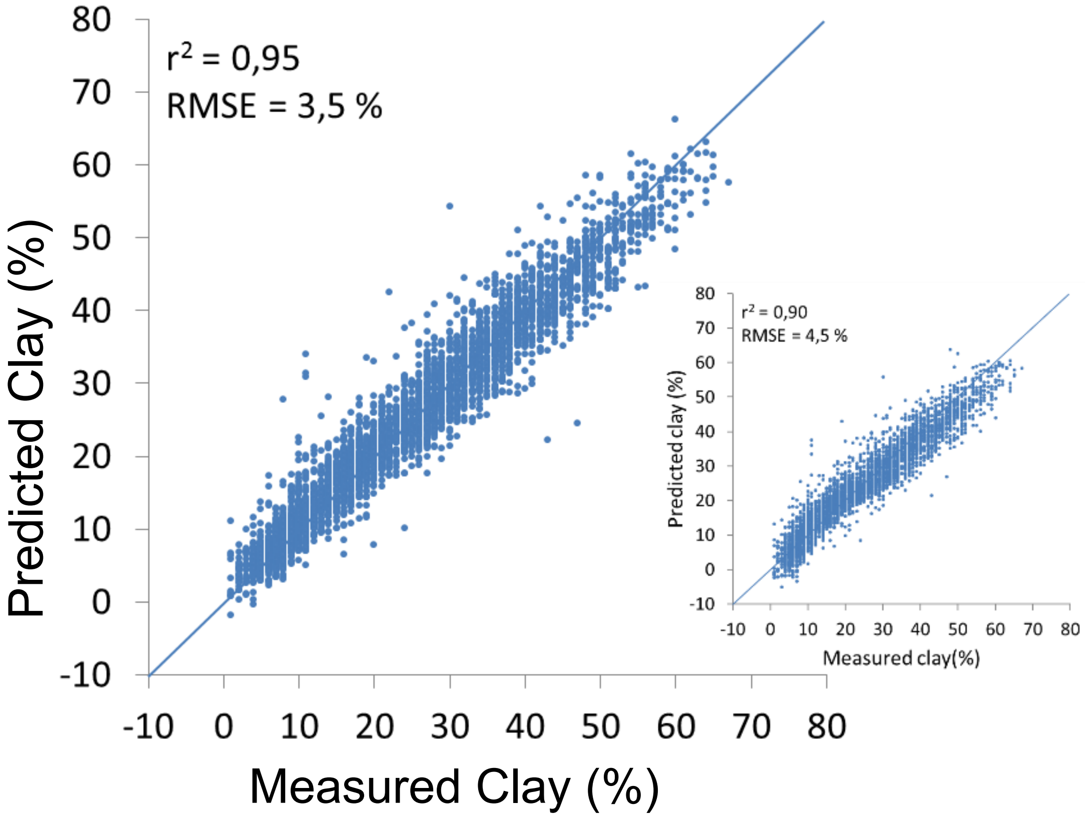
The Swedish national SSL ~12 000 samples Agricultural top soil

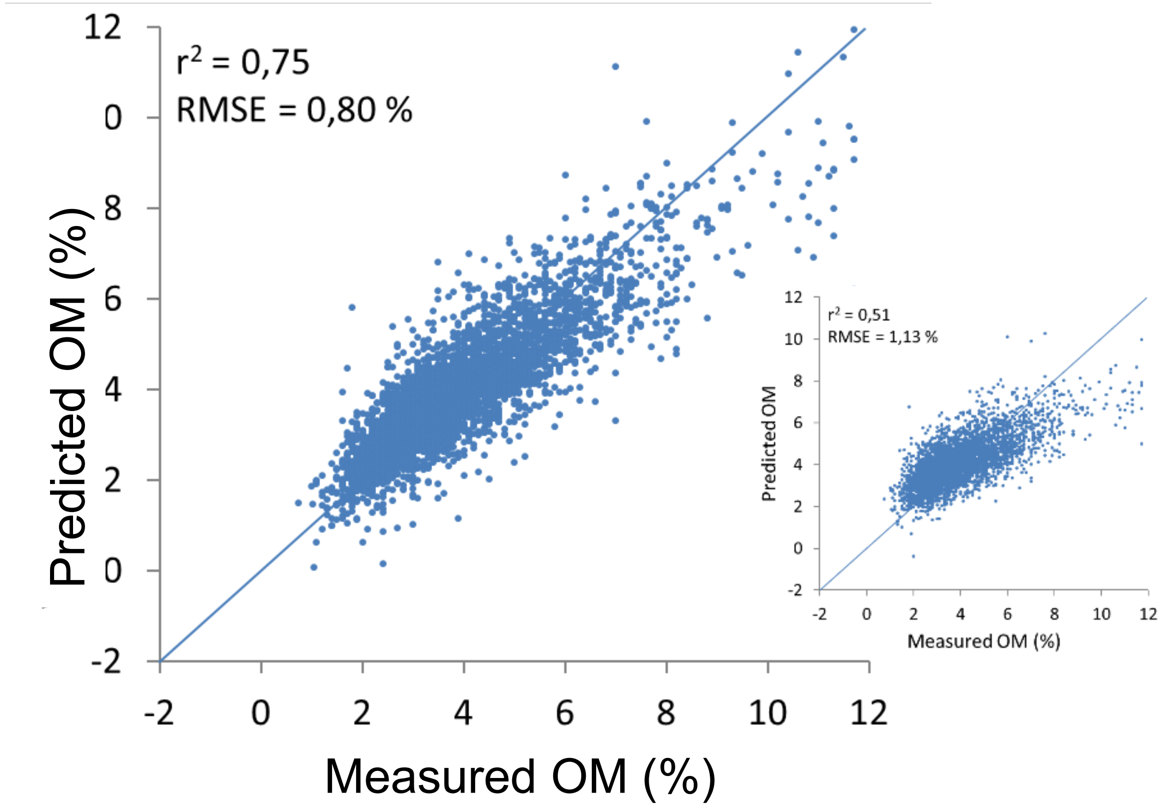


x



~25 million data points



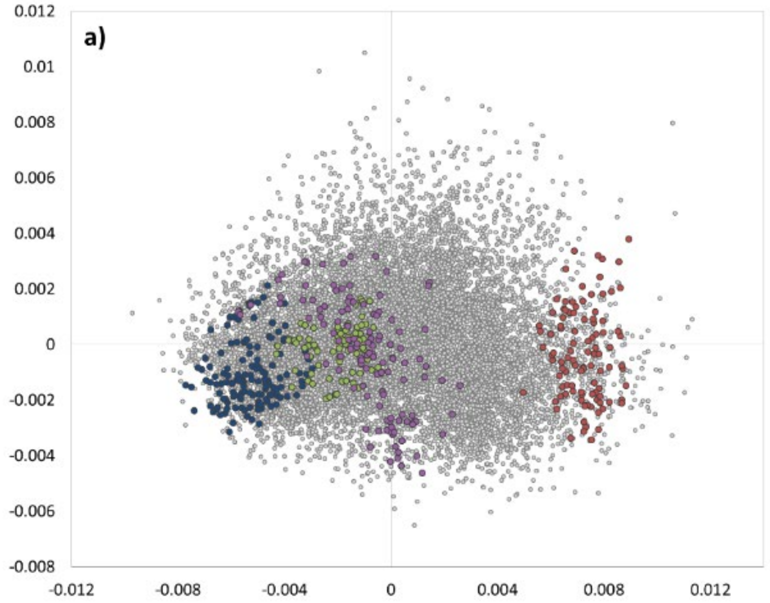


J

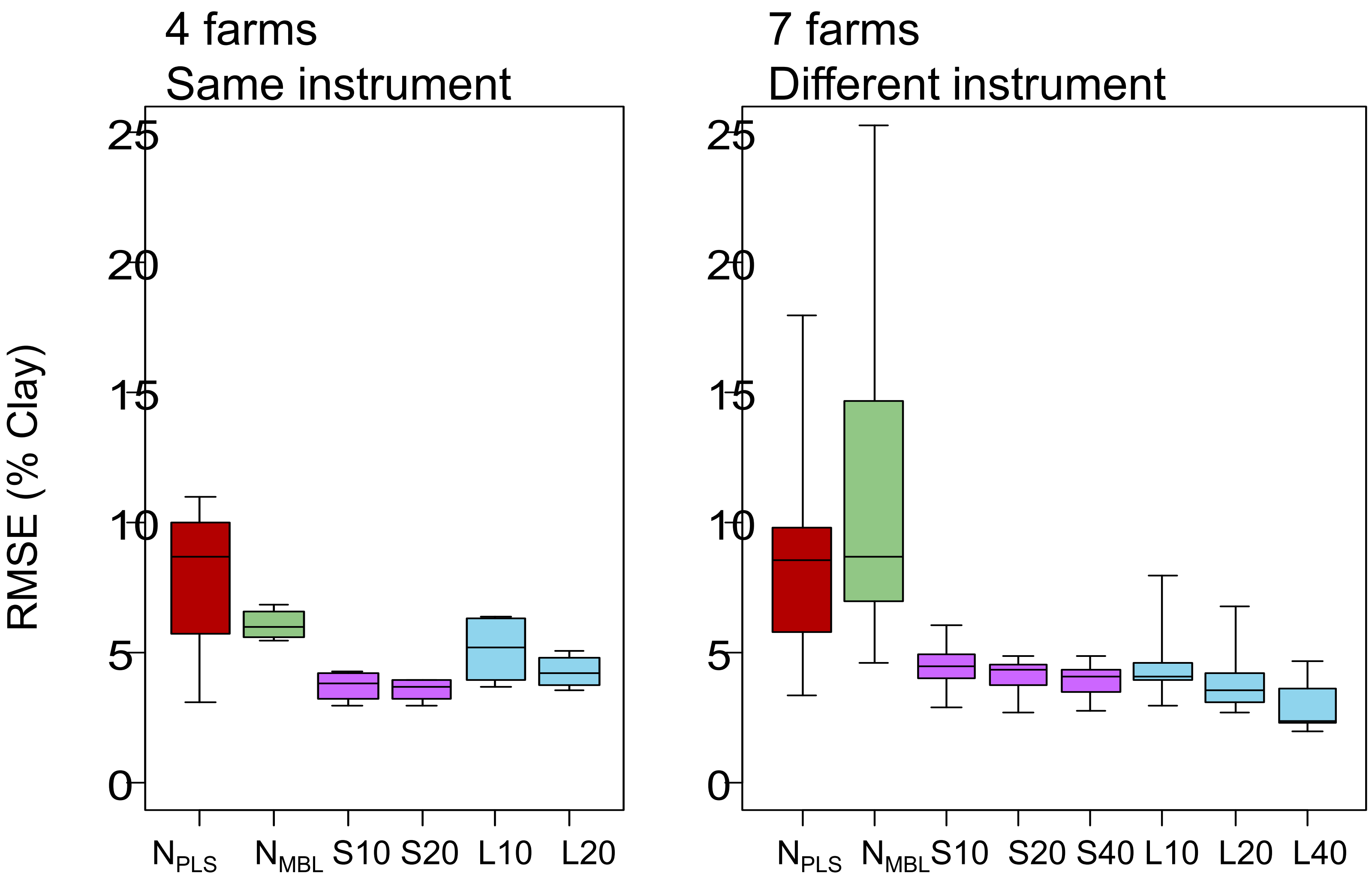
Memory Based Learner PLS in soil spectroscopy

- For each sample to predict, a number of similar samples in the SSL are chosen for the calibration.
- Principal components distance using the spectral features

- *Ramirez-Lopez et al. 2013. The spectrum-based learner: A new local approach for modeling soil vis–NIR spectra of complex datasets. Geoderma, 195–196, p. 268-279*

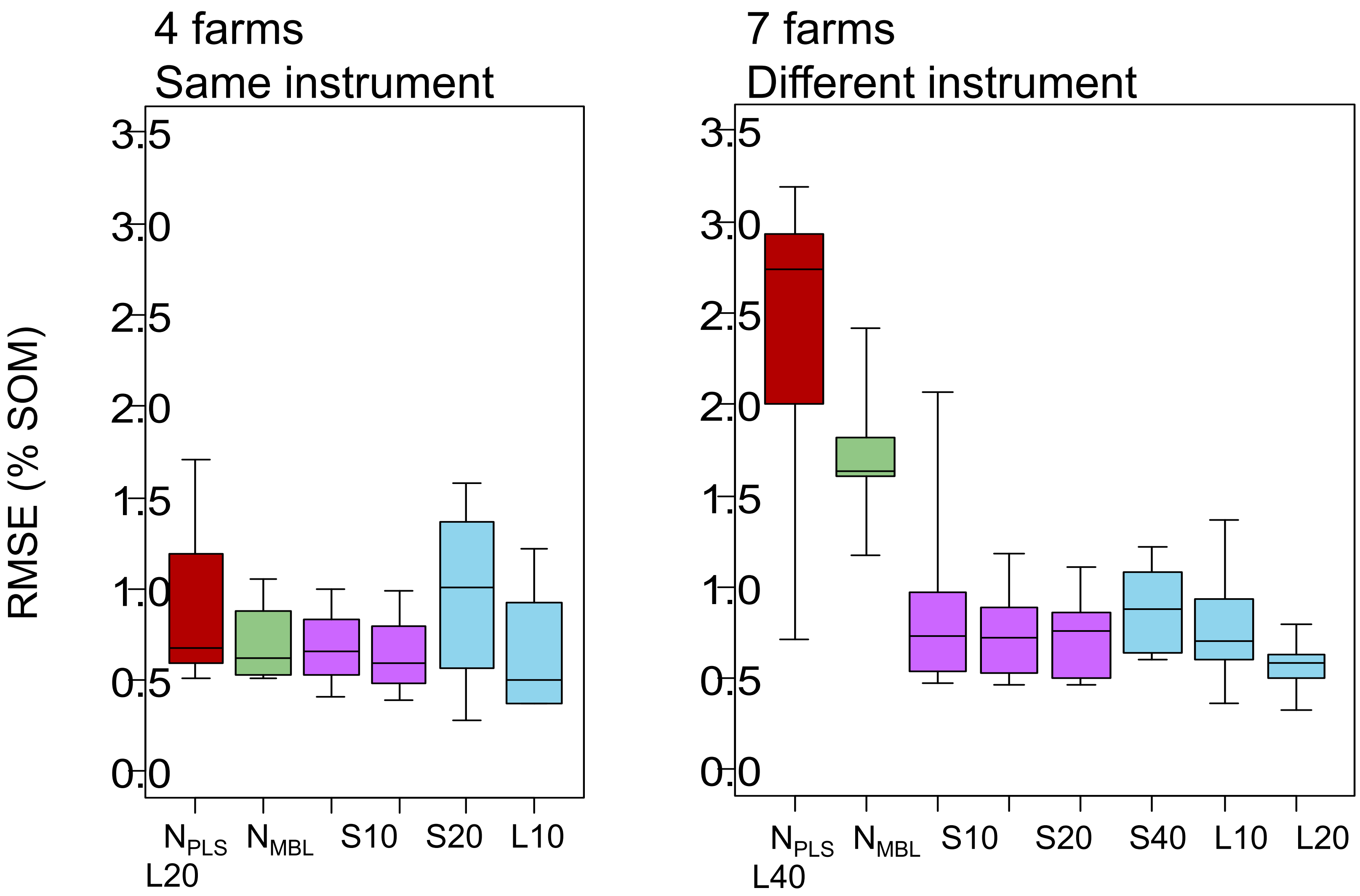


Distribution of Farm RMSEP - Clay Content



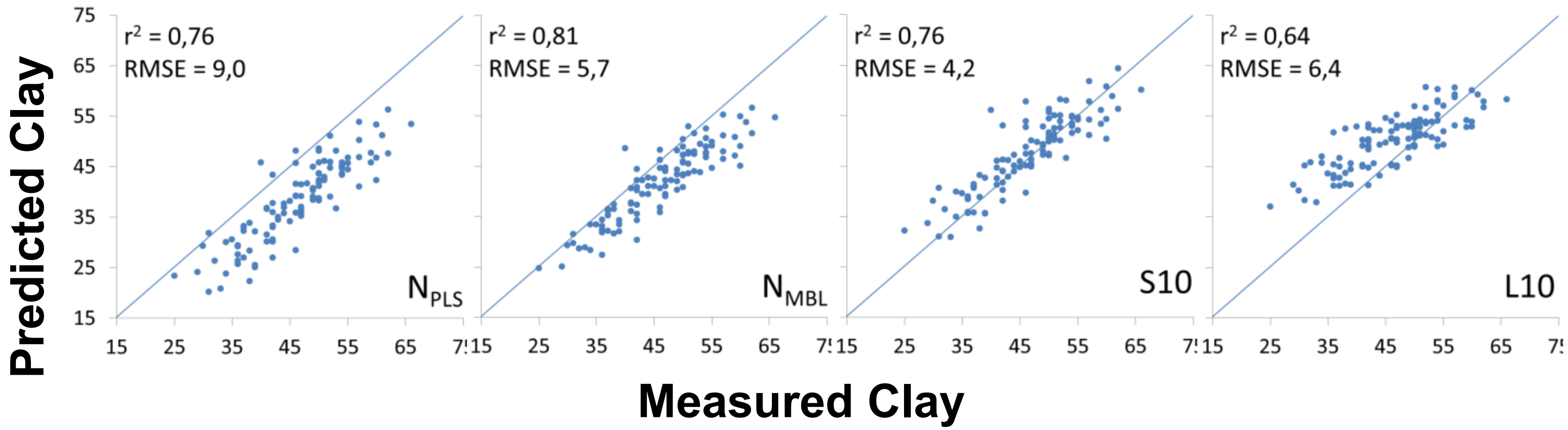
J

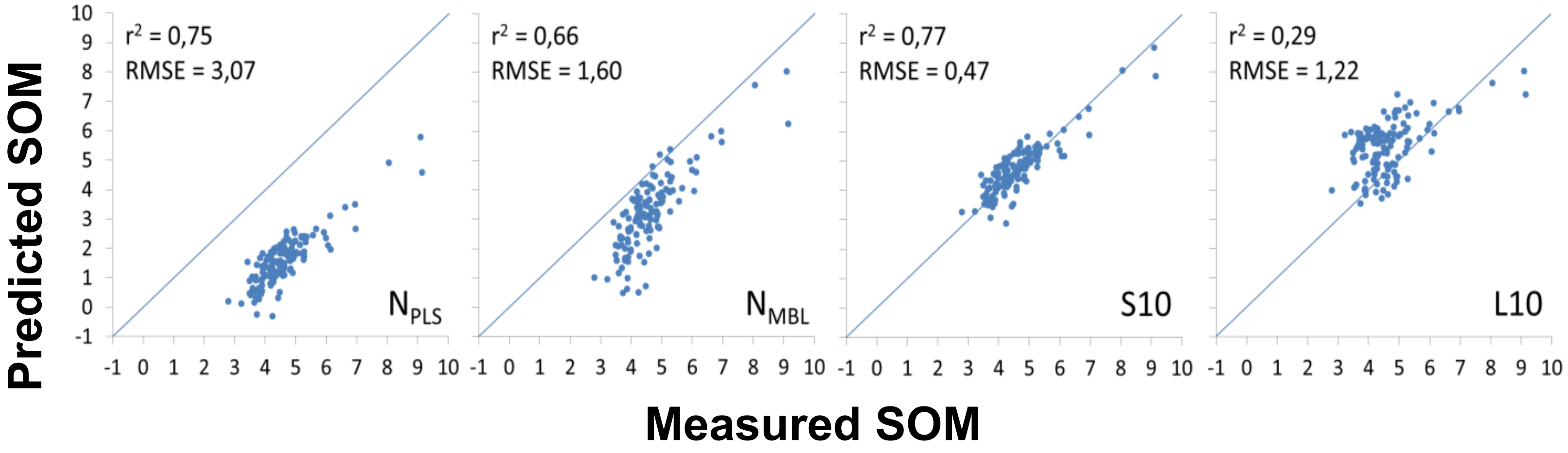
Distribution of Farm RMSEP – SOM



J

Exampel of local predictions





Conclusions

- Large systematic errors with any national model
- Instrument transfer enhanced these systematic errors
- Spiking with N=10 reduced the systematic error to correspond to a local farm calibration with N=40
- Spiking reduced the instrument transfer effect substantially
- Wet chemistry on 10 local samples is affordable on most farms
- **Combining an SSL with few local samples combine the robustness and precision of many samples with the accuracy of local samples**