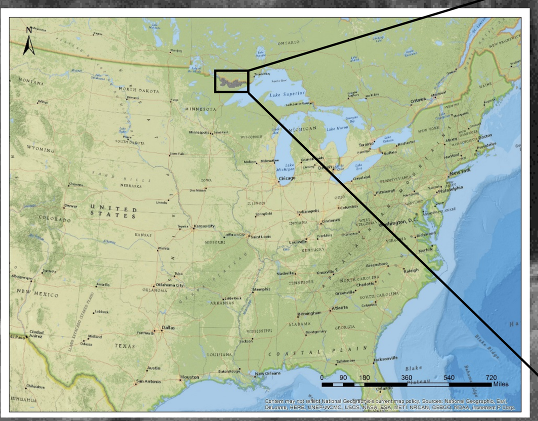
The Boundary Waters Canoe Area Wilderness (BWCA) Digital Soil Mapping Project

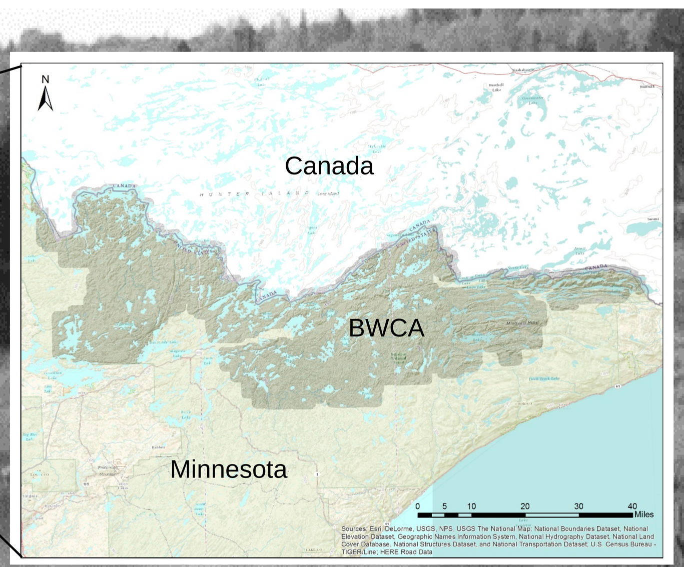
S. Kienast-Brown, T. D'Avello, J. Brennan, D. Evans, A. Koch, Z. Libohova, R. Long, J. Philippe, R. Risley, L. Schmitt, *USDA-NRCS,* R. Vaughn, *USFS*

INTRODUCTION

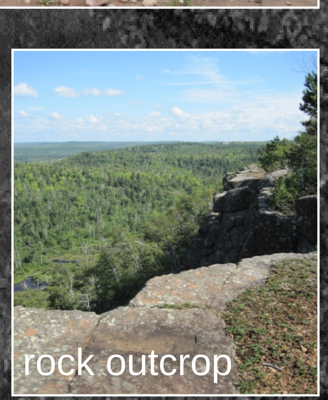
The BWCA wilderness area is approximately 1.1 million acres and was established in 1978 within the boundaries of the Superior National Forest in Northeastern Minnesota forming a border between Minnesota and Canada. This area is now located within North Central Glaciated Soil Survey Region 10. The BWCA is only accessible by foot or paddle; no road access exists. Management concerns in the BWCA are focused around recreational activities that include fishing, hunting, camping, canoeing, dog sledding, hiking, and skiing.

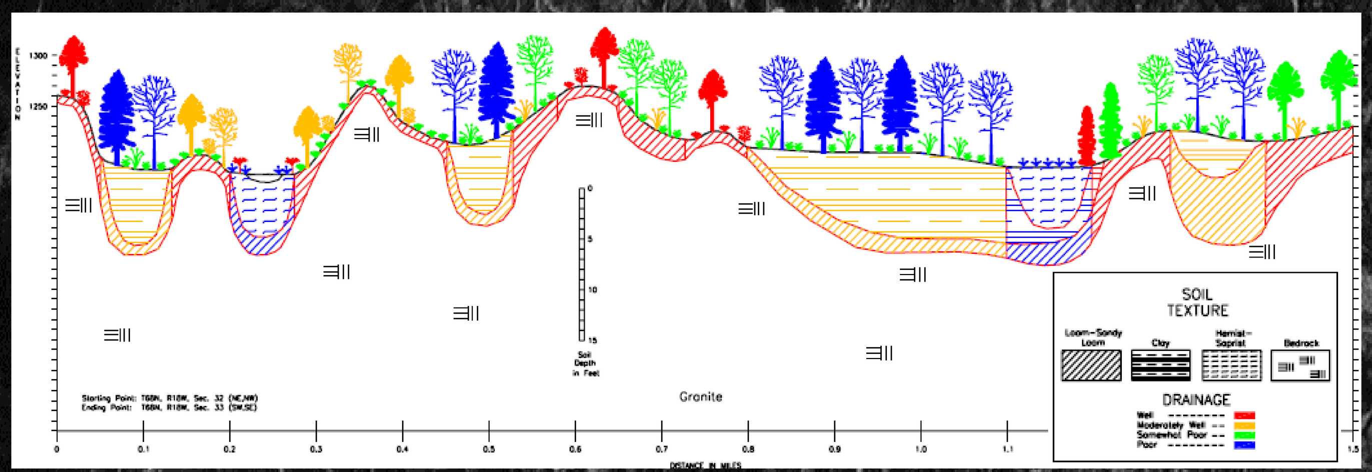
Remote wilderness lands present a unique opportunity and challenge for resource inventory. Digital soil mapping (DSM) is an effective method for producing soil survey products in remote areas with access limitations. In 2012, the USDA-Forest Service (USFS) and USDA-Natural Resources Conservation Service (NRCS) began a collaborative soil survey project over 595,000 acres of unmapped remote wilderness land in the BWCA.





The BWCA is a unique landscape carved over the last 2 million years by large continental glaciers. The landscape is a complex mosaic of open water, wetlands and bogs, forest, and rock outcrop. Four main parent materials occur in the BWCA: Rainy Lobe till, lacustrine Agassiz sediments, outwash deposits, and organic deposits. Soil depth classes range from bedrock exposure to very deep, and soil moisture classes range from ponded to well-drained.





Soil Forming Environment

Climate: frigid MAAT = 34.75˚F (1.5˚C); MAP = 27" (685mm)
Organisms: boreal forest
Relief: ~relation to depth and drainage
Parent Material: glacial drift over Precambrian bedrock
Age: ~12000 BP

METHODS

Environmental Covariates

- Multiple derivatives from 5m LiDAR and peak-of-green Landsat 5 imagery representing SCORPAN factors in the BWCA
- Covariate selection completed using Optimum Index Factor and random forests variable importance analysis

Final Predictor Variables

LiDAR	Landsat
Slope	Landsat band 1
Relative position	Landsat band 5
Curvature	Brightness, greenness, wetness band 1
Surface area factor	Brightness, greenness, wetness band 2
Depression cost surface	Principle component 6
Canopy height	NDVI

Training Data

- Conditioned Latin hypercube sampling method applied with 7 of the predictor variables listed above, plus geomorphons
- Sample locations constrained within 0.2 miles of trails or shorelines and contiguous travel routes to maximize access
- 214 pedon description sample points collected and used for modeling
- Field sampling proved difficult due to access, terrain, and conditions
- Optimum number of samples per class was not achieved

Classes

- 45 classes based on soil series identified initially using local expert tacit knowledge
- Thinned to 15 classes based on parent material, depth, and wetness with exploratory pre-map process (ISODATA unsupervised classification and ArcSIE rule-based classification)
- Final class separability evaluated with random forests class collapsing analysis

Final 11 Classes

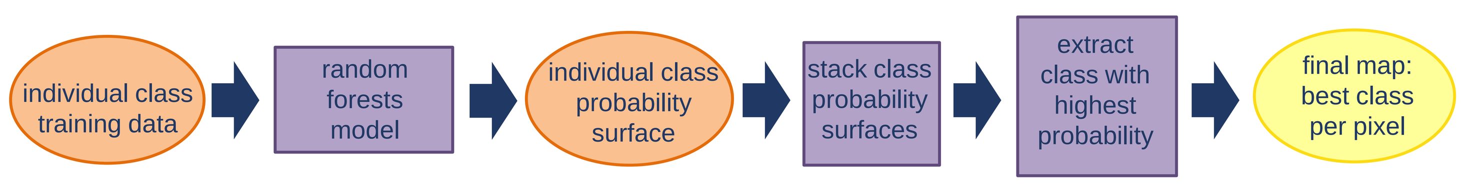
Original classes	# observations	Final classes	# observations
deep dry till	45	deep dry till	45
deep wet till	24	deep wet till	24
outwash	8	outwash	8
moderately deep dry till	25	moderately deep dry till	25
shallow dry till	21	shallow dry till	21
very shallow dry till	27	very shallow dry till	27
wet lacustrine	16	wet lacustrine	16
wet lacustrine, mantled	11	wet lacustrine, mantled	11
dry lacustrine	4	dry lacustrine	8
dry lacustrine, mantled	4		
dysic organics	6	organics	11
euic organics	3		
lithic organics	2		
mod dep wet till	3	mod deep wet till/shallow wet till	11
shallow wet till	8		
15 classes	207	11 classes	207

Modeling

- Rule-based, logistic regression, tree-based, and ISODATA unsupervised classification methods were applied and evaluated
  - Rule-based and tree-based methods for all classes
  - Logistic regression for lacustrine class
  - Unsupervised classification for dysic/euic organics class
- Qualitative review by method and class conducted by local experts

Random Forests Modeling

- Due to limited training data and poor model performance modeling all 11 classes simultaneously, each class was modeled separately
- Decreased out-of-bag error (OOB) from 58-75% to 4-22%
- Probability surfaces were predicted for each class
- Class with highest probability assigned to pixel to generate final map (best class per pixel)



Validation

- Field observations were collected based on a stratified random sampling design
- Areas of opportunity were selected based on access and diverse location
- Points were stratified randomly across classes within areas of opportunity
- Accuracy assessment confusion matrix was calculated using 207 observations, reporting user's and producer's accuracy by class, and overall map accuracy

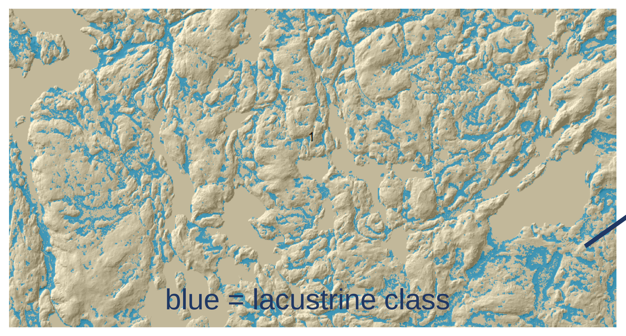
RESULTS

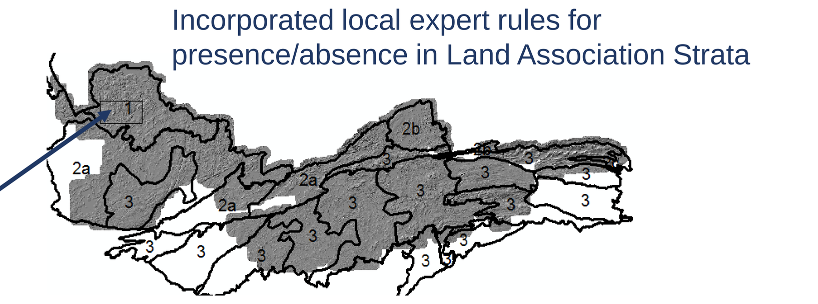
Environmental Covariates

- Logistic regression and tree-based methods used spectral and terrain variables
- Rule-based classification used terrain variables only
- Unsupervised classification used spectral variables only

Logistic Regression

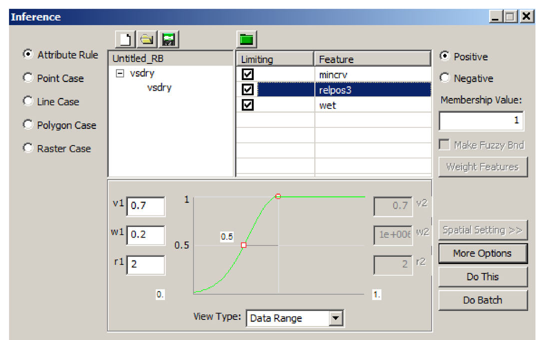
- Logistic regression using binary approach for lacustrine class

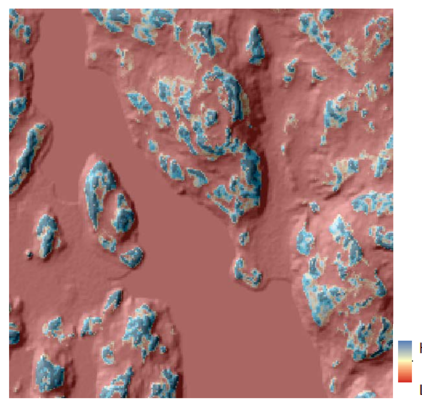




Rule-Based Classification

- ArcSIE used to create rules and fuzzy membership layer for each class

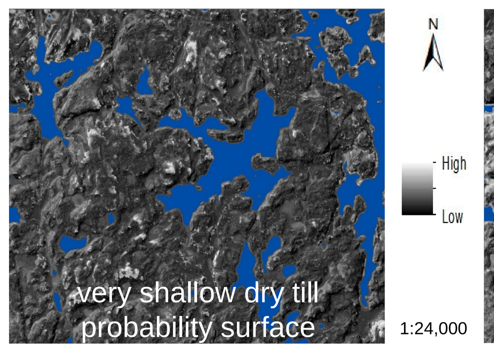


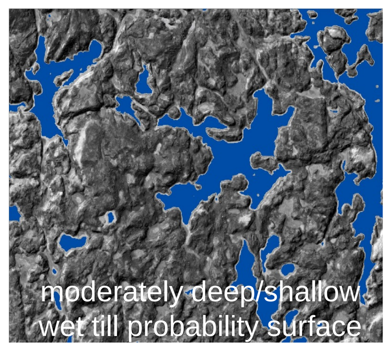


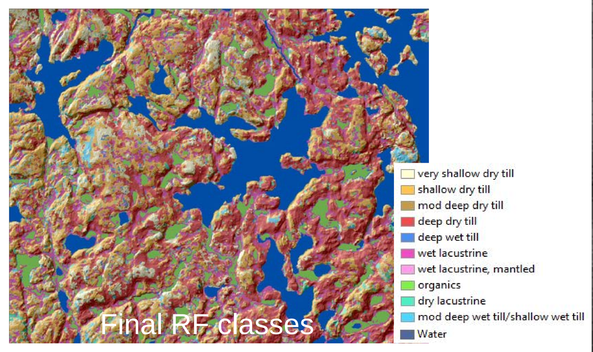
Example rules and fuzzy membership output for very shallow dry till class

Random Forests Modeling

- Modeled each class separately in a binary approach to produce probability surfaces and assign best class per pixel



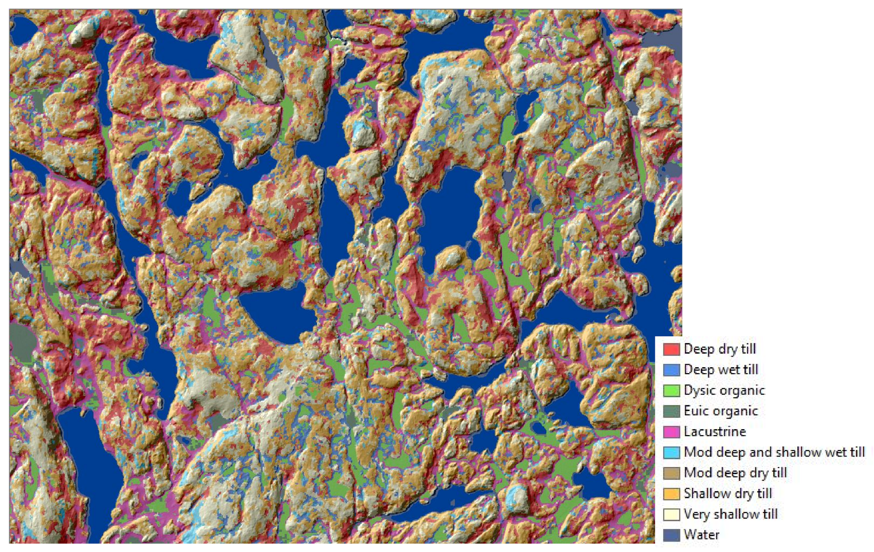




Final Thematic Map

- Hybrid assemblage based on qualitative review of local experts

Class	Source
deep dry till	random forests (RF)
moderately deep dry till	RF
deep wet till	RF
moderately deep and shallow wet till	RF
very shallow dry till	rule-based (RB)
Shallow dry till	RB
dysic organic	unsupervised
euic organic	unsupervised
lacustrine	logistic regression
water	LiDAR break lines



Validation (preliminary results)

- 64% overall accuracy for final hybrid map
  - Overall user's error = 4%; Overall producer's error = 36%

DISCUSSION

Application of DSM proved successful for predicting soil classes based on parent material, depth, and wetness in a remote wilderness area where extensive soils information was previously lacking. Challenges of access and field conditions impacted collection of training and validation points, as expected. Random forests offered a robust option for predicting classes with limited training data by modeling each class separately and using class probabilities to determine the "best class per pixel". This method reduced OOB error significantly from models predicting all classes simultaneously. Overall performance and individual class accuracy for all approaches could be improved with additional training data in low-performing classes. Results from random forests, rule-based classification, logistic regression, and ISODATA unsupervised classification were combined to produce the final hybrid raster soil class map for the BWCA. Much potential exists for application of predictive modeling in soil survey for initial, update, and disaggregation projects, and particularly for remote areas with limited resources for a traditional soil mapping approach.

ACKNOWLEDGEMENTS
Colby Brungard, Brooke Fonnesbeck, *Utah State University*
Bob Benton, Kevin Megown, Carston Stam, *USFS Remote Sensing Applications Center*
Utah State University