Predictive Soil Health Economic Calculator (P-SHEC)

Technical Guide

Version 2.0

08/29/2025

American Farmland Trust

Authors

Chellie H. Maples, Meng Li, Ben Wiercinski, Aysha Tapp Ross, Bonnie McGill, & Michelle Perez

Acknowledgements

The authors express their gratitude to Florence Schwartz, Lori Metz, Candiss Williams, Karl Anderson, Laura Starr and Amanda Branham for their support conceptualizing the tool and Carson Sievert for developing the RShiny application. This work was supported by the United States Department of Agriculture's Natural Resources Conservation Service (Cooperative Agreement No. NR223A750010C003).

Contents

1. Introduction

1.1. Overview

The Predictive Soil Health Economic Calculator (P-SHEC) is designed to provide farmers and farm advisors with estimates for how soil organic matter (SOM), crop productivity, and farm profitability may change with the adoption of soil health practices in average and drought conditions. P-SHEC asks users to provide information about their crops, input prices, acreage, county, and state. Users then select the soil health practices they are interested in and specify how they would manage them over the next ten years. Practices available include planting cover crops, transitioning to no-till or reduced tillage, and nutrient management. The tool then provides both annual recurring cost changes and cumulative 10-year SOM, yield, drought resistance, and net income changes for adopting the selected practices in their county.

P-SHEC estimates the long-term yield changes of adopting soil health practices. Figure 1 shows the flow the tool uses, from selecting the soil health practices to the final combined results. Once users select soil health practices, the tool splits into two analytical pathways: the recurring, budget costs calculation and the long-term predictive models. The recurring costs analysis estimates the economic impact of changes from budget categories such as machinery, inputs, and learning costs associated with adopting soil health practices using state production budgets. The long-term predictive modeling uses COMET-Planner dataset (http://comet-planner.com/) to estimate a county-level change in soil organic carbon (SOC) due to specific practice changes, converts SOC to SOM, and then uses the county-specific yield-SOM and drought-SOM models to estimate how crop yield could change, at the estimated change in SOM, due to the practice change. The tool then combines the results from the recurring budget analysis with the long-term modeling to provide final results.

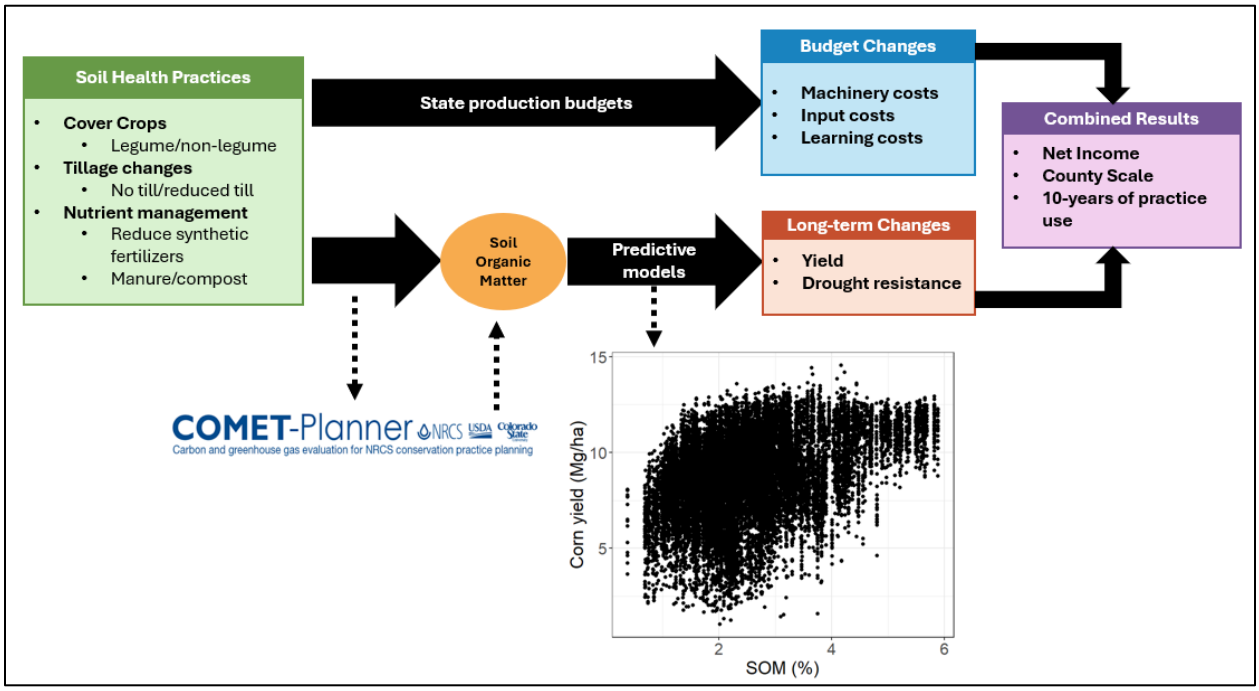


Figure 1. A flow chart displaying a general overview of how the P-SHEC predicts the on-farm economic impacts of soil health practice adoption.

P-SHEC's long-term modeling approach can be broken down into a data processing phase and a Yield and SOM modeling phase (Figure 2). In the data processing phase, the tool compiles and cleans multiple national datasets, including 24 years (2000-2023) of county-level crop yields from USDA-National Agricultural Statistics Service (NASS) (USDA NASS 2024b) as well as SOM and other soil data from gSSURGO (Soil Survey Staff 2024), climate data from PRISM (PRISM Climate Group 2025), nitrogen input data from Zhang, Cao, and Lu (2021), and drought related data from USDA-Risk Management Agency (RMA) Cause of Loss payment data (USDA RMA 2024a; 2024b) and University of Nebraska Lincoln Drought Severity and Coverage Index (Akyuz 2017)—all averaged across cropland at the county level. In the second phase, the cleaned datasets are used to train machine learning models for yield-SOM and drought-SOM relationships to predict the impacts of increased SOM on yield.

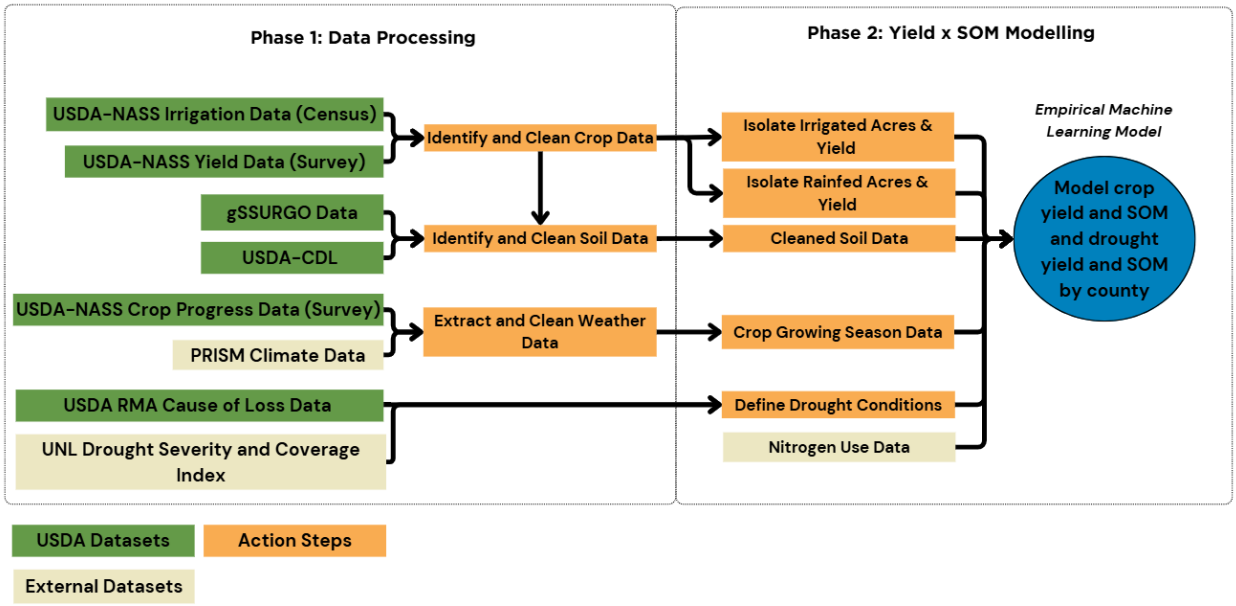


Figure 2. A flow chart of P-SHEC data sources and modeling steps.

Currently, the practices available in P-SHEC align with the crop management practices available in COMET-Planner (Table 1). In addition, P-SHEC supports ten different row crops, with crop availability varying by county, reflecting the availability of county-level yield data from USDA NASS (Table 1 and Appendix Figure 1). It provides results for one or more counties in 41 of the conterminous U.S. states (Figure 3). The detailed data sources used in the tool are provided in Appendix File 1.

Table 1. Practices and crops available in P-SHEC.

Soil Health Practices	Crops (availability varies by county):	
Planting cover crops (legume or non-legume)	Barley	Sorghum
Changing to no-till or reduced tillage	Corn	Soybean
Reducing synthetic fertilizer inputs	Cotton	Sugar Beet
Replacing part of N budget with manure or compost	Oats	Tobacco
	Peanut	Wheat

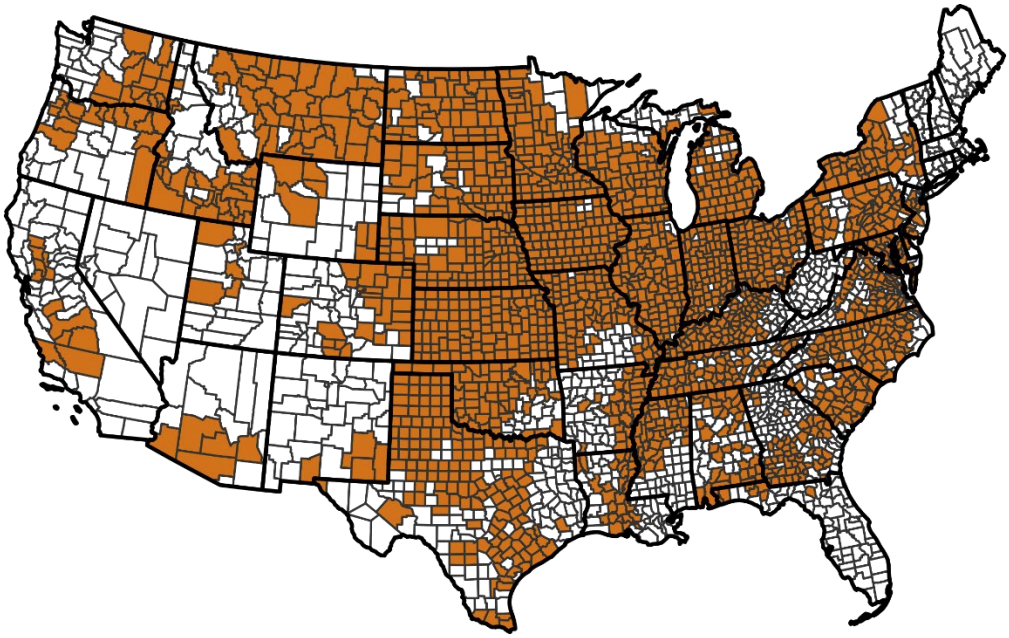


Figure 3. Geographic coverage of P-SHEC across the conterminous United States. *Shaded counties (n = 1,962) in 41 states contain at least one of the 10 supported crops—corn, soybeans, winter wheat, cotton, barley, oats, sorghum, peanuts, sugar beets, and tobacco. Crop availability varies by county reflecting the availability of county-level yield data from USDA NASS* (USDA NASS 2024b).

1.2. Assumptions and limitations

The data underlying the tool have several limitations, which affects the crops and counties included in the tool. P-SHEC estimates are not available in all states, territories, or counties in the U.S. (Figure 3). Crop selection is limited to those with available county-level yield data, broad geographic coverage encompassing multiple counties with a range of SOM levels, and inclusion in COMET-Planner's soil carbon estimation framework. These criteria ensure the predictive models are built on robust data.

P-SHEC estimates assume that the selected soil health practice will be implemented continuously for ten years. This is timeframe was chosen to align with COMET-Planner. Thus, COMET-Planner estimates, although provided as annual rates, are the result of 10 years of modeling divided into ten annual changes. **Even though a producer may want to know how a soil health practice could affect their crop yield yearly, American Farmland Trust (AFT) decided that the most scientifically appropriate use of the underlying data is to provide a long-term (10 year) prediction rather than an annual projection.**

The tool does not quantify other ecological or economic benefits of adopting soil health practices outside of changes in SOM. The relationship between SOM and yield is where most adequate data are available, but research shows that soil health practices have many other benefits, which could directly or indirectly support crop yields. These benefits include weed management, yield resilience to wet weather extremes, beneficial biodiversity, reduction in soil nitrate losses, and more (Basche et al. 2016; Basche and DeLonge 2019; Gentry et al. 2025; Isbell et al. 2017; O'Brien et al. 2022; Obour et al. 2021). Agroecosystems, defined by U.S. Environmental Protection Agency (EPA) as a dynamic association of

crops, pastures, livestock, other flora and fauna, soils, water, and the atmosphere, can also support many important ecosystem services beyond producing food and fiber, potentially including water infiltration and groundwater recharge, wildlife habitat, climate change mitigation, and maintaining cultural heritage (Daugstad et al. 2006; Robertson et al. 2014).

One might ask whether changes in SOM are an accurate predictor of changes in crop yield. SOM and SOC are convenient and common measures for characterizing soil health. In an agroecosystem, SOM and SOC are the product of plant community productivity and diversity as well as soil microbial and macroinvertebrate community structure and function, which are themselves a product of the underlying mineralogy, climate, land use history, and soil management (Cotrufo and Lavallee 2022). SOM, including plant root exudates, is itself a source of nutrients and structure to the soil food web, but the amount of SOM and how it changes over time is an indicator of the delicate balance between processes building up SOM (e.g., organic matter inputs, soil aggregate formation and stabilization, microbial biomass, microbial nitrogen availability) and processes consuming SOM (mainly mineralization of soil organic matter via microbial respiration and soil erosion) (Cotrufo and Lavallee 2022; Hartmann and Six 2022).

SOM is also key to good soil porosity and reducing compaction, which improve water infiltration and enable water storage throughout the soil profile. These functions can buffer crop yields during droughts—another reason SOM can be an indicator of crop yields (Basche et al. 2016). These functions are dynamic and change in response to many factors, e.g., residue management, nutrient inputs, disturbance, and climate, which makes it difficult to isolate the individual impact of a practice change on SOM over time.

For now, P-SHEC assumes an otherwise stationary SOM baseline based on the county average SOM. The tool uses the county-level crop-specific average SOM to estimate the baseline yield for each crop. Yield changes are then calculated based on estimated SOM changes, which are derived from COMET-Planner (Figure 1). Like P-SHEC, COMET-Planner also assumes a fixed soil carbon baseline, defined as the starting point of SOC before any soil health practices are implemented and only provides estimates of SOC changes due to the adoption of soil health practices. The stationary baseline assumption simplifies the modeling processes, but it can also introduce bias if actual SOM at a specific farm differs substantially from the county average or if SOM is changing over time due to other factors (management, climate, etc.).

P-SHEC assumes that the contribution of each crop is additive, and it provides estimates for rotations involving multiple crops by stacking the predicted outcomes of individual crops. This approach makes the tool straightforward to use and interpret. The order in which crops are added to the rotation does not alter the final estimates. In practice, however, crop rotation may affect SOM accumulation based on the diversity and sequence of rotation crops and consequently influence crop yields (Oldfield et al. 2022). Incorporating these interactions presents an exciting opportunity to further improve the accuracy of the tool in the future.

More detailed measurements of each of the processes mentioned above could potentially provide higher resolution indications of how crop yield will fare, however those data are not *(yet)* available at the county or finer scale across the US. AFT recognizes this is the first iteration of P-SHEC. As more fine-scale crop yield and soil data become available, P-SHEC could be improved.

1.3. P-SHEC history

P-SHEC is the product of a cooperative agreement that began in 2022 between AFT and USDA NRCS Ecological Sciences and Soil Health Divisions. Prior to P-SHEC, AFT developed the Retrospective Soil Health Economic Calculator (R-SHEC), also in collaboration with NRCS, which is available at https://farmlandinfo.org/rshec-toolkit/. AFT has produced dozens of case studies of "soil health successful" farmers from across the US (https://farmlandinfo.org/publications/soil-health-case-studies/) with the aim of demonstrating how soil health practices can make good business sense for farmers. The Predictive-SHEC grew from AFT's and NRCS's shared desire to estimate for farmers who have not yet adopted soil health practices but are interested in how they may benefit economically from soil health practices adoption.

The science behind P-SHEC was greatly influenced by the work of Kane et al. (2021), which looked at how SOM buffered crop yields from losses during droughts nationwide. AFT started with a linear regression, like Kane et al. (2021), and an MS Excel-based tool for corn, soybeans, and wheat on rainfed croplands, and two sets of machinery costs from the University of Illinois (Farmdoc) and Iowa State University. The linear regression worked well at the national level, but it did not perform well for building a relationship between SOM and yield at the county level. So, AFT transitioned to a machine learning model (described in detail in Section 3). This tool applied to 34 states and about 1,300 counties. The tool's Excel file was inconveniently large due to the immense amount of data the tool uses.

With time and NRCS support, the AFT team further improved the model, expanded the tool to seven more crops (Table 1), incorporated irrigated cropping systems (explained in Section 3) and added state-specific machinery costs—using region-specific estimates when state-based data were unavailable or insufficient. In 2025, AFT transitioned the tool to a web application, available here: AFT's P-SHEC tool. This resulted in the current tool, which reaches 10 row crops across 41 states and 1962 counties (Figure 3), with improved model performance and a more user-friendly interface.

All data and source code for the model are available at the following Github web site: https://github.com/American-Farmland-Trust/AFT-P-SHEC.git. Please reach out to the AFT project team with any questions or suggestions on the data and code.

2. Methodology and Data: Recurring Cost Analysis

2.1. Partial budget analysis overview

In P-SHEC, a partial budget analysis (PBA) is used to calculate the recurring costs and benefits associated with adoption of soil health practices. A PBA is an economic tool used to evaluate the financial impact of small changes in a farm or business operation, focusing only on the costs and benefits that changed because of a specific decision, rather than analyzing the entire operation (e.g., enterprise budget analysis). Key components of a PBA are additional costs, reduced revenue, additional revenue, and reduced costs. From these four financial categories, net change in profit ($\Delta Profit$) can be calculated as:

$$\Delta Profit = (Additional\ Revenue + Reduced\ Costs) - (Additional\ Costs + Reduced\ Revenue) \quad (1)$$

The first sum in Equation (1) denotes the total financial gains associated with the proposed change, while the second sum denotes the total financial losses associated with the proposed change. If the change in profit is positive, it is expected that the practice adoption is associated with increasing overall

profitability, whereas if the change in profit is negative, it is expected that the practice adoption is associated with decreasing overall profitability.

In the context of partial budget analysis, a T-chart is typically used to list and compare the added revenues and reduced costs on one side (left columns in Figure 4), and the added costs and reduced revenues on the other (right columns in Figure 4). The T-chart format helps visualize the financial impact of a proposed change by organizing the gains and losses for easy comparison.

Increases in Net Income

Increase in Income

ITEM	PER ACRE	ACRES	TOTAL
Increase in net income from adding double crop triticale (2.5 dry ton/ac)	$154	1,080	$166,656
Yield increase in hay due to interseeding triticale into alfalfa (+1 dry ton/ac)	$154	338	$51,960
Yield increase in corn silage (+1.6%)	$22	950	$21,208
Grazing 100 acres multi-species cover crop	$150	100	$15,000
Total Increased Income			**$254,824**

Decrease in Cost

ITEM	PER ACRE	ACRES	TOTAL
Machinery cost savings due to reduced tillage alfalfa & no-till corn silage	$62	1,400	$86,883
One less week of irrigation on corn silage	$7	950	$6,384
Eliminate gully-fixing caused by erosion	$2	1,700	$3,738
Total Decreased Cost			**$96,351**
Annual Total Increased Net Income			**$351,175**
Total Acres in this Study Area			**1,700**
Annual Per Acre Increased Net Income			**$207**

Decreases in Net Income

Decrease in Income

ITEM	PER ACRE	ACRES	TOTAL
Decrease in net income from shifting 300 acres alfalfa to corn silage	$88	300	$26,297
Total Decreased Income			**$26,297**

Increase in Cost

ITEM	PER ACRE	ACRES	TOTAL
Cover crop and interseeding triticale costs (seed, planting machinery, fertility)	$106	508	$53,692
Soil health practices learning activities (100 hrs/yr + $3000 conference fees)			$5,923
Total Increased Cost			**$59,615**
Annual Total Decreased Net Income			**$85,911**
Total Acres in this Study Area			**1,700**
Annual Per Acre Decreased Net Income			**$51**

Figure 4. Example of Partial Budget Analysis T-Chart.
Source: Heglar Creek Farms, Idaho AFT-NRCS Soil Health Economics Case Study. Available here.

2.2. Regional cost data motivation

Geographically granular cost data (e.g., state or region level) is used in the P-SHEC tool. Localized data increases the accuracy of production estimates by capturing geographical variations in machinery costs. Additionally, it allows for the inclusion of more specific input options, such as machinery types and fertilizer applications, which can vary significantly across regions and influence cost outcomes.

Cost variations are influenced by several region-specific factors, especially differences in purchase prices. Costs also differ non-linearly based on implement sizes, meaning larger or more specialized equipment does not scale proportionally in cost or efficiency. Additionally, variations in equipment performance and usage rates, local topography, and the average size and shape of fields all contribute to the complexity of estimating costs accurately at a broader scale.

2.3. Regional cost data sources and relevant calculations

To provide accurate cost data, an extensive literature review was conducted to collect different regional datasets for farm machinery cost estimates. Each regional dataset includes variable estimates for purchase price, salvage value, interest expense, repairs and maintenance, ownership expense, fuel expense, and labor expense estimated for that specific region. Additionally, performance rates for self-propelled machinery and implements were provided to calculate the per acre cost of the equipment used for each operation. These performance rates were then multiplied by the performance rates of the tractor, resulting in the cost of using the tractor for that operation. Estimates were then added together for a final total cost per acre estimate. Combine machinery estimates often differ based on the size of an

individual operation. For this reason, the hours used per year parameter, that is used to estimate other machinery costs will often lead to a drastic overestimation of combine machinery costs.

In order to provide comparable estimates across all regional cost estimates, an average of estimated acres of use per year was taken from Lattz and Schnitkey (2023a) for each of the combines represented. This allowed the creation of a linear relationship between horsepower and acres used per year since horsepower is often an indicator for production ability with combines. This linear relationship was then used to calculate the estimated acres of use per year for combines in other regional datasets.

Five state-level budget datasets were used: Georgia (Smith 2024), Illinois (Lattz and Schnitkey 2023b), Louisiana (Deliberto and Hilbun 2023), Minnesota (Lazarus 2024), and Mississippi (Mississippi State University Department of Agricultural Economics 2024). Because data are not available for all states, data from a representative state was used as a substitute where necessary, as detailed in the Appendix Table 1.

2.4. Excluded cost and benefit categories

The P-SHEC tool focuses specifically on estimating cost changes and yield effects associated with the adoption of practices that influence SOM. The tool does not include other potential costs and benefits, such as improvements in water quality, increased water retention, reduced erosion, or other ecosystem services that may result from the adoption of soil health practices. Additionally, it does not provide any estimation of public costs and benefits, such as those related to environmental policy, public health, or long-term societal gains or losses, due to a lack of data and the difficulty in quantifying these outcomes at the societal level.

3. Methodology and Data: Long-Term Analysis

3.1. Overview

To estimate the long-term benefits of adopting soil health practices, the current P-SHEC tool focuses primarily on two aspects: changes in crop yield and enhanced yield resistance to drought. Both of these are estimated based on changes in SOM resulting from the long-term adoption of soil health practices.

The tool uses national datasets (Appendix File 1) encompassing gradients of SOM and county-level crop yields across the conterminous U.S. These datasets enabled AFT to generate robust machine learning models that quantify the relationship between SOM and yield, while accounting for other environmental covariates. Once generated, these models are used to predict changes in crop yield associated with changes in SOM due to soil health practices adoption. A similar modeling approach is also applied to evaluate improvements in drought resistance.

3.2. Data collection and processing

Developing the predictive models requires integration and analysis of multiple national datasets. This section provides detailed information on the data sources and outlines the data processing steps used in preparing these datasets for subsequent modeling efforts.

3.2.1. SOM changes associated with soil health practices

To estimate changes in SOM associated with soil health practices, changes in soil organic carbon (dSOC) estimates were obtained from COMET-Planner (http://comet-planner.com/). COMET-Planner provides county-specific estimated dSOC between a baseline scenario and soil health practice implementation

using a sample-based metamodeling approach. These estimates were originally generated from COMET-Farm (https://comet-farm.com), which is based on the DayCent Model (Parton et al. 1998). The outputs were then averaged across all samples to generate a mean dSOC estimate for implementation of each soil health practice at the county level for mineral soils.

COMET-Planner's estimated dSOC is converted to changes in SOM (dSOM) in P-SHEC using the following steps. COMET-Planner reports dSOC changes in CO_2-equivalent units (MT CO_2e/ac/year). We first converted the CO_2 mass to carbon mass for 30 cm depth modeled and normalized by the county-mean soil bulk density data derived from gSSURGO (Soil Survey Staff 2024) to arrive at Mg SOC/Mg soil/year. County-mean bulk density ("dbthirdbar_r") was calculated for areas within the county identified as croplands in the 2024 National Cultivated Layer of the Crop Data Layer (CDL) dataset (USDA NASS 2024a), which represents cultivated lands based on the most recent five-year period (2020-2024). Finally, we used the van Bemmelen factor (assumes that SOC makes up 58% of the mass of SOC) to convert dSOC to Mg dSOM/Mg soil/year and multiplied by 100 to express SOM as a percentage of soil mass.

3.2.2. Yield data

We extracted county-level yield data following the modified pipeline outlined in Kane et al. (2021). Briefly, yield data were downloaded for each crop (Table 1) across all conterminous U.S. counties for the period spanning 2000 to 2023 from NASS survey data (USDA NASS 2024b). Data were limited to the years 2000-2023 to minimize the potential confounding effects of improvements in crop cultivars and management (Kane et al. 2021) and to align with other data sources, such as the crop frequency layers and drought index detailed below. Inclusion criteria for counties required a minimum of 15 of 24 years of yield data of the respective crop. AFT chose 15 years to ensure that counties included had a relatively consistent production of the target crop while also allowing for a relatively broad coverage of states and counties in the tool (i.e., 23 out of 24 years was too geographically limiting).

For each crop, yield data were detrended on a county basis to remove the long-term mean yield changes and variability caused by technological and climate changes. This step was done by fitting a locally weighted regression model with county yield as the response variable and year as the predictor. Models were fitted using a ten-fold cross validation with the span parameter constrained to a range of 5-10 years and the degree parameter set to either 1 (linear) or 2 (polynomial) (Lu et al. 2017; Kane et al. 2021). The residual yield from the yield trend was then added to the mean county yield across years to generate the detrended yields.

We also collected the percentage of irrigation area of each county to be included in generating the model, as the relationship between SOM and yield may be confounded by the effect of irrigation. To do this, we retrieved U.S. Census of Agriculture data on the total number of crop acres harvested and the total number of irrigated crop acres harvested in each county for the years 2002, 2007, 2012, 2017, and 2022 (USDA NASS 2024b). We calculated the percentage of crop-growing acres that were irrigated in each county for every census year and then computed the average irrigation percentage over the five census years. In addition, counties were classified as rainfed if, on average, 5% or less of their crop-growing acres were irrigated, while all other counties were classified as irrigated. The final data set included in the yield analysis is detailed in Appendix Table 2.

3.2.3. Soil and nitrogen input data

County-level soil data were extracted from the Gridded Soil Survey Geographic (gSSURGO) database (Soil Survey Staff 2024) following a modified pipeline outlined from Kane et al. 2021). Soil variables extracted at the depth of 0-30 cm included sand, silt, clay, organic matter (OM), available water capacity (AWC), available water storage (AWS), cation exchange capacity (CEC), pH, soil order, wilting point, slope, and drainage classes. AWS was directly extracted from gSSURGO's Value Added Look Up Table (Value1), which was pre-summarized by depth. The rest of the variables were extracted by weighting the percentage of the component of the map unit at 0-30 depth.

To calculate the county-level average for soil variables, we first excluded non-crop growing areas for each specific crop within each county. The CropScape Crop Frequency Layer (CFL) provides raster data at 30-m resolution, detailing crop distribution for corn, cotton, soybeans, and wheat between 2008 to 2023 (Han et al. 2012). We used this layer to filter for pixels that had 3 or more years of growing the respective crop, to filter for areas growing the crop with relative consistency, and not opportunistically. Note this filtering criterion is less conservative than the criterion for including a county's yield data, where a county needed to have 15 of 24 years of crop yield data, as described in Section 3.2.2. After this filtering process, we computed the mean of each soil variable for the resulting crop-growing areas for corn, cotton, soybeans, and wheat within each county.

For all other crops where a CropScape CFL layer was not available, we created unique frequency data layers. We uploaded the 30-m national CDL (USDA NASS 2024a) for the years 2008-2023 into ArcGIS and isolated the respective crop from each layer. To expedite the process, we developed an ArcGIS model (Figure 5) using the "Iterate Rasters" function to iterate through the annual CDL layers and the "Reclassify" function to extract and isolate each crop. In each resulting binary raster, pixels representing the crop in question were assigned a value of 1, while all other areas received a value of 0. This process produced 16 annual individual layers for each crop. Subsequently, we used the "Raster Calculator" geoprocessing tool in ArcGIS to produce a georeferenced data layer indicating how frequently a crop was grown over the conterminous U.S. from 2008-2023. This custom-generated frequency layer was used to filter crop-growing pixels for processing the soil data for barley, sorghum, oats, peanuts, sugar beets, and tobacco.

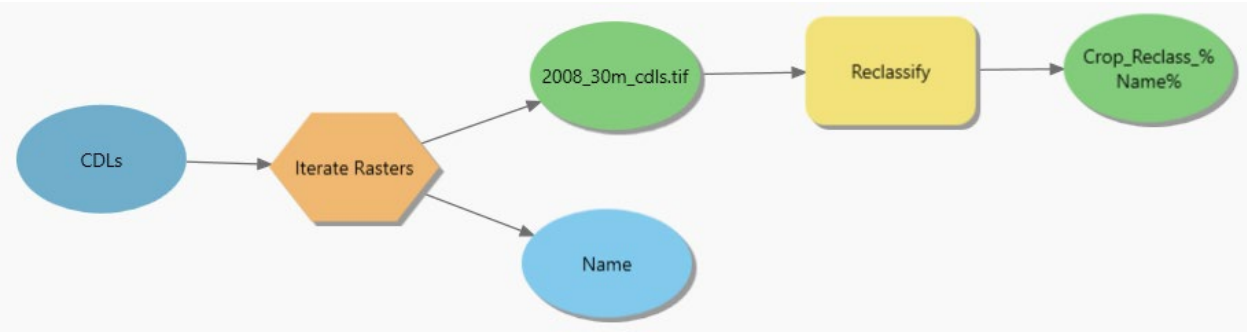


Figure 5. ArcGIS model used to create the crop frequency data layers.

When processing gSSURGO soil data, we noticed that some counties have crop growing areas with suspiciously high SOM (>60%). These areas led to some counties having a mean SOM greater than 10%, which had a large impact on the downstream yield and SOM relationship. To further clean the soil data and to align with COMET-Planner (which assumes mineral soils), we removed pixels with SOM greater

than 10% before calculating the county mean SOM. Most counties remained unchanged, and only 3% of counties had more than 10% of soil data removed by this extra cleaning step. The county mean SOM of the cleaned dataset ranged from 0 to 7%, which is much smaller than the uncleaned dataset (0-35%).

Yield is not only a product of SOM, but also closely related to nitrogen (N) fertilizer rates. The machine learning model accounts for county average N fertilizer rate as one of the independent variables for predicting yield. AFT obtained county-level N input data for corn, winter wheat, cotton, barley, and sorghum from Zhang, Cao, and Lu (2021). This dataset synthesizes multiple databases and provides an estimate of the total N input derived from chemical N fertilizer use, manual N application, atmospheric N deposition, and N residues from soybean fixation from the previous year on a county scale during 1970-2019. For the downstream analysis, we calculated the average county-level mean N input by taking the mean of N inputs from the years 2000 to 2019. For oats, tobacco, and sugar beets, since crop-specific N input data are not available, we used the county-level N use estimates on crops from the Nutrient Use Geographic Information System NuGIS database (IPNI 2024). For soybeans and peanuts, we assumed zero N input as both are leguminous crops.

3.2.4. Crop growing season data and climate data

Weather also plays a role in determining crop yields, and this is accounted for in the machine learning model. To obtain climate data, we first extracted daily precipitation, mean, minimum, and maximum temperature data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) weather dataset dataset (2000–2023) at a 4 km spatial resolution (PRISM Climate Group 2025). These data were then aggregated to the county level to produce daily county-level climate variables such as daily precipitation and mean temperature.

Next, to isolate the crop growing period, we applied two approaches: 1) state-level crop-specific calendars (Sacks et al. 2010), defined as the period between planting and harvest, were used to calculate the plant-to-harvest cumulative precipitation and mean daily temperature for each crop in each year; and 2) crop growing seasons, based on crop phenological stages, were used to calculate Growing Degree Days (GDD). Specifically, we defined the growing season from planting to maturity for corn, sorghum, barley, peanuts, and oats; from planting to dropping leaves for soybeans; from planting to bolls opening for cotton; and from transplanting to harvesting for tobacco. Crop growth progress data was extracted when 50% of a state was planted or matured based on the state-level weekly progress reports. These crop growth progress data were extracted from USDA NASS for all available states and averaged to generate state level averages (USDA NASS 2024b). GDD were then calculated for each crop in each county using the crop-specific upper and lower temperature thresholds (Schillerberg and Tian 2023; King and Tarkalson 2017; Oakes et al. 2020; Viator et al. 2005) (see Table 2 for detailed temperature thresholds).

Table 2. Temperature thresholds used to calculate Growing Degree Days (GDD) for crops.

Crop	Base temperature (°C)	Upper temperature (°C)
Corn	10	30
Soybeans	10	30
Sorghum	10	30
Sugar beets	10	30
Wheat	0	30
Barley	0	30
Oats	0	30
Cotton	15.5	30
Peanuts	13	30
Tobacco	10	32.2

3.2.5. Cause of loss insurance data and DSCI data

To identify drought conditions, two sources of data were used: County-level crop insurance data from USDA Risk Management Agency (RMA) and the Drought Severity and Coverage Index (DSCI) (Akyuz 2017) spanning from 2000 to 2023.

The county-level crop insurance data was used to calculate loss cost ratio (LCR) and loss acres ratio (LAR). LCR was defined as the ratio of total indemnities due to a weather-related cause over total liabilities for a specific crop in the county for a given year. LAR was calculated by taking the ratio of total acres with weather related indemnities over the total insured acres for a specific crop in the county for a given year. The weather-related indemnities data was from the RMA Cause of Loss (COL) (USDA RMA 2024a), and the total liabilities and total insured acres were from the RMA Summary of Business (SOB) datasets (USDA RMA 2024b). LCR reflects the ratio of actual payment for a specific weather-related cause, while LAR reflects the proportion of insured acres in the county with a loss (Aglasan et al. 2023). In the tool, we calculated both LCR and LAR for each county-year. After consulting with RMA staff, we combined drought and heat losses together to use as the drought loss, as there may be instances in which a particular loss was the result of both drought and heat, and it may be hard to distinguish drought and heat for a particular loss.

In addition to the crop insurance data, we also used DSCI, which provides drought levels to a single value on a weekly basis for an area (Akyuz 2017). County-level DSCI was calculated as the average of weekly mean DSCI values during the growing season for each crop. Due to varied planting and mature dates in different states for different crops, crop growing season data described in the previous section was used to extract state-level planting and maturity date for each crop, which was further used to extract DSCI values.

3.3. SOM-yield relationship

3.3.1. Model development

To obtain an accurate prediction of the relationship between SOM and crop yield, we selected crops based on a series of criteria to ensure adequate temporal and geographic coverage for statistical robustness: 1.) Only crops with county-level yield data from USDA NASS were considered. Crops with data reported only at the state level, such as cranberries in Massachusetts, were excluded. 2.) We

included only counties that had at least 15 years of yield data available between 2000 and 2023 (out of 24 years total) to ensure consistent crop production. 3.) Crops were required to have data from at least 50 counties nationwide and a minimum of 1,000 yield observations to support statistically meaningful analysis. 4.) Only crops and practices recognized in COMET-Planner were retained. Crops such as rice, which involve distinct cultivation methods and are excluded from COMET-Planner, were also excluded from P-SHEC. 5) CDL data or CFL data are available for each selected crop to enable soil data extraction from relevant crop-growing areas. The final data set included in the yield analysis are detailed in Appendix Table 2.

A series of regression models were tested, and we selected the model with the best yield predictions following the tidymodels framework in R (Kuhn and Wickham 2020). We tested a range of simple and machine learning regression models, including least absolute shrinkage and selection operator (Lasso) regression, generalized additive model, classification and regression tree, random forest (RF), eXtreme gradient boosting (XGB), and support vector machines (SVM) regression models. In all models, the detrended yield was used as the dependent variable, and SOM and other environmental covariates were used as the independent variables (Appendix Table 3). State was also included in the model to account for potential geographic differences in management and environment that are not accounted for by other variables. All numeric predictors were normalized to have a standard deviation of one and a mean of zero. Models were fitted with two-way interactions with state as either fixed or random. For the machine learning models, all variables were used, and potential parameters were tuned on a grid = 25. The data were divided into 80% training and 20% testing subset, and models were fitted using a ten-fold cross validation with five repeats to reduce the potential model overfitting. Model performance was evaluated using root mean squared error (RMSE), mean absolute error (MAE), and R-square. For all selected crops, XGB model showed the average lowest RMSE, and was used in county yield predictions (Appendix Table 4).

AFT checked to see if crop yields in irrigated counties required a different model from rainfed counties. For each crop we tested two modeling strategies: 1.) a single model applied to both irrigated and rainfed counties, or 2.) separate models for irrigated and rainfed counties. According to the model performance on predicting county-level yields, applying one model showed a greater prediction accuracy than the separated models. Therefore, a single XGB model was used for all selected crops.

3.3.2. Preparing results for P-SHEC

To estimate the relationship between SOM and crop yield, yields were predicted using a gradient of SOM values, while holding other variables constant for each county. Yield predictions were generated using Accumulated Local Effects (ALE) to quantify how incremental changes in SOM (in 0.01-unit steps) influence predicted yield, across the full range of SOM values observed for each crop. ALE plots are particularly well-suited for this purpose, as they provide unbiased effect estimates even when input variables are correlated (Molnar 2025). ALE curves were generated separately for each county for each crop to capture the SOM-yield relationships.

Prediction intervals were generated to quantify the uncertainty associated with predicting individual data points. A 90% prediction interval was generated for every individual yield prediction using a bootstrap method with 1000 bootstrap models.

3.4. SOM-drought resistance relationship

The relationship between SOM and drought resistance is defined as the deviation of drought-impacted yields from the expected long-term yield trend, per unit change in SOM due to the implementation of soil health practices.

3.4.1. Yield change

Yield change due to drought in any given county year is the difference between the observed yield and yield trend (expected yield in a normal year based on the long-term yield trend, as described in section 3.2.2), divided by the yield trend as shown in Equation (2). The yield change percentage reflects the yield departures as a proportion of their long-term yield trend (Li et al. 2019). A negative yield change indicates a yield loss in a given county-year. Crop yield changes were calculated for each county from 2000-2023 for all crops using:

$$Yield\ change\ (\%) = \frac{\text{Observed yield - Yield trend}}{\text{Yield trend}} * 100\% \quad (2)$$

3.4.2. Identify drought county-years

To ensure that the SOM-drought resistance dataset only included county-years where drought was the primary cause of yield loss, we included only county-years that were drought-affected using either of the following two criteria: 1) the county's drought-specific LCR was greater than the combined total of all other weather-related causes for that year and the LAR exceeded 0.2, indicating substantial loss area; or 2) the county's DSCI was greater than 400, meaning that the county was, on average, under extreme drought conditions during the crop growing season. Relying solely on LCA and LAR can sometimes result in an overly narrow dataset, as certain drought-related losses may not be reported to RMA. To address this limitation, we additionally incorporated county-years with a DSCI index greater than 400, capturing potential drought impacts during the growing season that did not meet the LCA/LAR criteria. The 400 threshold, though somewhat arbitrary, was selected to reflect conditions of extreme drought at the county level. The final SOM-drought resistance data set is detailed in Appendix Table 2.

3.4.3. Model development

To obtain the prediction of the relationship between SOM and crop yield drought resistance, we applied a similar modeling pipeline as described in 3.3.1, testing a series of regression models and selecting the model with the best performance. Yield change (%) was used as the response variable, while four soil variables (after removing highly correlated ones), climate variables, nitrogen input, and DSCI were used as predictors (all variables are listed in Appendix Table 3). Due to the relatively small sample size for each crop, a single model was developed across all crops, excluding sugar beets and tobacco, which had limited observations under drought and negatively impacted model performance. The XGB model showed the best model performance and was used for final drought resistance predictions (Appendix Table 4).

3.4.4. Results in the tool

To estimate the relationship between SOM and crop yield drought resistance, yield changes were predicted using a gradient of SOM values, while holding all other environmental variables consistent at their county-level means and holding DSCI to 300, representing an average of severe to extreme drought for each county. Yield change estimations were generated for each 0.01 increment in SOM, spanning from the minimum to the maximum values observed, using ALE as described in 3.3.2.

3.4.5 Drought frequency prediction

To account for the likelihood that some years will have normal conditions while others may experience severe drought, the probability of a county experiencing drought over a 10-year period was used to adjust potential economic outcomes related to yield changes. Markov transition probabilities (Markov, 1907) were estimated at the county level for each crop-growing season to capture these dynamics. A Markov transition probability matrix for severe drought describes the probabilities of transitioning between severe drought or normal weather conditions (states) specified time steps. The time steps for this analysis are crop specific growing seasons. Crop growing seasons are described in section 3.2.4. The DSCI drought index was used to estimate the steady-state Markov probability that a county will experience severe drought conditions within crop-specific seasons, based on data from 2000 to 2023. The Markov steady-state transition probabilities are used to calculate the expected value of the change in yield revenue under increasing levels of SOM.

3.5. Model assumptions and limitations

We have made every effort to minimize bias and improve the robustness of the models. However, several limitations remain and should be carefully considered when interpreting the results.

1. Although we included a diverse set of variables representing soil, climate, and landscape characteristics, this list is not exhaustive. Selection was guided by two key considerations: (a) we prioritized variables with high data quality and broad spatial coverage at the county level, and (b) we aimed to reduce multicollinearity and avoid overfitting, especially since some crops had a limited number of observations, making it difficult to support complex models.
2. The modeling approach focuses on estimating the effect of changes in SOM while holding all other variables constant at their county-level means. This assumption isolates the influence of SOM but does not reflect real-world scenarios where multiple variables often change simultaneously. For example, an increase in SOM may also improve soil water holding capacity. To partially account for such correlations, we used the ALE estimation for more robust interpretation.
3. Some crops, such as peanuts and tobacco, had relatively few observations, which reduced the predictive performance of the models for these crops. As a result, point estimates for these crops may be less reliable. To address this, we calculate prediction intervals for all estimates to offer a more objective understanding of the model's uncertainty.
4. Although the data indicate a general positive correlation between SOM levels and yield, there are instances where this relationship may not hold. In particular, in cases where low levels of SOM are associated with high yields, the model may not produce a positive outcome for these specific counties. This occurs because the model attempts to align with the overall trend between SOM and yield, while simultaneously accounting for the county-specific low SOM and high yields.
5. When crop yields are already near their potential under the given environment, further increases in SOM may not lead to higher yields. However, increasing SOM may still offer other benefits, like increasing yield stability and enhancing drought resistance.

4. Long-term Economic Results

4.1. Calculation of 10-year change in yield revenue

County-level, crop-specific SOM–yield model outputs were used to estimate long-term changes in yield revenue resulting from shifts in SOM levels. Baseline SOM values for each county and crop were obtained from the gSSURGO database (Soil Survey Staff 2024). When users include multiple rotation crops with different acreages, we compute a weighted baseline SOM according to user-specified crop acreages. This acreage-weighted average applies to all crops in the user-defined rotation, allowing a single, consistent SOM baseline, to which estimated 10-year SOM increases from soil health practice implementation (via the COMET-Planner dataset) were added to estimate updated crop-specific county-level SOM values. The acreage-based weighting approach was necessary because COMET-Planner provides SOM projections at the county level, but not by crop.

It was assumed that, without the adoption of soil health practices, baseline SOM levels would remain constant. Baseline yield revenue for crop *j* was calculated by multiplying the baseline yield ($Y_{j,base}$; units/acre) associated with baseline SOM by the crop price (P; $/unit). For each of years 1 through 10, yield revenue for crop *j* was recalculated using the projected yield at the updated SOM level (Y_{jt}; $t = 1, \dots, 10$), multiplied by the crop price (P). The difference between the updated and baseline yield revenues (ΔR) was calculated annually and aggregated to estimate the impact of SOM changes (Equation 3). A positive (negative) difference indicated an increase (decrease) in yield revenue due to changes in SOM.

$$\Delta R = \sum_{t=1}^{10} Y_{jt} \cdot P - Y_{j,base} \cdot P \quad (3)$$

4.2. Drought resistance economic analysis

The estimated county-level crop yield changes (%) associated with severe drought at the baseline SOM level ($\%Yield\ Change_{base,j}$) and at increasing SOM levels ($\%Yield\ Change_{tj}$; see section 3.4.1) were used to calculate the change in drought-related revenue per acre (ΔDR) (Equation 4).

$$\Delta DR = \sum_{t=1}^{10} \left(Y_{tj} \cdot P\right) \cdot \%Yield\ Change_{tj} - \left(Y_{base,j} \cdot P\right) \cdot \%Yield\ Change_{base,j} \quad (4)$$

The estimated change in total revenue (ΔTR), accounting for both the probability of severe drought (P_{dj}) (see section 3.4.5 for drought probability prediction methods) and the probability for normal conditions ($1 - P_{dj}$) is (Equation 5):

$$\Delta TR = \sum_{j=1}^{J} \left(1 - P_{dj}\right) \cdot \Delta R_j + P_{dj} \cdot \Delta DR_j \quad (5)$$

4.3. Combining recurring cost changes and longer-term data

For this analysis, the recurring difference in costs, as calculated through the partial budget analysis (see Section 2.1), are held constant and applied each year. This approach assumes that baseline management practices would continue annually in the absence of the soil health practice(s), resulting in a consistent year-over-year cost difference. The recurring costs are aggregated over a ten-year period and combined with the change in total revenue (Equation 5) to determine the change in net income. The change in net income is then discounted to its present value at a discount rate selected by the user.

4.4. How to properly interpret the data

The P-SHEC tool displays multiple ways to visualize results. Results are broken down in both graphical and tabular forms, with the option to break down results by practice. Additionally, users can scroll through recurring costs, a long-term predicted yield, yield resistance to severe drought, and an overall results section. The following example from Adair County, Iowa, shows estimated results and how users might interpret the results. The below example features a corn and soybean rotation where the user is evaluating the adoption of cover crops with a 50% nitrogen reduction and a transition from intensive tillage to no-till or strip till.

Figure (6) shows the recurring costs associated with the two selected soil health practices. The recurring current costs of $44 and the recurring planned of $77 are the yearly estimated changes in costs the user can expect from adopting these practices for each year over the 10-year timeline used in the tool. These results can be further broken down by changes due to each of the soil health practices. The tool also provides a bar graph to visualize the magnitude of these recurring changes.

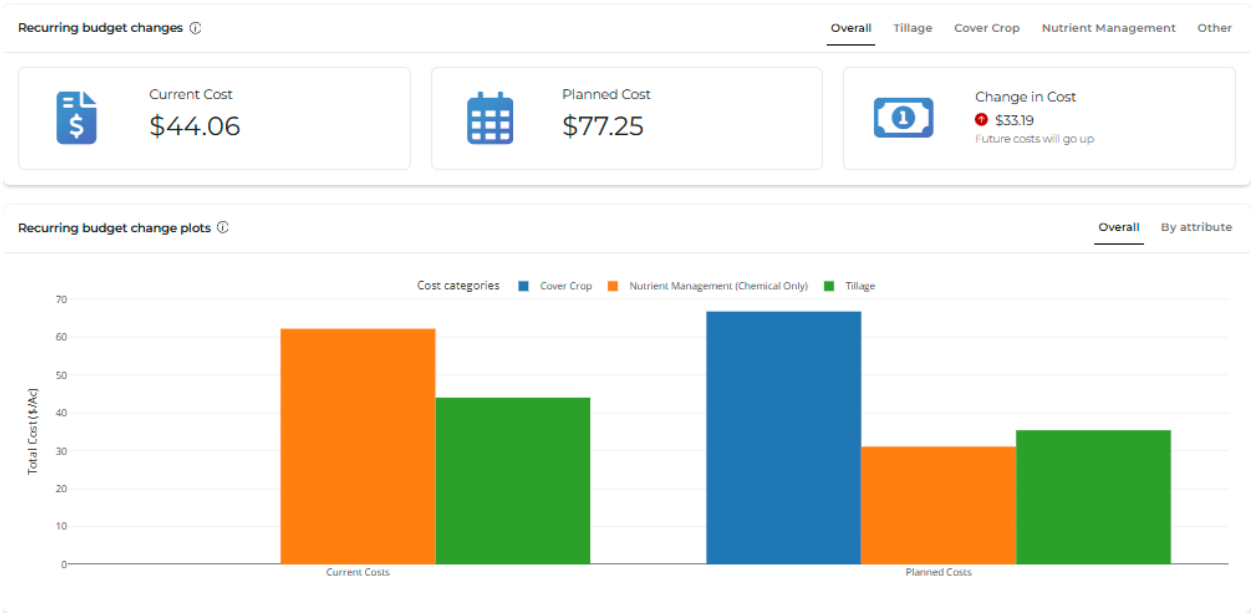


Figure 6. Example of Recurring Costs Section for a user in Adair County, Iowa exploring adoption of cover crops and no-till in a corn-soybean rotation.

The modeled long-term changes due to the increase in SOM from the adoption of the two soil health practices are shown in Figure (7). Results are separated by crops, in this example corn and soybeans. The table first displays on the left the starting, Baseline SOM, which is the county average SOM for soils in that county that grow each crop. The second column, the Updated SOM, represents Baseline SOM plus the increases in SOM over 10 years due to the addition of soil health practices. The third column, Predicted Yield, is the yield prediction at the new, updated SOM, whereas the fourth column, baseline yield, is the starting yield predicted at the baseline SOM. The fifth column, Yield Revenue Difference, is the percent change from the Predicted Yield to the Baseline Yield.

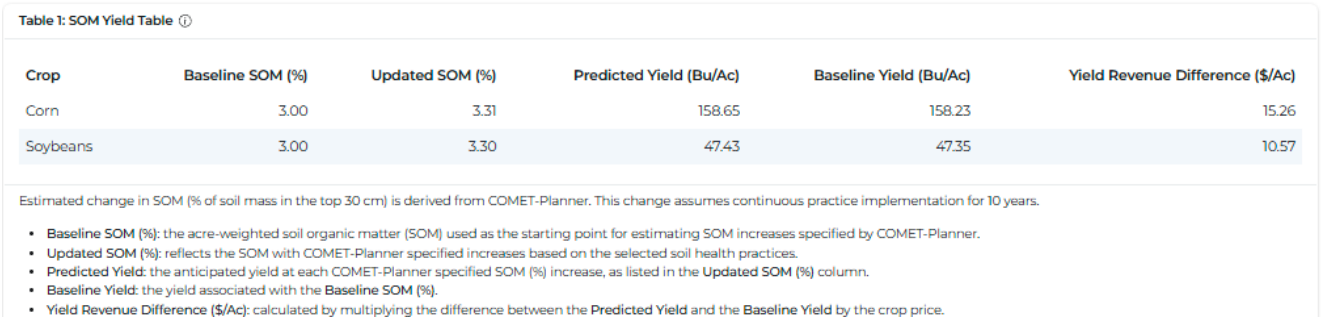


Table 1: SOM Yield Table ⓘ

Crop	Baseline SOM (%)	Updated SOM (%)	Predicted Yield (Bu/Ac)	Baseline Yield (Bu/Ac)	Yield Revenue Difference ($/Ac)
Corn	3.00	3.31	158.65	158.23	15.26
Soybeans	3.00	3.30	47.43	47.35	10.57

Estimated change in SOM (% of soil mass in the top 30 cm) is derived from COMET-Planner. This change assumes continuous practice implementation for 10 years.

- **Baseline SOM (%)**: the acre-weighted soil organic matter (SOM) used as the starting point for estimating SOM increases specified by COMET-Planner.
- **Updated SOM (%)**: reflects the SOM with COMET-Planner specified increases based on the selected soil health practices.
- **Predicted Yield**: the anticipated yield at each COMET-Planner specified SOM (%) increase, as listed in the **Updated SOM (%)** column.
- **Baseline Yield**: the yield associated with the **Baseline SOM (%)**.
- **Yield Revenue Difference ($/Ac)**: calculated by multiplying the difference between the **Predicted Yield** and the **Baseline Yield** by the crop price.

Figure 7. Example of Long-Term Yield Modeling Results Table.

The final results table, Figure 8, shows the overall results of the tool, which combines recurring costs and benefits with the long-term yield changes and drought resistance changes. These results are shown as Net Present Values (NPV) of the cumulative changes over 10 years. NPV is used to account for the time value of money. The primary result is the Net Income change, which for this example is a loss of $209 per acre over ten years. This is the total income change over ten years from adopting the two soil health practices and is shown per acre. Below, estimates of Crop Yield Revenue changes and Costs changes are displayed. In this example, Crop Yield Revenue increases by $26 per acre over ten years while the Yearly Costs increase by $270 per year. Given the 10-year planning timeline, the discount rate can be adjusted to account for the user's preferences. The tool starts with a default 3% discount rate. Entering a higher rate would mean that a user places higher value on upfront changes to net income, whereas a lower discount rate means that current and future net income changes are valued similarly.

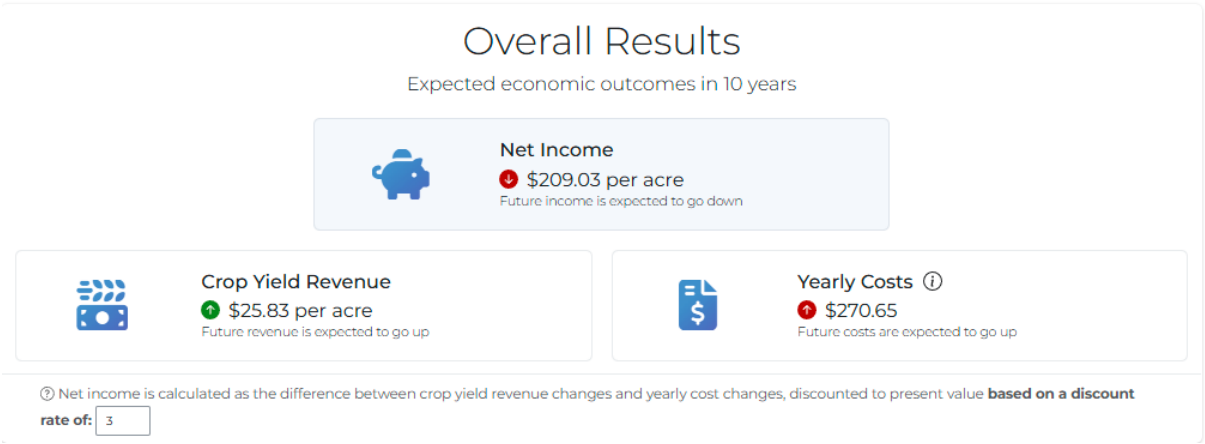


Figure 8. Example of the Overall Results Table.

5. Pilot Testing Process

AFT and NRCS established a pilot process after the completion of the first draft of the tool. The aim of the process was to improve the user experience of the tool. Twelve pilot testers were selected to

participate in the process because they represented potential user groups. The testers included farmers, extension professionals, NRCS staff and conservation advisors.

NRCS conducted the pilot testing process. Testers were interviewed over Microsoft Teams and asked to complete one of three scenarios in the P-SHEC tool. As they were filling out the example, testers provided feedback on their experience. Once complete, the testers were asked a series of questions about the tool. They were also emailed an online survey after the pilot so that they could give anonymous feedback.

Overall, the tool received positive feedback from the testers, but they did highlight important places to improve the user experience. The 12 testers rated the overall usability of the tool 4.1 out of 5, with 9 out of the 12 reviewers stating that they would recommend the tool for future use. Importantly, of the three that did not recommend the tool, one experienced technical issue with the tool that were later resolved, while another stated they would not recommend the tool because they primarily worked in livestock and nut orchard systems, which are both outside of the scope of this tool.

In the post-test questions and survey, multiple themes emerged. First, users wanted more detailed instructions and user tips within the tool. At times, testers were unsure of how to enter specific information or why they were being asked certain questions. To rectify these concerns, we have included additional user tips and text boxes within the tool. The pilot process also revealed areas where testers could misinterpret results of the tool. To ensure accurate interpretation of results, we simplified tables, visually highlighted key results, and added more text to assist in interpretation of results.

6. Knowledge Gaps and Theoretical Documentation

6.1. Exploratory approaches not included in final analysis

Several approaches and data sources were explored during the development of the model but ultimately not included in the final version, due to data limitations, model performance issues, or alignment with project objectives. While these elements were considered valuable, they did not meet the criteria necessary for robust, interpretable modeling at the national scale.

1) Crops with insufficient data: We evaluated the inclusion of beans and sunflowers, given their agricultural importance. However, these crops are reported as aggregated categories that combine multiple subtypes in the USDA NASS dataset. For instance, the "Beans" category includes 12 different types, with the most commonly reported type accounting for only 786 county-level yield observations. The minimum requirement for inclusion in the model was 1,000 county-level yield observations to ensure robust results. Similarly, sunflowers are divided into oil-type and non-oil-type, but neither had sufficient data for separate modeling. Including these crops would either compromise model reliability due to limited observations or result in generalized predictions that average across distinct crop types, reducing their usefulness for farm-level decision-making.

2) Limited geographic coverage and specialty crops: Some major agricultural states had limited representation in the final dataset for various reasons. States in New England and Nevada (Figure 2), for example, were excluded because their agricultural production is primarily driven by dairy or livestock, rather than row crops. Additionally, specialty crops with limited geographic distribution, such as sugarcane in Florida (Appendix Figure 2) or processing tomatoes in California (Appendix Figure 3), were not included. These crops often lack broad spatial coverage and are not well-represented in national

datasets, making them less suitable for inclusion in a generalized modeling framework. For inclusion in the model, we required county level yield data, as opposed to state-level yield data, and a minimum of 50 counties with yield observations.

Overall, the tools focus on row crops as opposed to tree crops and livestock and the data limitations when dealing with specialty crops led to states such as California, Texas, Florida and the New England states having limited coverage in the tool.

6.2. Hypothetical data sources that would improve the tool

During the development of the tool, we identified several data sources that could potentially improve the tool:

1) Improving spatial resolution and site-specific estimations: Currently, the tool operates at the county level, which is primarily driven by the resolution of USDA NASS crop yield data, the most reliable yield data source offering both high quality and broad spatial coverage. As a result, higher-resolution datasets, such as gSSURGO soil data (30 m resolution) and PRISM climate data (4 km resolution), were aggregated to the county level to align with the yield data. While sufficient for county level analysis, this limits the tool's ability to generate sub-county or field-level predictions that could provide more localized insights for producers.

There have been discussions around the potential development of georeferenced farm-level databases within USDA NASS by adopting geospatial conventions such as the Common Land Unit (CLU), which is currently used by the Farm Service Agency (FSA) (National Academies of Sciences and Medicine 2017). If such a database becomes available in the future, the integration of sub-county or field-level yield data could significantly enhance the tool's ability to generate site-specific recommendations for producers.

In addition, with the rapid advancement of remote sensing and crop modeling, alternative approaches for estimating sub-county yield have become more feasible. For example, the Scalable Crop Yield Mapper (SCYM) uses satellite data to estimate crop yields at a 30 m resolution (Lobell et al. 2015). SCYM has been validated using over one million ground-truth fields across the U.S. Corn Belt and has demonstrated performance, capturing 40% and 27% of pixel-level yield variation and 69% and 63% of county-level variation for corn and soybean, respectively (Deines et al. 2023). While promising, such approaches still suffer from high variability at fine scales and may require further calibration to complement survey-based data sources effectively.

Although the P-SHEC tool currently provides predictions at the county level, other data sources may offer sub-county insights that help explain within-county variability. One such resource is the National Commodity Crop Productivity Index (NCCPI) (Sinclair, Jr. et al. 2009), which is integrated into the gSSURGO database and provides 30-meter resolution estimates of potential crop productivity at a national scale for corn, soybeans, cotton, and small grains. The NCCPI is developed using soil and environmental variables similar to those used in P-SHEC.

In a preliminary analysis conducted in Pennsylvania, we observed similar spatial patterns between the NCCPI corn index and P-SHEC's county-level corn yield predictions. We also found a positive correlation between the coefficient of variation in NCCPI at the county level and the predicted yield changes from P-SHEC (correlation = 0.4, $p = 0.005$). This suggests that counties with greater within-county soil variability may see greater potential yield benefits from adopting soil health practices. These findings offer a

promising approach to linking P-SHEC's county-level predictions to sub-county variation. However, the directionality of the relationship remains unclear; it is not yet evident whether lower NCCPI values are associated with greater yield increases or vice versa. Further analysis is needed to better understand and interpret these patterns.

2) Incorporating additional co-benefits of soil health practices: As discussed in the introduction, the current version of P-SHEC focuses primarily on the economic outcomes related to changes in SOM and crop yields. However, research consistently demonstrates that soil health practices offer additional ecological benefits, which may directly or indirectly influence farm profitability. For example, soil health practices can positively influence soil nutrient dynamics, potentially reducing the need and cost for synthetic fertilizers.

In particular, biological N fixation from legume cover crops can contribute N to subsequent cash crops, reducing the amount of N fertilizer needed (Coombs et al. 2017). Similarly, no-till systems can build SOM in the surface soil by increased residue cover, thereby reducing soil losses and increasing potential mineralizable nitrogen available for crop uptake (Mahal et al. 2018). Additionally, using specific soil health indicators, such as soil biological activity, can help guide N fertilizer use to optimize economic returns and reduce environmental impacts (Franzluebbers et al. 2022).Soil health practices have been reported to improve water infiltration and increase water holding capacity, potentially lowering irrigation needs and costs (Abdallah et al. 2021). Such economic and ecological co-benefits were also observed in AFT's soil health case studies.

Nevertheless, systematically quantifying these effects requires comprehensive, consistent, and nationally representative datasets, as well as careful consideration to accurately isolate and predict individual impacts. Future integration of modeling efforts e.g., (Huang et al. 2020), satellite-derived data (e.g., Volk et al. 2024), and economic analyses could further help quantify the cost-saving benefits of reduced fertilizer and irrigation use associated with adopting soil health practices. Incorporating these additional co-benefits would significantly enhance the comprehensiveness and utility of the P-SHEC tool.

3) Integrating flexible management scenarios and dynamic baselines: Currently, the estimated changes in SOM resulting from soil health practices in P-SHEC are derived from COMET-Planner. COMET-Planner is based on a process-based model and provides a limited set of management practice options primarily intended for general planning purposes. As COMET-Planner continuously updates and expands its datasets, incorporating more flexible and detailed management scenarios could significantly enhance the tool's functionality and relevance to diverse farming systems.

In addition, integrating SOC estimations from multi-model ensemble (MME) approaches using dynamic baselines may further improve P-SHEC's accuracy. Compared to individual process-based models, MMEs combine several process-based models, capturing differences in model structure and a broader range of environmental conditions and management practices, thereby offering more robust and accurate soil carbon predictions across varying scenarios (Basso et al. 2025). Furthermore, P-SHEC's current baseline yield projections assume a static "business-as-usual" trajectory of SOM based on historical averages. This steady-state assumption can introduce bias if actual SOM trends under conventional management diverge from those averages. Implementing dynamic baselines that adjust for site-specific factors and evolving management practices could provide more precise and reliable SOM and yield estimates (Basso et al. 2025), substantially enhancing the predictive accuracy of the tool.

References

Abdallah, Ahmed M., Hanuman S. Jat, Madhu Choudhary, Emad F. Abdelaty, Parbodh C. Sharma, and Mangi L. Jat. 2021. "Conservation Agriculture Effects on Soil Water Holding Capacity and Water-Saving Varied with Management Practices and Agroecological Conditions: A Review." *Agronomy* 11 (9): 1681. https://doi.org/10.3390/agronomy11091681.

Akyuz, F. A. 2017. "Drought Severity and Coverage Index." United States Drought Monitor. https://droughtmonitor.unl.edu/About/AbouttheData/DSCI.aspx.

Basche, Andrea D., Thomas C. Kaspar, Sotirios V. Archontoulis, et al. 2016. "Soil Water Improvements with the Long-Term Use of a Winter Rye Cover Crop." *Agricultural Water Management* 172 (July): 40–50. https://doi.org/10.1016/j.agwat.2016.04.006.

Basso, Bruno, Tommaso Tadiello, Neville Millar, et al. 2025. "A Multi-Model Ensemble Reveals Soil Carbon Gains from Regenerative Practices in the U.S. Midwest Cropland." Preprint, February 8. https://doi.org/10.1101/2025.02.04.636509.

Coombs, Claire, John D. Lauzon, Bill Deen, and Laura L. Van Eerd. 2017. "Legume Cover Crop Management on Nitrogen Dynamics and Yield in Grain Corn Systems." *Field Crops Research* 201 (February): 75–85. https://doi.org/10.1016/j.fcr.2016.11.001.

Cotrufo, M. Francesca, and Jocelyn M. Lavallee. 2022. "Soil Organic Matter Formation, Persistence, and Functioning: A Synthesis of Current Understanding to Inform Its Conservation and Regeneration." In *Advances in Agronomy*, vol. 172. Elsevier. https://doi.org/10.1016/bs.agron.2021.11.002.

Daugstad, Karoline, Katrina Rønningen, and Birgitte Skar. 2006. "Agriculture as an Upholder of Cultural Heritage? Conceptualizations and Value Judgements—A Norwegian Perspective in International Context." *Journal of Rural Studies* 22 (1): 67–81. https://doi.org/10.1016/j.jrurstud.2005.06.002.

Deines, Jillian M., Kaiyu Guan, Bruno Lopez, et al. 2023. "Recent Cover Crop Adoption Is Associated with Small Maize and Soybean Yield Losses in the United States." *Global Change Biology* 29 (3): 794–807. https://doi.org/10.1111/gcb.16489.

Deliberto, Michael A., and Brian M. Hilbun. 2023. *Projected Costs and Returns: Crop Enterprise Budgets for Corn Production in Louisiana, 2023*. Technical Report Agricultural Economics Information Report Series No. 365. Department of Agricultural Economics and Agribusiness, Louisiana State University Agricultural Center. https://www.lsuagcenter.com.

Franzluebbers, Alan J., Robert Shoemaker, Jeff Cline, et al. 2022. "Adjusting the N Fertilizer Factor Based on Soil Health as Indicated by Soil-test Biological Activity." *Agricultural & Environmental Letters* 7 (2): e20091. https://doi.org/10.1002/ael2.20091.

Han, Weiguo, Zhengwei Yang, Liping Di, and Richard Mueller. 2012. "CropScape: A Web Service Based Application for Exploring and Disseminating US Conterminous Geospatial Cropland Data Products for Decision Support." *Computers and Electronics in Agriculture* 84 (June): 111–23. https://doi.org/10.1016/j.compag.2012.03.005.

Hartmann, Martin, and Johan Six. 2022. "Soil Structure and Microbiome Functions in Agroecosystems." *Nature Reviews Earth & Environment* 4 (1): 4–18. https://doi.org/10.1038/s43017-022-00366-w.

Huang, Guorui, Arjen Y. Hoekstra, Maarten S. Krol, et al. 2020. "Water-Saving Agriculture Can Deliver Deep Water Cuts for China." *Resources, Conservation and Recycling* 154 (March): 104578. https://doi.org/10.1016/j.resconrec.2019.104578.

IPNI, International Plant Nutrition Institute (IPNI). 2024. "A Nutrient Use Information System (NuGIS) for the U.S." http://www.ipni.net/nugis.

Kane, Daniel A, Mark A Bradford, Emma Fuller, Emily E Oldfield, and Stephen A Wood. 2021. "Soil Organic Matter Protects US Maize Yields and Lowers Crop Insurance Payouts under Drought." *Environmental Research Letters* 16 (4): 044018. https://doi.org/10.1088/1748-9326/abe492.

King, Bradley, and David Tarkalson. 2017. “Irrigated Sugarbeet Sucrose Content in Relation to Growing Season Climatic Conditions in the Northwest U.S.” *Journal of Sugar Beet Research*, January 1, 60–74. https://doi.org/10.5274/jsbr.54.1.60.

Kuhn, Max, and Hadley Wickham. 2020. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*. https://www.tidymodels.org.

Lattz, Dale, and Gary Schnitkey. 2023a. *Machinery Cost Estimates: Harvesting*. Technical Report. Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign. https://farmdoc.illinois.edu/assets/management/machinery-costs/harvest_operations_2023.pdf.

Lattz, Dale, and Gary Schnitkey. 2023b. *Machinery Cost Estimates: Summary*. Technical Report. Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign. https://farmdoc.illinois.edu/handbook/machinery-cost-estimates.

Lazarus, William F. 2024. *Machinery Cost Estimates*. Technical Report. University of Minnesota Extension, Department of Agricultural and Natural Resource Systems. https://z.umn.edu/machdata.

Li, Yan, Kaiyu Guan, Gary D. Schnitkey, Evan DeLucia, and Bin Peng. 2019. “Excessive Rainfall Leads to Maize Yield Loss of a Comparable Magnitude to Extreme Drought in the United States.” *Global Change Biology* 25 (7): 2325–37. https://doi.org/10.1111/gcb.14628.

Lobell, David B, David Thau, Christopher Seifert, Eric Engle, and Bertis Little. 2015. “A Scalable Satellite-Based Crop Yield Mapper.” *Remote Sensing of Environment* 164 (July): 324–33. https://doi.org/10.1016/j.rse.2015.04.021.

Lu, Junyu, Gregory J. Carbone, and Peng Gao. 2017. “Detrending Crop Yield Data for Spatial Visualization of Drought Impacts in the United States, 1895–2014.” *Agricultural and Forest Meteorology* 237–238 (May): 196–208. https://doi.org/10.1016/j.agrformet.2017.02.001.

Mahal, Navreet K., Michael J. Castellano, and Fernando E. Miguez. 2018. “Conservation Agriculture Practices Increase Potentially Mineralizable Nitrogen: A Meta-Analysis.” *Soil Science Society of America Journal* 82 (5): 1270–78. https://doi.org/10.2136/sssaj2017.07.0245.

Mississippi State University Department of Agricultural Economics. 2024. “Mississippi State Budget Generator.” https://www.agecon.msstate.edu/whatwedo/budgets/generator/index.php.

Molnar, Christoph. 2025. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 3rd ed. https://christophm.github.io/interpretable-ml-book.

National Academies of Sciences, Engineering and Medicine. 2017. “A Vision of NASS in 2025.” In *Improving Crop Estimates by Integrating Multiple Data Sources*. The National Academies Press. https://doi.org/10.17226/24892.

Oakes, J.C., M. Balota, D.L. Jordan, A.T. Hare, and A. Sadeghpour. 2020. “Peanut Response to Seeding Density and Digging Date in the Virginia-Carolina Region.” *Peanut Science* 47 (3): 180–88. https://doi.org/10.3146/PS20-16.1.

Oldfield, Emily E., Mark A. Bradford, Abigail J. Augarten, et al. 2022. “Positive Associations of Soil Organic Matter and Crop Yields across a Regional Network of Working Farms.” *Soil Science Society of America Journal* 86 (2): 384–97. https://doi.org/10.1002/saj2.20349.

Parton, William J., Melannie Hartmann, Dennis Ojima, and David Schimel. 1998. “DAYCENT and Its Land Surface Submodel: Description and Testing.” *Global and Planetary Change* 19 (1–4): 35–48. https://doi.org/10.1016/S0921-8181(98)00040-X.

PRISM Climate Group, Oregon State University. 2025. “PRISM Climate Data.” https://prism.oregonstate.edu.

Robertson, Philip G., Katherine L. Gross, Stephen K. Hamilton, et al. 2014. “Farming for Ecosystem Services: An Ecological Approach to Production Agriculture.” *BioScience* 64 (5): 404–15. https://doi.org/10.1093/biosci/biu037.

Sacks, William J., Delphine Deryng, Jonathan A. Foley, and Navin Ramankutty. 2010. “Crop Planting Dates: An Analysis of Global Patterns.” *Global Ecology and Biogeography* 19 (5): 607–20. https://doi.org/10.1111/j.1466-8238.2010.00551.x.

Schillerberg, Tayler, and Di Tian. 2023. “Changes in Crop Failures and Their Predictions with Agroclimatic Conditions: Analysis Based on Earth Observations and Machine Learning over Global Croplands.” *Agricultural and Forest Meteorology* 340: 109620. https://doi.org/10.1016/j.agrformet.2023.109620.

Sinclair, Jr., H. Raymond, Robert R. Dobos, Sharon W. Waltman, Karl W. Hipple, and Joseph V. Chiaretti. 2009. “A MODEL FOR EVALUATING AND COMPARING SOIL AND SITE FACTORS AFFECTING PRODUCTIVITY OF DISTURBED AND UNDISTURBED SIMILAR AND DISSIMILAR SOILS.” *Journal American Society of Mining and Reclamation* 2009 (1): 1248–75. https://doi.org/10.21000/JASMR09011248.

Smith, Amanda R. 2024. *2024 Estimate of Machinery Operations*. Technical Report. Department of Agricultural and Applied Economics, University of Georgia Cooperative Extension Service.

Soil Survey Staff. 2024. *Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States*. October 28. https://www.nrcs.usda.gov/resources/data-and-reports/gridded-soil-survey-geographic-gssurgo-database.

USDA NASS. 2024a. “Cropland Data Layer - Metadata.” https://www.nass.usda.gov/Research_and_Science/Cropland/metadata/meta.php.

USDA NASS. 2024b. “USDA NASS Quick Stats Agricultural Database.” https://quickstats.nass.usda.gov/.

USDA RMA, Risk Management Agency. 2024a. “Cause of Loss.” https://www.rma.usda.gov/tools-reports/summary-business/cause-loss.

USDA RMA, Risk Management Agency. 2024b. “Summary of Business.” https://www.rma.usda.gov/tools-reports/summary-of-business/state-county-crop-summary-business.

USGS. 2024. “Protected Areas Database of the United States (PAD-US) 4.” U.S. Geological Survey Gap Analysis Project (GAP). https://doi.org/10.5066/P96WBCHS.

Viator, Ryan P., Russell C. Nuti, Keith L. Edmisten, and Randy Wells. 2005. “Predicting Cotton Boll Maturation Period Using Degree Days and Other Climatic Factors.” *Agronomy Journal* 97 (2): 494–99. https://doi.org/10.2134/agronj2005.0494.

Volk, John M., Justin L. Huntington, Forrest S. Melton, et al. 2024. “Assessing the Accuracy of OpenET Satellite-Based Evapotranspiration Data to Support Water Resource and Land Management Applications.” *Nature Water* 2 (2): 193–205. https://doi.org/10.1038/s44221-023-00181-7.

Zhang, Jien, Peiyu Cao, and Chaoqun Lu. 2021. “Half-Century History of Crop Nitrogen Budget in the Conterminous United States: Variations Over Time, Space and Crop Types.” *Global Biogeochemical Cycles* 35 (10): e2020GB006876. https://doi.org/10.1029/2020GB006876.

Appendix

Appendix Table 1. Machinery cost data sources and state coverage

Data Source	State Covered
GA	FL, GA, SC
IL	AZ, CA, CO, DE, IA, ID, IL, IN, KS, KY, MD, MI, MO, MT, NC, ND, NE, NJ, NM, NY, OH, OK, OR, SD, TX, UT, VA, WA, WI, WV, WY
MS	AL, AR, MS, TN
LA	LA
MN	MN

Note: Data sources are: Georgia (Smith 2024), Illinois (Lattz and Schnitkey 2023b), Louisiana (Deliberto and Hilbun 2023), Minnesota (Lazarus 2024), and Mississippi (Mississippi State University Department of Agricultural Economics 2024).

Appendix Table 2. Final data availability for each selected crop, including the number of counties, states, and total observations used in the analysis.

Crop	Description[1]	Yield Analysis			Drought Analysis	
		Number of Counties	Number of States	Number of Observations	Number of States	Number of Observations
Corn	Grain	1595	36	34282	35	5380
Soybeans		1371	30	30107	30	4678
Wheat	Winter Wheat	1081	32	22041	29	2454
Cotton	Upland cotton	308	17	6730	17	1397
Oats		320	14	6065	13	647
Sorghum	Grain	189	10	3540	9	1324
Barley		147	14	2751	13	422
Peanuts		96	8	1929	8	335
Sugar beets		81	9	1499	NA[2]	NA
Tobacco	Air-cured light burley (Type 31)	60	2	1024	NA	NA

Note: [1] Detailed short description used to extract data from USDA NASS (USDA NASS 2024b). [2] Sugar beets and tobacco had a low number of observations and were therefore excluded from the drought analysis model.

Appendix Table 3. Variable used in the final models for yield and drought analyses.

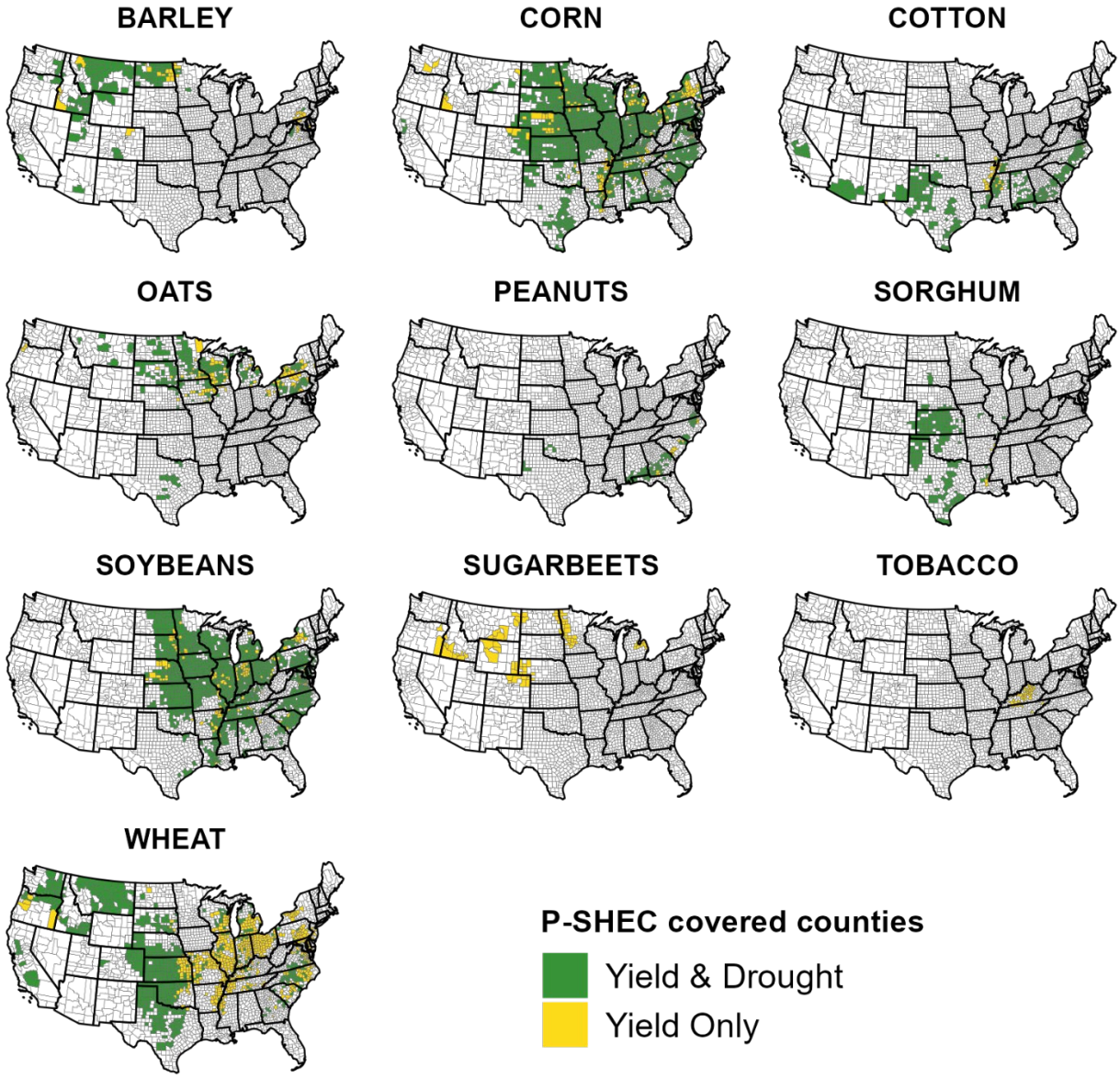
Analysis	Dependent variable	Independent variables
Yield analysis	Crop de-trended yield	State, soil organic matter (%), irrigation area (%), clay (%), slit (%), sand (%), available water capacity (%), available water storage (mm), wilting point (%), cation exchange capacity (meq 100 g soil-1), H3O+ concentration (mol), nitrogen input (kg N ha-1 year-1), growing season precipitation, growing season mean temperature, growing degree days, drainage class, slope
Drought analysis	Yield change under drought	State, soil organic matter (%), mean irrigation area (%), clay (%), sand (%), H3O+ concentration (mol), nitrogen input (kg N ha-1 year-1), growing season precipitation, growing season mean temperature, growing degree days, drainage class, slope, growing season DSCI

Appendix Table 4. The final model evaluation metrics on test data for each model

Analysis	Crop Model[A]	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)	R^2
Yield analysis	Corn	1.04	0.79	0.76
	Soybeans	0.30	0.23	0.78
	Wheat	0.44	0.34	0.85
	Cotton	7.71	5.96	0.76
	Oats	0.58	0.44	0.60
	Sorghum	0.78	0.63	0.61
	Barley	0.52	0.41	0.88
	Peanuts	22.9	18.5	0.53
	Sugar beets	0.16	0.12	0.63
	Tobacco	9.60	6.95	0.56
Drought analysis	Corn, Soybeans, Wheat, Cotton, Oats, Sorghum, Barley, Peanuts	14.57	11.16	0.37

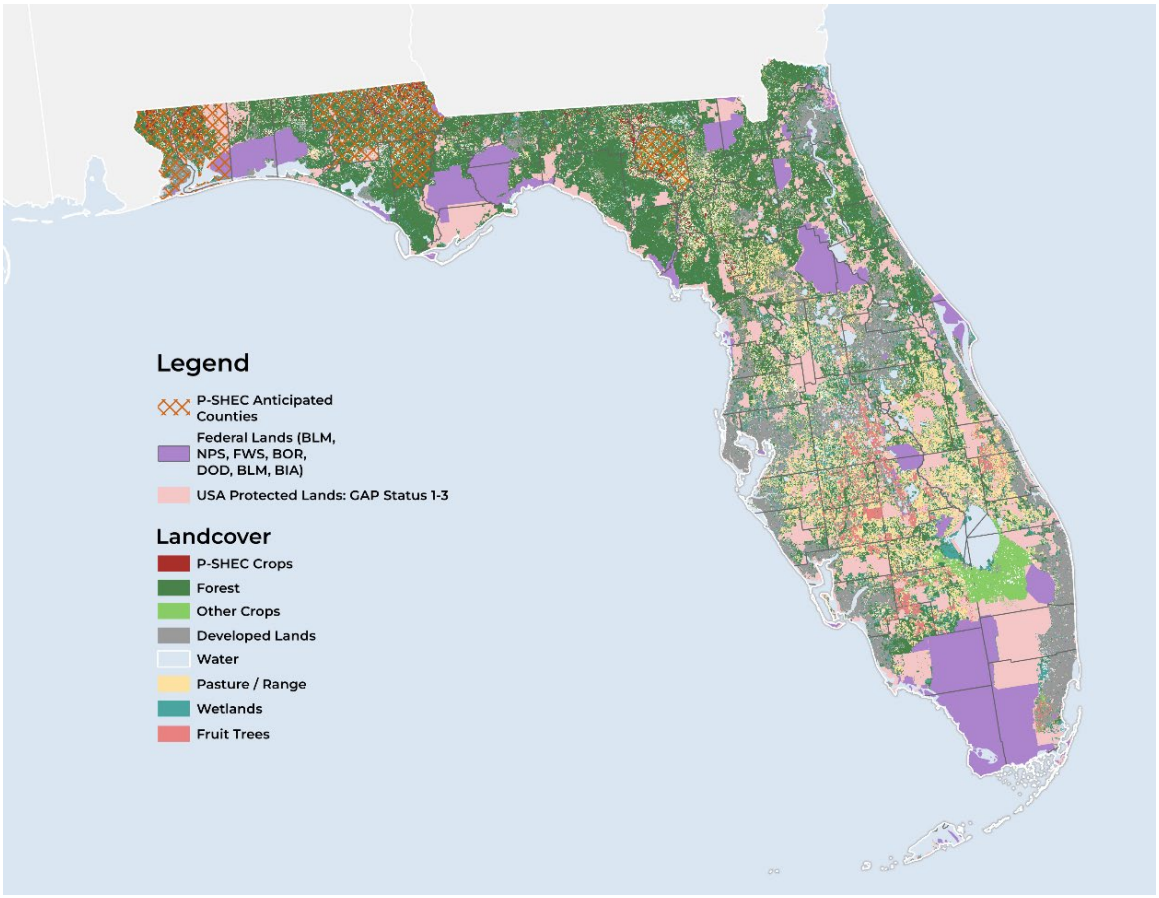
[A]XGBoot model was used for all individual crop models and the drought model.

Appendix Figure 1. Geographic coverage of P-SHEC across U.S. counties for 10 row crops.



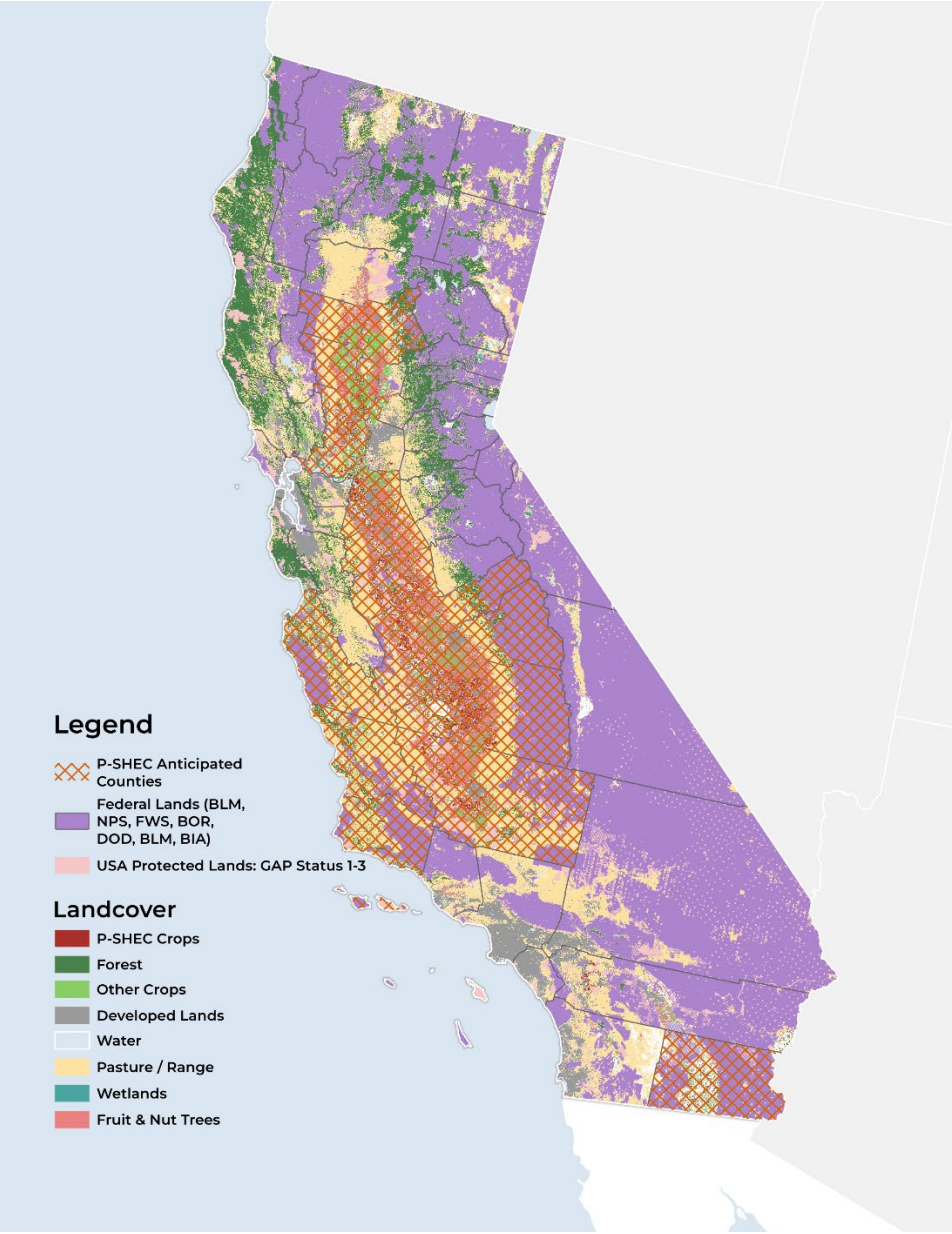
Note: Shaded areas represent counties covered by P-SHEC tool. Green indicates that both long-term yield predictions and drought yield loss estimates are available for the crop in that county. Yellow indicates that only long-term yield predictions are available. Yield data is from USDA National Agricultural Statistics Service (USDA NASS 2024b).

Appendix Figure 2. Map of P-SHEC coverage in Florida, showing included counties and land cover types.



Source: USDA Cropland Data Layer (USDA NASS 2024a) and Protected Areas Database of the U.S. (USGS 2024)

Appendix Figure 3. Map of P-SHEC coverage in California, showing included counties and land cover types.



Source: USDA Cropland Data Layer (USDA NASS 2024a) and Protected Areas Database of the U.S. (USGS 2024)

Appendix File 1. All raw data sources used in the P-SHEC tool

Data	Data Sources	Note
	Datasets used in machine learning models	
Soil organic matter changes	Comet-Planner https://comet-planner-domestic.uc.r.appspot.com/	Version 3.1, Build 1, updated in Dec. 2023
Crop yield	National Agricultural Statistics Service (NASS) https://quickstats.nass.usda.gov/	County-level yield data in the survey years of 2000-2023
Irrigation area	National Agricultural Statistics Service (NASS) https://quickstats.nass.usda.gov/	County-level percentages of crop-growing acres irrigated with records from the census years of 2002, 2007, 2012, 2017, and 2022
Crop progress weekly data	National Agricultural Statistics Service (NASS) https://quickstats.nass.usda.gov/	State-level crop weekly progress reports with records from the survey years 2000-2023
Crop calendar data	Sacks et al. (2010) (https://sage.nelson.wisc.edu/data-and-models/datasets/crop-calendar-dataset/)	State-level crop-specific calendars defined as the period between planting and harvest
Soil data	Gridded Soil Survey Geographic (gSSURGO) Database	
Crop frequency layer	National Agricultural Statistics Service (NASS) https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php	2024 national Crop Frequency Layer data were used for corn, cotton, soybeans, and wheat
Cropland data layer	National Agricultural Statistics Service (NASS) https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php	National 30-meter Cropland Data Layer data from 2008 - 2023 were used for oats, sorghum, barley, peanuts, sugar beets, and tobacco
National cultivated layer	National Agricultural Statistics Service (NASS) https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php	2024 national cultivated layer was used
Climate data	PRISM datasets https://prism.oregonstate.edu/	Daily precipitation, mean, minimum, and maximum temperature data from 2000–2023 at a 4 km spatial resolution
Nitrogen data - 1	Zhang et al., (2021)	County-level mean nitrogen (N) input data for corn, winter wheat, cotton, barley, and sorghum across the years from 2000 to 2019
Nitrogen data - 2	Nutrient Use Geographic Information System (NuGIS) https://nugis.tfi.org/	County-level N fertilizer application data was used for oats, sugar beets, and tobacco
Crop insurance data	1. Summary of Business (SOB): https://www.rma.usda.gov/tools-	Risk Management Agency data from 2000-2023

	reports/summary-of-business/state-county-crop-summary-business 2. Cause of loss historical (COL): https://www.rma.usda.gov/tools-reports/summary-business/cause-loss	
Drought index	Drought Severity and Coverage Index (DSCI) https://droughtmonitor.unl.edu/About/AbouttheData/DSCI.aspx	DSCI data from 2000-2023

Datasets used in the economic analysis

Equipment cost	1. University of Georgia Cooperative Extension Service: "2024 Estimated Cost per Acre of Specified Field Operations in Georgia". Revised and updated by A. R. Smith (February 2024). Available: agecon.uga.edu/content/dam/caes-subsite/ag-econ/documents/extension/budgets/2024-budgets/2024-Estimate-Machinery-Operations.pdf 2. Louisiana State University Department of Agricultural Economics and Agribusiness: "Projected Costs and Returns Crop Enterprise Budgets for Corn Production in Louisiana, 2023". Agricultural Economics Information Report Series No. 365 by M. A. Deliberto and B. M. Hilbun. Available: lsuagcenter.com/~/media/system/6/9/8/f/698f02b591ffe4242d3c86a3186e5177/2023 corn enterprise budgets final pdf.pdf 3. University of Minnesota Extension: "Machinery Cost Estimates". W.F. Lazarus. July 2024. Available: https://wlazarus.cfans.umn.edu/william-f-lazarus-farm-machinery-management 4. Mississippi State University Department of Agricultural Economics: "Mississippi State Budget Generator, V. 6.0". The generator was used with 2024 Row Crop Budget files. Available: Budget Generator | Agricultural Economics | Mississippi State University 5. University of Illinois at Urbana-Champaign Department of Agricultural and Consumer Economics/University of Illinois Extension: "Machinery Cost Estimates: Forage Field Operations". September 2023. Available:	

	https://farmdoc.illinois.edu/handbook/forage-operations 6. University of Illinois at Urbana-Champaign Department of Agricultural and Consumer Economics/University of Illinois Extension: "Machinery Cost Estimates: Tractors". September 2023. Available: https://farmdoc.illinois.edu/handbook/tractor-costs 7. University of Illinois at Urbana-Champaign Department of Agricultural and Consumer Economics/University of Illinois Extension: "Machinery Cost Estimates: Field Operations". September 2023. Available: https://farmdoc.illinois.edu/handbook/field-operations 8. University of Illinois at Urbana-Champaign Department of Agricultural and Consumer Economics/University of Illinois Extension: "Machinery Cost Estimates: Harvesting". September 2023. Available: https://farmdoc.illinois.edu/handbook/harvest-operations	
Fertilizer prices	Iowa State University Extension and Outreach, Ag Decision Maker: "Estimated Costs of Crop Production in Iowa-2024". File A1-20. Prepared by A. Plastina. Available: extension.iastate.edu/agdm/crops/pdf/a1-20-2024.pdf	
Crop Prices	National Agricultural Statistics Service (NASS)	Monthly average prices received for each crop included in the analysis were compiled and averaged to calculate the annual average price received for each crop