A probabilistic crop-footprint approach to characterize potential pesticide use sites for endangered species assessments at the national scale

Katie Budreski[1], Michael Winchell[1], Lauren Padilla[1], JiSu Bang[2], Richard Brain[2]

STONE ENVIRONMENTAL INC

1. Stone Environmental, Inc. 2. Syngenta Crop Protection, LLC

syngenta

Abstract: A methodology for developing probabilistic crop use site footprints to estimate the likelihood of pesticide use was tested at the national scale and compared to alternative methods. The probabilistic aspect of the approach accounts for annual crop rotations and the uncertainty in remotely sensed crop and land cover datasets. The crop use site footprints currently recommended by EPA (and followed by the FIFRA Endangered Species Task Force (FESTF)) are derived purely from the National Land Cover Dataset (NLCD) Cultivated Cropland and/or Pasture/Hay classes from a single snapshot in time to represent the crop use site footprint for an agricultural pesticide. This current approach is overly conservative in the representation of individual crops, does not use the best available crop data, represents a single point in time, and does not account for the uncertainty in land cover dataset classifications. The probabilistic crop use site footprint approach incorporates best available information from the National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) for the most recent 5 years, the National Land Cover Dataset (NLCD), the 2007 NASS Census of Agriculture, and multiple years of NASS Quick Stats. The approach accounts for misclassification of crop classes in the CDL by incorporating accuracy assessment information by state, year, and crop. The NLCD provides additional information to influence the CDL crop probability through an adjustment based on the NLCD accuracy assessment data using the principles of the Bayes' Theorem. Finally, crop probabilities are scaled at the state level by comparing against NASS surveys (Census of Agriculture and Quick Stats) of reported planted acres by crop. In an application of the new method, the probabilistic crop footprint for soybean resulted in a national soybean acreage that is within the error bounds of the average reported NASS yearly soybean acreage over the same time period, whereas the standard method using only NLCD resulted in an acreage that is over four times the survey acreage. When the probabilistic crop use site footprint for soybean was used in a co-occurrence analysis with endangered species locations (MJD), the number of affected species was reduced by nearly half when compared to the standard NLCD method (from 511 to 276 species). The probabilistic crop use site footprint methodology allows for a more comprehensive and representative understanding of the pesticide exposure risk to endangered species and can be used in subsequent co-occurrence, proximity, and downstream dilution analyses.

Introduction

Background

- An important component of endangered species assessments (ESAs) is the definition of crop footprints that represent potential sites for pesticide applications based on uses approved on the pesticide label.
- In many previous ESAs, crop footprints have been based on generalized land use, such as the National Land Cover Dataset (NLCD) cultivated crops and pasture/hay classes, without information concerning individual crop locations or historic use. Using generalized land use information is overly conservative and often not representative of potential or historic use of a pesticide.

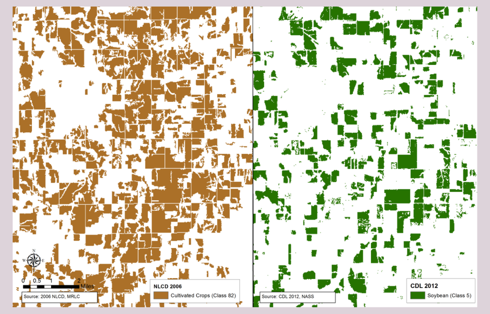


Figure 1. Comparison of generalized land cover data (NLCD 2006, Cultivated Crops Class 82, Left) with the more detailed land cover data that is now available for the entire US (CDL 2012, Soybean, Class 5, Right).

- The National Academy of Sciences (NAS) has made recommendations to the EPA and the Services (US Fish & Wildlife and National Marine Fisheries) to include the best available data in ESAs and have listed the Cropland Data Layer (CDL) as a source of best available crop/land cover information (NAS 2013). In general, the NAS committee recommended incorporating probabilistic approaches within ESAs to address uncertainty.
- In this assessment, a methodology for developing probabilistic crop footprints to estimate the likelihood of pesticide use was tested at the national scale and compared to alternative methods. The methodology incorporated the most up-to-date and best available crop specific location information including the 2008- 2012 CDL, land cover information from the 2006 NLCD, 2007 Census of Agriculture (AgCensus) and 2008-2012 NASS Quick Stats.

Study Area

- The continental United States was evaluated to test the proposed approach.
- Soybean was used as the example target crop for pesticide applications, which is the crop with the second largest acreage planted in the United States (approximately 77 million acres planted annually, USDA NASS 2014).

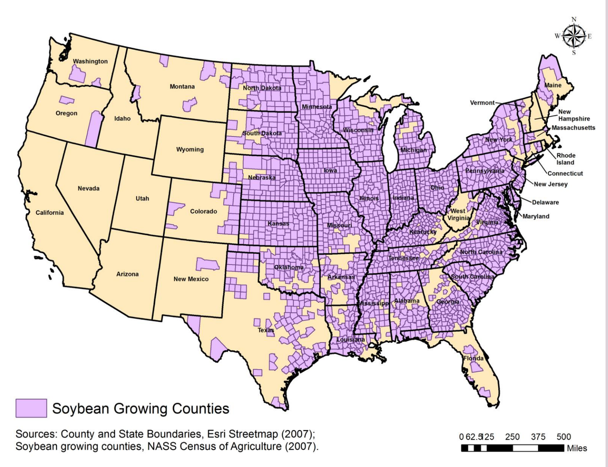


Figure 2. The study area includes soybean growing areas in the continental US.

Objective

- Develop a probabilistic methodology that uses publically available, high resolution geospatial datasets to create refined crop footprints representing potential pesticide use sites for endangered species assessments. This crop footprint development method is referred to as the 5-Year CDL, Bayesian Probability method.

Methodology

Datasets

The following datasets were used to develop the potential pesticide use footprint.

- CDL Crop Classifications: The Cropland Data Layer (CDL) from the National Agricultural Statistics Service (NASS) (USDA 2008-2012) was used to locate crop areas of interest. The CDL dataset has 141 classes of crop and other land uses. Five years of data were incorporated into the analysis to account for crop rotation and classification errors of omission. The approach also accounts for potential errors of commission and misclassification of crop classes in the CDL by incorporating accuracy assessment information by state, year, and crop.

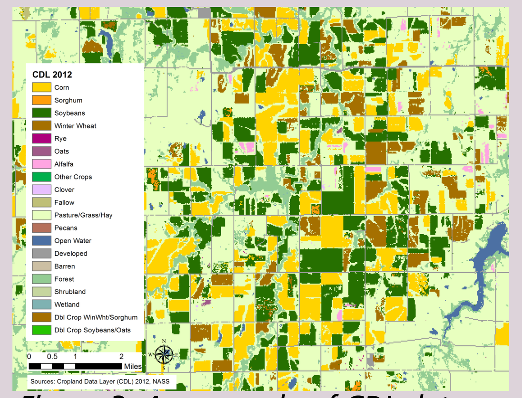


Figure 3. An example of CDL data within Kansas

- NLCD Land Cover: The NLCD provides additional information to influence the CDL crop probability through an adjustment based on the NLCD accuracy assessment data using the principles of Bayes' Theorem. It is assumed that the probability of a soybean pixel actually being soybean may be higher, lower, or unchanged depending upon which NLCD class the soybean pixel coincides with. For example, soybean pixels that co-occur with NLCD class 'Cultivated Crops' would have a higher probability of actually being soybean. Conversely, soybean pixels that co-occur with NLCD medium and high intensity developed classes would have a lower probability of actually being soybean.

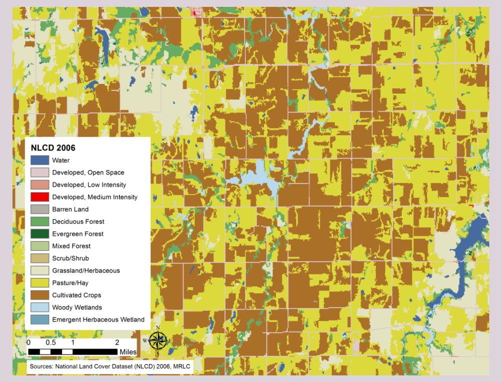


Figure 4. NLCD 2006 data within Kansas

- 2007 Ag Census and NASS Quick Stats: To account for 'omission errors', the 2007 Census of Agriculture and NASS Quick Stats reported acreages by state and crop were incorporated. The NASS reports a confidence interval for each crop by year. If the commission adjusted CDL acreage fell within the 90% confidence interval, no further action was taken. For commission adjusted CDL acreage that was outside the confidence interval, the ratio of NASS state acreage ("state adjustment ratio") to "commission adjusted" CDL acreage for each year was calculated.

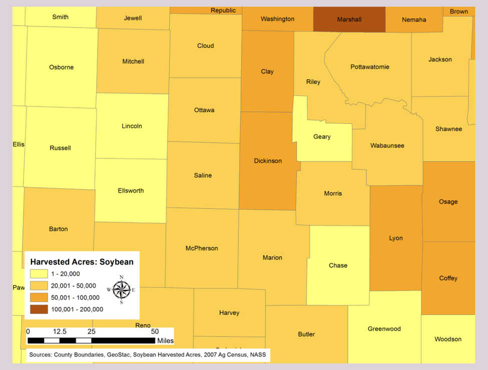
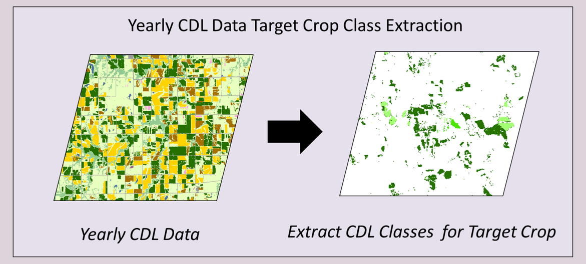


Figure 5. 2007 AgCensus harvested acres for soybean within Kansas.

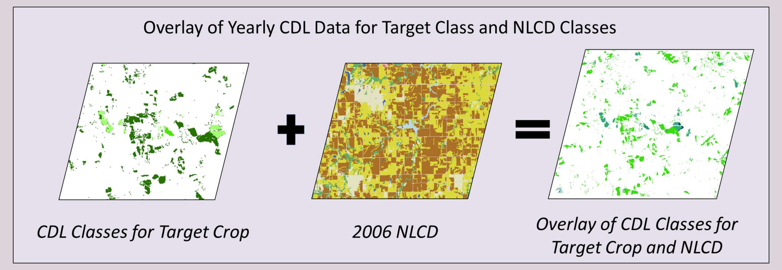
Probabilistic Crop Footprint Analysis Approach

The spatial extent of crop footprints was characterized for soybean.

- *Step 1, Base-layer Development: A CDL crop footprint, by year, was developed where each pixel (representing a 30x30 m grid cell) preserved the original CDL crop class for the target crop for each year from 2008 to 2012. Soybean (Class 5), Dbl Crop Soybeans/Cotton (Class 239), Dbl Crop Soybeans/Oats (Class 240), and Dbl. Crop Barley/Soybeans (Class 254) pixels were extracted.*



- *Step 2, 2006 NLCD Overlay:* An overlay of the yearly CDL soybean footprint and 2006 NLCD was created. All original classes were preserved.



- *Step 3, Commission Adjustment:* The Bayesian probability was calculated for each pixel based on CDL accuracy by year, state, and class; and NLCD accuracy by NLCD class. This is referred to as the 'Commission Adjustment'.

$$p(soy|nlcd) = p(soy) * p(nlcd|soy) / [p(soy) * p(nlcd|soy) + p(soy') * p(nlcd|soy')]$$

where:
- $p(soy|nlcd)$ was the posterior probability that a pixel is soybean
- $p(soy)$ was the probability a given soybean pixel in the class was correctly classified (i.e. CDL user's accuracy),
- $p(soy')$ was the probability a soybean pixel in the class was a false positive (i.e. CDL commission error),
- $p(nlcd|soy)$ was the conditional probability of mapping a pixel to an NLCD class, given that the pixel is actually correctly identified as soybean, and
- $p(nlcd|soy')$ was the conditional probability of mapping a pixel to an NLCD class, given that the pixel is incorrectly identified as soybean.

- *Step 4, Omission Adjustment:* Probability adjusted acreages were calculated and compared to the 2007 Ag Census or NASS Quick Stats acreages for soybean. If the probability adjusted acreages were outside of the state yearly NASS acreages, probabilities were adjusted. This is referred to as the 'Omission Adjustment'.
- *Step 5, Final Yearly Crop Footprint:* The 'Commission Adjusted' raster and the 'Omission Adjusted' raster were merged to create final yearly probabilistic crop footprints.
- *Step 6, Combine Yearly Crop Footprints:* The average probability was calculated on a pixel based over the 5 years.
- *Step 7, Final Crop Footprint Check:* For original CDL soybean pixels with a final probability of 0, assign 0.0001. This step ensures that no original soybean pixels are removed from further analysis, even though the calculated probability of being soybean is zero.

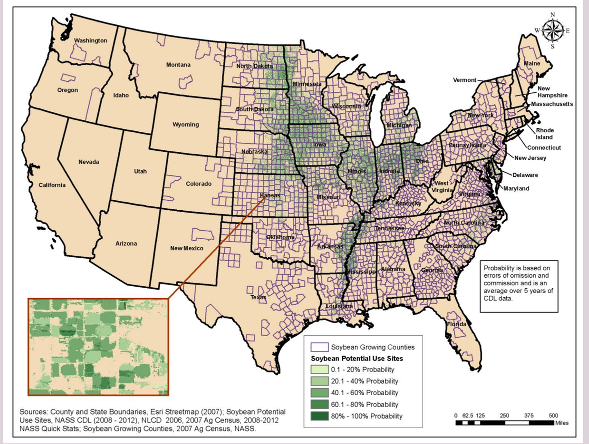


Figure 6. Final 5-Year CDL, Bayesian Probability Crop Footprint

Results and Discussion

Probabilistic Crop Footprint Acreage Comparison

- The 5-Year CDL, Bayesian Probability method was compared with four other methods.
  1) NLCD 2006, Cultivated Crop Class
  2) 1-Year CDL, All Soybean Classes
  3) 5-Year CDL, All Soybean Classes
  4) 5-Year CDL, All Soybean Classes, Probability Based on # of Years Soybean is Present (e.g., soybean in 1 of 5 years would equal a probability of 0.20).
- The Bayesian probability method incorporates CDL data along with 2006 NLCD and USDA NASS estimates of soybean acreage. The results are similar to the 5-Year CDL, Simple Probability method, but more conservative. It is the only method that results in acreage within the confidence bounds of the NASS reported acreage.

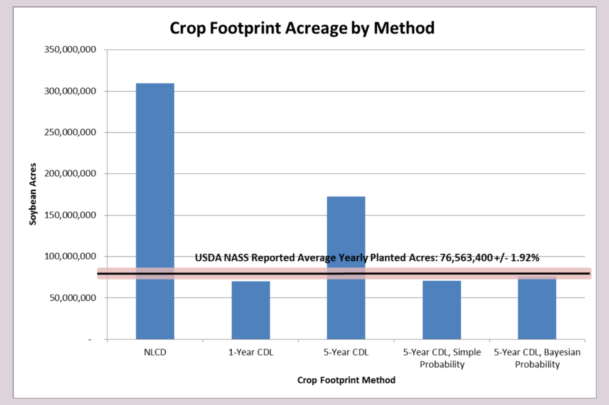


Figure 7.Comparison of soybean acres in the continental US using five different crop footprint methods and comparison with USDA NASS reported average yearly planted acreage for soybean.

- In Figure 8, the NLCD method (a) ignores crop-specific data, resulting in a crop footprint that covers nearly the entire area. The 1-Year CDL method (b) results in the smallest crop area, has evidence of 'spurious' pixels around field edges, and does not account for crop rotation. The 5-Year CDL (c), assigns all pixels an equal probability of being soybean in any single year, although does exclude crop areas that are never soybean, as opposed to the NLCD method. The 5-Year CDL, Simple Probability method (d), accounts for crop rotation, the probability of pixels being soybean in any single year, and errors of omission; however ignores errors of commission and accuracy information available for the CDL. The 5-Year CDL, Bayesian Probability method (e) accounts for crop rotation, the probability of pixels being soybean in any single year, errors of omission and commission and incorporates accuracy information available for the CDL.

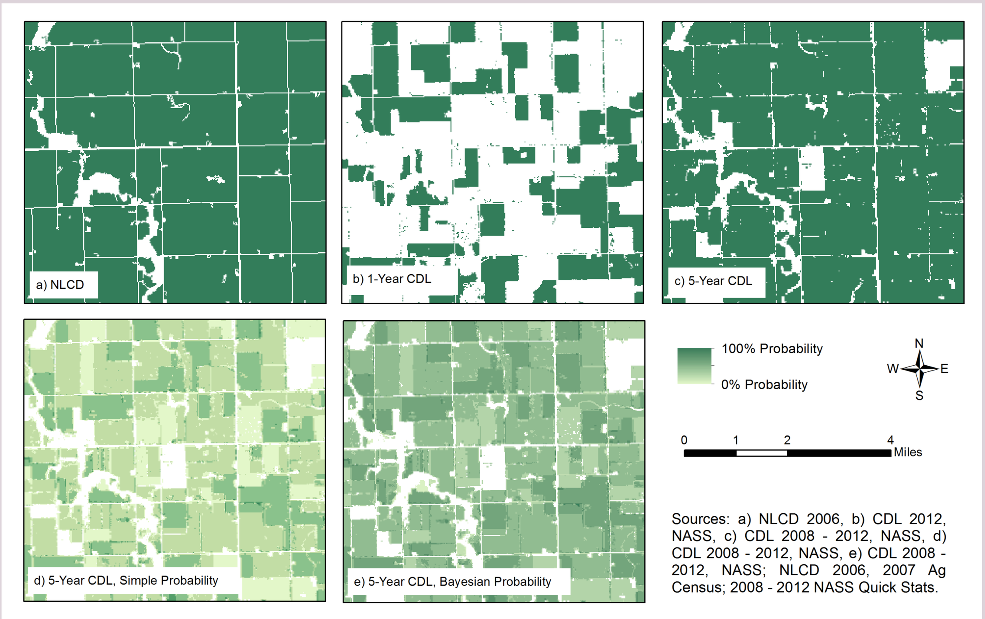


Figure 8. Comparison of crop footprint extents at the field level in Dallas County, Iowa for five different crop footprint methods.

Results and Discussion, Cont.

Co-occurrence Analysis

- A co-occurrence analysis is often implemented in risk assessments for endangered species to understand the extent of overlap of potential pesticide use sites and habitat areas. For probability-based crop footprints, overlap area was adjusted based on probability at the pixel level.
- Using the 5-Year CDL, Bayesian Probability generated crop footprint, the co-occurrence with current (1980-present) endangered species elemental occurrences (EOs) is 518,427 acres compared to 11,736,734 acres when using the crop footprint based on 2006 NLCD cultivated crops class.
- Based on the 5-Year CDL, Bayesian Probability method, there are 139 EOs (2.4% of EOs that overlap with soybean), that have less than or equal to 0.0001% of their total area in soybean and 59% that have less than 1% of their total area in soybean. This compares to only 19% with less than 1% of the total area in soybean for the NLCD method.

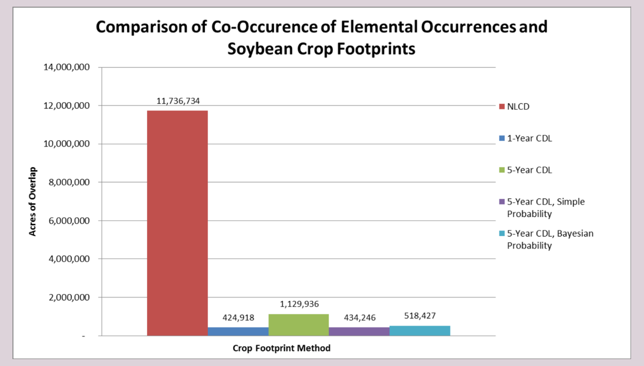


Figure 9. Comparison of the co-occurrence of elemental occurrences and soybean crop footprints by footprint method.

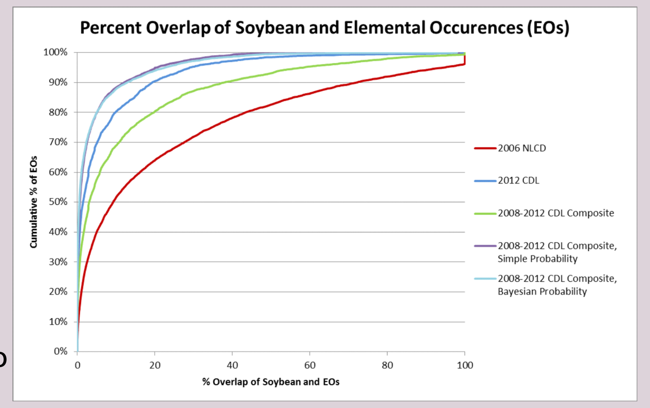


Figure 10. Cumulative distribution of overlap of soybean and elemental occurrences (EOs) by crop footprint method. Only EOs that overlap with soybean are illustrated.

Summary

- A methodology was developed to use publically available, high resolution datasets to refine crop footprints representing potential pesticide use sites for ESAs.
- Best available, crop-level and yearly land cover data is utilized.
- There are known and quantifiable uncertainties in land cover datasets that can be accounted for using probabilistic methods.
- Using yearly data can help understand and account for crop rotation and changes in land use over time.
- The probabilistic crop footprints represent the same amount of acreage as NASS reported acreages by state.
- The probabilistic crop footprints reduce the anomalous influence of 'spurious' pixels without removing them from the analysis.
- Approaches that use probabilistic crop footprints to represent the likelihood of species overlap with potential pesticide use are currently being developed in accordance to the NAS panel recommendations.

Acknowledgements

- The funding for the project was provided by Syngenta Crop Protection, LLC.
- For additional information: Michael Winchell, Stone Environmental, 802-229-1882, mwinchell@stone-env.com.

References

Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham. 2011. Completion of the 2006 National Land Cover Database for the conterminous United States. PE&RS 77:858-864.

USDA National Agricultural Statistics Service 2007. 2009. Census of Agriculture [Internet]. [cited 2014-03-25]. Available from: http://www.agcensus.usda.gov/Publications/2007/Full_Report/. USDA-NASS, Washington, DC, USA.

USDA (United Stated Department of Agriculture). National Agricultural Statistics Service Cropland Data Layer. 2008 - 2012. Published crop-specific data layer [Online]. Available at http://nassgeodata.gmu.edu/CropScape/ (accessed August 2012). USDA-NASS, Washington, DC.

USDA National Agricultural Statistics Service Quick Stats. 2008-2012b. Published crop-specific statistics [Internet]. [cited 2014-03-25]. Available at http://www.nass.usda.gov/Quick_Stats/. USDA-NASS, Washington, DC.

National Academy of Sciences , Committee on Ecological Risk Assessment under FIFRA and ESA, Board on Environmental Studies and Toxicology, Division on Earth and Life Studies, National Research Council of the National Academies. 2013. Assessing Risks to Endangered and Threatened Species from Pesticides, The National Academies Press, Washington, D.C., www.nap.edu.