Remote Sensing of Environment 232 (2019) 111286

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Using the Landsat archive to map crop cover history across the United States

David M. Johnson*

U.S. Department of Agriculture, National Agricultural Statistics Service, United States of America

ARTICLE INFO

Keywords:
Landsat
Google Earth Engine
Croplands
Classification
Time series
Retrospective

ABSTRACT

Landsat Thematic Mapper has been collecting multispectral imagery at 30 m resolution globally since 1984. One utility of the data has been for detailed mapping of agricultural regions and seasonal identification of crops grown within them. However, the ability to do so has only been applied sporadically and eluded widespread adoption due to cost of the imagery, burdensome preprocessing requirements, and computing not being up to the task. These hurdles have become much reduced of recent with the free and open distribution of the Landsat imagery, emphasis on ready-to-use surface reflectance data products, and distributed high performance computing infrastructures available online in "the cloud." As such, this work leverages these aspects and investigates the ability to retrospectively map summer crops over the United States (US) annually from 1984 to 2007. Google's Earth Engine Internet-based analytical platform containing the historical Landsat archive in surface reflectance format was used as a foundation for the classification work. Robust 30 m Cropland Data Layer (CDL) information from US Department of Agriculture (USDA) for years 2008 through 2011 were leveraged to train rule-based classifiers which were applied back through time to each year 1984 through 2007. Focus crops were corn, soybeans, and winter wheat – the three largest by area in the US. A large sampling of highly intensive counties throughout the country were prototyped for generation of the 24 years of historical crop cover. For validation, crop area statistics were calculated for each county-year and compared to survey-based information existing from the USDA. Results were muted overall with the average crop area coefficient of determination (R^2) correlations for the years 1984–2007 found to be 0.192, 0.159, and 0.142 for corn, soybeans, and winter wheat, respectively. Furthermore, the standard deviations were variable at 0.132, 0.177, and 0.133, also respectively. While unimpressive, it was found as a benchmark that the R^2 between the 2008 through 2017 CDL classifications were only 0.478, 0.686, and 0.726 and thus a suggestion that the USDA area statistics are an imperfect measure of map accuracy. Deletion of approximately one third to one half of the grossest 1984–2007 outlier years from the historical outputs pulled the correlations to the benchmark standard. Qualitatively, most of the remaining years classified looked of high quality and were believed useful as field-level thematic crop area maps. These historical cropland maps could provide the ability to better detail the role farming has played on the broad US landscape over recent decades.

1. Introduction

The ability to identify and map the location and extent of crops through time is societally important for many reasons. The most fundamental is being able to monitor and forecast the amount of food production across given regions (Gallego et al., 2010; Thenkabail et al., 2012), ranging to the more holistic in trying to understand how agricultural fields change in size (Ferguson et al., 1986; Yan and Roy, 2016) and interface with and impact the surrounding ecosystems now and will into the future (Ramankutty and Foley, 1999; Bruinsma, 2003; Fritz et al., 2015). Space-based remotely-sensed data from Landsat (Loveland and Dwyer, 2012; Goward et al., 2017), which began providing 30 m resolution multispectral data in 1984, has been the most suitable platform for observing and tracking regional areas of croplands across the globe. However, hurdles including cost, accessibility, calibration consistency, and lack of computing prowess have hindered the full exploitation of the imagery even though it has been around for decades and sophisticated land cover change methodologies have been brought forth (Sexton et al., 2013; Zhu and Woodcock, 2014; Vogelmann et al., 2016). Recently though those barriers have been falling given the Landsat archive being made freely available (Wulder et al., 2012), consolidated (Wulder et al., 2016) and distributable over the Internet, processed to surface reflectance (Claverie et al., 2015), and analyzable with high performance, scalable "cloud" computing with systems like

* Corresponding author at: 1400 Independence Ave SW, Washington, DC 20250, United States of America.
E-mail address: dave.johnson@nass.usda.gov.

https://doi.org/10.1016/j.rse.2019.111286
Received 31 October 2018; Received in revised form 31 May 2019; Accepted 25 June 2019

Google Earth Engine (GEE; Gorelick et al., 2017). This convergence of factors now finally allows for massive amounts of Landsat data to be ingested and analyzed across large geographies over dozens of years, and as such global 30 m Landsat-based change products leveraging GEE do now exist for forests (Hansen et al., 2013) and surface waters (Pekel et al., 2016). Nothing of this type or scale has been produced for croplands, however.

Therefore, *the purpose of this research is to investigate and prototype the ability to map at the field-level the summer crops over the United States (US) annually from 1984 through 2007*. If successful this would mimic Cropland Data Layer (CDL) classifications (Johnson and Mueller, 2010; Boryan et al., 2011) that already exist from 2008 to present and significantly extend back in time the synoptic knowledge of what crops were growing where across the US. Having this 30-plus year history is important because it would provide a field-level understanding of how agricultural land cover uses, and impacts to the surrounding landscape, have evolved due to farming. Furthermore, and more vitally, lessons learned from retrospective mapping can provide insight toward methods for real-time crop mapping both in the US and abroad. This has big relevancy for the rapid and objective estimation of the food supply from local to global scales.

A methodological foundation for retrospective mapping exists given the Landsat 5 mission had a 25 + year lifespan and remained operational through 2011 with the last four years overlapping with the first four epochs of national-level CDLs. Furthermore, Landsat 7 supplements that dataset having become available beginning in 1999 and continuing to operate (Markham and Helder, 2012; Kovalskyy and Roy, 2013). Combined they provide a unique opportunity to derive generalized spectral and temporal signatures for the creation of land cover type decision rules that can be applied to classify imagery annually back to 1984. The entire Landsat archive is extremely large but the online GEE platform manages a copy of the data alongside a freely accessible and scalable high performance image processing computing infrastructure developed for research purposes. Without GEE, the undertaking of this type of effort would be a huge undertaking.

Crop extent datasets do already exist over the US, and throughout the globe, but are sporadic in their era, purpose, and quality. The flagship crop cover product over the US is that of the Department of Agriculture's CDL. It has provided 30 m resolution summer circa information across the country since 2008. For certain states the time series extends even further back in time (in particular Arkansas, Illinois, and Mississippi go back to 2000, and North Dakota 1997). The focus of the CDL is for documenting crop area on the major commodities like corn, soybeans, and wheat, but in total over 100 crops, and even generalized non-cropped categories, are mapped. Accuracies are usually very good (Lark et al., 2017), especially for the dominant categories in intensively farmed regions. This is particularly true since 2008 at which time input data became substantial and the methodology standardized. The National Land Cover Database (NLCD) has epochs for years 1992, 2001, 2006, 2011 (Homer et al., 2015) but only contains a single combined class incorporating all crop types. It does compliment the CDL well though given they share the same map projection and grid system, have synergetic land cover classes, and the older products have been improved upon with newer methods to better allow for change analysis. In terms of global 30 m products there are the GlobeLand30 (Jun et al., 2014) and the Global Food Security Analysis-Support Data at 30 m (Massey et al., 2017). They are each single year efforts, circa 2010 and 2015 respectively, providing information that delineates crop areas from non-crop areas. Thus, there are no specifics about what crops are grown where and no ability to provide insight about changes that have occurred through time.

2. Data

The primary imagery dataset exploited for this work is that from Landsat 5 Thematic Mapper (TM) instrument spanning effectively the entire archive from 1984 through late 2011. This remarkably long time series provided 28 summer crop seasons in which imagery data were collected. Landsat 7 Enhanced Thematic Mapper Plus Data (ETM +) was also integrated for the years 1999, the time it was commissioned, through fall of 2011. This provided supplemental imagery for the latter 13 growing season. For both the TM and ETM + products only the "Tier 1" products were used as they represent the highest quality imagery meeting certain radiometric and geometric requirements. Furthermore, the data used were those processed to surface reflectance which help minimize atmospheric effects. The surface reflectance was calculated using the Landsat Ecosystem Disturbance Adaptive Processing System algorithm (Masek et al., 2006; Maiersperger et al., 2013) The processing of surface reflectance, and overall management of the data, was all maintained with GEE and referenced as LANDSAT/LT05/C01/T1_SR and LANDSAT/LE07/C01/T1_SR for the analysis coding. To reduce the computational burden, only the red, green, near-infrared, and two shortwave channels (bands 2, 3, 4, 5, and 7) were utilized.

USDA CDL information from the years 2008 through 2011 was relied upon as a surrogate data for training the land cover classifier. Again, these four years were important because they overlapped with the last four years of Landsat 5 TM data. The CDL information is also inherent within GEE and referenced as USDA/NASS/CDL/{YYYY} where YYYY is the year of interest. The category values of the CDL are the same within GEE as the native product which, for reference, are 1 = corn, 5 = soybeans, and 24 = winter wheat. Category 26 is also important in this context and tracked as it represents double-crop winter wheat/soybeans (meaning a field had winter planted wheat and then soybeans harvested within the same calendar year).

A few ancillary gridded layers were also included as predictor variables alongside the Landsat imagery. Those included 30 m elevation from the US Geological Survey's (USGS) National Elevation Dataset (NED; Gesch et al., 2002), impervious and percent tree cover from the 2011 NLCD, and USDA cultivated crop layer from 2013 (the earliest available). Respectively, they were referenced within GEE as USGS/NED, USGS/NLCD/NLCD20011, USDA/NASS/CDL/2013. All of these, like Landsat TM and ETM + are natively gridded to 30 m. It should be noted the inclusion of these extra data layers is debatable but they were added by USDA for their contemporary CDL production, as known to improve the land cover manifestation, so were included here as well for consistency.

3. Methods

The crop type classification methodology, all of which was performed in GEE, can be summarized as: 1) Landsat 5 surface reflectance data, and for years available Landsat 7 also, were first cloud screened and then median composited by spectral band into four seasonal epochs (roughly, northern hemisphere's winter, spring, summer, and fall), 2) these four seasonal composites for each year 2008 through 2011 were "stacked" and intersected with the respective CDLs and random samples drawn for each year across the scene for all cover classes available, 3) the four years of training samples were combined to provide a generalized distribution of predictor cases, 4) the samples were trained using a Classification and Regression Tree (CART) classifier (Breiman et al., 1984) to derive a ubiquitous set of decision tree rules, and 5) the rules were applied to each of the 1984 through 2007 annual four season composite stacks to derive a land cover classification for each of the years. Analysis was constrained to the county level so as to not extrapolate across the landscape to where the crop timing and management might vary. For validation a final step was to calculate the area of particular crop within the county and compare that to the corresponding USDA published planted area statistics (USDA, 2017).

Step one, the creation of seasonal cloud-free median imagery composites, first involved selecting optimal time span widths. Ideally, a dense time-series of composites would be created throughout the growing season to best capture the differences in phenology across the crops (USDA, 2010) as this, alongside multispectral information, aids in crop type discrimination (Song et al., 2017; King et al., 2017). However, given the fixed 16-day revisit rate of Landsat, and there being a certain probability of clouds or other atmospheric effects on a given day, too narrow in time of compositing time windows would often lead to areas of missing data. Through trial and error testing, it was found that a 64-day composite period (equivalent to four overpasses from a single Landsat sensor) was a reasonable compromise between having a dwell time long enough to have usually collected at least one cloud free scene yet not too long as to overly integrate through time what could be rapidly changing ground cover.

Next was to determine the start and end points for the entire growing season using the 64-day time windows. Through inspection it was estimated that four spans starting March 6 (day of year 65) and running through November 16 (day of year 320) generally fit the growing season well throughout the US. This left seasonal breakpoints at May 9, July 12, and September 14 given the 64 day constraint. The span midpoints, and thus the most representative single date of the composite times, were ultimately April 7, June 10, August 13, and October 16. Imagery prior to or after the full growing season dates were found to be overly compromised with clouds, snow, and/or low sun angle so not integrated into the analysis.

Fig. 1 depicts the bounds of the four time window ranges and summarizes US crop phenology as described via the normalized difference vegetation index (NDVI) as acquired from the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard the Terra satellite (adopted from Johnson, 2016). Ultimately the 64-day composites were generated after first removing any individual scene pixels flagged in the surface reflectance metadata layer highly probable as clouds, cloud shadow, or snow and then running a median calculation on those pixels left. This effectively reduced the best parts of several images over a 64 day period to one, spectral band by band. Areas which were inclined to be less cloudy or farther north (where there is more overlap between Landsat paths) tended to have cleaner looking composite results. Typically, there were found to be several Landsat images going into each composite although many of the individual scenes may not have covered in entirety the study area of interest. The years 1999 forward, with the inclusion of the Landsat 7 in addition to Landsat 5, had even more availability of imagery. Often a dozen scenes or more were available for these later years and thus there was a better chance of a composite output being generated without missing data or obvious scene edges. Fig. 2 provides a visual example of the seasonal composites created for Cedar County, Iowa years1984 and 2007.

While the ultimate goal would be to generate a US-level classification for all the years 1984–2007, pragmatically this was still a computation challenge even within GEE. Furthermore, it is best to constrain samples, and the training rules derived from them, to localized areas since cropping practices and timing vary across the vast US. Thus, the sampling and land cover classification analysis were constrained to the county-level. This was also a natural fit because detailed USDA crop area statistics exist at that scale as well.

To generate enough statistics to provide insight to how well the methodology would work nationally, a random sampling of 30 counties for each of the crops of corn, soybeans, and winter wheat was gathered. There are a total of 3007 counties in the US but many of them do not contain crops or have small areas that do not warrant statistical significance. So, to get the sample counties for each commodity they were first randomly ordered and next stepped through to assess which contained a full history of yearly planted areas statistics (albeit allowing for a single missing value for the 1984 through 2007 period). A map of the counties randomly selected is shown in Fig. 3 and listed in Table 1. Of note, Carroll, Maryland was randomly selected for both soybeans and winter wheat and Marion, Tennessee for both corn and soybeans. Thus, while there were 90 total analyses there were only 88 unique counties investigated.

For each of the 90 county-crop study sites, pixel-level training samples for the years 2008 through 2011 were chosen at a rate of one tenth of 1%. First though the boundary of the county was extended outward by 3 km beforehand to aid the classifier along the county edges. For a typical county there would be a few thousand random pixel-level samples chosen per year with the total amount ultimately a function of the county area. After combining the four years of training

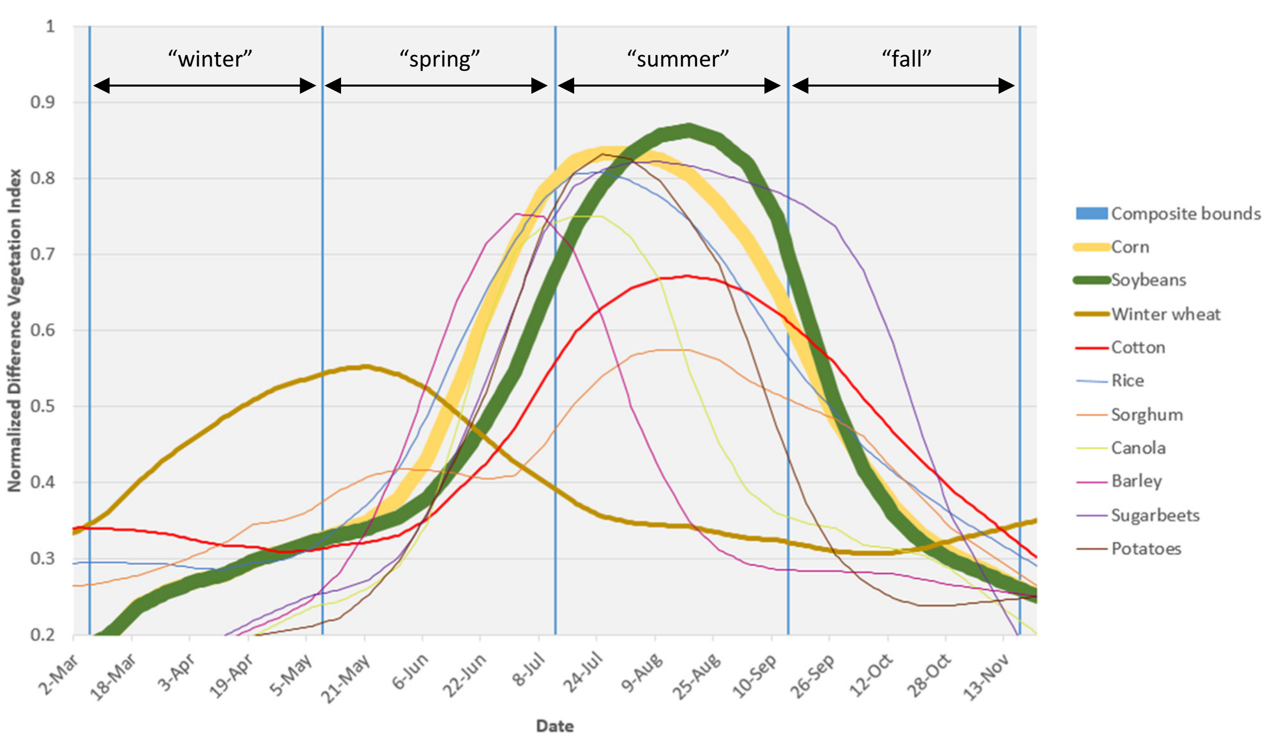

Fig. 1. Average NDVI profiles for major commodities over US with 64-day compositing window bounds. Several crops are shown even though the emphasis here is ultimately corn, soybeans, and winter wheat. The thickness of the line depicts the relative planted area.

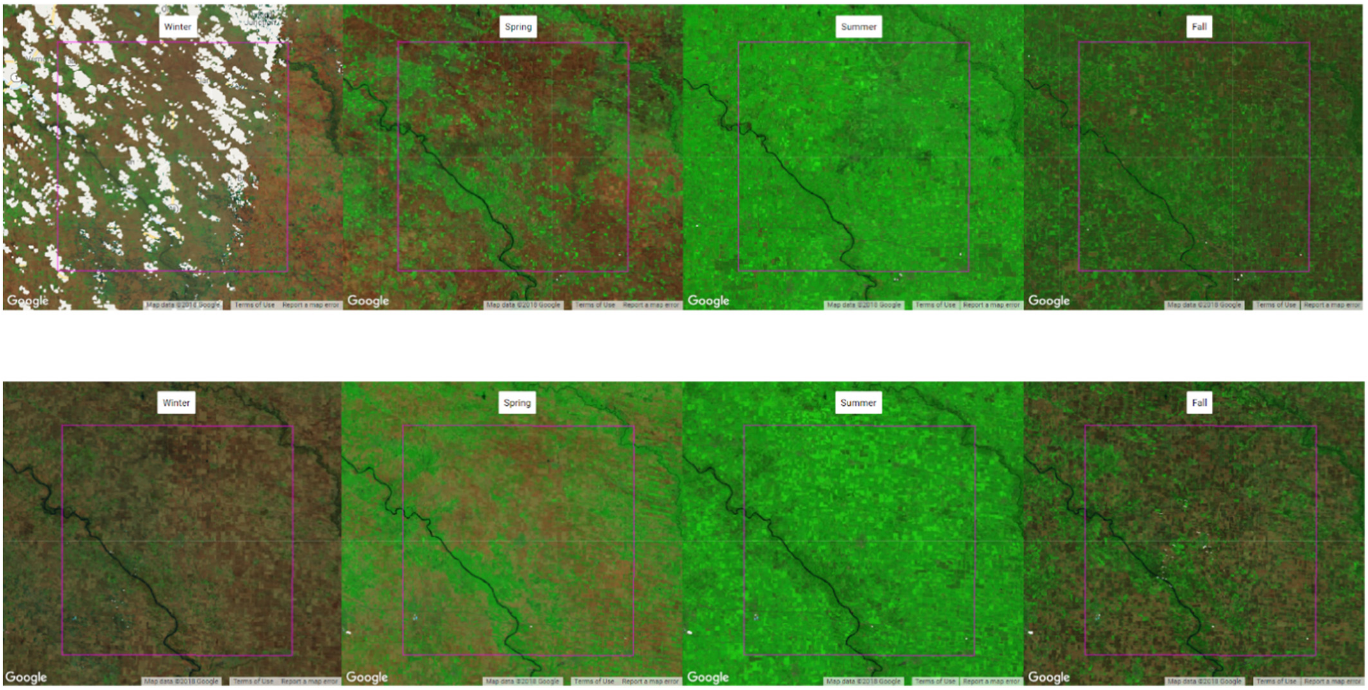

Fig. 2. Example 64-day generated median composites for four seasonal windows for the initial analysis year 1984 (top) and the final analysis year 2007 (bottom) for Cedar County, Iowa. Left to right the scenes are representative of the dates April 7, June 10, August 13, and October 16.

points together a dataset with roughly 10 to 20 thousand points existed.

Next these training points were fed to the classifier. GEE houses several classifiers and after trial and error the CART was decided upon given its output results which in most cases were similar to the Random Forest option but noticeably faster in computation. Other options of the dozen or so were also investigated but gave poor results or were too computationally intensive. The use of CART within GEE has been shown to be the best by others too (Shelestov et al., 2017).

Subsequently, the CART decision rules as derived from the combined 2008 through 2011 training samples were applied to each of the

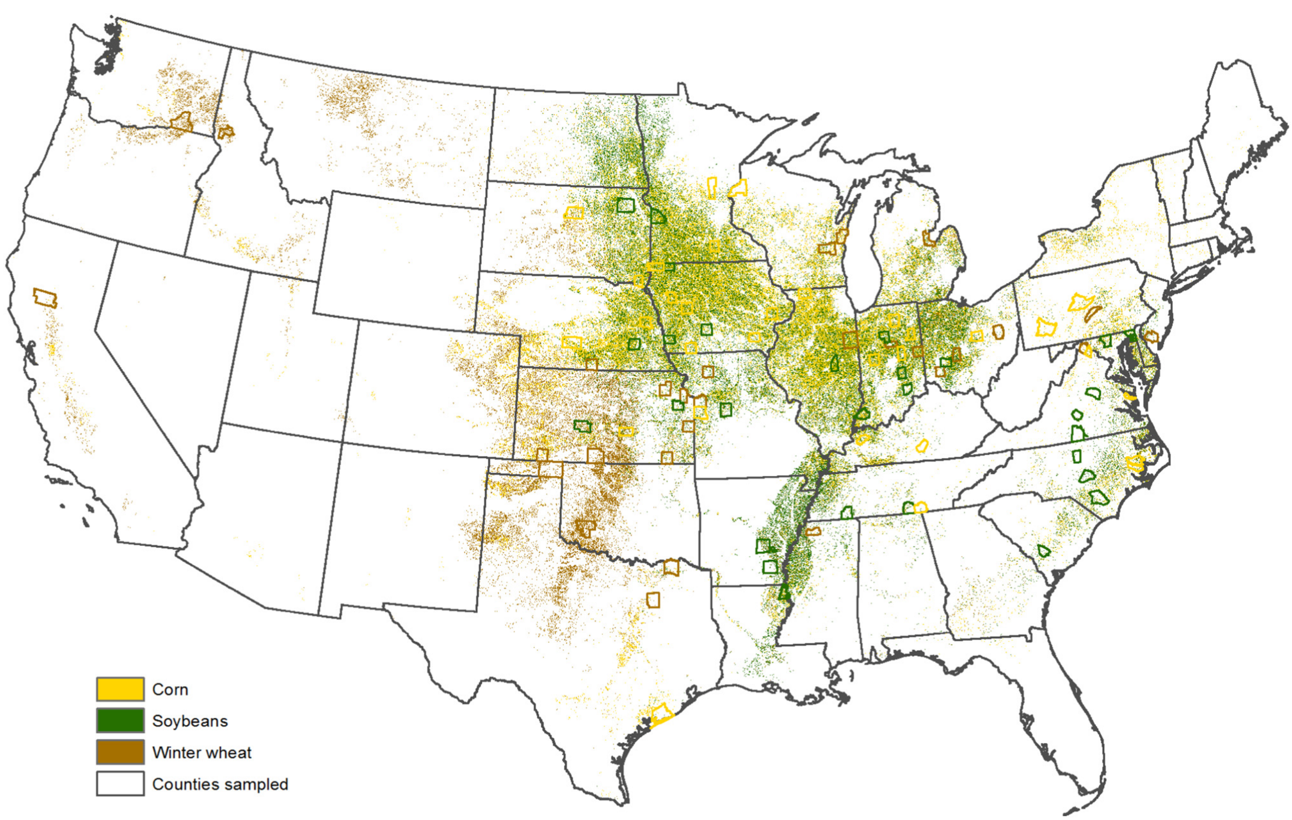

Fig. 3. Location of the counties randomly sampled alongside the 2017 distribution of corn, soybeans, and winter wheat areas. The county boundary color represents the crop type focused on for that area. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
List of counties randomly selected and analyzed by crop of interest.

Sample county	Corn	Soybeans	Winter wheat
1	Cedar, Iowa	Cecil, Maryland	Montgomery, Kansas
2	Centre, Pennsylvania	Louisa, Virginia	Linn, Kansas
3	Kosciusko, Indiana	Jefferson, Arkansas	Iroquois, Illinois
4	Potter, South Dakota	Halifax, Virginia	Juniata, Pennsylvania
5	Dodge, Nebraska	Cass, Indiana	Fond du Lac, Wisconsin
6	Beaufort, North Carolina	Lac qui Parle, Minnesota	Jackson, Missouri
7	Madison, Indiana	Douglas, Kansas	Randolph, Indiana
8	Knox, Ohio	East Carroll, Louisiana	Carroll, Maryland
9	Stephenson, Illinois	Madison, Iowa	Franklin, Nebraska
10	Ida, Iowa	Piatt, Illinois	Jackson, Kansas
11	Mille Lacs, Minnesota	Appomattox, Virginia	Bay, Michigan
12	Webster, Kentucky	Gibson, Indiana	Howard, Indiana
13	Cass, Missouri	Shelby, Indiana	Madison, Ohio
14	Middlesex, Virginia	Greene, Ohio	Walla Walla, Washington
15	Carroll, Iowa	Pettis, Missouri	Cumberland, New Jersey
16	Westmoreland, Pennsylvania	Day, South Dakota	Daviess, Missouri
17	Adair, Kentucky	Jennings, Indiana	Glenn, California
18	Dawson, Nebraska	Osceola, Iowa	Berkeley, West Virginia
19	Clay, South Dakota	Carroll, Maryland	Seward, Kansas
20	Wells, Indiana	Bamberg, South Carolina	Kiowa, Oklahoma
21	Jefferson, West Virginia	Franklin, Tennessee	Warren, Ohio
22	Harvey, Kansas	Pawnee, Kansas	Lewis, Idaho
23	Lyon, Iowa	Harnett, North Carolina	Leavenworth, Kansas
24	Matagorda, Texas	Orange, North Carolina	Lamar, Texas
25	Jefferson, Iowa	Drew, Arkansas	Manitowoc, Wisconsin
26	Waseca, Minnesota	Seward, Nebraska	Kaufman, Texas
27	Montgomery, Indiana	Mills, Iowa	Barber, Kansas
28	Burnett, Wisconsin	Bladen, North Carolina	Tuscarawas, Ohio
29	Marion, Tennessee	McNairy, Tennessee	Beaver, Oklahoma
30	Taylor, Iowa	Marion, Tennessee	Tate, Mississippi

years 1984 through 2007. This generated a 24-year time series of classifications. While the focus here was on the major crops, it should be noted the classifications contained a host of cover types (e.g. grassland, urban, forest etc.) based on what was found in the reference CDLs. All classification analysis was performed during April 2018 and so any imagery and tools were dependent on what existed in GEE at that time.

Then finally, for each of the classifications the area of the crop type of interest was calculated within GEE and compared to USDA statistics for the corresponding years. Note, if a county had an area of soybeans/ winter wheat double-crop then that area was also including in the soybean or winter wheat total. The 24 year area comparisons by county were ultimately interpreted together give a sense of the correlation and a proxy of validation. Also, area totals for the selected counties were also calculated from the ten years of CDLs 2008 through 2017 and compared to the corresponding USDA statistical estimates. This was done to provide a benchmark as to the expected relationship strength since even with near-perfect classifications they may not perfectly match the officially reported which has uncertainties of its own.

4. Results

Visual inspection was relied upon to first qualitatively assess the results since no field-level crop type validation exists back in time. Fig. 4 shows the 24 years of corn classifications for Cedar County, Iowa – the first that was randomly sampled for that crop. Subjectively, the visual results show similar patterning for each of the years. Some years like 1987, 1988, 1991, and 2005 show a lot of corn being predicted while years like 2000 and 2001 show much less. Note, 1985 did not have enough imagery to complete the classification and thus it was unable to be derived. The results for the first soybean county randomly selected, Cecil County in the State of Maryland, and that for winter wheat, Montgomery County in Kansas are shown in Figs. 5 and 6, respectively. Those examples contained solid imagery for all years and thus none are blank. The spatial patterning is again similar across images but some years do show notable more area classified than others. For the other 29 county scenarios per crop and not shown, interpretation of map output ranged from quite reasonable for some years and counties to being grossly incoherent for others. Problems included either obvious over or under classification for one or more cover types or a noisy, speckled look. Closer inspection of the errors showed that they were worst when the summer composite were muddled due to lack of cloud free imagery. The most common year for which there was no classification possible was 1985 but occasionally other years would be blank as well. Additionally of note, most classifications appeared to show smaller field sizes in the earlier years.

The qualitative analysis is insightful but provided little reflection on the mapping accuracy. Ideally, field or pixel-level validation information would exist back through time but the availability of this is unlikely. However, what does exist are county-level crop acreage statistics from USDA and so those were used as a comparator. Fig. 7 shows the classified crop area as generated in GEE to that published from the USDA for the initially sampled corn, soybean, and winter wheat counties (again, Cedar, Cecil, and Montgomery). For context the CDL and USDA estimates are also shown for years 2008 through 2017. Note, there are a couple of years with missing data. Fig. 8, in turn, shows the scatterplots and linear fits for those area relationships. The corn sampled county shows a mild positive relationship but the soybean and winter wheat examples effectively show none. Expanding to the full 30 sample counties for each crop is summarized in Table 2 with the R^2 correlations.

The R^2 averages for the 30 counties were 0.192 for corn, 0.159 for soybeans, and 0.142 for winter wheat. Those values are weak. They provide relative suggestion that corn is performing better than soybeans, which in turn is performing better than winter wheat. The R^2 range was quite variable across the samples with standard deviations of 0.132, 0.177, and 0.133, for corn, soybeans, and winter wheat, respectively. For corn the range was 0.001 to 0.529, soybeans 0.000 to 0.830 and winter wheat 0.001 to 0.524. The best performing counties were Lyon, Iowa for corn, Day, South Dakota for soybeans, and Glenn,

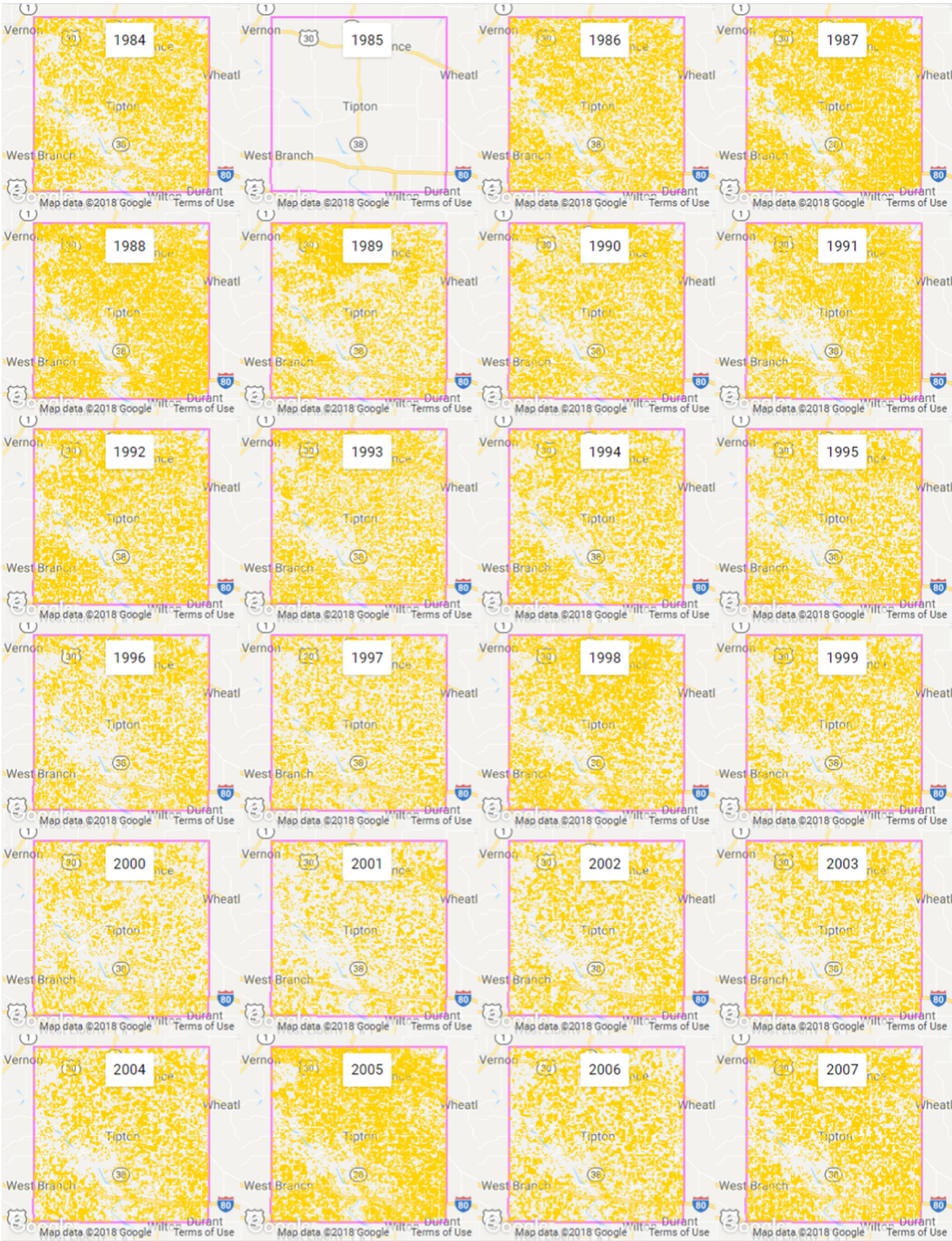

Fig. 4. Cedar County, Iowa - 24 years of predicted corn areas (yellow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

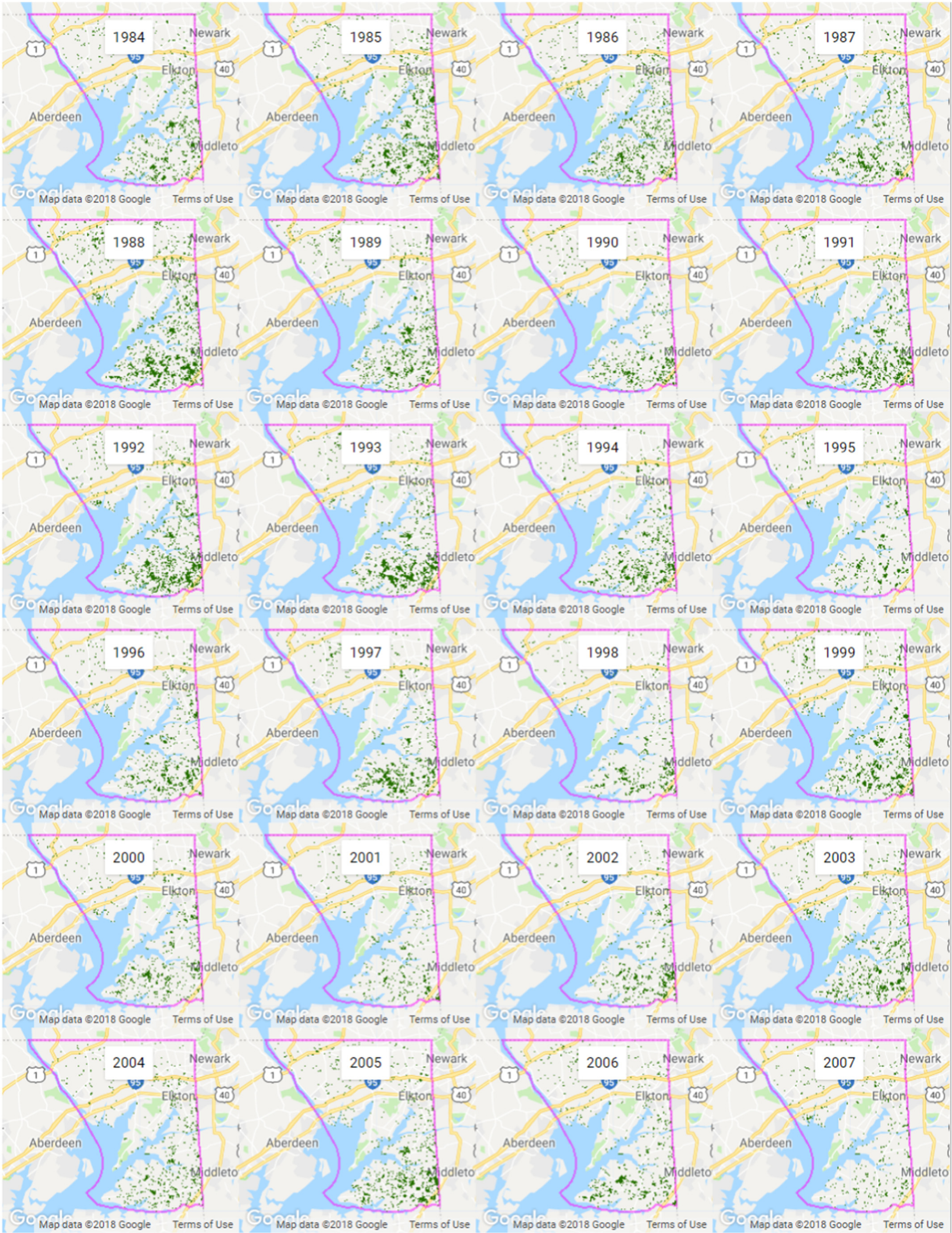

Fig. 5. Cecil County, Maryland – 24 years of predicted soybean areas (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

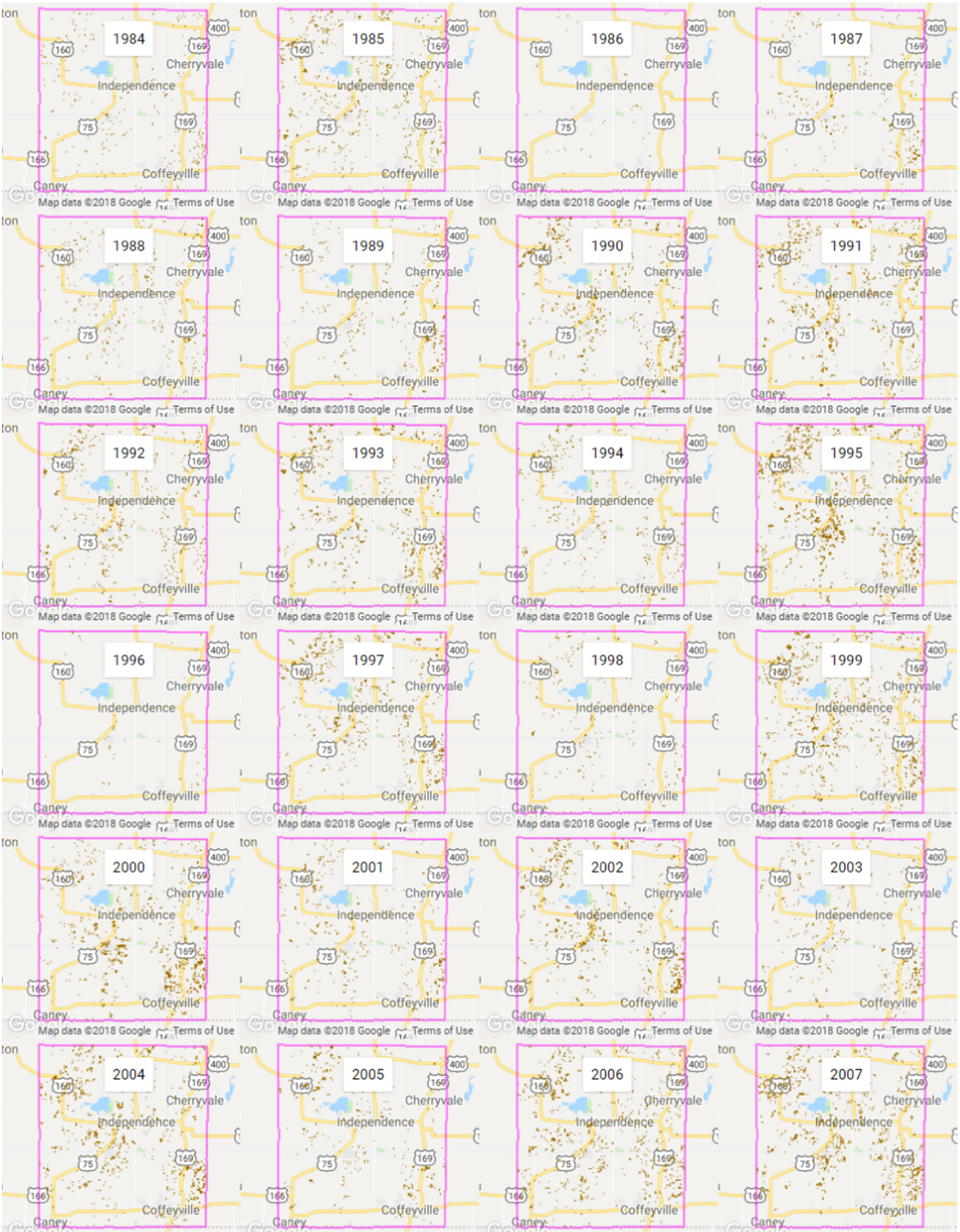

Fig. 6. Montgomery County, Kansas – 24 years of predicted winter wheat areas (brown). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

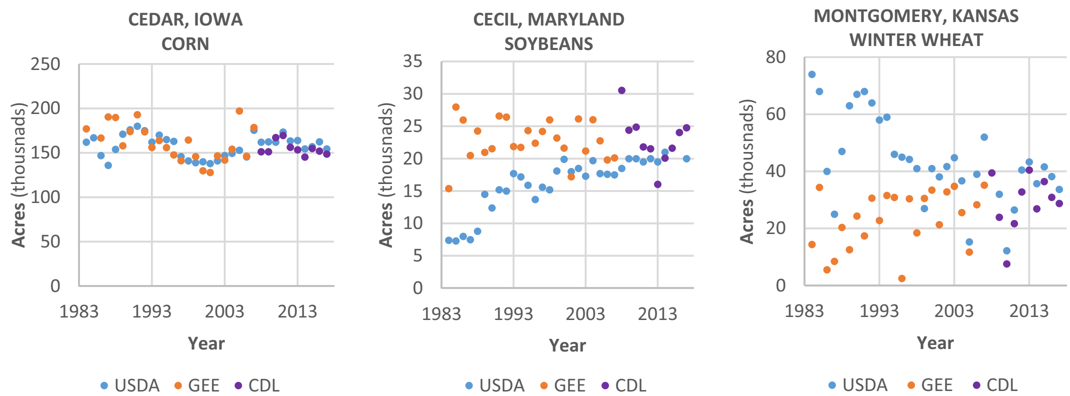

Fig. 7. Comparison of 1984 through 2007 crop area (acres) from USDA to that generated with GEE for the first counties sampled from corn, soybeans, and winter wheat. Statistics for 2008 through 2017 CDL area information versus USDA also provided for context.

California for winter wheat. Conversely, the poorest relationship county samples were Marion, Tennessee for corn, Douglas, Kansas for soybeans, and Warren, Ohio for winter wheat. The distribution of the results were positively skewed and more so for soybeans and winter wheat which had median R^2 values of 0.124 and 0.109, respectively. The mean R^2 for corn was 0.214.

The linear model slope and intercept for each sample can also be useful for understanding utility of the output. In short, the pixel area summation showed a nearly universal positive correlation to USDA published statistics over the 24 year period. Twenty-seven of the 30 were positively sloped for corn, 23 of the 30 for soybeans, and 28 of the 30 for winter wheat. On average the linear regression slope for corn, soybeans, and winter wheat was respectively 0.575, 0.206, and 0.310. Ideally, these would be close to 1.000 so the lacking slope is a reflection of the GEE analysis over-classifying in the large area counties and/or under-classifying in the small area ones. The average respective intercepts were 15,973, 40,745, 24,514 acres reinforcing this finding. Also more generally, on average it was found that corn and winter wheat tended to under estimate the predicted crop area while soybeans overestimated.

A follow-up analysis to look at counties with markedly increased or decreased acreage over the 24 year period was also undertaken to better understand if a long-term change detection methodology, which is commonly sought after, is worthwhile. It was also hypothesized that there could be better performance of the acreage estimation in areas which contained a large spread of values. Thus, for each of the three commodities the USDA county-level area statistics were ranked to select the five greatest counties of increase and the five of greatest decrease. Stratification was employed to ensure that no two samples came from the same state to increase the variety. In total 30 more county-crop comparisons were created. Their locations are shown in Fig. 9. Table 3 lists those 30 additional counties analyzed and the R^2 relationship

Table 2
Correlation analyses by crop-county for 1984 through 2007 GEE derived areas versus USDA area statistics.

Sample county	Corn R^2	Soybeans R^2	Winter wheat R^2
1	0.204	0.001	0.024
2	0.060	0.101	0.307
3	0.247	0.003	0.007
4	0.236	0.147	0.129
5	0.306	0.242	0.106
6	0.154	0.266	0.294
7	0.141	0.000	0.167
8	0.090	0.309	0.002
9	0.073	0.095	0.107
10	0.120	0.001	0.190
11	0.260	0.065	0.272
12	0.382	0.198	0.100
13	0.023	0.036	0.057
14	0.033	0.250	0.078
15	0.268	0.013	0.173
16	0.002	0.830	0.342
17	0.288	0.226	0.524
18	0.194	0.161	0.004
19	0.379	0.447	0.305
20	0.153	0.265	0.123
21	0.002	0.061	0.001
22	0.298	0.157	0.134
23	0.529	0.292	0.054
24	0.019	0.000	0.328
25	0.299	0.075	0.266
26	0.233	0.159	0.006
27	0.224	0.000	0.001
28	0.276	0.039	0.046
29	0.001	0.005	0.111
30	0.279	0.339	0.011
Average	0.192	0.159	0.142

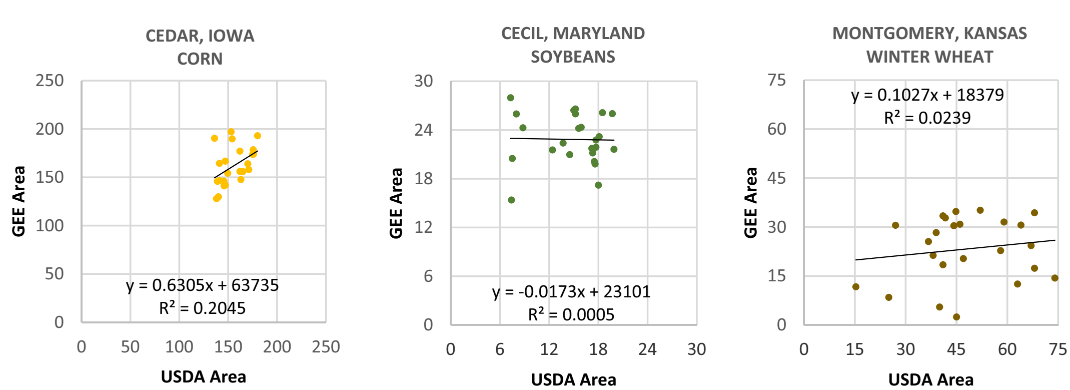

Fig. 8. USDA versus GEE areas (in thousands of acres) for initial randomly selected county for corn, soybeans, and winter wheat.

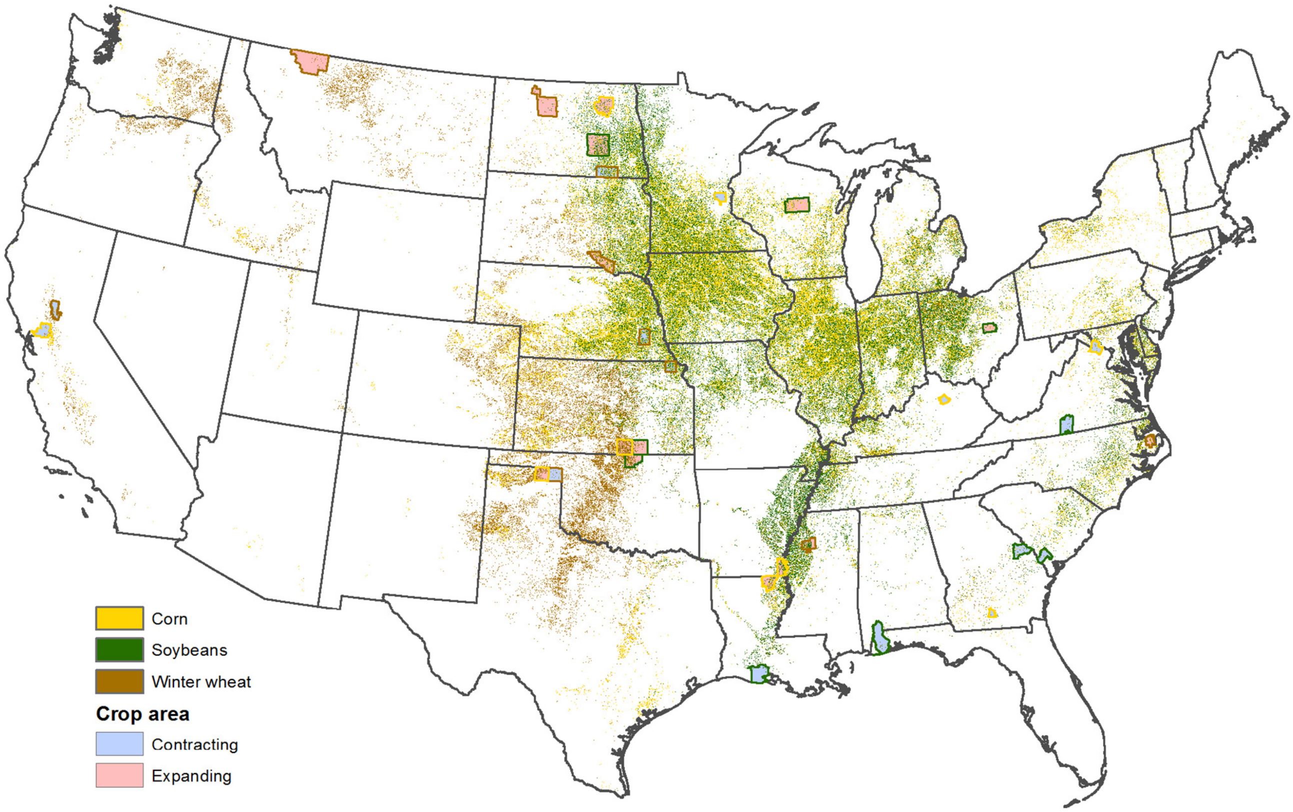

Fig. 9. Counties sampled with rapidly expanding (pink interior) or contracting (blue interior) area for corn, soybean, or winter wheat. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

metrics that resulted.

The qualitative map assessment of those results appeared to be somewhat better to the analysis of the counties selected randomly. Quantitatively, the R^2 estimates for the counties chosen with rapidly contracting or expanding crop areas showed better performance on average being 0.355, 0.298, and 0.270, for corn, soybeans, and winter wheat, respectively. However, there were fewer samples to draw a strong conclusion and there is still a lot of variably amongst the 30 results which averaged 0.308 in aggregate. Granted, this was nearly twice as good as for the purely random 90 samples which in total averaged 0.164. This improved average is particularly driven higher from the examples of counties with expanding corn area of the 24 years because Ochiltree, Texas and Morehouse, Louisiana had very good correlation results (0.754 and 0.774). Conversely, Ramsey, North Dakota and Sumner, Kansas were still poor (0.022 and 0.001). The other crops and counties had a range of R^2 values that fell in between these extremes.

5. Discussion

An R^2 of 1.000 would indicate a perfect relationship between the area of the mapping effort and what the USDA estimated through its survey programs. However, the vast majority of the counties tested failed to come close, and this in spite of many of the retrospective classifications looking quite reasonable upon inspection. Confounding the accuracy assessment is there being uncertainties in the USDA statistics in addition to those inherent to the crop classifications. In other words, an R^2 of 1.000 is not a fair benchmark because a perfect relationship is not ever likely.

Measures of crop area uncertainty at the county-level are not provided alongside the annual USDA statistics even though there has been a push to do so recently (National Academies of Science, 2017).

Table 3
Historic crop mapping area performance versus USDA statistics for counties with rapidly contracting (top) and expanding (bottom) crop area.

County	Corn	R^2	Soybeans	R^2	Winter wheat	R^2
1	Solana, California	0.617	Baldwin, Alabama	0.607	Sutter, California	0.229
2	Cook, Georgia	0.348	Burke, Georgia	0.401	Brown, Kansas	0.437
3	Bourbon, Kentucky	0.428	Vermillion, Louisiana	0.617	Lancaster, Nebraska	0.674
4	Loudoun, Virginia	0.238	Pittsylvania, Virginia	0.022	Lipscomb, Texas	0.228
5	Anoka, Minnesota	0.183	Hampton, South Carolina	0.072	Dickey, North Dakota	0.196

County	Corn	R^2	Soybeans	R^2	Winter wheat	R^2
1	Ochiltree, Texas	0.754	Cowley, Kansas	0.116	Tallahatchie, Mississippi	0.430
2	Ramsey, North Dakota	0.022	Stutsman, North Dakota	0.401	Glacier, Montana	0.071
3	Morehouse, Louisiana	0.774	Kay, Oklahoma	0.017	Tyrrell, North Carolina	0.277
4	Sumner, Kansas	0.001	Marathon, Wisconsin	0.382	Charles Mix, South Dakota	0.154
5	Chicot, Arkansas	0.183	Coshocton, Ohio	0.348	Ward, North Dakota	0.001

However, a sense of the accuracy of the data can be gleaned from USDA macro-level survey information, the 2017 Census of Agriculture (CoA), and internal research efforts. The national-level sampling error generates coefficients of variation (CV), or relative standard deviation, of roughly 1.0 to 2.0% for corn, soybeans, and winter wheat. This is summarized in both the annual Crop Summary and Small Grains results in addition to the CoA, all available via Quick Stats (USDA, 2017). Smaller units of area results in larger uncertainties and as such, state-level corn average CVs average 11.2, 12.1, 11.0% for corn, soybean, and winter wheat, respectively. States with significant area devoted to the commodities usually have error rates closer to the national-level than the state-level.

At the county-level the 2017 CoA reports average CVs of 22.6, 21.3, and 22.1%. Again, respectively for corn, soybeans, and winter wheat. In terms of the annual county-level information (i.e. not the CoA), the values computed directly from surveys show the average errors for corn and soybeans to also be just above 20.0% for the heavily farming intensive states of Iowa, Illinois, and Indiana (Erciulescu et al., 2018). For all states in the nation, the CV estimates from the county-level surveys are closer to 30.0% on average while winter wheat errors are upwards of 40.0%. However, other sources of USDA information help benchmark and reconcile the raw county-level survey data so errors of the published values are likely closer to that of the CoA's CVs of slightly over 20.0%. As pertaining to the sampled 30 sampled corn, soybean, and winter wheat counties, the CoA CVs were 20.2, 22.0, and 21.6%. This closely mimics the errors found for all the counties average across the nation. Finally, the initially examined three counties of Cedar, Iowa for corn, Cecil, Maryland for soybeans, and Montgomery, Kansas for winter wheat had CVs of 15.7, 8.2, and 20.2%, respectively.

In summary, the CVs for the planted area statistics for which the hindcasted CDLs are being compared to have estimated inherent average errors just above 20.0%. In other words, a typical county-level area statistic for corn, soybeans, or winter wheat is likely only within 20.0% of the actual area about two-thirds of the time. The data collection and dissemination methods have been fairly stable so those suggested error rates are likely consistent through the era of this analysis back to 1983. With this validation data weakness in mind, and to test and quantify what an expected good R^2 relationship should be, the existent CDL classifications 2008 through 2017 were compared in their area relationship to those of the corresponding USDA statistics. This was done for the same 30 samples for each crop with the results are shown in Table 4.

The average CDL to USDA area statistics for corn had an R^2 relationship 0.477, far short of a perfect 1.000. Soybeans and winter wheat were much stronger at 0.686 and 0.726, however. The range of individual values was very large spanning 0.001 to 0.918 for corn, 0.003 to 0.967 for soybeans, and 0.088 to 0.994 for winter wheat. In short, some of the counties had very strong CDL to USDA area correlations but many showed little relation. Whether this is a function of errors in the CDL classifications, that of the USDA crop statistics, or both, is unknown.

Differencing the benchmarked R^2 versus that historically observed through the GEE maps, and also listed in Table 4, suggests that the classification efforts are better than they might first appear. For example, while the historical mapping average R^2 was 0.192 for corn that is only 0.285 less than the 0.447 benchmark. This is much better than if using 1.000 as the goal. The soybean and winter wheat differences are significantly greater though at 0.527 and 0.584, respectively. These relative differences suggest that the mapping efforts for corn are actually the best of the three even though the pure correlation average is the lowest.

Many of the classifications were obvious outliers. Reasons could be many but were most likely a result the poor image quantity or quality for a particular year. If one were allowed to remove the most egregious outlier years the R^2 relationships do improve and in many cases to the point of matching or beating the benchmark. An iterative final analysis was done to remove the largest outliers for each county-crop sample until the remaining classifications resulted in a R^2 of benchmarks for corn, soybeans, and wheat of 0.477, 0.686, and 0.726, respectively. On average it took the removal of the seven worst classifications to bring the R^2 from 0.192 to meet the 0.477 benchmark for corn. In other words, this suggests about two-thirds of the historically generated classifications for corn were actually quite robust. Soybeans and winter wheat had higher thresholds to obtain and it took on average the removal of twelve of the classifications, or half, to obtain that benchmark for both cases. So, this suggests on average every other year one can expect a useful classification for these two crops.

Table 4
Correlation analyses by crop-county CDL areas versus USDA area statistics. This is for years 2008 through 2017 for which CDLs already exist and can be directly compared. The second column for each commodity is the subtraction of the 2008 through 2017 CDL versus USDA R^2 (in Table 2) and this 1984 through 2007 GEE to USDA R^2.

Sample county	CORN R^2	vs. '84–'07	Soybeans R^2	vs. '84–'07	Winter wheat R^2	vs. '84–'07
1	0.544	0.339	0.236	0.747	0.957	0.582
2	0.009	−0.051	0.842	0.741	0.889	0.933
3	0.395	0.149	0.941	0.939	0.874	0.868
4	0.904	0.668	0.894	0.235	0.488	0.359
5	0.517	0.211	0.922	0.680	0.932	0.825
6	0.582	0.428	0.811	0.545	0.994	0.700
7	0.521	0.381	0.003	0.003	0.972	0.805
8	0.471	0.381	0.935	0.626	0.088	0.086
9	0.441	0.369	0.213	0.118	0.859	0.752
10	0.810	0.690	0.855	0.855	0.939	0.749
11	0.184	−0.076	0.699	0.634	0.814	0.542
12	0.736	0.354	0.951	0.753	0.711	0.612
13	0.599	0.576	0.740	0.704	0.919	0.862
14	0.104	0.071	0.755	0.505	0.412	0.334
15	0.833	0.565	0.905	0.893	0.737	0.564
16	0.205	0.204	0.953	0.123	0.745	0.403
17	0.428	0.141	0.959	0.733	0.736	0.211
18	0.677	0.484	0.550	0.389	0.460	0.456
19	0.441	0.062	0.003	−0.444	0.771	0.465
20	0.829	0.675	0.967	0.702	0.839	0.716
21	0.001	−0.001	0.406	0.345	0.895	0.894
22	0.918	0.620	0.923	0.766	0.608	0.474
23	0.647	0.118	0.685	0.394	0.725	0.671
24	0.787	0.768	0.646	0.646	0.949	0.621
25	0.041	−0.258	0.867	0.793	0.848	0.582
26	0.378	0.146	0.621	0.462	0.659	0.653
27	0.774	0.550	0.418	0.418	0.127	0.126
28	0.003	−0.273	0.660	0.621	0.465	0.419
29	0.280	0.279	0.845	0.840	0.791	0.681
30	0.259	−0.020	0.382	0.043	0.585	0.574
Average	0.477	0.285	0.686	0.527	0.726	0.584

The results were dissected in other manners to look for association trends through time and geography. They are summarized here. Averaging all of the sample correlations by year showed no suggestion of some time periods performing better than others. Albeit, 1985 could be deemed worst because was unable to consistently produce classifications due to lack of imagery that year. It was hypothesis that the years for which Landsat 7 was available (1999 forward) would show better classification results, but they appeared to perform the same as the Landsat 5 only classifications. With the SLC-off problem it was further hypothesized that the scenes with the inclusion of Landsat 7 data might actually be hindering. However, there were four years with Landsat 7 functioning correctly (1999–2002) and those years showed no better R^2 results compared to the rest. Finally, it was also thought years, particularly 1995, when Landsat 5 had drifted from its prescribed 9:45 a.m. overpass (Zhang and Roy, 2016) would perform worse but there was no supportive evidence.

Spatially, it was hypothesized that particular areas of the country might provide better results than others. This could be explained as function of cloud cover, field sizes, USDA statistics quality, etc.

However, there was no geographic patterning or clustering of the correlations to suggest some areas worked better than others. This was true for all three crops. Related, it was considered that counties with large crop areas might perform better (as there should be lower USDA statistical uncertainty) than the small ones, but there was no trend found in that respect either.

Finally, it was theorized that composite windowing length, and the available imagery going into it could influence the quality of the results. This was also investigated through a couple of means. First, some of the counties sample fell by chance in areas with full overlap from adjacent Landsat scenes and thus this scenario would provide up to twice as much input imagery available to build the composite. Five county-crop samples by chance had complete adjacent scene overlap: Middlesex, Virginia for corn, Cass, Indiana and Seward, Nebraska for soybeans, and Franklin, Nebraska and Glenn, California for winter wheat. Those counties did not show superior correlation performance however.

Adjustments to the compositing window length could also be seen as a way to improve the classifications. Parallel testing was undertaken using a 32-day window, thus yielding eight total imagery composites per year, with mixed results. Some of the GEE classifications to USDA statistics did improve but at the cost of a much greater likelihood of having no classification results at all due to lack of complete imagery composites. This was true throughout all of the study samples and even including those in high overlap path areas. On the flipside, 128-day windows (thus only two composites per year) were also constructed and classifications tested but those results were definitely poorer on average. There was likely just too little phenological information across the season to differentiate the cover types and/or overstretching of the compositing because of the four month bounding window. These findings helped reinforced that the 64-day compositing window chosen was probably the best compromise but in cases where tighter time ranges can be used there could be some gain.

6. Conclusions

The overall effectiveness of extending the classification methodology back annually though the Landsat archive to derive retrospective mapping of corn, soybeans, and winter wheat areas was found to be variable by year and location. In some cases the output maps looked quite reasonable and the quality reinforced through the favorable comparison with historical USDA crop area statistics. Other cases though the maps produced were incorrect due to either obvious crop area bias or spatial noise. This was not believed random but rather largely a function of the quantity and quality of imagery available. It was found to be particularly problematic if the data were deficient during the peak of the growing season. These issues where imagery is lacking are very difficult to overcome but for some instances tuning the compositing time windows could help. Ultimately though, prudent removal of up to half of the poor output years, those which are likely noise, should leave a remaining time series dataset which is quite robust. In turn, that what is left should still provide as uniquely useful dataset for long-term, and sometimes inter-annual, assessment of crop area extent and change at a 30 m resolution.

The stated intension of this research was to construct historical crop maps, but the forward looking utility is to construct classifications in real-time area estimation for major commodities. This has been a long held promise since the beginning go of the Landsat program but has remained elusive beyond research settings with strong computing infrastructure. Now though the online availability of spatially and temporal rich Landsat datasets, exploitable elegantly with high performance computing platforms like GEE, has brought forth a new era which has been demonstrated here. And while the outcomes shown were not consistently strong, thus providing skepticism toward the accuracy of operational area estimation products, it is believed that the even better calibrated and temporally richer contemporary datasets that were not available in the past will improve forward looking map products. In particular the Copernicus Sentinel-2 tandem satellites in conjunction with Landsat 8, and forth coming Landsat 9, will provide many more times the data that was available from a sole Landsat 5 and/or mostly hobbled SLC-off Landsat 7. Also, a longer period (i.e. more than four years) and one that is adjacent in time (i.e. the years just before) in which to derive training samples could result in a more representative decision tree dataset to improving classifications. And finally, the aim here was that of identify crops which by their nature are highly dynamic and probably the most challenging land cover types to document. Thus, it is assumed if the methods employed here would see much more favorable results over more generalized cover types.

Acknowledgements

This research was supported by the intramural research program of the U.S. Department of Agriculture, National Agricultural Statistics Service, Research and Development Division. The findings and conclusions in this publication have not been formally disseminated by the U. S. Department of Agriculture and should not be construed to represent any agency determination or policy. Thanks to Mark Friedl for recommendations regarding classification validation and Noel Gorelick for GEE coding advice. Special thanks for internal review and editing by Rick Mueller and the comments and guidance as provided by the anonymous peer reviewers and editors of Remote Sensing of Environment.

References

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 Int. 26, 341–358.
Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Chapman and Hall/CRC.
Bruinsma, J., 2003. World Agriculture: Towards 2015/2030: An FAO Perspective. Earthscan, Rome.
Claverie, M., Vermote, E., Franch, B., Masek, J., 2015. Evaluation of the Landsat-5 TM and Landsat-7 ETM+ surface reflectance products. Remote Sens. Environ. 169, 390–403.
Erciulescu, A., Cruze, N., Nandram, B., 2018. Model-based county level crop estimates incorporating auxiliary sources of information. J. R. Stat. Soc. A. Stat. Soc. 182.
Ferguson, M., Badhwar, G., Chhikara, R., Pitts, D., 1986. Field size distributions for selected agricultural crops in the United States and Canada. Remote Sens. Environ. 19, 25–45.
Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrechy, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M.C., Gong, Peng., Abdel Aziz, S., Cipriani, A., Cumani, R., Cecchi, G., COnchedda, G., Ferreira, S., Gomex, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D.R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhoc, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., Van Der Velde, M., Dunwoody, A., Kraxner, F., Obersteiner, M., 2015. Mapping global cropland and field size. Glob. Chang. Biol. 21 (5), 1980–1992.
Gallego, J., Carfagna, E., Baruth, B., 2010. Accuracy, objectivity and efficiency of remote sensing for agricultural statistics. In: Agricultural Survey Methods. John Wiley & Sons, Ltd, pp. 193–211.
Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., Tyler, D., 2002. The National Elevational Dataset. Photogramm. Eng. Remote. Sens. 68 (1), 5–11.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.
Goward, S., Williams, D., Arvidson, T., Rocchio, L., Irons, J., Russel, C., Johnston, S., 2017. Landsat's Enduring Legacy; Pioneering Global Land Observations From Space. American Society for Photogrammetry and Remote Sensing, pp. 586.
Hansen, M., Potapov, P., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S., Goetz, S., Loveland, T., Kommareddy, A., Egorov, A., Chini, L., Justice, C., Townshend, J., 2013. High-resolution global maps of 21-st-century forest cover change. Science 342, 850–853.
Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-representing a decade of land cover change information. Photogramm. Eng. Remote. Sens. 81, 345–354.
Johnson, D., 2016. A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products. Int. J. Appl. Earth Obs. Geoinf. 52, 65–81.
Johnson, D., Mueller, R., 2010. The 2009 Cropland Data Layer. Photogramm. Eng. Remote. Sens. 76, 1202–1205.
Jun, C., Ban, Y., Li, S., 2014. Open access to Earth land-cover map. Nature 514, 343.

King, L., Adusei, B., Stehman, S., Potapov, P.V., Song, X.-P., Krylov, A., Bella, C.D., Loveland, T.R., Johnson, D.M., Hansen, M.C., 2017. A multi-resolution approach to national-scale cultivated area estimation of soybean. Remote Sens. Environ. 195, 13–29.
Kovalskyy, V., Roy, D., 2013. The global availability of Landsat 5 TM and Landsat 7 ETM + land surface observations and implications for global 30 m Landsat data product generation. Remote Sens. Environ. 130, 280–293.
Lark, T., Mueller, R., Johnson, D., Gibbs, H., 2017. Measuring land-use and land-cover change using the U.S. Department of Agriculture's Cropland Data Layer: cautions and recommendations. Int. J. Appl. Earth Obs. Geoinf. 62, 224–235.
Loveland, T., Dwyer, J., 2012. Landsat: building a strong future. Remote Sens. Environ. 122, 22–29.
Maiersperger, T., Scaramuzza, P., Leigh, L., Shrestha, S., Gallo, K., Jenkerson, C., Dwyer, J., 2013. Characterizing LEDAPS surface reflectance products by comparisons with AERONET, field spectrometer, and MODIS data. Remote Sens. Environ. 136, 1–13.
Markham, B., Helder, D., 2012. Forty-year calibrated record of earth-reflected radiance from Landsat: a review. Remote Sens. Environ. 122, 30–40.
Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.-K., 2006. A Landsat surface reflectance dataset for North America, 1990–2000. IEEE Geosci. Remote Sens. Lett. 3, 68–72.
Massey, R., Sankey, T.T., Yadav, K., Congalton, R.G., Tilton, J.C., Thenkabail, P.S., 2017. NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) @ 30m for North America: Cropland Extent Product (GFSAD30NACE). NASA EOSDIS Land Processes DAAC.
National Academies of Sciences, Engineering, and Medicine, 2017. Improving Crop Estimates by Integrating Multiple Data Sources. The National Academies Press, Washington, DC.
Pekel, J.-F., Cottam, A., Gorelick, N., Beward, A., 2016. High resolution mapping of global surface water and its longterm changes. Nature 540, 418–422.
Ramankutty, N., Foley, J.A., 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992. Glob. Biogeochem. Cycles 13, 997–1027.
Sexton, J., Urban, D., Donohue, M., Song, C., 2013. Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record. Remote Sens. Environ. 128, 246–258.
Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., Skakun, S., 2017. Exploring Google Earth Engine platform for big data processing: classification of multi-temporal satellite imagery for crop mapping. Front. Earth Sci. 5, 17.
Song, X.-P., Potapov, P.V., Krylov, A., King, L., Bella, C.D., Hudson, A., Khan, A., Adusei, B., Stehman, S.V., Hansen, M.C., 2017. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. Remote Sens. Environ. 190, 383–395.
Thenkabail, P.S., Knox, J.W., Ozdogan, M., Gumma, M.K., Congalton, R.G., Wu, Z., Milesi, C., Finkral, A., Marshall, M., Mariotto, I., You, S., Giri, C., Nagler, P., 2012. Assessing future risks to agricultural productivity, water resources and food security: how can remote sensing help? Photogramm. Eng. Remote. Sens. 78, 773–782.
USDA, National Agricultural Statistics Service, 2010. Field crop usual planting and harvesting dates. In: Agricultural Handbook Number 628, Retrieved October 30, 2018, from. http://usda.mannlib.cornell.edu/usda/current/planting/planting-10-29-2010.pdf.
USDA, National Agricultural Statistics Service, 2017. Quick stats. Retrieved May 30, 2019. http://www.nass.usda.gov/Quick_Stats/.
Vogelmann, J., Gallant, A., Shi, H., Zhu, Z., 2016. Perspectives on monitoring gradual change across the continuity of Landsat sensors using time-series data. Remote Sens. Environ. 185, 258–270.
Wulder, M., Masek, J., Cohen, W., Loveland, T., Woodcock, C., 2012. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. Remote Sens. Environ. 122, 2–10.
Wulder, M., White, J., Loveland, T., Woodcock, C., Belward, A., Cohen, W., Fosnight, G., Shaw, J., Masek, J., Roy, D., 2016. The global Landsat archive: status, consolidation, and direction. Remote Sens. Environ. 185, 271–283.
Yan, L., Roy, D.P., 2016. Conterminous United States crop field size quantification from multi-temporal Landsat data. Remote Sens. Environ. 172, 67–86.
Zhang, H., Roy, D., 2016. Landsat 5 Thematic Mapper reflectance and NDVI 27-year time series inconsistencies due to satellite orbit change. Remote Sens. Environ. 186, 217–233.
Zhu, Z., Woodcock, C., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171.