

What Is Driving Non-Reported National Agricultural Statistics Service Yields?



**By W. Hence Duncan,
Christopher N. Boyer,
S. Aaron Smith, and
Eunchun Park**

W. Hence Duncan is a Graduate Research Assistant in the Department of Agricultural and

Resource Economics at University of Tennessee. Christopher N. Boyer is a Professor in the Department of Agricultural and Resource Economics at University of Tennessee; contact him at cboyer3@utk.edu. S. Aaron Smith is an Associate Professor in the Department of Agricultural and Resource Economics at University of Tennessee. Eunchun Park is an Assistant Professor in the Department of Agricultural Economics and Agribusiness, University of Arkansas.

Acknowledgements

We thank the editor and reviewers for providing feedback to improve this manuscript. This research was supported by Agriculture and Food Research Initiative Competitive Grant no. 2021-67023-33819 from the United States Department of Agriculture National Institute of Food and Agriculture.

Abstract

The number of non-reported county yields by the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) is increasing. This article explores factors that impact county corn

and soybean yields that are not reported by USDA NASS. Factors such as county, land coverage, and average farm size are used to explain the likelihood of a yield being reported. We find that counties that have a high number of acres concentrated in a few farms may not have a NASS yield reported due to NASS reporting requirements.

INTRODUCTION

The United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) is a government agency primarily focused on collecting and publishing agricultural statistics. One of the most utilized statistics collected by USDA NASS are annual county crop yield averages, which are frequently used by numerous groups such as government agencies, researchers, market analysts, and producers to provide much needed information about U.S. food and fiber production for policy development, farm insurance programs, farm disaster payment calculations, and decision-making, modeling the impact of several factors on production.

USDA NASS yields are collected as a part of an annual survey administered by NASS, which includes all states except Alaska and Hawaii. The county yield estimates are partially determined by self-reported average field-based estimates from producers. The survey, which starts in November and ends mid-January, is collected through mail, phone interviews, in-person interviews, and electronically via email, with most responses collected through phone calls. Statisticians review the survey results to identify and analyze outliers, such as extreme values, data entry errors, or inconsistencies with historical patterns, before inputting the data into a computer system for further analysis. Once outliers are corrected, the data is summarized by county and released (USDA NASS, 2023a).

However, over the last 15 years, there has been a decline in survey responses and the number of county

yields being reported, which has resulted in growing concerns about yield accuracy (Johansson et al., 2017; Schnepf, 2017). The declining response rates create a challenge for USDA because it will not report yield data if the identity of the respondent can be revealed. Thus, NASS will not publish data for counties where small numbers of producers or acres in production may disclose the business of individual producers. NASS requires at least 30 producers or 25% of harvested acres to be reported to release a yield for a county. Figure 1 shows the increasing number of county level crops yields not reported by USDA NASS.

Along with researchers using this data (Lusk, 2016), the USDA relies on NASS surveys for a variety of payment calculations and policy recommendations. For example, NASS yields have been used to calculate Agricultural Risk Coverage-County (ARC-CO) payments, which compensate farmers when actual county crop revenue falls below guaranteed levels. Additionally, NASS yield data has been utilized in determining indemnity payments under the Federal Crop Insurance Program (Rejesus, Goodwin, and Coble, 2010) and in calculating farm disaster payments for programs like the Wildfire and Hurricane Indemnity Program (WHIP), where yield losses due to natural disasters are compared to historical averages to establish payment amounts. The rising number of missing NASS yields resulted in the Agricultural Improvement Act of 2018, switching from using NASS yields as the preferred yield for the ARC-CO payment calculations to USDA Risk Management Agency (RMA) yields as the primary yield used to calculate payments.

Rejesus, Coble, and Knight (2010) found that when reference yields for RMA were based on NASS data, it was not a true representation of yield information for the producers enrolled in the Federal Crop Insurance Program. As these yields were not updated, the problem increased due to technology in the agricultural sector improving (Rejesus, Coble, and Knight 2010). As a result of the study, it was recommended to use a reference yield calculation based on RMA yield (Rejesus, Coble, and Knight, 2010). Similarly, Li et al. (2020) examined the variability and reliability of a yield estimator based on NASS yields compared to RMA data for corn, soybeans, and wheat. Their paper used NASS and RMA yields from the years 1991-2015 to examine the feasibility of using RMA yields, rather than NASS yields to calculate ARC-CO payments. They found no major difference between the two yield values. Using RMA yield data for ARC-CO payments also resulted in less variability between nearby counties, possibly because RMA data reflects consistent, field-level records from insured farms.

While studies have attempted to estimate missing yield (Ishee, 2020; Park, Harri, and Coble, 2022), no study has attempted to try to understand the factors driving the missing yields. One hypothesis is that farm consolidation could result in fewer farms within a county, which might not meet the threshold for reporting yields. Therefore, the objective in this study is to determine if a county landscape and average size farm impact the likelihood of a NASS yield being reported. The results could directly impact how NASS could adjust its reporting requirement to adjust for larger and fewer farms in a county.

DATA

Data on county yield values for corn and soybeans across the U.S. was sourced from USDA NASS from 2011 to 2022 (USDA NASS, 2024). Additionally, land cover information was obtained from USDA CroplandCROS for all states (USDA, 2024). This land cover data from USDA CroplandCROS was then integrated with the USDA NASS yield data to form separate datasets for corn and soybeans. These datasets were refined by removing records of the observed crops that occupied less than a thousand acres in a county, as per USDA CroplandCROS data. Subsequently, data on the average farm size for each county, crop, and year was incorporated, derived from the USDA RMA summary of business statistics data (USDA RMA, 2024).

From USDA CroplandCROS, percentages of a county landscape by different land cover classification were calculated for the following classifications: soybeans, corn, cotton, pasture/hay (combined variable of the two classifications), developed (i.e., residential, commercial, or industrial uses), and forest. These percentages were calculated by taking the classifications and dividing them by the sum of all the classifications for a county. This calculation was done for each county by year. A USDA RMA summary of business statistics data was used to calculate a proxy of average farm size by dividing the total insured acres in a county for a given year by the number of insurance policies issued in that county during the same year, serving as an indirect measure of average farm size. A county was marked having a missing NASS yield value if the county had a thousand acres or more of the observed crop in a year according to USDA CroplandCROS and did not report a NASS yield that year.

Figure 2 displays the average corn acres planted by county from 2011-2022 for the U.S. from USDA CroplandCROS data. This figure shows that counties with greater corn acres are concentrated in the upper

Midwest and Northern Plains. Figure 3 displays the average soybean acres planted by county from 2011-2022 for the U.S. Soybean acres are concentrated along the Mississippi River, Northern Plains, and Midwest. Figure 4 shows the percentage of years where a NASS corn yield was reported by county from 2011-2022 for the U.S. A visual inspection between Figure 2 and Figure 4 suggests there is a relationship in yield report rate and corn acres planted. Figure 5 shows the percentage of years where a NASS soybean yield was reported by county from 2011-2022 for the US. The areas where counties have higher report rates are along the Mississippi River, Northern Plains, and Midwest, and this is also where soybean planted acres are more concentrated.

Table 1 displays the summary statistics of the variables observed for corn and used in the model, where the variables are the percent of the county that is in each crop, and average farm size is scaled. The summary stats indicate forest, and pasture/hay provide the most land cover in counties used for the corn analysis. On average, corn acres cover about 12% of the land within a county, and soybeans cover 11% of the land within a county; the average corn farm size was 134 acres. The summary statistics of the variables observed for soybeans are displayed in Table 2. The percentage of land cover in a county was similar for soybeans, with both corn and soybeans covering 13% of the land area for the soybean data. The average soybean farm size was 148 acres, and according to 2022 USDA Census data, the average harvested crop farm was 158 acres in 2022 (USDA National Agricultural Statistics Service, 2022), which is slightly higher than our average in our data.

METHODS

A logit model was utilized to determine how a county's landscape and the average size of corn and soybean farms influence the probability of a NASS yield being reported. A logit model is a type of statistical analysis used to predict the likelihood of an outcome when the dependent variable is binary. In this case, we define a NASS yield as being equal to 1 and 0 if it is not reported. The findings are expressed in terms of odds ratios, which quantify how a change in an independent variable affects the odds of a NASS yield being reported, either increasing or decreasing these odds by a specific percentage. This approach is particularly suited for binary outcomes like this model, where the independent variables include the percentage of the county that is in soybean acres, in corn acres, in cotton

acres, in pasture and hay acres, in developed acres, and in forest acres; the average farm size proxy scaled by a thousand; the average farm size proxy squared scaled by a thousand; and a fixed effect for state and year. The model and average marginal effects were calculated in R using the margins package.

RESULTS

The results from the logistic regression analysis aimed to explore the impact of county landscape and average farm size on the likelihood of a NASS yield being reported are summarized in Table 3.

In the case of corn, the proportion of county landscape used for corn production exhibited a strong positive relationship ($p < .001$) with the likelihood of a NASS corn yield being reported. Additionally, the proportion of county landscape dedicated to corn production, along with percentages of landscape in cotton and pasture/hay, average farm size, and average farm size squared, all showed significant relationships ($p < .001$). Similarly, in the soybean model, the proportion of county landscape allocated to soybean production had a significant relationship ($p < .001$). In both models, the percentage of the county that was developed was not significant for a NASS yield being reported. In both models, state and year fixed effects were incorporated into the model to control variations across different states and years. The average marginal effects are shown in Table 4. For example, a 1% increase in the percentage of the landscape in corn results in a 0.82% increase in the likelihood of a NASS corn yield. For the soybean model, a 1% increase in the percentage of the landscape in soybeans results in a 0.86% increase in the likelihood of reporting a NASS soybean yield.

In both models, the proxy of average farm size and the proxy of average farm size squared were found to be significant, indicating a positive influence on the likelihood of reporting a NASS yield until a certain threshold. Figure 6 illustrates the predicted probability curve, showing that for corn, the likelihood starts decreasing after 228.33 acres, and for soybeans, after 253.13 acres. This could indicate farm consolidation could negatively impact the likelihood of NASS yields being reported, which would suggest that farms continue to consolidate USDA NASS and revisit their criteria of reporting yields in a county to avoid not reporting counties with a reportable level of acres but not enough farms.

CONCLUSION

The increasing number of counties non-reporting a USDA NASS yield is increasing and is causing concern for researchers, government agencies, and market participants. This research seeks to discover factors that are associated with corn and soybean yields not being reported. By investigating the impact of county landscape and average farm size on the likelihood of reporting NASS yields for corn and soybeans, this study contributes valuable insights to the existing literature.

The results of the logit model indicate that a county's landscape impacts the likelihood of a NASS yield being reported. Counties that have a higher percentage of their landscape in agricultural production are more likely to have a NASS yield reported. It also indicates that average farm size plays a role in the likelihood of a NASS yield being reported as well. Counties that have a high number of acres concentrated in a few farms may not have a NASS yield reported, due to large farm size resulting in the county having fewer than 30 producers or 25% of harvested acres responding to the survey. This could continue to be an issue as we continue to see farm consolidation across the U.S. Having a county that produces many acres for either crop or not reporting a yield could result in a reporting bias, which could occur because gaps in a county NASS yield history impact the historical county average. One county not receiving enough survey responses could also impact state and national averages as well.

Addressing the challenges posed by the increase in non-reported yields is crucial for policymakers, researchers, and market participants to make informed decisions and foster a more resilient and sustainable agricultural sector. Continued efforts to improve data collection methods and enhance the accuracy of yield reporting are imperative to ensure the reliability of NASS data.

REFERENCES

- Ishee, Z.S. 2020. "Recovering Missing Yield Values: Ramifications for the USDA's ARC-CO Program." Master's Thesis. Department of Agricultural Economics at Mississippi State University. <https://scholarsjunction.msstate.edu/td/3737/>.
- Johansson, R., A. Effland, and K. Coble. 2017. "Falling Response Rates to USDA Crop Surveys: Why It Matters." <https://farmdocdaily.illinois.edu/2017/01/falling-response-rates-to-usda-crop-surveys.html>.
- Li, X., Z. Guo, Y. Huang, and X. Zheng. 2020. "Comparing Survey-Based and Program-Based Yield Data: Implications for the U.S. Agricultural Risk Coverage-County Program." *Geneva Papers on Risk and Insurance Issues and Practice* 45(1): 184–202. <https://doi.org/10.1057/s41288-019-00148-4>.
- Lusk, J. 2016. "From Farm Income to Food Consumption: Valuing USDA Data Products." Council on Food, Agricultural, and Resource Economics (C-FARE) reports. <https://ageconsearch.umn.edu/record/266593/?ln=en&v=pdf>.
- Park, E., A. Harri, and K.H. Coble. 2022. "Estimating Crop Yield Densities for Counties with Missing Data." *Journal of Agricultural and Resource Economics* 47(3): 634–S10. 10.22004/ag.econ.313319.
- Rejesus, R., B. Goodwin, K. Coble, and T. Knight. 2010. "Evaluation of the Reference Yield Calculation Method in Crop Insurance." *Agricultural Finance Review* 70: 427–445. <https://doi.org/10.1108/00021461011088530>.
- Schnepf, R. 2017. *NASS and U.S. Crop Production Forecasts: Methods and Issues*. Specialist in Agricultural Policy, Congressional Research Service. <https://sgp.fas.org/crs/misc/R44814.pdf>.
- USDA RMA. 2024. *Summary of Business*. <https://www.rma.usda.gov/>.
- USDA NASS. 2022. *Census of Agriculture: 2022 Full Report* 1(2). https://www.nass.usda.gov/Publications/AgCensus/2022/Full_Report/Volume_1,_Chapter_2_US_State_Level/st99_2_001_001.pdf.
- USDA NASS. 2023a. *Survey Methods*. https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Agricultural_Yield/index.php.
- USDA NASS. 2023b. *Quick Stats*. www.nass.usda.gov/AgCensus.

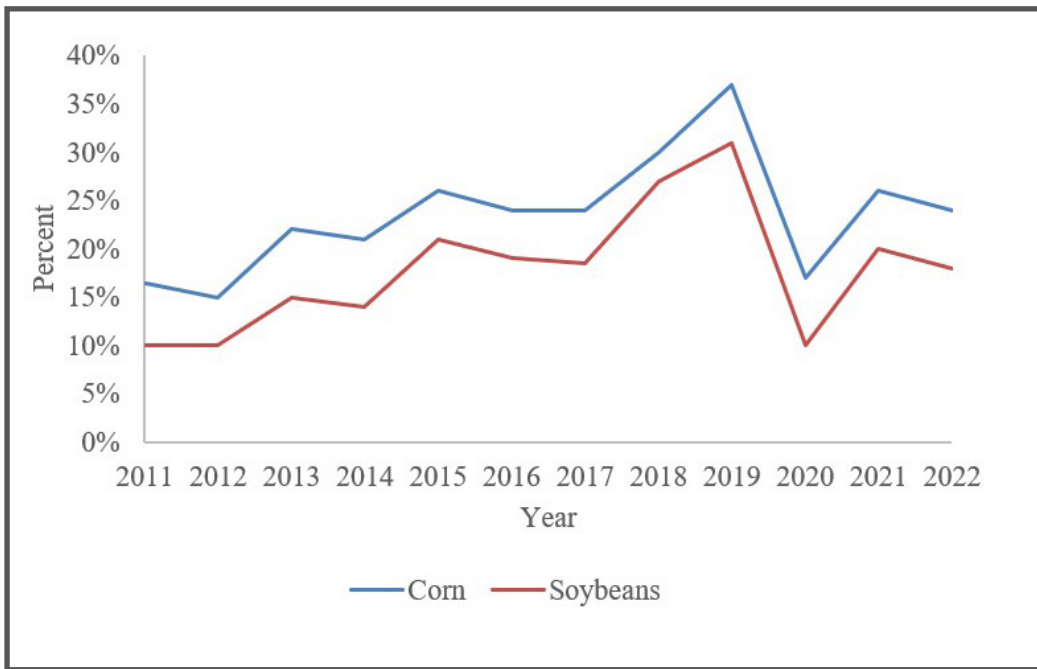


Figure 1. Percentage of missing NASS county corn and soybean yields response, 2011–2022

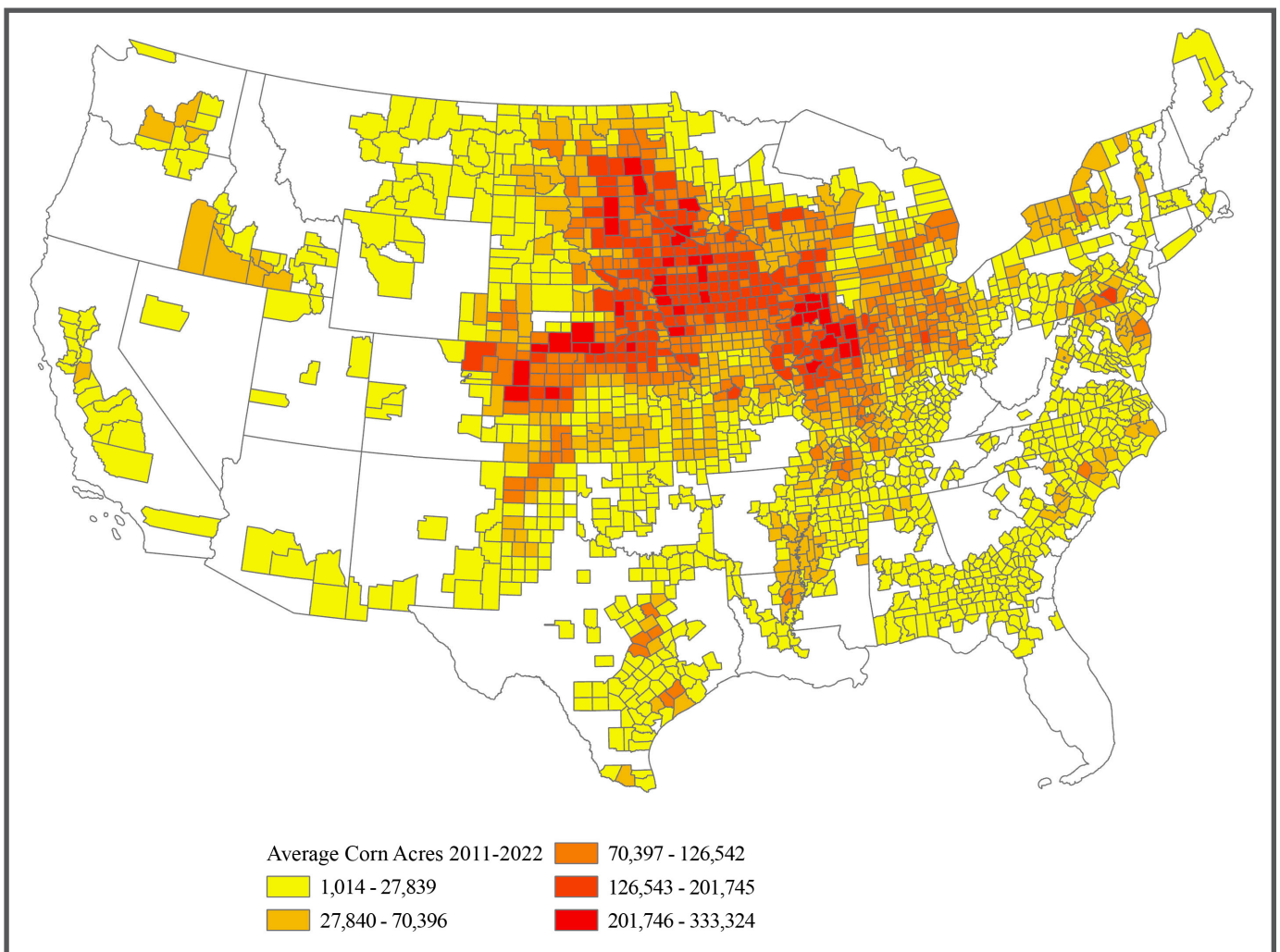


Figure 2. Average planted corn acres 2011–2022 USDA CroplandCROS

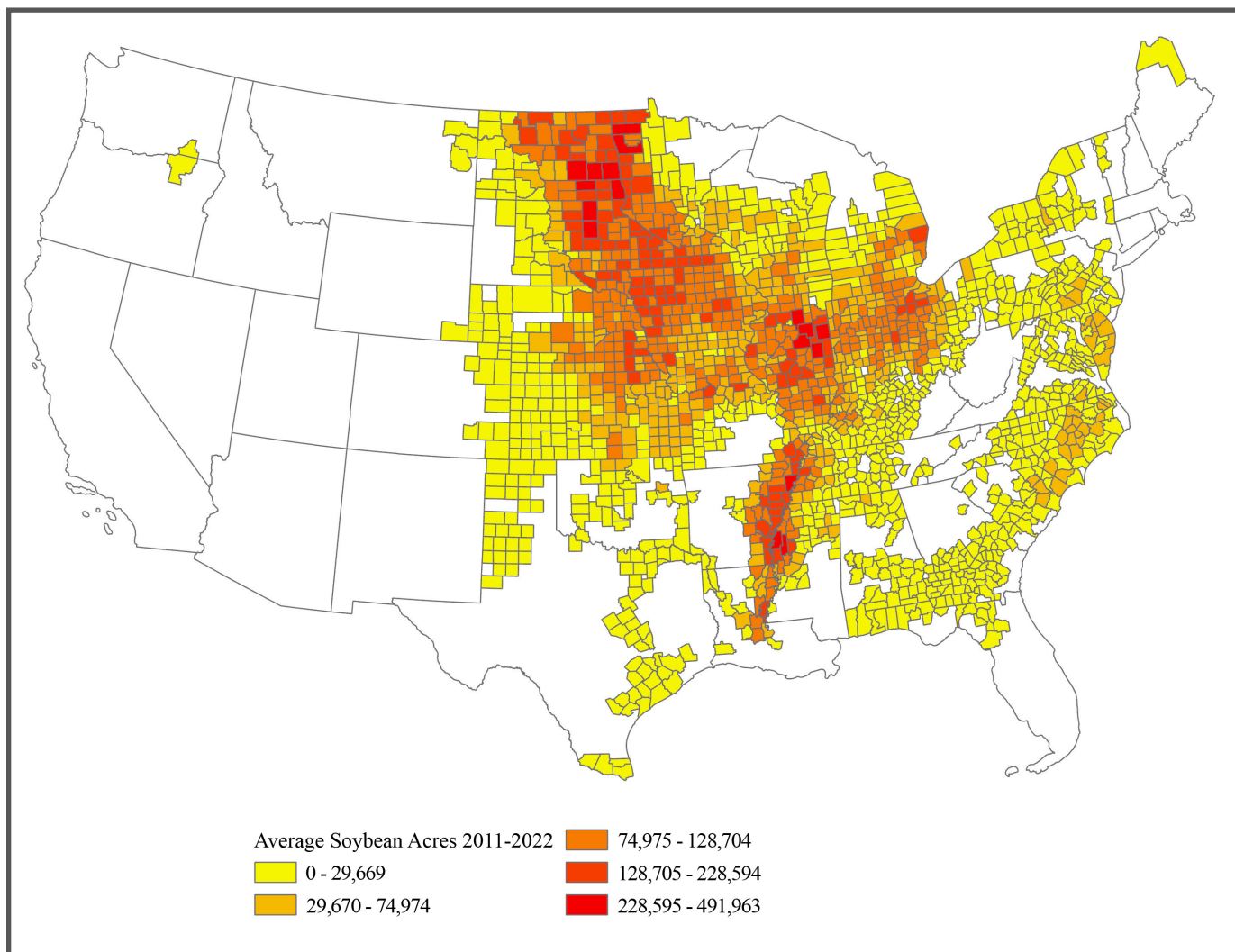


Figure 3. Average planted soybean acres 2011–2022 USDA CroplandCROS

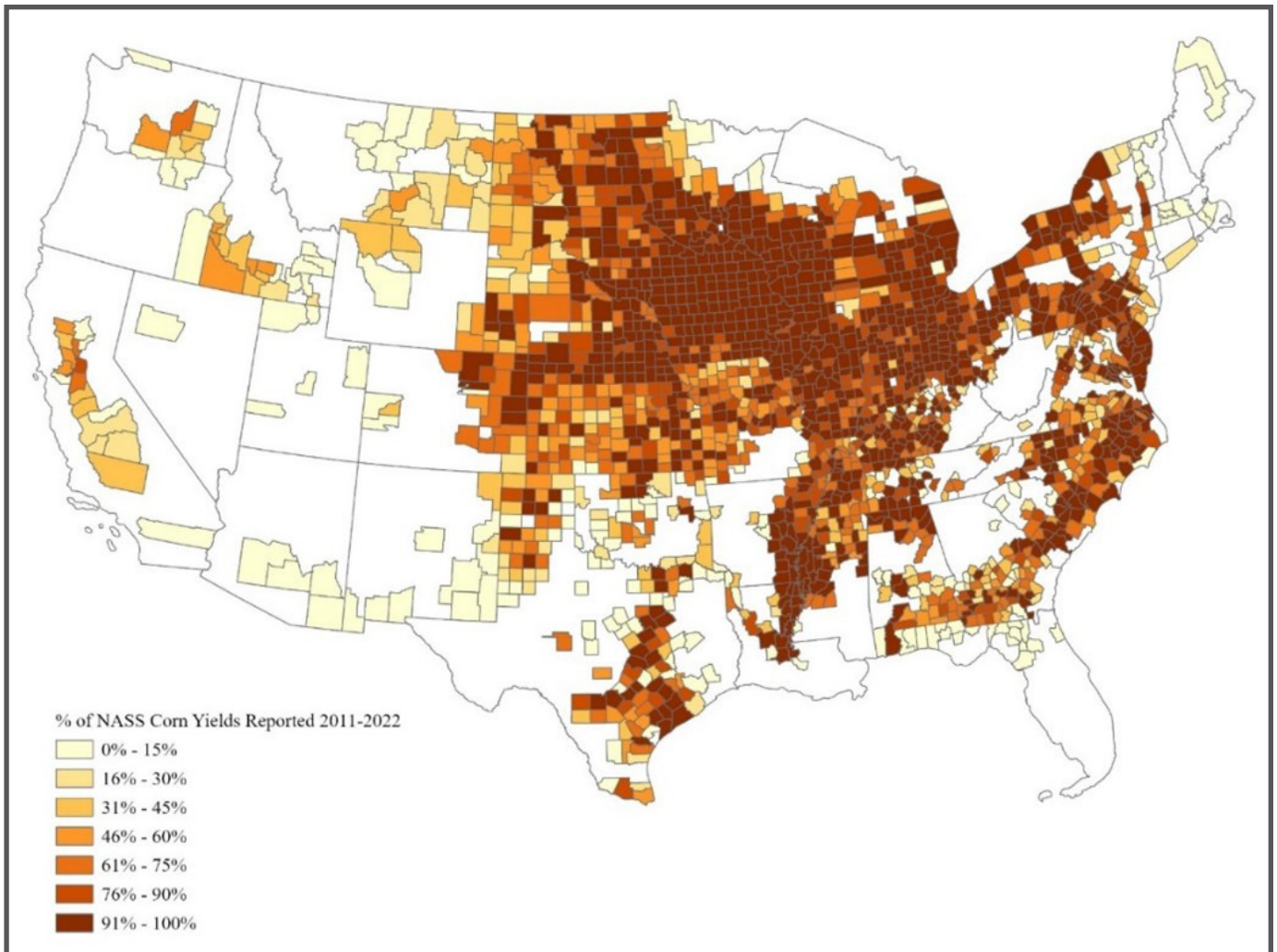


Figure 4. Percentage of NASS corn yields reported 2011–2022

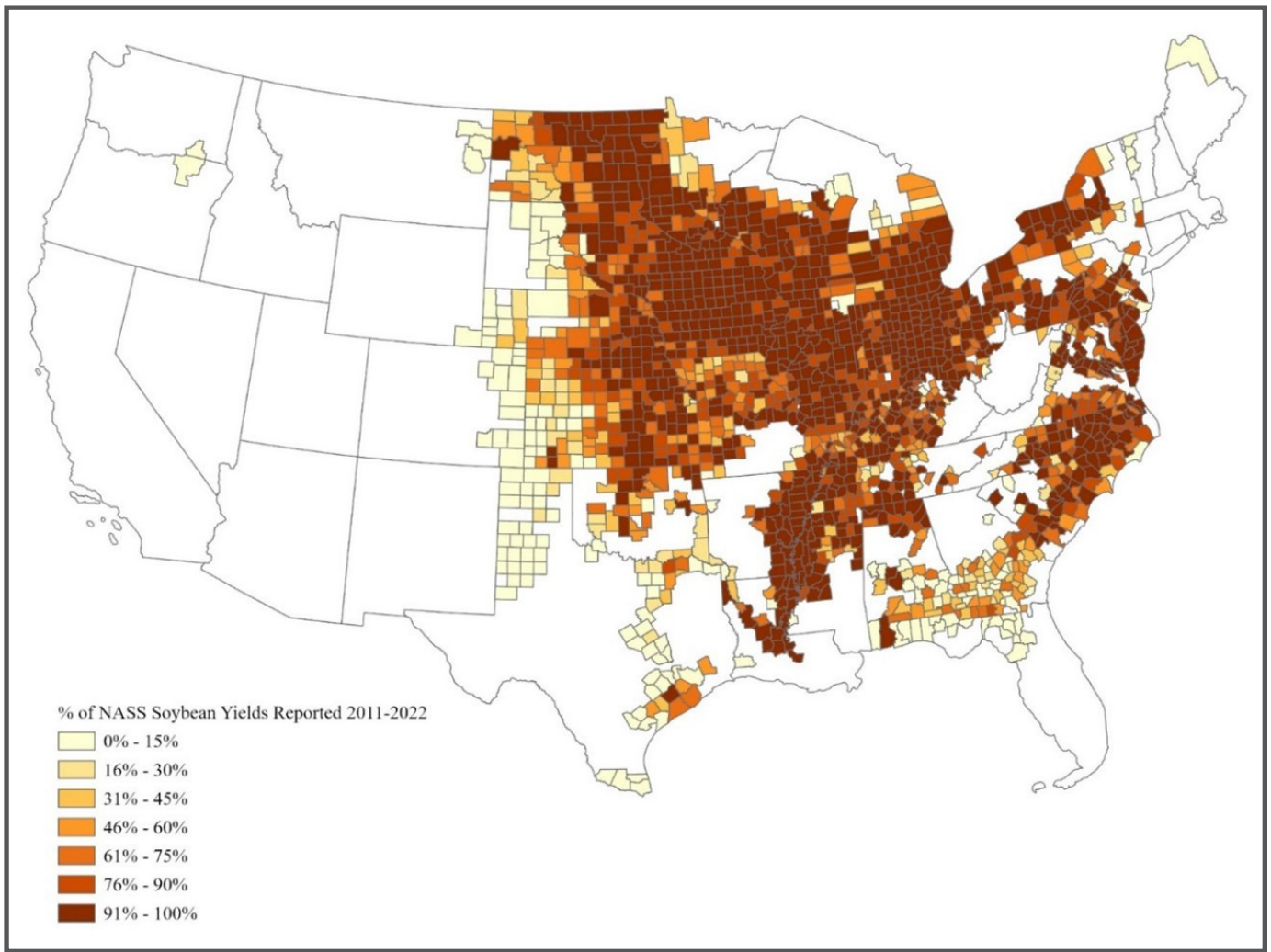


Figure 5. Percentage of NASS soybean yields reported 2011–2022

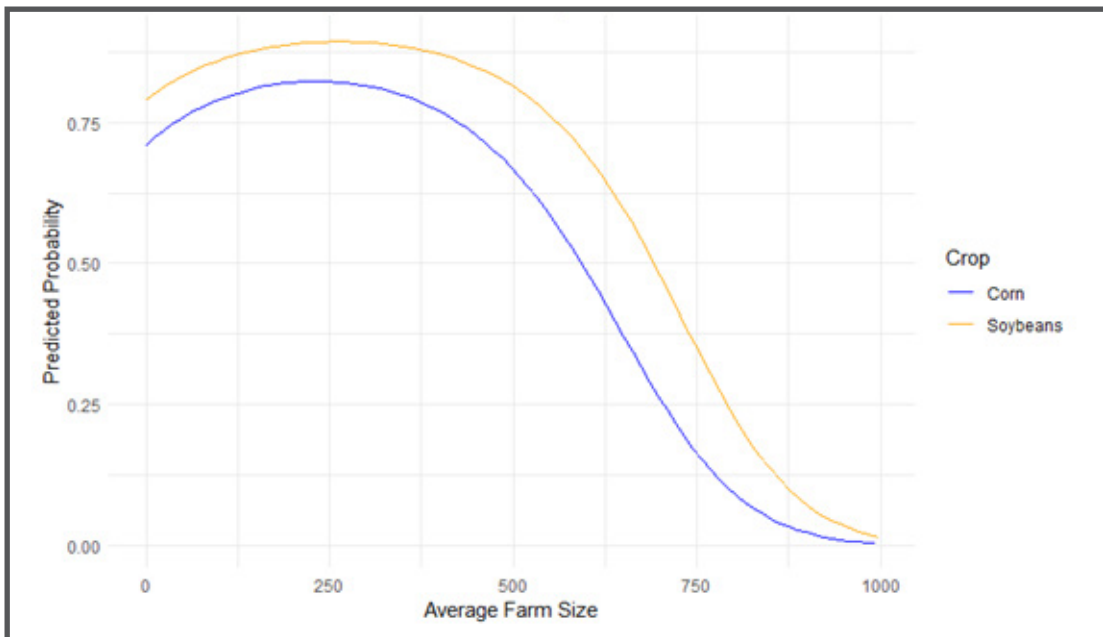


Figure 6. Predicted probability of a NASS yield being reported in a county by farm size

Table 1. Summary Statistics of Independent Variables for Corn Model from 2011 to 2022

Variable	Average	Standard Deviation	Minimum	Maximum
Percentage of County in Beans	0.11	0.13	0.00	0.68
Percentage of County in Corn	0.12	0.14	0.00	0.63
Percentage of County in Cotton	0.01	0.05	0.00	0.71
Percentage of County in Pasture or Hay	0.22	0.19	0.00	0.97
Percentage of County that is Developed	0.08	0.08	0.00	0.85
Percentage of County in Forest	0.23	0.22	0.00	0.83
Average Farm Size (in 1,000 acres)	0.13	0.08	0.00	1.17

Table 2. Summary Statistics of Independent Variables for Soybean Model from 2011 to 2022

Variable	Average	Standard Deviation	Minimum	Maximum
Percentage of County in Beans	0.13	0.13	0.00	0.68
Percentage of County in Corn	0.13	0.14	0.00	0.63
Percentage of County in Cotton	0.01	0.04	0.00	0.70
Percentage of County in Pasture or Hay	0.20	0.17	0.00	0.97
Percentage of County that is Developed	0.08	0.08	0.00	0.85
Percentage of County in Forest	0.24	0.22	0.00	0.83
Average Farm Size (in 1,000 acres)	0.14	0.09	0.00	1.33

Table 3. Logit Model Results for NASS Yield Reporting by County Landscape

Variable	Corn	Soybeans
	Estimate	Estimate
Intercept	1.013***	1.31***
Percentage of County in Beans	2.189***	7.486***
Percentage of County in Corn	6.144***	1.341**
Percentage of County in Cotton	1.878***	3.886**
Percentage of County in Pasture or Hay	0.918***	0.5*
Percentage of County that is Developed	-0.013	0.511
Percentage of County in Forest	-0.065	-0.079
Average Farm Size (in 1,000 acres)	5.48***	5.569***
Average Farm Size (in 1,000 acres) Squared	-0.012***	-0.011***

*, **, *** represent significance at the 10%, 5%, and 1% levels, respectively.

Table 4. Average Marginal Effects of NASS Yield Reporting

Variable	Corn	Soybeans
	Estimate	Estimate
Percentage of County in Beans	0.2904***	0.8602***
Percentage of County in Corn	0.8149***	0.1613**
Percentage of County in Cotton	0.2492***	0.5548**
Percentage of County in Pasture or Hay	0.1217***	0.0629*
Percentage of County that is Developed	-0.017	0.030
Percentage of County in Forest	-0.0086	-0.015
Average Farm Size (in 1,000 acres)	0.7269***	0.00007***
Average Farm Size (in 1,000 acres) Squared	-0.016***	-0.0013***

*, **, *** represent significance at the 10%, 5%, and 1% levels, respectively.