HHS Public Access

Author manuscript

J Expo Sci Environ Epidemiol. Author manuscript; available in PMC 2024 February 24.

Published in final edited form as:

J Expo Sci Environ Epidemiol. 2024 March ; 34(2): 294–307. doi:10.1038/s41370-022-00467-0.

Evaluating the Accuracy of Satellite-Based Methods to Estimate Residential Proximity to Agricultural Crops

Carly Hyland[1], **Kathryn McConnell**[2], **Edwin DeYoung**[3], **Cynthia L. Curl**[1]

[1]School of Public and Population Health, Boise State University, Boise, ID, USA

[2]Yale School of the Environment, New Haven, CT, USA

[3]Department of Geosciences, Boise State University, Boise, ID, USA

Abstract

Background: Epidemiologic investigations increasingly employ remote sensing data to estimate residential proximity to agriculture as a means of approximating individual-level pesticide exposure. Few studies have examined the accuracy of these methods and the implications for exposure misclassification.

Objectives: Compare metrics of residential proximity to agricultural land between a groundtruth approach and commonly-used satellite-based estimates.

Methods: We inspected 349 fields and identified crops in current production within a 0.5 km radius of 40 residences in Idaho. We calculated the distance from each home to the nearest agricultural field and the total acreage of agricultural fields within a 0.5 km buffer. We compared these groundtruth estimates to satellite-derived estimates from three widely used datasets: CropScape, the National Land Cover Database (NLCD), and Landsat imagery (using Normalized Difference Vegetation Index thresholds).

Results: We found poor to moderate agreement between the classification of individuals living within 0.5 km of an agricultural field between the groundtruth method and the comparison datasets (53.1–77.6%). All satellite-derived estimates overestimated the acreage of agricultural land within 0.5 km of each home (average =82.8–148.9%). Using two satellite-derived datasets in conjunction resulted in substantial improvements; specifically, combining CropScape or NLCD with Landsat imagery had the highest percent agreement with the groundtruth data (92.8–93.8% agreement).

Significance: Residential proximity to agriculture is frequently used as a proxy for pesticide exposure in epidemiologic investigations, and remote sensing-derived datasets are often the only practical means of identifying cultivated land. We found that estimates of agricultural

Corresponding Author: Dr. Carly Hyland, carlyhyland@boisestate.edu, 1910 University Drive, Boise, ID 83725, 208-426-3924.

Author Contributions:

CH: Data Collection, Formal Analysis, Writing – Original Draft; KM, Conception of Data Analysis, Formal Analysis, Writing – Reviewing and Editing; ED, Formal Analysis, Writing – Reviewing and Editing; CLC: Conceptualization, Writing – Reviewing and Editing.

Competing Interests

The authors declare no conflicts of interest.

proximity obtained from commonly-used satellite-based datasets are likely to result in exposure misclassification. We propose a novel approach that capitalizes on the complementary strengths of different sources of satellite imagery, and suggest the combined use of one dataset with high temporal resolution (e.g., Landsat imagery) in conjunction with a second dataset that delineates agricultural land use (e.g., CropScape or NLCD).

Keywords

pesticides; exposure modeling; geospatial analyses

Introduction

Research suggests that individuals living in agricultural areas have higher levels of pesticide exposure than the general population (1–4). Residential proximity to pesticide-treated fields may be adversely associated with a range of health outcomes, including pre-term birth (5), fetal growth parameters (6), birth defects (7, 8), and other birth outcomes (9, 10); childhood (11–16) and adult cancer (17–19); respiratory problems, such as asthma (20); adverse neurodevelopment, including decreased cognition (21, 22) and increased emotional and behavioral problems (23) and traits related to Autism Spectrum Disorder (24, 25); and neurodegenerative disorders such as Parkinson's disease (26–30). While previous epidemiologic studies have focused primarily on farmworkers and their families (31), non-farmworkers living in areas of intensive agricultural pesticide use may also be exposed to pesticides via spray drift during their application and via inhalation, dermal exposure, or ingestion from volatilization after application (32).

While biomonitoring is often considered the gold standard for exposure assessment, there are various challenges that can limit the utility of biological monitoring for pesticides, including the short half-lives of many pesticides (33), as well as the cost and availability of analytical methods. To combat these limitations, an increasing number of exposure assessment and environmental epidemiology studies have employed Geographic Information Systems (GIS) and remote sensing data to assess residential proximity to agricultural pesticide use. Specifically, studies have used various geospatial data sources including aerial photography (34, 35), publicly available pesticide use reports (16, 21–25), land use surveys/crop maps developed from satellite imagery (e.g., Landsat, Sentinel, state-wide crop maps) (36–39), and satellite imagery combined with publicly available pesticide use data (40–44) to estimate agricultural pesticide use near study participants' homes. Such data sources have many advantages and can be used in both prospective exposure assessment when methods like biomonitoring may not be available or are cost-prohibitive, as well as for retrospective exposure reconstruction, particularly for diseases with long latency periods. Remote sensing-derived products also offer a low-cost method to estimate use of multiple pesticides near individuals' homes, facilitating analyses examining the health effects of potential exposure to mixtures of pesticides (23), a growing field in environmental epidemiology (45, 46). Previous studies have employed remote sensing-derived datasets to estimate residential proximity to agricultural land as a proxy for individual-level pesticide exposure to investigate associations with numerous health outcomes (44), including various

types of cancer (47–49); asthma (50); birth defects (51, 52), neural tube defects (8), and birth weight (53); child neurodevelopment (54); and child blood pressure (55).

Despite the many advantages of employing satellite-derived crop estimates to approximate pesticide exposure, some limitations exist. These data do not directly describe the type or quantity of pesticides applied, and instead provide an indirect measure of potential exposure in the form of crop locations. The varying spatial and temporal resolution of satellite-derived crop datasets mean they may not reflect sub-pixel changes in land cover. Further, more recent land cover changes (i.e., urbanization, market-driven change in crop types, rotation of non-permanent crops) many not be immediately reflected (31, 36, 56). Previous studies have also largely relied on aggregated annual exposure models, which cannot capture seasonal changes in pesticide use or allow for the examination of critical periods of exposure (56). Due to these limitations, previous analyses have often crudely classified participants as living in close proximity or far from agricultural land (e.g., < vs. ≥ 0.5 km), which may result in misclassification error (31) that could bias epidemiologic effect estimates (56). Additionally, studies have found that examining the total acreage of agricultural fields within a particular buffer is a more reliable indicator of potential exposure to pesticides than the distance to the nearest agricultural field (37, 41); however, this metric is rarely used and itself may be incorrectly characterized due to the aforementioned limitations.

Given the many benefits of satellite-derived data products, including their low cost, ease of estimating potential exposure to multiple pesticides, and ability to reconstruct past exposures for diseases with long latency periods (e.g., in case-control studies), it is imperative to examine the accuracy of these method and the implications for exposure assessment and epidemiologic studies. The goal of this analysis was to compare estimates of 1) the proximity to the nearest agricultural field and 2) the total acreage of agricultural fields near the homes of 40 participants from a study in Idaho's Treasure Valley between a "groundtruth" gold standard method with other commonly used satellite-derived datasets. The results of this analysis can be used to improve future exposure assessment and epidemiologic studies employing satellite-derived datasets to estimate study participants' proximity to agricultural land as a proxy for pesticide exposure.

Methods

Study Background

This analysis originated from a larger study examining dietary and agricultural contributions to exposure to the herbicide glyphosate among 40 pregnant women living in southern Idaho. Participants were recruited during their first trimesters of pregnancy from Idaho Women, Infant, and Children (WIC) clinics in the Southwestern, Central, and South-Central Health Districts from January-June, 2021 and were followed until they gave birth (August-December, 2021). For the current analysis, we compare ground truth observations of agricultural fields within a 0.5 km buffer of each participant's home during the growing season (August 2021), and compare these metrics of residential proximity to cultivated agricultural crops with estimates taken from satellite imagery and satellite-derived data products.

Groundtruth Based Crop Locations

We plotted the address of each of the participant's home(s) where they reported living any time during the study period and visually identified all potential agricultural fields within a 0.5 km radius of each home in Google Earth in August 2021 (imagery captured from Google Earth between July 2018 and August 2020). We identified "potential fields" as areas of green or brown that did not contain homes, buildings or other structures, or were not obviously occupied by other use (e.g., baseball fields). One researcher (CH) went to each of the potential fields in August 2021, visually inspected it, and classified the area as: 1) a grass field, 2) under development/developed land, 3) not a field (e.g., a lawn, dirt, soccer field), 4) a dormant field with nothing currently planted, 5) inaccessible (e.g., inaccessible without trespassing on private property), or 6) an agricultural field currently being used for crop production. For each agricultural field in production, the specific crop that was growing was identified.

If no potential fields were identified within a 0.5 km radius, we inspected potential fields within a 0.75 km and 1.0 km radius, as necessary, in order to determine the distance from the participant's home to the closest field. If there were no agricultural fields with a 1.0 km radius of the participant's home, the nearest field was listed as ">1 km".

Selection of Comparison Geospatial Data Sources and Metrics

In order to compare metrics from our groundtruth approach with metrics from satellite-derived crop estimates, we selected the data sources that have been most frequently used in previous studies examining residential proximity to agricultural fields, including the CropScape Cropland Data Layer from the US Department of Agriculture (37, 38, 57), Landsat satellite imagery (40–42, 44, 58), and the National Land Cover Database (NLCD) (48, 59). We converted our ground truth data to a raster layer (with 30 m resolution) for direct comparison of each of the data sources across two metrics: 1) distance to the nearest agricultural field, and 2) the total acreage of agricultural fields within a 0.5 km buffer. Each of these datasets has a 30 m spatial resolution (60–62), meaning each pixel was 900 m^2.

Classification of Cultivated Land

To calculate groundtruth metrics, we used ArcGIS Pro 2.8.0 and its World Imagery base map imagery, which is derived from various data sources (63). First, we plotted the latitude and longitude of each home based on the geocoded street address. All addresses were self-reported by participants and verified by researchers during field work. Within a 0.5 km radius buffer around each home, we manually plotted polygons corresponding to crops identified during the groundtruth. Using this groundtruth layer, we then calculated the distance from each home point to the nearest crop (referred to as "crop distance") and total acreage of crops within a 0.5 km radius of each home point (referred to as "crop acreage").

We used R Studio (R Version 3.6.2) to calculate satellite-derived crop estimates from each of the three comparison data sources. We used the sf, raster, nngeo, and tidyverse packages for analysis (64–67). For each dataset, we imported the participants' geocoded addresses, plotted 0.5 km buffers around each address, and calculated crop distance and crop acreage. Crops are defined as follows for each of the three remote sensing-derived data sources:

The Cropland Data Layer (CDL), also referred to as CropScape, is a crop-specific land cover map using satellite imagery and agricultural ground truth that is produced by the National Agricultural Statistics Service (NASS) of the US Department of Agriculture (USDA) (68). We downloaded the CropScape data layer representing the state of Idaho for the year 2020 (the most recent year available) using the *CropScapeR* (69) package in R. The 2020 CDL was produced using satellite imagery from the Landsat 8 OLI/TIRS sensor, the Disaster Monitoring Constellation (DMC) DEIMOS-1, the ISRO ResourceSat-2 LISS-3, and the ESA SENTINEL-2 sensors collected during the growing season (70). We calculated the crop acreage and crop distance for all cultivated crops (field crops, vegetable crops, fruits, nuts). CDL codes corresponding to cultivated crops that we used are available in the supplementary material (Table S1).

NLCD is a Landsat-based land cover database that provides descriptive data for land-cover classes (e.g., urban, cultivated land, forest) (61). The Multi-Resolution Land Characteristics (MRLC) provides public access to the NLCD, which is released every 2–3 years (71). We downloaded the latest available NLCD dataset from 2019 (referred to as NLCD19), covering southern Idaho from https://www.mrlc.gov/viewer/. NLCD19 contains 34 different products that characterize land cover and land cover change from 2001–2019 (61). We calculated the metrics of interest using NLCD pixels classified as 82, corresponding to cultivated crops.

For Landsat-based crop estimates, we derived Normalized Difference Vegetation Index (NDVI) values and established a cutoff threshold to classify cultivated and non-cultivated pixels. NDVI is a measure of greenness that has been widely used for a variety of applications in remote sensing, including for pesticide exposure assessment (42) and detection of herbicide applications (72), and has been shown to be effective in differentiating different types of land cover, such as dense forest, non-forest, and agricultural fields (73). NDVI values range from −1.0 to +1.0, with higher NDVI values corresponding to dense vegetation or agricultural crops during their peak (74).

To establish an NDVI threshold above which pixels would be classified as cultivated, we first downloaded Landsat scenes from the month in which the groundtruth data were collected (August) and from the year in which the most recent NLCD data were available (2019). We utilized scenes from Landsat 8 OLI Collection 1 Analysis Ready Data (ARD). Because a single Landsat scene was not able to capture the entire study area, we downloaded separate scenes for participants clustered in the Western Idaho (e.g., Nampa/Caldwell area; captured August 9, 2019) and in Central Idaho (e.g., Twin Falls area; captured August 2, 2019). Due to cloud cover, we were not able to select scenes covering both regions from exactly the same dates. We then calculated NDVI for each of the two scenes, and calculated the 75^{th} and 90^{th} percentiles of NDVI within pixels designated as crops by NLCD (Figures S1 and S2 for Western and Central Idaho, respectively). “Landsat75” and “Landsat90” are subsequently used as distinct thresholds, above which we classify pixels as cultivated land and below which we classify pixels as not cultivated land. We applied these NDVI thresholds to Landsat 8 ARD scenes from 2021 captured on July 29, 2021 for Nampa and September 1, 2021 for Twin Falls. We selected these scenes because they were the closest available to the dates in which the groundtruth data were collected (August 2-August 31, 2021 for Nampa and August 18–19, 2021 for Twin Falls) and had minimal cloud coverage.

Prior to calculating NDVI for these scenes, we removed a small number of clouded pixels in Google Earth Engine using the Quality Assurance band.

Comparison of Cultivated Land Between Groundtruth and Satellite-Derived Estimates

To compare each of the three satellite-derived crop measures with our groundtruth measures, we compared the average acres of cultivated land estimated within a 0.5 km buffer of the participants' homes with the average acres estimated from the groundtruth data by dividing the absolute difference of the estimates by the average of the two estimates, aggregated across all home buffers, and multiplying by $100:100 \times \frac{|acres_{comparison} - acres_{ground-truth}|}{\frac{acres_{comparison} + acres_{ground-truth}}{2}}$.

For each of the five data sources (groundtruth, CropScape, NLCD, $Landat_{75}$, and $Landsat_{90}$), we created a raster layer with all of the area within 0.5 km of each participant-home classified dichotomously as cultivated crops (areas containing any agricultural fields currently being used for agricultural crop production; coded as "1") or non-cultivated area (e.g., barren land, water, grass/pasture, development, or other area containing land that was not currently being used for crop production; coded as "0"). Using the groundtruth data as a gold standard, we assessed the accuracy of each of the satellite-derived crop measures in designating areas as cultivated vs. non-cultivated land by calculating the following metrics: 1) overall percent agreement (i.e., the percent of pixels within a 0.5 km buffer of all participant-homes in which the satellite-based dataset agreed with the groundtruth data's designation of an area as cultivated vs. non-cultivated); 2) sensitivity (i.e., "true positive", percent of acreage within a 0.5 km buffer of all 49 participant-homes that each satellite-based dataset identified as cultivated that was designated as cultivated based on the groundtruth data); and 3) specificity (i.e., "true negative", percent of acreage within a 0.5 km buffer of all participant-homes that each satellite-based dataset identified as non-cultivated that was designated as non-cultivated based on the groundtruth data). Potential fields that were inaccessible were coded as "0" (non-cultivated) for the groundtruth method. We conducted a sensitivity analysis in which inaccessible areas were removed from the analysis.

In addition to comparing the groudtruth data with each of the four datasets individually, we also compared the sensitivity and specificity of the groundtruth data with a combination of each of the satellite-derived datasets. First, we considered any area to be cultivated land if *either* of the comparison datasets designated it as cultivated land. Second, we considered an area to be cultivated land if *both* comparison datasets designated it as cultivated land.

Results

The 40 participants in this study lived in a total of 49 different homes during the study period. We identified a total of 349 potential fields within a 0.5 km buffer of these homes using Google Earth; of these 55 (15.8%) were inaccessible. Of the 294 fields that were accessible, 27 (7.7%) were grass fields, 22 (6.3%) were developed or under development, 147 (42.1%) were another type of non-agricultural field (e.g., soccer field), and 13 (3.7%) were dormant; the remaining 85 (24.4%) were identified as agricultural fields (Table 1).

Table 2 and Figure 1 show the distance to the nearest cultivated agricultural field and the average acreage of cultivated fields estimated within 0.5 km of the 49 participant-homes from each dataset, as well as the percent difference in the estimates from the groundtruth data and the satellite-derived datasets. Each of the four satellite-derived datasets underestimated the crop distance (i.e., placed the nearest field closer to the participant's homes than it actually was) and overestimated the crop acreage around participant's homes. Of the four satellite-based datasets, NLCD Crop showed the greatest agreement with the groundtruth data in terms of residential proximity to the nearest field. Each of the Landsat-derived NDVI metrics estimated that over 90% of homes had an agricultural field within 100 meters, whereas our groundtruth data estimated that only 6% of homes had an agricultural field within 100 meters. We estimated an average of 13.6 acres of cultivated land within 0.5 km of each of the homes from the groundtruth data; NLCD was the closest to this estimation (32.8 acres; 82.8% difference) and $Landsat_{75}$ was the farthest (85.3 acres; 148.7% difference). While we estimated that about 51% of participant-homes had an agricultural field within 0.5 km, the satellite-derived datasets estimated there was an agricultural field within 0.5 km of 69–98% of homes, depending on the dataset.

Using the groundtruth distance to the nearest agricultural field, we categorized participants as living within 0.5 km and 1.0 km of a field from each of the four datasets, which represent commonly used metrics to categorize participants as living near or far from agricultural pesticide use (8, 10, 38, 40, 44, 56, 58), and compared how often each satellite-derived dataset agreed with the groundtruth data (Table 3). Following previous research with Interclass Correlation Coefficients (ICCs) (75), we considered values of less than 0.5 to be indicative of "poor" agreement, values between 0.5 and 0.75 indicative of "moderate" agreement, values between 0.75 and 0.9 indicative of "good" agreement, and values greater than 0.9 indicative of "excellent" agreement. At 0.5 km, we observed moderate agreement between CropScape (67.3%) and both Landsat-based metrics (53.1%) with the groundtruth data, and good agreement with the NLCD Crop method (77.6%). At 1.0 km, we still observed moderate agreement for each of the Landsat-derived estimates (65.3%), but good agreement with CropScape (81.6%) and NLCD Crop (85.7%). We observed the lowest agreement between both Landsat-based metrics and the groundtruth method regarding the percentage of participants that would be categorized as "near field" at both 0.5 km and 1.0 km, and the highest agreement with the NLCD Crop method.

Table 4 shows the percent agreement in the acreage of cultivated and non-cultivated land, as well as the sensitivity and specificity between our groundtruth method and each of the satellite-derived comparison datasets. NLCD and CropScape had the highest percent agreement with the groundtruth data; they each agreed with the designation of cultivated vs. non-cultivated land from the groundtruth method for about 88% of the total acreage. These datasets also had the highest sensitivity (84.4% and 83.3%, respectively) and specificity (88.5% and 88.2%, respectively). The Landsat-derived $NDVI_{75}$ estimate had the lowest agreement with the groundtruth data (56.2%) and had very low specificity (54.6%). Increasing the NDVI threshold with the $Landsat_{90}$ analysis increased the percent agreement with the groundtruth method to 75.9%, while also increasing the specificity significantly to 76.5%. However, increasing the NDVI threshold for $Landsat_{90}$ also decreased the sensitivity from 79.6% to 68.5%.

Figures 2–7 demonstrate the areas that each of the datasets designated as cultivated land for participants living in areas of high-density development (Figures 2 and 3), medium-density development (Figures 4 and 5), and low-density development (Figures 6 and 7). We use these figures to demonstrate some of the strengths and weaknesses of each dataset across different levels of urbanization.

As shown in Figures 2 and 3, CropScape and NLCD tended to agree with our groundtruth results in areas of high-density development with little to no cultivated land. In Figure 2, the participant lived in an urban area with one small grass field to the southwest of their home and no cultivated agricultural fields within a 0.5 km buffer. Thus, the true amount of cultivated land in this buffer was 0 acres^2. CropScape and NLCD each designated a small area that was actually just grass as cultivated crops; when we examined historical images of this area from Google Earth, we found that this area was, in fact, previously an agricultural field but has since been converted to uncultivated land. Because Landsat images are captured more frequently than CropScape or NLCD data layers are produced, both Landsat-derived crop estimates correctly identified this area as non-cultivated, but also tended to mis-designate sections of grass/yard as fields as cultivated. Increasing the NDVI cutoff from the 75th to the 90th percent (thus going from Landsat$_{75}$ to Landsat$_{90}$) decreased the amount of area incorrectly designated as cultivated crops, but still overestimated the acreage within the buffer that was considered cultivated. Similarly, in Figure 3, CropScape and NLCD were in almost perfect agreement with the groundtruth data that there were no cultivated crops within 0.5 km of the participant's home in an urban area. However, Landsat$_{75}$ and Landsat$_{90}$ estimated 101.1 and 38.0 acres of crops, respectively, all of which appears to be small sections of grass/yards.

Figures 4 and 5 illustrate common situations in areas of medium-intensity development. For Figure 4, groundtruthing showed that there was just one cultivated agricultural field within 0.5 km of the participant's home, in the extreme northwest corner of the buffer. GoogleEarth images showed four other possible fields, however groundtruthing confirmed that these areas had been developed since the GoogleEarth images were captured. CropScape designated the large majority of these recently developed areas as cultivated land, and NLCD designated one of them as such, which was likely due to urbanization that had occurred since the most recent release of CropScape and NLCD data. While the NDVI threshold approach correctly identified these areas as non-cultivated land (likely because these scenes are more recent), they also designated a baseball field and various sections of grass/yard as cultivated fields, regardless of the NDVI cutoff (though the overestimation of agricultural land was, of course, greater when the 75th percentile of the NDVI was used). In Figure 5, the participant lived in an area surrounded by multiple currently cultivated agricultural fields and one large area that was a new suburban housing development. Both CropScape and NLCD incorrectly designated this central housing area, including the actual location of the participant's home, as cultivated crops. In this scenario, Landsat imagery, and particularly when the NDVI cutoff was set at the 90th percentile, was better at estimating the cultivated acreage in this buffer and was closest to on-the-ground conditions because the Landsat images are more recent and were able to detect the new development, and because this area was so large (28 acres). Notably, there was one field near this participant's home that was inaccessible, and each of the comparison datasets designated this area as a cultivated field. However, this

area was relatively small (8 acres), and each dataset except for $Landsat_{90}$ still would have overestimated the amount of agricultural land within the buffer even if we had designated it as a cultivated crop.

As highlighted in Figures 6 and 7, all datasets tended to over-estimate the acres of agricultural crops in rural regions in which there was substantial non-crop vegetation. However, the number of acres that was estimated in the Landsat methods was much closer to, and sometimes even smaller than, the acres estimated from CropScape and NLCD in these rural areas. In Figure 6, the participant lives near two relatively large cultivated fields and several smaller non-cultivated fields (e.g., grass fields) to which agricultural pesticides are not likely to be applied. Similar to previous findings, Figure 6 shows that CropScape and NLCD identified non-agricultural fields as cultivated land. While the Landsat-based methods correctly identified these as non-cultivated areas, all of the datasets misclassified small patches of non-agricultural fields such as yards and baseball fields. Figure 7 similarly demonstrates that all datasets designated areas we identified as grass to be cultivated fields. Notably, none of the datasets captured the full areas we identified as cultivated fields; in Figures 7B and 7C, CropScape and NLCD missed some area to the north of the participant's home, whereas in Figures 7D and 7E, the Landsat-derived estimates missed some area to the south of the participant's home.

As demonstrated, each of the satellite-derived comparison datasets has various strengths and weaknesses. For example, the use of Landsat-based NDVI threshold values for crop detection avoids misclassifying recently urbanized areas, a limitation of NLCD and CropScape data. However, because NDVI is designed to measure general vegetation levels instead of specifically identifying cultivated land, this approach mis-identified small areas of grass such as yards as crops.

One approach to maximizing the strengths of each dataset is to consider two datasets in conjunction. Table 5 shows the percent agreement, sensitivity, and specificity when comparing our groundtruth data to a combination of two of the satellite-derived crop estimation methods. First, we considered an area to be cultivated land if *either* of the two satellite-derived datasets designated those pixels as cultivated crops. We observed significantly higher sensitivity (ranging from 93.2–95.1%), but lower specificity (ranging from 48.3–84.9%). By using either dataset to designate an area as cultivated land, we inherently increased the acreage considered to be true positives and false positives, thereby increasing the sensitivity but decreasing the specificity. We found that the combination of CropScape with NLCD or NLCD with $Landsat_{75}$ had the highest sensitivity (95.1% and 95.4%, respectively). Second, we considered an area to be cultivated land if *both* of the satellite-derived datasets designated that area as a field, and observed decreased sensitivity (ranging from 58.5–75.5%) but significantly increased specificity (91.2–96.2%). Similarly, by requiring two datasets to designate an area as cultivated land, we increased both the true negatives, thereby increasing the specificity, while also increasing the false negatives, thereby decreasing sensitivity. The combination of either CropScape or NLCD with $Landsat_{90}$ had the highest specificity (96.2% each); however, sensitivity values for both combinations were below 60%. Combining CropScape and NLCD had just slightly lower specificity (91.2%), while still maintaining relatively high sensitivity (75.5%).

Our results were similar and our overall interpretations did not change in sensitivity analyses in which we excluded inaccessible fields (Tables S2 and S3).

Discussion

In this analysis, we compared residential proximity to agricultural fields between a “gold standard” groundtruth method with four analyses from three commonly-used remote sensing and remote sensing-derived datasets, CropScape, NLCD, and Landsat (using NDVI cutoffs from the 75th and 90th percentiles). We observed moderate or good agreement between the classification of individuals living within 0.5 km or 1.0 km of an agricultural field, commonly used metrics to classify participants as living “near” or “far” from pesticide use, between our groundtruth method and the satellite-derived estimates. The comparison satellite-derived datasets tended to overestimate the total acreage of agricultural land within 0.5 km of each home, a metric that has been shown to be a better predictor of pesticide exposure than just distance alone (31, 41). We found that using two of the satellite-derived datasets in conjunction increased the sensitivity or specificity, depending on whether one or both datasets were required to designate an area as cultivated land; this has implications for different applications of remote sensing-derived crop datasets. Given the many strengths of employing remote sensing-derived crop estimates in pesticide exposure assessment and epidemiology studies, including their relatively low financial and logistical burdens, we advocate for these methods to continue to be used while properly acknowledging their strengths and limitations.

Many studies crudely classify study participants as living near or far from pesticide use based on whether they live within a certain distance of an agricultural field, such as 0.5 km or 1.0 km. Evidence suggests that variables such as the total acreage of crops within a particular buffer zone may be better predictors of pesticide exposure than dichotomously classifying participants as near versus far field or continuously examining their proximity to the closest field (31, 41); however, these data are not always available. Our results suggest that using any of the satellite-derived methods to classify participants as “near” or “far” field based on the presence of a field within 0.5 km will result in exposure misclassification. More specifically, the satellite-derived methods identified fields both in much closer proximity to participant-homes, and also a greater number of fields within 0.5 km of participant-homes than we did in our groundtruthing, which results in an over-estimation of the number of “near-field” participants. We found significantly better agreement between our groundtruth data and the comparison datasets at 1.0 km. CropScape and NLCD each agreed with the groundtruth data regarding the presence of an agricultural field within 1.0 km of participant’s homes over 80% of the time; these appear to be the most reliable datasets and metric if crude near vs. far-field classification is to be used. While there is no scientific consensus regarding the best buffer distance to use to assess nearby agricultural pesticide use, previous meta-analyses and other studies have shown that pesticide concentrations in environmental samples collected from homes decrease with increasing distances from pesticide-treated fields (3, 76). We chose to characterize fields within a 0.5 km buffer of homes, as this has been used in previous studies as an intermediate distance for nonspecific pesticide applications (40, 56).

In addition to underestimating the distance to the nearest agricultural field, we found that the satellite-derived datasets overestimated the total acreage of agricultural land near participant's homes. The primary reasons for this are that: 1) CropScape and NCLD fail to capture recent changes in land use and urbanization due to the lag time in their public release, and 2) the inability of NDVI to fully differentiate between vegetation generally and crops specifically, which limits the utility of Landsat imagery paired with NDVI data. The CropScape data layer is updated annually (60, 77), whereas the NLCD is available every 2–3 years (78) and new Landsat scenes are available approximately every 16 days (79). Here, we discuss some of the strengths, limitations, and implications of our findings for each method.

Individually, NLCD had the highest sensitivity and specificity, as well as the highest agreement regarding the presence of an agricultural field within 0.5 km or 1.0 km of the home when compared with the groundtruth data. Notably, we did not include the NLCD class "hay/pasture" (code 81) as cultivated crops in our analysis. While hay is often produced from cultivated alfalfa, and alfalfa is a common crop in Idaho, we found that areas we identified as alfalfa from the groundtruth data were often correctly classified as "cultivated crops" (code 82) in NLCD. Therefore, including the hay/pasture category as cultivated land would have resulted in an even greater overestimation of the acreage of agricultural land near participant's homes. We recommend that future investigations similarly exclude the hay/pasture category.

CropScape had slightly lower overall agreement with the groundtruth data than NLCD, but still had relatively high sensitivity, specificity, and agreement regarding the existence of crops within 0.5 km or 1.0 km of participant's homes.

One of the primary strengths of the Landsat-based method was the more accurate detection of urbanizing areas as non-cultivated fields. Agricultural land in Idaho, and particularly within the Treasure Valley where many of our study participants live, has been increasingly converted for housing development due to rapid population growth (see Supplementary Material Figure S3) (80), and datasets such as CropScape and NLCD may not accurately reflect recent changes in land use due to the infrequent nature in which new data are made available. However, one of the primary limitations of the NDVI threshold approach was the frequent misclassification of small areas of grass (e.g., lawns) as agricultural fields. Because we found that $Landsat_{75}$ designated a large acreage of yards as cultivated fields, we aimed to examine the impacts on specificity by increasing the NDVI threshold from the 75^{th} to the 90^{th} percentile. We found that even when we selected the 90^{th} NDVI percentile, the specificity was still lower than either CropScape or NLCD. In this study area, where water is widely used on yards and non-agricultural fields (e.g., soccer fields, school yards), this method does not appear to be able to differentiate agricultural fields from other greenspace, no matter what cutoff is chosen. This is not unexpected, as NDVI is a general measure of vegetation greenness, and not a measure of agricultural lands in particular. By definition, the sensitivity of $Landsat_{75}$ will always be higher and the specificity will always be lower than $Landsat_{90}$. By increasing the NDVI threshold that was required to designate an area as cultivated crop from $Landsat_{75}$, $Landsat_{90}$ decreased the acres of truly non-cultivated fields that was designated as cultivated, but also removed some of the acres of truly cultivated fields that were designated as cultivated.

Overall, CropScape and NLCD were highly sensitive at identifying “unexposed” participants in urban areas with no nearby agricultural fields. The NDVI approach designated numerous small sections of grass or yards in urban/suburban areas as cultivated fields, which would result in a consistent overestimation of exposure among truly non-exposed participants. CropScape and NLCD tended to overestimate cultivated fields in areas of rapid urbanization, as the most recently released data products lag by several years, and as such do not accurately capture recent urbanization. These datasets also were less specific in more rural areas where there is substantial green space (e.g., lawns, soccer fields, school fields) that was mistaken as cultivated land. The NDVI threshold approach was much better at not designating urbanizing areas as cultivated fields, as images are captured every two weeks, but also had low specificity. Increasing the NDVI threshold from the 75th percentile to the 90th percentile in Landsat images decreased the number of acres Landsat incorrectly identified as cultivated land, but also decreased the method’s ability to capture cultivated land.

These findings highlight the utility of using multiple datasets in conjunction to estimate residential proximity to agricultural land in order to maximize the aforementioned strengths while minimizing limitations of each method. While decisions of which specific datset to select depend on the goals of the investigator, including whether they would like to maximize sensitivity or specificity and the geographic landscape (e.g., level of urbanization) near participant’s homes, we advocate for the use of either CropScape or NLCD *and* Landsat$_{75}$ in order to achieve maximal sensitivity, specificity, and percent agreement and to leverage the strengths of each dataset. By combining these datasets, the investigator would benefit from the sensitivity of CropScape/NLCD while also increasing the specificity of identification of cultivated lands in urbanizing areas from the Landsat-based method. Additionally, requiring *both* datasets to designate an area as cultivated land mitigates concerns regarding the Landsat-based methods’ identification of lawns and small areas of grass as cultivated lands, as CropScape and NLCD each had high sensitivity. While using CropScape or NLCD *and* a Landsat-based method decreases the sensitivity compared to classifying an area as cultivated if *either* designates it as a crop, the trade-off is significantly increased specificity and overall percent agreement. Additionally, because the satellite-based methods consistently over-estimated the number of agricultural fields and total acreage of agricultural land near participant’s homes, decreasing the sensitivity may actually result in more accurate exposure classification. We advocate specifically for the use of CropScape or NLCD Crop with Landsat$_{75}$, rather than Landsat$_{90}$, in order to balance specificity (94.5–94.6%, respectively), maintain relatively high sensitivity (68.3–68.5%, respectively, compared to <60% for Landsat$_{90}$), and overall percent agreement (92.8–92.9%, respectively). Because the percent agreement statistic is more heavily driven by specificity than sensitivity, as non-cultivated land comprises significantly more of the acreage near participant’s homes than cultivated land, investigators may also consider conducting sensitivity analyses using a combination of other methods, such as NLCD or Landsat$_{75}$, which had the highest sensitivity (95.4%). However, while the time and resources needed to use two datasets in conjunction were relatively minimal for this small study, we acknowledge that combining datasets may not always be feasible in studies with a much larger number of participants. However, advances in geospatial cloud computing, such as the

Google Earth Engine platform, may help overcome barriers to processing high volumes of land cover imagery.

This analysis has various strengths and limitations. Importantly, we identified at least one potential field within 0.5 km of 13 of the 49 homes that we were not able to access because it was surrounded by private property, resulting in a potential under-estimation of agricultural land near these homes. However, our overall interpretations were qualitatively similar in sensitivity analyses where we excluded potential fields that were inaccessible; thus, we do not anticipate that not being able to access a relatively small percentage of the total potential field identified had a large impact on our analysis. Additionally, while we believe the results of this analysis have implications for studies elsewhere in the United States, further research is needed to analyze the performance of remote sensing-derived data products for exposure assessment in different countries, which may have distinct crop types, levels of greenness, and land use patterns.

Our analysis describes the benefits and potential biases of using commonly-used remote sensing-derived data sources to identify crop locations in exposure studies. We highlight three publicly-available data sources that can be used at minimal cost and computational expense. While we selected these data given their broad accessibility to researchers, there are a range of alternative data sources and land use classification techniques that could enhance future exposure studies.

There are a growing number of fine-scale remote sensing and aerial imagery data products available that could be used to identify crop locations with greater precision. The National Aerial Imagery Program (NAIP) operated by the USDA Farm Service Agency, for instance, provides coverage of the United States at spatial resolutions ranging from .5 to 2 meters (81). At its finest resolution, NAIP imagery is sixty times more finely resolved than Landsat imagery, meaning that it can be used to more precisely identify both small patches and the true boundaries of crop regions. In addition to public data sources such as NAIP, imagery produced by private companies are available with global coverage, and at very high spatial and temporal resolutions. These data sources are being actively incorporated into agricultural and land use research (82–84). However, privately-produced remote sensing data can come at high financial and computational costs, and thus may not be a feasible data source for use in all research contexts.

In addition to more finely-resolved spatial data, there are a range of machine learning techniques that can be used to classify pixels in remote sensing imagery. We chose not to employ machine learning in this investigation, as our goal was to compare groundtruth findings with methods that have commonly been adopted in this field to approximate pesticide exposure and can be used with minimal computational expense. However, future research should evaluate the extent to which unsupervised (no training data used) and supervised (training data used) algorithms can be utilized in an exposure assessment context, as machine learning techniques have been successfully used to classify land cover types with high precision in geographical research (85–88). These techniques would very likely offer an improvement in crop identification accuracy over our simpler to use but less discriminating NDVI threshold approach. Combining machine learning techniques with high-resolution,

high-frequency remote sensing imagery could overcome many of the limitations we identify with CropScape, NLCD, and Landsat-derived NDVI. Data available at higher spatial resolution could more accurately discriminate between non-agricultural and agricultural vegetation as well as the boundaries between agricultural and non-agricultural land covers. Data available at a higher temporal frequency could identify recent land cover changes associated with urbanization that take CropScape and the NLCD longer to detect. Applying machine learning techniques to these data could, in turn, yield more accurate crop estimates than NDVI metrics alone.

This study also has a number of strengths. It is the first to our knowledge to compare remote sensing-derived crop estimates that have been widely used to assess residential proximity to agricultural fields with a “gold standard” groundtruth approach and provides important information for the future use of these datasets. Although our sample size contained just 49 homes, we groundtruthed 349 potential fields and were able to capture homes in a range of rural, semi-urban, and urban areas with different levels of urbanization over time. While this study was conducted solely in Idaho, the weaknesses that we noted (e.g., the time-lag in which data become available, the rapid land-use changes from farmland (89), and the detection of lawns as agricultural land) are likely universal and applicable to other locations in the United States. Notably, this analysis lays the groundwork to conduct more accurate studies examining associations with residential proximity to agricultural land in areas where publicly available pesticide use data are not available. California is the only state in the U.S. with comprehensive agricultural pesticide use data reported annually, further highlighting the importance of understanding the strengths and weaknesses of different satellite-based datasets in studies that must rely on residential proximity to agricultural fields as a proxy for pesticide exposure. Additionally, in order to develop more sophisticated exposure models that could incorporate consideration of factors such as wind speed and direction, it is critical that the actual location of agricultural fields be correctly identified. Thus, choosing the most reliable sources of data to designate agricultural fields becomes even more critical. The current analysis can be used to contribute to research examining potential pesticide exposure among the millions of residents living in agricultural communities in states where public pesticide use reporting data are not available, which remains severely understudied.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We gratefully acknowledge all of the study participants.

Funding

Research reported in this publication was supported by the National Institute of Environmental Health Sciences of the National Institutes of Health under Award Number K01ES028745. The content of this manuscript, including all findings and conclusions, is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability Statement

To protect the privacy of our participants, data from this analysis are not available.

Abbreviations

CDL	Cropland Data Layer
GIS	Geographic Information Systems
MRLC	Multi-Resolution Land Characteristics
NASS	National Agricultural Statistics Service
NDVI	Normalized Difference Vegetation Index
NLCD	National Land Cover Database
USDA	United States Department of Agriculture
WIC	USDA's Special Supplemental Nutrition Program for Women Infants Children

REFERENCES

1. Hyland C, Laribi O. Review of take-home pesticide exposure pathway in children living in agricultural areas. Environ Res. 2017;156:559–70. [PubMed: 28437652]
2. Lopez-Galvez N, Wagoner R, Quiros-Alcala L, Ornelas Van Horne Y, Furlong M, Avila E, et al. Systematic Literature Review of the Take-Home Route of Pesticide Exposure via Biomonitoring and Environmental Monitoring. Int J Environ Res Public Health. 2019;16(12).
3. Deziel NC, Freeman LEB, Graubard BI, Jones RR, Hoppin JA, Thomas K, et al. Relative Contributions of Agricultural Drift, Para-Occupational, and Residential Use Exposure Pathways to House Dust Pesticide Concentrations: Meta-Regression of Published Data. Environ Health Perspect. 2017;125(3):296–305. [PubMed: 27458779]
4. Deziel NC, Friesen MC, Hoppin JA, Hines CJ, Thomas K, Freeman LE. A review of nonoccupational pathways for pesticide exposure in women living in agricultural areas. Environ Health Perspect. 2015;123(6):515–24. [PubMed: 25636067]
5. Larsen AE, Gaines SD, Deschênes O. Agricultural pesticide use and adverse birth outcomes in the San Joaquin Valley of California. Nature Communications. 2017;8(1):302.
6. Gemmill A, Gunier Robert B, Bradman A, Eskenazi B, Harley Kim G. Residential Proximity to Methyl Bromide Use and Birth Outcomes in an Agricultural Population in California. Environ Health Perspect. 2013;121(6):737–43. [PubMed: 23603811]
7. Rappazzo KM, Warren JL, Meyer RE, Herring AH, Sanders AP, Brownstein NC, et al. Maternal residential exposure to agricultural pesticides and birth defects in a 2003 to 2005 North Carolina birth cohort. Birth Defects Research Part A: Clinical and Molecular Teratology. 2016;106(4):240–9. [PubMed: 26970546]
8. Rull RP, Ritz B, Shaw GM. Neural Tube Defects and Maternal Residential Proximity to Agricultural Pesticide Applications. Am J Epidemiol. 2006;163(8):743–53. [PubMed: 16495467]
9. Carmichael SL, Yang W, Ma C, Roberts E, Kegley S, English P, et al. Joint effects of genetic variants and residential proximity to pesticide applications on hypospadias risk. Birth Defects Research Part A: Clinical and Molecular Teratology. 2016;106(8):653–8. [PubMed: 27098078]
10. Meyer Kristy J, Reif John S, Veeramachaneni DNR, Luben Thomas J, Mosley Bridget S, Nuckols John R. Agricultural Pesticide Use and Hypospadias in Eastern Arkansas. Environ Health Perspect. 2006;114(10):1589–95. [PubMed: 17035148]

11. Carozza SE, Li B, Wang Q, Horel S, Cooper S. Agricultural pesticides and risk of childhood cancers. Int J Hyg Environ Health. 2009;212(2):186–95. [PubMed: 18675586]
12. Gómez-Barroso D, García-Pérez J, López-Abente G, Tamayo-Uria I, Morales-Piga A, Pardo Romaguera E, et al. Agricultural crop exposure and risk of childhood cancer: new findings from a case–control study in Spain. International Journal of Health Geographics. 2016;15(1):18. [PubMed: 27240621]
13. Rull RP, Gunier R, Von Behren J, Hertz A, Crouse V, Buffler PA, et al. Residential proximity to agricultural pesticide applications and childhood acute lymphoblastic leukemia. Environ Res. 2009;109(7):891–9. [PubMed: 19700145]
14. Hyland C, Gunier RB, Metayer C, Bates MN, Wesseling C, Mora AM. Maternal residential pesticide use and risk of childhood leukemia in Costa Rica. Int J Cancer. 2018;143(6):1295–304. [PubMed: 29658108]
15. Lombardi C, Thompson S, Ritz B, Cockburn M, Heck JE. Residential proximity to pesticide application as a risk factor for childhood central nervous system tumors. Environ Res. 2021;197:111078.
16. Park AS, Ritz B, Yu F, Cockburn M, Heck JE. Prenatal pesticide exposure and childhood leukemia - A California statewide case-control study. Int J Hyg Environ Health. 2020;226:113486.
17. Jones RR, Yu C-L, Nuckols JR, Cerhan JR, Airola M, Ross JA, et al. Farm residence and lymphohematopoietic cancers in the Iowa Women's Health Study. Environ Res. 2014;133:353–61. [PubMed: 25038451]
18. Carles C, Bouvier G, Esquirol Y, Piel C, Migault L, Pouchieu C, et al. Residential proximity to agricultural land and risk of brain tumor in the general population. Environ Res. 2017;159:321–30. [PubMed: 28837904]
19. El-Zaemey S, Heyworth J, Fritschi L. Noticing pesticide spray drift from agricultural pesticide application areas and breast cancer: a case-control study. Australian and New Zealand Journal of Public Health. 2013;37(6):547–55. [PubMed: 24892153]
20. Raanan R, Gunier Robert B, Balmes John R, Beltran Alyssa J, Harley Kim G, Bradman A, et al. Elemental Sulfur Use and Associations with Pediatric Lung Function and Respiratory Symptoms in an Agricultural Community (California, USA). Environ Health Perspect.125(8):087007.
21. Coker E, Gunier R, Bradman A, Harley K, Kogut K, Molitor J, et al. Association between Pesticide Profiles Used on Agricultural Fields near Maternal Residences during Pregnancy and IQ at Age 7 Years. Int J Environ Res Public Health. 2017;14(5).
22. Gunier RB, Bradman A, Harley KG, Kogut K, Eskenazi B. Prenatal Residential Proximity to Agricultural Pesticide Use and IQ in 7-Year-Old Children. Environ Health Perspect. 2017;125(5):057002.
23. Hyland C, Bradshaw PT, Gunier RB, Mora AM, Kogut K, Deardorff J, et al. Associations between pesticide mixtures applied near home during pregnancy and early childhood with adolescent behavioral and emotional problems in the CHAMACOS study. Environmental Epidemiology. 2021;5(3).
24. Sagiv SK, Harris MH, Gunier RB, Kogut KR, Harley KG, Deardorff J, et al. Prenatal Organophosphate Pesticide Exposure and Traits Related to Autism Spectrum Disorders in a Population Living in Proximity to Agriculture. Environ Health Perspect. 2018;126(4):047012.
25. Shelton JF, Geraghty EM, Tancredi DJ, Delwiche LD, Schmidt RJ, Ritz B, et al. Neurodevelopmental Disorders and Prenatal Residential Proximity to Agricultural Pesticides: The CHARGE Study. Environ Health Perspect. 2014;122(10):1103–9. [PubMed: 24954055]
26. Brouwer M, Huss A, van der Mark M, Nijssen PCG, Mulleners WM, Sas AMG, et al. Environmental exposure to pesticides and the risk of Parkinson's disease in the Netherlands. Environment International. 2017;107:100–10. [PubMed: 28704700]
27. Costello S, Cockburn M, Bronstein J, Zhang X, Ritz B. Parkinson's Disease and Residential Exposure to Maneb and Paraquat From Agricultural Applications in the Central Valley of California. Am J Epidemiol. 2009;169(8):919–26. [PubMed: 19270050]
28. Wang A, Costello S, Cockburn M, Zhang X, Bronstein J, Ritz B. Parkinson's disease risk from ambient exposure to pesticides. European Journal of Epidemiology. 2011;26(7):547–55. [PubMed: 21505849]

29. Wang A, Cockburn M, Ly TT, Bronstein JM, Ritz B. The association between ambient exposure to organophosphates and Parkinson's disease risk. Occupational and Environmental Medicine. 2014;71(4):275. [PubMed: 24436061]
30. Manthripragada AD, Costello S, Cockburn MG, Bronstein JM, Ritz B. Paraoxonase 1, agricultural organophosphate exposure, and Parkinson disease. Epidemiology (Cambridge, Mass). 2010;21(1):87–94. [PubMed: 19907334]
31. Dereumeaux C, Fillol C, Quenel P, Denys S. Pesticide exposures for residents living close to agricultural lands: A review. Environment International. 2020;134:105210.
32. Teysseire R, Manangama G, Baldi I, Carles C, Brochard P, Bedos C, et al. Assessment of residential exposures to agricultural pesticides: A scoping review. PloS one. 2020;15(4):e0232258–e. [PubMed: 32343750]
33. Barr DB. Biomonitoring of exposure to pesticides. Journal of Chemical Health and Safety. 2008;15(6):20–9.
34. O'Leary ES, Vena JE, Freudenheim JL, Brasure J. Pesticide exposure and risk of breast cancer: a nested case-control study of residentially stable women living on Long Island. Environ Res. 2004;94(2):134–44. [PubMed: 14757376]
35. Brody JG, Vorhees DJ, Melly SJ, Swedis SR, Drivas PJ, Rudel RA. Using GIS and historical records to reconstruct residential exposure to large-scale pesticide application. J Expo Anal Environ Epidemiol. 2002;12(1):64–80. [PubMed: 11859434]
36. Rull RP, Ritz B, Shaw GM. Validation of self-reported proximity to agricultural crops in a case–control study of neural tube defects. J Expo Sci Environ Epidemiol. 2006;16(2):147–55. [PubMed: 16047039]
37. Plascak JJ, Griffith WC, Workman T, Smith MN, Vigoren E, Faustman EM, et al. Evaluation of the relationship between residential orchard density and dimethyl organophosphate pesticide residues in house dust. J Expo Sci Environ Epidemiol. 2019;29(3):379–88. [PubMed: 30254255]
38. Warren JL, Luben TJ, Sanders AP, Brownstein NC, Herring AH, Meyer RE. An evaluation of metrics for assessing maternal exposure to agricultural pesticides. J Expo Sci Environ Epidemiol. 2014;24(5):497–503. [PubMed: 24149974]
39. Avruskin GA, Meliker JR, Jacquez GM. Using satellite derived land cover information for a multi-temporal study of self-reported recall of proximity to farmland. J Expo Sci Environ Epidemiol. 2008;18(4):381–91. [PubMed: 17805231]
40. Ward MH, Nuckols JR, Weigel SJ, Maxwell SK, Cantor KP, Miller RS. Identifying populations potentially exposed to agricultural pesticides using remote sensing and a Geographic Information System. Environ Health Perspect. 2000;108(1):5–12.
41. Ward MH, Lubin J, Giglierano J, Colt JS, Wolter C, Bekiroglu N, et al. Proximity to crops and residential exposure to agricultural herbicides in iowa. Environ Health Perspect. 2006;114(6):893–7. [PubMed: 16759991]
42. Maxwell SK, Airola M, Nuckols JR. Using Landsat satellite data to support pesticide exposure assessment in California. International Journal of Health Geographics. 2010;9(1):46. [PubMed: 20846438]
43. Maxwell SK. Downscaling Pesticide Use Data to the Crop Field Level in California Using Landsat Satellite Imagery: Paraquat Case Study. Remote Sensing. 2011;3(9).
44. VoPham T, Wilson JP, Ruddell D, Rashed T, Brooks MM, Yuan JM, et al. Linking pesticides and human health: a geographic information system (GIS) and Landsat remote sensing method to estimate agricultural pesticide exposure. Appl Geogr. 2015;62:171–81. [PubMed: 28867851]
45. Gibson EA, Goldsmith J, Kioumourtzoglou M-A. Complex Mixtures, Complex Analyses: an Emphasis on Interpretable Results. Current environmental health reports. 2019;6(2):53–61. [PubMed: 31069725]
46. Hamra GB, Buckley JP. Environmental Exposure Mixtures: Questions and Methods to Address Them. Curr. 2018;5(2):160–5.
47. Brody JG, Aschengrau A, McKelvey W, Rudel RA, Swartz CH, Kennedy T. Breast cancer risk and historical exposure to pesticides from wide-area applications assessed with GIS. Environ Health Perspect. 2004;112(8):889–97. [PubMed: 15175178]

48. VoPham T, Brooks MM, Yuan JM, Talbott EO, Ruddell D, Hart JE, et al. Pesticide exposure and hepatocellular carcinoma risk: A case-control study using a geographic information system (GIS) to link SEER-Medicare and California pesticide data. Environ Res. 2015;143(Pt A):68–82. [PubMed: 26451881]
49. Marusek JC, Cockburn MG, Mills PK, Ritz BR. Control selection and pesticide exposure assessment via GIS in prostate cancer studies. Am J Prev Med. 2006;30(2 Suppl):S109–16. [PubMed: 16458785]
50. Bukalasa JS, Brunekreef B, Brouwer M, Koppelman GH, Wijga AH, Huss A, et al. Associations of residential exposure to agricultural pesticides with asthma prevalence in adolescence: The PIAMA birth cohort. Environment International. 2018;121:435–42. [PubMed: 30266014]
51. Meyer KJ, Reif JS, Veeramachaneni DNR, Luben TJ, Mosley BS, Nuckols JR. Agricultural pesticide use and hypospadias in eastern Arkansas. Environ Health Perspect. 2006;114(10):1589–95. [PubMed: 17035148]
52. Ochoa-Acuña H, Carbajo C. Risk of limb birth defects and mother's home proximity to cornfields. Sci Total Environ. 2009;407(15):4447–51. [PubMed: 19427676]
53. Xiang H, Nuckols JR, Stallones L. A geographic information assessment of birth weight and crop production patterns around mother's residence. Environ Res. 2000;82(2):160–7. [PubMed: 10662530]
54. Friedman E, Hazlehurst M, Loftus C, Karr C, McDonald K, Suarez-Lopez J. Residential proximity to greenhouse agriculture and neurobehavioral performance in Ecuadorian children. Int J Hyg Environ Health. 2019;223.
55. Suarez-Lopez J, Hong V, McDonald K, Suarez-Torres J, López D, Cruz F. Home proximity to flower plantations and higher systolic blood pressure among children. Int J Hyg Environ Health. 2018;221.
56. Rull RP, Ritz B. Historical pesticide exposure in California using pesticide use reports and land-use surveys: an assessment of misclassification error and bias. Environ Health Perspect. 2003;111(13):1582–9. [PubMed: 14527836]
57. Ochoa-Acuña H, Carbajo C. Risk of limb birth defects and mother's home proximity to cornfields. Sci Total Environ. 2009;407(15):4447–51. [PubMed: 19427676]
58. Xiang H, Nuckols JR, Stallones L. A Geographic Information Assessment of Birth Weight and Crop Production Patterns around Mother's Residence. Environ Res. 2000;82(2):160–7. [PubMed: 10662530]
59. Jones RR, Yu CL, Nuckols JR, Cerhan JR, Airola M, Ross JA, et al. Farm residence and lymphohematopoietic cancers in the Iowa Women's Health Study. Environ Res. 2014;133:353–61. [PubMed: 25038451]
60. United States Department of Agriculture (USDA) National Agricultural Statistics Service. Reserach and Science. CropScape and Cropland Data Layers - FAQs 2021 [Available from: https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php.
61. Multi-Resolution Land Characteristics Consortium (MRLC). Land Cover 2019 [Available from: https://www.mrlc.gov/data/type/land-cover#:~:text=The%20National%20Land%20Cover%20Database,land%20cover%20and%20associated%20changes.
62. United States Geological Survey (USGS). What are the band designations for the Landsat satellites? 2021 [Available from: https://www.usgs.gov/faqs/what-are-band-designations-landsat-satellites.
63. Esri. World Imagery Basemap [Available from: https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9.
64. Pebesma E. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 2018;10(1):439–46.
65. Hijmans RJ, van Etten J. raster: Geographic analysis and modeling with raster data. R package version 2.0–12. 2012 [Available from: http://CRAN.R-project.org/package=raster.
66. Dorman M, Rush J, Hough I, Russel D, Ranghetti L, Benini A, et al. nngeo: k-Nearest Neighbor Join for Spatial Data [Available from: https://cran.r-project.org/web/packages/nngeo/index.html.

67. Wickham A, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. Journal of Open Source Software. 2019;4(43):1686.
68. Boryan C, Yang Z, Mueller R, Craig M. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. Geocarto International. 2011;26(5):341–58.
69. Chen B. CropScapeR: Access Cropland Data Layer Data via the 'CropScape' Web Service. R package version 1.1.1. 2020 [
70. United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS). 2020 Idaho Cropland Data Lyaer | NASS/USDA 2020 [Available from: https://www.nass.usda.gov/Research_and_Science/Cropland/metadata/metadata_id20.htm.
71. Yang L, Jin S, Danielson P, Homer C, Gass L, Bender SM, et al. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. ISPRS Journal of Photogrammetry and Remote Sensing. 2018;146:108–23.
72. Pause M, Raasch F, Marrs C, Csaplovics E. Monitoring Glyphosate-Based Herbicide Treatment Using Sentinel-2 Time Series—A Proof-of-Principle. Remote Sensing. 2019;11(21):2541.
73. Pettorelli N, Vik JO, Mysterud A, Gaillard JM, Tucker CJ, Stenseth NC. Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol Evol. 2005;20(9):503–10. [PubMed: 16701427]
74. United States Geological Survey (USGS). NDVI, the Foundation for Remote Sensing Phenology 2018 [Available from: https://www.usgs.gov/special-topics/remote-sensing-phenology/science/ndvi-foundation-remote-sensing-phenology#:~:text=NDVI%20values%20range%20from%20%2B1.0,(approximately%200.2%20to%200.5).
75. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–63. [PubMed: 27330520]
76. Gibbs JL, Yost MG, Negrete M, Fenske RA. Passive Sampling for Indoor and Outdoor Exposures to Chlorpyrifos, Azinphos-Methyl, and Oxygen Analogs in a Rural Agricultural Community. Environ Health Perspect. 2017;125(3):333–41. [PubMed: 27517732]
77. Jin S, Homer C, Yang L, Danielson P, Dewitz J, Li C, et al. Overall Methodology Design for the United States National Land Cover Database 2016 Products. Remote Sensing. 2019;11(24).
78. Homer C, Dewitz J, Jin S, Xian G, Costello C, Danielson P, et al. Conterminous United States land cover change patterns 2001–2016 from the 2016 National Land Cover Database. ISPRS Journal of Photogrammetry and Remote Sensing. 2020;162:184–99. [PubMed: 35746921]
79. United States Geological Survey. What are the acquisition schedules for the Landsat satellites? [Available from: https://www.usgs.gov/faqs/what-are-acquisition-schedules-landsat-satellites.
80. Dolven RI. Urban Sprawl and Farmland Protection: Responding to Changes in Idaho's Treasure Valley. Idaho L Rev. 2021;269.
81. Maxwell A, Warner T, Vanderbilt B, Ramezan C. Land Cover Classification and Feature Extraction from National Agriculture Imagery Program (NAIP) Orthoimagery: A Review. Photogrammetric Engineering & Remote Sensing. 2017;83:737–47.
82. Cheng Y, Vrieling A, Fava F, Meroni M, Marshall M, Gachoki S. Phenology of short vegetation cycles in a Kenyan rangeland from PlanetScope and Sentinel-2. Remote Sensing of Environment. 2020;248:112004.
83. Breunig FM, Galvão LS, Dalagnol R, Dauve CE, Parraga A, Santi AL, et al. Delineation of management zones in agricultural fields using cover–crop biomass estimates from PlanetScope data. International Journal of Applied Earth Observation and Geoinformation. 2020;85:102004.
84. Turker M, Ozdarici A. Field-based crop classification using SPOT4, SPOT5, IKONOS and QuickBird imagery for agricultural areas: a comparison study. International Journal of Remote Sensing. 2011;32(24):9735–68.
85. Kussul N, Lavreniuk M, Skakun S, Shelestov A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. IEEE Geoscience and Remote Sensing Letters. 2017;14(5):778–82.

86. Neetu Ray SS. Exploring Machine Learning Classification Algorithms for Crop Classification Using Sentinel 2 Data. Int Arch Photogramm Remote Sens Spatial Inf Sci. 2019;XLII-3/W6:573–8.

87. Zurqani HA, Post CJ, Mikhailova EA, Cope MP, Allen JS, Lytle BA. Evaluating the integrity of forested riparian buffers over a large area using LiDAR data and Google Earth Engine. Sci Rep. 2020;10(1):14096.

88. Duda T, Canty M. Unsupervised classification of satellite imagery: Choosing a good algorithm. International Journal of Remote Sensing. 2002;23(11):2193–212.

89. United States Department of Agriculture (USDA) Economic Research Service. Major Land Uses [Available from: https://www.ers.usda.gov/topics/farm-economy/land-use-land-value-tenure/major-land-uses/.

Impact Statement:

The goal of this analysis was to compare commonly-used satellite-based estimates of residential proximity to agricultural crops with estimates from a “gold standard” groundtruth approach. The results of this analysis suggest that datasets such as CropScape and the National Landcover Database (NLCD) have higher sensitivity than Landsat-based methods, but the latter is better at identifying developed areas as non-cultivated land. Remote sensing datasets are increasingly being employed to examine residential proximity to agricultural land as a proxy for pesticide exposure; we advocate for the use of CropScape or NLCD in conjunction with a Landsat-based classification method to minimize exposure misclassification.

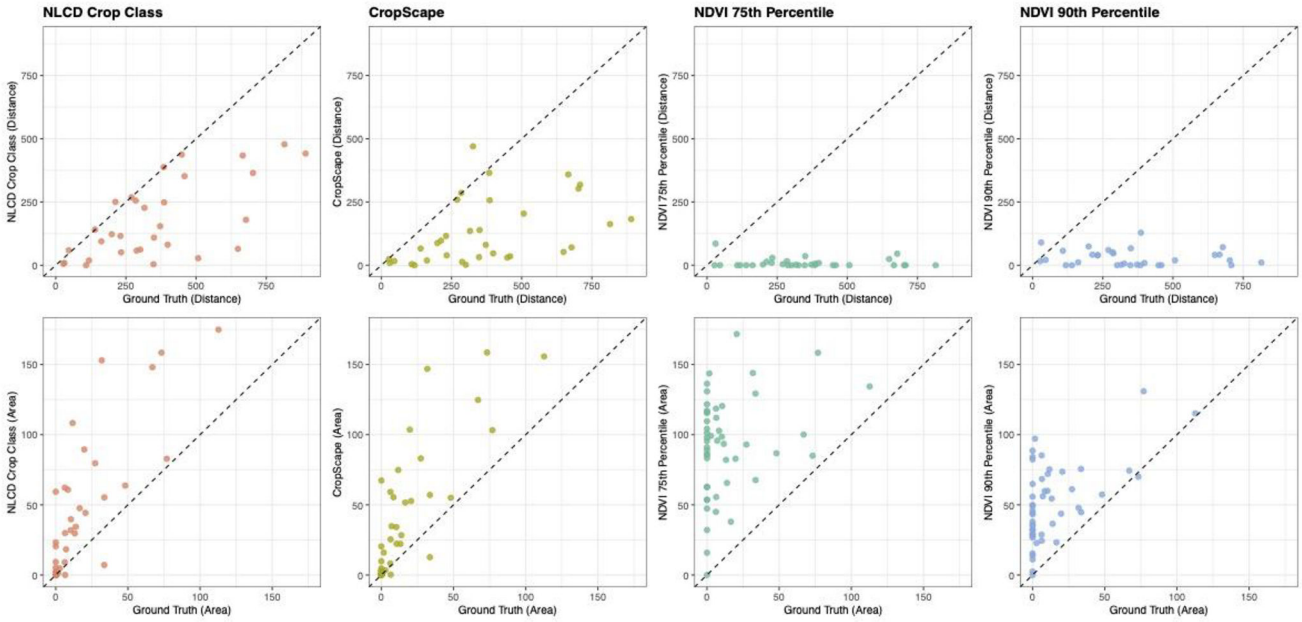

Figure 1.

Crop distance and crop acreage estimated from ground-truth and satellite-based comparison methods (ground-truth distances > 1,000 m excluded)

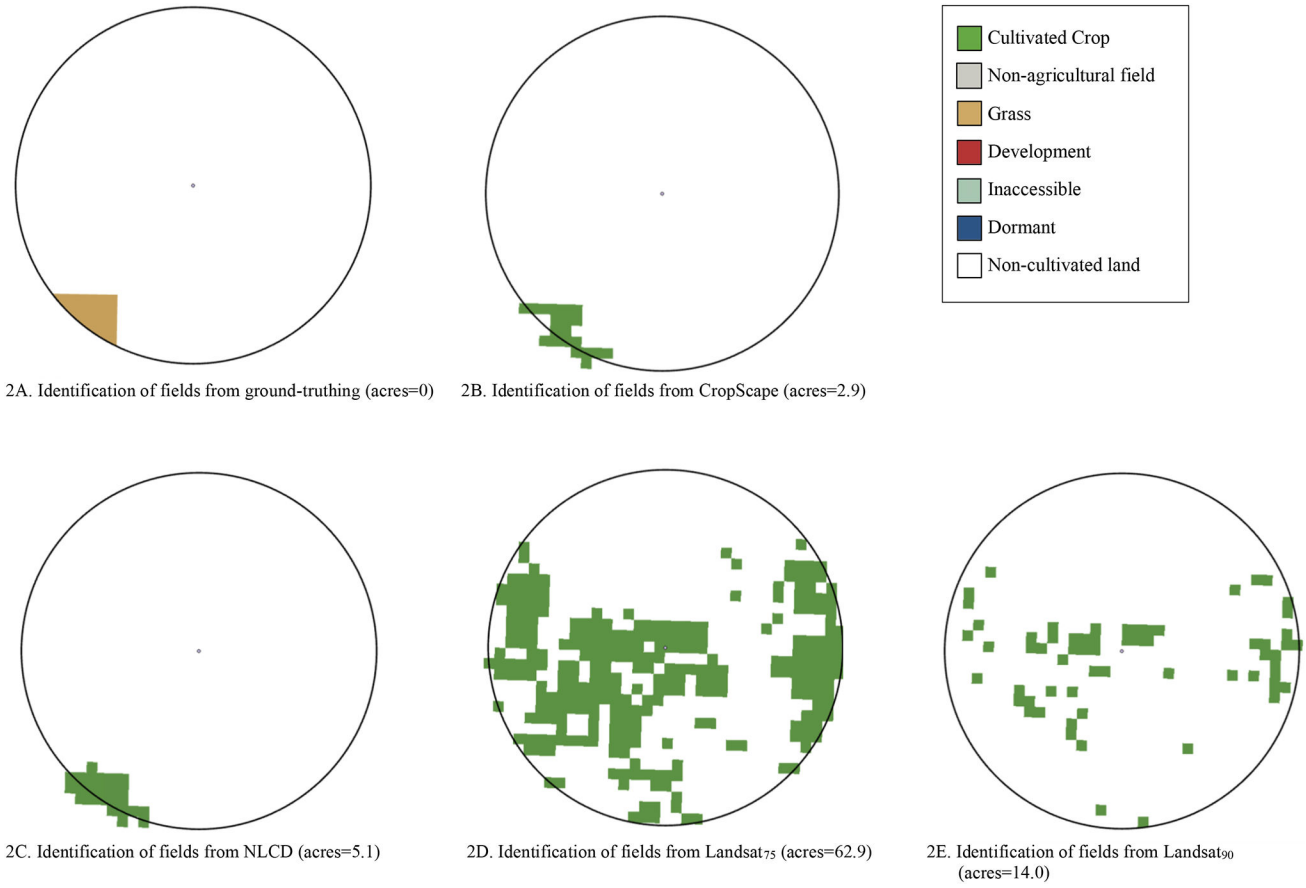

Figure 2.
Identification of fields from all methods (participant living in area of high-density development)

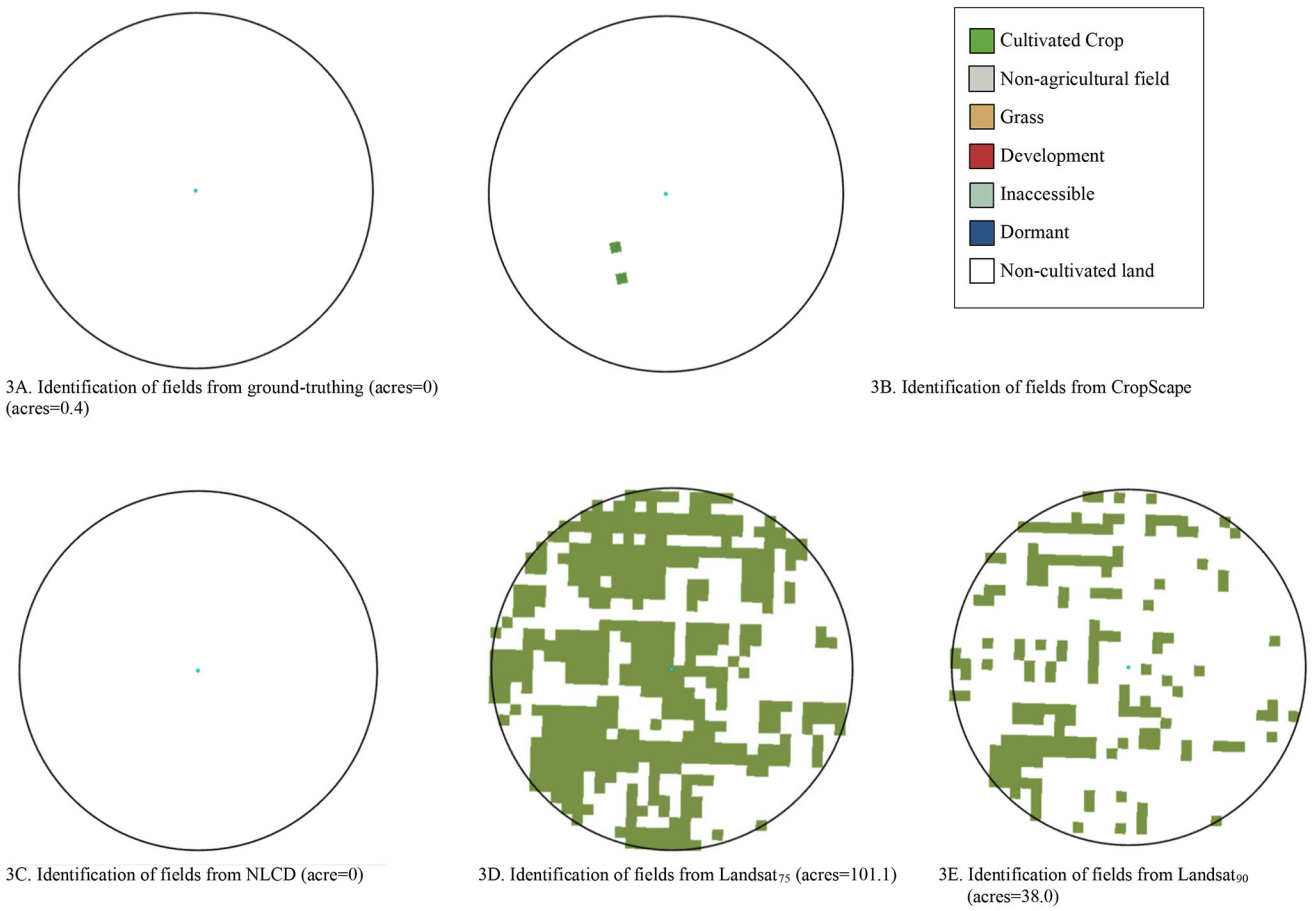

Figure 3.

Identification of fields from all methods (participant living in area of high-density development)

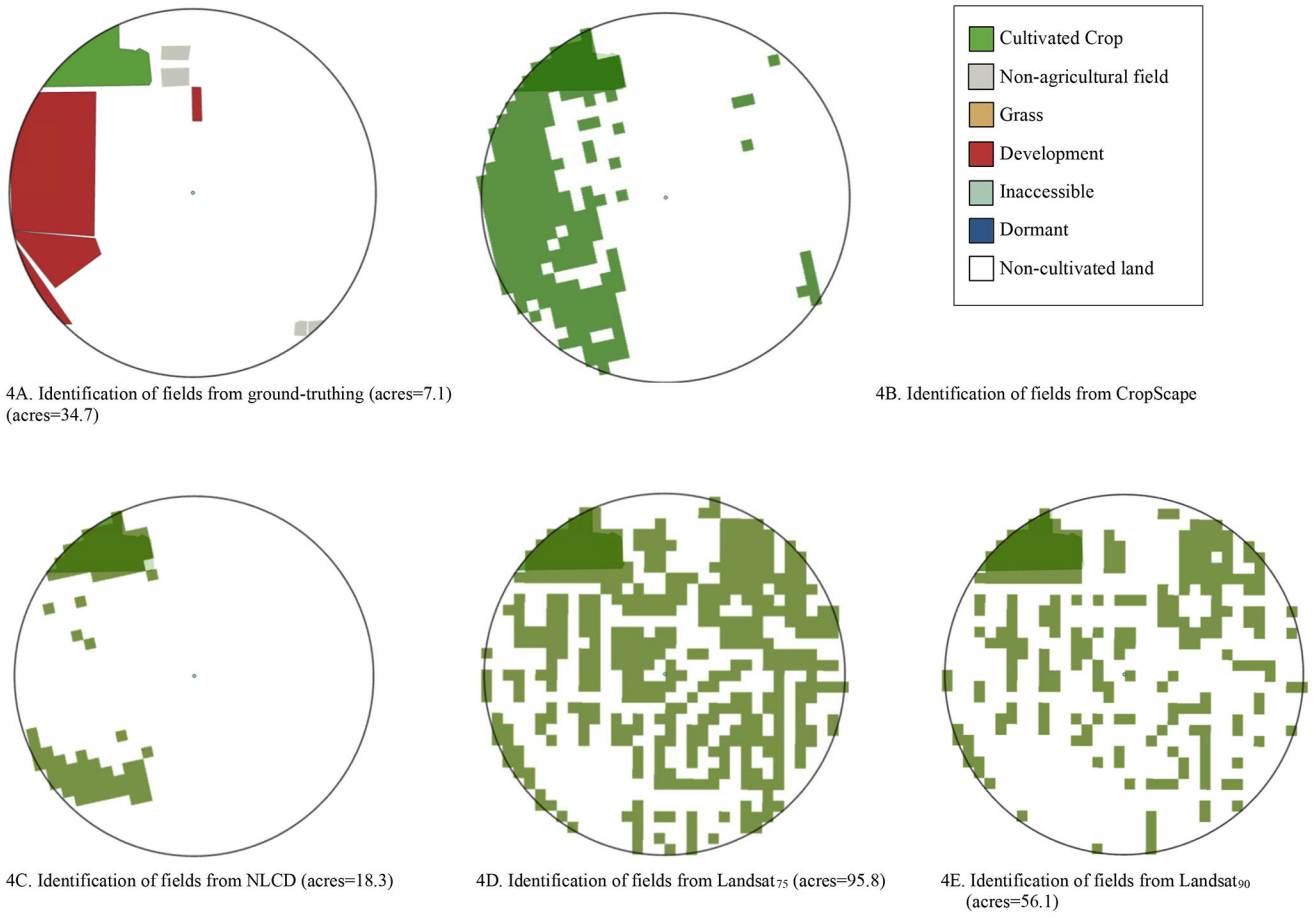

Figure 4.
Identification of fields from all methods (participant living in area of medium-density development)

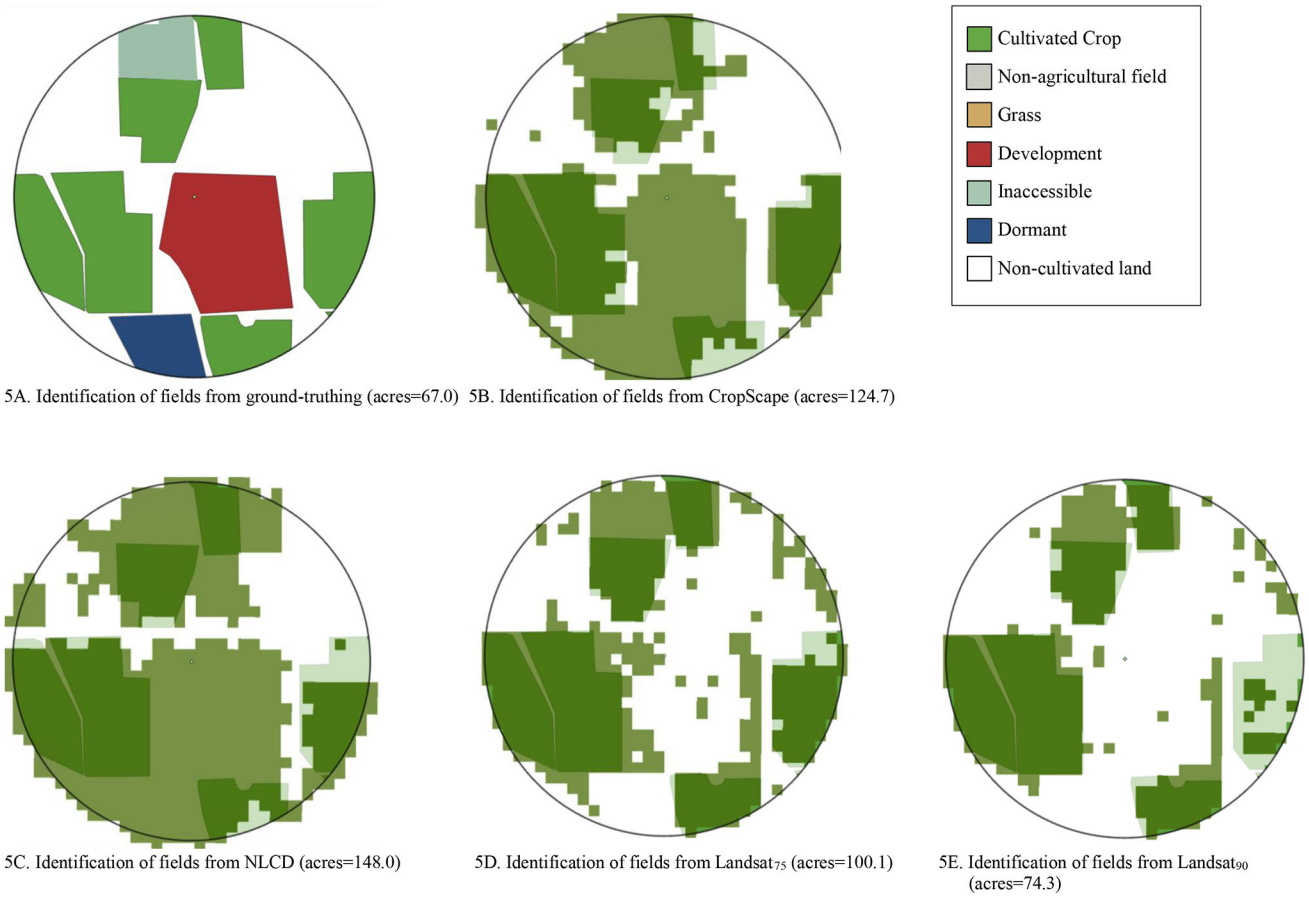

Figure 5.
Identification of fields from all methods (participant living in area of medium-density development)

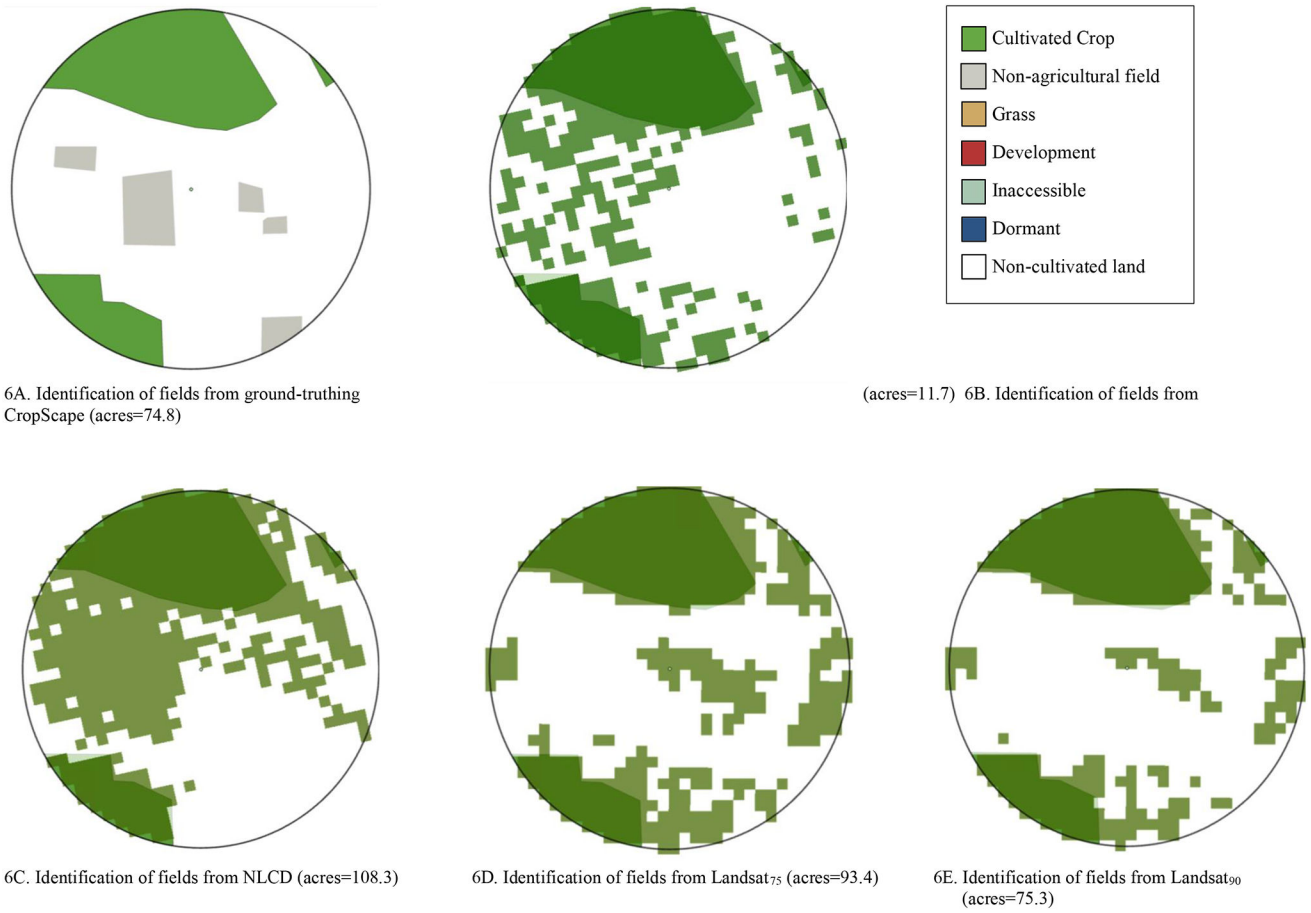

Figure 6.
Identification of fields from all methods (participant living in area of low-density development)

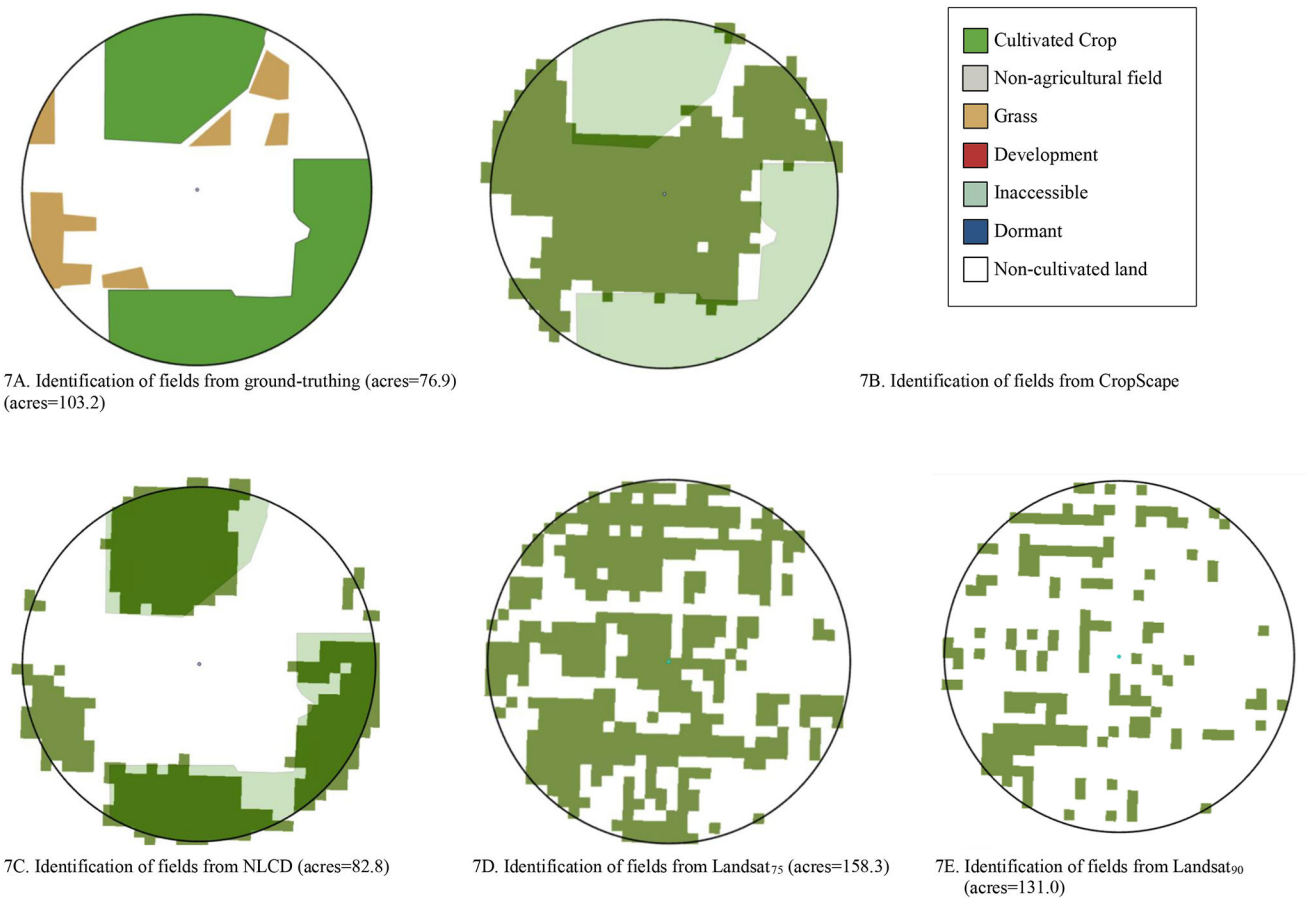

Figure 7.
Identification of fields from all methods (participant living in area of low-density development)

Table 1.

Identification of potential fields with a 0.5 km buffer of participant's homes (*n*=349)

Identification	*n* (%)
Agricultural field[1]	85 (24.4)
Alfalfa	36 (42.4)
Corn	23 (27.1)
Mint	3 (3.5)
Onion	3(3.5)
Soybeans	5 (5.9)
Straw/hay	6 (7.1)
Sugarbeets	4 (4.7)
Wheat	5 (5.9)
Non-agricultural field	147 (42.1)
Inaccessible	55 (15.8)
Grass Field	27 (7.7)
Developed/under development	22 (6.3)
Dormant	13 (3.7)

[1]Number and percentage of specific crops represent the proportion of fields within the "agricultural field" category

Table 2.

Distance to nearest agricultural crop and average acres of agricultural crops within 0.5 km of each participant-home from each method

	Ground-truthing	CropScape	NLCD Crop	Landsat-derived $NDVI_{75}$	Landsat-derived $NDVI_{90}$
Distance to nearest agricultural field (*n* [%])					
≤ 100 m	3 (6.1)	20 (40.8)	13 (26.5)	48 (98.0)	46 (93.9)
> 100 m– ≤ 500 m	22 (44.9)	21 (42.9)	21 (42.9)	0 (0.0)	2 (4.1)
> 500 m – ≤ 1,000 m	8 (16.3)	1 (2.0)	2 (4.1)	0 (0.0)	0 (0.0)
> 1,000 m	16 (32.7)	7 (14.3)	13 (26.5)	1 (2.0)	1 (2.0)
Average acres	13.6	33.8	32.8	92.5	50.9
Percent difference in average acres from ground-truthing data	-	85.3	82.8	148.7	115.7

Table 3.

Percent agreement of existence of agricultural crops within 0.5 km and 1.0 km buffers of participant's homes between ground-truthing method and satellite-derived crop estimates

CropScape		**NLCD Crop**		**Landsat-derived $NDVI_{75}$**		**Landsat-derived $NDVI_{90}$**	
Percent agreement within 0.5 km	Percent agreement within 1.0 km	Percent agreement within 0.5 km	Percent agreement within 1.0 km	Percent agreement within 0.5 km	Percent agreement within 1.0 km	Percent agreement within 0.5 km	Percent agreement within 1.0 km
67.3%	81.6%	77.6%	85.7%	53.1%	65.3%	53.1%	65.3%

Table 4.

Sensitivity and specificity in pixels of cultivated crops and non-cultivated land within 0.5 km of each participant-home between ground-truthing and satellite-derived estimates

Method	Overall agreement[1] (%)	Sensitivity (%)	Specificity (%)
CropScape	87.9	83.3	88.2
NLCD Crop	88.2	84.4	88.5
Landsat-derived $NDVI_{75}$	56.2	79.6	54.6
Landsat-derived $NDVI_{90}$	75.9	68.5	76.5

[1]Overall agreement in area designated as cultivated crops vs. non-cultivated area

Table 5.

Sensitivity and specificity in pixels of cultivated crops and non-cultivated land within 0.5 km of each participant-home between ground-truthing and combinations of satellite-derived estimates

Method	Overall agreement[1] (%)	Sensitivity (%)	Specificity (%)
CropScape or NLCD Crop[2]	85.4	95.1	84.9
CropScape or Landsat-dervied $NDVI_{90}$[2]	51.3	94.5	48.3
CropScape or Landsat-derived $NDVI_{90}$[2]	70.1	93.2	68.4
NLCD Crop or $Landsat_{75}$[2]	51.6	95.4	48.5
NLCD Crop or $Landsat_{90}$[2]	70.4	93.3	68.7
CropScape and NLCD Crop[3]	90.7	75.5	91.2
CropScape and Landsat-dervied $NDVI_{75}$[3]	92.8	68.3	94.5
CropScape and $Landsat\text{-}derived_{90}$[3]	93.8	58.5	96.2
NLCD Crop and $Landsat_{75}$[3]	92.9	68.5	94.6
NLCD Crop and $Landsat_{90}$[3]	93.8	59.6	96.2

[1] Overall agreement in area designated as cultivated crops vs. non-cultivated area

[2] Area considered cultivated land if *either* comparison method designated it as cultivated land

[3] Area considered cultivated land if *both* comparison methods designated it as cultivated land