How to Collect Data on Agricultural Nutrient Management Practices: Survey Results from Iowa

Joint Statistical Meetings, 2024

August 8, 2024

A CSSM collaboration with INREC, and Dept. of Agriculture, ISU

Overview

Background & Motivation

Why nutrient management practice?

- Nitrogen(N) & Phosphorus(P)) flux to water resources potentially....
  - impacts human health and aquatic system integrity
  - causes hypoxia in Gulf of Mexico via increased algae growth
    - which decreases oxygen levels in water system
- Conventional row-crop agriculture: identified as a leading source of nutrient loading to the Gulf of Mexico.
- The 2008 Gulf Hypoxia Action Plan was introduced to
  - mitigate nutrient loading on Gulf of Mexico

Why nutrient management practice?

- Nitrogen(N) & Phosphorus(P)) flux to water resources potentially....
  - impacts human health and aquatic system integrity
  - causes hypoxia in Gulf of Mexico via increased algae growth
    - which decreases oxygen levels in water system
- Conventional row-crop agriculture: identified as a leading source of nutrient loading to the Gulf of Mexico.
- The 2008 Gulf Hypoxia Action Plan was introduced to
  - mitigate nutrient loading on Gulf of Mexico

Why nutrient management practice?

- Nitrogen(N) & Phosphorus(P)) flux to water resources potentially....
  - impacts human health and aquatic system integrity
  - causes hypoxia in Gulf of Mexico via increased algae growth
    - which decreases oxygen levels in water system
- Conventional row-crop agriculture: identified as a leading source of nutrient loading to the Gulf of Mexico.
- The 2008 Gulf Hypoxia Action Plan was introduced to
  - mitigate nutrient loading on Gulf of Mexico

Gulf of Mexico

Figure: Geographical position Gulf of Mexico

Why nutrient management practice in Iowa?

- Natural question to ask
  - being so far from the gulf, how Iowa plays a potential part in the contamination of northern gulf of Mexico?
- **Answer is.....**

One Mississippi

Two Mississippi

Why nutrient management practice in Iowa?

- Natural question to ask
  - being so far from the gulf, how Iowa plays a potential part in the contamination of northern gulf of Mexico?
- **Answer is.....**

One Mississippi

Two Mississippi

Why nutrient management practice in Iowa?

- Natural question to ask
  - being so far from the gulf, how Iowa plays a potential part in the contamination of northern gulf of Mexico?
- **Answer is.....**

One Mississippi

Two Mississippi

Why nutrient management practice in Iowa?

- Natural question to ask
  - being so far from the gulf, how Iowa plays a potential part in the contamination of northern gulf of Mexico?
- **Answer is.....**

One Mississippi

Two Mississippi

Its Mississippi....

How Iowa responded?

- Iowa Nutrient Reduction Strategy (INRS):
  - Developed from partnership of Iowa Department of Agriculture, and College of Agriculture and Life Sciences at Iowa State University
- The INRS directs efforts to reduce nutrients in surface water from both point and non-point sources

How Iowa responded?

- Iowa Nutrient Reduction Strategy (INRS):
  - Developed from partnership of Iowa Department of Agriculture, and College of Agriculture and Life Sciences at Iowa State University
- The INRS directs efforts to reduce nutrients in surface water from both point and non-point sources

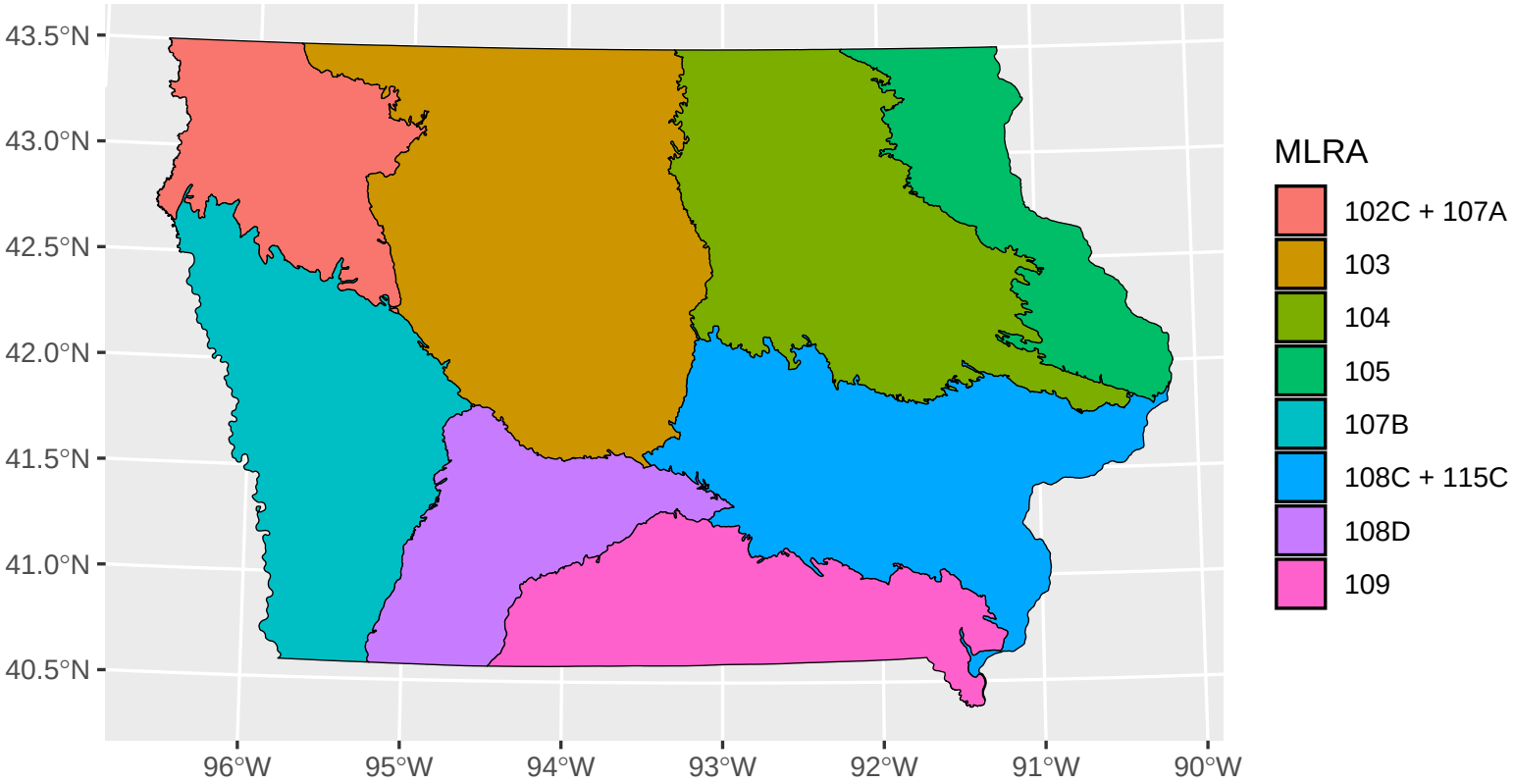
Objective

- To collect data on nutrient practices throughout Iowa using a balanced sampling design
- Using the finite sample data, extrapolate the results of nutrient practices for entire state
- Discuss results of all the surveys from 2017 to 2022

Survey design & sampling methodology

Study region

- Entire state of Iowa
- Across all 8 MLRAs(Major Land Resource Areas)



The finite population

- The available list of 586 agricultural retailers from Iowa
- Available information for retailers
  - Address
  - Company affiliation
- From the given address we geocoded the corresponding lat-long information

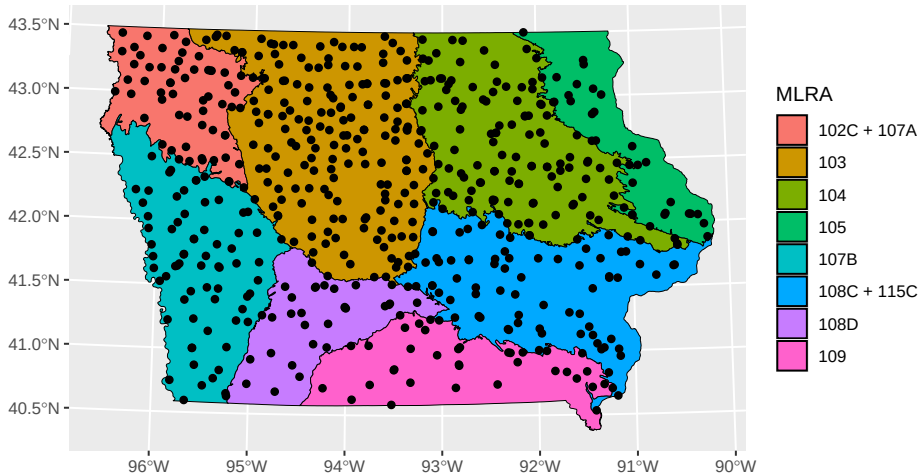


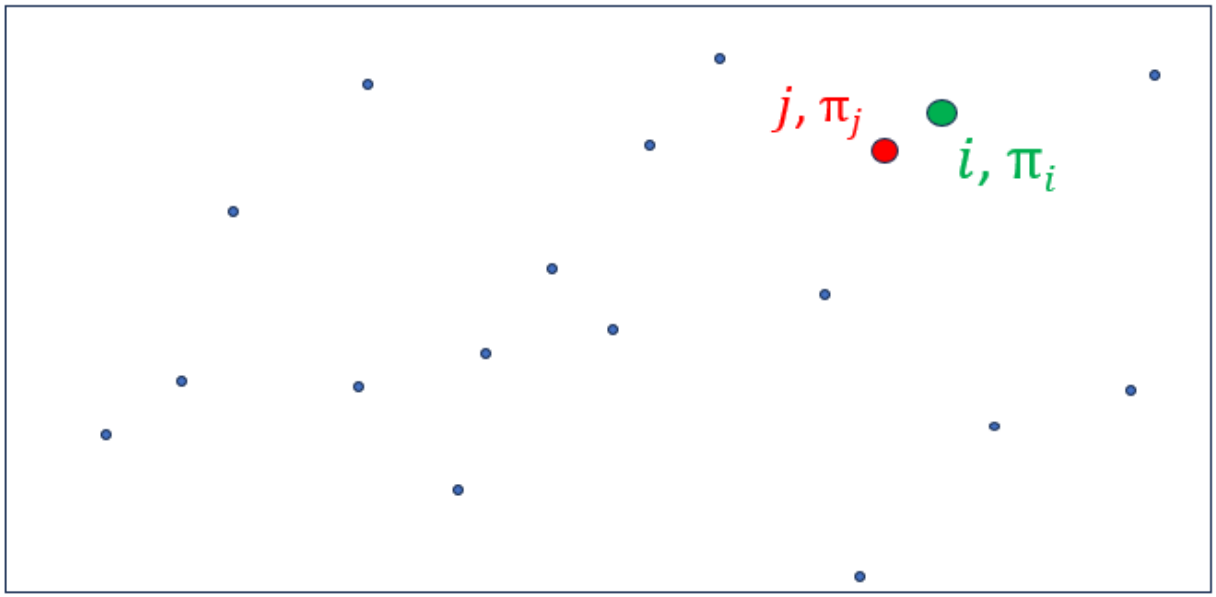
Figure: All retailer locations from 2022 survey

Sampling method : LPM

- Local Pivotal Method(LPM)[Grafström et al. (2012)] is used for obtaining a balance design
- The main idea is to create a strong negative correlation between the inclusion indicators of units that are close in distance.
- Consider notations
  - U be a spatial population of N units
  - Each unit i has a known location and a prescribed inclusion probability $0 < \pi_i \leq 1$
  - Expected sample size : $n = \sum_{i \in U} \pi_i$

LPM algorithm

- **Step 1:** Randomly choose one unit i
- **Step 2:** Choose unit j, a nearest neighbour to i. If two or more units have the same distance to i, then randomly choose any one of them with equal probability



LPM algorithm

- **Step 3:**
  - If j also has i as its nearest neighbor, then update the inclusion probabilities(π_i', π_j') as follows:
    - If $\pi_i + \pi_j < 1$,

$$(\pi_i', \pi_j') = \begin{cases} (0, \pi_i + \pi_j) & \text{w.p.} \quad \frac{\pi_j}{\pi_i + \pi_j} \\ (\pi_i + \pi_j, 0) & \text{w.p.} \quad \frac{\pi_i}{\pi_i + \pi_j} \end{cases}$$

and if $\pi_i + \pi_j \geq 1$, then

$$(\pi_i', \pi_j') = \begin{cases} (1, \pi_i + \pi_j - 1) & \text{w.p.} \quad \frac{1-\pi_j}{2-\pi_i-\pi_j} \\ (\pi_i + \pi_j - 1, 1) & \text{w.p.} \quad \frac{1-\pi_i}{2-\pi_i-\pi_j} \end{cases}$$

  - Else, return to Step 1
- An unit, say i, is called finished if $\pi_i' = 0$ or $\pi_i' = 1$. Then it will be removed from the population and cannot be chosen again.
- This updating procedure is repeated with the updated inclusion probabilities until all units are finished.

LPM algorithm

- **Step 3:**
  - If j also has i as its nearest neighbor, then update the inclusion probabilities(π_i', π_j') as follows:
    - If $\pi_i + \pi_j < 1$,

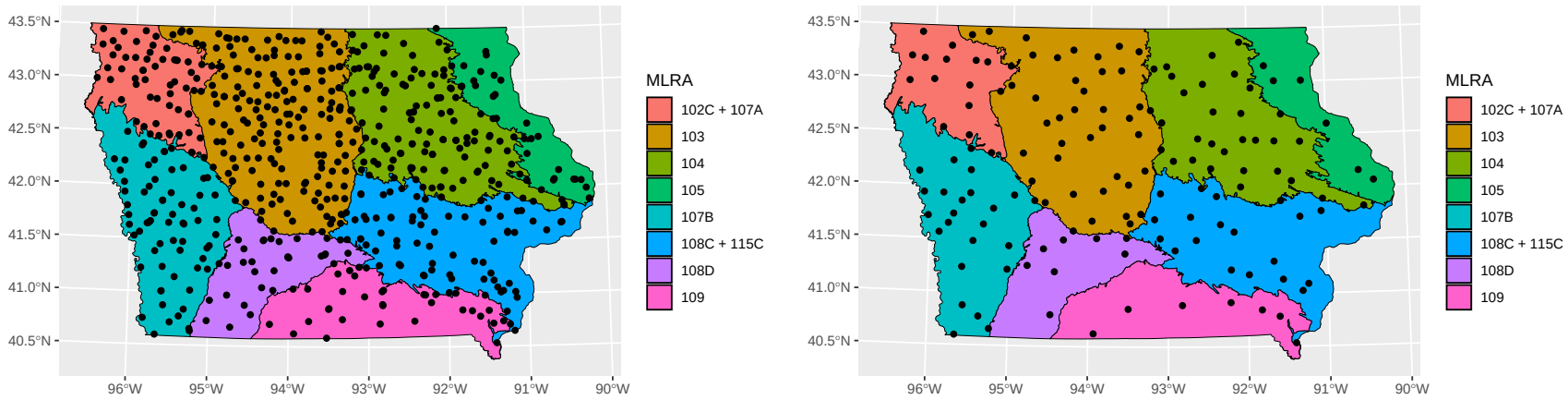
$$(\pi_i', \pi_j') = \begin{cases} (0, \pi_i + \pi_j) & \text{w.p.} \quad \frac{\pi_j}{\pi_i + \pi_j} \\ (\pi_i + \pi_j, 0) & \text{w.p.} \quad \frac{\pi_i}{\pi_i + \pi_j} \end{cases}$$

and if $\pi_i + \pi_j \geq 1$, then

$$(\pi_i', \pi_j') = \begin{cases} (1, \pi_i + \pi_j - 1) & \text{w.p.} \quad \frac{1-\pi_j}{2-\pi_i-\pi_j} \\ (\pi_i + \pi_j - 1, 1) & \text{w.p.} \quad \frac{1-\pi_i}{2-\pi_i-\pi_j} \end{cases}$$

  - Else, return to Step 1
- An unit, say i, is called finished if $\pi_i' = 0$ or $\pi_i' = 1$. Then it will be removed from the population and cannot be chosen again.
- This updating procedure is repeated with the updated inclusion probabilities until all units are finished.

Selected sample of 150 Iowa ag retailers



(a) (b)

Figure: a) All retailer locations from 2022 survey; b) 150 sampled retailer locations

Data collection

What data are collected and sent to CSSM?

- NASS CDL cropland totals of corn & soy acreages for all individual MLRAs
- Summary(min, max, avg, quantiles, SE etc.) of the amount of surveyed field size for individual MLRAs
- For each individual MLRAs, percentage of land where
  - different variant of Commercial N are applied
  - different variant of Commercial P are applied
  - Fall anhydrous amonia are applied
  - different variant of bio manures are applied
  - manures are used over different period of year
  - different tillage practices are adopted prior to corn and soy

Estimation and extraploration of results

Estimation

- Comprises of two different type of estimation(based on data reported as actual numbers or percentages)
  - Case 1
    - Actual amount of N & P applied(lb/ac) on fields with continuous corn and corn on rotation practices
  - Case 2
    - Percentage(or, proportions) of land covers where different categories/variables of interest are applied

Estimation : Case 1

- Notations
  - $h : h = 1, \cdots, 8$: denote each of the MLRAs as one stratum
  - S_h : Total corn and soy acres from h
  - m_h : Number of retailers sampled from h
  - q_h : Average surveyed field size from each of the selected m_h fields in h
- Define,
  - $r_h = \frac{m_h q_h}{S_h}$: Ratio of surveyed acreages over the total acreages in h

Estimation : Case 1

- Consider estimation of Commercial only N-rate in fields with only continuous corn production
- Survey component variables available from data are
  - n_h : sample size(the number of retailers in h, reported using Commercial N for fields with continuous corn)
  - $\bar{y}_h$: sample mean(mean commercial N rate on fields with continuous corn from h)
  - s_h : sample standard deviation(SD of commercial N rate on fields with continuous corn from h).

Estimation : Case 1

- Consider estimation of Commercial only N-rate in fields with only continuous corn production
- Survey component variables available from data are
  - n_h : sample size(the number of retailers in h, reported using Commercial N for fields with continuous corn)
  - $\bar{y}_h$: sample mean(mean commercial N rate on fields with continuous corn from h)
  - s_h : sample standard deviation(SD of commercial N rate on fields with continuous corn from h).

Estimation : Case 1

- Consider estimation of Commercial only N-rate in fields with only continuous corn production
- Survey component variables available from data are
  - n_h : sample size(the number of retailers in h, reported using Commercial N for fields with continuous corn)
  - $\bar{y}_h$: sample mean(mean commercial N rate on fields with continuous corn from h)
  - s_h : sample standard deviation(SD of commercial N rate on fields with continuous corn from h).

Estimation : Case 1

- Consider estimation of Commercial only N-rate in fields with only continuous corn production
- Survey component variables available from data are
  - n_h : sample size(the number of retailers in h, reported using Commercial N for fields with continuous corn)
  - $\bar{y}_h$: sample mean(mean commercial N rate on fields with continuous corn from h)
  - s_h : sample standard deviation(SD of commercial N rate on fields with continuous corn from h).

Estimation : Case 1

- Then, the mean commercial-only N-rate applied on the fields with continuous corn production for the entire state of Iowa can be estimated as,

$$\hat{\bar{y}}_U = \frac{\sum_{h=1}^{8} \bar{y}_h N_h}{\sum_{h=1}^{8} N_h},$$

where $N_h = r_h n_h$, with the standard error of the estimate is given by,

$$SE(\hat{\bar{y}}_U) = \frac{\sum_{h=1}^{8} \frac{N_h^2 \left(1 - \frac{n_h}{N_h}\right) s_h^2}{n_h}}{\left(\sum_{h=1}^{8} N_h\right)^2}$$

Estimation : Case 2

- **Objective**
  - To appropriately estimate the proportions for the categories of interest
  - To extrapolate those estimated proportions on the entire state level with the help of NASS CDL based statewide corn and soy acreages data.
- For example, consider the estimation of proportions of farmland in Iowa where Fall Anhydrous is applied.

Estimation : Case 2

- **Objective**
  - To appropriately estimate the proportions for the categories of interest
  - To extrapolate those estimated proportions on the entire state level with the help of NASS CDL based statewide corn and soy acreages data.
- For example, consider the estimation of proportions of farmland in Iowa where Fall Anhydrous is applied.

Estimation : Case 2

- Notations
  - p_U : proportions of farmland in Iowa where Fall Anhydrous is applied
  - n_h : number of retailers who applied Fall Anhydrous in the h-th stratum
  - n_{hT} : total numbers of retailers from MLRA h
  - $p_h = \frac{n_h}{n_{hT}}$
- Then the estimate of p_U would be,

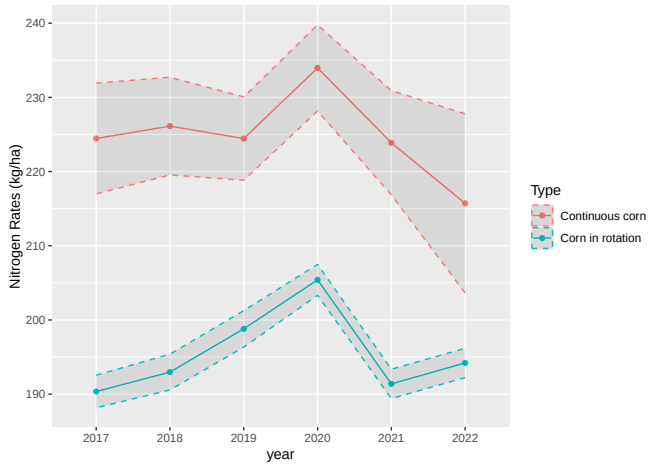
$$\hat{p}_U = \frac{\sum_{h=1}^{8} p_h N_h}{\sum_{h=1}^{8} N_h},$$

and the SE of the estimator will be,

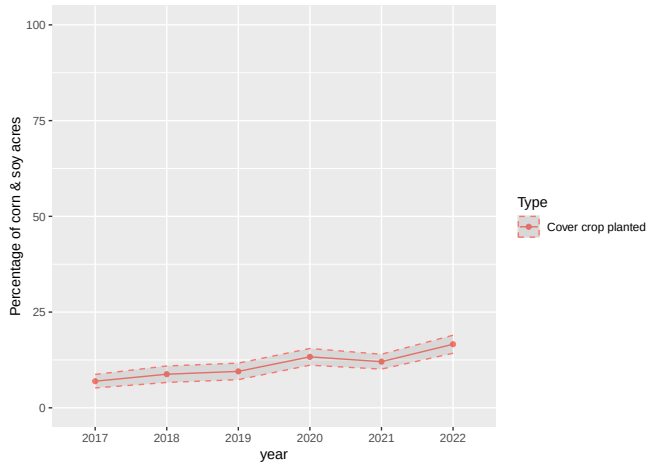
$$SE(\hat{p}_U) = \frac{\sum_{h=1}^{8} \frac{N_h^2 \left(1 - \frac{n_h}{N_h}\right) s_h^2}{n_h}}{\left(\sum_{h=1}^{8} N_h\right)^2},$$

with s_h being the sample standard deviation calculated for the h-th stratum.

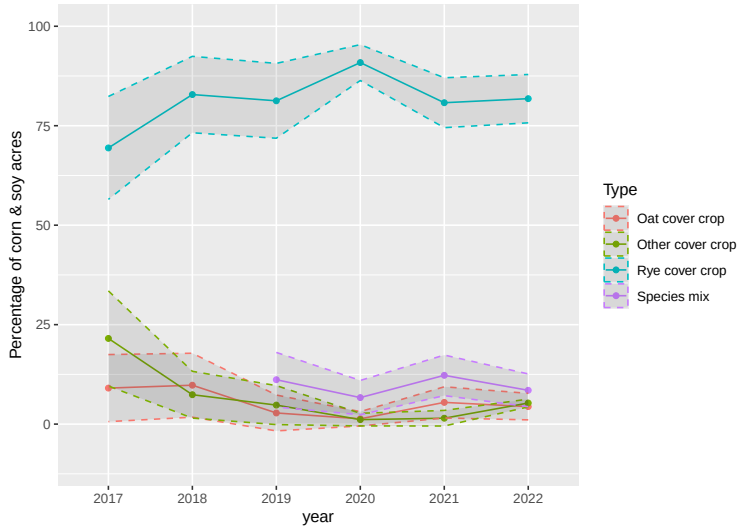
Results for statewide extrapolations



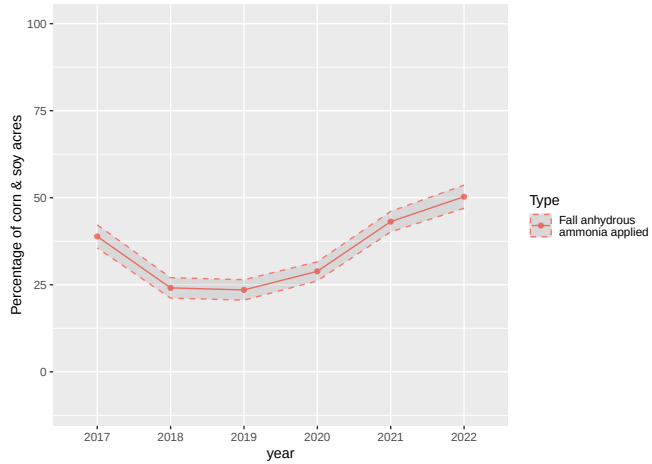
(a)



(b)

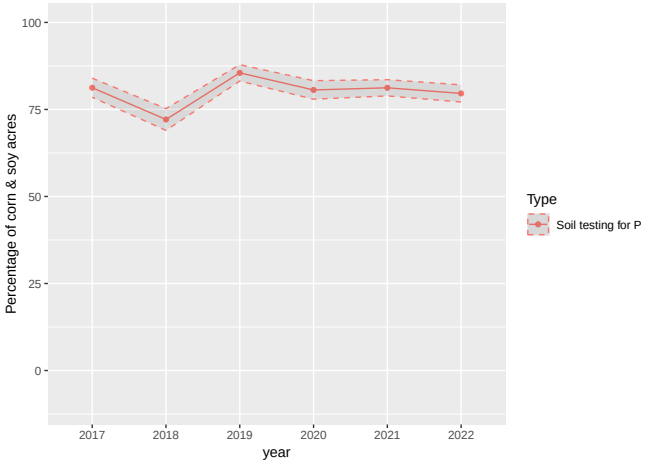


(c)



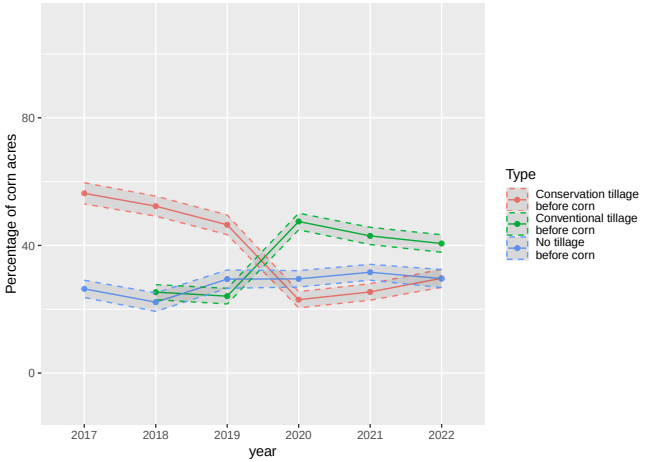
(d)

Results for statewide extrapolations

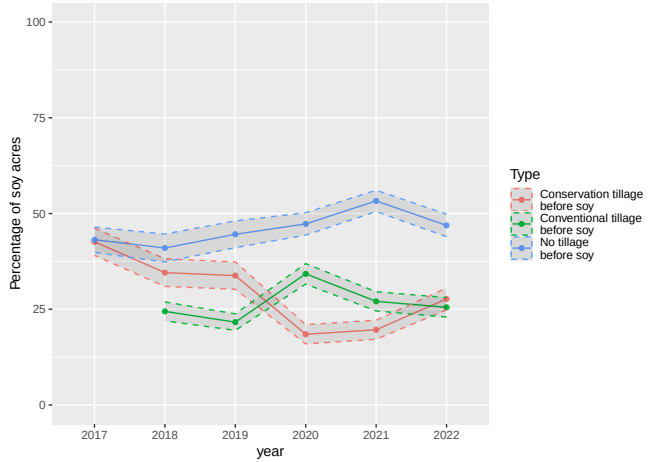


(a)

(b)



(c)



(d)

Road ahead

Illinois on board too

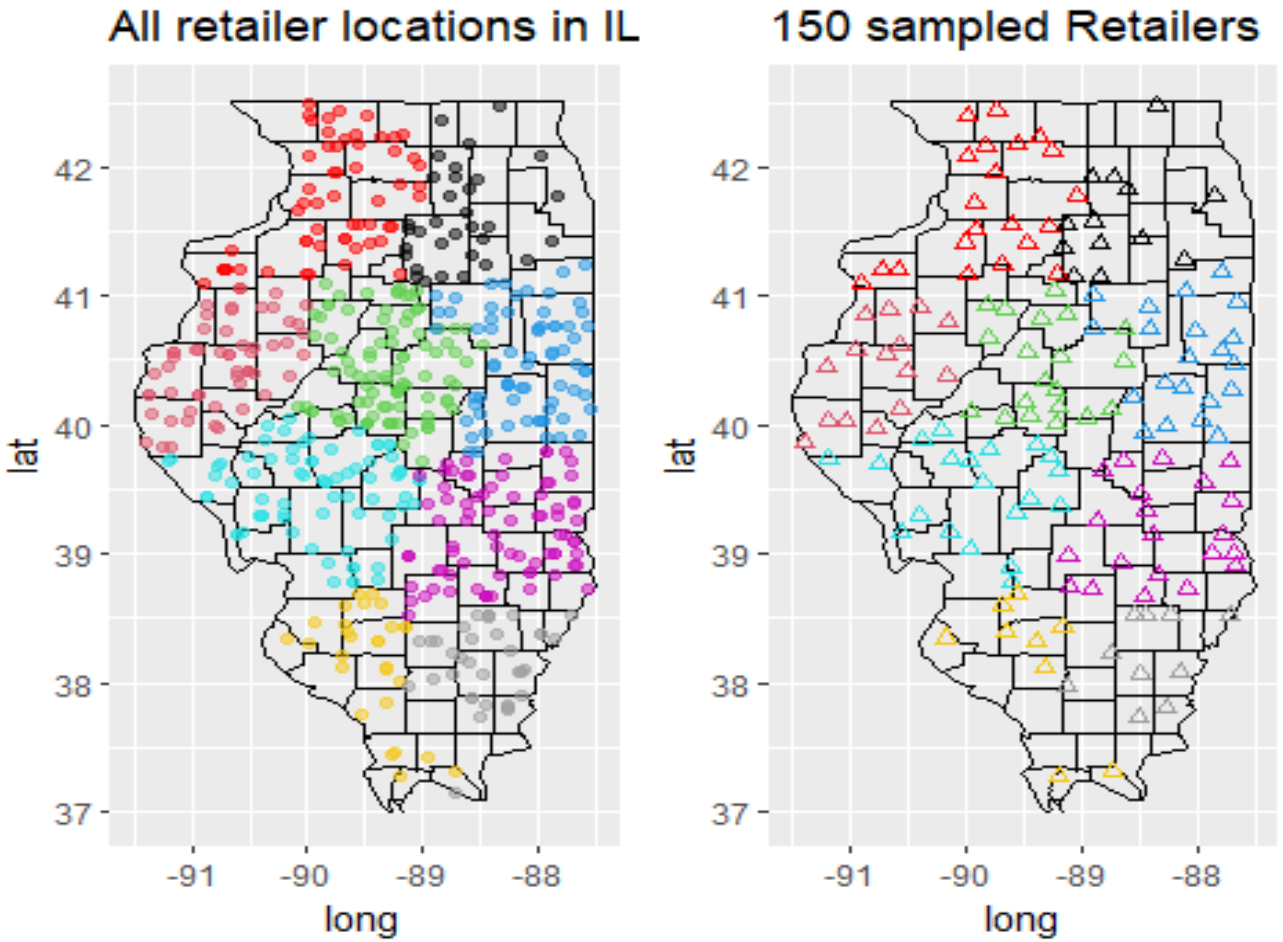


Figure: Selection of 150 sampled retailers from Illinois in 2023

Acknowledgement

This is a joint work with my Co-Authors:

- Zhengyuan Zhu [Department of Statistics, CSSM; Iowa State University]
- Ben Gleason[INREC]
- Rob Davis [INREC, Iowa State University]
- Matthew J Helmers [Department of Agricultural and Biosystems Engineeriing, Iowa State University]
- Thomas M Isenhart [Department of Natural Resource Ecology and Management, Iowa State University]

Reference

Deville, J.-C. and Tille, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika*, 85(1):89–101.

Grafström, A., Lundström, N. L. P., and Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2):514–520.

Thank You!!