Satellite detection of cover crops and their effects on crop yield in the Midwestern United States

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence
https://creativecommons.org/licences/by/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the article online for updates and enhancements.
Satellite Detection of Cover Crops and Their Effects on Crop Yield in the Midwestern United States

Christopher A Seifert*, George Azzari2, David B Lobell2

1 Department of Earth System Science, Stanford University, Stanford, CA 94305, USA
2 Department of Earth System Science and Center on Food Security and the Environment, Stanford University, Stanford, CA 94305, USA

Abstract

The practice of planting winter cover crops has seen renewed interest as a solution to environmental issues with the modern maize- and soybean-dominated row crop production system of the U.S. Midwest. We examine whether cover cropping patterns can be assessed at scale using publically available satellite data, creating a classifier with 92.2% accuracy (κ = 0.72). We then use this classifier to examine spatial and temporal trends in cover crop occurrence on maize and soybean fields in the Midwest since 2008, finding that despite increased talk about and funding for cover crops as well as a 77% increase in cover crop acres planted from 2008-2016, increases in winter vegetation have been more modest. Finally, we combine cover cropping with satellite-predicted yields, finding that cover crops are associated with decreased maize and soybean production and poor soil quality. Despite this association, when controlling for invariant soil quality using a panel regression model, we find modest benefits of cover cropping, with average yield increases of 0.88% for maize and 0.12% for soy. Give these slight impacts on yields, greater incentives or reduced costs of implementation are needed to increase adoption of this practice for the majority of maize and soybean acres in the U.S.

Introduction

Cover crops are nothing new to the American farm, with experiments growing “other plants” between one year’s maize harvest and the next year’s maize planting dating back to at least the 1890s [1]. More modern interest in the practice, however, stems from farmers who wish to alleviate soil compaction, reduce soil erosion and reduce soil nitrogen loss [2], and has been accompanied by a large increase in funding available for farmers to adopt cover crops [3]. In reaction to this increased demand, new cover crop varieties and blends have been brought to market [4] and new equipment has been devised to lower costs associated with cover crop adoption [5].
Despite these innovations, cover crops remain the domain of a minority of farmers in the Midwest who try them only on a portion of their land. As of 2006 fewer than 20% of farmers had used cover crops and those farmers had only tried them on only 6% of their farm[6]. In 2012, the U.S. Census of Agriculture found 1.44 million ha of cover crops planted across nine Upper Midwest states responsible for a collective 51.2 million ha of maize and soybean planted [7]. Major agronomic and operational challenges with the practice include getting the cover crop established, dealing with the increased time and labor requirements of running an operation that includes cover cropping, and figuring out what species might be the most beneficial in a given area [2]. There may also be allelopathic effects of some cover crop species such as cereal rye on cash crops such as maize[8-9].

Meanwhile, economic challenges include cost, a cover crop budget might run as high as 136 dollars per hectare[10] as well as structural factors (e.g. lack of use for biomass created by a cover crop due to absence of livestock on the farm[11]). Thus, even if yields are comparable, economic returns for cover crop adopters can be harmed[12]. These challenges, combined with recent negative returns for maize and soybean farming operations [13] have highlighted the critical importance of quantifying any potential economic returns to cover crop adoption.

The typical method for assessing these returns has been through field trials. A metastudy of such trials by Tonitto et al. (2006) [14] found a roughly equal distribution of slightly positive and slightly negative yields for non-leguminous cover crops while a separate metastudy by Miguez and Bolero (2005) found a distribution of yields with more negative skew[15]. Similarly, more recent trials by
the Practical Farmers of Iowa found decreases of as much as 1.4 Mg/ha and
increases of as much as 0.5 Mg/ha bushel for cover crops on maize and -0.2 to 0.5
Mg/ha yield differences for cover crops on soybean [13]. Adding insight for
potential mechanisms, Reese et al. (2014) [17] found that cover crops reduced
yields when moisture was scarce in the production cycle following, but had no effect
otherwise. While these sorts of studies have a strength that they allow for strong
claims of causality, they are subject to limited applicability based on site-specific
weather, soil qualities, and management practices where they took place.

As an alternative, the combination of satellite-based classifiers and satellite-
based yield assessments offers the potential to assess the yield effects of cover crops
at scale. Such assessments have been applied to limited areas and timescales, with
Hively et al. (2015)[14] running one such assessment for Southeastern Pennsylvania
for 2010-2013 and the Environmental Working Group producing one for Iowa,
Illinois and Indiana for 2015[3]. These assessments have the potential for wider
applicability and insight into patterns at larger spatial and temporal scales but
cannot definitively prove causality. However, even if claims at causality are limited,
scalable mapping of cover crops alone can show whether funding to programs like
EQIP is having an effect on adoption and show the way for scalable verification of
cover crop establishment.

In this study, we set out so use such a classifier to examine not only trends
and patterns in cover crop adoption but also yield effects of cover crop adoption
over time with the hope that the information provided can be used to more
thoughtfully construct government, nonprofit, and private sector programs in relation to the practice.

Methods

Data Processing

Raw data from 24 different farm operators and/or landowners or agencies across eight states were acquired for the purposes of the study. While cash crop production years from 2007-2017 were represented, 83% of the total number of field-years came from the last three years of the interval. Georeferenced, atmospherically corrected, cloud masked Landsat 5, 7 and 8 imagery available through the Google Earth Engine (GEE) platform [19] served as the primary source of remotely sensed data for the analysis in addition to MODIS data which were used to get average values of vegetative indices for the dates of the Landsat imagery. In addition, weather data were compiled for each image using the Daymet gridded daily weather dataset [20], the vapor pressure data of which were transformed to vapor pressure deficit (VPD) using [21] and [22] and aggregate weather data were acquired from PRISM [23].

For each year, imagery from the postharvest-planting interval was composited for the study region - Ohio, Michigan, Wisconsin, Indiana, Illinois, Iowa, Minnesota and the portions of South Dakota and Nebraska where cover crop termination at planting is an accepted practice [24]. The interval was defined as defined as November 15 - the 90th percentile date for the latest 0C freeze in the 1981-2010 climatological history for Illinois, the state with the latest average last
freeze in the study area [25] - to April 3rd, a week before the usual beginning of
maize planting in the study region [26].

This resulted in a set of ten images with 23 bands. The subset of bands used
for classification is outlined in Table 1. Here it should be noted that despite the
slight difference in Landsat 5/7 and 8 sensor platforms, Li et al. (2013) [27] found
that they can be used as complementary data for similar purposes and Flood (2014)
[28] found small mean absolute differences (.014) between Normalized Difference
Vegetation Index (NDVI) measurements across the sensors. Growing degree days
(GDD) for the final band were based on a 0°C base temperature given its
appropriateness for many cover crop varieties (e.g. wheat, rye, oats) [29-30].

<table>
<thead>
<tr>
<th>Band</th>
<th>Math Applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>Median</td>
</tr>
<tr>
<td>NIR</td>
<td>Median</td>
</tr>
<tr>
<td>SWIR1</td>
<td>Median</td>
</tr>
<tr>
<td>NDVI</td>
<td>Median, Minimum, Maximum</td>
</tr>
<tr>
<td>Number of Images</td>
<td>Count</td>
</tr>
<tr>
<td>Landsat-MODIS NDVI difference</td>
<td>Median-Mean, Minimum-Mean, Maximum-Mean</td>
</tr>
<tr>
<td>GDD</td>
<td>NDVI image dates</td>
</tr>
</tbody>
</table>

Table 1: Bands used in cover crop classifier

In order to avoid the potential for mixed pixels across classes, once the final
imagery was compiled, pixels for cover-cropped and non-cover-cropped field areas
were sampled from each image. Sampled pixels came only from areas greater than
30 meters (one Landsat pixel) from a field boundary. Data from states bordering
study area (e.g. Missouri and Pennsylvania) were dropped at this step as well.

71,524 hectare-years of data (794,890 pixels) survived the buffer.
To avoid issues with spatial autocorrelation the pixels were grouped by fields and split into training and test samples. Pixels from 80% of the fields were used as training, pixels from the remaining 20% of fields were used to test the classification. Random forest was chosen as the classification method given examples of high performance on imbalanced training datasets [31-32]. Parameters were set as 128 trees given [33] and 4 variables per tree (rounding the square root of predictors), and a minimum of 5 pixels per node (no decision should be made on less than a ~.45 hectare, or roughly one-acre, basis). This classifier was then used to categorize a masked version of the original images’ pixels as cover cropped or not. The mask here was based on which pixels were classified as maize or soybean pixels in the following year’s cropland data layer (CDL) [34] so as to avoid potential contamination issues with winter wheat and winter rye, which may have a very similar spectral appearance. This eliminated years without a national CDL at the time of the analysis (i.e. 2017 and 2007).

The accuracy of the classifier was gauged according to the number of correct classifications on the test dataset/total number of pixels in the dataset in addition to Cohen’s Kappa (\(\kappa\)) given in equation 1:

\[
(1) \quad k = 1 - \frac{(1 - p_o)/(1 - p_e)}
\]

Where \(p_o\) is the accuracy and \(p_e\) is the probability of chance agreement.

In order to examine the influence of cover crops on yields, we sampled 7500 pixels classified as cover cropped and 7500 pixels classified as non-cover cropped by county (fewer than 7500 cover cropped pixels were used in counties with fewer
than that many pixels classified as cover crop). Linear regression in the R programming language [35] was used to examine correlations between yield and a set of weather soil covariates, in addition to cover cropping. The yield estimates for this analysis were obtained from analyses based on Landsat using the scalable crop yield mapper (SCYM) for maize and soybean [36]. The original implementation of SCYM covered Iowa, Illinois, and Indiana, and thus our soybean yield analysis is limited to those states. For maize, a more recent version [37] was used that all states in the current study area except Nebraska. The accuracy of SCYM has been assessed at both field and county level, as reported elsewhere [37], [38], [39], [40]. Overall, SCYM captures a significant fraction of variability in yields across fields, and any errors should be unrelated to cover crop status as no images from the off-season are used in yield estimation.

Specifically, we estimated the following regression:

\[
Y_i = U_i \alpha + T_i \beta + V_i \delta + W_i \phi + X_\tau + \epsilon
\]

Where \( Y \) is yield in year \( i \), \( U_i \) is cover cropping status as a factor, \( T_i \) is the year, \( V_i \) is a set of weather covariates, \( W_i \) is National Commodity Crop Productivity Index Version 2 (NCCPI) for maize and soybean as derived from the SSURGO data, \( X \) is water holding capacity, derived from the same and \( \epsilon \) is the error term. The (related) model examining patterns by year simply dropped the year term. Yield differences between cover cropped and conventional fields can therefore be thought of as the difference in predicted yields between cover cropped and conventional fields, taking all other covariates into account.
Although equation (2) attempts to control for weather and soil differences between fields with and without cover crops, we recognize that many aspects of soil in particular are hard to capture with existing datasets. Therefore, we also considered a panel regression model to examine how cover crops influenced yields in the same area(s) over time:

\[ y_i = U_i \alpha + T_i \beta + V_i \delta + F + \epsilon \]

Where \( y \), \( U \) and \( V \) are as above and \( F \) is a pixel-level fixed effect.

**Results**

**Classification**

The classifier built for determining cover crop presence/nonpresence had an out of sample accuracy of 92.2%. Given the high prevalence of non-cover-cropped fields, a more appropriate measure of performance, Cohen's Kappa, registered 0.72, slightly above the middle of the interval of 'substantial agreement' given by [41]. This compares favorably to the [3] work with an accuracy of 72% with no Kappa released. Variable importance plots for the identically parameterized classifier built using the R programming language [35] ranked maximum NDVI and the number of GDDs from November 15 to the image date of the maximum NDVI image as the most important variables.

**Trends in Cover Cropping Over Time**
As shown in panel A of figure 1, funding for cover crops has increased drastically over the study area. The largest increases have come in Indiana and Ohio where funding has increased from a nominal amount to 6.1 and 5.2 million dollars respectively. Funding in Iowa also increased to 3.8 million dollars and was above any state other than Indiana for all years except 2015 Ohio passed it as well. While increases were significant in all states other than South Dakota in the dataset, funding amounts remained below $2 million in those states.

Likewise, chatter about cover crops has also increased. As seen in Fig 1 B, farmer social media posts mentioning cover crops increased in their relative frequency from roughly two per thousand to over four per thousand over the study period.
Running the classifier across imagery from the 2008-2016 crop production years resulted in the pattern shown in Fig 1 C. Here, the increase in cover cropping is apparent, with the lowest amount of cover crop area in 2009 at approximately 2.7 million ha, or 6.6% of the total acres planted to maize and soybean. The lowest in percentage terms was in 2008, at approximately 2.9 million ha or 5.7% of maize and soybean planted. The highest number of cover crop acres was in 2016 with 5.2 million planted to cover crops, or roughly 9.3% of the maize and soybean area. The trend in cover crop plantings over this nine year period is significant at p<.1.

Concern with multiple-year adoption of the practice lead to a similar analysis for pixels in which at least two continuous years are cover cropped, showing a similar pattern with an increase from .11 to .35 million hectares in that category, or 0.27 to 0.62% of cropped area.

Along with this increase in cover crop plantings, there has also been an increase in winter NDVI of maize and soybean fields as shown in Fig 1 D. This trend is also significant (p < .05) and bears similarity to the work of Hively et al (2015) who found distinctly increasing vegetative groundcover trends in Southeastern Pennsylvania. In order to see how much of this trend could be due to cover cropping, we compared it with the trend in NDVI on pasture over the same time period, on the assumption that median NDVI of a strongly established cover crop present over the entire winter would have a similar magnitude. Based on that standard, a large gap remains. Indeed, once normalized for the trend in winter pasture NDVI over the study period, the significance in wintertime maize and soybean NDVI disappears.
As shown in Fig 2, summing the satellite-predicted cover crop acres by county shows Southern Indiana, Southern Illinois and Southern and Eastern Ohio as hotspots for cover cropping activity with Central to Western Wisconsin and Central to Western Michigan serving as secondary hotspots in northern areas and Western Iowa having relatively high adoption in some years in absolute but not percentage terms. Broadly, the areas of adoption and non-adoption are similar to the single-year maps produced by USDA ERS (2012) and [3]. However, key differences with respect to those maps are the higher numbers reported here especially in Western Iowa.
Given the concern about consistent adoption in the absence of subsidy as described by [3], we also define cover cropping as having adopted the practice for two years and map in panels C and D above. The analysis presented in these panels should be less vulnerable to idiosyncratic errors for any single year in the classifier. Again, areas of Southern Illinois, Indiana and Ohio are highlighted. Here, however the stark difference in adoption between those areas and northern areas are readily apparent. This large difference in multi-year adoption vs single year adoption in northern areas may be driven by rotation effects, as the shorter growing season only allows for cover cropping behind early harvested crops.

**Differences Between Cover Cropped and Non-Cover-Cropped Fields**

Using a linear model with controls for weather and soil quality and a time trend, cover cropped areas had 0.21 Mg/ha lower yields for maize and 0.11 Mg/ha lower yields for soybean overall relative to conventional fields. Results of similar models built for each individual year are shown in panel A of Fig 3. For maize, the largest yield differences were seen for the crops harvested in 2014 and 2015 with yield decreases of as much as 0.50 Mg/ha. For soybean, the largest yield decreases

![Figure 3](image-url)
were in 2014 and 2012 with 0.19 and 0.15 Mg/ha. The gap between cover cropped and conventional yields increased with year for maize, but not soy.

Cover cropping was also associated with poorer quality soils, as seen in panel B. Pixels in the lowest 10% for soil quality (as measured by the National Commodity Crop Productivity Index) were 32% and 53% more likely than those at the median to be cover cropped for maize and soybean respectively. Pixels in the highest 10% of soil quality were 20% and 40% less likely than the median maize and soybean pixel to be cover cropped. This is despite the fact that soil quality was not included in the classifier used to determine presence/non-presence of cover crops. This pattern even carries over into states with high general soil quality (e.g. Iowa).

**Effects of Cover Cropping on Yield By Years Cover Cropped**

![Figure 4](image-url)

Figure 4: Yield effects of cover cropping for maize after one year (A) maize after three years (B) soybean after one year (C) soybean after three years (D).
While NCCPI serves as the best currently available nationwide soil quality index, it cannot account for soil properties such as compaction or nutrient depletion which vary at a scale far below what is mapped in SSURGO [35]. The index also cannot account for farmer practices which are likely to vary at the field or subfield level. The results of the panel regression created to eliminate the effects of such latent variables are shown in Fig 4. Overall, this model showed a 0.88% increase for maize yields with cover cropping and a 0.12% increase for soybean yields in areas that followed the practice.

The model used here also allowed for the examination of trends over space, though not over time. Panel A of Fig 4 shows that on maize fields, cover crops were most beneficial in Minnesota and Wisconsin where benefits were 0.23 and 0.14 Mg/ha respectively and least beneficial in Illinois and Indiana with 0.01 and 0.02 Mg/ha increase in yields with cover cropping. As shown in panel C, for soybean, benefits or losses were small, regardless of state, ranging from a loss of 0.003 Mg/ha in Iowa to a gain of 0.015 Mg/ha in Illinois.

Panels B and D of the figure show model outcomes comparing areas that had been cover cropped for at least three continuous years to areas not cover cropped. Here again Minnesota and Wisconsin show large benefits, for maize as illustrated in panel B; however, Indiana, and Michigan show yield decreases with long-term cover cropping (South Dakota was excluded here as <10 ha of maize and soybean area had followed the practice for three or more years in the dataset). For soybean, benefits or losses due to cover cropping again hover around zero, with -0.17 Mg/ha difference in Indiana and 0.54 Mg/ha difference in Iowa.
Discussion

New tools for processing large swaths of satellite imagery in bulk have helped overcome the limitations of previous imagery-limited approaches [43] and have created the opportunity to map cover crops reliably at the region scale. The approach taken here shows that while discussion of, funding for, and planting of cover crops have all increased significantly over the past decade, the practice is still only used on a small minority of maize and soybean fields in the heart of the Midwest.

Often, cover crops are put forward to farmers as a way of improving soil productivity [36] and 66% of growers agree or strongly agree that cover crops can have that effect [45]. Thus, it is perhaps unsurprising that patterns in cover crop planting show that they are more likely to be used in poor soils than in good ones. While soil quality can be partly observed at the 30-meter scale of analysis used here, elements of it cannot. Thus, it is also perhaps unsurprising that even with weather and soil controls, cover cropping is associated with poorer maize and soybean yields relative to conventional practice.

The real test of the effects of cover cropping on yields is whether cover cropped years of the same exact areas have higher yields than non-cover cropped years, controlling for weather. Here, cover crops show modest benefits of 0.88% and 0.12% for maize and soy. These benefits are comparable to [46] which found a 1.3% increase in maize yields, however, they are lower than the 3.8% soybean yield benefit reported. The SARE and CTIC numbers, however, were from a self-reported
survey of growers following the practice so respondents may be vulnerable to a choice supportive bias [47] of cover crop adopters.

Cover cropping for multiple years does not appear to add a clear direction to the effects of the practice. For maize, the four states that benefit most with one year of cover cropping also benefit the most with three years; however large losses from cover cropping become apparent in other areas, potentially due to imperfect controls. Overall, however, the data appear to indicate a lack of generalized economically large yield benefits to cover crops in the Midwest.

Caveats for the yields effects seen here include the fact that this analysis did not take into account cover crop type, found to have differential effects on yield by [48] and cultivar within a specific crop type, found to have differential effects by [49]. Additionally, yield outcomes in the analysis may be skewed toward zero by the inclusion of more cover cropped fields than actually present, adding noise to the regressions, or in a worst case, the classifier may in some places be detecting something other than cover crops such as poor weed management that has negative yield effects of its own. Yield predictions for cover cropped fields may similarly be skewed by the prevalence of the practice on silage and/or seed maize fields versus maize for grain, especially in northern sections of the study area where maize grown for silage is more common [7].

Despite these caveats, we believe that this is a valuable vein for future analysis, especially as both high-frequency satellite data necessary become more readily available and soil quality estimations improve with the additions of
NCCPIv3, Fragile Soil Index and soil susceptibility to compaction from future SSURGO versions.

References


24. NRCS. Cover Crop Termination Zones. 2013th ed. USDA NRCS ESD;


46. SARE, CTIC. 2016-2017 Cover Crop Survey. Sustanible Agriculture Research & Education Program; Conservation Technology Information Center; 2017 Sep p. 46.
