An introduction to soil spectroscopy

Bo Stenberg

Soil and Environment
Swedish university of Agricultural Sciences
Soil Visible and near infrared (vis-NIR) Spectroscopy

Reflectance

Wavelength (nm)
Vis-NIR spectroscopy

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.
Example for liquid H$_2$O

Symmetric stretching $\nu_1$

Bending $\nu_2$

Asymmetric stretching $\nu_3$

<table>
<thead>
<tr>
<th>Absorption</th>
<th>Frequency</th>
<th>Wavelength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundamentals</td>
<td>$\nu_1$, $\nu_2$, $\nu_3$</td>
<td>$\lambda_1$, $\lambda_2$, $\lambda_3$</td>
</tr>
<tr>
<td>1$^{\text{st}}$ Overtone</td>
<td>$2\nu_1$, $2\nu_2$, $2\nu_3$</td>
<td>$\frac{1}{2} \lambda_1$, $\frac{1}{2} \lambda_2$, $\frac{1}{2} \lambda_3$</td>
</tr>
<tr>
<td>Combinations</td>
<td>$\nu_1 + \nu_2$ or $\nu_2 + \nu_3$</td>
<td>$1/(1/\lambda_1 + 1/\lambda_2)$ or $1/(1/\lambda_2 + 1/\lambda_3)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fundamental</th>
<th>1$^{\text{st}}$ Overtone</th>
<th>Combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_1$</td>
<td>(\ nm) 3050</td>
<td>(\ nm) 1525</td>
</tr>
<tr>
<td>$\nu_2$</td>
<td>(\ nm) 6080</td>
<td>(\ nm) 3040</td>
</tr>
<tr>
<td>$\nu_3$</td>
<td>(\ nm) 2870</td>
<td>(\ nm) 1435</td>
</tr>
</tbody>
</table>
What kind of information can be found in the vis-NIR?

• Strong absorption by water (1400 & 1900 nm)
• Absorption 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

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

William Herschel discovered NIR radiation in 1800
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)

\[ H_2O \text{ bend} \& \text{OH stretch} \]

combination of fundamentals at 6080nm & 2870nm
\((1645 \text{ cm}^{-1} \& 3480 \text{ cm}^{-1})\)

\[ \log_{10}(1/R) \]

\[
\begin{align*}
1st overtone of OH stretch fundamental at 2870nm (3480cm^{-1}) \\
H_2O \text{ bend} \& \text{OH stretch combination of fundamentals at 6080nm \& 2870nm} \\
(1645 \text{ cm}^{-1} \& 3480 \text{ cm}^{-1})
\end{align*}
\]
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

**Kaolinite**

**Montmorillonite**

**Illite**

**High Carbonate**

**High Gypsum**

**Goethite**

CaSO₄•2H₂O
Natural soils – a mixture of everything

Organic soil

Comparably young illitic soil.

Older more weathered soil, containing kaolinite and iron oxides.

Reflectance

Up to ~2000 data points

Wavelengths nm
Effect of organic matter on soil spectra

![Graph showing absorbance (Log 1/R) vs. wavelength (nm) with different soil conditions: Organic, Low SOC; High Clay, and Low SOC; High Sand.](image)
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$_3^-$ methyl group
- CH$_3$-CH$_2^-$ ethyl group
- H$_3$C-CH$_2^-$ propyl group
- CH$_3$-CH(CH$_3$)$_2^-$ isopropyl group
Effect of structure
- Sieving and grinding

Raw vis-NIR reflection spectra

- Milled sand
- Sieved sand
- Milled clay
- Sieved clay

Reflectance vs Wavelength
Scatter correction through transformation
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 read 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

• Calibration samples must be representative

• Validation with independent samples is very important

Unknown samples
Prediction model relating: Y-space to X-space
Dependent to independent
Clay to spectra

Validation

Prediction with deviation
Statistical estimates of model performance

- Validation statistics

• $R^2 = \text{Explained variance by model}$

• RMSE = Root Mean Squared Error

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}}$$

• 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

$\begin{align*}
\text{Measured Clay} (\%) & \quad r^2 = 0.95 \\
\text{Predicted Clay} (\%) & \quad \text{RMSE} = 3.5 \% \\
\text{Measured OM} (\%) & \quad r^2 = 0.75 \\
\text{Predicted OM} (\%) & \quad \text{RMSE} = 0.80 \%
\end{align*}$

~25 million data points
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)
Large scale calibration for predictions at the farm scale

Stenberg / Minasney 2013, Pedometron

Wetterlind & Stenber 2010, EJSS 61
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 samples only
### National and farm data sets

<table>
<thead>
<tr>
<th>FieldSpec 460-2500 nm</th>
<th>FOSS 1300-2400 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ha 97ha</td>
<td>Vä 125ha</td>
</tr>
<tr>
<td>Kä 62ha</td>
<td>Es 110ha</td>
</tr>
<tr>
<td>Br 69ha</td>
<td>St 148ha</td>
</tr>
<tr>
<td>Sj 78ha</td>
<td>As 165ha</td>
</tr>
<tr>
<td>Sweden</td>
<td>Li 135ha</td>
</tr>
<tr>
<td></td>
<td>Ma 42ha</td>
</tr>
<tr>
<td></td>
<td>Sjö 130ha</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% SOM</th>
<th>% Clay</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>2</td>
</tr>
</tbody>
</table>

- Sediment concentrations in Sweden:
  - FieldSpec 460-2500 nm: Ha 97ha, Kä 62ha, Br 69ha, Sj 78ha
  - FOSS 1300-2400 nm: Vä 125ha, Es 110ha, St 148ha, As 165ha, Li 135ha, Ma 42ha, Sjö 130ha

- Sediment properties:
  - FieldSpec 460-2500 nm: % SOM = 0.1
  - FOSS 1300-2400 nm: % Clay = 2
Instrument transfer

- Piecewise Direct Standardization
- 18 standard samples

Original FOSS
Transformed FOSS
FieldSpec

Graphs showing absorbance and wavelength comparison between FOSS and FieldSpec.
The Swedish national SSL ~12 000 samples
Agricultural top soil

\[ r^2 = 0.95 \]
\[ \text{RMSE} = 3.5 \% \]

\[ r^2 = 0.75 \]
\[ \text{RMSE} = 0.80 \% \]

\[ r^2 = 0.51 \]
\[ \text{RMSE} = 1.13 \% \]
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

Distribution of Farm RMSEP - Clay Content

4 farms
Same instrument

7 farms
Different instrument

RMSE (% Clay)

N_{PLS}  N_{MBL}  S10  S20  L10  L20

N_{PLS}  N_{MBL}  S10  S20  S40  L10  L20  L40
Distribution of Farm RMSEP – SOM

4 farms
Same instrument

7 farms
Different instrument

RMSE (% SOM)
Example of local predictions

- Predicted Clay vs Measured Clay
  - $r^2 = 0.76$
  - RMSE = 9.0
  - $r^2 = 0.81$
  - RMSE = 5.7
  - $r^2 = 0.76$
  - RMSE = 4.2
  - $r^2 = 0.64$
  - RMSE = 6.4

- Predicted SOM vs Measured SOM
  - $r^2 = 0.75$
  - RMSE = 3.07
  - $r^2 = 0.66$
  - RMSE = 1.60
  - $r^2 = 0.77$
  - RMSE = 0.47
  - $r^2 = 0.29$
  - RMSE = 1.22
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