**Update Mapping Sub-Team Meeting**

**01/29/2019 – 11:00 AM to 1:00 PM Central**

**Month’s topic:** Implemented Sampling Design Local Approaches, Successes & Lessons Learned

**Roll Call**

1. Joe Brennan, St. Paul, MN
2. Dave White, Las Cruces, NM
3. Chance Robinson, Stephenville, TX
4. Jessica Phillipe, Saint Johnsbury, VT
5. Jamin Johanson, Dover-Foxcroft, ME
6. Stephen Roecker, Indianapolis, IN
7. Betsy Schug, Fergus Falls, MN
8. Kyle Thomson, Bismarck, ND
9. Brianna Wegner, Bismarck, ND
10. Jordaan Thompson, Fargo, ND
11. Tyson Morley, Altus, OK
12. Lynn Loomis, Marfa, TX
13. George Otto, Tuskegee, AL
14. Tiffany Smith, Waynesville, NC
15. Matthew Duval, Raleigh, NC
16. Jedd Bodily, Price, UT
17. Suzann Kienast-Brown, Bozeman, MT
18. Tom D’ Avello, Morgantown, WV
19. Chad Ferguson, Lincoln, NE
20. Linda Harring, Madison, WI
21. Wade Bott, SSS, Bismarck, ND
22. Jocelyn Wardrup, Univ. of Delaware

**Overview of Sampling Design and Considerations** (Suzann Kienast-Brown)

Suzann provided overview of the month’s topic from Intro to DSM course

* How do we address issues of modelling “small classes” with limited observations? (Stephen)

Suzann shared the following link which attempt to address class imbalance:

* [A Collection of Oversampling Techniques for Class Imbalance Problem Based on SMOTE](https://cran.r-project.org/web/packages/smotefamily/smotefamily.pdf)
* [smoteRegress: SMOTE algorithm for imbalanced regression problems](https://rdrr.io/cran/UBL/man/smoteRegress.html)
* [Classification of rare land cover types](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5784906/)
* [To Combat Multi-Class Imbalanced Problems by Means of Over-Sampling Techniques](https://www.computer.org/csdl/trans/tk/2016/01/07163639-abs.html)
* [Neighborhood Size of Training Data Influences Soil Map Disaggregation](https://www.mattlevi.com/uploads/5/1/0/1/51012121/levi_2017_neighborhoodsizetrainingdatasoilmapdisaggregation_sssaj.pdf)

**Initial Soil Survey Examples**

**Boundary Waters Canoe Area Wilderness, Minnesota** (Tom D’ Avello)

* Training points selected using conditioned Latin hypercube sampling
* 214 training points collected in 15 target classes

**Bob Marshall Wilderness, Montana** (Suzann Kienast-Brown)

* Extensive subsets by Parent Material
* Gower’s Similarity Index will be implemented

**White Mountains National Forest, New Hampshire** (Jessica Phillipe)

* Previous work in other projects
* Stratified grid sampling
* Random catena / stratified random
* Stratified random
* “edge data” made my modelling difficult
* Jamin asked for clarification of catena approach
  + Each catena has its own Random Forest model in current approach.

**Update Soil Survey Examples**

**MLRA 84B – Paluxy Erosional Hillslopes, Central Part, Texas** (Chance Robinson)

* Implemented a manual method of selecting replacement cLHS points
  + Plan to utilize Gower’s similarity index for future projects
* Testing for short-range field variability
  + Implemented RACA protocol and randomly selected subset of 10% of observations
  + Due to nature of aggregated classes modeled in soil survey, plan to modify and reduce this dataset for future projects

**MLRA 102A Fergus Falls Till Plain** (Jordaan Thompson)

* Testing the “offset plot radius” of the polypedon
* Collecting additional observation in next component

**MLRA 42 - Update of the Animas Valley Playas and adjoining map units** (Dave White)

* Training points selected using conditioned Latin hypercube sampling with cost
  + Land ownership access important in developing training dataset

**MLRA 90A Mille Lacs Uplands** (Betsy Schug)

* Utilize historical documentation reviewing locations closely
  + Progressive Soil Survey Documentation
  + Laboratory Analysis
* Purposive Geomorphic Traverse Documentation in Initial Soil Survey Area
  + 3-5 positions per site
  + Stratified by Surficial Geology Units & Old (1941) Soil Survey data
  + Targeting voids in point distribution from geological survey documentation and existing soil survey pedons
* Documentation Standards
  + Assure all documentation could be keyed out consistently for most important features of interest in the project
    - Aquic conditions
    - Mantle thickness
    - Organic soils
    - Document Plot Radius in Field
      * Radius with which the class concept is constant
* Sampling Plan in Update Area (Targeting 100 descriptions)
  + Stratified by Land Cover
  + cLHS applied to each strata for:
    - 12 sites in Cropland Strata
    - 28 sites in Forest Strata
  + Additional upslope and downslope offsets (one or two for site)
  + cLHS used most significant predicting terrain covariates to properties from existing soil map
* 282 total training points

**Closing Remarks – Follow-up Discussion in February.**

**2-year MLRA Field Project Template and Raster Product Updates at next meeting.**

**Next meeting on February 26, 2019**