

Using a Mobile Mapping Instrument to Evaluate a Permanent Grid Sampling Frame

Denise A. Abreu, Linda A. Lawson, Claire Boryan, Michael Gerling, Rick Hardin, Michael Hyman

National Agricultural Statistics Service

1400 Independence Ave, SW Washington DC 20250, denise.abreu@nass.usda.gov

Key Words: mobile mapping, area frame, grid cells, data collection, GIS

Abstract

The National Agricultural Statistics Service (NASS) conducts the June Agricultural Survey (JAS), which is based on an area frame. Segments of land comprise the sampling units for the JAS. Every year, 20 percent of the sampled segments are rotated out and a new rotation of segments is introduced. Building and constructing an area frame is expensive and time consuming. The agency is evaluating the use of a permanent grid sampling frame as a cost saving initiative. NASS is also investigating use of mobile mapping technology during data collection. Currently, JAS sampled segment boundaries follow roads or other physical features to match the infrastructure on the ground. Field enumerators use an aerial photo to locate and interview all operators within the segment boundary. NASS is evaluating the modernization of its data collection effort through replacing the use of the aerial photo and paper questionnaire with the use of a mobile mapping instrument. To test this concept of permanent grids in conjunction with the mobile mapping instrument, enumerators in North Carolina, Pennsylvania and South Dakota visited with farm operators during Summer 2014. This paper documents the challenges faced as well as the lessons learned from this test and discusses future plans for NASS's data collection activities, including the use of Farm Service Agency (FSA) common land units (CLUs).

I. Introduction and Background

In recent years, the National Agricultural Statistics Service (NASS) has been making strides to improve its data collection processes and to evaluate a number of cost saving initiatives. The June Agricultural Survey (JAS) is NASS's largest survey, which is based on an area frame. This is also NASS's only survey collected via in-person interviews utilizing pencil and paper. Frame construction and face-to-face data collection efforts are expensive and time consuming processes. It has been proposed to replace the current area frame with a permanent grid sampling frame. Currently, segments of land comprise the sampling units for the JAS. In the current segment creation process, segment borders are adjusted to follow physical features on the ground (i.e., an edge of a field, a road, a river, etc.). Determination and preparation of segments is labor intensive and expensive with overall costs around 2.6 million dollars. The proposed grid frame, on the other hand, has units having roughly equal-sized and shaped areas called grid cells or grids. This permanent frame was developed based on the Public Land Survey System (PLSS), which lacks any physically identifiable boundaries. Figure 1 shows the grid frame concept (outlined in red) compared with a traditional JAS segment (outlined in purple).

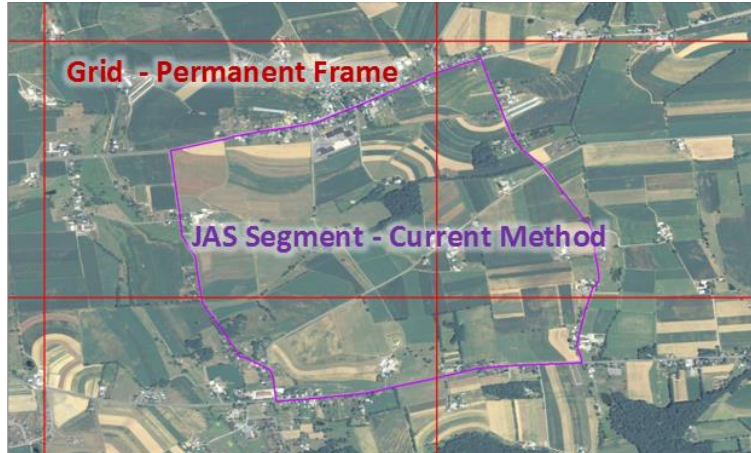


Figure 1: Grid cell (outlined in red) vs. JAS segment (outlined in purple)

Because grid cells do not follow the infrastructure on the ground and often cut across fields, a mobile mapping instrument was developed to use Geographic Information Systems (GIS) technology to calculate the acreage of fields surveyed. In 2012, a team of researchers from NASS and Iowa State University's (ISU's) Center for Survey Statistics and Methodology, developed a prototype mobile mapping instrument, called Geographic Information Running Area Frame Forms Electronically (GIRAFFE) (Gerling, Lawson, Weaber, Dotts, Vardeman, Wilson, 2015). The instrument was designed to operate on an iPad and can be utilized to collect data for either grid cells or traditional JAS segments. A series of studies are being conducted to evaluate whether or not NASS could implement the permanent grid frame operationally. Phase I was the actual development of the instrument in 2012. The objective of the 2013 study (or Phase II) was to determine whether the mobile mapping instrument's calculated GIS acreages were comparable to the acreages reported by farmers during the JAS before moving forward with testing grid cells. Enumerators were provided the completed aerial photos from previously collected data and tasked with replicating the field boundaries within the mobile mapping instrument. Results indicated that there was strong agreement between JAS farmer reported and GIS acreage in the 90 JAS segments and 2,246 fields in the three states participating in the study (Boryan, Lawson, Abreu, Weaber, Gerling, Hyman, Hardin, 2016). Based on this result, the next study (called Phase III here) was initiated to test collecting data using grid cells utilizing the mobile mapping prototype instrument. This paper documents the challenges faced as well as the lessons learned from Phase III and discusses future plans for NASS's data collection activities, including the use of Farm Service Agency (FSA) common land units (CLUs). First, the JAS and permanent grid frame concept are presented.

II. NASS Area Frame Construction and the June Agricultural Survey (JAS)

NASS's June Agricultural Survey (JAS) is based on an area frame, which ensures complete coverage of all land in the US. First, all land in a state is stratified using GIS technology, such as satellite imagery, aerial photography, and a GIS layer comprised of land and crop types, known as the Cropland Data Layer (CDL) (Boryan & Yang, 2012). This step is a manual process where Primary Sampling Units (PSUs) are digitized (electronically identified using GIS software) and classified into the defined strata for a state. A sample of PSUs is selected and smaller, similar-sized segments (each about a square mile (640acres)) of land are delineated within these selected PSUs. In general, staff divide a PSU of four square miles into four segments, one square mile each. Next, one segment is randomly chosen from within each sampled PSU (see Figure 2). This process avoids segment delineation for non-selected PSUs, thereby saving resources. In the current sampling scheme, the JAS replaces the oldest 20% of the segments with new segments rotated in each year. Eight staff working year-round are required to select the incoming rotation for the sample. Also, in the preparation of JAS segments, segment boundaries are adjusted (moved) to natural boundaries that can be easily identified outdoors, such as roads, ditches, edges of fields, rivers, tree lines, etc. This "tweaking" of boundaries is also a labor-intensive process. A state receives a completely new area frame sample approximately every fifteen years. This on-going process takes twenty-five staff with salary and benefits totaling about 2.5 million dollars and another 100,000 dollars in equipment, software, printing, and mailing of materials. (See Cotter et. al 2009

for further details on the JAS design).

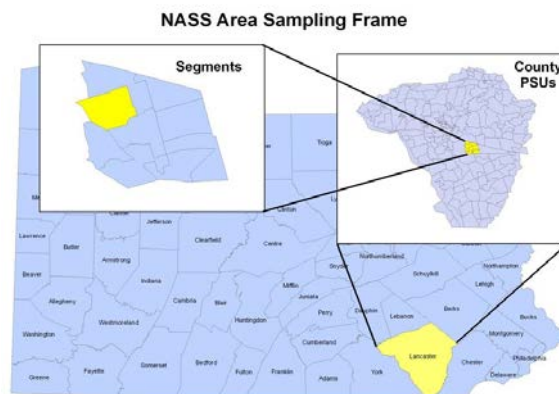


Figure 2: NASS area sampling frame for Pennsylvania

JAS segments (outlined in red in Figure 3) are pre-screened during the month of May prior to the June data collection period. During pre-screening, field enumerators divide each segment into tracts of land (outlined in blue in Figure 3). Each tract represents a unique land operating arrangements. Enumerators do not interview tract operators during pre-screening. Instead they are provided with a hard copy aerial photo, on which a segment is outlined, and a screening form, which contains screening questions that determine whether or not each tract has agricultural activity. Each screened tract is classified as agricultural or non-agricultural. Actual JAS data collection is conducted during the first two weeks of June. At this time, field enumerators return to only those tracts classified as agricultural. Those operations (tracts) that qualify as agricultural are subsequently interviewed using the JAS questionnaire, which collects detailed agricultural information about the operator's land, both inside and outside the segment. Enumerators complete a separate paper questionnaire for each agricultural operation within the segment. Farm operators identify all field boundaries (outlined in red in Figure 4) on the aerial photo and report acreage and the crop planted or other land use of each individual field (pasture, woods, wasteland, etc.).

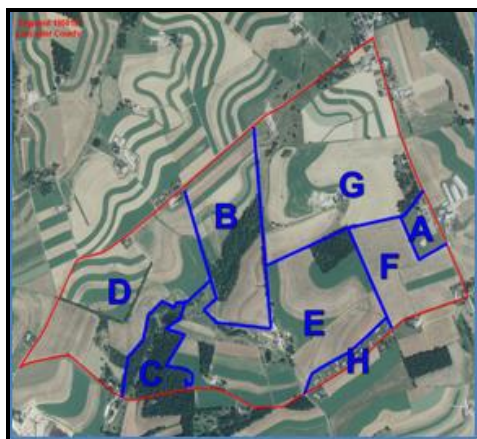


Figure 3: The area outlined in red is the segment. Tracts are outlined in blue and labeled with letters.



Figure 4: Tracts are outlined in blue and labeled. Individual fields are outlined in red within the tracts and labeled with numbers.

III. Theoretical Grid Sampling Frame

A permanent grid frame with units having roughly equal-sized and shaped areas, and thus lacking physically identifiable boundaries, was developed based on the Public Land Survey System (PLSS). PLSS is a surveying method used over large parts of 30 states in the United States to spatially identify parcels of land, especially in rural and undeveloped areas. Land was divided into (mostly) rectangular areas ranging from a 24 mile by 24 mile quadrangle down to a one mile by one mile square section (See Figure 5). Hence, the United States could be divided into roughly 1 mile x 1 mile squares called grid cells or grids. The sampling process could then be automated to handle stratification and sample selection of these grids using NASS's current methods. The field enumerator would then be responsible for collecting all agricultural data within the defined grid cells. This would reduce the resources required in the preparation for the JAS.

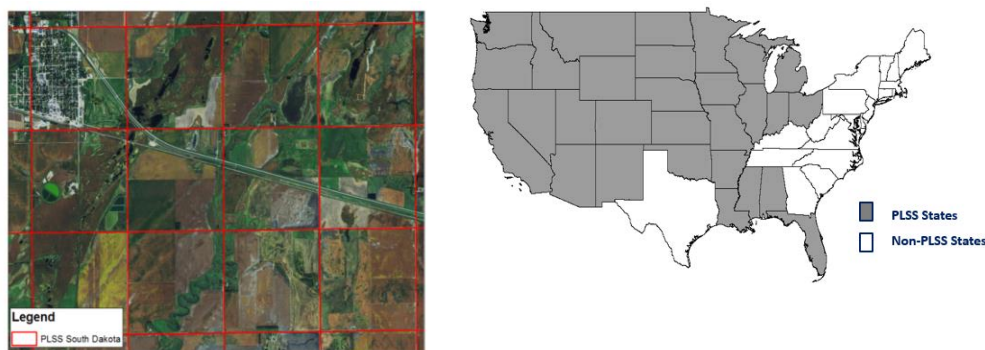


Figure 5: Grid concept close up of a specific area on the left. States covered under the PLSS layer on the right.

A challenge with grid cells is that fields may not be fully contained within a grid cell boundary. In these instances, information must be collected for the portion of the field that lies within the grid cell or partial fields (See Figure 6). In the figure below, the pink boundary identifies the road, while the red line determines the actual grid cell boundary. In this case, the enumerator will need to collect information on the portion of land to the south of the road (labeled as partial field in Figure 6). This may be difficult for an agricultural operator to report correctly, especially by only viewing a printed aerial photo. Thus, having a mobile mapping instrument incorporating GIS information in a geospatial display, combined with tools to delineate fields and tracts within the grid cell, could be used to eliminate the need for agricultural operators to report acreage for land within the grid cell.



Figure 6: Partial field identified at the left bottom corner of the grid cell.

IV. Prototype Mobile Mapping Instrument

In 2012, a team of researchers from NASS and Iowa State University's (ISU's) Center for Survey Statistics and Methodology, developed a prototype mobile mapping instrument called Geographic Information Running Area Frame Forms Electronically (GIRAFFE) (See Figure 7). GIRAFFE was developed to aid in the process of evaluating this permanent grid sampling frame, reducing costs, improving data quality and lengthening the data collection window (Gerling, Lawson, Weaver, Dotts, Vardeman, Wilson, 2015).

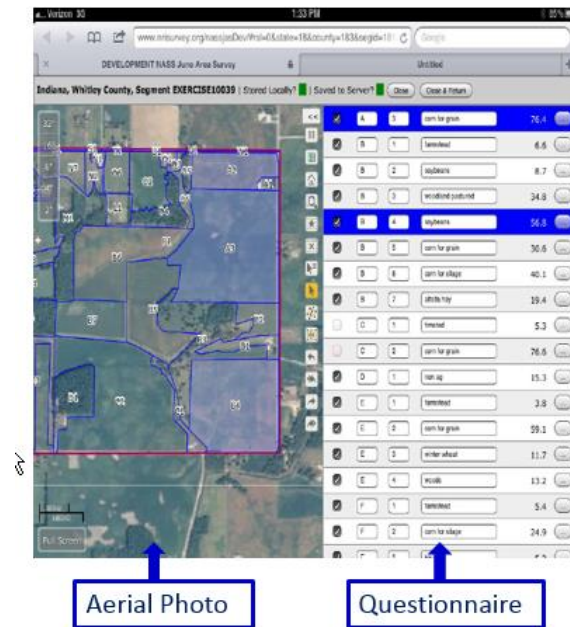


Figure 7: Mobile mapping instrument

The prototype JAS instrument was designed to run within the Safari browser of an iPad. It was essential that the instrument function without an Internet connection as a substantial amount of the JAS data collection occurs in rural areas that tend to have intermittent signal. Prior to data collection, enumerators run a cache routine to store the needed imagery in the iPad's memory. The instrument has two main parts (Figure 7). The left side of the screen contains the aerial imagery, which can also be run in full screen mode, and the right side of the screen displays general field information that replaces the Section D (Attachment A) part of the paper questionnaire (see Lawson, Abreu, Boryan, Gerling, and Hardin, 2015, for more information on the specifics of the mobile mapping instrument as well as improvements made to date).

V. Phase III Study -- Field Data Collection & Results

A random sample of 60 grid cells was selected in North Carolina (NC), Pennsylvania (PA) and South Dakota (SD) (20 per state). SD was selected because it is part of the PLSS. Because NC and PA are not part of the PLSS program, a fishnet was created using ArcGIS software and grid cells selected. Figure 8 shows the location of the grid cells in each of the participating states.

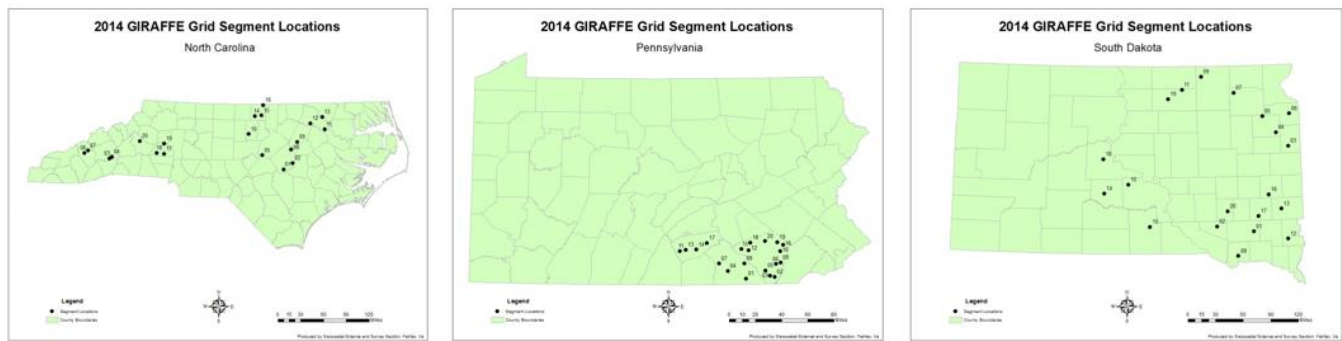


Figure 8: Selected grid cells in participating states.

Enumerators identified a total of 917 tracts. Then, they visited farm operators in all 457 tracts classified as agricultural. Enumerators were required to complete a questionnaire for each grid cell. This form collected information pertaining to that specific grid cells and all the tracts contained within (See Attachment B for questionnaire). They recorded any issues encountered with the mobile mapping instrument, challenges met while enumerating the grid cells, and time spent with and without the farm operator.

Enumerators were asked to record their experience and any challenges encountered while enumerating the grid cells, by indicating ‘yes’, ‘no’ or ‘sometimes.’ In addition, they explained any issues they encountered. Table 1 displays the results of the question pertaining to challenges the enumerators faced while enumerating the grid cells.

Table 1: Did You Encounter Any Issues While Enumerating the Grid Cells?

	North Carolina (non-PLSS)		Pennsylvania (non-PLSS)		South Dakota (PLSS)		ALL 3 States	
	Number of Agricultural Tracts	Percent	Number of Agricultural Tracts	Percent	Number of Agricultural Tracts	Percent	Number of Agricultural Tracts	Percent
Yes	14	12.2	9	4.7	17	21.8	40	10.4
Sometimes	20	17.4	29	15.2	4	5.1	53	13.8
No	81	70.4	153	80.1	57	73.1	291	75.8
Total*	115	100.0	191	100.0	78	100.0	384	100.0

*There were 73 instances of non-response that are not included in the total. Counts by state: NC-21, PA-48 and SD-4.

The results indicate that enumerators encountered issues while enumerating grids over 20 percent of the times. Even though non-PLSS states were expected to report numerous issues because they were not part of the PLSS, surprisingly SD (a PLSS state) reported significantly more issues with the enumeration (answered ‘yes’ 21.8% vs. 12.2% in NC and 4.7% in PA). The primary problem was that the grid cells did not match the infrastructure on the ground exactly. As an example, in figure 9, the SD grid boundary does not align exactly with the PLSS infrastructure consisting of roads, etc. As a consequence, the edges of fields are excluded from the left side of the grid cell and small slivers of fields are included on the right side of the grid cell. In these cases, enumerators were required to contact additional operators and to complete a survey form based upon a very small

portion of a field. Enumerators also reported that farm operators had difficulty estimating acreage for just the part of the field inside the grid cell boundary. PA and NC enumerators indicated that their main issue was identifying clear boundaries for the grid cells (See Figure 10). In addition, they also reported that grid cell boundaries cut across numerous crop fields.



Figure 9: SD grid cell with slivers highlighted in purple.

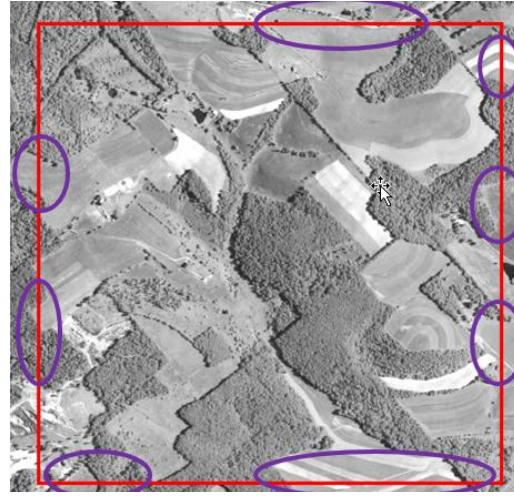


Figure 10: Grid cell with difficult to boundaries. Ovals shows grid boundary cutting across multiple fields.

Once enumerators contacted farm operators, they would collect crop specific data on the Section D portion of the instrument and then they proceeded to delineate all fields in the aerial section of the instrument. Fields were not drawn, they were created by splitting the areas within the grid cell. Table 2 displays the number of fields per tract by state and all states combined. PA has more tracts than both SD and NC. It is important to note that this was a result of the counties that were available for sampling in PA. This is not representative of all the counties in the state. However, because there were many more tracts with more fields than the other states, the PA enumerators took more time to draw off or delineate the fields on the mobile mapping instrument. Figure 11 shows an example of the number of fields delineated for each state. This helps contrast how many more fields needed to be delineated in PA as compared to the other two states.

Table 2: Number of Fields per Tract by State

	North Carolina		Pennsylvania		South Dakota		ALL	
	Number of Agricultural Tracts	Percent	Number of Agricultural Tracts	Percent	Number of Agricultural Tracts	Percent	Number of Agricultural Tracts	Percent
1 Field	41	30.1	44	18.4	24	29.3	109	23.9
2 Fields	19	14.0	51	21.3	16	19.5	86	18.8
3-4 Fields	28	20.6	47	19.7	14	17.1	89	19.5
5-7 Fields	23	16.9	28	11.7	13	15.8	64	14.0
8-10 Fields	8	5.9	32	13.4	8	9.8	48	10.5
Over 10 Fields	17	12.5	37	15.5	7	8.5	61	13.3
Total	136	100.0	239	100.0	82	100.0	457	100.0

North Carolina



Pennsylvania



South Dakota

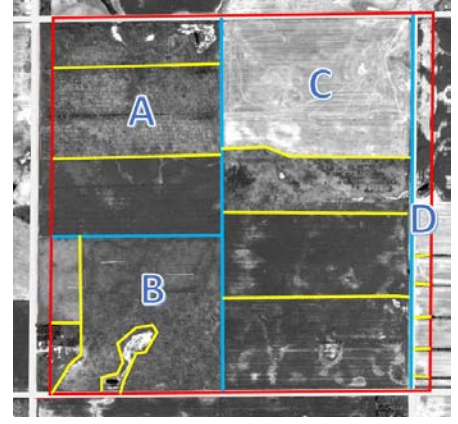


Figure 11: Tracts (blue) and fields (yellow) contained within a grid cell in NC, PA and SD

The time taken by enumerators to enter agricultural grid cells into the mobile mapping instrument and the time spent with the tract operator were examined for each state. Zeros were removed from the data because they were recorded when no interview took place. Figure 12 describes the enumeration times spent with the tract operator for each of the three states. The black vertical line represents the mean enumeration time with the operator for that particular state.

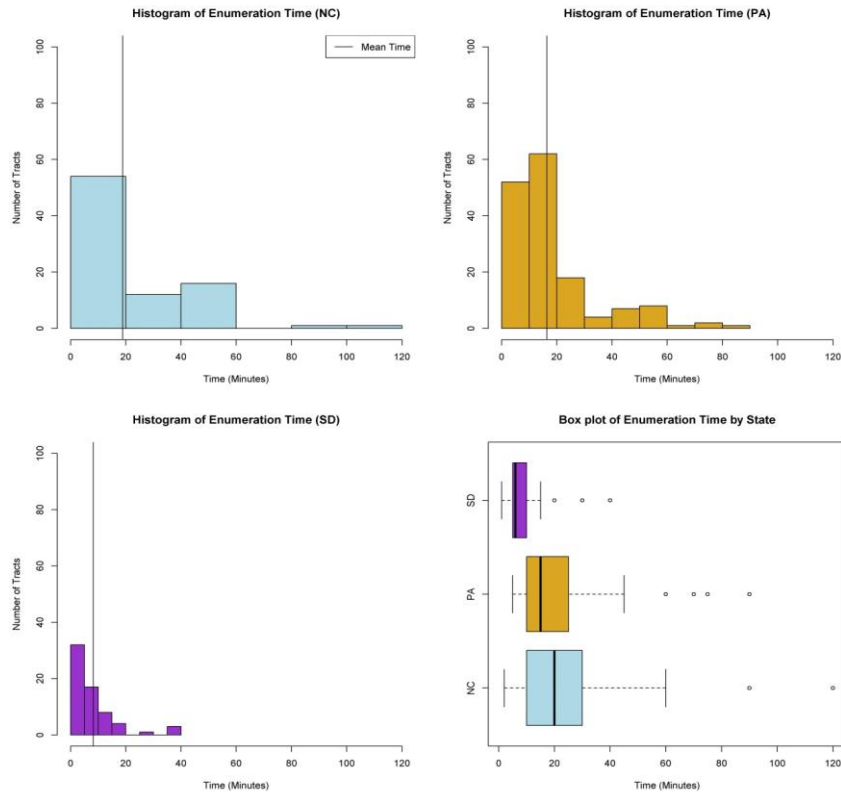


Figure 12: The time (with operator) taken in each state to input tracts into mobile mapping instrument. Blue represents times for enumerators in NC. Yellow represents times for enumerators in PA and purple represents times for enumerators in SD.

Based on a one-way analysis of variance, the mean time enumerators spent with the tract operator (see Table 3) was not the same for all states ($p < 0.0001$). Fisher's Least Significant Difference was then used to conduct pairwise comparisons between states (see Table 4). Only those times for which an interview was completed with an operator were included in the analysis. The mean enumeration times with operators of tracts differed significantly between North Carolina and South Dakota and between Pennsylvania and South Dakota; they did not differ significantly between tracts in North Carolina and Pennsylvania. The average time taken to enumerate tracts in South Dakota with the tract operator is significantly less than that for the other two states. The variability of this time is also significantly smaller in South Dakota than for the other two states.

Table 3: Summary Statistics for Enumeration Time Spent with the Tract Operator for Each State.

State	Number of Tracts	Mean Time (minutes)	Standard Deviation (minutes)
North Carolina	84	24.5	20.1
Pennsylvania	155	20.8	16.8
South Dakota	65	9.7	8.8
All States	304	19.5	17.3

Table 4: P-values for Two Sample T-test Comparing Enumeration Times Spent with the Tract Operator for Each Combination of Two States

<i>P-values</i>			
State	North Carolina	Pennsylvania	South Dakota
North Carolina	NA	0.156	< 0.0001
Pennsylvania	0.156	NA	< 0.0001
South Dakota	< 0.0001	< 0.0001	NA

Linear regression was used to examine the potential association between number of fields in a tract and enumeration time. Although the correlation was not strong ($R^2 < 0.35$ for all states), the time spent with an operator did tend to go up as the number of fields within a tract increased. The estimated slopes from the linear regression coefficients (coefficient associated with fields) from each state are shown in Table 5. The estimated increase in the number of minutes, on average, for each additional field within a tract was 1.7, 1.5, and 0.6 minutes for North Carolina, Pennsylvania, and South Dakota, respectively; all of the slopes were significantly different from 0. A plot of enumeration time with the tract operator versus the number of fields in the tract is shown in Figure 13 along with regression lines for each state.

Table 5: Coefficients and R^2 Values for Regression of Time with Operator vs. Number of Fields

	NC	PA	SD
Beta (Fields)	1.749	1.508	0.6382
<i>p</i> -value	<0.0001	<0.0001	0.0012
R^2	0.34	0.25	0.13

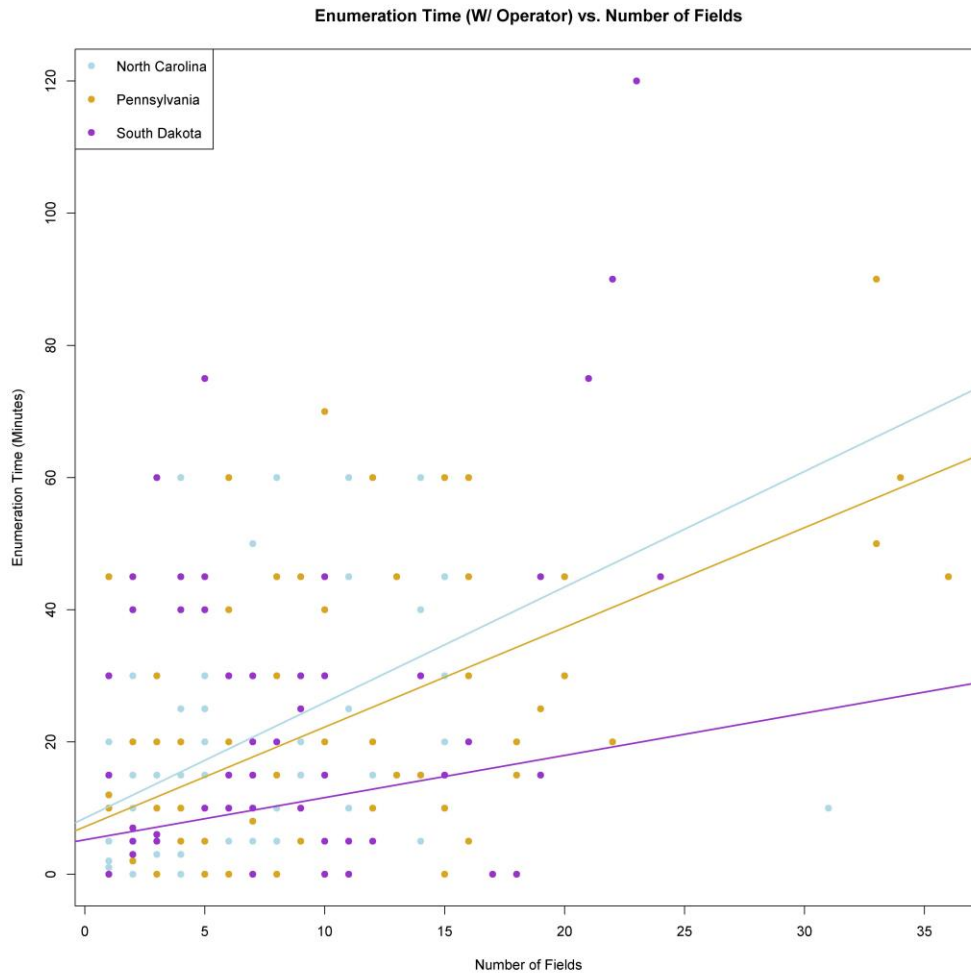


Figure 13: The times (with operator) taken by each enumerator to input tracts into mobile mapping instrument vs. the number of fields in the tract

VI. Conclusions and Future Research

The primary objective of this research was to determine whether grid cells could potentially replace traditional JAS segments. This was evaluated as a cost and resource saving initiative. The lack of physically identifiable boundaries for the grid cells presented a substantial problem for both enumerators and farmers. Enumerators struggled trying to identify where the imaginary grid boundaries fell in order to accurately calculate the acreage for the identified portion of the operation and the agricultural activity within that portion. In addition, farmers struggled when they were required to estimate acreage and production for a small sliver of a field that fell inside the grid boundaries. Also, small slivers added to both the burden of the enumerators who needed to contact all operators and to operators who had to report information on relatively small portions of land. The expectations were that utilizing the PLSS layer was going to be a significant saving initiative; however, even in PLSS states, the grid cells do not necessarily follow the infrastructure on the ground. Another lesson learned is that it took enumerators too much time to draw off the fields with the mobile mapping instrument for the approach to work operationally. This was especially difficult in PA, where enumerators had both many tracts within the grid cells as well as many small fields within each tract. Finally, although the time enumerators spent with the operator drawing off the fields and entering the data into the mobile mapping instrument was recorded, similar information on the current operational approach was not available

for comparison. For the most part, NASS combines the JAS pre-screening and data collection processes with other activities as a cost saving measure. Thus, the JAS estimates for enumeration are not isolated to just the JAS.

The use of grid cells instead of the area frame is not feasible for NASS. However, the mobile mapping instrument is still a promising tool for modernizing the agency data collection activities. Because the instrument can be used with both grid cells and traditional JAS segments, research should continue with the application being on JAS segments. One way to reduce the time taken to enumerate segments is to provide segments with pre-delineated field boundaries. Currently, NASS has an inter-agency agreement with the Farm Service Agency (FSA). FSA prepares common land unit (CLUs) GIS files. Using ArcGIS, NASS segments can be intersected with FSA CLUs and the field boundaries from FSA can be transferred to the JAS segments. Because FSA does not have 100% coverage of all the segments, the expertise of cartographers needs to be utilized to delineate fields for the areas not covered by FSA. The final product delivered to enumerators would be a segment with both FSA and cartographer-drawn field delineations.

VIII. References

Boryan, Claire, Lawson, L., Abreu, D.A., Weaber, J., Gerling, M., Hyman, M., and Hardin, R. (2016). Comparing Geographic Information System (GIS) Calculated Acres to Farmer Reported Acreage Utilizing a Mobile Mapping Instrument. Research and Development Division. RDD Research Report #RDD-16-XX. Washington, DC: USDA, National Agricultural Statistics Service. Pending internal review.

Boryan C., Yang Z., (2012) "A New Land Cover Classification Based Stratification Method for Area Sampling Frame Construction," Proc. in First Intl. Conf. on Agro-Geoinformatics, Shanghai, China.

Cotter, J. Davies, C., Nealon, J., & Roberts R. (2010). Area Frame Design for Agricultural Surveys in Agricultural Survey Methods (eds R. Benedetti, M. Bee, G. Espa and F. Piersimoni), John Wiley & Sons, TD, Chichester, UK, doi:10.1002/9780470665480.ch11

Gerling M., Lawson L., Weaber J., Dotts A., Vardeman A., Wilson E. (2015) "Field Data Collection Using Geographic Information Systems Technologies and iPads on the USDA's June Area Frame Survey", Research and Development Report RDD-15-03, United States Department of Agriculture, National Agricultural Statistics Service.

Lawson, Linda A., Abreu D.A., Boryan, C., Gerling, M., and Hardin, R., 2015. A Summary of Issues Associated with the Modernization of Data Collection. Proceedings of the 2015 Federal Committee on Statistical Methodology.

Attachment A

JAS QUESTIONNAIRE -- SECTION D

SECTION D – CROPS AND LAND USE ON TRACT

How many acres are inside this blue tract boundary drawn on the photo (map)?.....

.

Now I would like to ask about each field inside this blue tract boundary and its use during (year).

Field Number	01	02	03	04	05
1. Total acres in field	828 .	828 .	828 .	828 .	828 .
2. Crop or land use. [Specify]					
3. Occupied farmstead or dwelling	843 .				
4. Waste, unoccupied dwellings, buildings and structures, roads, ditches, etc.	841 .	841 .	841 .	841 .	841 .
5. Woodland	83_ .	83_ .	83_ .	83_ .	83_ .
NP = Not Pastured P = Pastured	<input type="checkbox"/> NP <input type="checkbox"/> P	<input type="checkbox"/> NP <input type="checkbox"/> P	<input type="checkbox"/> NP <input type="checkbox"/> P	<input type="checkbox"/> NP <input type="checkbox"/> P	<input type="checkbox"/> NP <input type="checkbox"/> P
6. Pasture					
Permanent (not in crop rotation)	842 .	842 .	842 .	842 .	842 .
Cropland (used only for pasture)	856 .	856 .	856 .	856 .	856 .
8. Idle cropland – idle all during (year)	857 .	857 .	857 .	857 .	857 .
9. Two crops planted in this field or two uses of the same crop.	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No
[Specify second crop or use.]					
Acres	844 .	844 .	844 .	844 .	844 .
10. Acres left to be planted	610 .	610 .	610 .	610 .	610 .
11. Acres irrigated and to be irrigated [If double cropped, include acreage of each crop irrigated.]	620 .	620 .	620 .	620 .	620 .
16. Winter Wheat	540 .	540 .	540 .	540 .	540 .
17. (include cover crop)	541 .	541 .	541 .	541 .	541 .
20. Oats	533 .	533 .	533 .	533 .	533 .
21. (include cover crop)	534 .	534 .	534 .	534 .	534 .
24. Corn[exclude popcorn and sweet corn]	530 .	530 .	530 .	530 .	530 .
25. For grain or seed	531 .	531 .	531 .	531 .	531 .
29. Other uses of grains planted (Abandoned, silage, green chop, etc.)	Use				
Acres
30. Hay	653 .	653 .	653 .	653 .	653 .
31. [Cut and to be cut for dry hay.]	656 .	656 .	656 .	656 .	656 .
33. Other Hay	654 .	654 .	654 .	654 .	654 .
34. Soybeans	600 .	600 .	600 .	600 .	600 .
35. Following another harvested crop	602 .	602 .	602 .	602 .	602 .
51. Other crops	---	---	---	---	---
Acres planted or in use

Attachment B

ENUMERATOR EVALUATION FORM

[illegible]