IMPLEMENTATION OF A NEW AUTOMATIC STRATIFICATION METHOD USING GEOSPATIAL CROPLAND DATA LAYERS IN NASS AREA FRAME CONSTRUCTION

Claire G. Boryan, Zhengwei Yang
USDA National Agricultural Statistics Service
3521 Old Lee Highway, Room 305, Fairfax, VA 22030, U.S.A.
Email: Claire.Boryan@nass.usda.gov

ABSTRACT

A new automatic stratification method utilizing USDA National Agricultural Statistics Service (NASS) geospatial Cropland Data Layers (CDLs) was recently implemented in NASS operations. Recent research findings indicated that using the CDL stratification method rather than visual interpretation of satellite imagery and aerial photography (traditional method) to define percent cultivation of land areas resulted in Area Sampling Frames (ASF) constructed with improved accuracy, objectivity and efficiency at reduced cost [3]. This paper describes an operational ASF construction process that integrates the automated CDL stratification results with traditional editing/review procedures, a hybrid approach. New 2013/2014 ASFs for South Dakota and Oklahoma were successfully built using the new operational process and illustrated significant improvements in frame accuracy, operational efficiency, and cost.

Index Terms— Area sampling frame (ASF), automated stratification, cropland data layer (CDL), cultivated layer, land cover-based stratification.

1. INTRODUCTION

The USDA National Agricultural Statistics Service (NASS) has produced state level Cropland Data Layers (CDLs) for major US agricultural states since 1970 and for all 48 conterminous states since 2009 [1]. NASS uses CDLs to produce acreage estimates of major crops in agriculturally intensive states using a linear model which regresses the CDL pixel data to segment summary data collected from the NASS June Area Survey (JAS) [2]. Boryan and Yang used CDL data in Area Sampling Frame (ASF) construction to automatically and objectively stratify land cover in the U.S. based on percent cultivation [3]. Research findings indicated that the automated CDL stratification significantly improved stratification accuracies in intensively cropped areas and performed less well in non-agricultural areas as compared with traditional stratification based on visual interpretation for five test states. The differences in accuracies were statistically significant at a 95% confidence level [3].

A model to create multi-temporal cultivated data sets to use as the input to the automated stratification procedure was suggested [4]. Based on this research, these cultivated data sets were made available to the public on the NASS online geospatial application CropScape [5].

NASS has implemented automated stratification into the NASS Area Frame operational process as the primary stratification method for building new state Area Frames as illustrated in flow chart Fig. 1. The implementation integrates the automated CDL stratification with the traditional method. This new operational process utilized results of both traditional and automatic CDL stratification methods and further refined frames with manual editing procedures. This paper describes implementation specifics and highlights gains in accuracy and efficiency that have been achieved with the adoption of the new automated method. NASS Area Frame construction techniques serve as a model around the world and this paper's goal is to describe the newest techniques utilized to update the NASS Area Frames.

2. SCOPE AND DATA

The implementation of automated stratification in NASS is for daily operation. The operational process is developed for constructing all State frames. In this paper, the state of South Dakota (SD), as shown in Fig. 2, is utilized as an illustration of the implementation of automated stratification in NASS. South Dakota is an agriculturally intensive state located in the North/Central portion of the U.S. which is a major contributor to U.S. agricultural production. Analysis conducted in NASS on the South Dakota Area Frame demonstrated the significant differences between 1) traditional and automated stratification results and 2) several major crops were not meeting target coefficients of variation

U.S. Government Work Not Protected by U.S. Copyright

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:52:43 UTC from IEEE Xplore. Restrictions apply.

(CV), which in combination highlighted the need for refinement of the South Dakota frame. In this experiment, the existing 2009 South Dakota Traditional Area Frame (shape file format) was used as the basis for the automated stratification procedure. Multi-temporal (2008-2012) CDLs and their derived cultivated data sets were used as the input to this process. Landsat data, National Agricultural Imagery Program (NAIP) aerial images, Farm Service Agency (FSA) Common Land Unit (CLU) data, topographic maps, U.S. Census data and state specific ancillary data sets were used in the South Dakota Area Frame refinement.

3. IMPLEMENTATION INTO NASS OPERATIONS

Based on the research conducted by Boryan and Yang [3], [4], and others' [6], NASS has implemented CDL automated stratification into Area Frame operations. When new frames were built in the past, all analyses were conducted from scratch with 100% of the PSUs requiring digitization and all stratum definitions determined using visual interpretation. This traditional method, which was subjective and very expensive, required, on average, five full time employees for six months per state [6].

To improve objectivity, efficiency and reduce the cost of construction, the new automatic CDL stratification method was developed to stratify every PSU based on its percent cultivation. The new automatic CDL stratification method improved stratification accuracy significantly for highly cultivated agricultural areas as compared with the traditional method [3]. However, the new automatic CDL method is still an error-prone method. To further improve the accuracy of new frames, the area frame construction process was rejuvenated by integrating both methods into a new process.

In the new process, the old Area Frames are currently being reused instead of starting from scratch. The CDL automated stratification is used as the primary stratification tool. In this process, PSUs that have potentially changed stratum definitions are identified by comparing the existing old Area Frames with the result of the CDL automated method. Non-matching PSUs (those PSU stratum definitions that are not consistent between the two methods) are targeted for manual inspection. While automated stratification alone achieved higher accuracies than the traditional stratification in the highly cultivated strata in previous research [3], NASS staff determined that improved accuracies could be achieved with additional manual inspection and editing of PSUs, a procedure coined "geospatial stratification" which is a hybrid process incorporating both stratification results followed by a targeted manual stratification refinement.

There are three primary reasons that the results of automated stratification alone can be improved using manual inspection and editing. First, it is common that PSUs, particularly in stratum 12 (>50-75% cultivated) and stratum 20 (15-50% cultivated) cannot be broken down into homogeneous segments. The cultivated land tends to be clustered within the PSU. Consequently, when the PSUs are automatically stratified, although accurate (meet stratum definition) at the PSU level, they cannot be broken down into homogeneous segments. Second, boundary and population changes occur which require manual editing of PSUs to update the frame to more accurately reflect current conditions. Third, the CDLs, although highly accurate (85-95% for major crops in major producing states) are not 100% accurate and local areas can be impacted by CDL errors of omission and commission that can be identified and accounted for through manual inspection.

The specific steps used to conduct traditional stratification and to conduct CDL automated stratification were listed in [6] and [3] respectively. The flowchart Fig. 1 illustrated the operational area frame production process, which integrated the CDL automated stratification method into the process as a primary stratification method. The PSUs, of the existing Area Frame produced using the traditional stratification method, are reused in automated stratification. Once automated stratification is completed, the manual procedures, termed "geospatial stratification", for frame refinement are followed to compare the new results with the existing traditional frame and identify PSUs that need to be improved through PSU editing. The geospatial stratification steps to conduct an integrated stratification, manual inspection and editing of PSUs are given as follows:

1) Input automatically stratified frame
2) Overlay satellite imagery and NAIP data
3) Obtain Farm Service Agency (FSA) Common Land Unit (CLU) data as ground truth to verify changes in the ASF
4) Examine PSUs that don't match traditional strata definitions and exceed PSU size tolerances
5) Identify areas with boundary (roads, railroads, river/streams) and population changes (based on auxiliary data).
6) Identify localized areas with CDL errors of omission and commission using FSA CLU data, NLCD, 2006, as well as satellite imagery and aerial photography based visual inspection.
7) Manually edit the original ASF PSUs using ESRI's Arc GIS based on the results of steps 4-6.
8) Rerun the automated stratification procedure on the newly updated ASF as a final review step to identify PSUs in which the CDL automated results do not match with the current stratum definitions.
9) Conduct a final review of these non-matching PSUs to determine the appropriate stratum definition for each.

This integrated process is currently being used in NASS area frame operation.

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:52:43 UTC from IEEE Xplore. Restrictions apply.

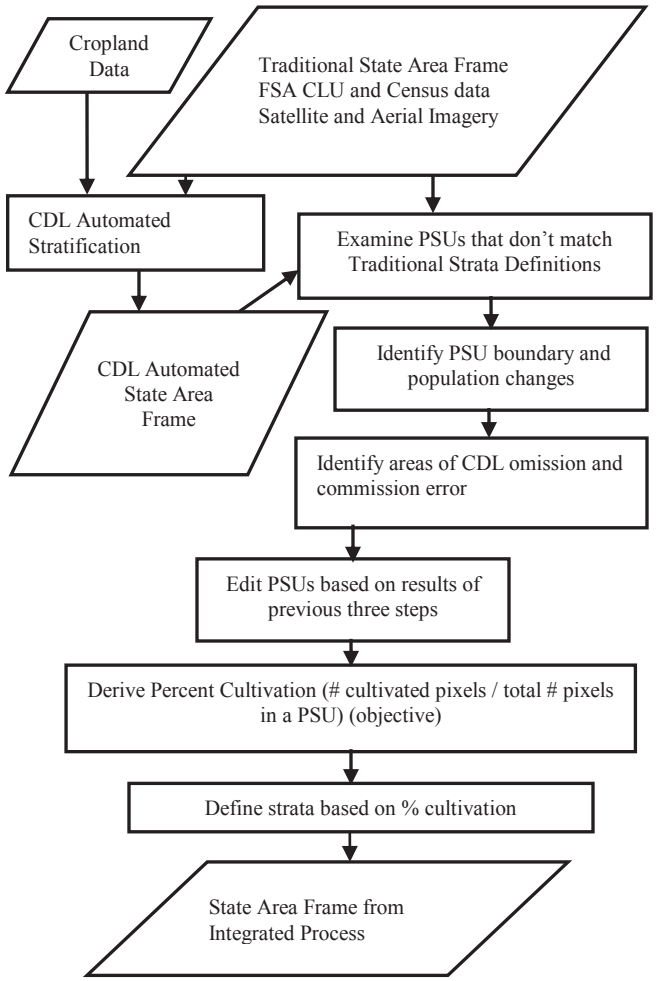

Figure 1. Flowchart illustrated integrated process based on automated stratification, of existing traditional Area Frames, and manual editing/ review of PSUs.

4. Results And Discussion

NASS successfully rebuilt Area Sampling Frames of South Dakota and Oklahoma. For example, in rebuilding the South Dakota State Area Frame, the original frame, which was stratified using the traditional stratification method, was reused as a reference as shown in Fig. 2 (left). The new Area Frame was constructed as shown in Fig. 2 (center) by using the CDL automated stratification method on the original frame PSUs. It is observed that there are big differences in the highly cultivated areas between the two stratified frames from two different stratification methods. As shown in Fig. 2 (center), the frame created using the CDL automated method is detailed and more accurately reflects conditions on the ground in highly cultivated areas while the traditional method yields more variations in the low cultivated areas as shown in Fig. 2 (left). The frame generated with the CDL automated method could be used as

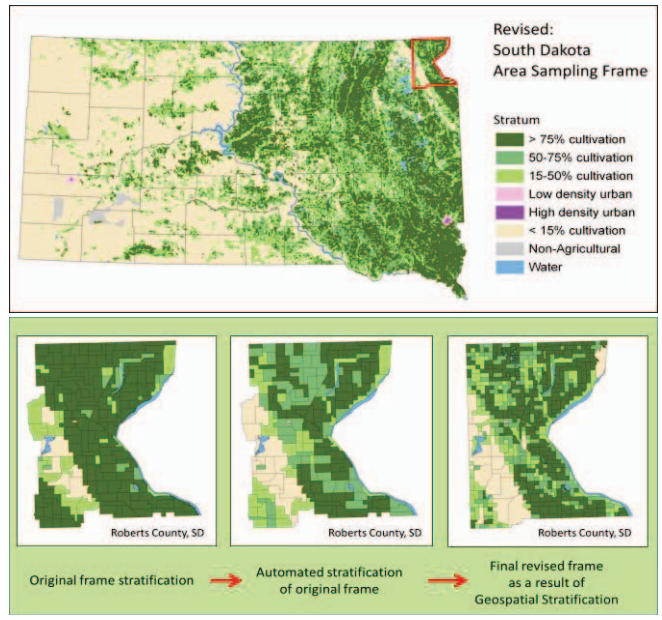

Figure 2. 2014 South Dakota Area Frame created using CDL based automated stratification techniques. Zoom of Roberts County, SD. The Area Sampling Frame (ASF) on the left was constructed using the traditional method (visual interpretation). The center ASF was constructed using automated stratification of the original frame. The right (ASF) was constructed using geospatial stratification which combines automated and manual stratification. PSUs are reduced in size which improves PSU homogeneity and manual editing improves frame accuracy.

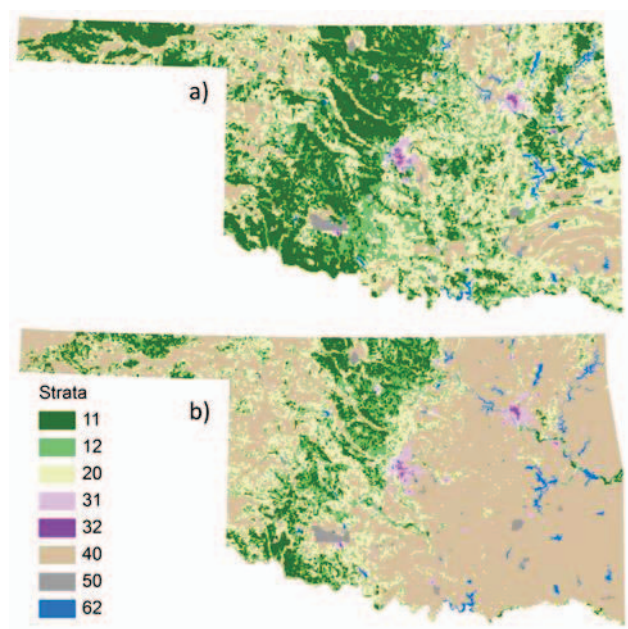

Figure 3. Oklahoma Area Sampling Frames (a) 2010 and (b) 2013.

an update of the South Dakota frame. However, the frame can be further improved by reducing PSUs' size making them more homogenous. Therefore, with this implemented new process, a number of manual procedures have been

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:52:43 UTC from IEEE Xplore. Restrictions apply.

applied on the frame derived with CDL automated method. The final refined new frame provides, as shown in Fig. 2 (right), much more details in both high and low cultivation areas. The stratification accuracy has been further improved as the PSUs' homogeneity has been significantly improved.

Fig. 3 illustrates the Oklahoma state frame updated using the new integrated process. As shown in Fig. 3 (a), the new 2013 frames have improved significantly as compared with old 2010 frame. Stratum 11 (greater than 75% cultivated) was significantly overestimated in the 2010 ASF. The analysis of the 2010 Oklahoma frame (traditional stratification) indicated an accuracy of only 34% of segments meeting stratum definition when compared with the 2012 June Area Survey (JAS) segment data. After rebuilding the 2013 Oklahoma frame, an accuracy of 79% was achieved using the 2013 JAS segment data as in situ validation. This example again shows the effectiveness of the implementation of the integration of the new CDL automated method with the old operational area frame production process.

Overall, with the implementation of the CDL automated method into NASS area frame operations, new state frames are being built at reduced cost with improved objectivity, efficiency and accuracy. For example, the 2010 Oklahoma area frame was constructed in 4552 employee hours using the NASS traditional method as compared with 2013 Oklahoma frame which was constructed in 1980 employee hours using the automated stratification method as the primary stratification method.

5. CONCLUSION

The implementation of the new CDL based automated stratification method into NASS Area Frame operations has brought about significant improvements. The accuracy of the new Oklahoma Area Frame increased from 34% to 79% accuracy with the updated ASF. New 2013/2014 ASFs for South Dakota and Oklahoma were successfully built using the new operational process, which integrated the CDL automated stratification method and manual stratification and illustrated significant improvement in accuracy, efficiency, and cost.

6. REFERENCES

[1] C. Boryan, Z. Yang, R. Mueller, and M. Craig, Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service Cropland Data Layer Program, *Geocarto International*, 2011, 26, (5): 341-358.

[2] C. D. Day, "A compilation of PEDITOR estimation formulas". NASS/RDD Research Paper RDD-02-03, USDA, NASS, Washington, D.C. January, 2002.

[3] C. Boryan, Z. Yang, L. Di, and K. Hunt, "A New Automatic Stratification Method for U.S. Agricultural Area Sampling Frame Construction Based on the Cropland Data Layer," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 1939-1404, May 2014, DOI: 10.1109/JSTARS.2014.2322584.

[4] C. Boryan., Z. Yang , and L. Di. "Deriving 2011 cultivated land cover data sets using usda national agricultural statistics service historic cropland data layers," Proc. of IEEE International Geoscience and Remote Sensing Symposium, July 22-27, 2012, Munich, Germany.

[5] W. Han, Z. Yang., Li Di., and R. Mueller, "CropScape: A Web service based application for exploring and disseminatingUS conterminous geospatial cropland data products for decision support." Computer and Electronics in Agriculture, Vol. 84, June, pp. 111–123.

[6] Cotter, J., Davies, C., Nealon, J. and 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, Ltd, Chichester, UK. doi: 10.1002/9780470665480.ch11

Authorized licensed use limited to: University of North Texas. Downloaded on December 18,2020 at 12:52:43 UTC from IEEE Xplore. Restrictions apply.