Journal of Data Science 22 (3), 423–435
July 2024

DOI: 10.6339/24-JDS1146
Data Science in Action

Evaluating a Method for Georeferencing Agricultural Fields☆

Robert L. Emmet[1,*], Kevin Hunt[1], Rachael Jennings[1], Kara Daniel[1], and Denise A. Abreu[1]

[1]*USDA National Agricultural Statistics Service, Washington, DC 20250, United States of America*

Abstract

The US Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) has begun a modernization effort to supplement survey data with non-survey data to improve estimation of agricultural quantities. As part of this effort, NASS has begun georeferencing farms on its list frame by linking geospatial data on agricultural fields with farm records on the list frame. Although many farms can be linked to geospatial data acquired by the Farm Service Agency (FSA), this linkage is not possible for farmers who do not participate in FSA programs, which may include members of some underrepresented groups in US agriculture. Thus, NASS has developed a georeferencing process for non-FSA farms, combining automatic and manual field identification, county assessor parcel data, record linkage, and classification surveys. This process serves the dual purpose of linking farms already on the list frame to geospatial data sources and identifying new farms to add to NASS's list frame. This report evaluates the output of the non-FSA georeferencing process for 11 states, with a focus on farms added to the list frame via georeferencing. Substantial percentages (>25% for each category) of the new farms added via georeferencing were urban or suburban farms, were small, had livestock, or were in counties with Amish settlements. The georeferencing process shows promise adding farms from these groups that have historically been less well covered in NASS surveys.

Keywords *area frame; georeferencing; list frame; record linkage; undercoverage*

1 Introduction and Background

The National Agricultural Statistics Service (NASS) conducts hundreds of list-based surveys each year. NASS has made a dedicated effort to georeference records on NASS's list frame. The agency maintains a list of farmers and ranchers from which the samples for all list-based surveys are selected. The list frame includes all known farms and potential farms in the United States (US). A farm is defined to be any place from which $1,000 or more of agricultural products were produced and sold or normally would have been sold during the year. This definition includes ranches and nurseries, as agricultural products include not only commodity crops (e.g., corn, soybeans), but also livestock and horticultural products. Farms on the list frame exclude landlord-only agricultural operations, where landowners rent out all of their agricultural land to operators, but do not raise or sell the agricultural products themselves.

Each record on the list frame is categorized as active, criteria record (i.e., potential farm),

☆Disclaimer: *The findings and conclusions in this report are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy.*

*Corresponding author. Email: Robbie.Emmet@usda.gov.

Received December 13, 2023; Accepted July 2, 2024

or inactive. Active list records are assumed to have a high probability of representing active farming operations. Criteria records are those whose involvement in agriculture is unknown. Inactive list records may be associated with, for instance, landlords, deceased operators, or farms no longer in business. Criteria records are periodically screened using the National Agricultural Classification Survey (NACS) to determine whether they are involved in agriculture. The list frame also includes control data, which are additional data on agricultural operations (e.g., crop acres, livestock) collected in previous surveys and used to inform sampling procedures for some NASS surveys.

Georeferencing involves linking non-survey, geospatial data (agricultural fields) to individual records on the list frame. This linkage between geospatial data sources (fields) and list frame records serves two major purposes. First, non-survey data can be used to inform and supplement survey data when imputing survey responses for individual list frame records. Non-survey geospatial data are used for manual imputation and may be used for automatic imputation of NASS's June Area Survey (Murphy et al., 2022); these data sources are also used for item- and record-level imputation in list-based surveys. Second, records found via georeferencing that are not currently on the list frame can be added to the list frame, thereby enhancing list frame coverage. The georeferencing process is conceptually similar to other methods for enhancement of address-based samples (Harter and English, 2018) in that objects with geographic coordinates in a region of interest (e.g., households in a sampled neighborhood and agricultural fields) are searched to identify records not on the list frame. However, NASS's georeferencing process does not require in-person enumeration.

NASS uses two primary geospatial data sources of agricultural fields for georeferencing the list frame: Common Land Units (CLUs) developed by the USDA's Farm Service Agency (FSA) and Crop Sequence Boundaries (CSBs) developed by NASS and the USDA's Economic Research Service (ERS).

Farmers participating in USDA programs or purchasing crop insurance must report their crop plantings each year. These reports include crop field locations, where each crop field boundary is defined by a CLU polygon (Heard, 2002; USDA NASS, 2023; Figure 1). FSA creates new CLUs each year and updates them with new data continuously throughout the growing season as farmers report to the over 2,300 FSA offices across the US.

CSBs are a publicly available spatial product automatically delineating agricultural fields using crop rotation information from NASS's Cropland Data Layer (CDL) (Hunt et al., 2023; Boryan et al., 2011; Figure 2).

Although georeferencing using CLUs can link geospatial data for many list frame records and has resulted in adding new records to the list frame, CLUs do not provide full spatial coverage of all US farms and may be missing some farm types and producer demographics at higher rates than others. In particular, only farms that participate in FSA farm assistance programs will be present in FSA CLUs. FSA programs may be more difficult to access for members of some underrepresented groups in US agriculture, including Latino immigrants, African-Americans, and Amish farmers (Escalante et al., 2006; Kraybill et al., 2013; Minkoff-Zern and Sloat, 2017). Many of these same groups, in addition to urban farmers, owners of small farms, and new or beginning farmers, may also have relatively low coverage on the list frame. Thus, to enhance list frame coverage for a variety of underrepresented producers, NASS should georeference farms that are not covered by FSA CLUs. To address the need to obtain geospatial data and improve coverage for farms that do not participate in FSA programs, NASS has developed a non-FSA georeferencing process combining geospatial data from CSBs, parcel data from county tax assessors' records, and data from the NACS.

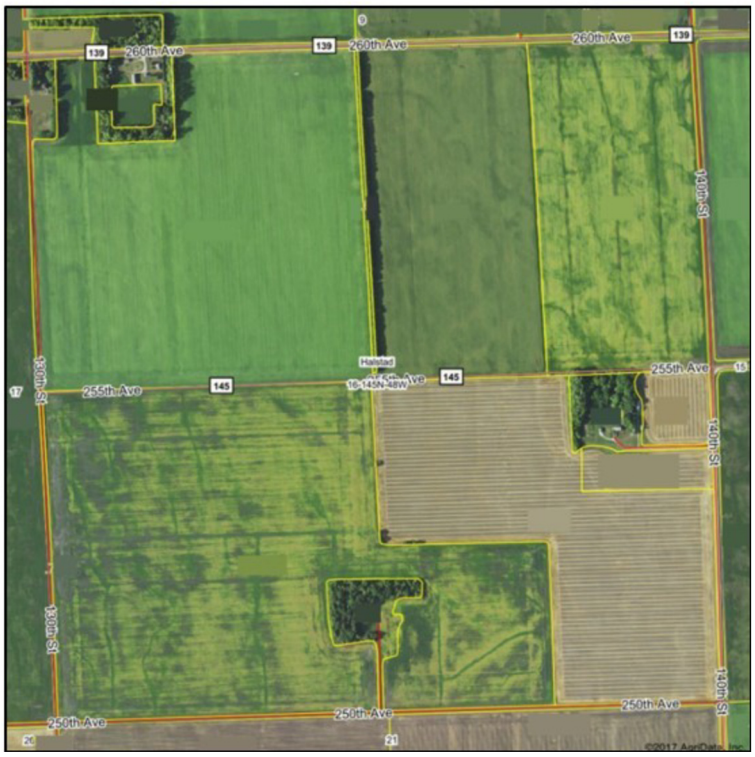

Figure 1: FSA CLUs are outlined in yellow. Each polygon represents an agricultural field, excluding non-agricultural areas such as farmsteads. Figure from AgriData Inc. (2008).

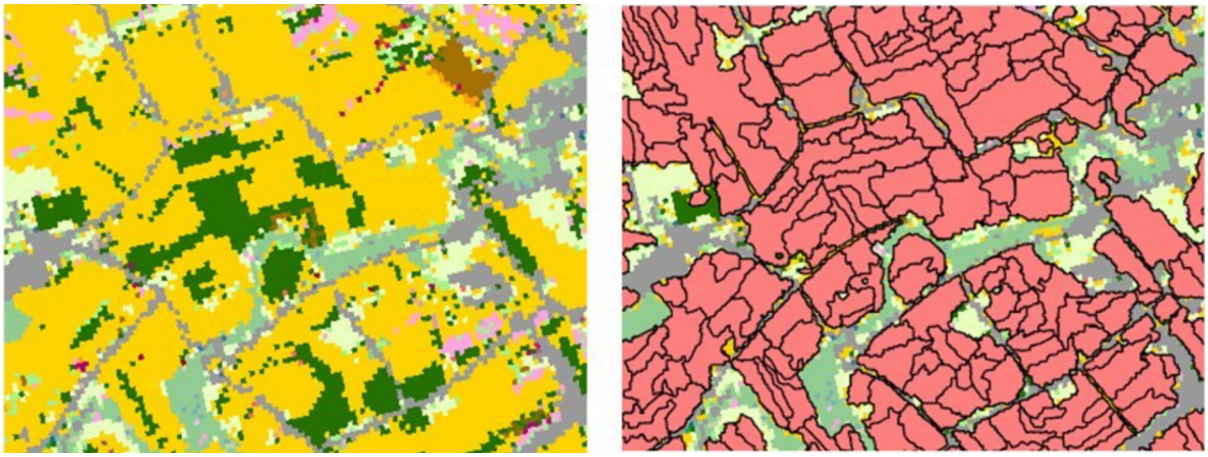

Figure 2: Example of Cropland Data Layer (CDL) (left) used to produce Crop Sequence Boundaries (CSBs) (right). CDL layers are stacked over multiple years, and areas of homogeneous crop sequences are combined to produce CSBs.

In this paper, the processes for identifying non-FSA fields through georeferencing, using record linkage to determine whether the non-FSA fields were on the NASS list frame, and determining which non-FSA fields not on the NASS list frame were associated with potential farms

are described. The primary objective of the study is to determine the characteristics of these potential farms. Section 2 outlines the study area, provides a general overview and terms associated with the georeferencing process, and describes the four phases of the process (agricultural field identification, geospatial overlap to retrieve landowner information from fields, record linkage, and NACS verification). Section 3 summarizes the results of the non-FSA georeferencing process for 11 US states with high corn and soybean production – Arkansas, Illinois, Indiana, Kansas, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. The results are presented with respect to characteristics of farms and potential farms added to the list frame via the georeferencing process. Implications of findings and limitations of the georeferencing process are outlined in Section 4.

2 Methods

2.1 Study Area

The study area consisted of 11 states: Arkansas (AR), Illinois (IL), Indiana (IN), Kansas (KS), Minnesota (MN), Missouri (MO), Nebraska (NE), North Dakota (ND), Ohio (OH), South Dakota (SD), and Wisconsin (WI)). These states were initially selected because they are among the states with the highest corn and soybean production in the United States (USDA NASS, 2023; Figure 3). Improving corn and soybean acreage and yield estimates has been a major research priority at NASS; therefore, georeferencing research and related research projects at NASS have generally begun in these states. In addition to their high corn and soybean produc-

Figure 3: Map of the continental United States, with states in the study area highlighted in dark gray.

tion, these states are notable for including 3 of the 5 states with the highest Amish populations, namely Indiana, Ohio, and Wisconsin (Young Center, 2022). Thus, the selection of these states provides an opportunity to assess whether the georeferencing process is effective at adding potential Amish farms to the list frame. Iowa was in the initial georeferencing study. However, it was excluded due to inconsistencies in data processing.

2.2 Overview of Non-FSA Georeferencing Process and Terminology

The georeferencing process requires linkages between multiple different data sources and procedures, each with different data structures and terminology. The non-FSA georeferencing process is performed in the following 4 phases: (1) identification of potential agricultural fields, which take the form of spatial polygons, specifically CSBs or hand-digitized fields (henceforth, agricultural fields); (2) geospatial overlap of agricultural fields with county assessor parcels that have both geospatial and tabular components to retrieve name and address information of the landowners parcels, henceforth, parcels; (3) record linkage of tabular parcel data to determine which landowners are already on the list frame; and (4) verification of agricultural activity through sending NACS questionnaires to landowners not found on the list frame during record linkage. Generally speaking, a record is equivalent to an operation. Records on the list frame may be identified by an operation ID or an operator ID for one or more of the operators associated with that record. To summarize the data sources and terms involved in the process, agricultural fields are identified, parcels (with landowner data) overlapping these fields are retrieved, and parcel data are either linked to existing records on the list frame or added as criteria records if they did not match any records on the list frame. Criteria records are sent a NACS questionnaire to determine whether agricultural activity is present and are determined to be "in-scope" (having agricultural activity) or not based on survey responses (Figure 4). In-scope records are assigned an active status and are subsequently eligible for sampling. Both records already on the list frame and in-scope criteria records are linked to specific non-FSA agricultural fields (i.e., CSBs or hand-digitized fields) as a result of this process. Thus, in addition to adding new records, the non-FSA georeferencing process allows operations to be linked to geospatial data associated with specific fields, whether CSBs or hand-digitized fields. Not every step in the process results

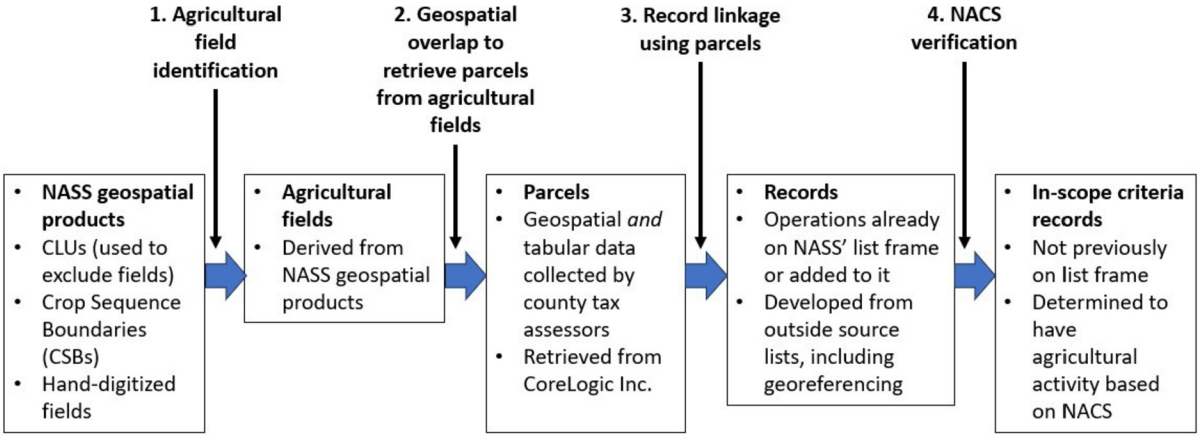

Figure 4: Map of the continental United States, with states in the study area highlighted in dark gray.

in a one-to-one link between data sources; agricultural fields may cover more than one parcel, and each parcel may be associated with more than one current or newly added list frame record. Additionally, parcels do not always represent all of the acres associated with a farm.

2.3 Georeferencing Phase 1: Agricultural Field Identification

The first step in the identification of potential non-FSA agricultural fields consists of intersecting FSA CLUs with CSB field polygons using ArcGIS Pro version 3.1. CSBs are polygons that combine areas of homogeneous crop sequences, using stacked Cropland Data Layer (CDL) pixels and road and rail networks to delineate boundaries of potential agricultural fields (Hunt et al., 2023). Base reference layers of political boundaries of counties (county borders), World Imagery (Esri, 2009), and a cultivated layer and pasture/hay mask created by NASS (Boryan et al., 2011) are additional sources used. To identify specifically non-FSA agricultural fields, CSBs and CLUs are loaded into ArcGIS and CSBs whose centroids overlapped a CLU are excluded. Zonal statistics are used with the cultivated layer and pasture/hay mask to calculate percentages of cultivated and pasture/hay within CSBs, and those with $\leqslant$10% cultivated or pasture/hay cover are excluded. NASS staff identify and hand-digitize any polygons that are not included in CSBs or CLUs and could be potential agricultural fields (Figure 5).

2.4 Georeferencing Phase 2: Geospatial Overlap of Agricultural Fields with County Assessors Parcels

The resulting set of agricultural fields (including both CSBs and hand-digitized) are intersected with county assessor parcels to retrieve individual landowner information. County assessor parcel information were obtained from CoreLogic Inc. (Figure 6). Parcels overlapping agricultural fields are retrieved using an API where possible, and using custom data requests from CoreLogic staff whenever the API is unable to retrieve data.

Figure 5: Example of spatial polygons used in identifying non-FSA potential agricultural fields. Non-FSA potential agricultural fields consist of CSBs (purple) and hand-digitized polygons (blue) that do not overlap with FSA CLUs (gray). Figure created by Avery Nagle.

Figure 6: Example of overlapping agricultural fields with parcels to retrieve parcel landowner data (name, address, etc.). Once an agricultural field is identified (light blue polygon (1)), parcels overlapping that field are retrieved (purple polygons (2)) and the retrieved overlapping parcels (light blue polygons (3)) are record linked using name, address, and other tabular data.

2.5 Georeferencing Phase 3: Record Linkage

Parcels are linked to the list frame using probabilistic record linkage (Broadbent and Iwig, 1999). Tabular parcel data used for record linkage include names, addresses (including city and state), and land use codes indicating land uses that fall into general categories (residential, commercial, vacant land, agricultural, etc.) as well as the percent cultivated data obtained from the CDL. Prior to record linkage, parcel information is processed through a filtering and reformatting program to eliminate parcels with landowner/organization names (e.g., cemeteries), land use codes and percent cultivated values indicating a low likelihood of agricultural activity. The remaining parcels are processed through six sub-matching routines, moving from matching with high-priority records (e.g., large farms) to low-priority records (e.g., non-farms) on the list frame.

2.6 Georeferencing Phase 4: Agricultural Activity Verification Using NACS

Landowner data (name, address) from parcels that are not on the list frame are added to it as criteria records until their involvement in agriculture can be established. Most of these records are added to the mailing list for the last NACS survey before the 2022 Census of Agriculture; not every record is included due to cost limitations. The sampling frame for NACS surveys includes agricultural operations and potential agricultural operations that have not yet been classified as farms or non-farms. This may have included some georeferenced records that were already on the list frame, and which were sent a NACS to update active status for reasons other than georeferencing.

2.7 Data Preparation

Data cleaning and analyses were largely conducted in R version 4.0.2 (R Core Team, 2020), with some queries conducted in SAS 9.4. The following packages were used: DBI version 1.1.0 (R-SIG-DB, Wickham, and Müller, 2019), tidyverse version 1.3.0 (Wickham et al., 2019), stringr version 1.4.0 (Wickham, 2019), sf version 0.9-4 (Pebesma, 2018), tidycensus version 0.9.9.5 (Walker, 2020), and stringdist version 0.9.5.5 (van der Loo, 2014).

For the 11 states in the study, agricultural fields from CSBs and hand-digitized polygons were linked to the list frame using agricultural field-level IDs. Because multiple landowner parcels were found in some agricultural fields, agricultural field ID alone was not sufficient to uniquely link operators on the list frame to individual parcels. The only exceptions were when fields contained exactly one parcel (one-to-one relationships). To address the one-to-many relationships between fields and parcels, a custom matching routine was developed to produce as many "clean" matches as possible. This matching routine included perfect matches, unambiguous (e.g., one-to-one) matches, and matches where names were similar but not identical (e.g., nicknames and suffixes) based on Jaro-Winkler string distances (van der Loo, 2014). This matching procedure does not contain every true match and may contain a small proportion of false matches.

2.8 Additional Data Sources

We investigated whether the georeferencing process was adding new records to the list frame from underrepresented groups. Because certain underrepresented groups in US agriculture, such as non-white, female, and small-scale livestock farmers, can be identified using NACS survey responses (or list frame data for records already on the list frame), the acquisition of additional data centered on two groups that the NACS survey does not allow us to separately identify, namely urban and suburban farmers and Amish farmers. Therefore, additional data sources were obtained to characterize the county-level presence of these underrepresented groups.

The National Bureau of Economic Research's (NBER) Census Core-Based Statistical Area (CBSA) to Federal Information Processing Series (FIPS) County Crosswalk (NBER, 2022) identifies the counties in the statistical area and whether they are central or outlying for each metropolitan and micropolitan statistical area based on 2010 US Census definitions. Data on whether the statistical area is metropolitan or micropolitan and whether the county is central or outlying are merged with the combined parcel and list frame data using 5-digit county FIPS codes, which include the state FIPS code and a state-specific county FIPS code.

Data on counties with known Amish settlements as of 2022 are maintained by the Young Center for Anabaptist and Pietist Studies at Elizabethtown College (Young Center, 2022). County names are cross walked with FIPS codes using R package tidycensus, and Amish population status of counties is merged with combined parcel and list frame data using FIPS codes.

3 Results

In the study area, there are 47,211 parcels for which clean matches to list frame records could be identified, representing 41,552 records. These 41,552 records are either active or criteria records on the list frame. Of the 41,522 records, 6,958 (17%) are associated with at least one hand-digitized agricultural field (records may be linked to multiple fields), 7,836 (19%) are already on the list frame, and 33,847 (82%) are included in NACS (Table 1). Of these 33,874 records included in NACS, 21,291 (63%) responded to NACS (Table 2), and of these 21,291, 4,287 (20%) were classified as in-scope on the NACS (Table 3).

7,836 records are already on the list frame. 214 (3%) had no sales during the year of the most recent survey response, and 498 (6%) had sales under $10,000. Records on the list frame are associated with relatively small operations with medians of 50 acres owned and 53 acres operated. 3,617 records on the list frame (46%) have some livestock (beef or dairy cattle, sheep, goats, pigs, horses, chickens, or turkeys), and 891 (11%) have some chickens.

Table 1: Records associated with outcomes of multiple phases of the georeferencing process.

		Total Records
Phase 1		
Crop Sequence Boundaries overlapped with FSA CLUs, hand-digitized agricultural fields created		NA (no records in this phase, only fields)
Phase 2		
Crop Sequence Boundaries	Hand-digitized agricultural fields	
34,594	6,958	41,522
Phase 3		
Not linked to the list frame	Linked to the list frame	
33,716	7,836	41,522
Phase 4		
Included in the NACS	Not included in the NACS	
33,874	7,678	41,522

Table 2: Responding status of records included in the NACS.

Respondents	21,291
Non-respondents	12,583
Total	33,874

Table 3: Classification status of records responding to the NACS.

In-scope	4,287
Out-of-scope	17,004
Total	21,291

There were 3,921 records that were not previously on the list frame and were classified as in-scope on the NACS – these are referred to as in-scope criteria records. Of the 3,921 in-scope criteria records, 2,410 (61%) had no sales during the NACS survey year and 999 (25%) had sales under $10,000. In-scope criteria operations tend to be relatively small, with medians of 11 acres owned and 10 acres operated. 1,873 in-scope criteria operations (48%) have some livestock, and a substantial proportion of these, 1,009 (26% of in-scope criteria operations), have some chickens.

Of these 3,921 in-scope criteria records, 3,160 (81%) are associated with parcels located in counties within either a metropolitan or micropolitan statistical area, while 2,213 (56%) are associated with parcels located in counties within a metropolitan statistical area. Of the 3,921 in-scope criteria records, 1,163 (30%) are associated with parcels located in counties with known Amish settlements as of 2022. There is substantial overlap between non-metropolitan and non-micropolitan areas and counties with Amish settlements among in-scope criteria parcels. Of the 761 in-scope criteria records that are not located in metropolitan or micropolitan statistical areas, 396 (52%) are associated with parcels located in counties with Amish settlements. In addition to

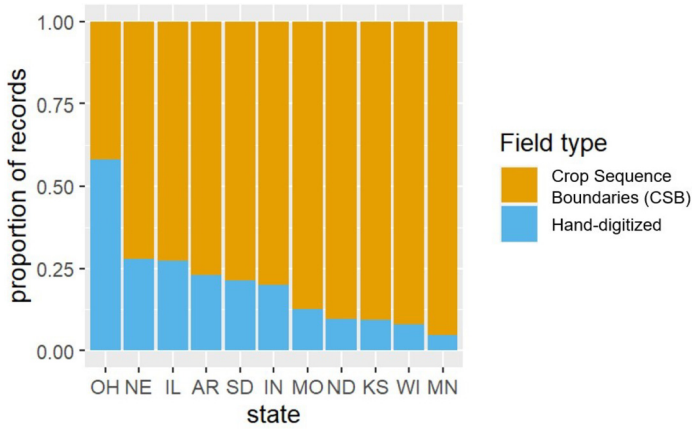

Figure 7: Proportion of in-scope criteria records that are associated with at least one hand-digitized agricultural field (blue) for each state in the study area. While the overall percentage of in-scope criteria records associated with hand-digitized polygons is low (19%), this percentage is markedly higher for certain states (e.g., Ohio). The other form of agricultural field, Crop Sequence Boundaries (CSB) (orange), consist of stacked homogeneous crop sequences.

large percentages of in-scope criteria parcels occurring in metropolitan areas, micropolitan areas, or counties with Amish settlements, many in-scope criteria records (736 out of 3,921, or 19%) are associated with hand-digitized agricultural fields, particularly in certain states including Ohio, Nebraska, and Illinois (Figure 7). Records associated only with CSBs and those associated with at least one hand-digitized agricultural field also seem to represent operations of different sizes; the median summarized average acres operated of records associated with in-scope criteria parcels was 11 acres for records from CSBs, but only 7 acres for records associated with hand-digitized agricultural fields.

4 Discussion

The georeferencing process in 11 major corn- and soy-producing US states resulted in the addition of 3,921 records to the list frame that were classified as in-scope on the NACS, indicating agricultural activity associated with the record. This number represents a small percentage of the active records on the list frame. However, the composition of these records is encouraging for using georeferencing as a list-building activity for specific types of farms with low list frame coverage. Even 3,921 records may greatly improve coverage for specific groups. In particular, the majority of in-scope criteria records (farms) in the study area are small farms, livestock farms, or urban or suburban farms, while a majority of in-scope records added from rural areas are located in the vicinity of Amish communities. These types of records may be difficult to add to the list frame using traditional list-building methods. Farms operated by non-white producers collectively make up less than 4% of the records newly added via georeferencing, and very few of the georeferenced records associated with non-white producers were on the list frame prior to georeferencing. While this may suggest more limited potential for enhancing coverage for certain underrepresented groups using georeferencing, the results of the process are likely to be different in states with higher numbers of underrepresented producers than the major corn- and soybean-producing states.

Records in counties with Amish communities are a surprisingly high proportion of in-scope criteria records (30%). Although not all of these records are necessarily associated with Amish farms, their proximity to Amish farms suggests that the georeferencing process may be useful for adding Amish farms to the list frame. Even though agriculture is declining as a proportion of Amish occupations (Lowery and Noble, 2000; Cross, 2014), the rapid growth of Amish populations and settlements, combined with net population declines in many rural areas of the United States, suggests that Amish farmers may be a growing proportion of farmers in the United States, particularly for some farm types and locations. For example, it was estimated in 2006 that one-eighth of dairy farms in the United States were operated by Amish farmers (Cross, 2006). Amish agriculture is characterized by small, diverse operations, such as dairy farms, produce farms, and organic farms (Kraybill et al., 2013). Amish agriculture is difficult to characterize on NASS's surveys, as they cannot be identified definitively using any survey data NASS collects. The US Census Bureau is prohibited by Title 13 of the US Code (13 U.S.C. §221, 1976) from requiring respondents to answer questions about religious beliefs or membership in a religious body. NASS also does not collect information regarding religion, which makes it impossible to identify Amish farms with certainty. However, potential Amish farms can be identified using proxy variables such as surname, location, horse ownership, farm production practices (e.g., electricity expenses), or lack of phone or Internet usage (Kent and Neugebauer, 1990; Cross, 2003; Holly et al., 2019). These variables are worth considering for inclusion in future coverage, response, and imputation models, if Amish farmers have different characteristics than non-Amish farmers in terms of these survey quality measures.

One limitation of the georeferencing process is that it links county assessor parcels with list frame records using data on landowners. This method may underrepresent groups in farming that rent land at higher rates, such as young or beginning farmers and some non-white farmers. If there was some method to deliver NACS surveys to renters of the landowner parcels, in addition to landowners, the representation of renters in the georeferencing process and on the list frame could be improved. Second, it is critical that list frame and parcel data be consistently matched based on the linkage data available on the list frame. A permanent, stable parcel ID is necessary for accurate linkage between the list frame and specific parcels. Finally, hand-digitizing agricultural fields is a time-consuming process that may be targeted or refined further to decrease the amount of search effort necessary. Geographical features or variables associated with high rates of in-scope records could be used to target specific sections of states. In addition to the variables explored in this analysis, more demographic variables and even topographic variables could be considered in future analyses of georeferenced records. Early results suggest that hand-digitizing plays a crucial role in the georeferencing process, as some states have added a large percentage of new records to the list frame via this process (Figure 7). Additionally, hand-digitized agricultural fields are associated with smaller farms that may be more difficult to add to the list frame using conventional list-building techniques.

The georeferencing process has resulted in the addition of 3,921 records to NASS's list frame to date. The addition of these records to the list frame is likely to improve list frame coverage of small and urban farms, and characteristics of georeferenced records and impacts on coverage should be monitored as georeferencing continues in the remaining states. Although the georeferencing process may benefit from certain refinements, such as more targeted hand-digitizing of agricultural fields, the initial georeferencing process for the study area in this paper has yielded promising results in terms of adding undercovered farm types to the list frame. If the georeferencing process is applied in the future, records added via georeferencing may contribute

to published Census of Agriculture and other NASS survey estimates, increasing coverage of estimates and potentially sample sizes.

Supplementary Material

The primary data used in this paper are CIPSEA protected and are not allowed to be distributed by law.

Acknowledgement

A. Dau, S. Dothage, L. Ebinger, K. McMillan, K. Mills, A. Nagle, and L. Lawson contributed to multiple aspects of the georeferencing process. M. Gerling, D. Miller, K. Turner, and L. Young provided valuable feedback on drafts of this manuscript.

References

AgriData, Inc (2008). FSA field boundaries (CLU) – May 2008. https://www.agridatainc.com/Home/Products/Mapping. Accessed: 2023-12-08.

Amendment to Title 13 of the US Code 13 USC §221 (1976).

Boryan C, Yang Z, Mueller R, Craig M (2011). Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer program. *Geocarto International*, 26(5): 341–358. https://doi.org/10.1080/10106049.2011.562309

Broadbent K, Iwig W (1999). Record linkage at NASS using automatch. In: *Proceedings of the 1999 FCSM Research Conference*.

Cross JA (2003). Amish surnames, settlement patterns, and migration. *Names*, 51(3/4): 193–214. https://doi.org/10.1179/nam.2003.51.3-4.193

Cross JA (2006). Restructuring America's dairy farms. *Geographical Review*, 96(1): 1–23. https://doi.org/10.1111/j.1931-0846.2006.tb00385.x

Cross JA (2014). Continuity and change: Amish dairy farming in Wisconsin over the past decade. *Geographical Review*, 104(1): 52–70. https://doi.org/10.1111/j.1931-0846.2014.12004.x

Escalante CL, Brooks RL, Epperson JE, Stegelin FE (2006). Credit risk assessment and racial minority lending at the Farm Service Agency. *Journal of Agricultural and Applied Economics*, 38(1): 1–15. https://doi.org/10.1017/S1074070800022070

Esri (2009). World Imagery. https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9#!. Accessed: 2023-09-18.

Harter R, English N (2018). Overview of three field methods for improving coverage of address-based samples for in-person interviews. *Journal of Survey Statistics and Methodology*, 6(3): 360–375. https://doi.org/10.1093/jssam/smx037

Heard J (2002). USDA establishes a common land unit. http://www.esri.com/news/arcuser/0402/usda.html.

Holly MA, Gunn KM, Rotz CA, Kleinman PJA (2019). Management characteristics of Pennsylvania dairy farms. *Applied Animal Science*, 35(3): 325–338. https://doi.org/10.15232/aas.2018-01833

Hunt KA, Abernethy J, Beeson P, Bowman M, Wallander S, Williams R (2023). Crop Sequence Boundaries: Delineated fields using remotely sensed crop rotations. In: *Proceedings of ICAS IX*, Washington, DC.

Kent RB, Neugebauer RJ (1990). Identification of ethnic settlement regions: Amish-Mennonites in Ohio. *Rural Sociology*, 55(3): 425–441.

Kraybill DB, Johnson-Weiner KM, Nolt SM (2013). *The Amish*. The Johns Hopkins University Press, Baltimore, MD.

Lowery S, Noble AG (2000). The changing occupational structure of the Amish of the Holmes County, Ohio, settlement. *Great Lakes Geographer*, 7(1): 21–36.

Minkoff-Zern L, Sloat S (2017). A new era of civil rights? Latino immigrant farmers and exclusion at the United States Department of Agriculture. *Agriculture and Human Values*, 34: 631–643. https://doi.org/10.1007/s10460-016-9756-6

Murphy T, Rosales A, Sartore L, Abreu DA (2022). Automatic imputation for an area survey. In: *Proceedings of Joint Statistical Meetings*, Seattle, WA, 1868.

National Bureau of Economic Research (2022). Census Core-Based Statistical Area (CBSA) to Federal Information Processing Series (FIPS) county crosswalk. https://www.nber.org/research/data/census-core-based-statistical-area-cbsa-federal-information-processing-series-fips-county-crosswalk. Accessed: 2023-03-02.

Pebesma E (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal*, 10(1): 439–446. https://doi.org/10.32614/RJ-2018-009

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

R Special Interest Group on Databases (R-SIG-DB), Wickham H, Müller K (2019). DBI: R database interface. R package version 1.1.0.

USDA National Agricultural Statistics Service (2023). Crop production 2022 summary. https://downloads.usda.library.cornell.edu/usda-esmis/files/k3569432s/9306v916d/wm119139b/cropan23.pdf. Accessed: 2023-08-23.

van der Loo MPJ (2014). The stringdist package for approximate string matching. *The R Journal*, 6(1): 111–122.

Walker K (2020). tidycensus: Load US Census boundary and attribute data as 'tidyverse' and 'sf'-ready data frames. R package version 0.9.9.5.

Wickham H (2019). stringr: Simple, consistent wrappers for common string operations. R package version 1.4.0.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. (2019). Welcome to the tidyverse. *The Journal of Open Source Software*, 4(43): 1686. https://doi.org/10.21105/joss.01686

Young Center for Anabaptist and Pietist Studies, Elizabethtown College (2022). Amish population in the United States by state, county, and settlement, 2022. https://groups.etown.edu/amishstudies/files/2022/07/Amish-Pop-2022_by_state_and_county.pdf. Accessed: 2023-02-08.