IOWA STATE UNIVERSITY
Department of Agronomy

Yones Khaledian and Bradley Miller

Soil Sample Optimization for Field Scale Map

Research Question

How many samples are needed to 'accurately' map soil in a field?

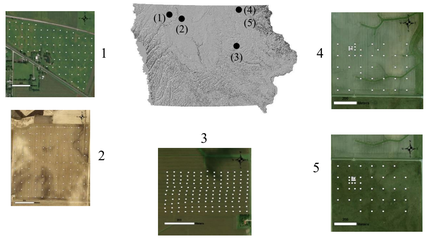
Objective

The main purpose of this research is to determine how accuracy of soil organic carbon (SOC) predictions change with different sampling designs and densities, based on spatial regression modelling.

Materials and Methods

To identify the optimal sample density using data from multiple fields in Iowa, USA (figure 1), we selected different subsets of those points by gradually reducing the number of model training points. In this study, we consistently used the same covariates (terrain derivatives) and model building method (Cubist to develop rule-based, multiple linear regression models) to predict soil organic carbon. The map accuracy of decreasing sample densities compared with the original sample density (grid pattern) using different sampling strategies (random, stratified by slope classes and digital hillslope positions). The Cubist algorithm was run 50 times to represent the standard deviation of the model results. The accuracy of the models was measured using the R^2 and RMSE of the cross-validation (~10 points).

Study Areas



Fields	Original Sampling Strategy	Land Cover
1	Regular Grid	Row Crops
2	Regular Grid	Row Crops
3	Regular Grid	Row Crops
4	Nested Grid	Row Crops
5	Nested Grid	Prairie

Figure 1. The locations of study areas and the original sampling strategies

Results

Our results indicated that for a regression-based model, there was high variability and no clear pattern in model accuracy (figure 2) and standard deviation (figure 3) as the quantity of randomly selected samples were reduced. In contrast, stratified strategies can maintain similar prediction accuracy with much fewer samples. The accuracy of maps based on stratified sampling by hillslope positions was higher than stratified by slope gradient.

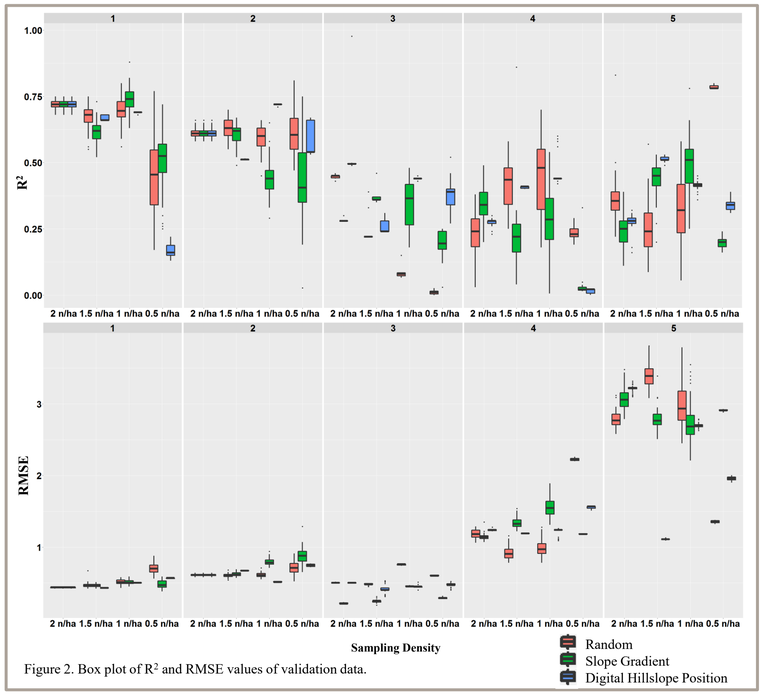


Figure 2. Box plot of R^2 and RMSE values of validation data.

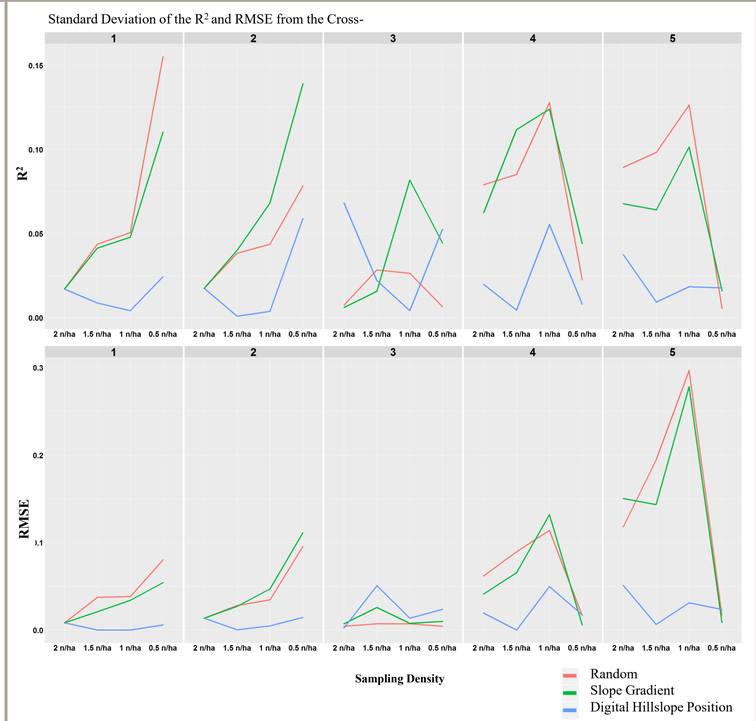


Figure 3. The variation of standard deviation of the R^2 and RMSE from the Cross-Validation.

Conclusions

- Grid sampling is not an optimized sampling design.
- More samples do not always improve the accuracy of spatial models.
- The R^2 of cross-validations were maintained until approximately 1 sample/ha.
- This suggests that in Iowa, the optimal sampling density is 1-1.5 samples/ha.

Geospatial Laboratory for Soil Informatics

Acknowledgements
We thank the Iowa Soybean Association, Dr. Michael Thompson, Dr. Skye Wills, and Dr. Lee Burras for the use of their data.