Sampling Discussion

8/2/19

1. Suzann Kienast-Brown
2. Jessica Philippe
3. Alex Stum
4. Stephen Roecker
5. Travis Nauman
6. Dave White
7. Jim Thompson
8. Tom D’Avello
9. Colby Brungard

Discussion items:

* cLHS in general
  + Use/implementation of the similarity buffers – is there a more efficient way to implement in R package?
  + Tom’s clustering approach as an alternative to similarity buffers
  + Thoughts/guidance on number of covariates compared to number of points generated
  + Thoughts/guidance on how many points should be generated
  + Thoughts on landform traverse or other supplemental samples to primary cLHS points
* cLHS in practice
  + What pitfalls are occurring in projects and how can we best provide guidance to avoid them?
  + Access constraints and ranking of points
  + Effective sampling approach if project isn’t positioned to generate cLHS points yet (no covariates developed, etc) and field season is upon them

Group Discussion:

* Tom – discuss Salmon-Challis NF project
  + Access was very limited – initial cLHS constrained to access areas
    - Many of the initial points were dropped as being too difficult for the crew to access
  + The goal was to find areas that are accessible and have high similarity to dropped points
  + Clustered dropped points into classes to reduce the number of unique instances
  + Create fuzzy membership layer for each class
  + Field crew uses fuzzy membership layer to identify sampling locations that are most similar to dropped points and more easily accessible
  + Fuzzy membership layers chosen because of contiguous extent over project area vs similarity index buffers already integrated into cLHS R package
* Stephen
  + Fuzzy clustering of covariates and design sampling on resulting clusters (McKenzie and Ryan)
  + This is similar to the method used in early NRCS projects – before cLHS
  + May be easier to implement/more flexible
  + Job aid with at least 2 models/approaches to choose from
* Colby
  + Divide the area based on covariate space (clustering)
  + Identify representative areas
  + Sample representative areas in a transect/traverse manner to capture variability
  + Some areas are inaccessible and won’t ever be sampled
    - Make it clear that this is a reality
    - Predictions in those areas will be extrapolations and may be more wrong than right
    - Implement a quantitative method to show where predictions are interpolations vs extrapolations
    - Maybe it’s ok to have high uncertainty in those areas – low management and use
  + Use cLHS points as centroids – compute similarity of every pixel from that centroid (this was the basic concept Toma applied in his fuzzy membership approach)
  + Identify locations that can be sampled considering access, but if not all points can be sampled, uncertainty will be high in those areas – is that acceptable? If not, those areas have to be sampled somehow
  + MESS algorithm in R – looks at the distribution of all observations in covariate space and measures similarity of each point to a set of reference points with respect to covariates – produces a map of similarity values and shows geographic area required for capturing variability
* Dave
  + Don’t run cLHS over entire project area – you get a lot of points that aren’t in great spots (tweeners)
  + Grouping SSURGO map units within a representative area and base sampling in those areas
* Travis
  + What is meant by “points are tweeners or are in a bad place”?
  + Two competing objectives: model that classifies soil types or typify the central concepts
    - People want ONE pedon to typify their central concept for series or map units
    - Methods based bias – how do you define the central concept if you haven’t observed the complete variability in the soil-landscape relationships?
  + Transitional approach – if you are sampling for DSM – use the non-biased approach and supplement with biased sampling
  + Beefed up stratified sampling used in BLM and other agency sampling methods
    - Look at cLHS with regard to these algorithms
    - How do we replace points we can’t get to?
* Jess
  + Stratify by parent material and model separately
  + People will still do what they want in practice in the field unfortunately (with regard to cLHS or other sample design)
  + Try to communicate in language they are familiar with (mapping orders for areas of high/low uncertainty)
  + The human factor is hard to control but hopefully through more training and knowledge this will improve
* Thoughts/guidance on number of covariates compared to number of points generated
  + Less is more – keep around 10-12 or less
  + Use soil-landscape relationship knowledge; keep it simple
  + More covariates require more training points in the modeling step – find the balance
  + Could look at standard set of covariates and consider basing guidance on that as a starting point
* Iterative sampling design workflow – improve subsequent sampling campaigns based on results and preliminary modeling from first sampling
* Flexible sampling and project design based on changing schedules, staff, etc
  + Include how to evaluate the samples you have to determine if they are capturing variability of covariates
  + Run cLHS twice – once with covariates, then second time on selected points/covariate values from first run
* Goals
  + Change the paradigm – this will happen through training and practice over time…don’t give up!
  + Reasonable sampling design considering access constraints and point replacement options
* Job aid
  + Contribute – Suz, Jess, Stephen, Tom, Alex, Dave (beg of Oct)
  + Review – Travis, Jim, Colby