Soil Survey of the Boundary Waters Canoe Area Wilderness (BWCAW)  
A Collaborative Project between the USDA-Forest Service and USDA-Natural Resources Conservation Service  

Joseph Brennan¹ Tom D’Avello² (tom.davello@wv.usda.gov) Danielle Evans¹ Suzann Kienast-Brown³  
Amy Koch⁴ Zamir Libohova² Robert Long⁵ Kevin Megown⁶ Jessica Philippe⁵  
Larissa Schmitt¹ Roger Risley¹ Robert Vaughn⁶  

1. USDA-NRCS Soil Survey Regional Office – 10 St. Paul MN  
2. USDA-NRCS National Soil Survey Center Lincoln NE  
3. USDA-NRCS Soil Survey Regional Office – 4 Bozeman MT  
4. USDA-NRCS Hilo HI  
5. USDA-NRCS St. Johnsbury VT  
6. USFS Salt Lake City UT
Soil Survey of the Boundary Waters Canoe Area Wilderness (BWCAW)
A Collaborative Project between the USDA-Forest Service and USDA-Natural Resources Conservation Service

BACKGROUND

- August 2012 – kick-off DSM project
- 1.1 million-acre Wilderness area established in 1978 within the Superior National Forest in Northeastern Minnesota
- NRCS MLRA Soil Survey Region 10 in St. Paul MN
  - Soil survey staff in Duluth MN
- Managed USFS and well known for recreational activities
  - Fishing, hunting, camping, canoeing, dog sledding, swimming, hiking, skiing
- Very limited access
  - No vehicles
  - Foot and paddle only with few exceptions
  - No cell phone service
- Wildlife!
MATERIALS AND METHODS

Soil = \( f(s, c, o, r, p, a, n) \)

- c – frigid MAAT=34.75F (1.5C) Precip=27” (685mm)
- o – boreal forest
- r – ~relation to depth and drainage
- p – drift over Precambrian BR
- a - ~12000 BP
MATERIALS AND METHODS

- Many covariates created and explored
- Covariate selection
  - Optimum Index Factor
  - Random forests variable importance
- Covariate predictor set
  - Terrain derivatives (5m LiDAR)
    - slope
    - minimum curvature
    - depression cost surface
    - downslope distance gradient
    - surface area factor
    - relative position (@ 2 neighborhoods)
    - canopy height (from 1m LiDAR)
    - landforms (geomorphons)
  - Spectral Derivatives (Landsat 5)
    - Landsat bands 1, 5
    - NDVI
    - brightness, greenness
    - principle component 6
MATERIALS AND METHODS

• Legend
  • Developed a data driven stratification/pre-map via unsupervised and supervised classification prior to field sampling
    • Determined number of classes data (covariates) would support
  • Initial observations and tacit knowledge yielded upwards of 45 soil series/classes
  • Exploratory pre-map process thinned to ~16 classes
  • Further refinement to 11 broad classes after field data collected, inclusive of water
    1. Very shallow dry till
    2. Shallow dry till
    3. Deep dry till
    4. Moderately deep dry till
    5. Deep wet till
    6. Moderately deep and shallow wet till
    7. Dysic organic
    8. Euic organic
    9. Lacustrine
   10. Eskers
   11. Water
MATERIALS AND METHODS

- Training points selected using conditioned Latin hypercube sampling
  - Access extremely limited
    - Potential sample areas created using a 0.2 mile buffer around navigable waters and trails
  - Attached selected covariates
    - Surface area factor, downslope distance gradient, relative position, minimum curvature, Landsat bands 1 & 5, NDVI, geomorphons
  - 214 training points collected in 15 target classes
MATERIALS AND METHODS

• Classification methods
  • Rule-based (ArcSIE) – knowledge-based inference to develop 9 classes
  • Unsupervised (ERDAS Imagine) – ISODATA clustering for dysic/euic organic soils
  • Supervised (ERDAS Imagine) – applying training data in two runs; 11 and 9 classes
  • Heads-up (ArcGIS) – automated identification of eskers was explored and abandoned in favor of heads-up digitization
  • Pre-defined class (ArcGIS) – water, as defined by break lines from LiDAR data used as “water”
MATERIALS AND METHODS

• Classification methods continued
  • Classification tree (R) – standard classification tree
  • Logistic regression (R) – primarily for modeling extent of lacustrine
    • Modeled class using a binary approach
    • Used probability to assign class to final map
  • Random forests (R)
    • Separability analysis & class collapsing
    • Modeled all 11 classes simultaneously
    • Modeled classes using binary approach
      • Used probability surfaces to assign classes to final map
RESULTS AND DISCUSSION

- **Classification methods** – Local experts qualitatively evaluated results from all classification methods
  - Each method independently
  - Each class, independent of method

- **Logistic regression**
  - Used spectral and terrain covariates
  - Lacustrine class result favored by local experts
  - Nagelgerke's "pseudo R squared“ = 0.53
  - Refined using rules defined by local experts according to Land Association Strata

---

1. 411m and below
2a. 442m and below
2b. 460m and below
2c. 469m and below
3. none, except incidentally at 442m and below

P >= 0.5
RESULTS AND DISCUSSION

• Classification methods
  • Rule-based
    • Used selected terrain covariates
    • ArcSIE extension used to create rules and a fuzzy membership layer for each class
    • Very shallow dry till and shallow dry till class results favored by local experts
RESULTS AND DISCUSSION

- Classification methods
  - Random Forests
    - Used spectral and terrain covariates
    - Separability analysis & class collapsing
      - 11 → 8 → 6 → 5 classes
      - Out-of-bag error (OOB) – 58-75%
    - Modeled all 11 classes simultaneously
      - OOB – 72-75%
    - Modeled each class separately (binary)
      - OOB – 4-22%
    - Predicted probability surface
    - Extracted class with maximum probability for each pixel for final map
  - Deep dry till, deep wet till, moderately deep/shallow wet till, and moderately deep wet till class results favored by local experts
RESULTS AND DISCUSSION

• Hybrid assemblage
  • Local experts qualitatively reviewed and selected the best representation of each class resulting in a hybrid raster map product

<table>
<thead>
<tr>
<th>Soil Type</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very shallow dry till</td>
<td>Rule base (RB)</td>
</tr>
<tr>
<td>Shallow dry till</td>
<td>RB</td>
</tr>
<tr>
<td>Deep dry till</td>
<td>Random Forest (RF)</td>
</tr>
<tr>
<td>Moderately deep dry till</td>
<td>RF</td>
</tr>
<tr>
<td>Deep wet till</td>
<td>RF</td>
</tr>
<tr>
<td>Moderately deep and shallow wet till</td>
<td>RF</td>
</tr>
<tr>
<td>Dysic organic</td>
<td>Unsupervised &amp; NWI</td>
</tr>
<tr>
<td>Euic organic</td>
<td>Unsupervised &amp; NWI</td>
</tr>
<tr>
<td>Lacustrine</td>
<td>Logistic</td>
</tr>
<tr>
<td>Eskers</td>
<td>Heads-up</td>
</tr>
<tr>
<td>Water</td>
<td>LiDAR break lines</td>
</tr>
</tbody>
</table>
RESULTS AND DISCUSSION

• **Accuracy Assessment (AA)** Tentative preliminary results for illustration only
  • 109 validation points collected

  11 classes -> 64% overall, KHAT = 0.6
  9 classes  -> 71% overall, KHAT = 0.67
  depth      -> 80% overall, KHAT = 0.54
  drainage   -> 84% overall, KHAT = 0.69
CONCLUSIONS

• Lessons learned

  • Even with constrained training sample areas, cost/sample was too high
  
  • In the future, consider clustered or multi-stage sampling strategy for areas with high cost/sample
  
  • All applied modeling methods are common in DSM community and found applicable
  
  • Rule-based methods are intuitive to soil scientists and offer the advantage of refinement as knowledge of soil-landscape relationships develops
  
  • Utilization of random forests in ‘binary’ mode is worth exploring as a suggested operating procedure for future projects with limited training data
  
  • Strive for developing expertise within project offices
REFERENCES

1. ArcSIE http://www.arcsie.com


5. ESRI http://www.esri.com/

6. ERDAS Imagine http://www.hexagongeospatial.com/products/remote-sensing

7. GRASS http://grass.osgeo.org/


12. R http://cran.r-project.org/

13. SAGA GIS http://www.saga-gis.org