

Spatially-Informed Sampling Approach for Cost-Effective Soil Health Assessment

Dipesh Giri, Purdue University (E-Mail: girid@purdue.edu);
Sajad Jamshidi, Purdue University; **Yichao Rui**, Purdue University; **Siddhartho S. Paul**, Purdue University

Introduction: Soil properties vary spatially with topography, parent material, cropping system, and management, creating a challenge for site-specific management. Effective sampling depends not only on how many samples are collected, but on how well those samples represent the full range of environmental covariates used as predictors. Conditioned Latin Hypercube Sampling (cLHS) uses simulated annealing to select locations that represent the multivariate distribution of the covariates. This study applies cLHS at two USGS edge-of-field water quality monitoring sites in Indiana (IN) and Wisconsin (WI).

Methodology:

- Covariates were derived from LiDAR-derived terrain attributes, SSURGO soil properties, Sentinel-2 indices, and prior soil map generated by kriging with external drift.
- All layers were resampled to a 5 m grid, producing initial stack of 80 covariates at WI and 88 at IN; reference sampling grids were set at 25 m (WI) and 30 m (IN)
- Redundant covariates were removed using a variance inflation factor filter ($VIF > 5$), retaining 11 covariates at IN and 7 at WI.
- cLHS was run repeatedly across a range of sample sizes, and Kullback-Leibler divergence was calculated between the sampled covariate distribution and the full field covariate population at each size.
- Optimal sample size was taken at the knee of KLD curve, where further reductions began to degrade feature-space coverage.

Results and Conclusion:

- Knee detection on the KLD curves gave optimal sizes of $n=50$ (IN) and $n=45$ (WI), reflecting differences in catchment size and covariate diversity.
- cLHS reduced sampling effort by ~70% at IN and ~60% at WI compared to full grid sampling while retaining feature-space coverage.
- This optimal sampling design provides a cost-effective foundation for digital soil mapping of key soil health indicators and delineation of critical source areas for nutrients

FOR FURTHER READING:

Minasny, B. & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences*, 32(9), 1378–1388.
Malone, B. P., Minasny, B. & Brungard, C. (2019). Some methods to improve the utility of conditioned Latin hypercube sampling. *PeerJ*, 7, e6451.

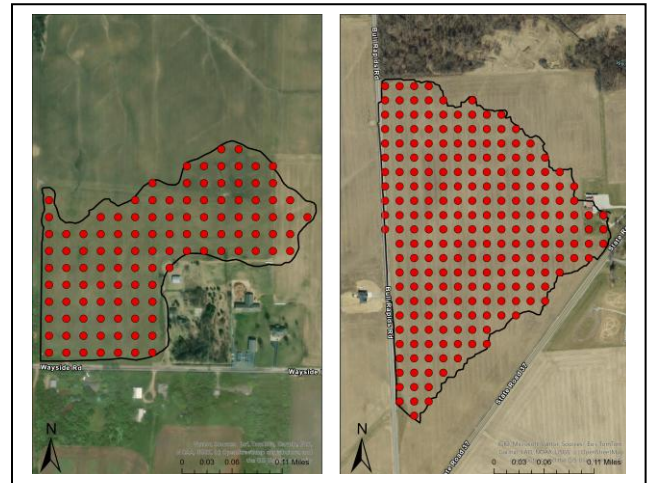


Figure 1. Spatial distribution of grid sampling locations (red circles) for WI & IN catchment.

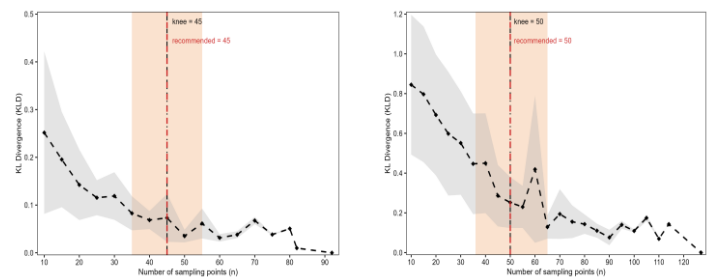


Figure 2. KL divergence as a function of sample size for WI & IN

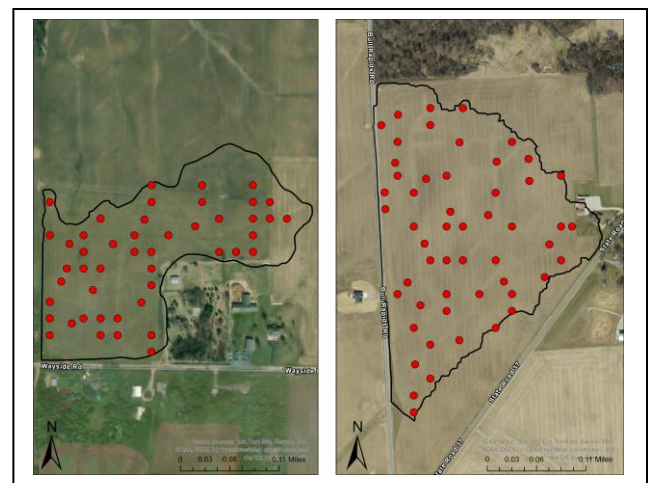


Figure 3. Spatial distribution of cLHS-selected sampling locations (red circles) for the WI & IN catchment