NCSRP – final report January 2025

Team members:

- Ignacio A. Ciampitti, Principal Investigator. Kansas State University
- Matthew Carroll, co-PI, PhD. Analytics & Insights Lead. Iowa Soybean Association Research Center for Farming Innovations

Collaborators:

- John P. Fulton, PhD, Professor, Food, Agricultural and Biological Engineering, Ohio State University.
- Shaun Casteel, PhD, Associate Professor, Extension Soybean Specialist, Purdue University.
- Peter Kovacs, PhD, Assistant Professor, Precision Cropping Systems, South Dakota State University.
- Andre de Borja Reis, PhD, Assistant Professor, Division of Plant Sciences, University of Missouri.
- Scott Nelson, Sr. Research Agronomist, Iowa Soybean Association.
- Mark Seamon, Research Director, Michigan Soybean Promotion Committee.
- Maninderpal Sing, PhD, Assistant Professor of Cropping Systems Agronomy, Michigan State University.
- Randy Pearson, PhD, Professor, Southern Illinois University of Edwardsville.
- Michael Ostlie, PhD, Precision Agriculture Specialists, North Dakota State University.
- Laila Puntel, PhD, Assistant Professor, Department of Agronomy, University of Nebraska-Lincoln.
- Laura Thompson, On-Farm Coordinator, Extension Educator, University of Nebraska-Lincoln.
- Guillermo Balboa, Research Assistant Professor, Department of Agronomy, University of Nebraska-Lincoln.

Project goals:

1. Develop a multistate database to allow upscaling of soybean quality predictions to regional levels and benchmark agronomic practices, soybean genetics, management, and environmental conditions that can lead to large-scale improvements in soybean quality.
2. Communicate the economic value of soybean quality mapping to farmers and agronomists through an online interactive simulation tool, technical publications, and social media.

Accomplishments for entire project

The team of all the collaborators from multiple states (Ohio, Indiana, South Dakota, Missouri, Iowa, Michigan, Illinois, North Dakota, Nebraska, Iowa, and Kansas), including John Fulton, Shaun Casteel, Peter Kovacs, Andre Borja Reis, Scott Nelson, Mark Seamon and Mani Sing, Randy Pearson, David Kramar and Michael Ostlie, and Guillermo Balboa, helped on collecting all field sites for 2022, 2023, and 2024 growing seasons.

All seeds were processed for seed quality traits, mainly protein and oil concentrations, from all fields were obtained and data share across all collaborators. Reports for each state were prepared every year to provide information on the soybean seed quality (mainly protein and oil) for each farmer field.

The process of seed and soil data collection was established as an initial characterization of farmer fields, exploring the within-field variation, developing clusters, and sending geo-locations for all the samples in each field across all states participating on this project.

Protocol for data collection and clustering

1) Select available images between "May" and "September" (growth time), from the last 3 years.
2) Build a database and apply Kmeans to find the best clustering.
3) Build a new database for each best cluster (e.g., 1, 2, 3).
4) Define the optimal number of samples based on geostatistical analysis through the parameters: "range" and "total area of the cluster". If the result is less than 2 samples per clustering, set 2 as the number of samples.

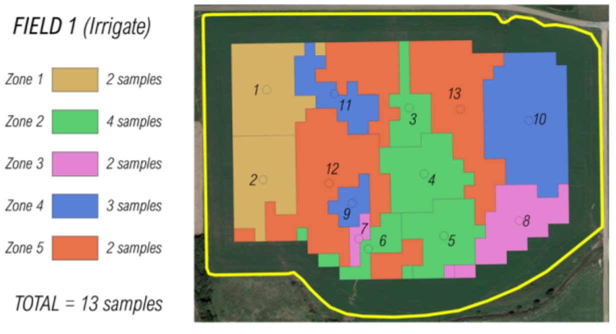
=====================================

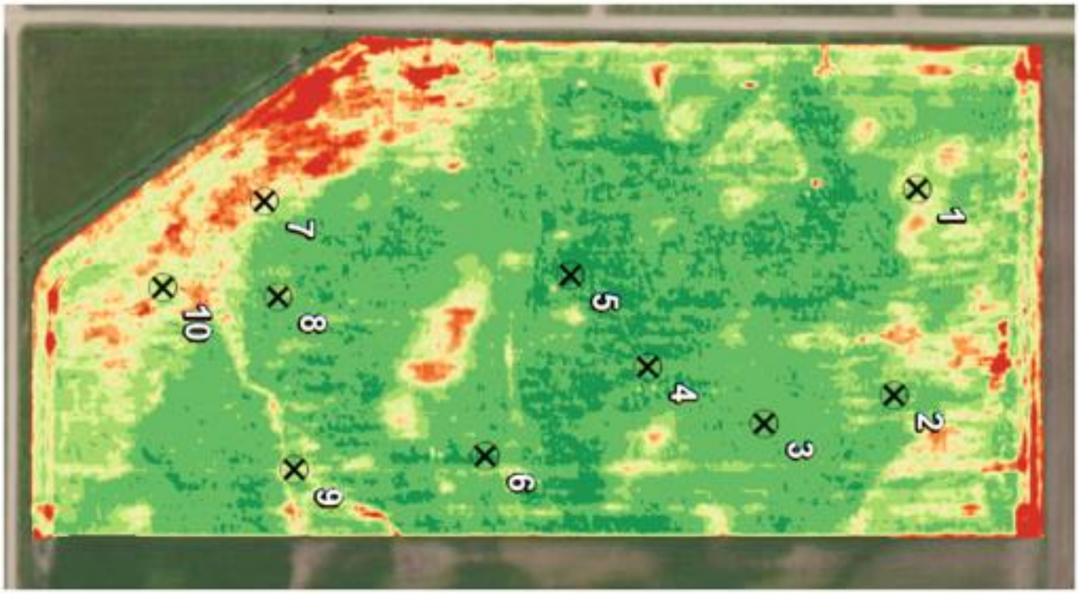
Field sampling protocol for mapping soybean seed quality

•Field selection and data collection (satellite imagery from 3 years)

•Clustering process, number and optimal management zones

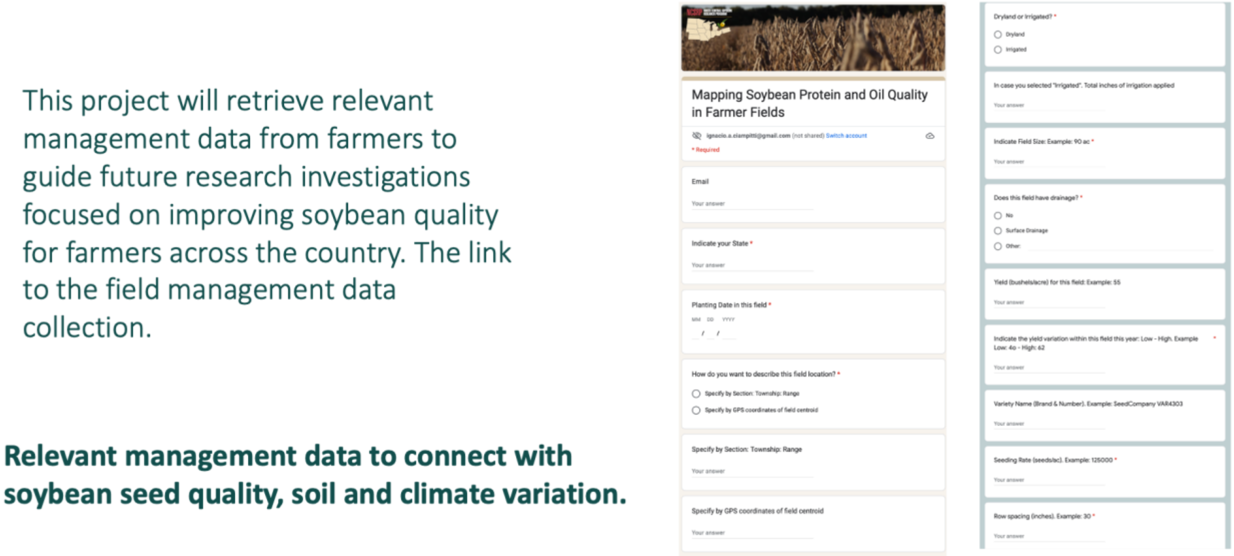
•Optimal number of soybean quality seed samples based on clusters





All field locations were received by early-to mid- summertime. The data was processed (integrating past yield, soil, satellite data). A clustering of field variation was developed (as shown above) and all field sampling was guided to collected variability of soybean seed quality.

All seed information was accompanied by the recording of relevant crop farming attributes via the collection of data utilizing a survey.



This project will retrieve relevant management data from farmers to guide future research investigations focused on improving soybean quality for farmers across the country. The link to the field management data collection.

Relevant management data to connect with soybean seed quality, soil and climate variation.

Mapping Soybean Protein and Oil Quality in Farmer Fields

ignacio.a.ciampitti@gmail.com (not shared) Switch account

* Required

Email

Your answer

Indicate your State *

Your answer

Planting Date in this field *

MM DD YYYY

__ / __ / ____

How do you want to describe this field location? *

○ Specify by Section: Township: Range

○ Specify by GPS coordinates of field centroid

Specify by Section: Township: Range

Your answer

Specify by GPS coordinates of field centroid

Your answer

Dryland or Irrigated? *

○ Dryland

○ Irrigated

In case you selected "Irrigated". Total inches of irrigation applied

Your answer

Indicate Field Size: Example: 90 ac *

Your answer

Does this field have drainage? *

○ No

○ Surface Drainage

○ Other:

Yield (bushels/acre) for this field: Example: 55

Your answer

Indicate the yield variation within this field this year: Low - High. Example Low: 4o - High: 62 *

Your answer

Variety Name (Brand & Number). Example: SeedCompany VAR4303

Your answer

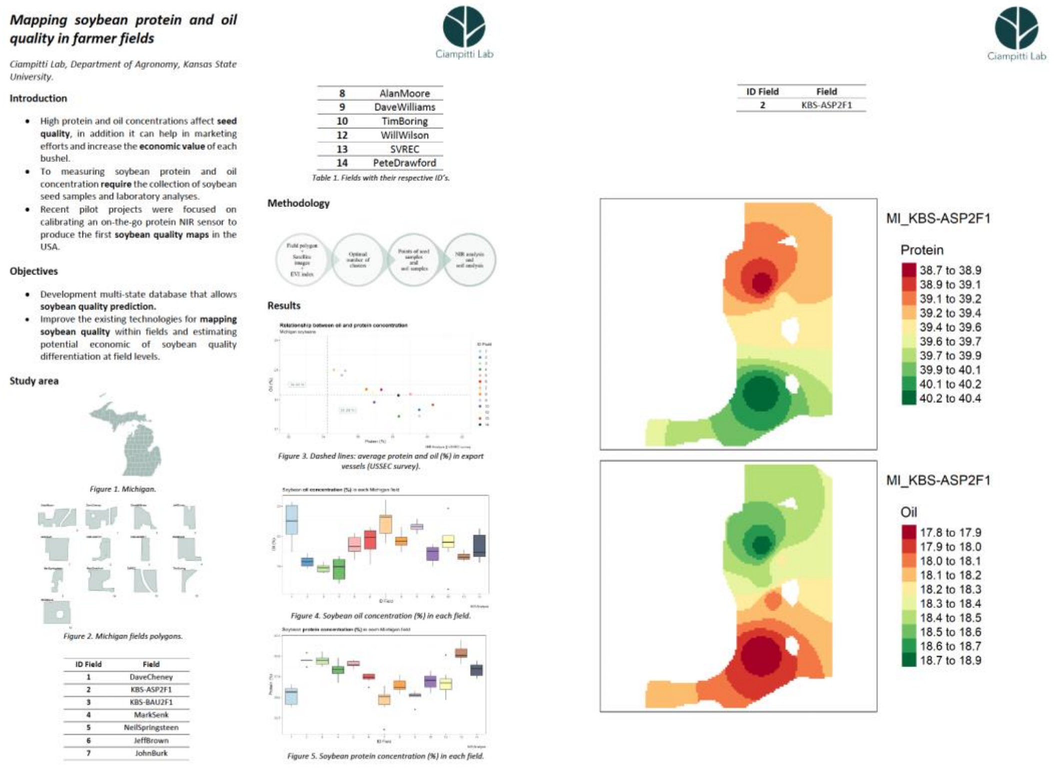
Seeding Rate (seeds/ac). Example: 125000 *

Your answer

Row spacing (inches). Example: 30 *

Your answer

Example of a report from a field at the end of the growing season:



Mapping soybean protein and oil quality in farmer fields

Ciampitti Lab, Department of Agronomy, Kansas State University.

Introduction

- High protein and oil concentrations affect **seed quality**, in addition it can help in marketing efforts and increase the **economic value** of each bushel.
- To measuring soybean protein and oil concentration **require** the collection of soybean seed samples and laboratory analyses.
- Recent pilot projects were focused on calibrating an on-the-go protein NIR sensor to produce the first **soybean quality maps** in the USA.

Objectives

- Development multi-state database that allows **soybean quality prediction.**
- Improve the existing technologies for **mapping soybean quality** within fields and estimating potential economic of soybean quality differentiation at field levels.

Study area

Figure 1. Michigan.

Figure 2. Michigan fields polygons.

ID Field	Field
1	DaveCheney
2	KBS-ASP2F1
3	KBS-BAU2F1
4	MarkSenk
5	NeilSpringsteen
6	JeffBrown
7	JohnBurk
8	AlanMoore
9	DaveWilliams
10	TimBoring
12	WillWilson
13	SVREC
14	PeteDrawford

Table 1. Fields with their respective ID's.

Ciampitti Lab

Methodology

Field polygon + Satellite images + EVI index

Optimal number of clusters

Points of seed samples and soil samples

NIR analysis and soil analysis

Results

Relationship between oil and protein concentration

Michigan soybeans

Figure 3. Dashed lines: average protein and oil (%) in export vessels (USSEC survey).

Soybean oil concentration (%) in each Michigan field

Figure 4. Soybean oil concentration (%) in each field.

Soybean protein concentration (%) in each Michigan field

Figure 5. Soybean protein concentration (%) in each field.

Ciampitti Lab

ID Field	Field
2	KBS-ASP2F1

MI_KBS-ASP2F1

Protein

38.7 to 38.9
38.9 to 39.1
39.1 to 39.2
39.2 to 39.4
39.4 to 39.6
39.6 to 39.7
39.7 to 39.9
39.9 to 40.1
40.1 to 40.2
40.2 to 40.4

MI_KBS-ASP2F1

Oil

17.8 to 17.9
17.9 to 18.0
18.0 to 18.1
18.1 to 18.2
18.2 to 18.3
18.3 to 18.4
18.4 to 18.5
18.5 to 18.6
18.6 to 18.7
18.7 to 18.9

For the last three growing seasons, 2022-2023-2024, a total of 394 fields with complete data on soybean seed quality and relevant crop management has been collected and compiled across the US soybean producing region. The states in the southern part of the US (Louisiana, Mississippi, and Alabama) were collected via a grant provided by the United Soybean Board (USB). The rest of the states are all the ones included in the current project funded by NCSRP.

Most recently, a manuscript was prepared to summarize all the information collected on this project. As mentioned before, we followed a standardized protocol across all farms for collecting representative seed and soil samples in-situ for analysis. In addition, we retrieved relevant crop management and yield data via surveys and linked all datasets with seasonal weather variables to develop a large on-farm database. The main objectives of the paper were to i) assess the importance of environmental variables in predicting seed oil and protein concentration and reported yields, ii) identify regions related to yield and seed quality, and iii) explore key predictors linked to these variables across regions to further understand seed oil and protein concentration differences across defined geographical regions.

A summary of the main soil and management variables (including crop phenology) across all states (North Dakota, South Dakota, Michigan, Ohio, Iowa, Nebraska, Kansas, Missouri, Illinois, Indiana, Louisiana, Mississippi, and Alabama) considering the median and the overall variation for each variable is introduced in a Table below.

An example of the first two growing seasons, n = 235 fields, is presented below. The distribution of values for yield, seed oil and protein concentrations were similar in both years. Overall, yield resulted in a median of 4.04 Mg ha-1 ranging between 1.88 and 5.38 Mg ha-1 as defined by the 95th percentile (P2.5 and P97.5). All in dry basis, oil concentration, reported on a dry basis, ranged between 17.9 and 22.9% with a median of 20.8%. Protein concentration had a median of 38.6% and ranged between 35.5 and 41.8%.

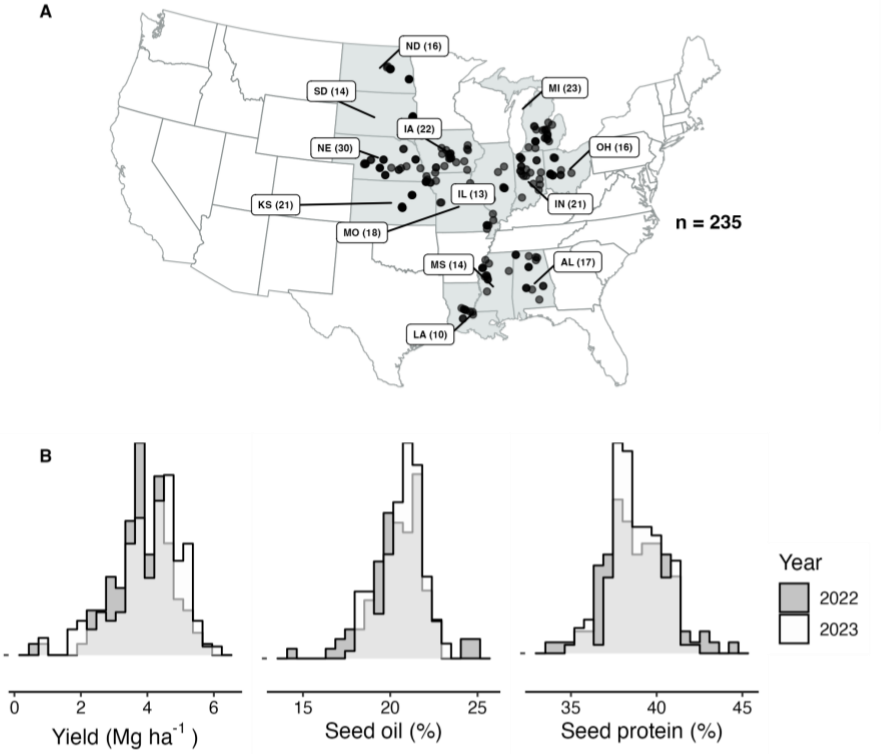


Figure 1. Location of the surveyed farmer fields (circles) in the US during 2022 and 2023 growing seasons, showing the total per state (A). Histogram displaying yield (Mg ha-1), seed oil and protein concentrations (%) in both 2022 and 2023 growing seasons (B).

Table 1. Summary of season characteristics, soil, and management variables for each state showing the median and range of observations for each variable.

	State													
	AL	IA	IL	IN	KS	LA	MI	MO	MS	ND	NE	OH	SD	Factor
vDays	38, [35, 60]	46, [20, 63]	48, [34, 59]	45, [36, 58]	42, [36, 56]	34, [29, 36]	47, [13, 71]	52, [29, 64]	38, [32, 46]	32, [26, 41]	47, [38, 60]	48, [12, 63]	37, [31, 41]	**Season**
rDays	33, [25, 37]	29, [19, 37]	32, [26, 36]	27, [24, 31]	24, [22, 37]	33, [31, 41]	26, [21, 29]	34, [25, 38]	38, [27, 42]	22, [18, 27]	28, [24, 37]	27, [21, 31]	26, [21, 32]	**Season**
sDays	41, [35, 50]	39, [34, 44]	43, [40, 45]	41, [36, 44]	34, [29, 44]	47, [30, 59]	39, [30, 43]	43, [34, 50]	44, [31, 50]	35, [28, 38]	36, [26, 43]	42, [40, 44]	36, [31, 40]	**Season**
MG	5, [4, 7]	3, [2, 4]	4, [3, 4]	3, [2, 4]	4, [4, 5]	5, [5, 6]	2, [2, 3]	4, [2, 5]	5, [5, 6]	0, [0, 1]	3, [2, 4]	3, [3, 4]	2, [1, 2]	**Season**
Clay (%)	18, [8, 56]	28, [17, 33]	18, [13, 38]	18, [3, 28]	29, [2, 42]	18, [8, 39]	11, [6, 27]	24, [12, 41]	46, [10, 78]	20, [12, 35]	27, [6, 35]	26, [16, 40]	26, [15, 28]	**Soil**
OM (%)	2.5, [0.8, 5.0]	3.6, [2.4, 5.2]	2.4, [1.9, 5.3]	2.9, [1.8, 4.9]	2.5, [0.8, 3.1]	1.5, [1.0, 2.6]	2.7, [1.6, 4.4]	1.8, [1.0, 3.7]	2.8, [2.1, 3.7]	3.5, [2.7, 5.0]	2.4, [0.9, 3.6]	3.2, [2.3, 4.1]	2.7, [2.0, 3.6]	**Soil**
pH	5.8, [5.1, 6.7]	6.3, [5.2, 7]	6, [5.6, 6.5]	6.2, [5.1, 6.9]	6.7, [4.9, 8.2]	6.6, [4.9, 8]	6.4, [5.6, 7.1]	6.2, [4.4, 7]	6.8, [5.8, 7.4]	7.1, [6.5, 8.1]	6.8, [6, 8.2]	6.7, [5.6, 7.1]	7.4, [5.8, 7.7]	**Soil**
P (ppm)[1,2]	69, [11, 350]	38, [9, 97]	30, [7, 176]	37, [14, 211]	27, [15, 79]	22, [8, 54]	40, [13, 189]	22, [11, 60]	40, [19, 82]	56, [8, 96]	19, [12, 53]	31, [13, 68]	12, [5, 41]	**Soil**
vPrec (mm)	137, [37, 244]	156, [68, 289]	113, [70, 195]	101, [28, 244]	146, [28, 326]	88, [61, 135]	109, [62, 227]	178, [73, 287]	189, [75, 304]	95, [56, 135]	132, [80, 277]	177, [54, 298]	85, [80, 115]	**Weather**
vT_{min} (°C)	19.4, [12.9, 22.6]	15.4, [12.3, 18.8]	16.6, [14.2, 21.0]	15.3, [13.3, 17.8]	20.4, [15.1, 21.4]	20.2, [16.3, 23.4]	13.3, [11.2, 18.6]	16.9, [12.9, 20.1]	19.6, [15.3, 22.5]	13.9, [11.0, 15.6]	14.4, [10.5, 16.9]	16.4, [14.5, 19.0]	16.7, [14.1, 18.3]	**Weather**
vT_{max} (°C)	31.5, [25.3, 35.0]	27.9, [26.4, 31.1]	28.6, [27.1, 33.3]	28.5, [26.3, 30.6]	33.0, [27.7, 35.5]	32.1, [28.1, 34.1]	27.1, [25.2, 29.5]	27.9, [24.0, 32.5]	30.9, [28.5, 34.1]	27.4, [25.4, 28.1]	28.4, [27.0, 32.0]	28.8, [27.2, 30.6]	28.2, [25.6, 30.5]	**Weather**
vVPD (kPa)	1.0, [0.5, 1.5]	0.8, [0.6, 1.1]	0.7, [0.6, 0.9]	0.7, [0.5, 0.9]	1.4, [0.8, 1.9]	0.8, [0.5, 1.1]	0.6, [0.6, 0.8]	0.7, [0.4, 1.0]	0.7, [0.5, 1.6]	0.8, [0.7, 0.9]	1.2, [0.8, 1.6]	0.7, [0.6, 0.8]	1.0, [0.8, 1.2]	**Weather**
rPrec (mm)	134, [27, 266]	44, [3, 175]	225, [58, 286]	102, [41, 368]	36, [25, 106]	106, [45, 154]	72, [30, 124]	94, [16, 339]	102, [75, 203]	33, [10, 53]	76, [22, 170]	80, [28, 155]	75, [40, 131]	**Weather**
rT_{min} (°C)	22.2, [18.0, 23.1]	17.9, [16.7, 21.6]	19.6, [17.0, 22.9]	17.6, [14.8, 20.7]	19.1, [17.5, 21.0]	23.0, [21.3, 23.7]	16.1, [14.3, 21.5]	20.0, [14.6, 21.4]	22.4, [20.8, 23.5]	13.8, [12.0, 17.3]	17.6, [15.1, 19.7]	19.0, [15.7, 20.5]	18.6, [17.8, 20.0]	**Weather**
rT_{max} (°C)	32.9, [30.5, 35.6]	30.4, [29.0, 34.8]	30.6, [29.2, 34.7]	28.9, [26.8, 31.6]	34.1, [31.0, 35.4]	33.8, [32.2, 35.3]	28.6, [27.0, 34.6]	31.6, [29.5, 35.2]	34.3, [32.5, 34.9]	27.9, [26.7, 29.0]	32.2, [29.8, 34.2]	29.1, [26.7, 31.4]	31.0, [30.3, 31.5]	**Weather**
rVPD (kPa)	1.0, [0.8, 1.4]	1.2, [0.8, 1.4]	0.9, [0.6, 1.1]	0.8, [0.6, 0.9]	1.7, [1.1, 1.8]	1.1, [0.9, 1.3]	0.8, [0.6, 1.2]	1.0, [0.7, 1.4]	1.4, [1.1, 1.5]	0.9, [0.8, 1.1]	1.5, [1.1, 1.8]	0.7, [0.6, 0.8]	1.2, [1.1, 1.6]	**Weather**
sPrec (mm)	171, [59, 342]	104, [52, 219]	90, [74, 183]	121, [50, 168]	42, [33, 153]	306, [112, 390]	100, [40, 192]	116, [38, 190]	192, [56, 323]	41, [9, 68]	48, [15, 127]	110, [71, 169]	124, [82, 172]	**Weather**
sT_{min} (°C)	20.6, [16.3, 22.7]	15.2, [14.4, 21.8]	17.4, [15.2, 20.7]	15.2, [12.8, 19.2]	15.5, [13.0, 18.5]	23.0, [22.0, 23.8]	14.2, [10.8, 23.0]	17.4, [13.8, 22.3]	22.0, [19.0, 23.1]	12.6, [10.3, 14.7]	15.1, [11.7, 17.3]	15.9, [11.2, 23.5]	15.6, [14.7, 16.5]	**Weather**
sT_{max} (°C)	30.5, [28.8, 33.3]	28.2, [27.3, 33.1]	28.5, [27.0, 32.3]	27.1, [25.6, 29.6]	31.0, [28.6, 34.7]	33.1, [31.7, 34.1]	26.6, [23.5, 34.5]	29.7, [27.7, 33.8]	32.3, [31.4, 34.4]	27.0, [26.3, 27.7]	31.4, [29.7, 32.2]	27.0, [22.9, 34.5]	28.2, [27.6, 28.6]	**Weather**
sVPD (kPa)	0.9, [0.8, 1.0]	1.0, [0.7, 1.5]	0.8, [0.6, 0.9]	0.8, [0.6, 0.9]	1.5, [1.2, 1.8]	0.8, [0.6, 1.1]	0.6, [0.5, 1.1]	1.0, [0.7, 1.2]	1.1, [0.8, 1.4]	0.9, [0.8, 1.0]	1.5, [1.1, 1.7]	0.7, [0.6, 1.0]	1.0, [0.9, 1.2]	**Weather**
sDaysT_{max}>30	28, [15, 46]	13, [6, 40]	13, [4, 35]	6, [2, 26]	24, [14, 34]	46, [29, 55]	4, [2, 34]	20, [10, 41]	39, [25, 46]	9, [4, 12]	24, [15, 30]	6, [2, 42]	10, [6, 14]	**Weather**
sDaysT_{min}<15	1, [0, 17]	21, [0, 26]	11, [1, 23]	23, [3, 31]	14, [4, 26]	0, [0, 2]	25, [0, 38]	12, [0, 27]	0, [0, 6]	28, [24, 36]	20, [10, 30]	18, [0, 36]	13, [9, 20]	**Weather**

Number of days between VE-R1, R1-R5, and R5-R7 (vDays, rDays, and sDays, respectively), maturity group (MG), Precipitation (Prec), minimum and maximum temperature (T_{min} and T_{max}, respectively), vapor pressure deficit (VPD), number of days during seed fill with maximum temperatures > 30°C, (sDaysT_{max}>30), number of days during seed fill with minimum temperatures < 15°C, (sDaysT_{min}<15).

[1] Weather was summarized in vegetative (v), reproductive pre-seed filling (r), and reproductive seed filling (s).

[2] Growth stages were simulated with DSSAT.

All the information collected was clustered in three main producing regions, with two of them pertaining to the North Central.

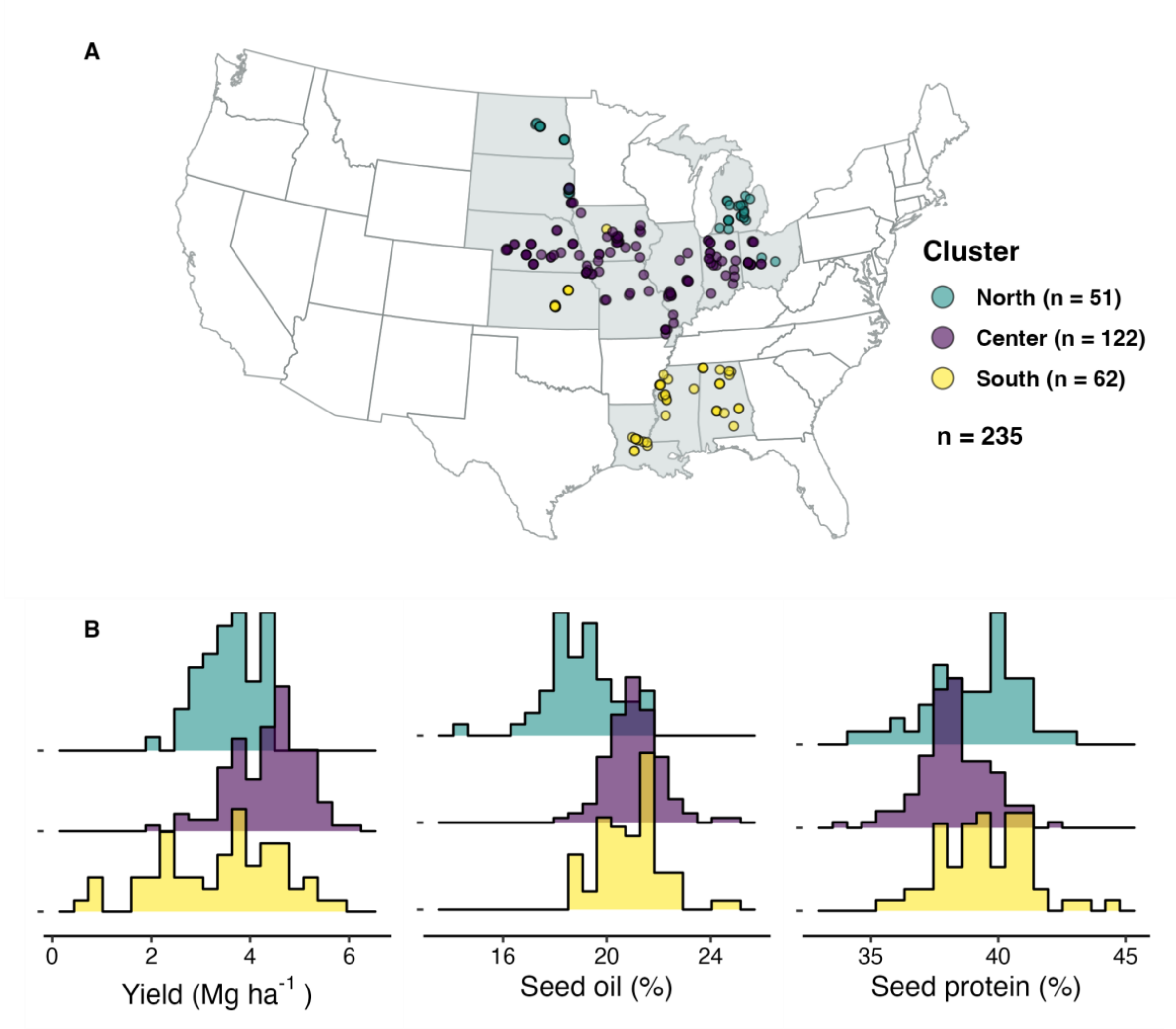
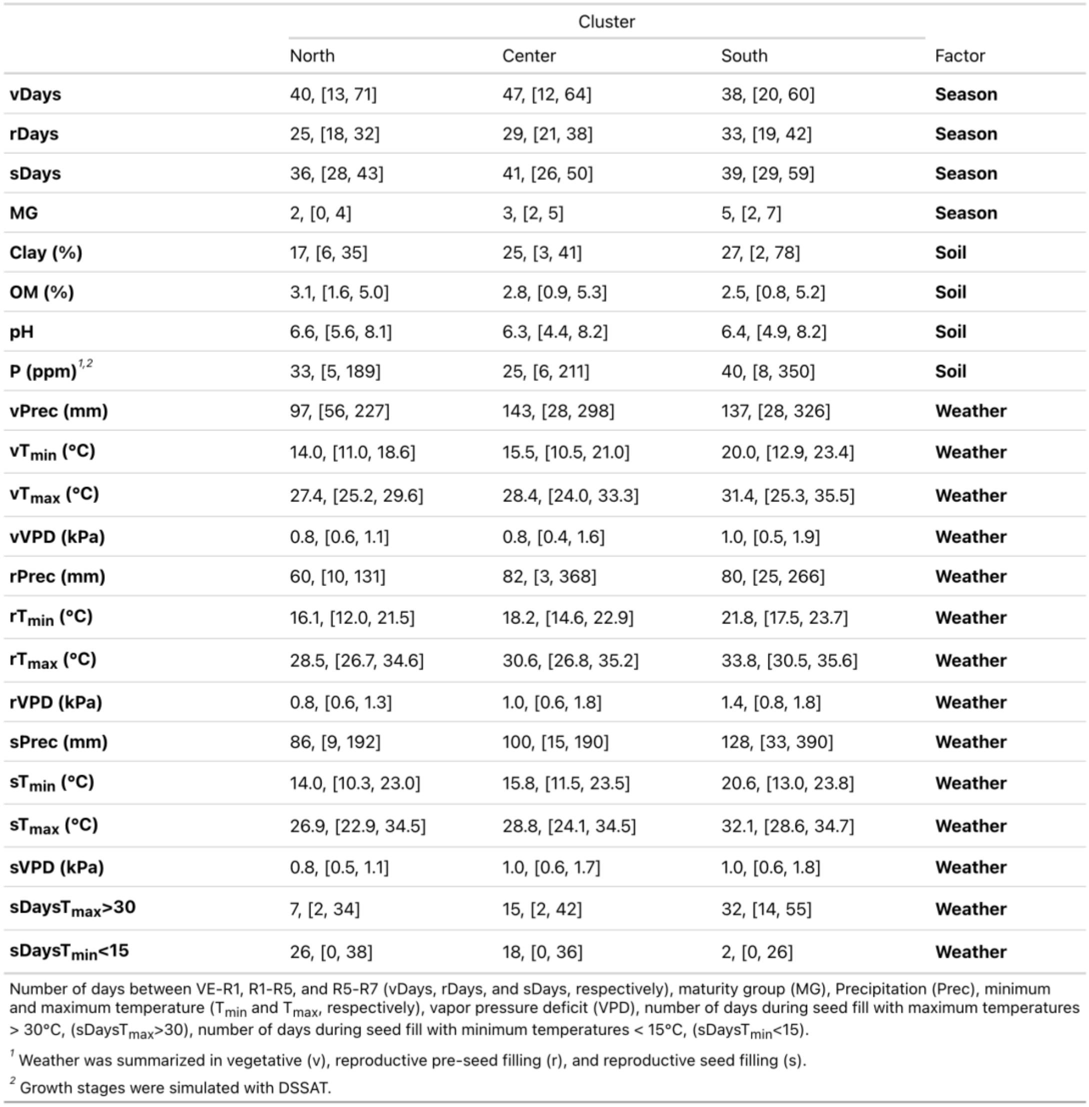


Figure 2. Location of the surveyed farmer fields (circles) in colored according to the cluster classification, showing the total per state (A). Histogram displaying yield (Mg ha-1), protein, and oil concentration (%) for each cluster (B).

Yield, oil and protein concentrations were significantly correlated with key variables influencing prediction accuracy (Fig. 1), with some of these variables exhibiting a strong link to regional differences (Fig. 3 A-I). Maximum temperatures during the early vegetative stages significantly limited yield (Fig. 4A), which was related to a decrease in the total season length, particularly in the southern region. The maturity group showed a regional pattern; with a positive yield response to longer crop maturity (and longer seasons) in the northern region, whereas in the southern region, the tendency was for higher yields with shorter crop maturity (MG 4) (Fig. 3B). No discernible correlation was identified between yield and VPD during seed fill (Fig. 3C). Regarding oil, a positive correlation was observed with increasing minimum temperatures during the pre-seed-fill and seed-fill periods and a northern-to-southern trend. Higher minimum temperatures in pre-seed-fill were linked to an extended R1-R5 period, which was also positively associated with oil concentration (Fig. 3D, E, F). Lastly, a positive association, albeit weak, was observed between protein concentration and precipitation in late reproductive stages (Fig. 3G, H, I).

A summary of growing season, soil and weather data for each cluster is presented below.

Table 2: Summary of season characteristics, weather, and soil variables for each cluster showing the median and range of observations for each variable.



	Cluster			
	North	Center	South	Factor
vDays	40, [13, 71]	47, [12, 64]	38, [20, 60]	**Season**
rDays	25, [18, 32]	29, [21, 38]	33, [19, 42]	**Season**
sDays	36, [28, 43]	41, [26, 50]	39, [29, 59]	**Season**
MG	2, [0, 4]	3, [2, 5]	5, [2, 7]	**Season**
Clay (%)	17, [6, 35]	25, [3, 41]	27, [2, 78]	**Soil**
OM (%)	3.1, [1.6, 5.0]	2.8, [0.9, 5.3]	2.5, [0.8, 5.2]	**Soil**
pH	6.6, [5.6, 8.1]	6.3, [4.4, 8.2]	6.4, [4.9, 8.2]	**Soil**
P (ppm)[1,2]	33, [5, 189]	25, [6, 211]	40, [8, 350]	**Soil**
vPrec (mm)	97, [56, 227]	143, [28, 298]	137, [28, 326]	**Weather**
vT_{min} (°C)	14.0, [11.0, 18.6]	15.5, [10.5, 21.0]	20.0, [12.9, 23.4]	**Weather**
vT_{max} (°C)	27.4, [25.2, 29.6]	28.4, [24.0, 33.3]	31.4, [25.3, 35.5]	**Weather**
vVPD (kPa)	0.8, [0.6, 1.1]	0.8, [0.4, 1.6]	1.0, [0.5, 1.9]	**Weather**
rPrec (mm)	60, [10, 131]	82, [3, 368]	80, [25, 266]	**Weather**
rT_{min} (°C)	16.1, [12.0, 21.5]	18.2, [14.6, 22.9]	21.8, [17.5, 23.7]	**Weather**
rT_{max} (°C)	28.5, [26.7, 34.6]	30.6, [26.8, 35.2]	33.8, [30.5, 35.6]	**Weather**
rVPD (kPa)	0.8, [0.6, 1.3]	1.0, [0.6, 1.8]	1.4, [0.8, 1.8]	**Weather**
sPrec (mm)	86, [9, 192]	100, [15, 190]	128, [33, 390]	**Weather**
sT_{min} (°C)	14.0, [10.3, 23.0]	15.8, [11.5, 23.5]	20.6, [13.0, 23.8]	**Weather**
sT_{max} (°C)	26.9, [22.9, 34.5]	28.8, [24.1, 34.5]	32.1, [28.6, 34.7]	**Weather**
sVPD (kPa)	0.8, [0.5, 1.1]	1.0, [0.6, 1.7]	1.0, [0.6, 1.8]	**Weather**
sDaysT_{max}>30	7, [2, 34]	15, [2, 42]	32, [14, 55]	**Weather**
sDaysT_{min}<15	26, [0, 38]	18, [0, 36]	2, [0, 26]	**Weather**

Number of days between VE-R1, R1-R5, and R5-R7 (vDays, rDays, and sDays, respectively), maturity group (MG), Precipitation (Prec), minimum and maximum temperature (T_{min} and T_{max}, respectively), vapor pressure deficit (VPD), number of days during seed fill with maximum temperatures > 30°C, (sDaysT_{max}>30), number of days during seed fill with minimum temperatures < 15°C, (sDaysT_{min}<15).

[1] Weather was summarized in vegetative (v), reproductive pre-seed filling (r), and reproductive seed filling (s).

[2] Growth stages were simulated with DSSAT.

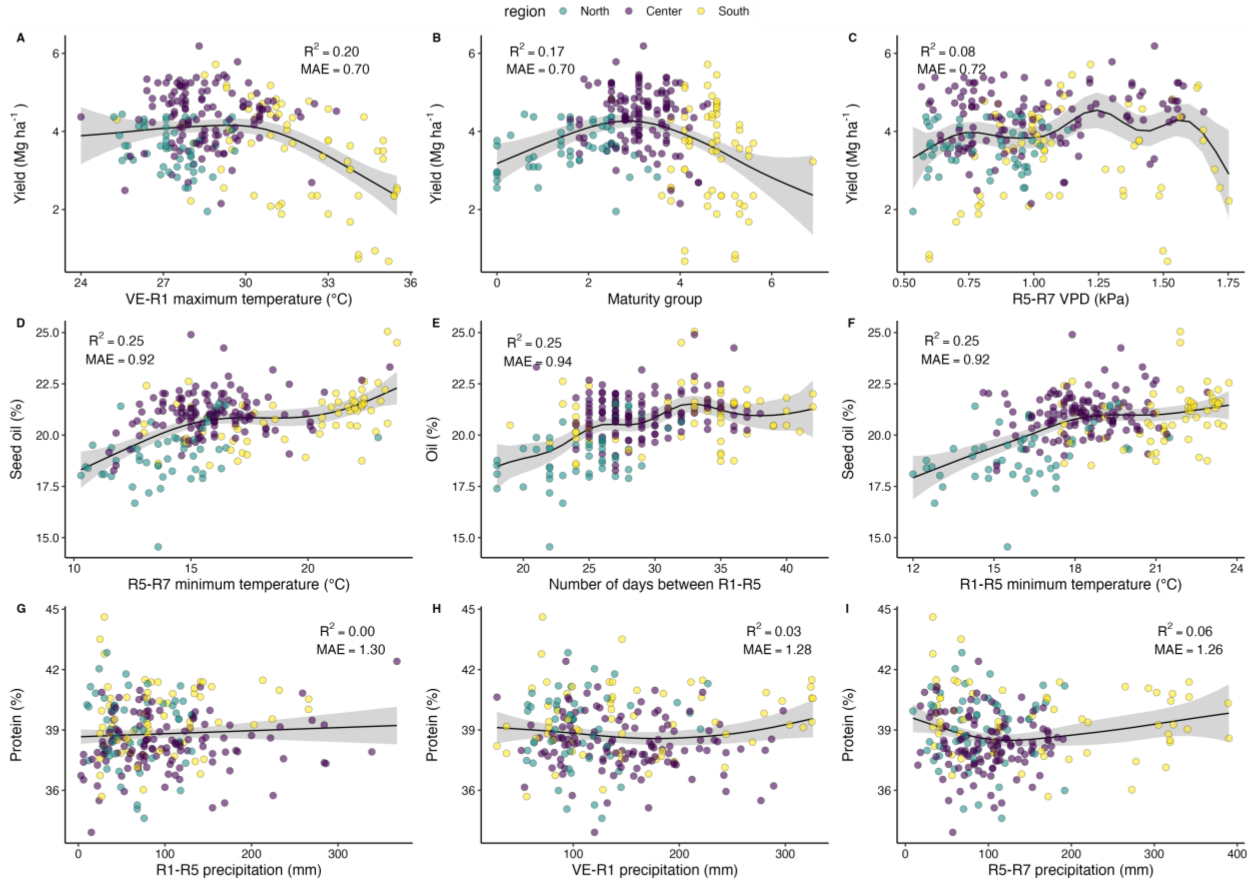


Figure 3. Soybean yield in relation to maximum temperature in VE-R1 (A), maturity group (B), and vapor pressure deficit (VPD, kPa) between R5-R7 (C). Additionally, soybean oil concentration in relation to minimum temperature between R5-R7 (D), number of days between R1-R5 (E), and minimum temperature between R1-R5 (F). Lastly, protein concentration in relation to precipitation (mm) between R1-R5 (G), VE-R1 (H), and R5-R7 (I). The colors represent three clustered regions: North (green), Center (purple), and South (yellow). Growth stages were simulated with DSSAT. The relationships were fitted using generalized additive models. The coefficient of determination (R2) and mean absolute error (MAE) are shown.

In summary, we used a robust protocol to collect representative seed and soil samples in 235 soybean farmer fields (processing only 2022-2023, and in next step now including 2024) across 13 US states. We collected yield and management data via survey, and combined growth models to summarize weather data during key crop phenological stages. The prediction of yield and oil concentration exhibited greater accuracy than that of protein concentration when seasonal variables related to weather, soil, and crop growth were considered. Yield, protein, and oil levels were within the ranges usually reported for soybean in the regions explored. However, higher protein levels in the north suggest a narrowing in the quality gap of soybeans between this region and the Corn Belt.

Summary of the project

- **Protein** ↑ in northern region (39.5%) compared to the main Corn Belt region (38.2%).

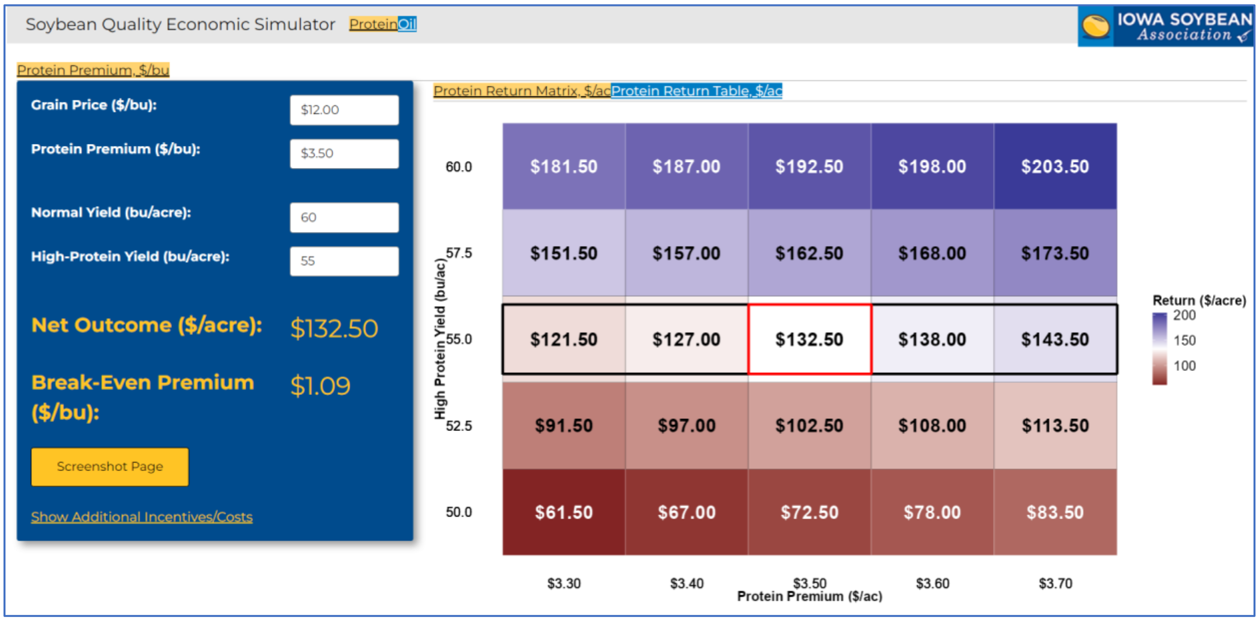
- Environmental descriptors predicted seed **yield** and **oil** with better accuracy than **protein** concentration

- Late planting in the south ↓ **yield** due to early season high T°C resulting in shorter growing season

- Seed **oil** in the north is more limited by low temperatures during seed filling

Developing a tool, the quality economic simulator

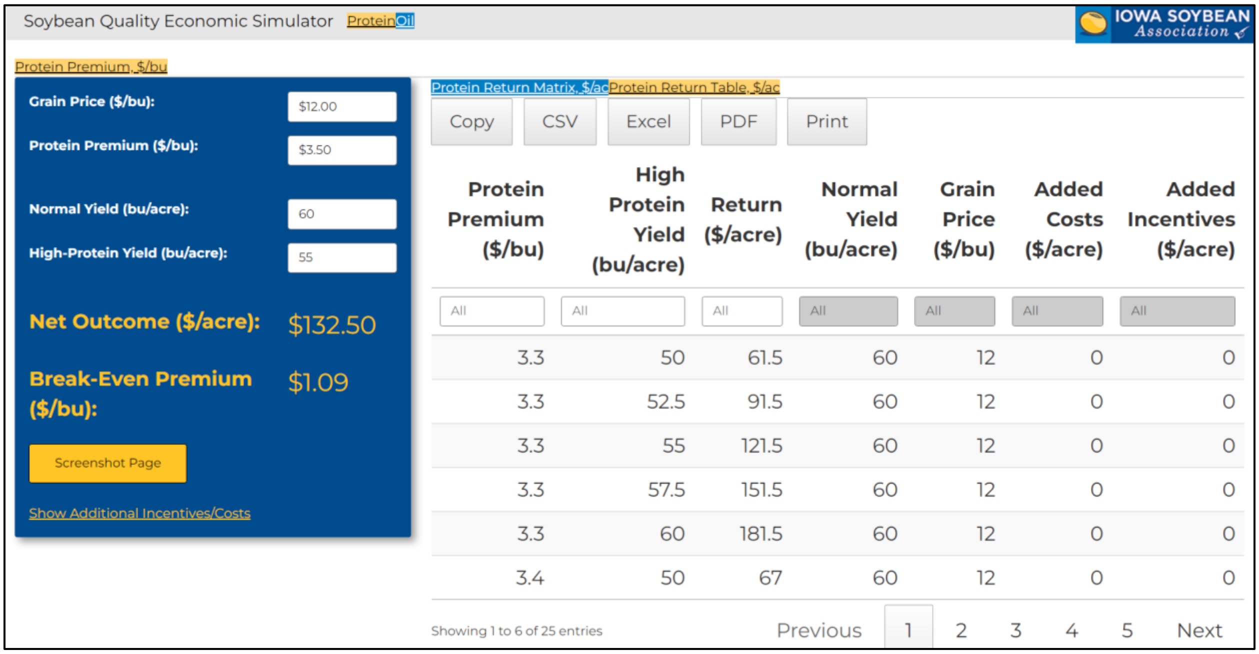
The soybean quality economic simulator has been updated and modified in two key areas. The first being that the oil quality portion of the tool has been built and is functioning well. The second update was to the existing user interface to make it more intuitive for users. Based on feedback from farmers the old version was difficult to understand what yield was used and how to add yield loss properly. We also added a break-even premium price so that farmers can quickly decide on if the premium they are receiving will have a positive ROI on their farm.



Grain Price ($/bu): $12.00
Protein Premium ($/bu): $3.50
Normal Yield (bu/acre): 60
High-Protein Yield (bu/acre): 55
Net Outcome ($/acre): $132.50
Break-Even Premium ($/bu): $1.09

High Protein Yield (bu/ac) \ Protein Premium ($/ac)	$3.30	$3.40	$3.50	$3.60	$3.70
60.0	$181.50	$187.00	$192.50	$198.00	$203.50
57.5	$151.50	$157.00	$162.50	$168.00	$173.50
55.0	$121.50	$127.00	$132.50	$138.00	$143.50
52.5	$91.50	$97.00	$102.50	$108.00	$113.50
50.0	$61.50	$67.00	$72.50	$78.00	$83.50

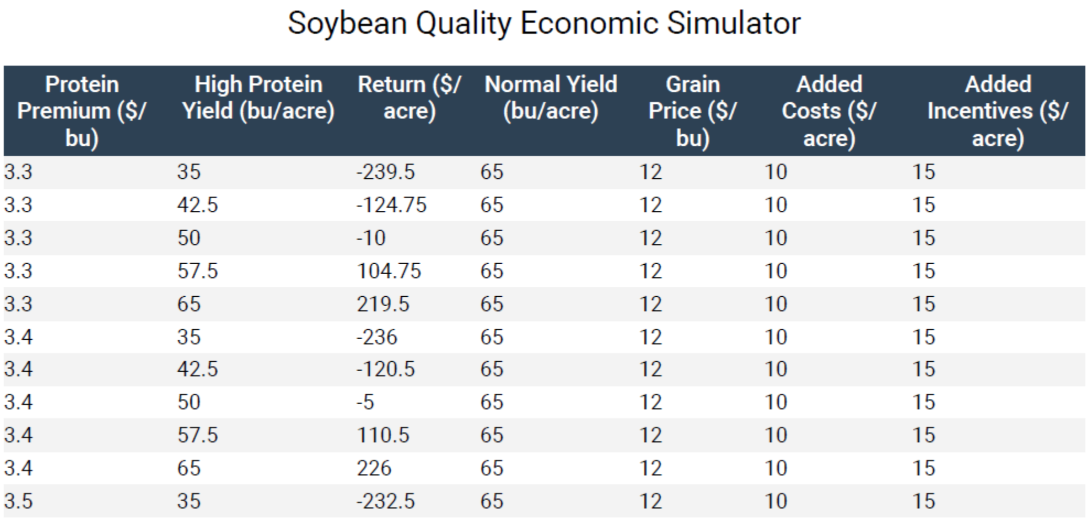
The last addition we made is that users can export their results in a pdf, csv, or XLSX document so they can save their work for future use.



Grain Price ($/bu): $12.00
Protein Premium ($/bu): $3.50
Normal Yield (bu/acre): 60
High-Protein Yield (bu/acre): 55
Net Outcome ($/acre): $132.50
Break-Even Premium ($/bu): $1.09

Protein Premium ($/bu)	High Protein Yield (bu/acre)	Return ($/acre)	Normal Yield (bu/acre)	Grain Price ($/bu)	Added Costs ($/acre)	Added Incentives ($/acre)
All	All	All	All	All	All	All
3.3	50	61.5	60	12	0	0
3.3	52.5	91.5	60	12	0	0
3.3	55	121.5	60	12	0	0
3.3	57.5	151.5	60	12	0	0
3.3	60	181.5	60	12	0	0
3.4	50	67	60	12	0	0

Showing 1 to 6 of 25 entries



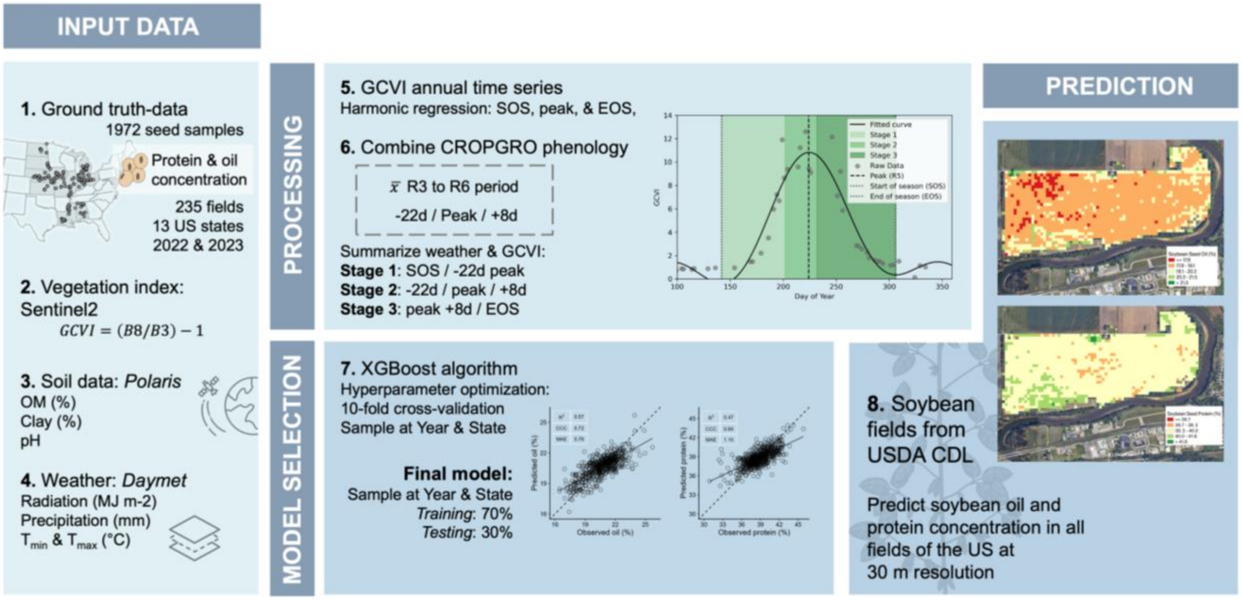
Soybean Quality Economic Simulator

Protein Premium ($/bu)	High Protein Yield (bu/acre)	Return ($/acre)	Normal Yield (bu/acre)	Grain Price ($/bu)	Added Costs ($/acre)	Added Incentives ($/acre)
3.3	35	-239.5	65	12	10	15
3.3	42.5	-124.75	65	12	10	15
3.3	50	-10	65	12	10	15
3.3	57.5	104.75	65	12	10	15
3.3	65	219.5	65	12	10	15
3.4	35	-236	65	12	10	15
3.4	42.5	-120.5	65	12	10	15
3.4	50	-5	65	12	10	15
3.4	57.5	110.5	65	12	10	15
3.4	65	226	65	12	10	15
3.5	35	-232.5	65	12	10	15

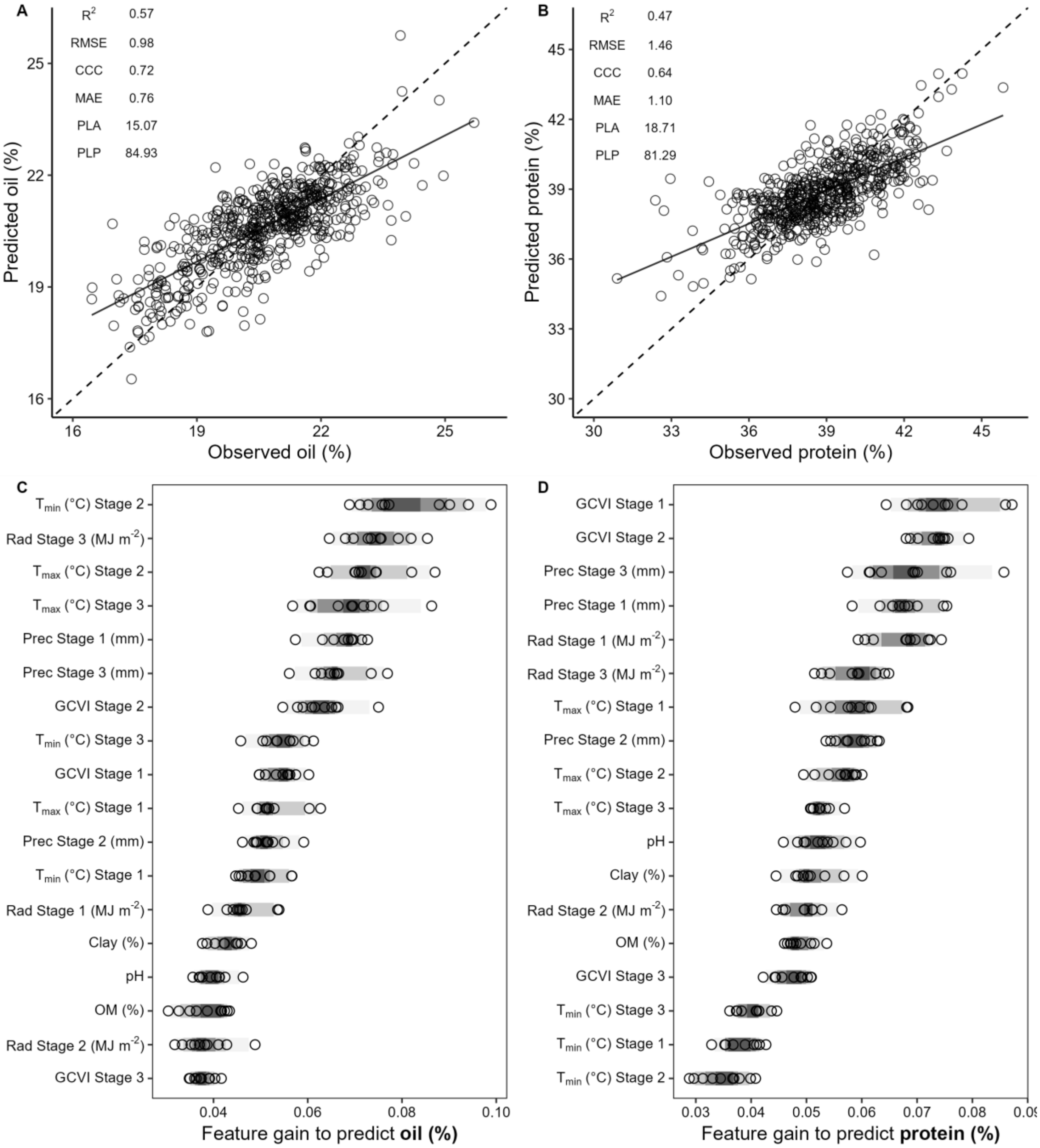
Users of the tool are also able to add any additional costs or savings that are a result from growing soybeans with an associated premium for oil and protein, such as increased seed and planting costs, or reduced financing options for inputs. This functionality can be accessed through selecting additional costs and incentives, and is capture as an aggregate sum in the downloaded pdf of added costs and incentives. Adding this utility allows farmers and other users to estimate total financial gain or loss of implementing a practice and can help in ensuring that all costs are estimated before implementing a new practice on their farm.

Next steps

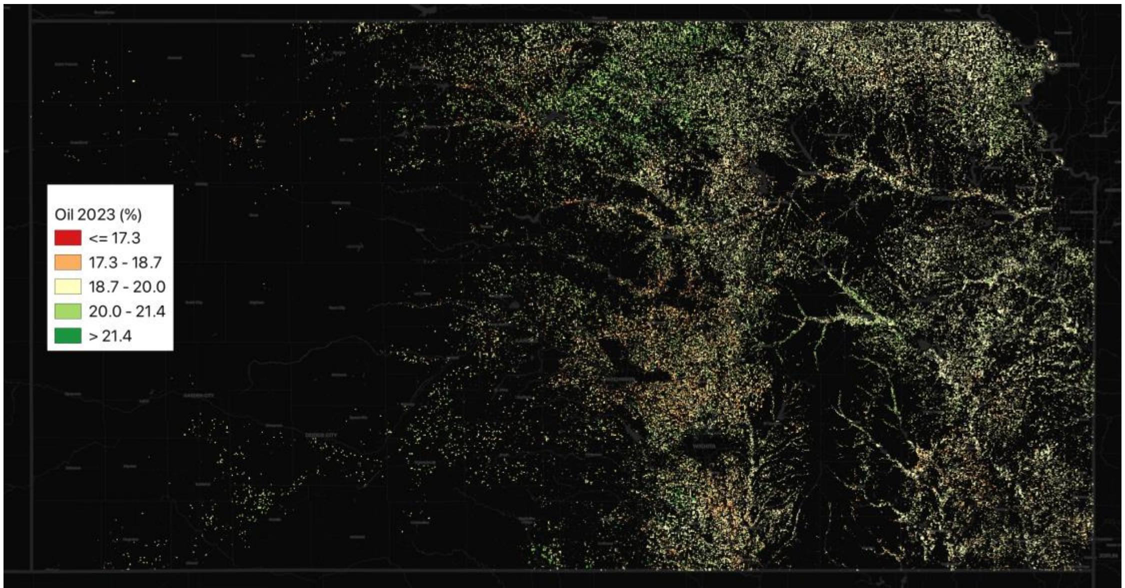
We are currently working on building prediction maps for quantifying soybean quality (mainly protein and oil) across the different soybean US producing regions. Using the collected database during 2022 and 2023, we trained predictive models using as inputs the green chlorophyl vegetation index (GCVI), soil data, and daily weather summaries. The main difference with the previous approach is the use of each of the individual points where seed samples were collected, therefore using 1972 seed samples over 235 fields.

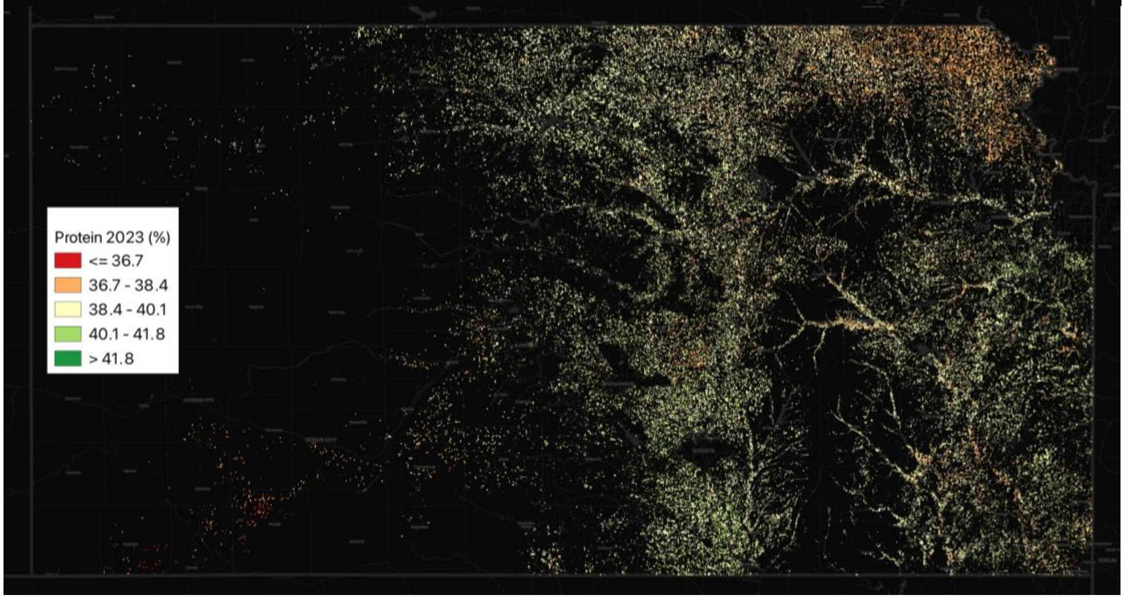


The final model predicting soybean oil concentration explained 57% of the variation (R^2) with a MAE of 0.76%, while the prediction of protein concentration explained 47% of the variation with a MAE of 1.10%, representing an improvement over previous effort. The three most significant variables for predicting oil were identified as the minimum temperature during the critical period surrounding the peak of GCVI (stage 2), radiation in the late season (stage 3), and maximum temperature during stage 2. In contrast, for protein, the GCVI during stage 1 and 2, and precipitation in stage 3 were determined to be the most influential predictors.



Following, we retrieved the soybean Crop Data Layer from USDA for the year 2023. In an example for Kansas, we can observe a tendence for lower protein and higher oil concentration in north-east Kansas in agreement with higher average precipitations and average yields.





In summary, we integrated a comprehensive ground truth dataset of on-farm soybean seed quality with satellite data on weather, soil, and GCVI to forecast soybean seed protein concentration across the United States during 2022 and 2023. This work illustrates the potential of collaborative research initiatives that integrate observed data across extensive regions. Additionally, this approach markedly enhances spatial resolution and prediction accuracy in comparison to previous endeavors employing interpolation of observations or the utilization of vegetation indices alone.

A few limitations of this study are linked to the use of harmonic regression to extract phenology features, which can result in an under- or overestimation of the beginning and end of the growing season and assume a symmetric relationship during the growing and senescence stages. Future endeavors should extend these approaches to yield predictions, enabling estimates of seed oil and protein production across diverse geographical regions.

Finally, we are working on validating the model predicting oil and protein for the 2024 season once the crop data layers (from USDA) are available.