RaCA SOC estimates from VNIR

Terry Loecke, PhD, U of Nebraska-Lincoln
Skye Wills, PhD, NSSC-Lincoln
Yufeng Ge, PhD, U of Nebraska-Lincoln
Goals

• Introduce / reintroduce RaCA
• Error propagation in RaCA
• Importance of measurement precision for other uses
• VNIR model performance and weaknesses
• Current modeling efforts
Hierarchical Sampling Design

• 17 Regions (MO)
• 8-20 soil groups w/in Regions
• 7 Landuses w/in groups w/in Regions
• 6553 sites
• 1 central pedon and 4 satellite pedons
• Sampled by horizon ~ 145,000 samples
• All samples scanned with VNIR
• Central pedons analyzed for TC, SOC, SIC
Spatial Distribution of 6418 sites and SOC Stocks 1 m
SOC stocks to 1m using RF-VNIR estimates
Random Forest Model Performance with independent samples

Table 7. Cross-validation and validation results of eight visible and near-infrared-based models developed to predict soil organic C (SOC). Calibration and validation sets had 6033 and 2580 samples, respectively, with observed SOC ranging from 0.0 to 63.5 and 0.0 to 63.3%, respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>Cross-validation</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$†</td>
<td>RMSPE‡ (%)</td>
</tr>
<tr>
<td>RF/First/11 †</td>
<td>0.93</td>
<td>2.500</td>
</tr>
<tr>
<td>RF/First/11/CS</td>
<td>0.93</td>
<td>2.464</td>
</tr>
<tr>
<td>RF/First/21</td>
<td>0.93</td>
<td>2.460</td>
</tr>
<tr>
<td>RF/First/21/CS</td>
<td>0.93</td>
<td>2.432</td>
</tr>
<tr>
<td>RF/Second/11</td>
<td>0.92</td>
<td>2.595</td>
</tr>
<tr>
<td>RF/Second/11/CS</td>
<td>0.92</td>
<td>2.585</td>
</tr>
<tr>
<td>RF/Second/21</td>
<td>0.93</td>
<td>2.403</td>
</tr>
<tr>
<td>RF/Second/21/CS</td>
<td>0.93</td>
<td>2.386</td>
</tr>
</tbody>
</table>

† $R^2$, coefficient of multiple determination.
‡ RMSPE, root mean squared prediction error.
§ RPD, residual prediction deviation.
†† RF, random forest; First, first derivative; Second, second derivative; 11, smoothing segment of 11 nm; 21, smoothing segment of 21 nm; CS, clipped spectra.

Random Forest Model Performance in RaCA

mineral non calcareous checked samples

EA measured SOC

RF predicted SOC

mineral non calcareous checked samples
~180 Mg C ha\(^{-1}\)
Random Forest Model Performance by Region

mineral non calcareous checked samples

166 58 7 51 69 2 75 10 63 193 27

%SOC model vs checks

MO
Random Forest Model Performance by Texture

mineral non calcareous checked samples

59 76 81 1 6 23 80 166 6 29 64 0 56 6 5
<table>
<thead>
<tr>
<th>Preprocessing</th>
<th>Model size</th>
<th>$R^2$</th>
<th>$RMSE_{CV}$</th>
<th>RPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>30</td>
<td>0.10</td>
<td>18.67</td>
<td>1.05</td>
</tr>
<tr>
<td>Moving average with window of 11 nm</td>
<td>30</td>
<td>0.88</td>
<td>6.83</td>
<td>2.88</td>
</tr>
<tr>
<td>1st derivative</td>
<td>30</td>
<td>0.87</td>
<td>6.97</td>
<td>2.82</td>
</tr>
<tr>
<td>2nd derivative</td>
<td>30</td>
<td>0.76</td>
<td>9.66</td>
<td>2.04</td>
</tr>
<tr>
<td>Savitsky Golay smoothing with 3$^{rd}$ polynomial &amp; 11 nm window</td>
<td>30</td>
<td>0.88</td>
<td>6.81</td>
<td>2.89</td>
</tr>
<tr>
<td>Savitzky Golay 1st derivative with 3$^{rd}$ polynomial &amp; 11 nm window</td>
<td>30</td>
<td>0.88</td>
<td>6.73</td>
<td>2.92</td>
</tr>
<tr>
<td>Savitzky Golay 2nd derivative with 3$^{rd}$ polynomial &amp; 11 nm window</td>
<td>30</td>
<td>0.87</td>
<td>6.97</td>
<td>2.83</td>
</tr>
<tr>
<td>Savitzky Golay 11 nm window &amp; SNV</td>
<td>30</td>
<td>0.90</td>
<td>6.24</td>
<td>3.16</td>
</tr>
<tr>
<td>Savitzky Golay &amp; SNV-detrend</td>
<td>30</td>
<td>0.90</td>
<td>6.29</td>
<td>3.13</td>
</tr>
<tr>
<td>Savitzky Golay &amp; MSC</td>
<td>30</td>
<td>0.76</td>
<td>9.63</td>
<td>2.04</td>
</tr>
<tr>
<td>Savitzky Golay with 21nm window &amp; SNV</td>
<td>30</td>
<td>0.90</td>
<td>6.26</td>
<td>3.14</td>
</tr>
</tbody>
</table>
Supported Vector Machine Model
Performance in RaCA

Nuwan Wijewardane, in prep.
Conclusions

• RaCA – CONUS SOC stocks and uncertainty
• VNIR model prediction precision is important for small scale detection of SOC stocks and concentration
• VNIR model bias in RaCA is reduced by including texture as covariate
• Support Vector Machine approach has promise for a global model.
RaCA partial uncertainty workflow
RaCA partial uncertainty workflow

Measurement/model precision

Balance precision

Method precision

Measurement precision

Multivariate analysis
PLSR, Random Forests, SVM, NN

Calibration
Validation

Reflectance (R) vs. Wavelength (nm)

- Silt loam
- Loam
- Sandy loam
- Loamy sand
Measurement/model precision

- dSOC%
- dBulk Density
- dHorizon depth

- Soil dry weight
- Soil volume

- Balance precision
- Method precision
- Measurement precision

RaCA partial uncertainty workflow
Spatial interpolation

- dSOC per LULC, Soil Group, Region
  - dSOC per Site
    - dSOC per pedon
      - dSOC per horizon
  - dSOC% (Measurement/model precision)
    - dBulk Density (Measurement precision)
      - Soil dry weight (Measurement precision)
        - Balance precision
      - Soil volume
        - Method precision
  - dHorizon depth (Measurement precision)
Spatial interpolation

dSOC per LULC, Soil Group, Region

dSOC per Site

dSOC per pedon

dSOC per horizon

dSOC%

dBulk Density

dHorizon depth

Measurement/ model precision

Soil dry weight

Balance precision

Method precision

Soil volume

Measurement precision

RaCA partial uncertainty workflow

Delta method vs. Hierarchical Bayesian
RaCA partial uncertainty workflow

Delta method vs. Hierarchical Bayesian
Components of a Semivariogram

- Range
- Sill
- Model Semivariogram
- Nugget

**Graph Details:**
- Variance on the y-axis.
- Distance (km) on the x-axis.
- Gamma (h) on the y-axis.