

## Health & Ecological Risk Assessment

# A Probabilistic Co-Occurrence Approach for Estimating Likelihood of Spatial Overlap Between Listed Species Distribution and Pesticide Use Patterns

Leif Richardson,\* † JiSu Bang, ‡ Katherine Budreski, † Jonnie Dunne, † Michael Winchell, † Richard A Brain, † and Max Feken ‡

†Stone Environmental, Inc, Montpelier, Vermont, USA

‡Syngenta Crop Protection, LLC, Greensboro, North Carolina, USA

### ABSTRACT

Characterizing potential spatial overlap between federally threatened and endangered (“listed”) species distributions and registered pesticide use patterns is important for accurate risk assessment of threatened and endangered species. Because accurate range information for such rare species is often limited and agricultural pesticide use patterns are dynamic, simple spatial co-occurrence methods may overestimate or underestimate overlap and result in decisions that benefit neither listed species nor the regulatory process. Here, we demonstrate a new method of co-occurrence analysis that employs probability theory to estimate spatial distribution of rare species populations and areas of pesticide use to determine the likelihood of potential exposure. Specifically, we 1) describe a probabilistic method to estimate pesticide use based on crop production patterns; 2) construct species distribution models for 2 listed insect species whose ranges were previously incompletely described, the rusty-patched bumble bee (*Bombus affinis*) and the Poweshiek skipperling (*Oarisma poweshiek*); and 3) develop a probabilistic co-occurrence methodology and assessment framework. Using the principles of the Bayes’ theorem, we constructed probabilistic spatial models of pesticide use areas by integrating information from land-cover spatial data, agriculture statistics, and remote-sensing data. We used maximum entropy methods to build species distribution models for 2 listed insects based on species collection and observation records and predictor variables relevant to the species’ biogeography and natural history. We further developed novel methods for refinement of these models at spatial scales relevant to US Fish and Wildlife Service (FWS) regulatory priorities (e.g., critical habitat areas). Integrating both probabilistic assessments and focusing on USFWS priority management areas, we demonstrate that spatial overlap (i.e., potential for exposure) is not deterministic but instead a function of both species distribution and land use patterns. Our work serves as a framework to enhance the accuracy and efficiency of threatened and endangered species assessments using a data-driven likelihood analysis of species co-occurrence. *Integr Environ Assess Manag* 2019;15:936–947. © 2019 SETAC

**Keywords:** Probabilistic risk assessment Co-occurrence analysis Endangered species risk assessment Pesticide risk assessment *Bombus affinis*

### INTRODUCTION

Section 7 of the US Endangered Species Act (ESA) requires federal agencies to submit a biological evaluation (BE) if their actions may affect species protected by the ESA (US Fish and Wildlife Service [USFWS] 2019). As the Agency charged with registering pesticide products under the Federal Insecticide, Fungicide, Rodenticide Act (FIFRA), the US Environmental Protection Agency (EPA) also assesses the potential risks to listed (threatened and endangered) species under the ESA. In 2013, the National Academy of

Sciences (National Research Council 2013) recommended using a 2-step BE process. First, the Agency makes a “no effect” or “may affect” determination for each listed species. The initial determination is based primarily on the degree to which there is spatial overlap between the pesticide use area and distribution of listed species, represented by species range and critical habitat maps. For example, in recent BEs issued by the USEPA, co-occurrence of species habitat and pesticide use area >0% (USEPA 2016a, 2016b) and ≥1% (USEPA 2016c) have both triggered “may affect” determinations. The second step is to refine a “may affect” determination to “may affect, not likely to adversely affect” or “may affect, likely to adversely affect.” These determinations are based on toxicity and exposure information using a weight-of-evidence approach. Either informal or

This article contains online-only Supplemental Data.

\* Address correspondence to leif@stone-env.com

Published 16 July 2019 on [wileyonlinelibrary.com/journal/ieam](http://wileyonlinelibrary.com/journal/ieam).

formal consultation is required with the USFWS and National Marine Fisheries Services (NMFS) in these cases (USFWS 2018d).

Geospatial datasets are used to both define pesticide use areas and species' ranges. The recent interim approach used by the USEPA and the USFWS and NMFS to establish the use area for registered pesticides incorporated multiple years of land-cover data which accounts for crop rotation and classification uncertainty, resulting in a multiyear composite crop footprint (USEPA 2016a, 2016b, 2016c). This method is deterministic, concluding that a given use pattern is present if it is detected in any of the multiple years of data. Similarly, a deterministic representation of species ranges is used in the co-occurrence analysis. In both instances, the input datasets are treated as binary presence or absence data, rather than as continuous variables with probability distributions.

Accuracy of deterministic co-occurrence analyses are affected by underlying errors and irregularities in datasets used to estimate crop footprints. For example, the US Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) (USDA NASS 2017a) features more than 110 land-cover classes, the accuracy of which varies among states, crop cover types, and years of the data product. For example, the accuracy for corn (CDL Class 1) in Alabama ranges approximately 86% to 94% across available data years (i.e., 2008–2017). Moreover, accuracy for the same class ranges about 70% to 98% across different states in the most recent data year. For each cover class, such errors may be related to both omission and commission. Errors of omission are those in which the cover type occurs but is misidentified as another class (e.g., when an area of corn field is classified as forest); errors of commission are those in which other cover types are misidentified as the type in question (e.g., when a forested area is erroneously typed as corn) (Campbell and Wynne 2011; USDA NASS 2017a).

Standard co-occurrence analysis may also be affected by limitations of deterministic species distribution data. Range maps are often obtained from aggregated occurrence records and derivative polygon data—for example, the FIFRA Endangered Species Task Force datasets (Brain et al. 2015) and designated critical habitat of listed species (USFWS 2018a). These products highlight areas of recent observation but may underestimate species distribution extents if they do not consider additional high-quality habitat where inventory has not been conducted. By contrast, they often feature polygons arbitrarily generalized to greater ecosystem or political boundaries, which may overestimate extent of occurrence. These shortcomings may diminish regulators' abilities to protect listed species (Flather et al. 1998); they could also lead to overly conservative management decisions that impact land use with no benefit to target species. These issues are of particular concern for listed species whose distribution and ecology are poorly known, or for whom detection is challenging, for example, terrestrial macroinvertebrates.

Several improvements to deterministic co-occurrence modeling have been proposed. In addition to the uncertainty

of crop rotation considered by the EPA within the first step of the BE process, uncertainties associated with misclassification errors within the source data can be addressed in co-occurrence analyses. Budreski et al. (2016) used a Bayesian approach to develop a probabilistic crop footprint to account for misclassification errors within best-available land-cover datasets. The methodology incorporated 5 y of the CDL dataset, which was used to determine the prior probability of crop presence, along with the National Land Cover Database (NLCD), which was incorporated to determine the posterior probability of crop presence. The method assumes that the percentage of actual area for crop of interest that maps to each NLCD class would be approximately the same as the percentage of cultivated crop (NLCD Class 82) reference area that maps to other NLCD classes. For example, if a CDL map pixel is classified as soybean (or any other crop) and the NLCD classification is open water, we assumed the probability that the pixel is correctly classified given the NLCD class is equal to the probability that open water will be misclassified as cultivated crop. A final step was a probability correction using the USDA NASS reported crop acreages by state and year. Since the development of these probabilistic approaches, high-resolution imagery and rapid remote-sensing processing have become more readily accessible (e.g., the European Space Agency's Sentinel-2 satellite imagery at 10-m resolution with a 5-day global revisit period became freely accessible in 2015). Research has shown that object-based image analysis (OBIA), which involves grouping similar adjacent pixels together as image objects representative of geographic features of interest, is an effective tool for delineating field boundaries and other image-based objects (Ma et al. 2017). Using high-resolution imagery and OBIA, uncertainties associated with omission errors, commission errors, as well as spatial resolution, can be further addressed in crop footprint development.

Species distribution modeling (SDM) provides an opportunity to move beyond deterministic geographical assessments of listed species' ranges (Hernandez et al. 2006). These statistical tools combine species presence and absence data with information on climate, geography, and ecology to make predictions about habitat suitability and species occurrence. These models have proven their utility for identifying and managing rare species populations (Richardson et al. 2019). They have advantages versus the use of static observation-based data in that they may identify areas of suitable habitat for which survey data are lacking. Such modeling approaches are an improvement compared with conservative estimates of distribution that generalize observations to encompassing political jurisdictions, such as county boundaries; however, their predictions can, in certain cases, also result in net increases in estimated species ranges, for example, expanding into suitable habitat previously not surveyed or occupied. Maxent, a leading SDM software platform, relies on presence-only species data and employs a probabilistic algorithm to predict species distributions (Phillips et al. 2017). Because Maxent can make accurate predictions

based on relatively small sets of occurrence data, such as those often encountered with rare species (Hernandez et al. 2006), it is well-suited to the modeling of listed species distributions.

Taken together, probabilistic approaches to use area delimitation and SDM have the potential to make major improvements to the regulatory process for pesticides for listed species. The purpose of this research is to demonstrate a method to combine parallel-use pattern and SDMs in a probabilistic framework that allows regulators an enhanced assessment of the likelihood of proximal overlap. We demonstrate these methods for a hypothetical soybean insecticide and 2 insect species currently protected by the endangered species act, the Poweshiek skipperling (*Oarisma poweshiek*) and the rusty-patched bumble bee (RPBB [*Bombus affinis*]), 2 species representative of the apparent ongoing decline in arthropod pollinators (Rhodes 2018). Our results are compared with those of standard deterministic assessments.

The Poweshiek skipperling historically occurred in mid-western tallgrass prairie and prairie fen habitats. Adult butterflies subsist on nectar from a variety of plants (Semmler 2010), but larvae feed on only a few species of host plants, especially matted muhly grass (*Muhlenbergia richardsonis*) and prairie dropseed (*Sporobolus heterolepis*) (Dupont 2011). Precipitous decline of Poweshiek skipperling populations was observed in the last decade of the 20<sup>th</sup> century (Selby 2005; Dupont 2011), and the species was listed as endangered by the USFWS in 2014 (USFWS 2018b). The Service designated approximately 25 900 acres of critical habitat in 56 units for the species in 2015 (USFWS 2015). The RPBB was historically one of the most common bumble bee species found in eastern North America, where it consumed nectar and pollen from a diverse array of host plants in a wide variety of habitats, including forests, agricultural lands, and areas of low-intensity development (Medler and Carney 1963; MacFarlane 1974; Williams et al. 2014; Richardson et al. 2019). The bee abruptly disappeared from nearly all of its former range in the mid to late 1990s, and was listed as endangered in 2017 (Colla and Packer 2008; USFWS 2017). A critical habitat designation for the RPBB has not been made, but the USFWS identifies 191 high potential zones (HPZ) centered on recent observations where actions by federal agencies require review under the assumption that the bee is highly likely to occur (USFWS 2017). For both insects, the causes of their declines remain somewhat unclear, and multiple, interacting threats, including pathogens, habitat loss, and pesticide exposure, are thought to have played a role.

## METHODS

### Crop footprint development

In this study, we model crop footprint of soybean, planted annually on approximately 84 million acres of US farm fields (USDA NASS 2017a), using 3 different methods: 1) a standard deterministic crop footprint, consistent with methods

currently used by EPA within the first step of the BE process (USEPA 2016a, 2016b, 2016c); 2) a probabilistic crop footprint method incorporating approaches developed by Budreski et al. (2016); and 3) modification of the probabilistic model with image objects by segmentation analysis of high-resolution sensing imagery to identify fields and other image objects.

**Deterministic crop footprint.** We used the most recent 5 y (2013–2017) of CDL data to create a 30-m<sup>2</sup> resolution soybean crop footprint. We included all cover classes that featured soybean agriculture: soybeans (class 5), double crop; soybeans/cotton (class 239), double crop; soybeans/oats (class 240), double crop; corn/soybeans (class 241), double crop; barley/soybeans (class 254). If a pixel was classified as 1 of these types in any of the years, it was given a value of 1; otherwise, the value was set to 0. This composite crop footprint thus captures all areas classified as soybeans by CDL in any 1 of 5 y (2013–2017).

**Probabilistic crop footprint.** Following methods outlined by Budreski et al. (2016), we generated posterior probabilities of crop presence based on information from the CDL dataset for 5 y (2013–2017), the NLCD (2011), and the NASS Quick Stats (2013–2017). This method assumes that the probability of a pixel accurately representing cover as soybean is higher, lower, or unchanged based on the NLCD class it coincides with. For example, CDL-based soybean pixels that coincide with NLCD-based deciduous forest pixels have a lower probability of actually being soybean on the ground than CDL-based soybean pixels that coincide with NLCD-based cultivated crop pixels.

A prior probability of crop presence was based on the CDL accuracy assessment information by state, year, and crop (USDA NASS 2017b). The prior probability was adjusted based on accuracy assessment information from the 2011 NLCD dataset (Wickham et al. 2017) by land-cover class using the principles of Bayes' theorem, which is used to update probabilities by incorporating additional evidence (Ott and Longnecker 2015). Lastly, crop probabilities were scaled at the state level by comparing against NASS surveys of reported planted acres by crop (USDA NASS 2017b).

Yearly probabilistic crop footprints were developed by first extracting crop classes for the crop of interest by year from 5 y of CDL data. All CDL crop classes for soybean described above were extracted to represent the initial crop footprint. Next, the CDL-based soybean data were overlaid with 2011 NLCD data. The CDL soybean class (i.e., 5 for soybean) and NLCD class (i.e., 82 for cultivated crops) were preserved for each pixel, allowing for the Bayes' calculation on a pixel basis.

The pixel-level probability by year was based on the reported accuracy for individual CDL soybean classes and NLCD classes, and is described in detail by Budreski et al. (2016). Briefly, following Bayes' theorem, the relationship between prior and posterior probability was defined as:

$$p(\text{soy}|\text{nlcd}) = \frac{p(\text{soy}) \cdot p(\text{nlcd}|\text{soy})}{p(\text{soy}) \cdot p(\text{nlcd}|\text{soy}) + p(\text{soy}') \cdot p(\text{nlcd}|\text{soy}')}$$

where  $p(\text{soy}|\text{nlcd})$  = posterior probability that a pixel is soybean;  $p(\text{soy})$  = probability a given soybean pixel in the class was correctly classified;  $p(\text{soy}')$  = probability a soybean pixel in the class was a false;  $p(\text{nlcd}|\text{soy})$  = conditional probability of mapping a pixel to any other NLCD class, given that the pixel is actually correctly identified as soybean; and  $p(\text{nlcd}|\text{soy}')$  = conditional probability of mapping a pixel to any other NLCD class, given that the pixel is incorrectly identified as soybean. The  $p(\text{nlcd}|\text{soy})$  (Supplemental Data Table S1) and  $p(\text{nlcd}|\text{soy}')$  (Supplemental Data Table S2) are derived from the published accuracy assessment information from the NLCD database (Wickham et al. 2017). When CDL user's accuracy and CDL commission error values were not reported in the published metadata (i.e., 0.001% of pixels), we applied a probability of 0.5 to  $p(\text{soy})$  and  $p(\text{soy}')$ .

Following initial posterior probability calculations, probabilities were adjusted to meet statewide reported crop acreages. Posterior probability-weighted statewide acreages of soybean were calculated and compared with NASS-reported soybean acreages by year (USDA NASS 2017b). Where posterior probability-weighted acreages were within the 90% confidence interval of NASS-reported soybean acreages by year, pixel-level posterior probabilities were unchanged. Where posterior probability-weighted acreages, by state, were above the 90% confidence interval of NASS-reported soybean acreages, pixel-level probabilities were adjusted down based on the ratio of NASS-reported state acreage to posterior probability-adjusted CDL acreage for the year and state. Where posterior probability-weighted acreages were below the 90% confidence interval of NASS-reported soybean acreages, nonsoybean pixels for the target year that are classified as NLCD class 82 pixels and soybean in at least 1 other year (2013–2017) were extracted. The posterior probability for these extracted pixels was adjusted from 0 to a probability calculated as Budreski et al. (2016):

$$\frac{([\text{NASS State acreage}] - [\text{posterior probability} - \text{adjusted CDL acreage}])}{([\text{NLCD Class 82 nonsoybean} - \text{acreage}])}$$

Finally, to account for crop rotation and variability in land use with time, adjusted posterior probability pixels were averaged across all target years (2013–2017). There are limited cases where pixels that were classified as soybean in at least 1 y have an average adjusted posterior probability of zero, for example, when  $p(\text{nlcd}|\text{soy})$  is zero (Supplemental Data Table S1). As a conservative measure, all pixels that were originally classified as soybean in any single year based on the CDL data and that have an average adjusted posterior probability of zero were assigned a probability of 0.0001. This final adjustment ensures that any pixels originally classified as soybean are included as potential pesticide-use sites in subsequent analysis.

*Probabilistic crop footprint with image objects.* As a case study, we sought to further improve the accuracy of the probabilistic crop footprint by summarizing 10-m resolution pixel-based probabilities within discrete delineated fields for 2 habitat zones in Illinois and Wisconsin where RPBB populations persist. We did so by downscaling the crop footprint model to 10-m resolution, then segmenting Sentinel-2 satellite imagery into image objects and comparing them to a subsample of manually delineated reference fields. For each year of the study for which Sentinel-2 imagery is available (2015–2017), we selected a single image featuring low percentage cloud cover and high contrast between apparent fields. The selected imagery was then processed by Farm-Shots.com into Soil-Adjusted Vegetation Index (SAVI) images (Huete 1988). Images were composited into a single-raster dataset for each study area and segmented using the segment mean shift tool in ArcGIS with multiple combinations of the spectral detail, spatial detail, and minimum-size parameters (ESRI 2017).

The reference field borders were derived by generating randomly placed points within the study area and manually delineating any crop fields where those points were generated, using the Sentinel-2 SAVI images as reference. The agreement between the reference and modeled boundaries was evaluated by comparing the degree of segmentation in the modeled boundaries to the reference fields, with reference crop fields being classified as “over-segmented” (i.e., uniform fields having been segmented into multiple polygons), “under-segmented” (i.e., multiple distinct fields having been combined as 1 polygon), or “match” (i.e., where results of segmentation closely mirror our manually delineated fields) (Figure 1). Validation of the field borders shows that 71% of delineated field borders either matches the reference fields or matches them and maps additional field subobjects as well.

The final segmentation results in image objects that represent fields, subfields, and other continuous land-use areas. While matching segmentation is desired, a final segmentation that reduced the amount of undersegmentation was selected. Assuming that the land use is consistent across image objects, and pixels within them have the same likelihood of depicting soybean agriculture, probabilistic predictions of soybean occurrence were averaged across all such pixels within image objects for the 2 selected RPBB habitat areas, which are described in the following section.

#### *Species distribution modeling*

We used Maxent software (Phillips et al. 2017) to construct probabilistic SDMs for each species. We then conducted a refined assessment of occurrence probability for the RPBB for 2 HPZs, 1 near Peoria, Illinois, and the other centered on Madison, Wisconsin. We chose to focus on these HPZs due to their inclusion of a range of land-cover types, including soybean agriculture. We focused our analysis of Poweshiek skipperling distribution on a large area of current habitat in North Dakota and South Dakota.

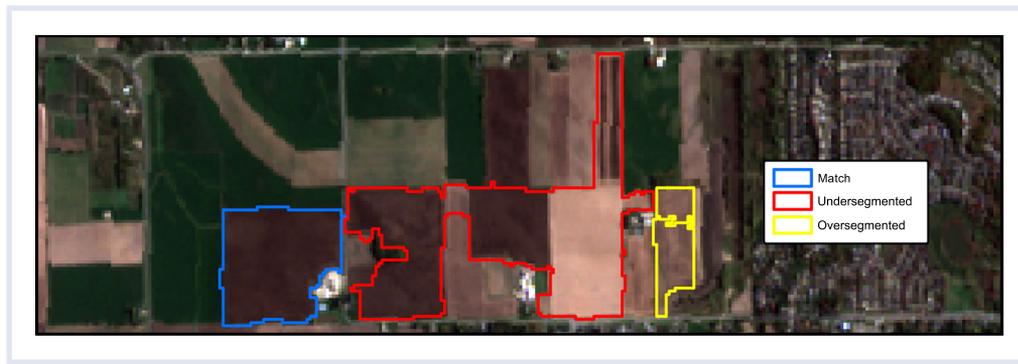


Figure 1. Example of image segmentation in which validation segment has been classified as match (blue), undersegmented (red), and oversegmented (yellow) in the validation process.

*Independent (predictor) variables.* We downloaded 19 global, gridded climate models (the Bioclim bioclimatic variables) and solar radiation grids for June to September from Worldclim at 30-arc-second resolution (i.e., pixel size  $\sim 1 \text{ km}^2$  at the equator). We obtained remote-sensing data products on vegetation and land surface attributes of potential importance to insects from the Copernicus Global Land Service, including dry matter productivity, fraction of green vegetation cover, leaf area index, normalized difference vegetation index (a measure of greenness of vegetation), and soil water index. From the US Geological Survey Land Cover Institute, we obtained digital elevation model and global land-cover grids (Supplemental Data Table S3).

As predictor variables, we also created SDMs for habitats and host plants of particular importance to the Poweshiek skipperling. We downloaded 7,093 occurrence records for its 2 primary larval host plants (*M. richardsonis* and *S. heterolepis*) (Selby 2005; Dupont 2011; Pointon 2015; USFWS 2018b) from the Global Biodiversity Information Facility (GBIF 2018a, 2018b), Integrated Digitized Biocollections (iDigBio 2018a, 2018b), and iNaturalist (iNaturalist 2018a, 2018b). After processing to remove records that had obvious data quality problems (e.g., erroneous location information) or lacked georeferencing, clipping to the Midwest US region where the butterfly occurs and filtering to retain only 1 record per collecting event, our dataset contained 33 *M. richardsonis* and 324 *S. heterolepis* records. To construct distribution models for the midwestern tallgrass prairie and prairie wetland habitats to which this species is restricted (Dupont 2011; USFWS 2018b), we downloaded 1720 iNaturalist records of observations of wild organisms for which the observer had recorded the words “tallgrass prairie” and “prairie fen,” respectively, in the description field of the record. After filtering as described above, our dataset contained 1160 tallgrass prairie and 50 prairie fen records. Distribution models for these host plants and habitats were constructed using Maxent software and roughly following the methods described below for the insect species that are the focus of this study.

Using ArcGIS 10.5.1, we clipped all predictor layers to eastern North America, which we defined as the states and provinces east of or crossed by the 100<sup>th</sup> meridian. All layers

were set to the WGS84 geographic projection and re-projected to 30 arc-second grid cell size.

*Dependent variables.* We used a subset of recent occurrence records from the Bumble Bees of North America database (Williams et al. 2014; Richardson 2018) to construct an SDM for the RPBB. Our initial dataset spans the period 1998 to 2018 (we excluded older records because the species’ range has shifted so dramatically) and was comprised of 1372 records. We used R package dplyr (Wickham et al. 2018) to filter this data, retaining only 1 record per collecting event (unique combination of observer or collector name, date, and geographic information). To reduce spatial bias in the distribution of RPBB observations (e.g., representing bias in areas where collectors work) we used the R package spThin (Aiello-Lammens et al. 2015) to further reduce the dataset so that all points were separated by  $\geq 10 \text{ km}$ . Our final dataset contained only 238 records of the original 1372.

We downloaded Poweshiek skipperling occurrence records from GBIF ( $n = 726$ ), iDigBio ( $n = 212$ ), Symbiota Collection of Arthropods Network (Gries et al. 2014) ( $n = 574$ ), and Biodiversity Information Serving Our Nation (BISON 2018) ( $n = 633$ ). We also digitized and georeferenced records from the literature ( $n = 242$ ) (Borkin 1995; Schlicht 1997; Swengel and Swengel 1997, 1999, 2014; Selby 2008; Schlicht et al. 2009; Semmler 2010; Swengel et al. 2010; Dupont 2011; Calhoun 2013; Slaughter et al. 2013; Pointon 2015).

*Model selection.* We modeled distribution of the RPBB and Poweshiek skipperling using Maxent software and following a set of best practice methods to select the best-fit model that uses the fewest predictor variables and minimizes correlation among these predictors (Yost et al. 2008; Araújo et al. 2019). For each species, we first constructed models containing all bioclimatic, geophysical, and land-surface predictor layers, plus the ecological variables described above. We used averaging of 5 models specifying replication via the cross-validation method and applied the cloglog transform to the output (Phillips et al. 2017). We then used the jackknife test of variable importance to

evaluate each predictor variable's contribution to the model, removing at each step the variable with lowest contribution to model performance, as well as any others with  $\leq 1\%$  contribution to performance. Employing expert biological opinion and following the principal of parsimony, we selected final models as those with the smallest number of predictor variables that were similar to full models in average training gain, and whose predictor variables were not strongly correlated (Pearson correlation coefficient  $\geq 0.8$ ). We created 10 repetitions of final models and used the average prediction as an estimate of likelihood of species occurrence in our co-occurrence analyses. As relative measures of model fit, we report the regularized training gain and area under the curve (AUC) statistics. In our co-occurrence analysis we used the full range of values in the species distribution predictions (i.e., probability 0–1); however, we note that Maxent outputs are often truncated at an arbitrary threshold below which the predictions are not considered meaningful. In estimating areal extent of our SDMs, we applied a threshold where pixels with predictions  $< 0.25$  were not considered.

*Species distribution model refinements.* Because range-wide SDMs may not capture fine-grained spatial structure to habitat suitability, we refined the prediction developing 10-m resolution models to better reflect the value of certain land-cover attributes to the RPBB, especially those related to edge habitats between adjacent cover types (Szymanski et al. 2016). For 2 HPZs in Illinois and Wisconsin we generated a high-resolution (10-m ground sample distance) land use/land cover (LULC) map using Sentinel-2 imagery taken during 6 dates in 2018. Mapping LULC consisted of segmenting imagery into self-similar “image objects” and selecting ancillary imagery for use in machine learning to detect statistical thresholds derived from the image objects that are useful in distinguishing LULC types. The image objects were derived from imagery acquired from 6 dates and processed into normalized difference water index, SAVI, and average brightness layers (McFeeters 1996) (Supplemental Data Table S4). These 3 were then composited and segmented using the segment mean shift tool in ArcGIS. Training samples to be used to classify the image objects were selected by randomly placing points within each study area and selecting proximate exemplary pixels that best matched the target categories (water, tree, crop, low vegetation, developed low, developed high). The training pixels were assigned values from the full stack of images collected for each HPZ. These values were then used to train a support vector machine classifier in ArcGIS, which builds a multidimensional spatial model based on the full stack of images and generates envelopes of the classes based on the training values. Image objects are then assigned LULC classes based on their similarity of spectral signatures to the multidimensional envelopes. We assessed the accuracy of the land cover classification by placing 400 points randomly within each study area and generating a confusion matrix between the map and the manually assigned class. Results

show that the overall accuracies for the Illinois and Wisconsin HPZs are 84% and 74%, respectively. We also attempted to classify woody and herbaceous wetlands, but due to the difficulty of sensing soil moisture under thick vegetative cover and assessing the accuracy of the classification from imagery interpretation alone, we instead mapped these features using National Wetland Inventory data (USFWS 2018c).

We then refined the SDM by downscaling to 10-m resolution and adjusting the value of the pixels within the HPZs based on the composition of LULC types and the density of borders between LULC types within each pixel. We first summarized LULC composition and border type density in 500-m neighborhoods (a distance from the nest at which most bumble bee foraging takes place) (Greenleaf et al. 2007); then we weighted the areas and distances based on the literature and our expert opinion (MacFarlane 1974; Richards 1978; Svensson et al. 2000; Kells and Goulson 2003), producing habitat value adjustment scores (Supplemental Data Table S5). For example, edge habitats we estimate to be of lesser value to RPBB, such as those where grasslands abut areas of intensive development resulted in adjustment of our predictions by +2%, whereas those known to be a key habitat component for this species, such as grassland-forest ecotones, resulted in pixel adjustment by +10% (Supplemental Data Table S5). The habitat value adjustment scores were limited to increasing or decreasing the values of the SDM by 1% for LULC cover classes and increasing it by up to 10% for border habitat density. In the Wisconsin HPZ this resulted in some pixels with probabilities  $> 1$ , so we normalized the SDM by dividing each pixel by the maximum probability.

#### *Co-occurrence analysis*

We completed co-occurrence analyses based on both deterministic and probabilistic methods. To do this, we down-scaled final models of crop footprint and species distribution to the lowest raster resolution used in either. We then calculated the product of the crop footprint and species distribution layers to determine overall probability of co-occurrence. For the deterministic approaches, the probability of pesticide use and species locations is based on presence (probability of 1) and absence (probability of 0), whereas the probabilistic methods result in probabilities ranging from 0 to 1. For the probabilistic approach, we also evaluated the distribution of crop footprint probability across species distribution probability. Because we did not calculate uncertainty parameters around our probabilistic estimates, this method resulted in single co-occurrence predictions for each pixel. To estimate the total area where listed species co-occur with soybean agriculture, we summed the product of co-occurrence probability and pixel area.

## RESULTS

### *Crop footprint comparison*

The average annual area of planted soybean in the United States from 2013 to 2017 was  $84\,181\,400 \pm 1\,786\,126$  acres

(USDA NASS 2017b). Using deterministic methods, we estimate soybean crop footprint is more than twice as large (184 292 493 acres) as the average amount of NASS-reported soybean acreage grown in the last 5 y. The weighted probabilistic crop footprint results in a total soybean acreage of 82 599 900 acres, 98.1% of the NASS statistic.

### Species distribution modeling

The full model for the RPBB had a regularized training gain of  $2.576 \pm 0.032$  (95% CI) and an AUC of  $0.967 \pm 0.007$  (Supplemental Data Figure S3 and Table S6). After 7 modeling iterations, the species' distribution was best described by a 6-variable final model, including soil type, soil moisture, temperature and precipitation in summer months, September solar radiation, and LULC (training gain =  $2.239 \pm 0.031$ ; AUC =  $0.959 \pm 0.014$ ; Table 1). LULC categories with the strongest predictive power (whether negative or positive) included those associated with developed areas, croplands, broadleaf forests, and water bodies. We predict RPBB occurrence in 56.6% (i.e., 5824 acres where mean prediction  $\geq 0.25$ ) and 46.6% (41 814 acres) of the area of our target HPZs in Illinois and Wisconsin, respectively (Figure 2).

The full model for distribution of the Poweshiek skipperling had a regularized training gain of  $3.554 \pm 0.040$  (95% CI) and an AUC of  $0.987 \pm 0.005$  (Supplemental Data Figure S3 and Table S6). The best-fit model was arrived at after 7 iterations included the predictor variables for both host plants and habitat types, as well as soil type, LULC,

temperature of the warmest quarter, precipitation of the wettest quarter, and vegetation greenness (training gain =  $3.267 \pm 0.053$ ; AUC =  $0.985 \pm 0.011$ ; Table 1).

### Comparison of co-occurrence results

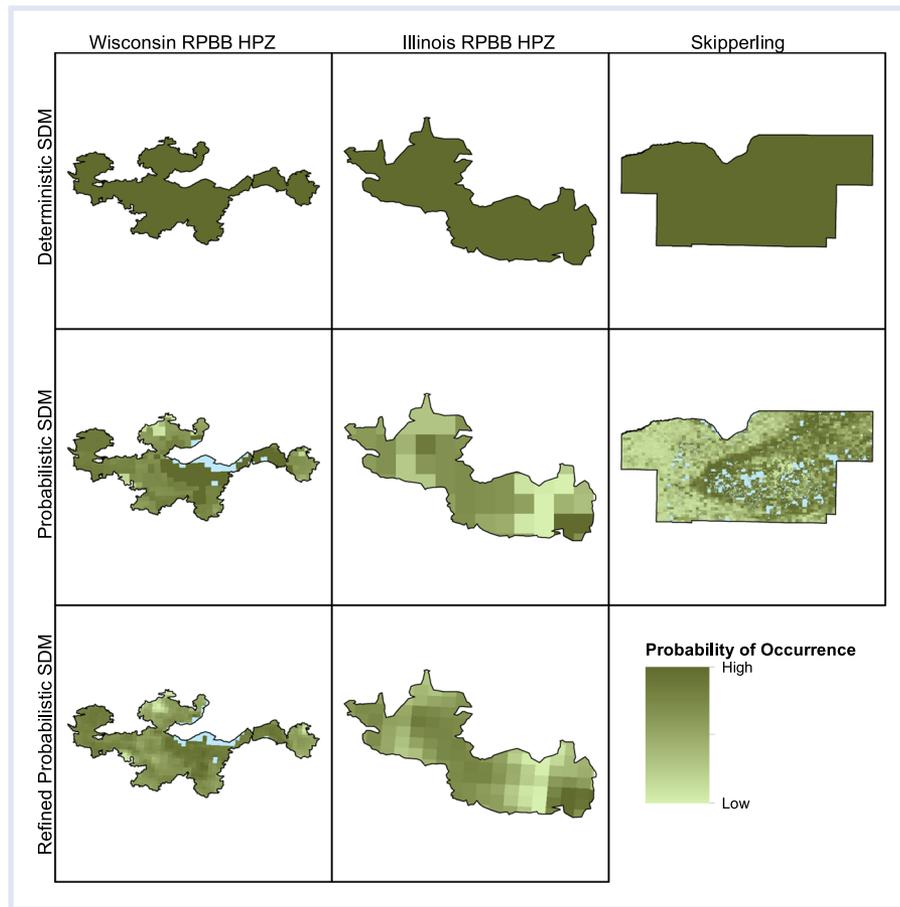
We evaluated species–pesticide use co-occurrence in a step-wise approach based on refinements to both constituent datasets (Table 2). The deterministic crop footprint and species ranges resulted in the highest co-occurrence across all species locations evaluated. For the Poweshiek skipperling, the deterministic approach resulted in 61.07% co-occurrence across the entire USFWS species range (5 311 447 acres). The co-occurrence reduced by more than half by incorporating the probabilistic crop footprint (27.70%; 2 409 154 acres) and further reduced to 9.07% (786 796 acres) by also incorporating the probabilistic SDM.

Similar to the Poweshiek skipperling, co-occurrence was highest using the fully deterministic approach for both the Illinois RPBB HPZ (23.54%; 1385 acres) and the Wisconsin RPBB HPZ (5.32%; 2559 acres) (Table 2). Co-occurrence results for the Wisconsin RPBB HPZ reduced incrementally with each refinement resulting in a co-occurrence of 1.25% (600 acres). Co-occurrence results for the Illinois RPBB HPZ followed a similar trend, except the final refinement incorporating both the probabilistic crop footprint with image objects and the adjusted probabilistic SDM. In this case, adjusting the SDM with land cover refinements resulted in a 5.73% co-occurrence (337 acres), an increase compared

**Table 1.** Biogeographic and ecological variables explaining best-fit species distribution models of the rusty-patched bumble bee and Poweshiek skipperling

Species	Variable	Percent contribution	Permutation importance
Rusty-patched bumble bee	Soil type	27.3	3.3
	Maximum temperature, warmest month (BIO5)	23.7	29.4
	Solar radiation (September)	20.0	42.5
	Precipitation, warmest quarter (BIO18)	11.4	19.4
	Soil water index	8.9	4.1
	Land use/land cover	8.6	1.3
Poweshiek skipperling	Prairie fen distribution	33.1	21.8
	<i>Sporobolus heterolepus</i> distribution	29.2	40.0
	<i>Muhlenbergia richardsonis</i> distribution	26.1	3.5
	Soil type	2.7	1.9
	Mean temperature, warmest quarter (BIO10)	2.1	3.3
	Tallgrass prairie distribution	1.9	4.2
	Land use/land cover	1.8	0.1
	Precipitation, warmest quarter (BIO16)	1.7	23.3
	Normalized Difference Vegetation Index	1.4	1.9

Note: Percent contribution is the relative increase in training gain of the model contributed by each variable. Permutation importance is a measure of model performance when each variable is randomly permuted on occurrence points (both presence points and background points) used to train the model. See Supplemental Data Table S3 for information on variables.



**Figure 2.** Species distribution models for rusty-patched bumble bee in 2 high potential zones in Illinois and Wisconsin and for the Poweshiek skipperling in a large area of current habitat in North Dakota and South Dakota. Maps show deterministic and probabilistic model outputs for each species, and refined probabilistic results for the RPBB. Scale: Wisconsin RPBB = 1:910 000; Illinois RPBB = 1:250 000; skipperling = 1:5 900 000. Blue = water bodies. SDM = species distribution model; RPBB = rusty-patched bumble bee; HPZ = high potential zone. Deterministic models were developed by US Fish and Wildlife Service.

with the probabilistic crop footprint and probabilistic SDM method (5.03% co-occurrence) and the probabilistic crop footprint with image objects and probabilistic SDM method (4.99% co-occurrence). Most of the area within the RPBB HPZs evaluated has a combined probability of less than 0.25 using the fully refined probabilistic method (Figure 3).

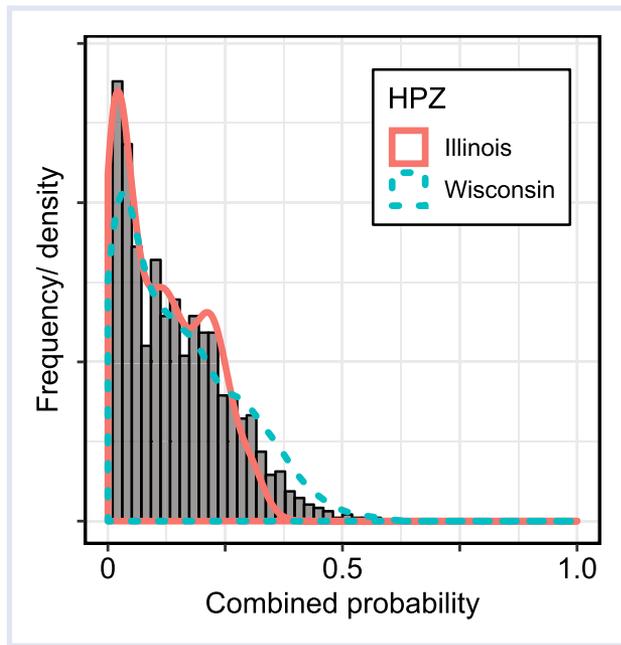
The distribution of co-occurrence was evaluated across soybean crop footprint probability and species presence

probability using the fully refined methodology within the RPBB HPZs (Figure 4). RPBB was significantly more likely to occur in the Wisconsin HPZ than the Illinois HPZ. The Illinois RPBB HPZ had a greater proportion of area with a soybean crop footprint probability >0.5 (7.2%) than the Wisconsin RPBB HPZ (0.1%). Using these distributions, the total area exceeding specific probability thresholds can be evaluated. For instance, 1.6% of the Illinois RPBB HPZ area and

**Table 2.** Percent co-occurrence between each of 2 endangered species and areas of soybean agriculture, using 5 combinations of deterministic and probabilistic modeling techniques

Species/location	Co-occurrence (%) by method				
	Deterministic crop footprint & deterministic species range	Probabilistic crop footprint & deterministic species range	Probabilistic crop footprint & probabilistic SDM	Probabilistic field footprint & SDM	Probabilistic field footprint & adjusted SDM
Poweshiek (entire range)	61.07	27.70	9.05	N/A	N/A
RPBB (IL HPZ)	23.54	11.13	5.03	4.99	5.73
RPBB (WI HPZ)	5.32	1.74	1.43	1.30	1.25

HPZ = high potential zone; IL = Illinois; RPBB = rusty-patched bumble bee; SDM = species distribution model; WI = Wisconsin.

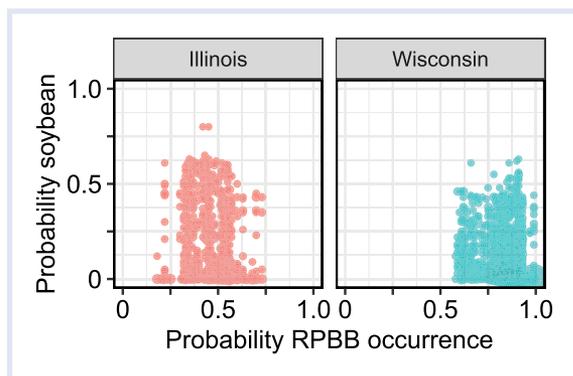


**Figure 3.** Kernel density estimates for probability of crop footprint and rusty-patched bumble bee co-occurrence in 2 US Fish and Wildlife Service high potential zones in Illinois and Wisconsin. Bars = histogram of data combining values for both HPZs. HPZ = high potential zone.

0.1% of the Wisconsin RPBB HPZ area have both soybean crop footprint and species presence probabilities >0.5.

## DISCUSSION

We have demonstrated that a probabilistic assessment of crop growing areas and species distributions can refine the co-occurrence analysis used in EPA Biological Evaluations and USFWS Biological Opinions to assess likelihood of co-occurrence. We report that both probability-weighted crop and species occurrence areas are substantially smaller than areas calculated using deterministic methods, and show consequent estimates of co-occurrence likelihood are lower on a per-area basis than estimates calculated using deterministic methods.



**Figure 4.** Co-occurrence distribution within selected rusty-patched bumble bee high potential zones using the probabilistic crop footprint and species distribution models. Values are scaled according to the cumulative area they represent in each of the 2 HPZs. HPZ = high potential zone; RPBB = rusty-patched bumble bee.

Using Bayesian approaches to detecting areas of soybean agriculture, we estimated the US soybean crop area at 98.1% of the area reported by annual USDA statistics, and within the error bounds of the reported area. By contrast, our deterministic assessment substantially overestimated soybean crop footprint relative to reported NASS statistics, a result that affects interpretation of co-occurrence assessments. This overestimate may be due to the fact that we accepted as soybean any of 5 cover types associated with that crop in any of 5 y. In reality, in the Midwest US, soybean is often grown in rotation within and between years with other crops (Karlen et al. 2006), reducing potential spatial and temporal overlap between a particular pesticide use and a listed species. The deterministic crop footprint may also exaggerate the extent of the use pattern by inclusion of frequent errors of commission and omission for soybean in the CDL (Budreski et al. 2016). Using a probabilistic approach and incorporating NLCD data as an additional line of evidence reduces such errors of commission, and averaging probabilities across image objects that are field or subfield areas reduces errors of omission (Supplemental Data Figures S1 and S2). Combining probabilistic crop footprint and species occurrence models, we conclude that co-occurrence probability is seldom above 0.5 (Figure 3) and is substantially lower within focal species areas than assessed by deterministic methods, at least for our example species (Table 1). It is interesting to note that our probabilistic co-occurrence is a product of 2 probability distributions (i.e., those for crop footprint and species distribution), but we have not explored consequences for risk assessment of relative magnitude of each parameter. For example, if probability of both crop and species occurrence = 0.5, then co-occurrence probability = 0.25, yet this same co-occurrence probability may also be obtained if 1 parameter = 0.9 and the other = 0.28. In evaluating probabilistic co-occurrence assessments, analysts should consider whether the underlying likelihood that a rare species occurs matters to the interpretation of combined probability. This may depend on biology of the species under review, or it could be influenced by confidence in either of the probabilistic assessments that underlie the co-occurrence analysis.

Our analysis is improved in accuracy compared with previous crop and species modeling efforts by use of novel methods. For crops, we extend the methodology of Budreski et al. (2016) by using segmentation of high-resolution remote-sensing imagery to identify boundaries and refine image objects such as fields. We also incorporate new methods to account for error and uncertainty in the land cover datasets on which these analyses depend.

For rare species, our models considered ecological parameters not typically investigated in risk assessments, including spatial distribution of food resources (i.e., host plants). Other studies have demonstrated the utility of incorporating species interactions in insect SDMs (Lemoine 2015), and our study suggests that co-occurrence assessments can be improved by better attention to such species life-history traits. We further refined distribution models for

the RPBB by incorporating information on spatial patterning of local land cover variables. Bumble bees are known to use diverse habitat types for nesting, forage, and hibernation habitats (Goulson et al. 2010; Richardson et al. 2019), and recent RPBB observations have been made in open woodlands and areas marked by human disturbance, including suburban parks, meadows, and gardens (USFWS 2018d). Available evidence suggests that the RPBB nests in crop fields, grasslands, parks, and areas of low-intensity development (Medler and Carney 1963; MacFarlane 1974). Bumble bees may be particularly attracted to ecotones or linear edge features where land use/land cover types meet (Richards 1978; Sjödin et al. 2008; Diaz-Forero et al. 2013), and there is evidence that edge types vary in their value to bumble bees (Svensson et al. 2000; Kells and Goulson 2003). Ecotones (e.g., field borders) are a fundamental component of landscapes dominated by agriculture, and our model refinements based on previous research and expert opinion improved co-occurrence predictions in these areas.

We highlight several caveats regarding interpretation of this study. First, the methods described here are more complex and data-intensive than those used in deterministic assessments. To address this, we recommend the development of standardized Bayesian tools for handling large spatial datasets and species occurrence data. Second, our SDMs were improved by incorporation of expert knowledge of the species' biology, but with the great diversity and number (>1500) of listed US species, it is not realistic to expect this will be possible in every co-occurrence analysis. Consequently, we recommend that the endangered species assessment is required to identify those aspects of species biology that, when considered, improve probabilistic co-occurrence analysis. Probabilistic co-occurrence modeling methods have clear utility in risk assessment, and we recommend their adoption to improve outcomes for the regulated community and rare species protection.

**Acknowledgment**—This research was supported by Syngenta Crop Protection LLC. We thank Farmshot.com for providing the satellite images for this study. We also thank R Branon Rodriguez for assistance with species occurrence data preparation.

**Disclaimer**—The authors declare no conflicts of interest. The peer-review process for this article was managed by the Editorial Board without the involvement of RA Brain.

**Data Accessibility**—All data and scripts generated in this analysis are available upon request from the corresponding author, Leif Richardson, at leif@stone-env.com.

## SUPPLEMENTAL DATA

**Figure S1.** Example of spurious pixels in Dane County, Wisconsin.

**Figure S2.** Example of omission errors in Dane County, Wisconsin.

**Figure S3.** Distribution of RPBB (A) and the Poweshiek skipperling (B).

**Table S1.** 2011 NLCD error matrix (Wickham et al. 2017) with assigned  $p(\text{nlcd}|\text{soy})$ , the conditional probability of mapping a pixel correctly classed as soybean to any NLCD class other than cultivated crops (which includes areas of soybean agriculture).

**Table S2.** 2011 NLCD error matrix (Wickham et al. 2017) with assigned  $p(\text{nlcd}|\text{soy})$ , the conditional probability of mapping a pixel to any NLCD class other than cultivated crops (to which soybean fields are assigned) when that pixel is incorrectly identified as soybean.

**Table S3.** Predictive layers used in modeling species distributions using Maxent software.

**Table S4.** Images used to train support vector machine learning algorithm for OBIA LULC mapping.

**Table S5.** Percent changes imposed on rusty-patched bumble bee species distribution model raster following segmentation and fine-grained land cover modeling using remote sensing data.

**Table S6.** Statistics for Maxent models of rusty-patched bumble bee and Poweshiek skipperling species distribution models.

## REFERENCES

- Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. 2015. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38(5):541–545.
- Araújo MB, Anderson RP, Barbosa AM, Beale CM, Dormann CF, Early R, Garcia RA, Guisan A, Maiorano L, Naimi B et al. 2019. Standards for distribution models in biodiversity assessments. *Science Advances* 5(1): eaat4858.
- [BISON] Biodiversity Serving Our Nation. 2018. BISON occurrence download: *Oarisma poweshiek*. Biodiversity Serving our Nation. [accessed 2018 Apr 30]. <https://bison.usgs.gov/#home>
- Borkin SS. 1995. 1994 Ecological studies of the Poweshiek skipperling (*Oarisma poweshiek*) in Wisconsin. Milwaukee (WI): Milwaukee Public Museum. Prepared for Wisconsin Department of Natural Resources.
- Brain RA, Teed RS, Bang J, Thorbek P, Perine J, Peranginangin N, Kim M, Valenti T, Chen W, Breton RL et al. 2015. Risk assessment considerations with regard to the potential impacts of pesticides on endangered species. *Integr Environ Assess Manag* 11(1):102–117.
- Budreski K, Winchell M, Padilla L, Bang J, Brain RA. 2016. A probabilistic approach for estimating the spatial extent of pesticide agricultural use sites and potential co-occurrence with listed species for use in ecological risk assessments. *Integr Environ Assess Manag* 12(2): 315–327.
- Calhoun JV. 2013. *Hesperia illinois* Dodge (Hesperiidae): An invalid neotype and the designation of a lectotype. *J Lepid Soc* 67(3):274–280.
- Campbell JB, Wynne RH. 2011. Introduction to remote sensing. New York (NY): Guilford Press. xxxii + 670 p.
- Colla S, Packer L. 2008. Evidence for decline in eastern North American bumblebees (Hymenoptera: Apidae), with special focus on *Bombus affinis* Cresson. *Biodivers Conserv* 17(6):1379–1391.
- Diaz-Forero I, Kuusemets V, Mänd M, Liivamägi A, Kaart T, Luig J. 2013. Influence of local and landscape factors on bumblebees in semi-natural meadows: A multiple-scale study in a forested landscape. *J Insect Conserv* 17(1):113–125.
- Dupont J. 2011. Minutes from the Poweshiek skipperling workshop. In: The Nature Conservancy of Canada, Manitoba Region, Winnipeg, Canada, 2011 March 24–25.
- [ESRI] Environmental Systems Research Institute. 2017. ArcGIS Desktop. Redlands (CA).
- Flatler CH, Knowles MS, Kendall IA. 1998. Threatened and endangered species geography. *BioScience* 48(5):365–376.

- [GBIF] Global Biodiversity Information Facility. 2018a. GBIF occurrence download: *Muhlenbergia richardsonis*. [accessed 2018 Sep 13]. <https://www.gbif.org/occurrence/download/0011630-180824113759888>
- [GBIF] Global Biodiversity Information Facility. 2018b. GBIF occurrence download: *Sporobolus heterolepis*. [accessed 2018 Sep 13]. <https://www.gbif.org/occurrence/download/0011631-180824113759888>
- Goulson D, Lepais O, O'Connor S, Osborne JL, Sanderson RA, Cussans J, Goffe L, Darvill B. 2010. Effects of land use at a landscape scale on bumblebee nest density and survival. *J Appl Ecol* 47(6):1207–1215.
- Greenleaf SS, Williams NM, Winfree R, Kremen C. 2007. Bee foraging ranges and their relationship to body size. *Oecologia* 153(3):589–596.
- Gries C, Gilbert EE, Franz NM. 2014. Symbiota—A virtual platform for creating voucher-based biodiversity information communities. *Biodivers Data J*. doi: 10.3897/BDJ.2.e1114. [accessed 2019 Oct 1]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4092327/>
- Hernandez PA, Graham CH, Master LL, Albert DL. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29(5):773–785.
- Huete AR. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sens Environ* 25(3):295–309.
- [iDigBio] Integrated Digitized Biocollections. 2018a. iDigBio occurrence download: *Sporobolus heterolepis*. [accessed 2018 Sep 25]. [www.idigbio.org/portal/recordsets/e2def7e2-1455-4856-9823-6d3738417d24](http://www.idigbio.org/portal/recordsets/e2def7e2-1455-4856-9823-6d3738417d24)
- [iDigBio] Integrated Digitized Biocollections. 2018b. iDigBio occurrence download: *Muhlenbergia richardsonis*. [accessed 2018 Sep 25]. <https://www.gbif.org/occurrence/download/0011630-180824113759888>
- iNaturalist. 2018a. iNaturalist occurrence download: *Sporobolus heterolepis*. [accessed 2018 Sep 13]. [www.inaturalist.org](http://www.inaturalist.org)
- iNaturalist. 2018b. iNaturalist occurrence download: *Muhlenbergia richardsonis*. [accessed 2018 Sep 13]. [www.inaturalist.org](http://www.inaturalist.org)
- Karlen DL, Hurlay EG, Andrews SS, Cambardella CA, Meek DW, Duffy MD, Mallarino AP. 2006. Crop rotation effects on soil quality at three northern corn/soybean belt locations. *Agron J* 98(3):484–495.
- Kells AR, Goulson D. 2003. Preferred nesting sites of bumblebee queens (Hymenoptera: Apidae) in agroecosystems in the UK. *Biol Conserv* 109(2):165–174.
- Lemoine NP. 2015. Climate change may alter breeding ground distributions of eastern migratory monarchs (*Danaus plexippus*) via range expansion of *Asclepias* host plants. *PLoS ONE* 10(2):e0118614.
- Ma L, Li M, Ma X, Cheng L, Du P, Liu Y. 2017. A review of supervised object-based land-cover image classification. *ISPRS J Photogramm* 130:277–293.
- MacFarlane RP. 1974. Ecology of Bombinae (Hymenoptera: Apidae) of Southern Ontario, with emphasis on their natural enemies and relationships with flowers [PhD thesis]. Guelph (ON): Univ Guelph.
- McFeeters SK. 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int J Remote Sens* 17(7):1425–1432.
- Medler JT, Carney DW. 1963. Bumblebees of Wisconsin. Wisconsin Agricultural Experiment Station. Madison (WI): Agricultural Experiment Station, University of Wisconsin. [accessed 2015 Jan 26]. <http://agris.fao.org/agris-search/search.do?recordID=US201300435555>
- National Research Council. 2013. Assessing risks to endangered and threatened species from pesticides. Washington (DC): National Academies Pr. 194 p. [accessed 2018 Oct 17]. <https://www.nap.edu/catalog/18344/assessing-risks-to-endangered-and-threatened-species-from-pesticides>
- Ott RL, Longnecker MT. 2015. An introduction to statistical methods and data analysis. Belmont (CA): Duxbury Pr. 1183 p.
- Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME. 2017. Opening the black box: An open-source release of Maxent. *Ecography* 40(7):887–893.
- Pointon H. 2015. Larval host plant selection and daily behavior of Poweshiek skipperling (*Oarisma poweshiek*) in Michigan [PhD thesis]. Kalamazoo (MI): Kalamazoo College.
- Rhodes CJ. 2018. Pollinator decline — An ecological calamity in the making? *Sci Progress-UK* 101(2):121–160.
- Richards KW. 1978. Nest site selection by bumble bees (Hymenoptera: Apidae) in southern Alberta. *Can Entomol* 110(03):301–318.
- Richardson LL. 2018. Bumble bees of North America occurrence records database. [accessed 2018 Oct 16]. [www.leifrichardson/bbna](http://www.leifrichardson/bbna)
- Richardson LL, McFarland KP, Zahendra S, Hardy S. 2019. Bumble bee (*Bombus*) distribution and diversity in Vermont, USA: A century of change. *J Insect Conserv* 23(1):45–62.
- Schlicht D, Swengel A, Swengel S. 2009. Meta-analysis of survey data to assess trends of prairie butterflies in Minnesota, USA during 1979–2005. *J Insect Conserv* 13(4):429–447.
- Schlicht DW. 1997. Surveys for the Dakota skipper in Minnesota. Minneapolis (MN): Minnesota Department of Natural Resources. 9 p.
- Selby G. 2005. Status assessment and conservation guidelines: Poweshiek skipperling (*Oarisma poweshiek* (Parker) (Lepidoptera: Hesperidae). Bloomington (MN): US Fish and Wildlife Service, Twin Cities Field Office. 53 p.
- Selby G. 2008. Buffalo Ridge II butterfly survey results. South Dakota permit application, Appendix D. [accessed 2018 May 21]. <https://www.google.com/search?q=oarisma+poweshiek+gis&aq=chrome.69f57.5468j0j7&sourceid=chrome&ie=UTF-8>
- Semmler SJ. 2010. The nectar sources and flower preferences of the Poweshiek skipperling (*Oarisma poweshiek*) in Manitoba [PhD thesis]. Winnipeg (MB): Univ Winnipeg.
- Sjödin NE, Bengtsson J, Ekbohm B. 2008. The influence of grazing intensity and landscape composition on the diversity and abundance of flower-visiting insects. *J Appl Ecol* 45(3):763–772.
- Slaughter B, Hyde D, Cuthrell D, Lee YM, Norris RA. 2013. The conservation and management of prairie fens and associated species — Accomplishments and lessons from the MNDR landowner incentive program 2004–2013. Michigan Natural Features Inventory. Publication no. 2013-16. 53 p.
- Svensson B, Lagerlöf J, Svensson B. 2000. Habitat preferences of nest-seeking bumble bees (Hymenoptera: Apidae) in an agricultural landscape. *Agri Ecosyst Environ* 77(3):247–255.
- Swengel AB, Swengel SR. 1997. Co-occurrence of prairie and barrens butterflies: Applications to ecosystem conservation. *J Insect Conserv* 12(2):131–144.
- Swengel AB, Swengel SR. 1999. Observations of prairie skippers (*Oarisma poweshiek*, *Hesperia dacotae*, *H. ottoe*, *H. leonardus pawnee*, and *Atrytone arogos iowa*) (Lepidoptera: Hesperidae) in Iowa, Minnesota, and North Dakota during 1988–1997. *Great Lakes Entomol* 32(4):267–292.
- Swengel AB, Swengel SR. 2014. Paradoxes of Poweshiek skipperling (*Oarisma poweshiek*) (Lepidoptera: Hesperidae): Abundance patterns and management of a highly imperiled prairie species. *ISRN Entomology* 216427:1–10.
- Swengel SR, Schlicht D, Olsen F, Swengel AB. 2010. Declines of prairie butterflies in the midwestern USA. *J Insect Conserv* 15(1–2):327–339.
- Szymanski J, Smith T, Horton A, Parkin M, Ragan L, Masson G, Olsson E, Gifford K, Hill L. 2016. Rusty-patched bumble bee (*Bombus affinis*) species status assessment. Bloomington (MN): US Fish and Wildlife Service. 100 p. <https://www.fws.gov/midwest/endangered/insects/rpbb/pdf/SSARreportRPBBwAdd.pdf>
- [USDA NASS] US Department of Agriculture National Agriculture Statistics Service. 2017a. Quick Stats. 2013–2017. Washington (DC). [accessed 2018 May 1]. [http://nassgeodata.gmu.edu/Quick\\_Stats](http://nassgeodata.gmu.edu/Quick_Stats)
- [USDA NASS] US Department of Agriculture National Agriculture Statistics Service. 2017b. Cropland data layer. 2008–2017. Washington (DC). [accessed 2018 May 1]. <http://nassgeodata.gmu.edu/Crop-Scape>
- [USEPA] US Environmental Protection Agency. 2016a. Biological evaluation chapters for chlorpyrifos ESA assessment. [accessed 2019 Jan 1]. <https://www.epa.gov/endangered-species/biological-evaluation-chapters-chlorpyrifos-esa-assessment>
- [USEPA] US Environmental Protection Agency. 2016b. Biological evaluation chapters for malathion ESA assessment. [accessed 2019 Jan 1]. <https://www.epa.gov/endangered-species/biological-evaluation-chapters-malathion-esa-assessment>
- [USEPA] US Environmental Protection Agency. 2016c. Biological evaluation chapters for diazinon ESA assessment. [accessed 2019 Jan 1]. <https://www.epa.gov/endangered-species/biological-evaluation-chapters-diazinon-esa-assessment>
- [USFWS] US Fish and Wildlife Service. 2015. Critical habitat for Dakota skipper and Poweshiek skipperling. [accessed 2019 Jan 2]. <https://www.fws.gov/midwest/endangered/insects/dask/finalch.html>

- [USFWS] US Fish and Wildlife Service. 2017. Rusty-patched bumble bee guidance on ESA implementation. [accessed 2019 Jan 2]. <https://www.fws.gov/midwest/endangered/insects/rpbb/>
- [USFWS] US Fish and Wildlife Service. 2018a. Endangered species—Critical habitat overview. [accessed 2019 Jan 1]. <https://www.fws.gov/endangered/what-we-do/critical-habitats.html>
- [USFWS] US Fish and Wildlife Service. 2018b. Poweshiek skipperling (*Oarisma poweshiek*)—Candidate species. [accessed 2018 Dec 28]. <https://www.fws.gov/midwest/endangered/insects/posk/index.html>
- [USFWS] US Fish and Wildlife Service. 2018c. National wetlands inventory. [accessed 2019 Jan 2]. <https://www.fws.gov/wetlands/>
- [USFWS] US Fish and Wildlife Service. 2018d. Rusty-patched bumble bee (*Bombus affinis*). Endangered Species Act Section 7(a)(2) voluntary implementation guidance. Version 2.1. [accessed 2018 May 1]. [https://www.fws.gov/midwest/endangered/insects/rpbb/pdf/Section\\_7\\_Guidance\\_RPBB\\_Ver2\\_1\\_16April2019.pdf](https://www.fws.gov/midwest/endangered/insects/rpbb/pdf/Section_7_Guidance_RPBB_Ver2_1_16April2019.pdf)
- [USFWS] US Fish and Wildlife Service. 2019. Endangered Species Act section 7 consultation. [accessed 2019 Jan 1]. <https://www.fws.gov/midwest/endangered/section7/index.html>
- Wickham H, Francois R, Henry L, Muller K. 2018. dplyr: A grammar of data manipulation. Version 0.7.6. [accessed 2019 Jan 2]. <https://CRAN.R-project.org/package=dplyr>
- Wickham J, Stehman SV, Gass L, Dewitz JA, Sorenson DG, Granneman BJ, Poss RV, Baer LA. 2017. Thematic accuracy assessment of the 2011 National Land Cover Database (NLCD). *Remote Sens Environ* 191: 328–341.
- Williams PH, Thorp RW, Richardson LL, Colla SR. 2014. Bumble bees of North America: An identification guide. Princeton, NJ: Princeton Univ Pr. 208 p.
- Yost AC, Petersen SL, Gregg M, Miller R. 2008. Predictive modeling and mapping sage grouse (*Centrocercus urophasianus*) nesting habitat using maximum entropy and a long-term dataset from southern Oregon. *Ecol Inform* 3(6):375–386.