A national probabilistic characterization of local crop proximity and density for refining US screening level exposure estimates of pesticides in surface water arising from agricultural use – Supplemental Information

Christopher Holmes[a, 1] Dean Desmarteau[b], Amy Ritter[b] and Paul Hendley[c]

[a] *Applied Analysis Solutions,* 535 McDonald Rd, Winchester, VA USA
[b] *Waterborne Environmental, Inc.,* 897B Harrison St. SE, Leesburg, VA USA
[c] *Phasera Ltd.* Bracknell, Berkshire, UK,

Methods and Materials

Spatial Unit of Analysis

The National Hydrography Dataset Plus (NHD+) [12] contains the 1:100,000 NHD attributes to enhance stream network navigation, analysis, and display; an elevation-based catchment for each flowline in the stream network; catchment characteristics; flow direction; flow accumulation and elevation grids for each flowline in the stream network.

Each catchment has a single flowing water body (either stream/river feature types or canal/ditch features in which flow direction could be determined), where the outflow at the catchment outlet reflects the direct runoff from the entire catchment. These data account for the entire US land area and comprise a range of very small units highly relevant to farming practices at the local scale. Based on the 2.2 million

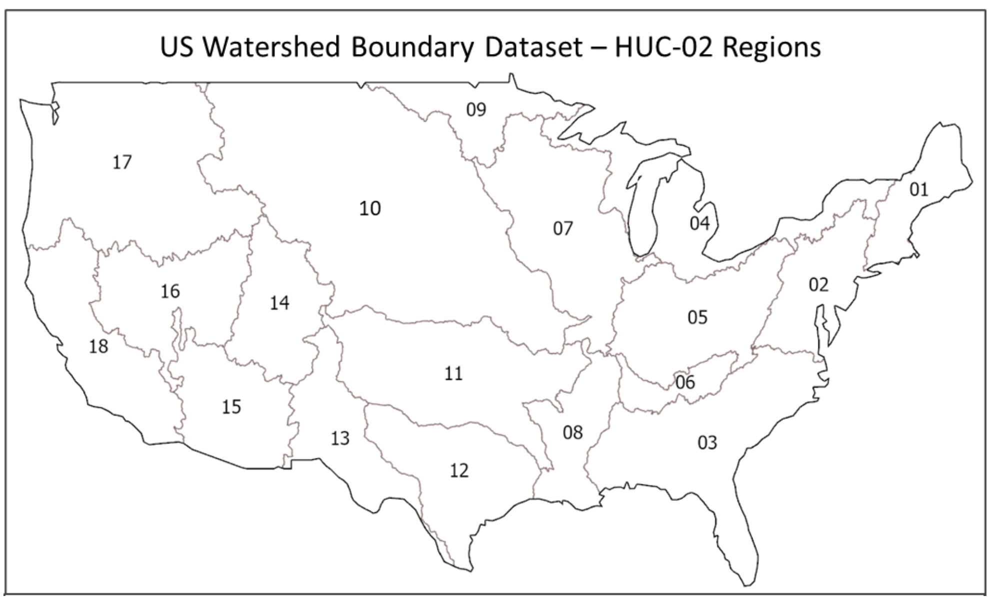

Figure S1. Map of HUC-02 units for conterminous US from the Watershed Boundary Dataset (WBD)

[1] Corresponding author: Christopher Holmes,
ChrisHolmes@AppliedAnalysis.solutions

catchments containing Cultivated Cropland across the US [4], 90% are 650 ha (2.5 mi^2) or smaller and 50% are smaller than 160 ha (0.62 mi^2). Table S1 provides a breakdown of catchment sizes and numbers by USGS two-digit Hydrologic Unit Code (HUC) boundaries from the USGS Watershed Boundary Dataset [18].

Table S1. Selected percentiles of NHD+ catchment area (km^2) for all agricultural catchments containing Cultivated Cropland in 2012[1]

HUC_02	Catchment area (km^2)											# of Catchments
	Min	1st	5th	10th	25th	50th	75th	90th	95th	99th	Max	
1	0.001	0.007	0.04	0.12	0.6	1.6	3.5	6.6	9.6	18.7	441	42,629
2	0.001	0.006	0.03	0.09	0.5	1.5	3.0	5.5	7.8	16.0	177	8,699
3	0.001	0.006	0.03	0.09	0.5	1.3	2.5	4.5	6.5	14.7	7984	265,573
4	0.001	0.008	0.05	0.14	0.7	1.8	3.8	7.4	11.1	23.2	170	82,480
5	0.001	0.011	0.08	0.21	0.8	1.7	3.1	5.4	7.7	16.8	162	147,817
6	0.001	0.006	0.04	0.11	0.6	1.4	2.6	4.2	5.6	9.4	59	49,416
7	0.001	0.009	0.06	0.16	0.7	1.7	3.3	5.9	8.5	18.5	144	153,724
8	0.001	0.005	0.03	0.06	0.3	1.1	2.2	4.1	6.0	14.1	374	110,657
9	0.001	0.005	0.04	0.11	0.6	2.0	5.7	13.4	22.1	52.7	435	21,109
10	0.001	0.007	0.04	0.13	0.6	1.7	3.4	6.5	9.8	23.4	1461	391,439
11	0.001	0.008	0.05	0.13	0.6	1.6	3.4	6.9	11.2	27.9	1209	176,167
12	0.001	0.011	0.08	0.26	1.3	4.0	8.6	16.1	23.1	50.0	1004	53,685
13	0.001	0.011	0.07	0.20	0.9	2.5	5.9	12.2	19.1	49.7	1026	49,597
14	0.001	0.007	0.04	0.12	0.7	2.1	4.7	8.8	12.7	25.5	337	72,959
15	0.001	0.008	0.05	0.14	0.7	2.1	4.3	8.0	11.6	25.0	1256	92,526
16	0.001	0.006	0.04	0.10	0.5	1.8	3.9	7.3	10.7	26.5	620	88,449
17	0.001	0.010	0.06	0.18	0.8	1.9	3.7	6.5	9.1	18.0	620	207,287
18	0.001	0.007	0.04	0.12	0.5	1.5	3.0	5.8	8.7	22.0	714	123,378
National	0.001	0.007	0.04	0.13	0.6	1.6	3.4	6.5	9.8	23.2	7984	2,217,591

[1] Agricultural catchments are defined as those containing at least some Cultivated Cropland as defined by USDA NASS [4]

Crop type and location

The NASS CDL program represents a cooperative venture between three USDA agencies (NASS headquarters, Foreign Agriculture Service group, and the Farm Service Agency Aerial Photography Field Office) plus in-state agreements among the Agricultural Statistics Service, the Department of Natural Resources and the Department of Agriculture.

USDA NASS includes over 30 individual vegetable and ground fruits in the CDL data product (USDA 2013). Classifying these individual vegetables and ground fruit from satellite imagery is challenging due to the large number of vegetable crops, the fact that they may rotate once or more during the year (i.e., three different vegetables grown on the same field in one calendar year), and that there may be a limited amount of training data available for classification of such diverse and possibly temporally transient crops. Consequently, all CDL vegetable and ground fruit crops were combined together into the ‘vegetables/ground fruit’ collective crop of interest (CoI) and the same was done for all tree nuts and tree fruit. **Error! Not a valid bookmark self-reference.** lists the relationship between the crops of interest (CoI), the CDL crop classes, and the USDA census of agriculture crop items.

Because the resulting total acres of vegetables/ground fruit in CA estimated from the NASS CDL data was less than (approximately 57% of the state total) the 2012 NASS Agricultural Census, the 5-year composite crop layer of vegetable and ground fruit was used for subsequent PCA calculations. The 5-year composite aggregates all areas that reported the vegetable and ground fruit classes in CDL in the last five years into a single data layer and represents 124% of the 2012 reported census acres in CA [2] and is therefore regarded as conservative.

In cases where all agricultural lands, regardless of crop type, needed to be identified, the 2012 NASS Cultivated Cropland layer [4] was used.

When catchment percent cropped acreages are estimated for crop groups like vegetables/ground fruit, tree nuts or orchard fruit, it gives rise to two sources of uncertainty. The first is the selection of the number of years of the coverage that are combined to generate the GIS layer. Either the 2012 single year crop coverage or a 5-year composite crop data set was selected. This ensured the PCA estimates would typically be an overestimate of the actual acres of the crop class in a region. The second source of uncertainty arises because the entire spatial extent of the crop class was used to estimate the PCAs while, in fact, a given crop (say onions or almonds) only represents a fraction of the crop complex acres (in the case of CA onions this is 4%). Since it was not possible to identify precisely which areas of the vegetable/ground fruit class were actually cropped to one specific type of vegetable, it was assumed that the entire area was that single crop for purposes of PCA estimation. Again, this is highly conservative and tends to overestimate the PCA values.

The goal was to select the most appropriate cropping year and spatial extent that provided the greatest number of NHD+ catchments while minimizing uncertainty introduced by CDL underestimates of cropping area or misclassification of crop (i.e., omission error). The 2012 USDA Census of Agriculture [2] was used as the measure of completeness for cropping area.

Table S2. Crosswalk table showing how CoIs, Census of Agriculture crops, and CDL crops relate to each other

Crop of Interest (CoI)	**USDA NASS Cropland Data Layer Classes**	**2012 Census of Agriculture Items**
Alfalfa	Alfalfa	Alfalfa Hay, Alfalfa Seed
Citrus	Citrus, Oranges	Citrus Fruit, All (Citrus, Other Citrus)
Corn	Corn, Pop or Orn Corn, Double Crop: Winter Wheat/Corn, Oats/Corn, Barley/Corn, Corn/Soybeans	Corn for Grain, Corn For Silage or Greenchop, Popcorn
Cotton	Cotton Double Crop: Lettuce/Cotton, Winter Wheat/Cotton, Soybeans/Cotton	Cotton, All (Pima, Upland)
Grass Seed	Sod/Grass Seed	Sod Harvested (Acres in the open), Grass Seed Crops, All
Peanuts	Peanuts	Peanuts for Nuts
Soybeans	Soybeans Double Crop: Winter Wheat/Soybeans, Soybeans/Cotton, Soybeans/Oats, Corn/Soybeans, Barley/Soybeans	Soybeans for Beans
Sunflower	Sunflower	Sunflower
Sweet Corn	Sweet Corn	Sweet Corn, Sweet Corn for Seed
Tree Fruits	Cherries, Peaches, Apples, Other Tree Crops, Pears, Prunes, Pomegranates, Nectarines, Plums, Apricots	All Non-Citrus Tree Fruits
Tree Nuts	Almonds, Pistachios, Pecans, Walnuts, Hazelnuts, Other Tree Crops	Almonds, Pistachio, Pecans- all, Walnuts, Hazelnuts
Vegetables, Ground Fruit	Dry Beans, Potatoes, Sweet Potatoes, Watermelons, Onions, Cucumbers, Chick Peas, Lentils, Peas, Tomatoes, Carrots, Asparagus, Garlic, Honeydew Melons, Broccoli, Peppers, Greens, Squash, Lettuce, Pumpkins, Double Crop Lettuce/Cantaloupe, Cabbage, Cauliflower, Celery, Radishes, Turnips, Eggplants, Gourds, Caneberries, Strawberries, Blueberries, Cranberries, Double Crop: Lettuce/Cotton, Lettuce/Durum Wheat, Lettuce/Barley, Misc Vegs and Fruits	Land in Berries, Land used for vegetables, Dry Edible Beans (Excluding Limas), Dry Edible Peas, Dry Lima Beans, Dry Southern Peas (Cowpeas), Lentils, Mustard Seed, Austrian Winter Peas, Dried Herbs, Dill for Oil, Mint for Oil
Wheat	Durum Wheat, Spring Wheat, Winter Wheat Double Crop: Winter Wheat/Corn, Lettuce/Durum Wheat, Durum Wheat/Sorghum, Winter Wheat/Sorghum, Winter Wheat/Cotton	Wheat for Grain, All (Winter Wheat, Spring Wheat, Durum Wheat, Other Spring Wheat)

Factors that were examined include:

- Regional preferences for cropping when crops not grown nationally. For example, because almost all almonds are grown in CA, only CA examined.
- Total acres of crop in CDL (2008 to 2012 and 5-year composite) compared to Census of Agriculture
- Omission error in the CDL as quantified in the USDA accuracy data [16]. A lower omission error means that fewer areas of CoI are missing (i.e. classified incorrectly as another land cover). Note that commission error was not assessed. This is because if land cover other than the CoI were classified as the CoI (i.e., commission error), this would only increase the potential for exposure, and hence this conservative aspect was accepted regardless of level of commission error.

All reasonable efforts were made to create a spatial land cover data layer for each crop that was representative of the area the crop was grown in the US, comprised at least the acreage reported in the 2012 USDA Census of Agriculture [2] for that spatial extent, and to quantify the aspects where these decisions added to the conservative nature of the land cover layer, and subsequent crop area calculations.

Table 1 in the main paper summarizes the acres from Census of Agriculture and final selected CDL data sets in terms of total acres and % of USDA census crop acres covered by the selected catchments. Aspects specific to particular CoI(s) are discussed below:

Alfalfa, corn, cotton, soybeans, wheat - The 2012 CDL and complete national extent for these crops was chosen because they have large national extents and the 2012 CDL represented between 97% and 143% of the national acres according to the USDA 2012 Census of Agriculture.

Citrus – Citrus is very regionally located, and FL encompassed 61% of the US citrus acres according to the USDA 2012 Census of Agriculture. FL is also one of the EPA Tier II scenarios for citrus. The 2012 CDL encompassed 181% of the census acres and was therefore selected as the FL citrus land cover since the CDL was clearly conservative in citrus representation in FL.

Grass seed – The “Field and grass seed crops, all” category from the USDA 2012 Census of Agriculture was examined for this crop. In 2012 OR cultivated 420,767 acres (55% of national total of 759,534), while the next highest state (MO) only had 76,749 acres. In addition, the CDL crop class containing grass seed (Class 59) also includes sod farms, so it was important to have a spatial extent that included large amounts of grass seed, but little sod. Because OR contained by far the largest state level grass seed, had <1800 sod acres, and is the location of the EPA Tier II scenario, it was selected as the spatial extent. Because OR was processed at a later date than the other crops, 2013 CDL was available and utilized for this crop.

Peanuts – The states of GA, FL and AL comprise 71% of the peanut acres in in 2012 and had reasonable omission errors in the 2012 CDL (10%-20%). TX, NC and FL each also had over 100,000 acres of peanuts in 2012, but 2012 CDL omission errors were deemed too large, ranging from 24% to 66%. Therefore, GA, FL and AL were selected as the spatial extent using the 2012 CDL, representing 119% of the 2012 census acres.

Sunflower – the 2012 national total did not fully represent the USDA national total (85%), while the 5-year composite was extremely over predictive (351%). Because over 78% of sunflower acres in 2012 were produced in ND and SD according to the Census of Agriculture, and the 2012 CDL omission error was relatively low for these states (11%), a hybrid approach was utilized in which the 2012 CDL for SD and ND was combined with the 5-year composite for the rest of the country. This resulted in the final spatial cropping data layer being 143% of the national total acres (from Census of Agriculture).

Sweet corn – The national extent was selected using the 5-year composite of CDL land cover in order to ensure complete coverage of all sweet corn growing areas. This is because the 2012 CDL comprised only a total of 53% of the national acres, while the 5-year composite accounted for 198% of the national acres.

Almonds – Almost all almond production in the US is located in CA, therefore the 2012 CA CDL (representing 114% of all tree nuts acres in CA) was used as the spatial and temporal extent. While we used a combined tree nut class (which also includes pecans, walnuts, and pistachios), the 2012 USDA Census of Agriculture shows that almonds account for only 62% of the total tree nuts acres in CA. Therefore, this is regarded as a conservative estimation of almond cropping extent and density.

Pecans – TX, GA and NM account for 57% of the national pecan acreage in the 2012 USDA Census of Agriculture. Of the remaining states, only OK produces more than 20,000 acres, however the CDL data do not represent this crop sufficiently (less than 13,000 of the more than 100,000 acres in OK), even using the 5-year CDL composite. Therefore, OK was not included in the spatial extent. TX and NM also contained less than the census acres in the 2012 CDL, therefore the 5-year composite was used for these states comprising 80% and 186% of census acres respectively. The 2012 CDL for GA encompassed 199% of the census acres, and was therefore selected, resulting in a hybrid CDL crop layer.

Lettuce – This crop is contained in the combined vegetables & ground fruit crop group composed of multiple vegetable and fruit CDL crop types. CA produces more vegetables than any other state and was selected as the state to examine for lettuce. Because the resulting total acres of vegetables/ground fruit in CA estimated from the CDL data was less than the 2012 USDA Agricultural Census total (57% of the state total), the 5-year composite crop layer of vegetable and ground fruit was used for subsequent PCA calculations. The 5-year composite aggregates all areas that reported the vegetable and ground fruit classes in CDL in the last five years into a single data layer and represents 124% of the 2012 reported census acres in CA and is therefore regarded as conservative. In addition, each of the individual crops is only a portion of the total vegetable and ground fruit crop class and would again be over-represented in the spatial land cover. Lettuce comprises only 21% of the total vegetables / ground fruit produced in CA reported in the 2012 census, which is additionally conservative. CA comprised more than 70% of lettuce acres harvested in the entire US in 2012.

Pepper – FL was selected as an additional state to examine vegetables production because it is also an important vegetable producing state, and the EPA Tier II pepper scenario is located in FL. The 2012 CDL was selected for FL even though it contained less than the 2012 census acres (24%) because the pepper crop represents such a small proportion of the vegetables & ground fruit in FL (5.7%), using the 5-year composite would vastly overestimate the pepper acres examined. The ~58,000 acres of CDL used in the analysis represent over 400% of the total FL pepper acres grown in 2012.

Potato - For potatoes, five of the top six potato producing states (based on 2012 Census of Agriculture) were chosen for landscape processing: ID, WA, WI, ME, and CO. Potatoes were included in the vegetables / ground fruit crop class, and the 2012 vegetables / ground fruit acres

Table S3. Summary of spatial extent and year of CDL for each crop based on national examination of CDL and USDA Census of Agriculture acreages

CDL Crop Group	US EPA Scenario Crop of Interest (CoI)	Spatial Extent	2012 USDA Census Harvested (acres) (crop group)	CDL 2012 (acres) (crop group)	5-yr Composite (acres) (crop group)	Hybrid (acres)	CDL 2012 % of 2012 Survey	5-yr composite % of 2012 Survey	Hybrid %	Year(s) selected	% of National CoI Acres Repres-ented	% of crop group that is CoI
Alfalfa	Alfalfa	National	16,710,820	16,165,805	38,176,095		97%	228%		2012	100%	
Citrus	Citrus	FL	539,908	976,906	1,872,681		181%	347%		2012	61.3%	
Corn	Corn	National	94,816,833	95,651,409	194,904,058		101%	206%		2012	100%	
Cotton	Cotton	National	9,384,080	13,451,958	25,452,190		143%	271%		2012	100%	
Grass Seed	Grass Seed	OR	420,767	420,122[2]	920,297		99%	218%		2013[2]	39.1%	
Peanuts	Peanuts	GA, FL, AL	1,146,206	1,368,424	3,214,760		119%	280%		2012	70.7%	
Soybeans	Soybeans	National	76,104,385	75,243,102	175,900,965		99%	231%		2012	100%	
Sunflower	Sunflower	National	1,876,890	1,594,947	6,584,195	2,676,850	85%	351%	143%	Hybrid[1]	100%	
Sweet Corn	Sweet Corn	National	571,611	301,398	1,129,230		53%	198%		Composite	100%	
Tree Nuts	Almonds	CA	1,496,610	1,704,659	3,129,560		114%	209%		2012	100%	62.5%
	Pecans	GA, TX, NM	330,314	322,999	726,851	453,758	98%	220%	137%	Hybrid[1]	57.1%	99.8%

Table S4. Summary of spatial extent and year of CDL for each crop based on national examination of CDL and USDA Census of Agriculture acreages (Contd.)

CDL Crop Group	US EPA Scenario Crop of Interest (CoI)	Spatial Extent	2012 USDA Census Harvested (acres) (crop group)	CDL 2012 (acres) (crop group)	5-yr Composite (acres) (crop group)	Hybrid (acres)	CDL 2012 % of 2012 Survey	5-yr composite % of 2012 Survey	Hybrid %	Year(s) selected	% of National CoI Acres Repres-ented	% of crop group that is CoI
	Pepper	FL	239,277	57,657	140,655		24%	59%		2012	16.6%	5.7%
	Potato	CO, ID, ME, WA, WI	1,779,460	1,519,791	4,595,914		85%	258%		2012	60.0%	39.1%
	Potato	ME	112,508	123,434	357,025		110%	317%		2012	5.3%	54.5%
Wheat	Wheat	National	49,038,649	55,399,808	134,930,046		113%	275%		2012	100%	

[1]A hybrid of 2012 and 5-yr composite was used. Sunflower: ND & SD were 2012, all other states composite. Pecans: GA was 2012 and TX, NM composite. All other states 2012.

[2]Grass Seed was added at a later time and 2013 was the latest available CDL at that time.

Table S4 summarizes the acres from 2012 Census of Agriculture and final selected CDL data sets in terms of total acres. Total CDL acres are presented for 2012 as well as 5-year composite, with the selected dataset in black text (and not selected in grey text). The last column in the table includes the percent of the collective crop group that is the COI for tree fruit, tree nuts, and vegetables/ground fruit.

represented between 169 to 317% of the 2012 census potatoes acres. These states represent 60% of the national potato acres. ND (ranked third) was excluded from the analysis because the ND 2012 CDL vegetables / ground fruit acres would have represented 1254% of the potato acres due to the large amounts of other vegetables produced in ND (e.g., peas, beans, sugar beets). Because potatoes are only 39% of the accumulated class of vegetables/ground fruit in these 5 states, the use of this spatial coverage to provide PCA information for the exposure assessment is, by definition, conservative. In order to examine the ME potato scenario more specifically, an additional spatial extent of only ME was created using the 2012 CDL (representing 110% of the 2012 census potato acres in ME).

Wheat – This wide-ranging crop is present in the majority of US states, although spring, durum and winter varieties may be more regional. Nationally, the 2012 CDL represented 113% of the total wheat acres in the 2012 USDA Census of Agriculture [2] and was selected to represent all wheat.

Crop frequency

One of the assumptions made in the standard Tier II modeling is the use of a 30-year simulation period in which the field being modeled is cropped to the crop of interest for 30 years continually. Spatial data from USDA NASS, the crop frequency data layer, provides an eight-year cropping history of each pixel covering four main crops (corn, cotton, soybeans, and wheat) [4] The frequency layer for each crop provides an integer value for each pixel, ranging from 0 to 8, indicating the number of years that pixel was the CoI from 2008 to 2015.

Using a national extent for each crop, the total acres for each of the years 1 to 8 were summarized and compared to the overall national total to determine what percentage of the cropped acres had 1, 2, …, 7, or 8 years of cropping between 2008-2015.

Error! Reference source not found. shows that, for example, only 20% of corn acres are cropped more than 4 years in the last 8 years (i.e., 5, 6, 7, or 8 years in 8). In other words, only 20% of corn acres were cropped more often than every other year on average over the last 8 years. Comparable metrics for cotton, soybeans and wheat are 29% of cotton, 15% of soybeans, and 17% of wheat acres are cropped more often than 4 in 8 years. Less than 10% of cotton acres were cropped to cotton in every one of the years 2008-2015 (i.e., comparable to the modeled Tier II scenario), and only 3% or less for the other crops.

Crop proximity to surface water

The proximity processing was performed in a raster GIS environment to enable a viable approach covering the 2.6 million catchments in the US. The source NHD+ flowing water features, originally supplied as lines and polygons, were converted to a raster dataset with a 10m resolution. This means that line features in the NHD+ had a minimum width of 10m, while polygon features (e.g., rivers) had a minimum width of 10m and a maximum width corresponding to the width of the polygon (+/- 5m due to rasterization).

The buffer function applied to the rasterized hydrology assigns a distance to each 10m pixel extending outward from the source features (i.e., water pixels). The closest pixels (i.e., the pixels adjacent to the water pixels, including diagonal) were assigned to the 0-10m proximity zone. In effect, the center point of these "0-10m buffer" pixels were either 10m from the center point of the closest water pixel if orthogonal, or 14m from the center of the water pixel if diagonal. Similar methodology was used for the 50m and 200m buffer distances. Because the pixel is the basic unit of raster datasets, they cannot be divided, hence the inclusion of diagonal pixels in which the center point was >10m. At least some portion of these pixels are within 10m of the water pixel.

Table S5. Percentage of crop area containing crop from 2008-2015 using USDA NASS crop frequency data

	Corn			**Cotton**		
	Acres	Class %	Cumulative %	Acres	Class %	Cumulative %
8 of 8 years cropped	4,316,940	2%	2%	2,413,481	9%	9%
7 of 8 years cropped	6,344,004	3%	5%	1,662,960	6%	14%
6 of 8 years cropped	11,243,007	5%	10%	1,751,670	6%	21%
5 of 8 years cropped	21,385,308	10%	20%	2,222,444	8%	29%
4 of 8 years cropped	58,909,308	27%	47%	2,891,029	10%	39%
3 of 8 years cropped	35,038,426	16%	64%	3,477,748	12%	51%
2 of 8 years cropped	31,452,001	15%	78%	4,564,060	16%	67%
1 of 8 years cropped	46,791,881	22%	100%	9,169,690	33%	100%
Total	215,480,874			28,153,082		

	Soybeans			**Wheat**	
	Acres	Class %	Cumulative %	Acres	Class %
8 of 8 years cropped	1,601,279	1%	1%	4,486,650	3%
7 of 8 years cropped	3,299,996	2%	2%	4,129,733	3%
6 of 8 years cropped	7,178,143	4%	6%	6,625,034	4%
5 of 8 years cropped	17,520,938	9%	15%	11,242,700	7%
4 of 8 years cropped	60,270,692	31%	46%	25,272,310	16%
3 of 8 years cropped	37,673,911	19%	65%	25,475,065	16%
2 of 8 years cropped	30,404,207	15%	80%	28,650,211	18%
1 of 8 years cropped	38,410,697	20%	100%	49,531,199	32%
Total	196,359,863			155,412,901	

Table S6. Complete listing of all NHDPlus flowlines (by FTYPE/FCODE) and whether they were included in the rasterization and proximity zone generation

FTYPE	FCODE	DESCRIPTION	LENGTH KM
Included in rasterization and proximity zone generation			
StreamRiver	46000	Stream/River	1,588
StreamRiver	46003	Stream/River: Hydrographic Category = Intermittent	3,213,025
460	46003	Stream/River: Hydrographic Category = Intermittent	3
Uninitialized	46003	Stream/River: Hydrographic Category = Intermittent	2
StreamRiver	46006	Stream/River: Hydrographic Category = Perennial	1,863,493
ArtificialPath	46006	Stream/River: Hydrographic Category = Perennial	39
CanalDitch	46006	Stream/River: Hydrographic Category = Perennial	2
ArtificialPath	55800	Artificial Path	277,973
StreamRiver	55800	Artificial Path	8
Artificial Path	55800	Artificial Path	7
Connector	55800	Artificial Path	1
		Total	5,356,141(94%)
Not included in rasterization or proximity zone generation			
Connector	33400	Connector	8,562
StreamRiver	33400	Connector	4
CanalDitch	33600	Canal/Ditch	283,435
ArtificialPath	33600	Canal/Ditch	10
Uninitialized	33600	Canal/Ditch	3
StreamRiver	33600	Canal/Ditch	1
CanalDitch	33601	Canal/Ditch: Canal/Ditch Type = Aqueduct	622
Pipeline	42800	Pipeline	37

Table S7. Complete listing of all NHDPlus flowlines (by FTYPE/FCODE) and whether they were included in the rasterization and proximity zone generation (contd.)

FTYPE	FCODE	DESCRIPTION	LENGTH KM
Not included in rasterization or proximity zone generation			
Pipeline	42801	Pipeline Type = Aqueduct; Relationship to Surface = At or Near	2,220
ArtificialPath	42801	Pipeline Type = Aqueduct; Relationship to Surface = At or Near	0
Pipeline	42802	Pipeline Type = Aqueduct; Relationship to Surface = Elevated	23
ArtificialPath	42802	Pipeline Type = Aqueduct; Relationship to Surface = Elevated	3
Pipeline	42803	Pipeline Type = Aqueduct; Relationship to Surface = Underground	7,226
ArtificialPath	42803	Pipeline Type = Aqueduct; Relationship to Surface = Underground	0
Pipeline	42804	Pipeline Type = Aqueduct; Relationship to Surface = Underwater	1
Pipeline	42805	Pipeline Type = General Case; Relationship to Surface = At or Near	5
Pipeline	42806	Pipeline Type = General Case; Relationship to Surface = Elevated	2
Pipeline	42807	Pipeline Type = General Case; Relationship to Surface = Underground	15
Pipeline	42809	Pipeline Type = Penstock; Relationship to Surface = At or Near	73
Pipeline	42811	Pipeline Type = Penstock; Relationship to Surface = Underground	15
Pipeline	42813	Pipeline: Pipeline Type = Siphon	156
Pipeline	42816	Pipeline: Pipeline Type = Aqueduct	2
Coastline	56600	Coastline	46,177
		Total	348,592 (6%)

Proximity zones were not created for flowline features in which flow direction was unknown (i.e., FlowDir = “Uninitialized”). Because these features are man-made (e.g., canals, ditches, pipelines) they often do not follow topographic features of the landscape, and a flow direction was not able to be determined based on elevation data in the generation of the NHD+ dataset, and as such do not have a separate catchment. Features with FlowDir = “Uninitialized” represented only 6% of the total flowline length in the US.

Table S6 lists the types of flowlines present in the NHD+, their total length, and whether they were included in the rasterization and proximity zone analysis. In this table "Artificial Path" represents the centerline of a river or other water feature represented by a polygon in the NHD+.

Because NHD+ catchments can be small in area, and flowlines may be near catchment boundaries, the proximity zone may extend outside the catchment boundary into an adjoining catchment. In these instances, the amount of proximity zone area extending outside is attributed to the neighboring catchment (and considered in the PCA for that catchment). In other words, metrics generated for a catchment are based only on the area within the catchment, regardless of source water feature of the proximity zones.

USEPA screening exposure approach

For a screening ecological exposure assessment, USEPA currently utilizes the Pesticide Root Zone Model [19], the Variable Volume Water Body Model [17] and AgDRIFT® model [10] to estimate surface water concentrations resulting from off-target mass loading following agricultural use of pesticides. PRZM is a field-scale one-dimensional flow and transport model used to simulate runoff and erosion masses of pesticide residues from a standard 10-ha field, providing edge-of-field loadings due to rainfall runoff/erosion events into USEPA's standard pond (1 ha x 2 m deep) that is modeled using the VVWM model. Additionally, the AgDRIFT® model is used to estimate off-target spray drift deposition onto the 1-ha water body adjacent to the field resulting from ground, airblast, or aerial applications. The spray drift deposition that occurs assumes that the wind is always blowing directly towards the pond for every application in a season. The pesticide mass loadings from PRZM, which includes drift fractions generated by the AgDRIFT model, are entered into the VVWM model to estimate aquatic concentrations on a daily timestep.

For each crop, USEPA has designed >100 "scenarios" intended to represent crop-specific landscape conditions vulnerable to chemical transport to aquatic ecosystems due to runoff and erosion. These scenarios identify a typical soil and slope used for the CoI in a particular region using locally appropriate crop timing (i.e., emergence and harvest), and a specified Solar and Meteorological Surface Observational Network (SAMSON) weather station [8] relevant to the soil/crop location. However, all scenarios utilize the same assumptions about field size, wind speed, and receiving water body characteristics. The scenarios provide inputs to the PRZM model that is typically run for a 30-year (1961-1990) period using the daily weather data.

A water body model (currently VVWM) estimates the fate and exposure of a compound after its entry into the aquatic system. VVWM outputs include annual maximum Estimated Environmental Concentrations (EECs) for 24-h, 96-h, 21-d, 60-d, and 90-d time-weighted averages, along with maximum instantaneous (peak) and mean annual EECs in the water column, sediment, and pore water compartments. The 30 years of daily exposure data are then analyzed to generate a distribution of 30 annual maximum time-weighted average EECs in the water body. The model output endpoint that the USEPA uses as an aquatic level of concern [11] is the *"…one-in-ten-year exceedance at a vulnerable use site that is representative of the 90th percentile of all sites across the United States where that specific type of application occurs."* Consequently, the model output is examined to identify a 1-in-10-year maximum concentration by selecting the 90th percentile value from the 30 annual maxima concentrations. Recent USEPA review of model daily time series has found the scenarios often represent a concentration at much lower occurrence frequency than the 90th percentile of the entire 30 years of 10,950 daily EECs, and in some cases greater than the 99.9th percentile [13].

Modified modeling approach

For this study, we used PRZM (version 3.12, Suarez 2005) for off-target runoff/erosion mass from the field and used the AgDRIFT (version 2.0.10) model to determine the drift fraction from applications. These versions of the models were both in use by USEPA in 2013. However, we utilized the AGRO-2014 (instead of EXAMS which was a precursor to the VVWM model) as the receiving water body model since it is more accurate in handling hydrophobic chemicals such as pyrethroids due to its enhanced capability of simulating sediment dynamic processes [7]. Since that time, USEPA adopted the use of the current VVWM model that addresses some of the aspects related to highly hydrophobic compounds that were accounted for in the AGRO-2014 model used in our modeling.

At the time this modeling was conducted, the product labels for all foliar pyrethroids required a 10-ft vegetative filter strip (VFS) between any area being treated and an adjacent water body. Standard PRZM/EXAMS or PRZM/VVWM modeling has no mechanism to account for the impact of a VFS in reducing the amount of chemical present in edge-of-field runoff or eroded sediment which reaches the receiving water body after passing through a VFS. With some pesticides this may not be a critical omission; however, the extreme hydrophobicity of pyrethroids means that most of the residues transported due to runoff will be adsorbed to soil particles. There has been considerable research recently in the US and Europe to design and validate models that can simulate this process. The best available model that has received extensive peer review is the Vegetative Filter Strip MODeling system (VFSMOD) [6].

For the present study, baseline modeling was conducted in 2013 and previous versions and/or different models were utilized; however, they were conceptually and functionally similar to current USEPA standard models described above. It is important to note that the utilization of the models in the present study to simulate the baseline EECs still provide a conservative estimate for a screening level risk assessment.

Sunflower scenario development

No USEPA Sunflower scenario existed, so to cover this pyrethroid-important crop, this scenario was generated based on the ND corn scenario. The USEPA ND corn scenario was used to develop a ND sunflower scenario because it was already parameterized for a row crop and was set in a county (Pembina County) which also has sunflower production, according to the 2007 Census of Agriculture (**Figure S2**)[1]. Additionally, the ND corn scenario was selected for sunflowers because there is substantial sunflower production on the Bearden soils used for the USEPA corn scenario. Based on the NRI database [14], the Bearden soil is one of the top 10 soils with sunflower acreage in ND.

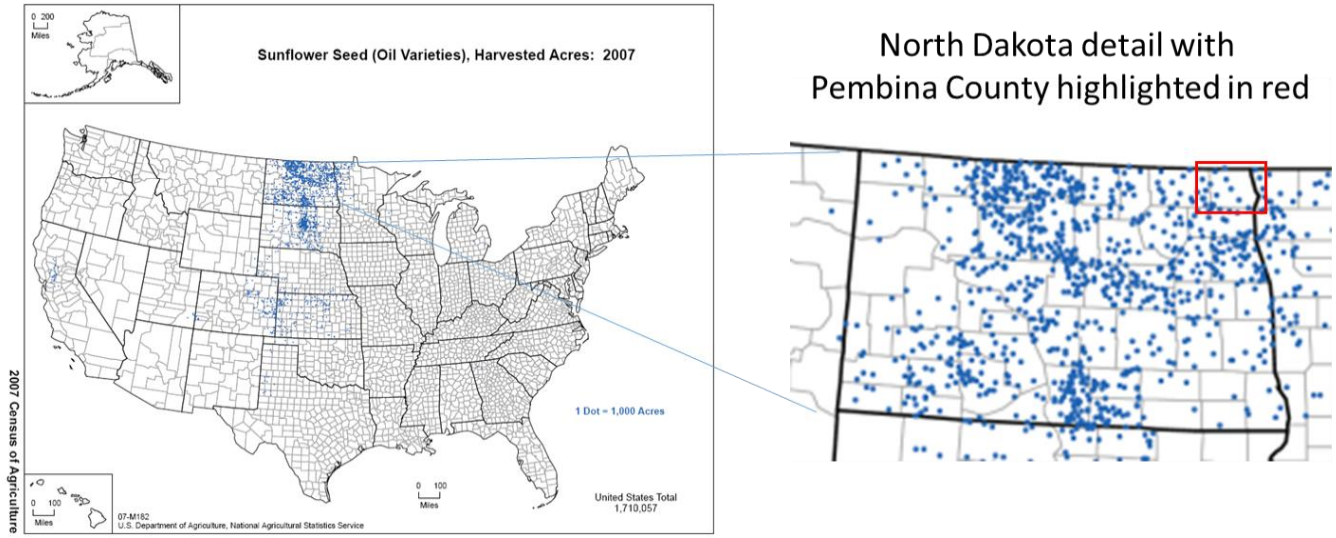

Figure S2. Map of sunflower production (2007) and location of Pembina County

All parameters built into the ND corn scenario were used for the ND sunflower scenario with the exception of the cropping dates (emergence, maturation and harvest). Sunflower crop parameters such as rooting depth, maximum crop canopy coverage and USLD factors were determined to be similar to the values for corn [3]a nd therefore were not modified for the ND sunflower scenario. The cropping dates for the ND sunflower scenario were based on guidance from available literature [3] and are compared with the ND corn cropping dates in the table below.

Table S8. USEPA ND Corn scenario cropping dates and associated developed ND Sunflower scenario dates

Scenario	Emergence Date	Maturation Date	Harvest Date
ND Corn (OP)	May 5	Aug. 5	Aug. 12
ND Sunflower	May 16	Sept. 5	Sept. 16

Details of Catchment Agronomic Distributional Analyses (CADA)

Table S9 provides an example of the CADA process. Refer to main paper Figure 2 and accompanying text.

Table S9. CADA approach combining 30 years of annual maxima EECs from the MS cotton baseline scenario modeling with PCA group and probabilities to create 300 representative probability weighted simulated year EECs.

Year Rank	Baseline EEC (μg/L)	PCA Percentile Group	PCA	CADA EEC (Baseline EEC x PCA as fraction) (μg/L)	Occurrence Probability of over 30 years	Simulated Year
1	0.0157	100	100%	0.0157	0.0003	1
		99	57.4%	0.0090	0.0003	2
		98	42.9%	0.0067	0.0007	3
		96	29.5%	0.0046	0.0010	4
		93	19.5%	0.0031	0.0010	5
		90	13.7%	0.0021	0.0008	6
		87.5	10.5%	0.0016	0.0042	7
		75	2.9%	0.0005	0.0083	8
		50	0.3%	0.0000	0.0083	9
		25	0.0%	0.0000	0.0083	10

Table S10. CADA approach combining 30 years of annual maxima EECs from the MS cotton baseline scenario modeling with PCA group and probabilities to create 300 representative probability weighted simulated year EECs.

Year Rank	Baseline EEC (µg/L)	PCA Percentile Group	PCA	CADA EEC (Baseline EEC x PCA as fraction) (µg/L)	Occurrence Probability of over 30 years	Simulated Year
2	0.0153	100	100%	0.0153	0.0003	11
		99	57.4%	0.0088	0.0003	12
		98	42.9%	0.0065	0.0007	13
		96	29.5%	0.0045	0.0010	14
…	…	…	…	…	…	…
29	0.0044	87.5	10.5%	0.0005	0.0042	287
		75	2.9%	0.0001	0.0083	288
		50	0.3%	0.0000	0.0083	289
		25	0.0%	0.0000	0.0083	290
30	0.0028	100	100%	0.0028	0.0003	291
		99	57.4%	0.0016	0.0003	292
		98	42.9%	0.0012	0.0007	293
		96	29.5%	0.0008	0.0010	294
		93	19.5%	0.0006	0.0010	295
		90	13.7%	0.0004	0.0008	296
		87.5	10.5%	0.0003	0.0042	297
		75	2.9%	0.0001	0.0083	298
		50	0.3%	0.0000	0.0083	299
		25	0.0%	0.0000	0.0083	300
				All 300 simulated years:	**1.0000**	

Results

Cropping density (PCA) in Proximity Zones

Figure S3 and Figure S4 illustrate the catchment-level distributions of PCA for each CoI.

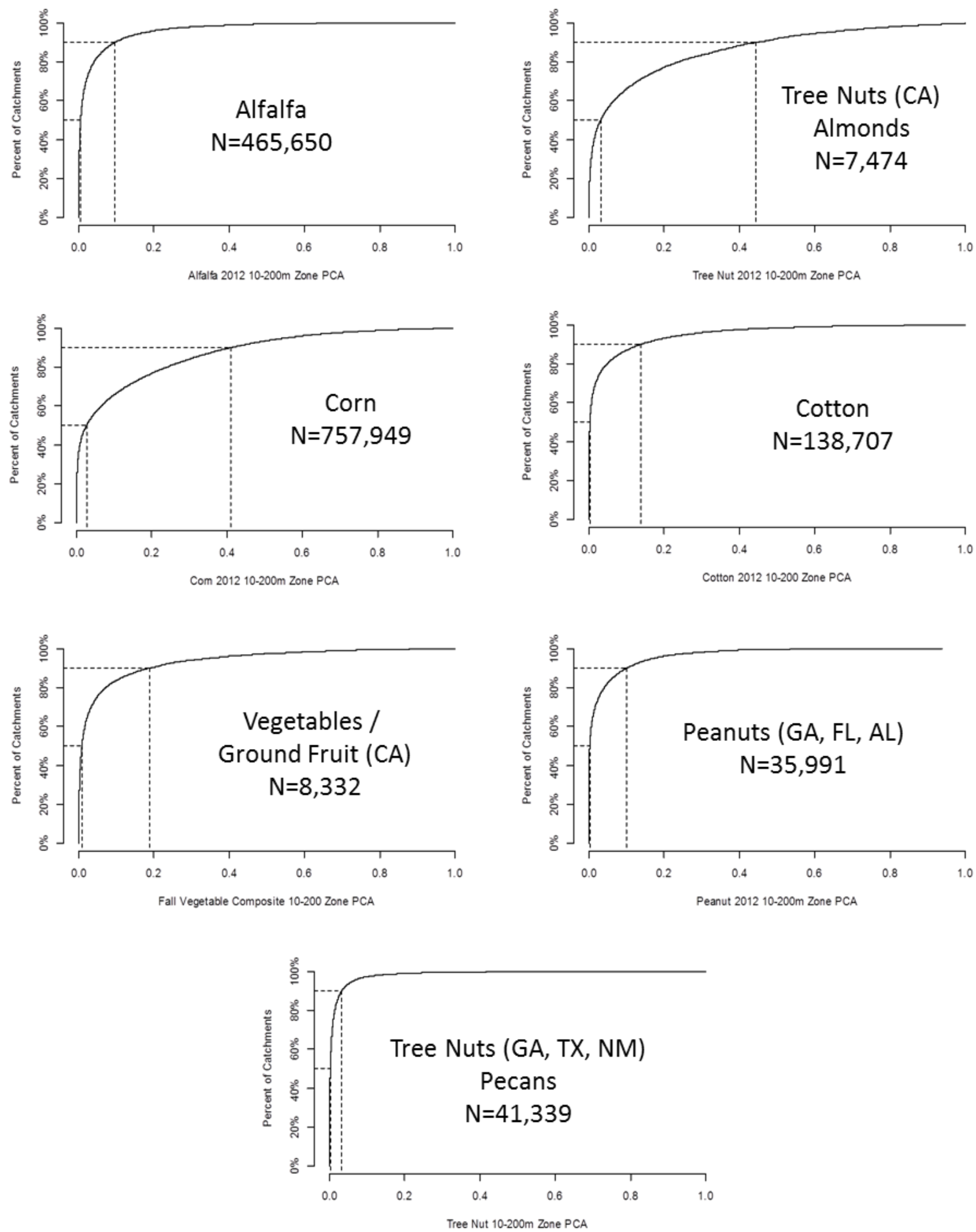

Figure S3. Catchment PCA distributions in the 10- to 200-m proximity zones around stream reaches in each NHD+ catchment for each CoI

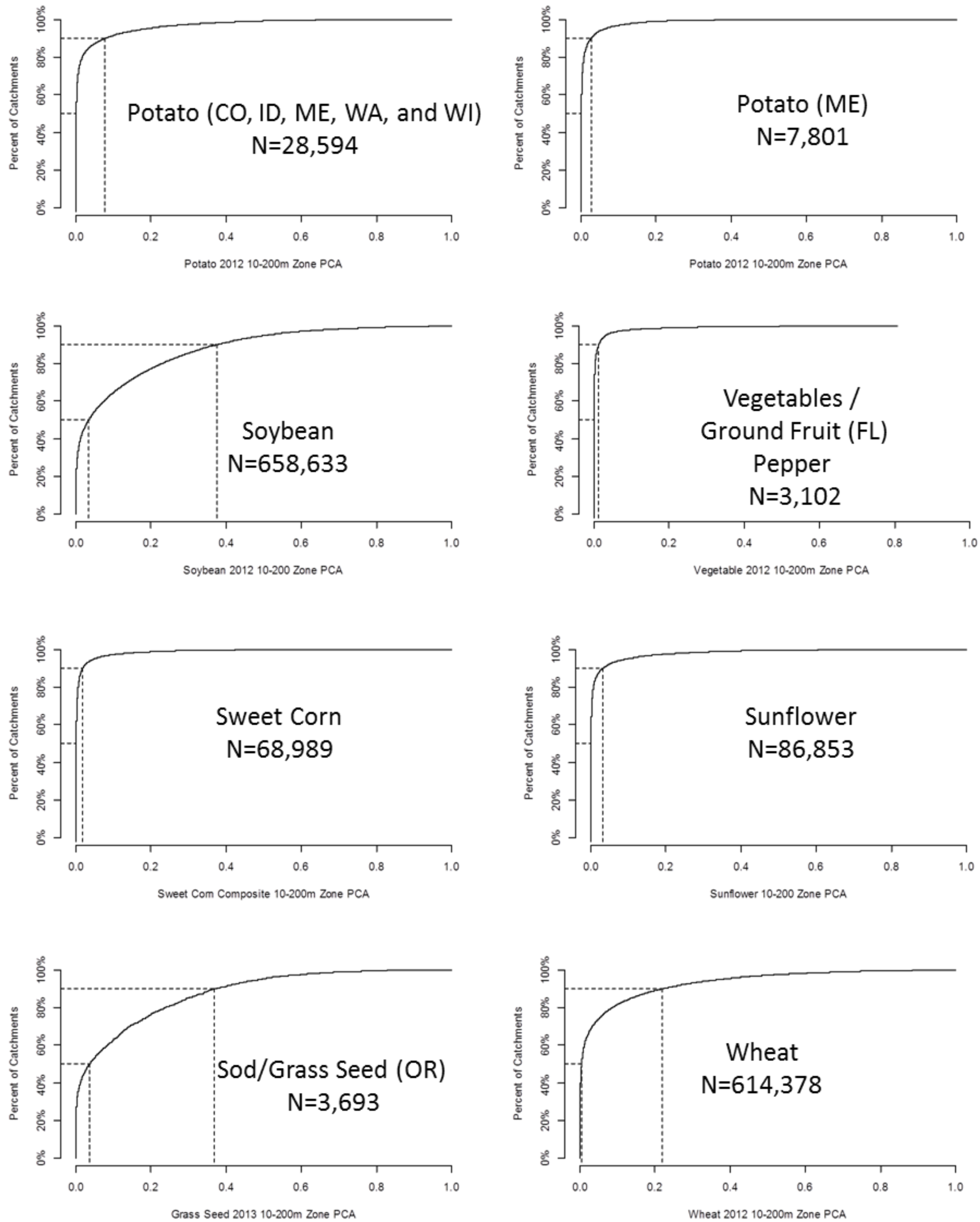

Figure S4. Catchment PCA distributions in the 10- to 200-m proximity zones around stream reaches in each NHD+ catchment for each CoI

CADA analyses - impact of PCAs on 21-day sediment baseline exposure assessments
Figure S5 illustrates the 21-day sediment data from the baseline EEC distribution compared with the output from the CADA approach for the MS cotton scenario for a single representative pyrethroid. The figure displays the distribution of 30 simulated annual maxima EECs from the baseline assessment based on the 100% cropped delivery area assumption (blue points) with the baseline EEC identified by point A (green). This presentation highlights the fact that the regulatory assumption is the 1-in-10 year value from the 30 years of modeling and so there are exposures higher than this regulatory concentration endpoint. The red line shows the distribution of the CADA simulated 300 water body yearly annual maxima obtained by applying the probability distribution of real-world PCAs. The purple arrow (at point A) shows the magnitude of reduction in probability of encountering the baseline EEC using the CADA approach, in this case 1-in-10 year (10% probability) is reduced to 0.067% probability. The horizontal grey arrow shows the extent the 1-in-10 year maximum EEC is reduced (i.e., the multiplication factor) by considering the impact of the crop-specific PCA on estimated aquatic exposures (in this case by a factor of 12). The orange shaded area identified by arrow B indicates that this probabilistic refinement does not negate the finding that concentrations greater than the 1-in-10 year baseline regulatory value may still occur. However, instead of exceedances regarded as occurring in two years out of every 30 (6.7% probability), their likelihood of occurrence is greatly reduced (in this case to less than 0.1% probability). The impact of this probabilistic approach applies throughout the distributions; for example, the horizontal blue arrow shows the reduction magnitude of the 50 percentile (1-in-2 year) EEC (in this case by a factor of 157).

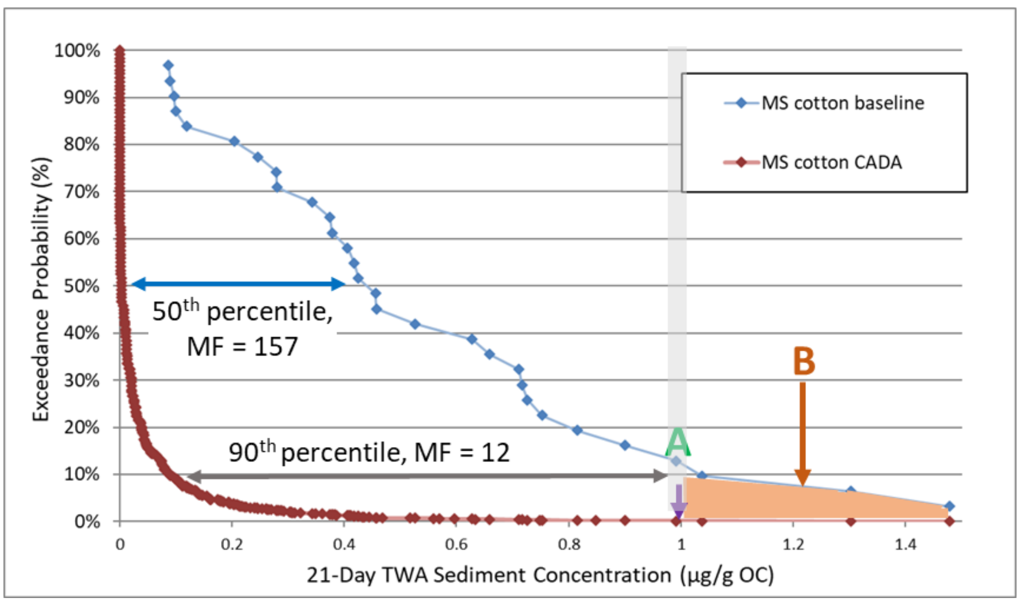

Figure S5. Results showing application of CADA (red points) to the distribution of 30 simulated annual maxima EECs for cotton from the baseline assessment (blue points) for water column with 50th and 90th percentile MFs illustrated (blue and grey horizontal lines, respectively). Vertical grey bar represents the single EEC that is selected for baseline scenario cases. Green arrow (A) shows the reduction in probability of exceeding the baseline EEC, and orange shaded area (B) illustrates concentrations greater than the baseline regulatory value may occur but with far lower probability.

Using the same approach, Table S11 reports the resulting MF values for the 90th and 50th percentile water body year EECs for a representative pyrethroid across the 15 CoIs and 18 USEPA scenarios. This table shows that 21-day sediment MFs for the 90th percentile range from 2.1 for CA almond (i.e., the CADA EEC must be multiplied by a factor of 2.1 to equal the standard CA almond scenario EEC) to 64 (GA pecan). As shown above, the influence of CADA on EECs is more pronounced for the 50th percentile

EEC, with MF values ranging from 6.2 (CA almond) to over 600 (FL pepper). Differences were apparent for the same crop between spatial extents, where the IL corn 50th percentile MF is 120% greater than IN corn and MS cotton 50th percentile MF is 43% greater than TX cotton. Clearly the impact of the real-world catchment cropping density is dependent upon both the crop and the national/regional scale as indicated by the shape of the distributions (Figure S5).

Table S11. 21-day sediment Multiplier Factors (MFs) as a result of applying catchment based PCA distributions to baseline scenarios

	21-d Sediment	
	50th Percentile MF	90th Percentile MF
PA alfalfa	89	13
CA almond	6.2	2.1
FL citrus	32	7.9
IL corn	16	4.4
IN corn	7.2	3.9
MS cotton	157	12
TX cotton	110	11
OR grass seed	9.0	2.5
CA lettuce	22	4.9
NC peanut	115	17
GA pecan	102	64
FL pepper	604	55
ID potato	107	8.0
ME potato	399	54
MS soybean	15	4.5
ND sunflower	304	37
OR sweetcorn	262	37
ND wheat	55	22

Sediment MF values compared to water column MFs indicate that while the impact of probabilistically applying the PCA distribution always had a similar directional effect, the magnitude and exact MFs can be influence by other environmental behaviors which have non-linear impacts on concentrations.

Table S12 shows the relative rankings of these CoIs using baseline inputs compared to those modified by the PCA distribution. The baseline ranking in this table relates only to our implementation for pyrethroids (including VFS requirements) and is not a general ranking for all chemicals. The table also shows the potential regulatory significance of conducting refined exposure assessments that examine sources of uncertainty such as cropping proximity and density, since the refined assessment might focus attention on a different use pattern (i.e., application to a specific crop) as deserving more regulatory attention.

Table S12. Ranking of crop scenarios for a representative pyrethroid showing 21-day sediment (ranked from highest to lowest 90th percentile EEC) comparing baseline approaches with results obtained by considering the impact of PCA in the 10-200m PZ (CADA).

	21-day Sediment	
Rank	**Baseline**	**CADA**
1 (highest)	MS cotton	IL corn
2	TX cotton	MS cotton
3	IL corn	TX cotton
4	OR sweet corn	CA lettuce
5	ME potato	PA alfalfa
6	NC peanut	CA almond
7	CA lettuce	IN corn
8	GA pecan	MS soybean
9	ND sunflower	OR grass seed
10	FL pepper	NC peanut
11	MS soybean	OR sweet corn
12	IN corn	ME potato
13	CA almond	ND wheat
14	ID potato	ID potato
15	OR grass seed	ND sunflower
16	ND wheat	GA pecan
17	PA alfalfa	FL pepper
18 (lowest)	FL citrus	FL citrus

Assumptions and potential sources of uncertainty

All baseline and refined risk assessments are based on a series of assumptions. A primary one is the conceptual model(s) for transport inherent in the model and the scenario used, while additional assumptions relate to defining model inputs. For baseline assessments under FIFRA, the assumptions and inputs are designed to be conservative and reflect parameters that will provide "reasonable worst case" exposures [11]. This study only examined the effect of replacing a 100% PCA assumption with distributions of CoI PCAs measured in areas proximate to NHD+ stream segments; all other inputs and assumptions were unchanged (see SI for list). Therefore, the key assumptions included:

- The spatial scale of NHD+ catchments is highly relevant for evaluating local farming landscapes. The comparison of catchment sizes to farm sizes indicates the NHD+ catchments cover a similar range of areas and the median catchment size (160 ha) is equivalent to the US average farm size in 2019 (180 ha). Catchment scale data in the NHD+ is used extensively by government agencies and regulatory bodies. However, catchment sizes vary considerably based on topography and will be more or less representative of a single or small set of farms accordingly.
- There are sufficient NHD+ catchments selected for each CoI to provide statistically meaningful crop-specific datasets. The range of populations of catchments (~3,000 - >750,000) support this assertion.
- PCAs measured near flowing water reaches are relevant to proximate areas near all types of water bodies. Because stream networks drain the great majority of the conterminous US, especially in areas commonly used for extensive agriculture, it is reasonable to assume this assumption is valid.
- Estimated PCAs in 200m zones within catchments are the most relevant to farm scale agriculture. Pyrethroid loading contributed by crop farther than 200m was considered much less impactful related to pesticide loadings.
- The assumption inherent in summarizing the PCA distributions into ten groups did not distort the findings. In fact, the assumption of using the maximum measured PCA for each group does impact the findings by exaggerating the calculated exposures and thus decreasing the estimated MF values.
- Using the average PCA within the 10-200m zone does not impact exposure transport. This is not a valid assumption since crop located farther away will typically contribute lower loadings compared to crop located closer to a waterbody. PCA data generated separately for the 10-50m and 50-200m zones indicate that median and 90th percentile PCAs in the 10-50m PZs are lower than the 50-200m PCAs for all crops except tree nuts (90th percentile) using 2012 CDL data. This indicates that our working assumption is realistic. Individual PZ PCA data tables are provided in SI. These data represent new/not previously available information for examining variability in PCA at various distances to flowing water for a wide set of crops.
- The national PCA distribution and subsequent MFs reflect the distributions found in localized areas where the crop may be grown. However, when this assumption is not met, the source of uncertainty is "one-tailed" (i.e., a measured PCA will never exceed the default assumption of 100%) and thus will always be a factor that reduces the probability of finding water bodies with pesticide concentrations approaching the baseline EEC.

References

1. *2007 Census of Agriculture*. (2007). United States Department of Agriculture - National Agricultural Statistics Service (USDA-NASS),. https://agcensus.library.cornell.edu/census_year/2007-census/ Accessed August 2022.

2. *2012 Census of Agriculture*. (2012). United States Department of Agriculture - National Agricultural Statistics Service (USDA-NASS),. https://agcensus.library.cornell.edu/census_year/2012-census/ Accessed August 2022.

3. Berglund, D. R. (2007). *Sunflower Production* (A-1331 (EB-25 Revised).). North Dakota Agricultural Experiment Station and North Dakota State University Extension Service Extension.

4. Boryan, C., Yang, Z., & Di, L. (2012, June 22–27). *Cultivated Land Cover Data Sets Using USDA National Agricultural Statistics Service Historic Cropland Data Layers.* [Symposium]. Proc. of IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany.

5. Boryan, C. G., Yang, Z., & Willis, P. (2014, August). US geospatial crop frequency data layers. *2014 The Third International Conference on Agro-Geoinformatics*. https://doi.org/10.1109/agro-geoinformatics.2014.6910657

6. Muñoz-Carpena, R., Parsons, J. E., & Gilliam, J. (1999). Modeling hydrology and sediment transport in vegetative filter strips. *Journal of Hydrology*, *214*(1–4), 111–129. https://doi.org/10.1016/s0022-1694(98)00272-8

7. Padilla, L. E., Winchell, M. F., & Jackson, S. H. (2015). Evaluation of AGRO-2014 for Predicting Hydrophobic Organic Chemical Concentrations in Ponds. *Journal of Environmental Quality*, *44*(5), 1568–1578. https://doi.org/10.2134 /jeq2015.03.0149

8. *Solar and Meteorological Surface Observation Network (SAMSON), 1961–1990* (1.0). (1993). [Software]. NOAA National Oceanographic and Atmospheric Administration, U.S. Department of Commerce. https://cfpub.epa.gov/si/si_public_record_Report.cfm?Lab=NERL&dirEntryID=16161

9. Suárez, L. A. (2005). *Model for Predicting Pesticide and Nitrogen Fate in the Crop Root and Unsaturated Soil Zones: Users Manual for Release 3.12.* (EPA/600/R-05/111). Environmental Protection Agency. https://cfpub. epa.gov/si/si_public_file_download.cfm?p_download_id=525014&Lab=NERL

10. Teske, M., Bird, S., Esterly, D., Ray, S., & Perry, S. (2003). *A User's Guide for AgDRIFT® 2.0.07: A Tiered Approach for the Assessment of Spray Drift of Pesticides* (No. 01–01). https://usermanual.wiki /Pdf/AgDriftusermanualpubFes2003.1946090729.pdf

11. United States Environmental Protection Agency. (2004). *Pesticide Root Zone Model (PRZM) Field and Orchard Crop Scenarios: Guidance for Selecting Field Crop and Orchard Scenario Input Parameters.*

12. U.S. Environmental Protection Agency and U.S. Geological Survey. (2005). *NHD Plus - NHDPlus Home*. NHDPlus Version 2. Retrieved August 31, 2022, from https://nhdplus.com/NHDPlus/

13. United States Environmental Protection Agency - Office of Pesticide Programs. (2019, December). *Creating new scenarios for use in pesticide surface water exposure assessments*. https://www.epa.gov/sites/default/files/2020-01/documents/creating-new-scenarios.pdf

14. United States Department of Agriculture. (1997). *National Resources Inventory | NRCS*. National Resources Inventory. Retrieved August 31, 2022, from http://www.nrcs.usda.gov/wps/portal/nrcs/main/national/technical/nra/nri/

15. United States Department of Agriculture. (2013). *CropScape - NASS CDL Program* (Version 2012) [National Agricultural Statistics Service Cropland Data Layer 2012. Published crop-specific data layer]. Retrieved on August 19, 2022 from https://nassgeodata.gmu.edu/CropScape/

16. *United States Department of Agriculture*. USDA. (2021.). Retrieved September 1, 2022, from https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php#Section3_22.0

17. *Watershed Boundary Dataset | U.S. Geological Survey*. (2013, September 17). [Dataset]. https://www.usgs.gov/national-hydrography/watershed-boundary-dataset

18. Young, D. (2019). *The Variable Volume Water Model* (USEPA/OPP 734S16002). U.S. Environmental Protection Agency - Office of Pesticide Programs.

19. Young, D., & Fry, M. (2020.). *PRZM5: A Model for Predicting Pesticides in Runoff, Erosion, and Leachate* (Version B). U.S. Environmental Protection Agency - Office of Pesticide Programs.